UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Cancer Sciences

The Impact of Germline Genotype on Breast Cancer Tumour Phenotype and Outcome

DOI: 10.5258/SOTON/D1618

by

Stephanie Leanne Greville-Heygate

ORCID 0000-0003-1516-3016

Thesis for the degree of Doctor of Medicine

November 2020

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Cancer Sciences

Thesis for the degree of Doctor of Medicine

THE IMPACT OF GERMLINE GENOTYPE ON BREAST CANCER TUMOUR PHENOTYPE AND OUTCOME

Stephanie Leanne Greville-Heygate

Breast cancer susceptibility gene panels are increasingly utilised in mainstream oncology diagnostic practice. This work describes the influence of commonly reported high and moderate penetrance genes on tumour histopathological phenotype, somatic mutational profile and clinical outcome for symptomatic, early onset breast cancer patients. It considers how genetic testing can be utilised to identify actionable risk and interpret Variants of Uncertain Clinical Significance (VUS).

Participants from the Prospective Outcomes in Sporadic Versus Hereditary Breast Cancer (POSH) Study (n=2744) were included. Tumour histopathological characteristics including grade, size, focality, hormone receptor status, nodal involvement and lymphovascular invasion were compared between gene carriers (BRCA1+, BRCA2+, PALB2+, CHEK2+, ATM+ and TP53+) and non-carriers. Kaplan Meier analysis was used to estimate differences between carriers and no-carriers for Overall Survival (OS) and Distant Disease-Free Survival (DDFS). A further sample with tumour sequence data and a greater range of onset ages was included for comparison from The 100,000 Genomes Project (100KGP) (Rare Disease, Familial Breast Cancer (n=826) and Cancer, Breast Cancer (n=2464)). Tumour Mutational Burden (TMB) and the presence of Single Base Substitution (SBS) Mutational Signatures were compared between gene carriers and non-carriers.

In the POSH study, 16.7% (453/2744) had a moderate or high penetrance variant. Hormone receptor status and tumour focality were significant independent predictors of *BRCA1+* and *BRCA2+*. *BRCA1+* were significantly more likely to present with a Triple Negative Tumour (TNT) (123/201 (61.2%) versus 417/2291 (18.2%)) (p<0.0001) and be localised (156/180 (86.7%) versus 1461/2085 (70.1%)) (p<0.0001)). *BRCA2+* were significantly more likely to be ER-positive (115/136 (84.6%) versus 1557/2279 (68.3%)) and multifocal at presentation (57/121 (47.1%) versus 624/2085 (29.9%) (p<0.0001)). Within 100KGP, *BRCA+* and *PALB2+* had a significantly higher TMB compared to non-carriers (*BRCA+*, 4.39 Mut/Mb and *PALB2+*, 6.39 Mut/Mb versus 2.51 Mut/Mb) (p=0.0433 and p=0.0066 respectively). *BRCA+* and *PALB2+* also had a significantly increased expression of SBS3 compared to non-carriers (31.14% and 32.14% versus 10.59%) (p<0.0001 and p=0.0047 respectively). *CHEK2+* presented with breast cancer that was significantly more likely to be ER-positive compared to non-carriers (p=0.0016). Survival analysis revealed that OS and DDFS was significantly worse in *CHEK2*+ versus *CHEK2*- (OS HR, 1.58 (95%CI, 1.01-2.48 (p=0.043))).

Tumour phenotypic characteristics including focality, hormone receptor status, TMB and SBS mutational signature contribute to estimating the likelihood of a *BRCA1, BRCA2* or *PALB2* germline variant and could be used to assist with the interpretation of a VUS. The utility of tumour phenotype in moderate risk gene carriers for likelihood prediction and variant interpretation is less clear.

Table of Contents

Table of Tables

Table of Figures

Academic Thesis: Declaration Of Authorship

I, Stephanie Greville-Heygate

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

The Impact of Germline Genotype on Breast Cancer Tumour Phenotype and Outcome.....

.....

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:
    a) Greville-Heygate SL. The Impact of Hereditary versus Sporadic Breast Cancer on Tumour Phenotype in A Young Prospective Cohort: *BRCA1* and *BRCA2*. 2017. Submitted in partial fulfilment of the requirements for the Degree of Master of Science. University of Southampton.
    b) Greville-Heygate SL, Eccles DM, Side LE. Invited Commentary: Hereditary Breast and Ovarian Cancer Testing in the Genomic Era. 2018. Journal of the American Medical Association Oncology 2019. 5(1):58-59.
    c) Pathogenic Variants in *CHEK2* Are Associated With an Adverse Prognosis in Symptomatic Early-Onset Breast Cancer. Greville-Heygate, SL, Maishman, T, Tapper, WJ, Cutress RI, Copson E, Dunning AM, Haywood L, Jones LJ and Eccles DM. Journal of Clinical Oncology, Precision Oncology. 2020 4: 472-485.

Signed: .....

Date: 4th November 2020.....

Acknowledgements

I gratefully acknowledge the support of my supervisors, Professor Diana Eccles and Dr William Tapper in the production of this work. I also acknowledge Mr Tom Maishman, Statistician, University of Southampton for this assistance with statistical analysis including, correlative statistics, logistic regression, survival analysis and some statistical tabulation.

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

Publications

1. Greville-Heygate, SL, Maishman, T, Tapper, WJ, Cutress RI, Copson E, Dunning AM, Haywood L, Jones LJ and Eccles DM. Pathogenic Variants in *CHEK2* Are Associated With an Adverse Prognosis in Symptomatic Early-Onset Breast Cancer. Journal of Clinical Oncology, Precision Oncology. 2020 4: 472-485
2. McCrorie A, Ashfield S, Begley A, Mcilmunn C, Morrison P, Boyd C, Eccles B, Greville-Heygate S, Copson ER, Cutress RI, Eccles DM, Savage KI, McIntosh SA. Multifocal Breast Cancers Are More Prevalent in *BRCA2* Versus *BRCA1* Mutation Carriers. The Journal of Pathology, Clinical Research. 2020 6(2): 146–153.
3. Spurdle AB, Greville-Heygate S, Antoniou AC, Brown MA, Burke LJ, Eccles DM et al. Towards Controlled Terminology for Reporting Germline Cancer Susceptibility Variants: An ENIGMA Report. 2019. Journal of Medical Genetics 56(6):347-357.
4. Greville-Heygate SL, Eccles DM, Side LE. Invited Commentary: Hereditary Breast and Ovarian Cancer Testing in the Genomic Era. 2019. Journal of the American Medical Association 5(1):58-59.
5. Copson ER, Maishman TC, Tapper WJ, Cutress RI, Greville-Heygate S, Eccles DM et al. Germline *BRCA* mutation and outcome in young-onset breast cancer (POSH): A Prospective Cohort Study. 2018. Lancet Oncology. 19(2): 169-180

Definitions and Abbreviations

ACCE	Analytic validity, Clinical validity, Clinical utility and Ethical implications model
ATM	Ataxia-Telangiectasia Mutated
BCAC	Breast Cancer Association Consortium
BOADICEA	Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
CGG	Cancer Genetics Group
CHEK2	Checkpoint Kinase 2
COCP	Combined Oral Contraceptive Pill
COGS	Collaborative Oncological Gene-Environmental Study
ESMO	The European Society of Medical Oncology
HRR	Homologous Recombination Repair
HRT	Hormone Replacement Therapy
NCCN	The National Comprehensive Cancer Network
NHS	National Health Service
NGS	Next Generation Sequencing
NICE	National Institute for Clinical Excellence
PALB2	Partner and Localizer of Breast Cancer 2
PRS	Polygenic Risk Score
SBS	Single Base Substitution
SNPs	Single Nucleotide Polymorphisms
TMB	Tumour Mutational Burden
TP53	Tumour Protein p53

Definitions and Abbreviations

UK-GTN	United Kingdom Genetic Testing Network
VUS	Variants of Uncertain Clinical Significance
Maximum Invasive Tumour Size	Maximum size of invasive tumour observed.
Overall Tumour Size	Size invasive tumour and surrounding DCIS.

Chapter 1 Literature Review

1.1 Background

Breast Cancer is the second most common cancer diagnosis in the developed world, affecting approximately 1 in 8 women over their lifetime.(1, 2) In 2015 there were 54,800 new cases of breast cancer and 11,500 directly attributable cancer related deaths in the UK.(1, 2) In the era of "Genomics" it is important to not only use germline genetic testing to identify heritable breast cancer risk but to understand the utility of this testing. We need to understand how breast cancer susceptibility genes influence tumour biology and how this can be exploited for the purposes of identifying actionable risk, variant interpretation and precision therapy. This review details the current landscape of genetic testing for heritable breast cancer susceptibility and how this translates into the aims and objectives of this thesis.

1.2 The Aetiology of Breast Cancer

A combination of both genetic and environmental factors are recognised as important aetiological risk factors in the evolution of disease.(3) Constitutional and environmental risk factors can be broadly described as reproductive factors, modifiable lifestyle factors and exogenous hormonal exposure (Table 1).(1, 4, 5) At a population level, exposure to these epidemiological risk factors may be implicated in up to 27% of female breast cancer cases within a Western population.(1, 6, 7) A combination of alcohol, obesity, physical exercise and post-menopausal hormone exposure represent some of the most important associations linked to 6.4%, 8.7%, 3.4% and 3.2% of cancers respectively.(1, 6, 7)

	Recognised Environmental Factors which Modify Breast Cancer Risk	
Factor	**Risk Increase**	**Risk Decrease**
Reproductive	Early menarche Late menopause First pregnancy after age 35 years	Increasing parity. Breast feeding post-partum.
Modifiable lifestyle	Post-menopausal obesity Alcohol	Increased physical activity
Exogenous hormone exposure	Combined oral contraceptive pill Hormone replacement therapy	

Table 1: Environmental Risk Factors for Breast Cancer Development

1.3 Genetics of Breast Cancer

In addition to the aforementioned constitutional and environmental factors, germline genetic factors have an important influence on the overall stratosphere of breast cancer risk for any one individual. Genetic risk factors for breast cancer development include a combination of high and moderate risk genotypes along with lower penetrance single nucleotide polymorphisms (SNPs).

1.3.1 High and Moderate Penetrance Genes

Approximately 5-10% of breast cancer predisposition is attributable to a higher penetrance single gene.(2, 3) These breast cancer susceptibility genes exhibit variable penetrance which are conveyed as estimations of absolute and relative risk derived from cohort or case control studies (Figure 1).(8, 9) Moderate risk genotypes are associated with an average relative breast cancer risk which is 2-4 times higher than the basal population risk.(10, 11) This equates to an approximate cumulative lifetime risk of between 17% and 30%.(10) High risk genotypes confer a relative risk of breast cancer which is on average 4 times higher than the population risk and a cumulative lifetime risk greater than 30%.(10)

Mutations in either the Breast Cancer 1 Gene (*BRCA1*) or Breast Cancer 2 Gene (*BRCA2*) are the main causal variants responsible for approximately 70% of heritable breast cancer attributable to a highly penetrant single gene alteration.(4, 12-14). The most frequently identified other intermediate and high risk genes are Partner and Localizer of Breast Cancer 2 (*PALB2*), Checkpoint Kinase 2 (*CHEK2*), Ataxia-Telangiectasia Mutated (*ATM*) and The Tumour Protein p53 (*TP53*) gene.(15) A study by Buys et al. which tested 35,409 women with a 25 gene cancer susceptibility panel found a strong genetic factor in 9.3% of women.(15) The most frequently identified non-*BRCA* genes included *CHEK2*, *ATM* and *PALB2* which represented 11.7%, 9.7% and 9.3% of all gene carriers respectively.(15)

1.3.2 Polygenic Risk

It is estimated that 10-30% of breast cancers demonstrate familial aggregation. It is well recognised that a family history of breast cancer increases risk with higher levels of risk correlating with the number of affected relatives.(16) This effect is often mediated through complex inheritance involving a combination of polygenic risk factors known as Single Nucleotide Polymorphisms (SNPs), and exposure to recognised environmental and constitutional risk factors.(17)

Genome Wide Association Studies (GWAS) have identified numerous lower penetrance SNPs that associate with breast cancer risk.(5) When utilised in isolation, these SNPs provide only a small, clinically uninformative increment in the basal cancer risk.(18) However, when considered collectively, they can produce a more pronounced effect which is expressed as a Polygenic Risk Score (PRS).(18) Polygenic risk scores are determined through the combined contribution of multiple SNP deemed to have a significant effect on disease (Genome Wide Significance $p<5x10^{-8}$) following SNP array.(19, 20) The number of SNPs including in any PRS can vary. Adding more SNPs into a PRS does not necessarily increase the accuracy. This is because larger GWAS are required to identify high numbers of contributory SNPs which can produce a consummate reduction in their overall influence on risk.(20)

Currently, polygenic factors explain approximately 18% of the familial aggregation of breast cancer susceptibility and a combination of PRS and environmental risk factor exposure has been shown to enhance breast cancer risk prediction with a greater magnitude of effect in ER-positive breast cancer (Table 2).(5, 21, 22) In 2018, Rudolph et al. observed the effect of PRS derived from 77 SNPs and environmental risk factors exposure including exogenous hormonal treatment, body mass index, lifetime alcohol intake and reproductive history on breast cancer risk. The sample population was ascertained from 28,239 female breast cancer cases and 30,445 controls of European descent within the BCAC consortium.(5) They identified that a combination of PRS and environmental risk factor exposure enhanced breast cancer risk prediction in a multiplicative manner.(5)

SNPs have also been shown to modify breast cancer risk amongst *BRCA1* and *BRCA2* gene carriers.(18) In 2017, Kuchenbaecker et al. observed the effect of PRS on breast cancer risk amongst 7797 *BRCA1* gene carriers and 4330 *BRCA2* gene carriers with breast cancer.(18) They identified that PRS modified breast cancer risk but with a lower magnitude of effect than that observed in the absence of a highly penetrant monogenetic risk factor.(18)

Population	Cases	Controls	SNPs	Breast Cancer Subtype	AUC*	HR (95% CI)	Study
European Ancestry (BCAC)	94,075	75,017	313	ER+	0.651	1.74 (1.66-1.82)	Mavaddat et al. 2019.(23)
				ER-	0.611	1.47 (1.37-1.58)	
European Ancestry (BCAC)	28,239	30,445	77	ER+ (90-95 percentile)		2.25 (2.04-2.47)	Rudolph et al. 2018.(5)
				ER- (90-95 percentile)		1.74 (1.50-2.03)	
Australia and North America (BCFR and kConFab)	1496	2,869	24	-	0.59	1.38(1.22-1.56)	Li et al. 2017.(24)
Global	7797 *BRCA1*		88	ER- (53 SNPs)		1.27 (1.23 to 1.31)	Kuchenbaecker et al. 2017(18)
(CIMBA)	4330 *BRCA2*		88	ER+ (87 SNPs)		1.22 (1.16 to 1.27)	
UK (PROCAS)	466	8897	18	-	0.67*	1.56 (1.38-1.77)	Van Veen et al. 2018. (19)

Table 2: Polygenic Risk Score and Breast Cancer Risk

Table detailing the influence of Polygenic Risk Score on Breast Cancer Risk Area Under the Receiver Operated Curve (AUC) calculations are used to determine whether the predicted risk is greater for cases than controls. An AUC greater than 0.5 indicates a discriminatory effect.(20) Breast Cancer Association Consortium (BCAC). Breast Cancer Family Registry (BCFR) and Kathleen Cunningham Consortium into Research on Familial Breast Cancer (kConFab). Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA). Predicting Risk of Cancer at Screening (PROCAS) study.*incorporating both mammographic density and SNP.

Overall, this demonstrates that the arbitrary division of the breast cancer susceptibility genes into high, moderate and low risk genotypes is not necessarily representative of the true complement of risk factors which an individual may possesses. Polygenic factors and recognised environmental aetiological exposures are also important modifiers of this risk. For example, an individual with a germline *CHEK2* mutation and a strong family history of cancer or a high polygenic risk score may have a higher relative risk of breast cancer than a *CHEK2* variant carrier in the absence of these factors. It demonstrates the importance of utilising monogenic, polygenic and environmental risk factors to determine a more comprehensive cancer risk stratification (Figure 1).

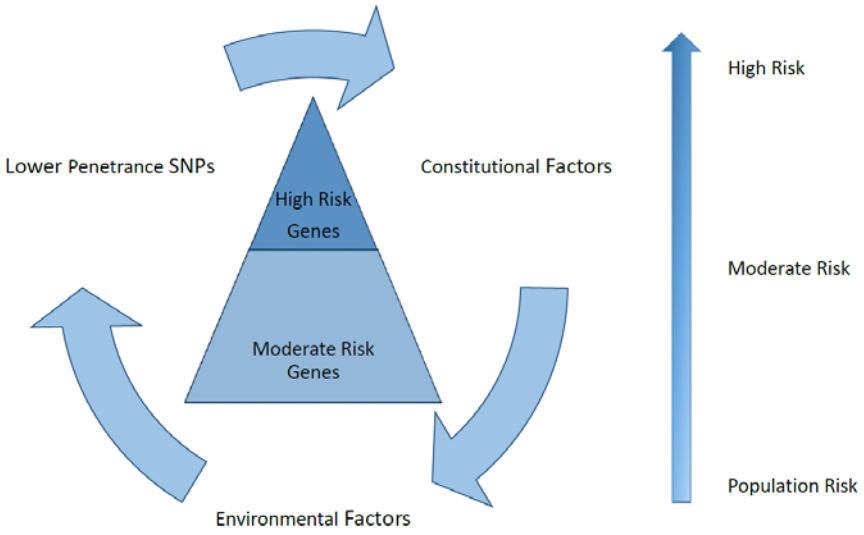


Figure 1: Genetics of Hereditary Breast Cancer

The combination of single gene and complex genetic factors alongside environmental exposures provides a more comprehensive stratification of personalised cancer risk than the utilisation of these factors in isolation.

1.4 Hereditary Breast Cancer

Hereditary Breast Cancer arises due to the presence of an inherited high or moderate penetrance gene which contributes significantly to overall breast cancer risk. The genes most frequently associated with non-syndromic, hereditary breast cancer include *BRCA1*, *BRCA2*, *PALB2*, *CHEK2* and *TP53*. They are described with reference to their structure, function, epidemiology and associated cancer risks.

1.4.1 Breast Cancer 1 and Breast Cancer 2 Genes

The Breast Cancer 1 (*BRCA1*) gene is composed of 24 exons and the Breast Cancer 2 (*BRCA2*) gene is composed of 27 exons.(25) The BRCA1 protein combines with other tumour suppressers, DNA damage sensors and signal transducers to form the *BRCA*-associated Genome Surveillance Complex.(4) BRCA2 interacts with RAD51, PALB2 and BRCA1.(26) They are specifically involved in the homologous recombination repair pathway which repairs DNA double strand breaks. In the absence of functional BRCA1 and BRCA2, cells use other more error prone mechanisms of DNA repair including non-homologous end joining. (25, 26) This can result in the progressive somatic accumulation of DNA damage and cancer evolution (Figure 2).(25, 26)

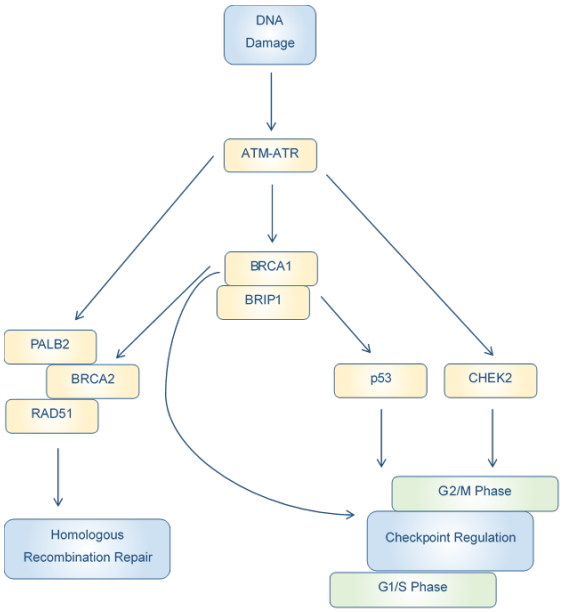


Figure 2: BRCA and the Cellular Response to DNA Damage

BRCA1 and *BRCA2* genes have a central role within the Homologous Recombination Repair Pathway. The presence of DNA damage results in activation of ATM and ATR. ATM and ATR phosphorylate BRCA1 to block cell cycle progression.(4, 27-33) They also phosphorylate PALB2 which promotes the localization of RAD51 for DNA repair through interaction with BRCA2. Genes are represented in yellow, pathways in blue and the cell cycle in green. G1 represents the phase between mitosis and DNA replication. G2/S refers to the synthesis phase when DNA replication occurs.

1.4.1.1 Epidemiology

Pathogenic variants in *BRCA1* and *BRCA2* are recognised at a global level (Table 2). The prevalence of *BRCA1* and *BRCA2* mutations varies depending on the tested cohort from 1.8% in sporadic breast cancer to 19% in unilateral triple negative breast cancer diagnosed under the age of 50 years (Table 3). (34, 35) This variation in prevalence reflects sample ascertainment. Higher prevalence is recognised in the context of a tumour phenotype associated with *BRCA1* such as TNT or in the context of Hereditary Breast and Ovarian Cancer.

Population (year)	Cohort	Prevalence	Sample Population (mutation carriers)	Authors
Swedish (2018) LIBRO1 Study	Unselected under 80yrs	1.8%	5099 (92)	Li et al.(34)
UK (2000-2008) POSH Study	Early onset breast cancer <40 yrs	12%	2733 (338)	Copson et al.(36)
German (2018) GC-HBOC*	Unilateral TNT (age 19-76) Unilateral TNT (age 20-29) Unilateral TNT (age 30-39) Unilateral TNT (age 40-49) Unilateral TNT (age 50-49) Unilateral TNT (age 60-69)	15.8% (14.7% *BRCA1* & 1.1% *BRCA2*) 32.9% (32.9% *BRCA1* & 0% *BRCA2*) 20.4% (19.4% *BRCA1* & 1.0% *BRCA2*) 11.6% (10.2% *BRCA1* & 1.4% *BRCA2*) 5.7% (4.9% *BRCA1* & 0.8% *BRCA2*) 6.9% (3.4% *BRCA1* & 3.4% *BRCA2*)	802 (127) 85 (28) 309 (63) 216 (25) 122 (7) 58 (4)	Engel et al. (35)
German (2017)	Familial	12.4%	581 (72)	Kraus et al. (37)
South African (2015)	TNT or pre-menopausal	12.1%	108 (13)	Francies et al.(38)
UK (2000)	Unselected	2%	1220 (24)	Anglian Breast Cancer Study Group (39)
American (2007) SEER study	Unselected White** Unselected Asian American Unselected Hispanic Unselected African American	4.0% 0.7% 0.5% 2.3%	549 (22) 444 (3) 393 (21) 341 (8)	John et al.(40)

Table 3: Epidemiology of *BRCA* Mutations

Summary table demonstrating BRCA1 and BRCA2 pathogenic variant prevalence with a comparison across cohorts based upon selection criteria. *German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC). **Sample including Ashkenazi Jewish descent.

1.4.1.2 *BRCA1* and *BRCA2* Associated Breast Cancer Risk

Pathogenic variants in *BRCA1* and *BRCA2* are associated with cumulative lifetime risks of breast cancer between 45-87%. (4, 25) In 2017, Kuchenbaecker et al. calculated the absolute cumulative lifetime risks of breast and ovarian cancer based upon prospective data obtained from 9856 *BRCA* mutation carriers derived from 3 consortia (The International *BRCA1*/2 Carrier Cohort Study (IBCCS), the Breast Cancer Family Registry (BCFR) and the Kathleen Cunningham Foundation (kConFab).(41)

They identified that the cumulative lifetime breast cancer risk was 72% (95% CI, 65%-79%) for *BRCA1* carriers and 69% (95% CI, 61%-77%) for *BRCA2* carriers to age 80 years.(41) The peak breast cancer incidence amongst *BRCA1* mutation carriers occurred between the ages of 41-50 years (28.3/1000 person years (95%CI, 23.1-34.7)). Conversely, the peak breast cancer incidence amongst *BRCA2* mutation carriers occurred between the ages of 51-60 years (30.6/1000 person

years (95% CI, 22.8-41.1)).(41) The ovarian cancer risks were 44% (95%CI, 36%-53%) and 17% (95% CI, 11%-25%) for *BRCA1* and *BRCA2* mutation carriers respectively.(41) The cumulative risk of contralateral breast cancer 20 years after the first diagnosis was 40% (95% CI, 35%-45%) for *BRCA1* mutation carriers and significantly lower in *BRCA2* mutation carriers 26% (95% CI, 20%-33%) (p=0.001).(41)

Similar findings were noted by Mavaddat et al. in 2013. They performed a large prospective cohort study of 988 *BRCA1* and *BRCA2* mutation carriers derived from the UK EMBRACE study. They identified that the cumulative lifetime risk of breast cancer by age 70 years was 60% (95% CI, 44% - 75%) for *BRCA1* mutation carriers and 55% (95% CI, 41% - 70%) for *BRCA2* mutation carriers.(42) The average cumulative risk of ovarian cancer by age 70 years was also similar to that observed by Kuchenbaecker et al. (59% (95% CI, 43%-76%) for *BRCA1* carriers and 16.5% (95% CI, 7.5%-34%) for *BRCA2* carriers.(42)

1.4.1.3 Modifiers of *BRCA1* and *BRCA2* Associated Breast Cancer Risk

The penetrance of pathogenic variants in *BRCA1* and *BRCA2* is dependent upon many of the aforementioned exogenous and endogenous factors along with the interaction with more common, lower penetrance breast cancer susceptibility alleles.(25) Kuchenbaecker et al. demonstrated that the *BRCA*-associated cancer risks were modified by the family history of malignancy. They demonstrated that the cumulative breast cancer risks significantly increased with the number of affected first and second degree relatives with breast cancer.(41) They also noted a variant position effect in relation to overall cancer risks that was independent of the family history of cancer.(41)

This observation may be attributable to the modifying effect of polygenic factors or SNPs that aggregate within a family. In 2020 Gallagher et al. observed the influence of an 86 SNP polygenic risk score on *BRCA*-associated cancer risk in a large cohort of European women.(43) They found that the highest percentiles of polygenic risks were associated with a significant increase in cancer risk beyond that conferred by the *BRCA* gene alteration in isolation.(43)

1.4.1.4 Other Associated Cancer Risks

Pathogenic variants in *BRCA2* have been seen in association with prostate cancer, pancreatic cancer and malignant melanoma.(44) In 2015, Mersch et al. described the observed number of cases in 1072 *BRCA1* and *BRCA2* mutation carriers. Significantly higher rates of pancreatic and prostate cancer were observed in *BRCA2* mutation carriers. Pancreatic cancer (SIR, 21.7 (95%CI, 13.1-34.0 (p<0.001)) and prostate cancer (SIR, 4.9 (95% CI, 2.0-10.1 (p=0.002)).(45)

1.4.2 Partner and Localizer of Breast Cancer 2 (*PALB2*)

Partner and Localizer of Breast Cancer 2 (*PALB2*) is a tumour suppressor gene composed of 13 exons which is integral to the BRCA mediated DNA Homologous Recombination Repair (HRR) pathway.(46) In 2019, Li et al. found that 16/24 breast tumours occurring in the context of a pathogenic germline mutation in *PALB2* demonstrated biallelic loss of *PALB2* either through Loss of Heterozygosity (LOH) or a somatic inactivating mutation.(47)

The PALB2 protein has several functional domains which exhibit differential protein binding (Figure 3).(46, 48) The N terminus holds a Coiled Coil (CC) domain which can bind to BRCA1, RAD51 recombinase.(46, 48) This domain can also self-associate with PALB2.(49) There are two DNA binding domains, and a small Chromatin Associated Motif. (46, 48) The C terminus contains a WD40 repeat domain which can bind with RAD51 and BRCA2.(46, 48)

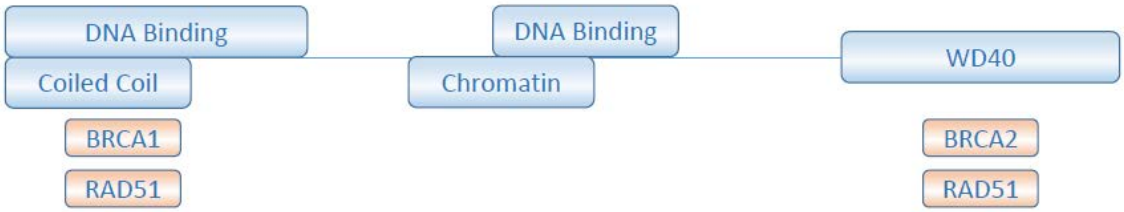


Figure 3: Structure of *PALB2*

The BRCA1 binding domain is located between amino acids 1 - 319 and the BRCA2 binding domain occurs in the WD40 motif located between amino acids 853 and 1186.(30)

PALB2 is activated by BRCA1 mediated phosphorylation following the activation of *ATM* in response to DNA double strand breaks (Figure 4).(46) This interaction occurs through the N terminus CC domain.(49) Activated PALB2 recruits BRCA2 through interaction with the WD40 domain. As such, BRCA1 and BRCA2 can exist in a protein complex linked by PALB2.(46, 50) PALB2 facilitates the nuclear localisation of BRCA2 and recruitment of RAD51 recombinase for HRR within the S and G2 phases of the cell cycle.(33, 48, 49, 51, 52) The DNA repair function of PALB2 is also facilitated through binding of single stranded DNA and chromatin to the DNA binding regions and chromatin motif.(46, 49) Heterozygous, pathogenic variants in PALB2 are associated with an increased predisposition to breast cancer whilst biallelic mutations are associated with Fanconi Anaemia. (30, 50)

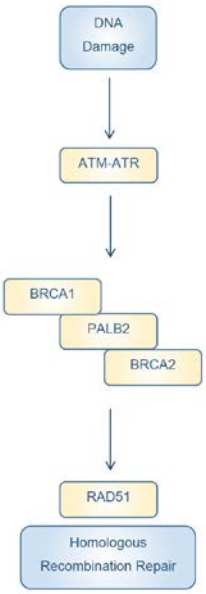


Figure 4: Mechanism of Action *PALB2*

The *PALB2* gene is involved within the Homologous Recombination Repair Pathway. Genes are represented in yellow and pathways in blue.

1.4.2.1 Epidemiology

An inherited pathogenic variant in *PALB2* is identifiable in approximately 0.66%-0.86% of breast cancer cases based upon unselected series worldwide (Table 4 and Appendix A).(50) The prevalence is recognised to be higher in familial breast cancer.(46, 50, 53) Several pathogenic variants are more frequently identifiable amongst specific ancestral groups and may represent founder mutations. This includes *PALB2* c.509_510delGA, p.(Arg170fs*14) amongst the Polish population; *PALB2* c.1592delT, p.(531fs*30) in Northern Europeans and c.3113G>A, p.(Trp1038)* amongst those of British descent including the UK, America, Australia and Canada.(54-57)

Population (year)	Cohort	Prevalence (%)	Authors (ref)
China (2017)	Familial	4/305 (1.31%)	Zhang et al.(58)
	Sporadic	11/1967 (0.56%)	
Malaysia (2017)	Unselected	4/467 (0.86%)	Yang et al.(59)
UK (2017)	Unselected	89/13087 (0.68%)	Decker et al.(56)
Finland (2017)	Familial	19/947 (2.0)	Wesola et al.(53)
	Unselected	8/1274 (0.6)	
Poland (2017)	Familial	7/460 (1.5)	Kluska et al. (60)
Poland and Ukraine (2017)	Mixed	4/427 (0.94%)	Myszka et al. (61)
Jamaica (2017)	Unselected	4/179 (2.23%)	Lerner-Ellis et al.(62)

Table 4: Global Prevalence of *PALB2* Pathogenic Variants

1.4.2.2 *PALB2* Associated Breast Cancer Risk

Protein truncating variants in *PALB2* are generally considered pathogenic or likely pathogenic.(63) This is because the Coiled Coil and WD40 domains located at the N and C Terminus are functionally important for *BRCA1* and *BRCA2* binding and downstream effect. Disruption of the terminal part of the WD40 repeat domain has been associated with breast cancer risk in the context of *PALB2*, c.3459 C>G, p.(Tyr1183*).(50)

The relative risk of breast cancer associated with rare truncating variants in *PALB2* has been determined by several large population based cohorts including both familial and unselected cases. Overall the relative risk of breast cancer is considered to be high-moderate, with an increase in risk that is at least 4 fold higher than the basal population level (Table 5).(30, 56, 57, 64) It is recognised that complex familial factors including genetic and environmental exposure are important modifiers of inherent genomic risk associated with *PALB2* carrier status. In 2014 Antoniou et al. determined the absolute lifetime risks of breast cancer to be 33% (95% CI, 25%-44%) in the absence of a family history and 58% (95% CI, 50%-66%) for a woman with two first degree relatives diagnosed under the age of 50 years.(30) More recently, in 2020, Yang et al. observed the relative and absolute cancer risks of *PALB2* pathogenic protein truncating variant carriers derived from 524 families.(20) The cohort was mixed and included both familial and unselected cases of breast and ovarian cancer.(20) The absolute breast cancer risk was 52.8% (95% CI 43.7%-62.7%) to age 80 years.(20) Birth cohort and family history were recognised modifiers of relative and absolute breast cancer risk.

Interpretation of missense variants and their association with *PALB2* related cancer risk is more challenging and potentially variant specific (Table 5).(61, 63) It has been shown that *PALB2*, c.2816T>G, p.(Leu939Trp) located in the WD40 *BRCA2* binding domain is not associated with a markedly elevated breast cancer risk.(57, 65) Decker et al. also found that rare missense variants may have no effect or only modestly increase breast cancer risk above the basal population level with *PALB2* missense variants located within the *BRCA1* binding domain having the strongest association with breast cancer risk.(56) An Australian based case control study of familial breast cancer published by Thompson et al. found no excess of *PALB2* missense variants overall in cases compared to controls.(64) One specific variant, c.1676A>G p.(Gln559Arg) was significantly enriched in cases and considered to represent a lower penetrance SNP (OR 1.24 (95 %CI, 1.09–1.47 (p = 0.002)).(64)

Population (year)	Cohort	*PALB2* Variants	Relative Cancer Risk (95% CI)	Sample Population (mutation carriers)		Authors
				Cases	Controls	
International (2020)	Mixed	Protein truncating	7.18 (5.82-8.85 ($p=6.5x10^{-76}$)	(976)		Yang et al. (20)
UK (2017) SEARCH Database	Unselected	Protein truncating	4.69 (2.27-9.68 ($p=6.9\times10^{-6}$))	13,087 (89)	5488 (8)	Decker et al. (56)
		Rare Missense	1.28 (0.95-1.73 (p=0.12))			
		Rare Missense**	1.76 (1.03-2.98 (p=0.047)			
International (2016) BCAC	Mixed	c.1592delT, p.Leu531fs	3.44 (1.39-8.52 (p=0.003))	42,671 (229)	42,164 (159)	COGS study (57)
		c.3113G>A, p.Trp1038*	4.21 (1.84-9.60 ($p=1.2x10^{-4}$))			
		c.2816T>G, p.Leu939Trp	1.03 (0.80 – 1.32 (p=0.82))			
Australian (2015)	Familial	Protein truncating	6.58 (2.3-18.9 (p=0.0001))	1996	1998	Thompson et al. (64)
		Missense	1.15 (1.02-1.32 (p=0.025))			
International (2014) *PALB2* Mutation	Familial	Protein truncating	9.47 (7.16-12.57)	362	-	Antoniou et al. (30)

Table 5: *PALB2* Associated Breast Cancer Risk

Evaluation of the published literature regarding *PALB2*-associated breast cancer risks. Mixed cohorts included both familial and unselected patient groups.**Rare missense variants located within the *BRCA1* binding domain (Amino acids 1-319).

1.4.2.3 Modifiers of *PALB2* Associated Breast Cancer Risk

Historically, there has been no observed difference in the associated risk of *PALB2* related breast cancer with age. Statistical modelling by Antoniou et al. found that age specific relative risk models were not significantly better than models that assumed a constant relative risk with age (p = 0.07). Cybulski et al. and Decker et al. also found no significant difference in the age of onset between *PALB2* mutation carriers and non-carriers.(56, 66) Cybulski et al. identified the relative risk of breast cancer in women to be 3.68 (95%CI 1.84-7.15) under the age of 50 years and 4.90 (2.53-9.49) for those diagnosed after the age of 50 years.(66). This is in contrast to high risk genes such as *BRCA1*, *BRCA2* and *TP53* where much of the elevation in breast cancer risk is most pronounced in the premenopausal years.(41) More recently, Yang et al report that under a linear trend model, the relative risk of breast cancer reduces with age from RR 13.10 at age 25 years to RR 4.69 at age 75 years.(20)

1.4.2.4 Other Associated Cancer Risks

The relatively low frequency of pathogenic *PALB2* mutations in combination with the rarity of the likely associated cancer types makes the estimation of other cancer risks challenging. (55) Potential associations have been observed with male breast cancer, ovarian, pancreatic, melanoma and prostate cancer but current evidence is insufficient to determine a definitive association.(30, 44, 57, 67-70)

Recent work by Yang et al. including 17906 individuals derived from 524 families with a known pathogenic truncating variant in *PALB2* observed significant associations with ovarian cancer, pancreatic cancer and male breast cancer risk.(20) The absolute ovarian cancer risk to age of 80 years was 4.8% (95% CI, 2.4%-9.7%).(20) The absolute pancreatic cancer risk to age of 80 years was 2.2% (95% CI, 1.2%-4.2%) for female gene carriers and 2.8% (95% CI, 1.5%-5.3%) for male gene carriers.(20) The absolute male breast cancer risk to age of 80 years was 0.9% (95% CI, 0.2%-4.9%).(20) There was no significant association with prostate or colorectal cancer risk. (20)

1.4.3 Checkpoint Kinase 2 (*CHEK2*)

Checkpoint Kinase 2 (*CHEK2*) is a serine/threonine kinase which functions as a tumour suppressor gene necessary for cell cycle checkpoint regulation, the inhibition of cellular proliferation and activation of DNA repair pathways.(71-73) The CHEK2 protein consists of multiple functional domains (Figure 5).(74) This includes an N-Terminal SQ/TQ rich domain, a Fork Head Associated domain (FHA) and Serine/Threonine protein kinase domain.(74, 75) It becomes activated by ATM mediated phosphorylation in response to DNA double strand breaks.(71, 76, 77) Activated CHEK2 phosphorylates p53, *BRCA1*, Cdc25A and Cdc25C facilitating cell cycle arrest during the G1 phase of mitosis, apoptosis and homologous recombination repair (figure 6).(71, 78) Pathogenic variants in *CHEK2* have been associated with an increased risk of breast cancer.(56, 57, 73, 79, 80)

Figure 5: Structure of *CHEK2*

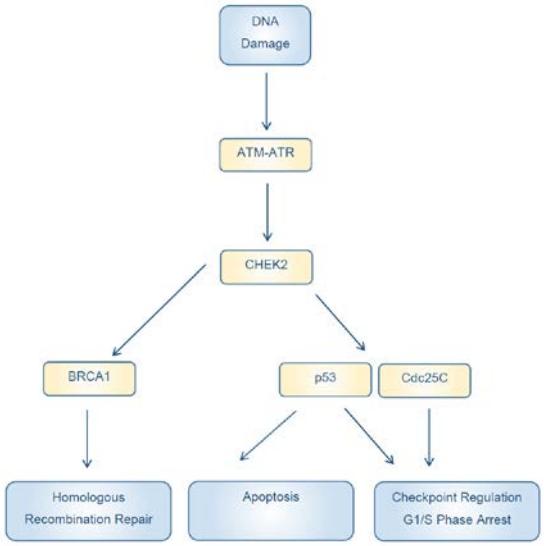


Figure 6: Mechanism of Action of *CHEK2*

CHEK2 becomes activated by *ATM* mediated phosphorylation in response to DNA double strand breaks.

1.4.3.1 Epidemiology

A number of studies have demonstrated that pathogenic variants in *CHEK2* are most prevalent amongst individuals of European descent.(73, 81) Several founder mutations exist in European populations including *CHEK2* c.444+1G>A, German and Polish; c.470T>C, p.(Ile157Thr), Slavic; c.1100delC, p.(Thr367Metfs), Northern European; del5395 (deletion of exons 9-10), Polish and c.1283C>T, p.(Ser428Phe), Ashkenazi Jewish.(82) Of these, *CHEK2* c.1100delC, p.(Thr367fs) is the most frequently identified and studied.(56, 79) A population based study within the UK involving 13087 breast cancer cases found that c.1100delC, p.(Thr367fs) was the prevalent rare truncating *CHEK2* variant, present in 81% (196/242) of *CHEK2*-associated breast cancer cases.(56) Schmidt et al. utilised genotyping data from 33 studies within the Breast Cancer Association Consortium (BCAC) to determine population frequencies of this specific variant across Europe.(72) The highest population frequencies were observed amongst individuals of Northern European descent (Table 6).(72)

Country	Frequency Rate	95% CI
The Netherlands	0.0134	0.0110-0.0162
Finland	0.0124	0.0100-0.0162
Denmark	0.0066	0.0053-0.0081
United Kingdom	0.0054	0.0047-0.0062
Poland	0.0024	0.0017-0.0036
Russia	0.0020	0.0010-0.0038

Table 6: Population Frequency of *CHEK2* c.1100delC in Europe

Schmidt et al. utilised genotyping data from 33 studies within the Breast Cancer Association Consortium (BCAC) to determine population frequencies of *CHEK2* c.1100delC across Europe.(72) Estimations of allele frequency were derived from 42,977 case and 44,777 controls.(72)

Variants in *CHEK2* are much less frequently seen amongst other worldwide populations. Studies have demonstrated that pathogenic variants in *CHEK2* are only found in a small proportion of *BRCA* negative Malaysian, Korean, Moroccan, Greek and Iranian patients with hereditary breast cancer.(76, 83-86) A further study observing the frequency of *CHEK2* pathogenic and likely pathogenic variants in 45,879 individuals that presented for gene panel testing regardless of their personal history of cancer between 2012 and 2015 found that the majority of the *CHEK2* variants were identified in Caucasians (75.9%) and individuals of Ashkenazi Jewish descent (11.1%).(81) This was in comparison to Asian (0.9%), African American (0.7%), Middle-Eastern (0.6%), Native-American (0.1%) and Hispanic (1.3%).(81)

1.4.3.2 *CHEK2* Associated Breast Cancer Risk

It is currently estimated that pathogenic variants in *CHEK2* are identifiable in 1%-2.5% of unselected female breast cancer cases and up to 4.9% of familial cases depending upon the population cohort.(56, 87, 88) Rare protein truncating variants can abolish the protein kinase activity resulting in loss of function and an increased cancer susceptibility.(75, 89) The risk associated with *CHEK2* c.1100delC, p.(Thr367Metfs) is generally considered a moderate risk increase 2-3 fold above baseline population risk (Table 7).(56, 57, 73, 79, 80) These relative and absolute breast cancer risk estimates derived for *CHEK2* c.1100delC, p.(Thr367fs) are likely to be applicable to other protein truncating variants in *CHEK2*.(72) In 2017 Decker et al. found that the aggregated risk estimate derived from the other rare protein truncating *CHEK2* variants was comparable to that associated with *CHEK2* c.1100delC, p.(Thr367fs).(56) Based upon the population risk of breast cancer, these relative risk figures equate to an estimated cumulative lifetime risk of 20-25% by the age of 70 years in the absence of a family history.(73, 78, 82)

Population (year)	Cohort	*CHEK2* Variants	Relative Risk (95% CI)	Sample Population (Mutation Carriers)		Authors
				Cases	Controls	
UK (2017) SEARCH Database	Unselected	All protein truncating	3.11 (2.15-4.69 ($p=5.6x10^{-11}$))	13,087 (213)	5488 (29)	Decker et al. (56)
		c.1100delC, p.Thr367Metfs	3.18 (2.01-4.92 ($p=6.1\times10^{-8}$))			
		Protein truncating (non c.1100delC)	2.83 (1.20-6.69, (p=0.020))			
		Rare Missense	1.36 (0.99 to 1.87 p=0.066))			
		Rare Missense (protein binding domain)	1.51 (1.02-2.24 (p=0.047))			
European (2016) BCAC	Mixed	c.1100delC, p.Thr367Metfs	2.26 (1.90-2.69 ($p=2.3\ x10^{-20}$))	44,777 (710)	42,977 (233)	Schmidt et al. (80)
International (2016) BCAC	Mixed	c.349A>G, p.Arg117Gly	2.03 (1.10 to 3.73) (p=0.020))	42,671 (261)	42,164 (204)	COGS study (57)
		c.538C>T, p.Arg180Cys	1.34 (1.06 to 1.70 (p=0.015))			
		c.715G>A, p.Glu239Lys	1.47 (0.60 to 3.64 (p=0.40))			
		c.1036C>T, p.Arg346Cys	3.39 (0.68 to 16.9 (p=0.11))			
		c.1312G>T, p.Asp438Tyr	0.87 (0.49 to 1.52 (p=0.62))			
European (2008) Meta-analyses	Unselected	c.1100delC, p.Thr367Metfs	2.7 (2.1-3.4)	26,488 (465)	27,402 (142)	Weischer et al. (73)
	Familial	c.1100delC, p.Thr367Metfs	4.8 (3.3-7.2)			

Table 7: *CHEK2* Related Breast Cancer Risks

Evaluation of the published literature regarding *CHEK2*-associated breast cancer risks. Mixed cohorts included both familial and unselected patient groups.

The association between rare missense variants in *CHEK2* and breast cancer susceptibility is more difficult to interpret. Missense variants have differential effects on protein function. This is demonstrable by functional yeast based assays which observe variable *in-vitro* responses to DNA damage for several missense alleles distributed across all domains of the *CHEK2* gene.(74) As such, missense variants in *CHEK2* may have variable effects on breast cancer risk depending on their location. Some will increase breast cancer risk but potentially to a lesser magnitude than protein truncating variants such as *CHEK2* c.1100delC, p.(Thr367fs) whilst others will exert minimal clinical effect. In 2016, The Collaborative Oncological Gene-Environmental Study (COGS) looked at six rare missense variants in *CHEK2* using data derived from 42671 invasive breast cancer cases from BCAC.(57) Only two missense *CHEK2* c.349A>G, p.(Arg117Gly) and *CHEK2* c.538C>T, p.(Arg180Cys) were significantly associated with breast cancer risk. Missense variants

located within the protein binding domains were also noted to confer a higher magnitude of effect.(57)

In 2004 Kilpivaara et al. observed the frequency of the missense founder mutation c.470T>C p.(Ile157Thr) in 1035 unselected breast cancer patients, 507 *BRCA* negative familial breast cancer patients and 1885 healthy controls derived from the Finnish Red Cross blood transfusion service.(75) It was present in 7.4% of unselected breast cancer patient, 5.4% of familial breast cancer and 5.3% of controls.(75) The frequency of this variant was significantly increased amongst unselected female breast cancer patients but the magnitude of risk was less than that seen in association with protein truncating variants (OR 1.43 (95% CI, 1.06-1.95 (p=0.021)).(75) A similar association was seen by Cybulski et al. (OR 1.5 (95% CI, 1.2-1.7)).(90) This lower breast cancer risk was evidenced by functional assays which demonstrated that *CHEK2* c.470T>C (p.Ile157Thr) produced a stable protein that was detectable immunohistochemically.(75) This protein was however able to form heterodimers with wild type *CHEK2* impairing the downstream DNA damage response and thus contributing to an increased breast cancer susceptibility.(75)

In contrast to the aforementioned variants, *CHEK2* c.1111C>T p.(His371Tyr), a novel recurrent missense variant within the activation loop of the kinase domain mutation has been shown to confer a 2.43 fold increased risk of breast cancer in women of Chinese descent.(89) This variant may represent a founder mutation in Asian populations.(56, 89) Overall this highlights that estimations of relative breast cancer risk associated with *CHEK2* missense variants may need to be variant and population specific due to their direct effect on protein function.

1.4.3.3 Modifiers of *CHEK2* Associated Breast Cancer Risk

1.4.3.3.1 Family History

Several studies have shown that the cumulative lifetime risk of breast cancer for a carrier of a pathogenic *CHEK2* variant is markedly higher in the context of a positive family history.(79) In 2008, a meta-analysis published by Weischer et al. found that the aggregated relative risk of breast cancer in individuals with *CHEK2* c.1100delC, p.(Thr367fs) and a family history was 4.8 (95%CI 3.3-7.2) compared to 2.7 (95% CI 2.1-3.4) in an unselected breast cancer population.(73) They estimated the cumulative lifetime risk of breast cancer amongst *CHEK2* c.1100delC, p.(Thr367fs) heterozygotes to be 37% in the context of a family history compared to 21% in the absence of a positive family history.(73)

In 2011, Cybulski et al. assessed the relative contribution of family history to *CHEK2*-associated breast cancer risk in the context of three truncating founder mutations c.444+1G>A, c.1100delC p.(Thr367Metfs) and del5395. (90) They utilised 7496 prospectively ascertained unselected

invasive breast cancer patients of Polish descent and compared them to a population based control.(90) The truncating variants were collectively associated with a 3.6 fold increase in the relative risk of breast cancer (95%CI 2.6-5.1).(90) This relative risk increased with a positive family history to 5.7 fold if one first degree relative was affected (95% CI 3.6-9.2) and 7.3 fold if both a first and second degree relative were affected (3.2-16.8).(90) This corresponded to an absolute lifetime risk of 34% and 44% respectively in the Polish population compared to 20% in a woman with no family history.(88, 90)

This excess risk in the context of a family history is probably in part attributable to complex polygenic factors including epigenetics, single nucleotide polymorphisms and shared familial environmental exposure.(88) Muranen et al. looked at the combined effects of Polygenic Risk Scoring (PRS) and *CHEK2* c.1100delC, p.(Thr367fs). They predicted a multiplicative relationship between *CHEK2* c.1100delC, p.(Thr367fs) and common low penetrance variants suggesting these may better help to risk stratify patients.(79)

1.4.3.3.2 Age

In recent years there has been some consideration that *CHEK2*-associated breast cancers may be associated with an earlier age of onset. Weischer et al. demonstrated within the BCAC cohort that *CHEK2* c.1100delC, p.(Thr367fs) carriers were on average 4 years younger at breast cancer diagnosis than non-carriers (p=0.001) and were more likely to be premenopausal (p=0.001).(91) Schmidt et al. and Decker et al. also demonstrated that the relative risk of *CHEK2*-associated breast cancer significantly reduced with age (Table 8).(56, 72)

Decker et al.		**Schmidt et al.**	
Age (yrs)	**OR (95% CI)**	**Age (yrs)**	**OR (95% CI)**
< 50	3.98 (2.62-6.21)	35-50	2.57 (1.83-3.59)
50-60	3.37 (2.24-5.22)	50-65	2.36 (1.80-3.10)
> 60	2.12 (1.35-3.41)	>65	1.40 (0.93-2.12)
$P_{trend} = 1.2x10^{-5}$		$P_{trend} = 0.014$	

Table 8: Relative Risk of *CHEK2* Associated Breast Cancer

Schmidt et al. and Decker et al. also demonstrated that the relative risk of *CHEK2*-associated breast cancer significantly reduced with increasing age. The analysis by Schmidt et al. included c.1100delC, p.(Thr367fs) whilst Decker et al. included several rare protein truncating variants.(56, 72)

1.4.3.3.3 Biallelic Mutations

Homozygous *CHEK2* c.1100delC, p.(Thr367fs) variants are associated with an elevated cancer risk in excess of that associated with *CHEK2* c.1100delC, p.(Thr367fs) heterozygosity.(82) In 2011,

Adank et al. identified 8 individuals homozygous for *CHEK2* c.1100delC, p.(Thr367fs) in a Dutch, familial non-*BRCA* cohort.(82) They estimated that *CHEK2* homozygosity would confer a 4 fold increased breast cancer risk within a familial cohort.(82) This risk was predicted to be 6 fold if extrapolated to a population based cohort.(82)

1.4.3.4 Other Associated Cancer Risks

A number of studies have tried to determine potential associations with pathogenic *CHEK2* mutations and other cancer risks including prostatic, ovarian and colorectal. Much of the current data is conflicting. Naslund-Koch et al. observed 86,975 patients between 2003 and 2010 with records linked to the Danish Cancer Registry. Of these, 670 (0.8%) were heterozygotes for *CHEK2* c.1100delC, p.(Thr367fs).(92) When correcting for age and sex they found an association between this genotype and the risk of sarcoma, stomach, renal and prostate cancer.(92) The relationship was less apparent after taking multiple tests into consideration. (92)

1.4.4 Ataxia-Telangiectasia Mutated (*ATM*)

The Ataxia-Telangiectasia Mutated (*ATM*) gene is a Serine Threonine Kinase composed of 66 exons composed of multiple functional domains (Figure 7).(93) It initiates the signalling cascade necessary for HRR, mediated by the phosphorylation and activation of downstream proteins including BRCA1, BRCA2, PALB2, CHEK2 and TP53 (Figure 8).(93-95) Pathogenic heterozygous mutations are associated with an increased risk of breast cancer. Homozygous or compound heterozygous mutations are associated with Ataxia Telangiectasia (AT).(93)

Figure 7: Structure of *ATM*

The *ATM* gene is a Serine Threonine Kinase composed of 66 exons composed of multiple function domains. It has a highly conserved FAT domain necessary for binding of regulatory proteins. The N terminus TAN domain influence protein, protein interactions.(29)

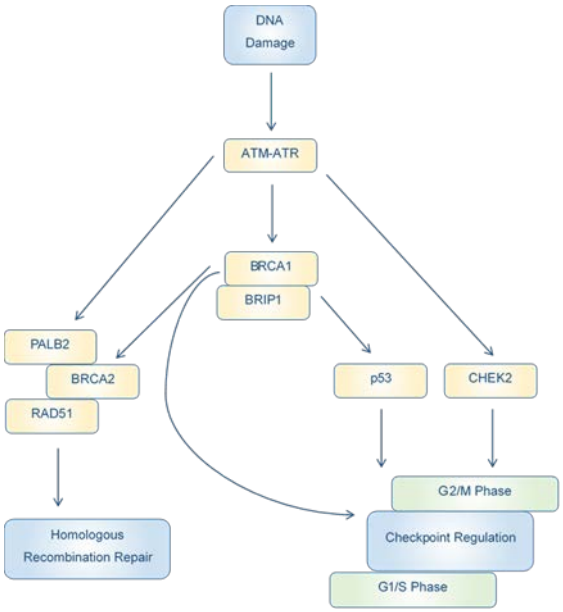


Figure 8: Mechanism of Action of *ATM*

1.4.4.1 Epidemiology

ATM is a globally important breast cancer susceptibility gene and pathogenic variants have been shown to contribute to breast cancer susceptibility in European, Chinese Han, Polish and South American populations.(96-98) The population frequency of *ATM* heterozygotes is 0.35-1% and this may be higher in the context of familial breast cancer.(95, 99) For example, Tavera et al. found the prevalence of pathogenic germline mutations to be 1.78% in the context of familial breast cancer derived from a Spanish population.(93)

1.4.4.2 *ATM* Associated Breast Cancer Risk

ATM is generally considered to be a moderate risk breast cancer gene. The outcome of several population and familial studies demonstrates that collectively, pathogenic protein truncating variants confer a relative breast cancer risk 2-3 fold above the population level (Table 9).(56, 57, 93, 100, 101) In 2016 Marabelli et al. performed a meta-analysis utilising 19 papers defining *ATM*-associated breast cancer risks.(28) The populations sampled were heterogeneous incorporating both sporadic and familial breast cancer cohorts derived from several different global centres.(28) They determined the cumulative risk of breast to be 6.02% by the age of 50 years and 32.83% by the age 80 years.(28)

Population (year)	Cohort	*ATM* Variants	Relative Breast Cancer Risk (95% CI)	Sample Population (gene carriers)		Authors (ref)
				Cases	Controls	
UK (2017) SEARCH Database	Unselected	Protein truncating	3.26 (1.82-6.46 ($p=2.1\times10^{-5}$))	13,087 (85)	5488 (11)	Decker et al. (56)
		Rare Missense	1.18 (0.99-1.40 (p=0.073))			
		Rare Missense (FAT and PI3K domains)	1.71 (1.12-2.61 (p=0.015)			
International (2016) BCAC	Mixed	c.7271T>G, p.Val2424Gly	11.0 (1.42-85.7 (p=0.0019))	42,671 (12)	42,164 (1)	COGS study (57)
International	Familial	Protein truncating	2.32 (1.12-4.83)	2531	2245	Tavtigian et al. (100)
		Rare Missense	18 (2.82 - 117)			
UK (2006) Multi-Centre	Familial	Protein Truncating Pathogenic Missense	2.37 (1.51-3.78 (p =0.0003))	443	521	Renwick et al. (101)

Table 9: *ATM* Related Breast Cancer Risks

Evaluation of the published literature regarding *ATM*-associated breast cancer risks. Mixed cohorts included both familial and unselected patient groups.

The full spectrum of mutations which are associated with *ATM* mediated cancer risk is not fully defined, in part because of the large gene size making the attribution of pathogenicity challenging. (100) The association between *ATM* missense variants and breast cancer risk is more variable. For some missense variants there are no or conflicting reports about pathogenicity.(102, 103) Decker et al found that the some missense variants in *ATM* have no clinical effect whilst other rare missense variants may elevate breast cancer risk above the basal population level.(56). In 2010 Fletcher et al. genotyped five SNPs with a MAF of 0.9%-2.6% in 26101 breast cancer cases and 29842 controls derived from the BCAC consortium.(104) They found that these SNPs conferred a small contribution to breast cancer risk OR 1.06 (p_{trend}=0.04).(104)

Other, rare missense variants have been identified to convey a higher risk of breast cancer than protein truncating variants. In 2016, the COGS study looked at one rare missense variant in *ATM*, c.7217T>G p.(Val2424Gly).(57) It was identified in 12 cases and 1 control and found to be significantly associated with breast cancer risk (OR 11.0 (95%CI, 1.42-85.7 (p=0.0012)).(57) Whilst, the number of gene carriers within the COGS sample population was small, Goldgar et al. also found this variant to have a particularly high associated breast cancer risk.(27) They studied 2,570 invasive breast cancers from familial breast cancer cohorts and 1448 controls for the presence of consensus splice site, truncating and evolutionary unlikely missense substitutions.(27) *ATM* c.7271T>G was the most frequently identified variant associated with an 8 fold increase in breast cancer risk by (p=0.0005).(27) Tavtigian et al. also found the associated cancer risk was higher for

evolutionarily unlikely rare missense variants (OR 18 (95%CI, 2.82-117).(100) One hypothesis for this particular elevation in risk associated with rare missense variants is a dominant negative effect.(105)

1.4.4.3 Other Associated Cancer Risks

There are no clear associations between pathogenic mutations in *ATM* and other cancer risks. The COGS study looked at *ATM* c.7217T>G p.(Val2424Gly) to determine whether there was an associated risk of prostate or ovarian cancer using cases derived from the PRACTICAL and OCAC consortia.(57) They found no association with this specific variant and these cancer risks.(57)

1.4.5 Tumour Protein p53 (*TP53*)

The Tumour Protein p53 (*TP53*) gene encodes p53 and consists of 11 exons. Exon 1 has two transcriptional start sites whilst exons 2-11 are coding.(106) There are several functional domains including an oligomerisation domain and the core DNA binding domain which is encoded by exons 5-8.(106, 107)*TP53* is known as the "Guardian of the Genome" and acts as an important cell cycle checkpoint regulator including for the initiation of DNA repair or apoptosis following DNA damage.(106) It is one of the most frequently encountered somatic mutations in cancer.(106, 107).

Most pathogenic variants are missense, representing 73% of all germline mutation types within the IARC database.(107) Missense variants in the DNA binding domain often effect the proteins quaternary structure or ability to bind DNA.(107) Codons 196 and 213 represent hotspots for nonsense mutations whilst 152, 209 and 241 are associated with indels.(107) Increasingly, the oligomerisation domain located at the C terminus (codons 323-356) is also recognised as a mutational hotspot.(107)

Pathogenic variants in *TP53* are associated with Li-Fraumeni Syndrome (LFS). This is a hereditary cancer predisposition syndrome associated with familial clustering and high lifetime risks for a wide variety of tumour types including Adrenal-Cortical Carcinoma (ACC), Central Nervous System (CNS) tumours, pre-menopausal breast cancers, osteosarcoma and soft tissue sarcomas (STS) including rhabdomyosarcoma.(108, 109) ACC, CNS and STS typically affect children with a further peak in the incidence of CNS and STS later in life.(107) Osteosarcomas are more classically observed in adolescence.(107) A number of diagnostic criteria exist for the identification of this condition including the Chompret Criteria and a broader classification of Li-Fraumeni-Like (LFL) which includes those families who do not meet the classical diagnostic criteria.(108, 110)

1.4.5.1 Epidemiology

In 2003, Lalloo et al. estimated the carrier frequency of pathogenic *TP53* variants as 1/5000 within the general population.(111) This estimate was derived from patients diagnosed with very early onset breast cancer (less than 30 years) and recorded in the North Western Cancer Registry, UK.(111) It has been estimated that up to 3-8% of unselected very early onset breast cancer will have a pathogenic variant in *TP53* with a higher prevalence in the context of a LFS/LFL pedigree (Table 10).(106, 112-117) In contrast, the estimated prevalence in patients diagnosed with breast cancer at any age and ascertained through familial or mixed cohorts is lower ranging from 0.5-2%.(118) This figure further reduces to less than 0.5% in patients diagnosed with breast cancer at any age and ascertained through gene panel testing in an unselected cohort. Buys et al. identified 61 pathogenic *TP53* variants in an unselected cohort of 35,409 women identified in this manner equating to a detection rate of 0.17%.(119)

A proportion of the *TP53* pathogenic variants identified outside the context of a LFS/LFL family history will represent de-novo variants.(106) In 2009, Gonzalez et al. determined the proportion of de-novo pathogenic *TP53* variants in a case series of 341 American patients with early onset breast cancer.(120) The estimated de-novo mutation rate was 5-20% was based upon a combination of molecular genetic testing and family history data.(120) As such, many Clinical Genetics centres now advocate genetic testing for pathogenic variants in *TP53* in all *BRCA* negative breast cancers diagnosed under the age of 31 years irrespective of family history.(115)

Age at Breast Cancer Diagnosis	Population	Sample Size	Prevalence	Study
Less than 30 years (unselected)	Canadian	28	5-8%	McCuig et al. 2012 (112)
Less than 35 years (unselected)	Irish	123	3%	O'Shea et al. 2018 (113)
Less than 31 years (unselected)	French	1730	6%	Bougeard et al. 2015(114)
Less than 35 years (unselected)	Malaysia	83	6%	Lee et al. 2012 (115)
Less than 30 years (unselected)	UK	100	4%	Lalloo et al. 2006 (116)
Less than 30 years (unselected)	Australian	52	3.8%	Mouchawar et al. 2010 (121)
Less than 40 years (familial cohort)	Australian	42	7%	Mouchawar et al. 2010 (121)
Less than 40 years (mixed cohort, HER2 positive)	USA	213	2.5%	Rath et al. 2013(117)
Any age (mixed cohort, HER2 positive)	USA	213	1.4%	Rath et al. 2013(117)
Any age (familial cohort)	USA	190	0.5%	Moran et al. 2017 (118)

Table 10: Prevalence of *TP53* Pathogenic Variants in Very Early Onset Breast Cancer

One founder mutation has been identified in the Southern and South-Eastern Brazilian population c.1010G>A, p.Arg337His with an associated allele frequency of 0.3%.(106, 110) It is found in 2.4% of women with breast cancer in Southern Brazil.(122) This is a lower penetrance variant which may not fulfil the classic LFS diagnostic criteria (although the full spectrum of LFS associated tumours can be seen).(110) It is estimated that, 15% of mutation carriers in this gene develop

cancer by the age of 30 years compared to 50% of classically pathogenic mutations.(109) The differential cancer susceptibility may be related to the impact upon DNA binding.(109) This founder mutation is located in the oligomerisation domain in which the arginine codes for an alpha-helix motif.(123, 124) The protein structure is very similar to wild type *TP53* but highly sensitive to alterations in pH resulting in a functional deficit and impaired oligomeristaion in a pH dependent manner.(123, 124) It is associated with the full spectrum of LFS associated tumours but predominately breast cancer, ACC and choroid plexus carcinomas.(124)

In 2017, Lolas et al. also identified a potential Palestinian founder mutation, c.541C>T p.(Arg181Cys).(125) The population cohort was enriched for a familial predisposition and included 453 women diagnosed primary invasive breast cancer at the age of 40 years or younger or with a family history of HBOC in another close relative.(125) They concluded that *TP53* p.Arg181Cys may produce a breast cancer predominant phenotype which is not diagnostic of classic LFS but similar to that of R337H observed within the Brazilian population.(125)

It raises the question about whether these specific variants should be the focus of targeted genetic testing in patients with pre-menopausal breast cancer derived from the representative populations at risk.(110, 122)

Specific polymorphisms which affect *TP53* function may also be associated with an altered cancer risk that is population and tumour type specific.(126, 127) For example, polymorphisms in the second nucleotide of codon 72, have been associated with increased susceptibility to breast cancer risk.(128) Variants at this location are associated with an alteration in the encoded amino acid to either Proline (R72P) or Arginine (R72R) which have differential structure and biological function effects.(128, 129)

Hossain et al. observed the effect of the Arg/Arg, Arg/Pro and Pro/Pro codon 72 genotype amongst 125 breast cancer patients and 125 age matched controls derived from the Bangladeshi population.(128) The Pro/Pro genotype was associated with a significant increase in breast cancer risk OR 2.5 (95% CI 1.19–5.33) (p = 0.0157).(128) In 2014, Goncalves et al. performed a meta-analysis looking at the breast cancer risk associated with the R72P and R72R polymorphisms in 25,629 cases and 26,633 controls derived from European, Asian, American and African populations with the highest representation for European and Asian studies.(130) They found an increased breast cancer susceptibility associated with the R72P allele except in the Asia population where the R allele was most strongly associated.(130) Damin et al. suggested that the population specific effect may be attributable to UV light exposure with an increased prevalence of the R allele with increasing latitude from the equator.(129)

More recently, a 16bp indel in intron 3 (rs17878362) has also been associated with an increased cancer risk in Indian, Mediterranean and Northern European populations.(127)

1.4.5.2 *TP53* Associated Breast Cancer Risk

In 2015 Bougeard et al. observed the *TP53*-associated cancer risk amongst 415 *TP53* mutation carriers in the French population.(114) The mean age of tumour onset was 24.9 years (25.9 years female and 22.7 years male).(114) A genotype-phenotype correlation was observed in relation to age of onset with dominant-negative missense mutations having the earliest age of onset (21.3 years).(114) In 2016, Mai et al. determined the cumulative cancer risks amongst 286 *TP53* positive individuals ascertained from the National Cancer Institute LFS cohort.(108) The cumulative cancer incidence was 50% by age 31 for female carriers and by age 46 for male carriers with the differential age effect influenced by the association with pre-menopausal breast cancer.(108) The absolute cancer risk was almost 100% by age 70 years regardless of sex and the risk of a second cancer was 49%.(108)

Breast cancer is the most frequently observed cancer amongst female *TP53* gene carriers and the risk is evident from age 20 years.(108, 114) *TP53*-associated breast cancers are typically pre-menopausal and are often early onset with a median age at diagnosis of 34 years and a peak incidence amongst women aged 30 years or younger.(106, 107).(107) Melham-Bertrandt et al. demonstrated that the likelihood of *TP53*-associated breast cancer was significantly associated with age and HER2 amplification. They estimated that the likelihood of having a *TP53* mutation and breast cancer decreased by 5% for every year of age.(131)

Determination of genotype, phenotype correlations in the context of *TP53* germline variants remains a particular focus of current research.(107) Birch et al. found that individuals with missense mutations in the DNA binding domain were significantly more likely to have a more penetrant cancer phenotype with earlier ages of onset.(132) There is some emerging evidence to suggest that specific *TP53* variants may be associated with a breast cancer only phenotype.(106, 107)

1.5 Genetic Testing for Hereditary Cancer Susceptibility

A genetic test is designed to identify a germline or somatic variant associated with a particular disease, in a defined population for a specified purpose.(8, 133, 134). Historically, genetic testing for a heritable susceptibility to breast cancer has focused on single gene sequencing in an iterative manner based upon the presenting cancer phenotype and family history. Often, access to targeted genetic testing for high risk breast cancer susceptibility has been dependent upon an

individual's personal and family history of breast and/or ovarian cancer and the outcome of breast cancer gene carrier probability models.(135)

The most commonly utilised models are the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) and the Manchester Score.(136-138) These scoring systems use several factors including the number of familial cancer cases, cancer type including hormone receptor status and age of onset to determine the likelihood of a gene alteration.(136-138) A likelihood score of greater than or equal to 10% is advocated by the National Institute for Clinical Excellence (NICE) as the threshold for genetic testing.(135) These models are heavily weighted towards the family history of cancer potentially reducing their efficacy in the identification of moderate risk genotype carriers where the cancer phenotype penetrance is recognised to be lower.

Advances in Next Generation Sequencing Technology (NGS) and a reduction in cost has widened access to cancer genomic technology in both mainstream cancer care and through private commercial organisations.(2) It has enabled a shift from single gene testing based upon carrier probabilities to more broad genomic tests including cancer susceptibility gene testing at the time of cancer diagnosis to identify germline cancer susceptibility. This includes targeted gene panels and whole exome/genome sequencing with targeted in-silico analysis. These germline genomic tests allow the simultaneous analysis of multiple cancer susceptibility genes (for example *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53*) in a similar timeframe and with comparable costs to conventional single gene testing.(118, 139) It has the potential to increase the diagnostic yield and identification of actionable risk.(3, 37) This testing can be cascaded throughout a family for the facilitation of primary and secondary prevention.

1.6 Gene Panel Assembly

Given the recognised benefits and limitations of gene panel testing, at present, there is considerable variability in hereditary breast cancer genetic testing between different commercial and healthcare organisations at a global level. In a utilitarian health service such as The National Health Service (NHS), any diagnostic tests and the related costs attributable to that test including primary and secondary risk intervention need to be evidence based, cost effective, sustainable and equitable at a population level.(133)

One framework which enables consideration of these factors is the ACCE Model. This assesses genetic tests on the basis of four components, Analytic validity, Clinical validity, Clinical utility and associated Ethical, legal and social implications (Figure 9).(9, 10, 139)

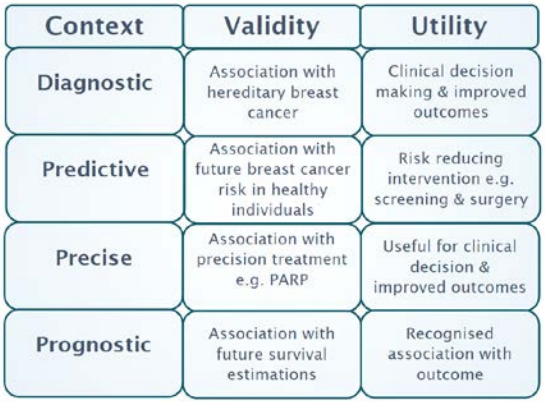


Figure 9: Framework for the Evaluation of Gene Tests

Assessment of the clinical validity and clinical utility of a genetic test is recommended to evaluate its suitability for diagnostic use. Central to the evaluation of utility is consideration of the context in which a result will be applied.(8, 9, 133)

1.6.1 Evaluation of Analytic Validity,

Analytic validity considers whether a test is accurately and reliably able to measure the genotype of interest.(9) Direct sequencing technologies have the potential to achieve a diagnostic accuracy close to 100%.(8) This is particularly important for cancer susceptibility genes which demonstrate both allelic and locus heterogeneity.(8, 9)

1.6.2 Evaluation of Clinical Validity and Utility

Clinical validity and utility considers whether the test will provide useful information for the care of patients.(8, 133) This includes the prediction of future cancer risks amongst healthy relatives with evidence based intervention strategies including screening and risk reducing surgery that are cost effective, sustainable and can be delivered equitably across the population.(133). A number of guidelines exist regarding the optimal management for individuals with moderate and high risk cancer susceptibility genotypes. These include publications from The National Institute of Clinical Excellence (NICE), The European Society of Medical Oncology (ESMO) and The National Comprehensive Cancer Network (NCCN).(11, 140, 141) In practice, the higher the estimations of cancer risk and the more effective and acceptable the options regarding that risk, then the higher the legitimacy of the test.(133)

1.6.3 Evaluation of Context

The purpose or context of the genetic test should also be considered critical to the evaluation process.(9, 133) As such we need to be clear about the clinical question that we are aiming to

answer from a gene panel test and whether it has the potential to fully or partially answer this question. Important considerations are whether the information derived from a gene panel test can be used to diagnose a cancer susceptibility syndrome or is accurately predictive of future cancer risk in an unaffected family member. Other considerations include whether the genotypic information can be used for primary cancer management decisions and for the determination of future prognosis (Figure 9).

1.7 Practice Standardisation for Oncogenetic Testing

The ACCE framework has been adapted by the United Kingdom Genetic Testing Network (UK-GTN) to create a dossier of approved genetic tests for utilisation within the NHS. More recently, this has been transferred to the National Genomic Medicine Test Directory.(142) In 2018, the UK Cancer Genetics Group (UK-CGG) also issued a Consensus Statement to provide guidance regarding the standardisation of cancer gene panel testing based upon these principles.(9) This advocated testing of *BRCA1*, *BRCA2*, *PALB2*, *CHEK2* (protein truncating variants), *ATM* (protein truncating variants and c.7271T>G), *TP53*, *PTEN* and *STK11* as a breast cancer susceptibility gene panel.(9)

1.8 Considerations and Limitations of Cancer Genomic Testing

This review has demonstrated that access to targeted genetic testing for high and moderate risk breast cancer susceptibility genes has often been dependent upon an individual's personal and family history of breast and/or ovarian cancer and the outcome of breast cancer gene carrier probability models.(135)

This approach may miss opportunities to identify actionable risk. It demonstrates the need for enhanced tools which can assist in identifying when to perform genetic testing particularly for lower penetrance, moderate risk genotypes were the family history may be discordant with the threshold for genetic testing defined by conventional methodologies such as the Manchester score or BOADICEA. This is also relevant for non-classical pedigrees which are small or male dominant and de-novo mutations which is a particular consideration for *TP53*.

Despite the recognised benefits of broader cancer genomic testing to identify germline cancer susceptibility variants, it is important to consider the true clinical utility of this information. This is particularly true for moderate risk genes when the associated cancer risk and optimal management guidelines are often less well characterised.(2, 3) In addition, identifying a moderate risk variant in the context of a family history of the same may not influence management.(143)

It is also important to consider the limitations of testing. The utilisation of broader techniques such as panel tests and exome sequencing with lower prior probabilities of genetic risk increases the potential for identifying Variants of Uncertain Clinical Significance (VUS). In 2016, the Wessex Regional Genetics Laboratory reported a 3% VUS rate following targeted testing for *BRCA1* and *BRCA2*. With a gene panel, the rate of VUS rises exponentially in relation to the number of genes tested with a modest increase in actionable risk.(144) It may create a significant downstream cost associated with VUS interpretation and potential for harm through misinterpretation and inappropriate intervention.(145, 146)

Gene panel tests also create the potential to identify variants in cancer susceptibility genes which are considered unrelated to the primary presentation.(3) The interpretation of genetic variants can be particularly challenging when the family history and primary malignancy is discordant with the variant identified. For example, does a pathogenic *TP53* gene alteration identified in an isolated pre-menopausal breast cancer patient constitute Li Fraumeni Syndrome?(147) These variants have the potential to be erroneously linked to cancer susceptibility, a recognised phenomenon even with well characterised genes.(145)

Given these potential pitfalls, we need to identify adjuncts to identify variant carriers and enhance variant interpretation within routine clinical practice. One such adjunct is tumour phenotype. This includes both tumour histopathology and somatic mutational profile.

