

# Circulating tumour DNA analysis in Advanced Breast Cancer

Belinda Kingston

Breast Cancer Now Centre  
The Institute of Cancer Research  
University of London

Submitted for the degree of PhD

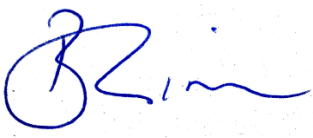
March 2022

Word count: 47,262

## Declaration of originality

I declare that the work presented in this thesis is my own.

Any work that is not mine has been referenced and any data originating from collaborative work has clearly been identified.

A handwritten signature in blue ink, appearing to read 'Belinda Kingston', is positioned above the printed name.

Belinda Kingston

## Abstract

Circulating tumour DNA (ctDNA) analysis shows great potential both as an approach to understand the evolution advanced breast cancer (ABC) and as a clinical tool to assist in ABC management. CtDNA analysis has not yet entered routine clinical practice, and important questions remain around the validity and applications of the approach.

The objectives of this work were to:

- validate the approach of ctDNA analysis
- establish the landscape of ABC according to ctDNA analysis
- utilise ctDNA analysis to identify biomarkers of response and resistance to therapy

CtDNA isolated from plasma samples from the plasmaMATCH trial underwent digital PCR and targeted sequencing. DNA extracted from paired tissues underwent targeted sequencing. A putative resistance mechanism was further investigated using transient transfection.

Two orthogonal ctDNA analysis techniques demonstrated a high level of agreement for targetable mutation status (kappa values 0.89 to 0.93). Sensitivity of ctDNA-assessed gene mutation status was high compared to the gold-standard tissue-based sequencing (88.2% to 96.8%), with specificity lower (40.0% to 98.5%), likely secondary to biological factors such as tumour heterogeneity and temporal mutation acquisition.

The landscape and clonal architecture of ABC was defined through analysis of a large cohort of patients (n=800). Significant and novel findings were identified in genes mutation patterns including dual *PIK3CA* mutagenesis, MAPK pathway and *ESR1* co-mutation, and APOBEC mutagenesis, which both enhances our understanding of the evolution of ABC and has prognostic and predictive relevance for patients on subsequent targeted therapy.

CtDNA analysis also revealed a novel putative resistance mechanism to fulvestrant, *ESR1* p.F404. Transient transfection data supported the role of this mutation in fulvestrant resistance, and suggested alternative treatment strategies in patients with this mutation which may prolong survival.

These findings enhance our understanding of ABC, and support the use of ctDNA analysis as a research and clinical tool.

## Publications

**Kingston, B.**, R. Cutts, M. Beaney, G. Walsh-Crestani, S. Hrebien, L. Kilburn, S. Kernaghan, L. Moretti, K. Wilkinson, I. Macpherson, R. Baird, R. Roylance, J. S. Reis-Filho, M. Hubank, I. Faull, K. Banks, I. Garcia-Murillas, J. Bliss, A. Ring and N. Turner (2021). "**Analysis of ctDNA in advanced breast cancer reveals polyclonal disease associated with adverse outcome.**" *Annals of Oncology* 32: S65-S66.

**Kingston, B.**, R. J. Cutts, H. Bye, M. Beaney, G. Walsh-Crestani, S. Hrebien, C. Swift, L. S. Kilburn, S. Kernaghan, L. Moretti, K. Wilkinson, A. M. Wardley, I. R. Macpherson, R. D. Baird, R. Roylance, J. S. Reis-Filho, M. Hubank, I. Faull, K. C. Banks, R. B. Lanman, I. Garcia-Murillas, J. M. Bliss, A. Ring and N. C. Turner (2021). "**Genomic profile of advanced breast cancer in circulating tumour DNA.**" *Nature Communications* 12(1): 2423.

Pascual, J., R. J. Cutts, **B. Kingston**, S. Hrebien, L. S. Kilburn, S. Kernaghan, L. Moretti, K. Wilkinson, A. M. Wardley, I. R. Macpherson, R. D. Baird, R. Roylance, M. Hubank, G. Walsh, I. Faull, K. C. Banks, R. B. Lanman, I. Garcia-Murillas, J. M. Bliss, A. Ring and N. C. Turner (2021). "**Abstract PS5-02: Assessment of early ctDNA dynamics to predict efficacy of targeted therapies in metastatic breast cancer: Results from plasmaMATCH trial.**" *Cancer Research*(81)(4 Supplement): PS5-02-PS05-02.

Turner, N. C., **B. Kingston**, L. S. Kilburn, S. Kernaghan, A. M. Wardley, I. R. Macpherson, R. D. Baird, R. Roylance, P. Stephens, O. Oikonomidou, J. P. Braybrooke, M. Tuthill, J. Abraham, M. C. Winter, H. Bye, M. Hubank, H. Gevensleben, R. Cutts, C. Snowdon, D. Rea, D. Cameron, A. Shaaban, K. Randle, S. Martin, K. Wilkinson, L. Moretti, J. M. Bliss and A. Ring. "**Circulating tumour DNA analysis to direct therapy in advanced breast cancer (plasmaMATCH): a multicentre, multicohort, phase 2a, platform trial.**" *The Lancet Oncology*(21): 1296-308.

Turner, N., **B. Kingston**, L. Kilburn, S. Kernaghan, A. M. Wardley, I. Macpherson, R. D. Baird, R. Roylance, P. Stephens, O. Oikonomidou, J. P. Braybrooke, M. Tuthill, J. Abraham, M. C. Winter, H. Bye, M. Hubank, C. Snowdon, D. Rea, D. Cameron, A. Shaaban, K. Randle, K. Wilkinson, L. Moretti, J. M. Bliss and A. Ring (2020). "**Abstract GS3-06: Results from the plasmaMATCH trial: A multiple parallel cohort, multi-centre clinical trial of circulating tumour DNA testing to direct targeted therapies in patients with advanced breast cancer (CRUK/15/010).**" *Cancer Research* 80(4 Supplement): GS3-06-GS03-06.

Macpherson, I., L. Kilburn, S. Kernaghan, A. Wardley, R. Baird, R. Roylance, P. Stephens, O. Oikonomidou, J. Braybrooke, M. Tuthill, J. Abraham, M. Winter, **B. Kingston**, K. Wilkinson, A. Ring, J. Bliss and N. Turner (2020). "**Abstract P1-19-04: Results from plasmaMATCH trial treatment cohort A: A phase II trial of extended-dose fulvestrant in patients with an ESR1 mutation identified via ctDNA screening (CRUK/15/010)**". *Cancer Research* 80(4 Supplement): P1-19-04-P1-19-04.

Wardley, A., L. Kilburn, S. Kernaghan, I. Macpherson, R. Baird, R. Roylance, P. Stephens, O. Oikonomidou, J. Braybrooke, M. Tuthill, J. Abraham, M. Winter, **B. Kingston**, K. Wilkinson, J. Bliss, A. Ring and N. Turner (2020). "**Abstract P1-19-07: Results from plasmaMATCH trial treatment cohort B: A phase II trial of**

**neratinib plus fulvestrant in ER positive breast cancer or neratinib alone in ER negative breast cancer in patients with a HER2 mutation identified via ctDNA screening (CRUK/15/010).**" Cancer Research 80(4 Supplement): P1-19-07-P1-19-07

Roylance, R., L. Kilburn, S. Kernaghan, A. Wardley, I. Macpherson, R. Baird, P. Stephens, O. Oikonomidou, J. Braybrooke, M. Tuthill, J. Abraham, M. Winter, **B. Kingston**, K. Wilkinson, A. Ring, J. Bliss and N. Turner (2020). "Abstract P1-19-11: Results from plasmaMATCH trial treatment cohort C: A phase II trial of capivasertib plus fulvestrant in ER positive breast cancer patients with an AKT1 mutation identified via ctDNA screening (CRUK/15/010)". Cancer Research 80(4 Supplement): P1-19-11-P1-19-11

Baird, R., L. Kilburn, S. Kernaghan, A. Wardley, I. Macpherson, R. Roylance, P. Stephens, O. Oikonomidou, J. Braybrooke, M. Tuthill, J. Abraham, M. Winter, **B. Kingston**, K. Wilkinson, N. Turner, A. Ring and J. Bliss (2020). "Abstract P1-19-14: Results from plasmaMATCH trial treatment cohort D: A phase II trial of capivasertib in patients with an AKT activation basket mutation identified via ctDNA testing or tumour sequencing (CRUK/15/010)." Cancer Research 80(4 Supplement): P1-19-14-P1-19-14

**Kingston, B.**, C. Bailleux, S. Delaloge, G. Schiavon, V. Scott, M. Lacroix-Triki, T. H. Carr, I. Kozarewa, H. Gevensleben, Z. Kemp, A. Pearson, N. Turner and F. André (2019). "Exceptional Response to AKT Inhibition in Patients With Breast Cancer and Germline PTEN Mutations." JCO Precision Oncology(3): 1-7.

Pearson, A., P. Z. Proszek, J. Pascual, C. Fribbens, M. K. Shamsheer, **B. Kingston**, B. O'Leary, M. T. Herrera-Abreu, R. J. Cutts, I. Garcia-Murillas, H. Bye, B. A. Walker, D. Gonzalez De Castro, L. Yuan, S. Jamal, M. Hubank, E. Lopez-Knowles, E. F. Schuster, M. Dowsett, P. Osin, A. Nerurkar, M. Parton, A. F. Okines, S. R. D. Johnston, A. Ring and N. C. Turner (2019). "Inactivating NF1 mutations are enriched in advanced breast cancer and contribute to endocrine therapy resistance." Clinical Cancer Research(26): 608-22.

## Acknowledgements

### **Nick Turner – Group Leader, Molecular Oncology Group, The Institute of Cancer Research**

Nick conceived the outline of this project and was the overall supervisor. He was the chief investigator of plasmaMATCH, and was instrumental in conceiving and setting up the trial and securing access to the samples for research use. Nick also set up the collaboration with Guardant Health to enable targeted sequencing to be undertaken on trial samples. During the project, Nick provided guidance and supervision, critiqued data analyses and provided advice on the final thesis write-up.

### **Isaac Garcia-Murillas – Staff Scientist, Molecular Oncology Group, The Institute of Cancer Research**

Isaac provided supervision and training in the laboratory, particularly in the arenas of DNA extraction and sequencing. Isaac has previously developed and formalised many of the laboratory protocols which were instrumental in the work undertaken in this project. Isaac also advised on many aspects of the project, assisted with troubleshooting and helped ensure the smooth running of the project.

### **Alex Pearson – Senior Scientific Officer, Molecular Oncology Group, The Institute of Cancer Research**

Alex provided supervision and training in the laboratory, particularly in the arena of tissue culture and transfection. Alex was instrumental in the work undertaken to investigate *ESRI* p.F404, advising on the experimental strategy of transient transfection, western blot techniques, and data analysis. Alex also performed many of the beta-galactosidase control assays when it was contraindicated for me to do so due to pregnancy.

### **Ros Cutts – Senior Bioinformatician, Molecular Oncology Group, The Institute of Cancer Research**

Ros was central in the bioinformatic processing of the sequencing data. Ros performed the alignment, copy number calling and variant calling for any raw sequencing data. She also helped adjust the in-house custom pipeline for tissue sequencing data, improving the accuracy of variant calling and reducing error. Ros also developed a custom annotation pipeline for the

sequencing data generated by Guardant Health that both reduced the number of germline calls, and enabled annotation of pathogenicity and targetability of alterations.

**Hannah Bye - Higher Scientific Officer, Centre for Molecular Pathology, The Institute of Cancer Research**

Hannah undertook all of the screening ddPCR assays for the plasmaMATCH trial. Hannah then analysed the results for a binary positive or negative result for each assay, which (after confirmation by a second assay and secondary reviewer) was fed back to the trial team to inform patient eligibility for trial enrolment. These assays results were transferred to the ICR for further analysis of exact mutation and allele frequency, which was work undertaken within this project.

**Sarah Hrebien – Higher Scientific Officer, Molecular Oncology Group, The Institute of Cancer Research**

Sarah assisted and guided in the construction of low-pass whole genome sequencing libraries, for which she had previously optimised a protocol. Sarah additionally was instrumental in sample management for plasmaMATCH, ensuring accurate recording and tracking of samples.

**Giselle Walsh – Scientific Officer, Molecular Oncology Group, The Institute of Cancer Research**

Giselle helped with plasma and DNA extraction, alongside helping with the aliquoting of plasma to be shipped to Guardant Health. Giselle also helped with sample retrieval and return from the -80° storage.

**Javier Pascual – Clinical Fellow, Molecular Oncology Group, The Institute of Cancer Research**

Javier helped with the final shipment of samples to Guardant Health when I was unable to do due to being on maternity leave.

**Adam Mills - Bioinformatician, Breast Cancer Now, The Institute of Cancer Research**

Adam helped run raw sequencing data from tissue sequencing through the in-house bioinformatic pipeline Veritas.

**Molecular Oncology Group, The Institute of Cancer Research**

Data derived from within this project was regularly discussed with team members within the Molecular Oncology team, from whom invaluable advice and expertise was gained. The supportive and friendly atmosphere created by the team helped make the fellowship an enjoyable experience.

## Funding

This work was funded by Cancer Research UK, who funded the fellowship (Grant number CRM120X), alongside Breast Cancer Now who fund the division of Breast Cancer Research within the Institute of Cancer Research.

The targeted plasma DNA sequencing was funded by Guardant Health.

plasmaMATCH was co-sponsored by the Institute of Cancer Research and the Royal Marsden National Health Service Foundation Trust. The trial is registered NCT03182634 and ISRCTN:16945804.

ICR-CTSU is supported by the Cancer Research UK core grant (C1491/A25351).

plasmaMATCH was supported by the National Institute for Health research (NIHR) Manchester Clinical Research Facility at The Christie Hospital, Manchester UK, and by the Cancer Research UK Cambridge Centre, the Cambridge NIHR Biomedical Research Centre and the Cambridge Experimental Cancer Medicine Centre, Cambridge UK, and the NIHR Biomedical Research Centre at University College London Hospital (UCLH), London UK. plasmaMATCH was supported at participating sites in England by the NIHR Clinical Research Network, in Scotland by the Chief Scientist Office, and in Wales by Health and Care Research Wales.



## Abbreviations

AutoDG – Automated droplet generator

$\beta$ -gal – beta-galactosidase

cfDNA – cell-free DNA

ctDNA – circulating tumour DNA

ddPCR – droplet digital PCR

DNA – deoxyribonucleic acid

dsDNA – double stranded DNA

ER+ – oestrogen receptor positive

ER- – oestrogen receptor negative

ERC – (o)estrogen receptor construct

FAM – 6-carboxyfluorescein

HEX – 5' hexachloro-fluorescein

HER2+ – HER2 over-expression and/or gene amplification

HER2- – lack of HER2 over-expression and/or gene amplification

HR+ – hormone receptor positive

HR- – hormone receptor negative

lpWGS – low pass whole genome sequencing

MAPK – mitogen-activated protein kinase

maxVAF – maximum variant allele frequency

mTOR – mammalian target of rapamycin

NGS – next generation sequencing

NTC – no template control

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PI3K - phosphatidylinositol-4,5-bisphosphate 3-kinase

qPCR – quantitative PCR

SERD – selective oestrogen receptor degrader

SNP – single nucleotide polymorphism

SNV – single nucleotide variant

VIC – VIC dye for probes

## Contents

Abstract.....	3
Publications.....	4
Acknowledgements.....	6
Funding .....	8
Abbreviations.....	9
1. Chapter 1. Introduction .....	23
1.1. Advanced Breast Cancer .....	23
1.2. Personalised therapy.....	23
1.3. Circulating tumour DNA as a clinical tool.....	24
1.3.1. ctDNA analysis an alternative to tissue-based sequencing.....	26
1.3.2. ctDNA analysis for the identification of targetable alterations .....	26
1.3.3. As a predictive and prognostic biomarker .....	27
1.3.4. To interrogate the advanced breast cancer genomic landscape .....	30
1.3.5. To identify mechanisms of resistance.....	31
1.3.6. Potential limitations and questions around the application of ctDNA analysis.....	31
1.4. plasmaMATCH.....	36
1.4.1. Study rationale .....	36
1.4.2. Study design.....	37
1.4.3. Use of plasmaMATCH samples within the fellowship .....	39
1.4.4. Cohort A: Patients with <i>ESR1</i> alterations for treatment with extended dose fulvestrant .....	40
1.4.5. Cohort B: Patients with <i>ERBB2</i> mutations for treatment with neratinib +/- fulvestrant .....	43
1.4.6. Cohort C: HR+ patients with <i>AKT1</i> mutation for treatment with capivasertib +/- fulvestrant .....	47

1.4.7.	Cohort D: Patients with AKT pathway activating mutations for treatment with capivasertib .....	50
1.5.	Summary .....	51
2.	Chapter 2. Methods and Materials .....	53
2.1.	Methods and materials within the plasmaMATCH study .....	53
2.1.1.	Study design.....	53
2.1.2.	Sample collection.....	55
2.1.3.	Plasma DNA extraction .....	56
2.1.4.	Tissue DNA extraction .....	58
2.1.5.	Quantification of DNA and PCR products .....	60
2.1.6.	Droplet digital PCR.....	62
2.1.7.	Tissue DNA library preparation and sequencing.....	65
2.1.8.	Tissue library assessment with an Agilent Bioanalyzer 2100 .....	69
2.1.9.	Tissue DNA library quantification.....	69
2.1.10.	Tissue sequencing on an Illumina MiniSeq .....	69
2.1.11.	Tissue bioinformatic pipeline and error correction .....	70
2.1.12.	Guardant360 Targeted Sequencing .....	71
2.2.	Low-pass Whole Genome Sequencing .....	78
2.2.1.	Library preparation and quantification .....	78
2.2.2.	Library sequencing.....	80
2.2.3.	Bioinformatic pipeline .....	80
2.3.	MCF-7 transient transfection and <i>ESR1</i> p.F404 investigation.....	80
2.3.1.	MCF-7 Culture.....	80
2.3.2.	Plasmid design, procurement and verification.....	80
2.3.3.	Plasmid Maxi-prep.....	84
2.3.4.	Estrogen Reporter Gene transfection protocol.....	86
2.3.5.	Drug Concentration assessment using cell viability assays.....	88

2.3.1.	Western Blot .....	88
2.4.	Statistical analysis .....	90
3.	Chapter 3. Validation of ctDNA analysis .....	91
3.1.	Introduction .....	91
3.2.	Hypothesis .....	92
3.3.	Aims .....	92
3.4.	Results .....	92
3.4.1.	Plasma droplet digital PCR compared to targeted sequencing of ctDNA .....	92
3.4.2.	Plasma ddPCR compared to paired tissue sequencing .....	97
3.5.	Discussion .....	105
3.6.	Conclusion.....	107
4.	Chapter 4. Clinico-pathological associations of ctDNA .....	108
4.1.	Introduction .....	108
4.2.	Hypothesis.....	108
4.3.	Aims .....	108
4.4.	Results .....	108
4.4.1.	Clinico-pathological associations of a negative ctDNA test.....	109
4.4.2.	Clinico-pathological associations of number of genome alterations and ctDNA purity	110
4.4.3.	Clinico-pathological associations of targetable mutations .....	111
4.4.4.	Organotropism of common breast cancer mutations .....	112
4.4.5.	ctDNA analysis of <i>ERBB2</i> copy number alteration and HER2 amplification.	113
4.5.	Discussion .....	114
4.6.	Conclusion.....	117
5.	Chapter 5. Breast cancer genomic landscape, clonal architecture, resistance mechanisms and mutational processes according to ctDNA analysis .....	118
5.1.	Introduction .....	118

5.2.	Hypotheses .....	120
5.3.	Aims .....	120
5.4.	Results .....	120
5.4.1.	Advanced breast cancer landscape.....	120
5.4.2.	Incidence and phenotype enrichment of SNVs and indels .....	121
5.4.3.	Clonal architecture in ctDNA .....	126
5.4.4.	Use of ctDNA to identify polyclonal genomic features.....	129
5.5.	Discussion .....	134
5.6.	Conclusion.....	137
6.	Chapter 6. CtDNA as a predictive and prognostic biomarker .....	139
6.1.	Introduction .....	139
6.2.	Hypothesis.....	140
6.3.	Aims .....	140
6.4.	Results .....	140
6.4.1.	ctDNA purity as a predictive and prognostic marker .....	140
6.4.2.	ctDNA as a biomarker of response to targeted therapy .....	145
6.5.	Discussion .....	158
6.6.	Conclusion.....	163
7.	Chapter 7. Use of ctDNA to investigate targeted therapy resistance mechanisms.....	164
7.1.	Introduction .....	164
7.2.	Hypothesis.....	165
7.3.	Aims .....	165
7.4.	Results .....	165
7.4.1.	Cohort A: extended dose fulvestrant.....	165
7.4.2.	Investigation of <i>ESR1</i> p.F404 as a potential resistance mechanism .....	168
7.4.3.	Cohort B: neratinib +/- fulvestrant.....	176
7.4.4.	Cohorts C and D: capivasertib +/- fulvestrant .....	180

7.5. Discussion .....	185
7.5.1. Putative resistance mechanisms to fulvestrant.....	185
7.5.2. Putative resistance mechanisms to neratinib.....	188
7.5.3. Putative resistance mechanisms to capivasertib .....	190
7.6. Conclusion.....	191
8. Chapter 8. Conclusion and Future Directions.....	193
References.....	201
Thanks.....	217

## List of Figures

### Chapter 1

Figure 1.1. Schematic representation of the systemic release, isolation and analysis of circulating tumour DNA (ctDNA) in a patient with advanced breast cancer. ....	25
Figure 1.2. The plasmaMATCH trial schema. ....	38
Figure 1.3. ctDNA testing approaches within plasmaMATCH. ....	39
Figure 1.4. The EGFR/HER receptor signalling and downstream pathway activation. ....	44
Figure 1.5. The PI3K/AKT signalling pathway. ....	48

### Chapter 2

Figure 2.1. ctDNA analysis approaches and trial schema of plasmaMATCH. ....	54
Figure 2.2. Volumes of Buffer ACL and carrier RNA required for processing 1ml, 2ml, 3ml, 4ml or 5ml plasma volume. ....	57
Figure 2.3. Reagent volumes required for extracting 1-5ml plasma. ....	57
Figure 2.4. OncoKB criteria needed to be met for assignment of oncogenicity for Variants of Possible Significance (VPS). ....	74
Figure 2.5. OncoKB levels of evidence required for assignment of targetability for oncogenic variants. ....	75
Figure 2.6. Association of allele frequency with gene copy number within 800 patients sequenced within plasmaMATCH. ....	77
Figure 2.7. Vector map of pcDNA3.1+C-DYK. ....	81
Figure 2.8. Estrogen Receptor construct sequencing confirmation of the insert. ....	82
Figure 2.9. Sanger sequencing result from the right primer for the ERE-Luc plasmid. ....	84
Figure 2.10. Six-well plate layout for ERC transfection. ....	86

### Chapter 3

Figure 3.1. Incidence and allele frequency of targetable mutations in plasmaMATCH. ....	93
Figure 3.2. Agreement between droplet digital PCR (ddPCR) and targeted sequencing (Targeted panel, TP) for targetable mutation status in 784 patients within plasmaMATCH. ....	94
Figure 3.3. Individual mutation agreement for actionable alterations. ....	95



Figure 3.4. DdPCR validation of <i>PIK3CA</i> and <i>ERBB2</i> mutations. ....	96
Figure 3.5. Agreement in sequencing calls made between the full and miniaturised (‘reduced’) volume libraries in test 1.....	98
Figure 3.6. Agreement in sequencing calls made between the full and miniaturised (‘reduced’) volume libraries in test 2.....	99
Figure 3.7. Pileup analysis of mutation calls across the 4 amplicons (1 – 4) tiling <i>ESR1</i> from primary triple negative breast cancer samples. ....	102
Figure 3.8. Base-specific mean error rate for the plasmaMATCH sequencing batch (A) and the TNBC sequencing batch (B).....	103
Figure 3.9. Analysis of mutation calls made by MuTect, VarDict and the agreement with each other and the pileup analysis for the sequencing run analysed.....	104
Figure 3.10. Agreement between ddPCR ctDNA testing and advanced disease tissue sequencing.....	105

## Chapter 4

Figure 4.1. Clinical and pathological associations of breast cancer mutation profile. ....	110
Figure 4.2. Association of clinical and pathological features with pathogenic alterations in the four targetable genes in plasmaMATCH: <i>PIK3CA</i> , <i>ESR1</i> , <i>HER2 (ERBB2)</i> and <i>AKT1</i> . ....	111
Figure 4.3. <i>ERBB2</i> mutation incidence in patients with HER2+ breast cancer, by line of therapy.....	112
Figure 4.4. Organotropism of mutations in ABC. ....	113
Figure 4.5. Adjusted <i>ERBB2</i> copy number (CN) in targeted sequencing.....	114

## Chapter 5

Figure 5.1. Mutational landscape of ABC defined by targeted sequencing .....	123
Figure 5.2. Incidence of gene mutations by breast cancer subtype .....	123
Figure 5.3. Copy number increase incidence by breast cancer subtype .....	124
Figure 5.4. Comparison of mutation incidence identified in the plasmaMATCH cohort via ctDNA sequencing, and primary breast cancer (TGCA) via tissue sequencing, by breast cancer subtype. ....	125

Figure 5.5. Comparison of mutation incidence identified in the plasmaMATCH cohort via ctDNA sequencing, and MSKCC via tissue sequencing, by breast cancer subtype.....	126
Figure 5.6. Cancer fractions of mutations in the most frequently altered genes .....	127
Figure 5.7. Cancer fractions of individual pathogenic hotspot mutations .....	128
Figure 5.8. Proportion of mutations that occur as a single versus multiple mutations per patient .....	128
Figure 5.9. Gene association analysis for the most frequent mutated genes .....	129
Figure 5.10. Gene association analysis for most frequent mutated genes in HR+ HER2- breast cancer .....	130
Figure 5.11. Incidence of MAPK pathway alterations in <i>ESR1</i> mutant and wild type populations.....	131
Figure 5.12. Overall survival (OS) in patients with HR+ HER2- disease who entered a treatment cohort in plasmaMATCH divided by combined <i>ESR1</i> and MAPK pathway mutation status. ....	132
Figure 5.13. <i>PIK3CA</i> mutagenesis in HR+ breast cancer.....	133
Figure 5.14. Proportion of <i>PIK3CA</i> mutations that occur at APOBEC consensus sites .....	133
Figure 5.15. Progression free survival (PFS) in patients on fulvestrant in treatment cohort A in plasmaMATCH by <i>PIK3CA</i> mutation status.....	134

## Chapter 6

Figure 6.1. Correlation of the low pass WGS purity estimate and targeted sequencing maximum VAF.....	142
Figure 6.2. Progression free survival in patients in cohort A on fulvestrant according to their adjusted <i>ESR1</i> clonality .....	143
Figure 6.3. Progression free survival in cohort A patients (n=67) according to ctDNA purity .....	144
Figure 6.4. Overall survival in cohort A patients (n=67) according to ctDNA purity.....	144
Figure 6.5. Waterfall plot for cohort A annotated with genomic sequencing availability, baseline and EOT genomic data. ....	147
Figure 6.6. Allele frequency and clonality of <i>ESR1</i> mutations within cohort A .....	147
Figure 6.7. Targetable <i>ESR1</i> alterations and response within cohort A .....	148
Figure 6.8. Baseline genomic status and response in cohort A .....	150

Figure 6.9. Progression free survival according to baseline <i>PIK3CA</i> and <i>TP53</i> mutation status .....	151
Figure 6.10. Progression free survival for patients enrolled into cohort B .....	152
Figure 6.11. Progression free survival in cohort B according to baseline <i>ERBB2</i> mutation characteristics.....	153
Figure 6.12. Progression free survival in cohort B according to baseline <i>PIK3CA</i> and <i>TP53</i> mutation status .....	153
Figure 6.13. Progression free survival in patients divided by baseline mutations in the TORC1 and MAPK pathways .....	154
Figure 6.14. Progression free survival for patients enrolled into cohorts C and D.....	155
Figure 6.15. Progression free survival in patients enrolled into cohorts C and D with an <i>AKT1</i> mutation, divided by clonal dominance of the <i>AKT1</i> mutation .....	156
Figure 6.16. Baseline incidence of gene alterations in patients enrolled based on an <i>AKT1</i> mutation .....	157
Figure 6.17. Association of MAPK alteration and response to capivasertib in patients enrolled based on an <i>AKT1</i> mutation .....	157

## Chapter 7

Figure 7.1. Baseline to EOT changes in gene mutation polyclonality in cohort A .....	166
Figure 7.2. Acquired alterations in cohort A .....	167
Figure 7.3. Amino acid structure of the wild type amino acid at <i>ESR1</i> p.F404, phenylalanine versus mutant variants isoleucine, valine and leucine. ....	168
Figure 7.4. Illustration of the ligand binding pocket of oestrogen receptor alpha bound to oestrogen.....	169
Figure 7.5. <i>ESR1</i> p.F404 base changes .....	170
Figure 7.6. <i>Cis/trans</i> analysis of the <i>ESR1</i> p.E380Q and p.F404 variants. ....	170
Figure 7.7. Activity of <i>ESR1</i> mutant variants in oestrogen-deprived compared to oestrogen-containing conditions, following prolonged oestrogen deprivation .....	172
Figure 7.8. Western blot of transfected MCF-7 .....	172
Figure 7.9. Activity of ERC transfected MCF-7 in oestrogen and fulvestrant compared to oestrogen alone .....	173
Figure 7.10. Cell viability curves for fulvestrant, AZD9833, tamoxifen and elacestrant.....	174

Figure 7.11. ERE-luciferase activity of each ERC in drug treated conditions relative to its activity in oestrogen alone .....	175
Figure 7.12. Baseline to EOT changes in gene mutation polyclonality, per gene, in cohort B. ....	177
Figure 7.13. Acquired alterations in cohort B.....	178
Figure 7.14. Baseline and EOT copy number within cohort B.....	179
Figure 7.15. Incidence of copy number gain in each gene from baseline to EOT in cohort B .....	180
Figure 7.16. Baseline to EOT changes in gene mutation polyclonality, per gene, in AKT1-directed and PTEN-directed patients .....	182
Figure 7.17. Acquired alterations in cohorts C and D .....	184

## List of Tables

### Chapter 1

Table 1.1. Multi-gene panels approved as companion diagnostic tests by the FDA in 2020.	27
Table 1.2. Studies and analyses within ABC investigating ctDNA as a prognostic or predictive marker.	30
Table 1.3. Studies comparing tissue and plasma ctDNA in ABC for mutation concordance.	35

### Chapter 2

Table 2.1. RNase P and triplex assays	62
Table 2.2. DdPCR assays used within plasmaMATCH to screen patients for targetable mutations.	63
Table 2.3. Master mixes for the respective assays employed within plasmaMATCH screening.	64
Table 2.4. Thermocycling conditions for mutation validation	65
Table 2.5. BCPv10 breast cancer panel gene and genome region inclusion	66
Table 2.6. Volume of components to be combined into a master mix prior to PCR amplification	66
Table 2.7. Thermocycling conditions for the first PCR amplification in the tissue sequencing protocol	66
Table 2.8. Thermocycling conditions for primer digestion with FuPa.	67
Table 2.9. Thermocycling conditions for NEBNext DNA end preparation	67
Table 2.10. Thermocycling conditions for PCR enrichment of indexed libraries	68
Table 2.11. Thermocycling conditions for KAPA library quantification.	69
Table 2.12. DNA end repair and A-tailing master mix constituents	78
Table 2.13. Thermocycling conditions following end repair and A-tailing	78
Table 2.14. Adaptor ligation master mix constituents	79
Table 2.15. Thermocycling conditions for whole genome library amplification	79
Table 2.16. Forward and reverse primer sequences for Sanger sequencing of the ERE-Luc plasmid.	83
Table 2.17. Master mix constituents of transfections using the in-house $\beta$ -gal DNA	87

Table 2.18. Master mix constituents of transfections using the maxi-prep of commercial $\beta$ -gal .....	87
Table 2.19. Primary and secondary antibody details for western blot.....	90

### Chapter 3

Table 3.1. Genes and genome regions covered by BCPv10.....	100
Table 3.2. Comparison of the mutation calls made by ddPCR (ABC-Bio Plasma) and by library preparation following the miniaturised protocol and the new targeted sequencing panel, BCPv10.....	101

### Chapter 4

Table 4.1. Clinico-pathological characteristics of patients with and without a ctDNA alteration .....	109
--	-----

### Chapter 7

Table 7.1. Drug concentration for ERC transfection experiments.....	175
---	-----

# 1. Chapter 1. Introduction

## 1.1. Advanced Breast Cancer

Breast cancer is both the commonest cancer and the commonest cause of cancer-related death in women worldwide<sup>1</sup>. In 2018, an estimated 2.1 million women will have been diagnosed with breast cancer, accounting for almost 1 in 4 of cancer cases in women<sup>1</sup>. In the UK while approximately 76% of women diagnosed with breast cancer survive more than 10 years from diagnosis<sup>2</sup> a substantial number of women will develop incurable advanced metastatic disease. With an estimated 11,500 women dying in the UK every year from the disease<sup>2</sup>, there is a pressing need to improve treatment outcomes in breast cancer.

There are several research approaches that can be taken to improve treatment outcomes in breast cancer. Firstly, resistance mechanisms can be identified and characterised so that new treatment strategies to avoid the resistance can be sought. A second approach is found within the identification of predictive and prognostic markers to differentiate which patients have the best chance of response to a therapy and which patients may benefit from an alternative treatment. Finally, the application of clinical tools which help guide therapeutic choices that prolong life can be investigated and implemented.

## 1.2. Personalised therapy

In the last 20 years the wider accessibility and affordability of genomic profiling has driven a move towards a 'personalised' approach to therapy. In breast cancer, the concept of selecting therapies to best suit the patient's cancer type was already familiar, with the relationship between hormone receptor positivity, as ascertained by immunohistochemistry (IHC), and likelihood of response to endocrine therapy identified in the 1970s<sup>3</sup>. With the advent of massively paralleled sequencing, however, it is now possible to delineate a patient's cancer at the genomic level, with the potential to match therapies directly to a patient's cancer profile through identification of genomic biomarkers that predict likelihood of response to a particular treatment, in a personalised therapy approach. In parallel, the advancement of genomic sequencing technology has furthered understanding of the genomic drivers of cancer, supporting the development of therapies that can be used to target these genomic aberrations.

The validity and importance of a personalised therapy approach was underlined by advances in lung cancer. The EGFR inhibitor gefitinib was found to be active in only 10-19% of patients with chemotherapy-refractory advanced non-small-cell lung cancer (NSCLC)<sup>4,5</sup>. Genomic sequencing revealed that the 8/9 patients who responded to therapy had a somatic mutations within the tyrosine kinase domain of *EGFR*<sup>6</sup>. Whilst early phase III trials of gefitinib in unselected patients with NSCLC demonstrated disappointing results<sup>7,8</sup>, later trials which analysed the influence of *EGFR*-mutant status on response identified this as an important predictive biomarker<sup>9</sup> leading to the licencing of the drug in this subpopulation.

With the accessibility of genomic sequencing coupled with increasing number of trials including biomarker analysis as an experimental endpoint, it is becoming increasingly evident that the ‘one-size-fits-all’ approach within oncological management is no longer fit for purpose. As we undergo the paradigm shift to a personalised-medicine approach, there is a pressing need to understand how we can best apply genomic sequencing in a clinical setting to optimise patient outcomes. Increasing numbers of studies and trials have identified confirmed responses in patients treated based on genomic sequencing of tumour tissue obtained through biopsy, and generally have supported a move towards personalising treatment at the genomic level<sup>10</sup>.

### 1.3. Circulating tumour DNA as a clinical tool

Circulating nucleic acids (cell free DNA, cfDNA) were first identified in human blood in 1948<sup>11</sup>. In the 1970s a comparative study of cfDNA in patients with and without a cancer identified higher levels in cancer patients, which fluctuated with treatment response and disease status<sup>12</sup>. Raised levels of cfDNA have also been identified in different physiological and pathological conditions, including exercise<sup>13</sup>, trauma<sup>14</sup>, cerebral infarction<sup>15</sup>, transplant<sup>16</sup> and active rheumatological conditions<sup>17,18</sup>. Foetal medicine spearheaded the clinical application cfDNA analysis, using the technology to identify parameters such as sex determination<sup>19</sup> and presence of aneuploidies<sup>20</sup>.

Cancer is a pathological condition fundamentally enabled by genomic instability leading to the acquisition of alterations which confer a selective advantage to the neoplastic cells<sup>21</sup>. Analysis of tumour-derived DNA is therefore key in understanding the pathological process and creating strategies to combat the disease. Tumour-derived DNA is released into the blood stream following tumour cell death<sup>22,23</sup>. This so-called circulating tumour DNA (ctDNA) is a component of the cfDNA in patients with cancer.



CtDNA consists of short strands of DNA of approximately 167bp in length<sup>24</sup>, corresponding to the length of DNA wrapped around a nucleosome. In 1989 it was demonstrated that ctDNA could be differentiated from cfDNA<sup>25</sup>. Following this, it was shown that key mutations could be identified in plasma through the analysis of ctDNA in patients with cancer<sup>26</sup>, giving an early insight into the potential clinical application of ctDNA analysis to non-invasively interrogate a patient's cancer profile and identify genomic features in an approach termed 'liquid biopsy' (Figure 1.1). It is now known that plasma ctDNA abundance is associated with disease features such as histological subtype, disease burden and stage<sup>27,28</sup>.

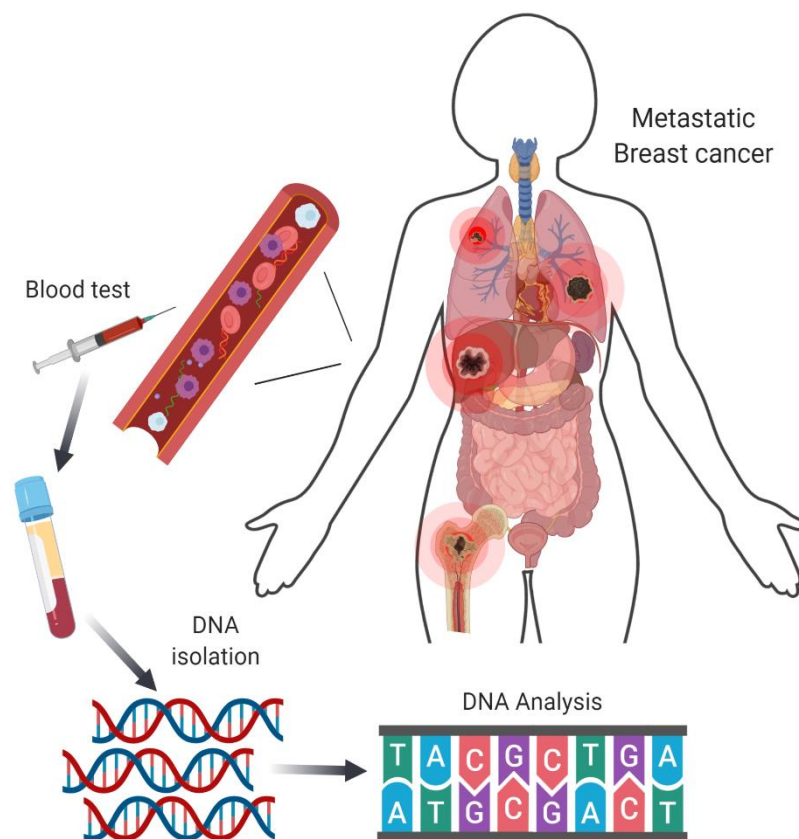


Figure 1.1. Schematic representation of the systemic release, isolation and analysis of circulating tumour DNA (ctDNA) in a patient with advanced breast cancer. Image created using BioRender software.

In the last decade, the application of ctDNA analysis within many areas of oncological management has been investigated. In early breast cancer, potential applications of ctDNA analysis include use as a tool for cancer screening and detection of minimal residual disease<sup>29,30</sup>, cancer localisation<sup>24,31</sup> and risk stratification and prognostication<sup>29,30</sup>. In advanced breast cancer (ABC) the potential applications are wide and include use as a biomarker of treatment response, in the identification of therapeutic targets, and to monitor for clonal

evolution and resistance to therapy<sup>11,32</sup>. This fellowship will focus on a number of these applications within ABC, as discussed below.

### 1.3.1.ctDNA analysis an alternative to tissue-based sequencing

Historically the genomic profile of a patient's cancer has been derived through sequencing of a tissue sample gained through a biopsy. Much of what is known about the genomic landscape of cancer comes through tissue sequencing within primary<sup>33</sup> and ABCs<sup>34-38</sup>. Tissue biopsy has several limitations. Firstly, and most importantly, biopsies are inherently invasive and incur risk for the patient<sup>39,40</sup>. Given that mutational profiles evolve over time, following metastatic relapse<sup>27,41-46</sup>, and with the selective pressure applied through treatments<sup>47</sup>, repeated genomic profiling is required. The prospect of repeated invasive biopsies quickly becomes unfeasible with the risks accumulating with each invasive biopsy. Furthermore, a range of factors may precluding successful tissue biopsy, including lack of accessible disease for biopsy, medical contraindications, biopsy failure and patient refusal<sup>32,40</sup>. CtDNA analysis via a liquid biopsy, however, is a non-invasive test that can be repeated throughout a patient's disease course. With a half-life of approximately 114 minutes<sup>48</sup>, ctDNA provides a genomic 'snapshot' of a patient's disease.

A second factor limiting the application of tissue biopsies is the existence of spatial tumour heterogeneity<sup>43,49-52</sup>. A single-site metastatic biopsy may miss clinically important genomic aberrations through selectively sampling a single area. Conversely, ctDNA is theoretically an admixture of tumour DNA derived from heterogeneous metastatic sites<sup>53,54</sup>, and therefore may circumvent tumour heterogeneity to give a full and representative profile of a patient's cancer.

### 1.3.2.ctDNA analysis for the identification of targetable alterations

The first ctDNA test was approved for clinical use by the FDA in 2016 for screening of *EGFR* mutations in non-small cell lung cancer to direct therapy with tyrosine kinase inhibitor erlotinib. Importantly, due to the lack of sensitivity of the tests, the FDA recommended a tissue biopsy in the instance where a ctDNA result is negative<sup>55,56</sup>. Following this, a second PCR-based assay was approved by the FDA for patients with ABC to identify patients with *PIK3CA*-mutant disease for treatment with PI3K-inhibitor alpelisib in combination with fulvestrant. Again, there was a requirement to undertake a tissue biopsy where the result of the ctDNA test was negative<sup>57</sup>.

In 2020, two commercial ctDNA multi-gene sequencing platforms (FoundationOne Liquid CDx and Guardant360 CDx) were licenced by the FDA for genomic profiling of solid tumours. In particular, the FDA approvals allow the ctDNA testing platforms to be used as a companion diagnostic for particular targetable mutations (Table 1.1).

<b>Blood Test</b>	<b>Cancer Type</b>	<b>Genetic Change</b>	<b>Corresponding Drug</b>
Guardant360 CDx	Non-small cell lung cancer	<i>EGFR</i> exon 19 deletions L858R mutation T790M mutation	Osimertinib
FoundationOne Liquid CDx	Non-small cell lung cancer	<i>EGFR</i> exon 19 deletions L858R mutation	Osimertinib Gefitinib Erlotinib
FoundationOne Liquid CDx	Prostate cancer	<i>BRCA1</i> and <i>BRCA2</i> alterations	Rucaparib
FoundationOne Liquid CDx	Ovarian	<i>BRCA1</i> , <i>BRCA2</i> mutation	Rucaparib
FoundationOne Liquid CDx	Lung (non-small cell)	<i>ALK</i> rearrangement	Alectinib
FoundationOne Liquid CDx	Breast	<i>PIK3CA</i> mutation	Alpelisib
FoundationOne Liquid CDx	Prostate	<i>BRCA1</i> , <i>BRCA2</i> , <i>ATM</i> mutation	Olaparib

Table 1.1. Multi-gene panels approved as companion diagnostic tests by the FDA in 2020.

### 1.3.3. As a predictive and prognostic biomarker

A number of studies and trials in ABC have investigated whether ctDNA can be used as a predictive or prognostic marker (Table 1.2). One of the earliest was Fribbens *et al*, who demonstrated that patients with baseline plasma *ESR1* alterations had a significantly shorter progression free survival (PFS) on subsequent exemestane therapy than those that were wild type<sup>58</sup>. It is now well established that the concentration of ctDNA in the bloodstream is prognostic<sup>11</sup>. However the majority of the data on baseline predictive biomarkers arises from tumour-based sequencing, and establishing the clinical utility of plasma-based alterations is an important unmet need.

<b>Marker type</b>	<b>Study/Trial</b>	<b>Summary</b>	<b>N assessable for ctDNA</b>	<b>Outcome</b>

			<b>biomarkers</b>	
Baseline Predictive Biomarker	SoFEA <sup>58</sup>	Prospective-retrospective study of patients randomised to exemestane versus fulvestrant-containing regimens (SoFEA)	161	In SoFEA, baseline presence of cfDNA detected <i>ESRI</i> mutations predicted for relative resistance to exemestane and relative sensitivity to fulvestrant.
	BOLERO-2 <sup>59</sup>	Retrospective analysis of baseline <i>ESRI</i> alterations and subsequent PFS and OS within the phase III BOLERO-2 trial (exemestane vs exemestane + everolimus)	541	cfDNA detected <i>ESRI</i> p.D538G was associated with a shorter PFS compared to wild type in the exemestane arm.
	MONALEE SA-2 <sup>60</sup>	Prospective analysis of baseline gene mutations and PFS within the phase III MONALEESA-2 trial (letrozole +/- ribociclib) in HR+ HER2- BC	427	Ribociclib + letrozole was associated with a greater PFS benefit compared to placebo + letrozole in patients with wild type versus altered genes involved in receptor tyrosine kinase signalling.
	MONALEE SA-3 <sup>61</sup>	Phase III trial of fulvestrant +/- ribociclib in HR+ HER2- BC	692	Consistent PFS benefit for ribociclib arm regardless of baseline alterations.
	BELLE-2	Phase III trial of fulvestrant +/- buparlisib	587	<i>PIK3CA</i> -mutant arm had longer median PFS on buparlisib.
	BELLE-3 <sup>62</sup>	Phase III trial of fulvestrant +/- buparlisib	348	In patients with <i>PIK3CA</i> -mutant positive ctDNA, buparlisib + fulvestrant significantly improved PFS vs placebo. However, the

				interaction test for <i>PIK3CA</i> status assessed by ctDNA was negative.
	Sharma <i>et al</i> <sup>63</sup>	Phase I/II randomised trial of Alpelisib + nab-paclitaxel in HER2- disease	42	Patients with <i>PIK3CA</i> -mutant disease demonstrated significantly longer median PFS.
	WJOG6110 B/ELTOP <sup>64</sup>	Randomised phase II trial investigating trastuzumab beyond progression with randomisation to Capecitabine or lapatinib	50	PFS and OS were relatively shorter in patients with <i>PIK3CA</i> mutations, irrespective of the treatment arm; however, in patients without <i>PIK3CA</i> mutations, lapatinib yielded relatively longer PFS and OS than Capecitabine.
	MONARCH 2 <sup>65</sup>	Phase III randomised trial of fulvestrant +/- abemaciclib in HR+ HER2- disease	Not stated	Patients with baseline <i>PIK3CA</i> -mutant and <i>ESR1</i> -mutant disease gained relatively more from the addition of abemaciclib than patients with wild type disease.
	SOLAR-1 <sup>66</sup>	Randomised phase III trial investigating the addition of alpelisib to fulvestrant in HR+ HER2- BC.	549	45% risk reduction in PFS for patients with ctDNA <i>PIK3CA</i> mutations (n=186); 20% for patients without (n=363).
	MONALEE SA-7 <sup>67</sup>	Randomised phase III trial tamoxifen/letrozole/anastrozole + gosereline +/- ribociclib in HR+ HER2- patients	565	Trend for more pronounced benefit with the addition of ribociclib in patients with altered <i>CCND1</i> , <i>GATA2</i> and genes involved in receptor tyrosine kinase signalling.

	Ma et al <sup>68</sup>	Non-randomised phase I trial of pyrotinib + capecitabine	28	Two or more mutations within the HER2 signalling pathway, PI3K/Akt/mTOR pathway and <i>TP53</i> were associated with significantly shorter PFS vs none or 1 (median, 15.8 vs. 26.2 months, p=0.006).
Prognostic Marker	Stover et al <sup>69</sup>	Analysis of ctDNA purity through analysis of somatic copy number alterations in the plasma of patients with advanced TNBC	164	Tumour fraction $\geq$ 10% associated with significantly worse survival.
	Clatot et al <sup>70</sup>	Retrospective analysis of <i>ESR1</i> D538G and p.Y537S/N/C in patients and the association with clinical outcome	144	Median OS was significantly lower in patients with an <i>ESR1</i> mutation compared to wild type.
	BOLERO-2 <sup>59</sup>	Retrospective analysis of baseline <i>ESR1</i> alterations and subsequent PFS and OS within the phase III BOLERO-2 trial (exemestane vs exemestane + everolimus)	541	Patients with an <i>ESR1</i> p.Y537S and/or p.D538G had a shorter OS than patients wild type both both.
	MONALEE SA-7 <sup>67</sup>	Randomised phase III trial tamoxifen/letrozole/anastrozole + gosereline +/- ribociclib. HR+ HER2- patients	565	Patients with <i>TP53</i> and <i>MYC</i> alterations had a poorer prognosis.

Table 1.2. Studies and analyses within ABC investigating ctDNA as a prognostic or predictive marker. HR+ HER2-, hormone receptor positive HER2 non-amplified. TNBC, triple negative breast cancer. BC, breast cancer

#### 1.3.4. To interrogate the advanced breast cancer genomic landscape

Several large-scale sequencing efforts have characterised the landscape of ABC from tissue-based sequencing<sup>34-38</sup>. The ABC landscape according to plasma-based sequencing has not yet

been established in a large-scale study<sup>71</sup>. Given the aforementioned limitations of tissue sequencing with regard to tumour heterogeneity and the potential for liquid biopsy to enter clinical practice, there is an unmet need to establish the landscape of breast cancer according to plasma ctDNA sequencing.

#### 1.3.5. To identify mechanisms of resistance

Through the selective pressure applied by therapy, subclones harbouring resistance mutations survive and proliferate in a clonal manner to dominate the disease profile. Tracking the emergence of known resistance mechanisms has demonstrated that these mutations can be identified in the cfDNA months prior to clinical progression and/or predicting poorer outcome on subsequent therapy in breast cancer<sup>72</sup>, colorectal cancer<sup>73</sup>, NSCLS<sup>74</sup> and ovarian cancer<sup>75,76</sup>, to name a few. In situations where the resistance mutations are known *a priori*, targeted detection using ctDNA testing approaches able to sensitively identify mutations present at low allele frequency (AF), as is likely to be the case with an emerging subclonal resistance mutation, can be used. Where resistance mutations are not known *a priori*, broader sequencing approaches may be required but, depending on the depth of sequencing, may not be able to identify mutations occurring at low allele frequency.

Few studies within breast cancer have used paired baseline and end-of-treatment (EOT) ctDNA sequencing as the primary method to investigate for novel putative resistance mechanisms to targeted therapy. Two studies have demonstrated *ERBB2* mutations and/or amplification in the plasma within patients who had progressed on HER2-directed tyrosine kinase inhibitors<sup>77,78</sup>. O'leary *et al* used paired baseline and EOT plasma samples in the PALOMA-3 trial to demonstrate evolution in the driver genes at progression in patients who initially responded to treatment with palbociclib +/- fulvestrant<sup>79</sup>. To date, the potential of ctDNA analysis to identify resistance mutations has not been fully reached with respect to targeted therapies in breast cancer.

#### 1.3.6. Potential limitations and questions around the application of ctDNA analysis

Multiple studies in ABC have compared tissue sequencing with contemporaneous ctDNA analysis to validate this approach (Table 1.3). The results demonstrate highly varying results that may be due to either biological or technical considerations. This variation highlights the potential limitations and pitfalls of ctDNA analysis and the use of tissue assessment as the gold standard measure.

Authors	Design	Sequencing/ ctDNA analysis technique	Number of assessable patients	Mutations compared	Results
Higgins <i>et al</i> , 2012 <sup>80</sup>	Part 1: Retrospective contemporaneous (same day) samples  Part 2: Prospective ctDNA with archival tissue samples	plasma: BEAMing  FFPE Tissue: Sequencing and BEAMing	Part 1: 49  Part 2: 51	<i>PIK3CA</i> gene: Ex 9 p.1633G> A E545K; Ex 20 p.3140A> G H1047R; Ex 20 p.3140A> T H1047L	Part 1: BEAMing of tissue and plasma showed 100% of patients had concordant results at the variant level.  Part 2: Sequencing of tissue and BEAMing of plasma: 72.5% of patients had concordant results at the variant level.  BEAMING of both tissue and plasma: 100% of patients had concordant results at the variant level.
Schiavon <i>et al</i> , 2015 <sup>81</sup>	Prospective	Plasma and tissue: ddPCR	31	<i>ESR1</i> mutations: p.L536R, p.Y537S, p.Y537N, p.Y537C, p.D538G	97% of patients had concordant results.
Maxwell <i>et al</i> , 2017 <sup>82</sup>	Prospective contemporaneous (within 1 month) samples	Plasma and FFPE tissue: Targeted sequencing	32	A number of commonly altered genes in breast cancer (not specified).	62% of variants were in agreement between plasma and tissue.



Kim <i>et al</i> , 2017 <sup>83</sup>	Retrospective non-contemporaneous samples	Plasma and tissue: Targeted sequencing	72	A number of genes (not specified)	75% agreement in variants detected overall, 100% for <i>PIK3CA</i> and <i>AKT1</i> mutations.
Chae <i>et al</i> , 2017 <sup>84</sup>	Retrospective non-contemporaneous samples	Plasma and tissue: Targeted sequencing	45	Between 45 and 67 genes	94.2% agreement (including concordant negative).
Adalsteinsson <i>et al</i> , 2017 <sup>85</sup>	Prospective	Plasma and tissue: Whole exome sequencing	41	All genes	88% of clonal and 47% of subclonal SNVs detected in the tumour were identified in cfDNA, while 88% of the clonal and 45% of the subclonal SNVs in the cfDNA were present in the tumour.
Urso <i>et al</i> , 2021 <sup>86</sup>	Prospective contemporaneous samples	Plasma and tissue: ddPCR	43	<i>ESR1</i> : p.Y537S, p.Y537C, p.Y537N, p.D538G, p.E380Q	88% of patients had concordant results.
Board <i>et al</i> , 2010 <sup>87</sup>	Archival metastatic and primary	Plasma and tissue: allele-specific PCR	41	<i>PIK3CA</i> : p.H1047R, p.H1047L, p.E545K, p.E542K	95% of patients had concordant results.
Rothé <i>et al</i> , 2014 <sup>88</sup>	Prospective study	Plasma and tissue: Targeted sequencing	17	50 genes	76% of patients had concordant results.

Takeshit a <i>et al</i> , 2017	Retrospective study	Plasma and tissue: ddPCR	35	<i>ESR1</i> : p.Y537S, p.Y537N, p.Y537C, p.D538G	74% of patients had concordant results.
BELLE- 2	Prospective trial, archival tissue	Plasma and tissue: Sanger sequencing	342	<i>PIK3CA</i> mutations within exons 1, 7, 9, or 20	77% of patients had concordant results.
BELLE- 3 <sup>62</sup>	Prospective trial, archival tissue	Plasma: BEAMing Fresh or archival tissue: PCR	256	<i>PIK3CA</i> hotspot mutations in exons 9 and 20.	83% of patients had concordant results.
WJOG61 10B/ELT OP <sup>64</sup>	Randomised phase II trial investigating trastuzumab beyond progression with randomisation to Capecitabine or lapatinib	Plasma: ddPCR  Tissue: ddPCR	26	<i>PIK3CA</i>	85% of patients had concordant results. Sensitivity 60%, specificity 100%
MONAR CH-2 <sup>89</sup>	Phase III randomised trial of fulvestrant +/- abemaciclib in HR+ HER2- disease	Plasma and archival tissue: ddPCR	Not stated	<i>PIK3CA</i> : p.E542K, p.E545K, p.H1047L, p.H1047R <i>ESR1</i> : p.D538G, p.Y537C, p.Y537N, p.Y537S)	<i>PIK3CA</i> concordance was 63%  <i>ESR1</i> concordance was 37%

Sharma <i>et al</i> <sup>63</sup>	Phase I/II randomised trial of alpelisib + nab-paclitaxel in HER2- disease	Plasma and tumour: next generation sequencing	42	<i>PIK3CA</i> activating mutations	70% concordance
Kuang <i>et al</i> <sup>90</sup>	Prospective study	Plasma: ddPCR  Tissue: ddPCR	23	<i>ESR1</i> p.Y537S, p.D538G, p.E380Q, p.Y537N, p.Y537C	100% sensitivity, 88% specificity

Table 1.3. Studies comparing tissue and plasma ctDNA in ABC for mutation concordance. HR+ HER2-, hormone receptor positive HER2 non-amplified.

The first major biological source of difference between tumour and ctDNA analysis can be ascribed to the existence of patients who shed low or undetectable levels of ctDNA into their circulation (“low-shedders”). It is known that certain patient groups have lower levels of ctDNA, such as those with a low burden of disease<sup>28</sup>, disease location<sup>28</sup> and disease features such as necrosis and neo-vascularisation<sup>27</sup>. These patients may be less amenable disease profiling by ctDNA testing. It is important that clinicians are able to evaluate in which patients ctDNA profiling via a liquid biopsy is best suited, and in which patients there are more likely to be false negative results.

A second major source of biological difference are variant features such as clonal dominance and temporal acquisition. Truncal and clonally dominant alterations are more likely to be concordant between tissue and plasma. Subclonal alterations, meanwhile, by their nature have a lower allele frequency and are more likely to be metastatic-clade, thus more likely to cause discordant results between plasma and tissue. These factors raise the argument that paired tissue sequencing may not be the appropriate benchmark with which to assess the analytical validity of ctDNA sequencing. Meanwhile the clinical relevance of subclonal alterations, whether they represent targetable alterations or potential resistance mechanisms, is not well understood.

More recently, the existence of clonal haematopoiesis (CH) has been identified as a biological factor which potentially limits the clinical application of ctDNA analysis. CH describes the accumulation of somatic mutations in hematopoietic cells that are clonally propagated from

their precursors<sup>91</sup>. Mutations typically occur in a number of genes, including *DNMT3A*, *TET2*, *ASXL1*, *JAK2*, *PPM1D*, *TP53* and *SF3B1*<sup>92</sup>, and confer increased lifetime risk to the development of a haematologic malignancy<sup>93</sup>. Presence of CH increases with age, with approximately 10% (or higher, in sensitive assays) of over 65-year-olds demonstrating the mutations<sup>93</sup>. The mutations are present in cfDNA, and, aside from occurring more frequently in certain genes, cannot be distinguished from tumour-derived somatic mutations and thus confound ctDNA analysis. One study which undertook paired sequencing of cfDNA and germline DNA with a large panel that included genes implicated in CH suggested that as many as 53% of alterations identified in cancer patients could be derived from CH<sup>94</sup>. One way to circumvent this issue is to simultaneously sequence the patient's germline DNA, in which the CH-mutations will also be identified and thus enable differentiation from the somatic tumour-associated alterations. Importantly, however, many of the genes associated with breast cancer are not mutated in CH (with the exception of *TP53*<sup>92</sup>), and so analyses involving these genes would not be expected to be significantly confounded by this issue.

Technical considerations limiting the application of ctDNA analysis often stem from a lack of consensus in methodology at each stage of a liquid biopsy<sup>95</sup>. From the patient side, there is much variation in sample collection and processing processes. Following this there are technical differences between DNA sequencing/testing platforms and their bioinformatic pipelines, with comparative studies demonstrating that the platforms vary widely in the mutations that are identified<sup>96,97</sup>. Finally, for variants with low allele frequency, stochastic sampling may result in variants being 'missed' by one ctDNA test whilst being identified in another.

## 1.4. plasmaMATCH

### 1.4.1. Study rationale

To bring ctDNA analysis into clinical practice, the application of the approach must be proven to be analytically valid, clinically valid and demonstrate clinical utility<sup>95</sup>. Analytical validity refers to ability of ctDNA testing to accurately detect the variant of interest. Clinical validity refers to the need for ctDNA testing to accurately report the tumour genomics<sup>95</sup>. Clinical utility refers to the requirement of ctDNA testing to have high levels of evidence to robustly demonstrate that it improves patient outcome<sup>95</sup>. Prior trials of targeted therapy recruited patients based on tissue sequencing results, with plasma ctDNA concordance of the targeted

mutation often a secondary or exploratory endpoint undertaken retrospectively. The plasmaMATCH trial was designed with the aim of investigating the activity of targeted therapy in patients selected purely based on ctDNA testing, in a move towards ascertaining the clinical validity and utility of the ctDNA testing. The primary objective of the trial was to assess the safety and activity profile of targeted therapies in patient subgroups identified by ctDNA screening<sup>98</sup>. With two orthogonal approaches to ctDNA testing, alongside paired tumour sequencing, the study design also allowed for further assessment of the analytical validity of ctDNA testing in clinical practice.

#### 1.4.2. Study design

PlasmaMATCH was a multiple parallel cohort, open label, multicentre phase IIa clinical trial aiming to provide proof of principle efficacy for designated targeted therapies in patients with ABC where the targetable mutation is identified through ctDNA screening (Figure 1.2). Eligibility criteria included women over 18 years with measurable metastatic or recurrent locally advanced breast cancer, which has progressed within the last 6 weeks. Patients must have received at least one line of therapy in the metastatic setting and/or relapsed within 12 months of completing (neo)adjuvant chemotherapy, and not received more than 2 lines of cytotoxic chemotherapy in the metastatic setting.

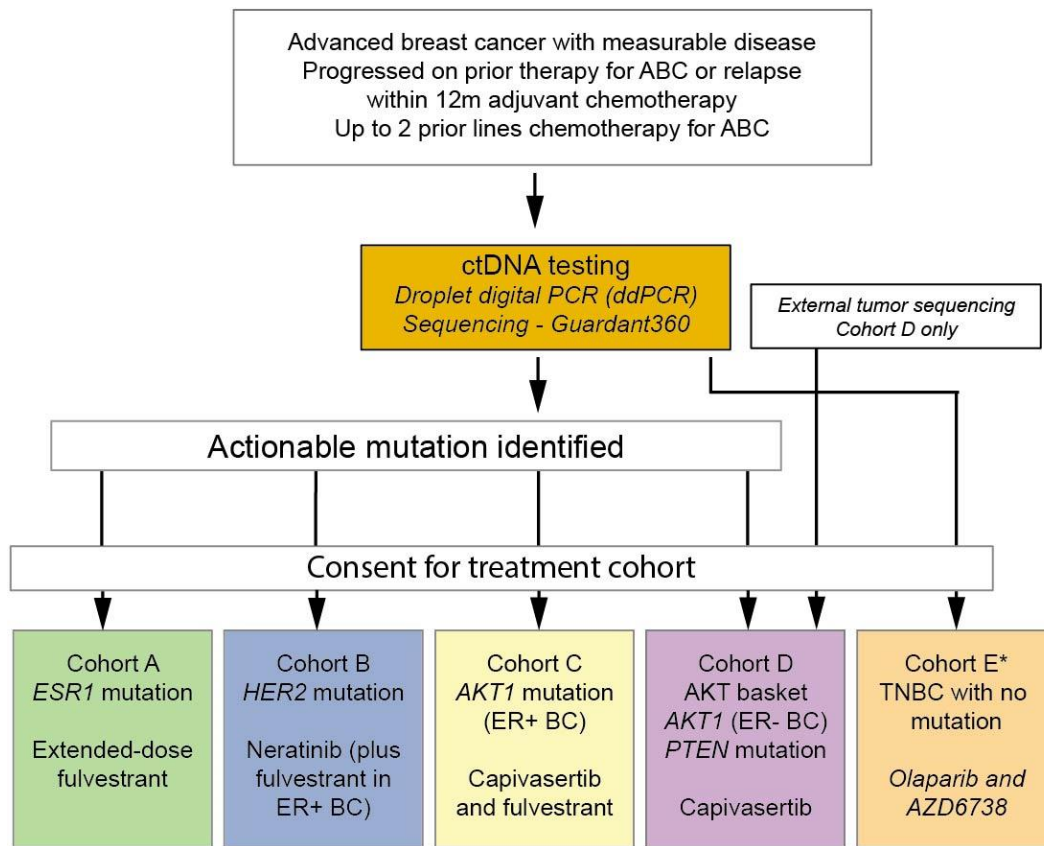


Figure 1.2. The plasmaMATCH trial schema.

Eligible women underwent ctDNA testing with one of two techniques, droplet digital PCR (ddPCR), and partway through the trial, targeted sequencing (Guardant360, Guardant Health, USA). Eligible patients found to have an actionable mutation in one of a number of genes (Figure 1.3) were able to enrol in the retrospective cohort for treatment with a matched targeted therapy (Figure 1.2). Patients enrolled into a cohort were required to provide a fresh metastatic biopsy or an archival metastatic tissue sample.

DNA testing	ctDNA		
	Droplet digital PCR (ddPCR)	Targeted Sequencing Guardant360	
Testing	Prospectively tested	Prospectively and retrospectively tested	
Location	Centralised at Centre for Molecular Pathology, Royal Marsden	Guardant Health	
n	1025	800 (364/436)	
Gene Inclusion	<i>PIK3CA</i> multiplex:	E542K	<i>AKT1 ALK APC AR ARAF ARID1A ATM BRAF BRCA1 BRCA2 CCND1 CCND2 CCNE1 CDH1 CDK4 CDK6 CDKN2A DDR2 EGFR ERBB2 ESR1 EZH2 FBXW7 FGFR1 FGFR2 FGFR3 GATA3 GNA11 GNAQ GNAS HNF1A HRAS IDH1 IDH2 JAK2 JAK3 KIT KRAS MAP2K1 MAP2K2 MAPK1 (ERK2) MAPK3 (ERK1) MET MLH1 MPL NF1 NFE2L2 NOTCH1 NPM1 NRAS NTRK1 NTRK3 PDGFRA PIK3CA PTEN PTPN11 RAF1 RB1 RET RHEB RHOA RIT1 ROS1 SMAD4 SMO SRY STK11 TERT TP53 TSC1 VHL</i>
		E545K	
		H1047R	
		H1047L	
	<i>ESR1</i> multiplexes:	E380Q	
		L536R	
		Y537C	
		D538G	
		S463P	
		Y537N	
	<i>AKT1</i> singleplex:	Y537S	
		E17K	
	<i>ERBB2</i> singleplex:	S310F	
		S310Y	
L755S			
V777L			
GSP insertion (P780_Y781insGSP)			
YVMA insertion (A775_G776insYVMA)			

Figure 1.3. ctDNA testing approaches within plasmaMATCH. Genes and alterations included with each approach are illustrated. Activating mutations in *ESR1*, *AKT1* and *ERBB2* allowed entry into a treatment cohort.

In order to identify the required number of patients for each cohort, the trial aimed to recruit over 1000 women with ABC for ctDNA testing for the actionable mutations.

#### 1.4.3. Use of plasmaMATCH samples within the fellowship

Screening plasma was processed by the Royal Marsden Biobank facility prior to extraction, quantification and mutation assessment using ddPCR for a binary positive or negative result at the Centre for Molecular Pathology, Royal Marsden, UK, as part of the trial. From July 2018, screening plasma was additionally shipped prospectively to Guardant Health, California, for

targeted sequencing using a 73-gene targeted panel. Positive actionable mutations identified by either platform allowed eligibility to Cohorts B-D. CMP ddPCR results only were reviewed for eligibility for Cohort A.

Further analysis on the baseline ddPCR results, including establishing the gene variant/s present and assessing the allele frequency, was completed as part of the fellowship. For samples shipped to Guardant Health, prospectively sequenced samples were shipped as part of the trial at the time of enrolment for ctDNA testing. For retrospectively tested samples, the samples were prepared and shipped as part of the fellowship. Additionally, C2D1 and EOT samples were prepared and shipped as part of the fellowship. Targeted sequencing results from all patients who underwent targeted sequencing at all timepoints by Guardant360 were analysed as part of the fellowship.

All tissue samples were processed and analysed as part of the fellowship.

#### 1.4.4. Cohort A: Patients with *ESRI* alterations for treatment with extended dose fulvestrant

##### 1.4.4.1. *ESRI* alterations

Approximately 75% of breast cancers are oestrogen receptor (ER) positive (HR<sup>+</sup>)<sup>99</sup> and express oestrogen receptor  $\alpha$ . The binding of oestrogen to an ER promotes expression of genes controlling cell proliferation and survival<sup>100</sup>. For the HR<sup>+</sup> breast cancer subtype, hormonal therapy is a cornerstone of therapy in both the early and advanced settings. Development of resistance to hormonal therapy is common, with a number of ER<sup>+</sup> breast cancers displaying intrinsic, or *de novo*, resistance, and many developing secondary resistance following hormonal therapy. Mechanisms of resistance include reduced ER expression<sup>101</sup>, enhanced signalling via the phosphatidylinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) pathways (including enrichment for *ERBB2* alterations, *NFI* loss of function alterations and *KRAS* and *EGFR* alterations)<sup>34,102</sup>, alterations within genes regulating gene transcription such as *MYC* and *CTCF*, and development of mutations within *ESRI*, which encodes the ER<sup>103</sup>.

