UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Cancer Sciences Unit

The role of the tumour suppressor gene, *TP53*, in determining breast tumour subtype

by

Katie Elizabeth Packwood

Thesis for the degree of Doctor of Philosophy

September 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Cancer Sciences Unit

Thesis for the degree of Doctor of Philosophy

THE ROLE OF THE TUMOUR SUPPRESSOR GENE, *TP53*, IN DETERMINING BREAST TUMOUR SUBTYPE

Katie Elizabeth Packwood

Background and hypothesis

A germline *TP53* mutation predisposes to young breast cancer and other tumours recognized clinically as Li-Fraumeni Syndrome. There is a growing recognition that HER2 amplified breast cancer is more frequent in *TP53* germline mutation carriers than amongst sporadic cases. Frequently HER2 amplification in pure ductal carcinoma *in situ* (DCIS) presenting in sporadic breast disease does not typically progress to HER2 amplified invasive breast cancer.

The hypothesis for this project is that an inherited *TP53* gene mutation is important in driving the HER2 amplified breast tumour subtype. I have explored the timing of loss of *TP53* function and the type of both inherited and acquired *TP53* mutation, (missense or nonsense mutation) in order to evaluate the importance of each in determining breast tumour subtype.

Materials and methods

The Cohort study Of p53 related Early onset breast cancer (COPE) cohort comprises 136 breast tumour formalin fixed paraffin embedded (FFPE) blocks from 45 patients with a germline *TP53* mutation. Full-face H&E sections were made for morphological review and cores were selected from the invasive tumour and DCIS for tissue microarrays (TMA) with immunohistochemistry staining.

A HaloPlex® targeted enrichment kit was used to characterise the genetic landscape in the COPE tumours and in a HER2 amplified control group of 9 young breast cancer cases drawn from the Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) cohort.

The early biochemical mechanisms involved in the development of these breast tumours were investigated using the MCF10A cell line for three-dimensional cell culture in order to study the breast glandular architecture and to mimic ductal carcinoma *in situ* (DCIS). Cells were grown as

spheroids on a reconstituted basement membrane in a growth arrest polarized fashion. The tumour suppressor genes *TP53* and *BRCA1* were transiently knocked down and stable over amplification of the oncogene *ERBB2* was performed on the MCF10A cell. This was to understand the influence these key genes have on the morphogenetic processes including luminal clearing and proliferation during the early stages of tumourgenesis.

Primary fibroblasts derived from associated HER2+ and triple receptor negative breast cancers (TNBC) were grown in culture and the expression of stromal markers investigated.

Results

Pathological analyses: I confirmed a high prevalence of HER2 amplified, high grade, ductal no special type tumours in *TP53* associated invasive breast tumours. HER2 overexpression was confirmed in 19/36 (52.8%) of *TP53* associated cases compared with 717/2956 (24%) from the large young onset POSH cohort (aged 40 or younger at diagnosis) (p=<0.001). I also noted that 13/36 (31.1%) were ER+/HER2+ tumours which were significantly higher than the POSH cohort (p=0.002). Frequent widespread high grade ductal carcinoma *in situ* (DCIS), a high frequency of sclerotic tumour stroma (80.6% of cases) and confirmed upregulation of TGFβ signalling was reported. All tumour cases showed abnormal p53 immunohistochemistry and patchy staining in the DCIS was suggestive of p53 signalling deregulation and stabilisation of mutant p53. Our work supports the hypothesis that a germline *TP53* mutation strongly predisposes to HER2 amplified, high grade, ER+ve tumour subtypes in contrast to triple negative breast cancers typically reported in *BRCA1* carriers.

Genomic analysis: Next generation sequencing (NGS) data from invasive tumour and DCIS samples revealed a loss of heterozygosity (LOH) of the wild type allele in 14/16 (87.5%) tumour samples and 4/4 (100.0%) DCIS samples. Clonality data suggests that cases typically acquire few somatic mutations and are clonally distinct with consistent widespread LOH of *TP53*. NGS data from HER2+ cases from POSH revealed a high variant allele frequency (VAF) of mutant *TP53* reads consistent with widespread clonal *TP53* loss and indicative of an early event.

Cellular biology: ErbB2 overexpression and a loss of *TP53* complement one another and drive a proliferative cell type losing cellular contact inhibition. This was further confirmed in three-dimension with ErbB2 overexpression and loss of *TP53* driving an aggressive invasive phenotype. Cells devoid of BRCA1 and p53 were shown to not cooperate suggesting why *TP53* carriers develop HER2+ and *BRCA1* carriers develop triple receptor negative tumours.

Conclusions: An early loss of *TP53* seems to be fundamental in driving a HER2 breast tumour phenotype. Early loss of *TP53* and HER2 overexpression cooperate and give DCIS lesions an invasive and selective advantage driving the evolution of Li-Fraumeni Syndrome associated breast

tumours. Patients with a missense *TP53* mutation seem to be predisposed to develop a sclerotic tumour stroma.

Table of Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, ...[please print name]

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

[title of thesis] .....

.....

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. [Delete as appropriate] None of this work has been published before submission [or] Parts of this work have been published as: [please list references below]:

Signed: .....

Date: .....

Acknowledgements

Undertaking this PhD has been an incredible experience and there are many people who have contributed to this. With Leicester City winning the Premier League title too, what a year it has been! I would like to sincerely thank;

My lead supervisor and mentor Diana Eccles for giving me the opportunity to work on such an exciting PhD project and her on going unfailing support and willingness to share her knowledge and expertise with me. I am extremely grateful to Diana who despite such a busy clinical and academic workload has always been so generous with her time and the guidance she has given me over the past few years has been absolutely invaluable.

Jeremy Blaydes for his advice and for sharing his scientific knowledge, particularly around the topics of cellular mechanisms and p53, which has helped me to take my project so far. Gareth Thomas for his guidance and expertise involved in the tissue microenvironment and for allowing me to join his lab team. Adrian Bateman for his support, clinical pathology knowledge and for leading the morphology review. Jon Strefford for his helpful supervision and genomic expertise.

Mat Rose-Zerilli who has given me so much of his time over the past few years. His on going belief, guidance and support has been invaluable during this project and to him I am truly thankful. I probably owe him at least a beer!

Matthew Sommerlad, Guy Martland, Emily Shaw, Kawan Moutasim for taking the time to investigate and take part in the morphology review. Reuben Pengelly for his bioinformatics expertise and Tom Maishman for his statistical knowledge.

Lily Barnett for her contribution with the DNA extractions during her summer project. You were a pleasure to have in the lab and share my expertise with.

Toby Mellows and Monette Lopez, for their technical support in pathology. Jon Ward, Jenny Norman, Karen Pickard, Dave Johnston for their support and excellent training in microscopy. Everyone who has given me their support in the Histochemistry Unit and the Bioimaging Unit.

The Experimental Pathology, Blaydes and Strefford lab group members past and present for their patience, guidance and support over the past few years. Hollie Robinson, Chris Hanley, Durga Sabnis, Charlie Birts, Sara Waise, Marina Parry, Chantal Hargreaves, Helen Parker, Veronika Jenei, Arindum Banerjee, Matt Darley, Max Mellone, Jason Fleming, Steve Frampton, Harindra Amarasinghe Kankanamge and Kate Latham.

Our collaborators Louise Jones and Linda Haywood, at the Bart's Cancer Institute, for giving me the opportunity to visit and learn how to construct tissue microarrays.

The Pathological Society of Great Britain and Ireland for funding this work.

The group of girls I started this incredible journey with: Faith Bateman, Lyndsy Ambler, Leanne Wickens, Abbie Mead, Caroline Barker, Dannielle Wellington, Liz Lemm, Ali Hill and Natt Day. We did it!

Last but certainly not least, my parents Jill and Steve, and sister Vicky for their on going and continued support over the years. I owe a lot to their encouragement and belief. My partner Ollie Hutton for his patience, understanding and ability to have a large gin and tonic waiting at home after a particularly long day in the lab! I would not have come this far without them!

Definitions and Abbreviations

ABC complex	Avidin-biotin block complex
ACC	Adrenocortical carcinoma
aCGH	Array based comparative genomic hybridization
ADH	Atypical ductal hyperplasia
ALH	Atypical lobular hyperplasia
α-SMA	Alpha smooth muscle actin
αvβ6	Alpha v beta 6
BMI	Body mass index
BRCA1	Breast cancer susceptibility gene 1
BRCA2	Breast cancer susceptibility gene 2
CAF	Cancer-associated fibroblast
CK5/6	Cytokeratin 5/6
CNAs	Copy number aberrations
CNVs	Copy number variants
COPE	Cohort study Of p53 related Early onset breast cancer
COSMIC	Catalogue of somatic mutations in cancer
COX-2	Cyclooxygenase-2
CRT	Cyclic reversible termination
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBD	DNA binding domain
DCIS	Ductal carcinoma *in situ*
ddNTPs	Dideoxynucleotides

dNTPs	Deoxynucleotides
dsDNA	Double stranded DNA
Ductal (NST)	Ductal (no special type)
DTT	Dithiothreitol
ECD	Enrichment control DNA
ECM	Extracellular matrix
EGFR	Epidermal growth factor receptor
ER	Oestrogen receptor
FACS	Fluorescence-activated cell sorting
FCS	Fetal calf serum
FFPE	Formalin fixed paraffin embedded
FISH	Fluorescence *in situ* hybridisation
gDNA	Genomic DNA
GII	Genomic instability index
GOF	Gain of function
H&E	Hematoxylin and eosin
HDs	Homozygous deletions
HER2	Human epidermal growth factor receptor 2
HGNC database	HUGO Gene Nomenclature Committee Database
IARC	International Agency for Research on Cancer
IDC	Invasive ductal carcinoma
IHC	Immunohistochemistry
Indels	Insertions/deletions
JAK2	Janus kinase 2
LAP	Latency-associated peptide

LB	Luria-Bertani
LCIS	Lobular carcinoma *in situ*
LCM	Laser capture microdissection
LFI	Li-Fraumeni Incomplete
LFL	Li-Fraumeni-like Syndrome
LFS	Li-Fraumeni Syndrome
LLC	Large latent complex
LOH	Loss of heterozygosity
LOXs	Lysyl oxidases
LTBP	Latent TGFβ-binding proteins
MDM2	Murine double minute 2
MEMo	Mutual Exclusivity Modules in cancer
NBF	Normal breast fibroblast
NHS	National Health Service
NGS	Next generation sequencing
OS	Overall survival
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Paired-end
POSH	Prospective study of Outcomes in Sporadic versus Hereditary breast cancer
PR	Progesterone receptor
pSMAD2/3	Phospho-SMAD2/3
QC	Quality control
qPCR	Quantitative polymerase chain reaction
SLC	Small latent complex

SNP	Single nucleotide polymorphism
TET2	Tet methylcytosine dioxygenase 2
TGFβ	Transforming growth factor beta
TILs	Tumour infiltrating lymphocytes
TMA	Tissue microarray
TNBC	Triple negative breast cancer
TP53	Tumour protein p53
VAF	Variant allele frequency
VHL	Von Hippel-Lindau tumour suppressor
WES	Whole exome sequencing
WHO	World Health Organisation
WT	Wild type
YBC	Young breast cancer

Chapter 1: Introduction

1.1 Breast cancer

1.1.1 Breast cancer: the facts

Annually, more than 1.3 million patients worldwide - nearly 50,000 in the UK alone - are diagnosed with breast cancer, making it the most common type of cancer and the second largest cancer related killer after lung cancer [1, 2]. The majority of breast cancer reported is diagnosed in women, contributing a third of all reported primary tumours reported in women in the UK. However 1% of breast cancers diagnosed are reported in men with 350 cases diagnosed in the UK every year [1, 2]. Between 2009 and 2011, 80% of breast cancer cases in the UK were reported in women over the age of 50, the majority of which were postmenopausal and 24% were in women 75 years of age or older [3-7]. However the remaining 20% of breast cancer cases are observed in younger patients often presenting with a more aggressive breast tumour subtype as well as a poorer prognosis. Frequently these are associated with genetic abnormalities and a family history of young onset breast cancer.

Through screening programmes and improved therapies, the rate of mortality has decreased in recent years with the five year survival increasing in the UK from 40% in 1954 to 85% in 2014 [2, 8]. Breast cancer is recognised now to be very heterogeneous and forms a collection of different types of malignancies rather than one disease with each subtype showing a range of responses to treatment and overall survival. Therefore different subtypes are treated according to whether there is only *in situ* disease, tumour grade, receptor status and if metastasises are present. Clinically, these subtypes are often defined through their level of co-expression of receptors oestrogen receptor (ER), progesterone receptor (PR) and overexpression of the human epidermal growth factor receptor 2 (HER2+). Breast tumours predominantly fall into four phenotypes: ER+/PR+/HER2-, ER-/PR-/HER2+, ER+/PR+/HER2+ or none of the above: triple receptor negative breast cancer (TNBC) ER-/PR-/HER2- [9].

1.1.2 Breast cancer subtypes and assessment

In the clinical setting, patients that present with suspected breast disease firstly have a biopsy taken. A pathologist reviews the sample and if disease is present the three surgery approaches are: full mastectomy, wide local excision and a quadrantectomy [10]. This tissue is then dissected and fixed in formalin fixed paraffin embedded (FFPE) blocks where hematoxylin and eosin (H&E) stained sections can be cut and morphology assessed. The pathologist comments on the clinical history, macroscopic features, and microscopic features including tumour type, grade, size, presence of *in situ* disease, vascular/lymphatic invasion and receptor status. The grade is determined by the

Nottingham Histologic Score system and relies on the formation of tubules (how much the tumour resembles the glandular architecture of normal breast tissue), the nuclear pleomorphism of the tumour cells (how large and irregular the cells look) and the mitotic count (the number of mitoses per 10 high power fields, i.e. the amount of mitotic activity). The receptor status of the tumour is determined using immunohistochemistry (IHC) to check for protein expression. Additionally for suspected HER2+ cases that score a borderline IHC score (2+), fluorescence *in situ* hybridisation (FISH) can be used as a measure of HER2 copy number. The receptor status of a tumour is important as this has an impact on possible treatment options. For example, the current national guidelines recommend a treatment of Tamoxifen post-surgery for those ER+ tumours with additional chemotherapy for those patients that are particularly at risk of disease reoccurrence [10]. HER2+ tumours are now treated with a drug that specifically targets and blocks the HER2 receptor, Trastuzumab (Herceptin).

Once the receptor status is reported for a case, the tumour can then be approximated to one of four groups: basal, HER2+, luminal A or luminal B. However clinically luminal A and luminal B subgroups are not often particularly differentiated. These groups are divided by their ER status and broken down further based on the remaining two receptors (see fig. 1)

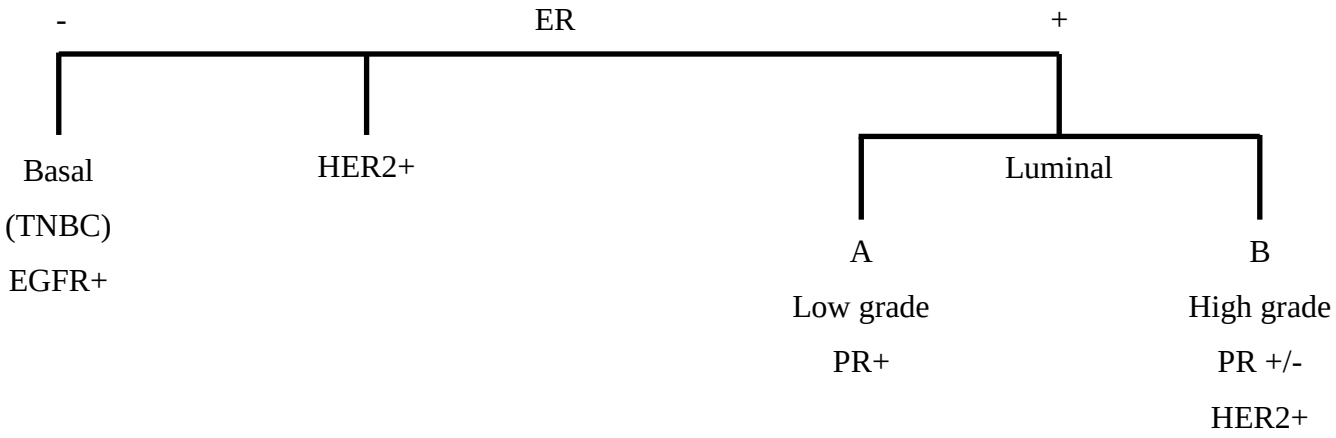


Figure 1 Breast cancer subtypes.

Breast cancers can be characterised into subtypes depending on the expression of the receptors ER, PR and HER2. These groups reveal different prognostic outcomes with basal and the HER2+ overexpressing tumours often associated with a poorer prognosis.

1.1.3 Breast morphology: The road to invasive carcinoma

Breast cancer is regarded as an extremely heterogeneous disease with diverse subtypes and prognoses, however the morphological processes by which these tumours evolve follow a similar path. Through epigenetic changes and/or an increase in genetic burden following the accrual of

somatic mutations and copy number changes, tissue undergoes changes from normal lobular tissue. Neoplastic benign disease, such as atypical ductal hyperplasia (ADH), develops through the accrual of additional genomic lesions with progressive genomic instability driving the formation of the invasive precursor lesion ductal carcinoma *in situ* (DCIS). Once malignant cells overcome this myoepithelial barrier, this ultimately leads to invasive carcinoma with metastatic potential (fig. 2) [11]. Malignant breast disease fall into one of four categories: no special type (ductal no special type (NST)), pure special type (90% purity), mixed tumour type (50-90% special type component) and other malignant tumour if there is only *in situ* disease present. For the invasive disease categories, these are split into their component types: ductal (NST), lobular, medullary like, mucinous and tubular/cribriform. The specific categories of invasive breast disease are derived from certain cell lineages. For example invasive ductal (NST) carcinoma typically evolves from DCIS originating from the ducts, with lobular often evolving from lobular carcinoma *in situ* (LCIS) derived from the breast lobules. If the primary tumour metastasises form secondary tumours, these often heavily resemble the primary tumour with the acquisition of further mutations and additional Darwinian evolution. Breast disease evolution is shown in fig. 2.