1.9 Breast Cancer Histopathology

Several breast cancer subtypes exist with differential incidence, prognosis and treatment responses. This grouping is based upon gene expression profiling and includes Luminal A, Luminal B, Basal-Like, and Human Epidermal Growth Factor (HER2) enriched.(148-150). The luminal subtype accounts for up to 60% of breast cancers and is characterised by the expression of genes related to oestrogen expression and other components of the luminal epithelium.(151) The basal-like subtype is observed in 15-20% of breast cancer and is associated with elevated expression of genes present in the basal myoepithelial cells.(151) HER2 enriched tumours represent 10-15% of breast cancers and are associated with high expression of genes located in the *HER2* amplicon.(151)

The described breast cancer subtypes roughly equate to recognised morphological and immunohistochemical tumour phenotypes. For example, Luminal A tumours are classically low grade, ER-positive, PR-positive and HER2-negative whilst Luminal B tumours are high grade and ER-positive. Basal like tumours are often Triple Negative (TN) which means they are ER-negative,

PR-negative and HER2-negative.(148-150) Increasingly, we are recognising associations between genotype and tumour histopathological phenotype.(152)

1.10 Germline Genetic Variation and Tumour Histopathology

Histopathological phenotype may serve as a useful predictor of germline mutational status which can assist in identifying when to perform genetic testing. For example, *BRCA1* associated tumours are classically high grade, basal and TN with a high mitotic index.(153, 154) *BRCA2* related tumours are less distinctive but are often higher grade with a luminal B phenotype and continuous pushing margins.(25, 153-156) TN breast cancer comprises less than 20% of all breast cancer diagnoses.(157) However, it is now recognised that early onset TN breast cancer is an indication for *BRCA* testing even in the absence of a family history. This form of tumour phenotyping may also serve as an adjunct to current methodologies for the interpretation of VUS.(154)

There is variability in the described histopathological tumour phenotype seen amongst *PALB2* mutation carriers. There is some association with ER-negative and triple negative disease but this is not a consistent finding across all studies and some report no association between *PALB2* genotype and ER status.(30, 53, 56, 57, 66) It remains to be determined whether this is a positional effect mediated in part by the location of the pathogenic variants within the gene such as the *BRCA1* or *BRCA2* binding domain (Table 11).

There is an evolving *CHEK2*-associated tumour phenotype. In general, they are significantly more likely to be Grade 2, ER and PR-positive compared to non-carriers (Table 11). (56, 72, 90, 91, 158) A number of studies also demonstrate a trend towards bilateral disease at presentation with higher levels of nodal involvement suggestive of a more aggressive underlying tumour biology. Decker et al. found that *CHEK2*-associated tumours were significantly more likely to be bilateral at presentation (OR=3.27 (95% CI 1.66 - 5.83) p=0.0014).(56) This was further supported by Kilpivaara et al. who also noted a strong association with bilateral disease and Cybulski who identified higher levels of nodal involvement.(90, 158)

There is minimal data regarding the *ATM*-associated histopathological tumour phenotype (table 8). Balleine et al. did not observe a clear *ATM*-associated phenotype amongst 25 breast cancer cases.(152) Decker et al. found that *ATM*-associated tumours were significantly more likely to be ER-positive than non-carriers.(56)

TP53 related breast cancers are associated with HER2 amplification and hormone receptor positivity.(106, 131, 159) The presence of a family history of breast cancer and a HER2-amplified

tumour can be predictive of germline *TP53* mutational status. In 2016, Eccles et al. analysed 591 patients with early onset breast cancer derived from the POSH cohort with HER2-amplified tumours. The combination of a BOADICEA score suggestive of a 10% threshold of a *BRCA* mutation in combination with a HER2-amplified tumour was predictive of a germline *TP53* mutation. In total, 7/59 (12%) patients who met the 10% BOADICEA threshold with early onset breast cancer and a HER2-amplified tumour carried a *TP53* gene alteration.(160) Conversely, the presence of a HER2-amplified tumour in the absence of meeting the diagnostic threshold for *BRCA* or LFS testing was poorly predictive of a germline *TP53* mutations 1/195 (0.5%).(160) A similar finding was noted by Rath et al.(117) They observed a low frequency of germline *TP53* mutations amongst unselected women with early onset breast cancer and HER amplification (2.5%, 95%CI 0.3% to 8.7%).(117) Of note, it has been suggested that the *TP53*, p.(Arg337His) immunophenotype may be less frequently HER2-amplified. Fitarelli-Kiehl et al. analysed immunophenotyping from 66 p.(Arg337His) mutation carriers compared to 12 carriers of other germline pathogenic variants and found that HER2 amplification was seen in 22.7% versus 75%. It remains to be determined whether this is a true association and if so whether it is a domain specific effect.(123)

Gene	Sample Population (Mutation Carriers)	Variant	Hormone Receptor Status Breast Cancer Cases		Odds Ratio (95% CI)	%	P Value	Authors (ref)
PALB2	13,087 (89)	Truncating	ER	ER-	5.58 (2.19-15.2)	*	0.55	Decker et al. (56)
				ER+	4.32 (2.07-10.5)	*		
			PR	PR-	6.21 (2.5-16.7)	*	*	
				PR+	4.16 (1.77-10.8)	*		
	42,671 (79)	c.1592delT p.Leu531Cysfs	ER	ER-	6.49 (2.17-19.4)	*	0.0023	COGS study (57)
				ER+	2.24 (1.05-7.24)	*		
		c.3113G>A p.Trp1038*	ER	ER-	*	*	0.15	
				ER+	*	*		
	12,529 (116)	c.172_175del4 p.Gln60Argfs	ER	ER- carrier	*	40	0.031	Cybulski et al. (66)
				ER- non-carrier	*	30		
			PR	PR- carrier	*	45	0.0004	
		c.509_510del p.Arg170Ilefs		PR- non-carrier	*	29		
			TNT	TNT carrier	*	34	<0.0001	
				TNT non-carrier	*	14		
	3927 (39)	c.1592delT p.Leu531Cysfs	ER	ER- carrier	*	46.7	0.0008	Heikkinen et al. (53)
				ER- non-carrier	*	20.9		
			PR	PR- carrier	*	56.7	0.0095	
		c.509_510del p.Arg170Ilefs		PR- non-carrier	*	33.8		
			TNT	TNT carrier	*	54.5	<0.0001	
				TNT non-carrier	*	12.2		
CHEK2	13,087 (213)	Truncating	ER	ER +	3.42 (2.33 - 5.21)	*	0.0032	Decker et al. (56)
				ER -	1.59 (0.80-3.00)	*		
			PR	PR +	3.87 (2.51-6.12)	*	0.18	
				PR -	1.75 (0.89-3.25)	*		
	44,777 (710)	c.1100delC, p.Thr367Metfs	ER	ER +	2.55 (2.10-3.10)	*	$9.9*10^{-6}$	Schmidt et al. (80)
				ER -	1.32 (0.93-1.88)	*		
			PR	PR +	2.51 (2.02-3.12)	*	$1.7x10^{-2}$	
				PR -	1.72 (1.29-2.30)	*		
	26,488 (465)	c.1100delC, p.Thr367Metfs	ER	ER+ carrier	*	63	<0.0001	Weischer et al. (91)
				ER+ non-carrier	*	57		
			PR	PR+ carrier	*	46	0.01	
				PR+ non-carrier	*	43		
	7931 (227)	Truncating	ER	ER+ carrier	*	69.4	0.002	Cybulski et al. (90)
				ER+ non-carrier	*	63.1		
			PR	PR+ carrier	*	77.8	<0.001	
				PR+ non-carrier	*	68.7		
ATM	13,087 (85)	Truncating	ER	ER+	3.19 (1.73 – 6.47)	*	0.11	Decker et al. (56)
				ER-	1.59 (0.80 – 3.00)	*		
			PR	PR-	*	*	*	
				PR+	*	*		
	24 (24)	Pathogenic protein truncating and missense	ER	ER+	100%	*	*	Weigelt et al.(161)

Table 11: Association Between Genotype and Tumour Histopathological Phenotype

Overview of several studies documenting the association between germline genotype and hormone receptor status.

1.11 Germline Genetic Variation and Somatic Molecular Profile

1.11.1 Tumour Mutational Burden

Breast cancer arises due to the progressive accumulation of somatic mutations. Each tumour will contain thousands of somatic mutations which are broadly divided into driver and passenger events.(162) Somatic driver mutations are implicit to cancer evolution and result in the development of fundamental biological capabilities or "hallmarks" which provide a survival advantage for clonal evolution.(163-166) These include sustained proliferation, evasion of growth suppression, the potential for invasion and metastasis, replicative immortality, the induction of angiogenesis and ability to evade apoptosis.(166)

Passenger mutations arise as a consequence of defective DNA repair and environmental exposure. They are not subject to selective pressures and do not confer a survival advantage but contribute to the overall Tumour Mutational Burden.(162, 164, 165) Genetic variants include single nucleotide variants, indels, copy number variants, amplifications, structural change and epigenetic modification.(167)

Tumour Mutational Burden (TMB) is defined as the total number of sequence variants per megabase of DNA and can include both synonymous and non-synonymous variants.(168, 169) Passenger mutations contribute to the overall TMB. A high TMBs or hyper-mutant state is the presence of greater than 10 somatic mutations per megabase of DNA (mut/Mb), whilst an ultra-hypermutant state is described as over 100 mut/Mb.(168-170)

Somatic mutational profiling allows assessment of the mutational burden associated with each tumour.(165) Childhood cancers and haematological malignancies have the lowest TMB.(165) This is because there are likely to be fewer cell divisions reducing the opportunity to acquire somatic mutations. Conversely, those associated with chronic environmental exposure such as lung cancer, melanoma and bladder cancer have a higher burden.(165, 168, 169) More recently, it has also been shown that somatic genomic instability can be also be acquired as a consequence of inherited germline variation including Mismatch Repair (MMR) gene deficiency and DNA Polymerase Epsilon (POLE) deficiency.(167-169)

The catalogue of mutations present in any cancer is indicative of the molecular processes and biological pathways affected by somatic mutation change that influence evolution.(171) It enables us to understand the mutational processes underlying tumorigenesis.(164, 165, 171)

The somatic mutational landscape of cancer is complex and the Catalogue of Somatic Mutations in Cancer (COSMIC) group has produced a Cancer Gene Census (CGC). This is a list of genes identified to be mutated in human cancer with a recognised function in cancer development. These are classified into tiers according to their functional significance. At present, there are over 500 Tier 1 breast cancer genes.(172) Tier 1 somatic variants have a strong clinical significance and proven activity relevant to breast cancer oncogenesis.(171) They may affect specific biological pathways such as MAPK or Fanconi Anaemia.(171). Several of these will be in Hallmark genes which confer fundamental biological capabilities and provide a survival advantage for clonal evolution. (171) Mutation types include structural rearrangements, copy number variation, indels and single base substitutions.(173) They occur gradually during tumour development resulting in clonal diversity.

The most frequently mutated genes in human breast cancer are *TP53*, *PIK3CA* and *GATA3*.(174) There is additional evidence to suggest somatic mutational patterning based upon the PAM50 molecular subtypes.(175) Luminal breast cancers frequently display somatic activation of PIK3-AKT signalling including PIK3CA, MAP3K1 and MAP2K4.(174, 175) Basal like tumours have a higher mutational burden compared to Luminal A/oestrogen receptor positive tumours.(174, 176) *HER2*-amplified tumours may also have a higher mutational burden.(174, 176)

1.11.2 Somatic Mutational Signature

A mutational signature is the imprint left on the cancer genome following exposure to a specific aetiological factors. A signature is the nucleotide base change and the flanking sequence bases.(164) There are six base substitutions (C>A, C>G, C>T, T>A, T>C, and T>G) and 16 possible flanking sequences resulting in 96 potential trinucleotide changes overall.(164) A computational framework is used to observe the relative proportions of these trinucleotide changes to determine the mutational signatures which are present.(164)

There are several recognised exogenous and endogenous risk factors and pathways for somatic mutation.(164) Exogenous, environmental factors include exposure to ultraviolet radiation or known carcinogens including benzo(a)pyrene diol epoxide within tobacco smoke and alkylating agents.(164, 165) Endogenous pathways includes the progressive spontaneous deamination of 5 methyl-cytosine with advancing age.(164) Further endogenous events include base incorporation during DNA polymerase mediated DNA replication or impaired DNA damage response pathways

such as Mismatch repair (MMR) or Homologous Recombination Repair (HRR).(164) These processes can be continuous such as the spontaneous deamination of 5 methyl-cytosine from the point of conception or episodic.(164, 177) Each of these exogenous and endogenous factors is associated with a specific mutational signature (Appendix B).(165)

In 2013, Alexandrov et al. analysed 4938362 mutations derived from 7042 cancers including breast cancer to determine whether particular mutation types were more frequently observed within differential cancer subtypes.(165) In 2020, Alexandrov et al. expanded this analysis and undertook somatic mutational analysis on 23829 cancer samples including 4645 set of whole genome somatic sequence data.(178) There are currently 49 biologically relevant signatures associated with Single Base Substitutions, 11 double base substitution signatures, 4 clustered base signatures and 17 small indel signatures.(178, 179) (180)

Mutational signature 3 is associated with DNA repair deficiency within the Homologous Recombination Repair pathways resulting in an excess of copy number variation within the tumour.(181) It is strongly associated with mutations in *BRCA1* and *BRCA2*.(165) Breast cancer occurring in association with an elevated somatic mutational burden and has been identified in association with signature 2 and occurs due to over activation of the APOBEC family of cytidine deaminases.(165) More recently, higher levels of *APOBEC*-related mutations have been demonstrated in HER2+ breast cancer.(181, 182)

Somatic mutational profiling also allows assessment of the relative proportions of each mutational signature within a tumour and provides an insight into the complex biology underlying tumorigenesis for any one cancer.(164, 165)

1.12 Prognosis in Association with High and Moderate Risk Genotypes

If germline genotype influences tumour biology, it also raises the question about whether it has an influence on patient outcome. In 2018, Copson et al. assessed the prospective survival amongst 338 early onset breast cancers occurring in association with a *BRCA* mutation compared to age matched controls within the POSH study. They found no significant difference in overall survival between *BRCA* mutation carriers and non-carriers.(36) There is limited prospective data detailing the association between other breast cancer susceptibility genotypes (*PALB2*, *CHEK2*, *ATM* and *TP53*) and prognosis.

PALB2-associated tumours may display a more aggressive tumour phenotype with a higher proliferation index (Ki67) and grade at presentation.(46, 53) Cybulski et al have provided the largest study looking at prospective outcomes in this group of patients. They found the crude 10

year survival to be significantly lower in *PALB2* carriers compared to normal population controls (48.0% (95%CI, 36.5-63.2) p<0.0001) versus (74.7% (95%CI, 73.5-75.8)).(66) This survival disparity was also present and significant at 5 years but not to the same magnitude.(66) The adjusted hazard ratio for death after correction for age at diagnosis, tumour size, nodal status, hormone receptor status and chemotherapy was 2.27 (95% C1 1.64-3.15) p<0.001.(66) The strongest predictor of an adverse outcome was a tumour size greater than 2cm at presentation.(66)

CHEK2 related breast cancer may also have an adverse prognostic phenotype. Schmidt et al. found that the risk of a second breast cancer was increased two-fold in patients with a *CHEK2* c.1100delC, p.(Thr367fs) germline mutation HR 2.1 (95% CI, 1.0 - 4.3 (p=0.049)).(183) They also had a worse recurrence-free survival (HR 1.7 (95% CI, 1.2-2.4) p=0.006) and breast cancer specific survival but this was not significant in the multivariable analysis.(183) Weischer et al found ER-positive *CHEK2* c.1100delC, p.(Thr367fs) heterozygotes had an increased risk of breast cancer specific death even after multi-variable analysis (HR 1.63 (95% CI, 1.24 to 2.15) p<0.001).(91)

Survival data is even more limited for *ATM* and *TP53*-associated breast cancer outcome. It has been suggested that the combination of specific rare missense variants and radiotherapy may have an adverse effect on the development of a second malignancy in the context of a germline *ATM* pathogenic variant.(184) Breast cancer occurring in the context of a pathogenic *TP53* variant is also associated with a high contralateral breast cancer risk.(106)

1.13 Research Question, Aims and Objectives

1.13.1 Research Question

The literature review has defined the current climate for genetic testing in the context of hereditary cancer susceptibility in the genomic era. It demonstrates the need for enhanced tools to identify carriers of monogenic factors in whom pedigree information may fail to meet conventional thresholds for genetic testing which are heavily weighted towards family history criteria. This includes reduced penetrance moderate susceptibility variants, de-novo variants and small, male dominant pedigrees. It also highlights the potential pitfalls related to the interpretation of VUS and the requirement for adjuncts to current methodologies for variant interpretation.

Tumour phenotype including tumour histopathology and somatic mutational profile has the potential to serve as an adjunct to current methodologies for the identification of actionable risk and interpretation of VUS but there is a need to build upon the currently available literature. This review has further demonstrated the paucity of prospective outcome data for non-*BRCA*

monogenic risk factors such as *PALB2*, *CHEK2*, *ATM* and *TP53* and the potential interplay between germline genotype, tumour biology and prognosis.

This research considers whether we can utilise tumour phenotype including histopathology and somatic mutational profile to identify individuals possessing a germline *BRCA1, BRCA2, PALB2*, *CHEK2*, *ATM* or *TP53* mutation. It will also consider whether this information can be further used to define the pathogenicity of VUS in *PALB2*, *CHEK2*, *ATM* and *TP53*, a question that will be increasingly important as we progress into the era of genomic testing for hereditary cancer susceptibility. This research will further consider whether germline genotype can influence patient prognosis.

1.13.2 Study Aims

1. The primary aim of this study is to provide a comprehensive overview of the histopathological tumour phenotype of *BRCA1, BRCA2, PALB2*, *CHEK2*, *ATM* and *TP53* mutation carriers and to determine which features (if any) best distinguish mutation carriers from non-carriers.
2. The secondary aim will be to determine whether germline *PALB2*, *CHEK2*, *ATM* and *TP53* genotype differentially influences breast cancer outcome.
3. The tertiary aim of this study is to provide an overview of the somatic mutation phenotype of *BRCA1, BRCA2, PALB2*, *CHEK2*, *ATM* and *TP53* mutation carriers and to determine which features (if any) best distinguish mutation carriers from non-carriers.

1.13.3 Study Objectives

The research objectives to achieve the aims of this study will include:

1. Manual curation of genomic data held within VCF files from the POSH study and application of ACMG guidelines to identify individuals with pathogenic or likely pathogenic germline variants in the cancer susceptibility genes *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53*.
2. Application of correlative statistics to compare tumour histopathological phenotype between gene carriers and non-carriers within the POSH study.
3. Survival analysis to compare outcome including Overall Survival and Distant Disease Free Survival between gene carriers and non-carriers within the POSH study.
4. Development of a bioinformatics pipeline to identify individuals with germline variants in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53* within The 100,000 Genomes Project. This

will include individuals recruited to the Rare Disease, Familial Breast Cancer Recruitment Domain and Cancer, Breast Cancer Recruitment Domain.

5. Development of a bioinformatics pipeline to identify individuals with somatic variants in target genes defined by the Cancer Gene Census within the Cancer, Breast Cancer Recruitment Domain of The 100,000 Genomes Project.
6. Evaluation of Tumour Mutational Burden within the Cancer, Breast Cancer Recruitment Domain of The 100,000 Genomes Project and application of correlative statistics to compare gene carriers with non-carriers.

1.13.4 Hypothesis

We hypothesise that germline mutations in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53* will be associated with characteristic tumour phenotypes predictive of carrier status. These genotypes may have differential effects on survival which are further modifiable by polygenic factors.

Chapter 2 Materials and Methods

2.1 Prospective Outcomes in Sporadic versus Hereditary Breast Cancer (POSH) Cohort

2.1.1 Sample Population

The sample population has been obtained from the Prospective Outcomes in Sporadic Versus Hereditary Breast Cancer (POSH) Study (UKCRN ID: 1137).(13, 14) POSH was a large multi-centre prospective cohort study which recruited 3095 women from 127 UK hospitals between 1st Jan 2000 and 31st January 2008.(13, 14) This study was designed to evaluate which factors influence prognosis and treatment response in women diagnosed with a primary invasive breast cancer under the age of 40 years and details of the study protocol are provided in Appendix C.(13, 14).

Inclusion required a diagnosis of primary invasive breast cancer at the age of 40 years or younger. Exclusion criteria included a previous cancer diagnosis with the exception of non-melanomatous skin cancer.(13, 14) Ethical approval was approved from the South and West Multi-Centre Research Ethics Committee (MREC 00/6/69). In total, 3021 of the 3095 participants recruited into the POSH study were eligible for further analysis.

2.1.2 Next Generation Sequencing

The POSH study obtained genomic DNA from whole blood lymphocytes. A customised gene panel was then created and the targeted genes, including *BRCA1, BRCA2, PALB2*, *CHEK2*, *ATM* and *TP53* were captured using an amplicon based library preparation system, the Fluidigm Access Array™.(185) NGS was performed using an Illumina platform and targeted the exonic regions and the exon/intron boundaries.(185, 186) Samples failed NGS either due to inferior quality or low concentration DNA.

The analysis population consisted of 2744 participants. In total, 277 (9%) were excluded due to missing genotyping data (n=159), M1 stage disease (n=74), age 41-50 years (n=42) or missing primary tumour data (n=2) (Figure 10).

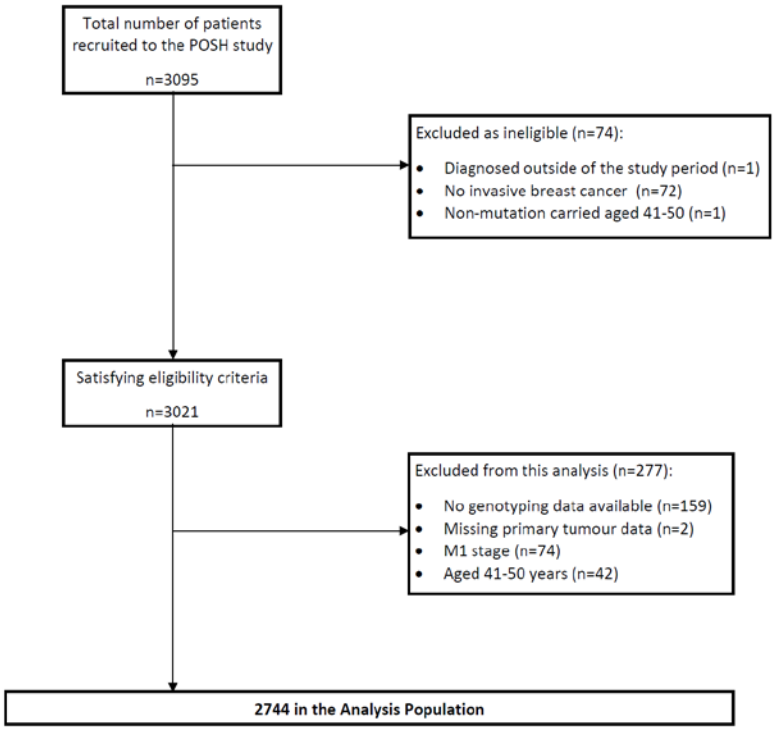


Figure 10: Sample Population

In total, 3021 of the 3095 participants recruited into the POSH study were eligible for further analysis. Of these, 2744 (91%) were included in the analysis population and 277 (9%) were excluded.

2.1.3 Bioinformatics Pipeline

2.1.3.1 Alignment and Annotation

The POSH study utilised two bioinformatics pipelines, Southampton and Cambridge for the identification of pathogenic and likely pathogenic variants in *BRCA1, BRCA2, PALB2*, *CHEK2*, *ATM* and *TP53*. Operating instructions for these pipelines are provided in Appendix D1.1 and Appendix D1.2.

As part of the POSH study bioinformatics analysis, raw NGS sequence data was aligned to the reference human genome (build GRCh37) using the Burrows Wheeler Aligner, BWA-MEM and this was stored in Binary Alignment/Map (BAM) file format.(187, 188) GATK was used for base quality recalibration and indel realignment. To ensure the accuracy of variant identification, reads with low mapping quality scores (less than phred 20), unmapped reads, failed primary alignments and reads failing platform or vendor quality checks were removed. Duplicate reads were kept because the amplicon based sequencing method generates legitimate duplicates.

SAMtools and GATK Unified Genotyper were used to identify variants and create a Variant Call File (VCF).(189-192) The Southampton bioinformatics pipeline used a combination of both SAMtools and GATK Unified Genotyper in a single sample analysis whilst the Cambridge pipeline used GATK Unified Genotyper in a multiple sample analysis. These pipelines have previously been validated and shown to have differential sensitivity and specificity (Table 12). SAMtools uses a probabilistic method for variant calling and computes the likelihood of a variant given each possible genotype.(189, 190). In general, this increases sensitivity and decreases specificity.(189, 190) GATK Unified Genotyper uses a Bayesian framework and error correction model based upon expected characteristics of human variation.(191, 192) It calculates a posterior probability for each genotype based upon prior probabilities and observed base quality resulting in high sensitivity and specificity.(193)

Pipeline	Sensitivity	Specificity
Southampton	91.0%	96.4%
Cambridge	83.7%	99.8%

Table 12: Sensitivity and Specificity of Bioinformatics Pipelines

Southampton which used single sample analysis had greater sensitivity and lower specificity whilst Cambridge which utilised multiple sample analysis has greater specificity but lower sensitivity.

Annotation of variants was performed using ANNOVAR.(194) This provided variant frequency and pathogenicity information derived from several databases of known population and disease causing variation including The 1000 Genomes Project, Database of Single Nucleotide Polymorphisms (dbSNP), Exome Sequencing Project (ESP), Exome Aggregation Consortium (ExAC) and Catalogue of Somatic Mutations in Cancer (COSMIC).A combination of these databases is used due to variation in the constituent data source, quality and associated strengths and limitations.

Variants in the VCF file were also annotated with respect to location and effect on coding sequence and ANNOVAR was also used to cross reference against predictors of functional significance including SIFT, PolyPhen2, LRT, PhyloP and AlignGVD. These use different algorithms to determine whether a variant is likely to have an effect on protein structure and/or function .(195)

2.1.3.2 Filtering

One of the key objectives of this work was to manually curate the annotated VCFs derived from the POSH study to identify pathogenic and likely pathogenic variants in the genes of interest (*BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53*).

The annotated VCFs were curated to list all called variants from 2744 participants in Human Genome Variation Society (HGVS) format according to the canonical transcript. Variants were manually reviewed and compared with assigned pathogenicity in ClinVar, CanVAR and The International Agency for Research on Cancer (IARC) databases. Variants with unreported pathogenicity including protein truncating variants and variants with a minor allele frequency (MAF) of less than 1% were classified using the American College of Medical Geneticists (ACMG) guidelines.(196-198)

The ACMG guidelines assign pathogenicity based upon several factors including the presence or absence within databases of population variation including gnomAD, case-control data, segregation analysis, family studies, functional data and *in-silico* predictions. A 5 tier classification system is used to categorise variants (Class 1 benign, Class 2 likely benign, Class 3 uncertain, Class 4 likely pathogenic and Class 5 pathogenic).(196) In the context of cancer susceptibility, only variants with a 95% or greater probability of being pathogenic are assigned as Class 4.

Variants classified as pathogenic or likely to be pathogenic (ACMG Class 4 and 5) were treated as mutation positive for the purposes of this analysis.(196) Class 1, Class 2 and Class 3 variants were defined as mutation negative.(196) Hypomorphic alleles were also considered mutation negative. In this study, any variant which conveyed less than a two-fold increase in breast cancer risk from the basal population level based upon case-control data was deemed hypomorphic.(75)

2.1.3.3 Validation

A further requirement was to validate filtered variant calls. In total, 1322 patients also underwent germline *BRCA1* and *BRCA2* testing either through collaborative research or NHS diagnostic testing. Variants identified by bioinformatics analysis were compared with those identified in other NHS diagnostic laboratories and research institutions for validation. In the absence of a prior genetic result, all pathogenic variants in *BRCA1* or *BRCA2* with a read depth less than or equal to 30 were sent to the Wessex Regional Genetics Laboratory for confirmatory testing. Confirmatory testing was completed by Sanger Sequencing.

Confirmatory testing of non-*BRCA* variants was completed within the department of Human Development and Heath (HDH) at the University of Southampton. The flanking genomic sequence within 51bp from the region of interest was provided for each variant of interest. Primer design, PCR amplification and gel extraction was completed within the department of HDH. Amplified genomic DNA was submitted to GATC, Eurofins Genomics for sequencing.(199) Sanger Sequencing data provided by the HDH team was reviewed and further analysis of the original variant call was conducted in IGV if required (Appendix E).

2.1.4 Clinical Data

Diagnostic histopathology reports were obtained as part of the POSH study for all patients including tumour size, stage, grade, focality ER, PR and HER2 receptor status. Additional immunohistochemistry on tissue microarrays derived from the proband was performed by the POSH study team to confirm the primary findings and supplement missing hormone receptor status.(13) BOADICEA without pathology adjustment was calculated using family history data. Prospective follow up data was obtained from patient records at 6 months, 12 months and then on an annual basis. This identified the date of recurrence, date of death and those participants lost to follow up.(36) At the time of this analysis, the median duration of follow up was 8.2 years (IQR 6.0-9.9 years).(36)

2.1.5 Statistical Analysis

The Statistical Analysis Plan (SAP) is provided in Appendix F. Correlative statistics were used to describe the cohort. Where appropriate, the Mann-Whitney test was used for continuous variables and the Pearson $\chi 2$ test for categorical variables to identify any specific differences between mutation carriers and non-carriers. Correlative statistics were performed using RStudio version 3.6.0 and a script is detailed in Appendix D.1.

A model selection process using multiple logistic regression and incorporating forward selection by way of likelihood ratio tests was utilised to determine which histopathological features were most predictive of germline mutational status. Odds ratios were used to further define these relationships and assign significance. Multiple imputation was incorporated into the model when required.

Logistic regression analysis was employed to identify the histopathological phenotypic characteristics Kaplan Meier curves were used to demonstrate Overall Survival (OS) and Distant Disease Free Survival (DDFS). Differential survival between *PALB2*, *CHEK2*, *ATM* and *TP53* gene carriers and non-carriers was compared using a univariable Cox regression model. Multivariable analyses (MVA) was also performed using Cox regression within the survival analysis.

2.2 The 100,000 Genomes Cohort

2.2.1 Sample Population

A second sample population has been derived from The 100,000 Genomes Project, a transformational project with the aim of sequencing the entire genome of 75,000 NHS patients

with rare diseases and cancer and an additional 25,000 tumour genomes.(200) Genomics England has approval from the HRA Committee East of England, Cambridge South (REC Reference 14/EE/1112).

Participants were recruited from 13 Genomic Medicine centres (GMCs) across England. Recruitment was broadly divided into the Rare Disease and Cancer Programme. Predefined inclusion and exclusion criteria were determined for each recruitment domain within these programmes following the study protocol (Appendix C).(201, 202) Cases have been derived from two domains within the 100,000 Genomes Project; Rare Disease, Familial Breast Cancer and Cancer, Breast Cancer.

Inclusion criteria for Rare Disease, Familial Breast Cancer was defined as a diagnosis of primary invasive breast cancer under the age of 50 years with 3 affected family members including first, second or third degree relatives with an average age at diagnosis below 60 years or a Manchester Score greater than or equal to 22. Participants were required to have NHS level diagnostic *BRCA1* and *BRCA2* genetic testing prior to recruitment.(202)

Inclusion criteria for Cancer, Breast Cancer was defined as a primary invasive breast cancer at any age if high quality DNA from both tumour and germline samples was available for analysis. Tumour material suitable for DNA extraction included Fresh Frozen Tissue (FFT) and optimised Formalin Fixed Paraffin Embedded (FFPE) tissue. This cohort represents unselected breast cancer.

In total 921 individuals were recruited into Rare Disease, Familial Breast Cancer and 3375 individuals were recruited into Cancer, Breast Cancer. The analysis population consisted of 826 participants within the Rare Disease, Familial Breast Cancer cohort and 2464 participants from within the Cancer, Breast Cancer cohort. In total, 95 (10.3%) were excluded from Familial Breast Cancer cohort and 911 (27.0%) were excluded from Sporadic Breast Cancer cohort the due to missing primary genome sequencing data. Subgroup analysis for Somatic Mutational Profile including both Tumour Mutational Burden and Somatic Mutational Signatures was performed on 1342 samples derived from the Cancer, Breast Cancer recruitment domain. (Figure 11).

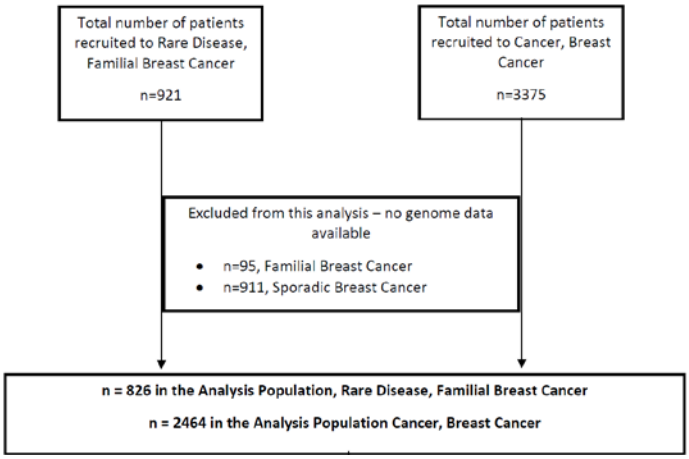


Figure 11: Sample Population

2.2.2 Next Generation Sequencing

Genomics England extracted DNA for germline analysis from whole blood lymphocytes. Whole genome sequencing was then performed on an Illumina platform using a Polymerase Chain Reaction (PCR) free methodology.(186) Germline sequencing aimed to achieve coverage of 95% of the autosomal human genome (build GRCh37 and build GRCh38) at the defined quality metrics of at least 15 independent observations with a read depth greater than Phred 30 and Mapability of greater than mapQ20.(201) Encrypted raw sequence data was transferred to the Genomics England Data Centre for access in the secure Research Embassy Environment.

DNA for somatic mutational analysis was derived from either Formalin Fixed Paraffin Embedded (FFPE) specimens or Fresh Frozen Tissue (FFT). The 100,000 Genomes protocol specified that a sample size of 60 microns was required for DNA extraction.(201) Tumour material fixed in formalin was to be processed within 24 hours to avoid DNA crosslinking and fragmentation. It was digitally optimised to ensure cellularity of the primary invasive tumour.(201) Whole genome sequencing was performed on an Illumina platform using a Polymerase Chain Reaction (PCR) free methodology and sequenced to a target depth of 75x.(186) Sequencing data was transferred to the Genomics England Data Centre for access in the secure Research Embassy Environment.(201)

2.2.3 Bioinformatics Pipeline

Genomics England aligned the raw sequence data to the reference human genome build GRCh37 (hg19) or build GRCh38 utilising external providers. This was supplied in the standard BAM file format.(201) Annotation was performed under the annotation pipeline of The 100,000 Genomes Project through the utilisation of external providers.(201) This included comparison with

databases of known population and disease causing variation including The 1000 Genomes Project.(201)

The annotated VCF files were analysed using Command Line on a Linux Terminal within the secure Genomics England Research Environment to identify germline and somatic variants of interest. Command Line was utilised to create file paths to access the genomic data for each of the eligible participants within the Rare Disease, Familial Breast Cancer and Cancer, Breast Cancer recruitment cohorts.

2.2.4 Bioinformatics Analysis: Germline

The annotated VCF derived from the genomic data of participants within Rare Disease, Familial Breast Cancer and Cancer, Breast Cancer recruitment domains of the 100,000 Genomes Project were filtered to identify variants within the genes of interest (*BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53*) through a bioinformatics pipeline. Three bioinformatics scripts (two shell scripts and one bash script) were created to analyse the annotated VCF files and identify variants.

These scripts had differential ability to identify variants. Overall, a combination of the 3 scripts maximised variant identification. An outline of the command for each of these scripts and function is detailed in Table 13, Table 14 and Appendix D.2.1 and Appendix D2.2. The scripts were actioned on a Linux terminal utilising a command line interface within the secure 100,000 Genomes Project Research Environment.

Filtered variants were listed with reference to their genomic coordinates and dbSNP accession number. Variants were manually curated in Libre Calc with conversion of the called variants from genomic coordinates or dbSNP accession number into standard HGVS nomenclature. Pathogenicity was assigned to the filtered variants using the standards defined in Chapter 2.1.2.1 Variants were visualised in the Integrative Genomics Viewer (IGV) for validation.(203)

Script Type	Command
Shell	cat CancerPatientPathsOctober2018_38 | while read pathToVCF build; do (echo $pathToVCF | cat; echo $build ; awk '(204)' <(gzip -dc $pathToVCF)) >> *PALB2*_build38.txt (echo $pathToVCF | cat; echo $build ; awk '{if($1=="chr22" && $2>28687743 && $2<28742422 && $8 ~ /*CHEK2*/ && $6 !~ 0 && $8 !~ /regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/) print $0}' <(gzip -dc $pathToVCF)) >> *CHEK2*_build38.txt (echo $pathToVCF | cat; echo $build ; awk '{if($1=="chr11" && $2>108222484 && $2<108369102 && $8 ~ /*ATM*/ && $6 !~ 0 && $8 !~ /regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/) print $0}' <(gzip -dc $pathToVCF)) >> *ATM*_build38.txt (echo $pathToVCF | cat; echo $build ; awk '{if($1=="chr17" && $2>7661779 && $2<7687550 && $8 ~ /*TP53*/ && $6 !~ 0 && $8 !~ /regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/) print $0}' <(gzip -dc $pathToVCF)) >> *TP53*_build38.txt done
Bash	#module load bcftools/1.9 while read -r vcf; do tabix -h $vcf -R regions38.txt | \ bcftools norm -m -any | \ bcftools view -f PASS -i 'MIN(FMT/DP)>10 & MIN(FMT/GQ)>15' | \ bcftools query -f '[%SAMPLE]\t%CHROM\t%POS\t%ID\t%REF\t%ALT\t%QUAL\t%FILTER\t[%GT]\t[%GQ]\t[%DP]\t%INFO/CSQT\n' | \ grep -f so_terms.txt >> results_c38.txt ; done < vcflist_c38 sed -i '1s/^/SAMPLE\tCHROM\tPOS\tID\tREF\tALT\tQUAL\tFILTER\tGT\tGQ\tDP\tCSQT\n/' results_c38.txt

Table 13: Summary of the Script Command for Germline Analysis

The scripts used for germline analysis linked variant data with the associated participant and identified variants in annotated VCFs aligned to build GRCh37 (hg19) or build GRCh38. Scripts produced in conjunction with Dr William Tapper.

Script Name	Script Type	Target Variant
Original Script	Shell	Designed to identify coding variants within *CHEK2*, *PALB2*, *ATM* and *TP53*. Filters applied within this script to filter out regulatory region variants, downstream gene variants and intronic variants.
Founder Script	Shell	Designed to identify Founder and specific mutations within *CHEK2*, *PALB2* and *ATM* (*CHEK2* c.1100delC, *PALB2* c.3113G>A, *ATM* c.7271T>G)
Genomics England	Bash	Coding variants within *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53*

Table 14: Summary of the Script Function for Data Analysis within The 100,000 Genomes Project

The Original Script was a shell script designed to identify all coding variants within *CHEK2*, *PALB2*, *ATM* and *TP53*. The Founder Script was a shell script used to identify recognised founder mutations and specific mutations within *CHEK2*, *PALB2* and *ATM*. The Genomics England script was a bash script modified from a Genomics England base script and used to identify pathogenic and likely pathogenic variants within *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *ATM* and *TP53*.

2.2.5 Somatic Mutational Analysis

2.2.5.1 Somatic Mutational Profile

The Catalogue of Somatic Mutations in Cancer (COSMIC), Cancer Gene Census was accessed to export a list of genes with recognised somatic mutations in breast cancer to create a candidate gene list of 39 significant, Tier 1, somatically mutated genes.(205) Tier 1 variants have proven activity relevant to cancer evolution. Several Tier 1 genes are associated with cancer Hallmark phenotypic effect. Hallmark genes confer fundamental biological capabilities which provide a survival advantage for clonal evolution. The candidate gene list is shown in Appendix G

A bash script was used to filter the annotated VCF files and identify somatic sequence variants within the genes of interest. A further script was used to remove variants with a low quality score and the output of this was placed into Mutation Annotation Format (MAF). MAF files are tab-delineated files that describe variants in a standardised format (sample ID, gene and variant protein effect).(204) The scripts required to identify somatic variants, filter and convert to MAF format were run on a Linux terminal utilising a command line interface within the secure 100,000 Genomes Project Research Environment. The script utilised is shown in Table 15 and Appendix D.2.3.

Command

```
#!/bin/bash
module load bcftools/1.9
while read -r vcf ; do
        echo $vcf >> results_somatic.txt
        tabix -h $vcf -R Genes.txt |
        grep -f so_terms.txt >> results_somatic.txt ;
done < VCFpathsSomatic_Cancer
sed -I 'ls/^/SAMPLE\tCHROM\tPOS\tID\tREF\tALT\tQUAL\tFILTER\tINFO\tFORMAT\n/' results_somatic.txt

awk '$7 !~ /BCNoiseIndel/ && $7 !~ /LowQscore/ && $7 !~ /QSI_ref/ && $7 !~ /RepetitiveRegion/'
results_somatic_NoMutation.txt > Qual_NoMutation.txt

awk -F "\t" '{if($0 ~ /^\//) {split($0,path,"/");} if($0 !~ /^\//) {ele=split($8,vinfo,"|"); for(i=0;i<ele;i++) {if(vinfo[i] ~ /ENST/ && (vinfo[i-
1] == "CHEK2" || vinfo[i-1] == "ATM"|| vinfo[i-1] ==  "ARID1A" || vinfo[i-1] == "GATA3" || vinfo[i-1] == "CCND1" || vinfo[i-1] ==
"CDKN1B" || vinfo[i-1] == "ETV6" || vinfo[i-1] == "SMARCD1" || vinfo[i-1] == "TBX3" || vinfo[i-1] == "BRCA2" || vinfo[i-1] == "RB1"
|| vinfo[i-1] == "AKT1" || vinfo[i-1] == "FOXA1" || vinfo[i-1] == "NTRK3" || vinfo[i-1] == "CDH1" || vinfo[i-1] == "CTCF" || vinfo[i-1]
== "PALB2" || vinfo[i-1] == "BRCA1" || vinfo[i-1] == "BRIP1" || vinfo[i-1] == "ERBB2" || vinfo[i-1] == "MAP2K4" || vinfo[i-1] ==
"NCOR1" || vinfo[i-1] == "PPM1D" || vinfo[i-1] == "TP53" || vinfo[i-1] == "KEAP1" || vinfo[i-1] == "BARD1" || vinfo[i-1] == "CASP8"
|| vinfo[i-1] == "SALL4" || vinfo[i-1] == "APOBEC3B" || vinfo[i-1] == "EP300" || vinfo[i-1] == "BAP1" || vinfo[i-1] == "MAP3K13" ||
vinfo[i-1] == "PBRM1" || vinfo[i-1] == "PIK3CA" || vinfo[i-1] == "POLQ" || vinfo[i-1] == "MAP3K1" || vinfo[i-1] == "ARID1B" ||
vinfo[i-1] == "ESR1" || vinfo[i-1] == "NOTCH1" || vinfo[i-1] == "IRS4")) print NR,"\t",path[8],"\t",vinfo[i-1],"\t",vinfo[i+1];}}}'
Qual_NoMutation.txt > formated_NoMutation.txt
```

Table 15: Script for Identifying Somatic Tier 1 Variants

Script produced in conjunction with Dr William Tapper.

LibreCalc was used to display the relative proportion of each variant type. Summary level data was exported from the Secure Research Environment using the Airlock file transfer application.

GenVisR was used within RStudio, version 3.4.4 to provide a graphical representation of the Somatic Mutational Profile. Samples were clustered based upon germline mutational status (No Mutation, *BRCA*, *PALB2*, *CHEK2*, *ATM*, *TP53*). GenVisR was used to create a hierarchical classification of all variants within the genes of interest based upon their recognised protein effect and plotted using the Waterfall function.(206) The RStudio Script used for production of the Waterfall Plots is shown in Appendix D.2.4.

2.2.5.2 Tumour Mutational Burden and Tumour Mutational Signature

The Tumour Mutational Burden in Mutations per Mb (TMB Mut/Mb) and the percentage prevalence of Single Base Substitution (SBS) Somatic Mutational Signatures 1-30 was available for 1342 participants within The 100,000 Genomes Project, Cancer, Breast Cancer cohort. This data was accessed through LabKey application within the secure research environment and annotated against gene carrier status established through germline analysis.

2.2.6 Clinical Data

Clinical phenotypic data including participant demographics, primary cancer diagnosis and age at onset was available through the Lab Key application within the secure Research Embassy Environment. Available, relevant clinical data could be linked to the genomic result using this resource. However, comprehensive tumour histopathology was unavailable for the majority of participants within the 100,000 Genomes Project at the time of this analysis.

2.2.7 Statistical Analysis

Statistical analysis was performed using RStudio version 3.4.3 within the 100,000 Genomes Project Secure Research Environment. Summary and correlative statistics were used to describe the cohort and compare age of onset, TMB and SBS Mutational Signatures between *BRCA*, *PALB2*, *CHEK2*, *ATM* AND *TP53* gene carriers and non-carriers. A two sided t-test was used to compare the age of onset between gene carriers and non-carriers across the familial breast cancer and unselected breast cancer cohorts. The Mann-Whitney test was used to identify any differences in TMB and SBS Mutational Signatures between gene carriers and non-carriers. The RStudio script used to complete this analysis is provided in Appendix D.2.5-D.2.7.

The application ggplot2 was used within RStudio version 3.4.3 to provide graphical representation of the data comparing age of onset, TMB and SBS Mutational Signature between gene carriers

and non-carriers. Operating instructions for this are also provided in Appendix D.2.5-D.2.7. Summary data and graphical representations were exported through the secure Airlock with the Research Environment after verification and independent review for anonymity.

.

Chapter 3 Results: Variant Detection

3.1 Pathogenic Variant Identification: A Comparison Across Cohorts

A pathogenic or likely pathogenic breast cancer susceptibility variant was identified in 453/2744 (16.5%) of all participants within the POSH cohort, (Table 16). This included *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *ATM* and *TP53*. In total, 6/453 (1.3%) of gene carriers within the POSH cohort were found to have multiple pathogenic variants which included a combination of moderate and high penetrance variants (Table 17). A summary of all pathogenic and likely pathogenic variants identified is provided in Appendix H.

A pathogenic or likely pathogenic breast cancer susceptibility variant was identified in 60/826 (7.3%) of all participants within The 100,000 Genomes Project, Familial Breast Cancer cohort. This sample was enriched for non-*BRCA* gene carriers and thus comprised *CHEK2*, *PALB2*, *ATM* and *TP53* gene carriers only.

The overall mutation detection rate for any pathogenic or likely pathogenic variant reduced to 50/2464 (2.0%) amongst unselected breast cancers derived from The 100,000 Genomes Project, Cancer, Breast Cancer recruitment domain.

Gene	Frequency 100,000 Genomes, Familial Breast Cancer (%) n=826	Frequency POSH Study, Breast Cancer Under 40 years (%) n=2744	Frequency 100,000 Genomes, Sporadic Breast Cancer n=2464
BRCA1+	-	201 (7.3%)	7 (0.3%)
BRCA2+	-	137 (5.0)	
CHEK2+	26 (3.1%)	53 (1.9)	20 (0.8%)
PALB2+	15 (1.8%)	31 (1.1)	8 (0.3%)
ATM+	11 (1.3%)	23 (0.8)	10 (0.4%)
TP53+	8 (1.0%)	15 (0.5)	5 (0.2%)
Total	**60 (7.3)**	**453 (16.7)***	**50 (2.0)**

Table 16: The Frequency and Percentage of Gene Carriers within the 100,000 Genomes Project, Familial Breast Cancer and Sporadic Breast Cancer

Summary statistics of all pathogenic and likely pathogenic variants identified within the genes *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *ATM* and *TP53* across all cohorts. *Within the POSH Study, 6 individual had multiple pathogenic variants. A total of 7 additional pathogenic/likely pathogenic variants were identified amongst 6 individuals.

3.1.1 *BRCA1* and *BRCA2* Variant Identification

Alterations in *BRCA1* and *BRCA2* represented the most prevalent pathogenic and likely pathogenic breast cancer susceptibility variants within the under 40 year (POSH) cohort. Overall, 338/2744 (12.3%) of individuals presenting with breast cancer under the age of 40 years had a class 5 or class 4 variant in either *BRCA1* or *BRCA2*. They represented 338/453 (74.6%) of all variant carriers within the POSH cohort. A pathogenic or likely pathogenic variant in *BRCA1* was identified in 201/2744 (7.3%) of participants and a pathogenic or likely pathogenic variant in *BRCA2* was identified in 137/2744 (5.0%) of participants. Of the 338 individuals with a *BRCA* mutation, 136/338 (40.2%) presented with a triple negative breast cancer and the majority of these, 123/136 (90.4%) were *BRCA1* related (Chapter 5, Table 25).

Figures were not available within the familial cohort of 100K to enable a direct comparison of *BRCA1* and *BRCA2* mutation prevalence. A class 4 or class 5 variant in *BRCA* was identifiable in only 7/2464 (0.3%) of unselected breast cancer cases within 100K.

3.1.2 *CHEK2*, *PALB2*, *ATM* and *TP53* Variant Identification

The highest proportion of non-*BRCA* variant carriers (*CHEK2*, *PALB2*, *ATM* and *TP53*) was present within the familial breast cancer group and the lowest proportion within unselected breast cancer. A pathogenic or likely pathogenic variant in the aforementioned genes was present in 60/826 (7.3%) of the familial breast cancer cohort, 122/2744 (4.4%) of the under 40 years cohort and 43/2464 (1.7%) of the unselected breast cancer cohort (Table 16, Figure 13 and Figure 14).

CHEK2 was the most prevalent single gene alteration within the sporadic breast cancer group. It was also the most common non-*BRCA* variant across all cohorts present in 23/826 (3.1%), 53/2744 (1.9%) and 20/2464 (0.8%) of the familial, under 40 years and sporadic breast cancer cohorts respectively. In comparison, a pathogenic or likely pathogenic variant in *PALB2* was identifiable in 15/826 (1.8%) of the familial cohort, 31/2744 (1.1%) of the under 40 cohort and 8/2464 (0.3%) of the unselected cohort. A pathogenic or likely pathogenic variant in *ATM* was identifiable in 11/826 (1.3%) of the familial cohort, 23/2744 (0.84%) of the under 40 cohort and 10/2464 (0.4%) of the unselected cohort (Table 16, Figure 13 and Figure 14).

Class 4 or class 5 variants in *TP53* were the least prevalent across all groups present in 8/826 (1.0%) of familial breast cancers, 15/2744 (0.5%) of under 40 breast cancer and 5/2464 (0.2%) of unselected breast cancers (Table 16, Figure 13 and Figure 14). Within the POSH study, 7/15 (46.67%) had protein truncating variants including the consensus splice and 8/15 (53.33%) had missense variants (Appendix H). Within The 100,000 Genomes Project unselected cohort, all

variants were missense variants. Within The 100,000 Genomes Project familial cohort, the majority of these were missense variants with one consensus splice site identifiable.

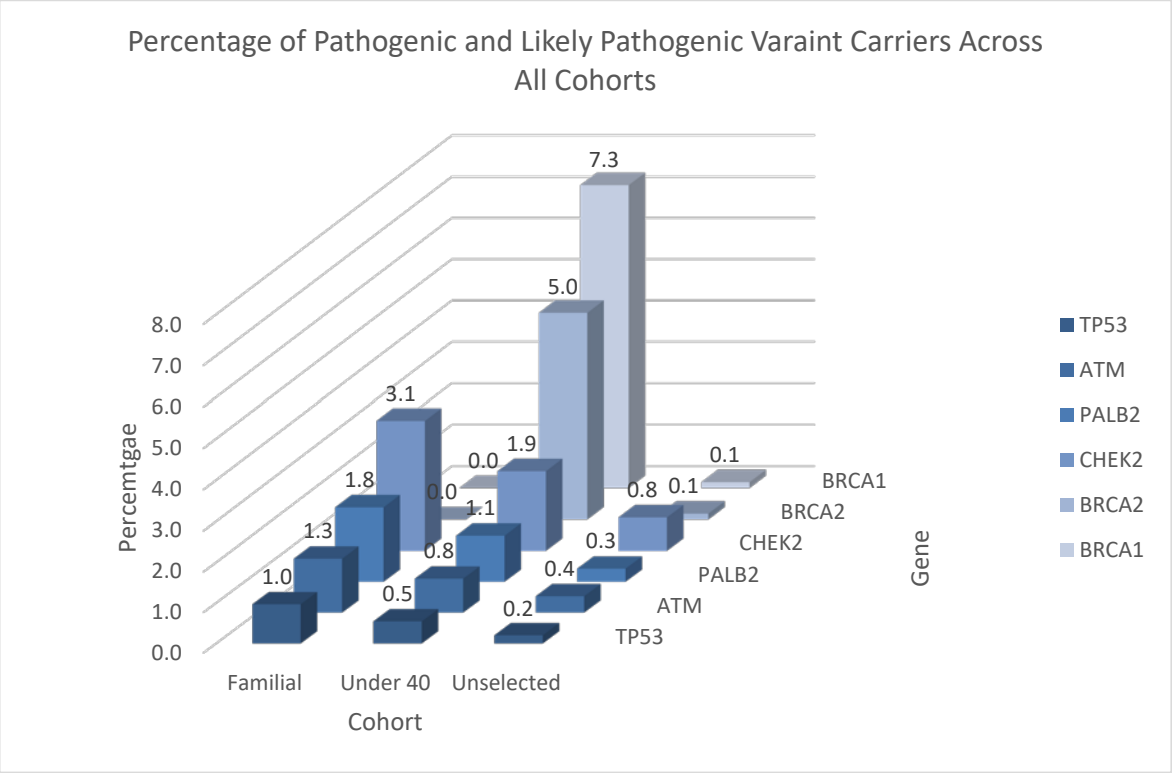


Figure 12: **Percentage of Pathogenic and Likely Pathogenic Variant Carriers Across All Cohorts**

Percentage of *CHEK2*, *PALB2*, *ATM* and *TP53* gene carriers across the breast cancer cohorts. Due to data restrictions the percentage of *BRCA1* and *BRCA2* variant carriers is shown collectively in this figure for Unselected Breast Cancer.

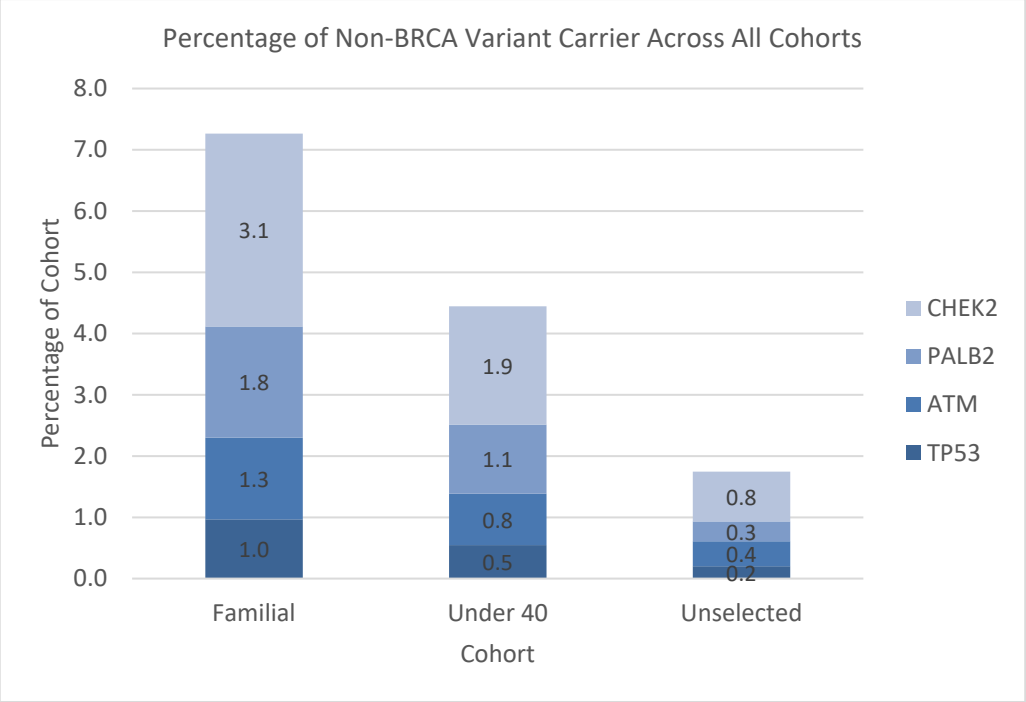


Figure 13: Percentage of Pathogenic and Likely Pathogenic Variant Carriers Across All Cohorts

Percentage of *CHEK2*, *PALB2*, *ATM* and *TP53* gene carriers across the breast cancer cohorts.

3.2 Founder Variant Identification

Overall, the founder mutation *CHEK2* c.1100delC, p.(Thr367fs) was the most frequently identified *CHEK2* variant across all cohorts and identified in 18/26 (69.2%) of the familial *CHEK2* variant carriers, 36/53 (67.9%) of the under 40 years *CHEK2* variant carriers and 16/20 (80.0%) of the sporadic *CHEK2* variant carriers. *PALB2* c. 3113G>A, p.(Trp1038Ter) was the most prevalent *PALB2* variant in those diagnosed with breast cancer under the age of 40 years (14/31 (45.2%)) compared to 3/15 (20.0%) of the familial breast cancer cases and 1/8 (12.5%) of the sporadic cases associated with *PALB2* pathogenic variants. The higher penetrance *ATM* c.7271T>G, p.(Val2424Gly) accounted for 4/11 (36.4%) of *ATM* variants in the familial breast cancer cohort, 5/23 (21.7%) within the under 40 years cohort and 3/10 (30%) of the *ATM* variants within the unselected cohort (Table 17 and Figure 14).

Gene	Proportion of Gene Carriers		
	100,000 Genomes, Familial Breast Cancer (%)	POSH Study, Breast Cancer Under 40 years (%)	100,000 Genomes, Sporadic Breast Cancer
CHEK2+ c.1100delC	18/26 (69.2%)	36/53 (67.9%)	16/20 (80%)
PALB2+ c.3113G>A	3/15 (20%)	14/31 (45.2%)	1/8 (12.5%)
ATM+ c.7271T>G	4/11 (36.4%)	5/23 (21.7%)	3/10 (30%)

Table 17: Founder Variant Identification: A Comparison Across Cohorts

Proportion of founder mutations in *CHEK2*, *PALB2* and *ATM* gene carriers across the breast cancer cohorts.

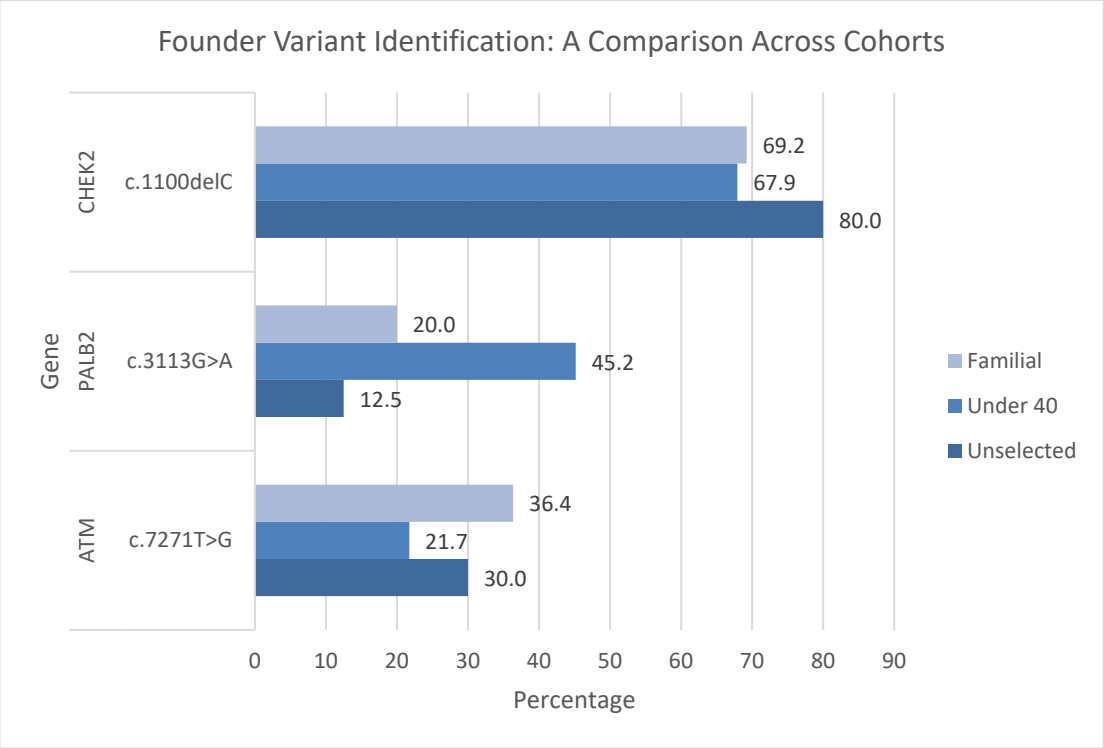


Figure 14: Founder Variant Identification: A Comparison Across Cohorts

3.3 Discussion

3.3.1 Variant Identification

This work represents a unique series detailing the prevalence of pathogenic and likely pathogenic variants in several high and moderate risk breast cancer susceptibility genes presented across three distinct breast cancer cohorts (familial, under 40 years and unselected breast cancers). A pathogenic or likely pathogenic breast cancer susceptibility variant was identified in 16.5% of individuals with early onset breast cancer compared to 2.0% amongst unselected breast cancers. A direct comparison was not possible for familial breast cancer as prior diagnostic *BRCA* testing was a pre-requisite for recruitment into The 100,000 Genomes Project, Familial Breast Cancer

recruitment domain. The identification of pathogenic variants in *BRCA1* and *BRCA2* amongst unselected breast cancers was lower in comparison to other unselected series.(34, 39, 40) It raises the possibility of selection bias. For example, individuals with screen detected breast cancer or those with a strong family history may have been recruited to the familial rather than unselected recruitment domain.

Within this series, pathogenic and likely pathogenic variants in *BRCA1* and *BRCA2* were the most common single gene alterations observed in the context of primary invasive breast cancer under the age of 40 years, present in 12.3% of cases. We assume the same to be true for familial breast cancer cases. This was not demonstrable within the current series due to the ascertainment of samples. However, the mutation detection of *BRCA* variants is consistent with Kraus et al. who observed the mutational detection frequency amongst 581 consecutive individuals with familial breast and or ovarian cancer following a 14 gene breast and ovarian cancer susceptibility panel.(37) In total 72/581 (12.4%) of the cohort carried a mutation in *BRCA1* or *BRCA2*.(37)

Within this study, the additional testing of *CHEK2*, *PALB2*, *ATM* and *TP53* produced a diagnostic uplift of 1.7%-7.3% depending upon the tested cohort. The prevalence of non-*BRCA* gene alterations was highest in the context of familial breast cancer and lowest in the context of sporadic breast cancer. Pathogenic and likely pathogenic variants in *CHEK2* were the most frequently identified non-*BRCA* variants across all cohorts occurring in 3.1%, 1.9% and 0.8% of familial, under 40 and sporadic breast cancer cases respectively. Pathogenic variants in *TP53* were the least prevalent occurring in 1.0%, 0.5% and 0.2% of the aforementioned cohorts. The identification of *TP53* variant carriers was lower within the under 40 (POSH cohort) compared to other studies of early onset breast cancer.(116, 117, 121) This may be attributable to the exclusion of individuals with isolated DCIS and previous malignancy.

A number of studies have described non-*BRCA* variant detection in the context of familial breast cancer. In 2017, Moran et al. evaluated the detection of non-*BRCA* genes amongst 190 familial breast cancer patients.(118) In total, 9/190 (4.7%) had pathogenic variants in *PALB2*, *CHEK2*, *ATM* and *TP53*.(118) *CHEK2* was the most prevalent variant identifiable in 5/190 (2.6%) of all breast cancer cases.(118) Pathogenic variants in *TP53* were identifiable in 1/190 (0.5%) of cases.(118)

In 2017, Kraus et al. observed the diagnostic uplift of testing *CHEK2*, *PALB2*, *ATM* and *TP53* as part of a multi-gene panel within a cohort fulfilling diagnostic *BRCA* testing criteria.(37) Overall, the diagnostic uplift of testing these genes was 3.5%. Pathogenic variants in *CHEK2* were most prevalent accounting for 10/581 (1.7%) of breast cancer cases.(37) Conversely, pathogenic variants in *TP53* were least prevalent and present in 2/581 (0.3%).(37) In total, 6/581 (1.0%) had a pathogenic variant in *PALB2* and 3/581 (0.5%) had a pathogenic variant in *ATM*.(37)

In 2017, Couch et al. analysed a population of 58798 consecutive patients referred for diagnostic genetic testing.(207) A number of different cancer panels were used for diagnostic purposes.(207) Pathogenic variants in *CHEK2* remained the most frequently identified non-*BRCA* gene present in 624/44220 (1.41%) of tested cases.(207) Pathogenic variants in *PALB2* were present in 416/45513 (0.91%) of cases.(207) Pathogenic variants in *ATM* were present in 446/44176 (1.01%) of cases and pathogenic variants in *TP53* were present in 91/58788 (0.15%) of cases.(207) This approximated to a diagnostic uplift of 3.7% for these non-*BRCA* variants.(207)

There is limited data regarding mutation detection of non-*BRCA* genes amongst early onset and unselected breast cancer cohorts. In 2017 Buys et al. reported on the mutational detection frequency following a 25 gene panel in an unselected cohort of 35,409 women with a single breast cancer diagnosis.(119) The diagnostic uplift of testing *CHEK2*, *PALB2*, *ATM* and *TP53* was 1103/35409 (3.1%).(119) Pathogenic variants in *CHEK2* were the most frequently identified present in 397/35409 (1.1%) of cases.(119) Pathogenic variants in *TP53* were the least prevalent present in 61/35409 (0.17%) of cases.(119)

Subgroup analysis observing the prevalence of founder mutations identified that *CHEK2* c.1100delC, p.(Thr367fs) was the most frequently identified *CHEK2* variant. This is consistent with much of the published literature. A population based study within the UK involving 13087 breast cancer cases found that c.1100delC, p.(Thr367fs) was the prevalent rare truncating *CHEK2* variant, present in 81% (196/242) of *CHEK2*-associated breast cancer cases.(56)

PALB2, c.3113G>A, p.(Trp1038Ter) had the highest percentage prevalence amongst those with early onset breast cancer within the POSH study. Its percentage prevalence was much lower in the other tested cohorts. Based upon this over-representation, it raises the question of whether *PALB2* c.3113G>A, p.(Trp1038Ter) is a particularly high penetrance variant. In 2010, Southey et al. reported on the breast cancer risks associated with this specific variant in an Australian cohort of 1403 probands enriched for early onset breast cancer irrespective of family history.(208) The Hazard Ratio for breast cancer risk was 30.1(95%CI, 7.5-120 ($p<0.0001$)).(208) This equated to an absolute lifetime risk of 91% (95% CI, 44-100) to age 70 years.(208) Although not directly comparable, this is higher than the cumulative life time risk of pathogenic variants in *PALB2* evaluated collectively demonstrated in the largest cohort of 524 families with pathogenic variants in this gene by Yang et al. in 2020.(20) The relative risk of breast cancer observed within this cohort was RR 7.18 (95% CI, 5.82-8.85 ($p=6.5x10^{-76}$)) with an absolute cumulative lifetime risk of 53% (95% CI, 44%-63%). (20) The prevalence of *PALB2* c.3113G>A, p.(Trp1038Ter), was also lower affecting 61/524 (11.4%) families.(20) However, it was still the most frequently identified pathogenic variant in *PALB2*.(20)

PALB2 c.3113G>A, p.(Trp1038Ter) is located within the β-propeller structure of the WD40 domain of the gene which incorporates a Nuclear Export Signal (NES).(46) In 2017, Pauty et al. observed that *PALB2* c.3113G>A, p.(Trp1038Ter) was associated with a significant increase in cytoplasmic localisation of the protein and reduced interaction with *BRCA2* and RAD51.(48) The other protein truncating variants displayed predominant nuclear localisation of the protein.(48) It suggests that pathogenic variants within the WD40 domain may have a differential effect on *PALB2* function. There are currently 17 instances of *PALB2* c.3113G>A, p.(Trp1038Ter) within the gnomAD database including 14 counts within the European (non-Finnish) population suggesting a founder effect.(209)

Subgroup analysis also demonstrated that *ATM* c.7271T>G, p.(Val2424Gly) was the most frequently identifiable pathogenic variant in this gene across all cohorts present in 20%-30% of *ATM* pathogenic variant carriers. Whilst pathogenic variants in *ATM* are generally associated with a moderately increased risk of breast cancer, *ATM* c.7271T>G, p.(Val2424Gly) is considered a high risk variant. In 2016, the COGS study determined the associated risk as OR 11.0 (95%CI, 1.42-85.7 (p=0.0012)).(57) Goldgar et al. also found this variant conferred an 8 fold increased risk in breast cancer amongst 2,570 breast cancer cases and 1448 controls.(27) Pathogenic variants in *ATM* are identifiable on average in less than 1% of breast cancer cases. They can also be more challenging to interpret as this is a large gene with multiple exons. Given the proportionate identification of this specific high risk variant, it may represent a candidate for targeted testing within a broader breast cancer susceptibility gene panel.

It is also important to consider that 1.3% of individuals presenting with symptomatic early onset breast cancer had multiple pathogenic variants. Buys et al. also reported that multiple pathogenic variants were identified in 1.3-3.3% of cases.(119) Within this study, each of the identified variants had a clinical actionability which was independent of the other variant. It highlights that for a small number of families two or even three high and moderate penetrance genetic factors may contribute to overall risk for specific individuals. This could be identified through breast cancer susceptibility gene panel testing amongst individuals presenting for genetic testing.

3.3.2 Gene Panel Testing for Hereditary Breast Cancer Susceptibility

Breast cancer associated mortality has reduced by 39% in the last four decades.(1) Advances in the treatment of breast cancer and improved detection have contributed to the observed improvement in survival metrics. Despite this, breast cancer remains a chronic disease for many individuals with the potential for late relapse and the associated short and long term health economic implications.(210) As such, primary prevention is increasingly considered as a valuable

tool in managing breast cancer risk to provide health economic benefit at the population level.(211)

There has been increasing interest in the use of genotypic information for precision prevention in the context of hereditary breast cancer.(211) More specifically, identifying those who are at the greatest genetic risk of breast cancer and may benefit from screening, chemoprevention and risk reducing intervention such as bilateral risk reducing mastectomy.(211)

A small number of studies have now shown that gene panel testing for hereditary breast cancer susceptibility can produce a cost effective improvement in Quality Adjusted Life Years (QALY) and Life Expectancy through the identification of actionable risk. (212, 213) In 2018, Manchanda et al. compared the lifetime cost and effect of testing all non-Jewish women aged over 30 years for *BRCA1*, *BRCA2*, *PALB2* and other ovarian susceptibility genes.(213) They demonstrated that both family history and population based screening were cost effective. The family history based Incremental Cost Effectiveness Ratio (ICER) was £7629.65/QALY versus a population based ICER of £21 599.96/QALY. (213) In 2017, Li et al. produced hypothetical modelling based upon the impact of gene panel testing on life expectancy and QALY. They identified that a gene panel test including *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *CDH1*, *STK11* and *PALB2* cost $48,328 per QALY in the under 40 year old group.(212)

In this study, the overall detection of pathogenic and likely pathogenic variants in *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *ATM* and *TP53* was greater than 10% in the context of early onset breast cancer. We assume the same to be true for familial breast cancer. This threshold of 10% is advocated by NICE as a reasonable parameter to initiate diagnostic genetic testing. As such, the mutational detection rate when testing women with early onset or familial breast cancer is likely to produce a health economic benefit for the detection of high risk genes in QALY secondary to the identification of actionable genetic risk as demonstrated in the context of early onset and familial breast cancer. (212, 213)

However, it is important to recognise that despite the potential health economic benefit of diagnostic genetic testing reported by Manchanda and Li et al., neither of these studies included the cost of VUS interpretation and the potential cost of misinterpretation and inappropriate medical intervention which may be cascaded across a family in their statistical modelling structure. (212, 213) This is particularly relevant as genomic technology is mainstreamed across medical specialities with less experience in variant interpretation. In addition, there have been no significant appraisals of the clinical utility or cost effectiveness of testing moderate risk genes which are often considered equivocal in terms of their ability to alter risk stratification in the

context of a pre-existing family history of breast cancer. This remains an area for future evaluation particularly with increasing utilisation of gene panel testing for heritable cancer susceptibility.

Overall, this demonstrates that whilst there is potential heath economic benefit for genetic testing for hereditary breast cancer, the potential negative effects of variant misinterpretation and inappropriate medical intervention must also be considered. It highlights the need for robust systems for centralised reporting and the curation of cancer susceptibility variants according to ACMG guidelines. It also demonstrates the potential utility of a multi- tier approach to cancer variant reporting in which results are issued with an interpretation of pathogenicity and clinical actionability to mitigate the potential for misinterpretation.(214) Adjuncts for variant interpretation such as tumour histopathology and somatic mutational profiling may also assist in the interpretation of isolated cancer susceptibility variants within a family.

3.4 Summary

As we progress into the genomic era, gene panel testing has the potential to be used increasingly to identify heritable risk with potential health economic benefit.(211) Overall we have shown that 16.7% of individuals with symptomatic early onset breast cancer will have a moderate of high penetrance gene variant compared to 2% of unselected breast cancers.

The outcome of this analysis supports the utilisation of gene panel testing to identify actionable risk within familial and early onset breast cancer cohorts but not amongst unselected breast cancers. Genetic testing in this context may produce a cost benefit for the identification of actionable risk amongst high penetrance gene carriers but further evaluation is required to determine whether the same is true for moderate penetrance genotypes.