The ER is a nuclear protein consisting of two transcriptional activation domains at the N- (ligand-independent) and C- (ligand dependent) terminus, a ligand-binding domain (LBD) within the C-terminal region and DNA-binding and hinge domains within the protein core<sup>103</sup>. When not bound to an activating ligand, ER is mostly inactive and bound by heat-shock



proteins (HSP)<sup>104</sup> including HSP90. Binding of an agonist, such as oestrogen, releases the bound HSP leading to receptor dimerization and folding of helix 12 over the LBD creating a hydrophobic groove into which co-activators bind<sup>104</sup>. Activating mutations, which cluster in the LBD of *ESR1*, promote agonist conformation of the ER receptor leading to ligand-independent ER activity and tumour growth<sup>105-110</sup>. The mutations are commonly found in patients following treatment with aromatase inhibitors in the metastatic setting<sup>81,111</sup>, and, less commonly, tamoxifen<sup>58</sup>, and arise due to the selective evolutionary pressure applied by the hormonal therapy promoting development of the resistance mutations<sup>103</sup>.

#### 1.4.4.2. Targeting of mutant *ESR1* with extended-dose fulvestrant

Fulvestrant is a selective oestrogen receptor degrader (SERD) that competitively inhibits the binding of oestradiol to ER<sup>112</sup>. The binding of fulvestrant to ER impedes receptor dimerization and nuclear localisation, thus preventing activation of estrogen response elements within the regulatory regions of oestrogen sensitive genes<sup>112</sup>. The fulvestrant-bound ER is also unstable, leading to increased degradation of the ER receptor. Early clinical data demonstrated that a single pre-operative dose of fulvestrant significantly reduced ER expression in a dose-dependent manner relative to placebo and, for the 250mg dose, relative to tamoxifen<sup>113</sup>. Subsequent clinical trials demonstrated non-inferior response of 250mg fulvestrant in hormone-sensitive or ER+ patients in the metastatic setting compared to anastrozole, tamoxifen, and exemestane<sup>114-119</sup>.

Clinical trial data and large-scale sequencing efforts indicate that while patients with baseline *ESR1* mutations have inferior response to aromatase inhibition<sup>58</sup>, sensitivity to fulvestrant is retained<sup>34,111,120</sup>. A combined analysis of the EFECT and SoFEA trials demonstrated that patients with a baseline *ESR1* mutation had an inferior PFS in patients treated with exemestane compared to fulvestrant<sup>120</sup>.

*In vitro* data indicate that while cells with activating mutations in the LBD of *ESR1* remain sensitive to fulvestrant, high doses are required suppress ER activity to the same degree as wild type *ESR1*<sup>105-107</sup>. *In vivo*, transcriptional response is significantly altered following high dose fulvestrant (500 mg on days 0, 14, 28, and monthly thereafter), but not low dose (250 mg/28 days)<sup>121</sup>. Clinical trial data supports the notion that higher dose fulvestrant is required for greater clinical effect. The phase III CONFIRM trial demonstrated a statistically significant longer median PFS in patients treated with higher dose fulvestrant (500mg IM days 0, 14, 28

and every 28 days thereafter) compared to the standard dose (250mg IM days 0, 14, 28 and every 28 days thereafter), the higher dose associating with superior clinical efficacy and longer overall survival<sup>122</sup>. In head-to-head analysis of higher-dose (500mg IM monthly following loading dose) fulvestrant against other endocrine therapies in the metastatic setting, fulvestrant demonstrated similar or improved treatment outcomes<sup>123,124</sup>.

This background data provided the rational for cohort A in plasmaMATCH. Patients found to have an activating *ESR1* mutation on ctDNA screening were eligible to enrol into cohort A, for treatment with extended-dose fulvestrant (twice as frequent standard dosing: 500mg fulvestrant (IM) on Cycle 1 Days 1, 8 and 15 and Cycle 2 onwards Days 1 and 15).

#### 1.4.4.3. Predictive and prognostic biomarkers of response to fulvestrant

Baseline expression of the cyclic dependent kinase 6 (CDK6), which promotes cell cycle entry from G1 to S phase through association with D-type cyclins and phosphorylation and inactivation of retinoblastoma (RB), has been found to associate with subsequent PFS on fulvestrant<sup>125</sup>. In a retrospective analysis of baseline tissue CDK6 expression, high baseline levels was associated with median time to progression of 2.5 months vs. 8.2 months in the low expression group (p=0.0006), with results from a validation cohort corroborating the results. The same difference was not seen in patients treated with tamoxifen or aromatase inhibitor therapy.<sup>125</sup>

A large-scale genomic sequencing effort analysed biopsies from patients taken prior to subsequent SERD therapy<sup>34</sup>. Patients with alterations causing enhanced signalling via the MAPK pathway had a shorter PFS on SERD therapy than patients without any of a MAPK pathway mutation, *ESR1* mutation or alteration affecting a transcription factor<sup>34</sup>, suggesting presence of baseline MAPK alterations is a predicts response to subsequent SERD therapy.

Analysis of the FERGI trial, and a combined analysis of the SoFEA and PALOMA3 trials demonstrated that *ESR1* mutant polyclonality and allele frequency are not associated with response to fulvestrant<sup>58,111</sup>. The specific *ESR1* mutation variant present at baseline, however, has been found to be associated with sensitivity to fulvestrant *in vitro* and *in vivo*, with p.Y537S showing the highest resistance to treatment<sup>126</sup>. The combined analysis of the SoFEA and PALOMA3 trials did not recapitulate this result, with no significant difference in PFS in patients with different baseline hotspot *ESR1* alterations treated with fulvestrant, albeit the analysis was limited by sample size (n=224)<sup>58</sup>.

#### 1.4.4.4. Putative resistance mechanisms to fulvestrant

In contrast to tamoxifen and aromatase inhibitor therapy, little is known about specific mechanisms of resistance to SERD, or fulvestrant therapy. *In vitro* data gained from six fulvestrant-resistance cell lines demonstrated no difference between the parental and fulvestrant resistant cell lines in their genomic profile, as assessed by a 51-gene targeted panel<sup>127</sup>. Conversely, expression analysis of 3 of the cell lines demonstrated enrichment of pathways involved in cell cycle progression and DNA replication and repair, including cycle checkpoint regulators such as CDK6, CCND3, CCNE1, CCNE2, CDK1, and CDK2<sup>127</sup>. Two other groups have also identified increased expression of CDK6 in fulvestrant resistant cell lines<sup>125,128</sup> which, together with a clinical association of baseline CDK6 expression with PFS in fulvestrant<sup>125</sup>, supports a central role of dysregulated cell cycle progression control in fulvestrant resistance.

Clinical trial data has implicated *ESR1* and *PIK3CA* alterations in fulvestrant resistance. Baseline and EOT sequencing analysis of the fulvestrant-only arm in PALOMA3 demonstrated enrichment for *ESR1* mutation p.Y537S at EOT. Patients with a p.Y537S mutation at baseline had a significantly shorter PFS than those who later acquired the alteration. Combined, this data suggests the *ESR1* alteration p.Y537S may confer resistance to fulvestrant therapy<sup>79</sup>. A further finding was enrichment of mutations within *PIK3CA* at EOT, suggesting activation the PI3K/AKT/mTOR pathway may confer resistance to fulvestrant<sup>79</sup>, and is supported by *in vitro* data<sup>102</sup>.

#### 1.4.5. Cohort B: Patients with *ERBB2* mutations for treatment with neratinib +/- fulvestrant

##### 1.4.5.1. *ERBB2* alterations

The *ERBB2* gene encodes HER2 (Human epidermal growth factor receptor 2), a transmembrane growth factor receptor from the epidermal growth factor receptor (EGFR) family. The EGFR receptors, HER1/EGFR, HER2, HER3 and HER4, share a common morphology, consisting of an extracellular domain (ECD), a transmembrane domain (TMD), an intracellular region which includes the juxtamembrane domain (JMD) and kinase domain (KD) and a C-terminal tail domain (CTD)<sup>129</sup>. Overexpression of HER2, also known as ERBB2, occurs in 15% of breast cancer patients and is associated with a more aggressive breast cancer phenotype<sup>130,131</sup>. The HER2 receptor is unusual within its family both for lacking a specific

activating ligand<sup>132</sup> and residing in an open conformation ready to form dimers as the preferred dimerization partner of the other members of the EGFR/HER family<sup>133</sup> (Figure 1.4). Homodimerisation with other HER2 or heterodimerisation with other members of the EGFR/HER family<sup>133</sup> activates its intracellular tyrosine kinase domain leading to enhanced signalling via the PI3K and MAPK pathways<sup>134</sup> (Figure 1.4). HER2 overexpression, which often occurs as a result of *ERBB2* amplification, leads to increased HER2 receptor dimerization<sup>135</sup>. This, in turn, promotes cell growth, survival and differentiation via activated downstream signalling<sup>136</sup>.

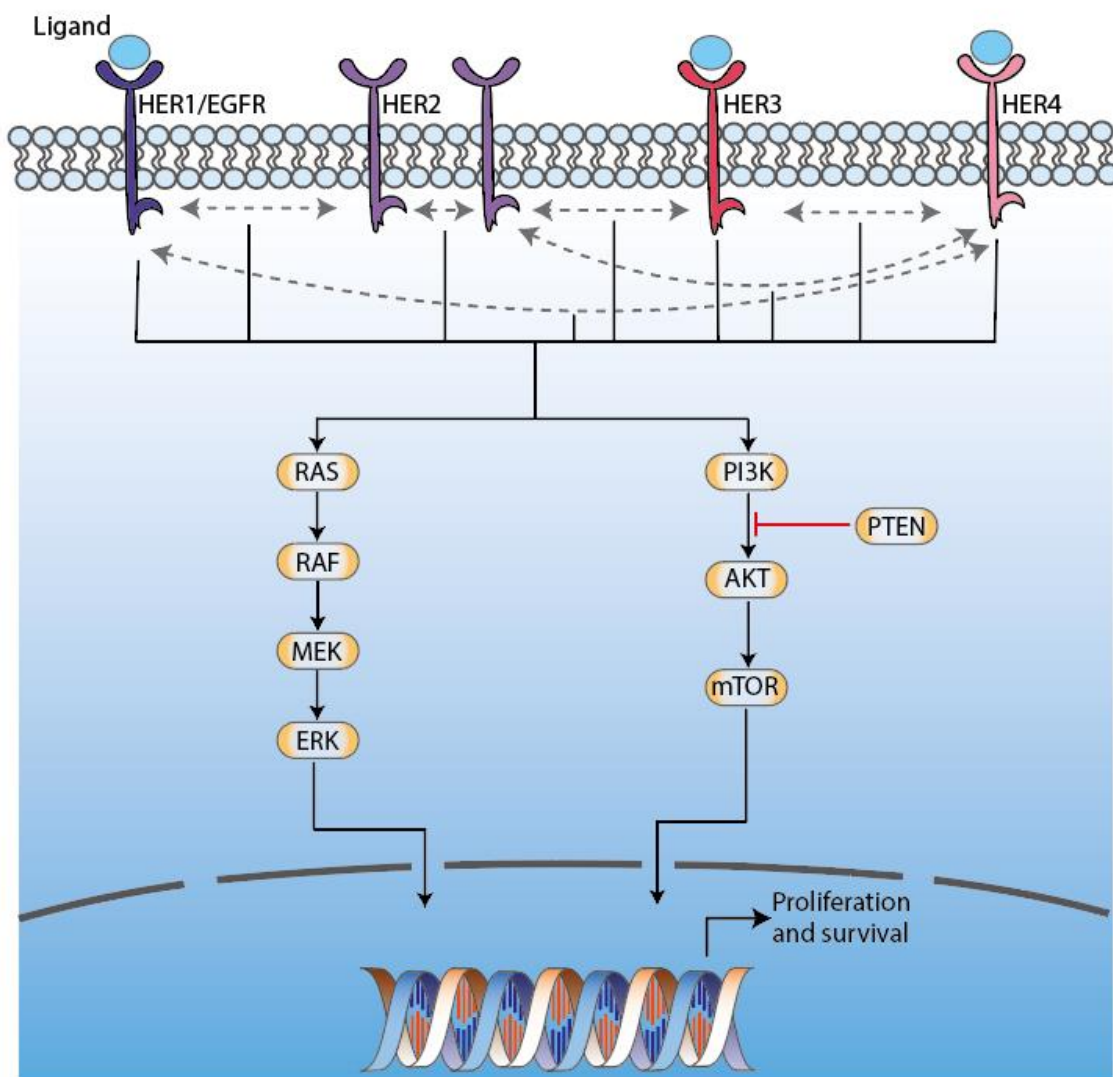


Figure 1.4. The EGFR/HER receptor signalling and downstream pathway activation. Ligand binding to EGFR/HER receptors promotes dimerization with other members of the EGFR/HER family, activating downstream signalling via MAPK and PI3K pathways promoting cell proliferation and survival. In the case of

HER2, it does not need activation via ligand binding and lies in the open confirmation ready to dimerise with other HER2 or other members of the EGFR/HER family. Image adapted from Goutsouliak *et al*, 2020<sup>137</sup>.

Recently, it has been demonstrated that approximately 2.4% percent of breast cancer patients harbour activating mutations in *ERBB2*<sup>138-140</sup>. It is had previously been thought that *ERBB2* mutations and HER2-amplification are mutually exclusive, based upon a low frequency of *ERBB2* mutations in HER2-amplified breast cancers within The Cancer Genome Atlas of primary breast cancers<sup>33</sup>. More recent studies have identified a higher rate (~7%) of *ERBB2* mutation in HER2-amplified metastatic breast cancers<sup>141</sup>. The most common mutations occur within the KD and ECD domains<sup>140</sup>, and have been causally linked with resistance to oestrogen therapy<sup>34,142,143</sup>. Activating mutations in *ERBB2* increase the flexibility of the HER2 protein, promoting binding with HER3 and subsequent activation of downstream signalling particularly via the PIK3CA/AKT/mTOR pathway<sup>142,144</sup>. Given the role of HER2 in promoting cell growth and survival, activating mutations in the gene are an attractive therapeutic target.

#### 1.4.5.2.Targeting of activating *ERBB2* mutations with neratinib

Neratinib is an irreversible tyrosine kinase inhibitor of HER1/EGFR, HER2 and HER4, with particular potency in inhibiting HER1/EGFR and HER2<sup>145</sup>. Early clinical work demonstrated neratinib decreased downstream MAPK and AKT phosphorylation and inhibited cell cycle progression in cancer cell lines, and repressed tumour growth in tumour xenografts<sup>145</sup>. This provided the rationale for early clinical trials of neratinib that showed promising results in both early and advanced breast cancer<sup>146-148</sup>. The MutHER phase 2 trial investigated the activity of neratinib in *ERBB2* mutant HER2 non-amplified breast cancer, demonstrating a clinical benefit rate of 31% amongst patients with a range of activating *ERBB2* alterations<sup>149</sup>. The SUMMIT trial also demonstrated the efficacy of neratinib in *ERBB2*-mutant metastatic breast cancers with an objective response rate of 32%<sup>150</sup>.

This background data provided the rational for cohort B in plasmaMATCH. Patients with an activating *ERBB2* mutation on ctDNA screening were eligible to enrol into cohort B, for treatment with neratinib (240mg PO OD), with the addition of fulvestrant (500mg IM on Cycle 1 Days 1 and 15 and Cycle 2 onwards Day 1) in HR+ patients.

#### 1.4.5.3. Predictive and prognostic biomarkers response to neratinib in *ERBB2*-mutant cancer

The SUMMIT trial specifically recruited patients with *ERBB2* mutations in a range of cancers for treatment with neratinib +/- fulvestrant<sup>150</sup>. The greatest activity was found in patients with breast cancer and in patients with known hotspot mutations in *ERBB2*<sup>150</sup>. The vast majority of patients in this trial had clonal *ERBB2* alterations, and none of the patients with subclonal *ERBB2* achieved clinical benefit<sup>77,150</sup>.

Co-mutations also appear to be predictive, with enrichment of *TP53* and *HER3* mutations in patients receiving no clinical benefit in the SUMMIT trial<sup>150</sup>. Within the breast cancer cohort of SUMMIT, *TP53* mutations were also associated with lack of clinical benefit<sup>77</sup>. Notably, breast cancer patients with multiple activating *ERBB2* alterations (mutations/amplification, 7 of 44 patients) did not achieve clinical benefit<sup>77</sup>. Integrating a trend identified within the SUMMIT breast cancer cohort with analysis of a broader cohort of prospectively sequenced tumours (n=29,373), the group suggest that additional alterations in *ERBB2* and *ERBB3* are selected for in a subset of patients, and that when present at baseline these tumours may be resistant to neratinib<sup>77</sup>.

#### 1.4.5.4. Putative resistance mechanisms to neratinib

Gatekeeper mutations are recognised as a common cause of resistance to targeted therapy. Neratinib is no exception, and the *ERBB2*<sup>T798I</sup> mutation was identified in a patient with *ERBB2*-mutant breast cancer following progression on neratinib. Isoleucine, which is bulkier than threonine, causes steric hindrance to neratinib binding<sup>151,152</sup> leading to drug resistance. Two patients in the SUMMIT trial, who enrolled patients based on a baseline activating *ERBB2* mutation, acquired this alteration by EOT<sup>77</sup>, suggesting it reflects a mechanism of resistance in patients with *ERBB2*-mutant positive breast cancer.

Data from the SUMMIT trial also demonstrated that of 22 assessable patients, 8 (36%) acquired an alteration (amplification or mutation) in *ERBB2*, of whom 7 derived clinical benefit from neratinib-containing therapy<sup>77</sup>. This high rate of *ERBB2* alteration acquisition is striking and may represent a mechanism of resistance.

Analysis of neratinib-resistant cell lines and patient-derived PDX models of neratinib resistance has suggested that enhanced TORC1 signalling is central to neratinib resistance<sup>153</sup>.

TORC1 integrates the proliferative signals from a number of pathways, including PIK3CA/AKT/mTOR and MAPK<sup>153</sup>, both of which are modulated by RAS. Knockdown of RAS (HRAS, NRAS and KRAS isoforms) led to reversal of neratinib resistance in drug-resistant cell lines, suggesting RAS-dependent TORC1 hyperactivity in neratinib resistance<sup>153</sup>. Neratinib resistance was overcome by concurrent application of TORC1 inhibitor everolimus<sup>153</sup>. The group also interrogated the baseline sequencing data of patients enrolled in SUMMIT, finding enrichment of activating mutations with the mTOR pathway in patients with *de novo* resistance to neratinib<sup>153</sup>. This was supported by cell culture data whereby MCF<sup>V777L</sup> cells with concurrent *PIK3CA* or *KRAS* mutations exhibited *de novo* resistance<sup>153</sup>.

#### 1.4.6. Cohort C: HR+ patients with *AKT1* mutation for treatment with capivasertib +/- fulvestrant

##### 1.4.6.1. *AKT1* alterations

The AKT kinases, which includes AKT1, AKT2 and AKT3, are key components of the PI3K/AKT pathway<sup>154</sup>. This pathway regulates many aspects of cell regulation including glucose metabolism, genome stability, growth, survival, proliferation and neo-vascularisation<sup>154</sup>. It is one of the most commonly activated pathways in cancer<sup>155</sup>, with loss of function mutations occurring in the negative regulators *PTEN* and *INPP4B*, and activating mutations/amplifications upstream in *ERBB2* and *FGFR1*, in *PIK3CA* (which encodes the p110 alpha catalytic subunit of PI3K), and its downstream effectors such as *AKT1* and *AKT2*<sup>155</sup>.

AKT1, the serine threonine kinase, contains an amino-terminal pleckstrin homology (PH) domain, a central catalytic domain, and a carboxy-terminal regulatory domain. PI3K activation, itself activated by the binding of ligands to receptor tyrosine kinases (RTKs), causes phosphorylation of phosphatidylinositol-3,4-bisphosphate (PIP<sub>2</sub>) to phosphatidylinositol-3,4,5-trisphosphate (PIP<sub>3</sub>) at the cell membrane (Figure 1.5). In turn, PIP<sub>3</sub> binds to the PH domain of AKT1 leading to recruitment of the kinase to the cell membrane<sup>156</sup>. Once at the cell membrane, the conformational change prompted by binding of PIP<sub>3</sub> to the PH domain allows phosphorylation of AKT1 at the regulatory amino acids Thr308 and Ser473 by the kinases PDK1 and PDK2 (containing mTORC2), respectively<sup>156</sup>, activating AKT1 downstream effectors.

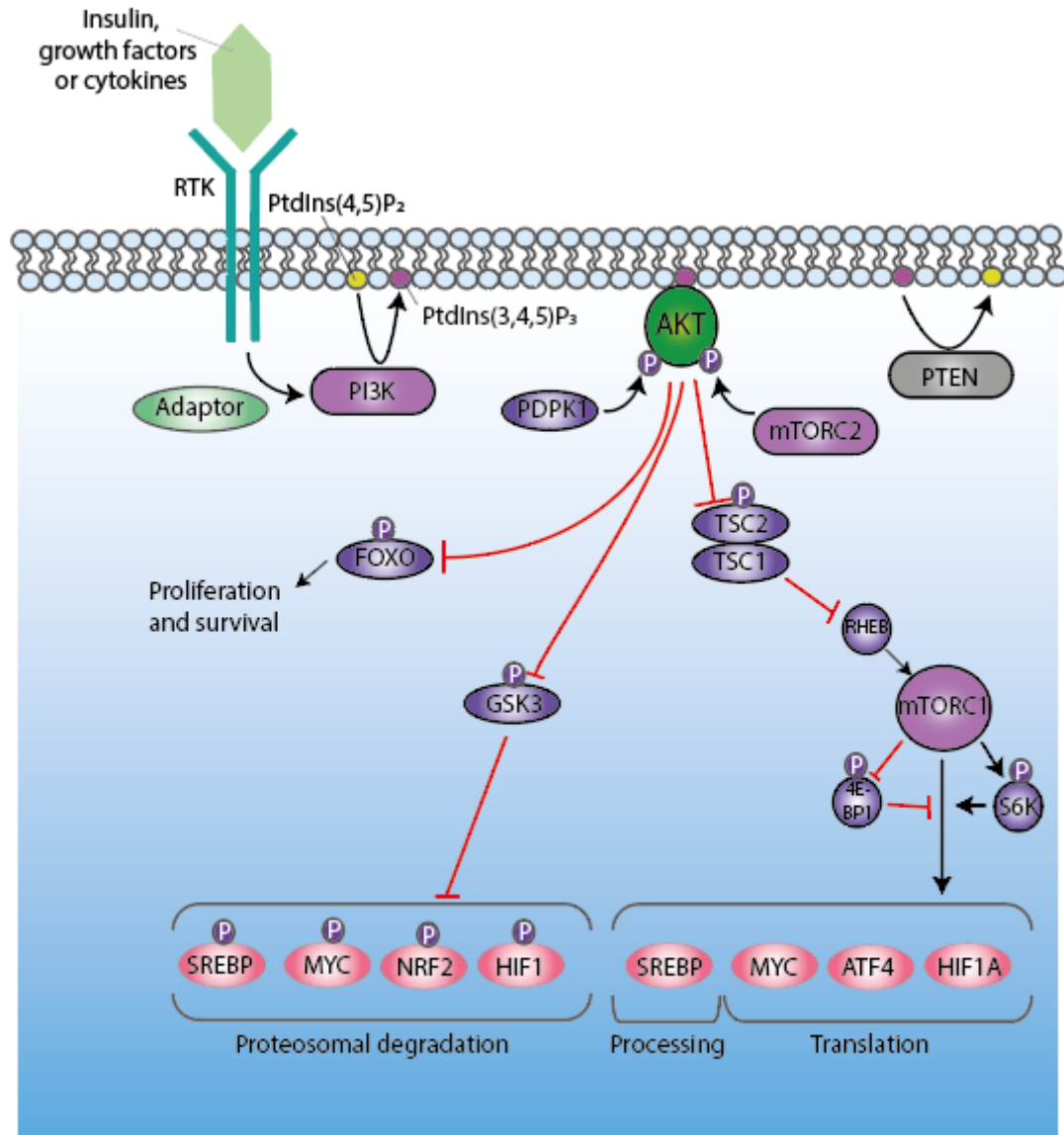


Figure 1.5. The PI3K/AKT signalling pathway. Ligand binding activates receptor tyrosine kinases (RTKs), leading to recruitment of PI3K to the lipid membrane. Here, PI3K phosphorylates phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P<sub>2</sub>, PIP<sub>2</sub>) to produce phosphatidylinositol 3,4,5-trisphosphate (PtdIns(3,4,5)P<sub>3</sub>, PIP<sub>3</sub>). PIP<sub>3</sub> then recruits AKT to the plasma membrane where it is phosphorylated by PDK1 and mTORC2 at the threonine and serine phosphorylation residues, respectively, activating the protein. Downstream signalling pathways are activated or inhibited, promoting metabolic changes, cell survival, growth and proliferation. Adapted from Hoxhaj *et al*, 2020<sup>157</sup>.

In 2007 Carpten *et al* identified the *AKT1* p.E17K mutation through sequencing of a range of tumour types, including breast cancer<sup>158</sup>. A G>A point mutation at nucleotide 49 causes substitution in amino acid from an acidic glutamic acid to leucine at amino acid 17. In the unbound state of wild type AKT1, the glutamic acid at amino acid 17 lies within the PIP<sub>2</sub>/PIP<sub>3</sub>



binding pocket (the PH domain), forming an ionic bond with the basic lysine at position 14. Alteration to lysine a position 17 prevents the formation of the ionic bond, leading to PI3K-independent activation. Overall, the E17K-mutant *AKT1* localises to the plasma membrane without stimulation, and has four-fold higher kinase activity than the wild type<sup>158</sup>. Sequencing studies have identified this alteration in around 4-6% of breast cancers<sup>159-161</sup>. The alteration was found to be mutually exclusive with *PIK3CA* alterations, suggesting that the alteration is singularly sufficient to activate the pathway<sup>160</sup>. Other activating *AKT1* alterations have been identified, albeit they are less common than the E17K hotspot<sup>162</sup>.

#### 1.4.6.2.Targeting of *AKT1* E17K with capivasertib

Capivasertib is an oral, potent, selective competitive ATP-inhibitor of all three isoforms of AKT (AKT1, -2 and -3), which has particular activity cancers with concurrent mutations/amplifications that would be anticipated to inhibit the PI3K/AKT pathway<sup>163-165</sup>. Phase I and II trials in patients with an *AKT1*-mutant positive breast cancer have demonstrated response rates of between 19% and 46%<sup>166-168</sup>.

Preclinical models have suggested that in ER+ breast cancer, inhibition of the PI3K/AKT pathway may lead to compensatory enhanced oestrogen-dependent signalling<sup>169</sup>. This prompted the investigation of combined blocking of the oestrogen dependent signalling and PI3K/AKT pathway signalling with fulvestrant and capivasertib. The phase II FAKTION trial demonstrated significantly improved PFS in patients who were treated with fulvestrant and capivasertib compared to those on with fulvestrant alone (10.3 months versus 4.8 months,  $p=0.0018$ )<sup>170</sup>, underlining a role for dual therapy in hormone positive breast cancer.

This background data provided the rational for cohort C in plasmaMATCH. Patients with HR+ disease found to have an activating *AKT1* mutation on ctDNA screening were eligible to enrol into cohort C, for treatment with capivasertib (400mg PO BD 4 days on, 3 days off within 28 day cycles) with the addition of fulvestrant (500mg IM on Cycle 1 Days 1 and 15 and Cycle 2 onwards Day 1).

#### 1.4.6.3.Predictive and prognostic biomarkers response to capivasertib

Two phase I/II randomised controlled trials have investigated the activity of capivasertib alongside paclitaxel in triple negative (TNBC) and HR+ breast cancer, respectively<sup>171,172</sup>. Whilst in HR+ breast cancer there was no there additional benefit of capivasertib in a *PIK3CA*-

mutant subpopulation<sup>171</sup>, patients with a *PIK3CA/AKT1/PTEN* mutation were demonstrated to gain additional benefit from capivasertib in TNBC<sup>172</sup>. The lack of benefit in *PIK3CA*-mutant HR+ breast cancer was suggested to be secondary to enhanced ER signalling prompted by PI3K inhibition<sup>169</sup>.

The phase II randomised FAKTION trial investigated the benefit of addition of capivasertib to fulvestrant in HR+ HER2- breast cancer, which in the overall population demonstrated that there was a significant improvement in median PFS from the combination. Interestingly, however, in the subgroup of patients with activating PI3K pathway alterations (including PTEN loss), this benefit was not significant<sup>170</sup>, arguing against enhanced activation of estrogen signalling as a mechanism of escape from capivasertib therapy. Thus while capivasertib appears to be more active in TNBC with PI3K pathway activation, the same does not appear to be true within HR+ breast cancer.

#### 1.4.6.4. Putative resistance mechanisms to capivasertib

Little is known of the mechanisms of resistance to capivasertib *in vivo*. *In vitro* data suggests that the SGK3 pathway, a PI3K-independent pathway that activates downstream mTORC1, becomes upregulated following prolonged treatment with an AKT inhibitor and could represent a mechanism of resistance<sup>173</sup>. However supportive patient-derived clinical data is lacking.

#### 1.4.7. Cohort D: Patients with AKT pathway activating mutations for treatment with capivasertib

##### 1.4.7.1. Activating AKT mutations

As discussed in section 1.4.6.1, mutations which enhance signalling via the PI3K-pathway are broad ranging, occurring at multiple points in the pathway. Mutations in *PTEN*, a negative regulator of the PI3K-pathway, have been identified in 9% of HR+ HER2- breast cancer and 15% of TNBC<sup>34</sup>. Inactivating mutations or homozygous deletion of *PTEN* lead to reduced degradation of PIP2 and PIP3 and subsequently enhance signalling via PI3K that is evident in a raised expression of p-AKT<sup>33,154,174</sup>. The clinical ramifications of this are particularly evident in patients with germline inactivating *PTEN* mutations, named Cowden's disease, a condition associated with increased lifetime risk of neoplasms<sup>175</sup>.

This background underlies the rationale for cohort D in plasmaMATCH. Patients with mutations expected to activate signalling via AKT were eligible, including:

- *AKT1* mutations identified in ctDNA in HR- breast cancer.
- *AKT1* mutations identified tumour sequencing in patients with HR+ or HR- breast cancer.
- *AKT2/3* E17K, *PIK3R1* or *PTEN* mutations or homozygous deletion of *PTEN* in ER positive or negative breast cancer identified in ctDNA or tumour sequencing.

Enrolled patients were treated with capivasertib 480mg PO BD 4 days on/3 days off within a 28-day cycle.

#### 1.4.7.2. Targeting of AKT-activating alterations with capivasertib

Several studies have investigated the use of capivasertib in patients with mutations anticipated to activate signalling via the PI3K pathway outside of *AKT1* p.E17K, with generally poor results. In one phase I study in patients with *PTEN*-mutant HR+ breast cancer, the response rate was low at 8% in the fulvestrant-naïve patients and 21% in fulvestrant pre-treated patients<sup>176</sup>. The authors noted that there was no apparent relationship between the type of *PTEN* alteration and clinical response<sup>176</sup>. A further phase I study of capivasertib monotherapy in patients with *PIK3CA*-mutated breast (HR+ or HER2+) and gynaecologic tumours failed to reach the required 20% response rate at the interim assessment, and recruitment was halted early<sup>165</sup>. Moreover, as discussed in section 1.4.6.3, patients with HR+ breast cancer with concomitant PI3K-pathway alterations do not appear to gain additional benefit from capivasertib over those wild type for PI3K pathway alterations within phase II trials, regardless of whether endocrine therapy was also given<sup>170,171</sup>.

In TNBC, there appears to be additional benefit of capivasertib in PI3K-pathway altered tumours. The randomised phase II PAKT trial demonstrated a median PFS of 5.5 months in patients with *PIK3CA/AKT1/PTEN*-non-altered tumours compared to 9.3 months in patients with *PIK3CA/AKT1/PTEN*-altered tumours<sup>172</sup>.

## 1.5. Summary

CtDNA analysis has great potential in the management of ABC in improving patient outcomes. The prospective applications of ctDNA analysis are wide, from the identification of targetable alterations, to investigation of predictive and prognostic biomarkers and delineation of the breast cancer mutational landscape and clonal architecture. The plasmaMATCH trial represents a novel trial design that creates an opportunity to not only validate the use of ctDNA

analysis, but also investigate these applications within a large number of patients and in a range of targeted therapies.

## 2. Chapter 2. Methods and Materials

### 2.1. Methods and materials within the plasmaMATCH study

#### 2.1.1. Study design

PlasmaMATCH (NCT03182634) was a multiple parallel cohort, open label, multi-centre, phase IIa study which aimed to establish the activity of targeted agents when the treatment target was identified using ctDNA testing<sup>177</sup>. Eligibility criteria included women over 18 years with measurable metastatic or recurrent locally advanced breast cancer, which had progressed within the previous 6 weeks. Patients must have received at least one line of therapy in the metastatic setting and/or relapsed within 12 months of completing (neo)adjuvant chemotherapy, and not received more than 2 lines of cytotoxic chemotherapy in the metastatic setting.

Eligible patients underwent ctDNA testing to identify a number of targetable mutations (Figure 2.1A), with positive mutation status making a patient eligible to enter a respective treatment cohort for treatment with a matched targeted therapy (Figure 2.1B). ctDNA testing was undertaken using two orthogonal approaches: droplet digital PCR (ddPCR) and error-corrected targeted sequencing (Guardant360, Guardant Health, USA). Mutation status from ddPCR allowed entry into cohorts A to E, while results from either ddPCR or targeted sequencing allowed entry into cohorts B to E. Alongside plasma ctDNA testing results, patients could also enter cohort D with mutation results from prior tissue sequencing or immunohistochemistry undertaken outside of the trial.

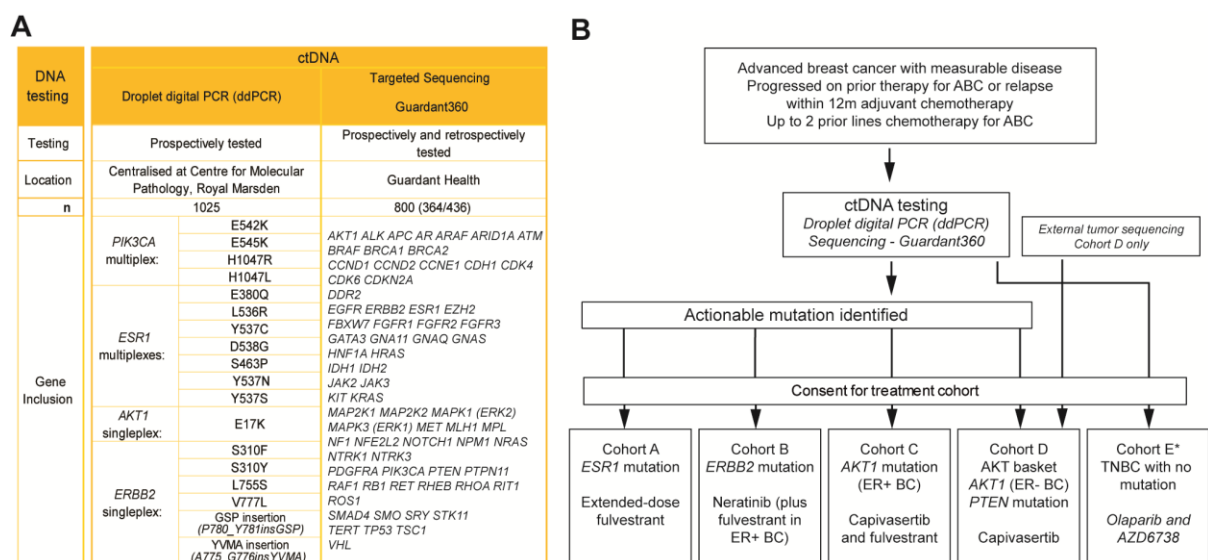


Figure 2.1. ctDNA analysis approaches and trial schema of plasmaMATCH. A, circulating tumour DNA (ctDNA) testing for patients in plasmaMATCH. Droplet digital PCR (ddPCR) tested for hotspot mutations in *ESR1*, *ERBB2* (*HER2*), *PIK3CA* and *AKT1* while targeted sequencing employed a duplex targeted sequencing 73/74 gene panel (Guardant360). B, Advanced breast cancer patients with disease progression after prior therapy, and following consent, were offered ctDNA testing by droplet digital PCR (ddPCR) for the presence of actionable mutations in *ESR1*, *ERBB2* (*HER2*), *PIK3CA* and *AKT1* or targeted sequencing. Patients were enrolled into one of the five interventional arms based on the mutations detected in plasma, the clinical results of which have been reported<sup>98</sup>. Cohort E is currently ongoing.

Prospective targeted sequencing was commenced partway through the trial, following a protocol amendment. For patients recruited before prospective targeted sequencing, remaining banked baseline plasma (remaining plasma from ddPCR ctDNA testing, or a pre-treatment CID1 sample) were shipped for targeted sequencing. Targeted sequencing results for patients with a retrospectively sequenced baseline sample were for research only and did not influence cohort eligibility.

Between 21<sup>st</sup> December 2016 and 26<sup>th</sup> April 2019, 1051 patients were registered into the study, of which 1044 entered via ctDNA testing and seven by prior tumour analysis. Cohorts A-D recruited to target with 84, 21, 18 and 19 patients enrolled in each cohort, respectively. Cohort E was opened later and will be reported at a later date, and does not form part of this thesis.

## 2.1.2. Sample collection

### 2.1.2.1. ctDNA screening sample collection and shipment

Eligible patients who had given written informed consent had 30 – 40 ml blood collected in 3–4 10ml cell-free DNA BCT Streck tubes. 30ml of the blood was shipped at ambient temperature to the central laboratory (Centre for Molecular Pathology, Royal Marsden Hospital) where it underwent plasma isolation, DNA extraction and ddPCR testing. Any remaining plasma was stored at -80°C prior to shipment on dry ice to the research laboratory (Institute of Cancer Research, London, UK). The remaining 10ml of blood was, following protocol amendment, shipped to Guardant Health (Redwood City, California, USA) to undergo targeted sequencing.

Blood samples shipped to the central laboratory were centrifuged at 1,600g for 10 minutes. The resulting plasma was isolated and underwent a second centrifuge at 1,600g for 10 minutes prior to aliquoting and storage at -80°C until further analysis. Blood samples shipped to Guardant Health were centrifuged at 1,600g for 10 minutes, and the resulting plasma further centrifuged at 3,220g for 10 minutes prior to aliquoting and storage at -80°C.

### 2.1.2.2. Baseline tissue sample collection and shipment

Patients enrolled into a treatment cohort within plasmaMATCH were required to provide a baseline pre-treatment tissue biopsy specimen. Fresh tumour biopsies of metastatic tissue were obtained under image guidance, where necessary. A 14G biopsy was used to obtain four research cores, two of which were placed in formalin containing 10% neutral buffered formalin before paraffin embedding, while the remaining two cores were frozen on dry ice prior and stored at -80°C until batched shipment to the research laboratory.

Where biopsy was contraindicated or not possible, an archival metastatic biopsy could be provided as an FFPE block or as five freshly prepared unstained 10 micron sections. Samples were shipped by local hospitals to the research laboratory at ambient temperature.

### 2.1.2.3. On-treatment blood sample collection and shipment

Patients enrolled into plasmaMATCH treatment cohorts underwent blood collection prior to each cycle of treatment and at the time of treatment discontinuation (end of treatment, EOT sample). Blood was collected in 2 x 10mL BD Vacutainer® EDTA tubes, inverted 8 to 10 times, and centrifuged within 1 hour of collection (4°C in a horizontal rotator for 2 minutes at

1600 RCF/2700 RPM). The resulting plasma was aliquoted into 4.5mL cryotubes and frozen at -80°C prior to batched shipment on dry ice to the research laboratory.

#### 2.1.2.4. Retrospective shipment of plasma to Guardant Health for targeted sequencing

For patients retrospectively sequenced within the ctDNA testing component of the trial, and for all in-cohort patients with available plasma at the C2D1 and EOT time points, plasma was shipped from the research laboratory to Guardant Health for targeted sequencing. The aim was to sequence one baseline sample (ctDNA testing or C1D1 time point), one C2D1 sample and one EOT sample for each in-cohort patient. The C1D1 time point was preferentially chosen as the baseline sample where there was available plasma. To prepare the plasma for shipment, plasma stored at the research laboratory was thawed at room temperature. Tubes were mixed by inversion before 2ml was aliquoted into a separate 4.5ml cryovial. Aliquots were stored at -80°C until batched shipment on dry ice to Guardant Health, USA.

#### 2.1.3. Plasma DNA extraction

##### 2.1.3.1. ctDNA screening samples at the CMP

DNA was extracted from 4 or 8mL of screening plasma using the automated QiaSymphony platform (Qiagen, Hilden, Germany) following the manufacturer's instructions, and eluted into 60 or 120uL Buffer, respectively. DNA was stored at -80C until further use.

##### 2.1.3.2. Samples processed at the research laboratory

Plasma underwent manual DNA extraction using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden) as per manufacturer's protocol as follows. Plasma was thawed at room temperature, and between 1 and 5ml aliquoted into a 50 ml centrifuge tube for extraction depending on downstream process requirements. Any samples with less than the needed volume of plasma were topped up to the required extraction volume with PBS (phosphate buffered saline). Buffer ACL and Carrier RNA (stock = 0.2ug/uL, dissolved in Buffer AVE) were combined in accordance to Figure 2.2 for the respective number of samples being extracted and mixed by inversion 10 times.



Number of samples	Buffer ACL (ml)					Carrier RNA in Buffer AVE (µl)
	1ml	2ml	3ml	4ml	5ml	
1	0.9	1.8	2.6	3.5	4.4	5.6
2	1.8	3.5	5.3	7.0	8.8	11.3
3	2.6	5.3	7.9	10.6	13.2	16.9
4	3.5	7.0	10.6	14.1	17.6	22.5
5	4.4	8.8	13.2	17.6	22.0	28.1
6	5.3	10.6	15.8	21.1	26.4	33.8
7	6.2	12.3	18.5	24.6	30.8	39.4
8	7.0	14.1	21.1	28.2	35.2	45.0
9	7.9	15.8	23.8	31.7	39.6	50.6
10	8.8	17.6	26.4	35.2	44.0	56.3
11	9.7	19.4	29.0	38.7	48.4	61.9
12	10.6	21.1	31.7	42.2	52.8	67.5
13	11.4	22.9	34.3	45.8	57.2	73.1
14	12.3	24.6	37.0	49.3	61.6	78.8
15	13.2	26.4	39.6	52.8	66.0	84.4
16	14.1	28.2	42.2	56.3	70.4	90.0
17	15.0	29.9	44.9	59.8	74.8	95.6
18	15.8	31.7	47.5	63.4	79.2	101.3
19	16.7	33.4	50.2	66.9	83.6	106.9
20	17.6	35.2	52.8	70.4	88.0	112.5
21	18.5	37.0	55.4	73.9	92.4	118.1
22	19.4	38.7	58.1	77.4	96.8	123.8
23	20.2	40.5	60.7	81.0	101.2	129.4
24	21.1	42.2	63.4	84.5	105.6	135.0

Figure 2.2. Volumes of Buffer ACL and carrier RNA required for processing 1ml, 2ml, 3ml, 4ml or 5ml plasma volume. Adapted from laboratory SOP 008: Manual extraction of cell free DNA from plasma.

The respective volume of Proteinase K and Buffer ACL/Carrier RNA, according to the volume of plasma extracted (Figure 2.3), was added to the plasma before vortexing the sample for 30 seconds.

Plasma (ml)	1	2	3	4	5
Proteinase K ul	100	200	300	400	500
ACL + CRNA mix (ml)	0.8	1.6	2.4	3.2	4
ACB (ml)	1.8	3.6	5.4	7.2	9

Figure 2.3. Reagent volumes required for extracting 1-5ml plasma . Adapted from laboratory SOP 008: Manual extraction of cell free DNA from plasma.

The sample was incubated in a water bath at 60°C for 30 minutes. Buffer ACB (Table 2.3) was added, the sample vortexed for 15-30 seconds, and incubated on ice for 5 minutes. QIAamp Mini columns were attached to VacConnectors on a QIAvac 24 Plus manifold. The sample was

added to the QIAamp Mini column and negative pressure applied to draw the sample through the column, collecting the DNA in the Mini column membrane. The column membrane was washed sequentially with Buffers ACW1 (600uL), ACW2 (750uL) and 96-100% ethanol (750uL) with negative pressure applied to draw each through the column membrane. The column was then removed from the manifold, centrifuged at 14000rpm for 3 minutes followed by incubation at 56°C for 3 minutes to dry the membrane. DNA was eluted by application of 25uL Buffer AVE to the membrane, incubation for 3 minutes at room temperature, and centrifugation at 14000rpm for 1 minute to collect the eluate. The last step was repeated to produce two 25uL DNA elutions that were transferred to barcoded tubes and stored at -20°C.

#### 2.1.4. Tissue DNA extraction

##### 2.1.4.1. FFPE tissue samples

FFPE tissue blocks and slides each had a haematoxylin and eosin–stained (H&E) slide prepared which a pathologist reviewed and annotated for tumour content. Tissue sections satisfying the criteria of a minimum of 20% tumour content underwent DNA extraction. Slides were cut from FFPE tissue blocks with the aim of extracting tumour tissue DNA and RNA from a total area of 100 mm<sup>2</sup> and stained with Nuclear Fast Red (NFR). Following the Allprep DNA/RNA kit protocol (Qiagen, Hilden, Germany), the NFR stained slides were needle microdissected to isolate the tumour-containing tissue and transferred to a microcentrifuge tube containing 10µL proteinase K, the tube vortexed and incubated at 56°C for 15 minutes and then on ice for 3 minutes. The tube was centrifuged at 20,000xg for 15 minutes to separate the supernatant from the pellet.

To extract RNA from the supernatant the supernatant was incubated at 80°C for 15 minutes before 320µL Buffer RLT and 1120µL 96-100% ethanol were added and the sample mixed by vortexing. The sample was added in 600µL increments to an RNeasy MinElute spin column and centrifuged to collect the RNA in the spin column until all the supernatant had passed through. The flow-through was discarded. The spin column was rinsed with 350µL Buffer FRN by centrifuging for 15 seconds at 10,000rpm and the flow-through discarded. For each sample, 10µL DNase I was combined with 70µL Buffer RDD in a microcentrifuge tube and mixed by inversion before 80µL was added directly to the spin column membrane. The spin column membrane was incubated at room temperature for 15 minutes. 500µL Buffer FRN was then added to the spin column and the column centrifuged for 15 seconds at 10,000rpm to wash the

Buffer through the membrane. The flow-through was re-applied to the membrane and centrifuged again for 15s at 10,000rpm and the flow-through discarded. The membrane was washed twice by the addition of 500µL Buffer RPE followed by centrifuging and discarding of the flow-through. The spin column was placed in a new 1.5ml Lo-bind DNA microcentrifuge tube. RNA was eluted by the addition of 30µL RNase-free water to the spin column membrane, incubation for 1 minute at room temperature and a centrifuge at full speed for 1 minute to collect the flow through. The eluate containing RNA was transferred to a barcoded tube and stored at -80 °C.

To extract the DNA from the pellet, the pellet was re-suspended in 180µL Buffer ATL and 40µL proteinase K, mixed by vortexing and then incubated overnight on a heat block at 56 °C with interval mixing (700rpm for 10 seconds every 90 minutes). The sample was then incubated on a heat block to 90 °C for 2 hours. Once cooled, 2µL RNase A was added and the sample incubated for 2 minutes at room temperature. 200µL Buffer AL was added to the sample, it was mixed by vortexing, following which 200µL ethanol (96 – 100%) was added, and the sample mixed by vortexing and centrifuged. The lysate is passed through a QIAamp MiniElute column by centrifuging at 8000rpm for 1 minute, with the DNA collected in the column membrane. The membrane is washed by adding and centrifuging sequentially 500µL Buffers AW1 and AW2, respectively, and the flow-through discarded each time. The spin column membrane was then dried completely by centrifuging at 13,000rpm for 5 minutes. DNA was eluted by the addition of 50µL of pre-warmed (42 °C) Buffer ATE, centrifuging and collection of the eluate. This process was repeated for a second elution, and the eluates transferred into barcoded tubes and stored at -20 °C.

#### 2.1.4.2. Fresh frozen tissue samples

The fresh frozen tissue had an H&E slide cut which a pathologist reviewed and annotated for tumour content. Tissues with a minimum of 20% tumour content were cut in 10µm slices on slides, with slides kept on dry ice until microdissection. Following the Allprep DNA/RNA Mini Kit protocol (Qiagen, Hilden, Germany), non-tumour tissue was removed and the remaining tumour tissue dissolved in 350µL Buffer RLT and transferred into QIAshredder spin column before centrifuging for 2 minutes at maximum speed to disrupt and homogenise the sample. The resulting sample was transferred to an AllPrep DNA spin column and centrifuged for 30 seconds at 10,000rpm. The spin column containing DNA was stored at 4 °C while the flow-through was processed to extract RNA.

For RNA extraction, 350µL 70% ethanol was added to the flow-through, pipetted to mix, and the sample transferred to an RNeasy spin column. The spin column was centrifuged to collect the RNA in the spin column membrane and the flow through discarded. 350µL Buffer RW1 was added to the column and passed through by centrifuging for 15 seconds at 10,000 rpm. 10µL DNase stock solution was added to 70 µL Buffer RDD per sample, and this was added to the spin column membrane and incubated at room temperature for 15 minutes. The membrane was then rinsed again with 350µL Buffer RW1 by centrifuging for 15 seconds at 10,000 rpm and the flow-through discarded. The membrane was then rinsed twice with Buffer RPE by centrifuging for 15 seconds and then 2 minutes, respectively, at 10,000 rpm with the flow-through discarded each time. The spin column membrane was dried fully by centrifuging at full speed for 1 minute. The spin column was placed in a new collection tube and RNA eluted by the addition of 30µL RNase-free water to the spin column membrane and centrifuging to elute the RNA. RNA was stored at -80 °C in barcoded fluidX tubes.

Genomic DNA was purified by sequentially washing the membrane with Buffers AW1 (500µL) and AW2 (500µL) by centrifuging and discarding of the flow-through. The spin column was placed in a new collection tube, and the DNA eluted by the sequential addition of 100µL Buffer EB, incubation for 1 minute at room temperature, and the sample centrifuged to produce two 100µL eluates. DNA was stored at -20 °C in barcoded tubes.

#### 2.1.5. Quantification of DNA and PCR products

##### 2.1.5.1. DNA quantification by the CMP

Extracted DNA was quantified using the Qubit High Sensitivity kit (Thermo Fisher Scientific, Waltham, USA) following the manufacturer's instructions.

##### 2.1.5.2. DNA and PCR products extracted at the research laboratory

DNA extracted from plasma, tissue and library PCR products were quantified using a Qubit dsDNA HS Assay Kit and Qubit® 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA) as per the manufacturer's instructions. The Qubit Fluorometer detects the absorbance of DNA bound fluorescent dye, allowing quantification of the DNA with reference to a reference range. The reference range was assessed prior to sample quantification by the addition of 10µL of standard #1 and standard #2 respectively to 190µL of working solution (1:200 dilution of dsDNA HS reagent to dsDNA HS Buffer). The standard/working solution mixes were vortexed

and incubated for 2 minutes at room temperature prior to measurement on the Qubit Fluorometer. To establish the concentration of samples, 1 $\mu$ L of the sample was combined with 199 $\mu$ L of working solution, vortexed and incubated at room temperature for 2 minutes before being quantified using the Qubit Fluorometer.

DNA extracted from plasma was alternatively quantified using ddPCR, with reference to single copy genes with consistently diploid copy number in breast cancer datasets. The original ddPCR quantification assay consisted of a 20X concentration of primers for the *RNaseP* gene and VIC dye-labelled probes (RNaseP assay). Later the assay was updated to contain three reference genes with VIC or FAM dye-labelled probes (*HOGA1*, *IRAK4*, *OR4C12*, ‘triplex’ assay) to improve confidence in the quantification by reducing the risk of the effect of copy number alteration events or acquisition of SNPs on the primer/probe region thus making the assay inaccurate.

The primer-probe assay (RNaseP or triplex) were thawed to room temperature. Master mixes containing assay, supermix and nuclease free water per reaction were combined (Table 2.1). The master mix was pipetted into wells containing the DNA to be quantified and mixed by vortexing. The samples were emulsified with 70 $\mu$ L droplet generation oil (Bio-Rad, Pleasanton CA) either manually using a QX200 manual droplet generator (Bio-Rad, Pleasanton, CA) or on a QX200 AutoDG<sup>TM</sup> Droplet Digital<sup>TM</sup> PCR System (Bio-Rad, Pleasanton, CA) for larger sample sets. The emulsification process partitions the mastermix and DNA molecules into approximately 14,000 micelles, effectively isolating single-unit PCR reactions. The emulsified samples are transferred into a 96-well plate and foil sealed. The samples then underwent PCR on a G-Storm Quad thermocycler (software version 3.3.0.0) with cycling conditions as described in Table 2.1.

Reagent	RNase P	Triplex Assay
	Volume per reaction ( $\mu$ L)	Volume per reaction ( $\mu$ L)
Primer Probe Mix	1.1	4.5
Droplet digital PCR supermix for	11	10
Nuclease free water	8.8	5.5
Total Volume	20.9	20
DNA template to be added	1.1	1

Thermocycling Conditions	105 °C lid microplate pressure	
	95 °C 10 minutes	
	95 °C 15 seconds	Cycled x40
	60 °C 1 minute	

	98 °C 10 minutes	
--	------------------	--

Table 2.1. RNase P and triplex assays . *Top*, Mastermix contents for the RNase P assay (left) and triplex assay (right). *Bottom*, thermocycling conditions for RNaseP and triplex assays.

Following PCR, the plates were read on a Bio-Rad QX-200 droplet reader. Data was analysed with QuantaSoft software (version v1.7.4) (Bio-Rad, Pleasanton, CA). The software plots channel 1 (FAM) and 2 (VIC) amplitudes of the droplets on an x- and y-axis, depending on the signal strength of the quenched fluorophores contained within each droplet. The droplet populations can then be identified visually, gated and quantified. The relative proportions of the droplets, which correspond directly to whether the droplet contains wild type, mutant, combination of wild type/mutant, or empty of DNA, are used to identify mutation status and infer allele frequency of positive mutations through Poisson regression. In the case of the quantitative assays (RNase P or triplex), the number of wild type probe positive droplets relative to empty droplets allows inference of the sample concentration, using the following formulas:

$$\text{Concentration (ng/uL)} = \frac{\text{Copies per 20uL well} * 3.3}{1000}$$

#### 2.1.6. Droplet digital PCR

##### 2.1.6.1. ctDNA screening for plasmaMATCH at CMP

A 0.5ml plasma-equivalent volume of extracted DNA was tested for hotspot mutation status using multiplex and singleplex ddPCR assays (Figure 2.1 and Table 2.2).

Gene	Variant	Single or Multiplex	Assay
<i>PIK3CA</i>	E542K (1624G>A)	Multiplex	
	E545K (1633G>A)	Multiplex	
	H1047R (3140A>G)	Multiplex	
	H1047L (3140A>T)	Multiplex	
<i>ESR1</i>	E380Q (c.1138G>C),	Multiplex	dHsaMDXE91450042
	L536R (c.1607T>G)	Multiplex	dHsaMDXE91450042
	Y537C (c.1610A>G)	Multiplex	dHsaMDXE91450042
	D538G (c.1613A>G)	Multiplex	dHsaMDXE91450042
	S463P (c.1387T>C)	Multiplex	dHsaMDXE65719815
	Y537N (c.1609T>A)	Multiplex	dHsaMDXE65719815
	Y537S (c.1610A>C)	Multiplex	dHsaMDXE65719815
<i>AKT1</i>	E17K (c.49G>A)	Singleplex	
<i>ERBB2</i>	S310F (c.929C>T)	Singleplex	
	L755S (c.2264T>C)	Singleplex	

	V777L (c.2329G>T)	Singleplex	
	P780_Y781insGSP	Singleplex	
	A775_G776insYVMA	Singleplex	

Table 2.2. DdPCR assays used within plasmaMATCH to screen patients for targetable mutations.

The DNA was combined with the respective assay, supermix and nuclease free water (Table 2.3) prior to droplet generation using an Automated Droplet Generator (QX200 AutoDG). The samples underwent 40 cycles of PCR on a thermal cycler (C1000 Touch Thermal Cycler, BioRad, Hercules, CA) prior to analysis on a QX200 Digital Reader (BioRad, Pleasanton, CA). The following criteria needed to be met to satisfy quality control:

- a minimum of 10,000 total droplets per well
- $\geq 2$  positive FAM droplets (mutant) per well
- $>300$  VIC droplets (wild type) per well

Positive samples were repeated, with two separate assays confirming a positive result. For the purposes of the trial screening result, two independent investigators verified a binary positive or negative ddPCR result. Outside of the trial and within this work, the exact mutation variant present (from the multiplex assays) were ascertained with reference to reference positive control plates based on the typical channel 1 and 2 amplitudes for the respective mutations, while the allele frequencies were calculated from the ratio mutant to wild type droplets using Poisson regression.

<b><i>PIK3CA</i> multiplex</b>	<b>Volume per reaction (uL)</b>
ddPCR Supermix for Probes	10
<i>PIK3CA</i> mplx Exon9 assay	0.5
<i>PIK3CA</i> mplx Exon20 assay	0.75
RNase-free Water	2.25
DNA	7.5
<b>Total</b>	<b>21</b>

<b><i>AKT1</i> assay</b>	<b>Volume per reaction (uL)</b>
ddPCR Supermix for Probes (no UTP)	10
<i>AKT1</i> E17K assay FAM	1
<i>AKT1</i> WT assay HEX	1
RNase-free Water	1.5
DNA	7.5
<b>Total</b>	<b>21</b>

<b><i>ERBB2 S310X</i></b>	<b>Volume per reaction (µl)</b>
ddPCR Supermix for Probes	10
<i>ERBB2 S310F</i> assay mix	1
RNase-free Water	2.5
DNA	7.5
<b>Total</b>	<b>21.0</b>

<b><i>ERBB2 L755S</i></b>	<b>Volume per reaction (µl)</b>
ddPCR Supermix for Probes	10
<i>ERBB2 L755S</i> assay mix	1
RNase-free Water	2.5
DNA	7.5
<b>Total</b>	<b>21.0</b>

<b><i>ERBB2 V777L</i></b>	<b>Volume per reaction (µl)</b>
ddPCR Supermix for Probes	10
<i>ERBB2 V777L</i> assay mix	1
RNase-free Water	2.5
DNA	7.5
<b>Total</b>	<b>21.0</b>

<b><i>ESRI multiplex 1</i></b>	<b>Volume per reaction (µl)</b>
ddPCR Supermix for Probes	10
<i>ESRI mplx1</i> assay	1
RNase-free Water	2.5
DNA	7.5
<b>Total</b>	<b>21</b>

<b><i>ESRI multiplex 2</i></b>	<b>Volume per reaction (µl)</b>
ddPCR Supermix for Probes	10
<i>ESRI mplx2</i> assay	1
RNase-free Water	2.5
DNA	7.5
<b>Total</b>	<b>21</b>

Table 2.3. Master mixes for the respective assays employed within plasmaMATCH screening.



### 2.1.6.2.ddPCR mutation validation in the research laboratory

Following the same principle as described in section 2.1.5.2, assays containing mutation specific primers and probes were combined with supermix, nuclease free water and DNA. The same procedure for ddPCR as described in section 2.1.6.1 was followed, but with themocycling conditions specific to the assay (Table 2.4). A single assay result demonstrating a positive result, and meeting the minimum quality criteria described in section 2.1.5.2 was required to identify a mutation as present or absent.

Thermocycling Conditions	105 °C lid microplate pressure	
	95 °C 10 minutes	
	95 °C 15 seconds	Cycled x40
	*54 °C 1 minute/+52 °C 1 minute	
	98 °C 10 minutes	

Table 2.4. Thermocycling conditions for mutation validation \* for multiplex assay containing *PIK3CA* E542K, *PIK3CA* E545K, *PIK3CA* H1047L, *PIK3CA* H1047R, +for multiplex assays containing *PIK3CA* Q546K, *PIK3CA* N345K, *PIK3CA* C420R, *PIK3CA* R88Q, *ERBB2* L755S, *ERBB2* V777L, *ERBB2* V842I, *ERBB2* S310Y, *ERBB2* G776V, *ERBB2* R678Q, *ERBB2* D769Y, *ERBB2* I767M.

### 2.1.7.Tissue DNA library preparation and sequencing

Tissue sequencing was undertaken using a custom-designed amplicon-based targeted panel, BCPv10. This panel includes genes and genome regions which are commonly mutated in breast cancer (Table 2.5). For each DNA sample 5ng DNA was pipetted into two 0.2ml PCR tubes or two wells of a 96 well plate. As the BCPv10 panel consists of two primer pools, two wells of 5ng DNA per sample is required for the first PCR step in the protocol. All DNA samples within the batch were made up to the same volume with the addition of nuclease free water. The DNA was concentrated by drying on a PCR machine set to 60 °C with the lid off and resuspended to the required volume with nuclease free water.

Gene Exons	Genome Regions
<i>CDH1</i>	<i>AKT1</i>
<i>GATA3</i>	<i>BRAF</i>
<i>MAP2K4</i>	<i>ERBB2</i>
<i>MAP3K1</i>	<i>ESR1</i>
<i>NF1</i>	<i>KIT</i>
	<i>KRAS</i>
	<i>PIK3CA</i>
	<i>PIK3R1</i>
	<i>RUNX1</i>
	<i>SF3B1</i>

	<i>TP53</i>
--	-------------

Table 2.5. BCPv10 breast cancer panel gene and genome region inclusion

5X AmpliSeq HiFi Mix (Ion AmpliSeq Library Kit 2.0, Life Technologies, Paisley, UK) and the primer pools were thawed on ice. Master mixes for primer pools 1 and 2, respectively, for the number of samples ( $n$ ) being prepared were made with the combination of 5X AmpliSeq HiFi Mix, the respective primer pool and nuclease free water according to Table 2.6. The master mixes were mixed by pipetting up and down 5 times before 4 $\mu$ L was pipetted into each respective well. The wells were covered, vortexed for 10 seconds and centrifuged at 1000rpm for 20 seconds.

Component	Volume for 1 well ( $\mu$ L)	Volume for $n$ wells
5X Ion AmpliSeq HiFi Mix	0.8	0.8 * $n$
2X Ion Primer Pool (Pool 1 or 2)	2	2 * $n$
Nuclease free water	1.2	1.2 * $n$
<b>Total Volume</b>	<b>4</b>	

Table 2.6. Volume of components to be combined into a master mix prior to PCR amplification

The reactions were run in a thermocycler (G-Storm Quad thermocycler, software version 3.3.0.0), undergoing PCR amplification with the cycling conditions as described in Table 2.7.

Stage	Step	Temp	Time
Hold	Activate enzyme	99°C	2 minutes
Cycle x20	Denature	99°C	15 seconds
	Anneal and extend	60°C	4 min
Hold	-	10°C	Up to 1 hour

Table 2.7. Thermocycling conditions for the first PCR amplification in the tissue sequencing protocol

The reactions were removed from the thermocycler and centrifuged at 1000rpm for 20 seconds to collect any droplets. The primer pools for each sample were combined into one well, the wells sealed, vortexed for 20 seconds and centrifuged at 1000rpm for 20 seconds. The primer sequenced were partially digested by the addition of 0.8 $\mu$ L FuPa Reagent (Ion AmpliSeq Library Kit 2.0, Life Technologies, Paisley, UK). Samples were mixed by pipetting 5 times. The wells were sealed and placed in the thermocycler with the cycling conditions as described in Table 2.8.

	Temperature	Time
Step 1	50°C	10 minutes
Step 2	55°C	10 minutes
Step 3	60°C	20 minutes

Step 4	10°C	Hold
--------	------	------

Table 2.8. Thermocycling conditions for primer digestion with FuPa.

The libraries next underwent the first of three bead ‘clean-ups’, which selects the correct size DNA libraries from the samples using magnetic beads to bind the charged library particles which are >100bp. AMPure XP Beads (Beckman Coulter, Wycombe, UK) are equilibrated to room temperature and vortexed to mix. The libraries were centrifuged at 1000rpm for 10 seconds to collect droplets, before 17.6µL AMPure XP Beads were added and the libraries mixed by pipetting 5 times. Libraries were incubated for 5 minutes at room temperature before being placed in a DynaMag 96 Side Magnet (Life Technologies, Paisley, UK) and incubated for a further 6 minutes. The supernatant was removed and discarded by pipetting, leaving the beads in the wells. 150µL of freshly prepared 70% ethanol was added to each well, the wells removed from the Magnet and the samples mixed by pipetting. The reactions were returned to the Magnet, incubated for 1 minute, and the supernatant removed and discarded by pipetting. The ethanol wash was repeated so that the beads underwent two ethanol washes. The wells containing the beads were dried at 37 °C on a thermocycler until the point that cracks could be seen in the bead deposits within each well. 21µL of low EDTA TE was added to each well and the sample mixed thoroughly by pipetting to homogenise the sample. The libraries were incubated for 3 minutes at room temperature before being placed back into the Magnet and incubated for 1 minute. 20µL of the supernatant from each well, which contains the libraries, was transferred into new wells by pipetting, leaving the beads undisturbed.

The DNA libraries next underwent end-repair, which phosphorylates the DNA ends and adds an adenosine residue to the 3’ end. For this, 1.2µL NEBNext ultra II End Prep Enzyme Mix (NEBNext® ultra™ II DNA Library Prep Kit for Illumina, New England Biolabs) and 2.8µL NEBNext ultra II End Prep Reaction Mix (NEBNext® ultra™ II DNA Library Prep Kit for Illumina) were added to each library and mixed by pipetting 10 times. The libraries were centrifuged at 1000rpm for 10 seconds to collect droplets, and placed in a thermocycler with the cycling conditions as described in Table 2.9.

	Temperature	Time
Step 1	20°C	30 minutes
Step 2	65°C	30 minutes
Step 3	10°C	Hold

Table 2.9. Thermocycling conditions for NEBNext DNA end preparation

The DNA libraries next underwent adaptor ligation. For this, a 1:10 dilution of the adaptors was prepared with 1 part Adaptor (NEBNext® ultra™ II DNA Library Prep Kit for Illumina) to 9 parts 10mM Tris-HCl, pH 8.0 with 10mM NaCl. For each library, 12µL of Ligation Master Mix (NEBNext® ultra™ II DNA Library Prep Kit for Illumina) and 0.4µL Ligation Enhancer (NEBNext® ultra™ II DNA Library Prep Kit for Illumina) were combined in a master mix and mixed well by pipetting. 12.4µL was added to each library, followed by 1µL of the diluted Adaptor. The libraries were mixed well by pipetting 10 times, and incubated at 20 °C for 15 minutes in a thermocycler. Following this, 1.2µL of USER™ Enzyme (NEBNext® Multiplex Oligos for Illumina) was added to each library and the libraries mixed by pipetting before incubation at 37 °C for 15 minutes on a thermocycler with the heated lid set to >47 °C.

The DNA libraries underwent a second bead ‘clean up’, following the same procedure as described earlier in this section, but this time with an input of 56µL Beads, and 10µL low EDTA TE with recovery of 8µL DNA library.

The DNA libraries were next ligated to indexes and underwent PCR enrichment. For this, a NEBNext Index/Universal Primer Mix plate (NEBNext® Multiplex Oligos for Illumina®) was thawed and centrifuged to collect the droplets. 10µL of index primer mix was added to each DNA library. The indexes added to each DNA library were recorded. 5µL NEBNext ultra II Q5 Master Mix (NEBNext® Multiplex Oligos for Illumina®) was added to each DNA library, and the samples mixed by pipetting 10 times before centrifuging to collect droplets. The reactions were placed in a thermocycler to undergo PCR enrichment with the cycling conditions as described in Table 2.10.

Step	Temp (°C)	Time	Cycles
Initial Denaturation	98	30 seconds	1
Denaturation	98	10 seconds	6
Annealing/Extension	65	75 seconds	
Final Extension	65	5 minutes	1
Hold	10	Hold	

Table 2.10. Thermocycling conditions for PCR enrichment of indexed libraries

The DNA libraries then underwent the final bead ‘clean-up’, following a similar procedure as described earlier in this section, but this time with an input of 33µL Beads and 53µL low EDTA TE, and recovery of 50µL of DNA library.

### 2.1.8. Tissue library assessment with an Agilent Bioanalyzer 2100

Libraries were checked for their presence and size distribution on an Agilent Bioanalyzer 2100 (Agilent, Stockport, UK) using a High Sensitivity DNA kit. Following the manufacturer's instructions, High Sensitivity DNA gel matrix and High Sensitivity DNA dye were equilibrated to room temperature before 385 $\mu$ L and 15 $\mu$ L, respectively, were combined by vortexing to form a gel-dye mix. The gel-dye mix was passed through the provided spin filter by centrifuging at 2240g for 15 minutes. DNA High-sensitivity DNA chips were loaded with the gel-dye mix, marker and ladder according to the manufacturer's instructions. 1 $\mu$ L of DNA library was pipetted into the test wells. The chip was vortexed at 2400rpm for 1 minute and then loaded onto an Agilent 2100 Bioanalyzer for analysis.

### 2.1.9. Tissue DNA library quantification

Libraries were quantified using the KAPA Illumina Library Quantification Kit (Roche, Burgess Hill, UK). Libraries were diluted 1:100,000 with Tris HCL. 6 $\mu$ L KAPA SYBR® FAST qPCR Master Mix containing Primer Premix was pipetted into each test well of a 96 well plate, with 4 $\mu$ L of library or standard added as appropriate. Libraries were run in duplicate while the standards were run in triplicate, with 6 negative template controls (NTCs) on each plate run. The well plate was loaded onto a QuantStudio 6 flex (ThermoFisher, Hemel Hempstead, UK) and run with conditions as described in Table 2.11.