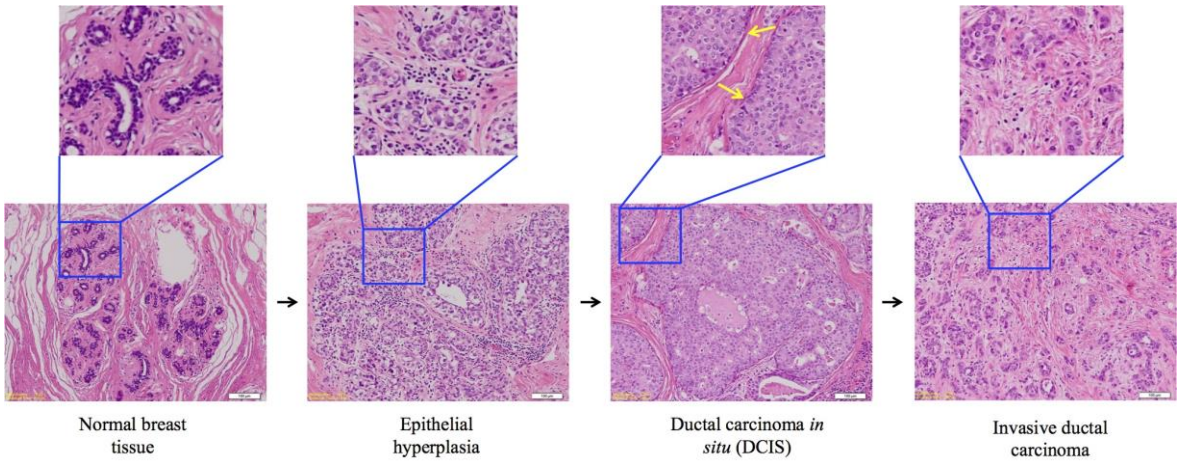


Figure 2 The development of invasive ductal carcinoma.

Through epigenetic changes and increased genetic instability normal tissue has the potential to become premalignant and then eventually malignant. Yellow arrows identify the myoepithelial barriers in the DCIS case.

The pathology of breast tissue has many different types of benign, *in situ* and invasive components adding to the complexity and the mechanistic behaviour of breast disease evolution. As well as atypical ductal hyperplasia (ADH) (proliferative lesion leading to enlargement of the ducts), there are other benign changes such as atypical lobular hyperplasia (ALH) (proliferative lesion leading to enlargement of lobules), columnar changes of the lining of the ducts, sclerosing adenosis (proliferative lesion of the lobules), apocrine changes (lesions in cells of an apocrine lineage),

radial scar/complex sclerosing lesion (area of particularly dense tissue), fat necrosis, cysts, fibrocystic disease (proliferative dense stroma associated with duct dilation and cyst formation) and fibroadenomas (encapsulated area involving lobules and fibrosis of the stroma) reported in breast tissue. Not all of these lesions are associated with an increased risk of disease progression with ADH and ALH described as having a higher risk of advancing as DCIS lesions and lobular carcinoma *in situ* (LCIS).

DCIS is described as a precursor to invasive breast carcinoma and is a neoplastic proliferation of the epithelial cells. At this stage of disease, malignant cells are restricted to the ducts by an intact basement membrane and a myoepithelial layer (shown in fig. 2). This lesion is graded low, intermediate or high with seven growth patterns that are typically presented in clinic including solid, comedo, cribriform, micropapillary, papillary, apocrine and flat. Typically solid and comedo DCIS is associated with high grade and flat is typically associated with low grade. One of the major obstacles currently in the clinic is the successful identification of which cases of DCIS will remain as *in situ* disease, and which of those 1% of cases will continue to progress to become invasive [12]. In recent years through widespread breast cancer screening programs, DCIS now represents 20-25% of detected malignant lesions diagnosed and the majority of these cases are now over treated with surgery followed by radiotherapy [12]. Unlike invasive tumours that typically evolve from a grade 1 to potentially a grade 3, DCIS does not continue to evolve from a low grade lesion to high grade. High grade DCIS is often associated with a higher probability of invading the surrounding breast tissue however a clear biomarker to assess invasive risk remains elusive [12, 13].

1.2 Li-Fraumeni Syndrome (LFS)

1.2.1 In the beginning

This syndrome was first described in 1969 by Li and Fraumeni after examining 600 medical records and 418 death certificates of children in the U.S. who died of a rhabdomyosarcoma [14]. Through examination of these medical records, four families were identified in this study with a high frequency of young onset malignancy including breast, acute leukaemia, lung, pancreas and skin carcinoma in close family members [14, 15]. Through additional epidemiological studies, including pathological examination and searches of the Cancer Family Registry of the National Cancer Institute, this initial observation was confirmed as a rare familial autosomal dominant cancer predisposition syndrome now been described as Li-Fraumeni Syndrome (LFS) [15-17].

1.2.2 Li-Fraumeni Syndrome: a syndrome definition

Li-Fraumeni Syndrome (LFS) is associated with a wide array of tumour types developing over a broad age range, including childhood. The characteristic or 'core' tumours associated with LFS include malignancies of the soft tissue and bone sarcomas, premenopausal breast cancers, adrenal cortical carcinomas, leukemias and brain tumours (specifically choroid plexus carcinomas). The more characteristic tumours make up around 77% of malignancies reported in LFS patients and occur at significantly earlier ages than in the general population [15, 18-20]. To a lesser extent, tumours such as lymphomas, lung cancer, melanoma and gastrointestinal malignancies have also been reported in LFS families [19, 20].

Unlike a lot of other diseases, LFS is identified as a syndrome diagnosis meaning that the diagnostic criteria are based clinically. Consequently the clinical definition of LFS is complicated, with varying levels of specificity and sensitivity associated with different classifications.

Criteria	**Description**
Classic (1988)	A proband with any bone or soft tissue sarcoma diagnosed before the age of 45 years, a first-degree relative with cancer before the age of 45 years and a first or second degree relative in the same lineage with a cancer under 45 years or a sarcoma diagnosed at any age [14, 15].
Li-Fraumeni-like Syndrome (LFL), Birch definition (1994)	A proband with any childhood cancer or sarcoma, brain tumour or adrenocortical carcinoma diagnosed before 45 years of age, a first or second degree relative with a typical LFS cancer (including sarcoma, breast cancer, adrenocortical carcinoma, brain tumour or leukaemia) diagnosed at any age and a first or second degree relative in the same lineage diagnosed with any form of cancer before the age of 60 [21].
Li-Fraumeni-like Syndrome (LFL), Eeles definition (1995)	Two first or second degree relatives with LFS associated tumours at any age [22].
Li-Fraumeni Incomplete (LFI) (1984)	A proband with breast cancer and a first-degree relative with a rhabdomyosarcoma [23, 24].

Chompret criteria, Bougeard revised (2008) and Tinat revised (2009)	A proband affected by a characteristic LFS tumour (including soft tissue or osteosarcomas, premenopausal breast cancer, adrenal cortical carcinomas (ACCs), brain tumours, leukemias and lung bronchoalveolar carcinomas) before the age of 46 years and at least a first or second degree relative affected by a characteristic LFS tumour (except breast cancer if the proband is affected by this type of cancer) before the age of 56 years or multiple primary tumours or a proband with numerous primary tumours (except multiple breast cancers) in which two fulfil the characteristic LFS tumour category and the first of these tumours was diagnosed before the age of 46 years or a proband with an ACC or choroid plexus carcinoma regardless of family history and age of onset [25-28].

Table 1 Different Criteria proposed for defining Li-Fraumeni Syndrome.

The criteria used to make a diagnosis of LFS are based on cancers observed in an individual and family members constituting a syndrome diagnosis.

Since the original 'classic' criteria are based on the description by Li and Fraumeni, other criteria have been proposed including a broader classification suggested in 1994 by Birch and colleagues, which describes Li-Fraumeni-like syndrome (LFL) including more cancer types [21]. To fall into this category, the proband must have been diagnosed with any childhood cancer or sarcoma, brain tumour or adrenocortical carcinoma before 45 years of age, a first or second degree relative with a typical LFS cancer at any age and a first or second degree relative in the same lineage with any form of cancer before the age of 60 [21]. Additionally a simpler criterion described as Li-Fraumeni Incomplete (LFI) was proposed in 1993 by Brugières and colleagues: a proband with breast cancer and a first degree relative with a rhabdomyosarcoma [24].

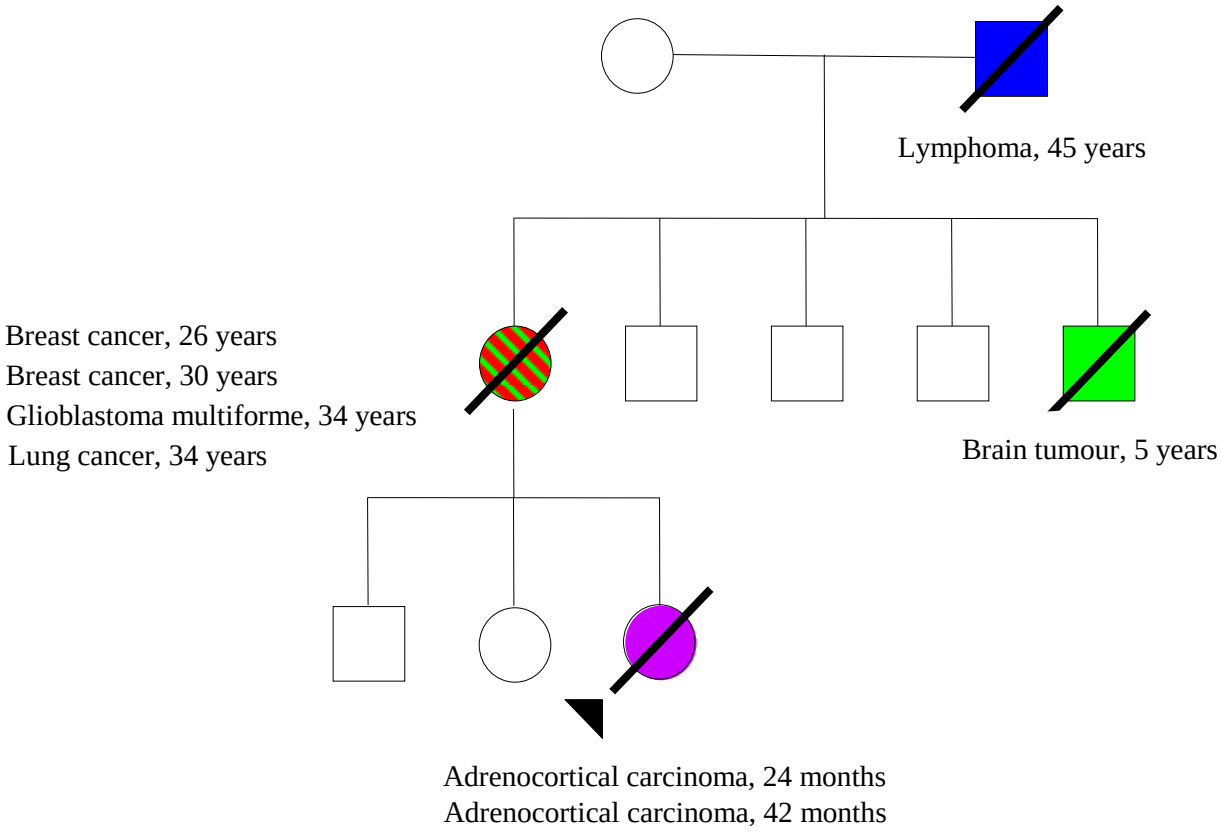


Figure 3 A classic Li-Fraumeni pedigree.

Filled squares and circles represent affected family members. A slash representing a deceased family member. This pedigree does not show the asymptomatic members with a germline *TP53* mutation.

In 1990 a further way of describing LFS was proposed by the French group following the discovery that in a large proportion of LFS cases, a germline *TP53* mutation was detected [29]. The group proposed the 'Chompret criteria' in order to identify those cases harbouring a germline mutation [25]. The criteria recommended *TP53* germline testing where a proband was affected by either a characteristic LFS tumour before the age of 36 years in addition to at least a first or second degree relative affected by a characteristic LFS tumour (except breast cancer if the proband is affected by this type of cancer) before the age of 46 years, or multiple primary tumours. As well as a proband with numerous primary tumours in which two are characteristic LFS tumours and the first of these tumours was diagnosed before the age of 36 years regardless of family history. In addition to a proband with adrenocortical carcinoma regardless of family history [25]. Using the Chompret criteria originally proposed, approximately 29-35% of families meeting these were shown to harbour a germline *TP53* mutation [19, 26]. The Chompret criteria has since been revised to increase the age of tumour onset to 46 years of age for the proband and 56 years of age for the first or second degree relative with an LFS tumour (including specific subset of paediatric malignancies including choroid plexus carcinoma) [26, 27]. Ruijs and colleagues undertook an independent study

and of the 105 families that were referred for *TP53* testing and met the revised Chompret criteria, 21% had a confirmed germline *TP53* mutation. 22/24 of those families with a germline mutation in this cohort met the Chompret criteria (sensitivity 92%) [30]. The remaining 2 families with a germline mutation were present in LFS suspected families but did not fulfil the revised Chompret, LFS or LFL criteria [30].

Chompret and colleagues believed that the stringent criteria of classic LFS and the more lenient criteria associated with LFL and LFI, did not allow for correction of selection bias which is why they proposed their relaxed criteria. From their work and using a less stringent criteria, they concluded that the cancer risks of these patients are very high despite the possibility of unaffected carriers falling into the criteria, there was no evidence of low penetrance mutations within these patients and that the proportion of *de novo* mutations is probably of considerable importance [31]. Unsurprisingly the more stringent the criteria, the higher the incidence of germline *TP53* mutation detection. Within these criteria groups, the detection of germline *TP53* mutations differed with around a 70% incidence in classic LFS, a 30-40% incidence in LFL and only 6% in LFI families [14, 18, 21, 31-35]. A likely possibility for this is that a germline *TP53* mutation does not necessarily guarantee the clinical diagnosis of LFS [31]. Furthermore if the *TP53* mutation is located in a malfunctioning regulatory region, a noncoding region or a large or whole exon deletion, it may not initially be detected. Therefore there is also a degree of sensitivity in the type of molecular test used and whether it is able to detect mutations across the entire gene or large variants including insertions/deletions (indels) which are often difficult to detect [31].

A flaw with this selection criteria to identify LFS and families harbouring a germline *TP53* mutation is that they rely on clinical data from a family pedigree. Consequently for a patient with a *de novo TP53* mutation it would be difficult to satisfy this criteria enabling testing. Chompret et al. showed that through examining the incidence of germline *TP53* mutations in childhood cancers either through the presence of several primary tumours in the proband, or a family history of cancer before the age of 46 years of age in a first or second degree relative, 24% (4/17) of the cohort were presumed to have a *de novo* mutation. Furthermore work published by Gonzalez and colleagues showed that out of 75 probands who tested positive for a germline *TP53* mutation, at least 7% harboured a *de novo* mutation with this figure possibly as high as 20% where no family history was available [36]. These studies show the importance of *de novo* mutations and that a lack of family history does not necessarily rule out a germline *TP53* mutation.

1.2.3 Cancer risk and increased incidence in females with LFS

The penetrance of cancer in germline *TP53* mutation carriers and LFS varies significantly from men to women, with a lifetime risk of 73% in men compared to nearly 100% in women when 13 families were investigated [31, 37]. The differences in lifetime risk observed between the sexes varied depending on age and during childhood the risk was 19% in males compared to 12% in

females [31]. However over time this risk dramatically increases in women and by the age of 16, 45 and 85 this risk was at 12%, 84% and 100% respectively compared to 19%, 41% and 73% in males [31]. From this study the higher penetrance in females was particularly obvious in the 16-45 age bracket, which represented 80% of cases. However work undertaken by Hwang and colleagues showed that LFS associated premenopausal breast cancer was only to some extent the reason for the differences in incidence between male and females germline carriers (see fig. 5) [38]. This was identified once sex specific tumours were removed from the statistics and the differences in risk were still apparent (fig. 5).

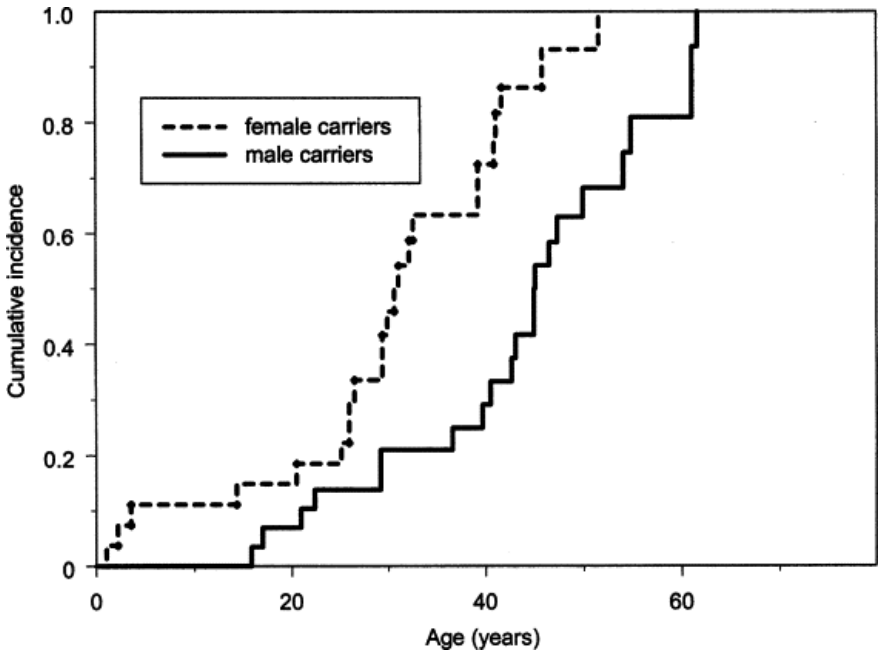


Figure 4 Kaplan-Meier of cumulative frequency of cancer in LFS patients taken from work published by Hwang and colleagues.