Despite the potential benefits of genetic testing, we must approach this technology with care particularly in relation to variant interpretation and ensuring appropriate medical intervention for the associated risk. Future advances in centralised variant databases and curation along with multi-tier reporting have the potential to mitigate some of these risk and truly harness the potential of genomic technology for heritable cancer susceptibility.

Chapter 4 Results: Age of Onset

4.1 POSH Cohort: Primary Invasive Breast Cancer Under 40 Years

The complete analysed cohort consisted of 2744 participants who were diagnosed with a primary invasive breast cancer under the age of 40 years. Breast cancer onset occurred at a significantly younger age in gene carriers compared to non-carriers. The median age at breast cancer diagnosis was 36.0 years (IQR, 33.0-38.0 years) in gene carriers versus 37.0 years (IQR 34.0-39.0 years) for non-carriers (p<0.0001) (Table 18, Figure 15 and Figure 16).

Sub-group analysis demonstrates that the genes associated with a significantly younger age at diagnosis were *BRCA1* and *TP53*. The median age at breast cancer diagnosis was 35.0 years (IQR 32.0-38.0) for *BRCA1* gene carriers and 33.0 years (28.5-34.5) for *TP53* gene carriers (p<0.0001). There was no significant difference in the age of breast cancer onset between *BRCA2*, *CHEK2*, *PALB2* and *ATM* gene carriers compared to non-carriers (Table 18 and Figure 15).

	Age in Years							
Summary Statistic	**Mutation-**	**Mutation+**	***BRCA1+***	***BRCA2+***	***CHEK2+***	***PALB2+***	***ATM+***	***TP53+***
Mean	35.7	34.9	34.3	35.3	35.6	35.5	36.7	31.3
Median	37.0	36.0	35.0	37.0	37.0	37.0	38.0	33.0
IQR	34.0-39.0	33.0-38.0	32.0-38.0	33.0-38.0	34.0-39.0	33.0-38.0	35.5-39.0	28.5-34.5
Minimum	18.0	20.0	22.0	21.0	20.0	29.0	29.0	22.0
Maximum	40.0	40.0	40.0	40.0	40.0	40.0	40.0	36.0
P Value†	-	p<0.0001	p<0.0001	p=0.3008	p=0.8971	p=0.5666	p=0.1911	p<0.0001

Table 18: Summary Statistics Age of Onset

Summary statistics comparing the average age of breast cancer onset across the POSH Cohort. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables. Mutation+, *BRCA1*+, *BRCA2*+, *CHEK2*+, *PALB2*+, *ATM*+ and *TP53*+ were compared against Mutation-.

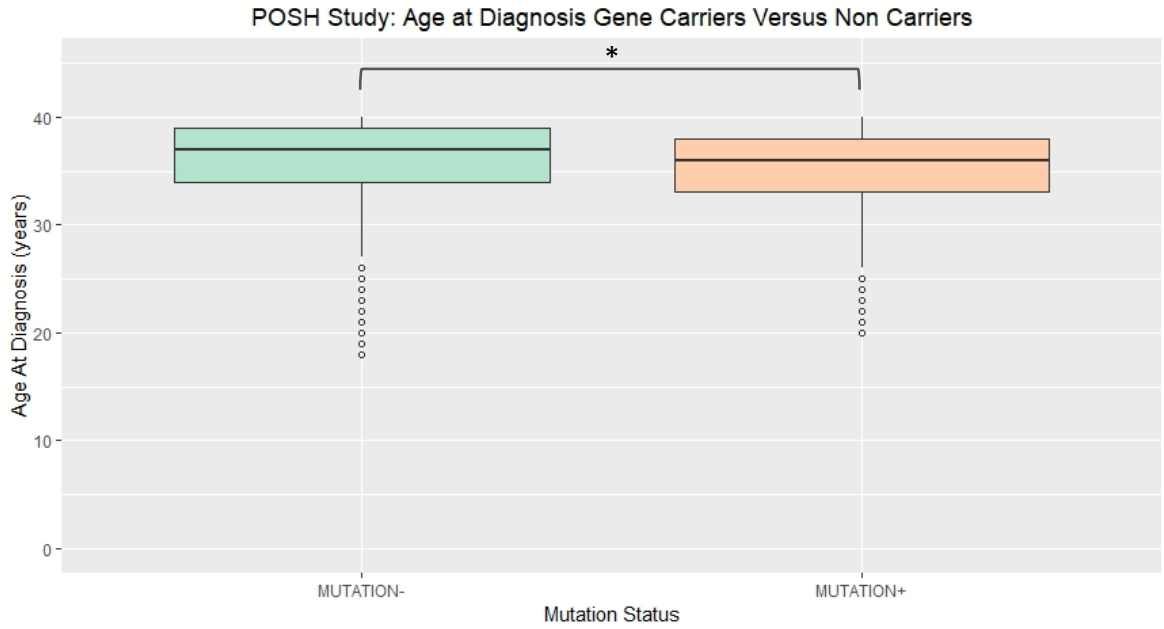


Figure 15: Age at Breast Cancer Diagnosis Gene Carriers versus Non-Carriers

Box plot demonstrating the age of breast cancer onset between gene carriers and non-carriers (*represents a statistically significant difference).

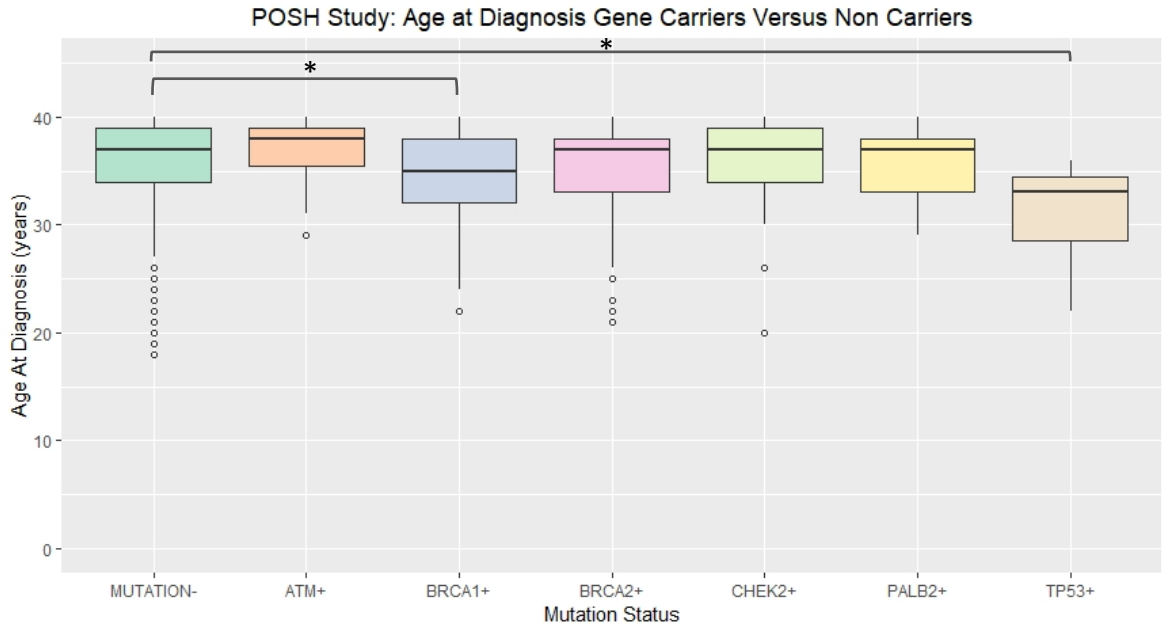


Figure 16: Age at Breast Cancer Diagnosis by Gene

Box plot demonstrating the age of breast cancer onset between individual gene carriers (*ATM*+, *BRCA1*+, *BRCA2*+. *CHEK2*+, *PALB2*+ and *TP53*+) compared to non-carriers (Mutation-) (*represents a statistically significant difference).

4.2 The 100,000 Genomes Project: Familial and Sporadic Breast Cancer

The complete analysed cohort within The 100,000 Genomes Project consisted of 826 participants diagnosed with familial breast cancer and 2464 unselected participants with breast cancer. This unselected cohort was most representative of sporadic breast cancer.

4.2.1 Age of Cancer Onset: Familial Breast Cancer

In the familial breast cancer cohort, there was no significant difference in the age of breast cancer onset between gene carriers and non-carriers. The median age at breast cancer diagnosis was 46.0 years (IQR, 42.0-51.5) amongst gene carriers and 46.0 years (IQR, 39.0-52.0) amongst non-carriers (p=0.832) (Table 20, Figure 17).

Sub-group analysis also demonstrated that age of breast cancer onset was not significantly different between *ATM*, *CHEK2*, *PALB2* and *TP53* gene carriers compared to non-gene carriers within the familial breast cancer cohort (Table 20 and Figure 18). The median age at breast cancer diagnosis was 47.0 years (IQR, 41.5-49.0 years) for *ATM* gene carriers, 45.0 years (IQR, 40.0-50.0 years) for *CHEK2* gene carriers, 52.0 (IQR, 43.0-57.3 years) for *PALB2* gene carriers and 43.5 years (IQR, 39.5-45.3) for *TP53* gene carriers (p=0.854, p=0.908, p=0.113, p=0.127 respectively) (Table 20 and Figure 18). *PALB2* gene carriers demonstrated a non-significant trend towards an older age at breast cancer diagnosis compared to other gene carriers within the familial breast cancer cohort.

4.2.2 Age of Cancer Onset: Unselected Breast Cancer

In the unselected breast cancer cohort, no significant difference in the age of breast cancer onset between gene carriers and non-carriers was observed. The median age of cancer onset was 59.0 years (IQR, 47.0-69.0 years) amongst gene carriers and 61.0 years (IQR, 51.0-70.0 years) for non-carriers (p=0.240) (Table 19 and Figure 17).

Sub-group analysis found that only carriers of a *BRCA* gene alteration (*BRCA1* and *BRCA2* combined) developed breast cancer at a significantly younger age compared to non-carriers. The median age at breast cancer diagnosis was 50.0 years (IQR, 47.5-53.0 years) for *BRCA* gene carriers versus 61.0 years (IQR, 51.0-71.0) for non-carriers (p=0.044) (Table 19 and Figure 18).

With the exception of this association, there was no significant difference in the age of breast cancer onset between individual gene carriers and non-carriers. The median age at breast cancer diagnosis was 61.0 years (IQR, 46.5-68.0 years) for *ATM* gene carriers, 64.5 years (IQR, 46.5-71.0 years) for *CHEK2* gene carriers, 57.0 (IQR, 51.0-67.5 years) for *PALB2* gene carriers and 64.0 years

(IQR, 46.0-68.0) for *TP53* gene carriers (p=0.496, p=0.911, p=0.742 and p=0.818 respectively) (Table 18 and Figure 20).

4.2.3 Age of Cancer Onset: A Comparison Across Cohorts

There was a significant difference in the age of breast cancer onset depending upon whether breast cancer was diagnosed in the context of familial or unselected breast cancer. This was true for both gene carriers and non-carriers. This is a notable finding as there was no significant difference in the age of cancer onset between gene carriers and non-carriers within the individual familial and unselected breast cancer recruitment cohorts.

The median age of cancer onset was significantly younger amongst gene-carriers identified in the context of familial breast cancer compared to unselected breast cancers (46.0 years (IQR, 42.0-51.5 years) versus 59.0 years (IQR, 47.0-69.0 years) (p<0.0001). The same was also true for non-carriers. The median age of cancer onset was 46.0 years (IQR, 39.0-52.0 years) amongst non-carriers identified in the context of familial breast cancer and 61.0 years (IQR, 51.0-70.0 years) for non-carriers identified as unselected breast cancer cases (p<0.0001)(Table 19 and Figure 17).

Subgroup analysis revealed a similar relationship for *CHEK2* and *TP53* gene carriers dependent upon the recruitment cohort. The median age of cancer onset was 45.0 years (IQR, 40.0-50.0 years) amongst *CHEK2* carriers identified in the context of familial breast cancer and 64.5 years (IQR, 46.5-71.0 years) for *CHEK2* carriers identified as unselected breast cancer cases (p=0.001) (table 16). The median age of cancer onset was 43.5 years (IQR, 39.5-45.3 years) amongst *TP53* carriers identified in the context of familial breast cancer and 64.0 years (IQR, 46.0-68.0 years) for *TP53* carriers identified as unselected breast cancer cases (p=0.042) (Table 19 and Figure 18). There was no significant difference in the age of breast cancer onset between *PALB2* or *ATM* gene carriers based upon the recruitment cohort. A comparison group for *BRCA* gene carriers was not available.

Rare Disease, Familial Breast Cancer: Age of Cancer Onset in Years							
Summary Statistic	**Mutation-**	**Mutation+**	***BRCA+***	***CHEK2+***	***PALB2+***	***ATM+***	***TP53+***
Mean	46.2	47.0	-	46.2	51.7	47.8	40.1
Median	46.0	46.0	-	45.0	52.0	47.0	43.5
IQR	39.0-52.0	42.0-51.5	-	40.0-50.0	43.0-57.3	41.5-49.0	39.5-45.3
Minimum	18.0	23.0	-	24.0	36.0	30.0	23.0
Maximum	78.0	77.0	-	75.0	77.0	73.0	48.0
p value†	-	0.832	-	0.908	0.113	0.854	0.127
Cancer, Breast Cancer: Age of Cancer Onset in Years							
Summary Statistic	**Mutation-**	**Mutation+**	***BRCA+***	***CHEK2+***	***PALB2+***	***ATM+***	***TP53+***
Mean	60.9	59.0	51.9	61.7	59.3	58.1	59.6
Median	61.0	59.0	50.0	64.5	57.0	61.0	64.0
IQR	51.0-71.0	47.0-69.0	47.5-53.0	46.5-71.0	51.0-67.5	46.5-68.0	46.0-68.0
Minimum	29.0	33.0	43.0	33.0	41.0	33.0	42.0
Maximum	94.0	92.0	69.0	92.0	80.0	88.0	78.0
p value†	-	0.240	0.044	0.911	0.743	0.496	0.818
P value ††	p<0.0001	p<0.0001	-	0.001	0.225	0.129	0.042

Table 19: Summary Statistics Age of Onset

Summary statistics comparing the average age of breast cancer onset across The 100,000 Genomes Project, Rare Disease, Familial Breast Cancer and Cancer, Breast Cancer domains. †Assessment of statistical significance was performed using the Mann-Whitney test (mutation+, *CHEK2*+, *PALB2*+, *ATM*+ and *TP53*+ compared against Mutation-). ††Assessment of statistical significance was performed using the Unpaired Two Sample T Test (mutation-, mutation+, *CHEK2*+, *PALB2*+, *ATM*+ and *TP53*+) Comparison between familial breast cancer and unselected breast cancer cohorts.

Figure 17: Age at Breast Cancer Diagnosis Gene Carriers versus Non-Carriers

Box plot demonstrating the age of breast cancer onset between gene carriers and non-carriers. Comparison between Familial Breast Cancer and Unselected Breast cancer cohorts within The 100,000 Genomes Project (*represents a statistically significant difference).

Figure 18: Age at Breast Cancer Diagnosis Gene Carriers versus Non-Carriers

Box plot demonstrating the age of breast cancer onset between individual gene carriers and non-carriers. Comparison between Familial Breast Cancer and Unselected Breast cancer cohorts within The 100,000 Genomes Project (*represents a statistically significant difference).

4.3 Discussion

Overall, the median age of breast cancer onset was 36.0 years amongst those recruited to the POSH study with symptomatic early onset breast cancer (diagnosis under the age of 40 years). Within this cohort, carriers of *BRCA1* and *TP53* pathogenic variants developed breast cancer at a significantly earlier age than non-carriers. This is consistent with much of the published literature. The median and inter-quartile range of breast cancer onset within The 100,000 Genomes project was over 31 years for both the familial and unselected recruitment domains. It means that TP53 gene carriers may have been screened out due to pretesting of very early onset breast cancers.

In 2017, Kuchenbaecker et al. prospectively observed the age of breast cancer onset amongst 9856 *BRCA1* and BRAC2 mutation carriers derived from three consortia (The International *BRCA1*/2 Carrier Cohort Study (IBCCS), the Breast Cancer Family Registry (BCFR) and the Kathleen Cunningham Foundation (kConFab).(41) The peak breast cancer incidence amongst *BRCA1* mutation carriers occurred between the ages of 41-50 years (28.3/1000 person years (95%CI, 23.1-34.7)). Conversely, the peak breast cancer incidence amongst *BRCA2* mutation carriers occurred between the ages of 51-60 years (30.6/1000 person years (95% CI, 22.8-41.1)).(41)

The age at breast cancer diagnosis also remains an important predictor of the likelihood of identifying a *TP53* gene alteration. It has been estimated that up to 3-8% of unselected very early onset breast cancers (diagnosed under the age of 30 years) will have a pathogenic variant in *TP53* with a higher prevalence in the context of a LFS/LFL pedigree (table 7).(106, 112-117) In contrast, Moran et al. 2017 and Rath et al. 2013 found the estimated prevalence in patients diagnosed with breast cancer at any age and ascertained through familial or mixed cohorts is lower ranging from 0.5-2%.(117, 118)

Most notable within this analysis is the recognition that the context in which breast cancer is diagnosed may represent an important aetiological risk factor and determinant of the age of breast cancer onset. Within The 100,000 Genomes Project, breast cancer cases identified in the context of familial breast cancer occurred at a significantly younger age compared to unselected cases. This was true for both gene carriers and non-carriers. It suggests that other weaker genetic factors are important modifiers of genetic risk even in the context of a moderate or high penetrance germline genotype.

A number of studies have observed the impact of polygenic factors on absolute cancer risks. It is well recognised that a family history of breast cancer increases absolute cancer risk with the level of risk rising incrementally with the number of affected relatives.(16) In 2020 Yang et al. observed that the family history of breast cancer modified *PALB2*-associated breast cancer risk.(20) The

absolute breast cancer risk to age 80 years increased from 52% (95% CI, 42%-62%) to 76% (95% CI, 69%-83%) in the presence of two affected first degree relatives. (20) They also observed that genetic models including a polygenic component provided a better fit to the observed risk of breast cancer than monogenic susceptibility in isolation.(20)

In 2017, Michailidou et al. determined that 18% of familial breast cancer susceptibility is attributable to common genetic variation identified through Genome Wide Association Studies (GWAS).(22) The Polygenic Risk Score (PRS) represents an objective measure of the relative contribution of these weaker genetic factors to the overall stratification of cancer risk for any individual. (16)

In 2017, Li et al. examined the utility of PRS in the context of non-*BRCA* familial breast cancer to determine whether a 24 SNP PRS could be used to prospectively stratify breast cancer risk amongst unaffected women with a family history of breast cancer.(24) In total, 1496 cases and 2869 controls were derived from two familial breast cancer cohorts within North America and Australia (The Breast Cancer Family Registry (BCFR) and the Kathleen Cuningham Consortium Foundation for Research into Familial Breast Cancer (kConFab)).(24) They identified a significant difference in breast cancer risk when comparing the highest and lowest quintiles of PRS (HR 3.18, 95%CI 1.84-5.23) $p=4.7x10^{-6}$. (24) This equated to absolute cumulative risks to age 70 years of 51% for women in the highest quintile and 21% in the lowest quintile. (24)

In 2019, Lakeman et al. used a validated 161 SNP PRS to determine the impact of PRS on breast cancer risk stratification in the context of *BRCA* negative familial breast cancer.(215) They compared PRS between 323 cases and 262 controls within a familial cohort and 357 breast cancer cases and 327 controls derived from an unselected cohort within the Dutch population. Overall, the mean PRS was higher in familial compared to unselected cases (0.70 (SD=0.90) versus 0.35 (SD=0.92)) indicating an excess of polygenic risk within the familial cohort.(215)

In 2017, Kuckenbaecker provided evidence for a modifying effect of PRS on absolute cancer risks in the context of high risk susceptibility secondary to a *BRCA1* or *BRCA2* gene alteration.(18) They observed PRS in 7797 *BRCA1* gene carriers and 4330 *BRCA2* gene carriers with breast cancer.(18) They demonstrated that estimations of absolute risk varied with standard deviations of PRS. *BRCA1* carriers with a PRS in the 90th centile had an estimated breast cancer risk of 39% by age 50 years whilst carriers at the 10th percentile had a risk of 21% by the age of 50 years.(18) Overall, the per standard deviation effect on breast cancer risk was smaller in *BRCA* gene carriers compared to other population based studies indicating a lesser effect in the context of a higher penetrance genotype.(18)

Several studies have also shown that PRS can optimise cancer risk stratification beyond the consideration of family history alone. In 2017, Li et al. concluded that the combination of BOADICEA and PRS may provide a more accurate estimation of breast cancer risk in comparison to the utilisation of family history in isolation.(24) In 2019, Lakeman et al. found that the addition utilisation of PRS changed screening advise beyond that which would have been give based upon family history alone.(215) In 2018, Van Veen et al. concluded that the utilisation of PRS based upon a subset of 18 SNPs enhanced breast cancer risk stratification beyond conventional statistical modelling (Tyrer Cuzick) and mammographic density alone.(19)

Collectively, these studies demonstrate that polygenic risk is an important determinant of absolute cancer risk even in the context of a higher penetrance genotype such as *BRCA*. We have shown that breast cancer cases identified in the context of familial breast cancer occurred at a significantly younger age compared to unselected cases. We can hypothesise that the polygenic factors are therefore important modifiers of both absolute cancer risks and the age of cancer onset. There is a paucity of published literature demonstrating the influence of PRS on the age of cancer onset. However, in 2012, Sawyer et al. observed the PRS amongst 1143 women with familial breast cancer. They found that a significantly higher proportion of women in the top quartile of polygenic risk developed very early onset breast cancer (OR 3.3 (95%CI 1.03-10.26) p=0.03).(216)

Risk estimations are a central to individualised decision making for risk management in the context of hereditary breast cancer susceptibility.(217) This includes the age to commence screening. Increasingly, age related stratifications of residual and contralateral risk are also utilised to support decision making around risk reducing surgery. This is of particular utility amongst women over the age of 50 years where the relative balance of risk versus survival benefit is sometimes less clear.

Currently, a variety of tools can be used to estimate absolute and age specific residual and contralateral breast cancer risks including data derived from prospective and retrospective cohort studies and statistical modelling tools such as CanRisk. The breast cancer risk stratification is largely based upon germline genotype and what we understand about the aggregated breast cancer risk across cohorts of gene carrier and non-carriers irrespective of family history. Statistical modelling tools such as CanRisk provides some weighting for the familial aggregation of cancer but SNP profiling has been consistently shown to improve this risk modelling.(19, 24, 215)

This work has shown that there was no significant difference in the age of breast cancer onset between gene carrier and non-carriers within the familial breast cancer cohort of 100K. There was however a significant difference in the age of breast cancer onset depending upon whether

cancer was identified in the context of a familial breast cancer. It raises the question about whether an individual at the highest quintile of PRS in a familial cohort should be offered high risk breast screening irrespective of whether they have a high penetrance breast cancer susceptibility genotype or not. It also raises the question of whether the age specific risk for an individual with a *BRCA1* gene alteration is the same if they have a strong family history of breast cancer versus no family history. As such, should SNP profiling to produce a PRS be introduced into routine testing to accentuate risk stratification and provide the opportunity for more personalised decision making around risk?

There are some limitations to this. Whilst, high and moderate risk breast cancer susceptibility genes have been identified alongside SNPs that confer sufficient risk to be incorporated into a PRS, the hereditary aspects driving the familial aggregation of many breast cancer remain largely undetermined.(16) There is also a lack of consensus over which SNPs confer the greatest utility within a PRS.(16)

4.4 Summary

Hereditary breast cancer susceptibility is a complex disorder influenced by a combination of strong and weak genetic factors which aggregate in families. This is true even in the presence of a high penetrance genotype such as *BRCA1*, *BRCA2* or *TP53*. The comprehensive analysis of each of these factors for any one individual has the potential to accentuate risk stratification including absolute and age specific cancer risks to achieve a more personalised approach to cancer risk management. It would be interesting to see how the percentile of PRS observed within POSH cohort of early onset breast cancer compares to the PRS within a cohort of unselected breast cancer cases. Future research may also consider whether polygenic modification of risk in the context of a high penetrance gene influences patient choices.

Chapter 5 Results: Tumour Histopathological Phenotype amongst Germline *BRCA1* and *BRCA2* Variant Carriers

5.1 Baseline Characteristics of the Cohort

The complete analysed cohort consisted of 2629 participants derived from the POSH study. This included 338 individuals with a pathogenic or likely pathogenic variant in *BRCA1* or *BRCA2* and 2291 mutation negative participants. Mutation negative was defined as being *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *ATM* and *TP53* negative (Table 20 and Table 21). Most recruits were Caucasian (2396/2594 (92.4%)) based upon self-reported ethnicity. There was missing ethnicity data for 35 individuals within this analysis.

A pathogenic *BRCA1* or *BRCA2* gene variant was found in 338/2629 (12.9%) of this study cohort and (338/2744 (12.3%)) of the whole cohort. Most of these variants were protein truncating. A pathogenic variant in *BRCA1* was found in 201/2629 (7.6%) of this cohort and 201/2744 (7.3%) of the whole cohort. A pathogenic variant in *BRCA2* was found in 137/2629 (5.2%) of this cohort and 137/2744 (5.0%) of the whole cohort.

There was no baseline difference in ethnicity or body mass index between *BRCA* mutation carriers and non-carriers (p=0.227 and p=0.539 respectively) (Table 21 and Table 22). On average, *BRCA* mutation carriers were significantly younger at diagnosis than non-carriers. The median age at breast cancer diagnosis was 36 years (IQR, 32-38 years) for *BRCA* mutation carriers versus 37 years (IQR, 34-39 years) for non-carriers (p<0.0001).

Subgroup analysis comparing *BRCA1* and *BRCA2* mutation carriers with non-carriers identified that *BRCA1* mutation carriers were significantly younger at diagnosis compared to non-carriers. The median age of breast cancer onset was 35 years (IQR, 32-38 years) for *BRCA1* carriers versus 37 years (IQR, 34-39 years) for non-carriers (p<0.0001). *BRCA2* mutation carriers were not significantly younger at diagnosis compared to non-carriers. The median age of breast cancer onset was 37 years (IQR, 33-38 years) for *BRCA2* carriers versus 37 years (IQR, 34-39 years) for non-carriers (p=0.301) (Table 20 and Table 21).

BRCA mutation carriers were significantly more likely to have a family history of breast cancer. In total, 209/322 (64.9%) of *BRCA* carriers had a family history of breast cancer compared to 670/2209 (30.3%) of non-carriers (p<0.0001). The median BOADICEA score for *BRCA* gene carriers

was also significantly higher than non-carriers (median score *BRCA1* 0.20 (IQR 0.06-0.58), *BRCA2* 0.11 (IQR 0.03-0.34) and non-*BRCA* 0.03 (IQR 0.02-0.05) (p<0.0001). However, it is notable that 69/193 (35.8%) of *BRCA1* mutation carriers and 44/129 (34.1%) of *BRCA2* mutation carriers had no family history of breast cancer (Table 20 and Table 21).

Characteristic		***BRCA1+*** (n=201)	***BRCA+*** **(ALL)** (n=338)	***BRCA-*** (n=2291)	**p-value†** ***BRCA1+ vs BRCA1-***	**p-value†** ***BRCA+*(ALL) vs *BRCA-***
Median age diagnosis (yrs)		35	36	37	p<0.0001	p<0.0001
	Range	22-40	21-40	18-40		
	IQR	32-38	32-38	34-39		
Body Mass Index (Total)		192 (100%)	325 (100%)	2203 (100%)	p=0.522	p=0.539
	Underweight/Healthy (<25)	114 (59.4%)	184 (56.6%)	1185 (53.8%)		
	Overweight (25-30)	47 (24.5%)	88 (27.1%)	603 (27.4%)		
	Obese (>30)	31 (16.1%)	53 (16.3%)	415 (18.8%)		
	Missing	9 (4.5%)	13 (3.8%)	88 (3.8%)		
Ethnicity		196 (100%)	330 (100%)	2264 (100%)	p=0.344	p=0.227
	Caucasian/white	178 (90.8%)	300 (90.9%)	2096 (92.6%)		
	Black	10 (5.1%)	16 (4.8%)	84 (3.7%)		
	Asian	5 (2.6%)	9 (2.7%)	70 (3.1%)		
	Other	3 (1.5%)	5 (1.5%)	14 (0.6%)		
	Missing	5 (2.5%)	8 (2.4%)	27 (1.2%)		
Family History		193 (100%)	322 (100%)	2209 (100%)	p<0.0001	p<0.0001
	No	69 (35.8%)	113 (35.1%)	1539 (69.7%)		
	Yes	124 (64.2%)	209 (64.9%)	670 (30.3%)		
	Missing	8 (4.0%)	16 (4.7%)	82 (3.6%)		
BOADICEA score					p<0.0001	p<0.0001
	Median	0.20	0.16	0.03		
	Range	0.01 to 1.00,	0.01 to 1.00,	0.00 to 0.95,		
	IQR	0.06 to 0.58	0.05 to 0.48	0.02 to 0.05		
	Missing	5 (2.5%)	10 (3.0%)	62 (2.7%)		

Table 20: Baseline Characteristics of the *BRCA1* Cohort

†Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

Characteristic		*BRCA2*+ (n=137)	*BRCA*+ (ALL) (n=338)	*BRCA*- (n=2291)	p-value† *BRCA2*+ vs *BRCA2*-	p-value† *BRCA*+(ALL) vs *BRCA*-
Median age diagnosis (yrs)		37	36	37	p=0.301	p<0.0001
	Range	21-40	21-40	18-40		
	IQR	33-38	32-38	34-39		
Body Mass Index (Total)		133 (100%)	325 (100%)	2203 (100%)	p=0.418	p=0.539
	Underweight/Healthy (<25)	70 (52.6%)	184 (56.6%)	1185 (53.8%)		
	Overweight (25-30)	41 (30.8%)	88 (27.1%)	603 (27.4%)		
	Obese (>30)	22 (16.5%)	53 (16.3%)	415 (18.8%)		
	Missing	4 (2.9%)	13 (3.8%)	88 (3.8%)		
Ethnicity		134 (100%)	330 (100%)	2264 (100%)	p=0.640	p=0.227
	Caucasian/white	122 (91.0%)	300 (90.9%)	2096 (92.6%)		
	Black	6 (4.5%)	16 (4.8%)	84 (3.7%)		
	Asian	4 (3.0%)	9 (2.7%)	70 (3.1%)		
	Other	2 (1.5%)	5 (1.5%)	14 (0.6%)		
	Missing	3 (2.2%)	8 (2.4%)	27 (1.2%)		
Family History		129 (100%)	322 (100%)	2209 (100%)	p<0.0001	p<0.0001
	No	44 (34.1%)	113 (35.1%)	1539 (69.7%)		
	Yes	85 (65.9%)	209 (64.9%)	670 (30.3%)		
	Missing	8 (5.8%)	16 (4.7%)	82 (3.6%)		
BOADICEA score					p<0.0001	p<0.0001
	Median	0.11	0.16	0.03		
	Range	0.01 to 0.99,	0.01 to 1.00,	0.00 to 0.95,		
	IQR	0.03 to 0.34	0.05 to 0.48	0.02 to 0.05		
	Missing	5 (3.6%)	10 (3.0%)	62 (2.7%)		

Table 21: Baseline Characteristics of the *BRCA2* Cohort

†Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

5.2 Tumour Histopathology

5.2.1 Baseline Tumour Grade, Size and Focality

BRCA-associated tumours were significantly more likely to be grade 3 at presentation compared to non-carriers (grade 3 268/326 (82.2%) versus 1278/2229 (57.3%) (p<0.0001)) (Figure 19 and Table 22).

5.2.1.1 *BRCA1*

There was no significant difference in the maximum invasive tumour size between *BRCA1* gene carriers and non-carriers (median 2.1cm (IQR 1.5-3cm) versus 2.2cm (IQR 1.5-3.3cm) (p=0.244)) (Table 22 and Figure 19). However, *BRCA1* associated tumours had significantly smaller maximum overall tumour sizes with significantly lower levels of in-situ disease compared to non-carriers. The median maximum overall tumour size for *BRCA1* mutations carriers was 2.2cm (IQR, 1.7-3.2cm) versus 2.7cm (IQR 1.8-4.0cm) for non-carriers (p<0.001). The median maximum in-situ tumour size for *BRCA1* mutation carriers was 1.4cm (IQR 0.3-2.5cm) versus 2.0cm (IQR 0.9-4.0cm) for non-carriers (p=0.043).

BRCA1 associated tumours were significantly more likely to be localised than non-carriers (156/180 (86.7%) versus 1461/2085 (70.1%) (p<0.0001)). *BRCA1* associated tumours also displayed significantly lower levels of nodal involvement and lymphovascular infiltration compared to non-carriers. In total, 129/201 (64.2%) of *BRCA1* associated tumours were N0 at presentation compared to 1084/2253 (48.1%) of non-carriers (p=<0.0001) and 116/190 (61.1%) of *BRCA1* associated tumours had no evidence of lymphovascular infiltration at presentation compared to 1106/2129 (51.9%) of non-carriers (p=0.016) (Table 22 and Figure 19).

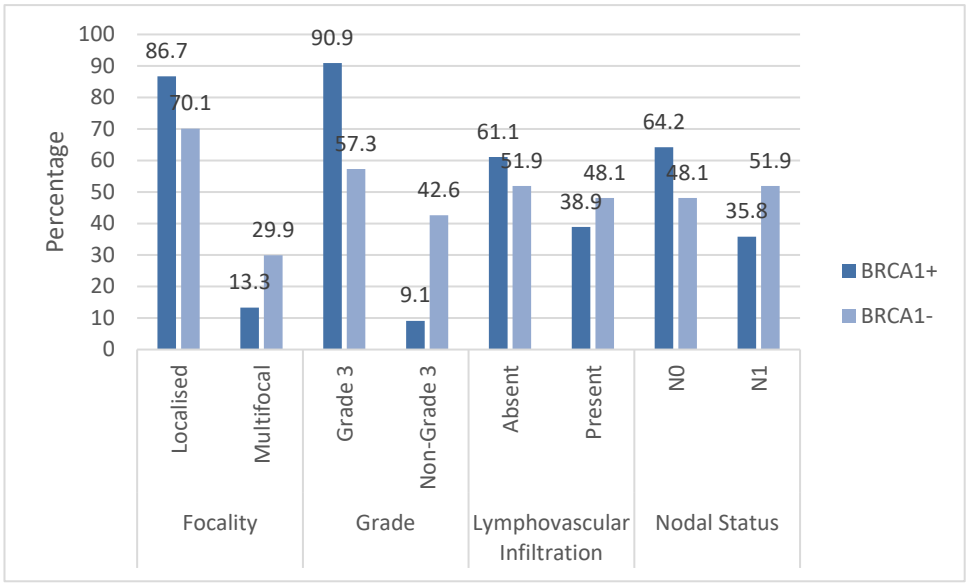


Figure 19: Baseline Histopathological Characteristics of the *BRCA1* Cohort

Comparison of the tumour focality, grade, nodal status and lymphovascular infiltration between *BRCA1* mutation carriers and non-carriers. Values represented as percentage of the cohort. *BRCA1* associated tumours were significantly more likely to be localised and displayed significantly lower levels of nodal involvement and lymphovascular infiltration.

Characteristic	*BRCA1*+ (n=201)	*BRCA*+ (ALL) (n=338)	*BRCA*- (n=2291)	p-value† *BRCA1*+ vs *BRCA1*-	p-value† *BRCA*+(ALL) vs *BRCA*-
Histological Grade (Total)	197 (100%)	326 (100%)	2229 (100%)	p<0.0001	p<0.0001
1	2 (1.0%)	2 (0.6%)	148 (6.6%)		
2	16 (8.1%)	56 (17.2%)	803 (36.0%)		
3	179 (90.9%)	268 (82.2%)	1278 (57.3%)		
Missing/not graded	4 (2.0%)	12 (3.6%)	62 (2.7%)		
Focality (Total)	180 (100%)	301 (100%)	2085 (100%)	p<0.0001	p=0.283
Localised	156 (86.7%)	220 (73.1%)	1461 (70.1%)		
Multifocal	24 (13.3%)	81 (26.9%)	624 (29.9%)		
Missing	21 (10.4%)	37 (10.9%)	206 (9.0%)		
Max invasive tumour size (cm)				p=0.244	p=0.956
Median	2.1	2.2	2.2		
Range	.1 to 14,	.05 to 14	0 to 17,		
IQR	1.5 to 3	1.5 to 3.1	1.5 to 3.3		
Missing	10 (5.0%)	24 (7.1%)	128 (5.6%)		
Max overall tumour size (cm)				p<0.001	p=0.423
Median	2.2	2.6	2.7		
Range	.1 to 14,	.06 to 15,	0 to 19,		
IQR	1.7 to 3.2	1.8 to 3.8	1.8 to 4		
Missing	10 (5.0%)	21 (6.2%)	104 (4.5%)		
Max in-situ tumour size (cm)				p=0.043	p=0.971
Median	1.4	2.0	2.0		
Range	.03 to 3.9,	.03 to 11.5,	0 to 19,		
IQR	.3 to 2.5	.8 to 3.8	.9 to 4.0		
Missing	182 (90.5%)	292 (86.4%)	2015 (88.0%)		
Pathological N stage (total)	201 (100%)	336 (100%)	2253 (100%)	p<0.0001	p=0.023
N0	129 (64.2%)	184 (54.8%)	1084 (48.1%)		
N1	72 (35.8%)	152 (45.2%)	1169 (51.9%)		
Missing	0	2 (0.6%)	38 (1.7%)		
Number of positive lymph nodes	201 (100%)	336 (100%)	2253 (100%)	p<0.0001	p=0.041
0	129 (64.2%)	184 (54.8%)	1084 (48.1%)		
1-3	43 (21.4%)	94 (28.0%)	764 (33.9%)		
4-9	14 (7.0%)	33 (9.8%)	273 (12.1%)		
10+	15 (7.5%)	25 (7.4%)	132 (5.9%)		
Missing	0	2 (0.6%)	38 (1.7%)		
Lymphovascular invasion (total)	190 (100%)	314 (100%)	2129 (100%)	p=0.016	p=0.251
Absent	116 (61.1%)	174 (55.4%)	1106 (51.9%)		
Present	74 (38.9%)	140 (44.6%)	1023 (48.1%)		
Missing	11 (5.5%)	24 (7.1%)	162 (7.1%)		

Table 22: Baseline Histopathological Characteristics of the *BRCA1* Cohort

Comparison of the tumour size, focality, grade and lymph node involvement between *BRCA1* mutation carriers and non-carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

5.2.1.2 *BRCA2*

There was no significant difference in the maximum invasive tumour size between *BRCA2* gene carriers and non-carriers (median 2.5cm (IQR, 1.7-3.2cm) versus 2.2cm (IQR, 1.5-3.3cm) (p=0.167)) (Table 23 and Figure 20). However, *BRCA2* associated tumours had a significantly greater maximum overall tumour size with a trend towards higher levels of in-situ disease compared to non-carriers. The median maximum overall tumour size for *BRCA2* mutations carriers was 3.2cm (IQR, 2.2-5.0cm) versus 2.7cm (IQR, 1.8-4.0cm) for non-carriers (p<0.001). The median maximum in-situ tumour size for *BRCA2* mutation carriers was 3.3cm (IQR, 1.4-5.0cm) versus 2.0cm (IQR 0.9-4.0cm) for non-carriers (p=0.104).

BRCA2 associated tumours were significantly more likely to be multifocal at presentation compared to non-carriers (57/121 (47.1%) versus 624/2085 (29.9%) (p<0.0001)). There was a trend towards N1 stage disease amongst *BRCA2* associated tumours compared to non-carriers (80/135 (59.3%) versus 1169/2253 (51.9%)) p=0.096. No significant difference was observed in the amount of lymphovascular infiltration between *BRCA2* gene carriers and non-carriers. In total 66/124 (53.2%) of *BRCA2* associated tumours had evidence of lymphovascular infiltration at presentation compared to 1023/2129 (48.1%) of non-carriers (p=0.262) (Table 23 and Figure 20).

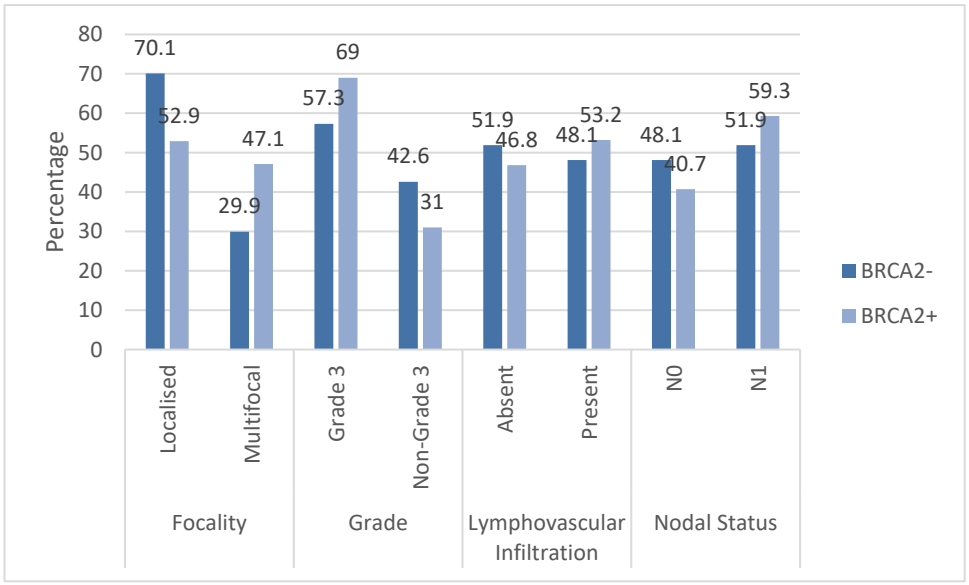


Figure 20: Baseline Histopathological Characteristics of the *BRCA2* Cohort

Comparison of the tumour size, focality, grade and lymph node involvement between *BRCA2* mutation carriers and non-carriers. † Samples derived from the POSH Cohort.

Characteristic	*BRCA2*+ (n=137)	*BRCA*+ (ALL) (n=338)	*BRCA*- (n=2291)	p-value† *BRCA2*+ vs *BRCA2*-	p-value† *BRCA*+(ALL) vs *BRCA*-
Histological Grade (Total)	129 (100%)	326 (100%)	2229 (100%)	p=0.002	p<0.0001
1	0	2 (0.6%)	148 (6.6%)		
2	40 (31.0%)	56 (17.2%)	803 (36.0%)		
3	89 (69.0%)	268 (82.2%)	1278 (57.3%)		
Missing/not graded	8 (5.8%)	12 (3.6%)	62 (2.7%)		
Focality (Total)	121 (100%)	301 (100%)	2085 (100%)	p<0.0001	p=0.283
Localised	64 (52.9%)	220 (73.1%)	1461 (70.1%)		
Multifocal	57 (47.1%)	81 (26.9%)	624 (29.9%)		
Missing	16 (11.7%)	37 (10.9%)	206 (9.0%)		
Max invasive tumour size (cm)				p=0.167	p=0.956
Median	2.5	2.2	2.2		
Range	.05 to 9.2,	.05 to 14	0 to 17,		
IQR	1.7 to 3.2	1.5 to 3.1	1.5 to 3.3		
Missing	14 (10.2%)	24 (7.1%)	128 (5.6%)		
Max overall tumour size (cm)				p<0.001	p=0.423
Median	3.2	2.6	2.7		
Range	.06 - 15,	.06 to 15,	0 to 19,		
IQR	2.2 to 5	1.8 to 3.8	1.8 to 4		
Missing	11 (8.0%)	21 (6.2%)	104 (4.5%)		
Max in-situ tumour size (cm)				p=0.104	p=0.971
Median	3.3	2	2.0		
Range	.06 to 11.5,	.03 to 11.5,	0 to 19,		
IQR	1.4 to 5	.8 to 3.8	.9 to 4.0		
Missing	110 (80.3%)	292 (86.4%)	2015 (88.0%)		
Pathological N stage (total)	135 (100%)	336 (100%)	2253 (100%)	p=0.096	p=0.023
N0	55 (40.7%)	184 (54.8%)	1084 (48.1%)		
N1	80 (59.3%)	152 (45.2%)	1169 (51.9%)		
Missing	2 (1.5%)	2 (0.6%)	38 (1.7%)		
Number of positive lymph nodes	135 (100%)	336 (100%)	2253 (100%)	p=0.404	p=0.041
0	55 (40.7%)	184 (54.8%)	1084 (48.1%)		
1-3	51 (37.8%)	94 (28.0%)	764 (33.9%)		
4-9	19 (14.1%)	33 (9.8%)	273 (12.1%)		
10+	10 (7.4%)	25 (7.4%)	132 (5.9%)		
Missing	2 (1.5%)	2 (0.6%)	38 (1.7%)		
Lymphovascular invasion (total)	124 (100%)	314 (100%)	2129 (100%)	p=0.262	p=0.251
Absent	58 (46.8%)	174 (55.4%)	1106 (51.9%)		
Present	66 (53.2%)	140 (44.6%)	1023 (48.1%)		
Missing	13 (9.5%)	24 (7.1%)	162 (7.1%)		

Table 23: Baseline Histopathological Characteristics of the *BRCA2* Cohort

Comparison of the tumour size, focality, grade and lymph node involvement between *BRCA2* mutation carriers and non-carriers. †Assessment of statistical significance

were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

5.2.2 Hormone Receptor Status

The majority of *BRCA* variant carriers developed HER2-negative tumour. The proportion of HER2-negative tumours was significantly greater in *BRCA* variant carriers compared to non-carriers ((275/301 (91.4%)) versus (1428/2021 (70.7%)) (p<0.0001)) (Table 24 and Figure 21).

5.2.2.1 *BRCA1*

BRCA1 related tumours were significantly more likely to be ER-negative compared to non-carriers (p<0.0001). In total, 151/200 (75.5%) of *BRCA1* associated tumours were ER-negative compared to 722/2279 (31.7%) of non-carriers. *BRCA1* related tumours were also significantly more likely to be PR-negative compared to non-carriers with 144/171 (84.2%) of *BRCA1* associated tumours demonstrating a PR-negative status compared to 764/1848 (41.3%) of non-carriers (p<0.0001). Overall, *BRCA1* carriers were significantly more likely to have a TNT compared to non-carriers 123/201 (61.2%) versus 417/2291 (18.2%) (p<0.0001) (Table 24 and Figure 21). An overview of variant identification in TNTs within the POSH cohort is provided in Appendix J.1.

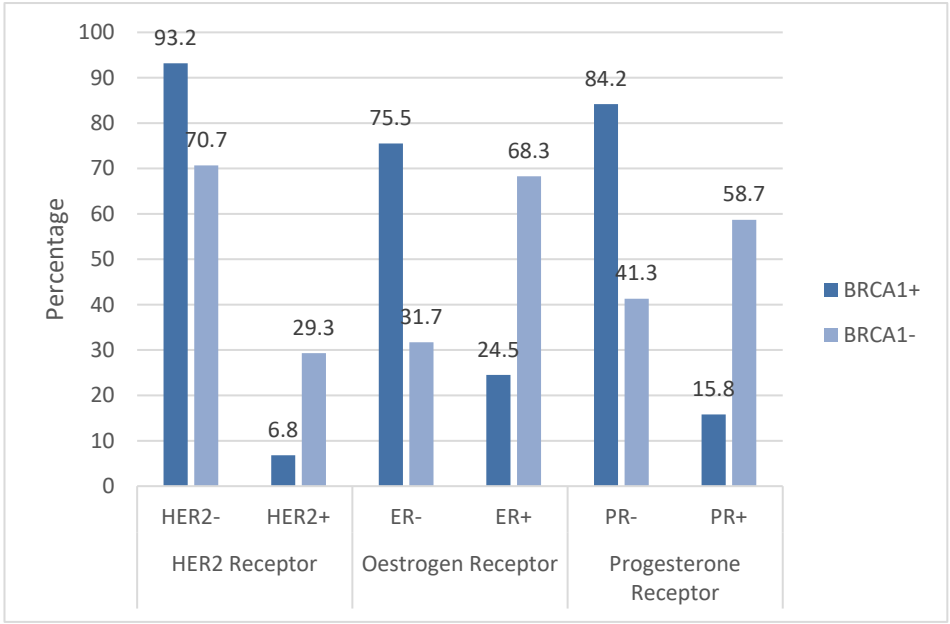


Figure 21: Baseline Hormone Receptor Status of the *BRCA1* Cohort

Comparison of ER, PR and HER2 receptor status between *BRCA1* mutations carriers and non-carriers. Values represented as percentage of the cohort.

Characteristic		*BRCA1*+ (n=201)	*BRCA*+ (ALL) (n=338)	*BRCA*- (n=2291)	p-value† *BRCA1*+ vs *BRCA1*-	p-value† *BRCA*+(ALL) vs *BRCA*-
ER status (total)		200 (100%)	336 (100%)	2279 (100%)	**p<0.0001**	p<0.0001
	Negative	151 (75.5%)	172 (51.2%)	722 (31.7%)		
	Positive	49 (24.5%)	164 (48.8%)	1557 (68.3%)		
	Missing	1 (0.5%)	2 (0.6%)	12 (0.5%)		
HER2 status (total)		176 (100%)	301 (100%)	2021 (100%)	**p<0.0001**	p<0.0001
	Negative	164 (93.2%)	275 (91.4%)	1428 (70.7%)		
	Positive	12 (6.8%)	26 (8.6%)	593 (29.3%)		
	Missing	25 (12.4%)	37 (10.9%)	270 (11.8%)		
PR status (total)		171 (100%)	278 (100%)	1848 (100%)	**p<0.0001**	p<0.0001
	Negative	144 (84.2%)	167 (60.1%)	764 (41.3%)		
	Positive	27 (15.8%)	111 (39.9%)	1084 (58.7%)		
	Missing	30 (14.9%)	60 (17.8%)	443 (19.3%)		
TNT status (total)		201 (100%)	338 (100%)	2291 (100%)	**p<0.0001**	p<0.0001
	Not TNT	78 (38.8%)	202 (59.8%)	1874 (81.8%)		
	TNT	123 (61.2%)	136 (40.2%)	417 (18.2%)		
	Missing	0	0	0		

Table 24: Baseline Hormone Receptor Status of the *BRCA1* Cohort

Comparison of ER, PR and HER2 receptor status between *BRCA1* mutation carriers and non-carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

5.2.2.2 *BRCA2*

BRCA2 related tumours were significantly more likely to be ER-positive compared to non-carriers. In total, 115/136 (84.6%) of *BRCA2* associated tumours were ER-positive compared to 1557/2279 (68.3%) of non-carriers (p<0.0001). *BRCA2* related tumours were also significantly more likely to be PR-positive compared to non-carriers with 84/107 (78.5%) of *BRCA2* associated tumours demonstrating a PR-positive status compared to 1084/1848 (58.7%) of non-carriers (p<0.0001). *BRCA2* associated tumours were not associated with a TNT phenotype. In total, 13/137 (9.5%) of *BRCA2* associated tumours were TNT at presentation compared to 417/2291 (18.2%) of non-carriers versus (p=0.009) (Table 25 and Figure 22).

Characteristic		***BRCA2*+** (n=137)	***BRCA*+ (ALL)** (n=338)	***BRCA*-** (n=2291)	**p-value†** ***BRCA2*+ vs *BRCA2*-**	**p-value†** ***BRCA*+(ALL) vs *BRCA*-**
ER status (total)		136 (100%)	336 (100%)	2279 (100%)	p<0.0001	p<0.0001
	Negative	21 (15.4%)	172 (51.2%)	722 (31.7%)		
	Positive	115 (84.6%)	164 (48.8%)	1557 (68.3%)		
	Missing	1 (0.7%)	2 (0.6%)	12 (0.5%)		
HER2 status (total)		125 (100%)	301 (100%)	2021 (100%)	p<0.0001	p<0.0001
	Negative	111 (88.8%)	275 (91.4%)	1428 (70.7%)		
	Positive	14 (11.2%)	26 (8.6%)	593 (29.3%)		
	Missing	12 (8.8%)	37 (10.9%)	270 (11.8%)		
PR status (total)		107 (100%)	278 (100%)	1848 (100%)	p<0.0001	p<0.0001
	Negative	23 (21.5%)	167 (60.1%)	764 (41.3%)		
	Positive	84 (78.5%)	111 (39.9%)	1084 (58.7%)		
	Missing	30 (21.9%)	60 (17.8%)	443 (19.3%)		
TNT status (total)		137 (100%)	338 (100%)	2291 (100%)	p=0.009	p<0.0001
	Not TNT	124 (90.5%)	202 (59.8%)	1874 (81.8%)		
	TNT	13 (9.5%)	136 (40.2%)	417 (18.2%)		
	Missing	0	0	0		

Table 25: Baseline Hormone Receptor Status of the *BRCA2* Cohort

Comparison of ER, PR and HER2 receptor status between *BRCA2* mutation carriers and non-carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

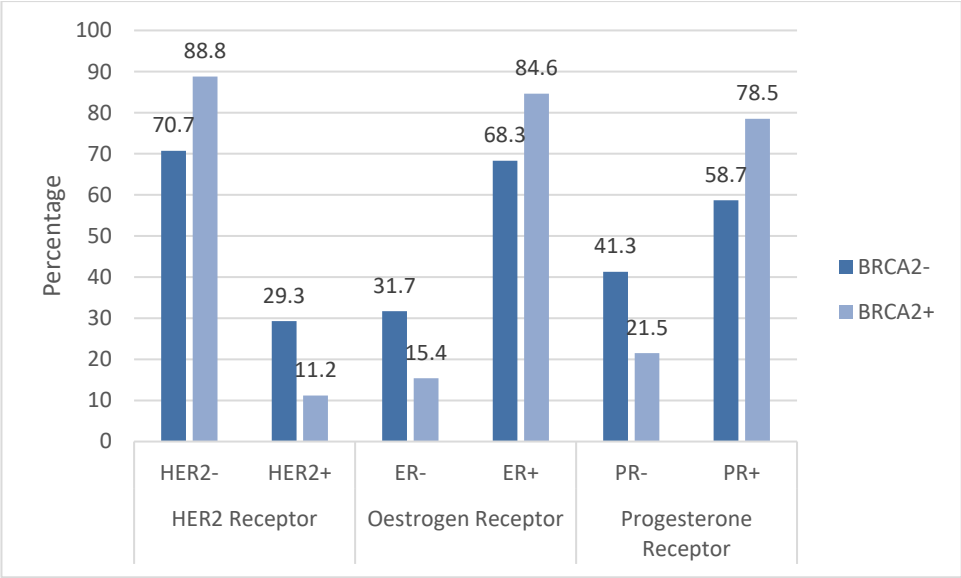


Figure 22: Baseline Hormone Receptor Status of the *BRCA2* Cohort

Comparison of ER, PR and HER2 receptor status between *BRCA2* mutations carriers and non-carriers. Values represented as percentage of the POSH cohort.

5.2.3 Histopathological Predictors

A model selection process was used to determine which histopathological characteristics to include in a multivariable logistic regression model. Multiple logistic regression incorporating forward selection by way of likelihood ratio tests was utilised to determine which histopathological features were better predictors of a *BRCA1* and *BRCA2* germline gene alteration (Table 26). Multiple imputation was utilised as 1705/2743 (37%) had missing data.

5.2.3.1 *BRCA1*

Tumour focality, hormone receptor status and family history were significant independent predictors of germline *BRCA1* mutation (Table 26). Overall, the highest probability of identifying a pathogenic *BRCA1* variant would be observed amongst those patients with a localised, ER-negative, PR-negative, HER2-negative tumour in the presence of a family history of breast cancer.

		BRCA1	
Phenotype	**Classification**	**Adjusted odds ratio [95% CI] †**	**Significant Factor**
Focality	Localised	1*	-
	Multifocal	0.53 (0.32-0.86)	Yes
Age at diagnosis (yrs)	18-25	1*	-
	26-30	0.81 (0.23-2.83)	-
	31-35	0.67 (0.20-2.21)	-
	36-40	0.42 (0.13-1.38)	-
BMI	Underweight/Healthy	1*	-
	Overweight	0.92 (0.66-1.28)	-
	Obese	0.84 (0.48-1.45)	-
Nodal Involvement	N0	1*	-
	N1	0.83 (0.57-1.22)	-
Oestrogen Receptor	Negative	1*	-
	Positive	0.35 (0.20-0.59)	Yes
Progesterone Receptor	Negative	1*	-
	Positive	0.30 (0.16-0.55)	Yes
HER2 Receptor	Negative	1*	-
	Positive	0.19 (0.10-0.37)	Yes
Ethnicity	White/Caucasian	1*	-
	Black	1.02 (0.80-1.30)	-
	Asian	1.00 (0.85-1.17)	-
	Other	1.03 (0.64-1.66)	-
Family History	No Family History	1*	-
	Family History	4.09 (2.92-5.75)	Yes

Table 26: Multiple Logistic Regression Analysis *BRCA1*

Multiple logistic regression analysis. †Analyses adjusted for hormone receptor status, invasive tumour size, nodal involvement, hormone receptor status, focality, age at diagnosis, BMI, ethnicity and family history. Grade was removed due to insufficient numbers. Tumour invasive size was fitted over overall and in-situ tumour size due to lower AIC in the complete case models, and TNT fitted over ER, PR and HER2 due to lower AIC in the complete case models. *Reference category.

5.2.3.2 *BRCA2*

Tumour focality, hormone receptor status and family history were significant independent predictors of a germline *BRCA2* mutation (Table 27). Overall, the highest probability of identifying a pathogenic *BRCA2* variant would be observed amongst those patients with a multifocal, ER-positive, PR-positive, HER2-negative tumour in the presence of a family history of breast cancer.

		BRCA2	
Phenotype	**Classification**	**Adjusted odds ratio [95% CI] †**	**Significant Factor**
Focality	Localised	1*	-
	Multifocal	1.90 (1.29-2.81)	Yes
Nodal Involvement	N0	1*	-
	N1	1.12 (0.77-1.62)	-
Oestrogen Receptor	Negative	1*	-
	Positive	1.18 (0.62-2.27)	-
Progesterone Receptor	Negative	1*	-
	Positive	2.49 (1.25-4.96)	Yes
HER2 Receptor	Negative	1*	-
	Positive	0.34 (0.20-0.60)	Yes
Family History	No Family History	1*	-
	Family History	3.86 (2.63-5.67)	Yes

Table 27: Multiple Logistic Regression Analysis *BRCA2*

Multiple logistic regression analysis. †Analyses adjusted for hormone receptor status, invasive tumour size, nodal involvement, hormone receptor status, focality, age at diagnosis, BMI, ethnicity and family history. Grade was removed due to insufficient numbers. Tumour invasive size was fitted over overall and in-situ tumour size due to lower AIC in the complete case models, and TNT fitted over ER, PR and HER2 due to lower AIC in the complete case models. *Reference category.

5.3 Discussion

This study represents one of the largest series comparing the histopathological phenotype of *BRCA1* and *BRCA2* mutation carriers with age matched controls (non-mutation carriers) diagnosed with a primary invasive breast cancer under the age of 40 years. Both cases and controls have been subject to the same genetic testing conditions.

5.3.1 *BRCA1* and Tumour Histopathology

BRCA1 mutations carriers were significantly younger at diagnosis than non-carriers. We have confirmed that *BRCA1* related tumours are significantly more likely to be grade 3, ER-negative, PR-negative and HER2-negative at presentation compared to non-carriers. They are also significantly more likely to present with an overall TN phenotype. These results are consistent with the established phenotype of *BRCA1*.(154, 218) Foulkes et al. found that the proportion of ER-negative tumours was higher in *BRCA1* mutation carriers than non-carriers across every age group.(149, 219) In 2012, Mavaddat et al. published one of the largest phenotypic series of 3797 *BRCA1* and 2392 *BRCA2* mutation carriers with primary invasive breast cancer.(219) They found

that the majority of tumours arising in *BRCA1* mutation carriers were ER-negative (78%), PR-negative (79%) and HER2-negative (90%). They also found that 68% of *BRCA1* associated tumours were TNT.(219)

This study has additionally found that *BRCA1* associated tumours are significantly more likely to be localised with lower levels of nodal involvement and lymphovascular infiltration at presentation compared to non-carriers. There was no significant difference in the maximum invasive tumour size between *BRCA1* gene carriers and non-carriers. However, *BRCA1* associated tumours demonstrated a smaller maximum overall tumour size and lower levels of in-situ disease compared to non-carriers. In 2007, Brekelmans et al. found that node negative tumours were significantly more common in *BRCA1* associated cancer than *BRCA2* (N0 *BRCA1* 63% versus N0 *BRCA2* 43%) ($p<0.001$).(153) However, the significance of focality, specifically the localisation of *BRCA1* associated tumours is a novel histopathological association. Further work could focus upon whether this is a product of TNT biology rather than the *BRCA1* germline variant itself.

5.3.2 *BRCA2* and Tumour Histopathology

BRCA2 mutations carriers were not significantly younger at diagnosis than non-carriers. Historically, the histopathological phenotype of *BRCA2* associated tumours is less well defined than *BRCA1*. We have shown that *BRCA2* associated tumours are significantly more likely to be grade 3 ER-positive, PR-positive and HER2-negative. Furthermore, *BRCA2* carriers with early onset breast cancer are significantly less likely to have a TN tumour compared to non-carriers

This is consistent with the results of many previous studies which have shown than *BRCA2* mutation carriers are more likely to have a high-grade tumours with a luminal B subtype.(154, 218, 220) Bane et al. found that *BRCA2* associated tumours were significantly more likely to be grade 3 than sporadic controls 60% versus 39% ($p<0.0001$).(220) They also found that HER2 expression was reduced compared to sporadic controls 6% versus 12%. Brekelmans et al. found that *BRCA2* associated tumours were significantly more likely to be oestrogen receptor positive than *BRCA1* associated tumours (84% versus 27% $p<0.001$).(153) In 2012, Mavaddat et al. demonstrated that 77% of tumours arising in *BRCA2* mutation carriers were ER-positive, 65% were PR-positive and 87% were HER2-negative.(219) In this cohort, only 16% of *BRCA2* associated tumours were triple negative.

This study has additionally found that *BRCA2* associated tumours were significantly more likely to be multifocal with a non-significant trend towards higher levels of nodal involvement compared to non-carriers. There was no significant difference in the maximum invasive tumour size between *BRCA2* gene carriers and non-carriers. However, *BRCA2* associated tumours had a significantly

greater maximum overall tumour size with a trend towards higher levels of in-situ disease compared to non-carriers. In 2018, Li et al. identified significantly higher levels of nodal involvement in *BRCA2* associated breast cancers $OR_{BRCA2 \text{ vs non-}BRCA}$ 2.71 (95% CI 1.31-5.62).(34) In 2017, Krammer et al. found that a significantly higher proportion of *BRCA2* mutation carriers presented with DCIS alone compared to *BRCA1* (15% (36/246) versus 9% (23/250) (p=0.0026)).(221) Some of these baseline differences in tumour morphology may reflect the difference in radiological appearance between *BRCA1* and *BRCA2* variant carriers that has previously been reported. In 2017, Ha et al. identified that *BRCA1* associated tumours are significantly more likely to appear benign on radiological assessment and have a well circumscribed margin compared to *BRCA2* associated tumours which more frequently have an indistinct margin (p=0.004).(222) In 2014, Yu et al. observed the histopathological phenotype of breast cancers in 181 *BRCA* mutation carriers of Korean descent and compared them to population controls. They identified that *BRCA2* associated tumours displayed significantly higher levels of nodal involvement with 45.5% demonstrating axillary nodal involvement versus 33.5% (p=0.002).(223)

5.3.3 Histopathological Predictors of Germline Mutations

A second aim of this study was to identify histopathological predictors of *BRCA1* and *BRCA2* carrier status. For the first time, we have identified that hormone receptor status, tumour focality and family history can serve as significant independent predictors of *BRCA1* and *BRCA2* carrier status. The use of focality as a significant independent predictor of germline *BRCA1* and *BRCA2* mutations is a novel concept.

When these features are identified in a sequential combination it can raise the probability of a *BRCA1* or *BRCA2* mutation to the NICE testing threshold of 10% amongst symptomatic early onset breast cancer. Overall, the tumour phenotype associated with the highest probability of identifying a pathogenic *BRCA1* variant would be observed amongst those patients with a localised, ER-negative, PR-negative, HER2-negative tumours in the presence of a family history of breast cancer. Conversely, the highest probability of identifying a pathogenic *BRCA2* variant would be observed amongst those with a multifocal, ER-positive, PR-positive, HER2-negative tumour in the presence of a family history of breast cancer.

Some of these results are consistent with Spurdle et al. who conducted a large-scale analysis of 4477 *BRCA1* mutation carriers, 2565 *BRCA2* mutation carriers and 47565 breast cancers in non-*BRCA* carriers in 2014.(154) They used ER status, age and grade to provide predictors of *BRCA1* and *BRCA2* mutations. They found that a combination of age, grade and ER receptor status

increased the likelihood of a pathogenic mutation than the presence of each factor in isolation.(154)

These phenotypic characteristics could be utilised as an adjunct to current probabilistic models of *BRCA* carrier risk including BOADICEA and the Manchester score which are heavily weighted towards family history (figure 7). They could also contribute to the evidence used for the interpretation of VUS and provide more accurate identification of those individuals who would benefit from germline genetic testing.(154) This is important as 35.8% of *BRCA1* mutation carriers and 34.1% of *BRCA2* mutation carriers had no family history of cancer. These individuals had the same histopathological phenotype as the *BRCA* mutation carriers with a family history but did not meet the *BRCA* testing threshold using conventional carrier probability models.

5.4 Summary

Overall, this work has provided the most comprehensive overview of tumour histopathology in early onset *BRCA*-associated breast cancer compared to non-*BRCA* carriers. It has more definitively described the *BRCA1* and *BRCA2* associated tumour phenotype and identified significant independent histopathological predictors of *BRCA1* and *BRCA2* mutation carriers including family history, hormone receptor status and focality. For the first time, we have defined the importance of tumour focality in the prediction of *BRCA* carrier status, and the novel histopathological feature of multifocality in association with *BRCA2*. It has also demonstrated that HER2 negativity an important independent predictor of *BRCA1* and *BRCA2* mutations. These factors can be incorporated into a carrier risk stratification model which can be used in combination with other carrier risk probability methods such as BOADICEA to determine *BRCA1* and *BRCA2* germline predisposition and improve the classification of VUS.

Chapter 6 Results: Tumour Histopathological Phenotype amongst Germline *CHEK2* Variant Carriers

6.1 Baseline Characteristics of the Cohort

The complete analysed cohort consisted of 2344 participants derived from the POSH study. This included 53 individuals with a tier 1 variant in *CHEK2* and 2291 variant negative participants. Variant negative was defined as being *CHEK2, PALB2*, *ATM*, *TP53*, *BRCA1* and *BRCA2* negative.

Most recruits were Caucasian 2146/2317 (92.6%). A pathogenic variant in *CHEK2* was found in 53/2344 (2.3%) of this study cohort (53/2744 (1.9%) of the whole cohort) with *CHEK2*, c.1100delC, p.(Thr367fs) being the most frequently identified and accounting for 36/53 (67.9%) of all *CHEK2* pathogenic variants (Table 28) (Appendix H). A further 28 individuals had Variants of Uncertain Significance in *CHEK2*. This included 27 missense variants and 1 in-frame deletion. These were defined as mutation negative (Appendix I).

The median age of cancer onset was 37 years (IQR 34-39 years) for *CHEK2* pathogenic and likely pathogenic variants carriers, and 37 years (IQR 34-39 years) for non-carriers. The majority of *CHEK2* carriers were Caucasian 50/53 (94.3%). There was no association between family history of breast cancer and median BOADICEA score between *CHEK2* variant carriers and non-carriers. In total, 36/51 (70.6%) of all *CHEK2* variant carriers had no family history of breast cancer. The median BOADICEA score was 0.03 for both *CHEK2* variant carriers and non-carriers ((*CHEK2* carriers, IQR 0.02-0.07) versus (*CHEK2* non-carriers, IQR 0.02-0.05) (p=0.86)) (Table 28). However, *CHEK2* variant carriers with invasive breast cancer were significantly more likely to be obese than non-carriers (28.3% versus 18.8%, p=0.039).

Characteristic		*CHEK2*+ (c.1100delC) (n=36)	*CHEK2*+ (other truncating) (n=17)	*CHEK2*+ (ALL) (n=53)	*CHEK2*- (n=2291)	p-value† *CHEK2*+ (c.1100delC) vs *CHEK2*+ (other)	p-value† *CHEK2*+(ALL) vs *CHEK2*-
Median age diagnosis (yrs)		35	38	37	37	p=0.31	p=0.90
	Range	20-40	26-40	20-40	18-40		
	IQR	34-38.5	35-39	34-39	34-39		
Body Mass Index (Total)		36 (100%)	17 (100%)	53 (100%)	2203 (100%)	p=0.97	p=0.039
	Underweight/Healthy (<25)	21 (58.3%)	10 (58.8%)	31 (58.5%)	1185 (53.8%)		
	Overweight (25-30)	5 (13.9%)	2 (11.8%)	7 (13.2%)	603 (27.4%)		
	Obese (>30)	10 (27.8%)	5 (29.4%)	15 (28.3%)	415 (18.8%)		
	Missing	0	0	0	88 (3.8%)		
Ethnicity		36 (100%)	17 (100%)	53 (100%)	2264 (100%)	p=0.68	p=0.40
	Caucasian/white	34 (94.4%)	16 (94.1%)	50 (94.3%)	2096 (92.6%)		
	Black	1 (2.8%)	1 (5.9%)	2 (3.8%)	84 (3.7%)		
	Asian	1 (2.8%)	0	1 (1.9%)	70 (3.1%)		
	Other	0	0	0	14 (0.6%)		
	Missing	0	0	0	27 (1.2%)		
Family History		35 (100%)	17 (100%)	51 (100%)	2209 (100%)	p=0.77	p=0.89
	No	24 (68.6%)	9 (52.9%)	36 (70.6%)	1539 (69.7%)		
	Yes	11 (31.4%)	8 (47.1%)	15 (29.4%)	670 (30.3%)		
	Missing	1 (2.8%)	0	2 (3.8%)	82 (3.6%)		
BOADICEA score						p=0.87	p=0.86
	Median	0.03	0.03	0.03	0.03		
	Range	0.01 to 0.76,	0.01 to 0.86,	0.01 to 0.86,	0.00 to 0.95,		
	IQR	0.01 to 0.07	0.02 to 0.07	0.02 to 0.07	0.02 to 0.05		
	Missing	1 (2.8%)	1 (5.9%)	2 (3.8%)	64 (2.7%)		

Table 28: Baseline Characteristics of the *CHEK2* Cohort

†Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

6.2 Tumour Histopathology

6.2.1 Baseline Tumour Grade, Size and Focality

Overall, *CHEK2*-associated tumours were significantly more likely to be grade 2 at presentation compared to non-carriers (Grade 2 28/52 (53.8%) versus 803/2229 (36.0%) (p=0.029)). There was no difference in baseline tumour size between *CHEK2* carriers and non-carriers (Table 29 and Figure 23). *CHEK2*-associated tumours also displayed significantly higher levels of nodal involvement compared to non-carriers. In total, 37/53 (69.8%) of *CHEK2* carriers presented with

N1 stage disease versus 1169/2253 (51.9%) of non-carriers (p=0.0098) (Table 29 and Figure 23). In addition, *CHEK2*-associated tumours demonstrated a trend towards multifocality at presentation compared to non-carriers (22/52 (42.3%) versus 624/2085 (29.9%) (p=0.055)).

We compared baseline tumour grade, size and focality between *CHEK2* c.1100delC carriers and all other truncating variant carriers and found no significant difference indicating a shared histopathological tumour phenotype between all protein truncating variants in *CHEK2*.

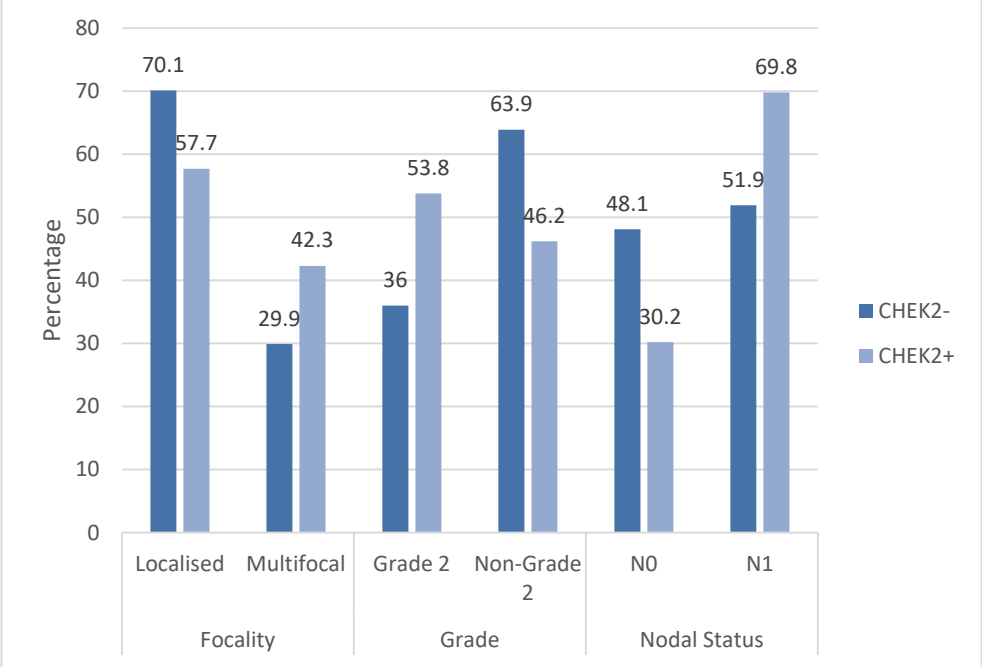


Figure 23: Baseline Histopathological Characteristics of the *CHEK2* Cohort

Comparison of the tumour focality, grade and lymph node involvement between *CHEK2* truncating variant carriers and non-*CHEK2* carriers. Values represented as percentage of the cohort. Samples derived from the POSH Cohort. Reprinted with permission. © (2020) American Society of Clinical Oncology. All rights reserved.(224)

Characteristic	*CHEK2*+ (c.1100delC) (n=36)	*CHEK2*+ (other truncating) (n=17)	*CHEK2*+ (ALL) (n=53)	*CHEK2*- (n=2291)	p-value† *CHEK2*+ (c.1100delC) vs *CHEK2*+ (other)	p-value† *CHEK2*+(ALL) vs *CHEK2*-
Histological Grade (Total)	35 (100%)	17 (100%)	52 (100%)	2229 (100%)	p=0.78	p=0.029
1	2 (5.7%)	1 (5.9%)	3 (5.8%)	148 (6.6%)		
2	20 (57.1%)	8 (47.1%)	28 (53.8%)	803 (36.0%)		
3	13 (37.1%)	8 (47.1%)	21 (40.4%)	1278 (57.3%)		
Missing/not graded	1 (2.8%)	0	1 (1.9%)	62 (2.7%)		
Focality (Total)	35 (100%)	17 (100%)	52 (100%)	2085 (100%)	p=0.91	p=0.055
Localised	20 (57.1%)	10 (58.8%)	30 (57.7%)	1461 (70.1%)		
Multifocal	15 (42.9%)	7 (41.2%)	22 (42.3%)	624 (29.9%)		
Missing	1 (2.8%)	0	1 (1.9%)	206 (9.0%)		
Max invasive tumour size (total)	35 (100%)	16 (100%)	51 (100%)	2163 (100%)	p=0.13	p=0.73
15mm or less	7 (20.0%)	4 (25.0%)	11 (21.6%)	561 (25.9%)		
>15mm to 20mm	8 (22.9%)	2 (12.5%)	10 (19.6%)	403 (18.6%)		
>20mm to 35mm	13 (37.1%)	7 (43.8%)	20 (39.2%)	733 (33.9%)		
>35mm to 50mm	1 (2.9%)	3 (18.8%)	4 (7.8%)	269 (12.4%)		
>50mm	6 (17.1%)	0	6 (11.8%)	197 (9.1%)		
Missing	1 (2.8%)	1 (5.9%)	2 (3.8%)	128 (5.6%)		
Pathological N stage (total)	36 (100%)	17 (100%)	53 (100%)	2253 (100%)	p=0.93	p=0.0098
N0	11 (30.6%)	5 (29.4%)	16 (30.2%)	1084 (48.1%)		
N1	25 (69.4%)	12 (70.6%)	37 (69.8%)	1169 (51.9%)		
Missing	0	0	0	38 (1.7%)		
Number of positive lymph nodes	36 (100%)	17 (100%)	53 (100%)	2253 (100%)	p=0.23	p=0.046
0	11 (30.6%)	5 (29.4%)	16 (30.2%)	1084 (48.1%)		
1-3	17 (47.2%)	6 (35.3%)	23 (43.4%)	764 (33.9%)		
4-9	5 (13.9%)	6 (35.3%)	11 (20.8%)	273 (12.1%)		
10+	3 (8.3%)	0	3 (5.7%)	132 (5.9%)		
Missing	0	0	0	38 (1.7%)		
Lymphovascular invasion (total)	32 (100%)	16 (100%)	48 (100%)	2129 (100%)	p=0.83	p=0.090
Absent	13 (40.6%)	6 (37.5%)	19 (39.6%)	1106 (51.9%)		
Present	19 (59.4%)	10 (62.5%)	29 (60.4%)	1023 (48.1%)		
Missing	4 (11.1%)	1 (5.9%)	5 (9.4%)	162 (7.1%)		

Table 29: Baseline Histopathological Characteristics of the *CHEK2* Cohort

Comparison of the tumour focality, grade and lymph node involvement between *CHEK2* truncating variant carriers and non-*CHEK2* carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort. Reprinted with permission. © (2020) American Society of Clinical Oncology. All rights reserved.(224)

6.2.2 Hormone Receptor Status

CHEK2 related tumours were significantly more likely to be ER-positive and PR-positive compared to non-carriers (Table 30 and Figure 24). In total, 47/53 (88.7%) of *CHEK2*-associated tumours were ER-positive compared to 1557/2279 (68.3%) of non-carriers (p=0.0016) and 33/42 (78.6%) of *CHEK2*-associated tumours were PR-positive compared to 1084/1848 (58.7%) of non-carriers (p=0.0094). *CHEK2* carriers were also significantly less likely to have a TN tumour compared to non-carriers (p=0.0022). In total, 1/53 (1.9%) of *CHEK2*-associated tumours had a TN phenotype compared to 417/2291 (18.2%) of non-*CHEK2* tumours (Table 30 and Figure 24). There was no significant association with HER2 receptor status and *CHEK2* genotype.