Step	Temperature (°C)	Time	Cycles
Initial activation	95	5 minutes	1
Denaturation	95	30 seconds	35
Annealing	60	45 seconds	

Table 2.11. Thermocycling conditions for KAPA library quantification

### 2.1.10. Tissue sequencing on an Illumina MiniSeq

DNA libraries were prepared for sequencing by diluting the stock library to 1nM concentration using 10mM Tris-HCl, pH 8.5 with 0.1% Tween 20. 5 $\mu$ L of each library was pooled into a single microcentrifuge tube. 5 $\mu$ L 0.1 NaOH was added, mixed by vortexing, and the sample incubated for 5 minutes at room temperature. 5 $\mu$ L 200 mM Tris-HCl pH 7.0 was added, the sample vortexed and further incubated for 1 minute at room temperature. The sample was diluted with Illumina Hybridization Buffer to achieve 1.8pM concentration and 5% PhiX

control was added before loading onto a MiniSeq mid output cartridge for sequencing on an Illumina MiniSeq platform (Cambridge, UK).

#### 2.1.11. Tissue bioinformatic pipeline and error correction

Sequencing data was analysed by an automated bioinformatic pipeline (VariTAS version 0.8.0) designed within the Breast Cancer Now Toby Robins Research Centre Bioinformatics Core Facility at the Institute of Cancer Research, UK, and facilitated by Dr Ros Cutts, Dr Syed Haider and Adam Mills. FASTQ files from the Illumina run input into the pipeline, which first aligns reads to human genome version 37 using BWA version 0.7.12<sup>178</sup> to produce BAM files. bedtools version 2.25.0 and Picard tools version 2.1.0<sup>179</sup> are used to calculate coverage and quality metrics. Variant calling is achieved using VarDict version 1.4.6<sup>180</sup> and Mutect2<sup>181</sup> software. Manual review of the variant calls made by VarDict revealed that the calls frequently did not align with the output from a pileup analysis (described below). The variant calls made solely by VarDict were therefore excluded from the pipeline. Variants identified by Mutect2 were required to meet the minimum criteria:

- 5 variant reads
- a minimum VAF of 0.01

In parallel, an analysis of mutation calls present was undertaken using a pileup analysis, which assesses the number of reference and altered reads present in the raw data. The following filters and criteria were set for the pileup analysis:

- germline SNPs removed with reference to 1000genomes database (<https://www.internationalgenome.org/data-portal/sample>)
- synonymous variants removed
- variant calls demonstrating strand bias (variant reads present predominantly only on one strand when compared to reference reads), which could result from FFPE artefact were removed
- where paired reads were overlapping, if the base calls disagreed, the base with the highest quality score was maintained, and where both bases had equal quality score both reads were discounted
- minimum depth of 100 at the locus
- minimum of 5 alternate reads
- minimum mapping read quality metric >30

- minimum allele frequency 2% (Chapter 3)
- variants within UTR3/5s, intronic, and ncRNA intronic regions removed

SNV mutations had to be called by either Mutect2 or the pileup analysis to be ascertained as present. Indels required identification by Mutect2. SNVs and indel calls were annotated using ANNOVAR<sup>182</sup>.

## 2.1.12. Guardant360 Targeted Sequencing

### 2.1.12.1. Guardant360 Targeted Panel

The Guardant360 targeted panel (Guardant Health, USA) is a clinical diagnostic panel designed for plasma-based sequencing that covers the protein coding regions, or exons, of 73 (version 2.10) or 74 genes (version 2.11) (Chapter 1, Figure 1.3). The panel identifies single nucleotide variants (SNVs), indels, copy number alterations and fusions. Validation data for the panel has been published<sup>183</sup>. In particular, 222 cancer samples with variants detected by Guardant360 underwent ddPCR validation, demonstrating a PPV of 99.6% (VAF 0.1 – 94%) and NPV of 97.8%<sup>183</sup>. Guardant Health states the panel is able to detect SNVs to 0.25% AF with 80% confidence<sup>184</sup>, and have presented data validating the panel to the FDA<sup>185</sup>.

### 2.1.12.2. Plasma DNA extraction and library preparation

Whole blood samples were shipped to Guardant Health prospectively within the trial to undergo ctDNA testing for cohort eligibility, and processed and stored as plasma as described in section 2.1.2.1. Plasma was also shipped to Guardant Health for sequencing retrospectively, as described in section 2.1.2.4.

At Guardant Health, plasma was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Inc) following the manufacturer's instructions. The samples were concentrated and size selected using Agencourt Ampure XP beads (Beckman Coulter) and quantified using a Qubit® 2.0 fluorometer. Between 5 and 30ng of DNA underwent library preparation. The libraries were labelled with non-random oligonucleotide barcodes (IDT, Inc) which ligate to each end of each single strand of double stranded DNA. This allows later bioinformatic identification of complementary strands, which can then be assessed and compared to differentiate error (which may be derived from sequencing error or DNA damage sustained during sample processing and library preparation) from true mutations. The libraries underwent amplification followed by hybrid capture (Agilent Technologies) using biotinylated

baits, which enriched the libraries for the target genes. Libraries were pooled and sequenced using paired-end synthesis (NextSeq 500 and/or HiSeq 2500, Illumina, Inc).

#### 2.1.12.3. Guardant360 bioinformatic pipeline

The Guardant360 bioinformatic pipeline uses CASAVA (version 1.8.4) to align the raw data to the reference genome. In-house custom scripts perform a number of steps, including trimming lower quality tails and removing lower quality 3' adapter sequences, unaligned sequences, adaptor contamination, PCR duplicates and spurious variants identified by decomplexing the dual barcoding approach. Variant calls (SNPs and SNVs) were made using proprietary software. Variant calls are compared to reference training sets of variant calls derived from healthy plasma which give loci-specific noise level, with variant calls required to achieve a signal level above that defined by the reference training set at the respective loci.

The Guardant360 panel ascertains copy number by counting the number of unique fragments covering a locus and normalising this to the modal number of unique fragments covering the genome, accounting for a diploid baseline. The resulting number of copies is compared to a reference training set of normal copy number profiles to calculate a Z-score, which estimates the likelihood of there being an amplification at that specific locus. Genes with a Z-score  $>2.5758$ , which approximates to the 99.5% confidence level, are called true amplifications<sup>184</sup>.

#### 2.1.12.4. Guardant360 validation

Guardant Health have examined the Guardant360 targeted sequencing panel to establish limits of detection, accuracy and precision of variant identification.

For SNVs, sequencing of plasma from 62 healthy donors established a limit of detection of  $\sim 0.04\%$  at a coverage of 5000, with noise below this limit preventing the detection of true variants<sup>183</sup>. The corresponding limit of detection (LOD) for indels and fusions was 0.02% and 0.04%, respectively<sup>183</sup>. Serial dilution experiments with germline and somatic variants titrated to achieve 5 VAFs around the estimated LOD across 753 observations demonstrated the SNV LOD<sub>95</sub> to be 0.3% for actionable alterations and  $\sim 0.4\%$  for variants of unknown significance<sup>183</sup>. At allele frequencies below the LOD<sub>95</sub> (0.05-0.25%), 63.8% of variants were identified (35/55) with a high accuracy (PPV) of 96.3%. When variants associated with clonal hematopoiesis were excluded, PPV rose to 99.96%<sup>183</sup>.



For indel detection, dilution experiments across an 8-point titration series with a total of 200 observations established a LOD<sub>95</sub> of 0.2% and 0.7% for driver and tumour suppression indels, respectively. The respective LOD<sub>95</sub> for fusion events, derived from titration experiments to generate 93 near-LOD observations, was 0.2%. The respective LOD<sub>95</sub> for CNV alterations, derived from 120 near-LOD dilutions, was 2.44 copies, with a PPV of 98.3%<sup>183</sup>.

#### 2.1.12.5. In-house bioinformatic pipeline

Sequencing results from Guardant Health were transferred to The Institute of Cancer Research as an encrypted .xlsx file. At the ICR, the results underwent further annotation in a process as follows. Data were converted from .xlsx to VCF format using custom R scripts and then to MAF format using `vcf2maf` (<https://github.com/mskcc/vcf2maf>) with flag `MSKCC` isoform overrides and were annotated with VEP version 96<sup>186</sup>. Likely germline calls were filtered out from the database based on a combination of VAF frequency around 50% +/-2% and a VAF >0.001% in the general population (Genome Aggregation Database, <https://gnomad.broadinstitute.org/>).

#### 2.1.12.6. Pathogenicity and targetability assessment

Mutations may be categorised into those that are likely pathogenic and those that are more likely to be ‘passenger’ mutations. To date, there is not a comprehensive understanding of all mutations within all oncological disease settings to enable definitive differentiation of mutations into those which are disease causing and those that are likely passenger mutations. There are, however, a number of databases emerging that aim to categorise mutations based on their pathogenicity<sup>187-190</sup>. We cross referenced our data with three resources, OncoKB, Cancer hotspots and Cosmic, to identify and differentiate the pathogenic mutations in the database from passenger mutations.

OncoKB is a precision knowledgebase published and maintained by the Memorial Sloane Kettering Cancer Centre (MSKCC) and used to annotate tumour and cfDNA mutations sequenced using the MSKCC-Impact platform<sup>189</sup>. The platform gathers mutation pathogenicity data from a number of sources, including cBioportal, CancerHotspots<sup>190</sup>, MSK-Impact, Cosmic<sup>191</sup>, BRCA exchange<sup>192</sup>, IARC TP53 database<sup>193</sup> and the published literature. A Clinical Genomics Annotation Committee (CGAC) review all available data manually to curate and assign levels of pathogenicity (Oncogenic, Likely Oncogenic and Neutral and inconclusive) based on pre-defined criteria (Figure 2.4).

## Sub-protocol 2.5: Assertion of the oncogenic effect of a VPS

Assertion of the oncogenic effect of an alteration (A-E) **requires at least 1 of criteria** from the corresponding evidence column.

Assertion	Definition	Criteria	Evidence (the alteration meets any of the following criteria)
<b>A. Oncogenic</b>	Strong evidence shows that the alteration is established in the literature as promoting cell proliferation or other hallmark of cancer as defined by Douglas Hanahan and Robert Weinberg (Hanahan and Weinberg, 2011).	1	Compelling experimental data (e.g., genetically engineered mouse data with the mutation) in one or more studies directly demonstrating that the alteration is oncogenic and is associated with at least one hallmark of cancer as defined by Hanahan and Weinberg
		2	The alteration is a known hotspot ( <a href="#">Chang et al., 2018</a> ) AND there is at least one experimental study suggesting the alteration is oncogenic.
		3	The alteration has been identified in a patient who responded to a targeted inhibitor, AND at least one experimental study provides strong evidence that the alteration is oncogenic.
		4	The alteration is classified as either known gain/loss/switch-of-function AND there is at least one experimental study suggesting the alteration is oncogenic.
<b>B. Likely Oncogenic</b>	Evidence suggests the alteration likely promotes cell proliferation or other hallmarks of cancer as defined by Douglas Hanahan and Robert Weinberg (Hanahan and Weinberg, 2011).	1	Representative experimental lines of data (e.g., downstream activation/inactivation of a signaling target/a hit in a high-throughput screen) in one or more studies pointing to possible oncogenic function or mutation associated with known germline syndrome.
		2	At least one experimental study provides reasonable evidence suggesting the alteration is oncogenic.
		3	The alteration is a known hotspot ( <a href="#">Chang et al., 2018</a> ) AND there are no known functional studies describing the oncogenic potential of the alteration.
		4	The gene is a tumor suppressor and the variant is a truncating mutation (i.e. nonsense/frameshift/deletion/splice site mutation)
		5	The mutation is a resistance mutation supported by demonstrating either patient and/or in vitro sensitivity/resistance to a targeted drug.
<b>C. Likely neutral</b>	Evidence suggests the alteration does not alter protein activity or does not confer growth or survival advantage when expressed in cells.	1	The mutation effect of the alteration is neutral or likely neutral.
		2	At least one experimental study provides reasonable evidence suggesting the alteration is likely neutral.
<b>D. Inconclusive</b>	There is conflicting and/or weak data describing the oncogenic effect of the mutant alteration	1	Conflicting data exists as to the oncogenic effect of the alteration.
		2	Data is limited to "weak" experimental data describing the oncogenic effect of the alteration (small, under-powered experimental studies in one or multiple publications).
		3	Data is limited to in silico studies that predict the oncogenic effect of the alteration.

Figure 2.4. OncoKB criteria needed to be met for assignment of oncogenicity for Variants of Possible Significance (VPS). From *OncoKB Curation Standard Operating Procedure v2.1*, July 2021, available at [OncoKB.org](https://oncoKB.org)<sup>194</sup>.

The CGAC also assign targetability by gathering variant targetability data from sources including the US Food and Drug administration (FDA), National Comprehensive Cancer Network (NCCN), National Institutes for Health (NIH), professional guidelines, conference abstracts and proceedings, ClinicalTrials.gov and published literature in pre-defined peer reviewed journals. This information is manually curated by members of the CGAC to identify targetable variants and the level of supporting evidence (Figure 2.5).



Figure 2.5. OncoKB levels of evidence required for assignment of targetability for oncogenic variants. From *OncoKB Curation Standard Operating Procedure v2.1*, July 2021, available at [OncoKB.org](https://oncobk.org)<sup>194</sup>

Cancer Hotspots is a resource also developed at MSKCC that adopts a computational approach to identifying recurrently mutated hotspots in the genome. It is based upon the rationale that recurrently mutated mutations/loci must harbour an evolutionary advantage to the cell and therefore are more likely to be pathogenic. The database was produced by analysis of the whole-exome and whole-genome sequencing of 11,110 tumour samples. The software identifies recurrently mutated loci, or ‘hotspots’, that are mutated more frequently than would be expected by chance having normalised for factors including nucleotide context mutability, gene-specific background mutation rates and major expected patterns of hotspot mutation emergence<sup>190</sup>.

Finally, Cosmic is an online data repository and resource that, by collating sequencing data from thousands of cancers, provides information on recurrently mutated areas in the genome alongside allowing identification of novel mutations<sup>191</sup>.

To identify and annotate data for likely pathogenic mutations, a bespoke pipeline was developed. Annotated MAF files were cross referenced with the OncoKB database using the OncoKB API. Cancer Hotspots was downloaded and variants were additionally cross referenced against this resource using custom R scripts. Mutations identified as Oncogenic, Likely Oncogenic or Predicted Oncogenic by OncoKB, or recurrently mutated mutations identified by Cancer Hotspots, were deemed pathogenic. Additionally, recurrent mutations in key breast cancer genes (*ESR1*, *ERBB2*, *PIK3CA*, *EGFR*, *RB1* and *FGFR2*) or splicing mutations were identified as pathogenic. The combined use of these resources, each of which approach the identification of pathological variants in different ways, aimed to allow identification of pathogenic mutations with high confidence.

#### 2.1.12.7. Gene pathway assignment

Genes were grouped into signalling pathways using the online KEGG database resource<sup>195</sup>.

MAPK pathway included mutations in the following genes (*EGFR*, *HRAS*, *KRAS*, *NRAS*, *ARAF*, *BRAF*, *RAF1*, *MAP2K1*, *MAP2K2*, *MAPK1*, *MAPK3*, *FGFR1*, *FGFR2* and *FGFR3*), fusions of *FGFR2* and *FGFR3*, and copy number changes (CN > 5) in the following genes: *BRAF*, *EGFR*, *FGFR1*, *FGFR2* and *KRAS*.

TORC1 pathway included mutations in the following genes (*PIK3CA*, *KRAS*, *AKT1*, *PTEN* and *TSC1*).

#### 2.1.12.8. Cancer fraction assessment

A number of variables dictate the variant allele frequency (VAF) of a mutation. In plasma, an important contributor is the concentration of tumour DNA relative to circulating free DNA, termed 'purity'. A second factor is the number of cancer cells harbouring the mutation relative to all the cancer cells present, with clonally dominant mutations expected to be in all or the majority of cancer cells or clones while subclonal alterations are only present in defined cell population. Finally, consideration must be made to the potential influence gene copy number alterations<sup>196</sup>. Analysis of our dataset, however, indicated that this was not a strong influence on VAF (Figure 2.6), and therefore VAF was not adjusted for the gene copy number.

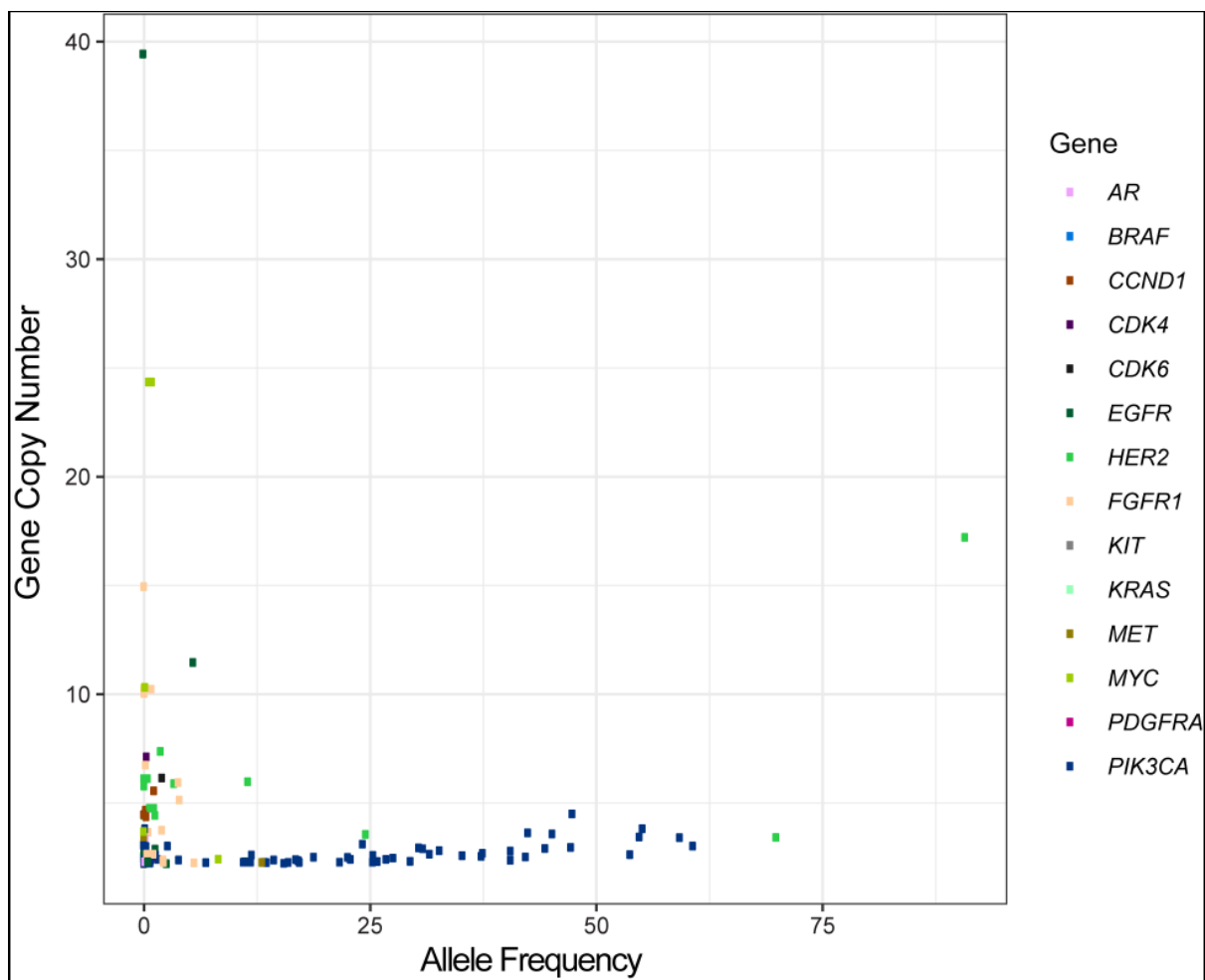


Figure 2.6. Association of allele frequency with gene copy number within 800 patients sequenced within plasmaMATCH. Lack of linear relationship indicates that copy number is not a strong contributor to mutation VAF. Figure courtesy of Dr Ros Cutts. *HER2*, *ERBB2*.

Under the assumption that the mutation with the highest allele frequency represents a truncal mutation present in every cancer cell, the clonality of other mutations can be calculated relative to the truncal mutation, and termed cancer fraction or clonal dominance. For each mutation, the cancer fraction/clonal dominance was calculated as the mutation VAF relative to the maximum somatic VAF (mVAF). Cancer fraction is either presented on a continuous scale, or categorised into clonally dominant (cancer fraction  $\geq 0.5$ ) versus subclonal (cancer fraction  $< 0.5$ ).

#### 2.1.12.9. Copy number adjustment

*ERBB2* copy number is influenced by the purity of the plasma sample. To account for this, the *ERBB2* copy number was adjusted relative to the sample mVAF using the following formula<sup>197</sup>:

$$\text{Adjusted pCN} = [\text{Observed pCN} - 2 * (1 - T\%)] / T\%$$

$$\text{Where } T\% = 2X \text{ mVAFmax} / 100$$

$$\text{pCN} = \text{patient Copy Number}$$

Where indicated, copy number was adjusted for other genes following the same method.

#### 2.1.12.10. APOBEC site identification

For categorical analysis of APOBEC consensus sites the following trinucleotides were included: T(C>G)T, T(C>G)A, T(C>A)[N] on both DNA strands.

## 2.2. Low-pass Whole Genome Sequencing

### 2.2.1. Library preparation and quantification

Libraries were prepared from between 0.7 to 93ng DNA in 50uL, made up to volume with nuclease free water. A Kapa Hyper Prep kit (Roche) was used to prepare the libraries. DNA first underwent end repair and A-tailing by the assembly of End Repair and A-tailing Buffer (Roche) and End Repair and A-tailing Enzyme mix with the DNA (Table 2.12)

Component	Volume (μL)
DNA	50
End Repair and A-tailing	7
End Repair and A-tailing	3
Total Volume	60

Table 2.12. DNA end repair and A-tailing master mix constituents

Libraries were vortexed and centrifuged before being incubated on a thermocycler (Table 2.13).

Step	Temp (°C)	Time (minutes)
End Repair and A-tailing	20	30
	65	30
HOLD	4	∞

Table 2.13. Thermocycling conditions following end repair and A-tailing

Libraries next underwent adaptor ligation. Adaptors (IDT, Illumina) were diluted in 10mM Tris-HCL before being combined in a microcentrifuge tube containing the DNA with the following components (Table 2.14). The tubes were incubated at 20 °C for 4 hours.

Component	Volume (μL)
End Repair and A-tailing product	60
Adaptor	5
PCR-grade water	5
Ligation Buffer	30
DNA ligase	10
Total volume	110

Table 2.14. Adaptor ligation master mix constituents

Following adaptor ligation, the libraries underwent a bead clean up. For this, 88μL KAPA Pure Beads were added to the adaptor ligated DNA and the sample mixed thoroughly by vortexing. Libraries were incubated for 15 minutes at room temperature, before being placed on a magnet to capture the beads. Once the liquid was clear, the supernatant was removed by pipetting. 200μL 80% ethanol was added whilst the sample remained on the magnet, incubated for 30 seconds, and then the supernatant removed by pipetting. The ethanol addition step was repeated a second time before the beads were dried at room temperature for 5 minutes. The samples were removed from the magnet and the beads resuspended in 25μL 10 mM Tris-HCL, pH 8.0-8.5), incubated for 2 minutes at room temperature, and the samples placed back on the magnet to capture the beads. Once the supernatant was clear, it was pipetted into a new tube to undergo library Amplification.

For the Amplification step, 25μL KAPA HotStart ReadyMix and 2μL KAPA Library Amplification Primer Mix were added to the adaptor ligated library, mixed by pipetting and centrifuged. The samples were placed in the thermocycler with the conditions as described in Table 2.15.

Step	Temp (C)	Duration	Cycles
Initial Denaturation	98	45 seconds	1
Denaturation	98	15 seconds	9
Annealing	60	30 seconds	
Extension	72	30 seconds	
Final Extension	72	1 minute	1
HOLD	4	∞	1

Table 2.15. Thermocycling conditions for whole genome library amplification

The libraries then underwent a bead clean up using the same methodology as described earlier in the section, but with 50uL bead input and a final elution volume of 30uL 10mM Tris HCL, pH 8.0-8.5. Libraries were checked for peak location on a Bioanalyzer (section 2.1.8) and quantified by qPCR (section 2.1.9).

### 2.2.2. Library sequencing

Libraries were supplied to the TPU (Tumour Profiling Unit, Institute of Cancer Research) at 4nM concentration made up to 20uL volume with Buffer EB. Libraries were quantified and quality checked in the TPU by Bioanalyzer and qPCR before pooling and sequenced on a NovaSeq with a 50bp PE S1 flowcell (v1). Libraries underwent PE50 cycles aiming to achieve a median depth of 0.5X. Sequencing data was obtained in the form of FASTQ files.

### 2.2.3. Bioinformatic pipeline

Fastq files were adaptor trimmed using cutadapt v3. Data was aligned to human reference genome hg38 (UCSC; GRCh38 Ensembl) using BWA v0.7.15. Duplicates were removed with MarkDuplicates from gatk v4.1.3.0. Sequencing metrics were generated using GATK CollectWgsMetrics. Reads were counted using a bin size of 1MB using hmmcopy v0.1.1 followed by Purity and ploidy estimation using ichorCNA<sup>85</sup> using only autosomal chromosomes with a maximum copy number set at 5.

## 2.3. MCF-7 transient transfection and *ESR1* p.F404 investigation

### 2.3.1. MCF-7 Culture

Genotyped MCF-7 cells were cultured in phenol-free RPMI media (Life Technologies) supplemented with 10% dextran/FBS (Life Technologies), 1nmol/L estradiol (Sigma), glutamine (Life Technologies), penicillin and streptomycin (Life Technologies) in T75 and T175 flasks and split before confluence. Mycoplasma testing of 1ml media was undertaken at least monthly.

### 2.3.2. Plasmid design, procurement and verification

#### 2.3.2.1. Estrogen Receptor Constructs

Estrogen receptor constructs (ERCs) were custom designed in-house before manufacturing by GenScript (The Netherlands). For the plasmid design process, the sequence for the *ESR1* alpha open reading frame (ORF), which encodes the estrogen receptor, was downloaded from Ensembl (<https://www.ensembl.org/index.html>). Base changes were made at the respective loci to achieve the following:



- No alteration (wild type *ESR1*)
- E380Q 1138G>C
- D538G 1613A>G
- F404L 1210T>C
- E380Q 1138G>C and F404L 1210T>C
- D538G 1613A>G and F404L 1210T>C

The sequences for the plasmid inserts were supplied to GenScript. The backbone plasmid, pcDNA3.1+C-DYK, was selected from the GenScript expression vector bank. pcDNA3.1+C-DYK (Figure 2.7) was selected for its compatibility with mammalian transfection and C-terminal DYK flag which would allow later assessment of the expression +/- localisation of the estrogen receptor construct. The plasmid contains a CMV promoter, suitable for mammalian expression, a Kozak sequence for ribosomal binding, and contains a neomycin resistance gene which allows selection of successfully transfected cells with Genticin. GenScript custom built the inserts which were then cloned into the backbone plasmid.

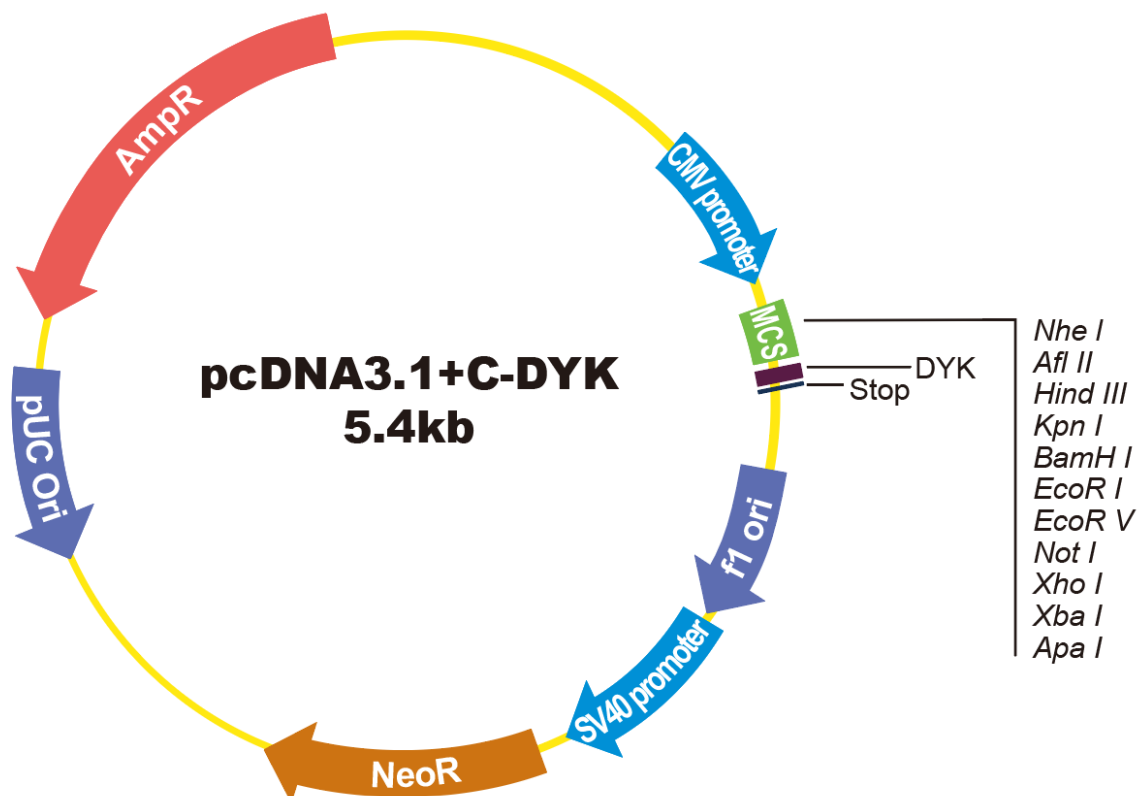


Figure 2.7. Vector map of pcDNA3.1+C-DYK. pcDNA3.1+C-DYK is the backbone vector into which the custom insert was applied.

The resulting ERCs underwent Sanger sequencing to confirm the insert sequences, with the sequencing files transferred to the ICR for confirmation (Figure 2.8). Six constructs were customised in total: wild type *ESR1*, the 5 variants as listed above, and an ‘empty vector’ without the *ESR1* insert. ERCs were shipped at ambient temperature to the ICR, where they were stored at  $-20^{\circ}\text{C}$ .



Figure 2.8. Estrogen Receptor construct sequencing confirmation of the insert. The reference wild type sequence is shown as the ‘Original Sequence’. The alterations within each construct are illustrated by a red box.

### 2.3.2.2.ERE-Luciferase

A plasmid expressing an estrogen-report element ligated to a firefly luciferase reporter (ERE-Luc) was kindly gifted to the laboratory by Dr Lesley-Ann Martin, ICR. The plasmid was originally constructed from the combination of pdGL3-Basic and pdEREGFP. The GFP was

removed from the pdEREGFP, and the luciferase element taken from pdGL3-Basic and spliced in to make pdERELuc (5.3kb).

Sanger sequencing was undertaken to confirm the presence of the estrogen response element. For this, primers were designed with reference to the firefly luciferase gene (sequence obtained from Promega: pGL3 Luciferase reporter vector, <https://www.promega.co.uk/products/luciferase-assays/genetic-reporter-vectors-and-cell-lines/pgl3-luciferase-reporter-vectors/?catNum=E1751>). The online web application Primer3web (<https://primer3.ut.ee/>) was used to identify suitable primer sites. The aim was to sequencing the regions immediately adjacent to the luciferase sequence, such that the reverse primer was situated at the 5' end of the sequence and the forward primer at the 3' (Table 2.16). The primer designs were then submitted to Sigma Aldrich (USA) for construction. The primers were delivered lyophilised at ambient temperature, and were reconstructed in nuclease free water to achieve 100µM concentration. 5pmol of each respective primer within 5uL nuclease free water was then combined with the ERE-Luc plasmid DNA in microcentrifuge tubes and shipped at ambient temperature to GENEWIZ for Sanger sequencing.

Primer	Sequence
Left	CCGACGATGACGCCGCGTGAACCTGCCCGCTGCCGTTGTTGTTTTGGAGCAC AGCGAAAGACGATGACGGANNAGAGATCGTGGATTACGTCGCCAGTCAAGT AACAAACCGCGAAA
Right	ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCCATTCTATCCTCTAG AGGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCT GGTTCTTGAACAATTGCTTTTACAGATGCACATATCGAGGTGAACATCACGT ACGCGGAATACTTCGAAATGTCCGTTCCGTTGGCAGAAGCTATGAAACGATAT GGGCTGAATACAAATCACAGAATCGTCG

Table 2.16. Forward and reverse primer sequences for Sanger sequencing of the ERE-Luc plasmid.

The Sanger sequencing results confirmed the presence of an ERE consensus sequence upstream of the luciferase gene (Figure 2.9).

```

TCGACGATTCTGTGATTTGTATTAGCCCATATCGTTTCATAGCTTCTGCCAACC
GAACGGACATTTTGAAGTATTCGCGGTACGTGATGTTACCTCGATATGTGCAT
CTGTAAAAGCAATTGTTCCAGGAACAGGGCGTATCTTTCATAGCCTTATGCA
GTTGCTCTCCAGCGTTCCATCCTCTAGAGGATAGAATGGCGCCGGCCTTTC
TTTATGTTTTGGCGTCTCCATTTTACCACAGTACCGGAATGCCAAGCTTGGA
TCCTCGAGATCTGCGGCACGCTGTTGACGCTGTTAAGCGGGTCGCTGCAGGGT
CGCTCGGTGTTGAGGCCACACGCGTACCTTAATGCGAAGTGGACCTCGG
ACCGCGCCGCCCGACTGCATCTGCGTGTTCGAATTCGCCAATGACAAGACGC
TGGCGGGGTTTGCCTGATCTCTAGAGGATCCTCTAGAGTGCACCTGCAGGCA
TGCAAGCTTCAAGTACAGTACCTGATCAAAGTAAATGTAACCTCAACCTGGA
AGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGT
TACCCAACTAATCGCCTTGACGACATCCCCCTTTCGCCAGCTGGCGTAATAG
CGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGC
GAATGGCGCCTGATGCGGTATTTCTCCTTACGCATCTGTGCGGTATTTACAC
CGCATAGGCCTCGTATACGCCTATTTTATAGGTTAATGTCATGATAAATGG
TTTCTTAGACGTCAAGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT
GTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGA
TAAATGCTTCAATAATATTGAAAAAGGAGATGAGTATTCAACATTTCCGCGT
TCGCCCTTATCCCTTTTTGCGGCATTTTGCCTTCTGTTTTGCTCACCCAGA
AACGCTGGTAAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTCGCCCGAAG
AACGTTTTCCATGATGAGCACTTTTAAAGTCTGCTNGTGGCGGTATTTCCCGT
ATTGACGCGGGCAGAACACTCGGTCGCGCATACCTATTTCCAGATGACTGGGT
GAAANCCCCCAGTCCNGAANCCTCTTAGAGGCCCTGACGTAGAAAATTTGCAG
GNGGCCAAACCTGNGGAAACCTGGGGCCATTNNTTGANNAAAATGGGGAANCN
AGAGNNNNNCTTTTTTCCNCGGGGAATT

```

Figure 2.9. Sanger sequencing result from the right primer for the ERE-Luc plasmid. Highlighted in yellow is the right primer. Highlighted in blue is the ERE palindromic consensus sequence, 5'-GGTCAAnnnTGACC-3'.

### 2.3.2.3.pSV-β-Galactosidase Control Vector

A β-Galactosidase Control Vector (β-gal) was kindly gifted by Dr Lesley-Ann Martin, ICR. Although the plasmid sequence was not available, transfections with this plasmid resulted in positive β-Galactosidase assay results, confirming the function of the plasmid. However maxi-prep of the plasmid resulted in low concentrations of the plasmid, potentially due to the low copy number replication apparatus. An initial transfection of β-gal alongside the ERCs and ERE-Luc demonstrated low concentrations of β-gal. The decision was made to purchase a commercial β-gal vector to use within the remaining ERC transfections. The β-Galactosidase Control Vector was therefore purchased from Promega (UK) and underwent maxi-prep to increase stocks.

### 2.3.3.Plasmid Maxi-prep

Maxi-prep of plasmid was undertaken using Hi-Speed Plasmid Kit (Qiagen). Competent E.Coli were thawed. 10ng of plasmid DNA was added to 50uL of competent cells and mixed by gentle flicking of the tube. The competent cell/DNA mixture was incubated on ice for 20-30 minutes before undergoing heat shock in a 42°C water bath for 60 seconds. Tubes were then incubated

on ice for 2 minutes. 1000uL of LB media was added to the bacteria and the tube placed on a shaking incubator at 37°C for 45 minutes. Following this, the transformed cells were pipetted and streaked onto a 10cm agar plate containing ampicillin to select for positively transformed cells. Plates were incubated at 37°C overnight.

The following day, colonies are selected and placed into 5ml of LB medium containing ampicillin within a flask prior to incubation for 8 hours at 37°C with vigorous shaking (300 rpm). The starter culture was then diluted 1/500 into 150ml ampicillin containing LB medium. The flask was then incubated at 37°C for 12-16 hours with vigorous shaking (300rpm). Cells were harvested by centrifugation at 6000xg for 15 minutes at 4°C. The pellet was resuspended in 10ml Buffer P1 followed by 10ml Buffer P2 and the sample mixed by inversion 4-6 times prior to incubation at room temperature for 5 minutes. 10ml of chilled Buffer P3 was added to the lysate and mixed by inversion 4-6 times. The lysate was then added to the barrel of a QIAfilter Cartridge and incubated at room temperature for 10 minutes.

10ml Buffer QBT was used to prepare a HiSpeed Maxi tip by allowing it to empty from the column by gravity flow. The cap was removed from the Quafilter outlet. A plunger inserted into the QIAfilter Cartridge and was used to filter the cell lysate into the HiSpeed Maxi tip. The lysate passed through the resin of the HiSpeed Maxi tip by gravity flow. The HiSpeed Maxi tip was then washed with 60ml Buffer QC. DNA was then eluted with 15ml Buffer QF. DNA was precipitated by the addition of 10.5 ml isopropanol, the solution mixed and incubated at room temperature for 5 minutes.

The syringe was removed from a 30ml syringe and the syringe attached to a QIAprecipitator Maxi Module. The eluate/isopropanol mixture was added to the syringe and a plunger used to filter the mixture through the QIAprecipitator Maxi Module. 2ml of 70% ethanol was then washed through the QIAprecipitator Maxi Module using the syringe. Subsequently the Module was dried by plunging air through it. The QIAprecipitator Maxi Module was then attached to the barrel of a 5ml syringe and 1ml of Buffer TE added. This was passed through the Module and the resulting eluate captured in a 1.5ml collection tube. This is repeated to achieve 2ml eluted plasmid DNA.

To quantify the DNA, DNA was diluted 1/100 and underwent Qubit quantification following the same procedure as described in section 2.1.5.2.

### 2.3.4. Estrogen Reporter Gene transfection protocol

Following optimisation, the following protocol was followed for transfection of ERCs, ERE-Luc and  $\beta$ -gal. On day 1, 250,000 MCF-7 cells were seeded into four 6-well plates in 3ml phenol-free RPMI media supplemented with 10% charcoal-stripped FBS, 1nmol/L estradiol, and glutamine (antibiotic-free DCC) (Figure 2.10 for plate layout).

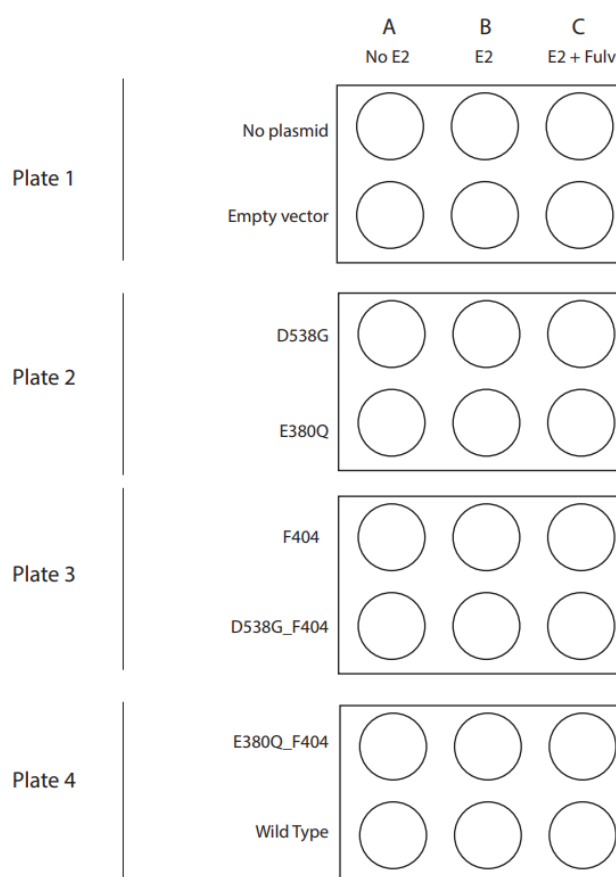


Figure 2.10. Six-well plate layout for ERC transfection. No E2, cells in estrogen free media prior to harvesting. E2, estrogen added for the last 24 hours of transfection. E2 + Fulv, estrogen and fulvestrant added for the last 24 hours of transfection.

On day 2 at 4pm, cells were transfected with ERCs, ERE-Luc and  $\beta$ -gal. For this, transfection master mixes were made with, optiMEM (Thermo Fisher Scientific), the plasmids and finally FuGENE (Promega) according to Table 2.17 for the first transfection with the in-house  $\beta$ -gal, and Table 2.18 for later transfections with maxi-prep DNA derived from commercial  $\beta$ -gal, with the aim of transfecting with a 2:1 lipid to plasmid ratio. The concentration of the plasmids was as follows:

- ERCs 1µg/µL
- ERE-Luc 0.69µg/µL
- In-house β-gal 0.07µg/µL
- Maxi-Prep of commercial β-gal 0.394 ug/uL<sup>+</sup> /0.284 ug/uL\*

Volumes of each respective plasmid were calculated to achieve 1µg of each plasmid per well, with a 1:1:1 ratio of each of ERC, ERE-Luc and β-gal (Table 2.17), with volumes scaled for the appropriate number of wells. Of note in later transfections (Table 2.18) the control plasmid free well transfection plan was altered such that the cells were not transfected with any plasmid, to improve the control, and the stock concentration of β-gal changed after maxi-prep.

ERC	For 1 well				
	Plasmid			OptiMEM	FuGENE
	ERC	ERE-Luc	β-gal		
1) Control	0	1.45	14.3	133.25	6
2) Empty vector	1	1.45	14.3	132.25	6
1) D538G	1	1.45	14.3	132.25	6
2) E380Q	1	1.45	14.3	132.25	6
3) F404	1	1.45	14.3	132.25	6
4) D538G & F404	1	1.45	14.3	132.25	6
5) E380Q & F404	1	1.45	14.3	132.25	6
6) Wild type	1	1.45	14.3	132.25	6

Table 2.17. Master mix constituents of transfections using the in-house β-gal DNA

ERC	For 1 well				
	Plasmid			OptiMEM	FuGENE
	ERC	ERE-Luc	β-gal		
1) Control	0	0	0	149	6
2) Empty vector	1	1.45	+2.54/*3.52	+144/*143	6
3) D538G	1	1.45	+2.54/*3.52	+144/*143	6
4) E380Q	1	1.45	+2.54/*3.52	+144/*143	6
5) F404	1	1.45	+2.54/*3.52	+144/*143	6
6) D538G & F404	1	1.45	+2.54/*3.52	+144/*143	6
7) E380Q & F404	1	1.45	+2.54/*3.52	+144/*143	6
8) Wild type	1	1.45	+2.54/*3.52	+144/*143	6

Table 2.18. Master mix constituents of transfections using the maxi-prep of commercial β-gal (first colony extraction with concentration 0.394 ug/uL<sup>+</sup>; second colony extraction with concentration 0.284 ug/uL\*).

At 9am on the third day, the plates underwent media change to remove the lipid, estrogen depletion and antibiotic selection for successfully transfected cells with gentamicin. Media was removed from each well, the well PBS washed, and then 3ml estrogen free media (phenol-free RPMI media supplemented with 10% charcoal-stripped FBS, glutamine, penicillin and

streptomycin) added. Wells in rows 2 to 8 (Figure 2.10) were dosed with 1µg/ml gentamicin to select for cells successfully transfected with an ERC. Plates were put on the shaker for 1 second to mix before being incubated for 7 hours at 37 °C.

After 7 hours, the media was suctioned off each well, and wells washed with PBS to remove dead cells. Media was added to each well to satisfy the estrogen and fulvestrant conditions as described in the plate layout (Figure 2.10). Namely, column 1 had 3ml of estrogen-free DCC. Columns 2 and 3 had 3ml estrogen-containing (1nmol/L) DCC added. Column 3 was also treated with 500nm fulvestrant. Plates were placed on the shaker before being incubated for 24 hours at 37 °C.

At 24 hours, lysates were made from each well. For this, 4 volumes of UF water were added to 1 volume of Reporter Lysis Buffer (Promega). The media was removed from the cells by flicking out the plates. The wells were washed twice in succession with phosphate buffered saline (PBS), with the remaining fluid removed by suction. 400µL of the RLB/water mix was added to each well and the plate rocked to ensure coverage. The plate was incubated for 15 minutes at room temperature with frequent rocking of the plate. Wells were then scraped and the cell lysate collected with a pipette into microcentrifuge tubes, and immediately put on ice. Cell lysates were vortexed for 10 seconds before centrifuging at full speed for 2 minutes at 4 °C. The resulting supernatant containing the lysate was pipetted into a new collection tube. Lysates were assayed immediately or stored at -80 °C until further analysis.

### 2.3.5. Drug Concentration assessment using cell viability assays

MCF-7 cells were plated in 96-well plates at 1000 cells per well. 24 hours later, wells were dosed with a range of drug concentrations of the respective drugs. At 6 days post drug application, the cells underwent CellTiter-Glo to establish cell viability. For this, media was flicked out of the plate. CellTiter-Glo buffer (Promega) was thawed and mixed 1:3 with water. 100µL CellTiter-Glo buffer mix was added to each well, the plate covered and incubated on an orbital shaker for 5 minutes. Luminescence was then measured.

### 2.3.1. Western Blot

#### 2.3.1.1. Lysate harvesting and protein quantification

The same protocol as described in section 2.3.4 was followed, except for a single media condition included: 'with oestrogen'. Lysates were prepared by mixing 10µl phosphate and



protease inhibitor solution with 5µl DTT and 985µl NP40 lysis buffer. This mixture was placed on ice alongside a 50ml falcon of PBS. Media was poured off the wells, and the plate placed on ice. The wells were washed with 1ml cold PBS and the wash removed. 1ml of cold PBS was added again to the wells and cells were then scraped and collected with a pipette into a microcentrifuge tube. 70µl of the NP40/DTT/inhibitor solution mix was added to each microcentrifuge tube, the tubes vortexed, and the samples incubated in ice for 15 minutes. Following this, the tubes were vortexed again and centrifuged at full speed for 10 minutes. The supernatant was removed and stored in a new microcentrifuge tube and the pellet discarded.

To ascertain protein concentration of the lysates, 2ml of Protein reagent dye was added to 8ml UF water. 5µl of each lysate (and 5µl of the NP40/DTT/inhibitor solution, as a control) was pipetted into a cuvette and 1ml of the protein reagent dye mix added and mixed by pipetting. Samples were read in the spectrophotometer at wavelength 595nm. The absorbance reading was converted to concentration for each sample. Lysates were frozen at -20°C until further analysis.

#### 2.3.1.2. Western blot

The lysate samples were first prepared. For this, lysate was mixed with Loading buffer, reducing agent and water with the aim of a final Western Blot sample protein volume of 17.5µg per sample (concentration 0.8µg/µL). Once mixed, samples were vortexed and heated on a heat block set to 75°C before being placed on ice.

To make the gel, 50ml MOPS was added to 950ml SQ water and mixed. A tank was filled with the MOPS solution before adding a 10 well cassette. Bubbles were flushed out of the cassette. Lanes 1 and 10 had 5µl and 3µl Full Range Rainbow Molecular Weight Marker added, respectively. Wells 2-9 had 25µl of each respective prepared lysate solution added. The lid was placed on the tank and the tank surrounded in ice before running the tank at 200V for 50 minutes.

The gel was then transferred to nitrocellulose. For this, the required equipment (4 sponges, nitrocellulose membrane, and two blotting paper) were pre-soaked in transfer buffer. The transfer stack mould was then put together with one sponge, one blotting paper, the gel, nitrocellulose membrane, one blotting paper and 3 sponges on top, followed by the mould lid. This was placed into the tank and the tank filled with transfer buffer. The lid was put on the tank and the tank run at 30V for 2 hours.

The resulting nitrocellulose paper was removed from the stack and blocked with 15ml 5% milk made up with TBST at 16 hours at 4°C on a rocker.

Following blocking, the blot was cut into two sections along the 50kDa ladder line. Protein antibodies were made according to Table 2.19. The antibody solutions were added to the blots and the blots incubated for 2 hours on a rocker. Blots were washed with three successive 5-minute washes on the rocker in TBST. Secondary antibody solutions were made up according to Table 2.19 for the respective primary antibody, and this added to the blots. Blots were incubated for an hour on the rocker in the secondary antibody solution before three 5-minute TBST washes. Anti- $\beta$  tubulin was used as a control antibody.

Primary Antibody	Code	Source	Concentration	MW (kDa)	Secondary antibody	Concentration
ESR1 alpha	D6R2W	Cell Signalling	1:1000	66	Rabbit mAb	1:2000
Anti-flag DYK	70765	Santa Cruz	1:200	60-80	Mouse mAb	1:2000
Anti- $\beta$ Tubulin	T8328	Sigma-Aldrich	1:1000	50	Mouse mAb	1:2000

Table 2.19. Primary and secondary antibody details for western blot.

Blots were imaged using ECL Prime Western Blotting Detection Reagent (RPN2232, GE Healthcare) on a LI-COR Imager (LI-COR Biosciences, Nebraska, USA).

Blots were stripped by rinsing in TBST, followed by the addition of stripping solution (1M Glycine + 10% Sodium dodecyl sulphate) and incubation for 10 minutes on the rocker at room temperature. The solution was removed and the blot rinsed in 1M Tris pH7.4 followed by a 5 minute rinse in TBST. Blots were either then blocked in TBST-5% non-fat dried milk for 16 hours prior to further antibody incubations, or stored in the refrigerator.

## 2.4. Statistical analysis

All statistical analyses were undertaken in R versions 3.5.2 or 4.0.5 and Graphpad Prism version 8.0.1.

## 3. Chapter 3. Validation of ctDNA analysis

### 3.1. Introduction

Genomic profiling has historically been achieved through tissue sequencing, and several large-scale tissue sequencing efforts have made significant progress in defining the genomic landscape of ABC<sup>34-38</sup>. For individual patients, however, temporal acquisition and loss of genomic alterations<sup>72,81</sup>, particularly from primary to advanced disease<sup>27,41-46</sup>, necessitates repeated biopsy to genomically profile their disease through time. Tissue biopsies are invasive<sup>39</sup> and limited to genomically sampling the biopsied lesion making them prone to sampling bias secondary to the existence of temporal heterogeneity<sup>49-51</sup>. These challenges could theoretically be overcome by analysis of ctDNA derived from plasma. ctDNA, which is an admixture of tumour DNA released from heterogeneous metastatic sites, may give a fuller picture of a patient's genomic profile. Crucially, ctDNA is gained through a non-invasive blood test, which allows repeated sampling over time in an approach that is more acceptable to patients, avoids risk, and allows genomic profiling in patients whose disease is otherwise not amenable to a biopsy. In an era where precision oncology is coming to the forefront of oncological management, the ability to non-invasively and repeatedly genomically profile a patient's cancer to direct therapeutic choice is critical.

Before bringing ctDNA analysis into medical management of ABC, the technology must be validated and the potential applications and limitations well defined and understood. We must also understand how ctDNA analysis relates to tissue analysis. Validating ctDNA analysis as an approach to genomically profile a patient's cancer is challenging, not least because the current gold standard of tissue biopsy has several limitations that make benchmarking difficult (Introduction sections 1.3.1 and 1.3.6). Historically, comparisons with paired tissue sequencing have demonstrated widely varying levels of agreement, ranging from 45 to 100% (Chapter 1, Table 1.3). There are many variables confounding the comparison, including the volume of ctDNA shedding of the tumour, the temporal acquisition of the samples, and the clonal dominance of the mutation.

The metric used to compare the mutation agreement also varies. Commonly 'concordance' is used to describe the proportion of patients with exact matching ctDNA and tissue sequencing results for one or more variants. Alternatively, sensitivity may be calculated within which the

tissue result is considered the gold standard approach, and describes the proportion of true positive results that are identified by ctDNA testing relative to all the positive calls made by ctDNA testing. Related to this is the positive predicted value, which informs physicians of how likely a ctDNA positive result would be to match the ‘gold standard’ tissue testing result. Conversely, specificity describes the proportion of true negative results that are identified by ctDNA testing relative to all the negative calls made by ctDNA testing. Related to this, the negative predictive value informs a physician on how likely a negative result is to represent a true negative result. The different metrics used can additionally make comparisons between studies and trials challenging.

In this Chapter, two approaches are taken to validate ctDNA analysis. Firstly, the results from two orthogonal methods of ctDNA analysis are compared. Secondly, a comparison of ctDNA to paired tissue is made with the aim of exploring the relationship between the two approaches across clinically actionable mutations.

## 3.2. Hypothesis

ctDNA can be analysed to accurately deduce a patient’s genomic tumour profile

## 3.3. Aims

- 1) Validate ctDNA sequencing through establishing the agreement between two orthogonal ctDNA analysis approaches
- 2) Optimise and validate a tissue next generation sequencing (NGS) panel and bioinformatic analysis
- 3) Validate ctDNA analysis through comparing the agreement between paired tissue and plasma sequencing

## 3.4. Results

### 3.4.1. Plasma droplet digital PCR compared to targeted sequencing of ctDNA

Of 1051 patients registered into plasmaMATCH, 1044 underwent ctDNA testing for actionable mutations. Droplet digital PCR (ddPCR) results were available for 1025 patients (Figure 3.1). In total, 800 patients underwent targeted sequencing with the Guardant360 targeted panel (364 prospective as part of the trial, and 436 sequenced retrospectively from a banked plasma

sample). Direct comparison of the targeted sequencing result with ddPCR was possible in 784 patients who underwent testing with both techniques.

*PIK3CA* alterations were common, occurring in 25.8% of the 1025 patients who underwent ddPCR. *ESR1*, *AKT1* and *ERBB2* alterations occurred in 27.7%, 4.2% and 2.7% of patients respectively (Figure 3.1). Mean allele frequency for each gene were significantly different, with *ESR1* alterations tending to occur at lower allele frequency (Figure 3.1).

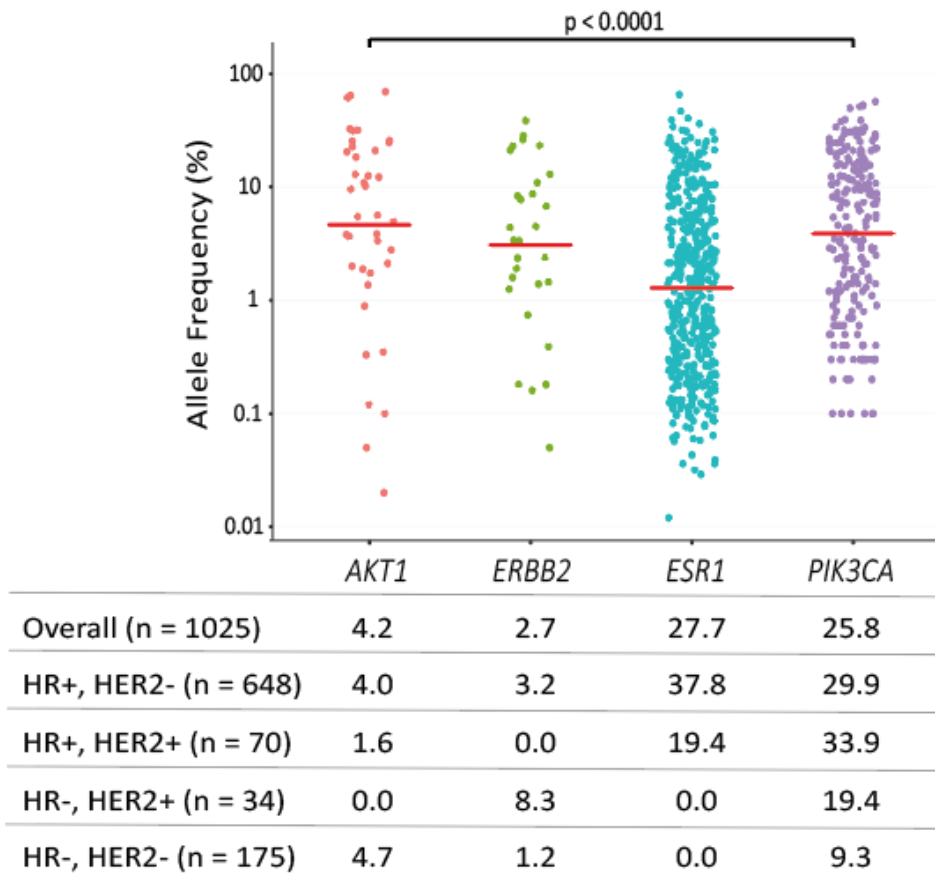


Figure 3.1. Incidence and allele frequency of targetable mutations in plasmaMATCH. *Top*, Allele fractions of potentially targetable mutations identified in 1025 patients with digital PCR ctDNA testing results. Comparison by Kruskal Wallis test. *Bottom*, Incidence of targetable mutations within each gene and breast cancer phenotype, respectively.

Overall, for each of the genes assessed in plasmaMATCH (Figure 3.1), the level of agreement for gene mutation status (mutant versus wild type) was high with kappa values ranging from 0.89 to 0.93 (Figure 3.2). Discordant calls had significantly lower allele frequency, less than 1%, as compared to concordant calls ( $p < 0.0001$ , Figure 3.2), in agreement with data published elsewhere<sup>97</sup>.

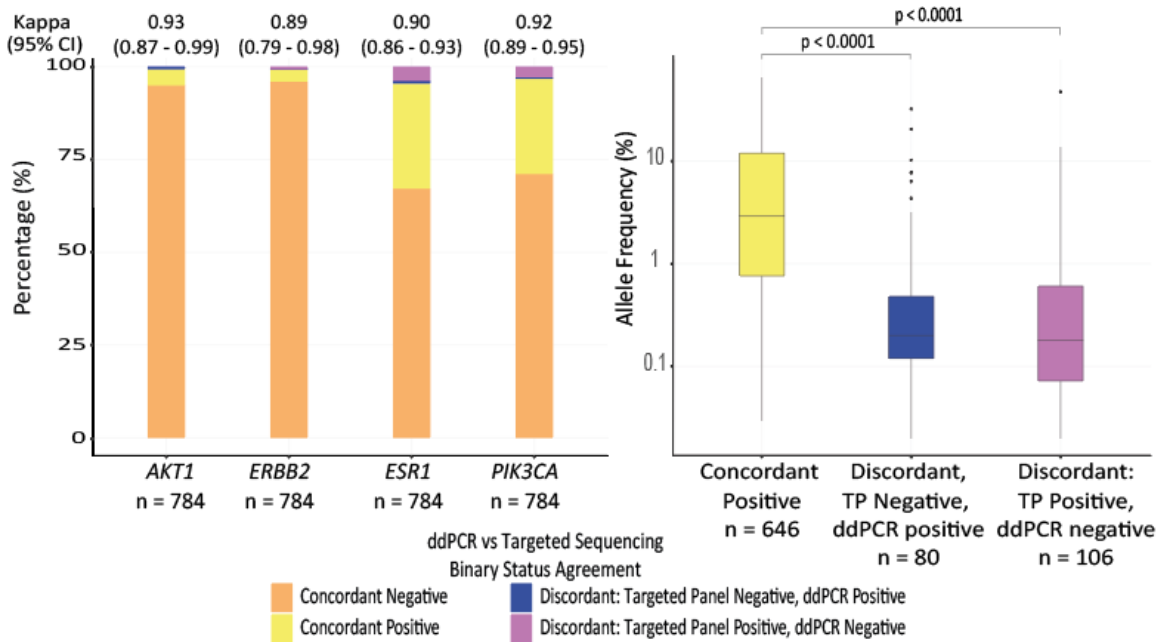


Figure 3.2. Agreement between droplet digital PCR (ddPCR) and targeted sequencing (Targeted panel, TP) for targetable mutation status in 784 patients within plasmaMATCH. *Left*, Agreement between ddPCR and targeted panel sequencing with kappa scores, *Right*, Allele frequency of concordant mutations compared to allele frequency of discordant mutations. Concordant positive allele frequency is the mean allele frequency from ddPCR and TP, otherwise the allele frequency is from the respective positive test. Comparison by Kruskal Wallis test.

When comparing the mutation status for the exact variants tested for each gene, the agreement level showed more variation (Figure 3.3A-D). For variants in *AKT1*, *ERBB2* and *PIK3CA*, agreement remained high with kappa values ranging between 0.80 and 1.0 (Figure 3.3B-D). For *ESR1*, the agreement was lower with kappa values ranging between 0.66 and 0.90 (Figure 3.3A). Overall, for all the variants tested, 77.6% of mutation calls made were concordant between the two techniques. For the remaining 22.4% of calls, 12.7% were made by targeted sequencing alone, and 9.6% were made by ddPCR alone (Figure 3.3E).

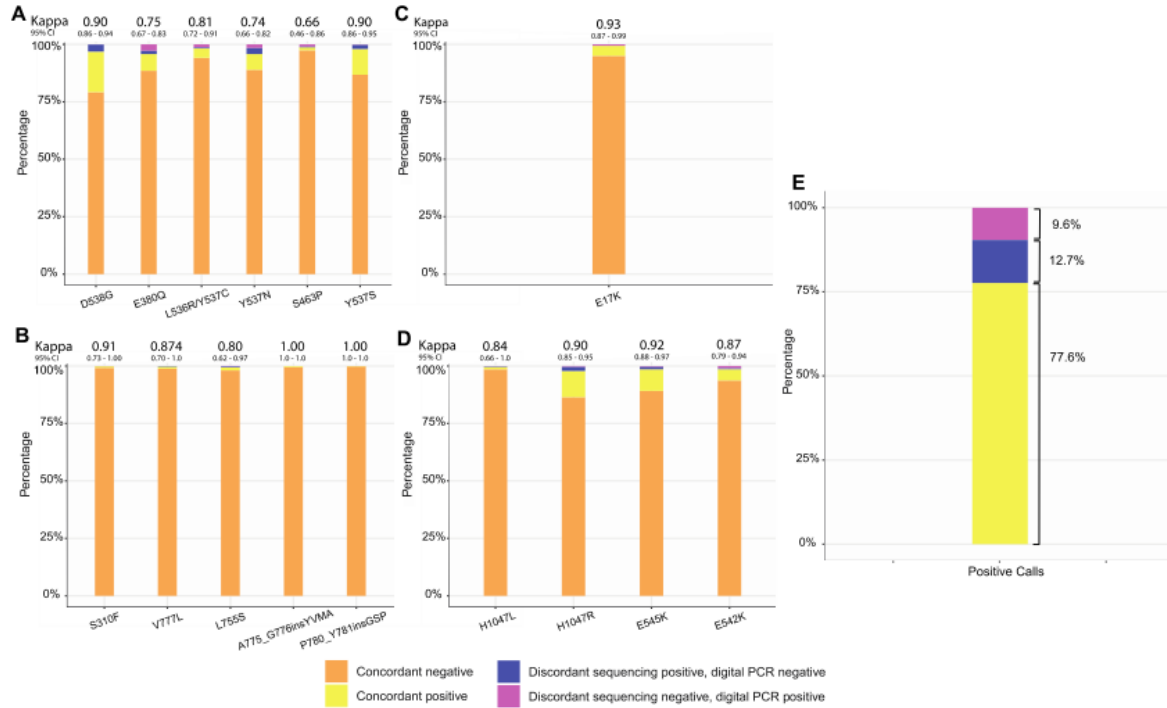


Figure 3.3. Individual mutation agreement for actionable alterations. A, *ESR1*, B, *ERBB2*, C, *AKT1*, D *PIK3CA*. E, Proportion of positive calls which were concordant or discordant.

#### 3.4.1.1.ddPCR validation of *PIK3CA* and *ERBB2* variants

To further validate targeted sequencing results, a number of *PIK3CA* and *ERBB2* variant calls by Guardant360 were validated by ddPCR. Of the *PIK3CA* variant calls, 18/22 calls were validated by ddPCR, while for *ERBB2*, 22/24 calls were validated (Figure 3.4). *PIK3CA* mutations had a lower correlation than *ERBB2* mutations ( $r=0.44$  vs  $r=0.76$ ), likely due to their significantly lower allele frequency (*PIK3CA* mean AF 0.75%, *ERBB2* mean 3.0%,  $p=0.003$ , Mann-Whitney U test).

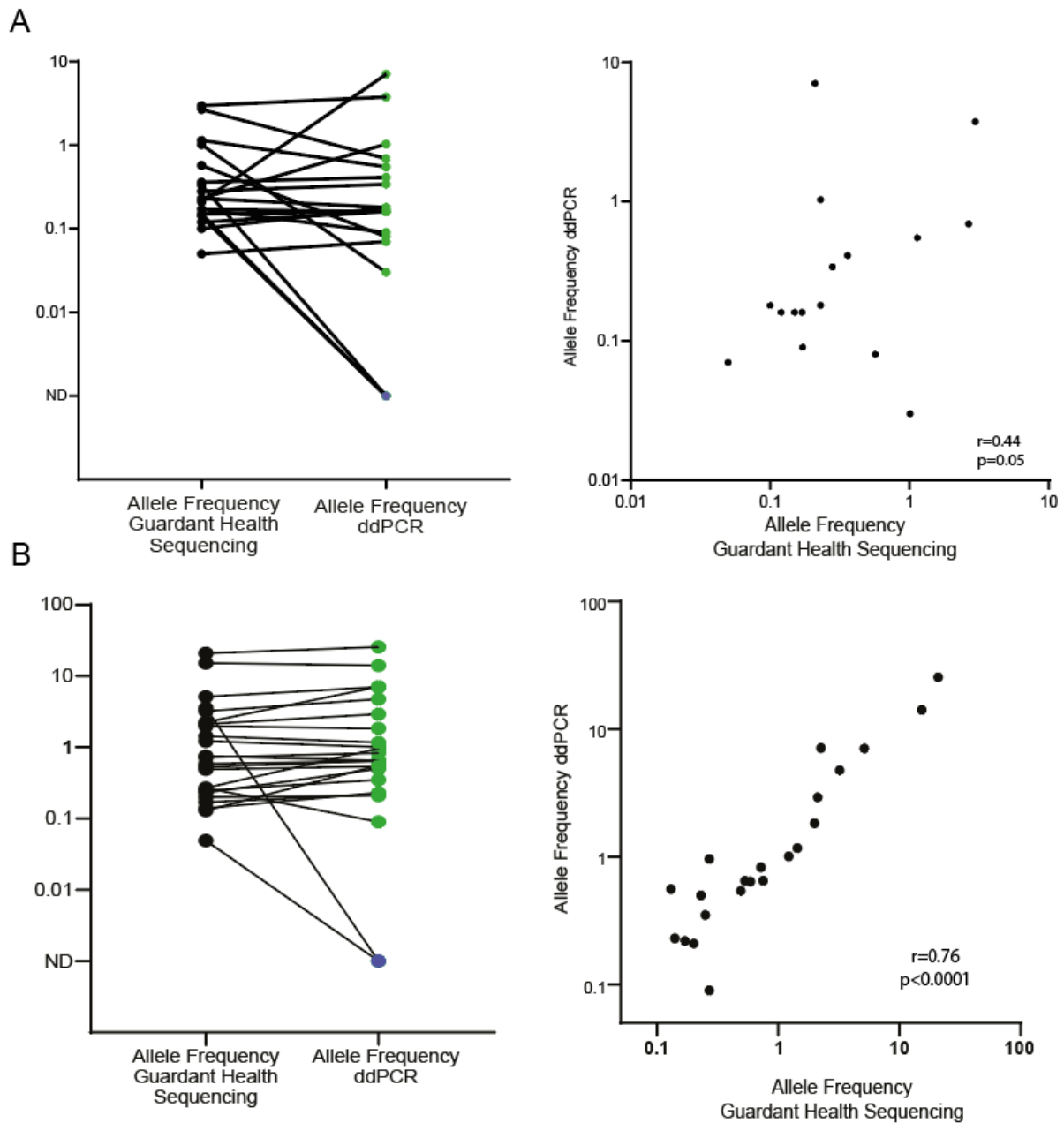


Figure 3.4. DdPCR validation of *PIK3CA* and *ERBB2* mutations. A, Association between allele frequency in ctDNA sequencing and validation analysis with plasma DNA droplet digital PCR (ddPCR), n=20 *PIK3CA* mutation assays. 16/20 (80.0%) of mutations were validated by ddPCR. Spearman correlation coefficient 0.44, p=0.05 (two-sided). ND, not detected. B, Association between allele frequency in ctDNA sequencing and validation analysis with plasma DNA droplet digital PCR (ddPCR), n=24 *ERBB2* mutation assays. 22/24 (91.7%) mutations were validated by ddPCR. Spearman correlation coefficient 0.76, p<0.0001 (two-sided). ND, not detected.