27 male and 29 female germline *TP53* mutation carriers from 7 kindreds were evaluated for the age at which they developed their first tumour with regular follow up until last contact was made or death for those asymptomatic gene carriers [38].

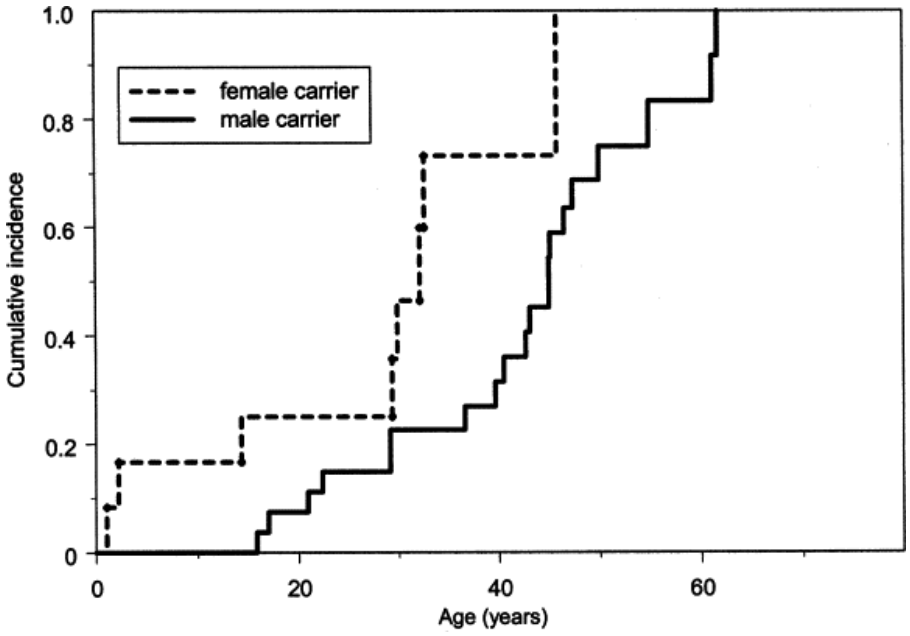


Figure 5 Kaplan-Meier of cumulative frequency of non-sex specific cancer in LFS patients taken from work published by Hwang and colleagues.

27 male and 12 female germline *TP53* mutation carriers from 7 kindreds were evaluated for the age at which they developed their first tumour with regular follow up until last contact was made, or death for those asymptomatic gene carriers. Sex specific cancers such as breast, ovarian and prostate were removed in order to show if sex specific tumours had any effect on the survival curves observed [38].

Hwang and colleagues found that once sex specific tumours were removed, by 20, 30, 40 and 50 years of age, the female carriers had a cumulative frequency risk of 18%, 49%, 77% and 93% respectively in comparison to a cumulative frequency risk in males of 10%, 21%, 33% and 68%. This risk was a lot lower in non-carrier control group 0.7%, 1.0%, 2.2% and 5.1% [38]. The group showed that using a Cox's proportional hazard model (fig. 5) that women had a 2.5-fold higher cancer risk than men, in addition to having an 8% increased cumulative frequency risk of childhood cancers (<20 years of age) [38]. The group concluded that the predicted average age of onset of LFS associated tumours was 29 years of age in women compared to 40 years of age for men [38]. This contradicts findings previously stated by Chompret and colleagues that to a large extent this difference in cancer incidence was because of a high prevalence of premenopausal breast cancer [31, 38]. However the fact that 21% (14/67) of malignancies identified in the Hwang study were breast carcinoma - a malignancy of female LFS patients only - shows the high penetrance of this tumour type in germline *TP53* carriers.

1.2.4 Genetic instability and other genetic modifiers in LFS patients

Despite the majority of LFS families harbouring germline *TP53* mutations, the actual clinical phenotype differs for each person. Even in the same family containing an identical germline *TP53* mutation, this early onset cancer syndrome differs in the time of onset, location, severity of disease and prognosis, suggestive of underlying modifier genes [39]. One polymorphism that has been suggested to play a role in LFS associated tumours is the MDM2-SNP309 polymorphism. This single nucleotide polymorphism (SNP) is located in the E3 ubquitin ligase and negative regulator of p53, murine double minute 2 (*MDM2*). This specific T>G variant lies within the gene promoter region, creating an improved binding site for the transcription factor Sp1 and consequently leading to higher levels of MDM2 mRNA expression [40]. This SNP was therefore shown to attenuate the p53 pathway increasing the rate of degradation of wild type p53. Bond et al. reported that *TP53* carriers with the G-allele developed tumours on average 7 or 10 years earlier than those patients that were homozygous for the T-allele of SNP309 [40]. The group hypothesised that as *TP53* germline carriers have only one copy of wild type functional p53, patients with the G-allele of SNP309 had a severely compromised p53 tumour-suppressing pathway [40]. Consequently resulting in higher mutation accumulation, poor DNA repair and earlier tumour development [40, 41]. Work published by Bougeard and colleagues supported previous findings by Bond et al. investigating the role of the MDM2-SNP309 T>G variant on tumour onset [42]. Additionally, they investigated the p53 codon 72 polymorphism - shown to have a higher binding affinity to MDM2 - in 61 French germline *TP53* patients to assess their contributing risk on age of tumour onset [40, 42, 43]. They found that the MDM2-SNP309 G allele with the Arg coding allele in p53 codon 72, were associated with earlier tumour onset [40, 42]. This was further confirmed by a Dutch and Finnish group [44].

In a follow up study by Bond and colleagues, this MDM2-SNP309 was examined further in LFS patients in a gender-specific and hormone-dependent manner, particularly in the context of oestrogen [45]. The group investigated this firstly because previous studies had shown that oestrogen signalling up regulated the levels of MDM2 expression in breast cancers expressing the oestrogen receptor (ER) [46-50]. Secondly, that the MDM2-SNP309 resides in the promoter region are responsible for hormone signalling pathways [51]. Additionally, the transcription factor Sp1 is well described in the literature as a transcriptional activator for various hormone receptors including the ER [52, 53]. Their hypothesis was that this SNP was associated with an earlier onset of tumourogenesis via hormone signalling in sex-specific tumours [45]. Their findings suggested that the G allele was only associated with an earlier age of tumour onset, in ER+ invasive ductal breast carcinomas [45].

With the heterogeneity of LFS in mind, Shlien et al. performed a genome-wide study characterizing the constitutional genetic variation present in LFS families and the abundance of DNA copy number variations (CNVs) [54]. These CNVs are segments of DNA at least 1 kb in length and are

present throughout the genome with deleted or duplicated regions being implicated in various diseases [54, 55]. What Shlien and colleagues discovered was that a germline *TP53* mutation increased the CNV frequency by around 3-fold [54]. They hypothesised that the high frequency of CNVs and genetic instability in these LFS patients provide an ideal base for larger somatic chromosomal abnormalities including deletions and duplications. This consequently provides the perfect genetic environment for the development of cancer [54].

All of these studies show the complexity of the relationships between the various genetic elements taking place in each of these LFS patients. It has become clear that there is far more to the genetic makeup of these patients than a germline *TP53* mutation and that contributes to the unique phenotype of an individual including the spectrum of tumours, the age of onset and the severity of the disease.

1.2.5 R337H: a unique case in Brazil

Despite precise *TP53* mutations driving no specific tumour type, there have been cases in Brazil where a mutation at R337H has been associated with a high incidence of adrenocortical carcinoma (ACC) in children with no family history [56]. From the 36 Southern Brazil ACC patients that were tested, 97% were shown to harbour this specific mutation. In this particular area of Brazil there is an incidence of the R337H *TP53* mutation 10-15 times higher than that in the United States.

This is a unique case only reported in South East Brazil other than one family in Portugal (with Brazilian ancestry) and a single patient with Portuguese ancestry in France [31, 57, 58]. This is believed to be due to a founder affect originating in Brazil [59, 60]. This mutation was tested to see if this was a common polymorphism in this specific region of Brazil where it was shown to be of low penetrance. The R337H mutation was therefore 10-20 times higher in Southern Brazil compared to other *TP53* mutations associated with LFS cancers [56, 61]. When patients from this area with a family history of LFS or LFL were investigated, 46.1% had a confirmed germline R337H *TP53* mutation [62]. Leading on from these studies, a neonatal screening program was initiated in which 171,649 newborns were screened in the state of Paraná Southeast Brazil, where a R337H mutation was confirmed in 0.27% (n = 461) of cases [63]. These carriers presented with a wide array of tumours including breast cancers, brain cancers, soft tissue sarcomas and adrenocortical carcinomas confirming that this mutation is not only responsible for the high incidence of adrenocortical carcinoma presented in children from the original report [61, 62]. The R337H *TP53* mutation is located in the tetramerization domain and was shown to have a different pattern of associated cancers with the frequency of ACC significantly higher than tumours presenting in patients with a mutation in the DNA binding domain. The International Agency for Research on Cancer (IARC) *TP53* database described that 65% (78 cases) of the tumour distribution for the R337H mutation were reported in the adrenal gland [64].

This unique mutation despite structurally being very similar to the wild type, differs in a pH sensitive manner which reduces the stability of this protein [65]. DiGiammarino and colleagues showed that in a high physiological pH environment; this histidine residue becomes deprotonated leading to loss of the stabilizing interhelix salt bridge [65]. This link with paediatric ACC can be attributed to the elevated pH within the adrenal gland and the extensive cellular changes which take place during both pre and post natal development via apoptosis [66-68].

1.3 The tumour suppressor gene: *TP53*

1.3.1 *TP53* guardian of the genome

The *TP53* gene located on chromosome 17p13.1 encodes the p53 protein which is also widely known as the "guardian of the genome" [69]. This essential tumour suppressor is involved in many pathways and regulatory processes implicated during prevention of cancer including cell cycle arrest, apoptosis, DNA repair, angiogenesis, metabolism and senescence in response to a number of genotoxic stresses and DNA damage (see fig. 6). The significance and clinical relevance of fully functioning wild type (WT) p53 in the prevention of cancer is exemplified by the fact that p53 mutations are present in more than fifty percent of all cancers and twenty-eight percent of breast cancers [70, 71]. Furthermore it has been suggested that even in tumours with WT p53, function is often compromised due to a fault in a regulator, for example MDM2, or a different regulatory mechanism [72].

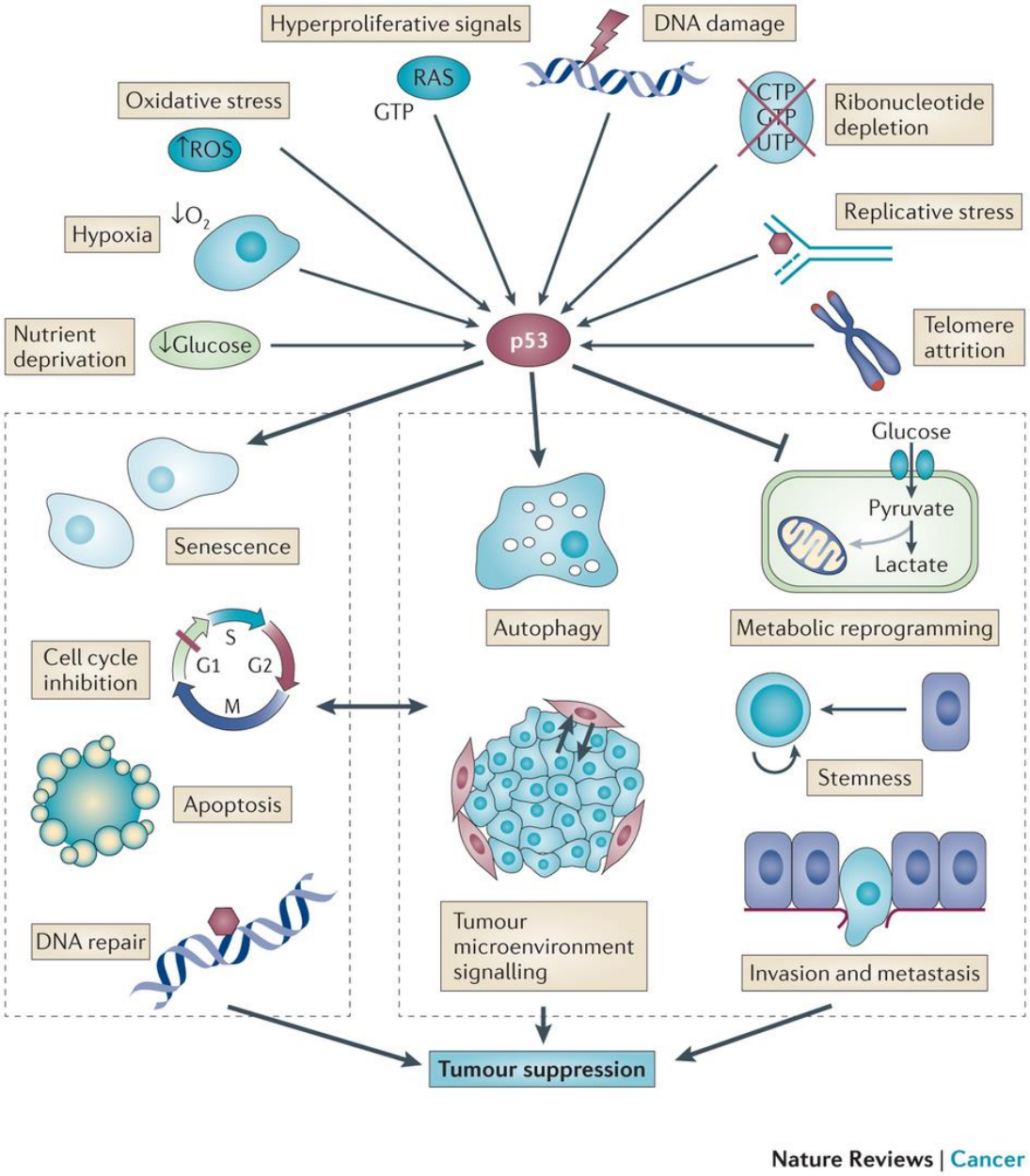


Figure 6 The many faces of p53

p53 is activated through various cellular stresses and drives many cellular responses including cell cycle inhibition, apoptosis, DNA repair, senescence, invasion and metastasis. Image taken from Nature Reviews [73].

1.3.2 The role of *TP53* in Li-Fraumeni Syndrome (LFS)

In 1990 it was discovered that in the majority of LFS cases the underlying genetic defect and cause for disease was as a result of germline mutations in the *TP53* tumour suppressor gene [29]. Under the classic definition around 70% of families harboured a germline *TP53* mutation (see section 1.2.2). Furthermore the observation that not all classic LFS families have a germline *TP53* mutation shows that the mutational screening may have disregarded additional alterations not present within the coding region but nevertheless fundamental to the regulation of the protein.

Alternatively another gene altogether may be responsible for the phenotype seen in some families. There have been cases of LFS initially attributed to germline mutations in the *CHEK2* gene but which were subsequently shown to be because of germline *BRCA2* mutations and it is now believed that the *CHEK2* checkpoint kinase, is not a major predisposing gene in this disease [74-76]. Although *CHEK2* mutations result in varying cancer types, these cancer types differ from those described in the clinical definition of LFS. Apart from breast, a large proportion of *CHEK2* related cancers were not common in typical LFS affected families including: prostate, colon, kidney and thyroid [76]. Soda et al. believe that *CHEK2* mutations have been present in LFS families because some of the variants seen may be breast cancer susceptible alleles [75].

1.3.3 The domains of the p53 protein

The p53 protein is comprised of 393 amino acid residues which fall into five functional domains: transactivation (1-50), proline rich (63-97), DNA binding (102-292), tetramerization (323-356) and a negative regulation domain (363-393) [77]. p53 is regulated largely through the amino and carboxyl termini of the protein through a variety of posttranslational modifications including acetylation, phosphorylation and ubiquitination. However, it is this central DNA binding domain and its interactions with downstream targets which activate and stimulate the various pathways p53 is associated with. The majority of somatic and germline *TP53* mutations are located in this domain critical to the activation of downstream targets described in the International Agency for Research on Cancer (IARC) *TP53* database.

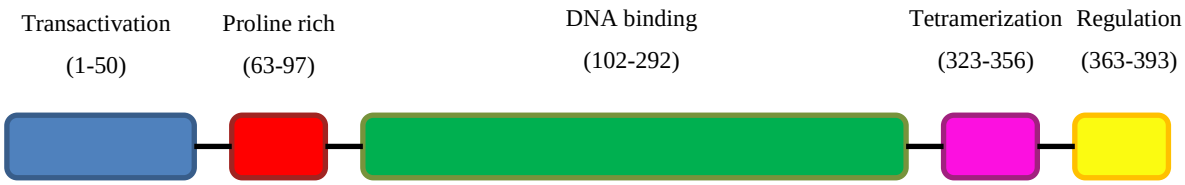


Figure 7 The domains of the p53 protein.