Characteristic	***CHEK2*+ (c.1100delC)** (n=36)	***CHEK2*+ (other truncating)** (n=17)	***CHEK2*+ (ALL)** (n=53)	***CHEK2*-** (n=2291)	**p-value†** ***CHEK2*+ (c.1100delC) vs *CHEK2*+ (other)**	**p-value†** ***CHEK2*+(ALL) vs *CHEK2*-**
ER status (total)	36 (100%)	17 (100%)	53 (100%)	2279 (100%)	p=0.94	p=0.0016
Negative	4 (11.1%)	2 (11.8%)	6 (11.3%)	722 (31.7%)		
Positive	32 (88.9%)	15 (88.2%)	47 (88.7%)	1557 (68.3%)		
Missing	0	0	0	12 (0.5%)		
HER2 status (total)	32 (100%)	16 (100%)	48 (100%)	2021 (100%)	p=0.21	p=0.22
Negative	22 (68.8%)	8 (50.0%)	30 (62.5%)	1428 (70.7%)		
Positive	10 (31.3%)	8 (50.0%)	18 (37.5%)	593 (29.3%)		
Missing	4 (11.1%)	1 (5.9%)	5 (9.4%)	270 (11.8%)		
PR status (total)	28 (100%)	14 (100%)	42 (100%)	1848 (100%)	p=0.43	p=0.0094
Negative	5 (17.9%)	4 (28.6%)	9 (21.4%)	764 (41.3%)		
Positive	23 (82.1%)	10 (71.4%)	33 (78.6%)	1084 (58.7%)		
Missing	8 (22.2%)	3 (17.6%)	11 (20.8%)	443 (19.3%)		
TNT status (total)	36 (100%)	17 (100%)	53 (100%)	2291 (100%)	p=0.14	p=0.0022
Not TNT	36 (100.0%)	16 (94.1%)	52 (98.1%)	1874 (81.8%)		
TNT	0	1 (5.9%)	1 (1.9%)	417 (18.2%)		
Missing	0	0	0	0		

Table 30: Baseline Hormone Receptor Status of the *CHEK2* Cohort

Comparison of ER, PR and HER2 receptor status between *CHEK2* truncating variant carriers and non-*CHEK2* carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort. Reprinted with permission. © (2020) American Society of Clinical Oncology. All rights reserved.(224)

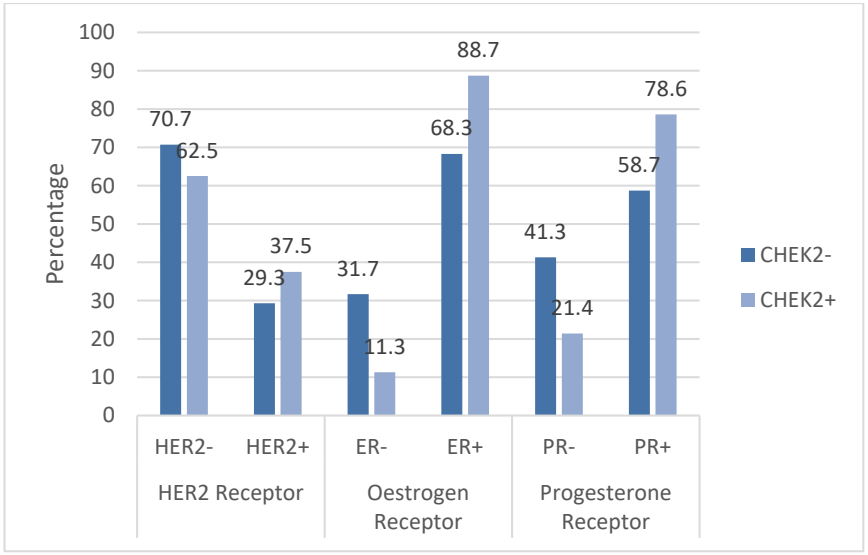


Figure 24: Baseline Hormone Receptor Status of the *CHEK2* Cohort

Comparison of ER, PR and HER2 receptor status between *CHEK2* truncating variant carriers and non-*CHEK2* carriers. Values represented as percentage of the cohort. Samples derived from the POSH Cohort. Reprinted with permission. © (2020) American Society of Clinical Oncology. All rights reserved.(224)

6.2.3 Histopathological Predictors of Germline Genotype

A model selection process using multiple logistic regression and incorporating forward selection by way of likelihood ratio tests was utilised to determine which histopathological features were better predictors of a *CHEK2* germline gene alteration (Table 31). Multiple imputation was utilised as 869/2344 (37%) had missing data. This analysis identified that only nodal involvement was a significant independent predictor of germline mutational status, N1 OR = 1.98 (95% CI, 1.09-3.60). Overall, the highest probability of identifying a pathogenic *CHEK2* variant would be observed amongst those patients with multifocal, N1 stage tumours demonstrating hormone receptor positivity (ER, PR and HER2).

		CHEK2	
Phenotype	**Classification**	**Adjusted odds ratio [95% CI] †**	**Significant Factor**
Nodal Involvement	N0	1*	-
	N1	1.98 (1.09-3.60)	Yes
Focality	Localised	1*	-
	Multifocal	1.02 (0.81-1.28)	-
Oestrogen Receptor	Negative	1*	-
	Positive	2.79 (0.81-9.54)	-
Progesterone Receptor	Negative	1*	-
	Positive	1.23 (0.48-3.15)	-
HER2 Receptor	Negative	1*	-
	Positive	1.15 (0.67-1.97)	-

Table 31: Multiple Logistic Regression Analysis

Multiple logistic regression analysis. †Analyses adjusted for hormone receptor status, tumour size, grade, nodal involvement, hormone receptor status, focality, age at diagnosis and lymphovascular infiltration. *Reference category.

6.3 Outcome

The histopathological analysis has utilised data derived from the POSH study. The complete analysed cohort consisted of 2397 participants. This included 53 individuals with a tier 1 variant in *CHEK2* and 2344 mutation negative participants.

The majority of patients received adjuvant chemotherapy. The most frequent regimen included anthracyclines with or without the additional of taxanes. There were no significant differences in the treatment received between *CHEK2* variant carriers and non-carriers (Table 32). However, a non-significant trend towards mastectomy was identified amongst *CHEK2* variant carriers. In total 36/53 (67.9%) of *CHEK2* variant carriers underwent mastectomy as the primary surgical intervention versus 1122/2291 (49.0%) of non-carriers (p=0.054).

Treatment	*CHEK2*+ (c.1100delC) (n=36)	*CHEK2*+ (other truncating) (n=17)	*CHEK2*+ (ALL) (n=53)	*CHEK2*- (n=2691)	p-value† *CHEK2*+ (c.1100delC) vs *CHEK2*+ (other)	p-value† *CHEK2*+(ALL) vs *CHEK2*-
Chemotherapy Timing (total)	36 (100%)	17 (100%)	53 (100%)	2291 (100%)	p=0.85	p=0.51
None	3 (8.3%)	2 (11.8%)	5 (9.4%)	262 (11.4%)		
Adjuvant	26 (72.2%)	11 (64.7%)	37 (69.8%)	1682 (73.4%)		
Neoadjuvant	7 (19.4%)	4 (23.5%)	11 (20.8%)	347 (15.1%)		
Palliative	0	0	0	0		
Missing	0	0	0	0		
Chemotherapy Regimen (total)	36 (100%)	17 (100%)	53 (100%)	2291 (100%)	p=0.18	p=0.70
None	3 (8.3%)	2 (11.8%)	5 (9.4%)	262 (11.4%)		
Anthracyclines	25 (69.4%)	7 (41.2%)	32 (60.4%)	1463 (63.9%)		
Anthracyclines and taxanes	7 (19.4%)	7 (41.2%)	14 (26.4%)	530 (23.1%)		
Taxanes only	0	1 (5.9%)	1 (1.9%)	21 (0.9%)		
Other	1 (2.8%)	0	1 (1.9%)	15 (0.7%)		
Missing	0	0	0	0		
Surgery Type (total)	36 (100%)	17 (100%)	53 (100%)	2291 (100%)	p=0.12	p=0.054
Breast Conserving Surgery	14 (38.9%)	3 (17.6%)	17 (32.1%)	1149 (50.2%)		
Mastectomy	22 (61.1%)	14 (82.4%)	36 (67.9%)	1122 (49.0%)		
Nodal surgery only	0	0	0	6 (0.3%)		
None	0	0	0	14 (0.6%)		
Missing	0	0	0	0		

Table 32: Treatment Characteristics of the *CHEK2* Cohort

Comparison of the treatment protocol in relation to genotype. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort. Reprinted with permission. © (2020) American Society of Clinical Oncology. All rights reserved.(224)

The median duration of follow up was 8.2 years. Contralateral breast cancers were more frequently observed in *CHEK2* carriers compared to non-carriers. A contralateral breast cancer was observed in 5/53 (9.4%) of *CHEK2* carriers at 10 years compared to 85/2291 (3.7%) of non-carriers. Of the 5 *CHEK2* variant carriers with contralateral breast cancer, 2 had bilateral disease at presentation and a further participant was found to have a contralateral breast cancer in the same year as their primary breast cancer diagnosis.

Subgroup analysis revealed that the observed increase in contralateral breast cancer risk observed amongst *CHEK2* variant carriers occurred in the context of familial breast cancer. In total, 3/15 (20.0%) of *CHEK2* variant carriers with a positive family history of breast cancer

developed a contralateral breast cancer compared to 1/36 (2.8%) *CHEK2* variant carriers without a family history. This difference was apparent in the first five years following breast cancer diagnosis. The contralateral breast cancer rates observed amongst *CHEK2* carriers without a family history of breast cancer where similar to the non-carriers with or without a family history (Table 33). Furthermore, 3/5 *CHEK2* carriers with contralateral disease had a family history of breast cancer and were obese.

Genotype	5 Years (%)	10 years (%)
CHEK2+ (ALL)	4 (7.5%)	5 (9.4%)
CHEK2+ (ALL) and FH-	0 (0%)	1 (2.8%)
CHEK2+ (ALL) and FH+	3 (20%)	3 (20%)
Mutation- (ALL)	62 (2.7%)	85 (3.7%)
Mutation- (ALL) and FH-	46 (3%)	59 (3.8)
Mutation- (ALL) and FH+	15 (2.2%)	24 (3.6%)

Table 33: Contralateral Breast Cancer Risk in Association with *CHEK2*

Contralateral breast cancer risk at both 5 and 10 years for *CHEK2* truncating variant carriers versus non-carriers. The presence of a family history (FH+) was associated with an increased contralateral breast cancer risk. Reprinted with permission.© (2020) American Society of Clinical Oncology. All rights reserved.(224)

Univariable analysis identified significantly worse Overall Survival (OS) in *CHEK2* variant carriers versus non-carriers (HR, 1.58 (95%CI, 1.01-2.48 (p=0.043))) (Figure 25). At 5 years, OS was 75.1% (95% CI, 60.9-84.7) amongst *CHEK2* mutation carriers versus 85.1% (95% CI, 83.5-86.5) in non-carriers. At 10 years, OS was 60.7% (95% CI, 42.5-74.8) amongst *CHEK2* mutation carriers versus 70.2% (95% CI, 67.8-72.5) in non-carriers. The observed difference in OS between *CHEK2* mutation carriers and non-carriers was maintained after adjustment for known prognostic factors including age at diagnosis, BMI, grade, maximum invasive size (cm), hormone receptor status, nodal involvement, ethnicity and taxanes in a multivariable analysis (HR 1.65 (95%CI, 1.05-2.59 (p=0.03)) (Appendix K.1.1)

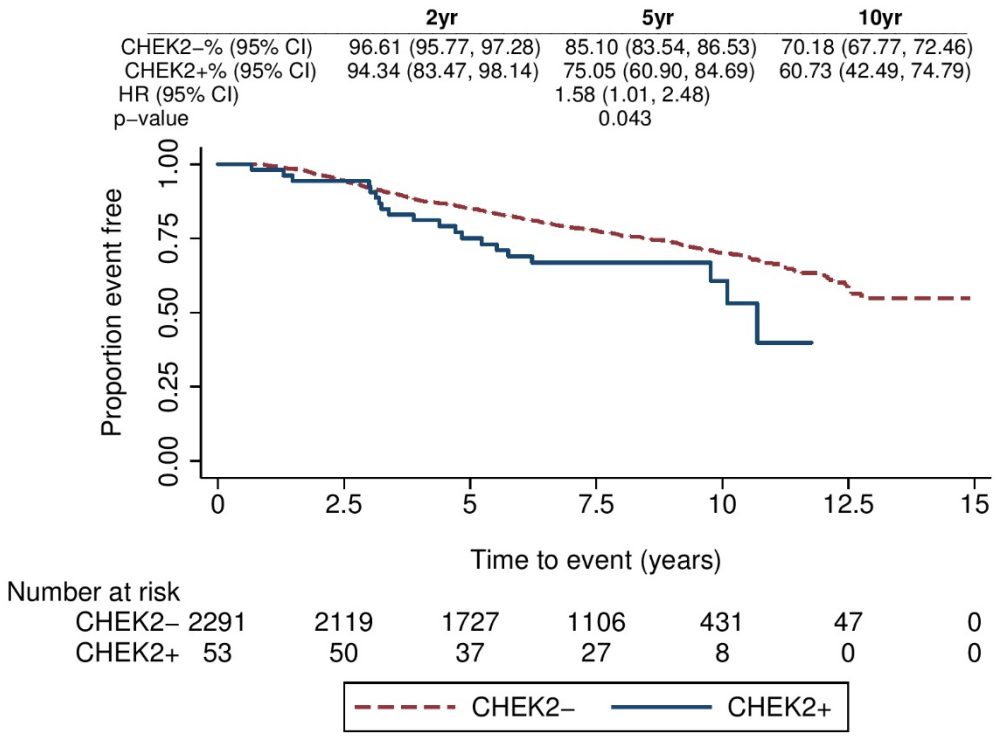


Figure 25: Kaplan Meier Plot of Overall Survival amongst *CHEK2* Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Overall Survival (OS) for *CHEK2* truncating variant carriers versus non-carriers following univariate analysis. Reprinted with permission. © (2020) American Society of Clinical Oncology.

Distant Disease Free Survival (DDFS) was also significantly worse in *CHEK2* mutation carriers versus non-carriers. Univariable analysis demonstrated a HR, 1.62 (95%CI, 1.06-2.48 (p=0.025)) (Figure 26). At 5 years, DDFS was 61.8% (95% CI, 47.2-73.4) amongst *CHEK2* mutation carriers versus 77.7% (95% CI, 75.9-79.4) in non-carriers. At 10 years, DDFS was 56.8% (95% CI, 41.8-69.3) amongst *CHEK2* mutation carriers versus 69.0% (95% CI, 66.7-71.2) in non-carriers. The observed difference in DDFS between *CHEK2* mutation carriers and non-carriers was also maintained after adjustment for known prognostic factors in a multivariable analysis (HR 1.60 (95%CI, 1.04-2.46 (p=0.033)) (Appendix K.1.2).

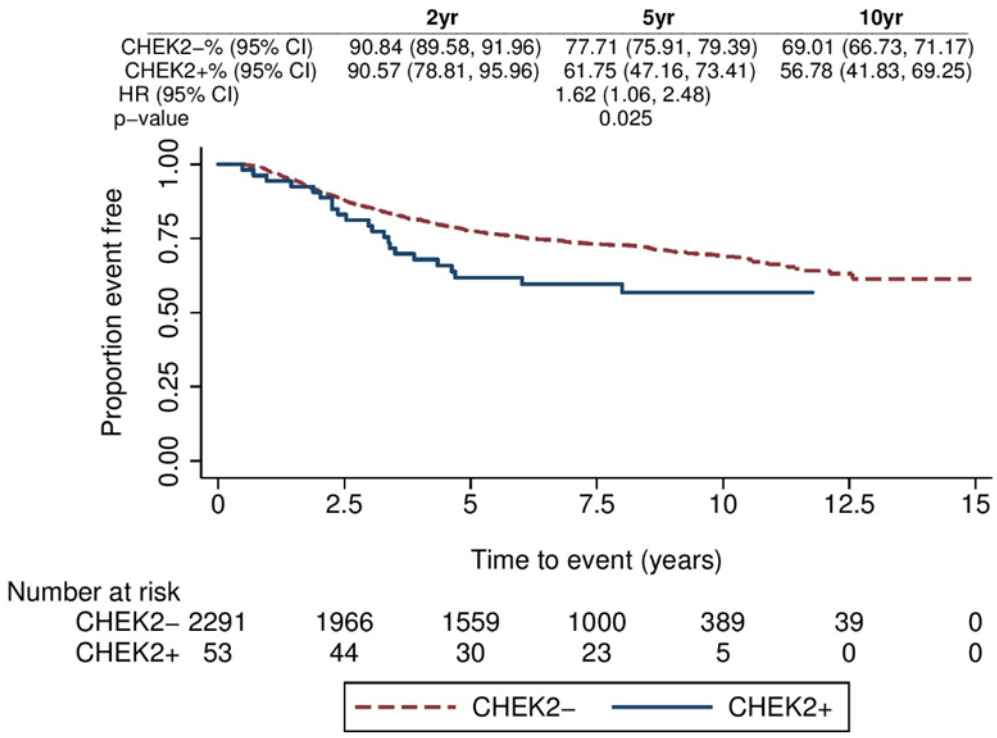


Figure 26: Kaplan Meier Plot of Distant Disease Free Survival amongst *CHEK2* Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Distant Disease Free Survival (DDFS) for *CHEK2* variant carriers versus non-carriers following univariate analysis. Reprinted with permission. © (2020) American Society of Clinical Oncology. All rights reserved.(224)

6.4 Discussion

6.4.1 *CHEK2* Variant Identification

Analyses of 2744 participants within the POSH cohort, has identified that pathogenic or likely pathogenic variants in *CHEK2* are present in 1.9% of unselected early onset breast cancers within a UK population. *CHEK2*, c.1100delC, p.(Thr367fs) is the most frequently identifiable, accounting for 67.9% of all truncating mutations. These figures are consistent with Decker et al. who identified a truncating *CHEK2* variant in 1.6% of unselected breast cancer cases within a large UK population based study.(56) *CHEK2* c.1100delC, p.(Thr367fs) was the most frequent variant accounting for 81% of all pathogenic or likely pathogenic variants.(56)

Within this cohort, family history did not reliably predict the presence of a germline *CHEK2* variant. This is consistent with a low-moderate overall increase in risk compared to population average.

Individuals with a germline *CHEK2* variant and invasive breast cancer were significantly more likely to be obese than non-carriers. It is well recognised that obesity is associated with an increased risk of post-menopausal breast cancer but it has not been associated with an increased risk in pre-menopausal breast cancer.(225) In 2017, The Premenopausal Breast Cancer Collaborative Group assessed the BMI associated breast cancer risk in 758,592 premenopausal women and found an inverse correlation between age and the associated risk. In this study, significantly more cancer patients with a *CHEK2* pathogenic variant were obese at presentation compared to non-carriers suggesting a potential synergistic interaction between *CHEK2* genetic risk and obesity.(5, 7) POSH study data has previously shown that BMI was associated with an adverse prognosis after breast cancer diagnosis, independent of other known risk factors, we therefore included BMI in the multivariable analysis.(226)

6.4.2 *CHEK2* and Tumour Histopathology

This work provides a unique series detailing the histopathological tumour phenotype associated with early onset breast cancer in the context of a *CHEK2* mutation compared with age matched controls. It has shown that early onset, *CHEK2*-associated breast cancers are significantly more likely to be Grade 2, ER and PR-positive with no difference in HER2 expression. They are not associated with a TNT phenotype. These results are consistent with other reports of the histopathological tumour phenotype associated with germline mutations in *CHEK2*.

Decker et al. found that *CHEK2*-associated tumours were significantly more likely to be ER-positive OR=3.42 (95% CI, 2.33 - 5.21 ($p=1.5x10^{-11}$)).(56) Cybulski et al. also found that *CHEK2*-associated cancers were significantly more likely to be oestrogen and progesterone receptor positive (69.7% versus 63.1% (p=0 .002)) and (77% versus 68.7% (p<0.001)) respectively.(90) Weischer et al. also found a significantly higher frequency of oestrogen and progesterone receptor positive breast cancers in c.1100delC carriers than non-carriers (63% versus 57% (p<0.001)) and (46% versus 43% (p=0.01)) respectively.(91) Couch et al. found no *CHEK2* pathogenic variants amongst 1,824 patients presenting with triple negative breast cancer.(157) The strong association of a germline *CHEK2* pathogenic variant with ER-positive disease may support the use of oestrogen receptor blockade such as Tamoxifen as chemoprophylaxis to modulate the *CHEK2* related breast cancer risk.(90)

Invasive breast cancers occurring in the context of a pathogenic *CHEK2* variant demonstrated a trend towards multifocality with significantly higher levels of nodal involvement at presentation. Our study confirmed the findings of Cybulski et al. who found that *CHEK2*-associated cancers demonstrated higher levels of nodal involvement. (90, 158) Multifocal tumour pathology is likely

to be one of the key drivers for more frequent mastectomy rather than breast conserving treatment amongst *CHEK2* carriers. The association between *CHEK2* genotype and multifocality is novel and highlights potential phenotypic similarity with *BRCA2* associated invasive breast cancers.

A further aim of this work was to identify histopathological predictors of *CHEK2* carrier status. Only nodal status was identified as a significant independent predictor of *CHEK2* genotype. Therefore, whilst there is an emerging histopathological *CHEK2* related tumour phenotype, this cannot be reliably used to identify *CHEK2* gene carriers versus non-carriers.

6.4.3 *CHEK2* and Outcome

We have determined that patients with a germline *CHEK2* pathogenic variant who develop breast cancer have an adverse outcome with reduced OS and DDFS compared to those without, a relationship which persists after adjustment for known prognostic factors. We noted that 4/74 individuals (5.7%) removed from the analysis because they presented with M1 disease, carried a pathogenic *CHEK2* variant compared with 74/3095 (2.4%) of the total patients in the POSH study.

Our results are consistent with Schmidt et al. who found that *CHEK2*, c.1100delC mutation carriers had a worse recurrence free and breast cancer specific survival than *CHEK2* non-carriers (HR 1.7 95% CI, 1.2-2.4 (p=0.006)) and (HR 1.4 95% CI, 1.0-2.1 (p=0.072)) respectively but after multi-variable analysis the difference was no longer statistically significant.(80) Wesicher et al. also found that ER-positive *CHEK2* c.1100delC carriers within the BCAC consortium had a significantly increased risk of breast cancer specific death which persisted after multi-variable analysis (HR 1.63 (95% CI, 1.24 to 2.15) p<0.001).(91) They also identified a 2.8 fold risk of a second breast cancer (HR 2.8 (95% CI, 2.00 - 3.83 p<0.001).(91)

The contralateral breast cancer rate amongst *CHEK2* pathogenic variant carriers was almost twice that of non-carriers at both 5 and 10 years. Although the absolute numbers of cases was small, we noted that *CHEK2* carriers with a family history, had a contralateral breast cancer rate more than five times higher than non-carriers, whereas *CHEK2* carriers without a family history had no increase in risk. Within the POSH cohort, family history was not an independent predictor of outcome. (227)

In 2004, De Bock et al. reported that at 5 years, a contralateral breast cancer had developed in 21% of *CHEK2* c.1100delC, p.(Thr367fs) carriers compared to 4% of non-carriers representing an almost 6-fold increase in risk.(228) Decker et al. found that *CHEK2*-associated tumours were significantly more likely to be bilateral at presentation (OR=3.27 (95% CI 1.66 - 5.83)

p=0.0014).(56) This was further supported by Kilpivaara et al. who also noted a strong association with bilateral disease at presentation.(90, 158) Our study notes the presentation with bilateral disease particularly in the context of other risk factors (obesity and family history). This may be reflective of the influence of polygenic factors in guiding outcome in the context of a moderate risk breast cancer susceptibility gene.

The importance of family history and hence polygenic factors in the *CHEK2*-associated breast cancer risk has been demonstrated by a number of studies. In 2011, Cybulski et al. observed identified a cumulative increase in breast cancer risk associated with the number of affected relatives and proximity of that relative to the proband. This equated to absolute lifetime risks of 20% in the absence of a family history, 28% for an affected SDR, 34% for a FDR and 44% if both a FDR and SDR were affected.(90) Weischer et al. also indicated that breast cancer risk was higher in the context of familial breast cancer RR 4.8 (95% CI, 3.3-7.2) equating to a lifetime risk of up to 37% in the presence of a family history.(73)

In 2017, Muranen et al. identified a multiplicative effect of lower penetrance SNPs on *CHEK2*-associated breast cancer risk in patients derived from the BCAC consortium.(79) It is possible to hypothesise that the increased breast cancer risk conferred by polygenic factors may be further applicable to patient outcome.

6.5 Summary

CHEK2 is commonly included in multigene panel testing, and most frequently identified in the context of a patient presenting with breast cancer. Although the numbers are small, we observed a higher likelihood of a contralateral breast cancer associated with a pathogenic *CHEK2* variant. The increased incidence appears to be confined to carriers with a family history of breast cancer. For carriers with no family history, the incidence of contralateral breast cancer is no greater than the incidence in a non-carrier population.

CHEK2 pathogenic variant carriers also presented with tumours more likely to metastasise, manifest as higher nodal involvement and poorer overall survival. This is in contrast to analysis in the same cohort which showed that prognosis was not altered in a multivariable analysis of *BRCA1* or *BRCA2* carriers.(36)

Including *CHEK2* genotyping as part of population risk stratified approaches to inform targeted screening and improve early diagnosis is aspirational. The current approach for managing moderate breast cancer risk within the UK is annual mammograms from the age of 40 years.(229) The use of chemoprophylaxis may be effective given the high proportion of hormone receptor

positive breast cancers.(229) However, neither measure has yet been tested in this particular group of patients.

Our study highlights the importance of including effective measures to address lifestyle risk factors, particularly around maintaining a healthy body weight, for premenopausal women at increased breast cancer risk.

In summary, this work describes the characteristics and clinical outcomes for patients who present with invasive early onset breast cancer and carry a *CHEK2* pathogenic variant. Since a pathogenic *CHEK2* variant is likely to be identified in approximately 2% of Caucasian breast cancers patients, including those aged 40 years or younger, clinicians should be aware of the adverse prognosis and the effect of family history on contralateral cancer risk in planning cancer treatment. Finally, in the context of both healthy population screening, and testing of cancer patients, supportive measures to mitigate risk should include addressing obesity and environmental factors in a multifactorial approach.

Chapter 7 Results: Tumour Histopathological Phenotype amongst Germline *PALB2* Variant Carriers

7.1 Baseline Characteristics of the Cohort

The *PALB2* histopathological analysis has utilised data derived from the POSH study. The complete analysed cohort consisted of 2322 participants who were diagnosed with a primary invasive breast cancer under the age of 40 years (Table 34). This included 31 individuals with a pathogenic or likely pathogenic variant in *PALB2* and 2291 variant negative participants. Variant negative was defined as *CHEK2*, *ATM*, *TP53*, *BRCA1* and *BRCA2* negative.

Most recruits were Caucasian 2124/2295 (92.5%) (n=27 did not record ethnicity). A confirmed *PALB2* pathogenic or likely pathogenic variant was found in 31/2322 (1.3%) of this study cohort (1.1% (31/2744) of the whole POSH cohort). The most frequently encountered *PALB2* variant was the founder c.3113G>A, p.(Trp1038Ter) identified in 14/31 (45.2%) of all carriers (Appendix H).

The median age of cancer onset was 37 years (IQR 33-38 years) for *PALB2* variant carriers, and 37 years (IQR 34-39 years) for non-carriers. The median age of cancer onset was 33 years (IQR 32-36) for individuals with a *PALB2* variant located within the *BRCA1* binding domain and 37 years (IQR 36-39) for individuals with a *PALB2* variant located within the *BRCA2* binding domain.

There was no significant association between a family history of breast cancer and *PALB2* carrier status. However, a higher proportion of *PALB2* variant carriers had a family history of breast cancer (12/27 (44.4%)) compared to variant negative individuals (682/2236 (30.5%)). Consistent with this observation, the median BOADICEA score was significantly higher in *PALB2* variant carriers than non-carriers but below the threshold of 0.10 (10%) utilised for diagnostic genetic testing (median score *PALB2* positive 0.06 (IQR 0.03-0.13) versus variant negative 0.03 (IQR 0.02-0.05) (p=0.036). There was no baseline difference in body mass index between *PALB2* mutation carriers and non-carriers (Table 34).

Characteristic	*PALB2* (*BRCA1* domain)* (n=7)	*PALB2* (*BRCA2* domain)* (n=16)	*PALB2*+ (ALL)** (n=31)	*PALB2*- (n=2291)	p-value† *PALB2*+ (*BRCA1* domain) vs *BRCA2* domain	p-value† *PALB2*+(ALL) vs *PALB2*-
Median age diagnosis (yrs)	33	37	37	37	p=0.073	p=0.57
Range	31-39	29-40	29-40	18-40		
IQR	32-36	36-39	33-38	34-39		
Body Mass Index (Total)	7 (100%)	16 (100%)	31 (100%)	2203 (100%)	p=0.61	p=0.97
Underweight/Healthy (<25)	3 (42.9%)	10 (62.5%)	16 (51.6%)	1185 (53.8%)		
Overweight (25-30)	2 (28.6%)	2 (12.5%)	9 (29.0%)	603 (27.4%)		
Obese (>30)	2 (28.6%)	4 (25.0%)	6 (19.4%)	415 (18.8%)		
Missing	0	0	0	88 (3.8%)		
Ethnicity	7 (100%)	16 (100%)	31 (100%)	2264 (100%)	p=0.11	p=0.36
Caucasian/white	5 (71.4%)	16 (100.0%)	28 (90.3%)	2096 (92.6%)		
Black	1 (14.3%)	0	1 (3.2%)	84 (3.7%)		
Asian	1 (14.3%)	0	1 (3.2%)	70 (3.1%)		
Other	0	0	1 (3.2%)	14 (0.6%)		
Missing	0	0	0	27 (1.2%)		
Family History	7 (100%)	14 (100%)	27 (100%)	2209 (100%)	p=0.15	p=0.11
No	4 (80.0%)	6 (42.9%)	15 (55.6%)	1539 (69.7%)		
Yes	1 (20.0%)	8 (57.1%)	12 (44.4%)	670 (30.3%)		
Missing	2 (28.6%)	2 (12.5%)	4 (12.9%)	82 (3.6%)		
BOADICEA score					p=0.682	p=0.0036
Median	0.06	0.08	0.06	0.03		
Range	0.01 to 0.24,	0.02 to 0.47,	0.01 to 0.65,	0.00 to 0.95,		
IQR	0.02 to 0.12	0.03 to 0.14	0.03 to 0.13	0.02 to 0.05		
Missing	1 (14.3%)	2 (14.3%)	3 (9.7%)	62 (2.7%)		

Table 34: Baseline Characteristics of the *PALB2* Cohort

*The *BRCA1* binding domain is located between amino acids 1 - 319 and the *BRCA2* binding domain occurs in the WD40 motif located between amino acids 853 and 1186.(30) *PALB2* (ALL) incorporates all pathogenic variants identified in *PALB2*. †Assessments of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

7.2 Tumour Histopathology

7.2.1 Baseline Tumour Grade, Size and Focality

Overall, a higher percentage of *PALB2*-associated tumours were grade 3 at presentation compared to non-carriers, however this relationship was non-significant (grade 3 22/31 (71.0%) versus 1278/2229 (57.3%) (p=0.18)). There was no difference in baseline tumour size, focality, nodal involvement or lymphovascular infiltration between *PALB2* variant carriers and non-carriers across the cohort (Table 35 and Figures 27 and 28).

Subgroup analysis which differentiated between pathogenic variants located within the *BRCA1* and *BRCA2* binding domains identified baseline differences in tumour histopathology. Tumours occurring in association with a variant in the *BRCA1* binding domain were significantly smaller than those associated with variants in the *BRCA2* binding domain. Overall 6/7 (85.7%) of tumours occurring in association with a variant in the *BRCA1* binding domain were less than 20mm. In comparison 13/16 (81.2%) of tumours occurring in association with a variant in the *BRCA2* binding domain were greater than 20mm (p=0.008) (Table 35 and Figures 27 and 28).

A higher proportion of tumours occurring in association with *PALB2* variants in the *BRCA1* binding domain were localised (5/6 (83.3%) versus 10/16 (62.5%)). A higher proportion of *BRCA1* binding domain tumours also presented with N0 stage disease (4/7 (57.1%) versus 4/16 (25.0%) (p=0.14)). Significantly lower levels of lymphovascular infiltration were identifiable amongst *BRCA1* binding domain associated tumours. In total, 0/7 (0%) of *BRCA1* binding domain associated tumours demonstrated lymphovascular infiltration versus 9/16 (56.2%) of *BRCA2* binding domain associated tumours (p=0.011) (Table 35 and Figures 27 and 28).

Characteristic	*PALB2* (*BRCA1* domain)* (n=7)	*PALB2* (*BRCA2* domain)* (n=16)	*PALB2*+ (ALL)** (n=31)	*PALB2*- (n=2291)	p-value† *PALB2*+ (*BRCA1* domain) vs *BRCA2* domain	p-value† *PALB2*+(ALL) vs *PALB2*-
Histological Grade (Total)	7 (100%)	16 (100%)	31 (100%)	2229 (100%)	p=0.62	p=0.18
1	0	0	0	148 (6.6%)		
2	1 (14.3%)	5 (31.2%)	9 (29.0%)	803 (36.0%)		
3	6 (85.7%)	11 (68.8%)	22 (71.0%)	1278 (57.3%)		
Missing/not graded	0	0	0	62 (2.7%)		
Focality (Total)	6 (100%)	16 (100%)	30 (100%)	2085 (100%)	p=0.35	p=0.99
Localised	5 (83.3%)	10 (62.5%)	21 (70.0%)	1461 (70.1%)		
Multifocal	1 (16.7%)	6 (37.5%)	9 (30.0%)	624 (29.9%)		
Missing	1 (14.3%)	0	1 (3.2%)	206 (9.0%)		
Max invasive tumour size (total)	7 (100%)	16 (100%)	31 (100%)	2163 (100%)	p=0.008	p=0.59
15mm or less	2 (28.6%)	3 (18.8%)	7 (22.6%)	561 (25.9%)		
>15mm to 20mm	4 (57.1%)	0	5 (16.1%)	403 (18.6%)		
>20mm to 35mm	0	9 (56.2%)	12 (38.7%)	733 (33.9%)		
>35mm to 50mm	1 (14.3%)	3 (18.8%)	6 (19.4%)	269 (12.4%)		
>50mm	0	1 (6.2%)	1 (3.2%)	197 (9.1%)		
Missing	0	0	0	128 (5.6%)		
Pathological N stage (total)	7 (100%)	16 (100%)	31 (100%)	2253 (100%)	p=0.14	p=0.74
N0	4 (57.1%)	4 (25.0%)	14 (45.2%)	1084 (48.1%)		
N1	3 (42.9%)	12 (75.0%)	17 (54.8%)	1169 (51.9%)		
Missing	0	0	0	38 (1.7%)		
Number of positive lymph nodes	7 (100%)	16 (100%)	31 (100%)	2253 (100%)	p=0.29	p=0.081
0	4 (57.1%)	4 (25.0%)	14 (45.2%)	1084 (48.1%)		
1-3	1 (14.3%)	7 (43.8%)	9 (29.0%)	764 (33.9%)		
4-9	2 (28.6%)	5 (31.2%)	8 (25.8%)	273 (12.1%)		
10+	0	0	0	132 (5.9%)		
Missing	0	0	0	38 (1.7%)		
Lymphovascular invasion (total)	7 (100%)	16 (100%)	30 (100%)	2129 (100%)	p=0.011	p=0.22
Absent	7 (100.0%)	7 (43.8%)	19 (63.3%)	1106 (51.9%)		
Present	0	9 (56.2%)	11 (36.7%)	1023 (48.1%)		
Missing	0	0	1 (3.2%)	162 (7.1%)		

Table 35: Baseline Histopathological Characteristics of the *PALB2* Cohort

Comparison of the tumour focality, grade and lymph node involvement between *PALB2* mutation carriers and non-carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort. *The *BRCA1* binding domain is located between amino acids 9 and 44 and the *BRCA2* binding domain occurs in the WD40 motif located between amino acids 853 and 1186. *PALB2* (ALL) incorporates all pathogenic variants identified in *PALB2*.

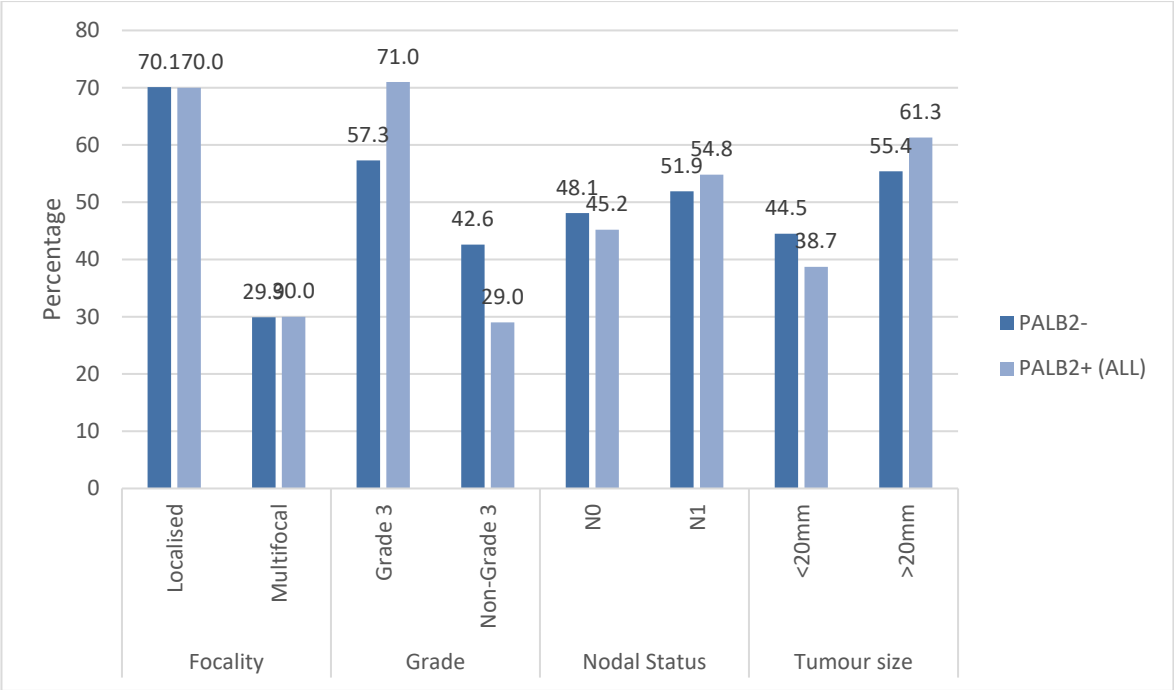


Figure 27: Baseline Histopathological Characteristics of the *PALB2* Cohort

Comparison of the tumour focality, grade, nodal status and tumour size between *PALB2* mutation carriers and non-carriers. Values represented as percentage of the cohort.

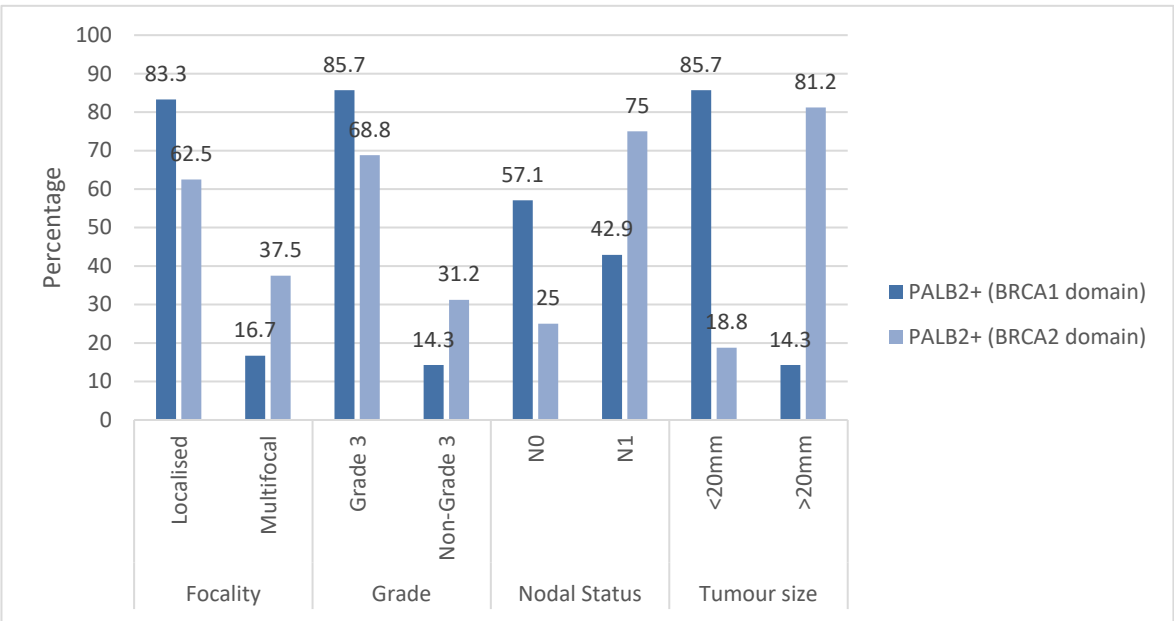


Figure 28: Baseline Histopathological Characteristics of the *PALB2* Cohort - Subgroup Analysis

Comparison of the tumour focality, grade, nodal status and tumour size between *PALB2* mutation carriers with variants in the *BRCA1* and *BRCA2* binding domains. Values represented as percentage of the cohort. *The *BRCA1* binding domain is

located between amino acids 9 and 44 and the *BRCA2* binding domain occurs in the WD40 motif located between amino acids 853 and 1186.

7.2.2 Hormone Receptor Status

There was no significant difference between ER or PR receptor status between *PALB2* mutation carriers and non-carriers. In total, 23/31 (74.2%) of the *PALB2*-associated tumours were ER-positive compared to 1557/2279 (68.3%) of non-carriers (p=0.48) and 16/24 (66.7%) of the *PALB2*-associated tumours were PR-positive compared to 1084/1848 (58.7%) of non-carriers (p=0.43). *PALB2*-associated tumours were significantly more likely to be HER2-negative with 23/26 (88.5%) *PALB2*-associated tumours demonstrating HER2 negativity compared to 1428/2021 (70.7%) of non-carriers (p=0.047) (Table 36, Figure 29 and Figure 30).

There was no significant difference in ER and PR receptor status in the subgroup analysis which compared variants within the *BRCA1* and *BRCA2* binding domains. However, a higher proportion of tumours located in the *BRCA1* binding domain were ER-negative tumours (3/7 (42.9%) versus 4/16 (25.0%)). In comparison, a higher proportion of tumours located in the *BRCA2* binding domain were ER-positive (12/16 (75.0%) versus 4/7 (57.1%)). This observation was not statistically significant (p=0.39).

PALB2 variants located within the *BRCA1* binding domain were significantly more likely to present with HER2-amplified tumours (3/6 (50.0%) versus 0/12 (0%) (p=0.0073)) (Table 36, Figure 29 and Figure 30). There was no association between *PALB2* variant status and the presence of a TNT phenotype. In total, 4/31 (12.9%) of *PALB2*-associated tumours had a TNT compared to 417/2291 (18.2%) of non-*PALB2* tumours (p=0.45) (Table 36, Figure 29 and Figure 30).

Characteristic	*PALB2* (*BRCA1* domain)* (n=7)	*PALB2* (*BRCA2* domain)* (n=16)	*PALB2*+ (ALL)** (n=31)	*PALB2*- (n=2291)	p-value† *PALB2*+ (*BRCA1* domain) vs *BRCA2* domain	p-value† *PALB2*+(ALL) vs *PALB2*-
ER status (total)	7 (100%)	16 (100%)	31 (100%)	2279 (100%)	p=0.39	p=0.48
Negative	3 (42.9%)	4 (25.0%)	8 (25.8%)	722 (31.7%)		
Positive	4 (57.1%)	12 (75.0%)	23 (74.2%)	1557 (68.3%)		
Missing	0	0	0	12 (0.5%)		
HER2 status (total)	6 (100%)	14 (100%)	26 (100%)	2021 (100%)	p=0.004	p=0.047
Negative	3 (50.0%)	14 (100.0%)	23 (88.5%)	1428 (70.7%)		
Positive	3 (50.0%)	0	3 (11.5%)	593 (29.3%)		
Missing	1 (14.3%)	2 (12.5%)	5 (16.1%)	270 (11.8%)		
PR status (total)	6 (100%)	11 (100%)	24 (100%)	1848 (100%)	p=0.59	p=0.43
Negative	3 (50.0%)	4 (36.4%)	8 (33.3%)	764 (41.3%)		
Positive	3 (50.0%)	7 (63.6%)	16 (66.7%)	1084 (58.7%)		
Missing	1 (14.3%)	5 (31.2%)	7 (22.6%)	443 (19.3%)		
TNT status (total)	7 (100%)	16 (100%)	31 (100%)	2291 (100%)	p=0.79	p=0.45
Not TNT	6 (85.7%)	13 (81.2%)	27 (87.1%)	1874 (81.8%)		
TNT	1 (14.3%)	3 (18.8%)	4 (12.9%)	417 (18.2%)		
Missing	0	0	0	0		

Table 36: Baseline Hormone Receptor Status of the *PALB2* Cohort

Comparison of ER, PR and HER2 receptor status between *PALB2* mutation carriers and non-carriers. *The *BRCA1* binding domain is located between amino acids 9 and 44 and the *BRCA2* binding domain occurs in the WD40 motif located between amino acids 853 and 1186. *PALB2* (ALL) incorporates all pathogenic variants identified in *PALB2*. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

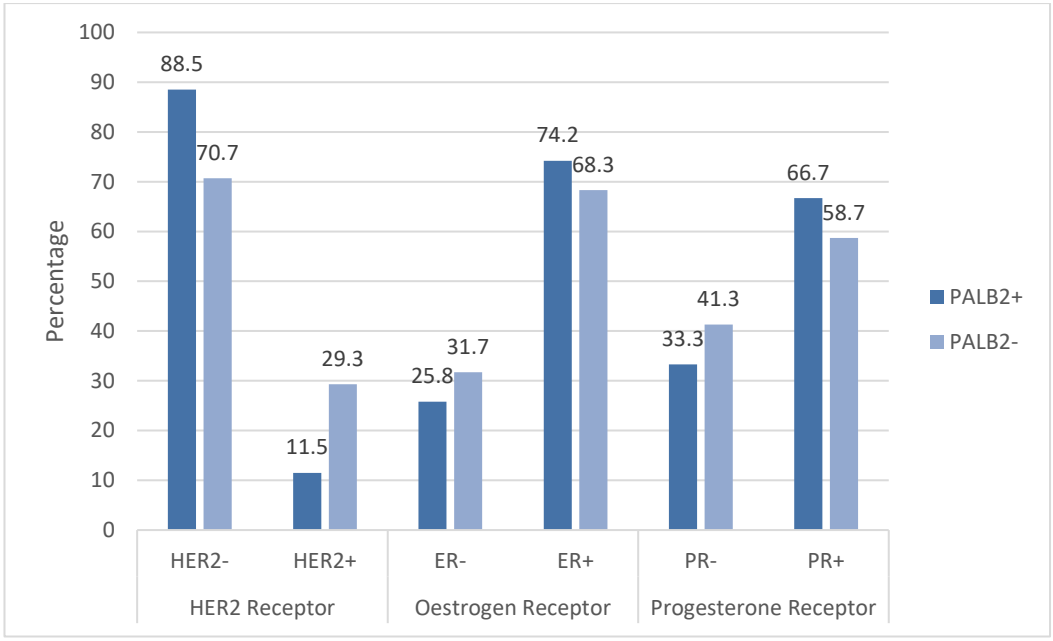


Figure 29: Baseline Hormone Receptor Status of the *PALB2* Cohort

Comparison of ER, PR and HER2 receptor status between *PALB2* mutations carriers and non-carriers. Values represented as percentage of the cohort.

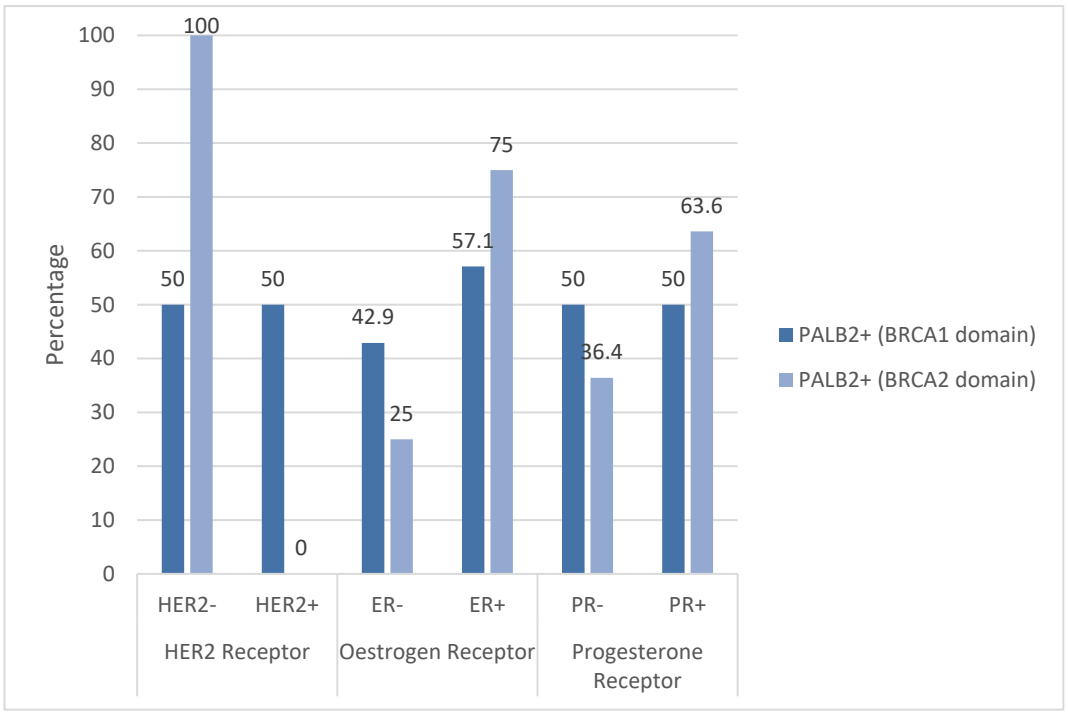


Figure 30: Baseline Hormone Receptor Status of the *PALB2* Cohort - Subgroup Analysis

Comparison of ER, PR and HER2 receptor status between *PALB2* mutation carriers with variants in the *BRCA1* and *BRCA2* binding domains. Values represented as percentage of the cohort. *The *BRCA1* binding domain is located between amino acids 9 and 44 and the *BRCA2* binding domain occurs in the WD40 motif located between amino acids 853 and 1186.

7.2.3 Histopathological Predictors

A model selection process using multiple logistic regression and incorporating forward selection by way of likelihood ratio tests was utilised to determine which histopathological features were better predictors of a *PALB2* germline gene alteration (Table 37). Multiple imputation was utilised as 860/2322 (37%) had missing data. Grade was removed from the analysis due to insufficient numbers. There were no significant independent predictors of germline mutational status. Lymphovascular infiltration, PR and HER2 receptor status contributed towards the prediction of a *PALB2* variant (Table 37).

Overall, the highest probability of identifying a pathogenic *PALB2* variant would be observed amongst those patients with PR-positive and HER2-negative tumours. There were no significant independent predictors of a germline *PALB2* mutation. Subgroup analysis of variants located within the *BRCA1* and *BRCA2* binding domains was not conducted due to insufficient numbers (Table 37).

		PALB2	
Phenotype	**Classification**	**Adjusted odds ratio [95% CI] †**	**p value**
Lymphovascular Infiltration	Absent	1*	-
	Present	0.99 (0.79-1.23)	-
Progesterone Receptor	Negative	1*	-
	Positive	1.19 (0.57-2.45)	-
HER2 Receptor	Negative	1*	-
	Positive	0.32 (0.09-1.05)	-

Table 37: Multiple Logistic Regression Analysis

Multiple logistic regression analysis. †Analyses adjusted for hormone receptor status, tumour size, nodal involvement, hormone receptor status, focality, age at diagnosis and lymphovascular infiltration. *Reference category.

7.3 Outcome

The majority of patients received adjuvant chemotherapy. The most frequent regimen included anthracyclines with or without the additional of taxanes. There were no significant differences in the treatment received between *PALB2* mutation carriers and non-carriers (Table 38). This was irrespective of the binding domain in which the variant was located. However, a higher proportion of *PALB2* carriers with variants located within the *BRCA1* binding domain were managed with breast conserving surgery. In total 6/7 (85.7%) of *PALB2* carriers with variants within the *BRCA1* binding domain received breast conserving surgery as the primary surgical intervention versus 6/16 (37.5%) of *BRCA2* binding domain carriers and 16/31 (51.6%) of all variant carriers (p=0.07).

Treatment	*PALB2* (*BRCA1* domain)* (n=7)	*PALB2* (*BRCA2* domain)* (n=16)	*PALB2*+ (ALL)** (n=31)	*PALB2*- (n=2291)	p-value† *PALB2*+ (*BRCA1* domain) vs *BRCA2* domain	p-value† *PALB2*+(ALL) vs *PALB2*-
Chemotherapy Timing (total)	7 (100%)	16 (100%)	31 (100%)	2291 (100%)	p=0.23	p=0.61
None	1 (14.3%)	0	2 (6.5%)	262 (11.4%)		
Adjuvant	5 (71.4%)	15 (93.8%)	25 (80.6%)	1682 (73.4%)		
Neoadjuvant	1 (14.3%)	1 (6.2%)	4 (12.9%)	347 (15.1%)		
Palliative	0	0	0	0		
Missing	0	0	0	0		
Chemotherapy Regimen (total)	7 (100%)	16 (100%)	31 (100%)	2291 (100%)	p=0.58	p=0.61
None	1 (14.3%)	0	2 (6.5%)	262 (11.4%)		
Anthracyclines	4 (57.1%)	11 (68.8%)	20 (64.5%)	1463 (63.9%)		
Anthracyclines and taxanes	2 (28.6%)	4 (25.0%)	8 (25.8%)	530 (23.1%)		
Taxanes only	0	1 (6.12%)	1 (3.2%)	21 (0.9%)		
Other	0	0	0	15 (0.7%)		
Missing	0	0	0	0		
Surgery Type (total)	7 (100%)	16 (100%)	31 (100%)	2291 (100%)	p=0.07	p=0.96
Breast Conserving Surgery	6 (85.7%)	6 (37.5%)	16 (51.6%)	1149 (50.2%)		
Mastectomy	1 (14.3%)	10 (62.5%)	15 (48.4%)	1122 (49.0%)		
Nodal surgery only	0	0	0	6 (0.3%)		
None	0	0	0	14 (0.6%)		
Missing	0	0	0	0		

Table 38: Treatment Characteristics of the *PALB2* Cohort

Comparison of the treatment protocol in relation to genotype. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort. *The *BRCA1* binding domain is located between amino acids 9 and 44 and the *BRCA2* binding domain occurs in the WD40 motif located between amino acids 853 and 1186.

The median duration of follow up was 8.2 years. No contralateral breast cancers were observed amongst *PALB2* variant carriers at 5 years compared to 62/2291 (2.7%) of non-carriers (Table 39). At 10 years, the rate of contralateral breast cancer was almost 2-fold higher amongst *PALB2* mutation carriers. A contralateral breast cancer was observed in 2/31 (6.5%) of *PALB2* mutation carriers at 10 years compared to 85/2291 (3.7%) of non-carriers. The observed increase in contralateral breast cancers was not elevated in the context of familial breast cancer.

Genotype	5 Years (%)	10 years (%)
PALB2+ (ALL)	0 (0%)	2 (6.5%)
PALB2+ (ALL) and FH-	0 (0%)	1 (6.7%)
PALB2+ (ALL) and FH+	0 (0%)	1 (8.3%)
Mutation- (ALL)	62 (2.7%)	85 (3.7%)
Mutation- (ALL) and FH-	46 (3%)	59 (3.8%)
Mutation- (ALL) and FH+	15 (2.2%)	24 (3.6%)

Table 39: Contralateral Breast Cancer Risk in Association with *PALB2*

Contralateral breast cancer risk at both 5 and 10 years for *PALB2* mutation carriers versus non-carriers.

Univariable analysis identified a trend towards improved OS amongst *PALB2* variant carriers versus non-carriers within the first 5 years after cancer diagnosis (HR, 0.72 (95%CI, 0.3-1.6 (p=0.430)) (Figure 31). This relationship was not significant. At 5 years, OS was 96.3% (95% CI, 76.5-99.5) amongst *PALB2* variant carriers versus 85.1% (95% CI, 83.5-86.5) in non-carriers. At 10 years, OS was 70.9% (95% CI, 39.4-88.1) amongst *PALB2* mutation carriers versus 70.4% (95% CI, 67.8-72.5) in non-carriers. The observed difference in OS between *PALB2* variant carriers and non-carriers was maintained after adjustment for known prognostic factors including age at diagnosis, BMI, maximum invasive size (cm), hormone receptor status, nodal involvement, ethnicity and taxanes in a multivariable analysis but remained non-significant (HR 0.76 (95%CI, 0.3-1.7 (p=0.51)) (Appendix K.2.1). Subgroup analysis was not performed due to low numbers.

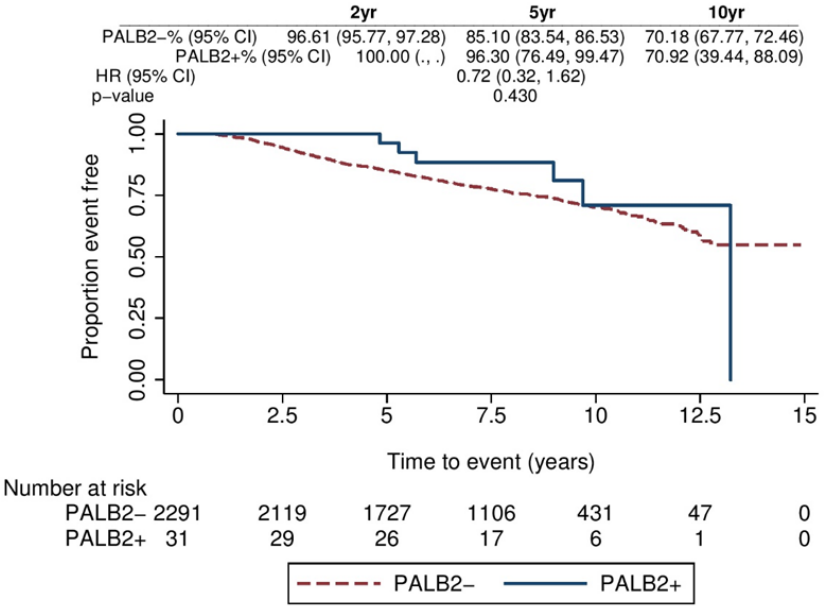


Figure 31: Kaplan Meier Plot of Overall Survival amongst *PALB2* Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Overall Survival (OS) for *PALB2* truncating variant carriers versus non-carriers following univariate analysis.

Univariable analysis identified a possible trend towards improved DDFS amongst *PALB2* variant carriers versus non-carriers (HR, 0.63 (95%CI, 0.3-1.4 (p=0.266))) (Figure 32). However the numbers of *PALB2* carriers are small and the difference is not significant. At 5 years, DDFS was 92.8% (95% CI, 74.3-98.2) amongst *PALB2* mutation carriers versus 77.7% (95% CI, 75.9-79.4) in non-carriers. At 10 years, DDFS was 79.8% (95% CI, 57.4-91.2) amongst *PALB2* mutation carriers versus 68.9% (95% CI, 66.7-71.2) in non-carriers. The observed difference in DDFS between *PALB2* mutation carriers and non-carriers was also maintained after adjustment for known prognostic factors in a multivariable analysis but remained non-significant (HR 0.66 (95%CI, 0.3-1.5) (p=0.32)) (Appendix K.2.2). Subgroup analysis was not performed due to low numbers.

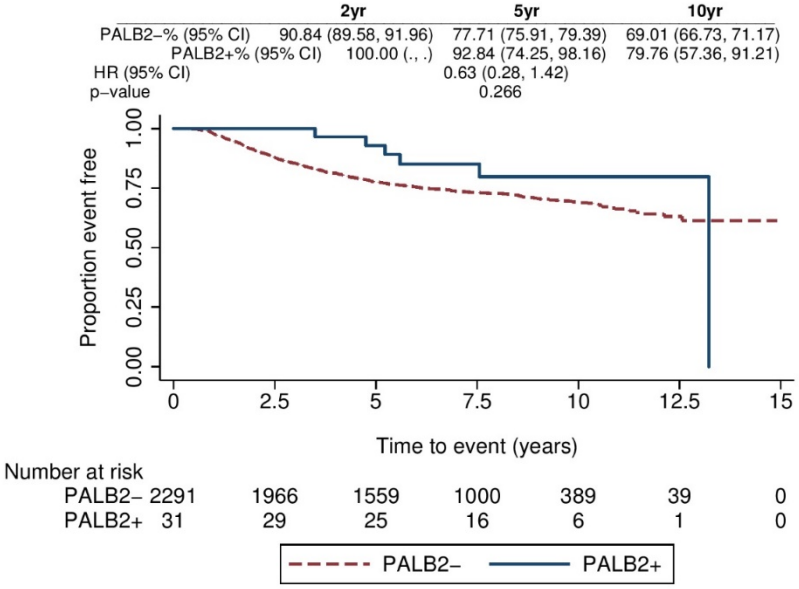


Figure 32: Kaplan Meier Plot of Distant Disease Free Survival amongst *PALB2* Mutations Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Distant Disease Free Survival (DDFS) for *PALB2* variant carriers versus non-carriers following univariate analysis.

7.4 Discussion

7.4.1 Baseline Characteristics of the Cohort

A confirmed *PALB2* pathogenic or likely pathogenic variant was found in 1.3% of this study cohort and 1.1% of the whole POSH cohort. The most frequently encountered *PALB2* variant was the founder mutation c.3113G>A, p.(Trp1038Ter). This is consistent with many previous studies that describe a prevalence of 1-2% depending upon the selection criteria for the cohort. These figures are consistent with Decker et al. who identified a pathogenic variant in *PALB2* amongst 0.68% of

unselected breast cancer cases within a large UK population based study. It is also consistent with the studies described in Chapter 1.4.2.(56)

There was no significant association between a family history of breast cancer and *PALB2* variant carrier status. However, a higher proportion of *PALB2* carriers had a family history of breast cancer compared to mutation negative individuals. Consistent with this observation, the median BOADICEA score was significantly higher in *PALB2* variant carriers than non-carriers but below the threshold for diagnostic genetic testing. This reflects the moderate to high increase in breast cancer risk conferred by pathogenic variants in *PALB2*. (20, 56, 57, 64)

There was no significant difference in the age of onset between *PALB2* variant carriers and non-carriers. This finding is consistent with much of the published literature which demonstrates that there is no difference in the associated risk of *PALB2*-associated breast cancer with age. Statistical modelling by Antoniou et al. found that age specific relative risk models were not significantly better than models that assumed a constant relative risk with age ($p = 0.07$). Cybulski et al. also found no significant difference in the age of onset between *PALB2* mutation carriers and non-carriers.(66) They found the relative risk of breast cancer in women was 3.68 (95%CI 1.84-7.15) under the age of 50 years and 4.90 (2.53-9.49) for those diagnosed after the age of 50 years.(66) Decker et al. also found no clear association between breast cancer risk and age in carriers of rare truncating variants in *PALB2*.(56) In contrast, Yang et al. have published the largest cohort of pathogenic *PALB2* variant carriers derived from 524 unselected families.(20) They identified that under a linear trend model, the relative risk of breast cancer reduces with age from RR 13.10 at age 25 years to RR 4.69 at age 75 years.(20) The ascertainment of participants within the POSH study would preclude the identification of differential pre and post-menopausal relative breast cancer risks.

This work has identified a non-significant trend towards a younger median age at cancer diagnosis amongst individuals with *PALB2* mutations located within the *BRCA1* binding domain. This observation has not been reported previously in the context of *PALB2* mutations. However, it is recognised that peak cancer incidence for *BRCA1* mutation carriers is at an earlier age compared to *BRCA2* gene carriers. In 2017, Kuchenbaecker et al. conducted the largest prospective cohort study of 6036 *BRCA1* and 3820 *BRCA2* mutation carriers.(41) They identified that the peak cancer incidence amongst *BRCA1* mutation carriers occurred between 41-50 years whilst the peak cancer incidence amongst *BRCA2* mutation carriers occurred between 51-60 years.(41) As such, the variation in age of onset between different *BRCA* binding domains within the *PALB2* gene may reflect a domain specific effect.

7.4.2 *PALB2* Variants and Tumour Histopathology

This work provides a unique series detailing the histopathological tumour phenotype associated with early onset breast cancer in the context of a *PALB2* pathogenic variant compared with age matched controls. There was no significant difference between ER or PR receptor status between *PALB2* variant carriers and non-carriers. In addition, there was no association between *PALB2* carrier status and the presence of a TN tumour phenotype.

The ER and PR receptor findings are consistent with the work of Decker et al. They identified 89 *PALB2* mutation carriers from 13, 087 breast cancer cases within a UK cohort and found no significant difference in the presenting ER phenotype (ER-positive OR=4.32 (95% CI 2.07-10.5) versus ER-negative OR=5.58 (95% CI 2.19-15.2) (p=0.55)).(56) Pathogenic variants in *PALB2* spanned the entire gene. This is also comparable to the work of Li et al. in 2019. They analysed 24 *PALB2*-associated invasive breast cancers.(47) It is also comparable to Lee et al. who in 2018 observed the histopathological tumour phenotype in 15 cases of *PALB2*-associated breast cancer.(230) In total, 7/15(46.7%) of *PALB2*-associated tumours were ER-positive and 8/15 (53.3%) were ER-negative.(230)

Subgroup analysis which differentiated between pathogenic variants located within the *BRCA1* and *BRCA2* binding domains identified that a higher proportion of individuals with alterations in the *BRCA1* binding domain presented with ER-negative tumours whilst a higher proportion of individuals with mutations in the *BRCA2* binding domain presented with ER-positive tumours. This relationship was non-significant, however, a possible interaction between the *PALB2 BRCA* binding domain and tumour histopathological phenotype has been described in other published literature.

Cybulski et al. tested 12,529 women with invasive breast cancer for two specific *PALB2* variants located within the *BRCA1* binding domain c.172_175delTTGT (p.Gln60Argfs) and c.509_510del (p.Arg170Ilefs).(66) They identified 116 *PALB2* mutation carriers. Individuals with wild type *PALB2* were significantly more likely to have ER-positive tumours then *PALB2* mutation carriers (70% versus 60% (p=0.031)). The same, and potentially stronger association was observed for Progesterone Receptor (PR) status. 71% of individuals with wild type *PALB2* had a PR-positive tumour compared to 55% of *PALB2* mutation carriers (p=0.0004)).(66) *PALB2* mutation carriers were also significantly more likely to have triple negative tumours (35% versus 14% (p<0.0001)).(66)

Heikkinen et al. observed the histopathological tumour phenotype amongst 27 individuals with *PALB2* c.1592delT.(53) A protein truncating variant proximal to the *BRCA2* binding domain. They found that *PALB2* mutation carriers were significantly more likely to have an ER-negative and PR-negative tumour than non-carriers with familial breast cancer ((ER-negative 46.7% versus 20.9%

(p=0.0008)) and (PR-negative 56.7% versus 33.8% (p=0.0095))).(53) They were also significantly more likely to have TN basal like tumours (31.8% versus 6.9% (p<0.0001).(53)

To an extent, the observed potential association with mutation domain and hormone receptor status is consistent with what is already understood about *BRCA1* and *BRCA2* associated hormone receptor status. *BRCA1* associated tumours are classically triple negative or ER-negative whilst *BRCA2* associated tumours are classically ER, PR-positive and HER2-negative.(219)

However, the potential association between hormone receptor status and *BRCA* binding domains is not reproducible across all *PALB2* studies. This may be reflective of an age related effect on hormone receptor status. In 2012, Mavaddat et al. observed the tumour histopathology in 3797 *BRCA1* mutation carriers and 2392 *BRCA2* mutation carriers.(219) They found that the proportion of ER-negative breast tumours significantly decreased with age at diagnosis amongst *BRCA1* gene carriers but increased with age at diagnosis among *BRCA2* carriers (p-trend=1.2×10−5 and p-trend=6.8×10−6 respectively).(219) They also found that the proportion of triple negative tumours decreased with age at diagnosis in *BRCA1* carriers.(219) The opposite was observed amongst *BRCA2* gene carriers.(219)

PALB2 c.3113G>A was over-represented within the POSH cohort and comprised the majority of mutations within the *BRCA2* binding domain. It is possible that the observed difference in the *BRCA1* and *BRCA2* binding domain histopathology is a variant specific (c.3113G>A) effect. The *PALB2* protein has a coiled-coil motif which is integral to the heterodimerisation and interaction with *BRCA1*.(46) It also has a WD40 domain which interacts with *BRCA2*.(46) Both the coiled-coil motif and WD40 β-propeller additionally interact with RAD51. (48) The WD40 domain has a β-propeller structure composed of several repeats of 40-60 amino acid residues which mask a Nuclear Export Signal (NES).(46) In 2017, Pauty et al. observed the functional effects of four protein truncating variants distributed across the *PALB2* gene (p.Arg170fs, p.Leu531fs, p.Gln775* and p.Trp1038*) with immunofluorescence.(48) They identified that c.3113G>A, p.(Trp1038*) was associated with a significant increase in cytoplasmic localisation and reduced interaction with *BRCA2* and RAD51.(48) The other protein truncating variants displayed predominant nuclear localisation of the protein.(48) They further demonstrated that any variant located in the WD40 repeat domain between amino acids 853-1186 produced the same mislocalisation effect.(48) However truncating variants proximal to and including amino acid 852 resulting in a reversion to nuclear localisation.(48)

Tumours associated with *PALB2* variants located within the *BRCA1* binding domain were significantly more likely to be HER2-amplified. In 2019, Li et al. conducted an analysis of 24 invasive breast cancers derived from 24 *PALB2* germline mutation carriers 1/24 tumours displayed HER2

receptor amplification and this was within the *BRCA2* binding domain.(47) In 2018 Lee et al. observed the HER2 receptor status in 15 *PALB2*-associated breast cancers and found that 4/15 were HER2-amplified.(230) The observed mutations were distributed between the *BRCA1* and *BRCA2* binding domains. From the current literature it is unclear whether HER2 amplification is more or less frequently observed in *PALB2* gene carriers than non-carriers.

A higher percentage of *PALB2*-associated tumours within this study cohort were grade 3 at presentation compared to non-carriers Whilst this relationship is non-significant, high grade at presentation is consistent with other tumours that occur in association with a germline gene alteration such as *BRCA1* or *BRCA2*.(41) There was no difference in baseline tumour size, focality, nodal involvement or lymphovascular infiltration between *PALB2* mutation carriers and non-carriers.

Subgroup analysis which differentiated between pathogenic variants located within the *BRCA1* and *BRCA2* binding domains identified baseline differences in tumour histopathology. Tumours occurring in association with a mutation in the *BRCA1* binding domain were significantly smaller than those associated with mutations in the *BRCA2* binding domain. A higher proportion of tumours occurring in association with *PALB2* mutations in the *BRCA1* binding domain were also localised with lower levels of nodal involvement and significantly lower levels of lymphovascular infiltration.