### 3.4.2. Plasma ddPCR compared to paired tissue sequencing

#### 3.4.2.1. Optimisation of library synthesis protocol for tissue sequencing

The laboratory utilises a hybrid protocol for tissue DNA sequencing that employs the initial library-synthesis steps from an Ampliseq platform protocol with the latter steps of an Illumina platform protocol bolted on. This approach allows Ampliseq multiplex libraries to be sequenced on an Illumina platform, which is thought to be superior to IonTorrent platforms with lower sequencing error. Sequencing at higher coverage gives greater ability to identify low-allele frequency mutations, at the expense of higher cost per base sequenced.

It was hypothesised that it would be possible to reduce the volumes of the reagents used in the library-synthesis protocol whilst maintaining the integrity and accuracy of the libraries produced. If proven, this would enable, a) a greater number of samples to be processed at reduced costs in library synthesis reagents, b) a potential reduction in sample input and c) to, conversely, maintain the same sample input but increase the efficacy of library preparation by maximising the interactions of molecules in a reduced volume. To test the hypothesis, libraries were produced following the laboratory protocol using the original, or ‘full’ volume reagents, and, in parallel, with reagents reduced to 0.4 of their original volume (‘miniaturised’). Libraries subsequently underwent multiplexed sequencing on an Illumina platform. This process was repeated twice in two tests, test 1 (8 DNA samples/libraries) and test 2 (16 DNA samples/libraries).

Analysis of the sequencing metrics revealed that the miniaturised libraries had similar read coverage (test 2, mean in full volume 1115 +/- 546 vs 1649 +/- 917 in miniaturised libraries,  $p=0.07$ , paired t test) or significantly higher (test 1, mean in full volume 1273 +/- 673 vs 2536 +/- 798 in miniaturised libraries,  $p=0.006$ , paired t test) than the full volume libraries. Test 1 demonstrated fewer variant calls in the miniaturised libraries compared to full volume libraries, likely secondary to higher coverage reducing the sequencing error rate (noise).

Analysis of the mutation calls (‘calls’) in test 1 demonstrated that 19 of 73 calls were made in both the full volume and miniaturised libraries (Figure 3.5). For the 40 calls made by the full volume and not by the miniaturised libraries, 17 were present in the raw sequencing data of the miniaturised libraries, and were overwhelmingly of allele frequency less than 3% (Figure 3.5). Review of the raw data revealed that many of these calls were not made by in the miniaturised libraries due to the higher coverage achieved in the miniaturised libraries. As a result, in the

miniaturised libraries many alterations did not meet the minimum bioinformatic criteria of >1% allele frequency and no strand bias, while they did in the full volume libraries secondary to the lower coverage in this sequencing run leading to a higher rate of erroneous calls. A further 11 mutation calls made singularly by the full volume libraries on review demonstrated strand bias such that the call should be considered erroneous (Figure 3.5). The miniaturised libraries were not without error, with a lower rate of erroneous calls compared to the full volume libraries, with five such calls. This demonstrates the importance of coverage in facilitating the bioinformatic pipeline to better differentiate error/noise from true calls, with the higher coverage in the miniaturised libraries assisting in reducing the error rate.

In total, there were 13 calls from the full volume sequencing set that were either present in the raw data of the miniaturised libraries but in an area of sequencing that appears error prone in the miniaturised library raw sequencing data which prevented the mutation call being made, or not present at all in the raw sequencing data. These calls are potential ‘missed’ calls by the miniaturised libraries and represent 22% of all the mutation calls made by the full volume sequencing data.

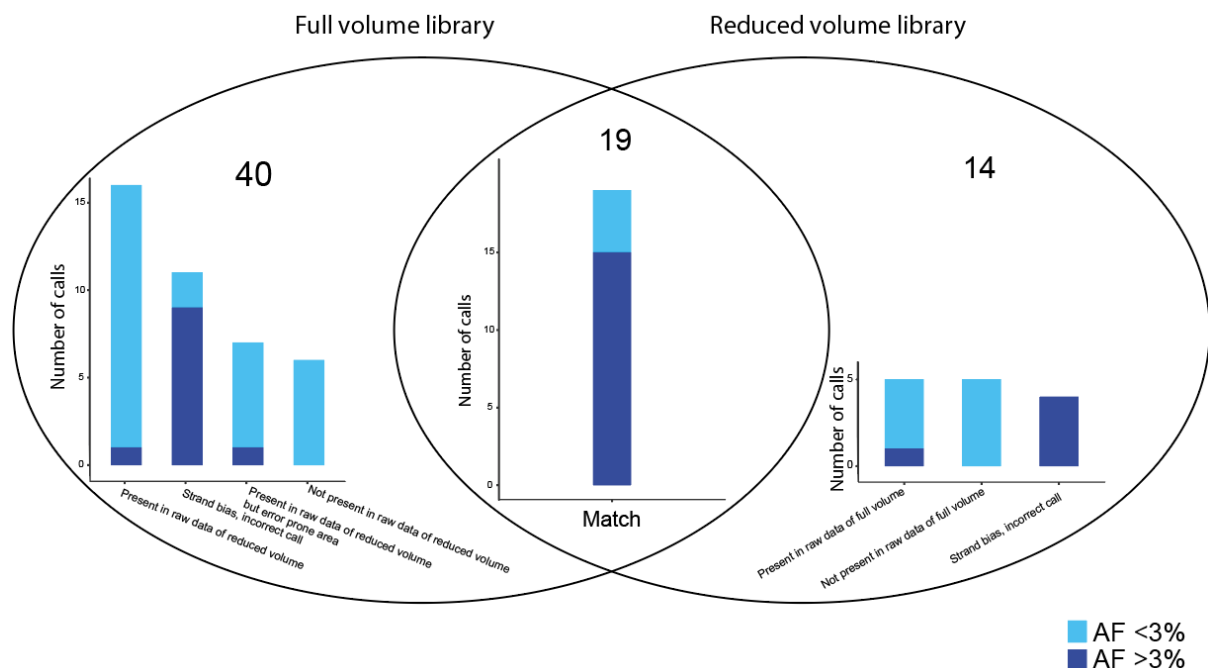


Figure 3.5. Agreement in sequencing calls made between the full and miniaturised (‘reduced’) volume libraries in test 1. AF, allele frequency.

For the second test, a minimum allele frequency for mutation calls was set at 5%. This test showed 100% concordance between the full and miniaturised libraries (Figure 3.6).

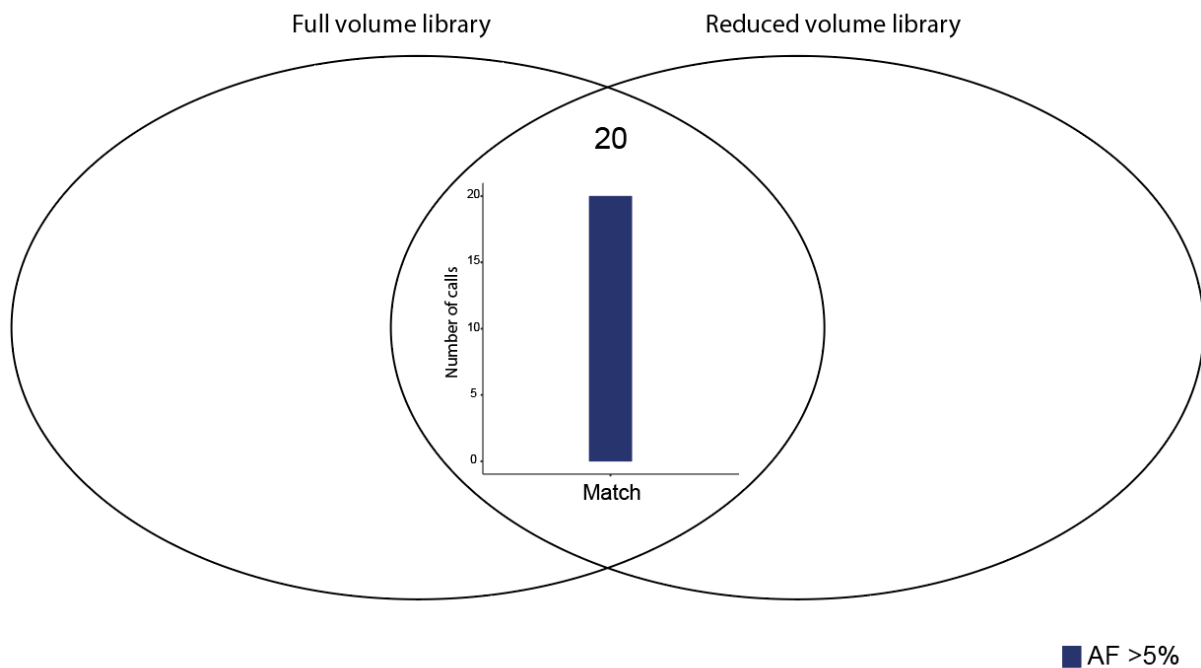


Figure 3.6. Agreement in sequencing calls made between the full and miniaturised ('reduced') volume libraries in test 2. AF, allele frequency.

When the second test results were compared to prior sequencing on a IonTorrent sequencing platform, the full volume and miniaturised libraries concurred with 100% (n=21 calls) of prior calls. One additional call was made in both the full and reduced volume sequencing relative to the prior IonTorrent sequencing, which was a splicing mutation in *MAP2K4*. The mutation appears robustly in the raw sequencing data in both the library sets, and as a splicing mutation may not have been captured by the prior IonTorrent sequencing.

Based on the results of test 1 and 2, the 'miniaturised' protocol for tissue library preparation was adopted for plasmaMATCH tissue sequencing.

#### 3.4.2.2. Optimisation of library targeted sequencing panel

The laboratory utilised an in-house designed Ampliseq amplicon panel, BCPv9, which includes amplicons covering breast cancer driver genes. The panel did not cover *NFI*, which recent data had suggested has an important role in breast cancer as a potential endocrine therapy resistance

mechanism<sup>198</sup>. The panel was therefore redesigned to include amplicons tiling *NF1*, resulting in a new amplicon panel, BCPv10 (Table 3.1).

<b>Genes</b>	<b>Genome Regions</b>
<i>CDH1</i>	<i>AKT1</i>
<i>GATA3</i>	<i>BRAF</i>
<i>MAP2K4</i>	<i>ERBB2</i>
<i>MAP3K1</i>	<i>ESR1</i>
<i>NF1</i>	<i>KIT</i>
	<i>KRAS</i>
	<i>PIK3CA</i>
	<i>PIK3R1</i>
	<i>RUNX1</i>
	<i>SF3B1</i>

Table 3.1. Genes and genome regions covered by BCPv10.

To validate the new panel, DNA samples remaining from previous ddPCR ctDNA testing within the ABC-Bio study (CCR3991) were identified. Libraries were synthesised following the previously validated ‘miniaturised’ protocol. Libraries were sequenced on an Illumina platform, and resulting mutation calls compared with prior cfDNA mutation calls made by ddPCR.

Overall, 19 DNA samples underwent library preparation, of which 16 produced a volume and quality of library that could undergo sequencing. The libraries underwent sequencing with a median coverage of 1819X per library (range 907 – 5050) and an average on-target percent of 76%. Mutation calls were compared to the mutations identified in prior ddPCR of the sample, with 9 mutation calls in agreement, and 3 mutation calls not identified by BCPv10 sequencing (Table 3.2). Review of the discordant calls revealed that two of the plasma calls were of low allele frequency (0.2 and 0.4%), below the bioinformatic detection level for targeted

sequencing using BCPv10. One mutation call had no coverage across the amplicon for that patient. Removing the two calls with low allele frequency, the adjusted sensitivity was 90%, and specificity was 100%. Based on these results, the sequencing panel BCPv10 was adopted for use in sequencing of tissue samples in plasmaMATCH.

	ABC Bio Plasma +	ABC Bio Plasma -
BCPv10 +	9	0
BCPv10 -	3	80

Table 3.2. Comparison of the mutation calls made by ddPCR (ABC-Bio Plasma) and by library preparation following the miniaturised protocol and the new targeted sequencing panel, BCPv10. +, positive mutation call, -, negative mutation call.

#### 3.4.2.3. Optimisation of tissue sequencing bioinformatic pipeline

An important consideration when analysing raw sequencing data is establishing the error rate of the sequencing approach employed. This knowledge enables a minimum allele frequency threshold to be set in the bioinformatic pipeline that avoids falsely identifying error as “true” mutation calls. To establish the error rate, which would inform where to set a threshold, 11 primary triple negative breast cancer tissue DNA samples were selected for sequencing, all of which were derived from FFPE tissue. Primary triple negative samples were selected due to the low likelihood of any *ESR1* mutations being present<sup>58</sup>, allowing assessment of the baseline error rate, or ‘noise’, present in FFPE tissue, which is anticipated to be higher than that of fresh frozen tissue. The DNA samples underwent library preparation using the miniaturised protocol. A pileup analysis was undertaken to identify mutation calls within *ESR1*. Pileup analysis revealed that error calls across *ESR1* amplicon coverage areas occurred at allele frequencies less than 1% (Figure 3.7).

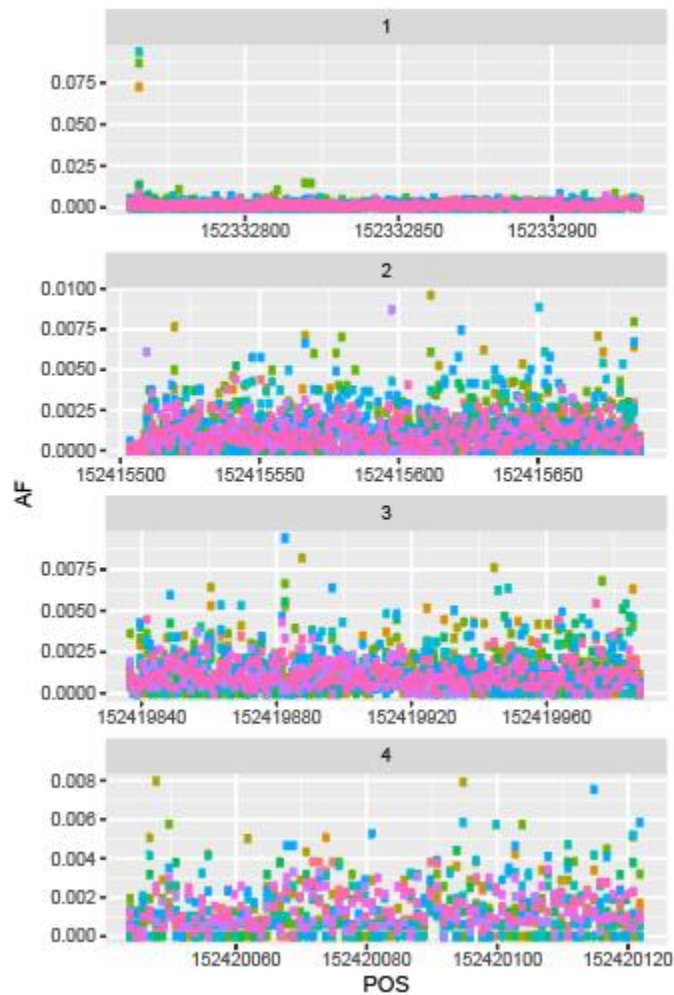


Figure 3.7. Pileup analysis of mutation calls across the 4 amplicons (1 – 4) tiling *ESRI* from primary triple negative breast cancer samples. Mutation calls from each sample are represented in a different colour. AF, allele frequency, POS, position across the amplicon (tiled across the x-axis). Figure courtesy of Dr Ros Cutts.

This result was corroborated by assessment of the allele frequency calls of a single batch of plasmaMATCH tissue sequencing, which included DNA derived from a mix of 9 FFPE and 7 fresh frozen tissue samples. Pileup analysis revealed that for both the TNBC set and the plasmaMATCH batch, the baseline error rate was consistently <1% (Figure 3.8). Error was higher for C>T alterations, which is to be expected due to FFPE artefacts likely caused by the chemical treatment of samples<sup>199-201</sup>. Additionally, some samples showed higher error rate than others, however the error rate stayed consistently below 1%. To achieve stringency and reduce the risk of false positive calls, a minimum allele frequency of 2% was set in the bioinformatic pipeline to identify mutation calls in the tissue sequencing dataset using pileup analysis.

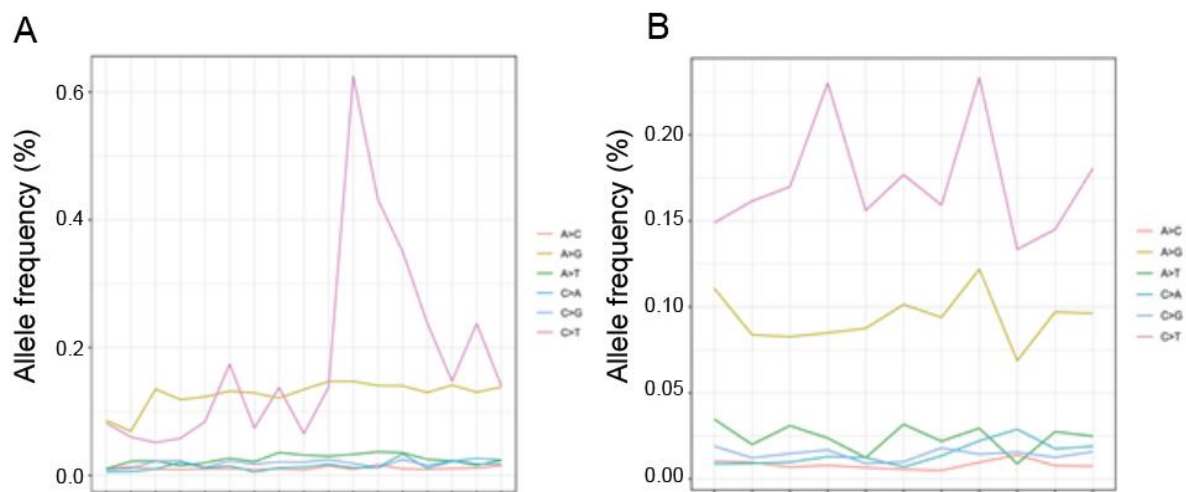


Figure 3.8. Base-specific mean error rate for the plasmaMATCH sequencing batch (A) and the TNBC sequencing batch (B). Samples are on the x-axis. Figure courtesy of Dr Ros Cutts.

Pileup analysis enables assessment of the raw data for the number of reference and alternate reads at every loci covered by the panel. The mutation calls made by pileup analysis were compared with those of MuTect and VarDict, which are the two mutation callers used within the in-house bioinformatic pipeline (VariTAS, Methods 2.1.11). The mutation calls made by VarDict were frequently not identified by the pileup analysis or by MuTect (Figure 3.9). SNV calls made by either pileup analysis or MuTect were therefore included whilst mutation calls made solely by VarDict were excluded. The presence of indels was established using MuTect.

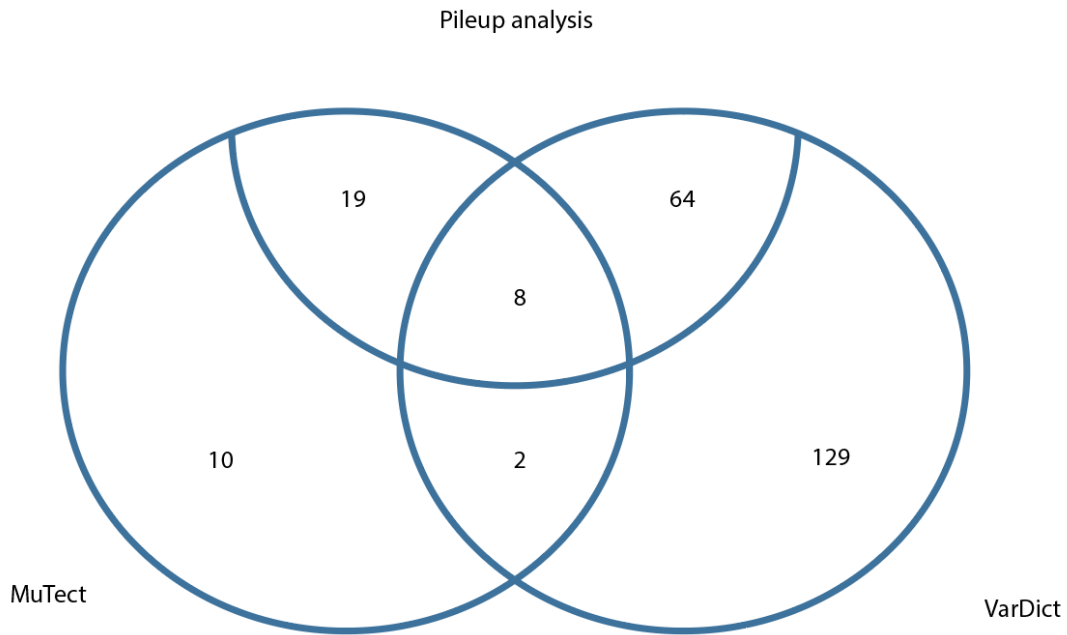


Figure 3.9. Analysis of mutation calls made by MuTect, VarDict and the agreement with each other and the pileup analysis for the sequencing run analysed. Data on mutations called by pileup alone, if any, was unavailable.

#### 3.4.2.4. Plasma ddPCR compared to tissue sequencing

The optimised tissue sequencing panel, protocol and bioinformatic pipeline were used to sequence tissue DNA from baseline and archival tissues obtained within the plasmaMATCH trial. In total, 80 patients had tissue sequencing results, of whom between 76 and 77 patients (one patient had a failed *AKT1* p.E17K ddPCR assay) had available plasma ddPCR and targeted sequencing results. When comparing binary gene level mutation status, the sensitivity of both ddPCR plasma testing and plasma targeted sequencing with the tissue sequencing was high, ranging between 88.2% and 96.8% (Figure 3.10). Specificity was lower for both ddPCR and targeted sequencing, ranging between 40.0% and 98.5%. When just contemporaneous samples were included (plasma and tissue taken within 60 days), specificity improved to between 57.1% and 100%.



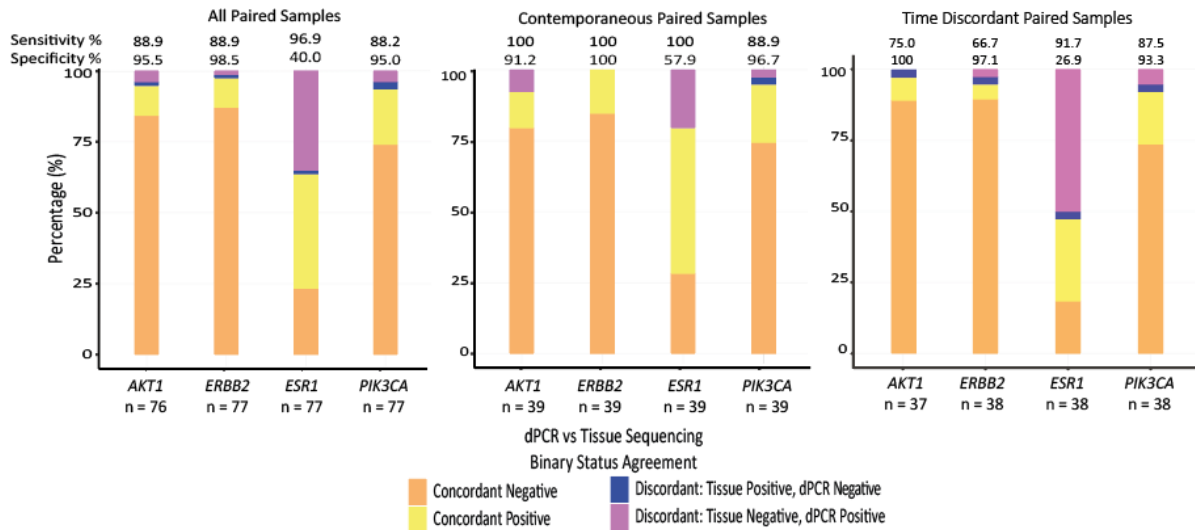


Figure 3.10. Agreement between ddPCR ctDNA testing and advanced disease tissue sequencing. Agreement in all paired samples (*left*), contemporaneous ctDNA and tissue samples (*middle*), and time discordant samples where tissue was taken  $\geq 60$  days prior to ctDNA sample (*right*).

### 3.5. Discussion

The aim of this chapter was to validate the use of ctDNA analysis to genomically profile a patient’s cancer. This was achieved by comparison of two orthogonal ctDNA testing approaches, and through comparison of ctDNA results with paired tissue sequencing.

The high level of agreement for the two orthogonal ctDNA testing approaches demonstrated here increases confidence in the validity of ctDNA analysis by either approach. Similarly, the sensitivity and specificity of ctDNA analysis, whether by ddPCR or targeted sequencing, was generally high when comparing to the ‘gold standard’ tissue sequencing approach. Notably, in both the ddPCR/targeted sequencing and ctDNA/tissue comparison, the agreement and specificity was lower for alterations within *ESR1* than other genes.

There are several factors contributing to the lower agreement and specificity for *ESR1*. Firstly, it is well established that *ESR1* alterations arise in the metastatic setting as a resistance mechanism to aromatase inhibitor therapy<sup>72</sup>. Time delays between sample collections are more likely to cause disparity in the mutation status due to the increased potential for temporal acquisition of the mutation. Indeed, when just plasma/tissue samples taken within 60 days of each other are considered, the *ESR1* specificity of ctDNA increased from 40% to 57.9%. A second important factor is the tendency for *ESR1* alterations to be subclonal and occur at low

allele frequency<sup>72</sup>. Subclonal alterations are more likely to be metastatic-clade, and subsequently ‘missed’ by single-site tissue biopsy while identified in the ctDNA. They are also more likely to occur at allele frequencies below the limit of detection, and be affected by the issue of stochastic sampling. Similarly, Ravazi *et al* found that the detection rate of mutations in the plasma was significantly correlated with the cancer cell fraction at which they occurred in the paired tumour<sup>94</sup>, demonstrating that for subclonal mutations detection is less accurate. Finally, the threshold of detection for alterations identified in tissue was set at a minimum of 2%. The patterns of *ESRI* mutations, in terms of inter-tumoural vs intra-tumoural heterogeneity, have not been well established. One study which profiled the genome of multiregional metastasis in 10 patients with ABC revealed that most mutations are either ‘stem’, or present in every cell, or ‘clade’, being common to more than one metastatic deposit rather than private to a singular deposit<sup>52</sup>. Whether this pattern exists with *ESRI* mutations is not well established. An autopsy study identified that in a patient with multiple *ESRI* mutations, they mostly occurred in different subclones, although one pair of mutations overlapped<sup>43</sup>. If *ESRI* mutations tend to be clonally dominant within a metastasis whilst absent in metastases occurring in parallel, then the mutation may be ‘missed’ by a single-site biopsy depending on the site biopsied. Conversely, if *ESRI* mutations tend to be homogeneously present at a subclonal level in disparate metastases, then a minimum allele frequency 2% may mean tissue biopsy is unable to detect the subclonal *ESRI* mutations. With the above factors particularly influencing the identification of *ESRI* mutations, it is likely that ctDNA testing represents the ‘gold-standard’ approach to identify *ESRI* mutations rather than single-site tissue biopsy.

One notable finding was that the binary, gene level agreement was higher than the exact mutation agreement for the ddPCR/targeted sequencing comparison. This was particularly marked for *ESRI* mutations. One consideration here is the way in which the ddPCR assessment is made using multiplex assays, with several *ESRI* mutations tested for in the same well. Differentiating which mutation is which is based upon the x/y location of the FAM (mutant) and HEX (wild type) droplets as compared to a reference well, with this method being prone to subjective interpretation of the result by the operator. Droplet migration, due to variation between assay batches, and droplet streaking, due to shredding of droplets, can infrequently occur, leading to a positive result being attributed to the incorrect variant. This is compounded by the nature of *ESRI* mutations to be polyclonal, making differentiating which variant is present more challenging when multiple variants are present. The clinical relevance of incorrect

variant call is not yet fully understood. There is some *in vivo* data suggesting that *ESR1* p.Y537S mutations exhibits a higher level of resistance to fulvestrant than other activating *ESR1* mutations<sup>126</sup>. It may become more apparent, for *ESR1* and mutations within other genes, that the exact mutation has prognostic and predictive significance. In this situation, the limitations of multiplex ddPCR assays mutation call should be reconsidered.

### 3.6. Conclusion

Overall, the high level of agreement between two orthogonal ctDNA analysis techniques coupled with reasonable sensitivity and specificity demonstrated in this Chapter reassures that ctDNA analysis is a valid approach for genomic profiling. Optimisation of the tissue sequencing protocol and bioinformatic pipeline enabled a greater number of samples to be analysed at greater depth and with increased accuracy of mutation calls, allowing for the comparison to be made. The data presented here suggests that analysis of ctDNA may be more limited in identifying mutations occurring at low allele frequency where sensitivity was lower. On the other hand, ctDNA analysis has potentially a greater ability to identify metastatic-clade subclonal alterations that are missed by single-site tissue biopsy, as suggested by the number of alterations that were identified in ctDNA but missed in tissue, particularly for *ESR1* mutations.

## 4. Chapter 4. Clinico-pathological associations of ctDNA

### 4.1. Introduction

To bring ctDNA analysis into clinical practice, we must define important clinical features which associate with the presence of ctDNA in the circulation. This will help physicians understand which patient groups may benefit from the test whilst highlighting the patient groups in whom the test may be inadequate. Of much clinical relevance is which patient groups are most likely to harbour a potentially therapeutically targetable mutation in their ctDNA, and to date this has not yet been well defined in ABC. With the plasmaMATCH dataset, we are uniquely situated to be able to investigate the associations between clinico-pathological features and ctDNA presence and nature.

### 4.2. Hypothesis

We can define the clinical and pathological features associated with ctDNA results

### 4.3. Aims

- 1) Define which patient groups are more likely to have detectable ctDNA
- 2) Understand the association between ctDNA purity and number of alterations with clinico-pathological features
- 3) Establish the clinico-pathological associations of targetable alterations
- 4) Investigate organotropism of common breast cancer mutations
- 5) Investigate whether ctDNA analysis can be used to identify *HER2* amplification

### 4.4. Results

Of 1051 patients enrolled into plasmaMATCH, baseline pre-treatment ctDNA targeted sequencing results were available for 800 patients (364 prospectively and 436 retrospectively). Of the patients with available targeted sequencing, 64.4% (n=515) had HR+ HER2- disease, 9.1% (n=72) were HER2+, and 17.3% (n=138) had TNBC.

#### 4.4.1. Clinico-pathological associations of a negative ctDNA test

Of the 800 patients with targeted sequencing results, 743 (92.9%) had an oncogenic alteration (SNV, indel, fusion and/or copy number change) identified. There was no significant difference in the proportion of patients with or without a ctDNA alteration according to breast cancer subtype, histological subtype or disease burden (Table 4.1). There was a significant difference in the number of prior lines of treatment in those with a ctDNA alteration compared to those without ( $p=0.01$ , Table 4.1), with patients who had a higher number of lines of prior treatment more likely to harbour a ctDNA alteration.

Clinical Characteristic		Number with alteration	Total number	Proportion with alteration (%)	P value
Breast Cancer Subtype	HR+ HER2-	484	515	94.0	0.10
	HR+ HER2+	40	46	87.0	
	HR- HER2+	22	26	84.6	
	TNBC	130	138	94.2	
	Unknown	67	75	89.3	
Histology	Ductal	534	577	92.5	0.83
	Lobular	75	79	94.9	
	Other	43	47	91.5	
	Not known	91	97	93.8	
Disease burden	Visceral	586	627	93.5	0.41
	Soft tissue/nodal	131	143	91.6	
	Bone	10	11	90.9	
	Not known	16	19	84.2	
Number of lines of prior treatment	0	67	77	87.0	0.01
	0-1	374	409	91.4	
	3-4	209	218	95.9	
	5+	93	96	96.9	

Table 4.1. Clinico-pathological characteristics of patients with and without a ctDNA alteration. p values from Chi-squared test.

#### 4.4.2. Clinico-pathological associations of number of genome alterations and ctDNA purity

The purity of a ctDNA sample, or the amount of ctDNA relative to the amount of cell free DNA (cfDNA), is an important consideration when reviewing targeted sequencing results. On a patient level, physicians must be able to review a targeted sequencing report with an understanding of how likely the lack of a mutation being present is due to clinico-pathological features versus true negativity. For the targeted sequencing data, the maximum variant allele frequency, or mVAF, acts as a proxy for tumour purity. This is under the assumption that the mutation with the greatest allele frequency represents a truncal mutation present in all cancer cells. The allele frequency of this variant can then be used as a ‘marker’ to denote the proportion of cancer DNA relative to germline DNA in the sample.

HR+ HER2- disease was associated with significantly more SNVs/indels than TNBC ( $p=0.007$ , Figure 4.1). Both the number of lines of treatment overall and number of lines of chemotherapy were also associated with increased number of SNVs/indels and mVAF (Figure 4.1). Metastatic disease site was also found to be a significant predictor for increased mVAF, with patients with visceral disease tending to have higher mVAF than those with soft tissue/nodal disease ( $q=0.002$ , Figure 4.1).

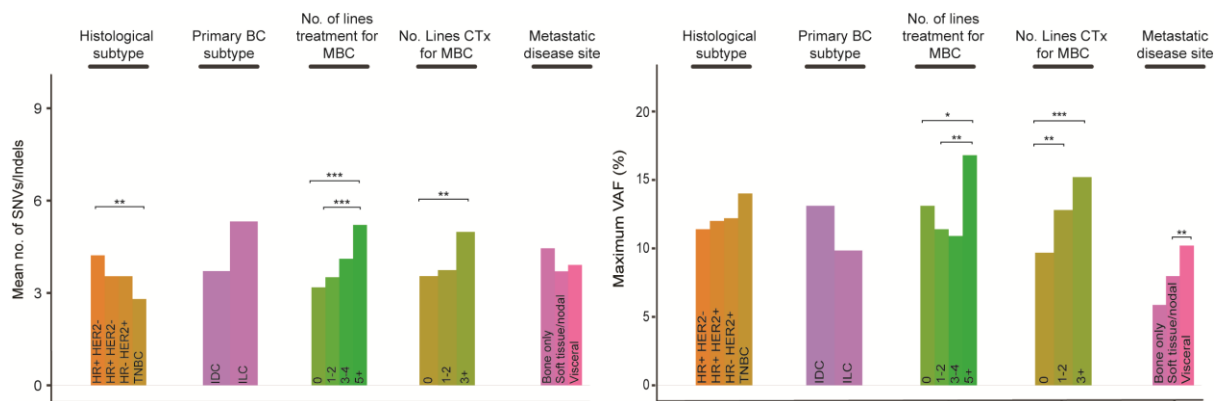


Figure 4.1. Clinical and pathological associations of breast cancer mutation profile. Association of number of mutations (SNVs/indels, left) and the maximum variant allele frequency (mVAF, right, as a proxy of ctDNA purity) with indicated clinical and pathological features. p values from pairwise two sided Kruskal–Wallis test with correction for multiple testing (number of mutations: HR+ HER2- vs TNBC  $q=0.008$ ; 0 vs  $\geq 5$  lines of treatment  $q=0.0005$ , 1–2 vs  $\geq 5$  lines of treatment  $q=0.0005$ ; 0 vs  $\geq 3$  lines of chemotherapy  $q=0.003$ . mVAF: 0 vs  $\geq 5$  lines of treatment  $q=0.03$ , 1–2 vs  $\geq 5$  lines of treatment  $q=0.006$ ; 0 vs 1–2 lines of chemotherapy  $q=0.003$ , 0 vs  $\geq 3$  lines of chemotherapy  $q=0.0003$ ; soft tissue/nodal vs visceral disease  $q=0.002$ ). MBC, metastatic breast cancer; CTx, chemotherapy; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma.

#### 4.4.3. Clinico-pathological associations of targetable mutations

The incidence of *PIK3CA* alterations was significantly associated with breast cancer phenotype and number of lines of treatment in the metastatic setting ( $p < 0.0001$  and  $p = 0.006$ , respectively, Figure 4.2). The incidence of *ESR1* alterations was similarly associated with breast cancer phenotype and number of lines of treatment in the metastatic setting ( $p < 0.0001$  and  $p < 0.0001$ , respectively), and additionally associated with disease site ( $p = 0.003$ ). *ERBB2* alterations were more common in lobular than in ductal breast cancer ( $p < 0.0001$ )<sup>202,203</sup>. Microsatellite instability (MSI-high) was identified in 1.1% of patients.

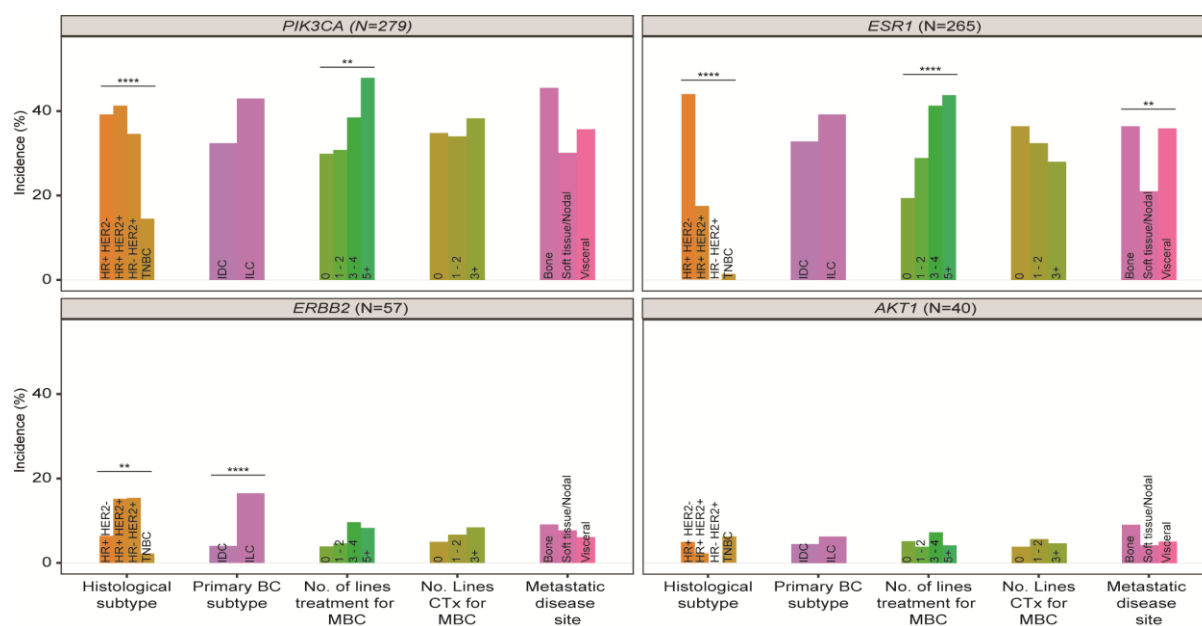


Figure 4.2. Association of clinical and pathological features with pathogenic alterations in the four targetable genes in plasmaMATCH: *PIK3CA*, *ESR1*, *HER2* (*ERBB2*) and *AKT1*. p values from Chi-squared test (*PIK3CA*: histological subtype  $p < 0.0001$ , lines of treatment  $p = 0.006$ ; *ESR1*: histological subtype  $p < 0.0001$ , lines of treatment  $p < 0.0001$ , disease site  $p = 0.003$ ; *HER2*: histological subtype  $p = 0.004$ , primary breast cancer subtype  $p < 0.0001$ ).

*ERBB2* mutations were found more commonly in *HER2+* breast cancers with increasing lines of *HER2*-directed therapy ( $p = 0.04$ , Figure 4.3).

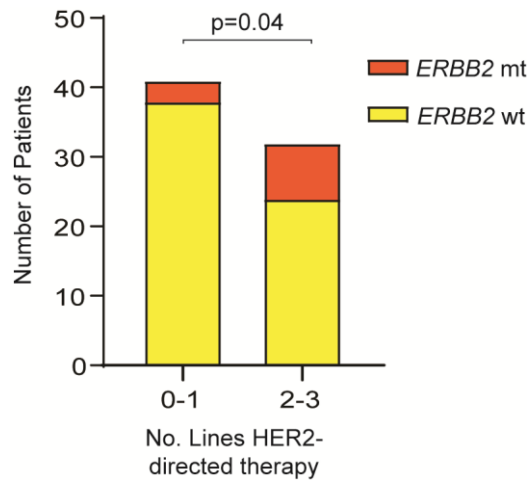


Figure 4.3. *ERBB2* mutation incidence in patients with HER2+ breast cancer, by line of therapy. 0–1 lines of therapy mutation incidence 7.3% (3/41) and 2–3 lines of therapy mutation incidence 25% (8/32)  $p=0.04$ , Chi-squared test. mt, mutant; wt wild type.

#### 4.4.4. Organotropism of common breast cancer mutations

In this cohort of 800 ABC patients, association between disease site and presence of gene alterations, or organotropism, was explored. Overall, alterations within *TP53*, *GATA3*, *ESR1* and *PIK3CA* showed a tendency for organotropism which was largely effected by breast cancer subtype associations (Figure 4.4). Bone disease demonstrated a positive association with *ESR1* and *GATA3* alterations ( $q<0.0001$  and  $q=0.0009$ , respectively), while *TP53* was negatively associated ( $q=0.02$ ). Liver disease was positively associated with *ESR1* alterations ( $q<0.0001$ ) but negatively associated with *TP53* alterations ( $q=0.002$ )<sup>204</sup>. In HR+ HER2- disease, *ESR1* mutations were positively associated with liver and bone disease ( $q=0.004$  and  $q=0.02$ , respectively, Figure 4.4). TNBC did not show a significant pattern of organotropism after correction for multiple testing (Figure 4.4).

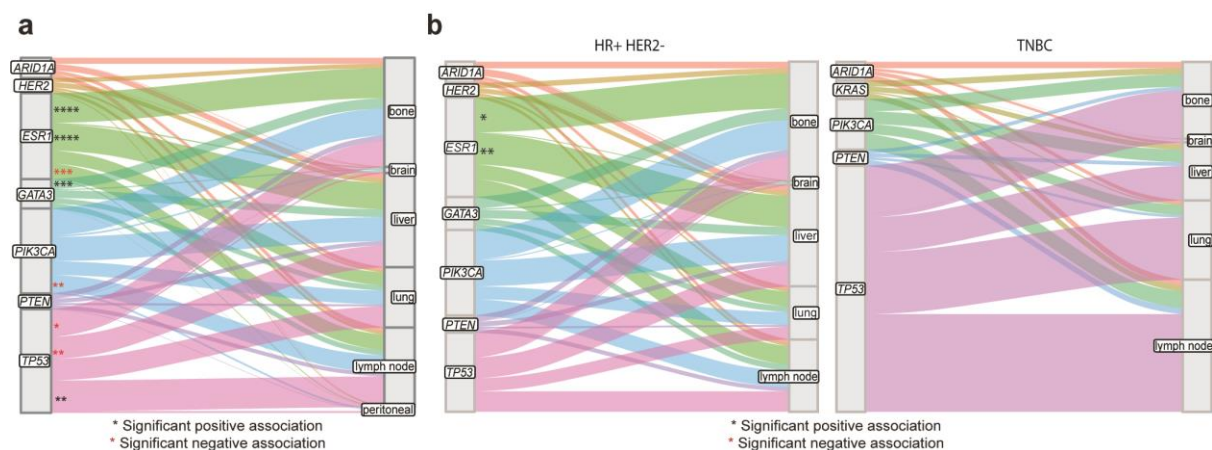




Figure 4.4. Organotropism of mutations in ABC. A, Association of mutations in indicated genes with sites of metastasis. p values from false discovery corrected Fisher's exact test (*ESR1*: bone  $q < 0.0001$ , liver  $q < 0.0001$ , lymph node  $q = 0.0001$ ; *GATA3*: bone  $q = 0.0009$ ; *PIK3CA*: lymph node  $q = 0.005$ ; *TP53*: lymph node  $q = 0.002$ , liver  $q = 0.002$ , bone  $q = 0.02$ ). B, Association of mutations in indicated genes with sites of metastasis in (left) HR+ HER2- and (right) TNBC. p values from false discovery corrected Fisher's exact test (HR+ HER2- *ESR1*: liver  $q = 0.004$ , bone  $q = 0.02$ ).

#### 4.4.5. ctDNA analysis of *ERBB2* copy number alteration and HER2 amplification

Currently HER2 amplification is established via immunohistochemistry (IHC) of a tissue biopsy in the primary breast cancer setting. However, approximately 21% of patients lose HER2 amplification on metastatic progression, while around 10% gain HER2 amplification<sup>205</sup>. Metastatic disease is not routinely biopsied to assess for change in HER2 amplification status despite there being significant therapeutic implications. ctDNA assessment of *ERBB2* amplification status would be advantageous in enabling a non-invasive approach to identify this important clinical parameter.

In the plasmaMATCH patients, the *ERBB2* amplification status identified in ctDNA was compared to the most recently available HER2 tissue biopsy status to assess sensitivity and specificity of ctDNA analysis. While the ctDNA assessment of *ERBB2* amplification status was found to be highly specific (98%), ctDNA had a sensitivity of just 50% (threshold copy number 2.2, Figure 4.5). The mean *ERBB2* copy number of HER2 positive/amplified cancers was significantly higher than those with HER2 negative/non-amplified disease (9.9 versus 2.1,  $p < 0.0001$ , Figure 4.5).

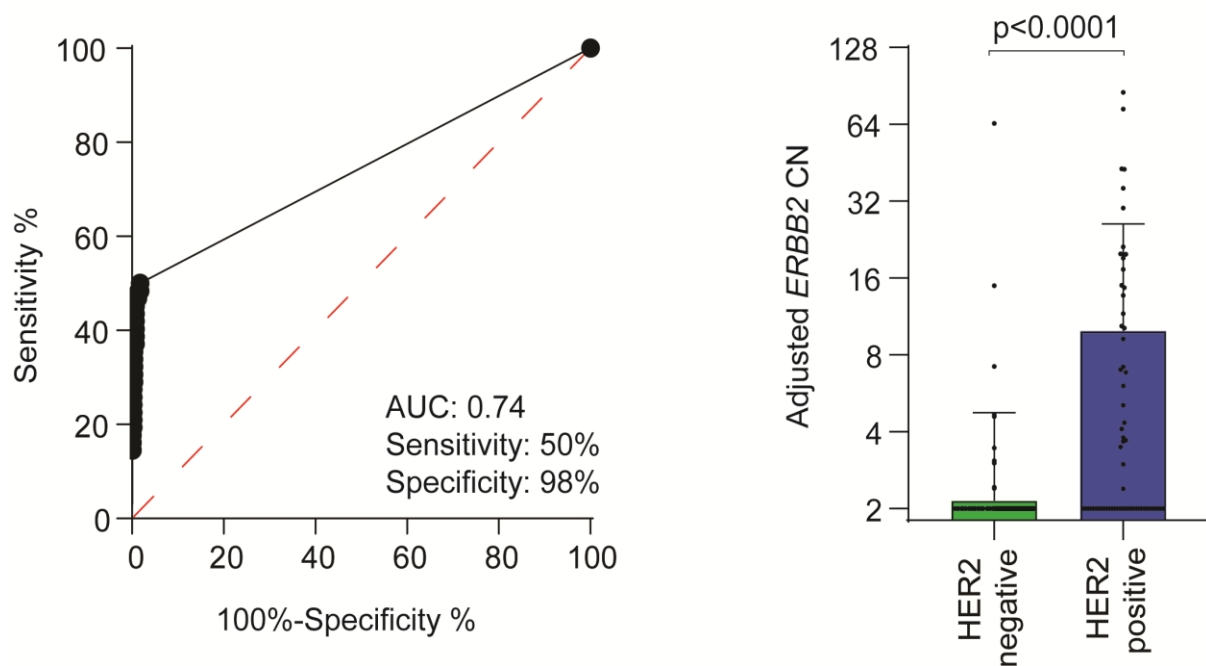


Figure 4.5. Adjusted *ERBB2* copy number (CN) in targeted sequencing, in patients with tissue assessed HER2+ (amplified, n = 72) and HER2- (non-amplified, n = 605) cancers. *Left*, receiver operator curve of adjusted *ERBB2* plasma copy number. *Right*, *ERBB2* plasma copy number adjusted for purity. Data are presented as mean + SD. The p value indicated is derived from a two-sided Mann–Whitney U-test.

## 4.5. Discussion

Analysis of this large dataset of 800 patients with ABC has revealed several important features that influence the clinical utility of the test. A major finding is the identification of a patient group less suited to this approach: those with less advanced disease. This is supported firstly by the increased number of patients with no identifiable alterations with fewer lines of prior treatment (Table 4.1). Secondly, both the mean number of SNVs/indels and the mVAF were significantly fewer in patients with fewer prior lines of treatment overall and chemotherapy lines (Figure 4.1). Finally, patients who had disease limited to soft tissue and nodal disease had significantly lower mVAF than patients with visceral disease (Figure 4.1). The amount of ctDNA identified in the blood is known to correlate with disease burden<sup>28,94,206</sup>. In early breast cancer, tumour purity has been found to correlate with tumour size and the Ki67 mitotic index<sup>207</sup>. The clinical data presented here corroborates this and identifies that genomic profiling via ctDNA is less suited to ABC patients with limited disease burden in whom the chance of a false negative result is increased.

A second finding was that, in this dataset, patients with HR+ HER2- disease had significantly more alterations than those with TNBC, which is converse to what has been previously published<sup>208</sup>. Prior work demonstrated that metastatic breast cancer has a higher tumour mutational burden than early breast cancer<sup>35,208</sup>, and that hypermutated tumours have a dominant APOBEC mutational signature<sup>208</sup>. In metastatic breast cancer, the APOBEC signature is particularly associated with HR+ breast cancer and also enriched in metastatic disease compared to primary disease<sup>38</sup>. Activation of APOBEC has been associated with secondary resistance to hormonal therapy<sup>209</sup>. Thus, this cohort of heavily pre-treated HR+ HER2- ABC patients could represent a group who have a highly mutated disease, potentially as a result of prior hormonal therapy and subsequent activation of APOBEC mutagenesis processes.

The clinico-pathological correlates of targetable mutations were explored using ctDNA analysis. The enrichment of *PIK3CA* and *ESR1* alterations in HR+ breast cancer is in concordance with prior studies<sup>33,36</sup>. In the case of mutations within *ESR1*, these alterations are known to arise as a result of prior hormonal therapy<sup>58,81</sup>, and are rare in TNBC<sup>81</sup>. *ESR1* mutations have also previously been associated with liver<sup>34,35</sup> and bone disease<sup>58</sup>. Meanwhile *ERBB2* mutations were associated with lobular breast cancer<sup>202,203</sup> and, in a novel finding, with HER2 positive/amplified disease with incidence increasing with number of lines of therapy (Figure 4.3). This finding suggests acquisition of *ERBB2* mutations as a mechanism of resistance to prior HER2 targeted therapies. This theory is supported by prior *in vitro* work which demonstrated that HER2 positive cells with activating *ERBB2* mutations were resistant to lapatinib but remained sensitive to neratinib<sup>210,211</sup>, and were enriched in metastatic HER2 amplified samples compared to primary<sup>211</sup>. Later *in vivo* work has corroborated these results<sup>141</sup>. This finding supports a role for neratinib as a novel treatment strategy for HER2+ resistant disease.

Microsatellite instability (MSI) was identified in 1.1% of the cohort, similarly to what has been identified in another breast cancer cohort<sup>38</sup>. Testing of ctDNA for MSI using Guardant360 has previously been validated against tumour tissue testing, with an overall accuracy of 98.4% and positive predictive value of 95%<sup>212</sup>. There is increasing evidence in other cancers, and particularly colorectal cancer, that patients with mismatch repair deficient cancers, or MSI high, have a high response rate to subsequent immunotherapy<sup>213-215</sup>. This is yet to be confirmed

in breast cancer cohorts, but if confirmed suggests that ctDNA analysis could be used to identify these patients.

Finally, the sensitivity and specificity of *ERBB2* copy number in ctDNA was assessed compared to patient's most recent HER2 tissue amplification status, identifying a sensitivity and specificity of 50% and 98% respectively. Copy number in plasma is a product of the number of copies present and the purity of the sample. Therefore a sample with high amplification but low purity could have the same amplification level as a low copy number in a high purity sample. Guardant Health apply a formula to account for this (Methods section 2.1.12.9), which takes into account the maximum VAF as a proxy measure of sample purity. Despite this adjustment, ctDNA assessment of copy number is limited by the issue of plasma purity. In ABC, approximately 65.5% of samples have a purity of less than 10% and 37.1% of patients having a purity of less than 3%<sup>85</sup>, demonstrating that low purity will influence a substantial proportion of patients and make copy number assessment challenging.

The data presented here suggests that ctDNA assessment only detects half of all true HER2 amplifications. It should be noted that the gold standard IHC testing has a false positive rate of around 6%, and a false negative rate of <2%<sup>216</sup>, which may account for a small degree of discrepancy. A further important factor is that approximately 10.3% of patients experience a change in HER2 status between primary and metastatic disease<sup>205</sup>. The comparison undertaken here was of ctDNA-defined *ERBB2* amplification versus the most recent tumour biopsy, which may not have been contemporaneous with the ctDNA sample. Thus there is potential for there to be a small degree of true discordance. In one large scale study (n=810 samples) which compared the concordance between HER2 status as defined by IHC with *ERBB2* amplification inferred from sequencing data the same tumour sample, the concordance was high at 98%<sup>34</sup>. This suggests the discrepancy between ctDNA assessed *ERBB2* copy number status versus tissue HER2 amplification status arises, in the most part, due to the poor ability to identify *ERBB2* copy number increase in ctDNA. Temporal changes in ctDNA status may have a small influence, while any purported lack of association between sequencing-defined *ERBB2* copy number change and tissue HER2 amplification status is not likely to be a factor. Nevertheless, in colorectal and gastric cancer, plasma *ERBB2* copy number has been found to positively associate with response to HER2 targeted therapy<sup>197,217</sup>. Conversely, in ABC the opposite was found, with high plasma *ERBB2* copy number associating with shorter median PFS on subsequent trastuzumab-based therapy<sup>218</sup>. This could be due to a more aggressive phenotype

of high *ERBB2* copy number disease, but it is important to note that disease burden and its influence on ctDNA purity will have been a confounder in this study. The data presented here suggests that while plasma *ERBB2* copy number could be used as a screening tool in ABC, it is not yet sensitive enough to replace the gold standard tissue-based IHC approach. The prognostic and predictive utility of ctDNA-assessed HER2 amplification in ABC also needs to be defined within a range of HER2-directed therapies in order to clinically validate the test.

#### 4.6. Conclusion

Analysis of the ctDNA results from 800 patients has enabled elucidation of clinico-pathological features associated with a positive result. Importantly, this data has highlighted the patients in whom the technology is less suited: those with a lower burden/less advanced disease where the risk of a false negative result is higher. The clinico-pathological associations of targetable alterations were also investigated, with certain patient groups more likely to harbour an actionable mutation. Broadly this aligned with what has been identified in tissue sequencing studies, however the finding of increased *ERBB2* mutations with more lines of HER2-directed prior therapy is a novel finding. This acquisition of *ERBB2* mutations potentially represents a mechanism of resistance to HER2 directed therapy in HER2+ breast cancer, emphasising the need to investigate whether these cancers would benefit specifically from HER2 tyrosine kinase inhibitors that inhibit mutant *ERBB2*<sup>150</sup>.

The data presented here does not support the widespread use of ctDNA analysis in the assessment of copy number change. For this important clinical parameter, it would still be necessary to perform a tissue-based biopsy.

A strength of this analysis is the number of patients that was analysed (n=800) and the comprehensively annotated dataset. Nevertheless, the findings presented here are limited to ABC patients and not applicable to patients with early disease. The analysis of clinico-pathological features is also more limited in patients with HR+ HER2+ and HR- HER2+ patients (n=46 and n=26, respectively), in whom the small numbers of patients will inevitably reduce the power of the analysis to identify clinically significant findings.

## 5. Chapter 5. Breast cancer genomic landscape, clonal architecture, resistance mechanisms and mutational processes according to ctDNA analysis

### 5.1. Introduction

The term ‘genomic landscape’ is a broad term encompassing genomic features of gene mutation frequency, copy number change, clonal architecture and disease heterogeneity. While the genomic landscape of primary breast cancer has been well defined<sup>219</sup>, it is now understood that the genomic features present in ABC can substantially differ from the early breast cancer<sup>42-46,220</sup>. With the increasing number of studies analysing sequencing data from large cohorts of patients with ABC<sup>35-38,94</sup>, our understanding of the genomic landscape and the factors that drive the evolution and progression of ABC is increasing. It is hoped that defining both the landscape of ABC and the processes that shape it will enable us to create better strategies to combat the disease.

Multiregional tumour biopsy studies derived from warm autopsies have provided valuable insights into the phylogeny of clonally heterogeneous metastatic disease from its primary breast cancer precursor<sup>42,43,52</sup>. Clones forming part or all of a patient’s metastatic deposits may develop from a single clone, or ‘metastatic precursor’, or may arise from multiple separate clones from within the primary tumour. Additionally, it has been identified that in metastatic disease, metastases can horizontally cross-seed other deposits<sup>42,43,52</sup>, illustrating the potential for non-linear development and propagation of metastatic disease which is characterised by increased genomic complexity and subclonal alterations.

Other studies have investigated the factors that drive the emergence of heterogeneous subclonal disease in ABC. One landmark study by Razavi *et al* performed prospective targeted sequencing of tumour samples derived from 1756 patients with metastatic breast cancer<sup>34</sup>. This study was able to identify certain genes that were enriched in ABC compared to early breast cancer, with gene enrichment associated with prior hormonal therapy treatment (namely, *ESR1*, *ERBB2* and *NF1* alterations<sup>34</sup>). Notably, the study identified that mutations occurring in *ESR1* and the MAPK pathway are mutually exclusive in the post-treatment tumour biopsy sequencing data studied, suggesting that these disparate routes to endocrine resistance do not overlap in

sequencing data derived from single-site tumour biopsies<sup>34</sup>. This phenomenon has also been noted in the development of resistance to PI3K inhibition, whereby disparate metastases in the same patient developed various PTEN alterations in parallel, which led to the same treatment-resistant phenotype<sup>221</sup>. This suggests that, through the evolutionary pressure exerted by anticancer therapy, individual tumour cells preferentially evolve a single route to treatment resistance, rather than multiple routes in parallel.

Further insights into the drivers of tumorigenesis in advanced disease have been gained from mutational signature analysis. With the advent of massively paralleled sequencing, genomic sequencing is increasingly economically viable leading to broad acquisition of a significant amount of data from large cohorts of patients. This has enabled broader analysis of mutational profiles of cancer samples outside of regions of interest that were previously focused on by necessity of limited sequencing. With this, it is now understood that mutations occurring outside of driver genes are not simply occurring by chance but are the result of specific biological mutational processes which leave distinctive ‘scars’ in the DNA<sup>222</sup>. Study of the scars, or mutational signatures, can give insight into the mutational processes which have been active in the development and evolution of the cancer<sup>222</sup>. Currently, 32 single base substitution signatures, 11 doublet base substitution signatures and 18 indel signatures have been described<sup>223</sup>. In primary breast cancer, these mutational signatures are not thought to relate to breast cancer subtype<sup>222</sup>. Genomic sequencing studies, however, have identified that the mutational signatures present tend to shift from early to metastatic breast cancer<sup>35,38</sup>. Of particular note in breast cancer is an association with mutational signatures 2 and 13<sup>224</sup>, which are associated with APOBEC enzymatic activity. This signature has been found to be enriched in metastatic breast cancer relative to primary breast cancer, and in hormone positive breast cancer<sup>35,38</sup>. APOBECs are a family of cytidine deaminases that use single stranded DNA as a substrate to deaminate cytosine to uracil<sup>222</sup>. There is some evidence to suggest that APOBEC-related mutagenesis is related to endocrine resistance<sup>209</sup>, and is associated with worse clinical outcomes<sup>35</sup>, which makes exploration of this mutational signature in breast cancer particularly interesting.

It is increasingly understood that ctDNA analysis may not only have the potential to characterise the genomic landscape of a patient’s ABC, but may indeed be superior in terms of detecting subclonal metastatic-clade disease<sup>43</sup>. In 2015 Murtaza *et al* demonstrated, through exome and targeted sequencing of multiple tumour sites and ctDNA samples from a single

patient over a three-year course of treatment, that the mutation allele frequency in the plasma reflects the clonal hierarchy inferred from multiregional tumour sequencing<sup>54</sup>. Furthermore, serial changes in the allele frequency of subclonal mutations correlated with the different treatment responses of the spatially distinct metastatic sites. This early finding underlines the potential of ctDNA analysis in being able to delineate the genomic landscape of a patient's cancer.

A major limitation to the use of ctDNA in assessing a patient's genomic landscape is the limited data available on the landscape of ABC according to ctDNA analysis derived from large cohorts of patients with ABC. This characterisation is required to contextualise patient results in the broader spectrum of advanced disease. Smaller studies, of 100 and 255 patients respectively, have made some progress<sup>225,226</sup>. This chapter describes the landscape of the ABC in a large cohort of 800 patients according to analysis of ctDNA, allowing the exploration of clonal architecture and mutational processes that shape clonal diversity.

## 5.2. Hypotheses

The landscape of ABC can be defined through analysis of ctDNA

## 5.3. Aims

- 1) Establish the incidence of SNVs, indels and copy number alterations according to breast cancer phenotype
- 2) Compare the incidence of alterations to that identified in tumour sequenced-cohorts in primary and ABC
- 3) Investigate patterns of clonal dominance and polyclonality in ctDNA
- 4) Investigate patterns of co-mutation and mutation mutual exclusivity

## 5.4. Results

### 5.4.1. Advanced breast cancer landscape

Between 21st December 2016 and 26th April 2019 1,051 patients enrolled into plasmaMATCH, of which 800 had available baseline pre-treatment ctDNA targeted sequencing results. Of the patients with available targeted sequencing, 64.4% (n=515) had HR+



HER2- disease, 9.1% (n=72) were HER2+, and 17.3% (n=138) had TNBC. Overall, 92.9% of patients had at least one alteration identified following ctDNA targeted sequencing.

#### 5.4.2. Incidence and phenotype enrichment of SNVs and indels

Overall, the genes most frequently altered were *TP53* (44.1%), *PIK3CA* (34.9%), *ESR1* (33.1%), *GATA3* (11.0%), *ARID1A* (7.8%) and *PTEN* (6.9%, Figure 5.1)

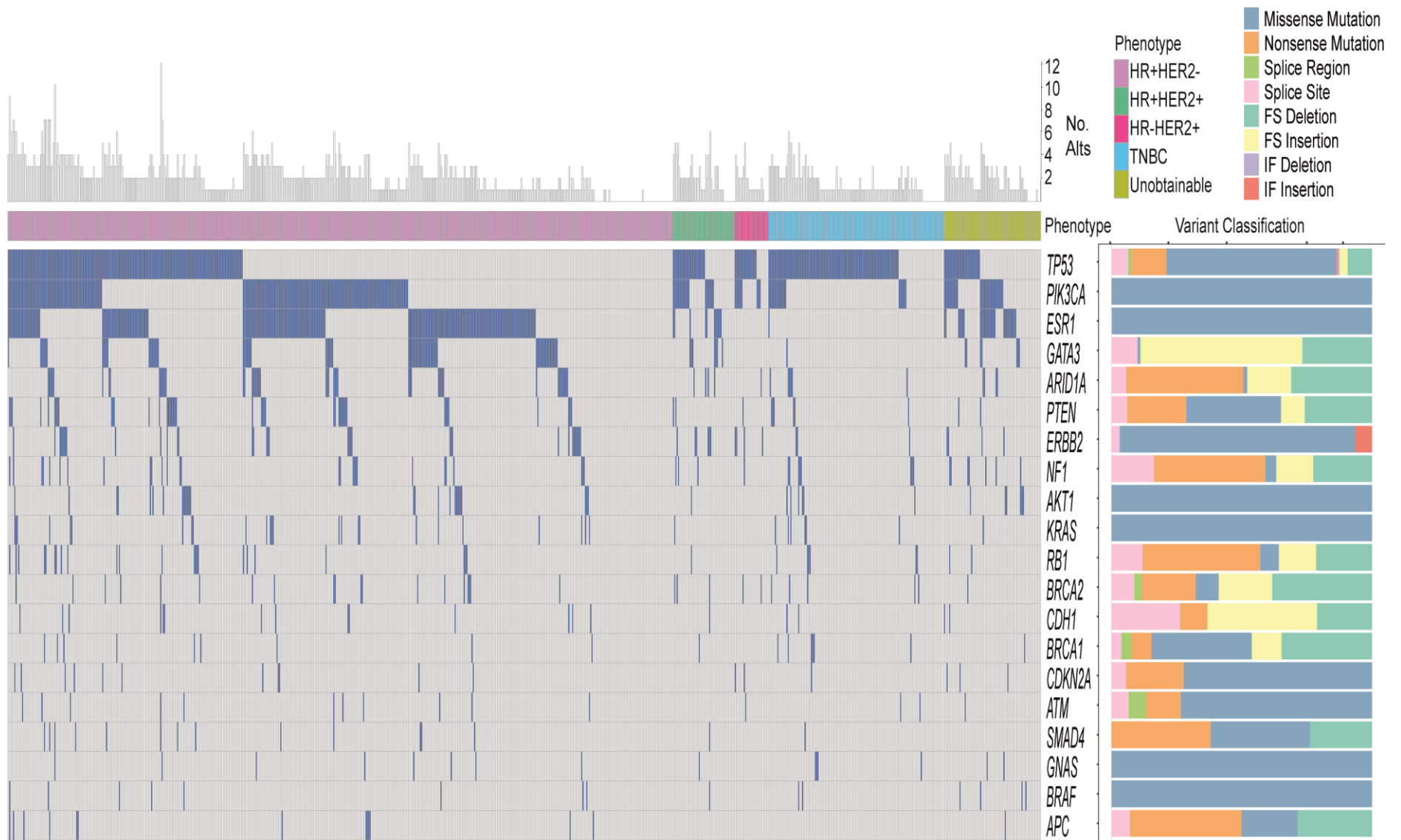


Figure 5.1. Mutational landscape of ABC defined by targeted sequencing of baseline ctDNA samples derived from 800 patients enrolled in plasmaMATCH. Each vertical column represented one patient. Pathogenic mutations (Methods section 2.2.2.1) are summarised by gene per patient. *Top*, total count of pathogenic mutations per patient. *Right*, variant classification of alterations for each gene. FS, frameshift; IF, in-frame.

The mutation incidence of *TP53*, *PIK3CA*, *ESR1* and *GATA3* varied across breast cancer subtypes (Figure 5.2). *ERBB2* mutations were more frequent in HER2+ disease compared to HR+ HER2- ( $q=0.05$ ) and TNBC ( $q=0.005$ ) disease (Figure 5.2). *ERBB2* mutations were more common in HER2+ breast cancers with increasing lines of prior HER2-directed therapy (Chapter 4, section 4.4.3).

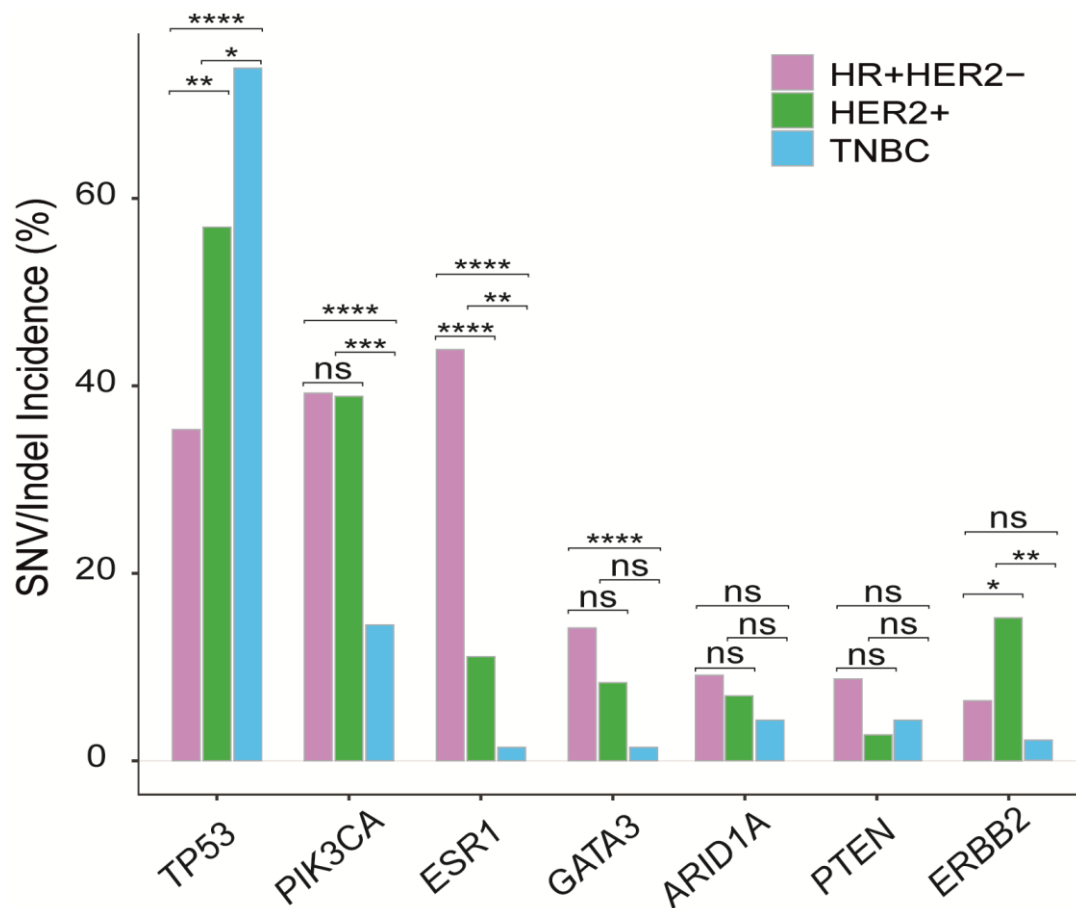


Figure 5.2. Incidence of gene mutations by breast cancer subtype (HR+ HER2-  $n=515$ , HER2+  $n=72$ , TNBC  $n=138$ ). Comparison with false discovery corrected Fisher's Exact tests (*TP53*: HR+ HER2- vs HER2+  $q=0.003$ , HR+ HER2- vs TNBC  $q<0.0001$ , HER2+ vs TNBC  $q=0.04$ ; *PIK3CA*: HR+ HER2- vs TNBC  $q<0.0001$ , HER2+ vs TNBC  $q=0.0008$ ; *ESR1*: HR+ HER2- vs HER2+  $q<0.0001$ , HR+ HER2- vs TNBC  $q<0.0001$ , HER2+ vs TNBC  $q=0.008$ ; *GATA3*: HR+ HER2- vs TNBC  $q<0.0001$ ; *ERBB2*: HR+ HER2- vs HER2+  $q=0.05$ , HER2+ vs TNBC  $q=0.005$ ). ns, not significant.