The p53 protein has five domains including the transactivation (1-50), proline rich (63-97), DNA binding (102-292), tetramerization (323-356) and a negative regulation domain (363-393) comprising a total of 393 amino acid residues. Adapted from Meek et al. [78].

1.4 The COPE Pilot study: A young breast cancer cohort with a germline *TP53* mutation

The pilot study was initiated in Southampton, investigating early onset breast cancer as a result of a germline *TP53* mutation [79]. Two female patient cohorts were recruited for this study. Group 1 containing 9 patients diagnosed with LFS associated with a pathogenic *TP53* gene mutation identified through one regional genetics service. 5/8 family pedigrees enclosed a minimum of one family member diagnosed with breast cancer where formalin fixed paraffin embedded (FFPE) material of the tissue was available. Immunohistochemistry (IHC) on whole tumour sections for ER, PR and HER2 were performed as part of the study with additional fluorescent *in situ* hybridization (FISH) for borderline HER2 IHC cases (2+). Patient group 2 consisted of 234 cases diagnosed ≤30 years of age that were recruited as part of the Prospective Study of Outcomes in Sporadic versus Hereditary Breast Cancer study (POSH) [80]. 216 of those cases had blood DNA available for germline *TP53* testing and were taken forward for the pilot study as a control group. IHC data was available for 231 of POSH patients where cases had been systematically stained and scored using tissue microarrays (TMAs) for receptor status.

12 tumour cases from group 1 had their pathological characteristics including grade and receptor status described locally and were compared to the carefully selected, similarly aged comparison group 2 from the POSH cohort where data were made available from the central review POSH steering group. The aim of this pilot study was to determine if there were specific pathological characteristics associated with breast cancers arising in patients harbouring a germline *TP53* mutation. Wilson et al. reported three key factors from this study:

- 83% showed HER2 amplification compared to 16% of young onset breast cancer cases from the POSH cohort.
- Large amounts of high grade DCIS.
- 67% of the *TP53* mutations were truncating.

HER2 amplification is often associated with a high chance of early relapse and a poor prognosis. This finding of HER2 amplification in these cases suggests that breast cancer developing in a patient with an inherited *TP53* mutation, is highly likely to present with a tumour displaying amplification of HER2. Furthermore the high incidence of truncating mutations was notable in contrast to the dominant missense mutations reported in the p53 IARC database. A lot of the data compiled from the pilot study are based on genetics reports as well as a morphology review investigating receptor status: ER (oestrogen receptor), PR (progesterone receptor) and HER2 (human epidermal growth factor 2). Table 2 shows the data from the pilot study.

Patient	Age at onset	ER	PR	HER IHC (FISH)	Grade	*TP53* mutation	Effect
1	35	+	+	3+	3	c.672+1G>T	Truncating
2	26	-	-	2+ (amplified)	3	C.112C>T (p.Q38X)	Truncating
3	24	+	+	3+	3	c.724T>C (p.C242R)	Missense
4	24	-	-	3+	3	c.743G>A (p.R248Q)	Missense
4 (C)	31	-	-	3+	3	As above	
5	28	+	+	3+	3	c.659A>G (p.Y220C)	Missense
5 (C)	28	+	+	2+ (normal range)	3	As above	
6	28	+	-	3+	3	c.625A>T (p.R209X)	Truncating
7	29	-	+	3+	3	c.919+1G>A	Truncating
8	24	+	+	3+	3	c.586C>T (p.R196X)	Truncating
8 (C)	27	+	-	-	2	As above	
9	22	+	+	3+	3	c.437G>A (p.W146X)	Truncating

Table 2 Morphology review of the COPE pilot cohort.

C, contralateral tumour; ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; FISH, fluorescent *in situ* hybridisation [79].

1.5 The Prospective Study of Outcomes in Sporadic and Hereditary Breast Cancer (POSH) cohort: The POSH study

The POSH study recruited 2956 female patients between January 1, 2000 and January 31, 2008 across 127 UK hospitals [81]. Patients were eligible for recruitment if they were diagnosed with an invasive breast cancer ≤40 years of age or if a known germline *BRCA1* or *BRCA2* mutation was present patients were accepted up to ≤50 years of age [80, 81]. For each patient details of the tumour pathology, stage of disease, treatment received and outcome were reported. The primary aim of this large study was to access whether the underlying genomic background affected the prognosis of young breast cancer patients [80].

13 histopathologists from the UK and Australia participated in the POSH pathology review with assessments taking place on either scanned virtual slide images (virtual microscopy) or glass slides (conventional microscopy) [82]. Two pathologists assessing features such as tumour grade and type, reviewed each case independently with agreement more consistent for features such as grade and vascular invasion but poor for more subtle features such as stroma [82]. ER, PR and HER2 receptor status for each patient's invasive disease was determined from the diagnostic pathology test reports [81]. Tissue microarray (TMA) data was obtained for 1336 of cases to confirm diagnostic pathology reports and supplement cases where diagnostic data was unavailable [81].

Following morphological review, the POSH cohort exhibited a median tumour diameter of 22 mm, of which 59% (1735/2956) were classed as grade 3 and 86% (2556/2956) reported as ductal (NST) histological type [81]. When receptor status was investigated using IHC, 66% (1947/2956) were scored as ER+, 45% (1342/2956) were PR+ and 24% (717/2956) of tumours were classed as HER2+ [81]. From those cases that were HER2+, 396 patients were selected to explore breast cancer susceptibility genes in young HER2+ invasive breast tumours [83]. The group found that in young HER2+ patients with no family history, there was a low probability of being a high-risk gene carrier (*BRCA1, BRCA2* and *TP53*) [83].

In addition to investigating overall features of young breast tumour biology such as receptor status on prognosis, such a large cohort has enabled less understood lines of enquiry to be investigated including ethnicity and obesity. Despite all patients recruited to POSH undergoing the same standard and access to health care, data from this large cohort revealed that Black ethnic groups were associated with larger tumour growth compared to White patient groups (26mm vs 22mm) [84]. Additionally, a significantly higher proportion of Black patients presented with triple negative invasive breast cancer and were associated with a significantly poorer 5 year overall survival (OS) compared to White patients [84]. Unsurprisingly, obesity (body mass index (BMI) of ≥30) in young patients was associated with a significantly lower 8 year OS (p=<0.001) when compared to patients of healthy weight [85]. A significant association was described between obesity and larger, high

grade (grade 3) tumour presentation [85]. Lastly, the obese patient group had a higher frequency of triple negative tumours reported when compared to the normal weight patient group (25% vs 18.3%, p=0.001) [85].

1.6 The genetics of breast cancer

It is well established that across all cancer types, somatic mutations are prevalent across the tumour genome. Typically it is an accrual of somatic lesions and increased genomic instability across the genome which eventually leads to the transformation from a benign lesion to carcinoma. Many of the somatic mutations in cancer cells have no effect on the many processes implicated in cancer and are known as "passenger" mutations. A subset of mutations or driver mutations, present the cell with a clonally selective advantage and represent critical steps driving oncogenesis. It is this clonal expansion of multiple generations harbouring and continuous acquisition of further mutations, which eventually leads to the development of cancer. This subset of mutations are key to the progression of cancer and it is the remaining passenger mutations which contribute to the genetic diversity of the disease.

In 2000 the first large scale breast genomics study based on gene expression profiling was published by Perou and colleagues [86]. This work began dissecting the complex genetic changes underpinning breast cancer and until this point, the driver mutations and mutational processes underlying these breast tumours were largely unknown. Since then there have been many other studies in larger cohorts investigating gene expression, somatic mutations and copy number changes present in this complex disease. The work published by Perou in addition to work by subsequent groups used the molecular changes to characterise breast cancer subtypes and through this work, the complexity of breast cancer started to become obvious. It was only during the late 1990's that clinical treatment was reviewed according to the different types of breast cancer and still at this point all breast cancer patients were treated with tamoxifen. Now it is widely accepted that this drug is only beneficial to breast tumours that are ER positive. Around 75% of breast tumours are ER+, 55% PR+ and 20-25% HER2+ [87, 88]. Furthermore around the same time it was determined that the genetic background also had a bearing on the developing tumour subtype, for example 8-16% of triple negative breast cancers can be attributed to *BRCA1* germline mutations [89-91]. This type is relatively less common than the luminal subtypes, even at the younger onset that *BRCA1* gene carriers typically develop breast cancer. Ideally, breast tumour morphology may help to reveal a particular genetic background profile for other susceptibility genes. Currently the triple negative subtype association with a *BRCA1* mutation is used to help select patients for genetic testing [92].

Perou and colleagues believed that to some degree the vast amounts of phenotypic diversity seen throughout breast tumours could in part be due to different patterns of gene expression [86]. They were the first group to use complementary DNA (cDNA) to characterise breast tumours based on gene expression signatures which were compared with gene expression from a pool of already characterised reference cell lines. Here the group studied the expression profile of 65 surgically removed breast tumour specimens from 42 different patients using cDNA microarrays covering 8102 genes. From this study 8 gene clusters were revealed: endothelial, stromal/fibroblast, breast basal epithelial, B cell, adipose-enriched/normal breast, macrophage, T-cell and a breast luminal epithelial cell gene cluster which they then clustered into 4 distinct subtypes. These were described as ER positive or luminal like, basal like, ErbB2 and normal breast tissue. Prior to this work, ER negative tumours were believed to represent a homogeneous entity. This early work molecularly characterising breast tumours, was the first attempt to reveal the complexity of breast cancer subtypes with ER negative tumours falling into basal-like and ERBB2 positive subtypes as two distinct diseases.

1.6.1 Early onset breast cancer and germline mutations

As previously stated 20% of reported breast cancer is observed in younger patients with often a more aggressive type of cancer associated with poorer prognosis [93]. These younger premenopausal women are more likely to be carriers of genetic abnormalities and have a family history of young onset breast cancer. Two genes closely linked to early onset breast cancer are the *BRCA1* (**br**east **ca**ncer susceptibility gene **1**) and *BRCA2* (**br**east **ca**ncer susceptibility gene **2**) genes both involved in DNA repair [94, 95]. In 1990 *BRCA1* was discovered to reside on chromosome 17q21 with the gene sequence determined in 1994 [95, 96]. Like *BRCA1*, *BRCA2* was localised to its chromosome 13q12-q13 in 1994 and then the gene was isolated as well as the protein in 1995 [94, 97]. Women and men who are high risk gene carriers have a higher risk of developing breast cancer at a much earlier age, which is why the POSH study (Prospective study of Outcomes in Sporadic versus Hereditary breast cancer) was set up, ascertaining women below age 41 with breast cancer to enrich the cohort for poor prognosis and genetic susceptibility [81].

1.6.2 Curtis: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups

The diversity of breast tumours is extensive with several approaches providing alternative means to classify these cancers including molecular pathology, histology, genetic and gene expression analysis [98]. Using gene expression analysis, there are currently five broad molecular subtypes described: basal-like, ErbB2 enriched, normal-like and luminal types A and B [86, 99]. Twelve years after the initial subtypes were described by Perou et al., these groups were further defined into ten subgroups when the genomic and transcriptomic architecture of two thousand breast

tumours was investigated by Curtis et al [100]. Each of these 10 subgroups is associated with a distinct copy number profile as a consequence of a specific set of genes deregulated driving tumourgenesis. The frequency and position of somatic copy number aberrations (CNAs) were assessed across the genome. A gain of copy number was identified as a red region and a loss of copy number was recognised as a blue region in the frequency plot (see fig. 8).

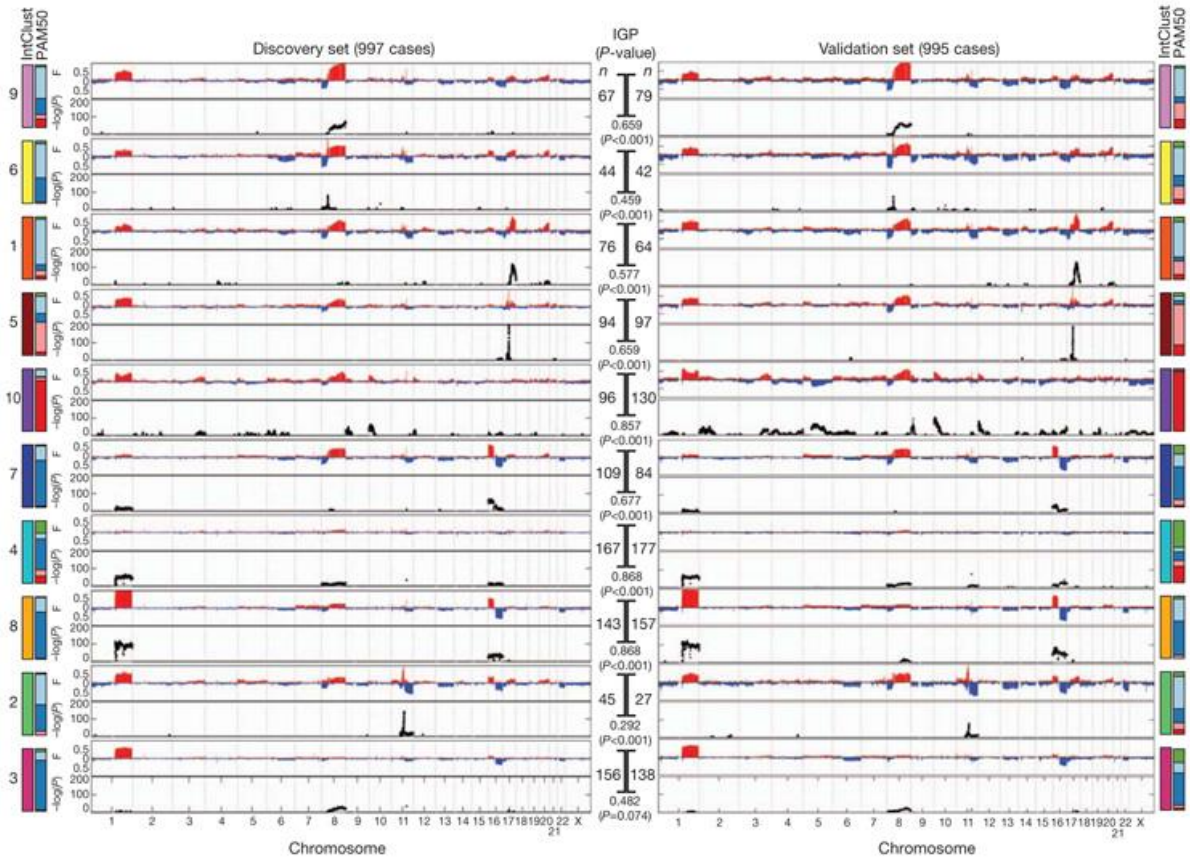


Figure 8 Ten integrative molecular subgroups of breast tumours.

10 distinct subgroups of breast cancer emerged based on the copy number profile obtained from 2000 breast tumours. A red region indicates a loss of copy number and a blue region specifies a gain of copy number. The bar to the left shows the intrinsic group as well as the PAM50 subgroups (dark blue: luminal A, light blue: luminal B, red: basal, pink: $HER2^{+}$ and green: normal) [100].

From analysis into the genomic landscape in 2000 breast tumours, Curtis et al. observed copy number variants (CNVs), single nucleotide polymorphisms (SNPs) and copy number aberrations (CNAs), were associated with atypical expression of around 40% of the genes investigated [100]. In comparison to the earlier Perou study, the advance in technology over the intervening decade is clear. The more recent study used larger numbers of cancers than Perou and they were able to use germline data for reference rather than cell lines to determine which variants were somatic mutations in their integrated analysis.

The majority of the atypical gene expression observed was made up by an array of both *cis*- and *trans*- acting CNAs. A *cis* acting genetic variant refers to a variant at a locus that affects the expression of the same gene. Whereas a *trans* acting variant has an effect on the expression of genes located at a different position in the genome. This finding was hardly surprising with somatically acquired CNAs already having been suggested to be a dominant feature in sporadic breast cancer however, the initial driver events in early tumorigenesis is largely an unknown territory. These difficulties are largely due to these early driver events occurring in conjunction with inherited CNVs as well as random non-pathogenic passenger alterations [101, 102]. Somatically mutated recessive tumour suppressor genes are particularly difficult to define with the majority of these mutations primarily detected in large homozygous deletions (HDs), often areas of the genome containing fragile sites [102]. These fragile sites are highly susceptible to agents such as replicative stress which result in chromosomal breakage in normal cells [103].

The major benefit of establishing these subgroups allows predictions to be made on successful treatments and prognosis. As well as examining the molecular architecture of two thousand tumours, work has continued to look at the breast cancer specific fifteen year survival and the different clinical outcomes of these ten subgroups. From plotting a Kaplan-Meier plot, patients with the poorest prognosis typically fall into clusters five or two whereas patients with a much better prognosis seem to reside in clusters three or four (fig.9).