There is a paucity of published literature to compare these observed baseline histopathological characteristics between *PALB2* variant carriers and non-carriers. However, the observed potential association with mutation domain and tumour histopathology is consistent with the *BRCA1* and *BRCA2* associated histopathology observed within this prospective cohort including smaller more localised tumours being observed in the context of a *BRCA1* mutation.

7.4.3 *PALB2* Variants and Outcome

Univariable analysis identified a non-significant trend towards improved OS amongst *PALB2* mutation carriers versus non-carriers within the first 5 years after cancer diagnosis although the numbers are small. Subgroup analysis was not possible due to the small sample size. *PALB2* carriers with variants located within the *BRCA1* binding domain were significantly more likely to be managed with breast conserving surgery which may be reflective of the trend towards smaller, more localised tumours observed within this sub-group.

There is limited available data regarding the prospective cancer outcome amongst individuals with a germline *PALB2* mutation. It has been suggested that *PALB2*-associated tumours may display a more aggressive tumour phenotype with a higher proliferation index (Ki67) and grade at

presentation.(46) Cybulski et al. have provided the largest study looking at prospective outcomes in this group of patients. They found the crude 10 year survival to be significantly lower in *PALB2* carriers compared to normal population controls (48.0% (95%CI 36.5-63.2)) versus (74.7% (95%CI 73.5-75.8) ($p<0.0001$)).(66) This survival disparity was also present and significant at 5 years but not to the same magnitude.(66) However, *PALB2* gene carriers with a breast cancer smaller than 2 cm had a 10-year survival of 82·4% (95% CI, 66·0–100·0%), compared with 32·4% (95%CI, 20·2–52·2%) for women with *PALB2*-associated cancers that were 2·0–4·9 cm in diameter.

7.5 Summary

This study represents one of the largest prospective cohorts observing *PALB2*-associated histopathology in early onset breast cancer. It has demonstrated a potential domain specific effect on tumour histopathology with mutations in the *BRCA1* binding domain producing a more "*BRCA1* like" tumour compared to *BRCA2* binding domain mutations producing a more "*BRCA2* like" tumour. Whilst the observed domain specific effect is novel, the observed *BRCA1* and *BRCA2* associated histopathological features are well characterised.

The current literature reporting histopathological tumour phenotype in association with pathogenic variants in *PALB2* is inconsistent as are reports of associated primary tumour sites. The identification of domain specific histopathological differences in *PALB2*-associated tumours in this work raises the question of whether there may also be domain specific cancer risks, in particular for those cancers recognised to occur in association with *BRCA2* gene alterations such as prostate and pancreatic cancer. It may explain some of the discordance between the non-breast cancer risks currently reported in association with *PALB2* mutations and suggest a potential area for meta-analysis of international pooled data.

Chapter 8 Results: Tumour Histopathological Phenotype amongst Germline *ATM* Variant Carriers

8.1 Tumour Histopathology

8.1.1 Baseline Characteristics of the Cohort

The histopathological analysis has utilised data derived from the POSH study. The complete analysed cohort consisted of 2344 participants. This included 23 individuals with a pathogenic or likely pathogenic variant in *ATM* and 2291 variant negative participants. Variant negative was defined as being *CHEK2*, *PALB2*, *TP53*, *BRCA1* and *BRCA2* negative.

Most recruits were Caucasian 2118/2287 (92.6%) There was missing ethnicity data for 27 individuals within this analysis. A pathogenic or likely pathogenic variant in *ATM* was found in 23/2314 (1.0%) of this study cohort (23/2744 (0.8%) of the whole cohort). The higher risk *ATM* founder mutation, *ATM,* c.7271T>G, p.(Val2424Gly) was identified in 5/23 (21.7%) of all participants (Table 40)(Appendix H). We additionally identified 13 individuals with VUS in *ATM* (Appendix I). As such 13/36 (36.1%) variants identified had uncertain clinical utility.

The median age of cancer onset was 38 years (IQR 35-39 years) for *ATM* variant carriers, and 37 years (IQR 34-39 years) for non-carriers. The majority of *ATM* carriers were Caucasian 22/23 (95.7%). *ATM* variant carriers were significantly more likely to have a family history of breast cancer (p=0.0036). In total (13/22 (59.1%)) of *ATM* carriers had a family history of breast cancer compared to 670/2291 (30.3%) of variant negative individuals. Consistent with this observation, the median BOADICEA score was higher in *ATM* variant carriers than non-carriers. However, it was below the threshold of 0.10 for diagnostic genetic testing (median score *ATM* positive 0.05 (IQR 0.02-0.18) versus variant negative 0.03 (IQR 0.02-0.05) (p=0.14). There was no baseline difference in body mass index between *ATM* variant carriers and non-carriers (Table 40).

Characteristic		*ATM*+ (n=23)	*ATM*- (n=2291)	p-value† *ATM*+ vs *ATM*-
Median age diagnosis (yrs)		38	37	p=0.19
	Range	29-40	18-40	
	IQR	35-39	34-39	
Body Mass Index (Total)		23 (100%)	2203 (100%)	p=0.97
	Underweight/Healthy (<25)	13 (56.5%)	1185 (53.8%)	
	Overweight (25-30)	6 (26.1%)	603 (27.4%)	
	Obese (>30)	4 (17.4%)	415 (18.8%)	
	Missing	0	88 (3.8%)	
Ethnicity (Total)		23 (100%)	2264 (100%)	p=0.83
	Caucasian/white	22 (95.7%)	2096 (92.6%)	
	Black	1 (4.3%)	84 (3.7%)	
	Asian	0	70 (3.1%)	
	Other	0	14 (0.6%)	
	Missing	0	27 (1.2%)	
Family History (Total)		22 (100%)	2209 (100%)	p=0.0036
	No	9 (40.9%)	1539 (69.7%)	
	Yes	13 (59.1%)	670 (30.3%)	
	Missing	1 (4.3%)	82 (3.6%)	
BOADICEA score				p=0.14
	Median	0.05	0.03	
	Range	0.01 to 0.67,	0.00 to 0.95,	
	IQR	0.02 to 0.18	0.02 to 0.05	
	Missing	1 (4.3%)	62 (2.7%)	

Table 40: Baseline Characteristics of the *ATM* Cohort

†Assessments of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

8.1.2 Baseline Tumour Grade, Size and Focality

There was no significant difference in baseline tumour grade, size, focality, nodal involvement or lymphovascular infiltration between *ATM* variant carriers and non-carriers across the cohort (Table 41 and Figure 33).

Characteristic		*ATM*+ (n=23)	*ATM*- (n=2291)	p-value† *ATM*+ vs *ATM*-
Histological Grade (Total)		23 (100%)	2229 (100%)	p=0.29
	1	3 (13.0%)	148 (6.6%)	
	2	10 (43.5%)	803 (36.0%)	
	3	10 (43.5%)	1278 (57.3%)	
	Missing/not graded	0	62 (2.7%)	
Focality (Total)		22 (100%)	2085 (100%)	p=0.85
	Localised	15 (68.2%)	1461 (70.1%)	
	Multifocal	7 (31.8%)	624 (29.9%)	
	Missing	1 (4.3%)	206 (9.0%)	
Max invasive tumour size (total)		21 (100%)	2163 (100%)	p=0.42
	15mm or less	9 (42.9%)	561 (25.9%)	
	>15mm to 20mm	3 (14.3%)	403 (18.6%)	
	>20mm to 35mm	7 (33.3%)	733 (33.9%)	
	>35mm to 50mm	1 (4.8%)	269 (12.4%)	
	>50mm	1 (4.8%)	197 (9.1%)	
	Missing	2 (8.7%)	128 (5.6%)	
Pathological N stage (total)		22 (100%)	2253 (100%)	p=0.80
	N0	10 (45.5%)	1084 (48.1%)	
	N1	12 (54.5%)	1169 (51.9%)	
	Missing	1 (4.3%)	38 (1.7%)	
Number of positive lymph nodes		22 (100%)	2253 (100%)	p=0.57
	0	10 (45.5%)	1084 (48.1%)	
	1-3	8 (36.4%)	764 (33.9%)	
	4-9	4 (18.2%)	273 (12.1%)	
	10+	0	132 (5.9%)	
	Missing	1 (4.3%)	38 (1.7%)	
Lymphovascular invasion (total)		22 (100%)	2129 (100%)	p=0.50
	Absent	13 (59.1%)	1106 (51.9%)	
	Present	9 (40.9%)	1023 (48.1%)	
	Missing	1 (4.3%)	162 (7.1%)	

Table 41: Baseline Histopathological Characteristics of the *ATM* Cohort

Comparison of the tumour focality, grade and lymph node involvement between *ATM* pathogenic variant carriers and non-carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

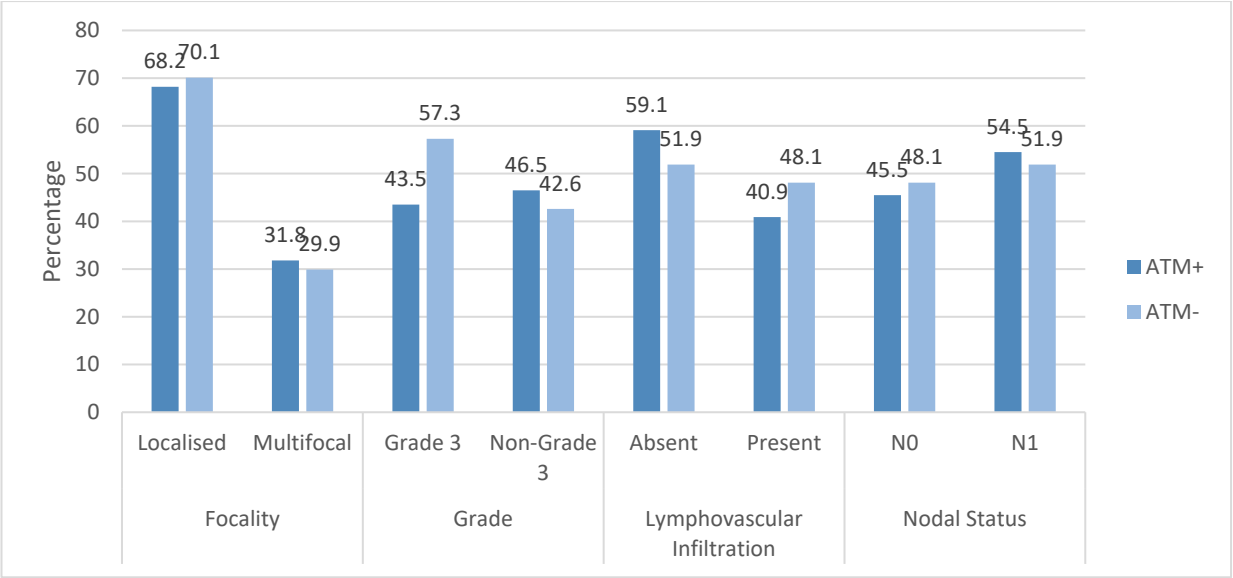


Figure 33: Baseline Histopathological Characteristics of the *ATM* Cohort

Comparison of the tumour focality, grade, lymph node involvement and tumour size between *ATM* mutation carriers and non-carriers. Values represented as percentage of the cohort.

8.1.3 Hormone Receptor Status

ATM-associated tumours were significantly more likely to be ER+ with 22/23 (95.7%) displaying ER+ tumours compared to 1557/2279 (68.3%) of non-carriers (p=0.0050) (Table 42, and Figure 34). There were no other significant difference between PR or HER2 receptor status between *ATM* variant carriers and non-carriers. However *ATM*-associated tumours were significantly less likely to have a TN tumour phenotype (p=0.024). In total 0/23 (0%) of the *ATM*-associated tumours had a TN phenotype compared to 1874/2291 (81.8%) of non-*TP53* tumours (Table 42 and figure 34).

Characteristic		*ATM*+ (n=23)	*ATM*- (n=2291)	p-value† *ATM*+ vs *ATM*-
ER status (total)		23 (100%)	2279 (100%)	p=0.0050
	Negative	1 (4.3%)	722 (31.7%)	
	Positive	22 (95.7%)	1557 (68.3%)	
	Missing	0	12 (0.5%)	
HER2 status (total)		20 (100%)	2021 (100%)	p=0.13
	Negative	11 (55.0%)	1428 (70.7%)	
	Positive	9 (45.0%)	593 (29.3%)	
	Missing	3 (13.0%)	270 (11.8%)	
PR status (total)		20 (100%)	1848 (100%)	p=0.14
	Negative	5 (25.0%)	764 (41.3%)	
	Positive	15 (75.0%)	1084 (58.7%)	
	Missing	3 (13.0%)	443 (19.3%)	
TNT status (total)		23 (100.0%)	2291 (100%)	p=0.024
	Not TNT	23 (100.0%)	1874 (81.8%)	
	TNT	0	417 (18.2%)	
	Missing	0	0	

Table 42: Baseline Hormone Receptor Status of the *ATM* Cohort

†Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

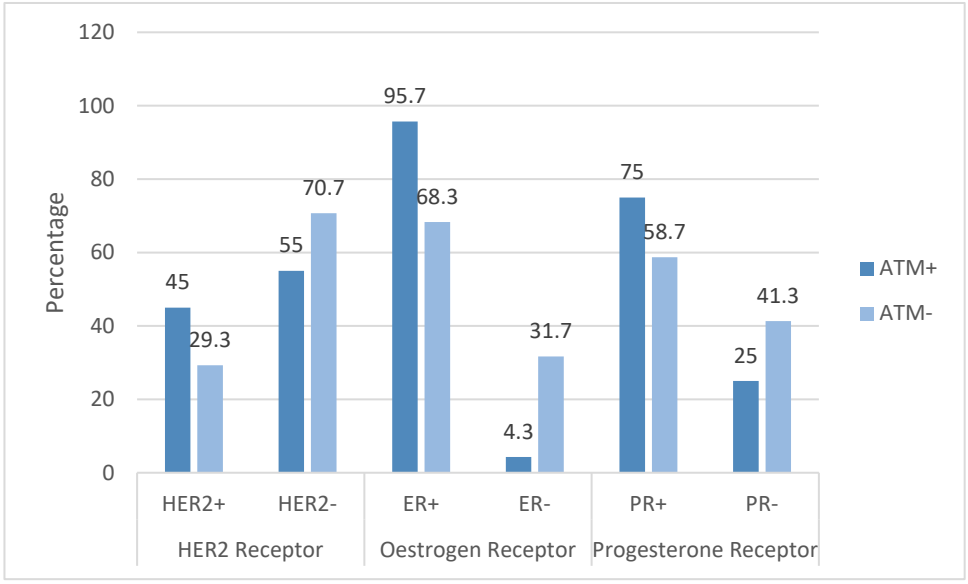


Figure 34: Baseline Hormone Receptor Status of the *PALB2* Cohort

Comparison of ER, PR and HER2 receptor status between *PALB2* mutations carriers and non-carriers. Values represented as percentage of the cohort.

8.1.4 Histopathological Predictors of Germline Genotype

A model selection process using multiple logistic regression and incorporating forward selection by way of likelihood ratio tests was utilised to determine which histopathological features were better predictors of a *ATM* germline gene alteration (Table 43). Multiple imputation was utilised as 855/2314 (37%) had missing data. This analysis identified that only oestrogen receptor status was a significant independent predictor of germline mutational status, ER+ OR = 10.21 (95% CI, 1.41– 74.25). Overall, the highest probability of identifying a pathogenic *ATM* variant would be observed amongst those patients with smaller, ER-positive and HER2 positive tumours.

		CHEK2	
Phenotype	**Classification**	**Adjusted odds ratio [95% CI] †**	**Significant Factor**
Age at diagnosis (years)	-	1.00 (0.97-1.04)	-
Tumour size (cm)	-	0.97 (0.80-1.18)	-
Oestrogen Receptor	Negative	1*	-
	Positive	10.21 (1.41-74.25)	Yes
HER2 Receptor	Negative	1*	-
	Positive	0.98 (0.72-1.33)	-

Table 43: Multiple Logistic Regression Analysis

Multiple logistic regression analysis. †Analyses adjusted for hormone receptor status, tumour size, grade, nodal involvement, hormone receptor status, focality, age at diagnosis and lymphovascular infiltration. *Reference category.

8.2 Outcome

The majority of patients received adjuvant chemotherapy. The most frequent regimen included anthracyclines with or without the addition of taxanes (Table 44). There were no significant differences in the baseline treatment received between *ATM* gene carriers and non-carriers. In total, 18/23 *ATM* gene carriers received adjuvant radiotherapy.

Treatment		*ATM*+ (n=23)	*ATM*- (n=2291)	p-value† *ATM*+ vs *ATM*-
Chemotherapy Timing (total)		23 (100%)	2291 (100%)	P=0.26
	None	5 (21.7%)	262 (11.4%)	
	Adjuvant	14 (60.9%)	1682 (73.4%)	
	Neoadjuvant	4 (17.4%)	347 (15.1%)	
	Palliative	0	0	
	Missing	0	0	
Chemotherapy Regimen (total)		23 (100%)	2291 (100%)	p=0.61
	None	5 (21.7%)	262 (11.4%)	
	Anthracyclines	13 (56.5%)	1463 (63.9%)	
	Anthracyclines and taxanes	5 (21.7%)	530 (23.1%)	
	Taxanes only	0	21 (0.9%)	
	Other	0	15 (0.7%)	
	Missing	0	0	
Surgery Type (total)		23 (100%)	2291 (100%)	p=0.27
	Breast Conserving Surgery	7 (30.4%)	1149 (50.2%)	
	Mastectomy	16 (69.6%)	1122 (49.0%)	
	Nodal surgery only	0	6 (0.3%)	
	None	0	14 (0.6%)	
	Missing	0	0	

Table 44: Treatment Characteristics of the *ATM* Cohort

Comparison of the treatment protocol in relation to genotype. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables.

The median duration of follow up was 8.2 years. Contralateral breast cancers were not more frequently observed in *ATM* gene carriers compared to non-carriers (Table 45). A contralateral breast cancer was observed in 1/23 (4.3%) of *ATM*, carriers at 10 years compared to 59/2291 (3.8%) of non-carriers.

Genotype	5 Years (%)	10 years (%)
ATM+ (ALL)	1 (4.3%)	1 (4.3%)
ATM+ (ALL) and FH-	0 (0%)	0 (0%)
ATM+ (ALL) and FH+	1 (7.7%)	1 (7.7%)
Mutation- (ALL)	62 (2.7%)	85 (3.7%)
Mutation- (ALL) and FH-	46 (3%)	59 (3.8%)
Mutation- (ALL) and FH+	15 (2.2%)	24 (3.6%)

Table 45: Contralateral Breast Cancer Risk in Association with *ATM*

Contralateral breast cancer risk at both 5 and 10 years for *ATM* mutation carriers versus non-carriers.

Univariable analysis identified a trend towards improved OS amongst *ATM* mutation carriers versus non-carriers within the first 5 years following cancer diagnosis (HR, 0.58 (95%CI, 0.22-1.56 (p=0.281)) (Figure 35). This relationship was non-significant. At 5 years, OS was 90.9% (95% CI, 68.1-97.6) amongst *ATM* mutation carriers versus 85.1% (95% CI, 83.5-86.5) in non-carriers. At 10 years, OS was 80.77% (95% CI, 56.4-92.4) amongst *ATM* mutation carriers versus 70.2% (95% CI, 67.8-72.5) in non-carriers. Following adjustment for known prognostic factors including age at diagnosis, BMI, maximum invasive size (cm), hormone receptor status, nodal involvement, ethnicity and taxanes in a multivariable analysis, this remained unchanged (HR 0.66 (95%CI, 0.25-1.77 (p=0.41)) (Appendix K.3.1).

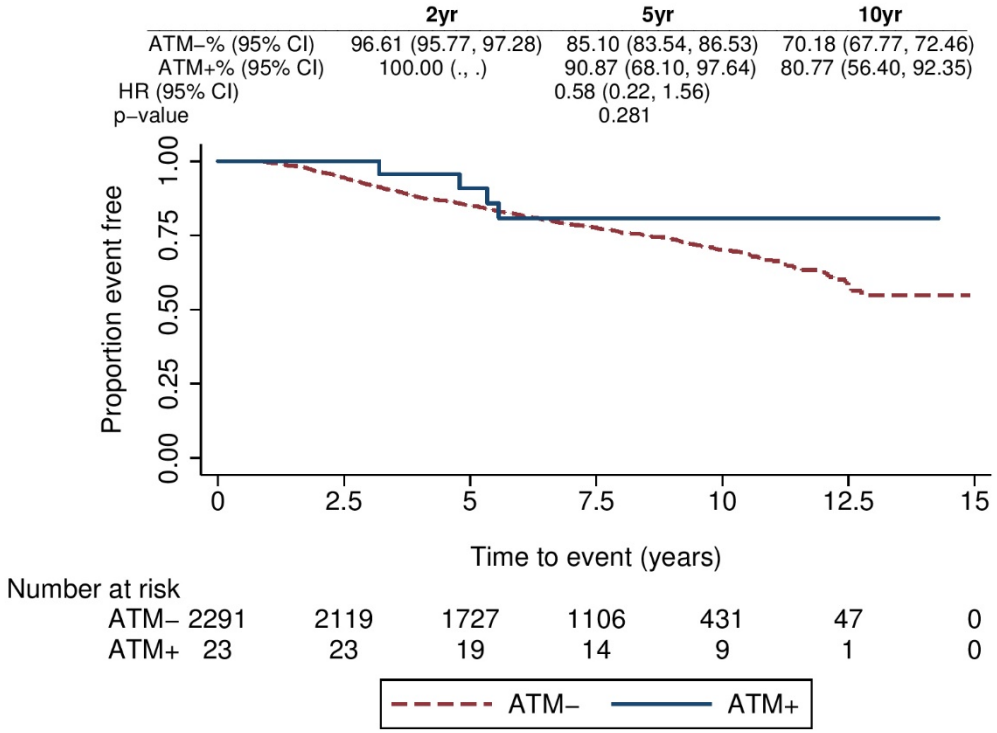


Figure 35: Kaplan Meier Plot of Overall Survival amongst *ATM* Mutation Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Overall Survival (OS) for *ATM* mutation carriers versus non-carriers following univariable analysis.

Univariable analysis identified a trend towards improved DDFS amongst *ATM* mutation carriers versus non-carriers (HR, 0.58 (95%CI, 0.22-1.56 (p=0.281))) (Figure 36). At 5 years, DDFS was 90.9% (95% CI, 68.1-97.6) amongst *ATM* mutation carriers versus 85.1% (95% CI, 83.5-86.5) in non-carriers. At 10 years, DDFS was 80.8% (95% CI, 56.4-92.4) amongst *ATM* mutation carriers versus 70.2% (95% CI, 66.8-72.5) in non-carriers. Following adjustment for known prognostic factors including age at diagnosis, BMI, maximum invasive size (cm), hormone receptor status,

nodal involvement, ethnicity and taxanes in a multivariable analysis, this remained unchanged (HR 0.58 (95%CI, 0.22-1.57 (p=0.29)) (Appendix K.3.2).

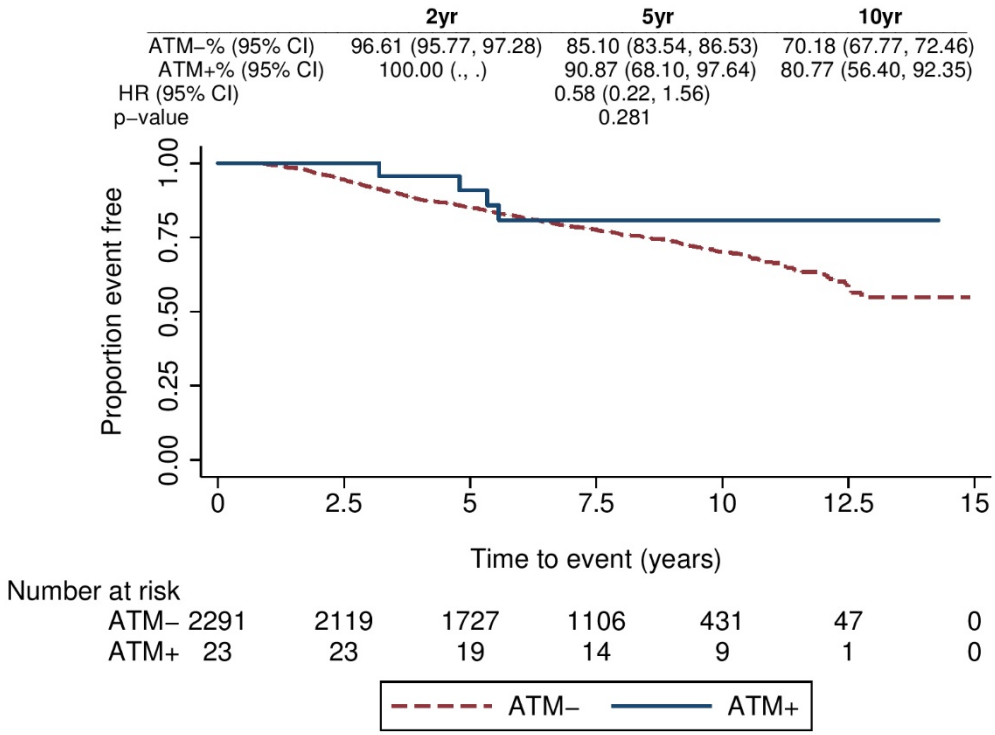


Figure 36: Kaplan Meier Plot of Distant Disease Free Survival amongst *ATM* Mutations Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Distant Disease Free Survival (DDFS) for *ATM* mutation carriers versus non-carriers following univariable analysis.

8.3 Discussion

8.3.1 Baseline Characteristics of the Cohort

A pathogenic variant in *ATM* was present in 1.0% of this study cohort and 0.8% of the whole cohort. The higher penetrance *ATM*, c.7271T<G (p.Val2424Gly) represented approximately 20% of all pathogenic *ATM* variants. This is consistent with the published literature which estimates that the population frequency of *ATM* heterozygotes is 0.35-1% and this may be higher in the context of familial breast cancer.(95, 99) In 2015, Mangone et al. tested 100 unselected individuals with breast cancer and 100 controls within the Brazilian Population. They identified 7 potentially pathogenic variants in *ATM*.(99) In 2019, Yang et al. tested 7657 unselected, *BRCA* negative individuals derived from a Chinese population for pathogenic variants in *ATM*.(231) Pathogenic variants were identifiable in 30/7657 (0.4%) of the cohort.(231)

It is notable that almost 40% of germline variants identified within the *ATM* gene in this study had unclear clinical significance. *ATM* is a large gene consisting of 66 exons and whilst the pathogenicity of protein truncating variants is interpretable, the association between *ATM* missense variants and breast cancer risk is more variable.(56, 93)

In 2016, Marabelli et al. performed a meta-analysis utilising 19 papers defining *ATM*-associated breast cancer risks.(28) The sample populations were heterogeneous incorporating both sporadic and familial breast cancers cohorts derived from several different global centres.(28) They determined the cumulative risk of breast to be 6.02% by the age of 50 years (95% CI=4.58%–7.42%) and 32.83% by the age 80 years (95% CI = 24.55%–40.43%) consistent with moderate risk susceptibility.(28) However, in 2016, the COGS study found that the *ATM*-associated relative risk may be higher for specific variants such as c.7271T>G OR 11.0 (95% CI 1.42-85.7) p=0.0012.(93) This is an observation that holds across multiple studies.(27, 100) This particular elevation in risk may be secondary to a dominant negative effect.(100) It highlights some of the challenges in sequencing *ATM* to identify heritable cancer susceptibility and stratify risk through mainstream diagnostic or population based testing. In support of this, the UK Cancer Genetics Group published guidance on breast cancer susceptibility gene testing in 2018.(143) They recommended that only protein truncating variants in *ATM* should be reported with the exception of c.7271T<G (p.Val2424Gly).(143)

Individuals with a pathogenic variant in *ATM* were significantly more likely to have a family history of breast cancer. However, the median BOADICEA score was below the threshold for diagnostic genetic testing which is consistent with moderate risk susceptibility.

8.3.2 *ATM* Variants and Tumour Histopathology

This work has shown that *ATM*-associated breast cancers are significantly more likely to be ER-positive and that ER positivity is a significant independent predictor of germline mutational status. There were no additional observed histopathological associations. There is a paucity of data regarding the *ATM*-associated histopathological tumour phenotype. In 2006, Balleine et al. described the tumour phenotype of 21 breast cancers occurring in association with a pathogenic variant in *ATM*.(152) They did not observe a clear *ATM*-associated tumour histopathological phenotype.(152) In 2017, Decker et al. found that *ATM*-associated tumours were more likely to be ER-positive than non-carriers.(56) There was however, no significant difference in the risk of ER-positive versus ER-negative disease. (ER-positive (OR=3.42 (95% CI 2.33 – 5.21) versus (ER-negative (OR=1.59 (95% CI 0.80 – 3.00) (p_{diff}=0.11)).(56) In 2019, Yang et al. described the histopathological tumour phenotype of 30 individuals with pathogenic variants in *ATM*

comparison to 7627 non-carriers.(231) They found that breast cancer occurring in association with a pathogenic variant in *ATM* was significantly more like to be ER-positive (93.1% versus 717.7% p=0.011), PR-positive (828.8% versus 64.5% p=0.040) and demonstrate lymph node invasion (44.8% versus 27.2% (p=0.034).(231) Whilst there is a paucity of literature, it demonstrates that tumour histopathology is not sufficiently reliable to differentiate *ATM* germline variant carriers from non-carriers.

8.3.3 *ATM* Variants and Outcome

Within this cohort, pathogenic variants in *ATM* were not associated with a significant difference in Overall Survival or Distant Disease Free Survival in comparison to non-carriers. The rates of contralateral breast cancer were also comparable between gene carriers and non-carriers.

There is limited data regarding germline variants in *ATM* and breast cancer prognosis. It has been suggested that the combination of specific rare missense variants and radiotherapy may have an adverse effect on the development of contralateral disease.(184) In 2017, Bernstein et al. reported on data from the WECARE Study which recruited 708 women with contralateral breast cancer and 1399 controls.(232) They summarised that women with pathogenic rare missense variants in *ATM* who also received radiotherapy had a significantly increased risk of Contralateral Breast Cancer RR = 2.8 (95% CI, 1.2-6.5; 1.0). Within this cohort 18 *ATM* variant carriers received adjuvant radiotherapy and there was one case of contralateral disease. This observation does not support an association between radiotherapy and contralateral breast cancer risk amongst germline *ATM* variant carriers.

8.4 Summary

Overall, this work has provided a comprehensive overview of tumour histopathology and outcome in early onset *ATM*-associated breast cancer compared to non-gene carriers. It demonstrates that ER status is a significant independent predictor of germline variants in *ATM*. However, in the absence of other discriminatory features, tumour histopathology cannot be used reliably to differentiate variant carriers from non-carriers. This observation in combination with the high rates of VUS observed within this cohort and the variability in cancer risk attributable to missense variation highlights the potential pitfalls of using *ATM* for diagnostic testing to identify actionable breast cancer susceptibility and stratify risk.

Chapter 9 Tumour Histopathological Phenotype amongst Germline *TP53* Variant Carriers

9.1 Baseline Characteristics of the Cohort

The histopathological analysis has utilised data derived from the POSH study. The complete analysed cohort consisted of 2306 participants. This included 15 individuals with pathogenic variants in *TP53* and 2291 mutation negative participants. Mutation negative was defined as being *CHEK2*, *PALB2*, *ATM*, *TP53*, *BRCA1* and *BRCA2* negative.

Most recruits were Caucasian 2110/2279 (92.6%). There was missing ethnicity data for 27 individuals within this analysis. A pathogenic variant in *TP53* was found in 15/2306 (0.7%) of this cohort and 15/2744 (0.5%) of the whole POSH study cohort (Table 46). Of these, 5/287 (1.7%) were observed in women with very early onset breast cancer (age 30 years or younger at diagnosis). In total 8/15 were pathogenic missense variants and 7/15 were protein truncating variants (Appendix H). On average, *TP53* mutation carriers were significantly younger at diagnosis that non-carriers. The median age of cancer onset was 33 years (IQR 28-35 years) for *TP53* mutation carriers, and 37 years (IQR 34-39 years) for non-carriers ($p<0.0001$).

There was no significant difference in the family history of breast cancer between *TP53* mutation carriers and non-carriers ($p=0.66$). In total, 9/14 (64.3%) of all *TP53* mutation carriers had no family history of breast cancer compared to 1548/2223 (69.6%) of non-carriers. Despite this, *TP53* mutation carriers had a significantly higher median BOADICEA score. The median BOADICEA score was 0.16 (IQR 0.03-0.65) for *TP53* gene carriers compared to 0.03 (IQR 0.02-0.05) for non-carriers ($p=0.0044$). There was no baseline difference in body mass index between *TP53* mutation carriers and non-carriers (Table 46).

Characteristic		*TP53*+ (n=15)	*TP53*- (n=2291)	p-value† *TP53*+ vs *TP53*-
Median age diagnosis (yrs)		33	37	p<0.0001
	Range	22-36	18-40	
	IQR	28-35	34-39	
Body Mass Index (Total)		15 (100%)	2203 (100%)	p=0.73
	Underweight/Healthy (<25)	7 (46.7%)	1185 (53.8%)	
	Overweight (25-30)	4 (26.7%)	603 (27.4%)	
	Obese (>30)	4 (26.7%)	415 (18.8%)	
	Missing	0	88 (3.8%)	
Ethnicity (Total)		15 (100%)	2264 (100%)	p=0.74
	Caucasian/white	14 (93.3%)	2096 (92.6%)	
	Black	0	84 (3.7%)	
	Asian	1 (6.7%)	70 (3.1%)	
	Other	0	14 (0.6%)	
	Missing	0	27 (1.2%)	
Family History (Total)		14 (100%)	2209 (100%)	p=0.66
	No	9 (64.3%)	1539 (69.7%)	
	Yes	5 (35.7%)	670 (30.3%)	
	Missing	1 (6.7%)	82 (3.6%)	
BOADICEA score				p=0.0044
	Median	0.16	0.03	
	Range	0.02 to 0.99,	0.00 to 0.95,	
	IQR	0.03 to 0.65	0.02 to 0.05	
	Missing	1 (6.7%)	62 (2.7%)	

Table 46: Baseline Characteristics of the *TP53* Cohort

†Assessments of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

9.2 Tumour Histopathology

9.2.1 Baseline Tumour Grade, Size and Focality

Overall, *TP53*-associated tumours were significantly more likely to be grade 3 at presentation compared to non-carriers (grade 3 14/15 (93.3%) versus 1278/2229 (57.3%) (p=0.019)). The majority of *TP53*-associated tumours were localised (10/13 (76.9%)) this was also true of breast tumours in non-carriers 1471/2098 (70.1%) p=0.59.

There was no significant difference in overall tumour size or focality between *TP53* mutation carriers and non-carriers (Table 47 and Figure 37). Assessments of tumour size included both invasive and in-situ disease. The median maximum invasive tumour size was 2.6cm (IQR, 1.9-3.4cm) for *TP53* variant carriers versus 2.2cm (IQR, 1.5-3.3cm) for non-carriers (p=0.363)). The median maximum in-situ tumour size was 0.54cm (IQR, 0.0-0.75cm) for *TP53* variant carriers versus 0.61cm (IQR 0.0-0.45cm) for non-carriers (p=0.9614). This equated to a median overall tumour size of 3.5cm amongst *TP53* variant carriers and 2.7cm amongst non-carriers (p=0.23).

There was also no significant difference in tumour focality between *TP53* variant carriers and non-carriers (Table 47 and Figure 37). In total, 10/13 (76.9%) of *TP53*-associated tumours were localised compared to 1461/2085 (70.1%) of non-carriers (p=0.59). However, significantly higher levels of nodal involvement were observed amongst the *TP53* carriers with 13/15 (86.7%) demonstrating N1 stage disease at presentation compared to 1169/2253 (51.9%) of non-carriers (p=0.0072). A non-significant trend towards higher levels of lymphovascular infiltration was also observed amongst the *TP53* gene carriers compared to non-carriers (10/14 (71.4%) versus 1023/2129 (48.1%) p=0.081).

Characteristic	*TP53*+ (n=15)	*TP53*- (n=2291)	p-value† *TP53*+ vs *TP53*-
Histological Grade (Total)	15 (100%)	2229 (100%)	p=0.019
1	0	148 (6.6%)	
2	1 (6.7%)	803 (36.0%)	
3	14 (93.3%)	1278 (57.3%)	
Missing/not graded	0	62 (2.7%)	
Focality (Total)	13 (100%)	2085 (100%)	p=0.59
Localised	10 (76.9%)	1461 (70.1%)	
Multifocal	3 (23.1%)	624 (29.9%)	
Missing	2 (13.3%)	206 (9.0%)	
Max invasive tumour size (cm)			p=0.363
Median	2.6	2.2	
Range	1.2-5.0	0-17.0	
IQR	1.9-3.4	1.5-3.3	
Missing	1 (6.7%)	128 (5.6%)	
Max overall tumour size (cm)			p=0.23
Median	3.5	2.7	
Range	1.7-5.0	0-19.0	
IQR	2.6-3.8	1.8-4.0	
Missing	1 6.7%)	104 (4.5%)	
Max in-situ tumour size (cm)*			p=0.9614
Median	0.54	0.61	
Range	0-2.4	0-14.6	
IQR	0-0.75	0-0.45	
Missing	1 (6.7%)	130 (5.7%)	
Pathological N stage (total)	15 (100%)	2253 (100%)	p=0.0072
N0	2 (13.3%)	1084 (48.1%)	
N1	13 (86.7%)	1169 (51.9%)	
Missing	0	38 (1.7%)	
Number of positive lymph nodes	15 (100%)	2253 (100%)	p=0.046
0	2 (13.3%)	1084 (48.1%)	
1-3	8 (53.3%)	764 (33.9%)	
4-9	4 (26.7%)	273 (12.1%)	
10+	1 (6.7%)	132 (5.9%)	
Missing	0	38 (1.7%)	
Lymphovascular invasion (total)	14 (100%)	2129 (100%)	p=0.081
Absent	4 (28.6%)	1106 (51.9%)	
Present	10 (71.4%)	1023 (48.1%)	
Missing	1 (6.7%)	162 (7.1%)	

Table 47: Baseline Histopathological Characteristics of the *TP53* Cohort

Comparison of the tumour focality, grade and lymph node involvement between *TP53* pathogenic variant carriers and non-carriers. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.*Maximum in-situ disease was derived from the maximum overall tumour size minus the maximum invasive tumour size.

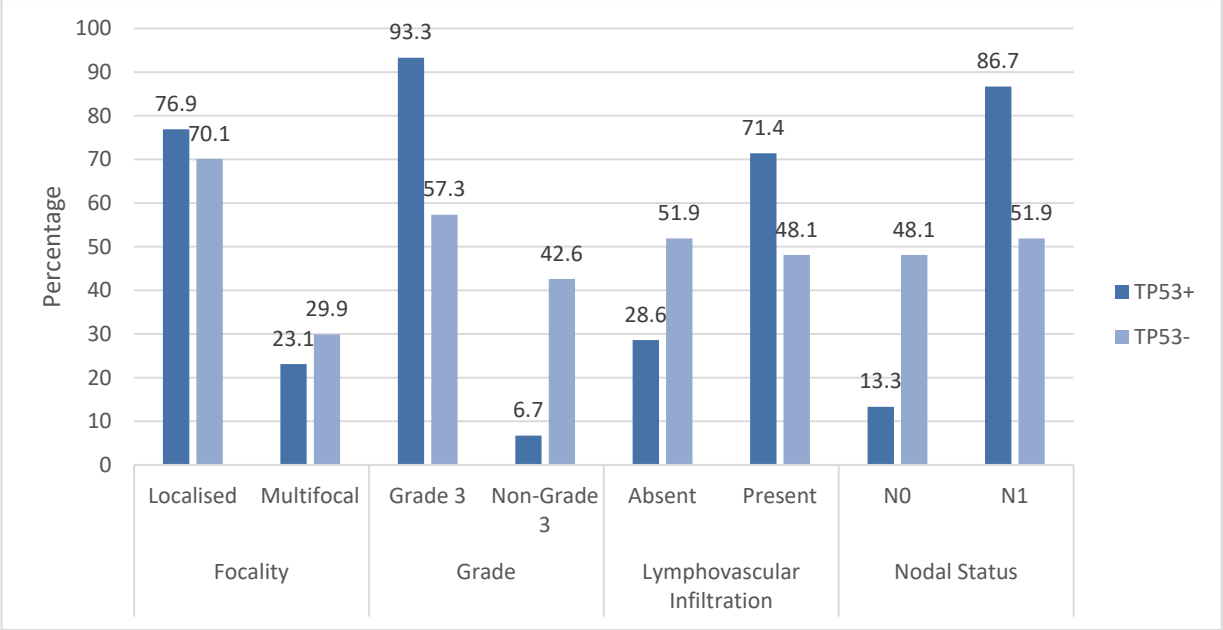


Figure 37: Baseline Histopathological Characteristics of the *TP53* Cohort

Comparison of the tumour focality, grade, lymph node involvement and lymphovascular infiltration between *TP53* mutation carriers and non-carriers. Values represented as percentage of the cohort.

9.2.2 Hormone Receptor Status

TP53-associated tumours were significantly more likely to be HER2-amplified with 11/12 (91.7%) demonstrating HER2 amplification compared to 593/2021 (29.3%) of non-carriers ($p<0.0001$) (Table 48 and Figure 38). This was no differential effect of variant type (truncating or missense) on the presence or absence of HER2 amplification (Table 49). In total 6/6 (100%) of individuals with a protein truncating variant presented with HER2-amplified breast cancer and 5/6 (83.3%) of individuals with missense variants presented with HER2-amplified breast cancer. The HER2 receptor status was missing for 3 variant carriers. One *TP53*-associated tumour was a TNT. This occurred in association with a pathogenic variant, *TP53* c.818G>A (p.Arg273His) located within the DNA binding domain of the *TP53* gene.

There was no significant difference between ER or PR receptor status between *TP53* mutation carriers and non-carriers. In total, 9/15 (60.0%) of the *TP53*-associated tumours were ER-positive compared to 1557/2279 (68.3%) of non-carriers ($p=0.49$) and 8/13 (61.5%) of the *TP53*-associated tumours were PR-positive compared to 1084/1848 (58.7%) of non-carriers ($p=0.83$). There was no association between *TP53* mutational status and the presence of a TN tumour phenotype ($p=0.25$). In total, 14/15 (93.3%) of *TP53*-associated tumours had a non-TN phenotype compared to 1874/2291 (81.8%) of non-*TP53* tumours (Table 48 and Figure 38).

Characteristic		*TP53*+ (n=15)	*TP53*- (n=2291)	p-value† *TP53*+ vs *TP53*-
ER status (total)		15 (100%)	2279 (100%)	p=0.49
	Negative	6 (40.0%)	722 (31.7%)	
	Positive	9 (60.0%)	1557 (68.3%)	
	Missing	0	12 (0.5%)	
HER2 status (total)		12 (100%)	2021 (100%)	p<0.0001
	Negative	1 (8.3%)	1428 (70.7%)	
	Positive	11 (91.7%)	593 (29.3%)	
	Missing	3 (20.0%)	270 (11.8%)	
PR status (total)		13 (100%)	1848 (100%)	p=0.83
	Negative	5 (38.5%)	764 (41.3%)	
	Positive	8 (61.5%)	1084 (58.7%)	
	Missing	2 (13.3%)	443 (19.3%)	
TNT status (total)		15 (100%)	2291 (100%)	p=0.25
	Not TNT	14 (93.3%)	1874 (81.8%)	
	TNT	1 (6.7%)	417 (18.2%)	
	Missing	0	0	

Table 48: Baseline Hormone Receptor Status of the *TP53* Cohort

†Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables. Samples derived from the POSH Cohort.

TP53 Variant		HER2 Status
Coding Change	**Protein Change**	
c.112C>T	p.Gln38X	Missing
c.437G>A	p.Trp146X	HER2+
c.524G>A	p.Arg175His	HER2+
c.524G>A	p.Arg175His	Missing
c.586C>T	p.Arg196X	HER2+
c.586C>T	p.Arg196X	HER2+
c.625A>T	p.Arg209X	HER2+
c.659A>G	p.Tyr220Cys	HER2+
c.672+1G>T	-	HER2+
c.725G>A	p.Cys242Tyr	HER2+
c.725G>A	p.Cys242Tyr	Missing
c.733G>A	p.Gly245Ser	HER2+
c.733G>A	p.Gly245Ser	HER2+
c.818G>A	p.Arg273His	TNT
c.919+1G>A	-	HER2+

Table 49: HER2 Receptor Status

Assessment of HER2 receptor status by variant type.

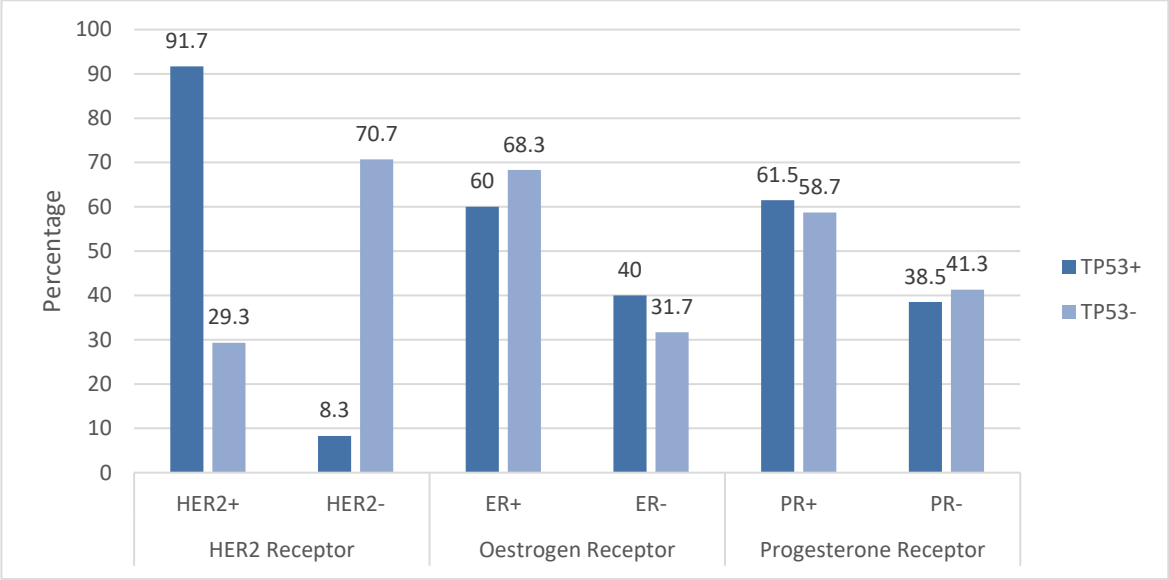


Figure 38: Baseline Hormone Receptor Status of the *TP53* Cohort

Comparison of ER, PR and HER2 receptor status between *TP53* mutations carriers and non-carriers. Values represented as percentage of the cohort.

9.2.3 HER2 Amplification and Variant Detection

Overall, HER2-amplified breast cancer cases represented 658/2744 (24.0%) of the whole POSH cohort and 90/287 (31.4%) of those diagnosed at the age of 30 years or younger. Pathogenic variants in *TP53* were identifiable in 10/658 (1.5%) of the unselected HER2-amplified cohort and 4/85 (4.7%) of those with HER2-amplified breast cancer diagnosed at the age of 30 years or younger (Figure 39 and Table 50).

In the absence of a significant family history of breast cancer (BOADICEA <10%), the variant detection rate for any breast cancer susceptibility gene was 33/526 (6.3%) across the whole HER2-amplified group. Pathogenic variants in *TP53* were identifiable in 4/526 (0.8%) of cases. Variants in other genes were identifiable in 29/526 (5.5%) of cases with *CHEK2* being the most frequently identifiable representing 12/526 (2.3%) of all gene carriers (figure 40 and table 48). Pathogenic variants in *BRCA1*, *BRCA2*, *PALB2* and *ATM* were identifiable in 5/526 (1.0%), 5/526 (1.0%), 1/526 (0.0%) and 6/526 (1.1%) of cases respectively (Figure 39 and Table 50).

Subgroup analysis observed variant identification amongst those individuals with very early onset (less than 30 years) breast cancer in the absence of a significant family history of breast cancer (BOADICEA <10%). The variant detection rate for any breast cancer susceptibility gene was 4/61

(6.6%). This was similar to variant identification in HER2-amplified breast cancer across the whole POSH cohort. Of these, pathogenic variants in *TP53* were identifiable in 1/61 (1.6%) of cases. Variants in other genes were identifiable in 3/61 (4.9%) of cases with *CHEK2* remaining the most frequently identifiable gene representing 2/61 (3.3%) of this group. Pathogenic variants in *BRCA1*, *BRCA2*, *PALB2* and *ATM* were identifiable in 1/61 (1.6%), 1/61 (1.6%), 0/61 (0.0%) and 0/61 (0.0%) of cases respectively (Figure 39 and Table 50).

It is notable that pathogenic variant identification increased considerably amongst HER2-amplified breast cancers occurring in the context of familial breast cancer. Amongst those individuals with a strong family history of breast cancer (BOADICEA ≥10%), the variant detection rate for any breast cancer susceptibility gene was 28/112 (25.0%). Pathogenic variants in *TP53* were identifiable in 6/112 (5.4%) of cases. Variants in other genes were identifiable in 22/112 (19.6%) with *BRCA1* and *BRCA2* being the most frequently identifiable representing 7/112 (6.3%) and 8/112 (7.1%) of all gene carriers. The overall detection of any *BRCA1* and *BRCA2* variants combined was 13/112 (11.6%) for all HER2-amplified breast cancers occurring in the context of familial breast cancer. Pathogenic variants in *PALB2*, *CHEK2* and *ATM* were identifiable in 1/112 (0.9%), 5/112 (4.5%), and 2/112 (1.8%) of cases respectively (Figure 39 and Table 50).

Subgroup analysis observed variant identification amongst HER2-amplified very early onset breast cancers (under the age of 30 years) occurring in the context of a strong family history of breast cancer. Overall, a pathogenic breast cancer susceptibility variant was identifiable in 7/24 (29.2%) of cases. Pathogenic variants in *TP53* were present in 3/24 (12.5%) of all variant carriers and represented 3/7 (42.9%) of variant carriers within this sub-group. This was the largest proportional representation of any gene within the HER2-amplified analysis. Variants in other genes were identifiable in 4/24 (16.7%) of cases with *BRCA2* and *CHEK2* being the most frequently observed, each accounting for 2/24 (8.3%) of cases. Pathogenic variants in *BRCA1*, *PALB2* and *ATM* were identifiable in 1/24 (4.2%), 0/24 (0.0%), and 0/24 (0.0%) of cases respectively (Figure 39 and Table 50).

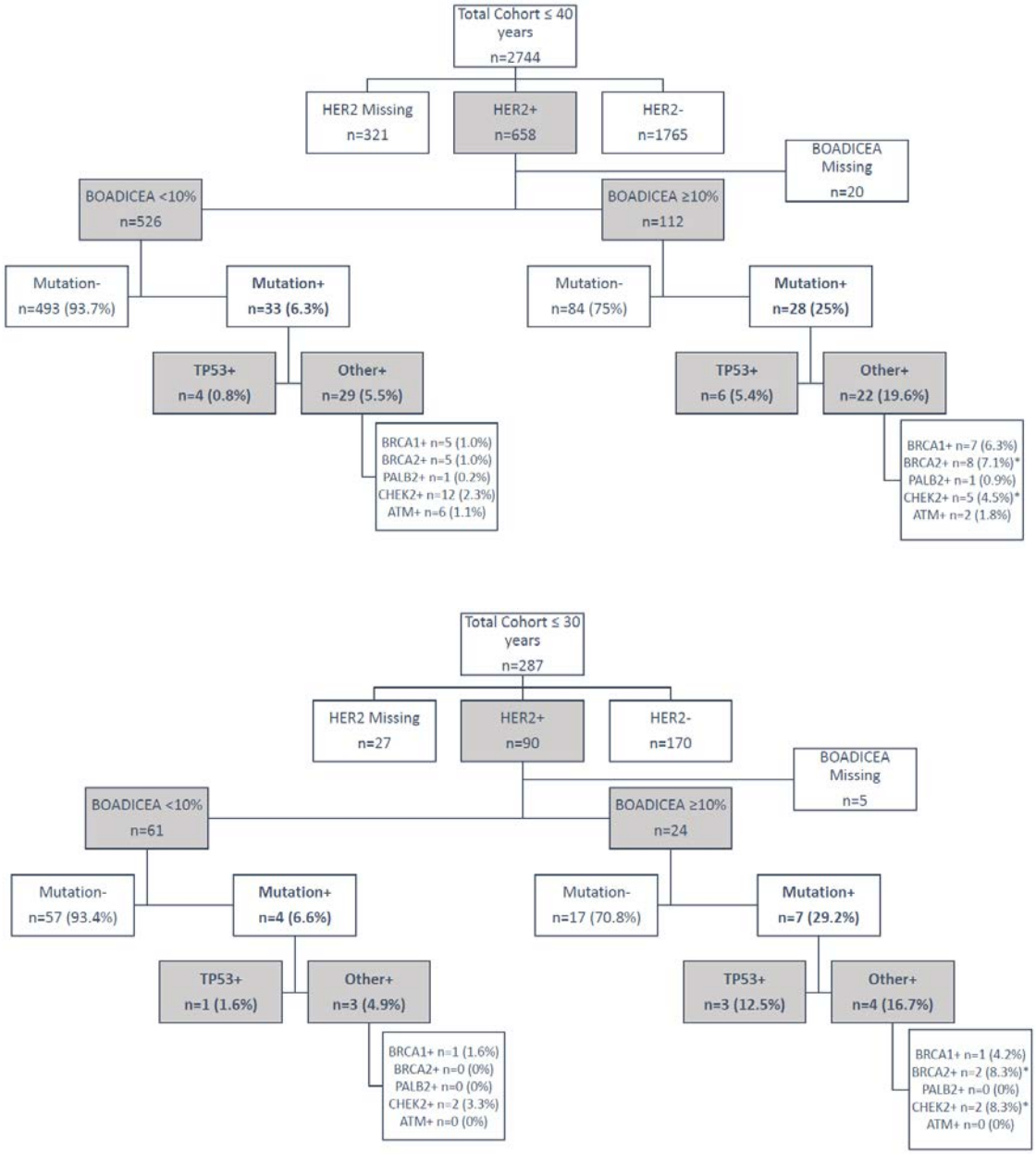


Figure 39: Variant Identification in HER2-amplified Breast Cancer Cases

A comparison of variant identification amongst HER2-amplified early onset and very early onset breast cancer cases. *One patient had two pathogenic variants; BRCA2 c.5682C>G p.Tyr1984X and CHEK2 c.1100delC.

Gene	HER2-amplified Breast Cancer N=658			
	BOADICEA <10% N=526		BOADICEA ≥10% N=112	
	≤40 years N=526	≤30 years N=61	≤40 years N=112	≤30 years N=24
TP53	0.8% (4/526)	1.6% (1/61)	5.4% (6/112)	12.5% (3/24)
BRCA1	1.0% (5/526)	1.6% (1/61)	6.3% (7/112)	4.2% (1/24)
BRCA2	1.0% (5/526)	0.0% (0/61)*	7.1% (8/112)	8.3% (2/24)*
PALB2	0.2% (1/526)	0.0% (0/61)	0.9% (1/112)	0.0% (0/24)
CHEK2	2.3% (12/526)	3.3% (2/61)*	4.5% (5/112)	8.3% (2/24)*
ATM	1.1% (6/526)	0% (0/61)	1.8% (2/112)	0.0% (0/24)
TOTAL	**6.3% (33/526)**	**6.6% (4/61)**	**25.0% (28/112)**	**29.2% (7/24)**

Table 50: Variant Identification in HER2-amplified Breast Cancer Cases

A comparison of variant identification amongst HER2-amplified early onset and very early onset breast cancer cases. *One patient had two pathogenic variants; *BRCA2* c.5682C>G p.Tyr1984X and *CHEK2* c.1100delC.

9.2.4 Histopathological Predictors

A model selection process using multiple logistic regression and incorporating forward selection by way of likelihood ratio tests was utilised to determine which histopathological features were better predictors of a *TP53* germline gene alteration (Table 51). Multiple imputation was utilised as 855/2306 (37%) had missing data. Grade was removed from the analysis due to insufficient numbers. Age at diagnosis, nodal status and HER2 receptor status were significant independent predictors of germline mutational status. Early age at diagnosis, N1 stage disease and HER2 amplification contributed towards the prediction of a *TP53* mutation (Table 51).

Overall, the highest probability of identifying a pathogenic *TP53* variant would be observed amongst those patients with an earlier age at diagnosis with localised, N1 tumours that were HER2-amplified. (Table 51).

		TP53	
Phenotype	**Classification**	**Adjusted odds ratio [95% CI] †**	**Significant Factor**
Age at diagnosis (years)	-	0.80 (0.72-0.90)	Yes
Nodal Involvement	N0	1*	-
	N1	5.13 (1.11-23.68)	Yes
Focality	Localised	1*	-
	Multifocal	0.57 (0.12-2.78)	-
Oestrogen Receptor	Negative	1*	-
	Positive	0.94 (0.49-1.79)	-
Progesterone Receptor	Negative	1*	-
	Positive	1.06 (0.57-1.96)	-
HER2 Receptor	Negative	1*	-
	Positive	25.37 (1.94-331.68)	Yes

Table 51: Multiple Logistic Regression Analysis

Multiple logistic regression analysis. †Analyses adjusted for hormone receptor status, tumour size, nodal involvement, hormone receptor status, focality, age at diagnosis and lymphovascular infiltration. *Reference category.

9.3 Outcome

The majority of patients received adjuvant chemotherapy. The most frequent regimen included anthracyclines with or without the additional of taxanes (Table 52). There were differences in the baseline treatment received. In total, 12/15 (80.0%) of *TP53* gene carriers received adjuvant chemotherapy versus 1682/2291 (73.4%) of non-carriers. A further 1/15 (6.7%) of gene carriers received palliative chemotherapy. No patients within the non-carrier group received palliative chemotherapy as the initial intervention. This observation was statistically significant $p<0.0001$ although the numbers are small. In total 11/15 (73.3%) of TP53 gene carriers received adjuvant radiotherapy.

A higher percentage of patients underwent mastectomy amongst the *TP53* mutation carriers. In total 11/15 (73.3%) of *TP53* mutation carriers underwent mastectomy as the primary surgical intervention versus 1122/2291 (49.0%) of non-carriers (p=0.21).

Treatment	*TP53*+ (n=15)	*TP53*- (n=2291)	p-value† *TP53*+ vs *TP53*-
Chemotherapy Timing (total)	15 (100%)	2291 (100%)	p<0.0001
None	0	262 (11.4%)	
Adjuvant	12 (80.0%)	1682 (73.4%)	
Neoadjuvant	2 (13.3%)	347 (15.1%)	
Palliative	1 (6.7%)	0	
Missing	0	0	
Chemotherapy Regimen (total)	15 (100%)	2291 (100%)	p=0.042
None	0	262 (11.4%)	
Anthracyclines	11 (73.3%)	1463 (63.9%)	
Anthracyclines and taxanes	3 (20.0%)	530 (23.1%)	
Taxanes only	0	21 (0.9%)	
Other	1 (6.7%)	15 (0.7%)	
Missing	0	0	
Surgery Type (total)	15 (100%)	2291 (100%)	p=0.31
Breast Conserving Surgery	4 (26.7%)	1149 (50.2%)	
Mastectomy	11 (73.3%)	1122 (49.0%)	
Nodal surgery only	0	6 (0.3%)	
None	0	14 (0.6%)	
Missing	0	0	

Table 52: Treatment Characteristics of the *TP53* Cohort

Comparison of the treatment protocol in relation to genotype. †Assessment of statistical significance were performed using the Mann-Whitney test for continuous variables and a Pearson $\chi 2$ test for categorical variables.

The median duration of follow up was 8.2 years. Contralateral breast cancers were more frequently observed in *TP53* carriers compared to non-carriers (Table 53). A contralateral breast cancer was observed in 3/15 (20.0%) of *TP53* carriers at 10 years compared to 85/2291 (3.7%) of non-carriers. Of the 3 *TP53* carriers with contralateral breast cancer, 1 had bilateral disease at presentation and another developed contralateral disease in the first 5 years following diagnosis.

Subgroup analysis revealed that the observed increase in contralateral breast cancer risk amongst *TP53* gene alteration carriers was higher in the context of familial breast cancer. In total, 2/5 (40.0%) of the *TP53* gene carriers with a positive family history of breast cancer developed a contralateral breast cancer compared to 1/9 (11.1%) of *TP53* gene carriers without a family history. This difference was apparent in the first five years following breast cancer diagnosis. The contralateral breast cancer rates observed amongst *TP53* carriers without a family history of breast cancer where still elevated compared to other gene carriers and non-carriers (Table 53).

Genotype	5 Years (%)	10 years (%)
TP53+ (ALL)	2 (13.3%)	3 (20%)
TP53+ (ALL) and FH-	0 (0%)	1 (11.1%)
TP53+ (ALL) and FH+	2 (40%)	2 (40.0%)
Mutation- (ALL)	62 (2.7%)	85 (3.7%)
Mutation- (ALL) and FH-	46 (3%)	59 (3.8%)
Mutation- (ALL) and FH+	15 (2.2%)	24 (3.6%)
PALB2+ (ALL)	0 (0%)	2 (6.5%)
PALB2+ (ALL) and FH+	0 (0%)	1 (8.3%)
CHEK2+ (ALL)	4 (7.5%)	5 (9.4%)
CHEK2+ (ALL) and FH+	3 (20%)	3 (20%)
ATM+ (ALL)	1 (4.3%)	1 (4.3%)
ATM+ (ALL) and FH+	1 (7.7%)	1 (7.7%)

Table 53: Contralateral Breast Cancer Risk in Association with *TP53*

Contralateral breast cancer risk at both 5 and 10 years for *TP53* mutation carriers versus non-carriers. Within the same cohort, the rate of contralateral breast cancer is 18% amongst *BRCA1* gene carriers and 12% amongst *BRCA2* gene carriers.(36)

Univariable analysis identified no significant difference in OS amongst *TP53* mutation carriers versus non-carriers within the first 5 years following cancer diagnosis (HR, 1.12 (95%CI, 0.47-2.72 (p=0.794)) (Figure 40). At 5 years, OS was 85.71% (95% CI, 53.94-96.22) amongst *TP53* mutation carriers versus 85.1% (95% CI, 83.5-86.5) in non-carriers. At 10 years, OS was 63.49% (95% CI, 27.33-85.29) amongst *TP53* mutation carriers versus 70.2% (95% CI, 67.8-72.5) in non-carriers. Following adjustment for known prognostic factors including age at diagnosis, BMI, maximum invasive size (cm), hormone receptor status, nodal involvement, ethnicity and taxanes in a multivariable analysis, an improvement in overall survival was observed but this remained non-significant (HR 0.86 (95%CI, 0.35-2.11 (p=0.75)) (Appendix K.4.1).

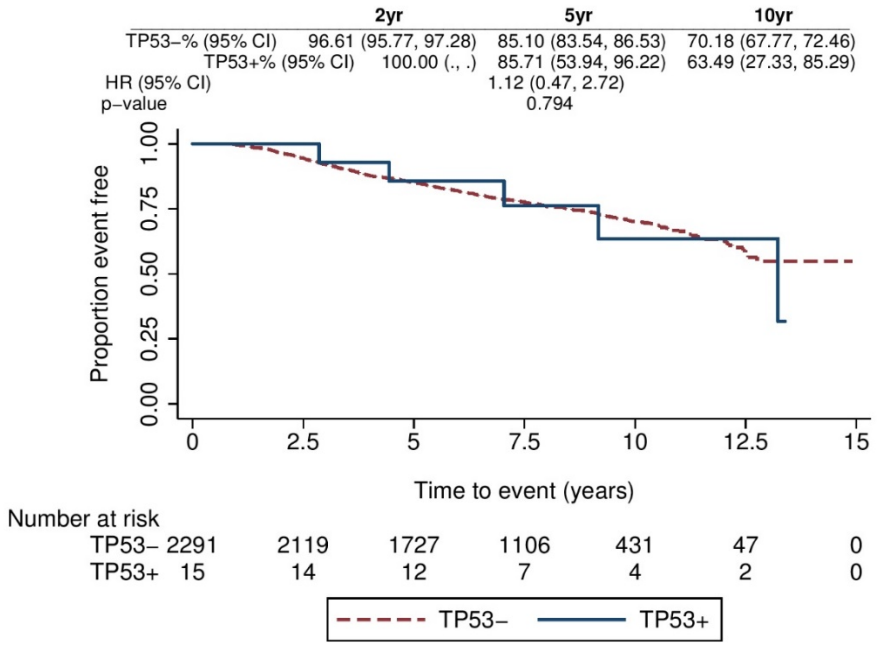


Figure 40: Kaplan Meier Plot of Overall Survival amongst *TP53* Mutation Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Overall Survival (OS) for *TP53* mutation carriers versus non-carriers following univariable analysis.

Univariable analysis identified no significant difference in DDFS amongst *TP53* mutation carriers versus non-carriers (HR, 1.33 (95%CI, 0.59-2.97 (p=0.490))) (Figure 41). At 5 years, DDFS was 79.0% (95% CI, 47.9-92.7) amongst *TP53* mutation carriers versus 77.7% (95% CI, 75.9-79.4) in non-carriers. At 10 years, DDFS was 57.6% (95% CI, 24.4-80.6) amongst *TP53* mutation carriers versus 69.0% (95% CI, 66.7-71.2) in non-carriers. Following adjustment for known prognostic factors including age at diagnosis, BMI, maximum invasive size (cm), hormone receptor status, nodal involvement, ethnicity and taxanes in a multivariable analysis, an improvement in overall survival was observed but this remained non-significant (HR 0.93 (95%CI, 0.41-2.12 (p=0.87)) (Appendix K.4.2).

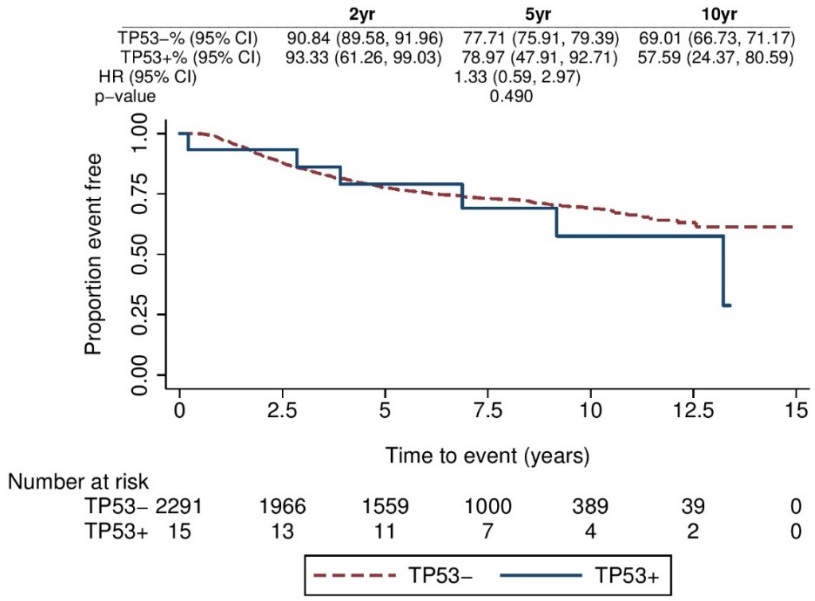


Figure 41: Kaplan Meier Plot of Distant Disease Free Survival amongst *TP53* Mutations Carriers versus Non-Carriers

Kaplan-Meier Plot demonstrating Distant Disease Free Survival (DDFS) for *TP53* mutation carriers versus non-carriers following univariable analysis.

9.4 Discussion

9.4.1 Baseline Characteristics of the Cohort

A pathogenic or likely pathogenic variant in *TP53* was found in 0.7% of this study cohort and 0.5% of the whole POSH cohort which represented unselected primary invasive breast cancer cases diagnosed under the age of 40 years. *TP53* mutation carriers were significantly younger at diagnosis in comparison to non-carriers and the prevalence of pathogenic variants in *TP53* was higher (1.7%) amongst very early onset breast cancers occurring at the age of 30 years or younger. This is consistent with Melhem-Bertrandt et al. who in 2012 concluded that the likelihood of identifying a pathogenic gene alteration reduced with increasing ages of cancer onset.(131)

The prevalence of *TP53* pathogenic variants is lower within this study cohort compared to the published literature where, it has been estimated that up to 3-8% of unselected very early onset breast cancers will have a pathogenic variant in *TP53* with a higher prevalence in the context of a LFS/LFL pedigree.(106, 112-117) However, many of these studies samples are ascertained from retrospective data or family studies. In 2006, Lalloo et al. observed the frequency of *TP53* pathogenic variants amongst 100 very early onset breast cancers ascertained from the North Western Cancer Registry.(116) In total, 4/100 (4%) had a pathogenic variant in *TP53*.(116) The

prevalence was lower in the absence of a family history of breast cancer 2/63 (3%).(116) In 2010, Mouchawar et al. observed the prevalence of *TP53* mutation carriers amongst 52 very early onset breast cancers unselected for family history from the Australian Breast Cancer Family Study.(121) They found that 2/52 (3.8%) has a pathogenic variant in *TP53*.(121) In 2013, Rath et al. observed the mutation detection amongst 213 women with HER2-amplified primary invasive breast cancer.(117) The prevalence of *TP53* pathogenic variants was 2/81 (2.5%) amongst women aged under the age of 40 years at diagnosis.(117) More recently, in 2019 Bakhuizen et al. observed *TP53* detection amongst 370 very early onset patients derived from a Dutch population.(233) They identified a *TP53* germline variant in 8/370 (2.2%) of unselected women.(233)

Breast cancer is the most frequently observed cancer amongst *TP53* gene carriers affecting approximately 60% of mutation carriers collectively.(114) Within this cohort, there was a significant association between *TP53* variant identification and a strong family history of breast cancer. However, 64.3% of variant carriers had no family history of breast cancer. This may reflect the broad range of cancer phenotypes associated with alterations in the *TP53* gene including adreno-cortical tumours, osteosarcoma, rhabdomyosarcoma and brain tumours.(114) It may also reflect the de-novo mutation rate observed amongst *TP53* gene carriers. In 2009, Gonzalez et al. determined the proportion of de-novo pathogenic *TP53* variants in a case series of 341 American patients with early onset breast cancer.(120) The estimated de-novo mutation rate was 5-20% was based upon a combination of molecular genetic testing and family history data.(120).

9.4.2 *TP53* Variants and Tumour Histopathology

We have shown that individuals with pathogenic variants in *TP53* are significantly more likely to present with high-grade tumours demonstrating nodal involvement and a trend towards lymphovascular infiltration. In 2013, Rath et al. reported on the tumour histopathological phenotype observed amongst *TP53* germline variants carriers derived from the CORIS database.(117) The majority of tumours, 69% were grade 3 at presentation.(117) In 2019, Packwood et al. through the COPE study observed tumour histopathological phenotype amongst 45 invasive breast cancers and 9 isolated cases of DCIS occurring in association with a germline pathogenic variant in *TP53*.(234) They found that 50% of tumours occurring in association with a *TP53* germline variant were grade 3 at presentation.(234) They also noted that vascular invasion was more frequent amongst those with a germline *TP53* variants (12/36 (33.3%)).(234)

There was no association with tumour focality or increased levels of in-situ disease amongst *TP53* variant carriers compared to non-carriers. The POSH study protocol excluded patients with isolated

DCIS.(13) However, this is an interesting observation as germline variants in *TP53* are significantly associated with isolated high grade DCIS in pre-menopausal breast cancer.(235)

TP53-associated tumours were significantly more likely to be HER2-amplified with 91.3% presenting with HER2 amplification compared to 29.3% of non-carriers. This was no differential effect of variant type (truncating or missense) on the presence of HER2 amplification. HER2 amplification associated with ERBB2 (HER2) gene amplification is present in 15-20% of invasive breast cancers.(236, 237) It can be associated with adverse prognosis and is the target of precision therapies including Trastuzumab.(236) Within the whole POSH cohort 24% of breast cancer were HER2-amplified and this increased to 31.4% amongst very early onset breast cancers.

The association between HER2 amplification and *TP53* germline mutations was first described by Wilson et al. in 2010.(159) They observed the proportion of breast cancers with HER2 amplification derived from 9 patients with pathogenic variants in *TP53* including individuals from the POSH cohort.(159) This was compared to non-gene carriers with very early onset breast cancer within the same cohort.(159) Overall, 83% of breast cancers occurring in association with a germline mutation in *TP53* demonstrated HER2 amplification compared to 16% of non-carriers.(159) In 2012, Masciari et al. observed the frequency of HER2-amplified invasive breast cancer and Ductal Carcinoma in Situ amongst 43 tumours derived from 39 *TP53* variant carriers.(238) They found that 26/32 primary invasive breast cancers were high grade and that 63% of breast cancers were HER2-amplified.(238)

In 2012, Melhem-Bertrandt et al. also observed the histopathological phenotype of *TP53*-associated breast cancers amongst 30 cases and 79 controls from individuals who met the NCCN family history criteria for *TP53* testing within an American population.(131) They found that tumours occurring in association with a *TP53* gene alteration were significantly more likely to be HER2-amplified ($p=0.0001$).(131) In total, 20/30 (67%) of breast cancers associated with *TP53* gene alterations were HER2-amplified compared to 20/79 (25%) of controls.(131) They concluded that the presence of HER2 amplification increased the OR of having a *TP53* gene alteration by 7 fold (OR 6.9, 95%CI 2.6-18.2 $p<0.0001$).(131) In 2013, Rath et al. also reported that 91% of breast cancers diagnosed under the age of 40 years and occurring in association with a *TP53* germline variant were HER2-amplified.(117) In 2019, Packwood et al. found that 20/32 (62.5%) of early onset breast cancers occurring in association with a *TP53* germline variant were HER2-amplified.(234)

9.4.3 HER2 Receptor Amplification and Variant Identification

There is limited literature detailing the mutational frequency of high and moderate penetrance genes in HER2-amplified breast cancer. One of the most comprehensive reviews was conducted by Eccles et al. in 2016 with a focus upon the high risk genes *BRCA1*, *BRCA2* and *TP53*.(160) Within this

cohort, age at diagnosis and HER2 receptor amplification were shown to be significant independent predictors of germline *TP53* variants. This work expands upon the analysis conducted by Eccles et al. to include other high and moderate penetrance genes. It also considers how the age of breast cancer onset influences variant detection.

Overall, pathogenic variants in *TP53* were identifiable in 1.5% of the whole, HER2-amplified cohort and 4.7% of those with HER2-amplified breast cancer diagnosed under the age of 30 years. Family history was an important modifier of the likelihood of identifying a pathogenic variant in *TP53*. In the absence of a strong family history of breast cancer, less than 1% of individuals with early onset (under 40 years) HER2-amplified breast cancer and less than 2% of individuals with very early onset (under 30 years) HER2-amplified breast cancer had a pathogenic variant in *TP53*.