#### 5.4.2.1. Incidence and phenotype enrichment of copy number alterations

The incidence of copy number alterations was compared across different breast cancer subtypes (Figure 5.3). TNBC was enriched for copy number alterations in *MYC*, *PIK3CA*, *EGFR*, *CCNE1*, *CDK6*, *BRAF* and *MET* ( $q=0.01$ ,  $q<0.0001$ ,  $q=0.001$ ,  $q<0.0001$ ,  $q=0.0003$ ,  $q=0.0002$  and  $q=0.01$ , respectively, Figure 5.3). In HR+ HER2- disease, copy number alterations in *FGFR1* and *CCND1* were more common ( $q=0.01$  and  $q<0.001$ , respectively). *HER2* amplification was detected in ctDNA of 1.7% of patients with HR+ HER2- and TNBC disease, potentially identifying the acquisition of *HER2* amplified disease in these patients.

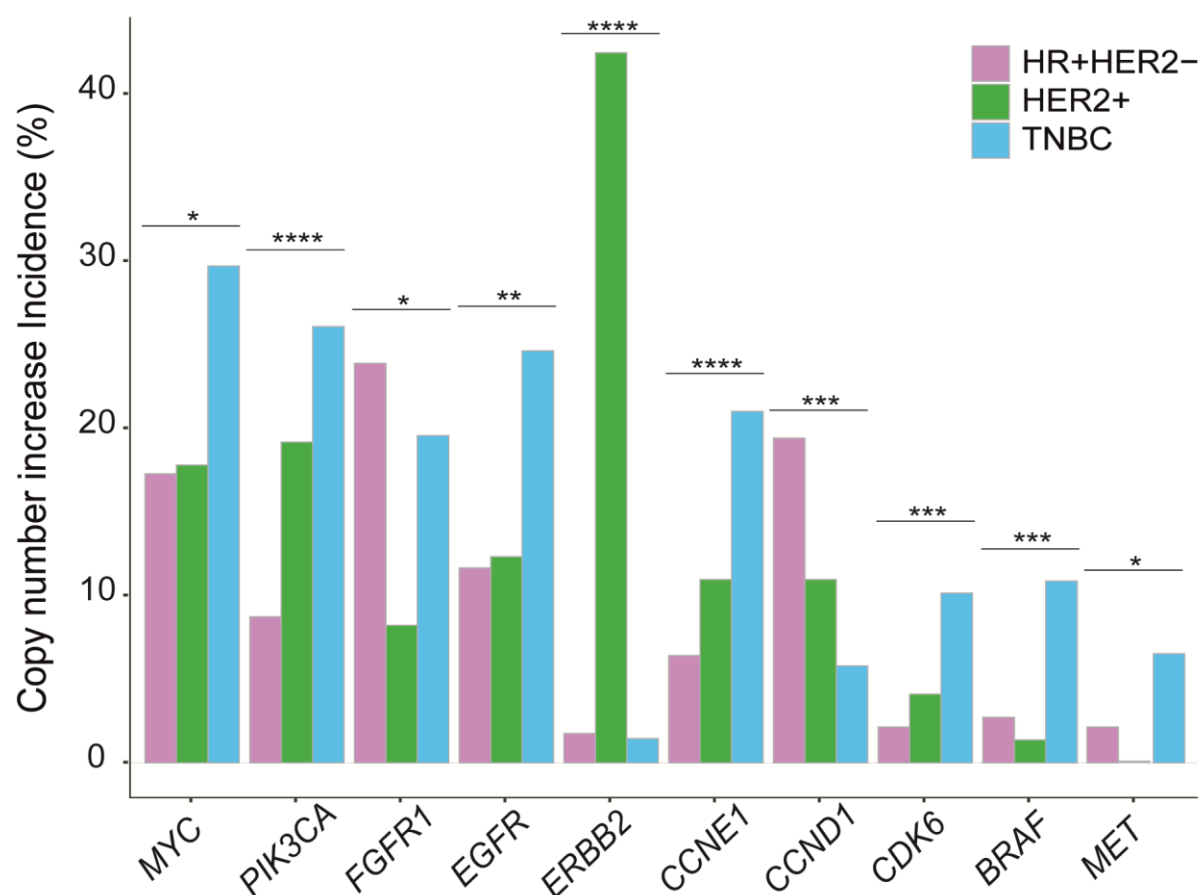


Figure 5.3. Copy number increase incidence by breast cancer subtype (HR+ HER2- n=515, HER2+ n=72, TNBC n=138). Comparison with false discovery corrected Chi-squared tests (*MYC*  $q=0.01$ , *PIK3CA*  $q<0.0001$ , *FGFR1*  $q=0.01$ , *EGFR*  $q=0.001$ , *ERBB2/HER2*  $q<0.0001$ , *CCNE1*  $q<0.0001$ , *CCND1*  $q<0.001$ , *CDK6*  $q=0.0003$ , *BRAF*  $q=0.0002$ , *MET*  $q=0.01$ ).

#### 5.4.2.2. Comparison of the plasmaMATCH plasma ctDNA landscape with advanced and primary breast cancer sequencing datasets

Large datasets of genomic sequencing data are publicly available for review and analysis via the online depositories, Genomic Data Commons Data Portal and cbiportal. We were able to access the genomic sequencing data of patients with early and ABC, respectively, gained from tissue sequencing. This sequencing data underwent the same bioinformatic pipeline as the Guardant360 data to remove germline mutations and identify pathogenic alterations (Chapter 2, section 2.1.12.5 and 2.1.12.6).

When compared to primary breast cancer, *PIK3CA* alterations were less common in metastatic HR+ HER2- disease ( $q < 0.0001$ ) in agreement with large metastatic tissue sequencing datasets<sup>34,36</sup>. Conversely, *ESR1* alterations and alterations in the epigenetic regulator *ARID1A* were enriched ( $q < 0.0001$  and  $q = 0.04$ , respectively)<sup>36,38,204</sup>. Additionally, *AKT1* alterations were enriched ( $q = 0.006$ ) within HR+ HER2- metastatic disease (Figure 5.4)<sup>204</sup>.

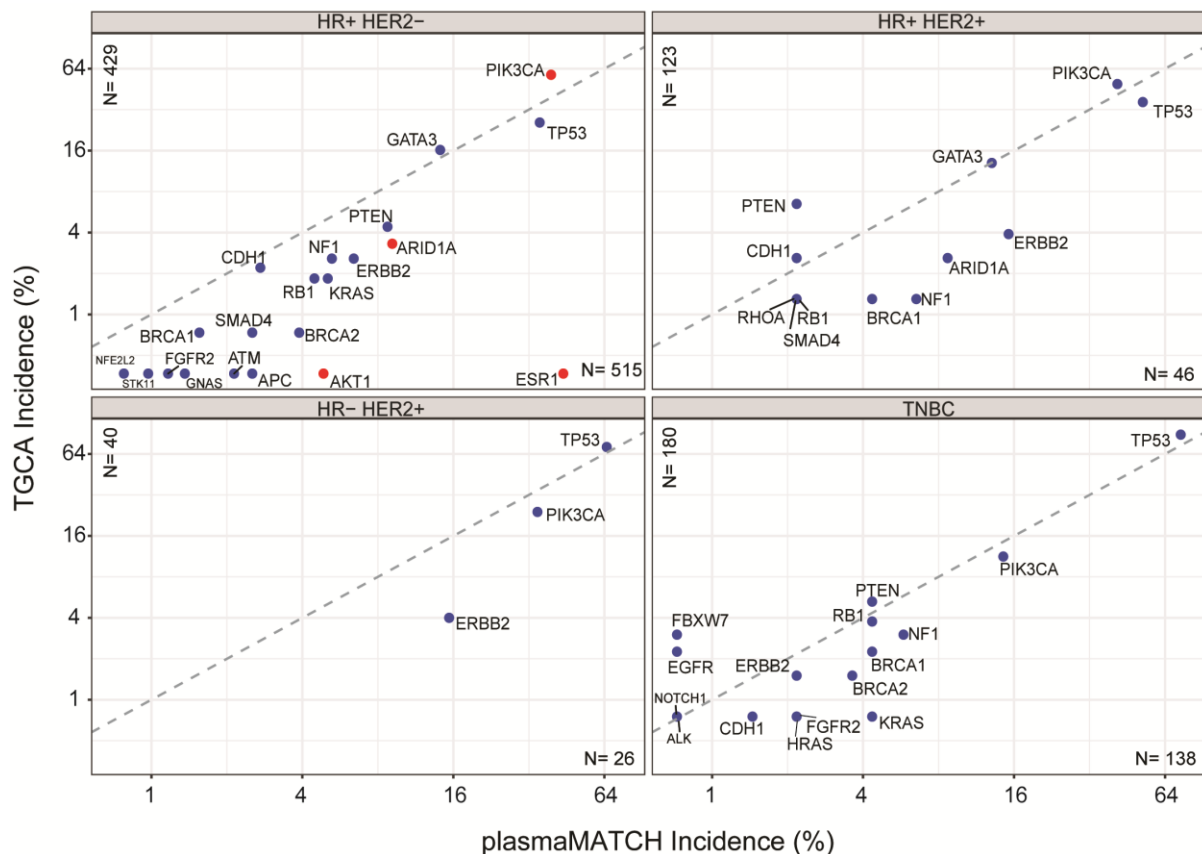


Figure 5.4. Comparison of mutation incidence identified in the plasmaMATCH cohort via ctDNA sequencing, and primary breast cancer (TGCA) via tissue sequencing, by breast cancer subtype. Red dots indicate genes that were significantly different in their mutation incidence after correction for multiple testing.

When compared to a large metastatic sequencing dataset<sup>34</sup> derived from tissue biopsy samples, mutations in *ESR1* and *TP53* were more common in HR+ HER2- disease within the plasmaMATCH dataset ( $q < 0.0001$  and  $q < 0.001$ , respectively, Figure 5.5). Otherwise, the mutational profile identified in ctDNA and tissue across different breast cancer phenotypes was broadly similar (Figure 5.5).

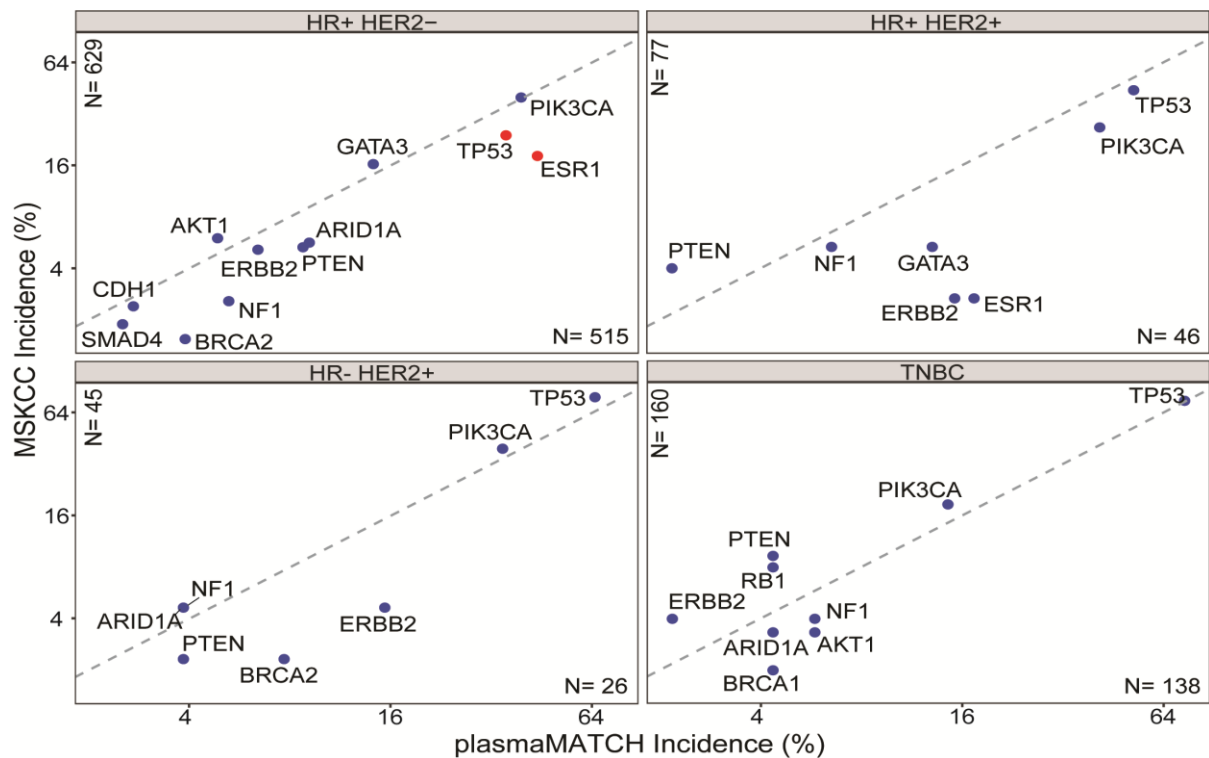


Figure 5.5. Comparison of mutation incidence identified in the plasmaMATCH cohort via ctDNA sequencing, and MSKCC via tissue sequencing, by breast cancer subtype. Red dots indicate genes that were significantly different in their mutation incidence after correction for multiple testing. Genes with an incidence of >1.5% in both datasets are presented.

### 5.4.3. Clonal architecture in ctDNA

Given the potential for ctDNA analysis to spatially dissect metastatic disease, the architecture of breast cancer according to ctDNA analysis could differ from that identified in tissue. To further investigate this, the clonal dominance of mutations and the tendency to occur as single or multiple mutations within a gene was interrogated.

#### 5.4.3.1. Patterns of clonal dominance

The cancer fraction of a mutation describes the clonal dominance of the mutation relative to the mutation with the highest AF in that patient sample. Mutations occurring at a high cancer

fraction are predicted to occur within a greater proportion of a patient's cancer cells than those that occur at a lower clonal fraction, and are more likely to be truncal mutations that could represent a driver mutation. Conversely, subclonal mutations, occurring at a lower cancer fraction, are more likely to be metastatic-clade.

The cancer fraction, or clonal dominance, of mutations within the most frequently altered genes in the cohort were calculated with reference to the alteration with the highest allele frequency in the sample (Chapter 2, section 2.1.12.8). Alterations in *AKT1*, *PIK3CA* and *GATA3* were significantly more clonally dominant ( $q < 0.0001$ ,  $q < 0.0001$  and  $q = 0.0003$ , respectively, Figure 5.6). Conversely, alterations in *ESR1*, *SMAD4* and *KRAS* were significantly more subclonal ( $q < 0.0001$ ,  $q = 0.02$ , and  $q < 0.0001$ , respectively, Figure 5.6).

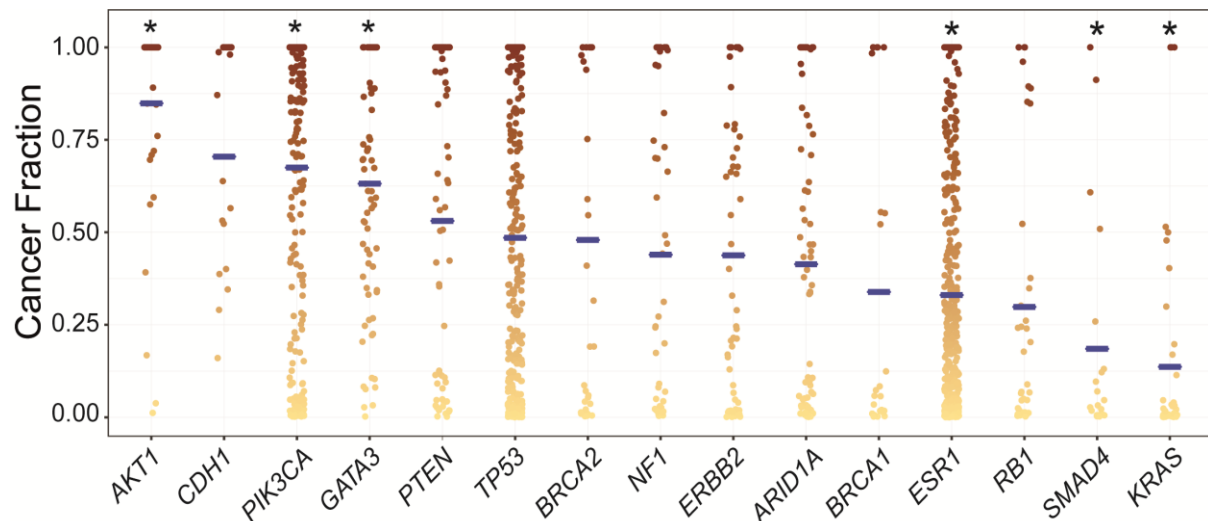


Figure 5.6. Cancer fractions of mutations in the most frequently altered genes, ordered by mean cancer fraction ( $n = 1974$  mutations with assessable cancer fractions). The mean value is indicated with a blue line. \*indicates significant difference in cancer fraction compared to remaining cases, false discovery corrected two-sided Wilcoxon signed-rank test (*AKT1*  $q < 0.0001$ , *PIK3CA*  $q < 0.0001$ , *GATA3*  $q = 0.0003$ , *ESR1*  $q < 0.0001$ , *SMAD4*  $q = 0.02$ , *KRAS*  $q < 0.0001$ ).

The clonal dominance of individual alterations within the targetable genes was next assessed. Alterations within *ESR1* and *PIK3CA* showed significant variance in the clonal dominance according to the specific mutation ( $p < 0.0001$  and  $p < 0.0001$ , Figure 5.7). In *ESR1*, alterations p.D538G and p.Y537S tended to be clonally dominant. In *PIK3CA*, alterations in p.N345K, p.H1047R and p.G1049R tended to be clonally dominant. In contrast, the *PIK3CA* hotspot mutations p.E545K and p.E542K tended to occur at relatively lower cancer fractions.

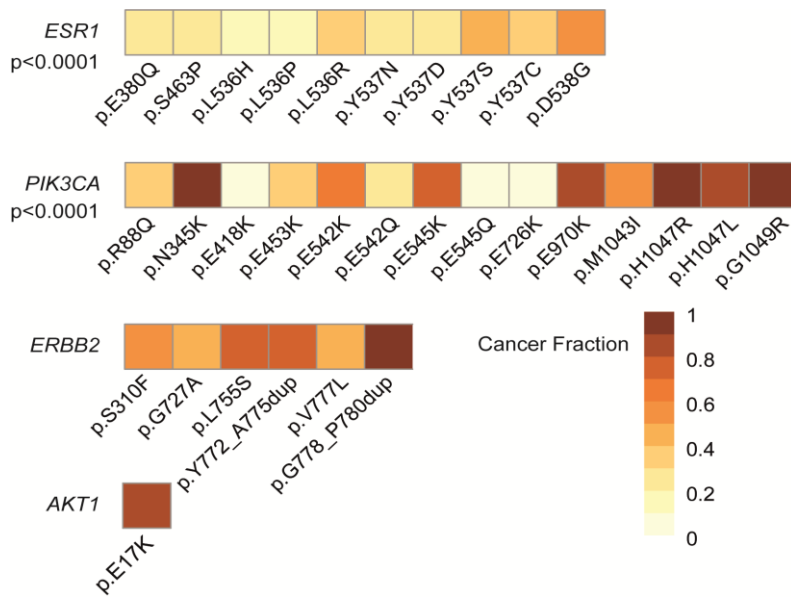


Figure 5.7. Cancer fractions of individual pathogenic hotspot mutations in the indicated targetable gene. The analysis includes hotspot mutations with a minimum of three mutations in the overall data set or for any indel. p value from two-sided Kruskal–Wallis test for variation in cancer fraction across mutations in gene (*ESR1* p<0.0001, *PIK3CA* p<0.0001).

#### 5.4.3.2. Patterns of single versus multiple mutation

Mutations were assessed for whether they occurred as the sole, or single gene mutation present or whether there were multiple mutations, per patient. Mutations in *AKT1*, *CDH1* and *GATA3* were significantly more likely to occur as single mutations. Conversely, alterations in *ESR1* were significantly more likely to occur as multiple ‘hits’ compared to alterations within other genes (Figure 5.8).

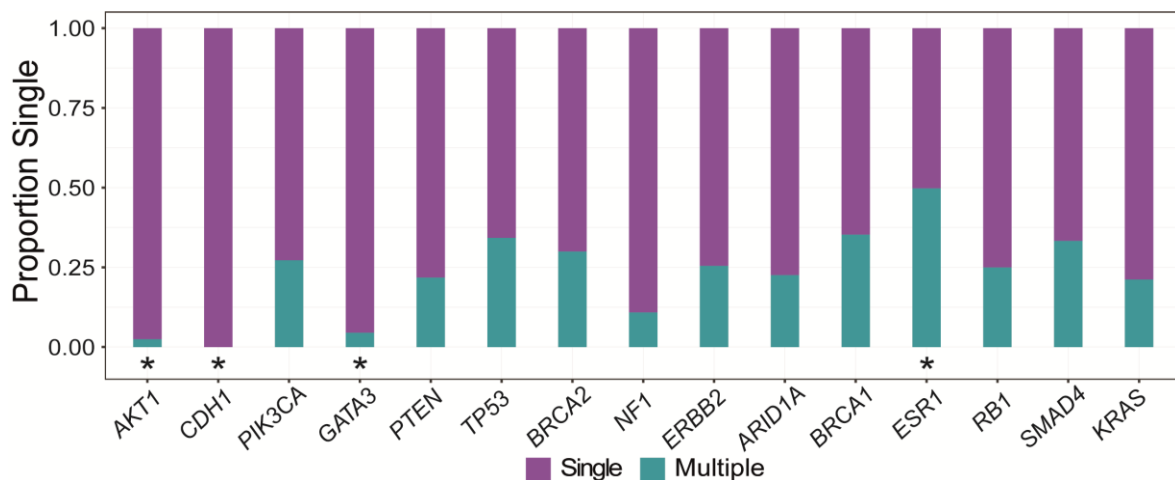


Figure 5.8. Proportion of mutations that occur as a single versus multiple mutations per patient in indicated genes. \*indicates significant difference in proportion to single to multiple mutations in the gene compared to remaining



cases, false discovery corrected Fisher's exact test (*AKT1* q=0.0009, *CDH1* q=0.05, *GATA3* q<0.0001, *ESR1* q<0.0001).

#### 5.4.4. Use of ctDNA to identify polyclonal genomic features

Gene alterations that significantly co-occur may arise through the combination of gene alterations providing the cancer cell an evolutionary advantage such as providing a mechanism of resistance to therapy. Conversely, gene alterations which do not tend to co-occur may represent a combination of cell alterations that produces a redundant or actively deleterious effect on the cell. Tissue biopsy studies have interrogated the co-enrichment and mutual exclusivity of gene alterations<sup>34,219,227</sup>. However, due to being limited to the mutations occurring at a single metastatic site, tissue biopsy studies may not fully delineate patterns of gene co-enrichment and exclusivity. Gene co-enrichment and exclusivity was therefore investigated via ctDNA analysis of the plasmaMATCH cohort.

##### 5.4.4.1. Gene co-enrichment and mutual exclusivity

Within the cohort, several gene pairs showed mutual exclusivity. *PIK3CA* and *AKT1* alterations (q=0.001)<sup>160</sup>, and *ESR1* and *TP53* alterations (q<0.0001) were mutually exclusive (Figure 5.9). Significant co-occurrence was identified in *NF1* and *TP53* (p=0.05), and *PIK3CA* and *ERBB2* (p=0.02, Figure 5.9).

	<i>TP53</i>	<i>PIK3CA</i>	<i>ESR1</i>	<i>GATA3</i>	<i>ARID1A</i>	<i>PTEN</i>	<i>ERBB2</i>	<i>NF1</i>	<i>AKT1</i>
<i>AKT1</i>		0.001							
<i>NF1</i>	0.047		0.023			0.032			
<i>ERBB2</i>		0.015		0.034					
<i>PTEN</i>		0.005							
<i>ARID1A</i>									
<i>GATA3</i>	0.004	0.044	0.003						
<i>ESR1</i>	0.000	0.015							
<i>PIK3CA</i>									
<i>TP53</i>									

Mutually Exclusive

q value < 0.05

p value < 0.05

Co-occurrent

q value < 0.05

p value < 0.05

Figure 5.9. Gene association analysis for the most frequent mutated genes (n=800) with overall Fisher's exact test p values. Green genes showing mutual exclusivity, and purple showing co-occurrence with dark colours indicating significance following false discovery correction.

In HR+ HER2- disease a tendency for mutual exclusivity was identified in *PIK3CA* and *GATA3* (p=0.03) and *ESR1* and *TP53* (p=0.002)<sup>35,198</sup>. Co-enrichment was identified in *PIK3CA* and *KRAS* alterations (p=0.02), and *PTEN* and *NF1* (p=0.02) alterations (Figure 5.10).

## HR+HER2-

	<i>ESR1</i>	<i>PIK3CA</i>	<i>TP53</i>	<i>GATA3</i>	<i>ARID1A</i>	<i>PTEN</i>	<i>ERBB2</i>	<i>NF1</i>	<i>KRAS</i>	<i>AKT1</i>	<i>RB1</i>
<i>RB1</i>											
<i>AKT1</i>											
<i>KRAS</i>		0.022									
<i>NF1</i>						0.023					
<i>ERBB2</i>				0.009							
<i>PTEN</i>			0.023								
<i>ARID1A</i>											
<i>GATA3</i>		0.028									
<i>TP53</i>	0.002										
<i>PIK3CA</i>											
<i>ESR1</i>											

Mutually Exclusive

q value < 0.05

p value < 0.05

Co-occurrent

q value < 0.05

p value < 0.05

Figure 5.10. Gene association analysis for most frequent mutated genes in HR+ HER2- breast cancer (n=515 patients) with Fisher's exact test p values. Green genes showing mutual exclusivity, and purple co-occurrence with dark colours indicating significance following false discovery correction.

### 5.4.4.2. MAPK pathway and *ESR1* alterations in HR+ BC

Ravazi *et al* previously demonstrated in a database derived from tissue sequencing that MAPK pathway alterations (Chapter 2, section 2.1.12.7) and *ESR1* mutations, both of which confer resistance to hormone therapy, tended to be mutually exclusive in HR+ breast cancer<sup>34</sup>. The association of these two was therefore investigated in this dataset derived from ctDNA sequencing. The analysis revealed that, in all patients and in patients with HR+ HER2- breast cancer, MAPK pathway alterations were more common in patients with an *ESR1* mutation (p=0.001 and p=0.02, respectively, Figure 5.11). Furthermore, patients with polyclonal *ESR1* mutations were more likely to harbour a MAPK pathway alteration than those who had a single alteration (p=0.02, Figure 5.11).

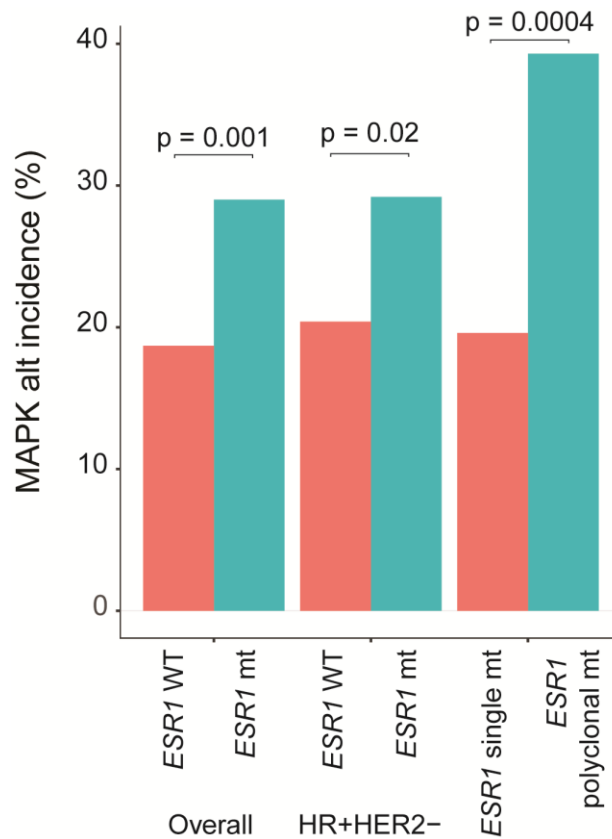


Figure 5.11. Incidence of MAPK pathway alterations in *ESR1* mutant and wild type populations. A comparison of the incidence in, *Left*, *ESR1* mutant (77/265) vs *ESR1* wild type cancers overall (100/535), *Middle*, *ESR1* mutant (66/226) vs wild type (59/289) in HR+ HER2- cancers, and *Right*, within *ESR1* mutant cancers between patients with single (27/138) and polyclonal *ESR1* mutations (50/127). p values from Fisher’s exact test.

The overall survival of patients with HR+ HER2- disease who entered a treatment cohort in plasmaMATCH was investigated according to whether they had an *ESR1* mutation, a MAPK pathway alteration, or both. Patients with HR+ HER2- disease (n=515) and concurrent MAPK and *ESR1* alterations (n=32) had a shorter median overall survival than patients wild type for both (n=26, Figure 5.12).

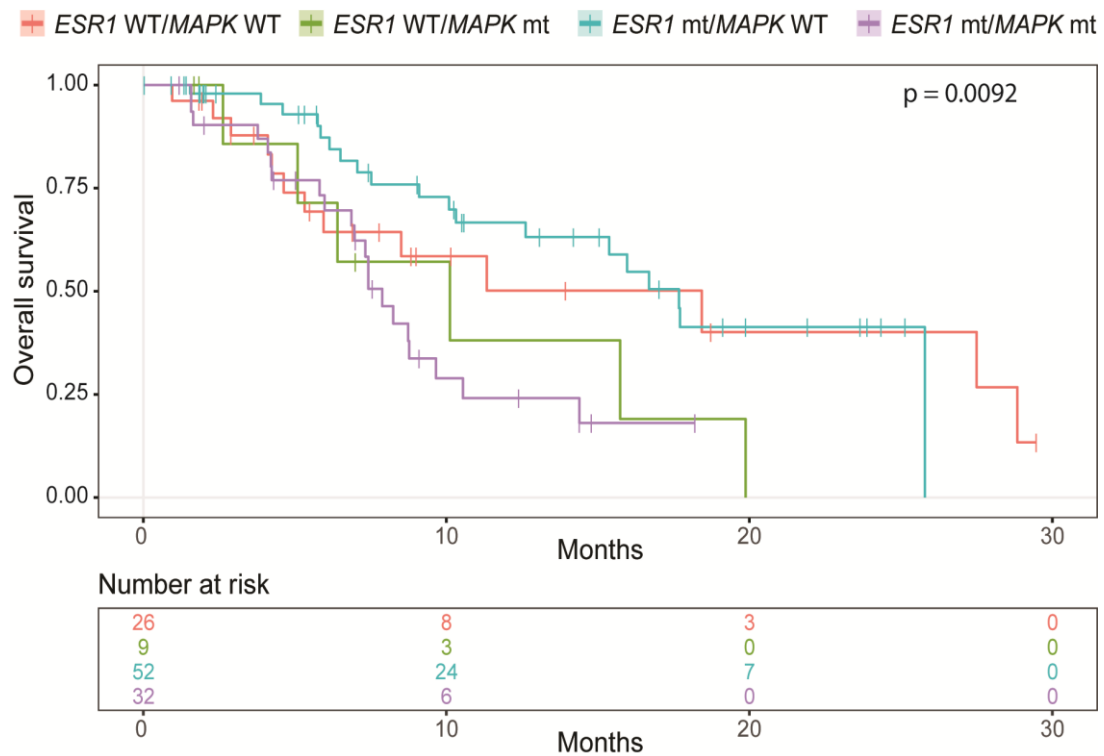


Figure 5.12. Overall survival (OS) in patients with HR+ HER2- disease who entered a treatment cohort in plasmaMATCH divided by combined *ESR1* and MAPK pathway mutation status. *ESR1* WT and MAPK WT, median 18.5 months, hazard ratio (HR) -. *ESR1* mt and MAPK WT, median 17.7 months, HR 0.82, 95% confidence interval (CI) 0.40 to 1.69. *ESR1* WT and MAPK mt, median 10.1 months, HR 1.65, 95% CI 0.56 to 4.88. *ESR1* mt and MAPK mt, median 7.9 months, HR 1.65, 95% CI 0.84 to 3.23. p value from log-rank test. HR > 1 indicates worse OS for that group. WT, wild type; mt, mutant.

#### 5.4.4.3. *PIK3CA* double mutants in HR+ BC

*PIK3CA* alterations were common, occurring in 34.9% of the plasmaMATCH cohort and enriched in HR+ HER2- disease (39.2%). Analysis of patients with *PIK3CA*-mutant HR+ disease revealed that having more than one *PIK3CA* mutation was frequent, occurring in 23% of patients.

Particular alterations within *PIK3CA* showed significant variation in their clonal dominance (Figure 5.7 and Figure 5.13). *PIK3CA* alterations occurring at low mean cancer fraction were frequently coincident with hotspot *PIK3CA* mutations, which themselves were commonly clonally dominant. *PIK3CA* mutations which occurred at low clonal dominance were frequently one of a number of *PIK3CA* mutations for the patient, and more likely to occur at sites associated with APOBEC mutagenesis (Figure 5.13, Chapter 2, section 2.1.12.10).

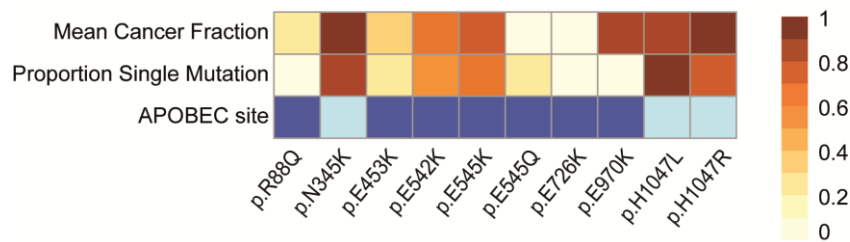


Figure 5.13. *PIK3CA* mutagenesis in HR+ breast cancer. Analysis of individual recurrent hotspot mutations in HR+ HER2- *PIK3CA* mutant disease (n=202) with mean cancer fraction, proportion of mutations detected as a single *PIK3CA* mutation, and indication of whether the mutation occurs at an APOBEC consensus site (dark blue). Mutations occurring at least 3 times in the HR+ HER2- disease dataset are included. Cancer fraction and proportion single mutations is lower at APOBEC sites,  $p < 0.001$ , Fisher's exact test both comparisons.

In HR+ HER2- *PIK3CA*-mutant disease, subclonal *PIK3CA* alterations were significantly more likely to occur at an APOBEC mutagenesis site ( $p < 0.0001$ ), while the same pattern did not occur in *PIK3CA*-mutant TNBC disease (not significant, Figure 5.14).

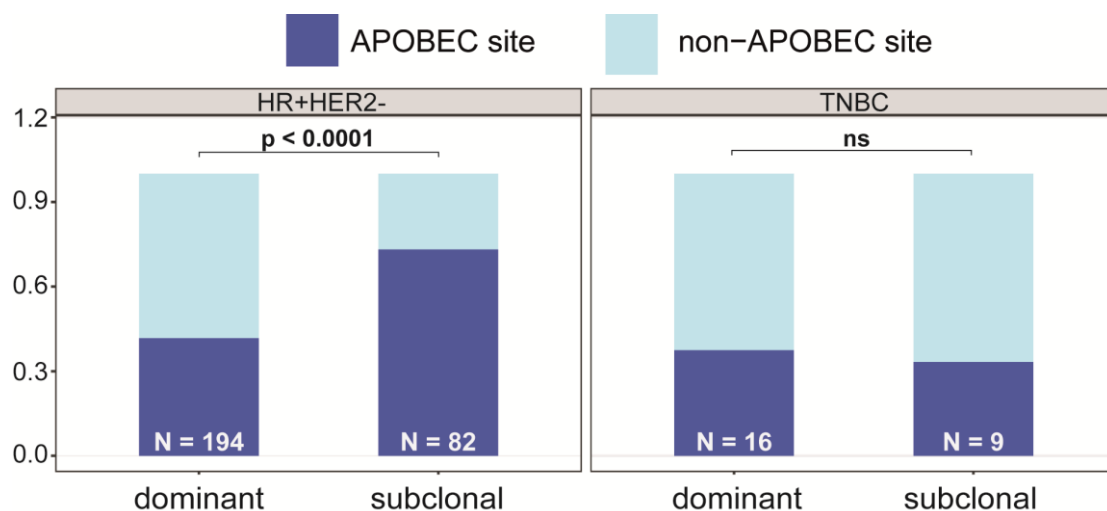


Figure 5.14. Proportion of *PIK3CA* mutations that occur at APOBEC consensus sites, by cancer subtype (HR+ HER2- n=197, TNBC n=21), and clonally dominant (n=194 mutations within HR+ HER2- breast cancers and n=16 within TNBC breast cancers) versus subclonal *PIK3CA* mutation (n=82 mutations within HR+ HER2- breast cancers and n=9 within TNBC breast cancers). p value from Fisher's exact test. ns, not significant.

Patients who entered cohort A for treatment with fulvestrant were compared for progression free survival (PFS) according to *PIK3CA* mutation status. Cohort A patients with polyclonal *PIK3CA* disease (n=7) had a shorter median PFS on fulvestrant than patients with WT (n=49) or one *PIK3CA* mutation (n=22) ( $p=0.0036$ , Figure 5.15).

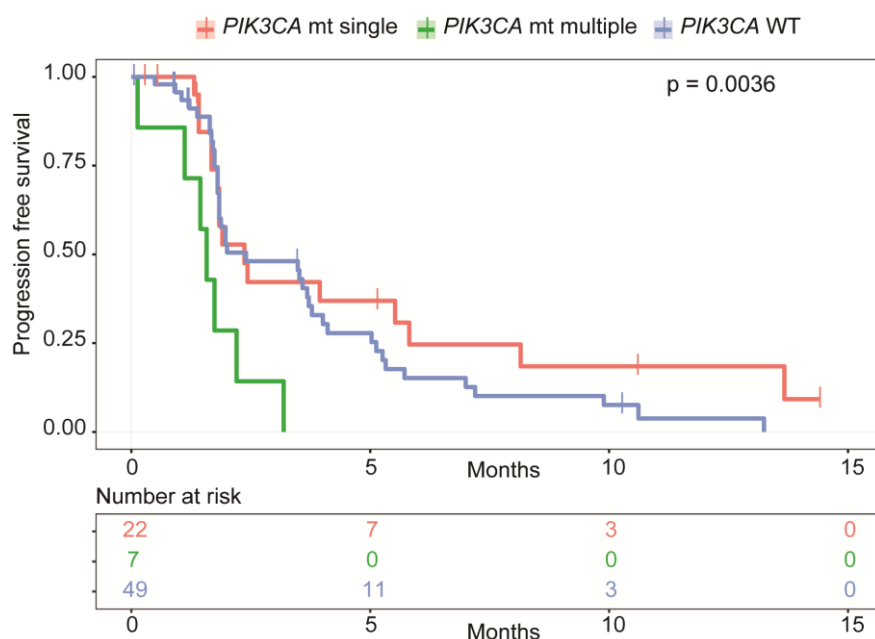


Figure 5.15. Progression free survival (PFS) in patients on fulvestrant in treatment cohort A in plasmaMATCH by *PIK3CA* mutation status (n=78). *PIK3CA* WT, median 2.4 months, HR -. *PIK3CA* single mt, median 2.4 months, HR 0.71, 95% CI 0.41 to 1.22. *PIK3CA* multiple mt, median 1.6 months, HR 3.15, 95% CI 0.88 to 11.33. p value from log-rank test. HR>1 indicates worse PFS for that group. WT, wild type; mt, mutant.

## 5.5. Discussion

Through baseline genomic sequencing of patients enrolled prospectively into plasmaMATCH, delineation of the landscape of ABC according to ctDNA analysis has been achieved. The incidence of the most commonly mutated genes, *TP53*, *PIK3CA*, *ESR1* and *GATA3*, varied between breast cancer subtype similarly to that identified in large published tissue sequencing datasets<sup>34,35,219</sup>.

The incidence of gene alterations was also broadly similar to both primary and metastatic tissue sequencing datasets (Figures 5.4 and 5.5). The plasmaMATCH dataset was enriched for *AKT1* mutations compared to primary disease. Prior data suggests *AKT1* mutations may be more frequent in relapsed breast cancer than primary breast cancer<sup>35,38,161</sup>.

In ABC, the HR+ HER2- plasmaMATCH cohort was enriched for alterations in *ESR1* and *TP53* compared to the MSKCC tissue sequencing dataset (Figure 5.5). As demonstrated in Figures 5.6 and 5.8, *ESR1* alterations are frequently subclonal and polyclonal, and more likely to be metastatic-clade. Secondary to this, *ESR1* mutations maybe ‘missed’ by single-site tissue biopsies whilst identified by ctDNA analysis. This supports the use of ctDNA sequencing to

circumvent the issue of tumour heterogeneity which limits tissue biopsies. Another possibility, however, is that there was enrichment for *ESR1* alterations within plasmaMATCH secondary to preferential recruitment to the trial of patients with prior aromatase inhibitor therapy exposure, in whom *ESR1* mutations are selected.

The second enriched mutation in the dataset, *TP53*, is a marker of poor prognosis<sup>227,228</sup>, and the higher incidence of these mutations may reflect that this cohort of patients have more aggressive disease. Other than these exceptions, the mutation incidence in the plasmaMATCH set was similar to that of primary and metastatic sequencing datasets, increasing confidence in the approach of ctDNA analysis in profiling the breast cancer landscape.

Analysis of the clonal dominance of alterations per gene (Figure 5.6) demonstrated that alterations in *AKT1*, *PIK3CA* and *GATA3* tend to be clonally dominant, which concurs with the generally accepted role of these gene alterations in being breast cancer drivers<sup>161,204,219,229</sup>, while mutations in *ESR1*, *SMAD4* and *KRAS* were more likely to be subclonal. As identified in this cohort, *ESR1* mutations are known to be generally subclonal<sup>36,72</sup>. Mutations in *SMAD4* are less well characterised and, alongside *KRAS* mutations, may arise as a result of clonal haematopoiesis<sup>230</sup>, which would align with their subclonal tendency identified here. Additionally, there is evidence to suggest *KRAS* alterations can arise as a mechanism of resistance to hormonal therapy<sup>34,72</sup>. If correct, then similarly to alterations within *ESR1* these mutations are highly likely to arise by chance in individual cancer cells and be selected through the evolutionary advantage conferred by the therapeutic resistance to become subclonal resistance mutations. Overall, this analysis demonstrates that analysis of ctDNA can elucidate the subclonal architecture of ABC, enabling greater understanding of the landscape of ABC.

Analysis of ctDNA also identified a number of genomic features with important therapeutic implications. Firstly, the increased incidence of *ERBB2* mutations within HER2+ disease (Figure 5.2) is a significant finding, and as identified in Chapter 4, could arise as resistance mechanism to prior HER2-directed therapy<sup>141,210</sup>. Prior work has demonstrated that *ERBB2*-mutant HER2+ disease remains sensitive to neratinib<sup>210,211</sup>, suggesting an alternative treatment strategy for these patients.

A second major finding was that, converse to what has been identified in a large tissue sequencing study<sup>34</sup>, *ESR1* and MAPK pathway mutations were found to be co-enriched (Figure 5.11). Both *ESR1* and MAPK pathway mutations are known to confer resistance to hormonal therapy<sup>34</sup>, and it is understood that they are mutually exclusive mechanisms of resistance<sup>34</sup>. The

finding of enrichment in the ctDNA dataset suggests that these mechanisms of resistance may be metastatic-clade, and that single-site tumour biopsy sequencing may have ‘missed’ the divergent routes of hormonal resistance present in the same individual. The potential for tumour biopsies to miss resistance mutations occurring in parallel could have significant therapeutic implications, and supports a role for ctDNA assessment of genomic resistance.

A third genomic feature identified was the presence of a high number of polyclonal *PIK3CA* mutations in HR+ breast cancer. Recent work by Vasani *et al* demonstrated that double *PIK3CA* mutations are common in breast cancer, and lead to increased signalling via the PI3K pathway with resulting increased downstream activity<sup>231</sup>. The rate of double *PIK3CA* mutation identified here was higher than was identified in multiple large tissue sequencing datasets (23% versus 8-13%)<sup>231</sup>. Similarly to the *ESR1*/MAPK pathway co-occurrence in ctDNA, this could be a result of ctDNA being better able to identify global heterogeneity. In support of this, Vasani *et al* found a similar phenomenon when they investigated the incidence of double *PIK3CA* mutation in the SANDPIPER trial using ctDNA analysis, again finding a higher rate of double mutation of 19%<sup>231</sup>. Importantly, these double mutants were found to predict for poorer treatment outcome on subsequent fulvestrant as demonstrated here (Figure 5.15), while, conversely, exhibited enhanced sensitivity to PI3K inhibition with alpelisib<sup>231</sup> demonstrating the clinical relevance of establishing *PIK3CA* polyclonality.

The data presented here demonstrates that in HR+ disease subclonal *PIK3CA* mutations commonly occur at sites that could be consistent with APOBEC mutagenesis (Figure 5.13 and 5.14). This suggests a possible mechanistic role of APOBEC in driving the development of polyclonal *PIK3CA* disease in HR+ HER- breast cancer<sup>35</sup> as a mechanism to enhance signalling via the pathway. Prior work has identified enrichment of APOBEC mutagenesis in HR+ HER2- advanced disease relative to primary disease<sup>35,36,38</sup>. APOBEC mutagenesis has additionally been associated with endocrine resistance<sup>209</sup>, with *PIK3CA* mutations occurring within the helical *PIK3CA* domain<sup>232,233</sup>, and in subclonal disease<sup>234</sup>. The finding here of enrichment for *PIK3CA* mutations at APOBEC sites specifically within the subclonal mutations of HR+ HER2- disease supports the hypothesis that APOBEC is activated later in carcinogenesis and is a driver of tumour evolution in this breast cancer subtype. It should be noted, however, that this investigation is much limited by the nature of the sequencing data it was derived from. Mutational signature analysis normally requires broad sequencing data including reads covering intronic areas in order to identify the classes of mutational signatures present. This data was derived from targeted sequencing data which is limited to regions included in the



panel and therefore may be biased in identifying the signatures present, yet concurs with previous data derived from broader sequencing approaches<sup>232-234</sup>.

Analysis of ctDNA demonstrated that variants within *ESR1* and *PIK3CA* show significant variation in their clonal dominance (Figure 5.7). The therapeutic implications of this are unclear. To date, clonality and allele frequency of *ESR1* has not been found to relate to response to fulvestrant<sup>98,111</sup>. Similarly, strong evidence of *PIK3CA* clonality correlating with PI3K inhibition response is lacking<sup>235</sup>, albeit most *PIK3CA* mutations are clonally dominant (Figure 5.6), precluding this analysis. Further investigation into importance of targetable mutation clonal dominance is required to establish if this is a clinically relevant parameter

A limitation of this analysis is highlighted by the finding of subclonal *SMAD4* and *KRAS* mutations, which may arise as a result of clonal haematopoiesis (CH, discussed in Chapter 1, section 1.3.6). Importantly, patient germline samples were not sequenced, precluding the ability to remove mutations caused by CH from the analysis. These mutations may confound the analysis if they are falsely attributed to being cancer derived, and represent a significant limitation to the project. It is thought that there is little overlap in the genes involved in CH and breast cancer, which mitigates the risk of confounding here. On a patient level, however, if germline sequencing has not been undertaken in parallel, the possibility of mutations arising from CH must be considered. Increased understanding of the occurrence and relevance of CH mutations is required in the future, with a view to creating criteria to enable confident and systematic removal of such mutations from any diagnostic report.

## 5.6. Conclusion

In this Chapter, the landscape of ABC has been defined in a large cohort of patients. The landscape described here broadly fits with that identified in tumour sequencing, with a few notable exceptions. In particular, *ESR1* mutations were enriched, and co-mutation of *ESR1* with MAPK pathway signalling and a high incidence of *PIK3CA* mutations. Taken together, these findings support the hypothesis that ctDNA analysis is able to sample intratumoral heterogeneity in a way that is not possible by tissue sequencing. The analysis of cancer fraction and polyclonality also demonstrates that it is possible to elucidate the clonal architecture of ABC through differentiating cancer driver mutations and identifying potential subclonal resistance mechanisms. This data supports the role of ctDNA analysis in genomically profiling ABC. Furthermore, ctDNA analysis has revealed several important and clinically relevant

findings that advance our understanding of breast cancer heterogeneity and, ultimately, has the potential to influence breast cancer management.

## 6. Chapter 6. CtDNA as a predictive and prognostic biomarker

### 6.1. Introduction

Predictive biomarkers predict a patient's outcome on a specific therapeutic intervention, while prognostic biomarkers give information on a patient's overall expected cancer outcome regardless of the therapy given<sup>236</sup>. Biomarkers of both types are increasingly being sought to enhance and optimise patient treatment in a personalised-medicine approach. The concept of biomarker driven therapeutic choice is not a new one in breast cancer, where anti-oestrogen therapies have, for many years, been directed to patients with hormone positive disease in the knowledge that these patients have a high rate of response. The established prognostic and predictive biomarkers in breast cancer are generally related to clinico-pathological characteristics such as disease stage and grade, and immunohistochemical features of the tumour such as breast cancer subtype, oestrogen and progesterone receptor positivity and HER2 staining. With increased accessibility of genomic sequencing coupled with advances in the understanding the prognostic and predictive genomic biomarkers, we are increasingly able to personalise therapeutic choices and prognosticate at the level of a patient's tumour genomic profile.

Tumour profiling is frequently used in other tumour types to identify molecular alterations that predict for response to a subsequent targeted therapy. The FDA has approved several such therapies, including inhibitors of EGFR, ALK and ROS1 in patients with NSCLC with alterations in the respective genes<sup>9,237,238</sup>, inhibitors of the MAPK signalling pathway in patients with *BRAF* mutated melanoma<sup>239-241</sup>, and EGFR-directed therapies in patients with *KRAS* or *NRAS* mutated colorectal cancer<sup>242</sup>. The use of plasma-based genomic profiling using ctDNA analysis is less well established. The first FDA approval for a plasma-based assay was in NSCLC, where the *cobas* EGFR mutation test v2 was approved to identify *EGFR* mutations when considering treatment with erlotinib<sup>243</sup>. Importantly though, in the case of a negative result a tumour biopsy is indicated<sup>243</sup>. Subsequently, the *therascreen* PIK3CA assay has been licenced by the FDA to identify *PIK3CA* mutations in ABC when considering treatment with PI3K inhibitor alpelisib<sup>57</sup>. Again, the same caveat of requirement of a tumour biopsy in the situation of a negative test was specified<sup>57</sup> due to the risk of a false negative result.

It is now established that the concentration of ctDNA in the blood is prognostic<sup>28,48,244,245</sup>. What is not well established is how to measure the ‘volume’, or purity of ctDNA in the plasma sample. Further research is needed to identify general prognostic and predictive biomarkers to specific therapies, which will greatly enhance our ability to direct therapy to patients who will respond whilst avoiding futile therapy in patients who might be best treated with an alternative therapy.

The plasmaMATCH trial investigated the utility of plasma ctDNA testing in identifying targetable alterations, bringing together the two questions of whether baseline genomic alterations can act as biomarkers to select subsequent therapy, and whether ctDNA analysis can be used to identify these genomic changes<sup>98</sup>. Through correlating the patient’s baseline genomic features with response on treatment and overall survival, it may be possible to identify further prognostic and predictive biomarkers in ctDNA for the three targeted treatments in the trial.

## 6.2. Hypothesis

ctDNA analysis can be used to identify prognostic and predictive biomarkers.

## 6.3. Aims

- 1) Investigate ctDNA purity as a biomarker of response and prognostic marker by comparing two different approaches of ctDNA purity assessment.
- 2) Associate baseline genomic features of patients in cohort A with response to fulvestrant to identify predictive biomarkers.
- 3) Associate baseline genomic features of patients in cohort B with response to neratinib +/- fulvestrant to identify predictive biomarkers.
- 4) Associate baseline genomic features of patients in cohorts C and D treated with capivasertib +/- fulvestrant with response to identify predictive biomarkers.

## 6.4. Results

### 6.4.1. ctDNA purity as a predictive and prognostic marker

CtDNA purity refers to the volume of ctDNA relative to the total volume of genomic DNA present in the plasma. The ctDNA purity of a sample can be measured in a number of different

ways, and there is no universally accepted approach. The adopted approach depends very much on the sequencing strategy employed. Specifically, where a more limited approach has been taken such as ddPCR or targeted sequencing, the allele frequency of a truncal mutation may be used to denote the purity. This approach is not without limitations. Specifically it relies on the assumptions that 1) there is a truncal mutation, 2) this truncal mutation has been identified by the sequencing strategy, 3) there is no copy number alteration influencing the allele frequency of this truncal mutation, and 4) this truncal mutation is present in every cancerous cell in the body. A second approach is to sequence and analyse heterozygous single nucleotide polymorphisms (SNPs). Cancer cell DNA exhibits frequent loss and gain of regions of DNA<sup>246</sup>. The proportion of DNA exhibiting this pattern of loss and gain of the SNPs (deviation away from 50% incidence of both alleles of the heterozygous SNP), can be distinguished and used to estimate the purity of the sample<sup>79</sup>. Finally, using a broad sequencing approach such as whole genome sequencing (WGS), the proportion of DNA demonstrating deviation away from a diploid copy number can be established to give an estimation of purity<sup>85</sup>. Adalsteinsson *et al* developed iChorCNA, a software program capable of quantifying the ctDNA content of cell free DNA using sequencing data derived from 0.1x (low pass, lp) WGS<sup>85</sup>. The benefit of this approach is that it avoids many of the assumptions made when using maxVAF to estimate purity, such as reliance on a truncal mutation to be present.

Two of the ctDNA purity assessment approaches, maxVAF and lpWGS, were compared to identify if either, or both, approaches could be used to calculate the purity-adjusted clonality of a targeted mutation. Patients enrolled into cohort A who had available targeted sequencing data and remaining DNA from the same sample had the DNA sample undergo library preparation followed by low pass (0.1x) whole genome sequencing (lpWGS) (n=67), with both approaches used to estimate sample purity. There was reasonable correlation between the two purity measurements, with a rho of 0.68 (p<0.0001, Figure 6.1). The general pattern was for lpWGS to estimate a higher purity than the maxVAF approach.

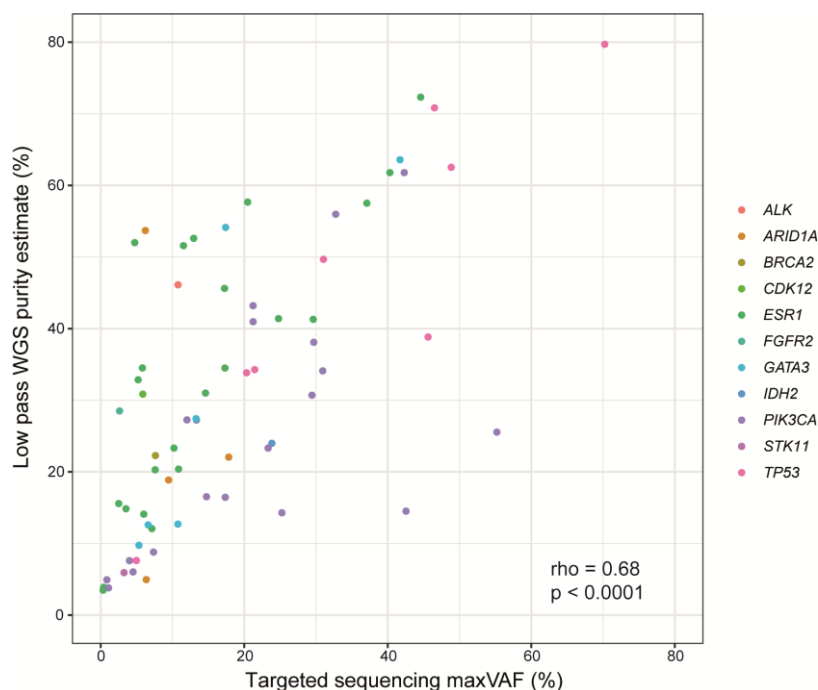


Figure 6.1. Correlation of the low pass WGS purity estimate and targeted sequencing maximum VAF (maxVAF) in 67 patients in cohort A,  $\rho = 0.68$ ,  $p < 0.0001$ . The dot colour represents the gene within which the alteration with highest allele frequency was identified. Correlation by Spearman's Rank method.

Under the assumption that patients with clonally dominant mutations will respond to a greater degree to targeted therapy than those with subclonal alterations, we aimed to ascertain whether the purity-adjusted *ESR1* allele frequency represented a predictive biomarker of response to subsequent fulvestrant therapy. This analysis was undertaken in cohort A due to the higher number of patients enrolled compared to cohorts B-D. The two methods of purity estimation, maxVAF or lpWGS, were compared to ascertain if either demonstrated a superior approach with which to adjust the *ESR1* clonal dominance when correlating with response to targeted therapy. The cumulative *ESR1* allele frequency (sum of the allele frequency of all pathogenic *ESR1* mutations present at baseline for each patient) was adjusted for the purity to estimate *ESR1* clonality using the following formula:

$$\text{Adjusted } ESR1 \text{ clonality} = \frac{\text{Cumulative } ESR1 \text{ allele frequency}}{\text{Purity}}$$

Where the *purity* estimate was provided by the maxVAF from targeted sequencing, or iChorCNA software from lpWGS. Patients were grouped according to whether they had clonal (adjusted clonality  $\geq 0.5$ ) or subclonal ( $< 0.5$ ) *ESR1*-mutant disease. This calculation is made under the assumption that activating *ESR1* mutations do not occur in the same clones but in

separate clones as mechanisms of resistance arising in parallel. In this situation, the summed allele frequency of the *ESRI* mutations would reflect all the *ESRI*-mutant positive disease. This assumption may not be biologically true, in which case this would not be an accurate method to assess *ESRI* clonality. A further assumption is that all of the included *ESRI* mutations are truly pathogenic rather than passenger mutations, which again may not be accurate.

Progression free survival (PFS) was not significantly different in either group when *ESRI* clonality was adjusted using either maxVAF or iChorCNA (Figure 6.2). Thus neither approach appeared to demonstrate that purity-adjusted *ESRI* clonality was predictive of response to fulvestrant.

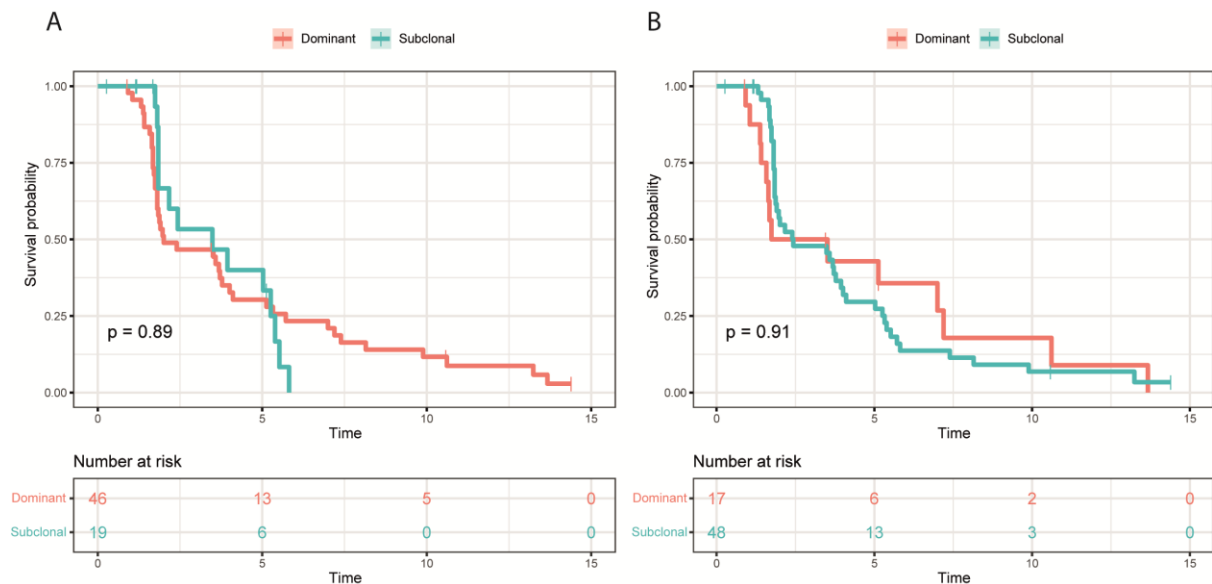


Figure 6.2. Progression free survival in patients in cohort A on fulvestrant according to their adjusted *ESRI* clonality. A, mVAF-adjusted and B, iChorCNA-adjusted *ESRI* clonality. Patients are grouped according to whether the *ESRI* mutations combined represent clonally dominant (clonality  $\geq 0.5$ ) or subclonal disease ( $< 0.5$ ). A, Dominant median 2.0 months, hazard ratio (HR) -. Subclonal median 3.5 months, HR 1.04 (95% CI 0.6 to 1.9). B, Dominant median 2.6 months, HR -. Subclonal median 2.4 months, HR 1.04 (95% CI 0.6 to 1.9). p value from log-rank test. HR  $> 1$  indicates worse PFS for that group.

The predictive value of ctDNA purity for progression free survival on fulvestrant was next established using the two purity assessment methods of mVAF and iChorCNA. For each approach, the percentage purity which identified the most significant prognostic differentiator was identified for both approaches. For mVAF, a ctDNA purity of 5.6% was most differentiating, compared to 16% for iChorCNA (Figure 6.3). Although there was a trend for

patients with low purity to have longer median PFS on fulvestrant, neither approach significantly predicted response to fulvestrant based on baseline ctDNA purity.

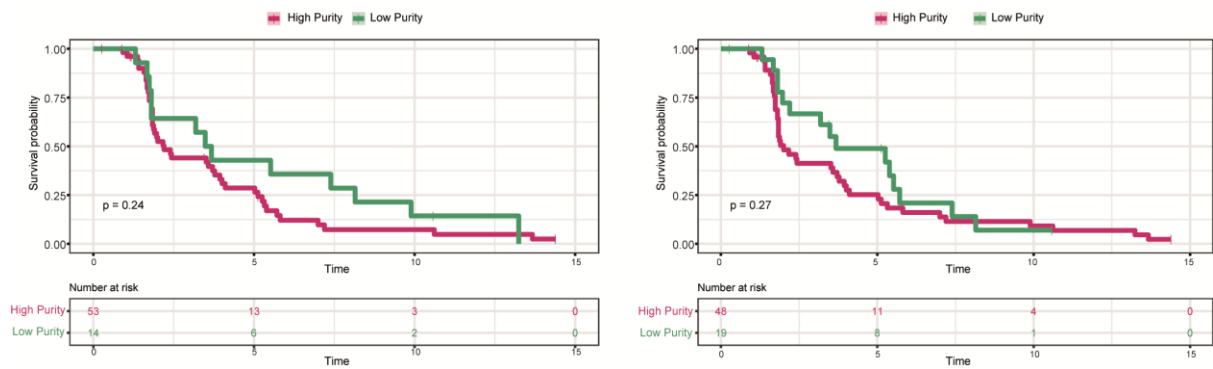


Figure 6.3. Progression free survival in cohort A patients (n=67) according to ctDNA purity as calculated by A, mVAF, ‘High Purity’ = mVAF  $\geq 5.6$  and B, lpWGS, ‘High Purity’  $\geq 16.0$ . A, High purity median PFS 2.2 months, Low purity median PFS 3.6 months, HR 0.71 (95% CI 0.39 to 1.27). B, High purity median PFS 2.0 months, Low purity median PFS 3.7 months, HR 0.73 (95% CI 0.41 to 1.28). p value from log-rank test. HR > 1 indicates worse PFS for that group.

The prognostic value of purity for overall survival (OS) was next established by comparing the two purity assessment methods. For each approach, the purity percentage which produced the most significant differentiation in prognosis was identified. An allele frequency of 5.6% represented mVAF with the greatest differentiator of prognosis (p=0.011, Figure 6.4). Meanwhile an iChorCNA purity estimate cutoff of 10.9% provided the strongest differentiator of prognosis (p=0.023, Figure 6.4).

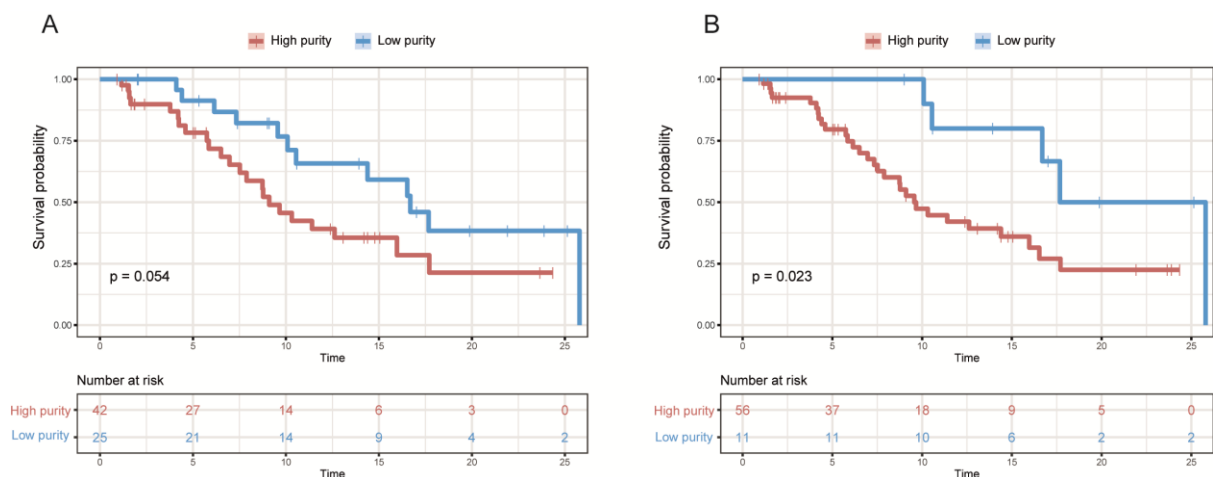


Figure 6.4. Overall survival in cohort A patients (n=67) according to ctDNA purity as calculated by A, maxVAF, ‘High Purity’ = mVAF  $\geq 5.6$  and B, lpWGS, ‘High Purity’  $\geq 10.9$ . A, High purity median OS 9.7 months, Low purity median OS 25.8 months, HR 0.39 (95% CI 0.19 to 0.80). B, High purity median OS 9.6 months, Low purity



median OS 21.7 months, HR 0.42 (95% CI 0.20 to 0.89). p value from log-rank test. HR > 1 indicates worse OS for that group.

#### 6.4.2.ctDNA as a biomarker of response to targeted therapy

Within the plasmaMATCH trial there are three targeted therapies utilised within cohorts A to D. This presents an opportunity to identify biomarkers of response for each targeted therapy. As cohort A enrolled the most patients this cohort would have the greatest power to identify significant findings, and will be focussed on here.

##### 6.4.2.1.plasmaMATCH Cohort A

Of 84 patients enrolled into cohort A, 80 started treatment and 74 had on treatment RECIST-assessable imaging (Figure 6.5). Patients were treated with extended-dose fulvestrant (500mg fulvestrant (IM) on Cycle 1 Days 1, 8 and 15 and Cycle 2 onwards Days 1 and 15). Baseline targeted sequencing data was available for 79 patients, and 69 patients had both baseline and end-of-treatment (EOT) sequencing. Three patients did not have detectable *ESR1* mutations by genomic sequencing, but had detectable targetable *ESR1* mutations by ddPCR. *ESR1* alterations were commonly subclonal, occurring at low allele frequencies, with 49.4% of patients having more than one plasmaMATCH targetable *ESR1* alteration (Figures 6.5 and 6.6).