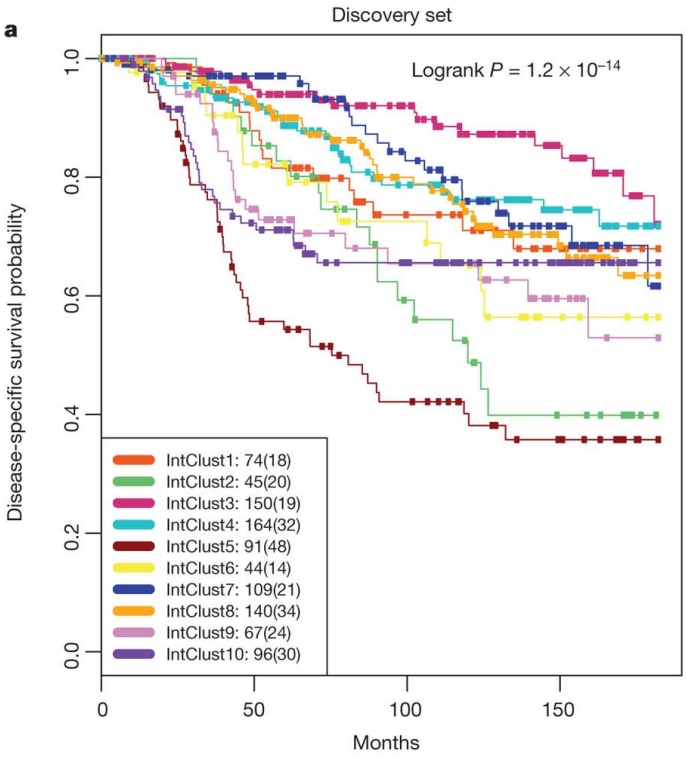


Figure 9 Kaplan-Meier plot of the ten subgroups with distinct clinical outcomes.

Follow up was continued up to 15 years and the disease specific survival was plotted for each of the subgroups. This Kaplan-Meier plot suggests that clusters 5 and 2 have the worse prognosis whereas clusters 3 and 4 have a much more favourable prognosis [100].

Cluster 2 was associated with the 11q13/14 amplicon resulting in ER-positive cis-acting luminal tumours. So instead of being associated with a specific oncogene (for example cluster 5's association with chromosome 17 and loss of *TP53*), Curtis and colleagues believed that this subgroup was driven by a cassette of genes. When cluster 4 was examined further, it became apparent why this subgroup had a particularly good prognosis. This CNA- devoid subgroup exhibited a strong immune and inflammatory response. These breast cancers encoded a *trans*-acting deletion hotspot localised to the TRG and TRA loci which as a result, was associated with an adaptive immune response module consequently leading to severe lymphocytic infiltration. Curtis and colleagues suggest that the presence of these mature T lymphocytes (containing a rearranged TCR locus) give rise to an immunological response to the cancer. Cluster 3 was also portrayed as a predictor of good prognosis predominantly containing luminal A cases with a copy number landscape containing very low genomic instability.

1.6.3 The Cancer Genome Atlas Network: Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network undertook a wider and less limiting approach in characterising breast tumours by analysing tumour samples over six different platforms [104]. From the 825 patients that were recruited for the study, their material was used to investigate copy number, DNA

methylation, gene expression, reverse-phase protein expression, exome sequencing and microRNA sequencing. Supporting work from previous studies, the tumours showed a substantial amount of molecular heterogeneity within the cohort and suggested the presence of four main subgroups of breast cancer. When the somatic mutations present in these tumours were examined, only 3 genes seemed to be recurrent and present in over 10% of cases: *TP53*, *PIK3CA* and *GATA3*. When individual subtypes were investigated, they did find specific mutations particularly in luminal A type tumours in specific genes such as *GATA3*, *MAP3K1* and *PIK3CA*. Furthermore, through the protein expression approach, the group were able to determine that specific pathways were involved in each of these subgroups as well as two unique expression profiles which they believed to be because of interactions taking place in the surrounding stroma.

From the somatic mutational studies ten novel significantly mutated genes were discovered in the cancers including *NF1*, *RUNX1*, *AFF2*, *TBX3*, *CCND3*, *PIK3R1*, *PTPN22*, *PTPRD*, *SF3B1* and *CBFB*. As well as somatic mutations in novel genes, the group also identified the majority of previously implicated breast cancer genes: *TP53*, *AKT1*, *GATA3*, *CDH1*, *PIK3CA*, *PTEN*, *RB1*, *MAP3K1*, *MLL3* and *CDKN1B*. Overall basal and HER2 overexpressing tumours harboured the highest mutation rate but luminal A and B had a much greater diversity of mutations. HER2 tumours notably had a smaller degree of mutational diversity with HER2 amplification (80%) and mutations in *TP53* (72%) and *PIK3CA* (39%) contributing enormously to the mutational spectrum and much lower frequencies of additional driver genes. Furthermore, the types of *TP53* mutations present in the tumours differed for each subgroup with luminal A and B containing mainly missense mutations and basal tumours were more susceptible to frame shift and nonsense mutations.

Using these different platforms, a profile was determined for each of the four subtypes: luminal A, luminal B, basal and HER2 amplified tumours. The luminal tumours showed the highest level of heterogeneity both in expression profile and the number of mutations compared to other types of breast tumours. In particular these subtypes contained a high number of *PIK3CA* mutations however when the other platforms were used this was not implicated in the activation of the PI3K pathway. Furthermore there was a large proportion of *MAP3K1* and *MAP2K4* mutations and these were shown by Mutual Exclusivity Modules in cancer (MEMo) analysis to activate p38-JNK1 pathway. This type of analysis is utilised in large cancer cohorts to investigate how genomic lesions converge onto similar biochemical pathways [105]. The p53 pathway was mostly intact in these tumours with few *TP53* mutations, low levels of *ATM* loss and *MDM2* amplification.

The HER2 positive tumours studied by the Cancer Genome Atlas Network revealed an added level of complexity suggestive of two clinically distinct HER2+ phenotypes. Not all of these HER2 tumours had overexpression of the HER2 amplicon associated genes that make up the HER2E mRNA expression profile. Furthermore some tumours that were not clinically described as HER2+ were shown to overexpress the HER2E mRNA category. The group discovered that only 50% of

those clinically HER2+ tumours had the HER2E mRNA expression profile with the remaining 50% much more closely resembling the profile of a luminal tumour. Tumours fitting the HER2E overexpression profile also revealed a higher expression of other receptor tyrosine kinases, a higher incidence of *TP53* mutations and were frequently ER-. In contrast those tumours that did not fulfill the HER2E overexpression profile, contained *GATA3* mutations and were predominately ER+. Overall this latter group presented a high incidence of *PIK3CA* mutations (39%), deletions of *PTEN* and *INPP4B* and a reduced overall number of mutations in *PTEN* and *PIK3R1*. Those HER2+ overexpressing the HER2E group of genes had a higher number of mutations and DNA amplifications of FGFRs, EGFR, cyclin D and CDK4 as well as greater genetic instability due to an increased proportion of aneuploidy.

The basal or triple negative breast cancer (TNBC) subtype showed the highest incidence of aberrant p53 signalling with 84% of tumours harbouring a *TP53* mutation. Furthermore, through MEMo analysis the group discovered that inactivation of *RB1* and *BRCA1* were also associated with this group in addition to having the highest PI3K/AKT pathway activation. After *TP53*, *PIK3CA* was the second most mutated gene in around 9% of tumours but the authors have suggested that the activation of this pathway in this subset of tumours, could also be because of amplification of genes involved in the pathway (*PIK3CA* (49%), *KRAS* (32%), *BRAF* (30%) and *EGFR* (23%)), or deletion of *PTEN* and *INPP4B*. Additionally, these tumours expressed keratins 5, 6 and 17 as well as high expression of genes such as *FOXM1* consistent with an enhanced cell proliferation signature. Using a PARADIGM analysis, hypoxia induction was also identified in these tumours through activation of the HIF1α/ARNT pathway.

This work has utilised an array of different technologies to significantly enhance current knowledge and provide a more detailed report of breast tumour heterogeneity.

1.6.4 Silwal-Pandit: *TP53* mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance

Silwal-Pandit and colleagues assessed the *TP53* status and prognostic significance of 1420 breast tumour which were separated by their PAM50 subtype and integrative clusters [70]. The tumour samples were from the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) cohort and using Sanger sequencing, had all the coding regions of *TP53* investigated. Overall the group found that 28.3% of these tumours had a mutation associated with a worse survival and was an independent indicator in ER+ breast tumours of poor prognosis. Furthermore, there was significant variation in the spectrum of *TP53* mutations between the subtypes including subtype specific modifications. For example basal-like and HER2-enriched tumours showed an enrichment of truncating mutations. In the luminal B, HER2+, and normal-like tumours an increased mortality was associated with any somatic *TP53* mutation. However this was not the case in luminal A and basal tumours where there was no significant affect. When these tumours were

arranged into their respective integrative clusters, comparable findings were made from groups IC1, IC4 and IC5 with patients in these clusters associated with an increased mortality (see fig. 8 for integrative clusters). Furthermore, the additional effect of loss of heterozygosity (LOH) of WT *TP53* and the amplification of the p53 negative regulator *MDM2* has a collective effect leading to a higher mortality.

When the types of *TP53* mutations were reviewed, a large proportion of these were single base substitution mutations (73.4%) with small deletions (18.7%), insertions (5.2%), complex (2.0%) and tandem mutations (0.7%) making up a smaller proportion of the mutation spectrum. Removing the 8 coding silent substitution mutations, the majority of the remaining mutations were G:C>A:T transitions (49.5%) occurring at CpG sites rather than A:T>T:A transversions (4.5%) which were present in the cohort far less frequently. From the five PAM50 subtypes, basal tumours were shown to have the highest proportion of these G:C>A:T transitions. In common with the data reported in the (IARC) *TP53* database, Silwal-Pandit and colleagues found that the majority of the mutations (81%) were located in the DNA-binding domain through exons 5-8, with exons 4 (9.6%) and 10 (6.5%) also contributing to a smaller degree. Mutational hotspots were present at codons 175, 179, 196, 213, 245, 248, 273, 278, 285 and 306 with hotspot codons 175, 245, 248 and 273 also reported in the IARC *TP53* database. Furthermore frameshift mutations differed from the remaining types of mutations (missense, nonsense, inframe, and splice mutations) that were mostly located in the DNA binding domain, whereas these frameshift mutations did not have any hotspot regions but were widespread throughout the gene.

The number of mutations varied from each intrinsic subgroup with basal (65%) and HER2+ (53%) subtypes containing the highest proportion of *TP53* mutations. Luminal B (25%), normal (11%) and luminal A (9.3%) had a much smaller proportion of tumours with a *TP53* mutation. Additionally 17.7% of ER+ and 15.6% of PR+ tumours had a *TP53* somatic mutation. Basal and HER2+ subtypes showed a significant proportion of mutations which were not missense and this can be compared to the luminal B group, where the majority of these mutations were DNA binding missense mutations. In respect to hotspots, luminal A tumours were shown to have a flat profile as opposed to hotspots whereas basal tumours had multiple hotspots present throughout the cohort. Different hotspots were present in the different PAM50 groups with hotspots present in basal tumours residing at codons 175 and 273, and the hotspot at codon 245 being prominent in HER2+, luminal B and basal tumours. Additionally, the specific nonsense R213* mutation was a hotspot in mostly tumours of basal origin. When these mutations were then arranged into the 10 integrative clusters, the spectrum of mutations demonstrated further variation with the 76.5% of the IC10 mainly basal group containing a *TP53* mutation. This can be compared to only 6.3% of IC3 tumours (majority luminal A), 53.6% of IC5 (mainly HER2+ tumours) and 48% of IC9 (mostly ER+/HER2-).

It is widely accepted that *TP53* mutations are linked with genomic instability and those tumours in the cohort which did have a *TP53* mutation were shown to have a significantly higher rate of genomic instability index (GII), particularly in the basal and HER2+ group [106]. Additionally the group discovered that there was a significantly higher incidence of *MDM2* amplification and *TP53* LOH (80.8%) in tumours with a *TP53* mutation, independent of the type of mutation. Once these tumours were arranged into their PAM50 subtypes, LOH was detected in 80% of IC1 but only 35% of IC10 tumours. Increased mortality was associated with *MDM2* amplification and/or *TP53* LOH in tumours that have lost their wild type *TP53*; however in tumours with mutant *TP53*, combined genomic abnormalities gave an increasingly worse prognosis with the addition of each mutational defect manifesting as an increased level of genomic instability.

This group found that basal breast tumours typically were enriched for frameshift and nonsense mutations and were not prognostic. In another study the *TP53* pathway was shown to be deregulated in most basal tumours but not necessarily through direct mutation of *TP53* [104]. Silwal-Pandit and colleagues have suggested that a compromised *TP53* pathway is required in the majority of basal tumours but this can be through genomic convergence onto this pathway as *TP53* mutation alone was not prognostic. This could indicate that the biologically different HER2+ tumours are losing p53 function solely through the loss of *TP53*. For those HER2+ tumours that are losing *TP53*, this was a marker of a poor prognosis suggesting that a combination of this genomic background is complementary and giving those clones a selective advantage.

In addition to altered gene expression and large-scale somatic variation analysis, it has also been demonstrated that there is a complex landscape of genomic rearrangements and fusion data [107]. Sequence data is just one dimension of multidimensional mutational landscape in a dynamic and rapidly moving area of technology and research. In order to truly understand the complex genomic landscape of breast cancer all of these approaches need to be taken into consideration.

1.7 Cancer evolution

It is well established as to how significant somatic mutations are in underpinning the progression of cancer and in the majority of solid tumours these genomic aberrations are present in their thousands across the genome [108, 109]. An essential step in determining tumour phenotype is Darwinian selection of driver events and specific clones [110]. Following selection of a clone, this then can expand and evolve either via a linear or branching evolutionally route [110, 111]. A linear route is identified through clonal development through expansion and accumulation of genomic lesions from the original clone leading to a lower level of tumour heterogeneity. In comparison branching evolution leads to greater tumour heterogeneity through the development of an array of clonally distant unique clonal populations (fig. 10).

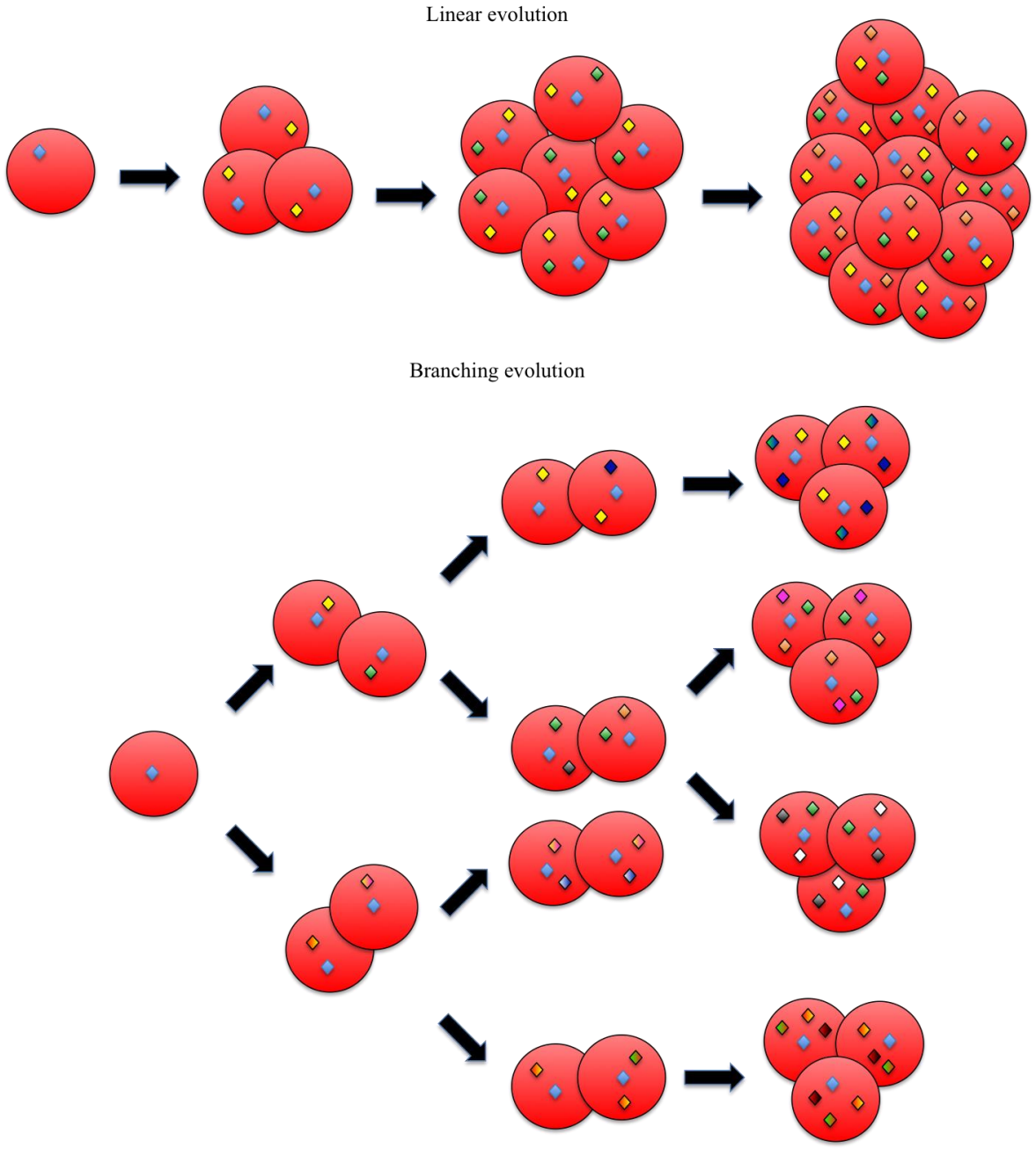


Figure 10 Linear and branching cancer evolution

Tumours can evolve via two routes. Different coloured diamonds represent a specific mutation. Mutations accumulate through clonal generations.