Many pathogenic variants in *TP53* are missense variants. Novel missense variants can be more challenging to interpret utilising ACMG guidelines, especially in the absence of a strong family history of cancer. Based upon this analysis, the likelihood of identifying Variation of Uncertain Clinical Significance amongst individuals with non-familial early onset and very early onset HER2-amplified breast cancer is greater than the likelihood of identifying clinically actionable variation in *TP53*. Therefore, whilst many studies advocate reflex testing for pathogenic variants in *TP53* amongst all very early onset breast cancers diagnosed under the age of 30 years, this work raises whether there is true clinical utility in this approach.(115)

The identification of other genes also remained below the threshold for diagnostic genetic testing in non-familial early onset and very early onset HER2-amplified breast cancer. Pathogenic variants in *CHEK2* were the most frequently identified variant amongst non-familial HER2-amplified breast cancers, accounting for 2.2% of very early onset and 3.3% of early onset breast cancers. It demonstrates that the likelihood of identifying a pathogenic variant in *CHEK2* is greater than the likelihood of identifying a pathogenic variant in *TP53* amongst symptomatic HER2-amplified early onset breast cancer occurring in the absence of a strong family history of breast cancer.

Overall, 37.5% of *CHEK2*-associated breast cancers were HER2-amplified within the POSH cohort at presentation. Whilst this is not significantly different to non-gene carriers overall (24% HER amplified), it is higher compared to *BRCA* gene carriers (8.6% HER2-amplified) and *PALB2* gene carriers (11.5% HER2-amplified). The identification of patients with a germline *CHEK2* pathogenic variant at the time of diagnosis is of potential importance to oncologists as they plan future treatment given the association between adverse outcome and higher rates of contralateral disease.(224)

Variant identification increased amongst HER2-amplified breast cancers occurring in the context of a strong family history of breast cancer (BOADICEA ≥ 10%). Overall, 25% of individuals presenting with HER2-amplified breast cancer and a strong family history of breast cancer had a pathogenic variant in a breast cancer susceptibility gene (*BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* or *TP53*). This increased to 29.2% amongst breast cancers diagnosed at the age of 30 years or younger.

Pathogenic variants in *TP53* were present in 5.4% of familial HER2-amplified early onset breast cancers diagnosed under the age of 40 years. Pathogenic variants in *BRCA1* and *BRCA2* were identifiable in 6.3% and 7.1% of cases respectively and accounted for 13.4% of pathogenic variants collectively. It means that *TP53* variant identification was equivalent to *BRCA1* and *BRCA2* detection amongst familial, early onset, HER2-amplified breast cancers. *BRCA* variant identification within this subgroup also reached the NICE agreed threshold for diagnostic genetic testing.(229)

Germline variants in the *BRCA* genes are not classically associated with HER2 receptor amplification. Within this cohort, HER2 receptor amplification significantly reduced the likelihood of identifying a germline pathogenic *BRCA* variant (chapter 5). This relationship is applied within the Manchester scoring System where the presence of HER2 receptor amplification reduces the prior probability of identifying a germline *BRCA* variant.(239) Despite the inverse relationship between HER2 receptor amplification and *BRCA* variant identification, this work highlights the importance of *BRCA* testing in HER2-amplified early onset breast cancers occurring in the context of a strong family history.

HER2 receptor amplification was the most discriminatory for identifying individuals with a pathogenic variant in *TP53* amongst very early onset breast cancers. More specifically, a constitution variant in *TP53* was identifiable in 12.5% of HER2-amplified breast cancers occurring under the age of 30 years in the context of a strong family history of breast cancer. Its relative prevalence within this subgroup was greater than all other variants including *BRCA1*, *BRCA2* and *CHEK2* which were present in 4.2%, 8.3% and 8.3% respectively. This is similar to the published literature. In 2015 Bougeard et al. observed that the *TP53* mutation detection rate in 333 women with primary invasive breast cancer and a family history suggestive of LFS who developed breast cancer before the age of 33 years was 14%.(114)

Whilst *TP53* was strongly associated with HER2 receptor amplification. There was no association with *PALB2* pathogenic variants and HER2-amplified breast cancer. *PALB2* variants were identifiable in less than 1% of HER2-amplified breast cancers irrespective of family history or age of breast cancer onset.

9.4.4 *TP53* Variants and Outcome

Germline variants in *TP53* are associated with high rates of contralateral disease. In total, 20% of individuals with *TP53*-associated breast cancer developed contralateral breast cancer by 10 years. This increased to 40% in the presence of a family history of breast cancer. This was the highest observed rate of contralateral disease seen in association with a breast cancer susceptibility gene alteration within this cohort. This is similar to the rates reported by Hyder et al. in 2020. (240) They observed the incidence of contralateral breast cancer amongst 47 *TP53* gene carriers who were diagnosed with primary invasive breast cancer under the age of 35 years.(240) The annual rate of contralateral disease was 7.03% amongst *TP53* gene carriers.(240) This was significantly higher than *BRCA1* and *BRCA2* gene carriers within the same cohort were the reported rate of contralateral disease was 3.57% and 2.63% per annum respectively. Individualised decision making around primary surgical intervention is an important factor within breast cancer care, particularly in the context of heritable susceptibility. The substantial cumulative risk of contralateral disease observed within this study suggest that bilateral mastectomy may be a consideration for women with early onset breast cancer and a germline variant in *TP53*.

We have shown no significant difference in OS and DDFS between *TP53* gene carriers and non-carriers. After correction for other known prognostic factors including tumour size and lymph node involvement there was a trend towards improved OS and DDFS but this remained non-significant. Whilst the overall numbers are small, the observed improvement following correction for other known prognostic factors may highlight the importance of early detection and intervention to improve survival amongst *TP53* gene carriers.

The impact of somatic *TP53* mutations and outcome is well characterised. In 2018, Meric-Bernstam et al. observed the effect of somatic *TP53* mutations on outcome amongst 257 individuals with metastatic breast cancer derived from The Cancer Genome Atlas. *TP53* mutations were associated with significantly worse Recurrence Free Survival ($p<0.001$), Progression Free Survival ($p<0.001$) and Overall Survival ($p=0.03$).(241) In 2006, Olivier et al. observed the impact of somatic *TP53* variants amongst 1794 individuals with primary breast cancer derived from a European population.(242) They found that *TP53* variants including missense variants in the DNA Binding Domain were associated with reduced survival and a 10 year mortality of 73.42 per 1000 persons ($p=0.0897$).(242) In 2014, Eikesdal et al. observed Recurrence Free Survival and Overall Survival amongst 90 individuals with Locally Advanced Breast Cancer and somatic *TP53* variants.(243) Overall, somatic *TP53* variants were associated with a significant reduction in Recurrence Free and Overall Survival ($p<0.001$).(243)

Despite the recognised associated between somatic *TP53* variants and adverse outcome, there is limited literature observing the effects of germline *TP53* variants on outcome. The number of cases represented within this analysis is too limited to provide definitive conclusions of the impact that germline variants may have on outcome. However, we could hypothesise that the association with HER2 receptor amplification and higher rates of contralateral disease compared to the cohort overall reflect more aggressive underlying tumour biology.

9.5 Summary

Analysis of this cohort suggests that all individuals with early onset HER2-amplified breast cancer should be tested for a panel of breast cancer susceptibility genes including *TP53* if the breast cancer is diagnosed in the context of familial breast cancer. The identification of pathogenic germline variants in *TP53* has increasing clinical utility for active cancer management, including primary surgical intervention and the facilitation of decisions regarding radiotherapy and chemotherapy due to the potential for secondary malignancies. It also has utility for primary prevention within the broader family and the facilitation of reproductive decisions. The UK Cancer Genetics Group now advocate annual Whole Body MRI screening as part of a broader screening initiative for individuals with LFS. NICE also advocate high risk breast screening from the age of 20 years with consideration of risk reducing breast surgery.(135)

Despite the apparent utility of testing in the context of HER2-amplified, familial breast cancer, caution must be applied when there is no family history of cancer. This is even true amongst those individuals with HER2-amplified breast cancer under the age of 30 years where reflex testing is the current standard across many Clinical Genetics Services. For this group, it is important to consider the potential psychological burden of identifying VUS and the pitfalls in determining absolute cancer risks in the absence of a classic LFS/LFL family history where the likelihood of this is potentially greater than finding actionable risk.

We have also shown that germline variation in *CHEK2* makes an important contribution to non-familial early onset HER2-amplified breast cancers and its identification may have utility in treatment planning secondary to the association of germline variants in this gene with adverse tumour biology and outcome.

The pathway which unites germline *TP53* variants with HER2 receptor amplification remains to be elucidated. In the future, this may serve as a target for therapeutic intervention or chemoprevention and somatic mutational profiling may enable the identification of this relationship and the somatic factors which drive tumour biology in HER2-amplified breast cancer.

Chapter 10 Somatic Mutational Analysis

The somatic mutational analysis was completed utilising data derived from The 100,000 Genomes Project, Cancer, Breast Cancer, Recruitment Domain. Whole genome tumour sequence data was available for 1342 individuals with invasive breast cancer of whom 49 were gene carriers and 1293 were mutation negative. A summary of this cohort is shown in (Table 54).

Gene Carrier Status	Number
Mutation-	1293
Mutation+	49
BRCA+	7
PALB2+	7
CHEK2+	20
ATM+	10
TP53+	5
Total	**1342**

Table 54: Summary of the Somatic Analysis Population

10.1 Somatic Mutational Profile

A pre-selected panel of 39 target genes was applied to 1342 genomes derived from The 100,000 Genomes Project, Cancer, Breast Cancer Domain. A summary of the target genes is shown in Appendix G. In total, 33101 somatically acquired single nucleotide variants were identified. Non-coding variants were most frequently identified accounting for 30125/33101 (91.01%) of all somatic variants within this cohort. Intronic variants represented the most prevalent subclass. In total, 28196 intronic variants were identified (Table 55).

Within the protein coding region 1486 missense variants, 630 frameshift variants, 281 stop gained variants, 148 synonymous variants, 98 in-frame deletions, 12 in-frame insertions, 2 start lost variants and 2 stop lost variants were identified. Coding sequence variants represented 2659/33101 (8.03%) of all somatic variants identified. Conversely splice region variants represented less than 1% of all variants identified (317/33101 (0.96%)) (Table 55).

Variant Type	Number	Percentage
Coding Sequence (Total)	2659	8.03
Stop Gained	281	0.85
Frameshift Variant	630	1.90
Start Lost	2	0.01
Stop Lost	2	0.01
Missense Variant	1486	4.49
In-frame Deletion	98	0.30
In-Frame Insertion	12	0.04
Synonymous Variant	148	0.45
Splice Region (Total)	317	0.96
Splice Acceptor Variant	98	0.30
Splice Donor Variant	65	0.20
Splice Region Variant	154	0.47
Non-Coding (Total)	30125	91.01
Intron Variant	28196	85.18
3' Untranslated Region Variant	1017	3.07
5' Untranslated Region Variant	112	0.34
Upstream Gene Variant	575	1.74
Downstream Gene Variant	225	0.68
Total	**33101**	**100**

Table 55: Somatic Mutational Analysis

Summary level data detailing the variant types identified during somatic mutational analysis.

Analysis of the coding region variants demonstrated that the 5 most prevalent somatically mutated genes were *PIK3CA*, *TP53*, *GATA3*, *CDH1* and *MAP3K1* (Figure 42). Of these, *PIK3CA*, *TP53*, *CDH1* and *MAP3K1* have a recognised Hallmark functions. When non-coding variants were included within this analysis, increased representation was identifiable from *NTRK3*, *ESR1*, *ARID1B* and *ETV6* (Figure 43). The numbers were too small to provide meaningful comparisons between variant carriers but the Waterfall plots are shown in Appendix L.1-L.4. Tumour histopathology was not available to compare this with somatic mutational profile.

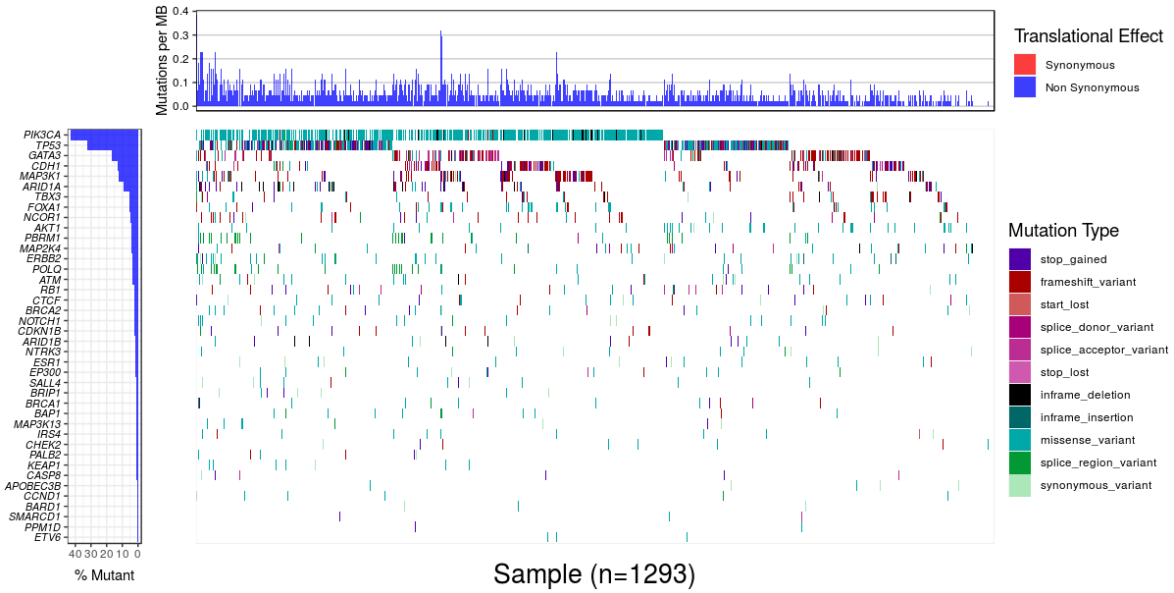


Figure 42: Somatic Mutation Profile Variant Negative Unselected Breast Cancer: Coding Variants

Waterfall plot illustrating the most prevalent somatic coding sequence variants identified in 1293 unselected breast cancers derived from The 100,000 Genomes Project.

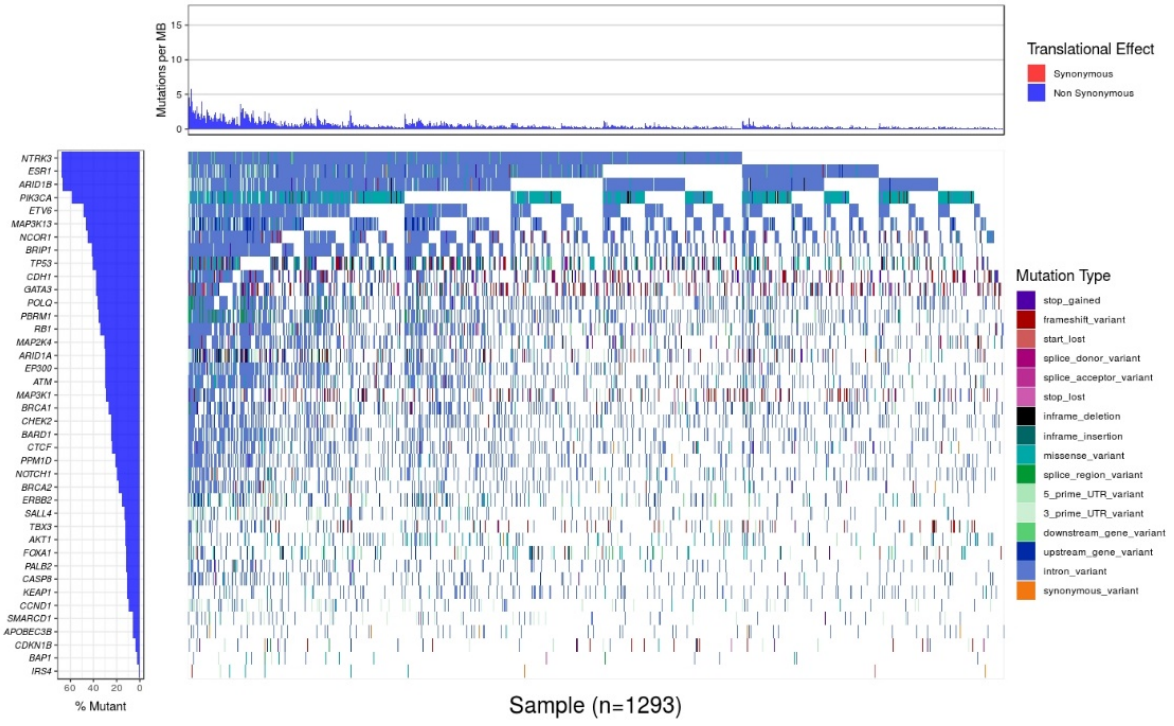


Figure 43: Somatic Mutation Profile Variant Negative Unselected Breast Cancer: Coding and Non-Coding Variants

Waterfall plot illustrating the most prevalent somatic coding and non-coding sequence variants identified in 1293 unselected breast cancers derived from The 100,000 Genomes Project.

10.2 Tumour Mutational Burden

The median Tumour Mutational Burden (TMB) was 2.51 (IQR, 1.87-4.25) Mutations per Megabase (Mut/Mb) amongst the 1293 individuals who were mutation negative within this cohort (Table 56 and Figure 44). The TMB was significantly higher amongst individuals with an identifiable germline variant (median 3.15 Mut/Mb (IQR, 2.38-5.17) (p=0.0169).

The significant difference in TMB between gene carriers and non-carriers was predominantly influenced by the contribution from *BRCA* and *PALB2* germline variant carriers. *BRCA* and *PALB2* variant carriers developed invasive breast tumours with a significantly higher TMB compared to non-carriers (p=0.0433 and p=0.0066 respectively). The median TMB for germline *BRCA* variant carriers was 4.39 Mut/Mb (IQR, 3.35-5.00). The median TMB for germline *PALB2* variant carriers was 6.39 (IQR, 3.96-7.66) (Table 56 and Figure 44).

There was no significant difference in the TMB between *CHEK2* or *ATM* gene carriers compared to non-carriers (p=0.4422 and p=0.6692 respectively). The median TMB was 2.49 Mut/Mb (IQR, 1.93-3.27) for *CHEK2* gene carriers and 2.70 Mut/Mb (IQR, 2.24-4.02) for *ATM* gene carriers. TMB was higher amongst *TP53* gene carriers compared to non-carriers (median 3.77 Mut/Mb (IQR, 2.58-6.41). However this relationship was non-significant (p=0.18).

Hypermutation (TMB greater than 10 mut/Mb) was not observed within the median or interquartile range of TMB for either gene carriers or non-carriers within this cohort. However, hypermutation was observed within the range of TMB amongst non-gene carriers (0.31-55.95 Mut/Mb) and *CHEK2* gene carriers (0.83-1570 Mut/Mb).

	Somatic Coding Variants Per Mb			
Gene	**Median**	**IQR**	**Range**	**P-value***
MUTATION-	2.51	1.87-4.25	0.31-55.94	-
MUTATION+	3.15	2.38-5.17	0.83-15.7	p=0.0169
BRCA+	4.39	3.36-5.00	3.04-6.16	p=0.0433
PALB2+	6.39	3.96-7.66	3.22-9.78	p=0.0066
CHEK2+	2.49	1.93-3.27	0.83-15.70	p=0.6692
ATM+	2.70	2.24-4.02	1.84-6.07	p=0.4422
TP53+	3.77	2.58-6.41	2.12-6.65	p=0.18

Table 56: Tumour Mutational Burden Gene Carriers Versus Non-Carriers

*Assessment of statistical significance using the Mann-Whitney U Test. Comparison made against Mutation-.

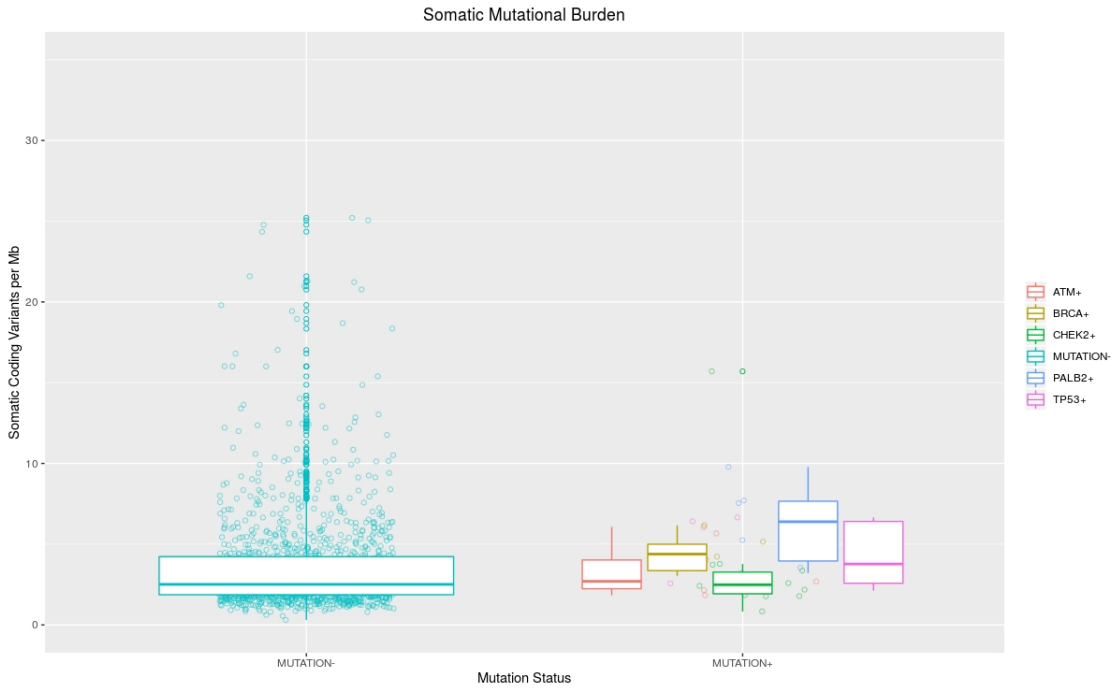


Figure 44: Tumour Mutational Burden described in Somatic Coding Variant per Megabase

Boxplot comparing the median and inter-quartile range of Tumour Mutational Burden observed between gene carriers (ATM+, BRCA+, CHEK2+, PALB2+ and TP53+) and non-carriers (Mutation-).

10.3 Somatic Mutational Signature

In total, the presence of 30 biologically relevant Single Base Substitution (SBS) Mutational Signatures was analysed and compared between gene carriers and non-carriers. Overall, 9 main SBS mutational signatures were expressed consistently amongst all breast cancers observed within this cohort. These mutational signatures were present concomitantly in each tumour. In-depth summary data of the proportional representation of each mutational signature is provided in Appendix M and Figures 45 and 46.

The 9 main SBS signatures observed amongst breast cancer cases included Signature 1 (SBS1), Signature 2 (SBS2), Signature 3 (SBS3), Signature 5 (SBS5), Signature 8 (SBS8), Signature 9 (SBS9), Signature 12 (SBS12), Signature 13 (SBS13) and Signature 16 (SBS16). Amongst non-gene carriers the percentage representation of these mutational signatures was as follows: SBS1 median 14.64 (IQR, 8.92-17.51); SBS2, median 2.88 (IQR, 1.72-6.17); SBS3, median 10.59 (IQR, 6.44-15.95); SBS5, median 28.08 (IQR, 11.41-38.30); SBS8, median 3.94 (IQR, 0.00-9.33); SBS9, median 4.67 (IQR, 2.87-6.51); SBS12, median 4.62 (IQR, 1.40-9.08); SBS13 median 1.78 (IQR, 0.88-4.65); SBS16 median 8.90 (IQR, 2.64-16.74) (Appendix L and Figure 45 and 46). In some tumours, a smaller

contribution (median less than 2%) was observed for SBS10, SBS11, SBS18, SBS25, SBS28 and SBS30.

The proportional representation of specific mutational signature was significantly different between germline variant carriers and non-carriers. The percentage expression of SBS1 was significantly reduced in gene carriers (10.81 Mut/Mb (IQR, 7.53-14.48)) compared to non-carriers (14.64 Mut/Mb (IQR, 8.92-17.51))(p=0.001235). Subgroup analysis revealed that this reduction in prevalence was attributable to germline variants in *BRCA* and *PALB2*. The median expression of SBS1 was 8.53 Mut/Mb (IQR, 6.59-9.99) for *BRCA* gene carriers (p=0.0245) and 6.71 Mut/Mb (IQR, 5.10-7.01) for *PALB2* gene carriers (p=0.0016) (Appendix L and Figure 45 and 46).

A similar observation was noted for SBS5 between germline gene carriers and non-carriers median 28.08 Mut/Mb (IQR, 11.41-38.30) versus 18.74 Mut/Mb (IQR, 0.00-30.11) (p=0.0021). This reduction in prevalence was attributable to germline variants in *PALB2* only. The median expression of SBS5 was 4.88 Mut/Mb (IQR, 0.00-17.36) for *PALB2* gene carriers (p=0.010) (Appendix L and Figure 45 and 46).

The percentage expression of SBS3 was significantly increased in gene carriers compare to non-carriers. This difference was due to a higher percentage prevalence of SBS3 amongst *BRCA* and *PALB2* gene carriers. The median percentage of SBS3 was 10.59 (IQR, 6.44-15.95) amongst non-gene carriers compared to 31.14 (IQR,20.71-35.17) amongst *BRCA* gene carriers and 32.14 (IQR, 31.02-37.71) amongst *PALB2* gene carriers (p<0.0001 and p=0.0047 respectively) (Appendix L and Figure 45 and 46).

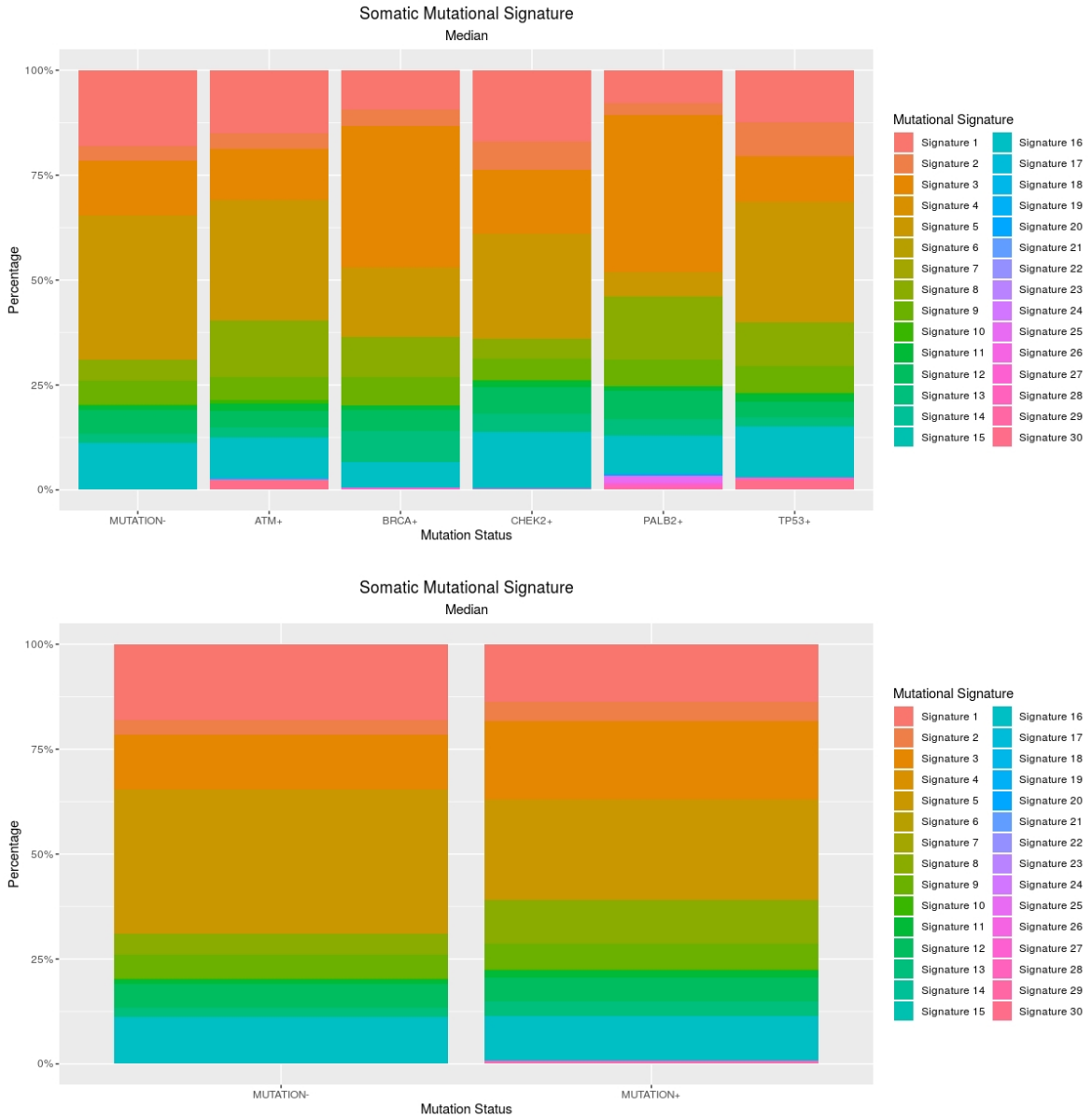
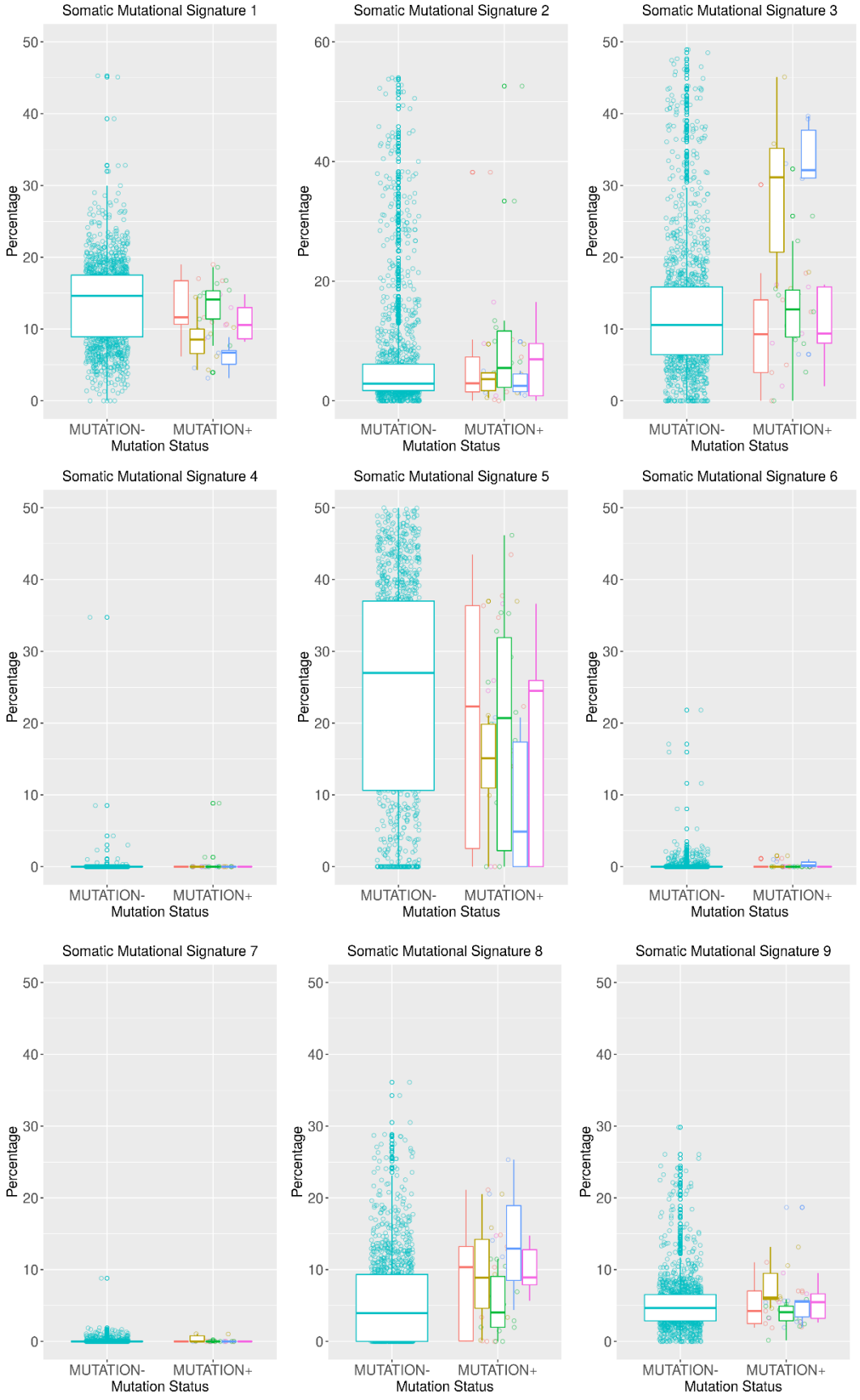
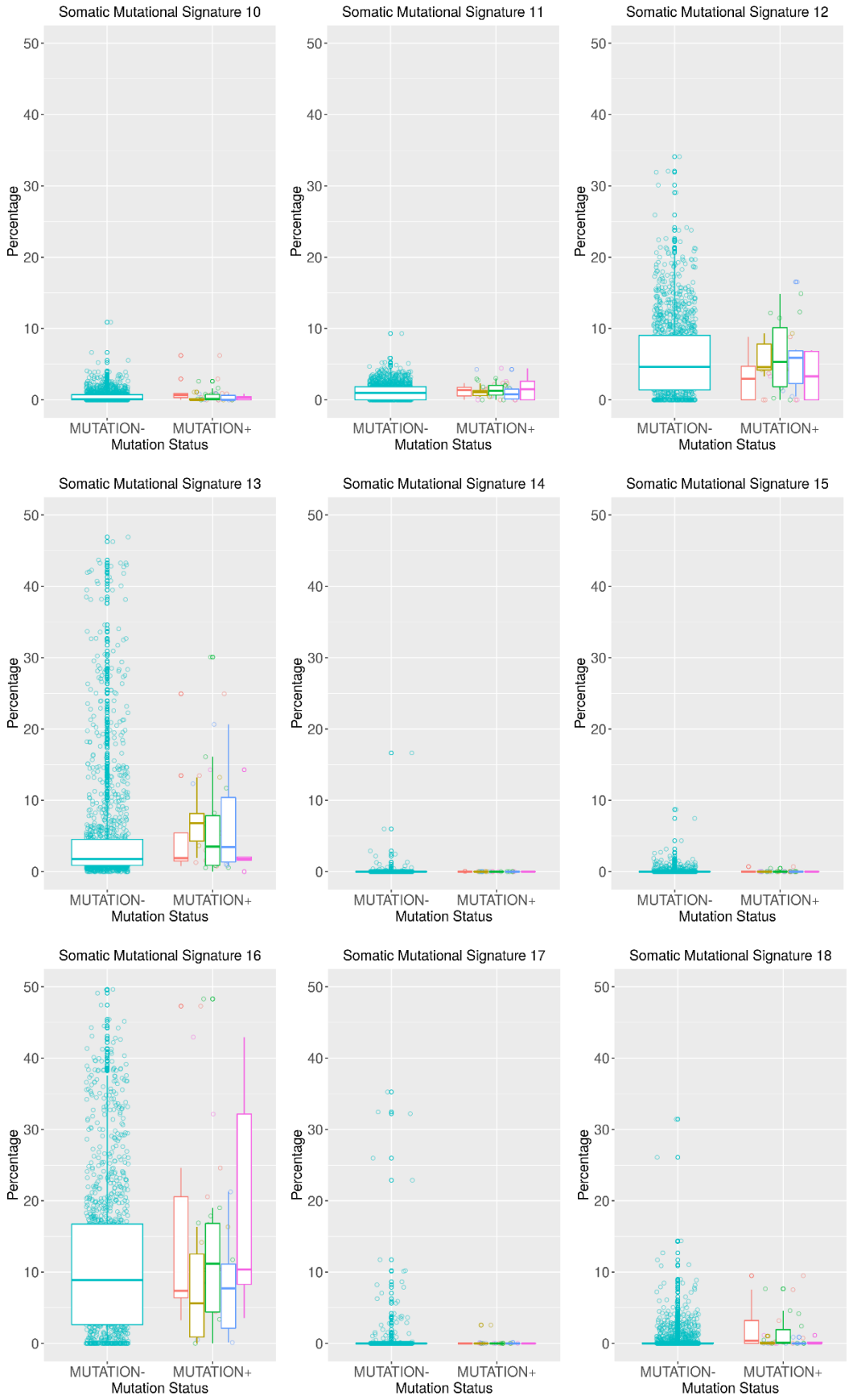
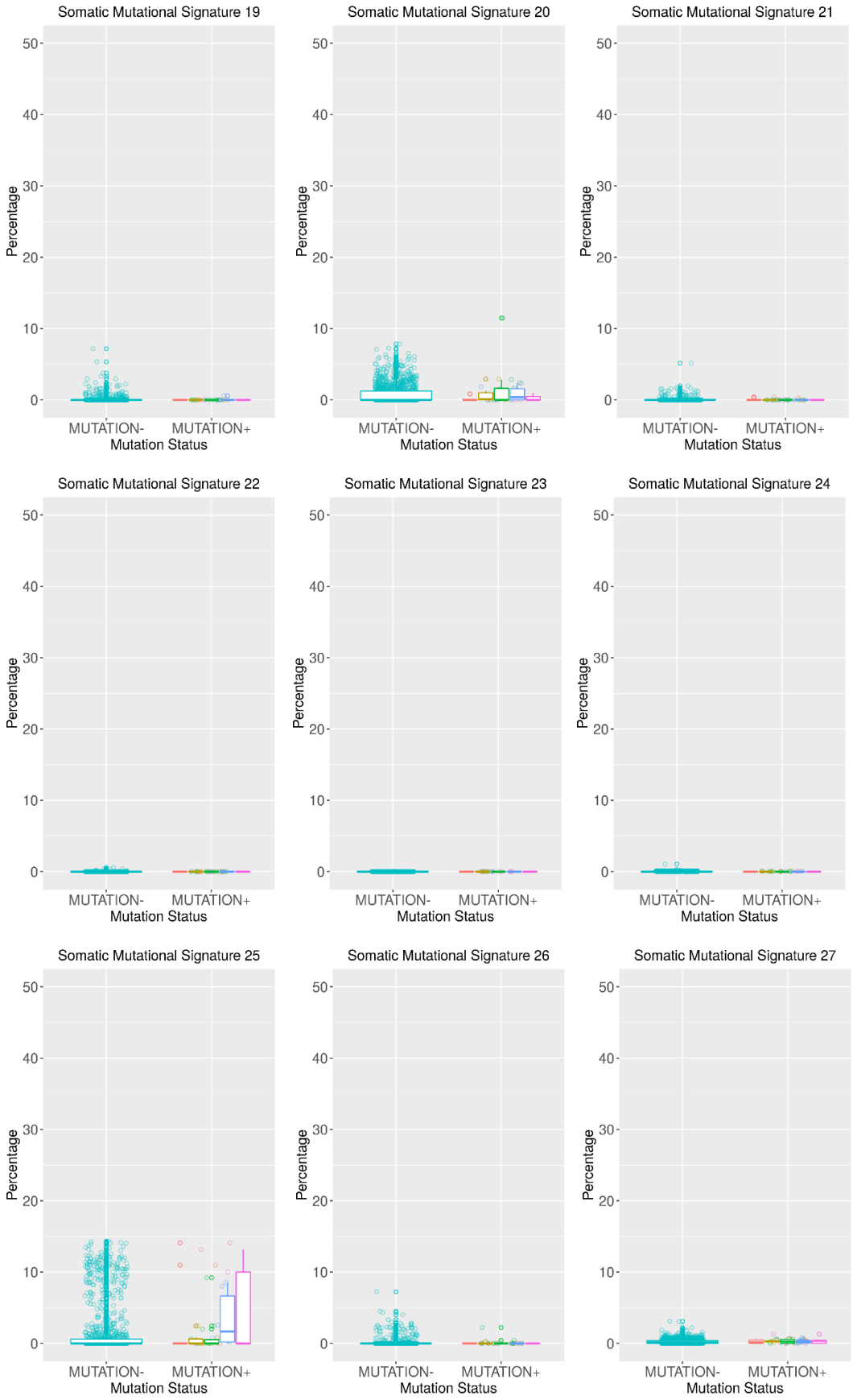


Figure 45: Single Base Substitution Somatic Mutational Signatures: A Comparison Between Gene Carriers and Non-Carriers.

Graphical representation of the differential expression of Single Base Substitution Somatic Mutational Signature 1-30 between gene carriers (Mutation+, ATM+, BRCA+, CHEK2+, PALB2+ and TP53+) compared to non-carriers (Mutation-).


Somatic Mutational Signature 1
Somatic Mutational Signature 2
Somatic Mutational Signature 3
Somatic Mutational Signature 4
Somatic Mutational Signature 5
Somatic Mutational Signature 6
Somatic Mutational Signature 7
Somatic Mutational Signature 8
Somatic Mutational Signature 9
Percentage
MUTATION-
MUTATION+
Mutation Status


Somatic Mutational Signature 10
Somatic Mutational Signature 11
Somatic Mutational Signature 12
Somatic Mutational Signature 13
Somatic Mutational Signature 14
Somatic Mutational Signature 15
Somatic Mutational Signature 16
Somatic Mutational Signature 17
Somatic Mutational Signature 18
Percentage
0
10
20
30
40
50
MUTATION-
MUTATION+
Mutation Status


Somatic Mutational Signature 19
Somatic Mutational Signature 20
Somatic Mutational Signature 21
Somatic Mutational Signature 22
Somatic Mutational Signature 23
Somatic Mutational Signature 24
Somatic Mutational Signature 25
Somatic Mutational Signature 26
Somatic Mutational Signature 27
Percentage
0
10
20
30
40
50
MUTATION-
MUTATION+
Mutation Status

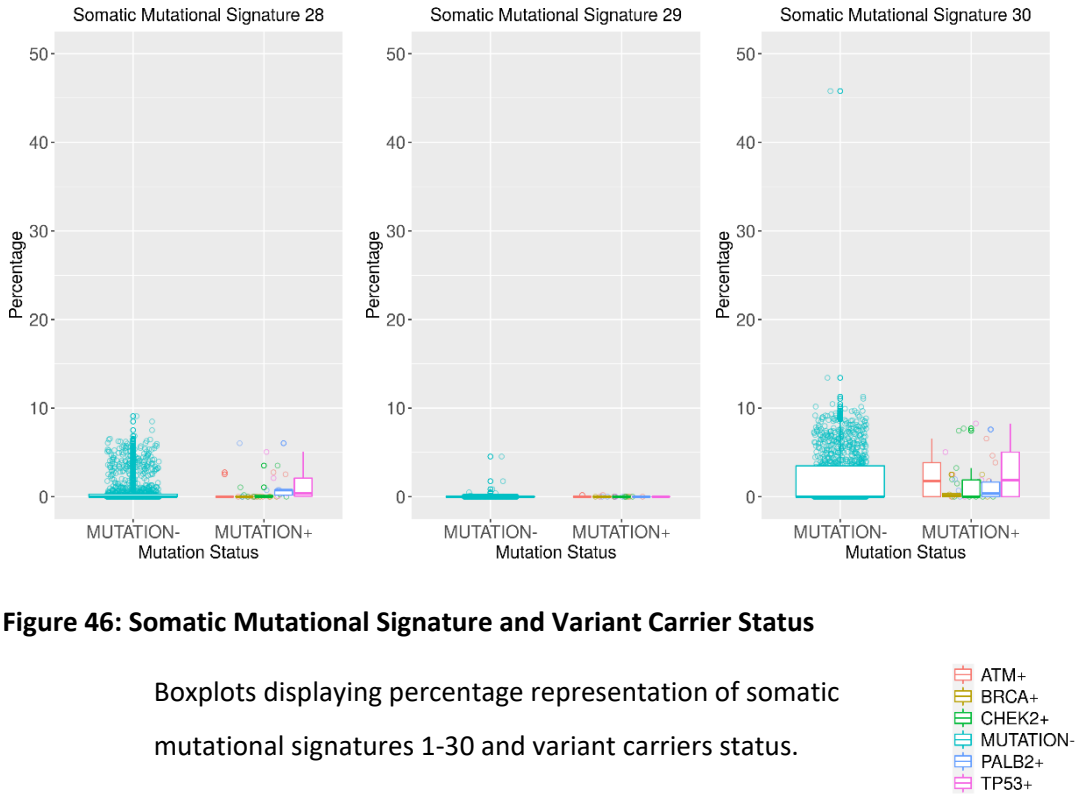


Figure 46: Somatic Mutational Signature and Variant Carrier Status

Boxplots displaying percentage representation of somatic mutational signatures 1-30 and variant carriers status.

10.4 Discussion

This work has observed the somatic mutational profile of a large number of unselected breast cancer cases and compared high and moderate penetrance gene carriers against non-carriers. We have demonstrated that somatic coding variants in *TP53*, *PIK3CA*, *GATA3*, *MAP3K1* and *CDH1* have a high prevalence in breast cancer. *TP53*, *PIK3CA*, *MAP3K1* and *CDH1* have recognised hallmark functions and as such, they are likely to represent key genetic components in breast cancer evolution. This is consistent with the published literature. In 2010, Kan et al. observed somatic mutations in 1507 genes across 183 breast cancers.(244) They identified that *PIK3CA* and *TP53* were the most frequently mutated in breast cancer.(244)

10.4.1 Tumour Mutational Burden

We have shown that the majority of invasive breast cancers are not associated with a high Tumour Mutational Burden (TMB) or hypermutation. This is reflected in the median and IQR of TMB observed within this cohort and is consistent with the published literature. In 2010, Kan et al. observed somatic mutations in 1507 genes across 441 primary invasive tumours comprising 183 breast cancers, 134 lung cancers, 58 ovarian, 58 prostate and 8 pancreatic with subgroup analysis based upon the underlying histopathological phenotype.(244) They identified 2576 somatic mutational events with an average of 1.8 somatic mutations per Mb of DNA analysed across all cancer types.(244) Invasive breast cancers displayed an average of 1.14 somatic mutations per Mb

of DNA.(244) In 2013, Alexandrov et al. observed the somatic mutational profile of 7042 cancers. The median TMB observed in breast cancers cases was approximately 1.0 Mut/Mb.(165)

A small proportion of breast cancers occurring in the absence of a germline variant were hypermutated with an average TMB greater than 10 Mut/Mb. In 2017, Chalmers et al. examined the Tumour Mutational Burden across 102,292 Cancer genomes derived from TCGA across a wide range of cancer types.(245) They identified that advancing age at diagnosis was associated with a significant increase in the TMB.(245) Within the 100,000 Genomes Project, the median age of breast cancer onset was 61.0 years (IQR, 51.0-70.0 years). Thus advancing age may explain, in part the broad range of TMB observed amongst unselected breast cancer cases.

Whilst the majority of invasive breast cancers were not hypermutated, individuals with germline variants in *BRCA* and *PALB2* demonstrate a significantly higher TMB compared to non-gene carriers. Alterations in the *BRCA1* and *PALB2* genes affect the Homologous Recombination Repair pathway. There is emerging literature to support this observation. In 2018, Thomas et al. observed the Tumour Mutational Burden amongst 930 primary breast tumours derived from TCGA.(246) They found the mean TMB amongst *BRCA* gene carriers to be 1.63/Mb which was considered to be TMB high and equivalent to the 80th percentile within the cohort overall.(246) Furthermore, in 2020, Mei et al. observed the TMB in 62 advanced breast cancer patients.(247) They found that patients with somatic mutations in DNA damage repair genes such as *BRCA* and *PALB2* had significantly higher levels of TMB.(247) In 2019, Lal et al. analysed 560 breast cancers with germline or somatic inactivation of either *BRCA1* or *BRCA2*. They demonstrated that individuals with *BRCA1* or *BRCA2* inactivated tumours had a significantly higher TMB compared to sporadic breast cancer.(248)

Similar observations have been noted in association with *PALB2* germline variants. In 2019 Li et al. conducted somatic mutation analysis of 16 invasive breast cancers derived from *PALB2* mutation carriers.(47) They demonstrated a median of 113.5 somatic mutations per case with a range of 59-269 and found that tumours displaying biallellic loss of *PALB2* had a higher somatic mutation rate (median 139.5, range 63–269 mutations per Mb).(47)

Overall, this supports the observation that germline variants associated with defects in the HRR pathway may produce an increase in TMB. The association between germline variation and increased TMB is well described for specific variant classes such as the DNA Mismatch Repair genes (MMR) and DNA polymerases including POLE. In 2017, Zehir et al. observed the TMB in 10,000 patients with metastatic cancer.(249) Hypermutation was observed in association with MMR and POLE signatures.(249) In 2017, Campbell et al. performed comprehensive pan cancer somatic mutational analysis of over 78,452 adult cancers and 2885 paediatric cancers.(170) The

median TMB in paediatric cancers was 2.50 Mut/Mb.(170) They found that all paediatric patients with hypermutant cancers and a strong replication repair mutational signature had a confirmed germline variant in an MMR gene or POLE.(170) Hypermutation was also observed in 17% of adult cancers and was associated with microsatellite instability and somatic variants in MMR genes and DNA polymerases such as POLE and POLD1.(170)

The magnitude of effect on TMB may be smaller for germline variants associated HRR deficiency compared with MMR deficiency or aberrant DNA polymerase activity. However, the overall number of variant carriers within this analysis is small. As such, larger sample sizes are required to draw definitive conclusions.

This potential association between elevated TMB and HRR deficit may have clinical relevance. Somatic mutations produce neo-antigens which can induce a T-cell mediated immune response.(169) A higher TMB is associated with higher levels of neoantigen presentation and immunogenicity which creates a selection pressure for the expression immune check point inhibitors such as anti PD-1 and PD-L1 within the tumour.(169) The expression of PD-1 and PDL-1 is linked with responsiveness to immune check point modulation therapy.(169) The use of Immune Check Modulation therapy such as Nivolumab is well described for MMR deficient colorectal cancer.(250) Indeed, TMB has been considered a biomarker for precision therapy as PDL-1 expression on immunohistochemistry alone is often unreliable.(251)

This work suggests that breast cancers that occur in association with a *BRCA* or *PALB2* germline variant may have immunogenicity which will enable responsiveness to immune checkpoint modulation. In support of this observation, in 2020, Mei et al. reported that TMB positively correlated with the percentage prevalence of Tumour Infiltrating Lymphocytes (TILs) amongst *BRCA* and *PALB2* gene carriers.(247) Furthermore, a meta-analysis published by Zou et al. in 2020 observed the efficacy of Immune Checkpoint Inhibitor Therapy in metastatic breast cancer across 27 studies.(252) They concluded that PDL-1 expression and high TIL levels may predict responsiveness to Immune Checkpoint Inhibitor Modulation.(252) In 2018, Matsuo et al. published a case series of 6 women with recurrent epithelial ovarian cancers occurring in the context of a germline *BRCA* variant.(253) They were managed with Nivolumab monotherapy and followed up for a period of 13.4 months. In total, 3 women demonstrated a complete response and 1 woman, a partial response.(253)

10.4.2 Somatic Mutational Signatures

This represents one of the first somatic mutational signature analyses comparing high and moderate penetrance variant carriers with non-carriers. The 9 main SBS mutational signatures

observed amongst breast cancer cases within this cohort included SBS1, SBS2, SBS3, SBS5, SBS8, SBS9, SBS12, SBS13 and SBS16. This is consistent with the published literature. In 2013, Alexandrov et al. analysed 4,938,362 mutations derived from 7042 cancers including breast cancer to determine whether particular mutation types were more frequently observed within different cancer subtypes.(165) In 2020, Alexandrov et al. expanded this analysis and undertook somatic mutational analysis on 23829 cancer samples including 4645 set of whole genome somatic sequence data.(178) Across these analyses, specific SBS mutational signatures were consistently identified amongst invasive breast cancer cases including SBS1, SBS2, SBS3, SBS5, SBS8, SBS9, SBS13, SBS17 and SBS18.(165, 178) In 2017, Polak et al. observed the somatic mutational signatures amongst 995 breast cancers derived from the Cancer Genome Atlas.(254) They identified 4 recurrent mutational signatures, SBS1, SBS2, SBS3, SBS6 and SBS13.(254) It suggests that there are shared mutational processes implicit in breast cancer evolution for both breast cancer variant carriers and non-carriers.(248)

SBS1 and SBS5 are observed across multiple cancer types.(180, 255) SBS1 is associated with the endogenous spontaneous deamination of 5-methylcytosine at CpG dinucleotides to thymine resulting in a G to T mismatch.(255) The biological process underlying signature 5 remains to be elucidated.(180, 255) The number of the single nucleotide conversions associated with SBS1 and SBS5 correlates with the number of cell divisions over time. (255) In 2015, Alexandrov et al. demonstrated the continuous “clockwise” accumulation of SBS1 and SBS5 somatic mutations and that this correlated with increasing age.(178, 255)

We have shown that SBS1 and SBS5 are identifiable in all breast cancers. However, a significantly lower contribution of SBS1 and SBS5 was observed amongst individuals with germline variants in *BRCA* and *PALB2*. This may be partly attributable to an age related effect. The median age at breast cancer diagnosis was significantly younger for *BRCA* gene carriers compared to non-carriers within this cohort (50.0 years (IQR, 47.5-53.0 years) for *BRCA* gene carriers versus 61.0 years (IQR, 51.0-71.0) for non-carriers (p=0.044)). As such, the time interval to acquire SBS1 would be reduced amongst *BRCA* gene carriers resulting in a lower overall contribution within the tumour. In 2019 Lal et al. observed the somatic mutation signatures in 560 breast cancer cases.(248) They also observed underrepresentation of signature 1 in *BRCA1* and *BRCA2* deficient tumours compared to sporadic tumours (6.0% *BRCA1*, 6.6% *BRCA2* AND 16% sporadic).(248)

There was no significant difference in the age of breast cancer onset between *PALB2* gene carriers and non-carriers and they still manifest significantly lower levels of SBS1 and SBS5. Therefore, it would also suggest that the continuous biological acquisition of somatic variation associated with SBS1 and SBS5 is less important in the process of tumorigenesis for individuals with mutations in

genes associated with HRR deficit.(178) It remains to be determined whether there are other biologically relevant mutational processes for individuals with HRR deficit. In 2019, Lal et al. concluded that *BRCA* deficient tumours displayed elevated levels of structural variation which would not be recognisable on SBS mutational signature analysis alone.(248)

SBS2 and SBS13 are frequently observed across a variety of cancer subtypes and are an important contributor to breast cancer evolution.(178, 180) Consistent with this, SBS2 and SBS13 were observed amongst both gene carriers and non-carriers within this cohort. These signatures are associated with the endogenous upregulation of APOBEC cytidine deaminase activity and occurs in episodic bursts.(180) The mechanism for APOBEC cytidine deaminase upregulation is not fully understood but proposed hypotheses include viral infection and tissue inflammation.(256)

SBS3 is an important focus for mutational signature analysis amongst individuals with breast cancer. This mutational signature is associated with Homologous Recombination Repair (HRR) deficit. If HRR is defective then non-homologous end joining is frequently exploited to repair DNA double strand breaks.(164) This is a more error prone repair mechanism and produces a characteristic pattern with indels at breakpoint junctions.(164)

This work has shown that SBS3 is observed in the majority of the breast cancers with a proportional representation of approximately 10% in the absence of an underlying germline variant. We have also shown that germline variants in *BRCA* and *PALB2* are associated with a significant increase in SBS3 compared to non-carriers with a proportional representation of approximately 30%.

The association between *BRCA* germline variants and SBS3 is well described. In 2016, Nik-Zainal et al. observed the somatic mutational profile derived from whole genome sequencing amongst 560 breast cancer cases and compared this to non-neoplastic tissue.(173) They found that expression of SBS3 was associated with germline, somatic or epigenetic inactivation of *BRCA1* and *BRCA2* function.(173) In 2017, Polak et al. observed the somatic mutational signatures amongst 995 breast cancers derived from the Cancer Genome Atlas.(254) A significant increase in SBS3 was observed amongst those individuals with a germline or somatic loss of function variant in *BRCA1* or *BRCA2* ($p<0.0001$).(254) In 2017, Davies et al. demonstrated that a combination of mutational signatures including SBS3 utilising the HRDetect model could predict germline or somatic inactivation of *BRCA1* or *BRCA2* with 98.7% sensitivity.(257)

The association with germline *PALB2* variants and SBS3 is more recently characterised with limited available literature. In 2017, Polak et al. observed SBS3 in association with germline variants in *PALB2*.(254) In 2018, Lee et al. analysed the somatic mutation signature amongst 15

invasive breast cancers associated with germline loss of function mutations in *PALB2* derived from familial cancer clinics in Australia.(230) They found that tumours occurring in the context of a *PALB2* germline variant were significantly more likely to demonstrate signature 3 than sporadic tumours.(230) However, they could not determine whether there was a differential effect of biallelic or monoallelic loss of *PALB2* on the relative prevalence of SBS3.(230) In 2019, Li et al. in 2019 sought to characterise the somatic mutational repertoire of germline *PALB2* mutation carriers amongst 24 invasive breast cancers derived from 14 distinct pathogenic gene alterations.(47) They found that only tumours with biallelic, not monoallelic loss of *PALB2* had a mutational signature (SBS3) consistent with HRR.(47)

Germline variants in *ATM*, *CHEK2* or *TP53* were not associated with a significant increase in signature 3. Polak et al. also found that SBS3 was not associated with germline variants in *CHEK2* or *ATM*.(254) There is limited additional literature available regarding somatic mutational signature expression amongst *CHEK2*, *ATM* and *TP53* gene carriers. In 2018, Weigelt et al. observed somatic mutational profile in 24 ATM-associated breast cancers and found them to be distinct from BRCA without high levels of SBS3 expression.(161) Within this series there were no further defining SBS mutational signatures which could be used to identify *CHEK2*, *ATM* or *TP53* gene carriers.

Whilst SBS8, SBS12 and SBS16 were consistently observed amongst breast cancers within this series, the underlying biological mechanism for expression remains to be determined and there were no significant differences between gene carriers and non-carriers.

10.5 Summary

The repertoire of somatic mutation in human cancer is vast including a combination of driver mutations which perpetuate the hallmark characteristics of cancer and passenger mutations which enable the elucidation of the molecular processes underlying tumorigenesis. The findings in this somatic mutational analysis complement and strengthen the current literature.

We have shown that 9 SBS Mutational Signatures are important for the evolution of breast cancer in both gene carriers and non-carriers. The molecular pathways underlying the evolution of *CHEK2*, *ATM* and *TP53*-associated cancer were not significantly different from sporadic breast cancer in this study. However, we have illustrated that tumours occurring in association with a germline variant in *BRCA* or *PALB2* have a significantly higher Tumour Mutational Burden and a significantly greater proportional representation of Signature 3 compared to non-carriers.

It suggests that particular molecular pathways (represented by the increased expression of SBS3 and reduction in expression of SBS1 and SBS5 amongst *PALB2* and *BRCA* gene carriers) are more important for the development of tumours occurring in association with Homologous Recombination Deficit compared to other high and moderate risk genes and non-carriers.

These differences can be exploited for therapeutic intervention including PARP inhibitors and PDL modulation therapy. Indeed, they may serve as biomarkers for therapeutic intervention. The presence of SBS3 in combination with a higher TMB may have utility in predicting germline variation and identifying actionable risk. It may also prove a useful adjunct for current classification systems to interpret VUS. These findings are particularly important as the utilisation of somatic testing increases in routine practice.

Chapter 11 Summary: Can Germline Genotype Influence Tumour Phenotype and Outcome

11.1 Variant Identification

As we progress into the genomic era, gene panel testing has the potential to be used increasingly to identify heritable risk with potential health economic benefit through primary prevention and precision management.(135, 211) This work represents a unique series detailing the prevalence of pathogenic and likely pathogenic variants in several high and moderate risk breast cancer susceptibility genes presented across three distinct breast cancer cohorts (familial, under 40 years and unselected breast cancers).

It has shown that 16.7% of individuals with symptomatic early onset breast cancer have a moderate or high penetrance gene variant compared to 2% of unselected breast cancers. It has also shown that pathogenic and likely pathogenic variants in *BRCA1* and *BRCA2* are the most common single gene alterations associated with heritable susceptibility and that the additional testing of *CHEK2*, *PALB2*, *ATM* and *TP53* produces a diagnostic uplift of 1.7%-7.3% depending upon the tested cohort.

Histopathological subtype specific analysis demonstrates that 25.4% of individuals with a symptomatic, early onset TNT will have a germline breast cancer susceptibility variant and the majority of these (86.6%) will be in *BRCA1*. In the absence of a TNT, germline breast cancer susceptibility variant identification amongst unselected early onset breast cancer is still 14.2%. The majority of these variants will be in *BRCA1* and *BRCA2* (9.3%) (Appendix J.1).

HER2-amplified breast cancers were present in 658/2744 (24.0%) of the POSH cohort. Pathogenic variants in *TP53* were identifiable in 1.5% of the whole, HER2-amplified cohort and 4.7% of those with HER2-amplified breast cancer diagnosed under the age of 30 years. Family history was an important modifier of the likelihood of identifying a pathogenic variant in *TP53*. In the absence of a strong family history of breast cancer, less than 2% of individuals with very early onset HER2-amplified breast cancer had a pathogenic variant in *TP53* compared to 12.5% in the context of familial breast cancer.

Current recognised indications for genetic testing including familial breast cancer, triple negative breast cancer occurring under the age of 60 years and very early onset (under 30 years) HER2-

amplified breast cancer. This is advocated by the Genomic Medicine Testing Directory and the published literature.(142)

The outcome of this analysis also supports the utilisation of gene panel testing to identify actionable risk within early onset breast cancers occurring under the age of 40 years but not amongst unselected breast cancers. This is an advancement on the current Genomic Medicine Testing Directory which only advocates testing unselected breast cancers that are diagnosed under the age of 30 years.(142) A small number of studies have now shown that gene panel testing for hereditary breast cancer susceptibility can produce a cost effective improvement in Quality Adjusted Life Years (QALY) and Life Expectancy through the identification of actionable risk. (212, 213) As such, genetic testing in this context may produce a cost benefit for the identification of actionable risk amongst high penetrance gene carriers.

Analysis of this cohort also suggests that all individuals with early onset (under 40 years) HER2-amplified breast cancer should be tested for alteration in the *TP53* if the breast cancer is diagnosed in the context of a strong family history of breast cancer. However, caution must be applied when there is no family history of cancer. This is even true amongst those individuals with HER2-amplified breast cancer under the age of 30 years where reflex testing is the current standard across many Clinical Genetics Services. For this group, it is important to consider the potential psychological burden of identifying variation of unclear significance and the pitfalls in determining absolute cancer risks in the absence of a classic LFS/LFL family history where the likelihood of this is potentially greater than finding actionable risk.

Despite the potential benefits of genetic testing, we must approach this technology with care, particularly in relation to variant interpretation and ensuring appropriate medical intervention for the associated risk. Future advances in centralised variant databases and curation along with multi-tier reporting have the potential to mitigate some of these risks and truly harness the potential of genomic technology for heritable cancer susceptibility.(214, 258)

11.2 Germline Genotype and Tumour Phenotype

This work has provided one of the most comprehensive overviews of tumour histopathology in early onset breast cancer comparing high and moderate penetrance gene carriers with non-carriers (Figure 47). It has also provided an overview of somatic mutational profile including TMB and SBS Mutation Signatures in relation to germline genotype (Figure 48).

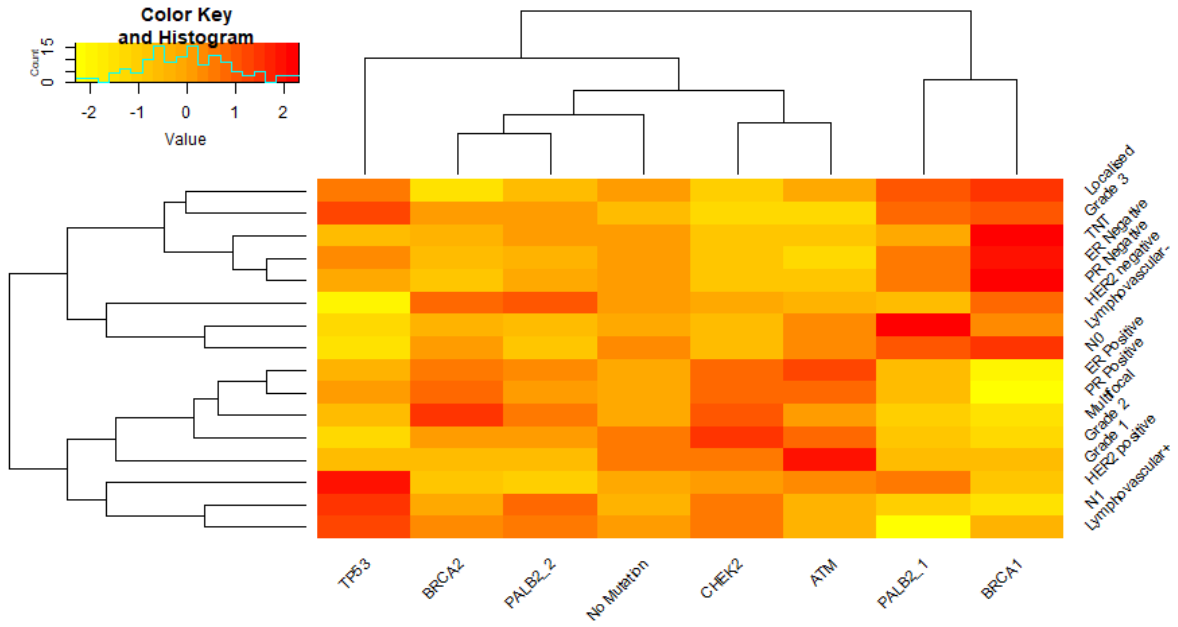


Figure 47: Dendogram Comparing Tumour Histopathological Phenotype Between Gene Carriers and Non-Carriers

Dendogram to demonstrate the hierarchical clustering between tumour histopathological phenotypic characteristics and germline genotype. PALB2_1 represents the PALB2 BRCA1 binding domain whilst PALB2_2 represents the PALB2 BRCA2 binding domain.

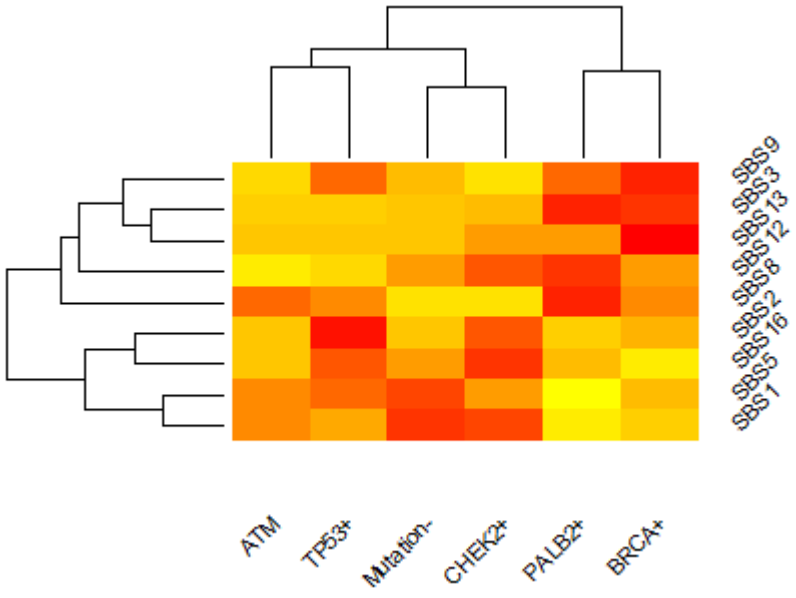


Figure 48: Dendogram Comparing Tumour SBS Mutational Signature Between Gene Carriers and Non-Carriers

Dendogram to demonstrate the hierarchical clustering between tumour SBS Mutational Signature and germline gene variants.

11.2.1 Homologous Recombination Repair Deficit and Tumour Phenotype

Pathogenic variants in *BRCA1*, *BRCA2* and *PALB2* are associated with aberrant Homologous Recombination Repair. As such, their histopathological characteristics and somatic mutational profile have been considered together for summary analysis.

11.2.1.1 Histopathology *BRCA1* and *BRCA2*

In accordance with the current literature, this work has shown that *BRCA1* related tumours are significantly more likely to be grade 3, ER-negative, PR-negative and HER2-negative at presentation compared to non-carriers. Conversely, *BRCA2* associated tumours are significantly more likely to be grade 3 ER-positive, PR-positive and HER2-negative (Figure 47).(149, 153, 154, 218, 219)

This work has also defined the importance of tumour focality in the prediction of *BRCA* carrier status. *BRCA1* associated tumours within this cohort were significantly more likely to be localised with lower levels of nodal involvement and lymphovascular infiltration at presentation compared to non-carriers with smaller maximum overall tumour size. Conversely, *BRCA2* associated tumours were significantly more likely to be multifocal with a non-significant trend towards higher levels of nodal involvement compared to non-carriers with a significantly greater maximum overall tumour size (Figure 47). The significance of focality, specifically the association between multifocality and *BRCA2* associated tumours is a novel histopathological association.(259)

11.2.1.2 Histopathology *PALB2*

Consistent with the published literature, this study observed no significant difference in hormone receptor status between *PALB2* variant carriers and non-carriers.(47, 56, 230) However, a higher proportion of individuals with *PALB2* variants within the *BRCA1* binding domain presented with ER-negative tumours whilst a higher proportion of individuals with *PALB2* variants in the *BRCA2* binding domain presented with ER-positive tumours. Tumours occurring in association with a mutation in the *BRCA1* binding domain were also significantly smaller than those associated with mutations in the *BRCA2* binding domain and a higher proportion were localised with lower levels of nodal involvement and significantly lower levels of lymphovascular infiltration (Figure 47).

There is a paucity of published literature to compare a domain specific effects on histopathological characteristics between *PALB2* variant carriers and non-carriers. However, a small number of studies have observed a higher proportion of ER-negative tumours associated with variants in the *BRCA1* binding domain or proximal to the *BRCA2* binding domain.(53, 66)

The observed potential association with mutation domain and tumour histopathology is consistent with what is already understood about *BRCA1* and *BRCA2* associated hormone receptor status and the associated histopathology observed within this prospective cohort.(219)

This work has also shown hierarchical similarity between *BRCA1* and variants in the *BRCA1* binding domain of *PALB2* versus *BRCA2* and variants in the *BRCA2* binding domain of *PALB2* (Figure 47). These histopathological features were not sufficiently predictive to differentiate gene carriers from non-carriers. The numbers are however small and this represents an area for further evaluation of the *PALB2* domain specific phenotype through meta-analysis of pooled international data.

11.2.1.3 Homologous Recombination Repair Deficit and Somatic Mutational Profile

This work has shown that the majority of invasive breast cancers are not associated with a high Tumour Mutational Burden (TMB) or hypermutation. This is reflected in the median and IQR of TMB observed within this cohort and is consistent with the published literature.(165, 244)

Whilst the majority of invasive breast cancers were not hypermutated, individuals with a germline variant in *BRCA* and *PALB2* had a significantly higher TMB compared to non-gene carriers. There is emerging literature to suggest that defects in the HRR pathway may produce an increase in TMB.(246-248) The association between germline variation and increased TMB is well described for specific variant classes such as the DNA Mismatch Repair genes (MMR) and DNA polymerases including POLE.(170, 249) The magnitude of effect on TMB may be smaller for germline variants associated HRR deficiency compared with MMR deficiency or aberrant DNA polymerase activity. However, the overall number of variant carriers within this analysis is small. As such, larger sample sizes are required to draw definitive conclusions.

The 9 main SBS mutational signatures observed amongst breast cancer cases within this cohort included SBS1, SBS2, SBS3, SBS5, SBS8, SBS9, SBS12, SBS13 and SBS16. This is consistent with the published literature.(165, 178, 254) It suggests that there are shared mutational processes implicit in breast cancer evolution for both breast cancer variant carriers and non-carriers (Figure 48).(248)

This work has demonstrated that germline variants in *BRCA* and *PALB2* are associated with a significant increase in SBS3 compared to non-carriers with a proportional representation of approximately 30% (Figure 48). The association between *BRCA* germline variants and SBS3 is well characterised and the association with germline *PALB2* variants and SBS3 is more recently described in a small number of publications.(47, 173, 230, 254)

Despite the association with germline variants in *BRCA* and *PALB2*, SBS3 is still observed in the majority of the breast cancers with a proportional representation of approximately 10% (Figure 48). As such, the proportional representation of SBS3 may be more important than its presence in isolation for the identification of germline variation.

We have shown that SBS1 and SBS5 are identifiable in all breast cancers. However, a significantly lower contribution of SBS1 and SBS5 was observed amongst individuals with germline variants in *BRCA* and *PALB2*. SBS1 is associated with the endogenous spontaneous deamination of 5-methylcytosine at CpG dinucleotides which occurs in a continuous manner with advancing age.(178, 255) The biological process underlying signature 5 remains to be elucidated but it is acquired in a similar continuous manner with advancing age.(180, 255) Lower levels of SBS1 and SBS5 expression is association with HRR deficit has been described in another publication and it would suggest that the continuous biological acquisition of somatic variation associated with SBS1 and SBS5 is less important in the process of tumorigenesis for individuals with mutations in genes associated with HRR deficit.(248)

Overall, the somatic mutational analysis suggests that it is the collective proportion representation of SBS Mutational Signatures that is more important that the consideration of each signature in isolation for the identification of germline variation and interpretation of VUS (Figure 48). In 2017, Davies et al. demonstrated that a combination of mutational signatures including SBS3 in a model known as HRDetect could predict germline or somatic inactivation of *BRCA1* or *BRCA2* with 98.7% sensitivity.(257) The utilisation of SBS Mutational Signatures and TMB to identify germline *BRCA* and *PALB2* gene carriers is an important focus for future work and validation. If validated, it could be used for the interpretation of VUS and has the potential to identify individuals with cryptic intronic variation which is not identifiable by conventional genetic testing through RNA seq testing amongst those with a somatic mutational profile suggestive of HRR deficit.

11.2.1.4 Using Tumour Histopathological Phenotype to Identify Germline Variants Associated with Homologous Recombination Repair Deficit

Overall, we have shown that tumour histopathology can be used to differentiate *BRCA* gene carriers from non-carriers. Hormone receptor status and tumour focality were significant independent histopathological predictors of *BRCA1* and *BRCA2* carriers. When these features are identified in a sequential combination it can increase the likelihood of a *BRCA1* or *BRCA2* variant. Overall, the tumour phenotype associated with the highest probability of identifying a pathogenic *BRCA1* variant would be observed amongst those patients with a localised, ER-negative, PR-negative, HER2-negative tumours in the presence of a family history of breast cancer. Conversely,

the highest probability of identifying a pathogenic *BRCA2* variant would be observed amongst those with a multifocal, ER-positive, PR-positive, HER2-negative tumour in the presence of a family history of breast cancer.