Radiologically evaluable for response

Non-evaluable

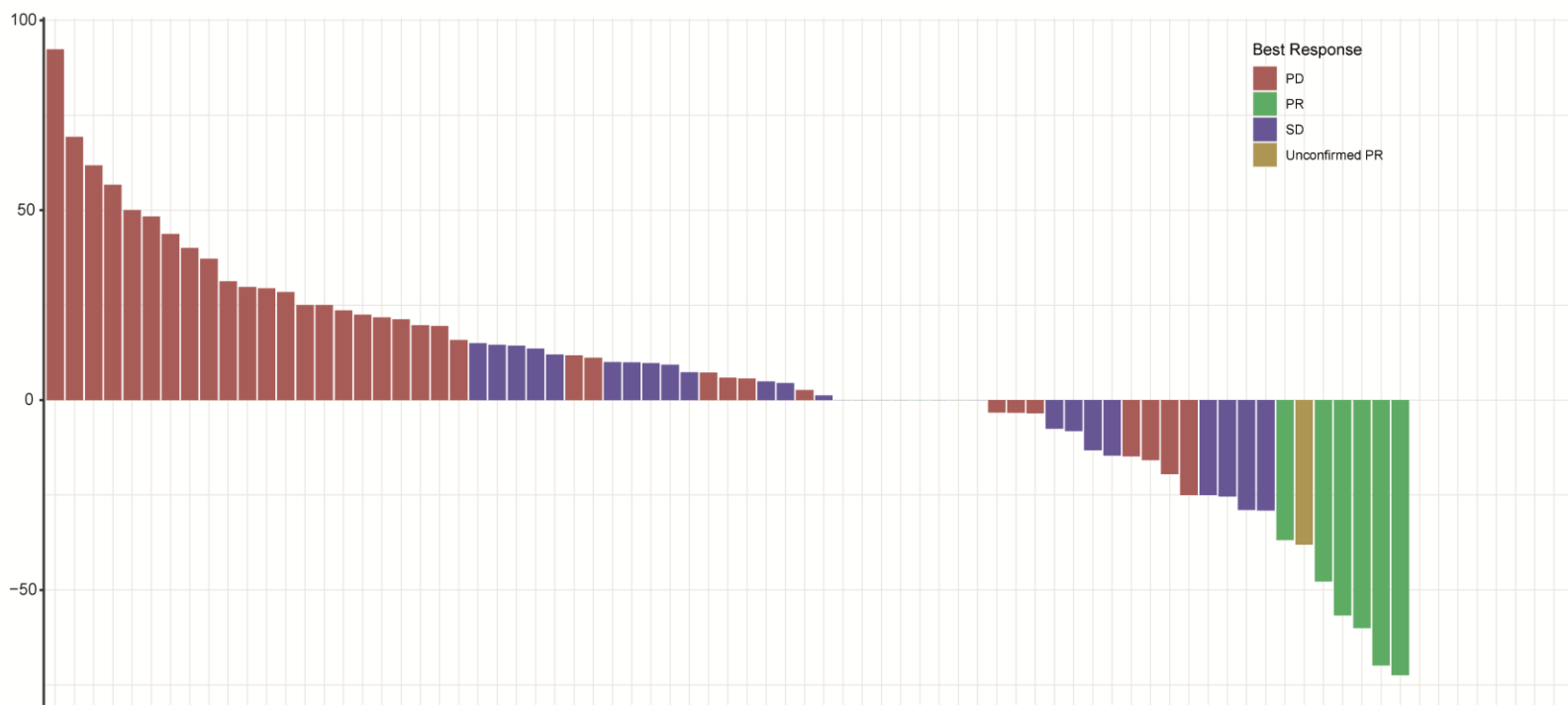
Sequencing availability

Baseline  
EOT

Baseline genomics

ESR1 clonality  
ESR1 alteration  
PIK3CA  
TP53

Change from Baseline (%)



Best Response  
 PD  
 PR  
 SD  
 Unconfirmed PR

Baseline  
 available  
 not available  
 EOT  
 available  
 not available  
 ESR1 clonality  
 Clonally Dominant  
 Subclonal  
 unascertainable  
 ESR1 alteration  
 p.D538G  
 p.E380Q  
 p.Y537C  
 p.Y537N  
 p.Y537S  
 polyclonal  
 unascertainable  
 wildtype  
 PIK3CA  
 negative  
 positive  
 unascertainable  
 TP53  
 0  
 1  
 unascertainable

EOT genomics

No. acquired alterations  
ESR1 F404 acquisition

No. acquired alterations  
 0  
 1-2  
 3-6  
 7+  
 F404 acquisition  
 F404 wt  
 F404 acquired  
 unascertainable

Figure 6.5. Waterfall plot for cohort A annotated with genomic sequencing availability, baseline and EOT genomic data.

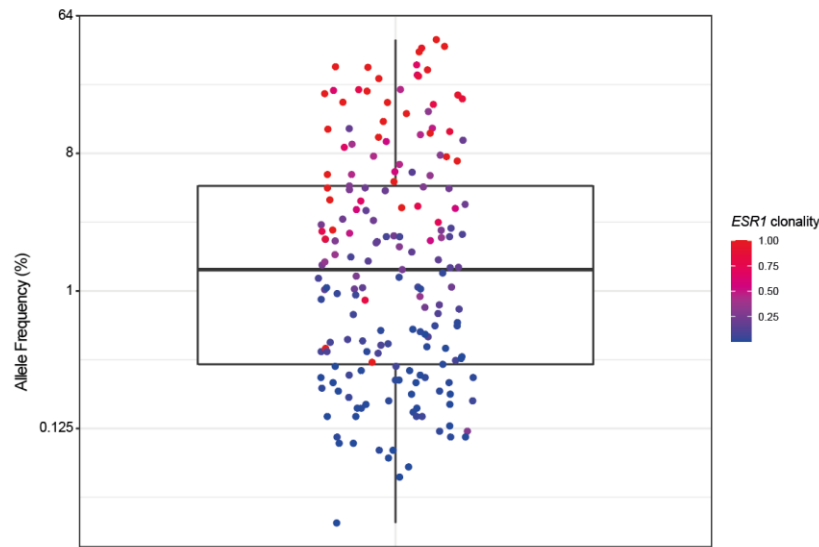


Figure 6.6. Allele frequency and clonality of *ESR1* mutations within cohort A (n=185 pathogenic *ESR1* alterations with assessable clonality).

The most frequent *ESR1* alteration in the cohort was p.D538G (n = 44, 55.7%), followed by p.Y537S (n = 34, 43.0%), p.E380Q (n = 22, 27.9%), p.Y537N (n = 22, 27.9%), p.Y537C (n = 11, 13.9%), p.L536R (n = 7, 8.9%) and p.S463P (n = 4, 5.1%) (Figure 6.7A). Analysis of change in clonal dominance from baseline to EOT demonstrated that p.E380Q, p.Y537C and p.S463P most frequently lost clonal dominance by EOT relative to baseline (p=0.002, p=0.046 and p=0.046, respectively), while p.D538G and p.Y537S most frequently gained or maintained clonal dominance (p=0.001 and p=0.030, respectively, Figure 6.7B). While targetable *ESR1* mutations were frequently polyclonal (Figure 6.5), mutations in p.D538G, p.Y537S and p.Y537C were significantly more likely to occur as monoclonal activating *ESR1* alterations than alongside other activating *ESR1* alterations (p=0.03, p=0.0003 and p=0.03, respectively, Figure 6.7C).

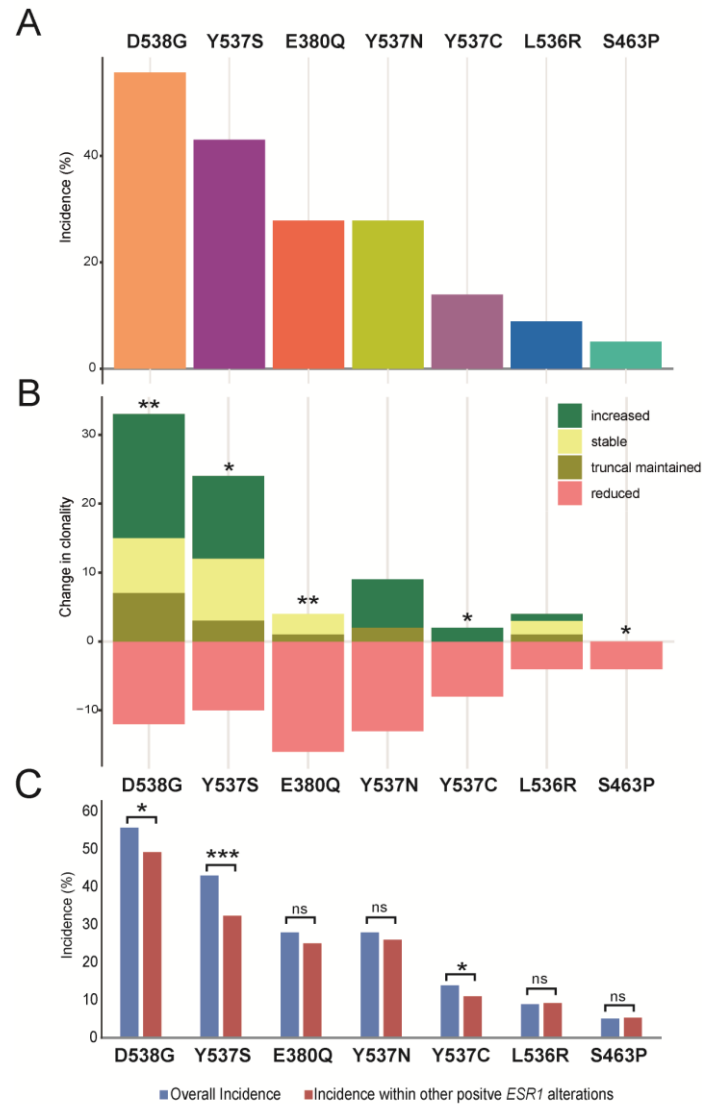


Figure 6.7. Targetable *ESR1* alterations and response within cohort A. A, incidence of baseline *ESR1* alterations within cohort A (n=79 assessable patients). B, Change in *ESR1* mutation clonality from baseline to EOT (n=69 assessable patients). For each *ESR1* mutation, the clonal dominance at baseline was compared to the clonal dominance at EOT. Those that lost clonal dominance had a clonal dominance ratio (EOT clonal dominance/Baseline clonal dominance) of <0.8. Those with a stable clonal dominance had a clonal dominance ratio of  $\geq 0.8$  and <1.2. Those gaining clonal dominance had a ratio  $\geq 1.2$ . *ESR1* mutations that are the truncal mutation with the highest allele frequency at both time points are identified. p.D538G p=0.001, p.E380Q p=0.002, p.Y537S p=0.030, p.Y537C p=0.046, p.S463P p=0.046. Comparison by Chi-square tests: combined increased/stable/truncal maintained versus reduced, for each *ESR1* variant versus all the others combined. C, Incidence of each activating *ESR1* mutation within the dataset and the incidence of the alteration amongst patients positive for another *ESR1* activating alteration. Mutations p.D538G, p.Y327S and p.Y537C occurred at a significantly higher incidence in the overall dataset than alongside another activating *ESR1* alterations. Comparisons by Fisher's Exact test, p.D538G p=0.03, p.Y537S p=0.0003, p.Y537C p=0.03.

Durable PFS was not consistently ascribed to a particular baseline *ESRI* genotype, with the infrequent durable PFS noted in patients with a range of baseline alterations (Figure 6.8A). Survival analyses demonstrate that patients with baseline p.Y537C alterations had a longer median progression free survival (5.6 versus 2.0 months, HR 2.8 (95% CI 1.3 to 5.9), Figure 6.8B). Conversely, patients with a baseline p.Y537S mutation had a shorter median PFS (1.8 versus 3.5 months, HR 0.53 (95% CI 0.33 to 0.86), (Figure 6.8B). Number of baseline targetable *ESRI* variants and cumulative *ESRI* clonality did not associate with response (Figure 6.8C)<sup>111</sup>.

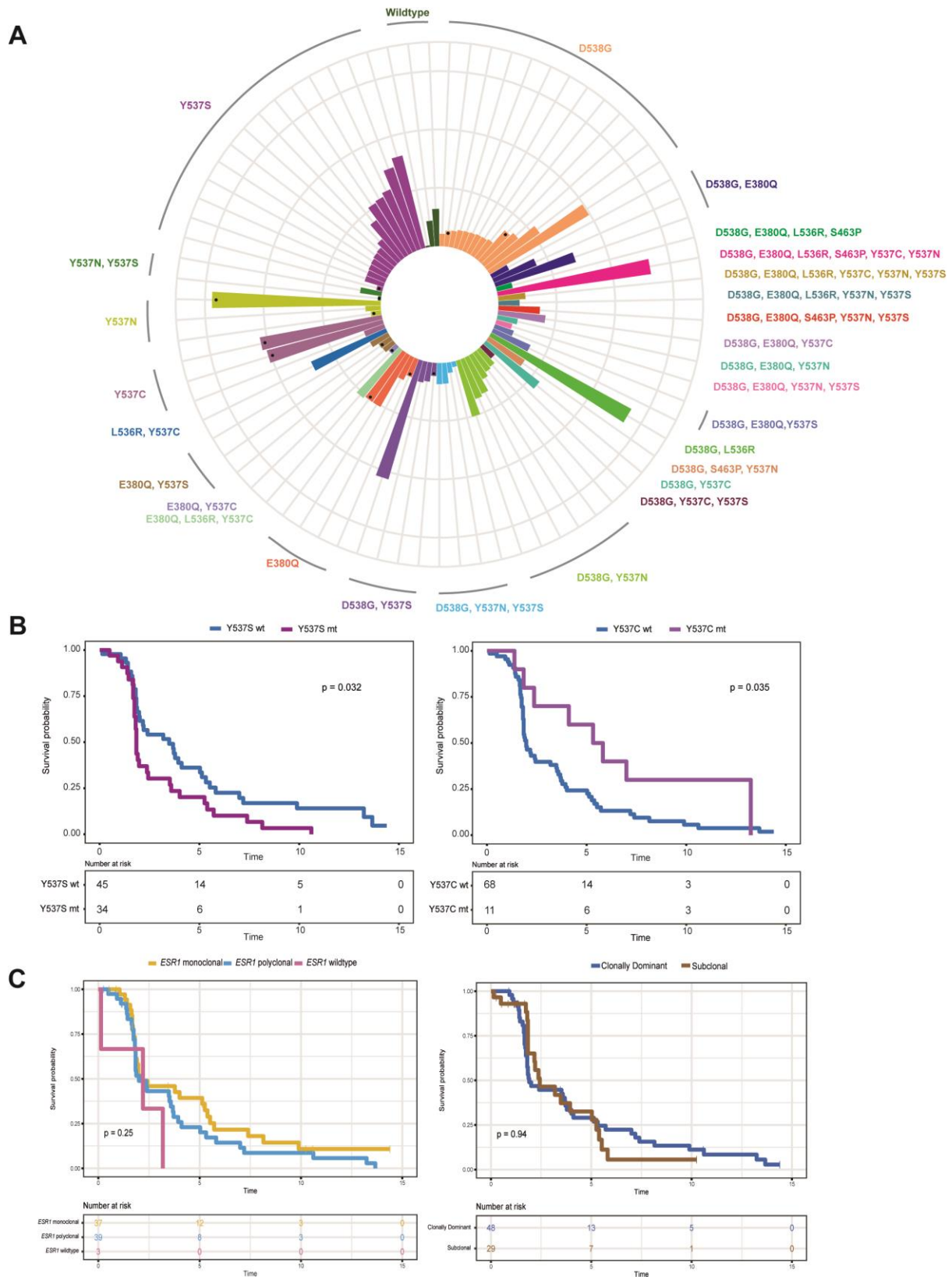


Figure 6.8. Baseline genomic status and response in cohort A. A, Patient progression free survival annotated with baseline *ESR1* status (n=79 patients). A black dot denotes patient censored at the time of data cut off. B, Progression free survival in patients within cohort A, divided by baseline p.Y537S mutation status (*left*) and p.Y537C mutation status (*right*). *left*, p.Y537S wild type, median 3.5 months, hazard ratio (HR) -; p.Y537S

mutant, median 1.8 months, hazard ratio 1.67 (95% confidence interval (CI) 1.00 to 2.78). *right*, p.Y537C wild type, median 2.0, HR -; p.Y537C mutant, median 5.6 months, HR 2.13 (95% CI 1.20 to 3.79). p values from log rank test. HR >1 denotes worse PFS for that group. WT, wild type; mt, mutant. C, Progression free survival in patients within cohort A, divided by left, *ESR1* mono/polyclonal status and right, clonally dominant versus subclonal baseline *ESR1* mutation status.

*PIK3CA*<sup>H11</sup> and *TP53* mutation status at baseline does not appear to associate with response (Figure 6.9).

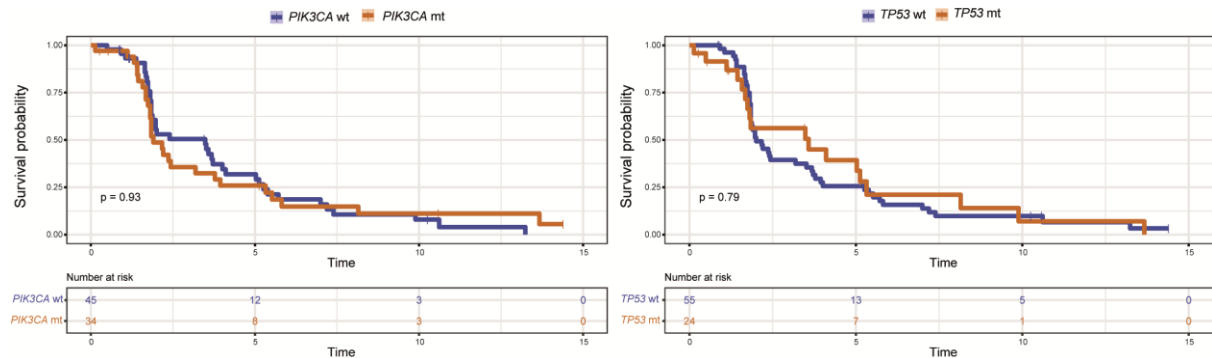


Figure 6.9. Progression free survival according to baseline *PIK3CA* and *TP53* mutation status. *Left*, *PIK3CA* mutation status. *Right*, *TP53* mutation status. *PIK3CA*-mutant median 1.9 months, *PIK3CA* wild type median 3.5 months (p=0.93). *TP53*-mutant median 3.6 months, *TP53* wild type median 2.0 months (p=0.79). wt, wild type. mt, mutant.

It was noted that there was a high incidence of *PIK3CA* double mutagenesis in patients with HR+ disease in the plasmaMATCH dataset of 800 patients (Chapter 5, section 5.4.4.3). It was postulated that this may be as a result of APOBEC-induced mutagenesis. To investigate any association of double *PIK3CA*-mutagenesis with response to subsequent fulvestrant treatment, the progression free survival of patients with wild type disease was compared to patients with a single mutation and those with more than one mutation (Chapter 5, section 5.4.4.3). Patients with multiple *PIK3CA* mutations had a significantly shorter PFS on fulvestrant than patients with wild type or single *PIK3CA* mutations (p=0.0036, Chapter 5, section 5.4.4.3).

#### 6.4.2.2.plasmaMATCH Cohort B

Cohort B enrolled 21 patients with *ERBB2* mutations, of whom 20 started treatment (17 patients with HR+ HER2- disease, 1 patient with HR+ HER2+, 2 patients with HR- HER2+ disease). Patients were treated with neratinib (240mg PO OD), with the addition of fulvestrant (500mg

IM on Cycle 1 Days 1 and 15 and Cycle 2 onwards Day 1) in hormone receptor positive patients. Of the patients who started treatment, all had baseline sequencing available.

The small number of patients enrolled within the cohort precludes detailed analysis of the baseline genomic correlates of response. Nevertheless, there was some evidence that *ERBB2* targeted alteration may be an important factor in response to neratinib (Figure 6.10). Three of four patients with p.V777L responded particularly well to neratinib, and were censored at the time of data cut off (Figure 6.10).

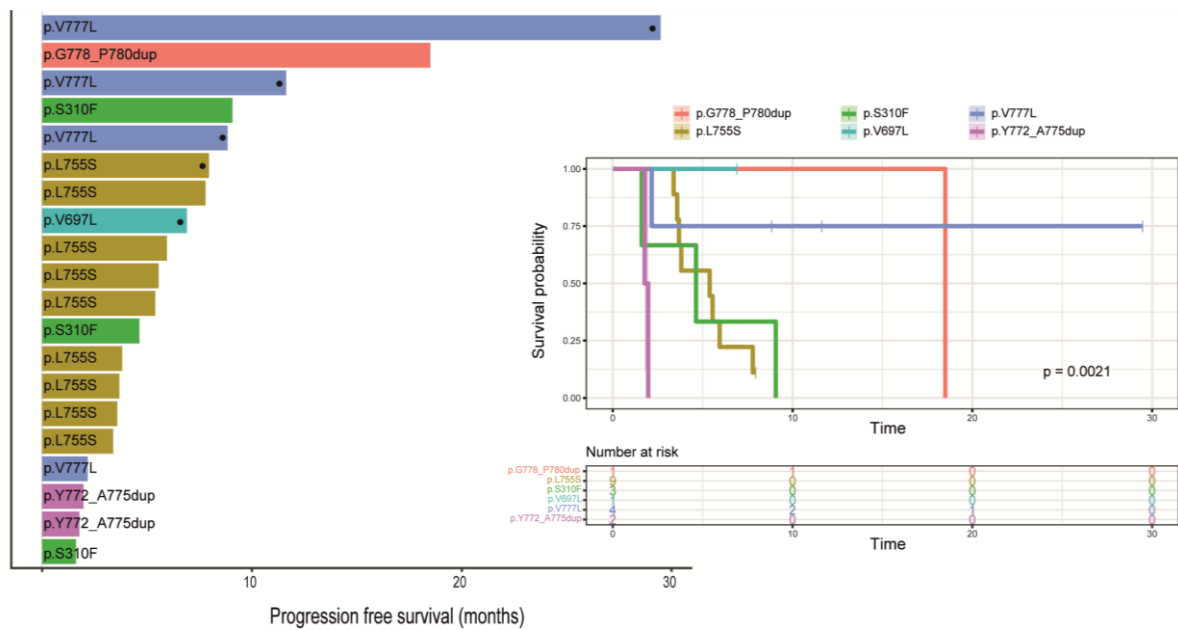


Figure 6.10. Progression free survival for patients enrolled into cohort B. *Left*, swimmer plot of progression free survival for each patient enrolled into cohort B, coloured by targeted *ERBB2* mutation. Censored patients are denoted by a black dot. *Right*, progression free survival in patients divided by targeted *ERBB2* mutation (median PFS: p.V697L undefined, p.V777L undefined, p.G778\_P780dup 18.5 months, p.L755S 5.4 months, p.S310F 4.6 months, p.Y772\_A775dup 1.9 months,  $p=0.0021$ ).  $p$  value from log-rank test.

Clonal dominance of the targeted mutation and whether the patient had a single pathogenic *ERBB2* mutation (monoclonal) or multiple (polyclonal) did not appear to associate with response (Figure 6.11). However this analysis was limited by the small number of patients in the cohort, with most patients having clonally dominant monoclonal mutations.



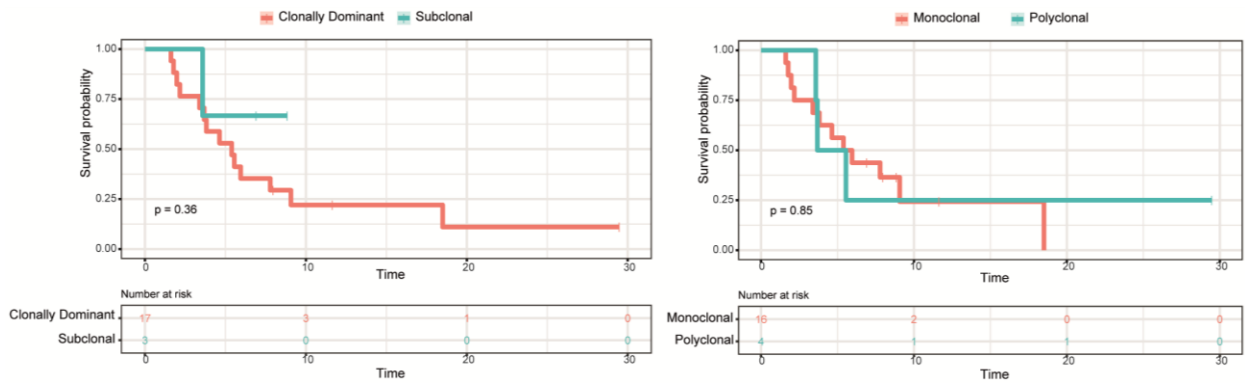


Figure 6.11. Progression free survival in cohort B according to baseline *ERBB2* mutation characteristics. *Left*, progression free survival in patients with clonally dominant targeted *ERBB2* mutations compared to those with subclonal alterations (clonal median PFS 5.4 months, subclonal median PFS undefined). *Right*, progression free survival in patients with monoclonal pathogenic *ERBB2* mutations compared to those with polyclonal disease (Monoclonal median PFS 5.7 months, polyclonal median PFS 4.6 months). p values from log rank test.

Baseline *PIK3CA* and *TP53* mutation status also did not appear to associate with response (Figure 6.12). Converse to what has been identified in a trial of neratinib in *ERBB2* mutant positive breast cancer elsewhere<sup>77</sup>, patients with a *TP53* mutation at baseline had a numerically longer PFS than those wild type.

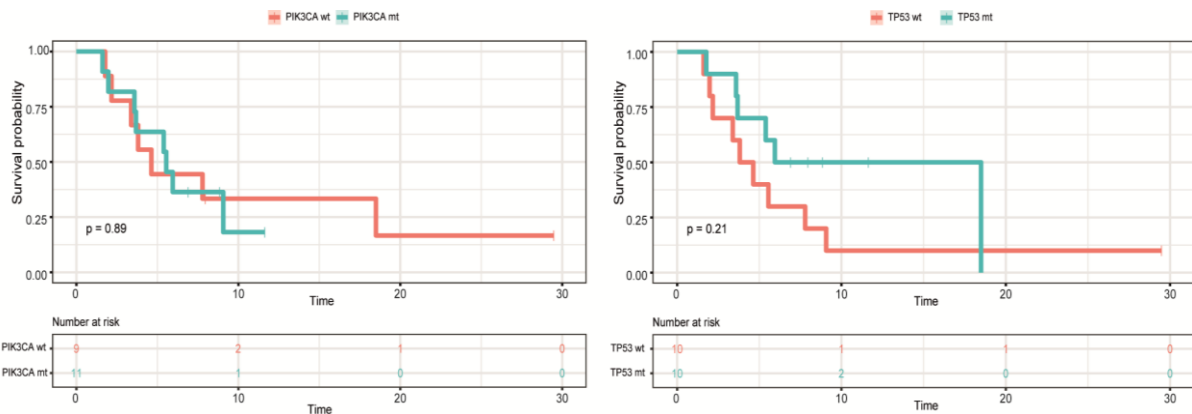


Figure 6.12. Progression free survival in cohort B according to baseline *PIK3CA* and *TP53* mutation status. *Left*, *PIK3CA* mutation status. *Right*, *TP53* mutation status. *PIK3CA* mutant median PFS 5.6 months vs *PIK3CA* wild type median PFS 4.6 months, HR 1.07 (95% CI 0.39 to 3.0). *TP53* mutant median PFS 12.2 months vs *TP53* wild type 4.2 months, HR 1.9 (95% CI 0.7 to 5.3). mt, mutant; wt, wild type.

Following prior data suggesting that activation of the TORC1 pathway (genes included in the panel: *PIK3CA*, *KRAS*, *AKT1*, *PTEN* and *TSC1*) drives resistance to neratinib<sup>153</sup>, this was investigated in cohort B. There was no evidence to suggest that patients with baseline mutations involved in the TORC1 pathway exhibited resistance to neratinib (Figure 6.13). Baseline

mutations involved in the MAPK pathway were also investigated for conferring resistance to treatment, with no evidence suggesting this was the case (Figure 6.13). It would be noted that this analysis is limited both by the small numbers of patients and the nature of a targeted panel sequencing, which limits the ability to assess all genes involved in the pathways.

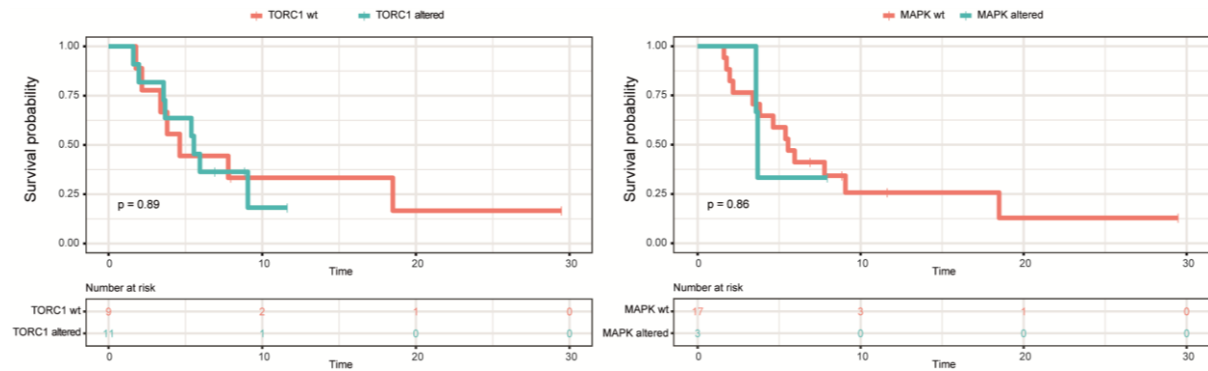


Figure 6.13. Progression free survival in patients divided by baseline mutations in the TORC1 and MAPK pathways. *Left*, TORC1 pathway. *Right*, MAPK pathway. TORC1 mutant median PFS 5.6 months, TORC1 wild type median PFS 4.6 months. MAPK mutant median PFS 3.7 months, MAPK wild type median PFS 5.6 months.

#### 6.4.2.4. plasmaMATCH cohorts C and D

Cohort C enrolled 18 patients (16 HR+ HER2-, 1 HR+ HER2+, 1 HR+ HER2 unknown) based on the presence of a plasma *AKT1* mutation (p.E17K in 17 and p.L52R in one), all of whom started treatment. Patients were treated with capivasertib (400mg PO BD 4 days on, 3 days off within 28 day cycles) with the addition of fulvestrant (500mg IM on Cycle 1 Days 1 and 15 and Cycle 2 onwards Day 1). Cohort D enrolled 19 patients, 13 with HR+ HER2- disease and 6 with TNBC disease all with activating AKT mutations (identified in plasma or tissue), all of whom started treatment. Overall, 6 patients were enrolled into cohort D based on an *AKT1* mutation (5 with p.E17K and 1 with p.L52R) and 13 based on a *PTEN* alteration (12 HR+ HER2-, 1 TNBC). Enrolled patients were treated with capivasertib 480mg PO BD 4 days on/3 days off within a 28 day cycle. All patients in cohort C had baseline sequencing, and 18/19 patients had baseline sequencing in cohort D.

Median PFS was significantly longer in the patients enrolled based on an *AKT1* mutation compared to those enrolled based on an AKT-activating mutation (in all cases, a *PTEN* variant; p=0.0029). Of patients enrolled with an *AKT1* mutation, 21% (5/24) had TNBC while the remainder had HR+ disease. Of the 13 patients enrolled based on *PTEN* alterations, one patient

(8%) had TNBC. The patient with TNBC enrolled with a *PTEN* alteration had a relatively long PFS of >22 weeks (Figure 6.14).

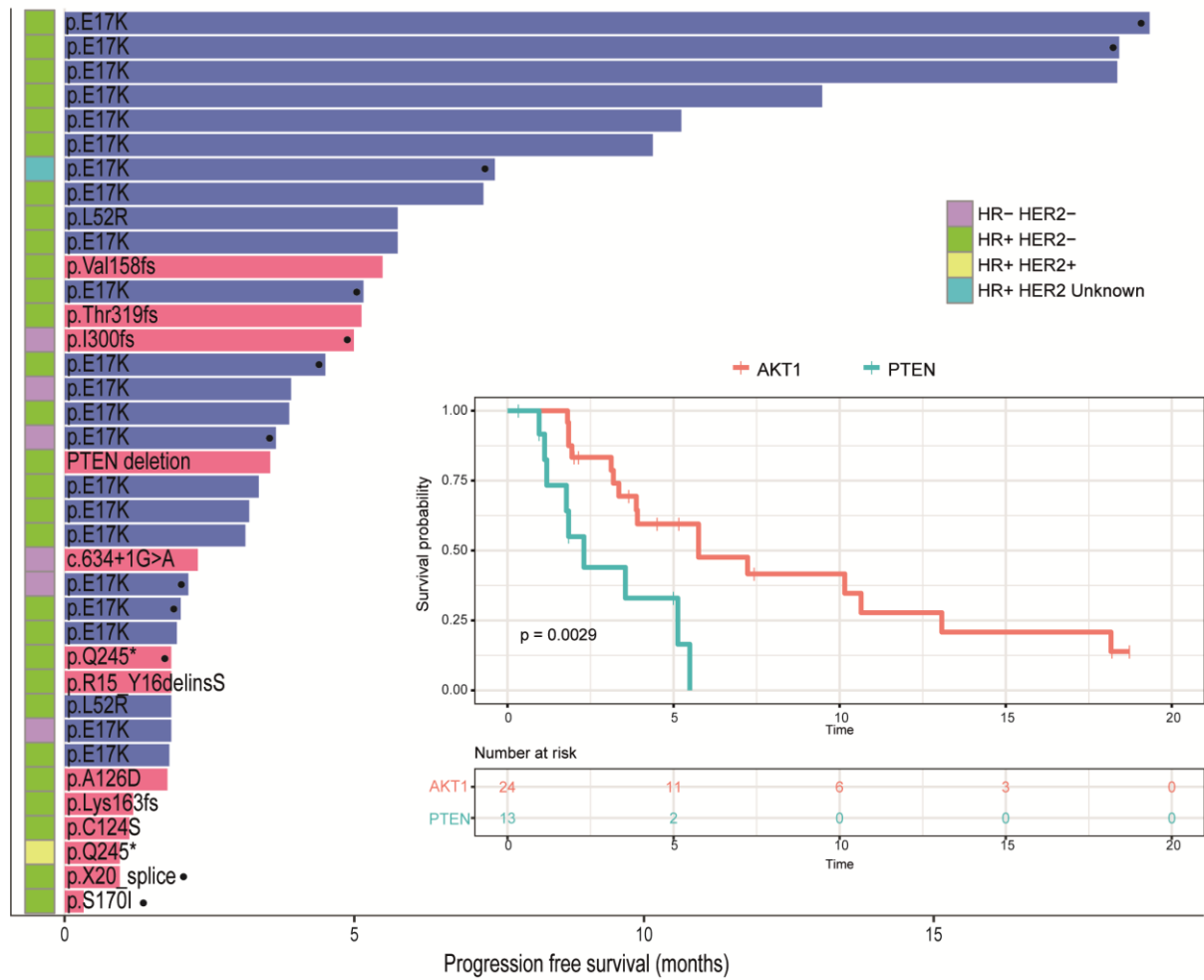


Figure 6.14. Progression free survival for patients enrolled into cohorts C and D. *Left*, swimmer plot of progression free survival for each patient enrolled into cohorts C and D, coloured by targeted gene (AKT1 = blue, PTEN = pink) and annotated with breast cancer phenotype. Censored patients are denoted by a black dot. *Right*, progression free survival in patients divided by targeted gene (median PFS: AKT1 5.7 months, PTEN 2.3 months,  $p=0.0029$ ).  $p$  value from log-rank test.

Of the 24 patients who entered cohorts C and D based on an *AKT1* mutation, 19 had an *AKT1* mutation detected by targeted sequencing, of which all were monoclonal and 17 were clonally dominant. Patients with a clonally dominant *AKT1* mutation at baseline had a significantly longer PFS than those with subclonal alterations ( $p=0.014$ , Figure 6.15).

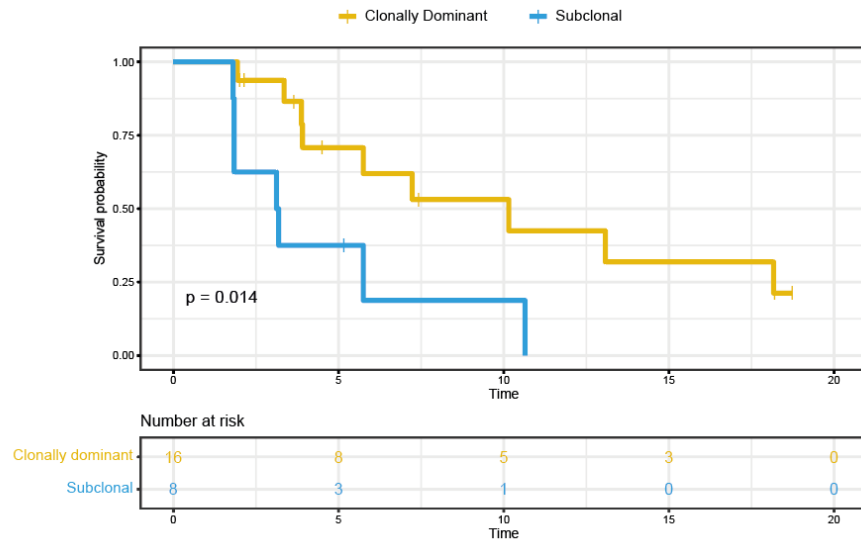


Figure 6.15. Progression free survival in patients enrolled into cohorts C and D with an *AKT1* mutation, divided by clonal dominance of the *AKT1* mutation. Clonally dominant median PFS 10.2 months, subclonal median PFS 3.2 months, HR 3.1 (95% CI 0.9 to 10.5). p value from log-rank test.

No genes showed a significant difference in their baseline incidence in *AKT1*-positive patients who received clinical benefit versus those that did not (Figure 6.16).

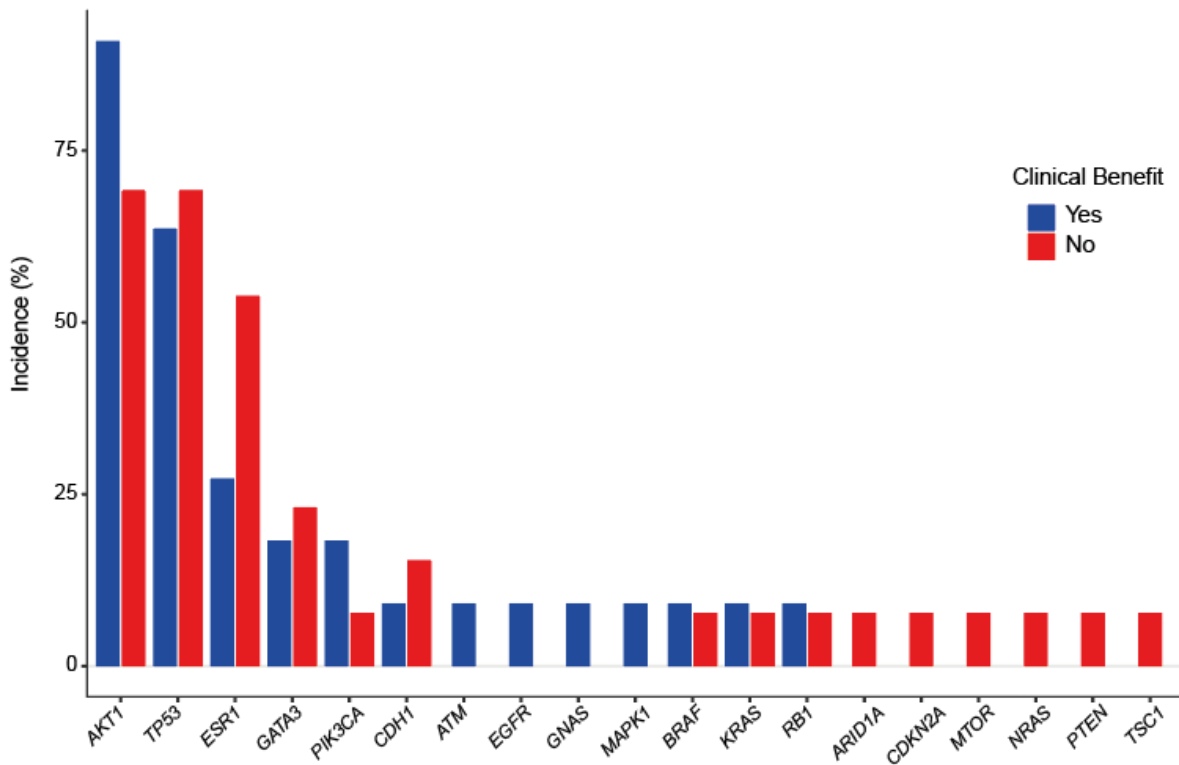


Figure 6.16. Baseline incidence of gene alterations in patients enrolled based on an *AKT1* mutation who achieved clinical benefit (n=11) versus those that did not (n=13). Comparison by Fisher's exact test.

As a pathway within which alterations can occur that confer resistance to other therapies<sup>34</sup>, the association between baseline MAPK pathway alteration and response was assessed in the *AKT1*-positive patients. Although there was a trend towards patients who gained clinical benefit from capivasertib to be MAPK wild type at baseline, this was not significant (Figure 6.17A). There was no significant difference in PFS in patients MAPK mutant positive at baseline versus wild type (Figure 6.17B).

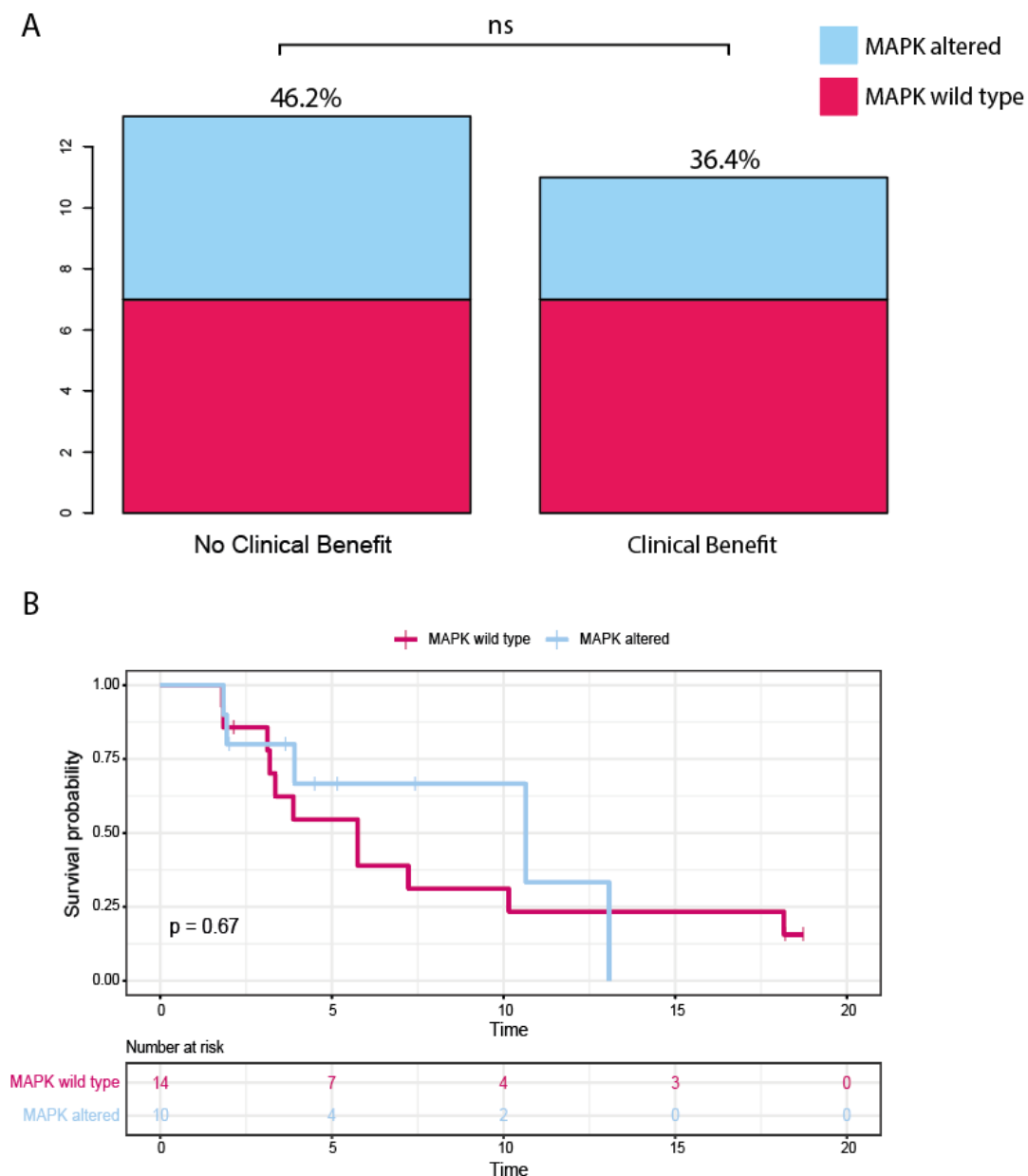


Figure 6.17. Association of MAPK alteration and response to capivasertib in patients enrolled based on an *AKT1* mutation. A, Incidence of baseline MAPK alteration in patients who gained clinical benefit from capivasertib

(n=11) versus those that did not (n=13). Comparison by Fisher's exact test. B Progression free survival of patients treated with capivasertib, divided by baseline MAPK alteration status. Median PFS MAPK altered 10.6 months, median PFS MAPK wild type 5.7 months, HR 1.3 (95% CI 0.45 to 3.47). p value from log rank test.

## 6.5. Discussion

In this Chapter the use of ctDNA as a prognostic and predictive biomarker was examined. CtDNA abundance is known to correlate with overall survival<sup>28,247</sup>, but the method of ctDNA purity assessment is not yet established. Here, two methods were compared to investigate their association with PFS and overall survival. mVAF and iChorCNA were both found to be prognostic for overall survival using different allele frequencies to denote high versus low purity (Figure 6.4). iChorCNA tended to estimate a higher purity in samples than mVAF. There are two potential reasons for this. Firstly, for heterozygous mutations that are present in every cancer cell where there is no copy number change, the mVAF would, at maximum, be 50% as there is a second un-mutated allele present. As iChorCNA purity is calculated in a different way, it is not limited to 50%, and could be anything up to 100%. Thus, it is to be expected that the differentiating mVAF cutoff is approximately half the value of iChorCNA, as it is based on data from a single allele. A second reason for the lower mVAF estimates is that use of mVAF to denote purity is based on the assumption that there is a truncal mutation and that this has been identified. If the truncal mutation is outside of the targeted panel utilised, mVAF would underestimate the purity. Nevertheless, both approaches appeared to be prognostic for overall survival. This data, and the respective purity cut-offs identified, require validation in a second dataset.

Conversely, ctDNA purity-adjusted *ESRI* clonal dominance did not appear to be a predictive biomarker of response to subsequent fulvestrant therapy (Figure 6.2). However there are important limitations of this analysis. Firstly, a major assumption was made regarding the biology of *ESRI* mutations when the mutations were summed together. The assumption here is that patients with polyclonal *ESRI* mutations most likely do not have activating *ESRI* mutations occurring in the same cell, creating redundancy, but instead develop activating *ESRI* mutations in parallel within separate clones. However, this may not be true, and it is possible that *ESRI* mutations co-exist within the same cell which would make summing the *ESRI* allele frequencies an erroneous approach. Secondly, this analysis relies upon there being a clonality-response relationship with *ESRI* mutations and response to fulvestrant. However this may not be true, and clinical data to date has not identified such a relationship<sup>111</sup>. Thus while no

association was identified here, this does not mean one does not exist either for *ESRI* mutations and fulvestrant or for other genomic alterations and targeted therapies.

The predictive value of baseline genomic alterations to specific targeted therapies within plasmaMATCH was next examined. In patients treated with fulvestrant within cohort A, several notable features were identified in the patterns of baseline *ESRI* mutations that associate with response. *ESRI* alterations exhibit patterns of mono- or polyclonality. Specifically, p.D538G, p.Y537S and p.Y537C appeared to be significantly more likely to occur in isolation than with other activating *ESRI* mutations (Figure 6.7C). *In vitro* data has demonstrated that these specific *ESRI* mutations are the most highly constitutively activating mutations in the absence of oestrogen<sup>111,248</sup>, potentially promoting the growth and clonal expansion of clones harbouring these mutations. Theoretically, cells harbouring alterations that are less constitutively active, such as p.L536R, may expand at a slower rate, supporting the outgrowth of a range of *ESRI* alterations within separate subclones which may have led to a heterogeneous *ESRI* mutant population and frequent co-mutation identified here. Given there are likely different underlying biological mechanisms for the development of certain *ESRI* mutation patterns, the drivers for this and clinical relevance is an area for further research.

Survival analysis demonstrated that patients with baseline p.Y537C alterations had a longer median progression free survival on fulvestrant (5.6 versus 2.0 months, HR 2.8 (95% CI 1.3 to 5.9), in agreement with *in vitro* data demonstrating a high sensitivity of cells harbouring this mutation to fulvestrant relative to other activating *ESRI* alterations<sup>110</sup>. Conversely, patients with a baseline p.Y537S mutation had a shorter median PFS (1.8 versus 3.5 months, HR 0.53 (95% CI 0.33 to 0.86) (Figure 6.8B). This concurs with prior data demonstrating p.Y537S is the most constitutively active variant, and exhibits greatest resistance to fulvestrant treatment *in vitro* and *in vivo*<sup>105,110,126</sup>, and in clinical trial data<sup>79</sup>. The change in clonal dominance patterns for the separate *ESRI* variants from baseline to EOT supports this data, with clones harbouring p.Y537S (and p.D538G) mostly gaining clonal dominance while those with p.E380Q, p.Y537C and p.S463P mostly losing clonal dominance (Figure 6.7B). This suggests that p.E380Q, p.S463P, and p.Y537C, may be highly sensitive to fulvestrant therapy, however due to their tendency to co-occur with other *ESRI* alterations (Figure 6.7C) this may not translate into being a baseline predictive factor for response to fulvestrant therapy.

Double *PIK3CA* mutations were found with a high incidence within the wider plasmaMATCH dataset (23% of patients with *PIK3CA*-mutant HR+ disease). Patients with double *PIK3CA*

mutations had a significantly shorter median survival on fulvestrant than those wild type or single-mutant positive ( $p=0.0036$ ). Vasan *et al* identified that double *PIK3CA* mutations lead to increased PI3K activity and enhanced pathway signalling<sup>231</sup>. Ultimately, this leads to cell proliferation and tumour growth<sup>231</sup>. Interestingly, patients with double *PIK3CA*-mutant disease were found to demonstrate enhanced sensitivity to subsequent PI3K inhibition<sup>231</sup>. The data presented here, conversely, suggests double *PIK3CA* mutations predict a worse response to subsequent fulvestrant. Thus, establishing *PIK3CA* mutation status appears to be an important clinical parameter that can predict response to subsequent therapy in hormone positive disease.

The baseline genomic features in cohort B were investigated for predictive biomarkers for response to neratinib. The small patients number reduces the power of the analysis to identify significant findings. Nevertheless, the data here indicated that baseline *ERBB2* variant targeted may be an important predictor of response, with patients with p.V777L having a longer median PFS (Figure 6.10). While responses have been noted in a basket study of *ERBB2*-mutant cancers<sup>166</sup>, and in breast cancers specifically<sup>77</sup>, patients with this baseline alteration were not identified as having a higher rate of response associated with it. One set of *in vitro* data demonstrated that MCF-7 cells harbouring a p.V777L mutation had a lower IC50 for neratinib than MCF-7 harbouring p.L755S<sup>140</sup>, however this was not recapitulated elsewhere<sup>144</sup>. Whether this translates into greater sensitivity of patients with this alteration to neratinib in the clinic is not established.

There was no supporting evidence here to suggest that *ERBB2* mutation clonality was an important factor in response to neratinib (Figure 6.11), but the analysis was limited by the number of patients. Within the SUMMIT trial cohort of *ERBB2*-mutant positive patients treated with neratinib plus fulvestrant, Smyth *et al* identified that none of the four breast cancer patients (of 44) with subclonal *ERBB2* alterations gained clinical benefit from neratinib therapy<sup>77</sup>. Similarly, there was no signal detected here to suggest that dual *ERBB2* mutation enhanced response to neratinib therapy (Figure 6.11), which is supported by data from the SUMMIT trial demonstrating that none of the four patients with dual concurrent *ERBB2* mutations gained clinical benefit<sup>77</sup>.

Cohorts C and D were combined to investigate for biomarkers of response to capivasertib therapy. It is clear that the patients enrolled based on an *AKT1* mutation gained greater benefit from capivasertib than those enrolled based on a *PTEN* aberration (Figure 6.14). These two groups are not directly comparable, with the patients in cohort C also receiving fulvestrant as



part of the regimen and also having a higher proportion of patients with TNBC who inherently have a more aggressive breast cancer subtype (21% versus 8%, respectively). However, the distinctly different median PFS demonstrated by the two groups prompted them to be considered separately.

For *AKT1*-directed patients, clonal dominance of the mutation was found to be a predictor of response to therapy, with patients with clonally dominant *AKT1* mutations having a significantly longer median PFS (Figure 6.15). This would concur with data published previously, which suggested that *AKT1* allelic imbalance (which can manifest as high allele frequency) is a predictor of response to capivasertib<sup>166</sup>. Analysis of a subset of patients within a phase 1 basket trial of capivasertib in *AKT1*-mutant cancer revealed that 57% of patients expressed allelic imbalance of *AKT1* p.E17K, of whom 48% had copy neutral loss of heterozygosity<sup>166</sup>. Patients demonstrating allelic imbalance of *AKT1* p.E17K had a significantly longer PFS on treatment than those without evidence of allelic imbalance (median PFS, 8.2 v 4.1 months, respectively; HR, 0.41; p=0.04). Most of the patients (92%) in the study had clonal *AKT1* alterations, precluding analysis of the relationship between clonality and response. However interestingly the authors analysed the genomic profile of eight separate metastatic sites in a patient with subclonal *AKT1* alteration prior to capivasertib treatment. They found that the lesion displaying the highest allele frequency of *AKT1* p.E17K also displayed the greatest response, suggesting p.E17K clonality is a predictor of response to capivasertib<sup>166</sup>. This supports the hypothesis that *AKT1* clonal dominance is a predictor of response to capivasertib.

The background rationale for inclusion of patients with PTEN aberrations centres on the knowledge that PTEN is a negative regulator of the PI3K/AKT pathway, with loss of PTEN expression leading to increased proliferation through enhanced signalling via the pathway<sup>174</sup>. In randomised phase II and III trials in metastatic prostate cancer, patients with PTEN-altered tumours had significant benefit from the addition of AKT-inhibitor ipatasertib to hormonal therapy compared to those in the overall intention-to-treat cohort, in which there was no significant benefit<sup>249,250</sup>. However this benefit does not seem to consistently translate to breast cancer. In HR+ breast cancer, the FAKTION trial identified that the benefit of the addition of capivasertib to fulvestrant was seen in the overall population but not in a subgroup which combined patients with PTEN-altered and PI3K-pathway altered tumours<sup>251</sup>, albeit the analysis may have been underpowered to detect a benefit in this group. A later phase I trial of capivasertib and fulvestrant HR+ patients with PTEN-altered tumours demonstrated a 24-week

clinical benefit rate of 17% in fulvestrant-naive and 42% in fulvestrant-pretreated patients<sup>176</sup>. In TNBC, data has been consistent for the application of capivasertib in *PIK3CA/AKT1/PTEN*-altered subgroups, with both the PAKT and LOTUS trials identifying enhanced response in patients with baseline *PIK3CA/AKT1/PTEN*-altered subgroups<sup>172,252</sup>, although it's not possible to differentiate the response rate based on PTEN-mutant groups specifically. The LOTUS trial did have a cohort of patients identified as 'PTEN-low' based on IHC criteria, but this group did not demonstrate a significant benefit from the addition of capivasertib<sup>252</sup>. In a preclinical study in gastric cancer, PTEN-null PDX models did not show any sensitivity to capivasertib unless treated in combination with taxotere where upon there was a synergistic effect<sup>253</sup>. It is notable that the one patient enrolled based on a *PTEN* alteration with TNBC had a relatively long response rate (>22 weeks).

A further factor to consider with PTEN alterations is that not all PTEN alterations result in loss of expression of a functional PTEN protein. Indeed, in the LOTUS trial, only 29% (14/48) of the patients who were PTEN-low on IHC had PTEN-alterations identified by NGS sequencing<sup>252</sup>. This would argue against the use of NGS to identify PTEN alterations as a biomarker for targeted therapy, even if the case for PTEN loss as a biomarker for response for capivasertib therapy was compelling.

Despite the aforementioned lack of data supporting the use of capivasertib in patients with PTEN aberrations, there are case reports and a small pilot study to suggesting that patients with germline *PTEN* alterations with the phenotypic changes associated with this, including cancers, exhibit high sensitivity to PI3K/AKT pathway inhibition<sup>254-258</sup>. Notably, two patients with Cowden's disease with associated breast cancers experienced exceptional responses to capivasertib<sup>258</sup>. This is despite the lack of data supporting the use of PTEN as a biomarker for AKT-directed therapy. However, these tumours have arisen due to the pro-tumorigenic environment created by a PTEN-null environment<sup>259,260</sup>, with the germline mutation being a fundamental clonal driving mutation. The tumour cells are therefore are likely to have hyperactive, and be dependent on, PI3K/AKT signalling<sup>154,260</sup>, underlying their sensitivity to the downstream AKT-inhibition. Thus, the evidence for PTEN alterations as a biomarker of response for capivasertib therapy is lacking, and likely depends upon the breast cancer subtype, treatment regimen, PTEN phenotype, and biological reliance of the cancer cell on PI3K/AKT signalling.

## 6.6. Conclusion

In this Chapter, ctDNA has been used to explore the predictive biomarkers of response to targeted therapies and prognostic markers. This analysis was limited by the small numbers of patients within each cohort, although the larger number in cohort A allowed for wider analysis of potential biomarkers. A further limitation is the immature dataset, with many patients being censored, which will inevitably have led to a higher number of type II errors.

Despite this, some interesting findings were made. Firstly, the purity of the ctDNA, as assessed through low-pass whole genome sequencing and maxVAF, was able to differentiate patients with a poorer prognosis. In cohort A, the baseline *ESR1* mutation being targeted was predictive of response, while the greater number of patients allowed for more in depth scrutiny of the patterns of co-mutation, with a suggestion that different *ESR1* mutations demonstrate disparate responses to fulvestrant treatment. Finally, in cohorts C and D, patients with clonally dominant *AKT1* mutation had significantly longer PFS on treatment than those with subclonal alterations. Meanwhile *PTEN* did not appear to be a biomarker for capivasertib therapy, while the data suggests that patients with a germline *PTEN* mutations with highly active PI3K/AKT signalling will benefit from therapy.

## 7. Chapter 7. Use of ctDNA to investigate targeted therapy resistance mechanisms

### 7.1. Introduction

Treatment resistance is a major issue in oncology, ultimately preventing cure of patients with ABC. Resistance arises in part due to the selective pressure applied by therapy, and can be considered as occurring early or late. Early resistance measures can arise from non-genomic cell modifications such as disruption in signalling feedback pathways and epigenetic changes<sup>261</sup>. Meanwhile prolonged treatment exposure may give rise to late, or acquired, resistance which can manifest as genomic changes affecting either the drug target itself or down/upstream of the target<sup>261</sup>.

Acquired resistance can arise through the evolutionary selection pressure applied by therapy<sup>72,262</sup>. This pressure favours the survival of clones which either already harbour a mutation that confers treatment resistance<sup>54,263</sup>, or supporting the acquisition and propagation of newly acquired resistance alterations<sup>72,264</sup>. In the presence of the applied therapy, these resistance mutations confer a survival advantage to the subclones harbouring the mutation, supporting clonal expansion and, ultimately, progressive disease.

Tissue biopsy sequencing studies have helped to elucidate genomic mechanisms of resistance. The use of ctDNA analysis to identify novel mechanisms of resistance, however, is less well defined. One major reason for this is that discovery of novel resistance mechanisms generally requires a broad sequencing approach such as whole exome sequencing. This approach requires a high purity of the sample, and given 65% of ABC patients have a purity below 10%<sup>85</sup>, considered to be the threshold at which one can perform ctDNA exome sequencing<sup>79,85</sup>, this makes novel mutation discovery by this method challenging. Furthermore, the cost involved in the depth and breadth of sequencing mean that this approach is not commonly used.

Increasingly, commercial targeted sequencing panels are becoming available which include genes commonly involved in cancer. Through a targeted sequencing approach, which is better able to cope with low purity samples as is frequent with plasma-derived ctDNA samples, it may be possible to identify putative resistance mechanisms, albeit discovery limited to the genes and genome regions included in the panel.

One potential benefit of using ctDNA analysis in resistance discovery is the ability of ctDNA to dissect spatial heterogeneity and sample heterogeneous metastatic sites for subclonal alterations<sup>54</sup>. Given that resistance mutations often arise subclonally, an approach which is not limited to sampling a single site of biopsy may, theoretically, have a greater chance of identifying emerging resistance mutations.

This Chapter describes the use of baseline and end-of-treatment (EOT) ctDNA testing using a targeted sequencing approach with the aim of identifying novel putative resistance mechanisms to a range of targeted therapies.

## 7.2. Hypothesis

Analysis of ctDNA using targeted sequencing can reveal putative resistance mechanisms.

## 7.3. Aims

- 1) Compare baseline and EOT sequencing to identify resistance mechanisms for each of the targeted therapies in plasmaMATCH
- 2) Further investigate any putative resistance mechanisms

## 7.4. Results

### 7.4.1. Cohort A: extended dose fulvestrant

Cohort A enrolled 84 patients for treatment with extended dose fulvestrant (500mg fulvestrant (IM) on Cycle 1 Days 1, 8 and 15 and Cycle 2 onwards Days 1 and 15). The majority of patients had HR+ HER2- disease (n=80, 95.2%), with two patients having HR+ HER2+ disease (3.6%) and one patient (1.2%) with HR+, HER2 unknown disease. Of the 84 patients, 80 commenced treatment and 74 had on treatment RECIST-assessable imaging (Chapter 6, Figure 6.5). 79 patients had baseline plasma sequencing results, and 69 had both baseline and end-of-treatment (EOT) sequencing (Chapter 6, Figure 6.5).

The majority of patients (n=50, 72.5%) maintained their poly- or monoclonal *ESR1* mutation pattern following fulvestrant treatment, with 5.8% (n=4) acquiring polyclonal disease and 4.3% (n=3) becoming *ESR1* wild type/having undetectable *ESR1* mutations. In *PIK3CA*, *TP53* and *GATA3*-mutant disease, the majority of patients maintained monoclonal disease (Figure 7.1).

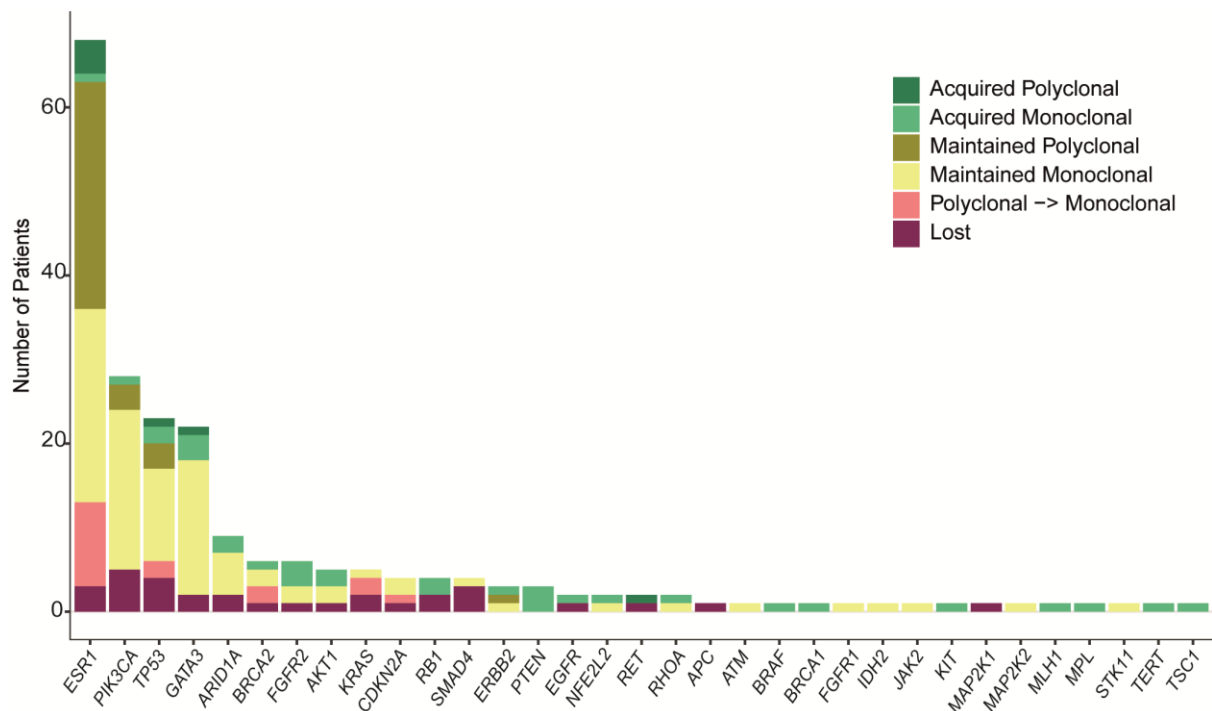


Figure 7.1. Baseline to EOT changes in gene mutation polyclonality in cohort A

Overall, 50.7% of patients (n=35) acquired alterations at the EOT time point, with 27.5% (n=19) acquiring more than one alteration. Acquired alterations occurred in genes involved in a range of signalling pathways, but most markedly in oestrogen and PI3K/AKT signalling pathways (Figure 7.2A). The genes most commonly found to have acquired alterations were *ESR1* (n=14 patients), *GATA3* (n=5 patients) and *TP53* (n=4 patients) (Figure 7.2A). The gene variant loci most frequently acquiring pathogenic alterations following treatment were all found within *ESR1*, namely p.L536P/H/R (n=9 alterations), p.Y537C/N/S (n=8 alterations), p.F404I/L/V (n=5 alterations), p.D538G (n=4 alterations) (Figure 7.2A). The maximum VAF (maxVAF) was not significantly different from baseline (mean 20.2 +/- SD 14.3) to EOT (mean 18.1 +/- SD 13.1) in patients with acquired *ESR1* alterations (Figure 7.2B). Potentially targetable alterations were identified in 24.6% (n=17), with 4.3% (n = 3) patients acquiring potentially targetable *PTEN* and *BRCA1/2* alterations, respectively (Figure 7.2C). Incidence of acquired alterations was not different in the patients gaining clinical benefit versus those not (Figure 7.2D).

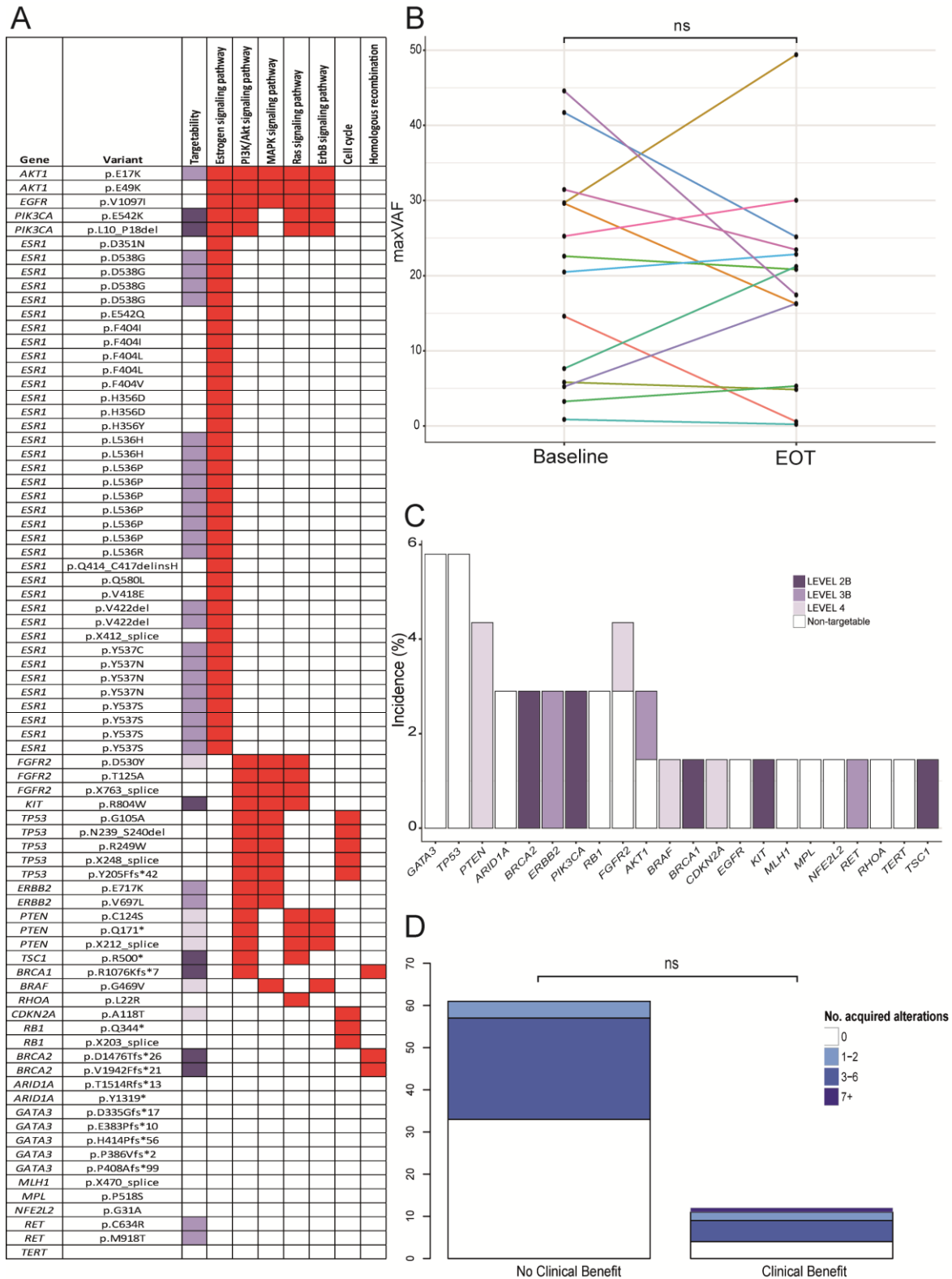


Figure 7.2. Acquired alterations in cohort A. A, all acquired mutations in cohort A, annotated with targetability (legend as per 7.2C) and signalling pathway involvement. B, Change in the maximum VAF (maxVAF) in patients with acquired *ESR1* alterations from baseline to EOT. Each coloured line represents the change in mVAF from

baseline to EOT in one patient. Comparison by paired t-test. C, Incidence of acquired gene mutations, coloured by targetability (Chapter 2, section 2.1.12.6). Excludes *ESRI* acquired mutations. D, Number of acquired mutations in patients gained clinical benefit (CR/PR/SD  $\geq$  24 weeks) vs not. Comparison by Chi-squared test.