The majority of published work investigating tumour evolution has been in blood cancers. Investigating tumour evolution is significantly easier in blood malignancies as samples can be obtained at various time points allowing genomic heterogeneity and the order of molecular events to be tracked with greater ease. As this is not possible in solid tumours, work in breast cancer has involved sequencing geographically distinct areas of tumour to interpret heterogeneity and early

genomic events via the variant allele frequency (VAF) and presence across various areas of the tumour.

Ortmann and colleagues were able to identify the significance of the order of molecular events in 48 myoproliferative neoplasms [112]. In this study they investigated the clinical significance of the order in which patients acquired a Janus Kinase 2 (*JAK2*) and *TET2* mutation. They showed that the order in which these genomic lesions were acquired affected response to therapy, age of presentation, clonal evolution of the neoplasm and the biology of stem and progenitor cells. In patients that acquired a *TET2* mutation first on average presented 10.46 years later in life with smaller homozygous subclones compared to that of a *JAK2* mutation. The group found that JAK2 first patients had a higher risk of thrombotic events and polycythemia vera. Their data showed the significance the order of genomic lesions can make in oncogenesis biology and giving select clones a growth advantage [112].

As previously discussed there have been many large-scale genomics studies in breast cancer but few studies have specifically investigated the cancer evolution and the order in which genomic lesions are acquired to lead to a specific phenotype. Yates et al. found that there is no strict order of genomic acquisition in breast cancers, going on to state that tumours are typically very clonal and diverse with a branching evolution [113]. Others groups have investigated the order of genomic lesions in breast cancer and reported similar genomic lesions in matched DCIS and invasive lesions [114]. This has been found on the transcriptomic level and across copy number aberrations (CNAs) [115]. CNAs that have been linked with disease progression from DCIS include *MYC, FGFR1* and *CCND1* [116-118].

Many groups have investigated the genomic landscape of DCIS and tumour samples using array based comparative genomic hybridization (aCGH) to look for chromosomal abnormalities. They found that in the majority of cases the invasive tissue showed very little differences compared to its matched DCIS lesions [119-121]. Additionally a similar finding was made when copy number was investigated, suggesting similar copy number changes between the DCIS lesions and invasive tissue. [121-125]. In fact Porter and colleagues found that the largest changes can be identified from the progression of normal tissue to DCIS [122]. Johnson et al. found that even once disease had progressed to become invasive, the DCIS lesions continued to evolve in parallel with the invasive ductal carcinoma (IDC) harbouring additional genomic lesions [123].

One CNA event that is important in driving an aggressive breast tumour subtype is *HER2* amplification. Various groups have found that HER2 overexpression is a common feature in DCIS, particularly in high grade lesions, but only a small number of invasive ductal carcinomas retain this feature [118, 126-128]. This suggests one of two things: that *HER2* amplification is lost during invasive progression or that the invasion derived from a clone that was negative for *HER2* amplification.

Bringing multiple studies together that have investigated the progression from DCIS to invasive ductal carcinoma, two hypothetical models have been proposed. Model A involves a convergence phenotype in which mutations, epigenetic changes or a combination of both are acquired which give the abnormal DCIS cells an advantage and ability to overcome the myoepithelial barrier [129]. The acquisition of numerous events would also give a possible explanation as to why negative results such as HER2 can be identified in the DCIS but not the invasive tumour. This would suggest that there are potentially many different combinations of genomic lesions which ultimately all lead to invasion. A second model proposed involves an evolutionary bottleneck. This model envisions the accumulation of many genomic lesions resulting in a heterozygous population with distinct clones [130, 131]. This model suggests that a specific subclone must have a selective advantage to become invasive due to an array of aberrations. Groups have found significant heterogeneity in DCIS lesions therefore there is evidence to support this model [132, 133].

Additionally groups have investigated the role germline mutations play in tumour evolution and heterogeneity. Fisher et al. investigated the genomic landscape and timing of genomic lesions in a patient with a germline mutation in the tumour suppressor gene *VHL* [134]. This young patient had developed four clear cell renal cell carcinomas that were removed from both kidneys. Using whole exome sequencing of these tumours, data revealed that they were clonally independent and harboured distinct secondary events despite identical histopathological characterisation. These tumours had few somatic mutations and seemed to converge upon the same PI3k-AKT-mOR signalling pathway following a linear evolutionary route.

1.8 The tumour microenvironment

1.8.1 Components of the tumour microenvironment

In addition to genetic instability, there has been increasingly more evidence to suggest that the intricate tumour-stromal interactions in the tumour microenvironment are essential to driving tumour progression. This complex system includes various cell types including fibroblasts, immune cells and endothelial cells that have been shown to differentially express various proteins which contribute to a dynamic extracellular matrix (ECM). The cross talk between the cellular components and extracellular matrix has been implicated in an array of cellular processes both driving and suppressing tumour progression.

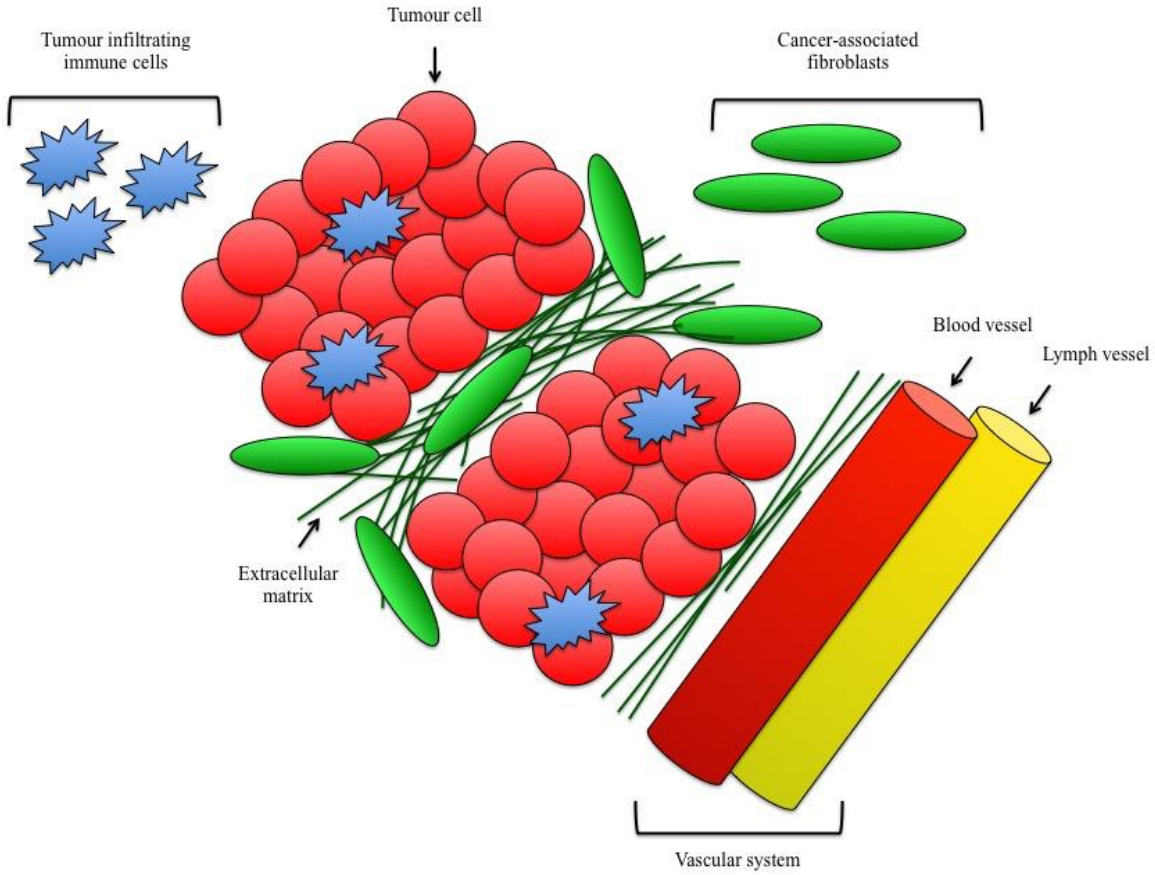


Figure 11 The tumour microenvironment

Components of the tumour microenvironment

Many groups have reported that a high tumour immune infiltration and in particular CD8+ tumour infiltrating lymphocytes (TILs), is often a marker of good prognosis and a better 5 year survival [135-139]. Recently there has been a revival in immunotherapy through research in melanoma where induction of the immune system was shown to reverse tumour progression [140]. On the other hand, tumours that have high incidence of cancer-associated fibroblasts (CAFs) and a reactive ECM have been reported as having a poorer prognosis [141-145]. CAFs are often the most prominent cell type present in the stroma and secrete various components that contribute to the ECM. Through the secretion of cytokines, growth factors, hormones and protease, particularly CAFs positive for alpha smooth muscle actin (α-SMA), have been shown to increase migration, invasion, proliferation, angiogenesis and inhibition of infiltration of lymphocytes through a barrier affect created by expression of collagen by these stromal cells [146-148]. A key pathway in which CAFs activate these processes is through transforming growth factor beta (TGFβ) signalling.

1.8.2 Transforming growth factor beta (TGFβ): The double edged sword

The cytokine transforming growth factor beta (TGFβ) is the key mediator of transforming growth factor beta (TGFβ) signalling. This cytokine can behave as both a tumour suppressor in early disease and a tumour promoter in later disease [149-151]. Tumour suppression is mediated through the activation of the cyclin-dependent kinase inhibitors (p21 and P15 Ink4b) [152]. Key to this switch in function, is the accumulation of mutations thus inhibiting these suppressive features [149]. As a result this cytokine can directly drive an invasive phenotype in the tumour cells and indirectly promote tumour progression via the tumour microenvironment.

In order to initiate this pathway TGFβ firstly needs to be activated. TGFβ is deposited in the extracellular matrix in its inactive latent form. Inactivation is achieved through binding to the latency-associated peptide (LAP) forming the small latent complex (SLC). These proteins then bind to one of four latent TGFβ-binding proteins (LTBP) overall forming the large latent complex (LLC) which is firmly anchored to the ECM via fibrillin-1 [152, 153]. Activation of TGFβ is achieved through mechanical release from this LLC complex and one way in which this can occur is through the integrin αvβ6. Integrin αvβ6 causes a conformational change through induction of mechanical stress on the latent TGF-beta1 complex and this in turn releases TGFβ [154].

Activation of transforming growth factor beta (TGFβ) signalling is initiated through binding of the cytokine transforming growth factor β1 (TGFβ1) to the type 2 TGFβ receptor (TGFBR2) resulting in the recruitment and phosphorylation of TGFBR1 [155]. Activation of this receptor results in a cascade of signalling via phosphorylation of the carboxyl terminal serine residues in SMAD2 or SMAD3 [156, 157]. Once phosphorylated, this SMAD protein oligomerizes with SMAD4 enabling nuclear translocation and binding to the SMAD-binding element inducing gene expression [158]. A simplified summary of this tumour promoting process is shown in fig. 12.

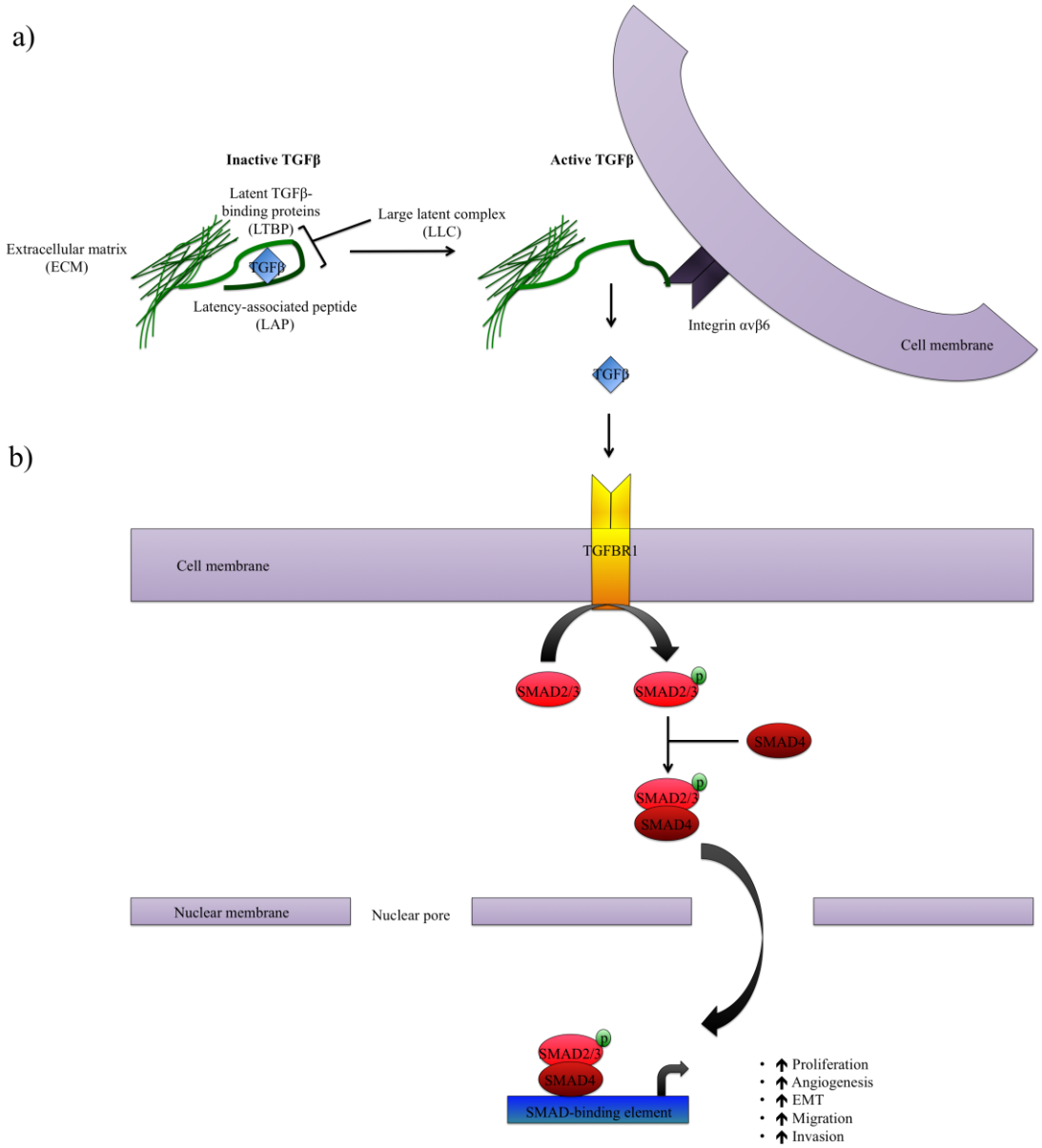


Figure 12 TGFβ activation and signalling

Integrin αvβ6 activates TGFβ driving tumour promoting TGFβ signalling. a) Activation of TGFβ through integrin αvβ6 mediated conformational change and mechanical stress on the latent TGF-β1 complex. b) Release of TGFβ induces phosphorylation cascade and activation of processes such as the epithelial-mesenchymal transition (EMT), migration, invasion, angiogenesis and proliferation. Modified from Pickup et al. and Wipff and Hinz et al. [154, 155]

Fibroblasts typically are involved in wound healing but during oncogenesis they often undergo differentiation through tumour cell interactions and activation of transforming growth factor beta (TGFβ) [159]. As previously stated, a key mechanism in which TGFβ is activated is through expression of integrin αvβ6 on the cell surface of tumour cells [159]. Once released, it is this cytokine that is involved in myofibroblast differentiation [160, 161]. This is typically identified through the formation of these stress fibres recognised via the expression of α-SMA [162]. It is

these myofibroblasts that secrete many of the pro-tumourgenic components to the ECM including collagen, a marker of a sclerotic tumour stroma [163].

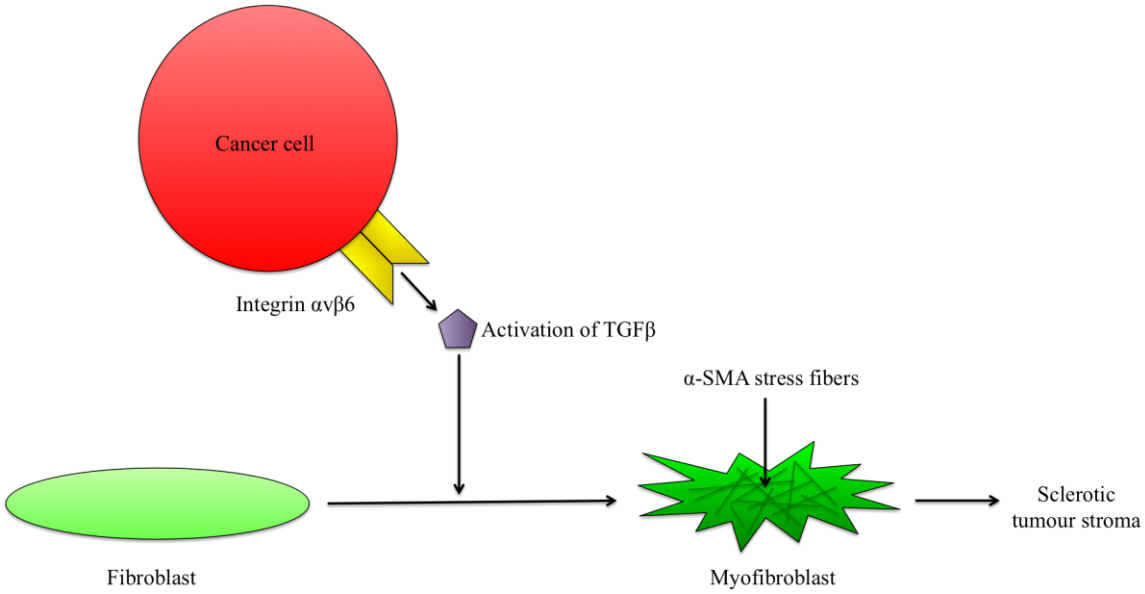


Figure 13 Myofibroblast differentiation through TGFβ activation

Integrin αvβ6 activates TGFβ which drives myofibroblast differentiation.