In 2014, Spurdle et al. conducted a large-scale analysis of 4477 *BRCA1* mutation carriers, 2565 *BRCA2* mutation carriers and 47565 breast cancers in non-*BRCA* carriers in 2014.(154) They used ER status, age and grade to provide predictors of *BRCA1* and *BRCA2* mutations. They found that a combination of age, grade and ER receptor status increased the likelihood of a pathogenic mutation than the presence of each factor in isolation.(154)

Tumour somatic mutational profile including TMB and SBS Mutational signature profile may also help to differentiate *BRCA* and *PALB2* gene carriers from non-carriers and may prove a useful addition to histopathological assessment.

Using these principles, tumour phenotypic characteristics including histopathology and somatic mutation profile including focality, hormone receptor status, TMB and SBS mutational signature could be incorporated into a carrier risk stratification model to determine the likelihood of *BRCA, BRCA2* and *PALB2* germline predisposition to provide more accurate identification of those individuals who would benefit from germline genetic testing.(154) This is important as 35.8% of *BRCA1* mutation carriers, 34.1% of *BRCA2* mutation carriers and 55.6% of *PALB2* mutation carriers had no family history of cancer. These individuals had the same histopathological phenotype as variant carriers with a family history but did not meet the genetic testing threshold using conventional carrier probability models.

These histopathological and tumour somatic characteristics could also contribute to the evidence used for the interpretation of VUS. For example, the CanVIG Specification for the ACMG Guidelines for variant interpretation provides the opportunity to utilise specific tumour histopathological phenotypic information as contributing evidence for the pathogenicity of variants using PP4.(258) If validated, SBS mutational signature and TMB may serve as useful phenotypic adjunct for classifying variants in genes associated with HRRD including *PALB2* and *BRCA*.

11.2.2 Tumour Phenotype *TP53*, *CHEK2* and *ATM*

This work has shown that individuals with pathogenic variants in *TP53* are significantly more likely to present with HER2-amplified, high-grade tumours demonstrating nodal involvement and a trend towards lymphovascular infiltration (Figure 47). This is consistent with the published literature. (159, 238) We have also shown that early age at diagnosis, N1 stage disease and HER2

amplification can be used to contribute towards the prediction of a *TP53* germline variant. Therefore, the presence of a HER2-amplified breast cancer may serve as an indication for *TP53* gene testing when considered in the context of the age at diagnosis, nodal status and broader family history of cancer. It may also serve as an adjunct for variant interpretation.

This work has shown that *CHEK2*-associated breast cancers are significantly more likely to be Grade 2, ER and PR-positive with no difference in HER2 expression. They are not associated with a TNT phenotype. Invasive breast cancers occurring in the context of a pathogenic *CHEK2* variant demonstrated a trend towards multifocality with significantly higher levels of nodal involvement at presentation. It highlights potential phenotypic similarity with *BRCA2* associated invasive breast cancers. Only nodal status was identified as a significant independent predictor of *CHEK2* genotype. Therefore, whilst there is an emerging histopathological *CHEK2* related tumour phenotype, this cannot be reliably used to identify *CHEK2* gene carriers versus non-carriers.(91, 157)

Consistent with the published literature, *ATM*-associated breast cancers were significantly more likely to be ER-positive.(152, 231) ER positivity was also a significant independent predictor of germline mutational status. However, there were no additional observed histopathological associations. As the majority of sporadic breast cancers are also oestrogen receptor positive, this feature cannot be used in isolation to predict gene carriers versus non-carriers.

There were no somatic mutational features which could be used to differentiate *CHEK2*, *ATM* or *TP53* gene carriers from non-carriers. However, it is notable that there was a higher representation of SBS mutational signatures associated with environmental exposure including SBS1, SBS2 and SBS5 amongst *TP53*, *CHEK2* and *ATM* gene carriers (Figure 48). This indicates that environmental exposures may be more important for cancer evolution amongst these gene carriers compared to individuals with germline variants associated with HRR deficit.

11.3 Multiple Pathogenic Variants

Overall, 1.3% of individuals presenting with symptomatic early onset breast cancer had multiple pathogenic variants. Buys et al. also reported that multiple pathogenic variants were identified in 1.3-3.3% of cases.(119) Within this study, each of the identified variants had a clinical actionability which was independent of the other variant. It highlights that for a small number of families two or even three high and moderate penetrance genetic factors may contribute to overall risk for specific individuals. This could be identified though breast cancer susceptibility gene panel testing amongst individuals presenting for genetic testing.

A summary of the clinical features for those individuals with multiple pathogenic variants is provided in Appendix J.2. Of these, 3/6 (50%) presented below the average age of diagnosis within the POSH cohort (age 26 years and age 33 years respectively). In total, 2/6 (33.3%) developed an ipsilateral recurrence and 2/6 (33.3%) developed a second primary malignancy. There is limited literature regarding patients with multiple pathogenic variants and whether we should manage people with multiple pathogenic variants differently to those with a single variant, in particular how multiple variants influence absolute risk calculations and whether it creates an increase on gene specific absolute cancer risks or the likelihood of developing a second primary cancer.

11.4 Germline Variants and Outcome

We have determined that patients with a germline *CHEK2* pathogenic variant who develop breast cancer have an adverse outcome with reduced OS and DDFS compared to those without, a relationship which persists after adjustment for known prognostic factors. Our results are consistent with Schmidt et al. and Wesicher et al. who also observed worse survival amongst *CHEK2* variant carriers .(80, 91) The contralateral breast cancer rate amongst *CHEK2* pathogenic variant carriers was almost twice that of non-carriers at both 5 and 10 years. Although the absolute numbers of cases was small, we noted that *CHEK2* carriers with a family history, had a contralateral breast cancer rate more than five times higher than non-carriers, whereas *CHEK2* carriers without a family history had no increase in risk. Within the POSH cohort, family history was not an independent predictor of outcome.(227)

Univariable analysis identified a non-significant trend towards improved OS amongst *PALB2* mutation carriers versus non-carriers within the first 5 years after cancer diagnosis although the numbers are small. Subgroup analysis was not possible due to the small sample size. However, *PALB2* carriers with variants located within the *BRCA1* binding domain were significantly more likely to be managed with breast conserving surgery which may be reflective of the trend towards smaller, more localised tumours observed within this sub-group.

Within this cohort, pathogenic variants in *ATM* were not associated with a significant difference in Overall Survival or Distant Disease Free Survival in comparison to non-carriers. The rates of contralateral breast cancer were also comparable between gene carriers and non-carriers.

Conversely, germline variants in *TP53* are associated with high rates of contralateral disease. In total, 20% of individuals with *TP53*-associated breast cancer developed contralateral breast cancer by 10 years. This increased to 40% in the presence of a family history of breast cancer. This was the highest observed rate of contralateral disease seen in association with a breast cancer

susceptibility gene alteration within this cohort. This is similar to the rates reported by Hyder et al. in 2020. (240) The substantial cumulative risk of contralateral disease observed within this study suggest that bilateral mastectomy may be a consideration for women with early onset breast cancer and a germline variant in *TP53*.

This work has shown no significant difference in OS and DDFS between *TP53* gene carriers and non-carriers. After correction for other known prognostic factors including tumour size and lymph node involvement. Whilst the overall numbers are small, the observed improvement following correction for other known prognostic factors may highlight the importance of early detection and intervention to improve survival amongst *TP53* gene carriers.

11.5 Strengths and Limitations

There are several strengths of this work. In the POSH study, patients were ascertained shortly after diagnosis through oncology clinics and presented with symptomatic rather than screen detected breast cancers. The clinical data available is comprehensive and allows multivariable data analysis. There was no systematic bias in selecting patients for the study or for genotyping and the cohort is representative of the general UK breast cancer population.(13) Furthermore, carriers of additional high-risk breast cancer susceptibility genes could be excluded from the analysis allowing a clean comparison between pathogenic variant carriers and non-carriers.

Whilst this work has many strengths the utilisation of second generation sequencing and an amplicon based targeted capture system within the POSH study limited the identification of CNVs. This is important as 5-10% of *BRCA1* and *BRCA2* mutations are dosage anomalies and this proportion is higher within specific populations.(260) Given this recognised limitation, patients with a high prior probability of a *BRCA* gene alteration based upon a strong family history or bilateral breast cancer were preselected for MLPA testing. Finally, although the initial sample size within the POSH study is large, the number of non-*BRCA* variants remains low due to the lower contribution of these variants to hereditary breast cancer. A larger sample size may help to strengthen associations between genotype and tumour histopathological phenotype particularly for non-*BRCA* genes such as *PALB2* and strengthen the validity of the findings.

Data derived from The 100,000 Genomes Project also had several strengths. Patients recruited to the Cancer, Breast Cancer Domain were also ascertained through oncology clinics following the detection of breast cancers whilst patients recruited via Rare Disease, Familial Breast Cancer required a strong family history of breast cancer as a prerequisite for recruitment. This created two distinct cohorts for analysis (unselected and familial breast cancer).

Although Whole Genome Sequencing was applied through The 100,000 Genomes Project, the bioinformatics pipeline was established to identify coding sequence variants for germline analysis. Furthermore, somatic analysis focused on Single Nucleotide Variation. It would be helpful to expand this to other types of variation including Copy Number Variation and Structural Change to better understand their relative contribution to breast cancer biology.

11.6 Future Research

This work, could also be expanded into older cohorts to see if the same relationships are observed. Previous studies suggest that the relative proportion of TN and ER and PR-negative *BRCA1* associated breast cancers reduces with age whilst the inverse relationship is seen with *BRCA2*.(35, 219) The proportion of HER2-negative tumours remains relatives stable in both *BRCA1* and *BRCA2* mutation carriers across all age groups.(138, 219)

This summary chapter has described how a combination of mutational signatures including SBS3 in a model known as HRDetect could predict somatic inactivation of *BRCA1* or *BRCA2* with 98.7% sensitivity.(257) Future research could focus upon whether a preselected combination of SBS mutational signatures, TMB and tumour histopathology could reliably identify germline variant carriers in *BRCA* and *PALB2*. If validated this analysis could be expanded to identify biologically relevant intronic variation through targeted RNA seq of those individuals with a somatic mutational profile of HRR deficit. It could also be used for the interpretation of VUS.

This work has shown that the context in which breast cancer is diagnosed (familial or unselected) may influence the age of breast cancer diagnosis with earlier age at diagnosis occurring in association with a strong family history of breast cancer. Polygenic risk may contribute to this observation and it would be useful to compare whether higher quintiles of PRS are observed in familial versus unselected breast cancer.

11.7 Conclusion

This work has shown that 16.7% of individuals with symptomatic early onset breast cancer will have a moderate or high penetrance gene variant compared to 2% of unselected breast cancers. Variants in *BRCA1* and *BRCA2* are the most common single gene alterations associated with heritable susceptibility and that the additional testing of *CHEK2*, *PALB2*, *ATM* and *TP53* produces a diagnostic uplift of 1.7%-7.3% depending upon the tested cohort.

This work has provided one of the most comprehensive overviews of tumour histopathology and somatic mutational profile comparing high and moderate penetrance gene carriers with non-

carriers. It has shown that tumour histopathology can be used to predict the likelihood of a germline *BRCA* or *TP53* variant carrier and this may serve as an adjunct to current methodologies for variant classification. TMB and SBS mutational signature profile may also identify HRR deficient tumours associated with germline *PALB2* and *BRCA* variants.

Genetics is increasingly utilised in cancer care for precision therapy and the identification of actionable risk. As we progress forwards with this technology including mainstream testing and somatic mutational profiling, tumour phenotype including both histopathology and somatic mutational profile will be important to develop our understanding of tumour biology as an adjunct to current utilised methodologies to identify actionable risk and interpret rare variation in high penetrance genes. Whilst the potential utility for high penetrance genes associated with HRR deficit is apparent, the utility of tumour phenotype for identifying moderate risk gene carriers and interpreting rare variation is less clear.

Appendix A Global Prevalence of *PALB2* Variants

Population	*PALB2* Variants		Case Frequency		Breast Cancer Cohort	Ref.
	Coding	Protein	Number	%		
UK	c.2718G>A c.3113G>A c.3116delA Other rare variants	p.Trp906* p.Trp1038* p.Asn1039Ilefs	89/13087	0.68	Unselected	(56)
Finland	c.1592delT	p.Leu531Cysfs	19/947	2.0	Familial	(53)
			8/1274	0.6	Unselected	
SW Poland & W Ukraine	c.509_510del	p.Arg170Ilefs	3/338	0.89	Familial Early onset	(61)
Poland	c.172_175delTTGT c.347insT c.509_510del	p.Gln60Argfs p.Leu116fs p.Arg170Ilefs	7/460	1.5	Familial Early Onset	(60)
Australia	c.172_175delTTGT c.196C>T c.522_523delAA c.577dupA c.693dupA c.758dupT c.860dupT c.1947dupA c.2386G>T c.2391delA c.2966_2967insCAACAAGT c.2982dupT c.3113G>A c.3116delA c.3256C>T c.3362delG c.3507_3508del c.3549C>G	p.Gln60Argfs p.Gln66* p.Arg175Thrfs p.Thr193Asnfs p.Gly232Argfs p.Ser254Ilefs p.Ser288Lysfs p.Glu650Argfs p.Gly796* p.Gln797Hisfs p.Glu990Asnfs p.Ala995Cysfs p.Trp1038* p.Asn1039Ilefs P.Arg1086* p.Gly1121Valfs p.His1170Phefs p.Tyr1183*	26/1996	1.3	Familial	(64)
China	c.715C>T c.1058delA c.1744C>T c.2323C>T c.2693G>A c.2748+1G>A c.2749-1G>C c.3114-1G>A c.3256C>T c.3271delC c.3507_3508delTC	p.Gln 251* p.Lys353fs p.Leu581* p.Gln775* p.Trp898* - - - p.Arg1086* p.Gln1091fs p.His1170fs	15/2279	0.66	Familial Unselected	(58)
Malaysia	c.1559+1G>A c.1691-2A>G -	- - p.Asn251X	4/467	0.86	Unselected	(59)

	c.2828delA	-				
Jamaica	c.43G>T c.109C>A c.758_759insT c.2052delC c.3166C>T	p.Glu15Ter p.Arg37Ser p.Ser254Ilefs p.Arg686Glyfs p.Gln1056Ter	5/179	2.8	Unselected	(62)

Appendix B Somatic Mutational Signatures

Single Base Substitution (SBS) Signature	Aetiological Description
Signature 1	Spontaneous deamination of 5-methylcytosine with advancing age (165, 178, 261)
Signature 2	APOBEC activity (165, 178)
Signature 3	Homologous Recombination Repair deficit (*BRCA1*, *BRCA2* and *PALB2*) (165, 178, 254, 262)
Signature 4	Benzo(a)pyrene diol epoxide within tobacco smoke (165, 178)
Signature 5	
Signature 6	Defective DNA Mismatch Repair (165, 178)
Signature 7a,b,c,d	Ultraviolet light exposure (165, 263, 264)
Signature 8	
Signature 9	Partial contribution from polymerase η activity (165, 178)
Signature 10a,b	*POLE* mutation (165, 178)
Signature 11	Temozolomide treatment (165, 178)
Signature 12	
Signature 13	APOBEC activity (165, 178)
Signature 14	*POLE* mutation and DNA Mismatch Repair deficiency (178)
Signature 15	Defective DNA Mismatch Repair (178)
Signature 16	
Signature 17	
Signature 18	Reactive oxygen species (178)
Signature 19	
Signature 20	*POLD1* mutation and DNA Mismatch Repair deficiency (178)
Signature 21	Defective DNA Mismatch Repair (178)
Signature 22	Aristolochic acid exposure (178)
Signature 23	
Signature 24	Aflatoxin exposure (178)
Signature 25	Chemotherapy (178)
Signature 26	Defective DNA Mismatch Repair (178)
Signature 28	
Signature 29	Tobacco chewing (178)
Signature 30	Defective base excision repair; *NTHL1* mutation (178)
Signature 31	Platinum compound chemotherapy (178, 263)
Signature 32	Azathioprine therapy (178)
Signature 33	
Signature 34	
Signature 35	Platinum compound chemotherapy (263)
Signature 36	Defective base excision repair secondary to inactivating germline or somatic mutations in MUTYH (265, 266)
Signature 37	
Signature 38	Ultraviolet light exposure (178, 263)
Signature 39	
Signature 40	
Signature 41	
Signature 42	Exposure to Haloalkanes (267)
Signature 44	Defective DNA Mismatch Repair (268)

Single Base Substitution Somatic Mutational Signatures:

Overview of the Single Base Substitution (SBS) Somatic Mutational Signatures described in cancer. The aetiological mechanism for several SBS mutational signatures remains to be elucidated and these are shown as blank within the table.

Appendix C Study Protocols

Hyperlink:

Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol

100,000 Genomes Project Protocol | Genomics England

Appendix D Bioinformatics Commands

D.1 The Prospective Outcomes in Hereditary versus Sporadic Breast Cancer Study

D.1.1 Southampton Bioinformatics Pipeline:

Produced by Dr William Tapper
Command for SAMtools Variant Calling
samtools view -bq 20 -F 772 $SAMPLE_callRGOrdered.bam | samtools mpileup -EDSgu -d 40000 -f hg19.fa -L 40000 -F 0.05 - | bcftools view -Nvg - > $SAMPLE_raw.vcf
samtools view takes the raw bam file and makes a new one according to options given.
-b = output bam format
-q = skip alignment with MAPQ <20
-F 772 = exclude unmapped reads, reads not primary alignment and read that fail platform/vendor quality checks
| = pipe filtered BAM file to samtools mpileup which compiles info for variant calling
-E = extended BAQ computation
-D = output per sample read depth
-S = output per sample phred-scaled strand bias p-value
-g = compute genotype likelihoods & output in binary call format .bcf
-u = similar to –g but output uncompressed for piping
-d = 40,000 at position read maximally INT reads per bam input
-f = reference human genome sequence
-L = 40,000 skip indel calling if average per sample read depth is above INT
-F = 0.05 controls when to initiate indel realignment, minimum freq 5%
| = pipe variant calling info to bcftools view for variant calling
-N = skip sites where REF is not A/C/G/T
-v = output potential variant sites only
-g = call genotypes at variant sites
Command for GATK Unified Genotyper Variant Calling
java -jar $GATK_PATHGenomeAnalysisTK.jar -T UnifiedGenotyper -R $REF_GENOME_PATHhg19.fa –I $SAMPLE_callRGOrdered.bam -o $SAMPLE_GATKraw.vcf -stand_call_conf 30.0 -stand_emit_conf 10.0

D.1.2 Cambridge Bioinformatics Pipeline

Produced by Dr Jamie Allen
Command for GATK Unified Genotyper Variant Calling
java -jar -Xmx248g GenomeAnalysisTK.jar\
-T UnifiedGenotyper \
-nt 2 \
-nct 16 \
-mbq 20 \
-contamination 0.05 \
-dt NONE \
-R /path/to/ref_genome.fa \
-I samples.list \
--dbsnp dbsnp_138.hg19.vcf \
-stand_call_conf 50.0 \

```
-stand_emit_conf 10.0 \
-dcov 1000 \
-L gene_region.interval_list \
--max_alternate_alleles 12 \
-glm BOTH \
-o gene_region.vcf
```

D.1.3 R-Studio Script: Statistical Analysis of the POSH Cohort

```
install.packages("gmodels")
library(gmodels)

CrossTable(BRCA_Focused_for_Totals$N_STAGE, BRCA_Focused_for_Totals$BRCA1_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$histgrade_cat, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$bmi_cat_by4, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$ETH_CAT, BRCA_Focused_for_Totals$BRCA1_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$FH_CAT, BRCA_Focused_for_Totals$BRCA1_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$FOCAL_CAT, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$POS_NODES_CAT, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$LYMPH_INV_CAT, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$ER_FINAL_CAT, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$PR_FINAL_CAT, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$HER2_FINAL_CAT, BRCA_Focused_for_Totals$BRCA1_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$TNT, BRCA_Focused_for_Totals$BRCA1_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
write.table(aggregate(BRCA_Focused_for_Totals$ageatdiag~BRCA_Focused_for_Totals$BRCA1_I
ND, FUN=summary), file="BRCA1_ageatdiagnosis.csv", sep =",", quote=FALSE, row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$BOD_AT_PRES~BRCA_Focused_for_Totals$BRC
A1_IND, FUN=summary), file="BRCA1_BOADICEA.csv", sep =",", quote=FALSE, row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_invasive~BRCA_Focused_for_Tota
ls$BRCA1_IND, FUN=summary), file="BRCA1_maxtumourinvasive.csv", sep =",", quote=FALSE,
row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_insitu~BRCA_Focused_for_Totals$
BRCA1_IND, FUN=summary), file="BRCA1_maxtumourinsitu.csv", sep =",", quote=FALSE,
row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_overall~BRCA_Focused_for_Total
s$BRCA1_IND, FUN=summary), file="BRCA1_maxtumouroverall.csv", sep =",", quote=FALSE,
row.names=F)
CrossTable(BRCA_Focused_for_Totals$N_STAGE, BRCA_Focused_for_Totals$BRCA2_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$histgrade_cat, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
```

```
CrossTable(BRCA_Focused_for_Totals$bmi_cat_by4, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$ETH_CAT, BRCA_Focused_for_Totals$BRCA2_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$FH_CAT, BRCA_Focused_for_Totals$BRCA2_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$FOCAL_CAT, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$POS_NODES_CAT, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$LYMPH_INV_CAT, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$ER_FINAL_CAT, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$PR_FINAL_CAT, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$HER2_FINAL_CAT, BRCA_Focused_for_Totals$BRCA2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$TNT, BRCA_Focused_for_Totals$BRCA2_IND, digits=1,
prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
write.table(aggregate(BRCA_Focused_for_Totals$ageatdiag~BRCA_Focused_for_Totals$BRCA2_I
ND, FUN=summary), file="BRCA2_ageatdiagnosis.csv", sep =",", quote=FALSE, row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$BOD_AT_PRES~BRCA_Focused_for_Totals$BRC
A2_IND, FUN=summary), file="BRCA2_BOADICEA.csv", sep =",", quote=FALSE, row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_invasive~BRCA_Focused_for_Tota
ls$BRCA2_IND, FUN=summary), file="BRCA2_maxtumourinvasive.csv", sep =",", quote=FALSE,
row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_insitu~BRCA_Focused_for_Totals$
BRCA2_IND, FUN=summary), file="BRCA2_maxtumourinsitu.csv", sep =",", quote=FALSE,
row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_overall~BRCA_Focused_for_Total
s$BRCA2_IND, FUN=summary), file="BRCA2_maxtumouroverall.csv", sep =",", quote=FALSE,
row.names=F)
CrossTable(BRCA_Focused_for_Pvalue$N_STAGE, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$histgrade_cat, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$bmi_cat_by4, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$ETH_CAT, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$FH_CAT, BRCA_Focused_for_Pvalue$BRCA1_IND, digits=1,
prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$FOCAL_CAT, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$POS_NODES_CAT,
BRCA_Focused_for_Pvalue$BRCA1_IND, digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq
= TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$LYMPH_INV_CAT, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$ER_FINAL_CAT, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$PR_FINAL_CAT, BRCA_Focused_for_Pvalue$BRCA1_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
```

```
CrossTable(BRCA_Focused_for_Pvalue$HER2_FINAL_CAT,
BRCA_Focused_for_Pvalue$BRCA1_IND, digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq
= TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$TNT, BRCA_Focused_for_Pvalue$BRCA1_IND, digits=1,
prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
wilcox.test(BRCA_Focused_for_Pvalue$ageatdiag~BRCA_Focused_for_Pvalue$BRCA2_IND,
paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$BOD_AT_PRES~BRCA_Focused_for_Pvalue$BRCA2_IND,
paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_invasive~BRCA_Focused_for_Pvalue$BRCA2
_IND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_insitu~BRCA_Focused_for_Pvalue$BRCA2_I
ND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_overall~BRCA_Focused_for_Pvalue$BRCA2_
IND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$ageatdiag~BRCA_Focused_for_Pvalue$BRCA1_IND,
paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$BOD_AT_PRES~BRCA_Focused_for_Pvalue$BRCA1_IND,
paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_invasive~BRCA_Focused_for_Pvalue$BRCA1
_IND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_insitu~BRCA_Focused_for_Pvalue$BRCA1_I
ND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_overall~BRCA_Focused_for_Pvalue$BRCA1_
IND, paired=FALSE)
CrossTable(BRCA_Focused_for_Pvalue$N_STAGE, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$histgrade_cat, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$bmi_cat_by4, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$ETH_CAT, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$FH_CAT, BRCA_Focused_for_Pvalue$BRCA2_IND, digits=1,
prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$FOCAL_CAT, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$POS_NODES_CAT,
BRCA_Focused_for_Pvalue$BRCA2_IND, digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq
= TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$LYMPH_INV_CAT, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$ER_FINAL_CAT, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$PR_FINAL_CAT, BRCA_Focused_for_Pvalue$BRCA2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$HER2_FINAL_CAT,
BRCA_Focused_for_Pvalue$BRCA2_IND, digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq
= TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$TNT, BRCA_Focused_for_Pvalue$BRCA2_IND, digits=1,
prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Totals$N_STAGE, BRCA_Focused_for_Totals$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
```

```
CrossTable(BRCA_Focused_for_Totals$histgrade_cat,
BRCA_Focused_for_Totals$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.C=TRUE,
prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$bmi_cat_by4,
BRCA_Focused_for_Totals$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.C=TRUE,
prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$ETH_CAT, BRCA_Focused_for_Totals$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$FH_CAT, BRCA_Focused_for_Totals$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$FOCAL_CAT, BRCA_Focused_for_Totals$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$POS_NODES_CAT,
BRCA_Focused_for_Totals$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.C=TRUE,
prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$LYMPH_INV_CAT,
BRCA_Focused_for_Totals$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.C=TRUE,
prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$ER_FINAL_CAT,
BRCA_Focused_for_Totals$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.C=TRUE,
prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$PR_FINAL_CAT,
BRCA_Focused_for_Totals$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.C=TRUE,
prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$HER2_FINAL_CAT,
BRCA_Focused_for_Totals$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.C=TRUE,
prop.r=FALSE,format="SPSS")
CrossTable(BRCA_Focused_for_Totals$TNT, BRCA_Focused_for_Totals$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.C=TRUE, prop.r=FALSE,format="SPSS")
write.table(aggregate(BRCA_Focused_for_Totals$ageatdiag~BRCA_Focused_for_Totals$BRCA1_O
R_2_IND, FUN=summary), file="BRCA_ageatdiagnosis.csv", sep =",", quote=FALSE, row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$BOD_AT_PRES~BRCA_Focused_for_Totals$BRC
A1_OR_2_IND, FUN=summary), file="BRCA_BOADICEA.csv", sep =",", quote=FALSE, row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_invasive~BRCA_Focused_for_Tota
ls$BRCA1_OR_2_IND, FUN=summary), file="BRCA_maxtumourinvasive.csv", sep =",",
quote=FALSE, row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_insitu~BRCA_Focused_for_Totals$
BRCA1_OR_2_IND, FUN=summary), file="BRCA_maxtumourinsitu.csv", sep =",", quote=FALSE,
row.names=F)
write.table(aggregate(BRCA_Focused_for_Totals$max_tumour_overall~BRCA_Focused_for_Total
s$BRCA1_OR_2_IND, FUN=summary), file="BRCA_maxtumouroverall.csv", sep =",", quote=FALSE,
row.names=F)
CrossTable(BRCA_Focused_for_Pvalue$N_STAGE, BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$histgrade_cat,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$bmi_cat_by4,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$ETH_CAT, BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$FH_CAT, BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE,format="SPSS")
```

```
CrossTable(BRCA_Focused_for_Pvalue$FOCAL_CAT,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$POS_NODES_CAT,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$LYMPH_INV_CAT,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$ER_FINAL_CAT,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$PR_FINAL_CAT,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$HER2_FINAL_CAT,
BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND, digits=1, prop.t=FALSE, prop.c=TRUE,
prop.r=FALSE, chisq = TRUE, format="SPSS")
CrossTable(BRCA_Focused_for_Pvalue$TNT, BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND,
digits=1, prop.t=FALSE, prop.c=TRUE, prop.r=FALSE, chisq = TRUE, format="SPSS")
wilcox.test(BRCA_Focused_for_Pvalue$ageatdiag~BRCA_Focused_for_Pvalue$BRCA1_OR_2_IND,
paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$BOD_AT_PRES~BRCA_Focused_for_Pvalue$BRCA1_OR_2_
IND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_invasive~BRCA_Focused_for_Pvalue$BRCA1
_OR_2_IND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_insitu~BRCA_Focused_for_Pvalue$BRCA1_O
R_2_IND, paired=FALSE)
wilcox.test(BRCA_Focused_for_Pvalue$max_tumour_overall~BRCA_Focused_for_Pvalue$BRCA1_
OR_2_IND, paired=FALSE)
```

D.1.4 R-Studio Script: Heatmap Histopathological Analysis of the POSH Cohort

```
data <- Heat_Map_noTNT[,2:8]
colnames(data) <- c(paste("No
Mutation"),paste("BRCA1"),paste("BRCA2"),paste("PALB2"),paste("CHEK2"),paste("ATM"),paste("
TP53"))
row.names(data) = sapply(Heat_Map_noTNT$`Histopathological Phenotype`,function(x)
strsplit(as.character(x), split = "\\\\" [[1]][1]))
data_mat <- data.matrix(data)
my_palette <- colorRampPalette(c("yellow","orange","red"))
data_scaled <- t(scale(t(data_mat)))
heatmap.2(data_scaled, col = my_palette(n=20), trace = "none", cexCol = 0.8, cexRow = 0.8,
srtRow = 45, srtCol = 45)
```

D.2 The 100,000 Genomes Project:

D.2.1 Shell Script Germline Analysis

Produced in conjunction with Dr William Tapper

Build 37

```
cat CancerPatientPathsOctober2018_37 | while read pathToVCF build;
```

```
do
( echo $pathToVCF | cat; echo $build ;
awk '{if($1==16 && $2>23614483 && $2<23652678 && $8 ~ /PALB2/ && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> PALB2_build37.txt
( echo $pathToVCF | cat; echo $build ;
awk '{if($1==22 && $2>29083731 && $2<29138410 && $8 ~ /CHEK2/ && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> CHEK2_build37.txt
( echo $pathToVCF | cat; echo $build ;
awk '{if($1==11 && $2>108093211 && $2<108239829 && $8 ~ /ATM/ && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> ATM_build37.txt
( echo $pathToVCF | cat; echo $build ;
awk '{if($1==17 && $2>7565097 && $2<7590868 && $8 ~ /TP53/ && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> TP53_build37.txt
done
```

Build 38

```
cat CancerPatientPathsOctober2018_38 | while read pathToVCF build;
do
( echo $pathToVCF | cat; echo $build ;
awk '{if($1=="chr16" && $2>23603160 && $2<23641310 && $8 ~ /PALB2/ && $6 !~ 0  && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> PALB2_build38.txt
( echo $pathToVCF | cat; echo $build ;
awk '{if($1=="chr22" && $2>28687743 && $2<28742422 && $8 ~ /CHEK2/ && $6 !~ 0 && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> CHEK2_build38.txt
( echo $pathToVCF | cat; echo $build ;
awk '{if($1=="chr11" && $2>108222484 && $2<108369102 && $8 ~ /ATM/  && $6 !~ 0 && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> ATM_build38.txt
( echo $pathToVCF | cat; echo $build ;
awk '{if($1=="chr17" && $2>7661779 && $2<7687550 && $8 ~ /TP53/  && $6 !~ 0 && $8 !~
/regulatory_region_variant/ && $8 !~ /downstream_gene_variant/ && $8 !~ /intron_variant/)
print $0}' <(gzip -dc $pathToVCF) ) >> TP53_build38.txt
done
```

Founder

D.2.2 Bash Script Germline Analysis

Build 37

```
module load bcftools/1.9
while read -r vcf; do
      tabix -h $vcf -R regions37.txt | \
      bcftools norm -m -any | \
      bcftools view -f PASS -i 'MIN(FMT/DP)>10 & MIN(FMT/GQ)>15' | \
      bcftools query -f
'[%SAMPLE]\t%CHROM\t%POS\t%ID\t%REF\t%ALT\t%QUAL\t%FILTER\t[%GT]\t[%GQ]\t[%DP]\t%
INFO/CSQT\n' | \
        grep -f so_terms.txt >> results_c37.txt ;
```

```
done < vcflist_c37

sed -i '1s/^/SAMPLE\tCHROM\tPOS\tID\tREF\tALT\tQUAL\tFILTER\tGT\tGQ\tDP\tCSQT\n/'
results_c37.txt
```

Build 38

```
module load bcftools/1.9
while read -r vcf; do
      tabix -h $vcf -R regions38.txt | \
      bcftools norm -m -any | \
      bcftools view -f PASS -i 'MIN(FMT/DP)>10 & MIN(FMT/GQ)>15' | \
      bcftools query -f
'[%SAMPLE]\t%CHROM\t%POS\t%ID\t%REF\t%ALT\t%QUAL\t%FILTER\t[%GT]\t[%GQ]\t[%DP]\t%
INFO/CSQT\n' | \
        grep -f so_terms.txt >> results_c38.txt ;
done < vcflist_c38

sed -i '1s/^/SAMPLE\tCHROM\tPOS\tID\tREF\tALT\tQUAL\tFILTER\tGT\tGQ\tDP\tCSQT\n/'
results_c38.txt
```

D.2.3 Bash Script Somatic Analysis

Produced in conjunction with Dr William Tapper

```
#!/bin/bash
module load bcftools/1.9
while read -r vcf ; do
        echo $vcf >> results_somatic.txt
        tabix -h $vcf -R Genes.txt |
        grep -f so_terms.txt >> results_somatic.txt ;
done < VCFpathsSomatic_Cancer
sed -I 'ls/^/SAMPLE\tCHROM\tPOS\tID\tREF\tALT\tQUAL\tFILTER\tINFO\tFORMAT\n/'
results_somatic.txt

awk '$7 !~ /BCNoiseIndel/ && $7 !~ /LowQscore/ && $7 !~ /QSI_ref/ && $7 !~ /RepetitiveRegion/'
results_somatic_NoMutation.txt > Qual_NoMutation.txt

awk -F "\t" '{if($0 ~ /^\//) {split($0,path,"/");} if($0 !~ /^\//) {ele=split($8,vinfo,"|");
for(i=0;i<ele;i++) {if(vinfo[i] ~ /ENST/ && (vinfo[i-1] == "CHEK2" || vinfo[i-1] == "ATM"|| vinfo[i-1]
==  "ARID1A" || vinfo[i-1] == "GATA3" || vinfo[i-1] == "CCND1" || vinfo[i-1] == "CDKN1B" ||
vinfo[i-1] == "ETV6" || vinfo[i-1] == "SMARCD1" || vinfo[i-1] == "TBX3" || vinfo[i-1] == "BRCA2" ||
vinfo[i-1] == "RB1" || vinfo[i-1] == "AKT1" || vinfo[i-1] == "FOXA1" || vinfo[i-1] == "NTRK3" ||
vinfo[i-1] == "CDH1" || vinfo[i-1] == "CTCF" || vinfo[i-1] == "PALB2" || vinfo[i-1] == "BRCA1" ||
vinfo[i-1] == "BRIP1" || vinfo[i-1] == "ERBB2" || vinfo[i-1] == "MAP2K4" || vinfo[i-1] == "NCOR1"
|| vinfo[i-1] == "PPM1D" || vinfo[i-1] == "TP53" || vinfo[i-1] == "KEAP1" || vinfo[i-1] == "BARD1"
|| vinfo[i-1] == "CASP8" || vinfo[i-1] == "SALL4" || vinfo[i-1] == "APOBEC3B" || vinfo[i-1] ==
"EP300" || vinfo[i-1] == "BAP1" || vinfo[i-1] == "MAP3K13" || vinfo[i-1] == "PBRM1" || vinfo[i-1]
== "PIK3CA" || vinfo[i-1] == "POLQ" || vinfo[i-1] == "MAP3K1" || vinfo[i-1] == "ARID1B" || vinfo[i-
1] == "ESR1" || vinfo[i-1] == "NOTCH1" || vinfo[i-1] == "IRS4")) print NR,"\t",path[8],"\t",vinfo[i-
1],"\t",vinfo[i+1];}}}' Qual_NoMutation.txt > formated_NoMutation.txt
```

D.2.4 R-Studio Scripts Waterfall Plot

```
waterfall(formated_NoMutation_Custom, fileType = "Custom", variant_class_order =
most_deleterious <-
c("stop_gained","frameshift_variant","start_lost","splice_donor_variant","splice_acceptor_varian
t","stop_lost","inframe_deletion","inframe_insertion","missense_variant","splice_region_variant"
,"5_prime_UTR_variant","3_prime_UTR_variant","downstream_gene_variant","upstream_gene_
variant","intron_variant","synonymous_variant"), mainRecurCutoff = 0)
```

D.2.5 R-Studio Scripts for Data Analysis: Age of Onset

```
> library(ggplot2)
> library(gmodels)
> library(RColorBrewer)

>write.table(aggregate(RareDisease_AgeofOnset20$Age_Of_Onset~RareDisease_AgeofOnset20$
Mutation, FUN = summary), file = “age_of_onset_raredisease_posneg.csv”, sep = “,”, quote =
FALSE, row.names =F)

> sink('pvalue_ageofonset.txt', append = TRUE)
df <- RareDisease_Cancer20[c("Age_Of_Onset","Domain","Gene")]
df$Domain <- factor(df$Domain, levels = c("Rare Disease","Cancer"))
df$Gene <- factor(df$Gene, levels = c("MUTATION-","ATM+",
“BRCA+”,"CHEK2+","PALB2+","TP53+"))
ggplot(data = df, aes(x=df$Gene, na.translate = FALSE, y=df$Age_Of_Onset, fill=df$Domain)) +
geom_boxplot() + labs(title = "Age at Diagnosis: A Comparison Across Cohorts", x = "Mutation
Status", y = "Age At Diagnosis (years)") + theme(plot.title = element_text(hjust = 0.5)) +
ylim(0,100) + scale_fill_discrete(name = "Cohort")
> View(RareDisease_Cancer20_noBRCA)
> df <- RareDisease_Cancer20_noBRCA[c("Age_Of_Onset","Domain","Gene")]
> df$Domain <- factor(df$Domain, levels = c("Familial Breast Cancer","Unselected Breast
Cancer"))
> df$Gene <- factor(df$Gene, levels = c("MUTATION-","ATM+","CHEK2+","PALB2+","TP53+"))
ggplot(data = df, aes(x=Gene, y=Age_Of_Onset, fill=Domain)) + geom_boxplot(outlier.shape = NA)
+ labs(title = "Age at Diagnosis: A Comparison Across Cohorts", x = "Mutation Status", y = "Age At
Diagnosis (years)") + theme(plot.title = element_text(hjust = 0.5)) + ylim(0,100) +
scale_fill_brewer(palette = "Pastel2", name = "Cohort")
sink('pvalue_agediagnosis.txt', append = TRUE)
> t.test(RareDisease_Cancer20$Age_Of_Onset~RareDisease_Cancer20$`ATM+`, alternative =
"two.sided", var.equal=FALSE)
> t.test(RareDisease_Cancer20$Age_Of_Onset~RareDisease_Cancer20$`CHEK2+`, alternative =
"two.sided", var.equal=FALSE)
> t.test(RareDisease_Cancer20$Age_Of_Onset~RareDisease_Cancer20$`PALB2+`, alternative =
"two.sided", var.equal=FALSE)
> t.test(RareDisease_Cancer20$Age_Of_Onset~RareDisease_Cancer20$`TP53+`, alternative =
"two.sided", var.equal=FALSE)
> t.test(RareDisease_Cancer20$Age_Of_Onset~RareDisease_Cancer20$`MUTATION-`, alternative
= "two.sided", var.equal=FALSE)
```

D.2.6 R-Studio Scripts for Data Analysis: Tumour Mutational Burden

```
> library(ggplot2)
> library(RColorBrewer)
> write.table(aggregate(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationburden_posneg.csv", sep = ",", quote = FALSE, row.names = F)
> write.table(aggregate(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationburden_indgenes.csv", sep = ",", quote = FALSE, row.names = F)
wilcox.test(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
wilcox.test(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
```

```
wilcox.test(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$`BRCA+`, paired=FALSE)
wilcox.test(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$`CHEK2+`, paired=FALSE)
wilcox.test(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$`PALB2+`, paired=FALSE)
wilcox.test(MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per
Mb`~MutationalBurden_MutationalSignature20$`TP53+`, paired=FALSE)
> ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Somatic Coding Variants Per Mb`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Burden", x = "Mutation Status", y = "Somatic Coding
Variants per Mb") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust
= 0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,35)
```

D.2.7 R- Studio Scripts for Data Analysis: Somatic Mutational Signature

```
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
1`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
2`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
3`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
4`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
5`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
6`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
7`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
8`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
9`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
10`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
11`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
```

```
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
12`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
13`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
14`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
15`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
16`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
17`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
18`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
19`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
20`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
21`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
22`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
23`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
24`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
25`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
26`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
27`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
28`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
29`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
```

```
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
30`~MutationalBurden_MutationalSignature20$`Mutation Status`, FUN = summary), file =
"mutationsignature_posneg.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)

write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
1`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
2`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
3`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
4`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
5`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
6`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
7`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
8`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
9`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
10`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
11`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
12`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
13`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
14`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
15`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
16`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
```

```
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
17`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
18`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
19`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
20`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
21`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
22`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
23`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
24`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
25`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
26`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
27`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
28`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
29`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)
write.table(aggregate(MutationalBurden_MutationalSignature20$`Signature
30`~MutationalBurden_MutationalSignature20$Gene, FUN = summary), file =
"mutationsignature_indgenes.csv", sep = ",", quote = FALSE, row.names = F, append = TRUE)

> sink('pvalue_posneg.txt', append = TRUE)
or
> sink('pvalue_posneg.csv', append = TRUE)

> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
1`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
2`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
3`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
```

```
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
4`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
5`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
6`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
7`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
8`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
9`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
10`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
11`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
12`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
13`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
14`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
15`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
16`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
17`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
18`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
19`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
20`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
21`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
22`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
23`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
24`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
25`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
26`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
27`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
28`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
29`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
30`~MutationalBurden_MutationalSignature20$`Mutation Status`, paired=FALSE)
```

```
> sink()

> sink('pvalue_ATM.txt', append = TRUE)
or
> sink('pvalue_ATM.cxv', append = TRUE)

> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
1`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
2`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
3`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
4`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
5`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
6`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
7`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
8`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
9`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
10`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
11`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
12`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
13`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
14`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
15`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
16`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
17`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
18`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
19`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
20`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
21`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
22`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
23`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
```

```
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
24`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
25`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
26`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
27`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
28`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
29`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> wilcox.test(MutationalBurden_MutationalSignature20$`Signature
30`~MutationalBurden_MutationalSignature20$`ATM+`, paired=FALSE)
> sink()
```

Individuals Box and Whisker Plots

```
png(file = "/home/sheygate/Project/somatic1.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 1`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 1", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic2.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 2`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 2", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic3.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 3`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 3", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
```

```
png(file = "/home/sheygate/Project/somatic4.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 4`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 4", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic5.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 5`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 5", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic6.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 6`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 6", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic7.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 7`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 7", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic8.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 8`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
```

```
shape = 1) + labs(title = "Somatic Mutational Signature 8", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic9.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 9`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 9", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic10.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 10`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 10", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic11.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 11`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 11", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic12.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 12`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 12", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
```

```
png(file = "/home/sheygate/Project/somatic13.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 13`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 13", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic14.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 14`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 14", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic15.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 15`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 15", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic16.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 16`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 16", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic17.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 17`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
```

```
shape = 1) + labs(title = "Somatic Mutational Signature 17", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic18.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 18`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 18", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic19.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 19`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 19", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic20.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 20`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 20", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic21.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 21`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 21", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
```

```
png(file = "/home/sheygate/Project/somatic22.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 22`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 22", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic23.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 23`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 23", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic24.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 24`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 24", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic25.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 25`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 25", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic26.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 26`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
```

```
shape = 1) + labs(title = "Somatic Mutational Signature 26", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic27.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 27`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 27", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic28.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 28`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 28", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic29.png", units = 'cm', width = 14, height = 16, res =
300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 29`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 29", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
png(file = "/home/sheygate/Project/somatic30.png", units = 'cm', width = 14, height = 16, res =
300, units = 'cm', width = 14, height = 16, res = 300)
ggplot(MutationalBurden_MutationalSignature20,
aes(x=MutationalBurden_MutationalSignature20$`Mutation Status`,
y=MutationalBurden_MutationalSignature20$`Signature 30`,
color=MutationalBurden_MutationalSignature20$Gene)) + geom_jitter(alpha = 0.5, width = .2,
shape = 1) + labs(title = "Somatic Mutational Signature 30", x = "Mutation Status", y =
"Percentage") + theme(legend.title = element_blank()) + theme(plot.title = element_text(hjust =
0.5)) + geom_boxplot(size=0.5, outlier.shape = 1) + ylim(0,60) + theme(axis.text =
element_text(size = 14)) + theme(plot.title = element_text(size = 14)) + theme(legend.text =
element_text(size = 14)) + theme(axis.title = element_text(size = 14))
dev.off()
```

Bar Plot Somatic Mutational Signature

```
> library(scales)

> df$`Mutation Status` <- factor(df$`Mutation Status`, levels = c("MUTATION-
","ATM+","BRCA+","CHEK2+","PALB2+","TP53+"))

> df <- mutationsignature_posneg_analysis[c("Signature","Mutation Status","Median")]

> head(df)

> df$Signature <- factor(df$Signature, levels = c("Signature 1","Signature 2","Signature
3","Signature 4","Signature 5","Signature 6","Signature 7","Signature 8","Signature 9","Signature
10","Signature 11","Signature 12","Signature 13","Signature 14","Signature 15","Signature
16","Signature 17","Signature 18","Signature 19","Signature 20","Signature 21","Signature
22","Signature 23","Signature 24","Signature 25","Signature 26","Signature 27","Signature
28","Signature 29","Signature 30"))

ggplot(data = df, aes(x=df$`Mutation Status`, y=df$Median, fill=df$Signature)) + geom_bar(stat =
"identity", position=position_fill()) + labs(title = "Somatic Mutational Signature", subtitle =
"Median", x = "Mutation Status", y = "Percentage") + theme(plot.title = element_text(hjust = 0.5),
plot.subtitle = element_text(hjust = 0.5)) + ylim(0,100) + scale_fill_discrete(name = "Mutational
Signature", na.translate = FALSE) + scale_y_continuous(labels = scales::percent_format())

> df <- mutationsignature_posneg_analysis[c("Signature","Mutation Status","Mean")]

> head(df)

> df$Signature <- factor(df$Signature, levels = c("Signature 1","Signature 2","Signature
3","Signature 4","Signature 5","Signature 6","Signature 7","Signature 8","Signature 9","Signature
10","Signature 11","Signature 12","Signature 13","Signature 14","Signature 15","Signature
16","Signature 17","Signature 18","Signature 19","Signature 20","Signature 21","Signature
22","Signature 23","Signature 24","Signature 25","Signature 26","Signature 27","Signature
28","Signature 29","Signature 30"))

> df$`Mutation Status` <- factor(df$`Mutation Status`, levels = c("MUTATION-","MUTATION+"))
> ggplot(data = df, aes(x=df$`Mutation Status`, y=df$Mean, fill=df$Signature)) + geom_bar(stat =
"identity", position=position_fill()) + labs(title = "Somatic Mutational Signature", subtitle =
"Mean", x = "Mutation Status", y = "Percentage") + theme(plot.title = element_text(hjust = 0.5),
plot.subtitle = element_text(hjust = 0.5)) + ylim(0,100) + scale_fill_discrete(name = "Mutational
Signature", na.translate = FALSE) + scale_y_continuous(labels = scales::percent_format())


df$`Mutation Status` <- factor(df$`Mutation Status`, levels = c("MUTATION-","MUTATION+"))
```

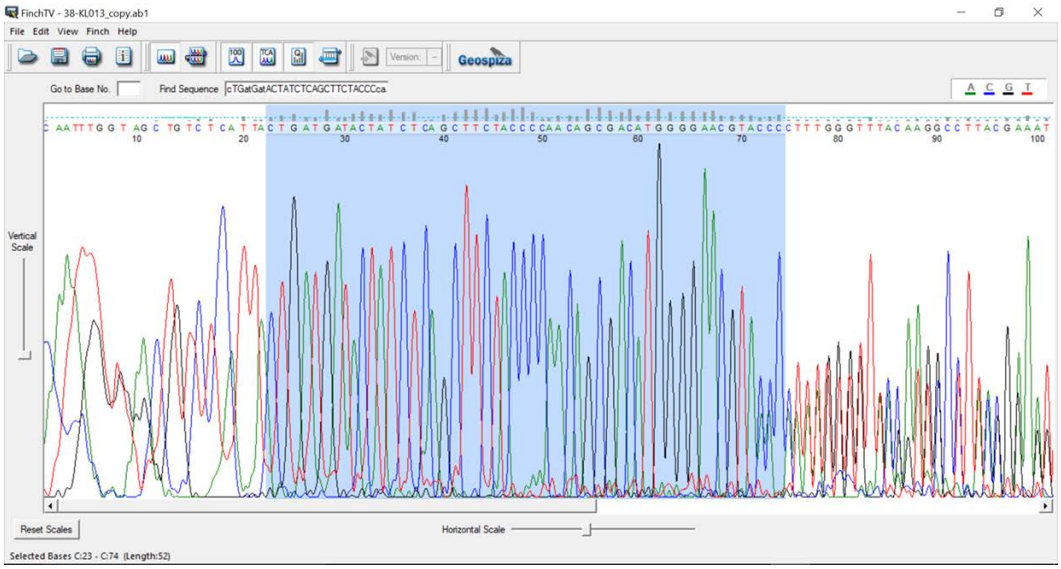
D.2.8 R- Studio Script Heatmap Somatic Mutational Signature

```
data <- mutationsignature_indgenes_analysis_focalSBS[,2:7]
colnames(data) <- c(paste("ATM"),paste("BRCA+"),paste("CHEK2+"),paste("Mutation-"),
paste("PALB2+"),paste("TP53+"))
```

Appendix D

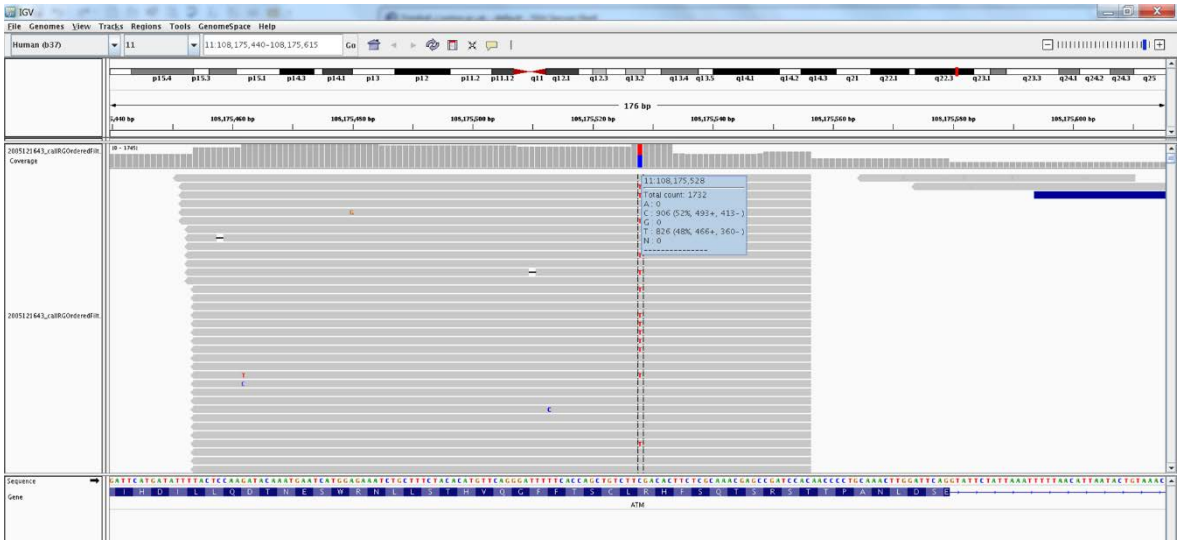
```
row.names(data) = sapply(mutationsignature_indgenes_analysis_focalSBS$Signature,function(x)
strsplit(as.character(x), split = "\\\\" [[1]][1]))
data_mat <- data.matrix(data)
my_palette <- colorRampPalette(c("yellow","orange","red"))
data_scaled <- t(scale(t(data_mat)))
data_scaled <- na.omit(data_scaled)
data_scaled[is.nan(data_scaled)] <- 0 (not used)
heatmap.2(data_scaled, col = my_palette(n=20), trace = "none", cexCol = 0.8, cexRow = 0.8,
srtRow = 45, srtCol = 45)
```

Appendix E Validation



Sequence Validation a.

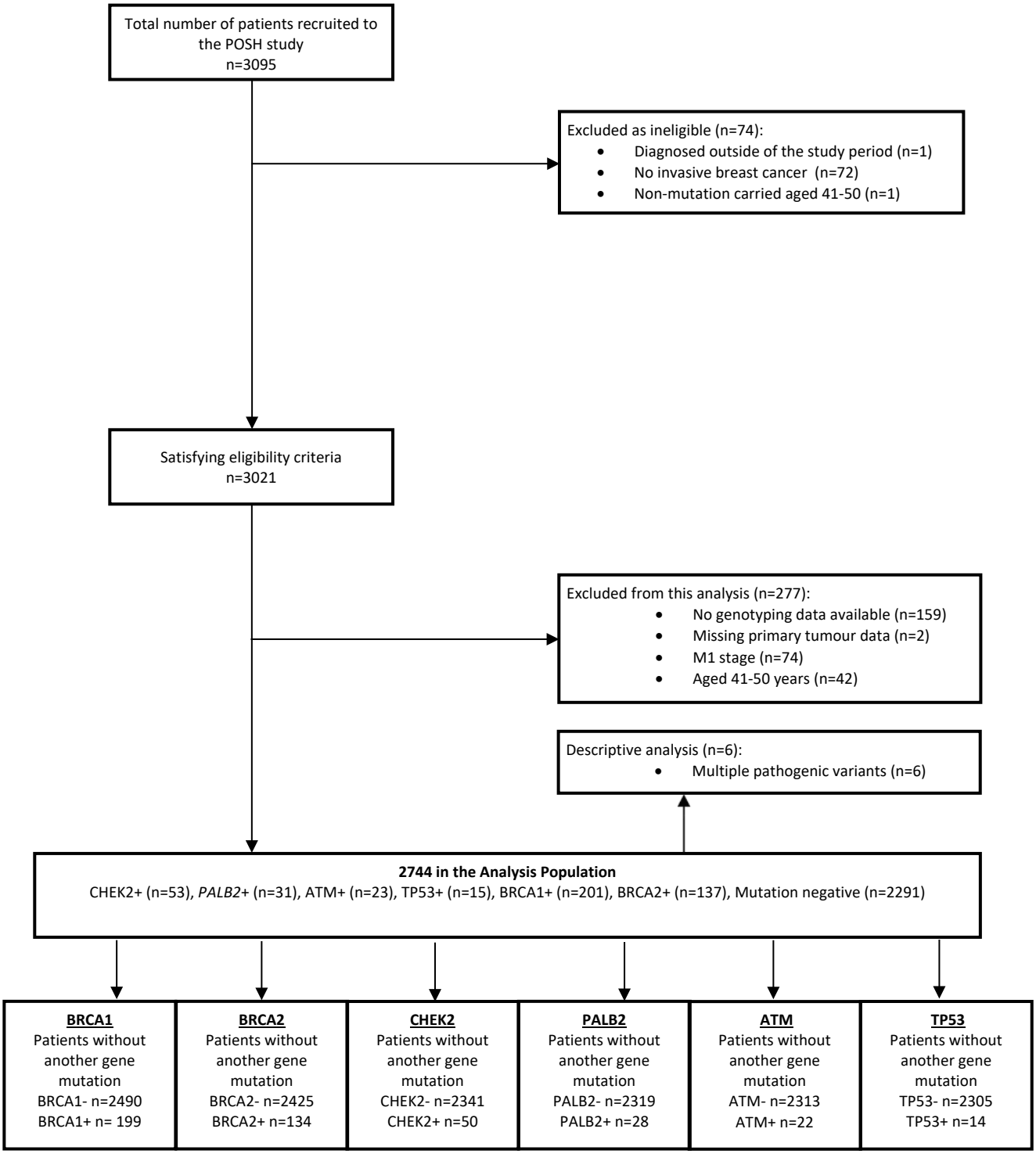
Example of validation sequence data visualised within FinchTV software.



Sequence Validation b.

Example of variant visualised within the Integrative Genomics Viewer.

Appendix F Statistical Analysis Plan (SAP)



Appendix G Somatic Analysis Target Gene List

Gene Symbol	Tier	Hallmark
AKT1	1	Yes
APOBEC3B	1	Yes
ARID1A	1	Yes
ARID1B	1	
ATM	1	Yes
BAP1	1	
BARD1	1	
BRCA1	1	Yes
BRCA2	1	Yes
BRIP1	1	Yes
CASP8	1	
CCND1	1	
CDH1	1	Yes
CDKN1B	1	
CHEK2	1	Yes
CTCF	1	
EP300	1	
ERBB2	1	Yes
ESR1	1	Yes
ETV6	1	
FOXA1	1	
GATA3	1	
IRS4	1	
KEAP1	1	
MAP2K4	1	Yes
MAP3K1	1	
MAP3K13	1	
NCOR1	1	Yes
NOTCH1	1	Yes
NTRK3	1	
PALB2	1	
PBRM1	1	
PIK3CA	1	Yes
PPM1D	1	
RB1	1	Yes

Appendix G

Gene Symbol	Tier	Hallmark
SALL4	1	
SMARCD1	1	
TBX3	1	
TP53	1	Yes

Summary of the most frequently identified genes with somatic point mutations in breast cancer. Top 20 refers to those which are in the 20 most frequently mutated genes in human breast cancer. Data derived from COSMIC and the Cancer Gene Census.(172)

Appendix H Class 4 and Class 5 Variants

The POSH Study

Gene	Coding Change	Protein Change	Total
CHEK2	c.1100delC	p.Thr367Metfs	36
CHEK2	c.1263delT	p.Ser422Valfs	3
CHEK2	c.283C>T	p.Arg95Ter	2
CHEK2	c.349A>G	p.Arg117Gly	5
CHEK2	c.405delA	p.Lys135Asnfs	1
CHEK2	c.409C>T	p.Arg137Ter	2
CHEK2	c.433C>T	p.Arg145Trp	1
CHEK2	c.58C>T	p.Gln20Ter	1
CHEK2	c.655delG	p.Glu219Asnfs	2
PALB2	c.1289_1290del	p.430_430del	1
PALB2	c.1675C>T	p.Gln559X	1
PALB2	c.1942_1949CA		1
PALB2	c.196C>T	p.Gln66X	1
PALB2	c.2050delC	p.Pro684fs	2
PALB2	c.2257C>T	p.Arg753X	1
PALB2	c.2324dupA	p.Gln775fs	1
PALB2	c.2718G>A	p.Trp906X	1
PALB2	c.31_32del	p.11_11del	1
PALB2	c.3113G>A	p.Trp1038X	14
PALB2	c.3115delA	p.Asn1039fs	1
PALB2	c.3256C>T	p.Arg1086X	1
PALB2	c.512_515del	p.171_172del	1
PALB2	c.619delC	p.Pro207fs	1
PALB2	c.706dupT	p.Phe236fs	1
PALB2	c.758dupT	p.Leu253fs	1
PALB2	c.944_953delCTTGTGGGCA	p.315_318del	1
ATM	c.1355delC	p.Thr452Asnfs	1
ATM	c.1564_1565delGA	p.Glu522Ilefs	2
ATM	c.170G>A	p.Trp57X	1
ATM	c.2098C>T	p.Gln700X	1
ATM	c.3802delG	p.Val1268X	1
ATM	c.4343T>A	p.Leu1448X	1
ATM	c.450_453del	p.150_151del	1
ATM	c.5623C>T	p.Arg1875X	2
ATM	c.6100C>T	p.Arg2034X	1

Appendix H

ATM	c.6916_6917delAG	p.Leu2307Cysfs	1
ATM	c.7271T>G	p.Val2424Gly	5
ATM	c.7456C>T	p.Arg2486X	1
ATM	c.7585dupA	p.Thr2529fs	1
ATM	c.7664_7665AGTGC		1
ATM	c.8307G>A	p.Trp2769X	1
ATM	c.8934_8935insCT	p.Thr2978fs	1
ATM	c.9022C>T	p.Arg3008Cys	1
TP53	c.112C>T	p.Gln38X	1
TP53	c.437G>A	p.Trp146X	1
TP53	c.524G>A	p.Arg175His	2
TP53	c.586C>T	p.Arg196X	2
TP53	c.625A>T	p.Arg209X	1
TP53	c.659A>G	p.Tyr220Cys	1
TP53	c.672+1G>T		1
TP53	c.725G>A	p.Cys242Tyr	2
TP53	c.733G>A	p.Gly245Ser	2
TP53	c.818G>A	p.Arg273His	1
TP53	c.919+1G>A		1

Appendix I Variants of Uncertain Significance

Gene	Coding Change	Protein Change	Frequency
CHEK2	c.1111C>T	p.His371Tyr	1
CHEK2	c.1283C>T	p.Ser428Phe	1
CHEK2	c.1427C>T	p.Thr476Met	2
CHEK2	c.190G>A	p.Glu64Lys	6
CHEK2	c.254C>G	p.Pro85Arg	4
CHEK2	c.470T>C	p.Ile157Thr	6
CHEK2	c.499G>A	p.Gly167Arg	1
CHEK2	c.539G>A	p.Arg180His	2
CHEK2	c.541C>T	p.Arg181Cys	1
CHEK2	c.715G>A	p.Glu239Lys	3
CHEK2	c.483_485delAGA	p.Glu161del	1
ATM	c.2250G>A	p.Lys750Lys	1
ATM	c.2593_2595del	p.865_865del	1
ATM	c.3118A>G	p.Met1040Val	7
ATM	c.5769_5771del	p.1923_1924del	1
ATM	c.7638_7646del	p.2546_2549del	3

Appendix J Overview of Variant Identification within the POSH Cohort

J.1 Triple Negative Breast Cancer

Within the whole cohort POSH cohort, TNTs were present in 559/2744 (20.4%) of women. Of these, 142/559 (25.5%) had a germline variant in a breast cancer susceptibility gene. The majority of variants 123/142(86.6%) were present in *BRCA1* (Table 26). Amongst the non-TNT, variant identification was 14.2% and pathogenic variants in *BRCA1* and *BRCA2* remained the most prevalent accounting for 202/311(65.0%) of all variant carriers.

Gene	Frequency (%)	
	TNT n=559	Non TNT n=2185
BRCA1+	123 (22.0)	78 (3.6)
BRCA2+	13 (2.3)	124 (5.7)
CHEK2+	1 (0.2)	52 (2.4)
PALB2+	4 (0.7)	27 (1.2)
ATM+	0 (0.0)	23 (1.1)
TP53+	1 (0.2)	14 (0.6)
Mutation -	417 (74.6)	1874 (85.8)
Total	**142 (25.4%)**	**311 (14.2%)***

Summary of all pathogenic and likely pathogenic variants identified within the genes *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *ATM* and *TP53* amongst individuals with TNTs and non TNTs. Samples derived from the POSH Study. *A total of 7 additional pathogenic/likely pathogenic variants were identified amongst 6 individuals.

J.2 Multiple Pathogenic Variants

	Variant 1	Variant 2	Variant 3
Individual 1	*BRCA2* c.5909C>A, p.(Ser1970X)	*ATM* c.170G>A p.(Trp57X)	-
Individual 2	*BRCA1* c.5095C>T p.(Arg1699Trp)	*PALB2* c.1675C>Tp.(Gln559X)	*TP53*, c.112C>T p.(Gln38X)
Individual 3	*PALB2* c.512_515del p.(171_172del)	*CHEK2* c.1100delC p.(p.Thr367Metfs)	-
Individual 4	*BRCA2* c.5682C>G p.(Tyr1894X)	*CHEK2* c.1100delC p.(p.Thr367Metfs)	-
Individual 5	*BRCA1* c.4574_4575delAA p.(1525_1525del)	*PALB2* c.3113G>A p.(Trp1038X)	-
Individual 6	*BRCA2* c.4478_4481delAAAG p.(1493_1494del)	*CHEK2* c.1100delC p.(p.Thr367Metfs)	-

Appendix K Multivariable Analysis

Produced by Dr Tom Maishman

K.1 Multivariable Analysis *CHEK2*

K.1.1 *CHEK2*: Overall Survival

Factor	Unadjusted Model	Adjusted Model (using multiple imputation)
*CHEK2*_IND		
*CHEK2*_IND-	1 (Ref. category)	1 (Ref. category)
*CHEK2*_IND+	1.58 (1.01, 2.48), 0.043	1.65 (1.05, 2.59), 0.03
Age at diagnosis		0.98 (0.96, 1.00), 0.076
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.14 (0.94, 1.39), 0.18
Obese		1.22 (0.98, 1.50), 0.071
Grade		
1		1 (Ref. category)
2		2.52 (1.33, 4.80), 0.0047
3		3.64 (1.92, 6.88), <0.0001
Max invasive size (in CM)		1.12 (1.08, 1.16), <0.0001
HER2 status		
Negative		1 (Ref. category)
Positive		1.01 (0.84, 1.21), 0.91
N stage		
N0		1 (Ref. category)
N1		2.31 (1.91, 2.81), <0.0001
ER_FINAL_CAT		
ER_FINAL_CAT-		1 (Ref. category)
ER_FINAL_CAT+(2 years)		0.33 (0.25, 0.44), <0.0001
ER_FINAL_CAT+(5 years)		1.22 (0.95, 1.58), 0.13
ER_FINAL_CAT+(10 years)		2.11 (1.47, 3.04), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.52 (1.07, 2.17), 0.021
Asian		1.15 (0.72, 1.82), 0.56
Other		0.63 (0.16, 2.53), 0.52
Taxanne indicator		
No		1 (Ref. category)
Yes		0.95 (0.78, 1.14), 0.56

K.1.2 *CHEK2*: Distant Disease Free Survival

Factor	Unadjusted Model	Adjusted Model (using multiple imputation)
*CHEK2*_IND		
*CHEK2*_IND-	1 (Ref. category)	1 (Ref. category)
*CHEK2*_IND+	1.62 (1.06, 2.48), 0.025	1.60 (1.04, 2.46), 0.033

Age at diagnosis		0.98 (0.96, 1.00), 0.12
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.10 (0.91, 1.32), 0.32
Obese		1.20 (0.98, 1.47), 0.072
Grade		
1		1 (Ref. category)
2		1.73 (1.05, 2.84), 0.031
3		2.34 (1.43, 3.84), 0.00074
Max invasive size (in CM)		1.13 (1.09, 1.17), <0.0001
HER2 status		
Negative		1 (Ref. category)
Positive		1.14 (0.96, 1.35), 0.14
N stage		
N0		1 (Ref. category)
N1		2.40 (2.00, 2.87), <0.0001
ER_FINAL_CAT		
ER_FINAL_CAT-		1 (Ref. category)
ER_FINAL_CAT+(2 years)		0.62 (0.51, 0.76), <0.0001
ER_FINAL_CAT+(5 years)		1.68 (1.29, 2.20), 0.00015
ER_FINAL_CAT+(10 years)		3.77 (2.16, 6.56), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.72 (1.24, 2.39), 0.0013
Asian		1.29 (0.85, 1.97), 0.23
Other		1.18 (0.44, 3.17), 0.74
Taxanne indicator		
No		1 (Ref. category)
Yes		0.86 (0.72, 1.04), 0.12

K.2 Multivariable Analysis *PALB2*

K.2.1 *PALB2*: Overall Survival

Factor	**Unadjusted Model**	**Adjusted Model (using multiple imputation)***
*PALB2*_IND		
*PALB2*_IND-	1 (Ref. category)	1 (Ref. category)
*PALB2*_IND+	0.72 (0.32, 1.62), 0.43	0.76 (0.34, 1.71), 0.51
Age at diagnosis		0.98 (0.96, 1.00), 0.054
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.16 (0.95, 1.41), 0.14
Obese		1.19 (0.96, 1.48), 0.11
Max invasive size (in CM)		1.12 (1.08, 1.16), <0.0001
HER2 status		
Negative		1 (Ref. category)
Positive		1.03 (0.85, 1.24), 0.78

N stage		
N0		1 (Ref. category)
N1		2.41 (1.98, 2.92), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.52 (1.06, 2.19), 0.022
Asian		1.04 (0.65, 1.64), 0.88
Other		0.68 (0.17, 2.75), 0.59
Taxanne indicator		
No		1 (Ref. category)
Yes		0.96 (0.79, 1.16), 0.69

*Grade removed due to insufficient numbers

K.2.2 *PALB2*: Distant Disease Free Survival

Factor	Unadjusted Model	Adjusted Model (using multiple imputation)*
*PALB2*_IND		
*PALB2*_IND-	1 (Ref. category)	1 (Ref. category)
*PALB2*_IND+	0.63 (0.28, 1.42), 0.27	0.66 (0.30, 1.48), 0.32
Age at diagnosis		0.98 (0.96, 1.00), 0.10
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.11 (0.92, 1.34), 0.27
Obese		1.18 (0.96, 1.45), 0.11
Max invasive size (in CM)		1.13 (1.09, 1.16), <0.0001
HER2 status		
Negative		1 (Ref. category)
Positive		1.15 (0.96, 1.37), 0.12
N stage		
N0		1 (Ref. category)
N1		2.46 (2.05, 2.95), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.73 (1.24, 2.41), 0.0013
Asian		1.20 (0.79, 1.83), 0.39
Other		0.94 (0.30, 2.95), 0.92
Taxanne indicator		
No		1 (Ref. category)
Yes		0.88 (0.74, 1.06), 0.18

*Grade removed due to insufficient numbers

K.3 Multivariable Analysis *ATM*

K.3.1 *ATM*: Overall Survival

Factor	Unadjusted Model	Adjusted Model (using multiple imputation)
*ATM*_IND		

*ATM*_IND-	1 (Ref. category)	1 (Ref. category)
*ATM*_IND+	0.58 (0.22, 1.56), 0.28	0.66 (0.25, 1.77), 0.41
Age at diagnosis		0.98 (0.96, 1.01), 0.13
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.14 (0.94, 1.38), 0.20
Obese		1.19 (0.96, 1.48), 0.11
Grade		
1		1 (Ref. category)
2		2.85 (1.45, 5.59), 0.0023
3		4.04 (2.06, 7.91), <0.0001
Max invasive size (in CM)		1.11 (1.07, 1.15), <0.0001
HER2 status		
Negative		1 (Ref. category)
Positive		1.01 (0.84, 1.21), 0.95
N stage		
N0		1 (Ref. category)
N1		2.38 (1.95, 2.90), <0.0001
ER_FINAL_CAT		
ER_FINAL_CAT-		1 (Ref. category)
ER_FINAL_CAT+(2 years)		0.34 (0.25, 0.45), <0.0001
ER_FINAL_CAT+(5 years)		1.23 (0.95, 1.59), 0.12
ER_FINAL_CAT+(10 years)		2.08 (1.45, 3.00), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.56 (1.09, 2.23), 0.015
Asian		1.17 (0.74, 1.85), 0.51
Other		0.78 (0.19, 3.14), 0.73
Taxanne indicator		
No		1 (Ref. category)
Yes		0.97 (0.81, 1.18), 0.78

K.3.2 *ATM*: Distant Disease Free Survival

Factor	**Unadjusted Model**	**Adjusted Model (using multiple imputation)**
*ATM*_IND		
*ATM*_IND-	1 (Ref. category)	1 (Ref. category)
*ATM*_IND+	0.54 (0.20, 1.46), 0.23	0.58 (0.22, 1.57), 0.29
Age at diagnosis		0.99 (0.96, 1.01), 0.24
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.09 (0.91, 1.32), 0.35
Obese		1.18 (0.96, 1.45), 0.11
Grade		
1		1 (Ref. category)
2		1.99 (1.17, 3.38), 0.011
3		2.67 (1.57, 4.53), 0.00027
Max invasive size (in CM)		1.12 (1.08, 1.16), <0.0001

HER2 status		
Negative		1 (Ref. category)
Positive		1.12 (0.94, 1.33), 0.20
N stage		
N0		1 (Ref. category)
N1		2.45 (2.03, 2.95), <0.0001
ER_FINAL_CAT		
ER_FINAL_CAT-		1 (Ref. category)
ER_FINAL_CAT+(2 years)		0.63 (0.52, 0.77), <0.0001
ER_FINAL_CAT+(5 years)		1.68 (1.29, 2.21), 0.00017
ER_FINAL_CAT+(10 years)		3.59 (2.09, 6.19), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.77 (1.27, 2.45), 0.00071
Asian		1.32 (0.86, 2.00), 0.20
Other		1.06 (0.34, 3.30), 0.92
Taxanne indicator		
No		1 (Ref. category)
Yes		0.89 (0.74, 1.07), 0.22

K.4 Multivariable Analysis *TP53*

K.4.1 *TP53*: Overall Survival

Factor	**Unadjusted Model**	**Adjusted Model (using multiple imputation)***
P53_IND		
P53_IND-	1 (Ref. category)	1 (Ref. category)
P53_IND+	1.12 (0.47, 2.72), 0.79	0.86 (0.35, 2.11), 0.75
Age at diagnosis		0.98 (0.96, 1.00), 0.073
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.16 (0.95, 1.40), 0.15
Obese		1.20 (0.97, 1.49), 0.097
Max invasive size (in CM)		1.12 (1.09, 1.16), <0.0001
HER2 status		
Negative		1 (Ref. category)
Positive		1.03 (0.85, 1.24), 0.78
N stage		
N0		1 (Ref. category)
N1		2.40 (1.97, 2.91), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.54 (1.07, 2.20), 0.02
Asian		1.09 (0.70, 1.71), 0.71
Other		0.69 (0.17, 2.78), 0.60
Taxanne indicator		
No		1 (Ref. category)
Yes		0.98 (0.81, 1.18), 0.80

*Grade removed due to insufficient numbers

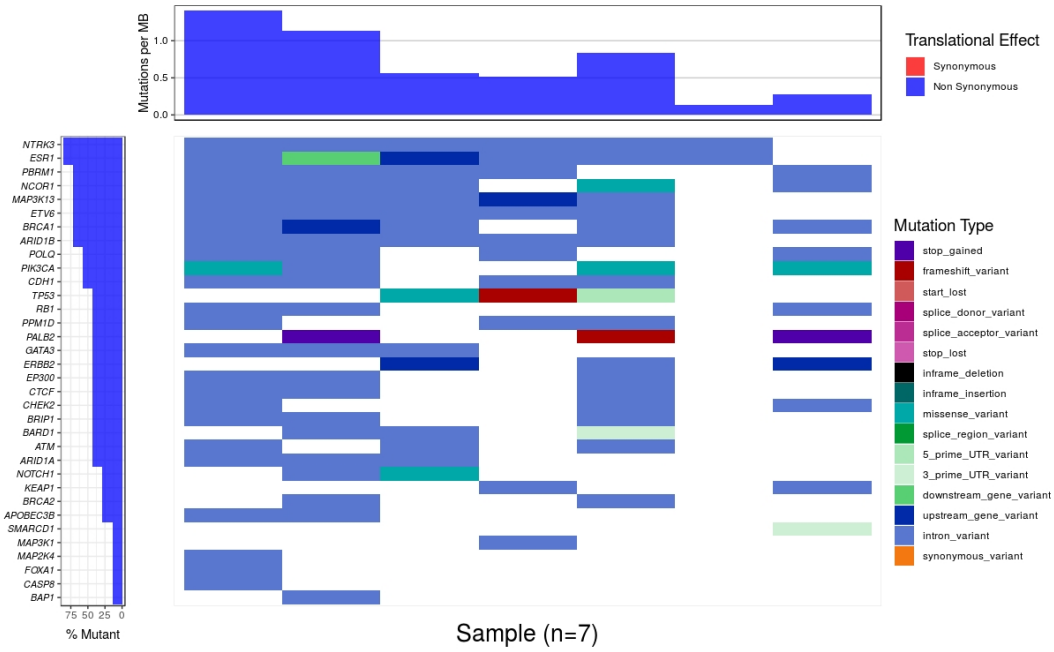
K.4.2 *TP53*: Distant Disease Free Survival

Factor	Unadjusted Model	Adjusted Model (using multiple imputation)*
P53_IND		
P53_IND-	1 (Ref. category)	1 (Ref. category)
P53_IND+	1.33 (0.59, 2.97), 0.49	0.93 (0.41, 2.12), 0.87
Age at diagnosis		0.98 (0.96, 1.00), 0.13
BMI		
Underweight/Healthy		1 (Ref. category)
Overweight		1.11 (0.92, 1.33), 0.29
Obese		1.20 (0.98, 1.47), 0.078
Max invasive size (in CM)		1.13 (1.09, 1.17), <0.0001
HER2 status		
Negative		1 (Ref. category)
Positive		1.15 (0.96, 1.37), 0.12
N stage		
N0		1 (Ref. category)
N1		2.45 (2.04, 2.94), <0.0001
Ethnicity		
White/Caucasian		1 (Ref. category)
Black		1.74 (1.25, 2.43), 0.0011
Asian		1.26 (0.83, 1.89), 0.28
Other		0.95 (0.30, 2.97), 0.93
Taxanne indicator		
No		1 (Ref. category)
Yes		0.89 (0.74, 1.07), 0.23

*Grade removed due to insufficient numbers

Appendix L Somatic Mutational Profile

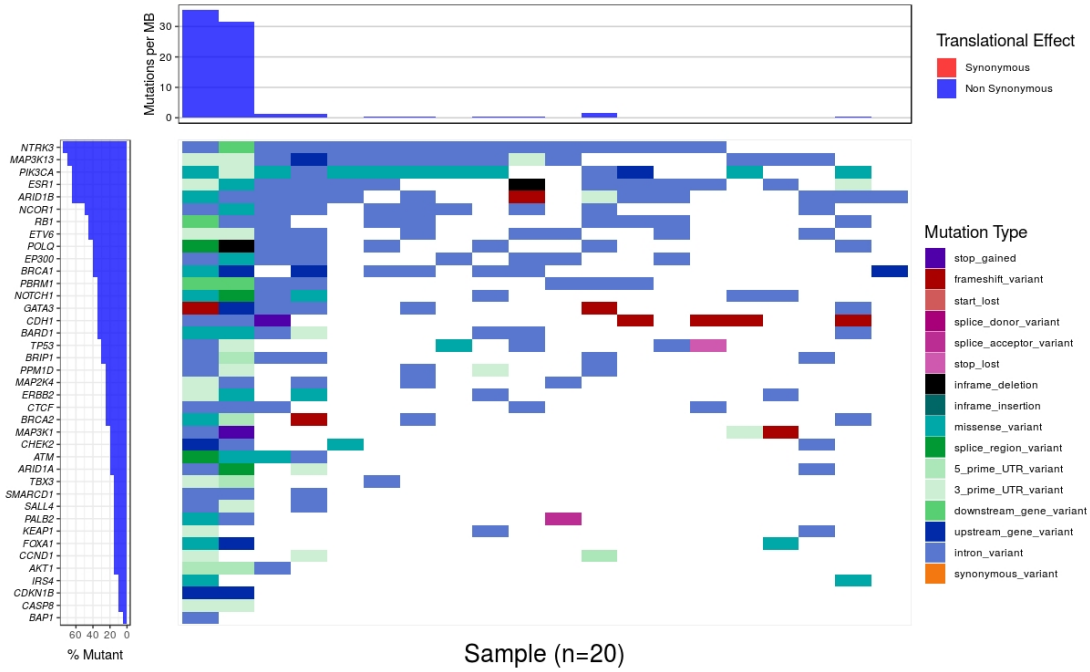
L.1 Somatic Mutation Profile *PALB2* Associated Unselected Breast Cancer: Coding and Non-Coding Variants



Somatic Mutation Profile *PALB2* Variant Carriers: Coding and Non-Coding Variants

Waterfall plot illustrating the most prevalent somatic coding and non-coding sequence variants identified in 7 breast cancers occurring in association with a *PALB2* germline variant derived from The 100,000 Genomes Project.