## 7.4.2. Investigation of *ESRI* p.F404 as a potential resistance mechanism

### 7.4.2.1. *ESRI* p.F404 background

The number of alterations acquired at the *ESRI* p.F404 locus prompted further investigation to ascertain whether these alterations represent putative resistance mechanism to fulvestrant therapy. Within cohort A, three patients (3/69, 4.3%, or 3/26, 11.5% of patients with PFS  $\geq$  16 weeks) developed one or more base changes at this locus to achieve an amino acid change of phenylalanine to one or more of isoleucine, valine and leucine. The latter three amino acids differ from phenylalanine in that they lack an aromatic ring (Figure 7.3).



Figure 7.3. Amino acid structure of the wild type amino acid at *ESRI* p.F404, phenylalanine versus mutant variants isoleucine, valine and leucine.

Review of the alteration location identified the p.F404 locus as being within the ligand binding domain of *ESRI* (Figure 7.4). Furthermore, structural studies of oestrogen binding suggest that there is a Pi-Stacking bond (attractive, noncovalent interaction between aromatic rings) between the aromatic ring of p.F404 and an aromatic ring on oestrogen. As fulvestrant has similar in structure to oestrogen, this prompted the hypothesis that an alteration at this locus to an amino acid lacking an aromatic ring may alter the binding efficiency of fulvestrant, and represent a mechanism of resistance to the drug.



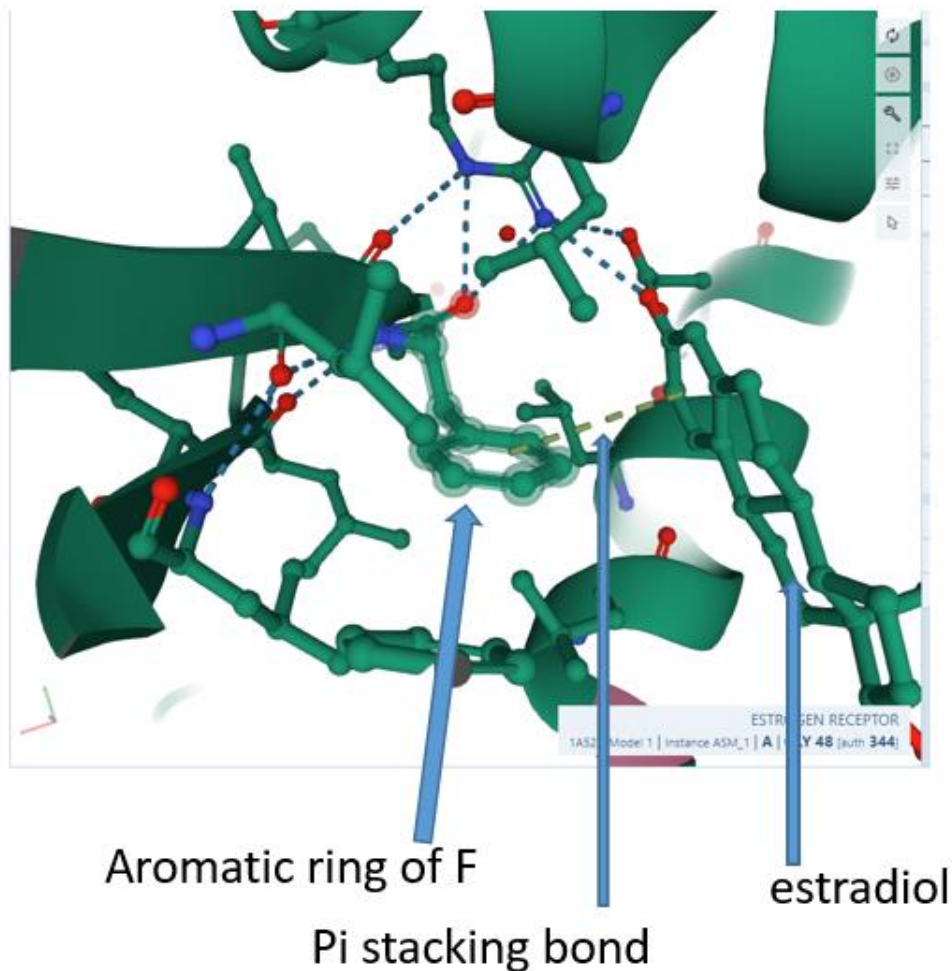


Figure 7.4. Illustration of the ligand binding pocket of oestrogen receptor alpha bound to oestrogen (estradiol). The light green dotted line represents a Pi-Stacking bond. Illustration from <https://www.rcsb.org/3d-view/1A52?preset=ligandInteraction&sele=EST>, PDB ID 1A52<sup>265</sup>. F, fulvestrant.

Review of the alteration incidence within the plasmaMATCH dataset revealed that one further patient within the cohort of 800 patients with targeted sequencing results had an *ESR1* p.F404 mutation (p.F404V). This patient had previously received fulvestrant. Of the three patients in cohort A with an *ESR1* p.F404 mutation, all had available baseline and EOT sequencing demonstrating that the mutation was acquired at the EOT time point (Chapter 6, Figure 6.5). In total, five different base changes were found to occur at locations 1 and 3 within the codon to create three different amino acid changes (Figure 7.5). All the mutation instances were subclonal.

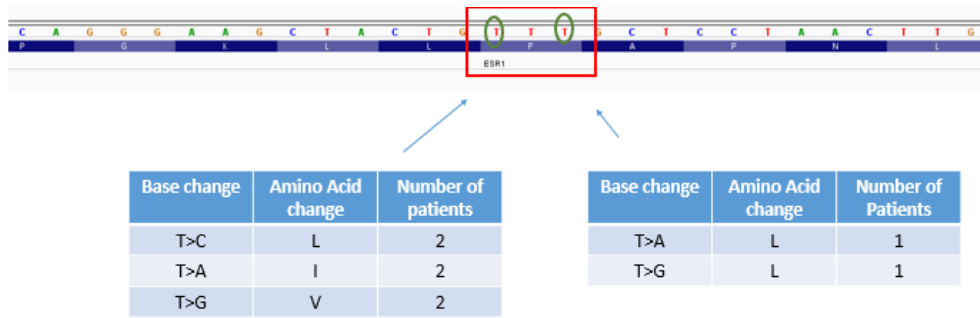


Figure 7.5. *ESRI* p.F404 base changes. The base changes occurred at location 1 and 3 (illustrated with green circles) within the *ESRI* p.F404 codon (illustrated with the red box).

The patients in cohort A who acquired the mutation all had a relatively good response to fulvestrant, with a minimum PFS of 17 weeks. Of note, all four patients demonstrated a co-mutation with *ESRI* p.E380Q (one of the in-cohort patients only had this additional *ESRI* mutation at baseline, not at EOT), and three patients were co-mutant with *ESRI* p.D538G. Due to their close genomic location and resulting likelihood of appearing on the same read, it was possible to assess whether the p.F404 variant/s occurred in *cis* or *trans* in the three patients with the co-existing p.E380Q mutation. Of the 7 different variants occurring in the 3 different patients, 6 occurred in *cis* with p.E380Q while one occurred in *trans* (Figure 7.6). All four patients had multiple activating *ESRI* variants which may have been in *cis* or *trans* with p.F404.

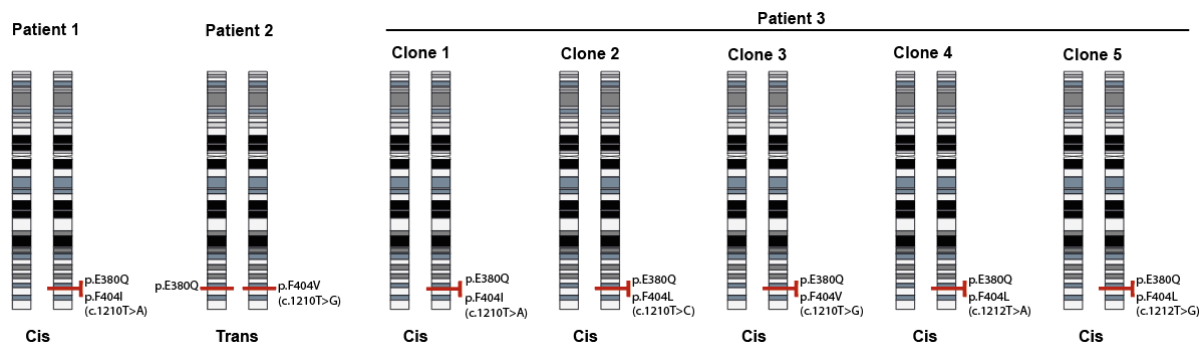


Figure 7.6. *Cis/trans* analysis of the *ESRI* p.E380Q and p.F404 variants.

#### 7.4.2.2. Transient transfection of *ESRI* p.F404L

To further investigate the *ESRI* p.F404 variant, a transient transfection approach was taken. Plasmids, or oestrogen receptor constructs (ERC), carrying the open reading frame of *ESRI* were custom designed to harbour wild type *ESRI*, *ESRI* single mutant for D538G, E380Q or F404L, or compound mutations D538G\_F404L or E380Q\_F404L, or the backbone empty

vector (Chapter 2, section 2.3.2.1). MCF-7 were co-transfected with the ERC alongside an ERE-luciferase (ERE-Luc) and beta-galactosidase ( $\beta$ -gal) plasmids (Chapter 2, sections 2.3.2.2 and 2.3.2.3). The ERE-Luc plasmid (Methods section 1.6.2.2) consists of the open reading frame of an oestrogen-response element linked to luciferase gene, which allows quantification of the oestrogen response activity of the cell through luciferase assay. The  $\beta$ -gal plasmid is a control plasmid allowing normalisation for factors such as number of cells and transfection efficiency.

Prior to assessment of ERC activity in oestrogen and drug-treated conditions, the activity of the ERCs in oestrogen-deprived conditions compared to oestrogen-containing was established. MCF-7 were cultured in an oestrogen-deprived media for 7 days prior to transfection. The cells were subsequently transfected in the absence of oestrogen. The transfection protocol described in Chapter 2, section 2.3.4 was followed, but with one set of transfected MCF-7 being changed into oestrogen-containing media 24 hours prior to harvesting, while a second set was maintained in oestrogen-free conditions. Cells were harvested for lysates, and the  $\beta$ -gal and oestrogen-treated empty vector normalised ERE-luciferase activity of each transfected condition compared (Figure 7.7). Results demonstrated suppression ERE-luciferase in maintained oestrogen-deprived conditions across all ERCs. D538G and E380Q demonstrated higher ERE-luciferase activity in oestrogen-deprived conditions than empty vector or wild type transfected MCF-7, in concordance with their known constitutively activating nature<sup>90,105,110,126</sup>, while F404L demonstrated lower activity in oestrogen deprived conditions compared to wild type (Figure 7.7). Notably this data is derived from just one replicate repeat assessment is required, and conclusions cannot be drawn from this data.

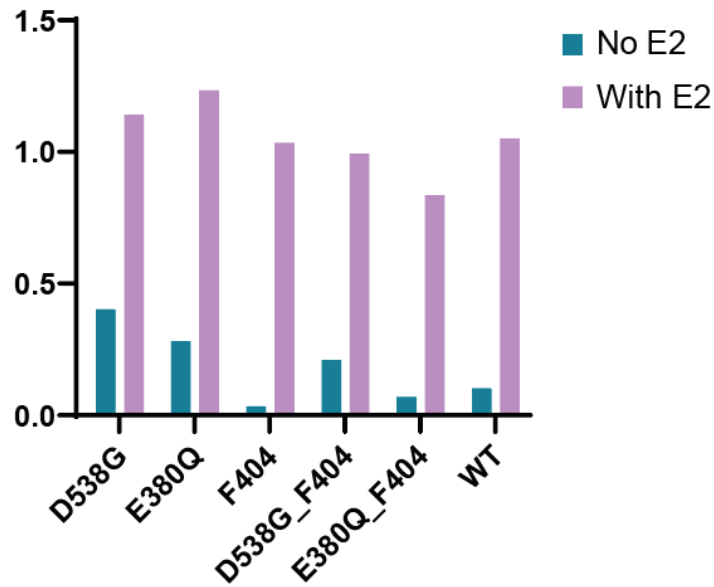


Figure 7.7. Activity of *ESR1* mutant variants in oestrogen-deprived compared to oestrogen-containing conditions, following prolonged oestrogen deprivation. Normalised ERE-luciferase activity (y-axis) of MCF-7 transfected with each respective ERC (x-axis) in maintained oestrogen deprived conditions compared to oestrogen-containing conditions. ERE-luciferase activity is normalised to beta-galactosidase, and then to the ERE-luciferase activity of E2-treated MCF-7 transfected with empty vector. n=1 replicate. E2, oestrogen

Next, to confirm that the transfected MCF-7 demonstrated the expected pattern of proteins according to the ERC transfected, western blot was undertaken. This confirmed that all the transfected MCF-7 expressed both DYK-tagged ESR1 and the endogenous ESR1 alpha (Figure 7.8). To verify these findings, this experiment requires repeating with two control conditions: non-transfected MCF-7 and MCF-7 transfected with empty-vector.

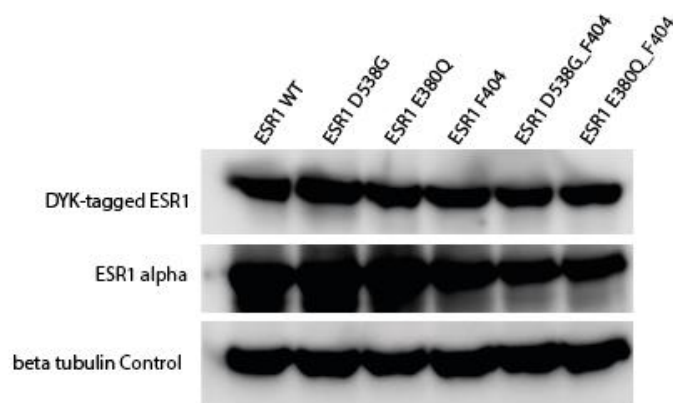


Figure 7.8. Western blot of transfected MCF-7, stained for ESR1 alpha, DYK-tagged ESR1 and beta-tubulin (control).

The comparative activity of each ERC in oestrogen compared to oestrogen plus fulvestrant was next established. Results demonstrate that the compound mutations (D538G-F404L and E380Q-F404L) confer relative resistant to fulvestrant, as assessed using the activity of the ERE-luciferase reporter, while the single mutations D538G, E380Q and F404L mutations do not demonstrate resistance to fulvestrant (Figure 7.9). Interestingly, the F404L transfected MCF-7 demonstrated relatively low activity in oestrogen compared to MCF-7 transfected with wild type or mutant *ESR1* (Figure 7.9).

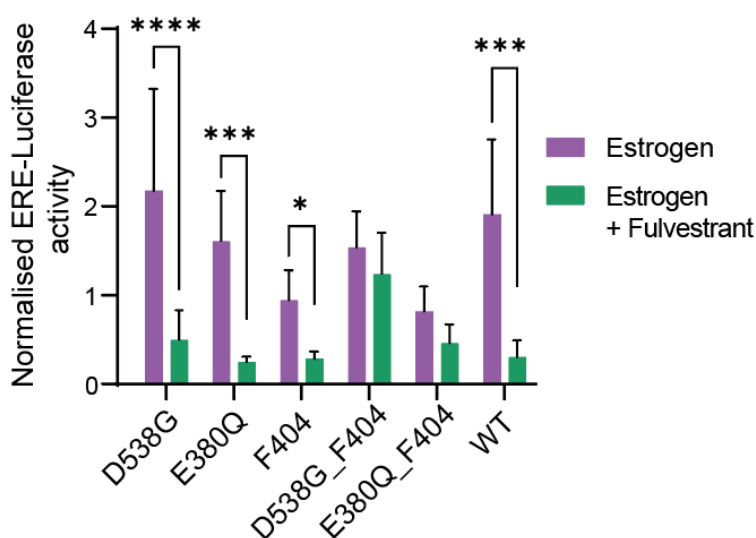


Figure 7.9. Activity of ERC transfected MCF-7 in oestrogen and fulvestrant compared to oestrogen alone. MCF-7 were transiently transfected with each oestrogen receptor construct (ERC), as annotated. ERE-luciferase activity, normalised to beta-galactosidase and ERE-luciferase activity of oestrogen-treated empty vector, is illustrated for each transfected ERC. Following transfection, cells were treated in oestrogen containing media +/- fulvestrant (500nM) for 24 hours before harvesting of the cell lysates. ERE-luciferase activity is then normalised to the ERE-luciferase activity of oestrogen-treated MCF-7 transfected with empty vector. Oestrogen-containing versus oestrogen plus fulvestrant ERE-Luc activity: D538G  $p < 0.0001$ , E380Q  $p = 0.0006$ , F404  $p = 0.0443$ , D538G\_F404  $p = ns$ , E380Q\_F404  $p = ns$ , wild type  $p = 0.0001$ . Comparison of normalised ERE-luciferase activity with and without fulvestrant for each ERC was achieved using a 2-way ANOVA. Results are derived from three biological replicates.

Finally, the activity of the individual and compound mutations in a range of anti-oestrogen drugs was assessed. Novel SERDs AZD9833<sup>266</sup> and elacestrant<sup>267</sup> were selected for testing following promising clinical trial data. Activity was also assessed with tamoxifen, as activating *ESR1* mutations do not tend to arise following treatment with tamoxifen suggesting that the compound mutation may equally not be effective in producing resistance to this drug.

To ascertain the correct experimental doses of AZD9833, elacestrant and tamoxifen cell viability assays were undertaken at a range of drug concentrations (Figure 7.10).

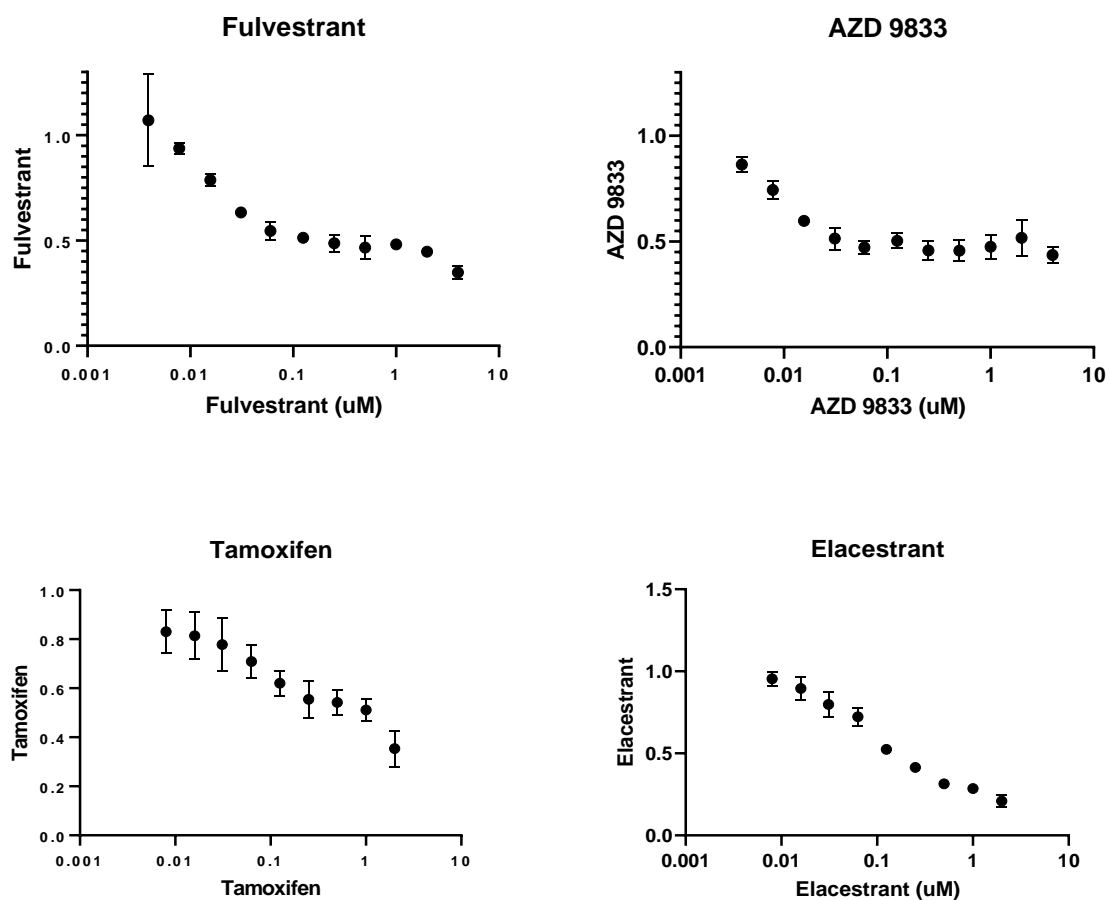


Figure 7.10. Cell viability curves for fulvestrant, AZD9833, tamoxifen and elacestrant, respectively. On the y-axis is luminescence, indicative of the quantity of ATP present enabling quantification of the number of viable cells present.

From the drug viability assays, drug concentrations were selected for use in MCF-7 ERC transfection (Table 7.1).

Drug	Concentration (nM)
Fulvestrant	500
AZD9833	50
Elacestrant	500

Tamoxifen	250
-----------	-----

Table 7.1. Drug concentration for ERC transfection experiments.

Transfected MCF-7 were maintained in oestrogen with or without drug treatment. The lysates were harvested and compared for ERE-luciferase activity. MCF-7 transfected with the compound mutations D538G\_F404L and E380Q\_F404L, and single F404L demonstrated a significantly higher level of ERE-luciferase activity in fulvestrant-treated conditions relative to their activity in oestrogen alone compared to wild type transfected MCF-7 ( $p=0.02$ ,  $p=0.003$  and  $p=0.02$ , respectively, Figure 7.11). Tamoxifen-treated E380Q\_F404L transfected MCF-7 demonstrated marginally higher ERE-luciferase activity in tamoxifen compared to wild type transfected tamoxifen-treated MCF-7 ( $p=0.02$ ), which concurs with prior *in vitro* data showing that p.E380Q mutants demonstrate relative resistance to tamoxifen compared to the parental line<sup>90</sup>. No other single or compound mutant transfected MCF-7 demonstrated significantly different drug-treated ERE-luciferase activity compared to the activity in wild type transfected MCF-7 treated with the same drug, demonstrating that the resistance of the compound mutations was singular against fulvestrant and not present to other tested therapies.

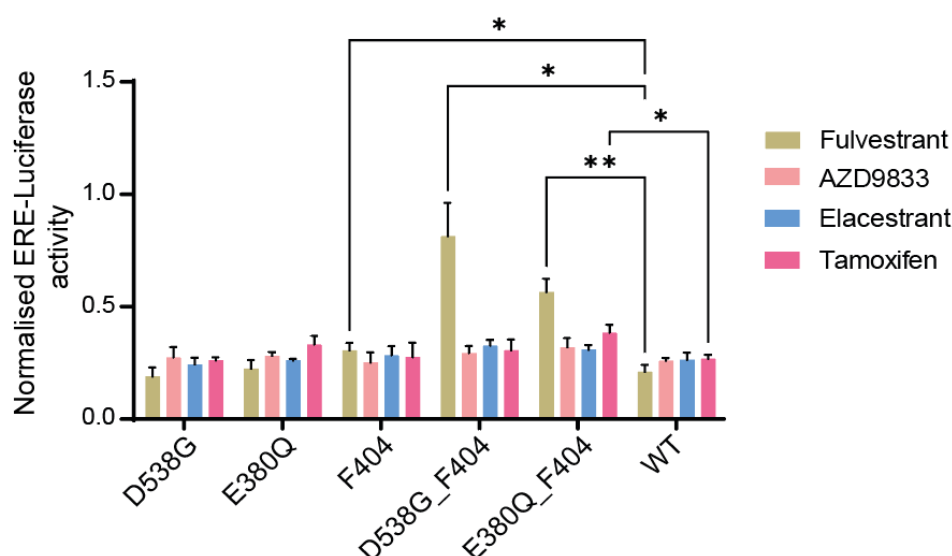


Figure 7.11. ERE-luciferase activity of each ERC in drug treated conditions relative to its activity in oestrogen alone. Activity in each of fulvestrant, AZD9833, elacestrant and tamoxifen was assessed. The activity of each ERC was compared to activity in *ESR1* wild type transfected MCF-7. D538G\_F404L, E380Q\_F404L and F404L alone all demonstrated significantly higher ERE-luciferase activity in fulvestrant-treated conditions than fulvestrant-treated wild type ( $p=0.02$ ,  $p=0.003$  and  $p=0.02$ , respectively). E380Q\_F404L also demonstrated

significantly higher activity in tamoxifen than wild type ( $p=0.02$ ). Results are derived from three biological replicates.

#### 7.4.3. Cohort B: neratinib +/- fulvestrant

Cohort B enrolled 21 patients, of whom 20 started treatment (17 patients with HR+ HER2- disease, one patient with HR+ HER2+ disease, and two patients with HR- HER2+ disease). Patients were treated with neratinib (240mg PO OD), with the addition of fulvestrant (500mg IM on Cycle 1 Days 1 and 15 and Cycle 2 onwards Day 1) in HR+ patients. Of the patients who started treatment, all had baseline sequencing available and 19 had EOT sequencing available.

The majority of patients (18/19, 94.7%) maintained a monoclonal or polyclonal pathogenic *ERBB2* alteration status throughout treatment (Figure 7.12). Two patients (2/19, 10.5%, both with HR+ HER2- disease) acquired polyclonal *PIK3CA* disease from monoclonal disease at baseline, one of whom developed three further pathogenic *PIK3CA* alterations additional to their baseline p.H1047R mutation. One patient maintained their polyclonal *PIK3CA* status. Five patients (26.3%) maintained polyclonal *TP53* mutation positive disease, while one further patient (5.3%) acquired polyclonal *TP53*-mutant disease and two patients (10.5%) lost their pathogenic *TP53* mutant status altogether (Figure 7.12).

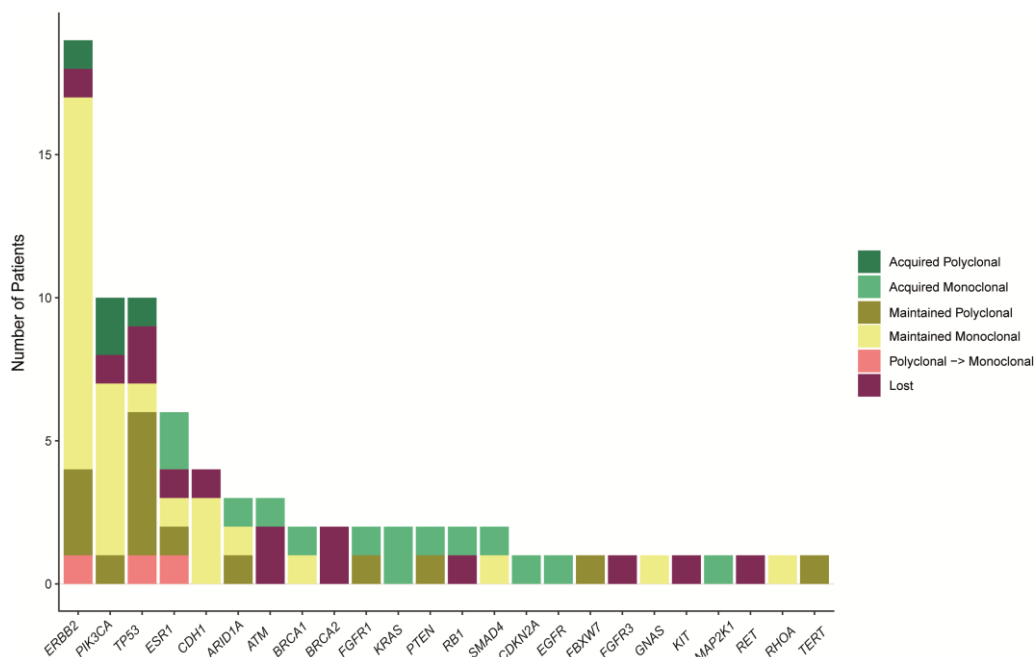




Figure 7.12. Baseline to EOT changes in gene mutation polyclonality, per gene, in cohort B.

Alterations were acquired in genes involved in a number of signalling pathways, but most markedly in the PI3K/AKT signalling and MAPK signalling pathways (Figure 7.13A). The most commonly acquired alteration was in *TP53*, with 26.3% of patients (n=5) gaining a non-targetable pathogenic alteration in the gene (Figure 7.13A and C) while the same number lost a *TP53* alteration. One patient appeared to have highly mutagenic *TP53* gene, acquiring 10 pathogenic *TP53* alterations in the course of neratinib treatment. Three patients (15.8%) gained pathogenic alterations in *ERBB2* and *ESR1*, respectively. The three patients who gained *ERBB2* mutations did not have a significantly different baseline to EOT change in maxVAF, albeit all were numerically higher suggesting higher purity samples at EOT (Figure 7.13B). Two patients gained potentially targetable alterations in *PIK3CA*, while one patient gained a potentially targetable alteration in *BRCA1* (Figure 7.13C). There was no significant difference in the number of acquired alterations between patients who received clinical benefit in cohort B (SD/PR for at least 24 weeks) compared to those who did not (Figure 7.13D).

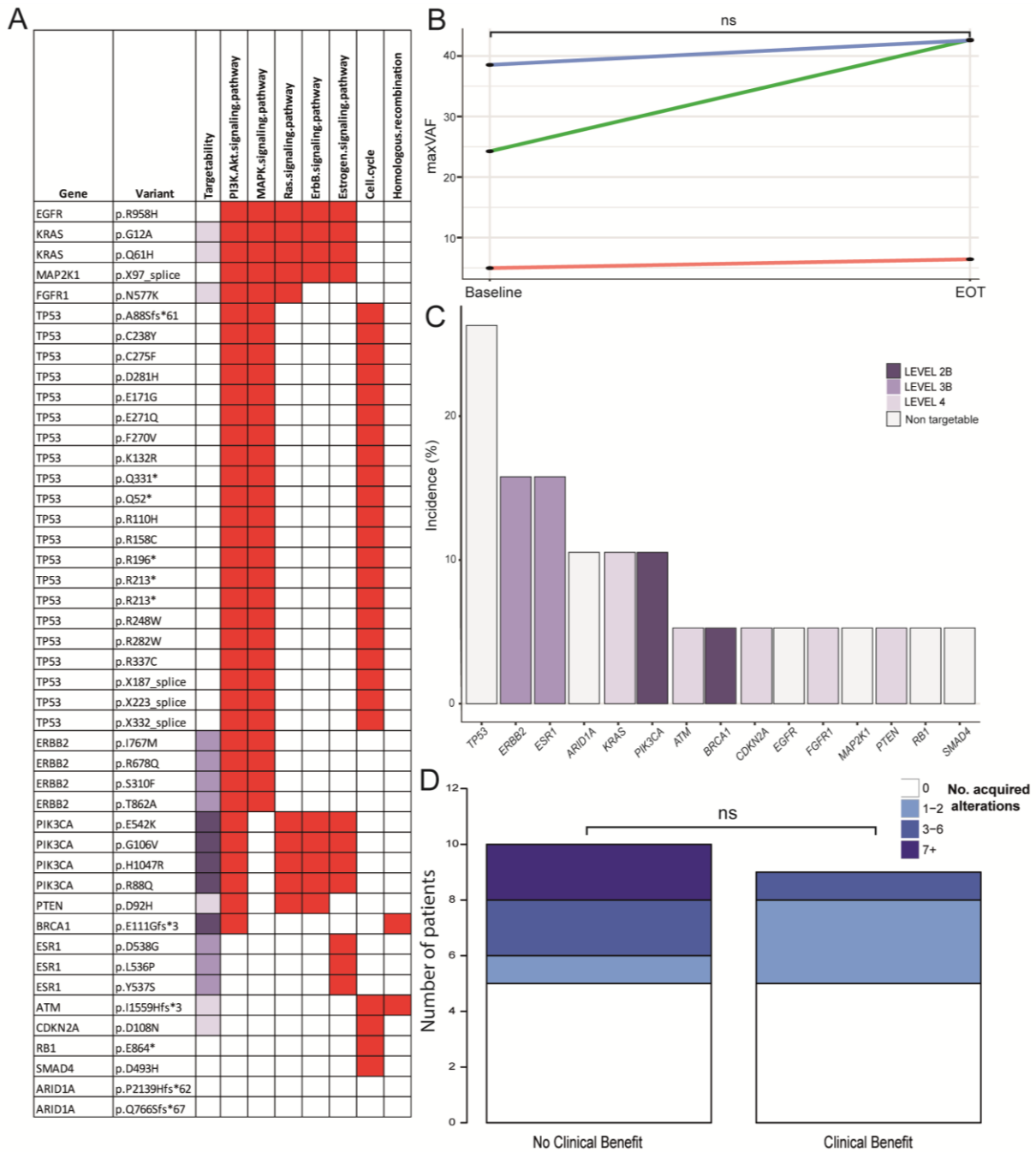


Figure 7.13. Acquired alterations in cohort B. A, List of acquired pathogenic alterations in Cohort B, annotated with level of evidence supporting targetability (key as per legend within 7.13C), and pathway association in red. B, Baseline to EOT maxVAF change in the three patients who acquired *ERBB2* mutations,  $p=0.74$ . Comparison by paired t-test. C, Incidence of acquired alterations within cohort B, coloured by level of evidence supporting targetability (Chapter 2, section 2.1.12.6). D, Comparison of number of acquired alterations between patients who gained clinical benefit in cohort B and those that did not. Comparison using Chi-squared test.

As copy number amplifications are common in the *HER2* gene, the incidence of copy number gains within this gene and other assessable genes within the cohort was assessed. Purity

adjusted gene copy numbers within cohort B were numerically higher at the EOT time point in *ERBB2*, *EGFR* and *MYC*, and numerically higher at baseline in *CCND1* (Figure 7.14).

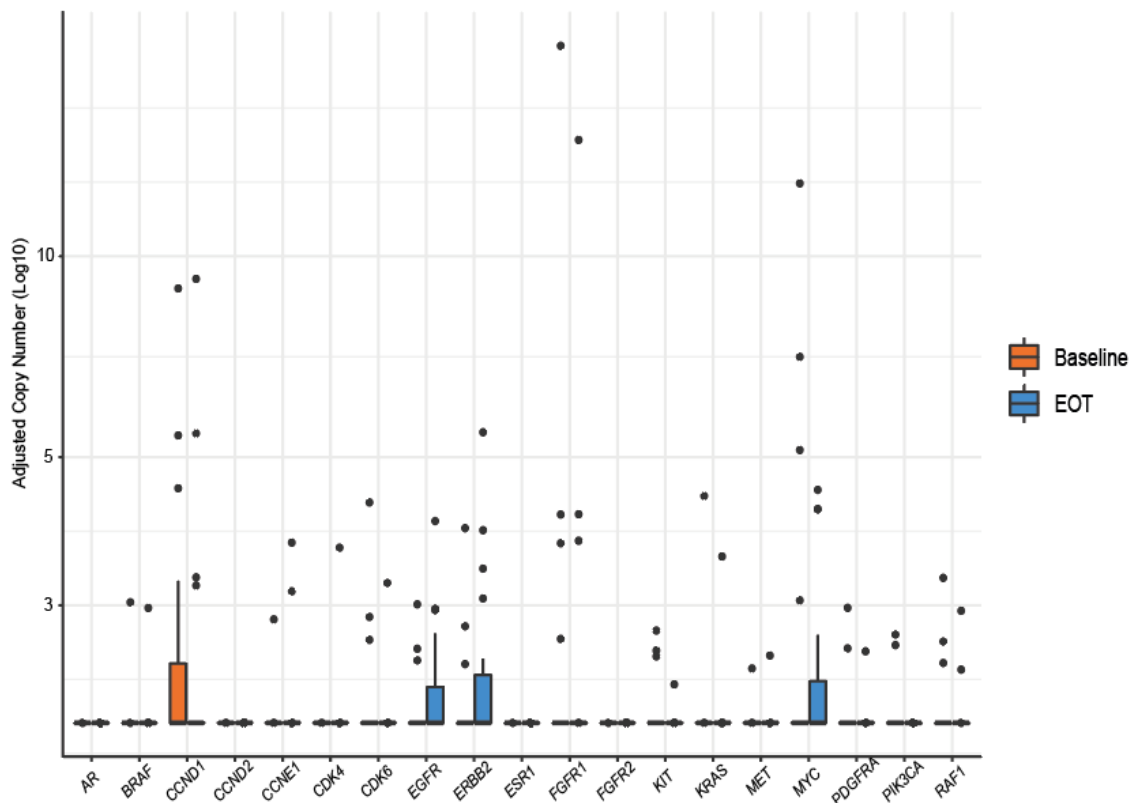


Figure 7.14. Baseline and EOT copy number within cohort B. Copy number for each patient was adjusted for purity of the sample (Chapter 2, section 2.1.12.9). Comparison with paired Wilcoxon signed rank test did not reveal and significant difference in baseline to EOT copy number values.

In total, 26.3% of cohort B (5/19 assessable patients) had copy number gains in *EGFR* and *ERBB2*, respectively, by EOT. Of the five patients with an increase in their *ERBB2* copy number from baseline to EOT, three (60%) were known to have HER2 amplified disease, whilst only one of these three patients had detectable ctDNA assessed *ERBB2* copy number increase at baseline. Of the two patients with *ERBB2* copy number gain with HER2 non-amplified disease, both had detectable *ERBB2* amplification at baseline which increased by the EOT timepoint.

*CCND1* copy number gains were also common, with 15.3% (3/19) of the cohort experiencing a copy number gain from baseline to EOT (Figure 7.15). There was no significant difference in maxVAF from baseline to EOT in patients with copy number gains in *EGFR*, *ERBB2*, *CCND1* and *CCNE1* (Figure 7.15), but many patients did have a numerical increase in their maxVAF, which would facilitate identification of copy number change in ctDNA.

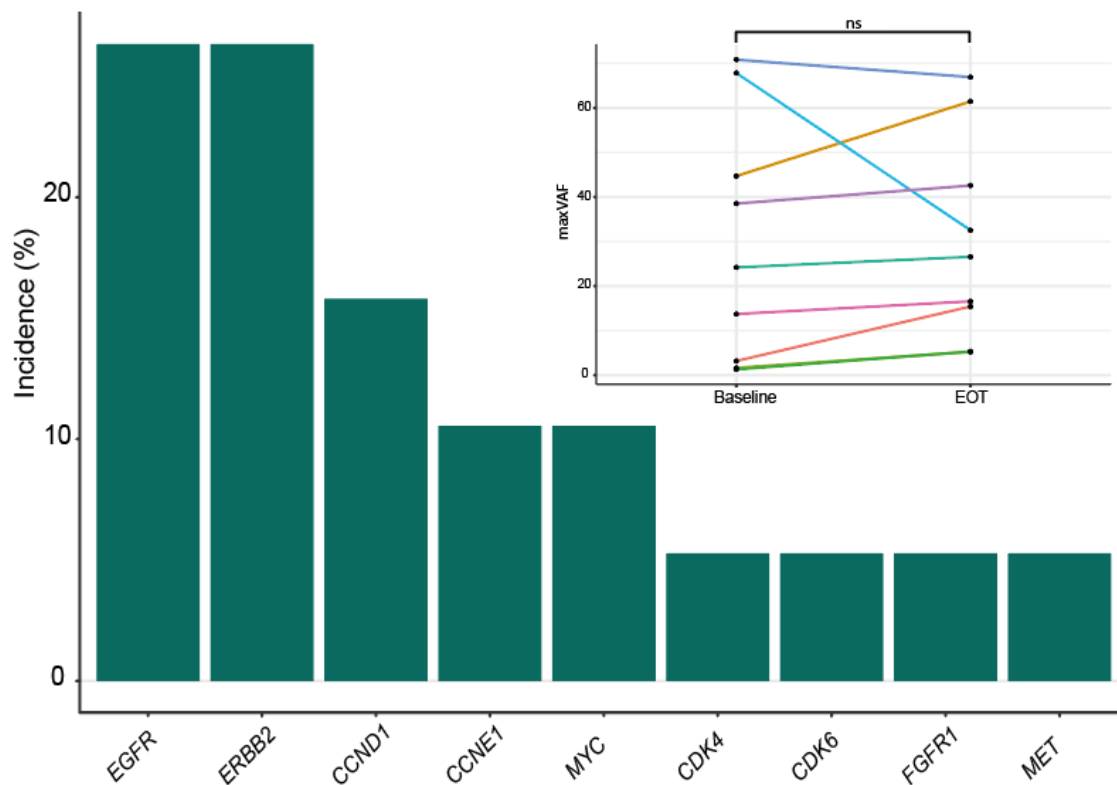


Figure 7.15. Incidence of copy number gain in each gene from baseline to EOT in cohort B. Included are patients where the copy number at EOT was at minimum  $>2.2$ . Copy number for each patient was adjusted for purity of the sample (Chapter 2, section 2.1.12.9). *Inset*, change in maxVAF from baseline to EOT in patients with an increase in copy number by the EOT time point in any of *EGFR*, *ERBB2*, *CCND1* and *CCNE1*,  $p=0.89$ . Comparison by paired t-test.

#### 7.4.4. Cohorts C and D: capivasertib +/- fulvestrant

Cohort C enrolled 18 patients (16 HR+ HER2-, 1 HR+ HER2+, 1 HR+ HER2 unknown) based on the presence of a plasma *AKT1* mutation (p.E17K in 17 and p.L52R in one), all of whom started treatment. Patients were treated with capivasertib (400mg PO BD 4 days on, 3 days off within 28 day cycles) with the addition of fulvestrant (500mg IM on Cycle 1 Days 1 and 15 and Cycle 2 onwards Day 1).

Cohort D enrolled 19 patients, 13 HR+ HER2- and 6 TNBC patients with activating *AKT* mutations (identified in plasma or tissue), all of whom started treatment. Overall, 6 patients were enrolled into cohort D based on an *AKT1* mutation (5 p.E17K and 1 p.L52R) and 13 based on a *PTEN* alteration. Enrolled patients were treated with capivasertib 480mg PO BD 4 days on/3 days off within a 28 day cycle.

All patients in cohort C had baseline sequencing, and 18 of 19 patients had baseline sequencing in cohort D. Seventeen patients in each cohort had available baseline and EOT sequencing.

While the overall response rate for cohort C was 22.2% (4/18), the overall response rate for cohort D was just 10.5% (2/19). However, the rate of response in the patients with an *AKT1* mutation in cohort D was higher, at 33.3% (2/6). For this reason, the following combined analysis of cohorts C and D was grouped according to whether the patients were enrolled based on an *AKT1* or *PTEN* alteration, and their breast cancer phenotype.

#### 7.4.4.1. Mutation acquisition and loss

The majority of HR+ *AKT1*-directed patients (11/18 assessable patients, 61.1%) maintained their monoclonal *AKT1* mutation through treatment and three patients (16.7%) lost or had undetectable *AKT1* mutation (Figure 7.16). HR- *AKT1*-directed patients had similar stability in the *AKT1* mutation status, four of five maintaining a monoclonal *AKT1* mutation (80%) and one patient (20%) losing *AKT1*-mutant status (Figure 7.16). No patients gained polyclonal *AKT1* disease. Co-mutation with *TP53* was common in the *AKT1*-directed group, with 66.7% (12/18) of HR+ patients and 80% (4/5) HR- patients having a baseline *TP53* mutation. By the EOT time point, 72.2% (13/18) of HR+ patients and 80% (4/5) HR- patients had a *TP53* mutation.

Acquisition of a monoclonal *ATM* mutation was relatively common in HR+ *AKT1*-directed patients, with three patients (3/18, 16.7%) patients gaining a monoclonal alteration and one patient (1/18, 5.6%) losing their *ATM*-mutant status (Figure 7.16).

*TP53*-mutant disease was also common in patients enrolled for *PTEN*-directed therapy at baseline (7/10, 70.0% HR+ and 1/1, 100% HR-) (Figure 7.16). No patients acquired *TP53* alterations, while two patients lost *TP53* mutation status (one HR+ and one HR-). The only acquired alterations in the *PTEN*-directed group (all in HR+ disease) were in *ARID1A* (1/10, 10%), *KRAS* (1/10, 10%) and *ESR1* (1/10, 10%).

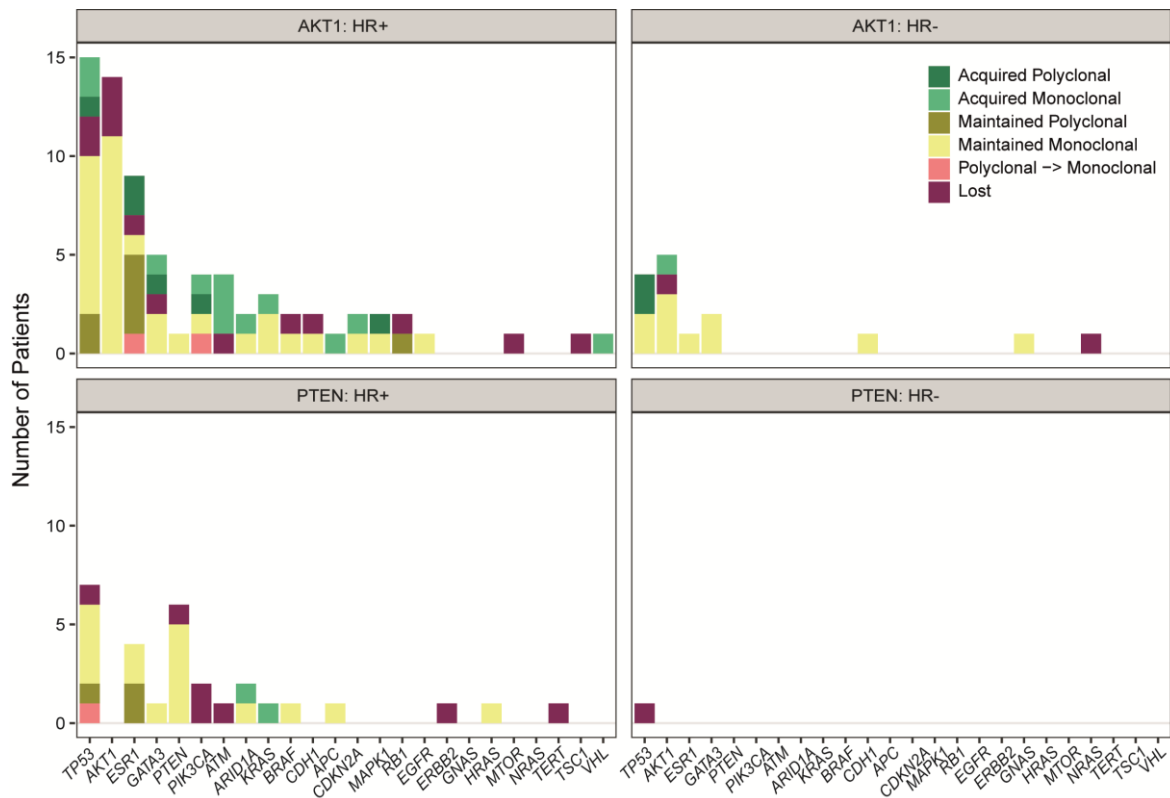


Figure 7.16. Baseline to EOT changes in gene mutation polyclonality, per gene, in AKT1-directed and PTEN-directed patients, divided by targeted gene and breast cancer phenotype. AKT1 HR+ n=18, AKT1 HR- n=5, PTEN HR+ n=10, PTEN HR- n=1.

Mutations were categorised as being involved in a range of signalling pathways (Figure 7.17A). In the HR+ AKT-directed group, alterations involved in the PI3K/AKT signalling pathway and cell cycle were the most commonly acquired both with 10 and 12 acquired mutations, respectively (Figure 7.17A). Acquisitions were also common in the MAPK signalling pathway (n=9) and oestrogen signalling pathway (n=11). Many of these were driven by acquisition of *TP53* alterations. This analysis should be interpreted with caution however as *TP53* is not an integral part of the MAPK and PI3K pathways. As a transcription regulator, *TP53* could be considered a peripheral element of many pathways. Notably, all patients in cohort C received fulvestrant in addition to capivasertib, and this may have driven the acquisition of *ESR1* alterations, all of which occurred in patients with HR+ disease. Acquisition of mutations in the RAS and ERBB signalling pathways was also frequent, with six acquisitions respectively attributable to these pathways (Figure 7.17A).

Acquisition of alterations in patients enrolled for PTEN-directed therapy was infrequent, likely due to the poor response to the therapy in this cohort and lack of time for the evolution of

genomic changes. Nevertheless, the acquisitions that did occur grouped mostly into the oestrogen signalling pathway in patients with HR+ disease (Figure 7.17A).

Acquisition of targetable alterations was infrequent in all groups (Figure 7.17B). The most clinically relevant acquisitions were in HR+ *AKT1*-directed group, with two patients (11.1%) acquiring targetable *PIK3CA* alterations (Figure 7.16 and 7.17B). Acquisition of potentially targetable alterations in *ESR1* occurred in three patients (16.7%). However, as these alterations were acquired in HR+ patients receiving fulvestrant, it is unlikely that these alterations are clinically targetable. Two patients (11.1%), both with HR+ disease on AKT1-directed therapy, gained *GATA3* mutations (Figure 7.17A and B). All three *GATA3* mutations were stop-gain mutations. Overall, there was no significant difference in the number of acquired alterations in the patients who gained clinical benefit from capivasertib +/- fulvestrant versus those that did not (Figure 7.17C and D).

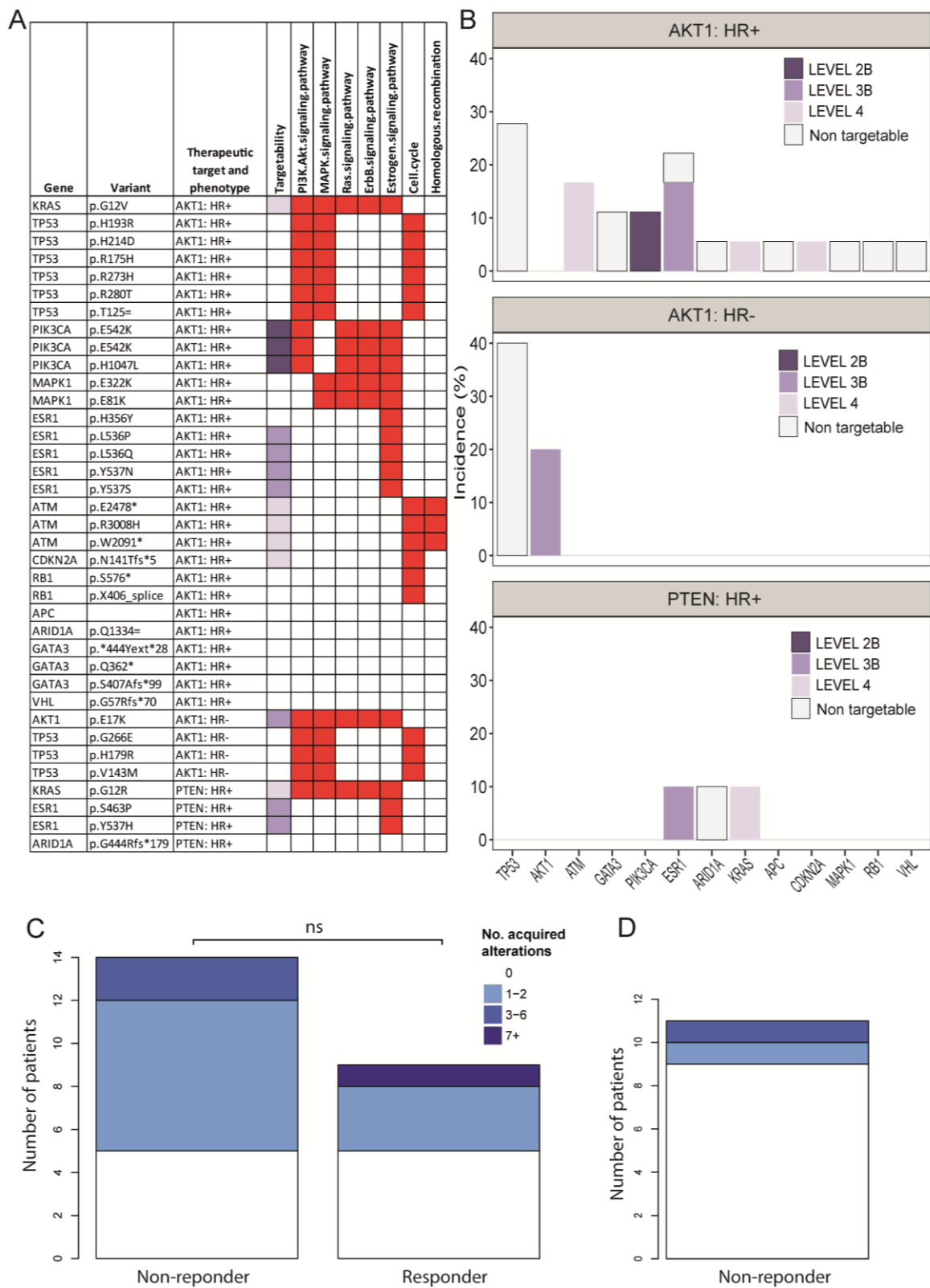


Figure 7.17. Acquired alterations in cohorts C and D. A, List of acquired pathogenic alterations in cohorts C and D, annotated with breast cancer phenotype and level of evidence supporting targetability (key as per legend within 7.17B), and pathway association in red. The gene alteration that the patient entered based on and patient phenotype is annotated as ‘Therapeutic target and phenotype’. B, incidence of acquired alterations divided by targeted gene alteration and breast cancer phenotype, coloured by level of evidence supporting targetability (Chapter 2, section



2.1.12.6). *Top*, patients enrolled based on an activating *AKT1* mutation with HR+ disease (n=18 assessable patients). *Middle*, patients enrolled based on an activating *AKT1* mutation with HR- disease (n=5 assessable patients). *Bottom*, patients enrolled based on a deleterious *PTEN* alteration with HR+ disease (n=10 assessable patients). No patients enrolled with a *PTEN* deleterious mutation and HR- disease acquired mutations. C, comparison of number of acquired alterations between patients who gained clinical benefit and those that did not in cohorts C and D who entered based on an activating *AKT1* mutation (n=23 assessable patients). Comparison using Chi-squared test. D, number of acquired alterations in the patients enrolled based on a deleterious *PTEN* alteration, none of whom gained clinical benefit (n=11 assessable patients).

## 7.5. Discussion

In this chapter the baseline and EOT plasma sequencing has been analysed to investigate for the acquisition of genomic changes that may represent resistance mechanisms. A novel putative resistance mutation *ESR1* p.F404 has been identified, while a number of other genomic acquisitions were identified which merit further investigation as potential mechanisms of resistance.

### 7.5.1. Putative resistance mechanisms to fulvestrant

Investigation of baseline to EOT ctDNA changes within cohort A revealed a number of interesting findings. The main finding was the identification of frequent acquisition of novel mutations at the *ESR1* p.F404 locus. This mutation was further investigated by transient transfection into MCF-7, with results suggesting that cells harbouring both an activating *ESR1* mutation alongside the p.F404L variant demonstrate resistance to fulvestrant (Figure 7.9). Importantly, the results suggest that presence of the p.F404L mutation alone is not sufficient to induce resistance, and indeed cells transfected with p.F404 alone appear to have lower activity in oestrogen than those transfected with wild type *ESR1* or mutant *ESR1* (Figures 7.7 and 7.9). It is hypothesised that the alterations at the p.F404 locus remove a pi-stacking bond between oestrogen and the *ESR1* receptor. In this scenario, in isolation, this mutation may reduce binding affinity of oestrogen and inhibit the receptor activity in oestrogen containing media. Conversely, in the presence of an activating mutation, the *ESR1* receptor exhibits ligand-independence, being constitutively active without the need for oestrogen binding. Thus, p.F404 mutations may only arise and propagate on the background of an activating mutation, where it confers a survival advantage to the cell in reducing the binding efficiency of fulvestrant whilst maintaining the ligand-independent activity courtesy of the activating mutation.

To date, p.F404 mutations have not been identified as a resistance mutation to fulvestrant. There are three reasons that this mutation may have been identified in the plasmaMATCH dataset. Firstly, patients were recruited to the cohort based on the presence of an activating *ESR1* mutation, and therefore, relative to other studies of fulvestrant, our cohort was enriched for patients with this genomic background. Secondly, plasmaMATCH was unique in using a higher dose of fulvestrant alongside a more intensive regimen, which may have added extra therapeutic pressure encouraging the development of this mutation. Finally, the sequencing approach used here may have helped identify this mutation. Specifically, this mutation arises subclonally, and the plasma-based sequencing may be better placed in this situation to identify subclonal variants than tissue-based sequencing, which would be limited to the tumour area sampled. The subclonal nature of the resistance mutation also means that it occurs at low allele frequency, and the targeted nature of the sequencing approach affords greater sensitivity to identify low allele frequency mutations than a broader sequencing approach such as exome or whole-genome sequencing. Additionally, the mutation arose in the same gene that was being targeted by the therapy, and was therefore included in the targeted sequencing panel. Had the mutation occurred a gene outside of *ESR1*, and not covered by the targeted panel, only a broader sequencing approach would have been able to identify it.

It is important to note that the patients who developed the mutation had the common feature of all harbouring a co-mutation with *ESR1* p.E380Q. Relative to other activating *ESR1* alterations, p.E380Q is not as strongly activating, and exhibits greater sensitivity to fulvestrant<sup>111,126</sup>. It is possible therefore that this mutation arose in these patients due to the prolonged progression free survival time on fulvestrant, allowing time for the genomic resistance mutation to occur and, through clonal selection, propagate<sup>47</sup>. Notably, all three patients who developed the mutation had a PFS >4 months, compared to a median PFS in the cohort of 2.2 months.

As patients increasingly have therapeutic choices driven by genomic characterisation, the number of ABC patients who receive fulvestrant based on the identification of activating *ESR1* mutations may increase. Therefore this putative resistance mechanism may be of increasing clinical relevance with the widespread use of fulvestrant in this patient group. The data presented in Figure 7.11 demonstrates that, for these patients, treatment with the novel SERDs AZD9833 and elacestrant, or indeed tamoxifen may circumvent this putative resistance mechanism and translate into longer progression free survival for patients.

The reduced ERE-luciferase activity of MCF-7 transfected with p.F404L in oestrogen is an interesting finding, and suggests that the mutation also disrupts the binding of oestrogen itself to the receptor. If true, there would only be a selective advantage to the cell of developing the mutation under two conditions. Firstly, there would also need to be an activating mutation present which negates the need for oestrogen binding. Secondly, there is the additional therapeutic pressure exerted by fulvestrant favouring the survival of cells where the binding of this drug is inhibited and, as a result, the receptor does not undergo fulvestrant-induced degradation. Of note this data was produced from one replicate and requires repeating to confirm.

The lack of resistance of the compound mutations to other SERDs and tamoxifen supports a mechanism of resistance operating through the distinct binding pattern of fulvestrant. However, structural modelling is required to investigate this further, which is currently underway. Additionally, this work will be complemented by investigation of the mutation using the orthogonal approach of CRISPR. The single and compound mutations will be stably expressed through CRISPR into MCF-7. These cells can then undergo cell viability experiments under conditions of oestrogen +/- fulvestrant or alternative drug to investigate for resistance. If the results concur with the transient transfection results, this would provide compelling evidence that the mutation represents a novel resistance mechanism to fulvestrant, with important clinical implications. This work is currently ongoing.

Outside of the *ESR1* p.F404 mutations, there were a number of other findings in cohort A. Firstly, we identified a high rate of acquisition of *ESR1* alterations (Figure 7.2A). This cannot be explained by an increased detection rate of existing *ESR1* subclones at the EOT time point secondary to higher disease burden/greater ctDNA purity as the patients who acquired *ESR1* alterations did not have significantly higher EOT mVAF (Figure 7.2B). Analysis of baseline and EOT sequencing within the PALOMA-3 trial (fulvestrant +/- palbociclib) also identified a higher incidence of *ESR1* mutations at EOT compared to baseline<sup>79</sup>. In particular, it was noted that p.Y537S appeared to be positively selected for through treatment<sup>79</sup>. Similarly to what was identified here (Chapter 6, Figure 6.8B), the group also found that there was a trend for patients with baseline p.Y537S to have shorter PFS than those wild type for the variant<sup>79</sup>, which is consistent with its known strong constitutive activity and relative resistance to anti-oestrogens<sup>104</sup> including fulvestrant<sup>110,248</sup>.

The most frequently acquired variant in our dataset was *ESR1* p.L536P (n=6 acquisitions). Interestingly, a trend toward higher incidence of this variant at EOT was also identified in patients treated with fulvestrant in the PALOMA-3 trial<sup>79</sup>, and in patients treated with fulvestrant + lerociclib in a phase I trial<sup>268</sup>. The p.L536 locus resides in the h11-h12 loop, near the amino terminus of h12 within *ESR1* alpha, and is a component of the ‘spring’ within the oestrogen receptor<sup>104</sup>. This spring consists of hydrophobic residues that keep the spring closed. Binding of an agonist receptor is required to overcome the hydrophobia and bend the spring into an agonist conformation. Mutation from a strongly hydrophobic residue to the less hydrophobic residue p.L536P lessens the strength of the spring and the ability of the *ESR1* receptor to remain in an ‘off’ conformation<sup>104</sup>. The sensitivity of this mutation to fulvestrant is not established, and it would be an interesting mutation to investigate for a potential role in fulvestrant resistance.

p.L536P and p.Y537S were just two of the frequently acquired *ESR1* mutations identified in this dataset, and the frequency of acquisition of other *ESR1* alterations is not explained. As demonstrated here and elsewhere<sup>111</sup>, it appears that having mono- versus polyclonal *ESR1* disease is not a predictor of response to fulvestrant therapy (Chapter 6, Figure 6.8C), and so the evidence does not support that the acquisition of these *ESR1* alterations is a resistance mechanism. One possibility is that, despite the increased dose of fulvestrant utilised, the poor oral bioavailability of fulvestrant did not maximally inhibit the subclones allowing further propagation of *ESR1* mutant subclones, supported by the poor response to fulvestrant overall in the cohort.

Outside of the oestrogen signalling pathway, within which mutations were frequently acquired in cohort A mostly due to the acquisition of *ESR1* alterations, the PI3K/AKT pathway also regularly acquired mutations (Figure 7.2A). *In vitro* data supports a role for activation of this pathway in fulvestrant resistance<sup>102,269-271</sup>. Furthermore, the PALOMA-3 trial demonstrated enrichment for *PIK3CA* mutations at the EOT time point in patients treated with fulvestrant +/- palbociclib<sup>79</sup>. While just two *PIK3CA* alterations were acquired in this cohort (Figure 7.2A), a large number of mutations implicated in the PI3K/AKT pathway were acquired, supporting a role for activation of this pathway in fulvestrant resistance.

#### 7.5.2. Putative resistance mechanisms to neratinib

Analysis of baseline and EOT sequencing within cohort B revealed a number of findings. One striking finding was the frequent acquisition of alterations within the PI3K/AKT pathway

(Figure 7.13A). In particular, two patients gained pathogenic *PIK3CA* mutations, with one of these gaining three additional alterations. It is understood that mutations in *ERBB2* constitutively activate the tyrosine kinase receptor, leading to enhanced signalling via PI3K and MAPK<sup>142</sup> (Chapter 1, Figure 1.4). It would seem plausible, therefore, that activating mutations in the PI3K/AKT pathway circumvent the inhibition of mutant *ERBB2* afforded by neratinib to generate a neratinib resistant phenotype. Prior *in vitro* work has demonstrated that *ERBB2*-mutant MCF-7 cells have enhanced signalling via the MAPK pathway, and that co-mutation with an activating *PIK3CA* mutation enhances this signalling<sup>144</sup>. Subsequent *in vitro* work has implicated activation of the TORC1 pathway in neratinib resistance<sup>153</sup>. The TORC1 pathway is downstream of the PI3K/AKT pathway, and activation via the PI3K/AKT pathway is expected to enhance signalling via the TORC1 pathway. The group performed knockdown of RAS, which modulates the activity of both the PI3K/AKT and MAPK pathways, finding that this reversed the neratinib resistant phenotype<sup>153</sup>. This is supported by the finding that patients with activating mutations within the mTOR pathway did not derive clinical benefit from neratinib<sup>153</sup>. Conversely, though, there was no apparent trend for acquisition of alterations activating the PI3K/AKT pathway in 16 patients analysed in the SUMMIT trial, which investigated the use of neratinib in patients with *ERBB2*-mutant cancer<sup>77</sup>. Additionally, in patients with early breast cancer treated with adjuvant neratinib in the ExteNET trial, patients with *PIK3CA*-mutated disease had a longer median absolute disease-free survival than those with *PIK3CA*-wild type disease, albeit the difference was not significant<sup>272</sup>. This does not support the hypothesis that *PIK3CA* alterations represent a mechanism of resistance.

A second finding was the frequent acquisition of additional *ERBB2* alterations within the cohort, with three patients (15.8%) acquiring *ERBB2* mutations and five patients (26.3%) having an increased purity-adjusted copy number at EOT compared to cohort (Figures 7.13 and 7.15). The three patients who acquired *ERBB2* mutations all had numerically higher maxVAF at EOT (albeit not significantly, Figure 7.13B), raising the possibility that clones carrying these mutations were pre-existing at baseline but became detectable with raised purity of the sample. If correct and not an artefact of increased purity at EOT, 42.1% of the cohort acquired an *ERBB2* alteration (mutation or copy number increase) that was additional to their baseline *ERBB2* mutation that they were recruited into the trial with. The SUMMIT trial similarly identified that 36% of patient's demonstrated acquisition of an *ERBB2* alterations<sup>77</sup>. The *ERBB2* mutations acquired in the plasmaMATCH cohort are hotspot mutations in the gene, but are not known gatekeeper resistance mutations as previously identified for neratinib

(p.T798I<sup>151</sup>). The plasmaMATCH cohort was also found to have a high rate of acquisition of raised *EGFR* copy number (26.3% patients). Mechanistically, ligand binding the EGFR receptor leads to dimerisation with ERBB2 (amongst other ERBB receptors) to activate downstream signalling, including via the PI3K pathway<sup>273</sup> (Chapter 1, Figure 1.4). Further research is required to understand the clinical relevance of second hit *ERBB2* mutations and acquired *ERBB2* and *EGFR* amplifications following neratinib treatment.

### 7.5.3. Putative resistance mechanisms to capivasertib

#### 7.5.3.1. *AKT1*-directed patients

Patients with an HR+ disease enrolled with an *AKT1* mutation treated with capivasertib had a notable rate of acquisition of *PIK3CA* alterations (n=2 patients, 11.1%, Figure 7.17). *AKT1* mutations and *PIK3CA* mutations are normally known for being mutually exclusive<sup>34,219,274</sup>, presumably because alterations in these two genes activate the same pathway (with *PIK3CA* situated upstream from *AKT1*) so a second mutation in this pathway would be expected to be redundant. However, in the situation of *AKT1* inhibition, it is mechanistically plausible that a *PIK3CA* mutation may enable cells to overcome the *AKT1* inhibition by activation of signalling elsewhere in the pathway. The FAKTION trial, which enrolled HR+ patients with ABC demonstrated that patients with *PIK3CA* mutations or *PTEN* null tumours did not benefit from the combination of capivasertib and fulvestrant, whereas those that did not have one of these alterations did significantly benefit<sup>251</sup>. In contrast, patients with a coincidence PI3K pathway alteration in one phase II study of single agent capivasertib had a significantly longer PFS than patients without<sup>166</sup>. Further research is required to understand the clinical relevance of *PIK3CA* mutation acquisition in capivasertib resistance.

Mutation acquisition was relatively frequent in the MAPK and RAS signalling pathways (Figure 7.17A). Prior *in vitro* data has suggested that signalling via the PI3K pathways and MAPK pathways occurs in parallel, with inhibition of one causing upregulation in the other<sup>275</sup>. As capivasertib specifically inhibits *AKT1* in the downstream segment of the pathway, theoretically activating mutations of the MAPK/RAS pathways could represent a mechanism of resistance to the therapy, with signalling upregulated in the complementary pathway. The data presented in Chapter 6 (Figure 6.17A and B) demonstrates that while there was no significant difference in the incidence of baseline MAPK alterations in those that benefitted from capivasertib versus those that didn't, there was a trend for higher incidence in those with

no benefit. This theory requires further investigation in HR+ breast cancer cell lines, and corroboration in clinical data derived from a larger set of patients.

In HR+ patients enrolled for AKT1-directed treatment, alongside frequent acquisition of *TP53* (27.8%), *ESR1* (16.7%) and *PIK3CA* (11.1%), *ATM* mutations were frequently acquired (n=3 patients, 16.7%). *ATM* is known as a tumour suppressor gene, with a major role in the repair of double strand breaks<sup>276</sup>. *ATM* also has an array of non-canonical modes of action, including the activation of AKT/PKB activity by stimulating the phosphorylation at Ser473 in AKT following insulin treatment<sup>277,278</sup>. Treatment of cells with two ATP-competitive inhibitors of AKT, which are similar to capivasertib, led to enrichment of *ATM* kinase compared to the change seen when the cells were treated with allosteric AKT inhibitors<sup>279</sup>. Further *in vitro* data has demonstrated that treatment of cancer cells known to have hyperactive AKT signalling with the *ATM*-inhibitor KU-55933 inhibited the phosphorylation of AKT, which in turn promoted cell cycle arrest in the G<sub>1</sub> phase and apoptosis<sup>278</sup>. This demonstrates the relationship between *ATM* and AKT in promoting cell survival. As a tumour suppressor, mutations in *ATM* generally promote loss of function of the protein, and it is notable that two of the three *ATM* mutation acquisitions in this cohort are truncating mutations which if expressed would be expected to inhibit protein function, thereby reducing activation of AKT and its downstream effectors.