1.8.3 The role of transforming growth factor beta (TGFβ) signalling in tumour progression

In addition to αvβ6's role in the activation of TGFβ and the tumour microenvironment, work in ductal carcinoma *in situ* (DCIS) progression has demonstrated its role in the cancer evolution from *in situ* disease to invasive ductal carcinoma [164, 165]. Work such as this is important as DCIS is frequently diagnosed through screening programs but only around 50 % of these cases will continue to progress and become invasive [166, 167]. Allen and colleagues found that αvβ6 was not expressed in benign malignancies. In high grade DCIS and DCIS with associated tumour, changes in the myoepithelial cells lining the DCIS underwent a switch from tumour preventing, to tumour promoting and matrix stiffening [165]. They concluded that αvβ6 may be promoting breast disease progression from DCIS to invasive ductal carcinoma. This group continued this work and found that high expression of αvβ6 was present in 15% to 16% of invasive ductal carcinoma across their two cohorts (>2000 women) [168]. In addition, co-expression of αvβ6 and HER2+ revealed the worst prognosis, suggesting cooperation of these proteins and a possible new target for those patients that become trastuzumab resistance [168].

In addition to integrin αvβ6, the p53 protein has been implicated in the expression of collagen genes in CAFs and driving a stromal response. Work published by Ghosh et al. discovered a further role for p53 in the suppression of collagen gene expression (COL1A2) through TGFβ stimulation [169]. From this they suggested that p53 is implicated in the regulation of fibrotic cellular pathways [169]. Murine models of the prostate used a $TgAPT_{121};p53^{+/-}$ model in which mice developed an extensive proliferative stromal reaction which was positive for α-SMA and S100A4, a specific marker for fibroblasts [170]. Tumour cells can also inhibit wild type p53 activation in CAFs through indirect cell contacts and p53 inhibition has been linked with immunosuppression [171, 172]. In comparison, work using a murine model in liver fibrosis and *in vitro* work investigating loss of PTEN function found the opposite affect. Kodama and colleagues found that p53 was actually driving fibrosis with the induction of *CTGF* expression [173]. Loss of PTEN was shown to stimulate Akt, SMAD3 and p53(Ser15) phosphorylation [174]. Despite opposing work, it is clear that p53 has a significant role in the adjacent stromal response in addition to driving oncogenesis in tumour cells.

Lyons et al. used a murine model and the MCF10A cell line - which resembles the breast ducts - to investigate this progression from DCIS to invasion [175]. They showed that the microenvironment was driving cells to transform and form large tumours overexpressing cyclooxygenase-2 (COX-2), which stimulates the deposition of collagen and the formation of a dense stroma. Through the use of non-steroidal anti-inflammatory drugs (NSAIDS), the investigators were able to partially block the formation of fibrillar collagen and overexpression of COX-2, leading to an inhibition of the invasive phenotype. Additionally Hu and colleagues also reported that the tumour promoting effects of fibroblasts are to some extent due to the overexpression of COX-2 in tumour cells, which led to an increase in invasion in the xenograft model of DCIS [176]. Further groups have investigated the role of lysyl oxidases (LOXs), a family of ECM modifying enzymes involved in the crosslinking of collagen, invasion and hypoxia induced metastasis [177, 178]. The role of the tumour microenvironment is becoming increasingly more evident in driving tumourogenesis.

Chapter 2: Methods and Patient cohorts

2.1 Patient and cohort groups

Five groups of female patients from two cohorts were recruited for this study. Group 1 consisted of patients with a confirmed germline *TP53* mutation and malignant breast disease. Patients that presented with pure DCIS were eligible for recruitment. This group of patients were recruited as part of the COPE cohort. Groups 2-5 were recruited as part of the POSH study in which eligibility required diagnosis of invasive breast disease <40 years of ages or <50 years of ages if a known *BRCA1* or *BRCA2* mutation was present [80]. Group 2 were carefully selected from POSH as a control group. Samples were selected for HER2+, matched DCIS and availability of germline data ruling out *BRCA1*, *BRCA2* and *TP53* mutations. Data for groups 3-5 were made available by the POSH steering group [80]. These *BRCA1* carriers, *BRCA2* carriers and young breast cancer (YBC) with no underlying germline mutation, were selected for this study because full morphological review had been completed on these cases as part of the POSH study [82]. For full details of the POSH cohort see section 1.5. The cohorts are summarised in table 3.

We recruited 59 patients to group 1 with a germline *TP53* mutation and after excluding 14 patients due to a lack of obtainable tumour material, 45 patients were taken forward. 136 breast cancer formalin fixed paraffin embedded (FFPE) tumour tissue blocks have been collected from *TP53* gene mutation carriers from across the UK and from international collaborators.

Cohort	**COPE**	**POSH**			
Total no. of patients	45	2956			
Group	1	2	3	4	5
Description	*TP53*	HER2+	*BRCA1*	*BRCA2*	YBC
No. of patients	45	55	60	61	98
Recruitment eligibility	Malignant breast disease, germline *TP53* mutation	HER2+ invasive disease with matched DCIS, diagnosis <40 years of age, no known germline mutation (*BRCA1*, *BRCA2* and *TP53*)	Invasive disease and <40 years of age at diagnosis. <50 years of age if a known *BRCA1* or *BRCA2* mutation is present.		
Morphological review as part of the COPE study*	✓	✓			
Morphological review as part of the POSH study**			✓	✓	✓

Table 3 Patient cohorts and recruitment eligibility

Patient selection and eligibility for groups 1-5: *TP53*, *BRCA1* and *BRCA2* gene carriers, HER2+ and YBC with no underlying germline mutation. *COPE morphological review histopathologists Dr Matthew Sommerlad, Dr Guy Martland and Dr Adrian Bateman. **POSH morphology review described by 13 histopathologists outlined in Shaw et al. [82].

2.2 Morphology and immunohistochemistry (IHC)

2.2.1 Hematoxylin and eosin (H&E) staining

4µm sections were cut using a microtome (Leica) from formalin fixed paraffin embedded (FFPE) breast and lymph node blocks that were then mounted on superfrost+ slides (Thermofisher Scientific). Slides were stained using a Hematoxylin and Eosin (H&E) stain using the automated CoverStainer (Dako). Slides were imaged using the Dotslide (Olympus).

2.2.2 Morphology review

All cases were reviewed independently by 2 histopathologist readers (Dr Matthew Sommerlad and Dr Guy Martland). A variety of pathological features including tumour type and grade, DCIS grade and type, stroma, vascular invasion, lymphocytic infiltration and any benign changes were reported (see appendix 7.1 for further details). For cases where readers reporting disagreed on pathological features, a consensus call was determined by a third consultant breast pathologist (Dr Adrian Bateman) resulting in one consensus report per patient. This was performed for groups 1 and 2.

2.2.3 Tissue microarray (TMA) construction

Tissue microarrays (TMA) were constructed taking 3 invasive cores and 2 DCIS cores where possible from donor blocks. The TMA was mapped using Excel to identify which core belonged to which patient and block. This information was inserted into the TMADesigner2 software (Alphelys) with the recipient block and initial donor block introduced into the tissue arrayer minicore 3 (Alphelys). These 1mm core areas were identified using the H&Es for each block which were scored and marked by a histopathologist (Dr Matthew Sommerlad or Dr Guy Martland). The blocks were manually scribed and using the TMADesigner2 software, the cores were marked. A 1mm minicore punch (Mitogen) removed the tissue from the donor block and this was inserted into a new recipient block. 4µm sections were cut using a microtome (Leica) for immunohistochemical analysis. Construction is shown in fig 14.

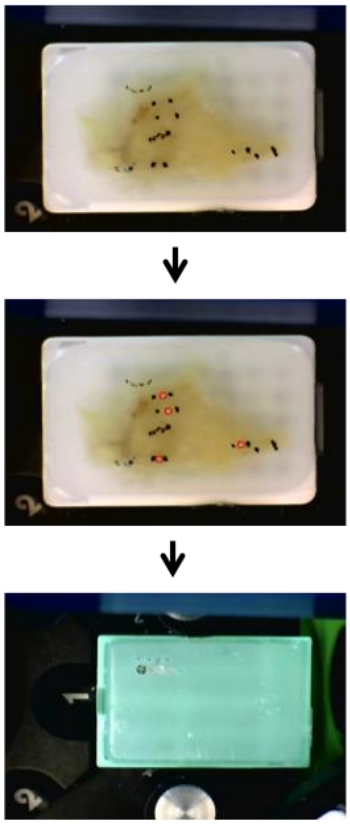


Figure 14 Construction of a Tissue Microarray (TMA)

Using the H&E as a guide, areas were marked on the block using a pen. The TMA2 designer software was used to electronically identify core positions using the drawn on areas as a model from the donor block. 1mm donor cores are taken and inserted into the recipient block.

2.2.4 Immunohistochemistry staining ER, PR, HER2, p53, integrin αvβ6, α-SMA and pSMAD2/3

4µm sections were cut from each TMA using a microtome (Leica). Immunohistochemistry (IHC) was used to determine the presence in cancer and DCIS cells for the oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth receptor 2 (HER2), p53, integrin alpha v beta 6 (αvβ6), alpha smooth muscle actin (α-SMA) and phospho-Smad2/3 (Ser465/467) (pSmad2/3). ER, PR, HER2 and p53 were stained using an automated system, the same system that is routinely used for clinical invasive breast samples. ER, PR and HER2 use the Roche (Ventana) equipment with the Ventana Benchmark XT staining platform and the Ultraview-Universal DAB detection Kit. p53 uses the Dako equipment with Dako PT link for antigen retrieval, a Dako autostainer Link 48 staining platform and the Envision FLEX detection system which is a complete kit and primarily requires the primary antibody.

αvβ6, α-SMA pSmad2/3 were stained manually with sections firstly deparaffinised in clearene and rehydrated through reduced concentrations of ethanol. Endogenous peroxidases were inhibited and the heat induced epitope retrieval (HIER) method was used for antigen retrieval for αvβ6 and α-SMA. A 20 minute avidin block followed by a biotin and protein block was administered prior to an overnight incubation with the primary antibody. Sections were washed and a biotinylated secondary antibody was incubated for 30 minutes. Sections were washed; an avidin-biotin block complex (ABC complex) was added for 30 minutes followed by the DAB substrate. Sections were dehydrated through increasing concentrations of ethanol, clearene and were mounted. Summary statistics were used to describe the characteristics of the cases.

Antibody	Dilution	Company
ER	HTU	Roche
PR	HTU	Roche
HER2	HTU	Roche
p53	1:30	Dako
αvβ6 integrin	1:1000	Santa Cruz
α-SMA	1:100	Dako
pSmad2/3	1:1000	Cell Signaling

Table 4 List of antibodies used for immunohistochemistry.

ER, PR and HER2 antibodies come ready to use as part of a Ventana Benchmark XT automated system. p53, integrin αvβ6, α-SMA and pSmad2/3 were diluted for use.

2.2.5 Immunohistochemistry evaluation

The ER, and PR status was evaluated using the Craig Allred scoring system [179]. This was scored based on both the percentage of cells that were expressing the receptor, as well as a score based on the intensity of the staining. A score of ≥3 or above was considered as positive. HER2 is scored between 0-3 and is scored based on the intensity of the staining when present in over 10% of the tumour. A score of 3+ is considered positive. p53 was scored using a semi-quantitative modified McCarthy 'H' score but this was scored to give a maximum score of 7 rather than 300 [180, 181]. This is scored based on the proportion of cancer cells staining positive 1= <25%, 2= 25-50%, 3 = 50-75%, 4= >75% and the strength of staining intensity 1= weak, 2= moderate, 3= strong. αvβ6 and α-SMA are scored based on the strength of staining intensity 1= weak, 2= moderate, 3= strong) as described by Marsh and colleagues for αvβ6 scoring [143]. Evaluation is shown in table 5 and 6.

Staining score	Proportion of positively stained cells	Intensity score	Intensity of positively stained cells
0	None	0	Absent
1	1%	1	Weak
2	1-10%	2	Intermediate
3	10-33%	3	Strong
4	33-66%		
5	>66%		

Table 5 Scoring system for ER, PR, HER2, integrin αvβ6 and α-SMA

ER and PR are scored using the Allred system which gives a quantitative score of 1-5 for the proportion of stained cells and 0-3 for staining intensity. HER2 is scored out of 0-3 based on intensity in over 10% of cells. Integrin αvβ6 and α-SMA are scored based on intensity: Absent to strong staining.

Staining score	Proportion of positively stained cells	Intensity score
0	None	0
1	<25%	1
2	25-50%	2
3	50-75%	3
4	>75%	

Table 6 Scoring p53 status

p53 is scored out of 7: proportion of cells staining positive 1= <25%, 2= 25-50%, 3 = 50-75%, 4= >75%; strength of staining intensity 1= weak, 2= moderate, 3= strong.

2.2.6 pSmad2/3 staining evaluation: Halo

TMA sections stained with pSmad2/3 were scanned at x40 magnification on an automated DotSlide system (Olympus). nDPI files of cores were uploaded to the HALO image analysis software (Indica Labs) and software was trained to identify epithelial tissue with classifiers added.

The classifier shows the area scored (red) and the white space excluded from statistical analysis (green). The output gives a score for each core for proportion of positive cells and staining intensity scoring cells as weak, moderate and strong. Cells scored strong were shown in red, moderate were stained in orange, weak were shown in yellow and negative cells were coloured blue. From this output file a scoring system similar to p53 was adapted.

2.2.7 Imaging of Hematoxylin and eosin (H&E) and tissue microarray (TMA) stained sections

H&E and TMA sections were imaged at a x10 or x20 magnification using the Dotslide (Olympus).

2.2.8 Morphology and immunohistochemical review statistics

IBM SPSS Statistics program was used for Pearson Chi Square, Fisher's Exact and a Wilcoxon signed rank tests. These tests were utilised depending on the number of cases and distribution of the data.

2.3 HaloPlex Target Enrichment System

2.3.1 HaloPlex design

HaloPlex Target Enrichment System (Agilent Technologies) is a targeted sequencing approach allowing a panel of genes to be specified into the kit design for the required DNA sequence. The kit was specific for DNA derived from formalin fixed paraffin embedded (FFPE) and was optimised for fragmented DNA expected from FFPE. This kit had around 6 times more amplicons than a standard kit designed for genomic DNA and the range of amplicons differs with a higher proportion of these towards the lower end of the 50-500bps spectrum.

2.3.2 Targeting the gene sequence

Various data mining approaches were used to investigate the genes most likely to be informative when designing the target region. Approaches included a literature search, databases such as the catalogue of somatic mutations in cancer (COSMIC) for mutations associated with specific breast cancer subtypes (including *in situ* disease, stromal genes, ER, PR and HER2 positive tumours suggested from the morphology review and IHC), the HUGO gene nomenclature committee (HGNC) database to identify associated genes and DAVID bioinformatics resources 6.7 to investigate genes that were mutated to a lesser extent in breast cancer, but were however involved in signalling pathways which are believed to be implicated in these tumours. Databases Ensembl

and the UCSC genome browser were utilised to examine coverage of these genes from the design output (Agilent Technologies).

2.3.3 Laser capture microdissection (LMD) and macrodissection

Samples marked by a histopathologist (Dr Matthew Sommerlad or Dr Guy Martland) were selected for ductal carcinoma *in situ* (DCIS) and invasive tumour to either undergoing laser capture microdissection (LCM) (Leica) or samples containing large areas of concentrated tumour were macrodissected using sterile conditions. 12µm thick sections for LCM or 15µm for macrodissection were cut using a microtome (Leica) and then mounted onto Arcturus PEN membrane glass slides (Life Technologies). For some samples it was possible for thicker 50µm sections to be cut using a microtome (Leica) and immediate manual dissection using sterile conditions. No staining was required for these samples. LCM and macrodissection sections were washed with xylene (Sigma), dehydrated with ethanol (70% and 100%) followed by staining with cresyl violet acetate (0.125% in 100% ethanol). LCM sections were marked using the Leica Laser Microdissection V 5.0 software (Leica) and cut using an ultraviolet cutting laser. Tissue was captured in the lid of a 0.5ml PCR tube (Greiner Bio-One) containing 50µl for LCM samples and 100µl of ATL buffer with 10% proteinase K (Qiagen) for macrodissected samples.

2.3.4 DNA extraction

DNA was extracted immediately following LCM or macrodissection using a FFPE DNA Tissue Kit (Qiagen) following the manufacturer's instructions. Unstained macrodissected samples followed the entire protocol with the xylene and ethanol washes. Stained samples followed a modified protocol following on from an overnight 56◦C lysis step (step 11, see manufacturer's instructions for full protocol).

2.3.5 DNA quantification: NanoDrop and Qubit

Samples were quantified using the NanoDrop 1000 spectrophotometer (Thermo Scientific) in which 1µl of sample DNA was used to determine the 260/280 ratio. The Qubit 2.0 Fluorometer (Life Technologies) and the Qubit dsDNA BR assay was (Life Technologies) used following the manufacturer's instructions to determine the concentration of dsDNA.