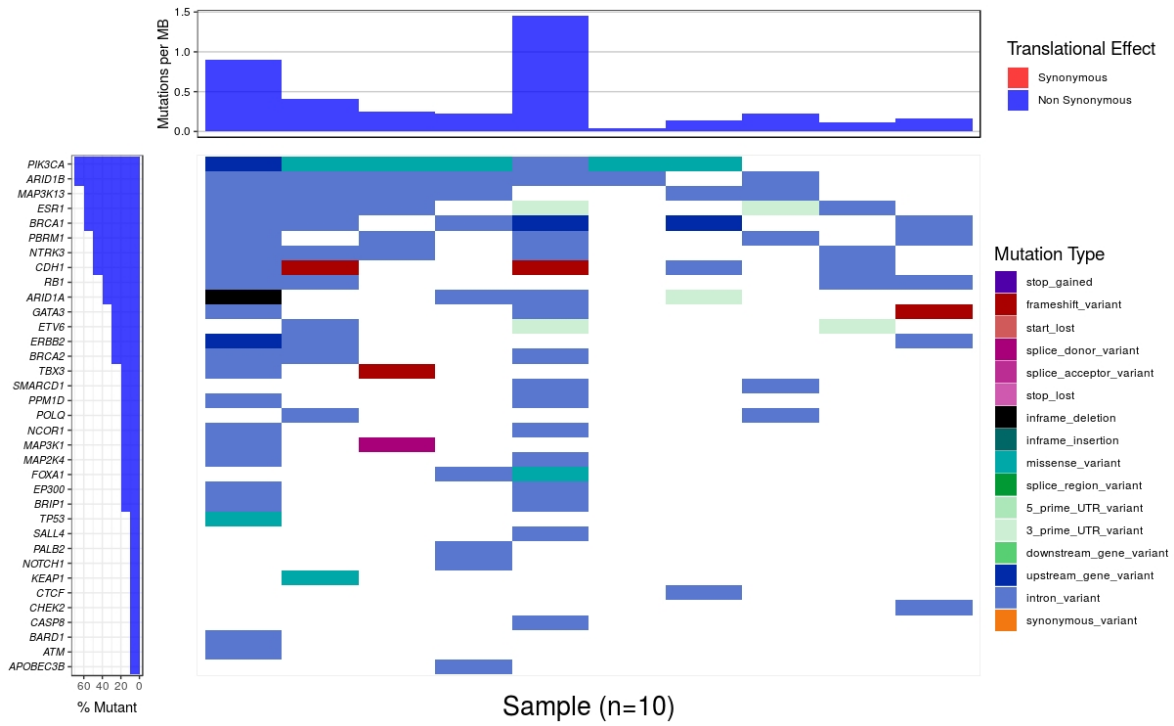
L.2 Somatic Mutation Profile Variant *CHEK2* Associated Unselected Breast Cancer: Coding and Non-Coding Variants



Somatic Mutation Profile *CHEK2* Variant Carriers: Coding and Non-Coding Variants

Waterfall plot illustrating the most prevalent somatic coding and non-coding sequence variants identified in 20 breast cancers occurring in association with a *CHEK2* germline variant derived from The 100,000 Genomes Project.

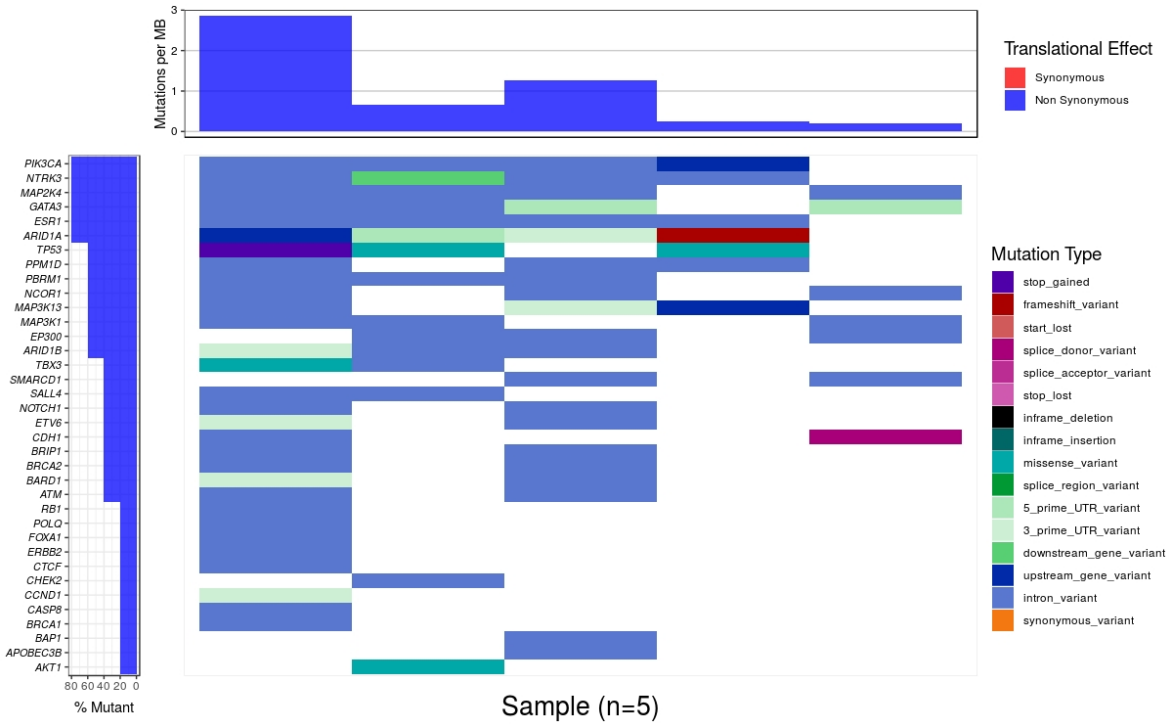
L.3 Somatic Mutation Profile *ATM* Associated Unselected Breast Cancer: Coding and Non-Coding Variants



Somatic Mutation Profile *ATM* Variant Carriers: Coding and Non-Coding Variants

Waterfall plot illustrating the most prevalent somatic coding and non-coding sequence variants identified in 10 breast cancers occurring in association with a *ATM* germline variant derived from The 100,000 Genomes Project.

L.4 Somatic Mutation Profile *TP53* Associated Unselected Breast Cancer: Coding and Non-Coding Variants



Somatic Mutation Profile *TP53* Variant Carriers: Coding and Non-Coding Variants

Waterfall plot illustrating the most prevalent somatic coding and non-coding sequence variants identified in 5 breast cancers occurring in association with a *TP53* germline variant derived from The 100,000 Genomes Project.

Appendix M Somatic Mutational Analysis

M.1 Somatic Mutational Signatures

Mutation Status	Median	Range	IQR	P-Value*
Signature 1				
MUTATION-	14.64	0.00-100.00	8.92-17.51	-
MUTATION+	10.81	3.16-18.98	7.53-14.48	0.001235
BRCA+	8.53	4.31-14.44	6.59-9.99	0.0245
PALB2+	6.71	3.16-8.87	5.10-7.01	0.001588
CHEK2+	14.10	3.94-18.60	11.40-15.32	0.4859
ATM+	11.61	6.20-18.98	10.66-16.72	0.5002
TP53+	10.57	8.22-14.82	8.64-12.99	0.2091
Signature 2				
MUTATION-	2.88	0.00-67.64	1.72-6.17	-
MUTATION+	3.59	0.00-52.60	1.46-9.51	0.5744
BRCA+	3.61	0.53-9.48	1.69-4.67	0.9205
PALB2+	2.51	0.92-9.89	1.53-4.49	0.6553
CHEK2+	5.49	0.00-52.60	2.25-11.68	0.1705
ATM+	2.91	0.00-38.20	1.50-7.32	0.8141
TP53+	6.93	0.00-16.50	0.84-9.57	0.8228
Signature 3				
MUTATION-	10.59	0.00-58.05	6.44-15.95	-
MUTATION+	14.78	0.00-45.12	8.39-29.32	0.01387
BRCA+	31.14	15.66-45.12	20.71-35.17	0.0008532
PALB2+	32.14	6.47-39.67	31.02-37.71	0.004651
CHEK2+	12.73	0.00-32.31	8.88-15.42	0.3601
ATM+	9.28	0.00-30.11	3.93-14.05	0.439
TP53+	9.35	2.05-16.18	8.03-15.86	0.8414
Signature 4				
MUTATION-	0.00	0.00-34.74	0.00-0.00	-
MUTATION+	0.00	0.00-8.83	0.00-0.00	0.01237
BRCA+	0.00	0.00-0.00	0.00-0.00	0.8153
PALB2+	0.00	0.00-0.00	0.00-0.00	0.8153
CHEK2+	0.00	0.00-8.83	0.00-0.00	<0.0001
ATM+	0.00	0.00-0.00	0.00-0.00	0.7738
TP53+	0.00	0.00-0.00	0.00-0.00	0.8316
Signature 5				
MUTATION-	28.08	0.00-61.66	11.41-38.30	-
MUTATION+	18.74	0.00-46.16	0.00-30.11	0.0021

Appendix M

Mutation Status		Median	Range	IQR	P-Value*
	BRCA+	15.08	0.00-36.97	10.95-19.84	0.1468
	PALB2+	4.88	0.00-20.78	0.00-17.36	0.01001
	CHEK2+	20.69	0.00-46.16	2.23-31.90	0.1551
	ATM+	22.32	0.00-43.47	2.53-36.35	0.5293
	TP53+	24.51	0.00-36.63	0.00-25.93	0.2432
Signature 6					
	MUTATION-	0.00	0.00-21.81	0.00-0.00	-
	MUTATION+	0.00	0.00-1.51	0.00-0.00	0.5273
	BRCA+	0.00	0.00-1.51	0.00-0.00	0.6697
	PALB2+	0.17	0.00-1.00	0.00-0.64	0.0048
	CHEK2+	0.00	0.00-0.00	0.00-0.00	0.1678
	ATM+	0.00	0.00-1.15	0.00-0.00	0.31
	TP53+	0.00	0.00-0.00	0.00-0.00	0.4099
Signature 7					
	MUTATION-	0.00	0.00-76.95	0.00-0.00	-
	MUTATION+	0.00	0.00-1.06	0.00-0.00	0.807
	BRCA+	0.00	0.00-1.06	0.00-0.77	0.05791
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.3844
	CHEK2+	0.00	0.00-0.21	0.00-0.00	0.8184
	ATM+	0.00	0.00-0.00	0.00-0.00	0.2866
	TP53+	0.00	0.00-0.00	0.00-0.00	0.4272
Signature 8					
	MUTATION-	3.94	0.00-36.10	0.00-9.33	-
	MUTATION+	8.17	0.00-25.31	2.83-12.74	0.002591
	BRCA+	8.90	0.16-20.53	4.63-14.21	0.1165
	PALB2+	12.93	4.39-25.31	8.50-18.92	0.006834
	CHEK2+	4.01	0.00-11.48	1.96-9.06	0.7858
	ATM+	10.33	0.00-21.12	0.07-13.22	0.275
	TP53+	8.92	5.65-14.73	7.89-12.77	0.0425
Signature 9					
	MUTATION-	4.67	0.00-29.83	2.87-6.51	-
	MUTATION+	4.90	0.16-18.67	3.08-6.01	0.7754
	BRCA+	6.09	4.62-13.15	5.87-9.48	0.03741
	PALB2+	5.55	1.95-18.67	3.43-5.66	0.7065
	CHEK2+	4.08	0.16-5.86	2.88-4.91	0.1366
	ATM+	4.23	1.89-11.03	2.48-7.02	0.9048
	TP53+	5.47	2.65-9.53	3.22-6.62	0.6443
Signature 10					
	MUTATION-	0.09	0.00-10.89	0.00-0.73	-
	MUTATION+	0.18	0.00-6.22	0.00-0.83	0.6558

Mutation Status		Median	Range	IQR	P-Value*
	BRCA+	0.00	0.00-1.09	0.00-0.16	0.31
	PALB2+	0.01	0.00-0.83	0.00-0.63	0.6749
	CHEK2+	0.13	0.00-2.61	0.00-0.76	0.8022
	ATM+	0.65	0.00-6.22	0.32-0.90	0.06983
	TP53+	0.30	0.00-0.87	0.00-0.44	0.9945
Signature 11					
	MUTATION-	0.98	0.00-9.30	0.00-1.84	-
	MUTATION+	1.22	0.00-4.43	0.47-2.01	0.3411
	BRCA+	1.08	0.42-2.33	0.60-1.31	0.7258
	PALB2+	0.79	0.00-4.26	0.12-1.52	0.844
	CHEK2+	1.28	0.00-3.03	0.67-2.04	0.3761
	ATM+	1.39	0.00-2.35	0.55-1.75	0.6475
	TP53+	1.51	0.00-4.43	0.00-2.61	0.6466
Signature 12					
	MUTATION-	4.62	0.00-55.35	1.40-9.08	-
	MUTATION+	4.48	0.00-16.53	1.56-6.98	0.5634
	BRCA+	4.59	3.30-9.33	4.20-7.83	0.6132
	PALB2+	5.88	0.48-16.53	2.29-6.87	0.7533
	CHEK2+	5.33	0.00-14.89	1.84-10.13	0.8885
	ATM+	2.98	0.00-8.83	0.00-4.71	0.1761
	TP53+	3.30	0.00-6.97	0.00-6.79	0.3304
Signature 13					
	MUTATION-	1.78	0.00-68.91	0.88-4.65	-
	MUTATION+	2.74	0.00-30.07	1.43-8.27	0.03463
	BRCA+	6.78	1.93-13.22	4.28-8.14	0.02651
	PALB2+	3.45	0.87-20.64	1.36-10.40	0.3317
	CHEK2+	3.50	0.00-30.07	0.89-7.85	0.3839
	ATM+	1.91	0.78-24.94	1.49-5.42	0.3709
	TP53+	1.80	0.00-14.28	1.62-2.04	0.9163
Signature 14					
	MUTATION-	0.00	0.00-16.63	0.00-0.00	-
	MUTATION+	0.00	0.00-0.04	0.00-0.00	0.8143
	BRCA+	0.00	0.00-0.00	0.00-0.00	0.7328
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.7328
	CHEK2+	0.00	0.00-0.00	0.00-0.00	0.6007
	ATM+	0.00	0.00-0.04	0.00-0.00	0.05518
	TP53+	0.00	0.00-0.00	0.00-0.00	0.7557
Signature 15					
	MUTATION-	0.00	0.00-8.70	0.00-0.00	-
	MUTATION+	0.00	0.00-0.70	0.00-0.00	0.2769

Appendix M

Mutation Status	Median	Range	IQR	P-Value*
BRCA+	0.00	0.00-0.00	0.00-0.00	0.405
PALB2+	0.00	0.00-0.00	0.00-0.00	0.405
CHEK2+	0.00	0.00-0.46	0.00-0.00	0.694
ATM+	0.00	0.00-0.70	0.00-0.00	0.9126
TP53+	0.00	0.00-0.00	0.00-0.00	0.4472
Signature 16				
MUTATION-	8.90	0.00-52.01	2.64-16.74	-
MUTATION+	8.45	0.00-48.30	3.58-16.72	0.6423
BRCA+	5.60	0.00-16.33	0.90-12.53	0.3238
PALB2+	7.73	0.12-21.25	2.13-11.13	0.6516
CHEK2+	11.18	0.00-48.30	4.40-16.82	0.579
ATM+	7.36	3.25-47.29	6.40-20.56	0.5205
TP53+	10.36	3.55-42.94	8.28-32.15	0.266
Signature 17				
MUTATION-	0.00	0.00-56.44	0.00-0.00	-
MUTATION+	0.00	0.00-2.57	0.00-0.00	0.7367
BRCA+	0.00	0.00-2.57	0.00-0.00	0.1091
PALB2+	0.00	0.00-0.12	0.00-0.00	0.1234
CHEK2+	0.00	0.00-0.00	0.00-0.00	0.4542
ATM+	0.00	0.00-0.00	0.00-0.00	0.5487
TP53+	0.00	0.00-0.00	0.00-0.00	0.6555
Signature 18				
MUTATION-	0.00	0.00-31.43	0.00-0.06	-
MUTATION+	0.00	0.00-9.49	0.00-0.78	0.01927
BRCA+	0.00	0.00-1.03	0.00-0.14	0.8592
PALB2+	0.00	0.00-0.88	0.00-0.00	0.5729
CHEK2+	0.07	0.00-7.65	0.00-1.93	0.03131
ATM+	0.36	0.00-9.49	0.00-3.20	0.02235
TP53+	0.00	0.00-1.13	0.00-0.18	0.6395
Signature 19				
MUTATION-	0.00	0.00-7.18	0.00-0.00	-
MUTATION+	0.00	0.00-0.57	0.00-0.00	0.1866
BRCA+	0.00	0.00-0.00	0.00-0.00	0.4636
PALB2+	0.00	0.00-0.57	0.00-0.00	0.4676
CHEK2+	0.00	0.00-0.00	0.00-0.00	0.2626
ATM+	0.00	0.00-0.00	0.00-0.00	0.3692
TP53+	0.00	0.00-0.00	0.00-0.00	0.5036
Signature 20				
MUTATION-	0.00	0.00-7.83	0.00-1.25	-
MUTATION+	0.00	0.00-11.47	0.00-0.87	0.5513

Mutation Status		Median	Range	IQR	P-Value*
	BRCA+	0.10	0.00-2.92	0.00-0.99	0.5539
	PALB2+	0.35	0.00-2.29	0.00-1.56	0.75
	CHEK2+	0.00	0.00-11.47	0.00-1.61	0.8974
	ATM+	0.00	0.00-0.83	0.00-0.00	0.1432
	TP53+	0.00	0.00-0.98	0.00-0.46	0.6011
Signature 21					
	MUTATION-	0.00	0.00-5.14	0.00-0.00	-
	MUTATION+	0.00	0.00-0.40	0.00-0.00	0.7624
	BRCA+	0.00	0.00-0.00	0.00-0.00	0.6212
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.6212
	CHEK2+	0.00	0.00-0.00	0.00-0.00	0.4495
	ATM+	0.00	0.00-0.40	0.00-0.00	0.007127
	TP53+	0.00	0.00-0.00	0.00-0.00	0.6521
Signature 22					
	MUTATION-	0.00	0.00-0.57	0.00-0.00	-
	MUTATION+	0.00	0.00-0.00	0.00-0.00	0.5779
	BRCA+	0.00	0.00-0.00	0.00-0.00	0.8317
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.8317
	CHEK2+	0.00	0.00-0.00	0.00-0.00	0.7431
	ATM+	0.00	0.00-0.00	0.00-0.00	0.7936
	TP53+	0.00	0.000.00	0.00-0.00	0.8467
Signature 23					
	MUTATION-	0.00	0.00-0.05	0.00-0.00	-
	MUTATION+	0.00	0.00-0.00	0.00-0.00	0.8639
	BRCA+	0.00	0.00-0.00	0.00-0.00	0.9547
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.9547
	CHEK2+	0.00	0.00-0.00	0.00-0.00	0.9231
	ATM+	0.00	0.00-0.00	0.00-0.00	0.9409
	TP53+	0.00	0.00-0.00	0.00-0.00	0.9603
Signature 24					
	MUTATION-	0.00	0.00-1.06	0.00-0.00	-
	MUTATION+	0.00	0.00-0.00	0.00-0.00	0.806
	BRCA+	0.00	0.00-0.00	0.00-0.00	0.9296
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.9296
	CHEK2+	0.00	0.00-0.00	0.00-0.00	0.8872
	ATM+	0.00	0.00-0.00	0.00-0.00	0.9113
	TP53+	0.00	0.00-0.00	0.00-0.00	0.9369
Signature 25					
	MUTATION-	0.00	0.00-14.32	0.00-0.60	-
	MUTATION+	0.00	0.00-14.09	0.00-2.11	0.3894

Appendix M

Mutation Status		Median	Range	IQR	P-Value*
	BRCA+	0.00	0.00-2.44	0.00-0.62	0.9393
	PALB2+	1.67	0.00-8.52	0.20-6.64	0.06678
	CHEK2+	0.00	0.00-9.22	0.00-0.53	0.9387
	ATM+	0.00	0.00-14.09	0.00-0.00	0.7794
	TP53+	0.00	0.00-13.15	0.00-10.01	0.4304
Signature 26					
	MUTATION-	0.00	0.00-7.25	0.00-0.00	-
	MUTATION+	0.00	0.00-2.22	0.00-0.00	0.9372
	BRCA+	0.00	0.00-0.25	0.00-0.00	0.4645
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.4799
	CHEK2+	0.00	0.00-2.22	0.00-0.00	0.359
	ATM+	0.00	0.00-0.00	0.00-0.00	0.3866
	TP53+	0.00	0.00-0.00	0.00-0.00	0.5191
Signature 27					
	MUTATION-	0.15	0.00-3.09	0.00-0.40	-
	MUTATION+	0.28	0.00-1.30	0.00-0.50	0.2793
	BRCA+	0.28	0.00-0.65	0.19-0.38	0.4116
	PALB2+	0.23	0.00-0.82	0.06-0.42	0.6188
	CHEK2+	0.25	0.00-0.67	0.00-0.57	0.5398
	ATM+	0.15	0.00-0.68	0.00-0.49	0.8482
	TP53+	0.35	0.00-1.30	0.00-0.40	0.7133
Signature 28					
	MUTATION-	0.00	0.00-9.09	0.00-0.28	-
	MUTATION+	0.00	0.00-6.03	0.00-0.45	0.7768
	BRCA+	0.00	0.00-0.00	0.00-0.00	0.09304
	PALB2+	0.74	0.00-6.03	0.18-0.76	0.0589
	CHEK2+	0.00	0.00-3.51	0.00-0.14	0.6365
	ATM+	0.00	0.00-2.75	0.00-0.00	0.5923
	TP53+	0.36	0.00-5.05	0.06-2.09	0.0368
Signature 29					
	MUTATION-	0.00	0.00-4.53	0.00-0.00	-
	MUTATION+	0.00	0.00-0.19	0.00-0.00	0.1937
	BRCA+	0.00	0.00-0.00	0.00-0.00	0.8406
	PALB2+	0.00	0.00-0.00	0.00-0.00	0.8406
	CHEK2+	0.00	0.00-0.00	0.00-0.00	0.7562
	ATM+	0.00	0.00-0.19	0.00-0.00	0.0003696
	TP53+	0.00	0.00-0.00	0.00-0.00	0.8549
Signature 30					
	MUTATION-	0.00	0.00-45.79	0.00-3.48	-
	MUTATION+	0.31	0.00-8.26	0.00-2.50	0.928

Mutation Status		Median	Range	IQR	P-Value*
	BRCA+	0.13	0.00-2.50	0.00-0.34	0.4689
	PALB2+	0.35	0.00-7.59	0.00-1.65	0.9233
	CHEK2+	0.00	0.00-7.70	0.00-1.89	0.6821
	ATM+	1.77	0.00-6.55	0.00-3.85	0.4217
	TP53+	1.88	0.00-8.26	0.00-5.03	0.4433

References

1. Cancer Research UK. Breast Cancer Statistics 2017 [Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer.
2. Tedaldi G, Tebaldi M, Zampiga V, Danesi R, Arcangeli V, Ravegnani M, et al. Multiple-gene panel analysis in a case series of 255 women with hereditary breast and ovarian cancer. Oncotarget. 2017;8(29):47064-75.
3. O'Leary E, Iacoboni D, Holle J, Michalski ST, Esplin ED, Yang S, et al. Expanded Gene Panel Use for Women With Breast Cancer: Identification and Intervention Beyond Breast Cancer Risk. Ann Surg Oncol. 2017.
4. Apostolou P, Fostira F. Hereditary breast cancer: the era of new susceptibility genes. Biomed Res Int. 2013;2013:747318.
5. Rudolph A, Song M, Brook MN, Milne RL, Mavaddat N, Michailidou K, et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. Int J Epidemiol. 2018;47(2):526-36.
6. Kluttig A, Schmidt-Pokrzywniak A. Established and Suspected Risk Factors in Breast Cancer Aetiology. Breast Care (Basel). 2009;4(2):82-7.
7. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br J Cancer. 2011;105 Suppl 2:S77-81.
8. Zimmern RL, Kroese M. The evaluation of genetic tests. J Public Health (Oxf). 2007;29(3):246-50.
9. Sanderson S, Zimmern R, Kroese M, Higgins J, Patch C, Emery J. How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. Genetics in Medicine. 2005;7(7):495-500.
10. Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372(23):2243-57.
11. Excellence NIfHa. Familial Breast Cancer: Classification Care and Managing Breast Cancer and Related Risks in People with A Family History of Breast Cancer. 2017 [Available from: https://www.nice.org.uk/guidance/cg164.
12. Economopoulou P, Dimitriadis G, Psyrri A. Beyond BRCA: new hereditary breast cancer susceptibility genes. Cancer Treat Rev. 2015;41(1):1-8.
13. Copson E, Eccles B, Maishman T, Gerty S, Stanton L, Cutress RI, et al. Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study. J Natl Cancer Inst. 2013;105(13):978-88.
14. Eccles D, Gerty S, Simmonds P, Hammond V, Ennis S, Altman DG, et al. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. BMC Cancer. 2007;7:160.
15. Buys SS, Sandbach JF, Gammon A, Patel G, Kidd J, Brown KL, et al. A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. Cancer. 2017.
16. Yanes T, Young MA, Meiser B, James PA. Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. Breast Cancer Res. 2020;22(1):21.
17. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet. 2016;17(7):392-406.
18. Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. J Natl Cancer Inst. 2017;109(7).
19. van Veen EM, Brentnall AR, Byers H, Harkness EF, Astley SM, Sampson S, et al. Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. JAMA Oncol. 2018;4(4):476-82.

20. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, et al. Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. J Clin Oncol. 2020;38(7):674-85.
21. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015;107(5).
22. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-4.
23. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2019;104(1):21-34.
24. Li H, Feng B, Miron A, Chen X, Beesley J, Bimeh E, et al. Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. Genet Med. 2017;19(1):30-5.
25. Larsen MJ, Thomassen M, Gerdes AM, Kruse TA. Hereditary breast cancer: clinical, pathological and molecular characteristics. Breast Cancer (Auckl). 2014;8:145-55.
26. Brough R, Bajrami I, Vatcheva R, Natrajan R, Reis-Filho JS, Lord CJ, et al. APRIN is a cell cycle specific BRCA2-interacting protein required for genome integrity and a predictor of outcome after chemotherapy in breast cancer. EMBO J. 2012;31(5):1160-76.
27. Goldgar DE, Healey S, Dowty JG, Da Silva L, Chen X, Spurdle AB, et al. Rare variants in the ATM gene and risk of breast cancer. Breast Cancer Res. 2011;13(4):R73.
28. Marabelli M, Cheng SC, Parmigiani G. Penetrance of ATM Gene Mutations in Breast Cancer: A Meta-Analysis of Different Measures of Risk. Genet Epidemiol. 2016;40(5):425-31.
29. Ahmed M, Rahman N. ATM and breast cancer susceptibility. Oncogene. 2006;25(43):5906-11.
30. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med. 2014;371(6):497-506.
31. Nguyen-Dumont T, Hammet F, Mahmoodi M, Tsimiklis H, Teo ZL, Li R, et al. Mutation screening of PALB2 in clinically ascertained families from the Breast Cancer Family Registry. Breast Cancer Res Treat. 2015;149(2):547-54.
32. Guo Y, Feng W, Sy SM, Huen MS. ATM-dependent Phosphorylation of the Fanconi Anemia Protein PALB2 Promotes the DNA Damage Response. J Biol Chem. 2015;290(46):27545-56.
33. Hartford SA, Chittela R, Ding X, Vyas A, Martin B, Burkett S, et al. Interaction with PALB2 Is Essential for Maintenance of Genomic Integrity by BRCA2. PLoS Genet. 2016;12(8):e1006236.
34. Li J, Wen WX, Eklund M, Kvist A, Eriksson M, Christensen HN, et al. Prevalence of BRCA1 and BRCA2 pathogenic variants in a large, unselected breast cancer cohort. Int J Cancer. 2019;144(5):1195-204.
35. Engel C, Rhiem K, Hahnen E, Loibl S, Weber KE, Seiler S, et al. Prevalence of pathogenic BRCA1/2 germline mutations among 802 women with unilateral triple-negative breast cancer without family cancer history. BMC Cancer. 2018;18(1):265.
36. Copson ER, Maishman TC, Tapper WJ, Cutress RI, Greville-Heygate S, Altman DG, et al. Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. The Lancet Oncology. 2018;19(2):169-80.
37. Kraus C, Hoyer J, Vasileiou G, Wunderle M, Lux MP, Fasching PA, et al. Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. Int J Cancer. 2017;140(1):95-102.
38. Francies FZ, Wainstein T, De Leeneer K, Cairns A, Murdoch M, Nietz S, et al. BRCA1, BRCA2 and PALB2 mutations and CHEK2 c.1100delC in different South African ethnic groups diagnosed with premenopausal and/or triple negative breast cancer. BMC Cancer. 2015;15:912.
39. Anglian Breast Cancer Study Group. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br J Cancer. 2000;83(10):1301-8.
40. John EM, Miron A, Gong G, Phipps AI, Felberg A, Li FP, et al. Prevalence of Pathogenic BRCA1 Mutation Carriers in 5 US Racial/Ethnic Groups. JAMA. 2007;298(24):2869-76.

41. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017;317(23):2402-16.
42. Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, et al. Cancer Risks for BRCA1 and BRCA2 Mutation Carriers: Results From Prospective Analysis of EMBRACE. JNCI: Journal of the National Cancer Institute. 2013;105(11):812-22.
43. Gallagher S, Hughes E, Wagner S, Tshiaba P, Rosenthal E, Roa BB, et al. Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. JAMA network open. 2020;3(7):e208501-e.
44. Zhen DB, Rabe KG, Gallinger S, Syngal S, Schwartz AG, Goggins MG, et al. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. Genet Med. 2015;17(7):569-77.
45. Mersch J, Jackson MA, Park M, Nebgen D, Peterson SK, Singletary C, et al. Cancers associated with BRCA1 and BRCA2 mutations other than breast and ovarian. Cancer. 2015;121(2):269-75.
46. Nepomuceno TC, De Gregoriis G, de Oliveira FMB, Suarez-Kurtz G, Monteiro AN, Carvalho MA. The Role of PALB2 in the DNA Damage Response and Cancer Predisposition. Int J Mol Sci. 2017;18(9).
47. Li A, Geyer FC, Blecua P, Lee JY, Selenica P, Brown DN, et al. Homologous recombination DNA repair defects in PALB2-associated breast cancers. NPJ Breast Cancer. 2019;5:23.
48. Pauty J, Couturier AM, Rodrigue A, Caron MC, Coulombe Y, Dellaire G, et al. Cancer-causing mutations in the tumor suppressor PALB2 reveal a novel cancer mechanism using a hidden nuclear export signal in the WD40 repeat motif. Nucleic Acids Res. 2017;45(5):2644-57.
49. Park JY, Zhang F, Andreassen PR. PALB2: the hub of a network of tumor suppressors involved in DNA damage responses. Biochim Biophys Acta. 2014;1846(1):263-75.
50. Southey MC, Teo ZL, Winship I. PALB2 and breast cancer: ready for clinical translation! Appl Clin Genet. 2013;6:43-52.
51. Luijsterburg MS, Typas D, Caron MC, Wiegant WW, van den Heuvel D, Boonen RA, et al. A PALB2-interacting domain in RNF168 couples homologous recombination to DNA break-induced chromatin ubiquitylation. Elife. 2017;6.
52. Wesola M, Jelen M. The risk of breast cancer due to PALB2 gene mutations. Adv Clin Exp Med. 2017;26(2):339-42.
53. Heikkinen T, Karkkainen H, Aaltonen K, Milne RL, Heikkila P, Aittomaki K, et al. The breast cancer susceptibility mutation PALB2 1592delT is associated with an aggressive tumor phenotype. Clin Cancer Res. 2009;15(9):3214-22.
54. Wojcik P, Jasiowka M, Strycharz E, Sobol M, Hodorowicz-Zaniewska D, Skotnicki P, et al. Recurrent mutations of BRCA1, BRCA2 and PALB2 in the population of breast and ovarian cancer patients in Southern Poland. Hered Cancer Clin Pract. 2016;14:5.
55. Southey MC, Winship I, Nguyen-Dumont T. PALB2: research reaching to clinical outcomes for women with breast cancer. Hered Cancer Clin Pract. 2016;14:9.
56. Decker B, Allen J, Luccarini C, Pooley KA, Shah M, Bolla MK, et al. Rare, protein-truncating variants in ATM, CHEK2 and PALB2, but not XRCC2, are associated with increased breast cancer risks. J Med Genet. 2017;54(11):732-41.
57. Southey MC, Goldgar DE, Winqvist R, Pylkas K, Couch F, Tischkowitz M, et al. PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet. 2016;53(12):800-11.
58. Zhang K, Zhou J, Zhu X, Luo M, Xu C, Yu J, et al. Germline mutations of PALB2 gene in a sequential series of Chinese patients with breast cancer. Breast Cancer Res Treat. 2017;166(3):865-73.
59. Yang XR, Devi BCR, Sung H, Guida J, Mucaki EJ, Xiao Y, et al. Prevalence and spectrum of germline rare variants in BRCA1/2 and PALB2 among breast cancer cases in Sarawak, Malaysia. Breast Cancer Res Treat. 2017;165(3):687-97.
60. Kluska A, Balabas A, Piatkowska M, Czarny K, Paczkowska K, Nowakowska D, et al. PALB2 mutations in BRCA1/2-mutation negative breast and ovarian cancer patients from Poland. BMC Med Genomics. 2017;10(1):14.

61. Myszka A, Nguyen-Dumont T, Karpinski P, Sasiadek MM, Akopyan H, Hammet F, et al. Targeted massively parallel sequencing characterises the mutation spectrum of PALB2 in breast and ovarian cancer cases from Poland and Ukraine. Fam Cancer. 2017.
62. Lerner-Ellis J, Donenberg T, Ahmed H, George S, Wharfe G, Chin S, et al. A high frequency of PALB2 mutations in Jamaican patients with breast cancer. Breast Cancer Res Treat. 2017;162(3):591-6.
63. Foo TK, Tischkowitz M, Simhadri S, Boshari T, Zayed N, Burke KA, et al. Compromised BRCA1-PALB2 interaction is associated with breast cancer risk. Oncogene. 2017;36(29):4161-70.
64. Thompson ER, Gorringe KL, Rowley SM, Wong-Brown MW, McInerny S, Li N, et al. Prevalence of PALB2 mutations in Australian familial breast cancer cases and controls. Breast Cancer Res. 2015;17:111.
65. Catucci I, Radice P, Milne RL, Couch FJ, Southey MC, Peterlongo P. The PALB2 p.Leu939Trp mutation is not associated with breast cancer risk. Breast Cancer Res. 2016;18(1):111.
66. Cybulski C, Kluźniak W, Huzarski T, Wokołorczyk D, Kashyap A, Jakubowska A, et al. Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. The Lancet Oncology. 2015;16(6):638-44.
67. Silvestri V, Zelli V, Valentini V, Rizzolo P, Navazio AS, Coppa A, et al. Whole-exome sequencing and targeted gene sequencing provide insights into the role of PALB2 as a male breast cancer susceptibility gene. Cancer. 2017;123(2):210-8.
68. Ramus SJ, Song H, Dicks E, Tyrer JP, Rosenthal AN, Intermaggio MP, et al. Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. J Natl Cancer Inst. 2015;107(11).
69. Borecka M, Zemankova P, Vocka M, Soucek P, Soukupova J, Kleiblova P, et al. Mutation analysis of the PALB2 gene in unselected pancreatic cancer patients in the Czech Republic. Cancer Genet. 2016;209(5):199-204.
70. Aoude LG, Xu M, Zhao ZZ, Kovacs M, Palmer JM, Johansson P, et al. Assessment of PALB2 as a candidate melanoma susceptibility gene. PLoS One. 2014;9(6):e100683.
71. Apostolou P, Papasotiriou I. Current perspectives on CHEK2 mutations in breast cancer. Breast Cancer (Dove Med Press). 2017;9:331-5.
72. Muranen TA, Blomqvist C, Dork T, Jakubowska A, Heikkila P, Fagerholm R, et al. Patient survival and tumor characteristics associated with CHEK2:p.I157T - findings from the Breast Cancer Association Consortium. Breast Cancer Res. 2016;18(1):98.
73. Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, Nordestgaard BG. CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. J Clin Oncol. 2008;26(4):542-8.
74. Roeb W, Higgins J, King MC. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. Hum Mol Genet. 2012;21(12):2738-44.
75. Kilpivaara O, Vahteristo P, Falck J, Syrjakoski K, Eerola H, Easton D, et al. CHEK2 variant I157T may be associated with increased breast cancer risk. Int J Cancer. 2004;111(4):543-7.
76. Jalilvand M, Oloomi M, Najafipour R, Alizadeh SA, Saki N, Rad FS, et al. An association study between CHEK2 gene mutations and susceptibility to breast cancer. Comp Clin Path. 2017;26(4):837-45.
77. Matsuoka S, Huang M, Elledge SJ. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. Science. 1998;282(5395):1893-7.
78. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55-9.
79. Muranen TA, Greco D, Blomqvist C, Aittomaki K, Khan S, Hogervorst F, et al. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. Genet Med. 2017;19(5):599-603.
80. Schmidt MK, Hogervorst F, van Hien R, Cornelissen S, Broeks A, Adank MA, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. J Clin Oncol. 2016;34(23):2750-60.
81. Leedom TP, LaDuca H, McFarland R, Li S, Dolinsky JS, Chao EC. Breast cancer risk is similar for CHEK2 founder and non-founder mutation carriers. Cancer Genet. 2016;209(9):403-7.

82. Adank MA, Jonker MA, Kluijt I, van Mil SE, Oldenburg RA, Mooi WJ, et al. CHEK2*1100delC homozygosity is associated with a high breast cancer risk in women. J Med Genet. 2011;48(12):860-3.
83. Kim H, Cho DY, Choi DH, Oh M, Shin I, Park W, et al. Frequency of pathogenic germline mutation in CHEK2, PALB2, MRE11, and RAD50 in patients at high risk for hereditary breast cancer. Breast Cancer Res Treat. 2017;161(1):95-102.
84. Mohamad S, Isa NM, Muhammad R, Emran NA, Kitan NM, Kang P, et al. Low prevalence of CHEK2 gene mutations in multiethnic cohorts of breast cancer patients in Malaysia. PLoS One. 2015;10(1):e0117104.
85. Marouf C, Hajji O, Diakite B, Tazzite A, Jouhadi H, Benider A, et al. The CHEK2 1100delC allelic variant is not present in familial and sporadic breast cancer cases from Moroccan population. Springerplus. 2015;4:38.
86. Apostolou P, Fostira F, Papamentzelopoulou M, Michelli M, Panopoulos C, Fountzilas G, et al. CHEK2 c.1100delC allele is rarely identified in Greek breast cancer cases. Cancer Genet. 2015;208(4):129-34.
87. Hallamies S, Pelttari LM, Poikonen-Saksela P, Jekunen A, Jukkola-Vuorinen A, Auvinen P, et al. CHEK2 c.1100delC mutation is associated with an increased risk for male breast cancer in Finnish patient population. BMC Cancer. 2017;17(1):620.
88. Adank MA, Verhoef S, Oldenburg RA, Schmidt MK, Hooning MJ, Martens JW, et al. Excess breast cancer risk in first degree relatives of CHEK2 *1100delC positive familial breast cancer cases. Eur J Cancer. 2013;49(8):1993-9.
89. Liu Y, Xu Y, Ouyang T, Li J, Wang T, Fan Z, et al. Association between CHEK2 H371Y mutation and response to neoadjuvant chemotherapy in women with breast cancer. BMC Cancer. 2015;15:194.
90. Cybulski C, Wokolorczyk D, Jakubowska A, Huzarski T, Byrski T, Gronwald J, et al. Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. J Clin Oncol. 2011;29(28):3747-52.
91. Weischer M, Nordestgaard BG, Pharoah P, Bolla MK, Nevanlinna H, Van't Veer LJ, et al. CHEK2*1100delC heterozygosity in women with breast cancer associated with early death, breast cancer-specific death, and increased risk of a second breast cancer. J Clin Oncol. 2012;30(35):4308-16.
92. Naslund-Koch C, Nordestgaard BG, Bojesen SE. Increased Risk for Other Cancers in Addition to Breast Cancer for CHEK2*1100delC Heterozygotes Estimated From the Copenhagen General Population Study. J Clin Oncol. 2016;34(11):1208-16.
93. Tavera-Tapia A, Perez-Cabornero L, Macias JA, Ceballos MI, Roncador G, de la Hoya M, et al. Almost 2% of Spanish breast cancer families are associated to germline pathogenic mutations in the ATM gene. Breast Cancer Res Treat. 2017;161(3):597-604.
94. Prodosmo A, Buffone A, Mattioni M, Barnabei A, Persichetti A, De Leo A, et al. Detection of ATM germline variants by the p53 mitotic centrosomal localization test in BRCA1/2-negative patients with early-onset breast cancer. J Exp Clin Cancer Res. 2016;35(1):135.
95. Prokopcova J, Kleibl Z, Banwell CM, Pohlreich P. The role of ATM in breast cancer development. Breast Cancer Res Treat. 2007;104(2):121-8.
96. Wang K, Ye Y, Xu Z, Zhang X, Hou Z, Cui Y, et al. Interaction between BRCA1/BRCA2 and ATM/ATR associate with breast cancer susceptibility in a Chinese Han population. Cancer Genet Cytogenet. 2010;200(1):40-6.
97. Gonzalez-Hormazabal P, Bravo T, Blanco R, Valenzuela CY, Gomez F, Waugh E, et al. Association of common ATM variants with familial breast cancer in a South American population. BMC Cancer. 2008;8:117.
98. Stredrick DL, Garcia-Closas M, Pineda MA, Bhatti P, Alexander BH, Doody MM, et al. The ATM missense mutation p.Ser49Cys (c.146C>G) and the risk of breast cancer. Hum Mutat. 2006;27(6):538-44.
99. Mangone FR, Miracca EC, Feilotter HE, Mulligan LM, Nagai MA. ATM gene mutations in sporadic breast cancer patients from Brazil. Springerplus. 2015;4:23.

100. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. Am J Hum Genet. 2009;85(4):427-46.
101. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 2006;38(8):873-5.
102. Mao C, Chung VC, He BF, Luo RC, Tang JL. Association between ATM 5557G>A polymorphism and breast cancer risk: a meta-analysis. Mol Biol Rep. 2012;39(2):1113-8.
103. Ding H, Mao C, Li SM, Liu Q, Lin L, Chen Q. Lack of association between ATM C.1066-6T > G mutation and breast cancer risk: a meta-analysis of 8,831 cases and 4,957 controls. Breast Cancer Res Treat. 2011;125(2):473-7.
104. Fletcher O, Johnson N, dos Santos Silva I, Orr N, Ashworth A, Nevanlinna H, et al. Missense variants in ATM in 26,101 breast cancer cases and 29,842 controls. Cancer Epidemiol Biomarkers Prev. 2010;19(9):2143-51.
105. Gatti RA, Tward A, Concannon P. Cancer risk in ATM heterozygotes: A model of phenotypic and mechanistic differences between missense and truncating mutations. Mol Genet Metab. 1999;68(4):419-23.
106. Schon K, Tischkowitz M. Clinical implications of germline mutations in breast cancer: TP53. Breast Cancer Res Treat. 2018;167(2):417-23.
107. Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, et al. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. Hum Mutat. 2016;37(9):865-76.
108. Mai PL, Best AF, Peters JA, DeCastro RM, Khincha PP, Loud JT, et al. Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. Cancer. 2016;122(23):3673-81.
109. Garritano S, Gemignani F, Palmero EI, Olivier M, Martel-Planche G, Le Calvez-Kelm F, et al. Detailed haplotype analysis at the TP53 locus in p.R337H mutation carriers in the population of Southern Brazil: evidence for a founder effect. Hum Mutat. 2010;31(2):143-50.
110. Andrade KC, Santiago KM, Fortes FP, Mambelli LI, Nobrega AF, Achatz MI. Early-onset breast cancer patients in the South and Southeast of Brazil should be tested for the TP53 p.R337H mutation. Genet Mol Biol. 2016;39(2):199-202.
111. Lalloo F, Varley J, Ellis D, Moran A, O'Dair L, Pharoah P, et al. Prediction of pathogenic mutations in patients with early-onset breast cancer by family history. The Lancet. 2003;361(9363):1101-2.
112. McCuaig JM, Armel SR, Novokmet A, Ginsburg OM, Demsky R, Narod SA, et al. Routine TP53 testing for breast cancer under age 30: ready for prime time? Fam Cancer. 2012;11(4):607-13.
113. O'Shea R, Clarke R, Berkley E, Giffney C, Farrell M, O'Donovan E, et al. Next generation sequencing is informing phenotype: a TP53 example. Fam Cancer. 2018;17(1):123-8.
114. Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, et al. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. J Clin Oncol. 2015;33(21):2345-52.
115. Lee DS, Yoon SY, Looi LM, Kang P, Kang IN, Sivanandan K, et al. Comparable frequency of BRCA1, BRCA2 and TP53 germline mutations in a multi-ethnic Asian cohort suggests TP53 screening should be offered together with BRCA1/2 screening to early-onset breast cancer patients. Breast Cancer Res. 2012;14(2):R66.
116. Lalloo F, Varley J, Moran A, Ellis D, O'Dair L, Pharoah P, et al. BRCA1, BRCA2 and TP53 mutations in very early-onset breast cancer with associated risks to relatives. Eur J Cancer. 2006;42(8):1143-50.
117. Rath MG, Masciari S, Gelman R, Miron A, Miron P, Foley K, et al. Prevalence of germline TP53 mutations in HER2+ breast cancer patients. Breast Cancer Res Treat. 2013;139(1):193-8.
118. Moran O, Nikitina D, Royer R, Poll A, Metcalfe K, Narod SA, et al. Revisiting breast cancer patients who previously tested negative for BRCA mutations using a 12-gene panel. Breast Cancer Res Treat. 2017;161(1):135-42.
119. Buys SS, Sandbach JF, Gammon A, Patel G, Kidd J, Brown KL, et al. A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. Cancer. 2017;123(10):1721-30.

120. Gonzalez KD, Buzin CH, Noltner KA, Gu D, Li W, Malkin D, et al. High frequency of de novo mutations in Li-Fraumeni syndrome. J Med Genet. 2009;46(10):689-93.
121. Mouchawar J, Korch C, Byers T, Pitts TM, Li E, McCredie MR, et al. Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. Cancer Res. 2010;70(12):4795-800.
122. Cury NM, Ferraz VE, Silva WA, Jr. TP53 p.R337H prevalence in a series of Brazilian hereditary breast cancer families. Hered Cancer Clin Pract. 2014;12(1):8.
123. Fitarelli-Kiehl M, Giacomazzi J, Santos-Silva P, Graudenz MS, Palmero EI, Michelli RA, et al. The breast cancer immunophenotype of TP53-p.R337H carriers is different from that observed among other pathogenic TP53 mutation carriers. Fam Cancer. 2015;14(2):333-6.
124. Giacomazzi J, Graudenz MS, Osorio CA, Koehler-Santos P, Palmero EI, Zagonel-Oliveira M, et al. Prevalence of the TP53 p.R337H mutation in breast cancer patients in Brazil. PLoS One. 2014;9(6):e99893.
125. Lolas Hamameh S, Renbaum P, Kamal L, Dweik D, Salahat M, Jaraysa T, et al. Genomic analysis of inherited breast cancer among Palestinian women: Genetic heterogeneity and a founder mutation in TP53. Int J Cancer. 2017;141(4):750-6.
126. Whibley C, Pharoah PDP, Hollstein M. p53 polymorphisms: cancer implications. Nature Reviews Cancer. 2009;9:95.
127. Sagne C, Marcel V, Amadou A, Hainaut P, Olivier M, Hall J. A meta-analysis of cancer risk associated with the TP53 intron 3 duplication polymorphism (rs17878362): geographic and tumor-specific effects. Cell Death Dis. 2013;4:e492.
128. Hossain A, Murshid GMM, Zilani MNH, Islam F, Sultana R, Sultana T, et al. TP53 codon 72 polymorphism and breast cancer risk in Bangladeshi population. Breast Cancer. 2017;24(4):571-8.
129. Damin AP, Frazzon AP, Damin DC, Roehe A, Hermes V, Zettler C, et al. Evidence for an association of TP53 codon 72 polymorphism with breast cancer risk. Cancer Detect Prev. 2006;30(6):523-9.
130. Goncalves ML, Borja SM, Cordeiro JA, Saddi VA, Ayres FM, Vilanova-Costa CA, et al. Association of the TP53 codon 72 polymorphism and breast cancer risk: a meta-analysis. Springerplus. 2014;3:749.
131. Melhem-Bertrandt A, Bojadzieva J, Ready KJ, Obeid E, Liu DD, Gutierrez-Barrera AM, et al. Early onset HER2-positive breast cancer is associated with germline TP53 mutations. Cancer. 2012;118(4):908-13.
132. Birch JM, Blair V, Kelsey AM, Evans DG, Harris M, Tricker KJ, et al. Cancer phenotype correlates with constitutional TP53 genotype in families with the Li-Fraumeni syndrome. Oncogene. 1998;17(9):1061-8.
133. Burke W, Zimmern R. Moving Beyond ACCE: An Expanded Framework for Genetic Test Evaluation. PHG Foundation; 2007.
134. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med. 2009;11(1):3-14.
135. The National Institute of Clinical Excellence. Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer 2015 [Available from: https://www.nice.org.uk/guidance/cg164.
136. Evans DG, Eccles DM, Rahman N, Young K, Bulman M, Amir E, et al. A new scoring system for the chances of identifying a BRCA1/2 mutation outperforms existing models including BRCAPRO. J Med Genet. 2004;41(6):474-80.
137. Evans DG, Lalloo F, Wallace A, Rahman N. Update on the Manchester Scoring System for BRCA1 and BRCA2 testing. J Med Genet. 2005;42(7):e39.
138. Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC, et al. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. Br J Cancer. 2014;110(2):535-45.
139. Feliubadalo L, Tonda R, Gausachs M, Trotta JR, Castellanos E, Lopez-Doriga A, et al. Benchmarking of Whole Exome Sequencing and Ad Hoc Designed Panels for Genetic Testing of Hereditary Cancer. Sci Rep. 2017;7:37984.

140. Paluch-Shimon S, Cardoso F, Sessa C, Balmana J, Cardoso MJ, Gilbert F, et al. Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO Clinical Practice Guidelines for cancer prevention and screening. Ann Oncol. 2016;27(suppl 5):v103-v10.
141. Network. NCC. Clinical Practice Guidelines in Oncology: Genetic/Familial High-Risk Assessment: Breast and Ovarian 2017 [Available from: https://www.nccn.org/professionals/physician_gls/pdf/genetics_screening.pdf.
142. NHS England. National Genomic Testing Directory 2019 [Available from: https://www.england.nhs.uk/publication/national-genomic-test-directories/.
143. Taylor A, Brady AF, Frayling IM, Hanson H, Tischkowitz M, Turnbull C, et al. Consensus for genes to be included on cancer panel tests offered by UK genetics services: guidelines of the UK Cancer Genetics Group. J Med Genet. 2018;55(6):372-7.
144. Maxwell KN, Hart SN, Vijai J, Schrader KA, Slavin TP, Thomas T, et al. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am J Hum Genet. 2016;98(5):801-17.
145. Eccles DM, Mitchell G, Monteiro AN, Schmutzler R, Couch FJ, Spurdle AB, et al. BRCA1 and BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. Ann Oncol. 2015;26(10):2057-65.
146. Evaluation of Genomic Applications in P, Prevention Working G. The EGAPP initiative: lessons learned. Genet Med. 2014;16(3):217-24.
147. Hermel DJ, McKinnon WC, Wood ME, Greenblatt MS. Multi-gene panel testing for hereditary cancer susceptibility in a rural Familial Cancer Program. Fam Cancer. 2017;16(1):159-66.
148. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol. 2011;5(1):5-23.
149. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010;363(20):1938-48.
150. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
151. Wittersheim M, Buttner R, Markiefka B. Genotype/Phenotype correlations in patients with hereditary breast cancer. Breast Care (Basel). 2015;10(1):22-6.
152. Balleine RL, Murali R, Bilous AM, Farshid G, Waring P, Provan P, et al. Histopathological features of breast cancer in carriers of ATM gene variants. Histopathology. 2006;49(5):523-32.
153. Brekelmans CT, Tilanus-Linthorst MM, Seynaeve C, vd Ouweland A, Menke-Pluymers MB, Bartels CC, et al. Tumour characteristics, survival and prognostic factors of hereditary breast cancer from BRCA2-, BRCA1- and non-BRCA1/2 families as compared to sporadic breast cancer cases. Eur J Cancer. 2007;43(5):867-76.
154. Spurdle AB, Couch FJ, Parsons MT, McGuffog L, Barrowdale D, Bolla MK, et al. Refined histopathological predictors of BRCA1 and BRCA2 mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. Breast Cancer Res. 2014;16(6):3419.
155. Waddell N, Arnold J, Cocciardi S, da Silva L, Marsh A, Riley J, et al. Subtypes of familial breast tumours revealed by expression and copy number profiling. Breast Cancer Res Treat. 2010;123(3):661-77.
156. Loman N, Johannsson O, Bendahl P, Dahl N, Einbeigi Z, Gerdes A, et al. Prognosis and clinical presentation of BRCA2-associated breast cancer. Eur J Cancer. 2000;36(11):1365-73.
157. Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, Miron P, et al. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. J Clin Oncol. 2015;33(4):304-11.
158. Kilpivaara O, Bartkova J, Eerola H, Syrjakoski K, Vahteristo P, Lukas J, et al. Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients. Int J Cancer. 2005;113(4):575-80.
159. Wilson JR, Bateman AC, Hanson H, An Q, Evans G, Rahman N, et al. A novel HER2-positive breast cancer phenotype arising from germline TP53 mutations. J Med Genet. 2010;47(11):771-4.

160. Eccles DM, Li N, Handwerker R, Maishman T, Copson ER, Durcan LT, et al. Genetic testing in a cohort of young patients with HER2-amplified breast cancer. Ann Oncol. 2016;27(3):467-73.
161. Weigelt B, Bi R, Kumar R, Blecua P, Mandelker DL, Geyer FC, et al. The Landscape of Somatic Genetic Alterations in Breast Cancers From ATM Germline Mutation Carriers. J Natl Cancer Inst. 2018;110(9):1030-4.
162. Royer-Bertrand B, Rivolta C. Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer. Cell Mol Life Sci. 2015;72(8):1463-71.
163. Consortium TITP-CAoWG. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82-93.
164. Van Hoeck A, Tjoonk NH, van Boxtel R, Cuppen E. Portrait of a cancer: mutational signature analyses for cancer diagnostics. BMC Cancer. 2019;19(1):457.
165. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
166. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
167. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet. 2003;33 Suppl:238-44.
168. Galuppini F, Dal Pozzo CA, Deckert J, Loupakis F, Fassan M, Baffa R. Tumor mutation burden: from comprehensive mutational screening to the clinic. Cancer Cell Int. 2019;19:209.
169. Buttner R, Longshore JW, Lopez-Rios F, Merkelbach-Bruse S, Normanno N, Rouleau E, et al. Implementing TMB measurement in clinical practice: considerations on assay requirements. ESMO Open. 2019;4(1):e000442.
170. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive Analysis of Hypermutation in Human Cancer. Cell. 2017;171(5):1042-56 e10.
171. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-24.
172. Catalogue of Somatic Mutations in Cancer. The Cancer Gene Census 2019 [
173. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47-54.
174. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.
175. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer cell. 2017;32(2):169-84.e7.
176. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature. 2014;512(7513):155-60.
177. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013;3(1):246-59.
178. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
179. Turner NC. Signatures of DNA-Repair Deficiencies in Breast Cancer. The New England journal of medicine. 2017;377(25):2490-2.
180. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell. 2019;176(6):1282-94 e20.
181. Pitt JJ, Riester M, Zheng Y, Yoshimatsu TF, Sanni A, Oluwasola O, et al. Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features. Nature communications. 2018;9(1):4181.
182. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 2013;45(9):970-6.
183. Schmidt MK, Tollenaar RA, de Kemp SR, Broeks A, Cornelisse CJ, Smit VT, et al. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. J Clin Oncol. 2007;25(1):64-9.

184. Broeks A, Braaf LM, Huseinovic A, Schmidt MK, Russell NS, van Leeuwen FE, et al. The spectrum of ATM missense variants and their contribution to contralateral breast cancer. Breast Cancer Res Treat. 2008;107(2):243-8.
185. Fluidigm. Fluidigm Access Array 2016 [Available from: https://www.fluidigm.com/assays.
186. Illumina Inc. Illumina Next Generation Sequencing Platforms 2016 [Available from: http://www.illumina.com/systems/sequencing-platform-comparison.html
187. Burrows Wheeler Aligner. The Burrows Wheeler Aligner 2010 [Available from: http://bio-bwa.sourceforge.net/.
188. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
189. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
190. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987-93.
191. The Broad Institute. GATK Unified Genotyper 2016 [Available from: https://software.broadinstitute.org/gatk/events/slides/1307/GATKwh1-BP-5-Variant_calling.pdf.
192. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8.
193. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11 0 1-33.
194. ANNOVAR. ANNOVAR 2016 [Available from: http://annovar.openbioinformatics.org/en/latest/.
195. Practice Guidelines for the Evaluation of Pathogenicity and the Reporting of Sequence Variants in Clinical Molecular Genetics., (2013).
196. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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):405-24.
197. The Global Alliance for Genomics and Health. BRCA Exchange 2016 [Available from: http://brcaexchange.org/.
198. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat. 2008;29(11):1282-91.
199. GATC. Eurofins Genomics: Eurofins Genomics; 2020 [Available from: https://www.eurofinsgenomics.eu/en/custom-dna-sequencing/gatc-services/.
200. Genomics England. The 100,000 Genomes Project 2018 [Available from: https://www.genomicsengland.co.uk/the-100000-genomes-project/.
201. Genomics England. The 100,000 Genomes Project Protocol v4. 2017.
202. Genomics England. The 100,000 Genomes Project Rare Disease Conditions Eligibility Criteria v1.9.0. 2018.
203. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
204. National Cancer Institute. The Cancer Genome Atlas Programme 2019 [Available from: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.
205. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45(D1):D777-D83.
206. Bioconductor. GenVisR 2019 [Available from: https://bioconductor.org/packages/release/bioc/html/GenVisR.html.
207. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol. 2017;3(9):1190-6.

208. Southey MC, Teo ZL, Dowty JG, Odefrey FA, Park DJ, Tischkowitz M, et al. A PALB2 mutation associated with high risk of breast cancer. Breast Cancer Res. 2010;12(6):R109.
209. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2020:531210.
210. Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, et al. 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. N Engl J Med. 2017;377(19):1836-46.
211. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. Nat Genet. 2018;50(9):1212-8.
212. Li Y, Arellano AR, Bare LA, Bender RA, Strom CM, Devlin JJ. A Multigene Test Could Cost-Effectively Help Extend Life Expectancy for Women at Risk of Hereditary Breast Cancer. Value Health. 2017;20(4):547-55.
213. Manchanda R, Patel S, Gordeev VS, Antoniou AC, Smith S, Lee A, et al. Cost-effectiveness of Population-Based BRCA1, BRCA2, RAD51C, RAD51D, BRIP1, PALB2 Mutation Testing in Unselected General Population Women. J Natl Cancer Inst. 2018;110(7):714-25.
214. Spurdle A, Greville-Heygate S, Antoniou A, Brown M, Burke L, Hoya M, et al. Towards controlled terminology for reporting germline cancer susceptibility variants: An ENIGMA report. Journal of Medical Genetics. 2019;56.
215. Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, Lee A, Vreeswijk MPG, Hollestelle A, et al. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. J Med Genet. 2019;56(9):581-9.
216. Sawyer S, Mitchell G, McKinley J, Chenevix-Trench G, Beesley J, Chen XQ, et al. A role for common genomic variants in the assessment of familial breast cancer. J Clin Oncol. 2012;30(35):4330-6.
217. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581-90.
218. Lakhani SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Osin PP, McGuffog L, et al. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. J Clin Oncol. 2002;20(9):2310-8.
219. Mavaddat N, Barrowdale D, Andrulis IL, Domchek SM, Eccles D, Nevanlinna H, et al. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). Cancer Epidemiol Biomarkers Prev. 2012;21(1):134-47.
220. Bane AL, Beck JC, Bleiweiss I, Buys SS, Catalano E, Daly MB, et al. BRCA2 mutation-associated breast cancers exhibit a distinguishing phenotype based on morphology and molecular profiles from tissue microarrays. Am J Surg Pathol. 2007;31(1):121-8.
221. Krammer J, Pinker-Domenig K, Robson ME, Gonen M, Bernard-Davila B, Morris EA, et al. Breast cancer detection and tumor characteristics in BRCA1 and BRCA2 mutation carriers. Breast Cancer Res Treat. 2017;163(3):565-71.
222. Ha SM, Chae EY, Cha JH, Kim HH, Shin HJ, Choi WJ. Association of BRCA Mutation Types, Imaging Features, and Pathologic Findings in Patients With Breast Cancer With BRCA1 and BRCA2 Mutations. AJR Am J Roentgenol. 2017;209(4):920-8.
223. Yu JH, Lee JW, Son BH, Kim SW, Park SK, Lee MH, et al. Characteristics of BRCA1/2 Mutation-Positive Breast Cancers in Korea: A Comparison Study Based on Multicenter Data and the Korean Breast Cancer Registry. J Breast Cancer. 2014;17(2):129-35.
224. Greville-Heygate SL, Maishman T, Tapper WJ, Cutress RI, Copson E, Dunning AM, et al. Pathogenic Variants in CHEK2 Are Associated With an Adverse Prognosis in Symptomatic Early-Onset Breast Cancer. JCO Precision Oncology. 2020(4):472-85.
225. Chen Y, Liu L, Zhou Q, Imam MU, Cai J, Wang Y, et al. Body mass index had different effects on premenopausal and postmenopausal breast cancer risks: a dose-response meta-analysis with 3,318,796 subjects from 31 cohort studies. BMC Public Health. 2017;17(1):936.

226. Copson ER, Cutress RI, Maishman T, Eccles BK, Gerty S, Stanton L, et al. Obesity and the outcome of young breast cancer patients in the UK: the POSH study. Annals of Oncology. 2015;26(1):101-12.
227. Eccles BK, Copson ER, Cutress RI, Maishman T, Altman DG, Simmonds P, et al. Family history and outcome of young patients with breast cancer in the UK (POSH study). Br J Surg. 2015;102(8):924-35.
228. de Bock GH, Schutte M, Krol-Warmerdam EM, Seynaeve C, Blom J, Brekelmans CT, et al. Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant. J Med Genet. 2004;41(10):731-5.
229. The National Institution for Health and Care Excellence. Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. 2013.
230. Lee JEA, Li N, Rowley SM, Cheasley D, Zethoven M, McInerny S, et al. Molecular analysis of PALB2-associated breast cancers. J Pathol. 2018;245(1):53-60.
231. Yang Z, Ouyang T, Li J, Wang T, Fan Z, Fan T, et al. Prevalence and characterization of ATM germline mutations in Chinese BRCA1/2-negative breast cancer patients. Breast Cancer Res Treat. 2019;174(3):639-47.
232. Bernstein JL, Concannon P. ATM, radiation, and the risk of second primary breast cancer. International Journal of Radiation Biology. 2017;93(10):1121-7.
233. Bakhuizen JJ, Hogervorst FB, Velthuizen ME, Ruijs MW, van Engelen K, van Os TA, et al. TP53 germline mutation testing in early-onset breast cancer: findings from a nationwide cohort. Fam Cancer. 2019;18(2):273-80.
234. Packwood K, Martland G, Sommerlad M, Shaw E, Moutasim K, Thomas G, et al. Breast cancer in patients with germline TP53 pathogenic variants have typical tumour characteristics: the Cohort study of TP53 carrier early onset breast cancer (COPE study). J Pathol Clin Res. 2019;5(3):189-98.
235. Petridis C, Arora I, Shah V, Megalios A, Moss C, Mera A, et al. Frequency of pathogenic germline variants in BRCA1, BRCA2, PALB2, CHEK2 and TP53 in ductal carcinoma in situ diagnosed in women under the age of 50 years. Breast Cancer Research. 2019;21(1):58.
236. Rakha EA, Pinder SE, Bartlett JM, Ibrahim M, Starczynski J, Carder PJ, et al. Updated UK Recommendations for HER2 assessment in breast cancer. J Clin Pathol. 2015;68(2):93-9.
237. Chen X, Yuan Y, Gu Z, Shen K. Accuracy of estrogen receptor, progesterone receptor, and HER2 status between core needle and open excision biopsy in breast cancer: a meta-analysis. Breast Cancer Res Treat. 2012;134(3):957-67.
238. Masciari S, Dillon DA, Rath M, Robson M, Weitzel JN, Balmana J, et al. Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort. Breast Cancer Res Treat. 2012;133(3):1125-30.
239. Evans DG, Harkness EF, Plaskocinska I, Wallace AJ, Clancy T, Woodward ER, et al. Pathology update to the Manchester Scoring System based on testing in over 4000 families. J Med Genet. 2017;54(10):674-81.
240. Hyder Z, Harkness EF, Woodward ER, Bowers NL, Pereira M, Wallace AJ, et al. Risk of Contralateral Breast Cancer in Women with and without Pathogenic Variants in BRCA1, BRCA2, and TP53 Genes in Women with Very Early-Onset (<36 Years) Breast Cancer. Cancers (Basel). 2020;12(2):378.
241. Meric-Bernstam F, Zheng X, Shariati M, Damodaran S, Wathoo C, Brusco L, et al. Survival Outcomes by TP53 Mutation Status in Metastatic Breast Cancer. JCO Precis Oncol. 2018;2018.
242. Olivier M, Langerod A, Carrieri P, Bergh J, Klaar S, Eyfjord J, et al. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. Clin Cancer Res. 2006;12(4):1157-67.
243. Eikesdal HP, Knappskog S, Aas T, Lonning PE. TP53 status predicts long-term survival in locally advanced breast cancer after primary chemotherapy. Acta Oncol. 2014;53(10):1347-55.
244. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010;466(7308):869-73.

245. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. 2017;9(1):34.
246. Thomas A, Routh ED, Pullikuth A, Jin G, Su J, Chou JW, et al. Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. Oncoimmunology. 2018;7(10):e1490854.
247. Mei P, Freitag CE, Wei L, Zhang Y, Parwani AV, Li Z. High tumor mutation burden is associated with DNA damage repair gene mutation in breast carcinomas. Diagn Pathol. 2020;15(1):50.
248. Lal A, Ramazzotti D, Weng Z, Liu K, Ford JM, Sidow A. Comprehensive genomic characterization of breast tumors with BRCA1 and BRCA2 mutations. BMC Med Genomics. 2019;12(1):84.
249. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23(6):703-13.
250. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz H-J, Morse MA, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. The Lancet Oncology. 2017;18(9):1182-91.
251. Heydt C, Rehker J, Pappesch R, Buhl T, Ball M, Siebolts U, et al. Analysis of tumor mutational burden: correlation of five large gene panels with whole exome sequencing. Sci Rep. 2020;10(1):11387.
252. Zou Y, Zou X, Zheng S, Tang H, Zhang L, Liu P, et al. Efficacy and predictive factors of immune checkpoint inhibitors in metastatic breast cancer: a systematic review and meta-analysis. Ther Adv Med Oncol. 2020;12:1758835920940928.
253. Matsuo K, Spragg SE, Ciccone MA, Blake EA, Ricker C, Pham HQ, et al. Nivolumab use for BRCA gene mutation carriers with recurrent epithelial ovarian cancer: A case series. Gynecol Oncol Rep. 2018;25:98-101.
254. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017;49(10):1476-86.
255. Gonzalez-Perez A, Sabarinathan R, Lopez-Bigas N. Local Determinants of the Mutational Landscape of the Human Genome. Cell. 2019;177(1):101-14.
256. Wellcome Trust Sanger Institute. COSMIC Catalogue of Somatic Mutations in Cancer 2020 [Available from: https://cancer.sanger.ac.uk/cosmic.
257. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23(4):517-25.
258. Garrett A, Callaway A, Durkie M, Cubuk C, Alikian M, Burghel GJ, et al. Cancer Variant Interpretation Group UK (CanVIG-UK): an exemplar national subspecialty multidisciplinary network. J Med Genet. 2020.
259. McCrorie AD, Ashfield S, Begley A, McIlmunn C, Morrison PJ, Boyd C, et al. Multifocal breast cancers are more prevalent in BRCA2 versus BRCA1 mutation carriers. J Pathol Clin Res. 2020;6(2):146-53.
260. Palma MD, Domchek SM, Stopfer J, Erlichman J, Siegfried JD, Tigges-Cardwell J, et al. The relative contribution of point mutations and genomic rearrangements in BRCA1 and BRCA2 in high-risk breast cancer families. Cancer research. 2008;68(17):7006-14.
261. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47(12):1402-7.
262. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012;149(5):979-93.
263. Boot A, Huang MN, Ng AWT, Ho SC, Lim JQ, Kawakami Y, et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. Genome research. 2018;28(5):654-65.

264. Saini N, Roberts SA, Klimczak LJ, Chan K, Grimm SA, Dai S, et al. The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. PLOS Genetics. 2016;12(10):e1006385.
265. Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, et al. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. EBioMedicine. 2017;20:39-49.
266. Pilati C, Shinde J, Alexandrov LB, Assié G, André T, Hélias-Rodzewicz Z, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J Pathol. 2017;242(1):10-5.
267. Mimaki S, Totsuka Y, Suzuki Y, Nakai C, Goto M, Kojima M, et al. Hypermutation and unique mutational signatures of occupational cholangiocarcinoma in printing workers exposed to haloalkanes. Carcinogenesis. 2016;37(8):817-26.
268. Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science. 2017;358(6360):234-8.