#### 7.5.3.2. *PTEN*-directed patients

Amongst eleven assessable patients who enrolled based on a *PTEN* aberration who were treated with capivasertib, there were only four mutation acquisitions (in *KRAS*, *ESR1* and *ARID1A*). The poor response rate coupled with small numbers precludes analysis into any potential resistance mechanisms within this subgroup. Further research is required to understand if patients with *PTEN*-loss do benefit from AKT1-directed therapy.

## 7.6. Conclusion

In this Chapter ctDNA analysis has been used to investigate for putative mechanisms of resistance to a number of targeted therapies. The most significant finding was that of *ESR1* p.F404 as a putative mechanism of resistance. This was further explored with transient transfection of MCF-7. This mutation was identified at a low allele frequency and was subclonal, which arguably would have made it easier to identify with a plasma targeted

sequencing approach over other broader sequencing approaches or from tissue biopsy sequencing. Further work is underway to investigate the relevance of this mutation. However, the data presented here suggests that this is a potential mechanism of resistance to fulvestrant that can be overcome with alternative SERDs or tamoxifen therapy. Identification of this mutation in patients would suggest that an alternative anti-oestrogen strategy, such as tamoxifen or a novel SERD, are indicated, with the aim of extending the progression free survival of patients on treatment.

Other interesting findings were also identified, which warrant further investigation. Of these, the acquisition of *ERBB2* and *EGFR* amplification in patients treated with neratinib was particularly notable. As identified and discussed in Chapter 4, copy number identification in plasma is specific but not sensitive. It is likely therefore that the patients that had identifiable copy number changes had true copy number alteration, and that further tissue-based assessment may identify this changes in more patients. The clinical relevance of the copy number gains requires further investigation, but mechanistically these copy number gains could represent resistance mechanisms.

A limitation of this work was the lack of response in cohort A and the PTEN-directed subgroup in cohort D, alongside the small numbers in cohorts B, C and D. This will limit the ability to identify new mutations or mechanisms of resistance and establish any significant changes. In particular, genomic mechanisms of resistance would be expected to take time to develop, and where patients have not responded to therapy these mechanisms may not have had time to arise and propagate. A further limitation is the use of a targeted panel for novel resistance mutation discovery. This approach will have been limited to identification of alterations within genes included in the panel. Conversely, the benefit of this approach is the ability to identify mutations at a low AF, which may be the case in resistance mutations that arise subclonally. A broader sequencing approach, such as exome sequencing, would enable a broader identification of genomic resistance mechanisms. Despite these limitations, ctDNA analysis using data from targeted sequencing has revealed a number of interesting findings, supporting the use of ctDNA analysis for this application.



## 8. Chapter 8. Conclusion and Future Directions

At the time of writing, ctDNA analysis is not funded by the NHS for use within ABC, and, consequently, it has not entered routine clinical practice. This is despite there being several potential clinical applications, a few of which have been explored and investigated within this thesis. There are a multitude of barriers preventing the routine clinical use of ctDNA analysis which must be overcome before adoption of the approach.

In 2018 The American Society of Clinical Oncology and College of American Pathologists published a joint review on the use of ctDNA analysis in patients with cancer<sup>95</sup>. The review proposes three categories with which to assess the utility of ctDNA analysis. The first is analytical validity, which describes the ability of a test to accurately and reliably detect variants of interest. The second term is clinical validity, which describes the ability of the test to differentiate patient groups according to their test results. The third term is clinical utility, which describes the presence of high levels of evidence demonstrating that the test improves patient outcomes. This thesis investigates a number of features within these parameters.

Firstly, Chapter 3 investigated the analytical and clinical validity of ctDNA analysis. Here, two orthogonal techniques of ctDNA analysis were compared, finding high levels of agreement particularly for binary gene-mutation status. This was supported by a high level of agreement between tissue and plasma for targetable alterations, particularly for clonally dominant/driving alterations such as those in *PIK3CA*. Although the data presented here are supportive of the analytical and clinical validity of ctDNA analysis, the analysis presented here may not be generalisable. The clinical validity of ctDNA analysis is dependent on a multitude of pre-analytical variables. As discussed in the ASCO/JCP joint review, there is much variation and lack of consensus regarding important factors including how blood is drawn and into what collection tube, the isolation of DNA from serum or plasma, sample transport, DNA purification, and bioinformatic analysis, all of which will have a significant impact upon the results of the test<sup>95</sup>. This highlights the question of whether the data presented here, which were gathered in the relatively tightly managed environment of a clinical trial, can be extrapolated into clinical practice and real-world scenarios.

To provide robust data supporting analytical and clinical validity and enable new technologies to move from research use to clinical practice, supportive data must be derived from several different sources arising within varied contexts. In the case of ctDNA testing, cross-platform

validatory testing has not provided strong evidence of analytical validity. In one comparative test of four commercially available next generation sequencing (NGS) platforms, 41 of 56 unique variants identified (68%) were discordant<sup>97</sup>. Sources of error included low allele frequency calls (<1%), mutational biases, novel somatic variants, background noise, bioinformatic filtering thresholds and germline variant calls<sup>97</sup>. In a further comparative study of two NGS platforms, including Guardant Health's Guardant360, only 33% patients (12/36) had totally concordant NGS results, while 17% patients were partially concordant<sup>96</sup>. A further consideration is that the cross-platform approach to validation is also liable to systematic errors producing erroneous results in all the tested assays which will falsely inflate performance metrics. This theoretically is less of an issue when orthogonal approaches have been used, such as presented here in Chapter 3 with targeted sequencing compared to ddPCR.

Most recently, the SEQC2 Oncopanel Sequencing Working Group published a rigorous assessment of the analytical performance of five commercially available ctDNA sequencing assays<sup>280</sup>. Through parallel analysis of the performance of multiple assays undertaken in a number of different laboratories with multiple replicates and multiple experimental conditions including DNA input, mutation allele frequency and DNA isolation method, the group were able to thoroughly interrogate the performance of NGS assays to account for pre-analytical and analytical factors. Reassuringly, at allele frequencies >0.5%, ctDNA mutations were detected with high sensitivity, accuracy and reproducibility by all the assays tested. Below this limit, sensitivity fell, and was highly dependent on the locus specific coverage and input DNA amount<sup>280</sup>. Reassuringly, false positive rates were low across all tested assays and pre-analytical variables such as variation in laboratory plasma extraction procedures and workflow were not found to be a significant cause of discordant results<sup>280</sup>.

Whilst the SEQC2 Oncopanel Sequencing Working Group analysis provides reassurance regarding the analytical validity of ctDNA analysis, it does highlight the poor performance of assays at allele frequencies <0.5%. Depending on the intended application of ctDNA analysis, this has varying implications. For clonal architecture assessment, such as undertaken in Chapter 5, there will not be sufficient resolution to define clonal architecture for low purity or significantly subclonal alterations. For novel resistance mutation discovery, as undertaken in Chapter 7, where mutations tend to arise subclonally, there may not be sufficient sensitivity to detect new mutations. It is notable that of the seven *ESR1* p.F404 mutations identified at the EOT time point, five were identified at an allele frequency <0.5%. Thus, it is possible that more patients in the cohort developed this particular alteration, and other resistance mutations that

were undetectable by NGS, highlighting the limitation of sequencing when it comes to resistance mutation discovery.

Future research needs to focus on increasing the sensitivity of NGS below 0.5%, which would vastly increase the applications of ctDNA analysis, both within metastatic disease, and more broadly to earlier disease settings or indeed to cancer screening, where allele frequencies are expected to be lower. Current NGS methods are fundamentally limited by the issue of purity, or amount of ctDNA within the cell-free DNA (cfDNA), and the technical approach of NGS that is based upon the random sampling of amplified fragments, leaving rare mutations prone to being missed through the effect of stochastic sampling. Potential methods include selective enrichment of ctDNA over cfDNA, which may be achieved by size selection<sup>281,282</sup> methylation status<sup>283</sup> or other biological features<sup>284</sup>.

The data presented in Chapter 4 highlights that certain patient groups are less well suited to ctDNA testing. This is likely due to these patients having less advanced/a smaller burden of disease or disease that ‘sheds’ less ctDNA due to biological processes not fully characterised and/or fully understood yet. Again, this will mean that ctDNA purity is the primary issue, and research to improve detection at low allele frequency may resolve this. However, there may still be tumours that do not shed ctDNA, such as those confined to the central nervous system, and for these patients analysis of ctDNA derived from the systemic circulation may not be the best approach that can be adopted. Enhanced understanding of the likelihood of a patient having detectable ctDNA would enable physicians to make more informed choices based on the results of the ctDNA test, and highlight the patients who are more at risk of a false negative result. One approach could be to routinely estimate and state the purity of the sample, perhaps through whole genome sequencing (Chapter 6, section 6.4.1), and to provide a percent confidence in the clinically important negative mutation calls.

The data in Chapter 4 also highlights that ctDNA testing is not sensitive in identifying *HER2* amplification status in breast cancer. It is understood that ctDNA testing is not yet sensitive enough at identifying copy number alterations and other important cancer-associated DNA changes such as translocations and structural variations<sup>280</sup>. This may be more pertinent in phenotype subsets in which these alterations are more frequent, such as TNBC, where copy number changes are common. Further research is needed to improve and validate the identification of these features in ctDNA.

As discussed in the ASCO/JCO joint review, once the analytical validity of ctDNA testing is addressed, the clinical validity and utility must be established. The clinical validity of ctDNA testing with reference to a number of targeted therapies in ABC was explored in Chapter 6, with the retrospective analysis of baseline prognostic and predictive factors and their association with response. Whilst a number of interesting findings were identified for each respective targeted therapy, with some warranting further investigation, this analysis also highlighted some important barriers to establishing the clinical validity and utility of ctDNA testing in selecting patients for targeted therapy.

Firstly, in a somewhat circular argument, the targeted therapy being retrospectively assessed must have a reasonable response rate in the cohort tested. Furthermore, there must be enough patients to provide the power to detect differences between certain patient groups based on ctDNA characteristics. Between these two elements, the analysis here was limited in its ability to detect biomarkers given the low response rate in cohort A and the small sample size in cohorts B-D. These challenges are not unique to plasmaMATCH, and data supporting the use of ctDNA to identify biomarkers of response to subsequent therapy in ABC is lacking. The most robust data to date comes from the SOLAR-1 trial, where a retrospective analysis established that patients who were ctDNA positive for *PIK3CA* mutations gained benefit from alpelisib whilst those that were negative did not<sup>66</sup>. Overall, for retrospective analyses such as the one undertaken in Chapter 6, targeted therapies with a good response rate, in a cohort large enough to detect differences, must be analysed.

A further consideration in the adoption of ctDNA testing to identify biomarkers for subsequent treatment is that there must also be a good understanding of both the pathogenicity and ‘targetability’ of alterations. Currently there is little consensus surrounding the assignment of pathogenicity and targetability, and for this thesis an approach was taken to combine the annotation provided by the publicly available annotation tools OncoKB<sup>189</sup> with Cancer Hotspots<sup>190</sup>, cross referenced against known recurrent mutations in the COSMIC online database<sup>285</sup>. This is a constantly developing field, however, with new data changing our understanding of which patients may benefit from a given targeted therapy. This poses challenges for the development of trials to investigate the efficacy of targeted therapies in patient groups defined by ctDNA results, which maybe be specifically targeted to narrow cohorts of patients with rare alterations, such as the case in plasmaMATCH<sup>98</sup>. Thus the clinical validity and utility of ctDNA testing as a biomarker of response to subsequent therapy is not yet established, and further research is required in this area.

Finally, to establish the clinical utility of ctDNA testing for guiding therapeutic choices, data from large clinical trials are needed. plasmaMATCH was a phase II trial and not randomised<sup>98</sup>, so is limited in its ability to contribute to demonstrating the clinical utility of ctDNA testing. There are an increasing number of retrospective cohort studies (often single site) describing the use of real-time ctDNA testing to direct patient care outside of the clinical trial settings, which demonstrate the feasibility of the approach<sup>286-288</sup>. However, phase III randomised trials with randomisation of patients to ctDNA driven choices versus standard care are lacking, and trial design and undertaking in the area of targeted therapy has proved challenging<sup>40,289</sup>. As a first step, such a trial would need to ethically satisfy the criteria that the clinical outcome on the targeted therapy could be expected to equal or exceed that of standard care. At the present time, there are few targeted therapies in ABC with validated biomarkers<sup>120,290</sup> that satisfy this criteria. With time, as new targeted therapies are developed and there is greater understanding of the biomarkers that support the use of these therapies, the design and undertaking of these trials will be able to validate the utility of ctDNA testing. Furthermore, if large-scale studies currently enrolling or directing patient care based on tumour sequencing results for matched targeted therapy such as NCI-MATCH<sup>291</sup>, AURORA<sup>292</sup>, IMPACT II<sup>289</sup> and TAPUR<sup>293</sup> are positive, the confidence in the approach of genomic personalisation of therapy will increase.

Outside of establishing the analytical and clinical validity and clinical utility, there are other important considerations around the application of ctDNA analysis highlighted by this thesis. Firstly, Chapter 5 explored the use of ctDNA analysis to dissect patterns of clonal dominance and subclonality. Notwithstanding the aforementioned issues with reduced sensitivity at low allele frequencies, the data presented here demonstrates compelling evidence that ctDNA can be used to dissect clonal architecture and sample subclonal disease at a cohort level. Confident differentiation of this at a patient level requires a broader sequencing approach with high depth, such as exome sequencing. The analysis presented here identified important subclonal features which have potentially large clinical repercussions. Firstly MAPK pathway mutations were found to frequently occur alongside *ESR1* mutations, and harbouring mutations in both predicted for shorter overall survival (Chapter 5, Figures 5.11 and 5.12). Secondly, dual *PIK3CA* mutations were frequent in HR+ disease and predict for shorter progression free survival on subsequent fulvestrant therapy (Chapter 5, section 5.4.4.3). Thirdly, in Chapter 7 a targeted sequencing approach was able to identify a putative resistance mechanism to fulvestrant, with the mutations arising being subclonal and below the limit of detection for many tumour sequencing panels. These observations support the use of ctDNA analysis over

the use of tumour biopsy analysis through facilitating the identification of clinically relevant mutations existing in parallel but, due to tumour heterogeneity and the metastatic-clade nature of subclonal resistance mutations, may not be identifiable by single-site tumour biopsy.

A second major observation in Chapter 5 is that targetable mutations tend to occur with distinct clonal dominance patterns (Chapter 5, Figure 5.7). The clinical relevance of this, however, is not clear. Within AKT1-directed therapy, Hyman *et al* demonstrated that response of a patients *AKT1* p.E17K mutant disease was proportional to allele frequency of the mutation within each metastatic deposit tested<sup>166</sup>. Within the SUMMIT trial of HER2 directed therapy with neratinib in *ERBB2*-mutant breast cancer, no responses were seen in patients with subclonal *ERBB2* alterations<sup>77</sup>. The association of clonal dominance and response within other targetable mutations has not yet been well defined, and again this is an area that warrants further research.

There is one potential application of ctDNA analysis in ABC that this thesis has not explored, namely the predictive and prognostic association of early ctDNA dynamic changes and response to targeted therapy. Changes in ctDNA early on treatment have been demonstrated to be predictive for response in several studies<sup>294-296</sup>. This information could be used to potentially stratify patients into expected response groups, with the possibility of escalating therapy in patients with poor predicted response whilst de-escalating therapy in those with good predicted response, in a personalised medicine approach. This approach has not yet been tested in clinical trials, and so the clinical utility has yet to be established.

There are several findings presented in this thesis that require further investigation. Firstly, the finding of MAPK pathway/*ESR1* co-mutation enrichment and its association with overall survival was novel and has potentially significant clinical implications. Interrogation of ctDNA sequencing-derived datasets from other patient cohorts would increase confidence in the clinical finding, whilst sequencing of sequential ctDNA samples alongside primary and metastatic tissues would shed light on the temporal and spatial pattern of mutation acquisition and allow analysis of RNA expression patterns of the oestrogen and MAPK pathways. *In vitro* experiments could include MAPK pathway knockdown in cells with activating *ESR1* mutations followed by clonogenic assays. Finally, patient derived models, if available, could be used to investigate whether dual inhibition with a combination of fulvestrant and a MAPK pathway inhibitor such as trametinib (MEK-inhibitor) can reverse the adverse phenotype associated with the compound mutations.

Secondly, the high rate of *ERBB2* mutation and copy number gain and *EGFR* copy number gain acquisition at the EOT of patients in cohort B on neratinib therapy is an interesting finding. Given the poor sensitivity of copy number identification in plasma, the finding of copy number changes needs validation in tissue sequencing datasets at baseline and on progression from neratinib therapy. In conjunction, *in vitro* experiments could examine whether copy number changes in *EGFR* and *ERBB2*, and additional acquired *ERBB2* mutations, represent resistance mechanisms. A clinical trial in patients with these additional alterations to tackle any putative resistance could examine the effect of dual HER-2 blockade, or enhanced EGFR blockade such as with the addition of cetuximab.

Finally, while the work presented in this thesis is supportive that *ESR1* p.F404 represents a novel resistance mechanism to fulvestrant, there are limitations of the data presented that require addressing, and further investigation is required. Specific areas that require further experiments including repeating the prolonged oestrogen deprivation transient transfection experiment (Chapter 7, Figure 7.7) to confidently define the behaviour of each ERC in the oestrogen deprived conditions, with a minimum of three replicates. The western blot (Chapter 7, Figure 7.8) also requires undertaking with a non-transfected ‘control’ condition to confirm that the exogenously transfected *ESR1* is expressed as expected. The presence of the mutation in the ctDNA of patients within whom it was identified should also be confirmed with an orthogonal ctDNA analysis approach, such as with ddPCR. To this end, ddPCR probes are currently being designed which will allow testing of the plasma for the *ESR1* p.F404 variant using a ddPCR approach. Finally, consideration could also be given to testing the transiently transfected mutation in other ER+ cell lines.

Outside of these further experiments, the findings would be significantly strengthened if they were confirmed through an orthogonal technique. As such, at the present time work is ongoing utilising CRISPR to stably introduce *ESR1* p.F404 +/- activating *ESR1* mutations into MCF-7 cells. This will allow for, at minimum, cell survival assays to establish whether these models also show resistance to fulvestrant and other therapies. Structural modelling may help define how the mutation interrupts the binding of oestrogen, fulvestrant and other therapies. This structural modelling, through defining how different therapies interact with the mutant ER, may highlight therapeutic approaches which would be expected to overcome the resistance that is seen with fulvestrant. Finally, interrogation of the baseline and EOT sequencing data from other clinical cohorts following treatment with fulvestrant would define whether this mutation

is found in cohorts outside of plasmaMATCH, and, as such, the clinical relevance of the finding to the broader population.

In summary, this thesis has explored the use of ctDNA analysis within ABC, making several clinically important observations in the course of the investigation. Following validation of ctDNA analysis (Chapter 3), several important findings were identified that have the potential to improve patient care. Firstly, the patient groups the technology was suited to best were defined, alongside the clinico-pathological characteristics of targetable alterations (Chapter 4). In the largest cohort studied to date, the clonal architecture and landscape of ABC was defined, with novel findings in the arena of the clonal architecture of resistance mutations that have direct clinical implications (Chapter 5). Prognostic and predictive biomarkers of selected targeted therapies were explored, with a number of findings that warrant further investigation (Chapter 6). Finally, ctDNA analysis was utilised to investigate resistance mechanisms, with the novel finding of a putative resistance mechanism, *ESRI* p.F404 mutations, further investigated and validated using *in vitro* techniques (Chapter 7). Importantly, the data presented here demonstrates that the resistance caused by this mutation may be overcome by alternative therapies, with direct implications for clinical practice and the potential to prolong patient's response to hormonal based therapies, delaying the commencement of chemotherapy.

While this thesis has demonstrated the wide utility of ctDNA analysis in ABC, the aforementioned broader limitations to the approach necessitate further research. This research needs to encompass both improving techniques around ctDNA analysis itself, with particular focus on improving sensitivity at low allele frequencies and identification of alterations outside of SNVs and indels, but also in related fields with the development and validation of new targeted therapies, identification of biomarkers of response, and greater understanding of the pathogenicity and targetability of mutations. With further development in these areas, ctDNA analysis shows great potential in changing the way ABC is managed in the clinic.



## References

1. Bray, F., *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**, 394-424 (2018).
2. Cancer Research UK. Breast Cancer Statistics. Vol. 2019.
3. Ravdin, P.M., *et al.* Prognostic significance of progesterone receptor levels in estrogen receptor-positive patients with metastatic breast cancer treated with tamoxifen: results of a prospective Southwest Oncology Group study. *Journal of Clinical Oncology* **10**, 1284-1291 (1992).
4. Kris, M.G., *et al.* Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *Jama* **290**, 2149-2158 (2003).
5. Fukuoka, M., *et al.* Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **21**, 2237-2246 (2003).
6. Lynch, T.J., *et al.* Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine* **350**, 2129-2139 (2004).
7. Giaccone, G., *et al.* Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 1. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **22**, 777-784 (2004).
8. Herbst, R.S., *et al.* Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 2. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **22**, 785-794 (2004).
9. Mok, T.S., *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361**, 947-957 (2009).
10. Tsimberidou, A.M., Fountzilias, E., Nikanjam, M. & Kurzrock, R. Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer treatment reviews* **86**, 102019 (2020).
11. Wan, J.C.M., *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* **17**, 223-238 (2017).
12. Leon, S.A., Shapiro, B., Sklaroff, D.M. & Yaros, M.J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer research* **37**, 646-650 (1977).
13. Breitbach, S., Sterzing, B., Magallanes, C., Tug, S. & Simon, P. Direct measurement of cell-free DNA from serially collected capillary plasma during incremental exercise. *Journal of applied physiology (Bethesda, Md. : 1985)* **117**, 119-130 (2014).
14. Rodrigues Filho, E.M., *et al.* Elevated cell-free plasma DNA level as an independent predictor of mortality in patients with severe traumatic brain injury. *J Neurotrauma* **31**, 1639-1646 (2014).
15. Tsai, N.W., *et al.* The value of serial plasma nuclear and mitochondrial DNA levels in patients with acute ischemic stroke. *Clinica chimica acta; international journal of clinical chemistry* **412**, 476-479 (2011).
16. Snyder, T.M., Khush, K.K., Valantine, H.A. & Quake, S.R. Universal noninvasive detection of solid organ transplant rejection. *Proceedings of the National Academy of Sciences* **108**, 6229-6234 (2011).

17. Raptis, L. & Menard, H.A. Quantitation and characterization of plasma DNA in normals and patients with systemic lupus erythematosus. *The Journal of clinical investigation* **66**, 1391-1399 (1980).
18. Leon, S.A., *et al.* DNA in synovial fluid and the circulation of patients with arthritis. *Arthritis and rheumatism* **24**, 1142-1150 (1981).
19. Hyett, J.A., *et al.* Reduction in diagnostic and therapeutic interventions by non-invasive determination of fetal sex in early pregnancy. *Prenatal Diagnosis* **25**, 1111-1116 (2005).
20. Lo, Y.M.D., *et al.* Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proceedings of the National Academy of Sciences* **104**, 13116-13121 (2007).
21. Hanahan, D. & Weinberg, Robert A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646-674 (2011).
22. Jahr, S., *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer research* **61**, 1659-1665 (2001).
23. Giacona, M.B., *et al.* Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas* **17**, 89-97 (1998).
24. Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57-68 (2016).
25. Stroun, M., *et al.* Neoplastic Characteristics of the DNA Found in the Plasma of Cancer Patients. *Oncology* **46**, 318-322 (1989).
26. Sorenson, G.D., *et al.* Soluble normal and mutated DNA sequences from single-copy genes in human blood. *Cancer Epidemiology Biomarkers & Prevention* **3**, 67-71 (1994).
27. Abbosh, C., *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446-451 (2017).
28. Bettegowda, C., *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224 (2014).
29. Garcia-Murillas, I., *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Science Translational Medicine* **7**, 302ra133-302ra133 (2015).
30. Tie, J., *et al.* Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* **8**, 346ra392 (2016).
31. Lehmann-Werman, R., *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A* **113**, E1826-1834 (2016).
32. Ignatiadis, M., Sledge, G.W. & Jeffrey, S.S. Liquid biopsy enters the clinic — implementation issues and future challenges. *Nature Reviews Clinical Oncology* **18**, 297-312 (2021).
33. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
34. Razavi, P., *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438.e426 (2018).
35. Bertucci, F., *et al.* Genomic characterization of metastatic breast cancers. *Nature* **569**, 560-564 (2019).
36. Lefebvre, C., *et al.* Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med* **13**, e1002201 (2016).

37. Zehir, A., *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nature medicine* **23**, 703-713 (2017).
38. Angus, L., *et al.* The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nature Genetics* **51**, 1450-1458 (2019).
39. Overman, M.J., *et al.* Use of research biopsies in clinical trials: are risks and benefits adequately discussed? *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**, 17-22 (2013).
40. André, F., *et al.* Comparative genomic hybridisation array and DNA sequencing to direct treatment of metastatic breast cancer: a multicentre, prospective trial (SAFIR01/UNICANCER). *The Lancet. Oncology* **15**, 267-274 (2014).
41. Yates, L.R., *et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **32**, 169-184.e167 (2017).
42. Brown, D., *et al.* Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nature Communications* **8**, 14944 (2017).
43. Savas, P., *et al.* The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program "CASCADE". *PLoS Med* **13**, e1002204 (2016).
44. Fumagalli, D., *et al.* Somatic mutation, copy number and transcriptomic profiles of primary and matched metastatic estrogen receptor-positive breast cancers. *Ann Oncol* **27**, 1860-1866 (2016).
45. Ng, C.K.Y., *et al.* Genetic Heterogeneity in Therapy-Naive Synchronous Primary Breast Cancers and Their Metastases. *Clin Cancer Res* **23**, 4402-4415 (2017).
46. Schrijver, W., *et al.* Mutation Profiling of Key Cancer Genes in Primary Breast Cancers and Their Distant Metastases. *Cancer research* **78**, 3112-3121 (2018).
47. Greaves, M. & Maley, C.C. Clonal evolution in cancer. *Nature* **481**, 306-313 (2012).
48. Diehl, F., *et al.* Circulating mutant DNA to assess tumor dynamics. *Nature medicine* **14**, 985-990 (2008).
49. Gerlinger, M., *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892 (2012).
50. Yates, L.R., *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature medicine* **21**, 751-759 (2015).
51. Jamal-Hanjani, M., *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* **376**, 2109-2121 (2017).
52. De Mattos-Arruda, L., *et al.* The Genomic and Immune Landscapes of Lethal Metastatic Breast Cancer. *Cell Rep* **27**, 2690-2708.e2610 (2019).
53. De Mattos-Arruda, L., *et al.* Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann Oncol* **25**, 1729-1735 (2014).
54. Murtaza, M., *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* **6**, 8760 (2015).
55. Administration, U.F.a.D. Premarket Approval (PMA): cobas® EGFR Mutation Test v2
56. Oxnard, G.R., *et al.* Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **34**, 3375-3382 (2016).
57. Administration., U.F.a.D. theascreen PIK3CA RGQ PCR kit. (2019).
58. Fribbens, C., *et al.* Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **34**, 2961-2968 (2016).

59. Chandarlapaty, S., *et al.* Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial. *JAMA oncology* **2**, 1310-1315 (2016).
60. Hortobagyi, G.N., *et al.* Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer. *Annals of Oncology* **29**, 1541-1547 (2018).
61. Neven, P., *et al.* Abstract PD2-05: Biomarker analysis by baseline circulating tumor DNA alterations in the MONALEESA-3 study. *Cancer research* **79**, PD2-05-PD02-05 (2019).
62. Di Leo, A., *et al.* Buparlisib plus fulvestrant in postmenopausal women with hormone-receptor-positive, HER2-negative, advanced breast cancer progressing on or after mTOR inhibition (BELLE-3): a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet. Oncology* **19**, 87-100 (2018).
63. Sharma, P., *et al.* Clinical and biomarker results from phase I/II study of PI3K inhibitor BYL 719 (alpelisib) plus nab-paclitaxel in HER2-negative metastatic breast cancer. *Journal of Clinical Oncology* **36**, 1018-1018 (2018).
64. Takano, T., *et al.* A randomized phase II trial of trastuzumab plus capecitabine versus lapatinib plus capecitabine in patients with HER2-positive metastatic breast cancer previously treated with trastuzumab and taxanes: WJOG6110B/ELTOP. *The Breast* **40**, 67-75 (2018).
65. Tolaney, S.M., *et al.* Abstract 4458: Clinical significance of *PIK3CA* and *ESR1* mutations in ctDNA and FFPE samples from the MONARCH 2 study of abemaciclib plus fulvestrant. *Cancer research* **79**, 4458-4458 (2019).
66. Juric, D., *et al.* Abstract GS3-08: Alpelisib + fulvestrant for advanced breast cancer: Subgroup analyses from the phase III SOLAR-1 trial. *Cancer research* **79**, GS3-08-GS03-08 (2019).
67. Bardia, A., *et al.* Abstract CT141: Genetic landscape of premenopausal HR+/HER2-advanced breast cancer (ABC) based on comprehensive circulating tumor DNA analysis and association with clinical outcomes in the Phase III MONALEESA-7 trial. *Cancer research* **79**, CT141-CT141 (2019).
68. Ma, F., *et al.* Safety, efficacy, and biomarker analysis of pyrotinib in combination with capecitabine in HER2-positive metastatic breast cancer patients: A phase I clinical trial. *Journal of Clinical Oncology* **37**, 1035-1035 (2019).
69. Stover, D.G., *et al.* Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **36**, 543-553 (2018).
70. Clatot, F., *et al.* Kinetics, prognostic and predictive values of ESR1 circulating mutations in metastatic breast cancer patients progressing on aromatase inhibitor. *Oncotarget* **7**, 74448-74459 (2016).
71. Hu, Z.Y., *et al.* Identifying Circulating Tumor DNA Mutation Profiles in Metastatic Breast Cancer Patients with Multiline Resistance. *EBioMedicine* **32**, 111-118 (2018).
72. Fribbens, C., *et al.* Tracking evolution of aromatase inhibitor resistance with circulating tumour DNA analysis in metastatic breast cancer. *Ann Oncol* **29**, 145-153 (2018).
73. Russo, M., *et al.* Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. *Cancer discovery* **6**, 147-153 (2016).
74. Oxnard, G.R., *et al.* Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. *Clin Cancer Res* **20**, 1698-1705 (2014).

75. Christie, E.L., *et al.* Reversion of BRCA1/2 Germline Mutations Detected in Circulating Tumor DNA From Patients With High-Grade Serous Ovarian Cancer. *Journal of Clinical Oncology* **35**, 1274-1280 (2017).
76. Lin, K.K., *et al.* BRCA Reversion Mutations in Circulating Tumor DNA Predict Primary and Acquired Resistance to the PARP Inhibitor Rucaparib in High-Grade Ovarian Carcinoma. *Cancer discovery* **9**, 210-219 (2019).
77. Smyth, L.M., *et al.* Efficacy and Determinants of Response to HER Kinase Inhibition in HER2-Mutant Metastatic Breast Cancer. *Cancer discovery* **10**, 198-213 (2020).
78. Ma, F., *et al.* ctDNA dynamics: a novel indicator to track resistance in metastatic breast cancer treated with anti-HER2 therapy. *Oncotarget* **7**, 66020-66031 (2016).
79. O'Leary, B., *et al.* The Genetic Landscape and Clonal Evolution of Breast Cancer Resistance to Palbociclib plus Fulvestrant in the PALOMA-3 Trial. *Cancer discovery* **8**, 1390-1403 (2018).
80. Higgins, M.J., *et al.* Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. *Clin Cancer Res* **18**, 3462-3469 (2012).
81. Schiavon, G., *et al.* Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* **7**, 313ra182 (2015).
82. Maxwell, K.N., *et al.* Comparative clinical utility of tumor genomic testing and cell-free DNA in metastatic breast cancer. *Breast cancer research and treatment* **164**, 627-638 (2017).
83. Kim, S.-B., Dent, R., Wongchenko, M.J., Singel, S.M. & Baselga, J. Concordance between plasma-based and tissue-based next-generation sequencing in LOTUS. *The Lancet Oncology* **18**, e638 (2017).
84. Chae, Y.K., *et al.* Concordance of Genomic Alterations by Next-Generation Sequencing in Tumor Tissue versus Circulating Tumor DNA in Breast Cancer. *Mol Cancer Ther* **16**, 1412-1420 (2017).
85. Adalsteinsson, V.A., *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications* **8**, 1324 (2017).
86. Urso, L., *et al.* ESR1 Gene Mutation in Hormone Receptor-Positive HER2-Negative Metastatic Breast Cancer Patients: Concordance Between Tumor Tissue and Circulating Tumor DNA Analysis. *Front Oncol* **11**, 625636 (2021).
87. Board, R.E., *et al.* Detection of PIK3CA mutations in circulating free DNA in patients with breast cancer. *Breast cancer research and treatment* **120**, 461-467 (2010).
88. Rothé, F., *et al.* Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. *Annals of Oncology* **25**, 1959-1965 (2014).
89. Tolaney, S.M., *et al.* Abstract 4458: Clinical significance of PIK3CA and ESR1 mutations in ctDNA and FFPE samples from the MONARCH 2 study of abemaciclib plus fulvestrant. *Cancer research* **79**, 4458-4458 (2019).
90. Kuang, Y., *et al.* Unraveling the clinicopathological features driving the emergence of ESR1 mutations in metastatic breast cancer. *NPJ breast cancer* **4**, 22 (2018).
91. Steensma, D.P., *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9-16 (2015).
92. Heuser, M., Thol, F. & Ganser, A. Clonal Hematopoiesis of Indeterminate Potential. *Dtsch Arztebl Int* **113**, 317-322 (2016).
93. Genovese, G., *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477-2487 (2014).
94. Razavi, P., *et al.* High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nature medicine* **25**, 1928-1937 (2019).

95. Merker, J.D., *et al.* Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **36**, 1631-1641 (2018).
96. Torga, G. & Pienta, K.J. Patient-Paired Sample Congruence Between 2 Commercial Liquid Biopsy Tests. *JAMA oncology* **4**, 868-870 (2018).
97. Stetson, D., *et al.* Orthogonal Comparison of Four Plasma NGS Tests With Tumor Suggests Technical Factors are a Major Source of Assay Discordance. *JCO precision oncology*, 1-9 (2019).
98. Turner, N.C., *et al.* Circulating tumour DNA analysis to direct therapy in advanced breast cancer (plasmaMATCH): a multicentre, multicohort, phase 2a, platform trial. *The Lancet Oncology*.
99. Li, C.I., Daling, J.R. & Malone, K.E. Incidence of invasive breast cancer by hormone receptor status from 1992 to 1998. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **21**, 28-34 (2003).
100. Green, K.A. & Carroll, J.S. Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state. *Nature Reviews Cancer* **7**, 713-722 (2007).
101. Lindström, L.S., *et al.* Clinically Used Breast Cancer Markers Such As Estrogen Receptor, Progesterone Receptor, and Human Epidermal Growth Factor Receptor 2 Are Unstable Throughout Tumor Progression. *Journal of Clinical Oncology* **30**, 2601-2608 (2012).
102. Miller, T.W., Balko, J.M. & Arteaga, C.L. Phosphatidylinositol 3-kinase and antiestrogen resistance in breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**, 4452-4461 (2011).
103. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nature reviews. Clinical oncology* **12**, 573-583 (2015).
104. Katzenellenbogen, J.A., Mayne, C.G., Katzenellenbogen, B.S., Greene, G.L. & Chandarlapaty, S. Structural underpinnings of oestrogen receptor mutations in endocrine therapy resistance. *Nat Rev Cancer* **18**, 377-388 (2018).
105. Toy, W., *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* **45**, 1439-1445 (2013).
106. Merenbakh-Lamin, K., *et al.* D538G Mutation in Estrogen Receptor- $\alpha$ : A Novel Mechanism for Acquired Endocrine Resistance in Breast Cancer. *Cancer research* **73**, 6856-6864 (2013).
107. Jeselsohn, R., *et al.* Emergence of constitutively active estrogen receptor- $\alpha$  mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* **20**, 1757-1767 (2014).
108. Weis, K.E., Ekena, K., Thomas, J.A., Lazennec, G. & Katzenellenbogen, B.S. Constitutively active human estrogen receptors containing amino acid substitutions for tyrosine 537 in the receptor protein. *Molecular endocrinology (Baltimore, Md.)* **10**, 1388-1398 (1996).
109. Zhang, Q.X., Borg, A., Wolf, D.M., Oesterreich, S. & Fuqua, S.A. An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. *Cancer research* **57**, 1244-1249 (1997).
110. Robinson, D.R., *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**, 1446-1451 (2013).
111. Spoerke, J.M., *et al.* Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant. *Nat Commun* **7**, 11579 (2016).

112. Osborne, C.K., Wakeling, A. & Nicholson, R.I. Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action. *British journal of cancer* **90 Suppl 1**, S2-S6 (2004).
113. DeFriend, D.J., *et al.* Investigation of a new pure antiestrogen (ICI 182780) in women with primary breast cancer. *Cancer research* **54**, 408-414 (1994).
114. Robertson, J.F.R., *et al.* Fulvestrant versus anastrozole for the treatment of advanced breast carcinoma in postmenopausal women. *Cancer* **98**, 229-238 (2003).
115. Howell, A., *et al.* Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 3396-3403 (2002).
116. Howell, A., *et al.* Comparison of Fulvestrant Versus Tamoxifen for the Treatment of Advanced Breast Cancer in Postmenopausal Women Previously Untreated With Endocrine Therapy: A Multinational, Double-Blind, Randomized Trial. *Journal of Clinical Oncology* **22**, 1605-1613 (2004).
117. Osborne, C.K., *et al.* Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: results of a North American trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 3386-3395 (2002).
118. Chia, S., *et al.* Double-blind, randomized placebo controlled trial of fulvestrant compared with exemestane after prior nonsteroidal aromatase inhibitor therapy in postmenopausal women with hormone receptor-positive, advanced breast cancer: results from EFECT. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**, 1664-1670 (2008).
119. Johnston, S.R., *et al.* Fulvestrant plus anastrozole or placebo versus exemestane alone after progression on non-steroidal aromatase inhibitors in postmenopausal patients with hormone-receptor-positive locally advanced or metastatic breast cancer (SoFEA): a composite, multicentre, phase 3 randomised trial. *The Lancet. Oncology* **14**, 989-998 (2013).
120. Turner, N.C., *et al.* ESR1 Mutations and Overall Survival on Fulvestrant versus Exemestane in Advanced Hormone Receptor-Positive Breast Cancer: A Combined Analysis of the Phase III SoFEA and EFECT Trials. *Clinical Cancer Research* **26**, 5172-5177 (2020).
121. Patani, N., *et al.* Differences in the transcriptional response to fulvestrant and estrogen deprivation in ER-positive breast cancer. *Clin Cancer Res* **20**, 3962-3973 (2014).
122. Di Leo, A., *et al.* Final overall survival: fulvestrant 500 mg vs 250 mg in the randomized CONFIRM trial. *J Natl Cancer Inst* **106**, djt337 (2014).
123. Robertson, J.F., *et al.* Activity of fulvestrant 500 mg versus anastrozole 1 mg as first-line treatment for advanced breast cancer: results from the FIRST study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **27**, 4530-4535 (2009).
124. Robertson, J.F.R., *et al.* Fulvestrant 500 mg versus anastrozole 1 mg for hormone receptor-positive advanced breast cancer (FALCON): an international, randomised, double-blind, phase 3 trial. *Lancet* **388**, 2997-3005 (2016).
125. Alves, C.L., *et al.* High CDK6 Protects Cells from Fulvestrant-Mediated Apoptosis and is a Predictor of Resistance to Fulvestrant in Estrogen Receptor-Positive Metastatic Breast Cancer. *Clinical Cancer Research* **22**, 5514-5526 (2016).
126. Toy, W., *et al.* Activating ESR1 Mutations Differentially Affect the Efficacy of ER Antagonists. *Cancer discovery* **7**, 277-287 (2017).

127. Kaminska, K., *et al.* Distinct mechanisms of resistance to fulvestrant treatment dictate level of ER independence and selective response to CDK inhibitors in metastatic breast cancer. *Breast Cancer Res* **23**, 26 (2021).
128. Giessrigl, B., *et al.* Fulvestrant induces resistance by modulating GPER and CDK6 expression: implication of methyltransferases, deacetylases and the hSWI/SNF chromatin remodelling complex. *British Journal of Cancer* **109**, 2751-2762 (2013).
129. Kovacs, E., Zorn, J.A., Huang, Y., Barros, T. & Kuriyan, J. A structural perspective on the regulation of the epidermal growth factor receptor. *Annu Rev Biochem* **84**, 739-764 (2015).
130. Witton, C.J., Reeves, J.R., Going, J.J., Cooke, T.G. & Bartlett, J.M. Expression of the HER1-4 family of receptor tyrosine kinases in breast cancer. *The Journal of pathology* **200**, 290-297 (2003).
131. Press, M.F., *et al.* HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **15**, 2894-2904 (1997).
132. Klapper, L.N., *et al.* The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor for multiple stroma-derived growth factors. *Proc Natl Acad Sci U S A* **96**, 4995-5000 (1999).
133. Graus-Porta, D., Beerli, R.R., Daly, J.M. & Hynes, N.E. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *The EMBO journal* **16**, 1647-1655 (1997).
134. Arteaga, C.L., *et al.* Treatment of HER2-positive breast cancer: current status and future perspectives. *Nature Reviews Clinical Oncology* **9**, 16-32 (2012).
135. Brennan, P.J., Kumogai, T., Berezov, A., Murali, R. & Greene, M.I. HER2/Neu: mechanisms of dimerization/oligomerization. *Oncogene* **19**, 6093-6101 (2000).
136. Yarden, Y. & Sliwkowski, M.X. Untangling the ErbB signalling network. *Nature Reviews Molecular Cell Biology* **2**, 127-137 (2001).
137. Goutsouliak, K., *et al.* Towards personalized treatment for early stage HER2-positive breast cancer. *Nature Reviews Clinical Oncology* **17**, 233-250 (2020).
138. Ross, J.S., *et al.* Relapsed classic E-cadherin (CDH1)-mutated invasive lobular breast cancer shows a high frequency of HER2 (ERBB2) gene mutations. *Clin Cancer Res* **19**, 2668-2676 (2013).
139. Ross, J.S., *et al.* Nonamplification ERBB2 genomic alterations in 5605 cases of recurrent and metastatic breast cancer: An emerging opportunity for anti-HER2 targeted therapies. *Cancer* **122**, 2654-2662 (2016).
140. Bose, R., *et al.* Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer discovery* **3**, 224-237 (2013).
141. Cocco, E., *et al.* Neratinib is effective in breast tumors bearing both amplification and mutation of ERBB2 (HER2). *Science signaling* **11**(2018).
142. Croessmann, S., *et al.* Combined Blockade of Activating *ERBB2* Mutations and ER Results in Synthetic Lethality of ER+/HER2 Mutant Breast Cancer. *Clinical Cancer Research* **25**, 277-289 (2019).
143. Nayar, U., *et al.* Acquired HER2 mutations in ER(+) metastatic breast cancer confer resistance to estrogen receptor-directed therapies. *Nat Genet* (2018).
144. Zabransky, D.J., *et al.* HER2 missense mutations have distinct effects on oncogenic signaling and migration. *Proc Natl Acad Sci U S A* **112**, E6205-6214 (2015).
145. Rabindran, S.K., *et al.* Antitumor activity of HKI-272, an orally active, irreversible inhibitor of the HER-2 tyrosine kinase. *Cancer research* **64**, 3958-3965 (2004).



146. Chan, A., *et al.* Neratinib after trastuzumab-based adjuvant therapy in patients with HER2-positive breast cancer (ExteNET): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet Oncology* **17**, 367-377 (2016).
147. Wong, K.K., *et al.* A phase I study with neratinib (HKI-272), an irreversible pan ErbB receptor tyrosine kinase inhibitor, in patients with solid tumors. *Clin Cancer Res* **15**, 2552-2558 (2009).
148. Burstein, H.J., *et al.* Neratinib, an irreversible ErbB receptor tyrosine kinase inhibitor, in patients with advanced ErbB2-positive breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 1301-1307 (2010).
149. Ma, C.X., *et al.* Neratinib Efficacy and Circulating Tumor DNA Detection of HER2 Mutations in HER2 Nonamplified Metastatic Breast Cancer. *Clin Cancer Res* **23**, 5687-5695 (2017).
150. Hyman, D.M., *et al.* HER kinase inhibition in patients with HER2- and HER3-mutant cancers. *Nature* **554**, 189-194 (2018).
151. Hanker, A.B., *et al.* An Acquired HER2(T798I) Gatekeeper Mutation Induces Resistance to Neratinib in a Patient with HER2 Mutant-Driven Breast Cancer. *Cancer discovery* **7**, 575-585 (2017).
152. Rexer, B.N., *et al.* Human breast cancer cells harboring a gatekeeper T798M mutation in HER2 overexpress EGFR ligands and are sensitive to dual inhibition of EGFR and HER2. *Clin Cancer Res* **19**, 5390-5401 (2013).
153. Sudhan, D.R., *et al.* Hyperactivation of TORC1 Drives Resistance to the Pan-HER Tyrosine Kinase Inhibitor Neratinib in HER2-Mutant Cancers. *Cancer Cell* **37**, 183-199.e185 (2020).
154. Altomare, D.A. & Testa, J.R. Perturbations of the AKT signaling pathway in human cancer. *Oncogene* **24**, 7455-7464 (2005).
155. Miller, T.W., Rexer, B.N., Garrett, J.T. & Arteaga, C.L. Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer. *Breast Cancer Res* **13**, 224 (2011).
156. Vivanco, I. & Sawyers, C.L. The phosphatidylinositol 3-Kinase–AKT pathway in human cancer. *Nature Reviews Cancer* **2**, 489-501 (2002).
157. Hoxhaj, G. & Manning, B.D. The PI3K–AKT network at the interface of oncogenic signalling and cancer metabolism. *Nature Reviews Cancer* **20**, 74-88 (2020).
158. Carpten, J.D., *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448**, 439-444 (2007).
159. Kim, M.S., Jeong, E.G., Yoo, N.J. & Lee, S.H. Mutational analysis of oncogenic AKT E17K mutation in common solid cancers and acute leukaemias. *Br J Cancer* **98**, 1533-1535 (2008).
160. Bleeker, F.E., *et al.* AKT1E17K in human solid tumours. *Oncogene* **27**, 5648-5650 (2008).
161. Rudolph, M., *et al.* AKT1 (E17K) mutation profiling in breast cancer: prevalence, concurrent oncogenic alterations, and blood-based detection. *BMC Cancer* **16**, 622 (2016).
162. Yi, K.H., Axtmayer, J., Gustin, J.P., Rajpurohit, A. & Lauring, J. Functional analysis of non-hotspot AKT1 mutants found in human breast cancers identifies novel driver mutations: implications for personalized medicine. *Oncotarget* **4**, 29-34 (2013).
163. Davies, B.R., *et al.* Preclinical pharmacology of AZD5363, an inhibitor of AKT: pharmacodynamics, antitumor activity, and correlation of monotherapy activity with genetic background. *Mol Cancer Ther* **11**, 873-887 (2012).
164. Tamura, K., *et al.* Safety and tolerability of AZD5363 in Japanese patients with advanced solid tumors. *Cancer Chemother Pharmacol* **77**, 787-795 (2016).

165. Banerji, U., *et al.* A Phase I Open-Label Study to Identify a Dosing Regimen of the Pan-AKT Inhibitor AZD5363 for Evaluation in Solid Tumors and in PIK3CA-Mutated Breast and Gynecologic Cancers. *Clin Cancer Res* **24**, 2050-2059 (2018).
166. Hyman, D.M., *et al.* AKT Inhibition in Solid Tumors With AKT1 Mutations. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **35**, 2251-2259 (2017).
167. Kalinsky, K., *et al.* Effect of Capivasertib in Patients With an AKT1 E17K-Mutated Tumor: NCI-MATCH Subprotocol EAY131-Y Nonrandomized Trial. *JAMA oncology* **7**, 271-278 (2021).
168. Smyth, L.M., *et al.* Capivasertib, an AKT Kinase Inhibitor, as Monotherapy or in Combination with Fulvestrant in Patients with AKT1 E17K-Mutant, ER-Positive Metastatic Breast Cancer. *Clinical Cancer Research* **26**, 3947-3957 (2020).
169. Bosch, A., *et al.* PI3K inhibition results in enhanced estrogen receptor function and dependence in hormone receptor-positive breast cancer. *Sci Transl Med* **7**, 283ra251 (2015).
170. Jones, R.H., *et al.* Fulvestrant plus capivasertib versus placebo after relapse or progression on an aromatase inhibitor in metastatic, oestrogen receptor-positive breast cancer (FAKTION): a multicentre, randomised, controlled, phase 2 trial. *The Lancet. Oncology* **21**, 345-357 (2020).
171. Turner, N.C., *et al.* BEECH: A dose-finding run-in followed by a randomised phase 2 study assessing the efficacy of AKT inhibitor capivasertib (AZD5363) combined with paclitaxel in patients with oestrogen receptor-positive advanced or metastatic breast cancer, and in a PIK3CA mutant sub-population. *Ann Oncol* (2019).
172. Schmid, P., *et al.* Capivasertib Plus Paclitaxel Versus Placebo Plus Paclitaxel As First-Line Therapy for Metastatic Triple-Negative Breast Cancer: The PAKT Trial. *Journal of Clinical Oncology* **38**, 423-433 (2020).
173. Bago, R., *et al.* The hVps34-SGK3 pathway alleviates sustained PI3K/Akt inhibition by stimulating mTORC1 and tumour growth. *The EMBO journal* **35**, 1902-1922 (2016).
174. Stambolic, V., *et al.* Negative Regulation of PKB/Akt-Dependent Cell Survival by the Tumor Suppressor PTEN. *Cell* **95**, 29-39 (1998).
175. Tan, M.H., *et al.* Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res* **18**, 400-407 (2012).
176. Smyth, L.M., *et al.* Selective AKT kinase inhibitor capivasertib in combination with fulvestrant in PTEN-mutant ER-positive metastatic breast cancer. *NPJ breast cancer* **7**, 44 (2021).
177. Turner, N.C., *et al.* Circulating tumour DNA analysis to direct therapy in advanced breast cancer (plasmaMATCH): a multicentre, multicohort, phase 2a, platform trial. *The Lancet Oncology* **21**, 1296-1308 (2020).
178. BWA.aln (v2). Vol. 2019.
179. Picard. Vol. 2019 (2019).
180. Lai, Z., *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic acids research* **44**, e108 (2016).
181. MuTect2. Vol. 2019.
182. ANNOVAR. Vol. 2019.
183. Odegaard, J.I., *et al.* Validation of a plasma-based comprehensive cancer genotyping assay utilizing orthogonal tissue- and plasma-based methodologies. *Clin Cancer Res* (2018).

184. Lanman, R.B., *et al.* Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA. *PLoS One* **10**, e0140712 (2015).
185. US Food and Drug Administration, F. SUMMARY OF SAFETY AND EFFECTIVENESS DATA (SSED). Vol. 2020 Guardant360® CDx FDA approval (2020).
186. McLaren, W., *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
187. Ramos, A.H., *et al.* Oncotator: cancer variant annotation tool. *Human mutation* **36**, E2423-2429 (2015).
188. Tokheim, C. & Karchin, R. CHASMplus Reveals the Scope of Somatic Missense Mutations Driving Human Cancers. *Cell Systems* **9**, 9-23.e28 (2019).
189. Chakravarty, D., *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO precision oncology* **2017**(2017).
190. Chang, M.T., *et al.* Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nature biotechnology* **34**, 155-163 (2016).
191. COSMIC – the Catalogue of Somatic Mutations in Cancer.
192. BRCA exchange.
193. IARC TP53 Database.
194. OncoKB Curation Standard Operating Procedure v2.1.
195. KEGG.
196. Zill, O.A., *et al.* The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients. *Clin Cancer Res* **24**, 3528-3538 (2018).
197. Siravegna, G., *et al.* Plasma HER2 (ERBB2) copy number predicts response to HER2-targeted therapy in metastatic colorectal cancer. *Clin Cancer Res* (2019).
198. Pearson, A., *et al.* Inactivating NF1 Mutations Are Enriched in Advanced Breast Cancer and Contribute to Endocrine Therapy Resistance. *Clinical Cancer Research* **26**, 608-622 (2020).
199. Do, H. & Dobrovic, A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem* **61**, 64-71 (2015).
200. Haile, S., *et al.* Sources of erroneous sequences and artifact chimeric reads in next generation sequencing of genomic DNA from formalin-fixed paraffin-embedded samples. *Nucleic acids research* **47**, e12 (2019).
201. Guo, Q., *et al.* The mutational signatures of formalin fixation on the human genome. *bioRxiv*, 2021.2003.2011.434918 (2021).
202. Desmedt, C., *et al.* Genomic Characterization of Primary Invasive Lobular Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **34**, 1872-1881 (2016).
203. Deniziaut, G., *et al.* ERBB2 mutations associated with solid variant of high-grade invasive lobular breast carcinomas. *Oncotarget* **7**, 73337-73346 (2016).
204. Cha, S., Lee, E. & Won, H.-H. Comprehensive characterization of distinct genetic alterations in metastatic breast cancer across various metastatic sites. *NPJ breast cancer* **7**, 93 (2021).
205. Schrijver, W., *et al.* Receptor Conversion in Distant Breast Cancer Metastases: A Systematic Review and Meta-analysis. *J Natl Cancer Inst* **110**, 568-580 (2018).
206. Thierry, A.R., *et al.* Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic acids research* **38**, 6159-6175 (2010).
207. Bredno, J., Lipson, J., Venn, O., Aravanis, A.M. & Jamshidi, A. Clinical correlates of circulating cell-free DNA tumor fraction. *PLoS One* **16**, e0256436 (2021).

208. Barroso-Sousa, R., *et al.* Prevalence and mutational determinants of high tumor mutation burden in breast cancer. *Annals of Oncology* **31**, 387-394 (2020).
209. Law, E.K., *et al.* The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Science advances* **2**, e1601737 (2016).
210. Xu, X., *et al.* HER2 Reactivation through Acquisition of the HER2 L755S Mutation as a Mechanism of Acquired Resistance to HER2-targeted Therapy in HER2(+) Breast Cancer. *Clin Cancer Res* **23**, 5123-5134 (2017).
211. Zuo, W.J., *et al.* Dual Characteristics of Novel HER2 Kinase Domain Mutations in Response to HER2-Targeted Therapies in Human Breast Cancer. *Clin Cancer Res* **22**, 4859-4869 (2016).
212. Willis, J., *et al.* Validation of Microsatellite Instability Detection Using a Comprehensive Plasma-Based Genotyping Panel. *Clin Cancer Res* **25**, 7035-7045 (2019).
213. Le, D.T., *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (New York, N.Y.)* **357**, 409-413 (2017).
214. Le, D.T., *et al.* Phase II Open-Label Study of Pembrolizumab in Treatment-Refractory, Microsatellite Instability-High/Mismatch Repair-Deficient Metastatic Colorectal Cancer: KEYNOTE-164. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **38**, 11-19 (2020).
215. André, T., *et al.* Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *New England Journal of Medicine* **383**, 2207-2218 (2020).
216. Rakha, E.A., *et al.* Updated UK Recommendations for HER2 assessment in breast cancer. *Journal of clinical pathology* **68**, 93-99 (2015).
217. Janjigian, Y.Y., *et al.* First-line pembrolizumab and trastuzumab in HER2-positive oesophageal, gastric, or gastro-oesophageal junction cancer: an open-label, single-arm, phase 2 trial. *The Lancet. Oncology* **21**, 821-831 (2020).
218. Ran, R., *et al.* Prognostic Value of Plasma HER2 Gene Copy Number in HER2-Positive Metastatic Breast Cancer Treated with First-Line Trastuzumab. *OncoTargets and therapy* **13**, 4385-4395 (2020).
219. Koboldt, D.C., *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
220. Brastianos, P.K., *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer discovery* **5**, 1164-1177 (2015).
221. Juric, D., *et al.* Convergent loss of PTEN leads to clinical resistance to a PI(3)K $\alpha$  inhibitor. *Nature* **518**, 240-244 (2015).
222. Nik-Zainal, S. & Morganella, S. Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clin Cancer Res* **23**, 2617-2629 (2017).
223. Cosmic Mutational Signatures. Vol. 2020 (2020).
224. Nik-Zainal, S., *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
225. Yi, Z., *et al.* Landscape of somatic mutations in different subtypes of advanced breast cancer with circulating tumor DNA analysis. *Scientific reports* **7**, 5995 (2017).
226. Davis, A.A., *et al.* Landscape of circulating tumour DNA in metastatic breast cancer. *EBioMedicine* **58**, 102914 (2020).
227. Ellis, M.J., *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353 (2012).
228. Olivier, M., *et al.* The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clinical Cancer Research* **12**, 1157-1167 (2006).

229. Usary, J., *et al.* Mutation of GATA3 in human breast tumors. *Oncogene* **23**, 7669-7678 (2004).
230. Acuna-Hidalgo, R., *et al.* Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *The American Journal of Human Genetics* **101**, 50-64 (2017).
231. Vasan, N., *et al.* Double PIK3CA mutations in cis increase oncogenicity and sensitivity to PI3K $\alpha$  inhibitors. *Science (New York, N.Y.)* **366**, 714-723 (2019).
232. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. & Fenton, Tim R. APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development. *Cell Reports* **7**, 1833-1841 (2014).
233. Wong, J.K.L., *et al.* Association of mutation signature effectuating processes with mutation hotspots in driver genes and non-coding regions. *Nature Communications* **13**, 178 (2022).
234. McGranahan, N., *et al.* Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* **7**, 283ra254 (2015).
235. Rodriguez-Freixinos, V., *et al.* Genomic heterogeneity and efficacy of PI3K pathway inhibitors in patients with gynaecological cancer. *ESMO Open* **4**, e000444 (2019).
236. Oldenhuis, C.N.A.M., Oosting, S.F., Gietema, J.A. & de Vries, E.G.E. Prognostic versus predictive value of biomarkers in oncology. *European Journal of Cancer* **44**, 946-953 (2008).
237. Kwak, E.L., *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* **363**, 1693-1703 (2010).
238. Shaw, A.T., *et al.* Crizotinib in ROS1-Rearranged Non-Small-Cell Lung Cancer. *New England Journal of Medicine* **371**, 1963-1971 (2014).
239. Chapman, P.B., *et al.* Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New England Journal of Medicine* **364**, 2507-2516 (2011).
240. Larkin, J., *et al.* Combined Vemurafenib and Cobimetinib in BRAF-Mutated Melanoma. *New England Journal of Medicine* **371**, 1867-1876 (2014).
241. Robert, C., *et al.* Improved Overall Survival in Melanoma with Combined Dabrafenib and Trametinib. *New England Journal of Medicine* **372**, 30-39 (2014).
242. De Roock, W., *et al.* Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *The Lancet. Oncology* **11**, 753-762 (2010).
243. US Food and Drug Administration, F. Pre-market Approval: Cobas EGFR MUTATION TEST V2.
244. Parkinson, C.A., *et al.* Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. *PLoS medicine* **13**, e1002198-e1002198 (2016).
245. Oshiro, C., *et al.* PIK3CA mutations in serum DNA are predictive of recurrence in primary breast cancer patients. *Breast cancer research and treatment* **150**, 299-307 (2015).
246. Beroukhi, R., *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899-905 (2010).
247. Dawson, S.J., *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**, 1199-1209 (2013).
248. Harrod, A., *et al.* Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. *Oncogene* **36**, 2286-2296 (2017).

249. Sweeney, C., *et al.* Ipatasertib plus abiraterone and prednisolone in metastatic castration-resistant prostate cancer (IPATential150): a multicentre, randomised, double-blind, phase 3 trial. *The Lancet* **398**, 131-142 (2021).
250. de Bono, J.S., *et al.* Randomized Phase II Study Evaluating Akt Blockade with Ipatasertib, in Combination with Abiraterone, in Patients with Metastatic Prostate Cancer with and without PTEN Loss. *Clinical Cancer Research* **25**, 928-936 (2019).
251. Jones, R.H., *et al.* Fulvestrant plus capivasertib versus placebo after relapse or progression on an aromatase inhibitor in metastatic, oestrogen receptor-positive breast cancer (FAKTION): a multicentre, randomised, controlled, phase 2 trial. *The Lancet Oncology* **21**, 345-357 (2020).
252. Kim, S.B., *et al.* Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet. Oncology* **18**, 1360-1372 (2017).
253. Li, J., *et al.* The AKT inhibitor AZD5363 is selectively active in PI3KCA mutant gastric cancer, and sensitizes a patient-derived gastric cancer xenograft model with PTEN loss to Taxotere. *Journal of translational medicine* **11**, 241 (2013).
254. Marsh, D.J., *et al.* Rapamycin treatment for a child with germline PTEN mutation. *Nature clinical practice. Oncology* **5**, 357-361 (2008).
255. Schmid, G.L., *et al.* Sirolimus treatment of severe PTEN hamartoma tumor syndrome: case report and in vitro studies. *Pediatric research* **75**, 527-534 (2014).
256. Iacobas, I., *et al.* Oral rapamycin in the treatment of patients with hamartoma syndromes and PTEN mutation. *Pediatric blood & cancer* **57**, 321-323 (2011).
257. Komiya, T., *et al.* A pilot study of sirolimus (S) in subjects with Cowden syndrome (CS) with germ-line mutations in PTEN. *Journal of Clinical Oncology* **31**, 2532-2532 (2013).
258. Kingston, B., *et al.* Exceptional Response to AKT Inhibition in Patients With Breast Cancer and Germline PTEN Mutations. *JCO precision oncology*, 1-7 (2019).
259. Alimonti, A., *et al.* Subtle variations in Pten dose determine cancer susceptibility. *Nat Genet* **42**, 454-458 (2010).
260. Papa, A., *et al.* Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell* **157**, 595-610 (2014).
261. Vasan, N., Baselga, J. & Hyman, D.M. A view on drug resistance in cancer. *Nature* **575**, 299-309 (2019).
262. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N Engl J Med* **368**, 842-851 (2013).
263. Diaz Jr, L.A., *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540 (2012).
264. Condorelli, R., *et al.* Polyclonal RB1 mutations and acquired resistance to CDK 4/6 inhibitors in patients with metastatic breast cancer. *Annals of Oncology* **29**, 640-645 (2018).
265. Sehnal, D., *et al.* Mol\* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic acids research* **49**, W431-W437 (2021).
266. Baird, R., *et al.* Abstract PS11-05: Updated data from SERENA-1: A Phase 1 dose escalation and expansion study of the next generation oral SERD AZD9833 as a monotherapy and in combination with palbociclib, in women with ER-positive, HER2-negative advanced breast cancer. *Cancer research* **81**, PS11-05-PS11-05 (2021).
267. Patel, H.K., *et al.* Elacestrant (RAD1901) exhibits anti-tumor activity in multiple ER+ breast cancer models resistant to CDK4/6 inhibitors. *Breast Cancer Research* **21**, 146 (2019).

268. Krastev, B., *et al.* 278MO cfDNA analysis from phase I/II study of lerociclib (G1T38), a continuously dosed oral CDK4/6 inhibitor, with fulvestrant in HR+/HER2- advanced breast cancer patients. *Annals of Oncology* **31**, S351-S352 (2020).
269. Huang, D., Tang, L., Yang, F., Jin, J. & Guan, X. PIK3CA mutations contribute to fulvestrant resistance in ER-positive breast cancer. *American journal of translational research* **11**, 6055-6065 (2019).
270. Wesseling-Rozendaal, Y., Stolpe, A.v.d., Holtzer, L., Inda, M.A.d. & Wiel, P.v.d. Abstract A131: Fulvestrant resistance in an MCF-7 model for breast cancer is associated with complete loss of ER pathway activity and gain of MAPK-AP1 pathway activity. *Molecular Cancer Therapeutics* **18**, A131-A131 (2019).
271. Miller, T.W., *et al.* Loss of Phosphatase and Tensin homologue deleted on chromosome 10 engages ErbB3 and insulin-like growth factor-I receptor signaling to promote antiestrogen resistance in breast cancer. *Cancer research* **69**, 4192-4201 (2009).
272. Chia, S.K.L., *et al.* PIK3CA alterations and benefit with neratinib: analysis from the randomized, double-blind, placebo-controlled, phase III ExteNET trial. *Breast Cancer Res* **21**, 39 (2019).
273. Hsu, J.L. & Hung, M.C. The role of HER2, EGFR, and other receptor tyrosine kinases in breast cancer. *Cancer metastasis reviews* **35**, 575-588 (2016).
274. Kingston, B., *et al.* Genomic profile of advanced breast cancer in circulating tumour DNA. *Nature Communications* **12**, 2423 (2021).
275. Crafter, C., *et al.* Combining AZD8931, a novel EGFR/HER2/HER3 signalling inhibitor, with AZD5363 limits AKT inhibitor induced feedback and enhances antitumour efficacy in HER2-amplified breast cancer models. *Int J Oncol* **47**, 446-454 (2015).
276. Cremona, C.A. & Behrens, A. ATM signalling and cancer. *Oncogene* **33**, 3351-3360 (2014).
277. Halaby, M.-J., Hibma, J.C., He, J. & Yang, D.-Q. ATM protein kinase mediates full activation of Akt and regulates glucose transporter 4 translocation by insulin in muscle cells. *Cellular Signalling* **20**, 1555-1563 (2008).
278. Li, Y. & Yang, D.-Q. The ATM Inhibitor KU-55933 Suppresses Cell Proliferation and Induces Apoptosis by Blocking Akt In Cancer Cells with Overactivated Akt. *Molecular Cancer Therapeutics* **9**, 113-125 (2010).
279. Kostaras, E., *et al.* A systematic molecular and pharmacologic evaluation of AKT inhibitors reveals new insight into their biological activity. *British Journal of Cancer* **123**, 542-555 (2020).
280. Deveson, I.W., *et al.* Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nature biotechnology* (2021).
281. Underhill, H.R., *et al.* Fragment Length of Circulating Tumor DNA. *PLOS Genetics* **12**, e1006162 (2016).
282. Mouliere, F., *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Science Translational Medicine* **10**, eaat4921 (2018).
283. Shen, S.Y., *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579-583 (2018).
284. Markus, H., *et al.* Refined characterization of circulating tumor DNA through biological feature integration. *Scientific reports* **12**, 1928 (2022).
285. Tate, J.G., *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research* **47**, D941-D947 (2018).
286. Zivanovic Bujak, A., *et al.* Circulating tumour DNA in metastatic breast cancer to guide clinical trial enrolment and precision oncology: A cohort study. *PLoS Med* **17**, e1003363 (2020).

287. Rothwell, D.G., *et al.* Utility of ctDNA to support patient selection for early phase clinical trials: the TARGET study. *Nature medicine* (2019).
288. Vidula, N., *et al.* Tumor Tissue- versus Plasma-based Genotyping for Selection of Matched Therapy and Impact on Clinical Outcomes in Patients with Metastatic Breast Cancer. *Clin Cancer Res* **27**, 3404-3413 (2021).
289. Vo, H.H., *et al.* Initiative for Molecular Profiling and Advanced Cancer Therapy (IMPACT2): Challenges and Opportunities in Conducting an MD Anderson Randomized Study in Precision Oncology. *Journal of Clinical Oncology* **39**, 3140-3140 (2021).
290. André, F., *et al.* Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *New England Journal of Medicine* **380**, 1929-1940 (2019).
291. Flaherty, K.T., *et al.* The Molecular Analysis for Therapy Choice (NCI-MATCH) Trial: Lessons for Genomic Trial Design. *JNCI: Journal of the National Cancer Institute* **112**, 1021-1029 (2020).
292. Zardavas, D., *et al.* The AURORA initiative for metastatic breast cancer. *Br J Cancer* **111**, 1881-1887 (2014).
293. Mangat, P.K., *et al.* Rationale and Design of the Targeted Agent and Profiling Utilization Registry (TAPUR) Study. *JCO precision oncology* **2018**(2018).
294. Hrebien, S., *et al.* Early ctDNA dynamics as a surrogate for progression free survival in advanced breast cancer in the BEECH trial. *Ann Oncol* (2019).
295. O'Leary, B., *et al.* Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. *Nat Commun* **9**, 896 (2018).
296. Pascual, J., *et al.* Abstract PS5-02: Assessment of early ctDNA dynamics to predict efficacy of targeted therapies in metastatic breast cancer: Results from plasmaMATCH trial. *Cancer research* **81**, PS5-02-PS05-02 (2021).



## Thanks

To:

Nick for the all the guidance, support, and confidence he has shown in me. Nick has been an inspiration throughout.

Isaac for the day-to-day support, guidance, and chats.

Alex for the endless patience with my questions, and for taking me under his wing late in the PhD to teach me all things tissue culture.

Ros for the endless bioinformatic and coding support, and for the margaritas in San Antonio.

Maite for taking on further experiments from my project when it was time to head back to doctoring.

The whole Molecular Oncology group, for creating a supportive, collaborative, and fun environment within which it was a pleasure to work in.

The patients, families, and research staff of plasmaMATCH, to whom I am greatly indebted.

Fiona, for always being interested, supportive, and proud.

Dominic, Sebastian and Otilie, for putting up with my evening and weekend trips to the lab, for the endless support and cheerleading, and for generally being all-round brilliant.