2.3.6 FFPE derived DNA quality assessment in preparation for HaloPlex target enrichment

In preparation for next generation sequencing, Agilent recommend a multiplex PCR-based quantification assay for FFPE samples to test DNA integrity (Agilent Technologies). Each sample was used as a template for the PCR amplification of two independent GAPDH amplicons. A 2100

Bioanalyser system with 2100 Expert Software (Agilent Technologies) was used for validation. The yield of amplicons from the FFPE material was compared to the yield obtained from intact reference DNA template at base pair fragments of 105 and 236. The FFPE to reference yield ratio served as a quantitative indicator of DNA integrity that was then used as a predictor for successful HaloPlex target enrichment. This score was used as a guide for the amount of DNA input for enrichment and the depth of sequencing required. Due to the poor quality of many of these archival samples, this protocol was modified to include samples where the largest amplifiable fragments were 105 bps. A summary of this scoring system is presented in table 7.

Sample integrity category	Average yield ratio	Recommended DNA input (ng)	Recommended additional sequencing
A	>0.2 (>20%)	200-500	1x-5x
B	0.05 to 0.2 (5% to	500-1000	5x-10x
C	<0.05 (<5%)	1000-2000	10x-100x
<C	N/A	2000+	10x-100x

Table 7 Recommendation for FFPE derived DNA for HaloPlex target enrichment.

The sample integrity categories were determined by the average yield ratio produced by the multiplex PCR-based quantification assay. Samples fell into one of four categories producing a guide for enrichment DNA input and additional sequencing.

2.3.7 Concentration of DNA samples for target enrichment

Vacuum concentration (Concentrator Plus, Eppendorf) was used to concentrate DNA samples to provide the required sample concentration for target enrichment.

2.3.8 HaloPlex Target Enrichment System for Illumina Sequencing

Patient sample DNA and the enrichment control DNA (ECD) underwent a 16 different restriction digest. The digest was validated running the ECD on a 2100 Bioanalyser system with 2100 Expert Software (Agilent Technologies). After validation all samples underwent a 16 hour hybridization specific to the kit size. Streptavidin beads (Agilent Technologies) were used to capture circularized target DNA-HaloPlex probe hybrids followed by a ligation step to close nicks. Eluted DNA libraries were amplified via a 16 cycle PCR reaction and purified using AMPure XP beads (Beckman Coulter Genomics). Successful target enrichment was confirmed using a 2100

Bioanalyser system with 2100 Expert Software (Agilent Technologies). Samples were pooled according to required additional sequencing and were directly sequenced on an Illumina HiSeq 2000 or NextSeq (Illumina) platform. A flowchart of HaloPlex target enrichment is shown in fig. 15.

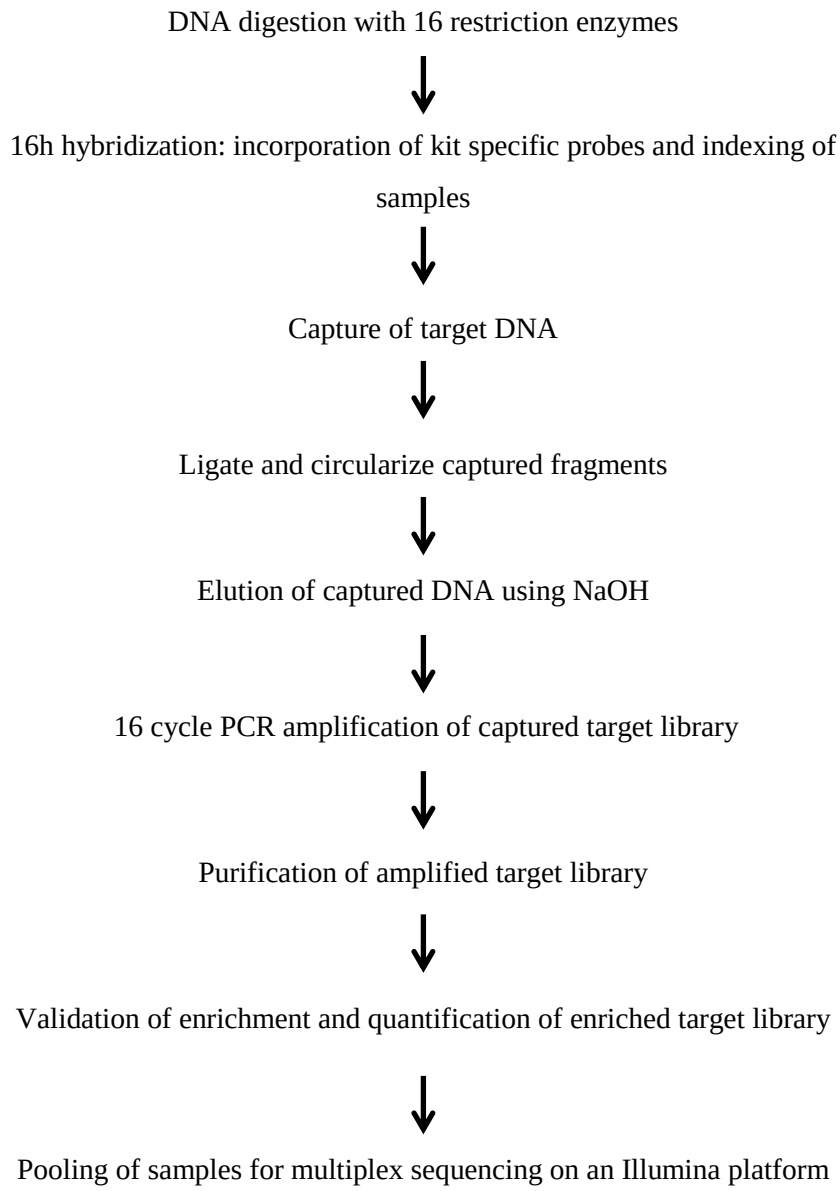


Figure 15 Flowchart for HaloPlex Target Enrichment System for Illumina sequencing.

Key steps involved in library preparation for HaloPlex Target Enrichment and sequencing. 16h; 16 hours.

2.3.9 Illumina sequencing

Libraries were sequenced at high depth to detect variants at variant allele frequencies (VAF) as low as 1%. Libraries were sequenced in 3 batches with batches 1 and 2 outsourced to High Throughput Genomics, Oxford Genomics Centre using a HiSeq 2000 sequencing system (Illumina) with 100 paired-end (PE) sequencing. Library 3 was sequenced internally on a NextSeq 500 system with 150 PE sequencing and a high output v2 kit.

2.3.10 Bioinformatics analysis and interpretation

Raw sequencing data was processed through an in-house cancer bioinformatics pipeline and annotated using VarScan2 caller for cancer samples by Dr Reuben Pengelly [182]. Variants underwent rational filtering including the removal of synonymous variants, germline variants present in dbSNP, 1000 Genomes, Exome Variant Server sequencing project and complete genomics 46 database and removal of suspect false positives in reference to published work by Fuentes et al. [183]. Variants were removed with a read depth below 100 reads and a VAF score of <5%. From the filtered list of variants, clinically relevant and mutations in *TP53* were visualised in Integrative Genomics Viewer (IGV) to see if these looked real or were more likely to be artefacts or sequencing errors [184, 185]. Variants called in repeat regions or reads with many bases that differed from the reference genome, were more likely to be false positives. DAVID pathway analysis was used for pathway interpretation [186, 187]. The Ease score is a modified Fisher's Exact P-Value adapted for gene-enrichment in annotation terms [186, 187].

2.4 Cell culture

2.4.1 MCF10A and MCF10A.ErbB2 cell culture

MCF10A and the virally modified cell line MCF10A.ErbB2 were cultured at 37◦C and 5% CO_2 in MCF10A growth media (Dulbecco's modified Eagle's/Hams F12 (1:1) media (Gibco-Invitrogen and Lonza-Biowhittaker) supplemented with 5% horse serum (Life technologies), 1X penicillin-streptomycin (Sigma), glutamine (Sigma), epidermal growth factor (20ng/ml) (peprotech), insulin (10µg/ml) (Sigma), hydrocortisone (0.5µg/nl) (Sigma) and cholera toxin (100ng/ml) (Sigma). A 5 minute Hank's Balanced Salt Solution (HBSS) (Life Technologies) wash was performed prior to trypsinization using 0.25% trypsin-EDTA solution (Sigma). The media was changed every 48-72 hours depending on the health and confluency of the cells.

Growth media	
Reagent	**Concentration**
Dulbecco's modified Eagle's/Hams F12 (1:1) media	1:1
Horse serum	5%
Penicillin-streptomycin	1X
Glutamine	1X
Epidermal growth factor	20ng/ml
Insulin	10µg/ml
Hydrocortisone	0.5µg/nl
Cholera toxin	100ng/ml

Table 8 Growth media used for two dimensional culture.

Growth media is used for cell maintenance and two-dimensional culture.

2.4.2 Transformation and plasmid DNA preparation

100ng of plasmids pMKO.1puro shRNA and pMKO.1puro p53 shRNA (gift from William Hahn, Addgene plasmid #10671), were transformed into XL-1 blue (*E.coli*) bacteria for 20 minutes on ice and were transformed using a heat shock method at 42◦C for 45 seconds. This was inoculated to a larger Luria-Bertani (LB*)* broth culture which was rocked at 200-250rpm, 37◦C for 1-1.5 hours. The culture was plated onto an LB ampicillin (100µg/ml) plate and colonies were grown overnight at 37◦C. A single colony was picked and inoculated in 5ml of LB culture with ampicillin (100µg/ml) as a mini prep at 37◦C all day. This was inoculated into 125ml LB and ampicillin (100µg/ml) culture at 37◦C overnight. Plasmid DNA was purified using a plasmid midi kit following the manufacturer's protocol (Qiagen).

The pBABEpuro-ErbB2 plasmid (gift from Matthew Meyerson, Addgene plasmid #40978) arrived in as a stab culture and was resuspended in 100µl LB broth. Plasmid DNA was amplified and purified as above. Plasmid maps are shown in fig. 16.

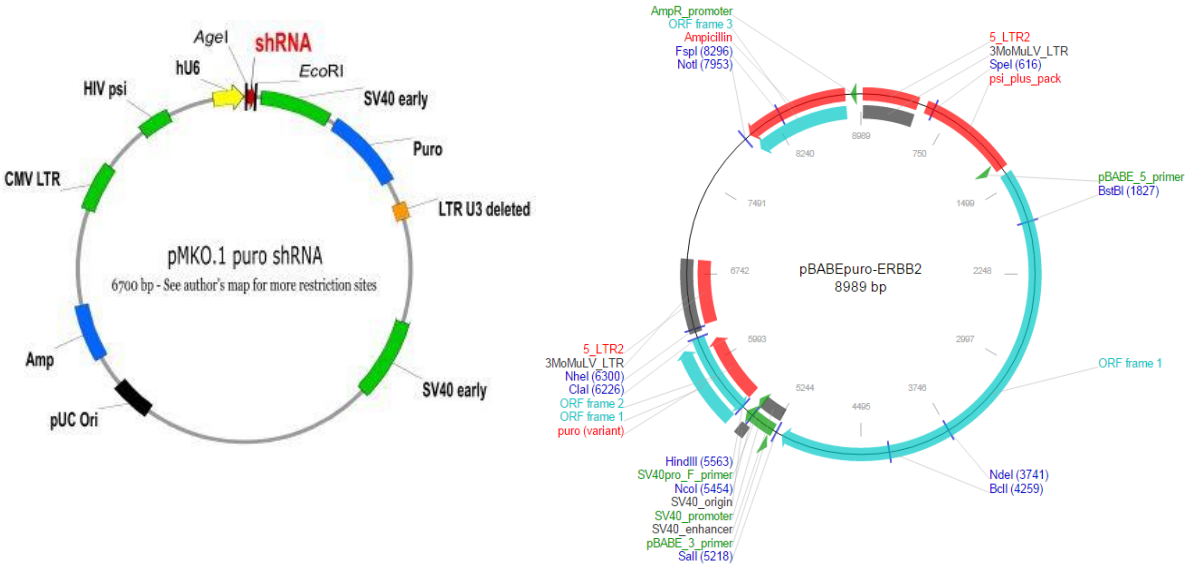


Figure 16 Plasmid maps.

Plasmids were designed by William Hahn (Addgene plasmid #10671) and Matthew Meyerson (Addgene plasmid #40978) from Addgene. Both plasmids have ampicillin resistance and the resulting cDNA can be expressed in mammalian cell line and for retroviral work.

2.4.3 Transfections

Cell lines were seeded at 250,000 cells per dish 10mm^2 24 hours prior to transfection for a confluency of around 60-70%. Both cell lines were transfected with either control SiRNA (Applied Biosystems), *TP53* SiRNA (Qiagen), *BRCA1* SiRNA (Life Technologies) or a combined double *TP53* and *BRCA1* knockdown. Following the manufacturer's guidelines, the transfection reagent were made up using Opti-MEM (Life Technologies), INTERFERin (Polyplus) and 25nM of each SiRNA. SiRNAs are shown in table 9.

SiRNA	Sense sequence	Antisense sequence	Company
TP53	GGAAAUUUGCGUGUGGAGUTT	ACUCCACACGCAAAUUUCCTT	Qiagen
BRCA1	CAGCUACCCUUCCAUCAUATT	UAUGAUGGAAGGGUAGCUGTT	Life Technologies
Control	N/A	N/A	Applied Biosystems

Table 9 List of SiRNAs for MCF10A and MCF10A.ErbB2 knockdown cultures.

TP53 and *BRCA1* were knocked down in MCF10A and MCF10.ErbB2 cultures. SiRNAs are listed with each sense sequence, antisense sequence and the company from which it was purchased.

2.4.4 Three-dimensional culture assays

Three-dimensional cultures for MCF10A and MCF10A.ErbB2 cell lines were seeded 48 hours post transfection using 8 well BD falcon culture slides (Scientific Laboratory Supplies) onto BD matrigel basement membrane complex (Scientific Laboratory Supplies) previously described in the literature [188]. Cells were seeded at 1000 cells per well and were maintained for 12 days in MCF10A assay media: Dulbecco's modified Eagle's/Hams F12 (1:1) (Gibco-Invitrogen and Lonza-Biowhittaker) supplemented with 2% horse serum (Life technologies), 1X penicillin-streptomycin (Sigma), glutamine (Sigma), epidermal growth factor (5ng/ml) (peprotech), insulin (10µg/ml) (Sigma), hydrocortisone (0.5µg/nl) (Sigma), cholera toxin (100ng/ml) (Sigma) and 2% BD matrigel basement membrane complex (Scientific Laboratory Supplies).

Assay media	
Reagent	**Concentration**
Dulbecco's modified Eagle's/Hams F12 (1:1) media	1:1
Horse serum	2%
Penicillin-streptomycin	1X
Glutamine	1X
Epidermal growth factor	5ng/ml
Insulin	10µg/ml
Hydrocortisone	0.5µg/nl
Cholera toxin	100ng/ml
BD matrigel basement membrane complex	2%

Table 10 Assay media used for three-dimensional culture.

Components of assay media used for three-dimensional culture

2.4.5 Proliferation assays

25,000 transfected cells (x3) were seeded into each well of a 6 well plate. 3 cell wells for each condition were typsinized using 0.25% trypsin-EDTA solution (Sigma) every 2 days and each well was counted using a CASYton (Roche). GraphPad Prism was used for graph production and statistics. Statistical analysis used was an independent t-test.

2.4.6 2D protein lysates

Two-dimensional cultures for MCF10A and MCF10A.ErbB2 cell lines were seeded up to 48 hours post transfection. Cells were washed with phosphate buffered saline (PBS), lysed using urea lysis buffer (7M urea, 25mM NaCl, 0.05% Triton X-100, 20mM HEPES, pH 7.6, 100mM Dithiothreitol (DTT)) and supernatants were collected for western blotting.

2.4.7 Western blotting

Lysates were quantified using the 1X Bradford protein assay (Biorad) and were loaded appropriately on either an 8% or 10% acrylamide concentration SDS-PAGE depending on protein

size. Proteins were transferred at 20V overnight onto a Hybond ECL nitrocellulose blotting membrane (GE Healthcare), blocked with 5% non-fat dried milk/0.1% Tween 20/PBS and probed with the relevant primary antibody overnight (see table 11). Membranes were probed with a secondary antibody for an hour and processed with the Supersignal West Pico Chemiluminescent substrate reagent (Thermo Scientific) and imaged using the Fluoro-S MultImager (Biorad). All blots were probed for Actin as a loading control. Antibodies are described in table 11.

Primary antibody	**Dilution**	**Company**	**Conditions**	**Secondary antibody**	**Company**
p53	1:1000	AbD Serotec	3% milk, 4°, O/N	Sheep anti-mouse-HRP	GE Healthcare
HER2	1:1000	Cell signalling	3% milk, 4°, O/N	Sheep anti-mouse-HRP	GE Healthcare
Actin	1:5000	Sigma	3% milk, 4°, O/N	Goat anti-rabbit-HRP	Sigma

Table 11 List of antibodies used for western blotting.

Antibodies were used to investigate the expression of these proteins in the MCF10A and MCF10A.ErbB2 cell lines. All secondary antibodies were used at a dilution of 1:2000. O/N; overnight

2.4.8 Immunofluorescence of 3D acinar culture

Cultures were washed (phosphate buffered saline (1X PBS)), fixed with 4% paraformaldehyde for 15 minutes, washed (1X PBS) and followed by a 100mM glycine/1X PBS and a 0.2% triton X-100/1X PBS 10 minute incubation. Acini were washed (1X PBS), blocked with 10% fetal calf serum (FCS)/1X PBS for 30 minutes and stained with phalloidin TRITC and DAPI (0.6% BSA/PBS) for 1 hour in the dark. Cultures were washed twice with 1X PBS and once with distilled water and mounted using fluorescent mounting medium (Dako). Slides were evaluated using fluorescent and confocal microscopy.

Antibody	Dilution	Company
Phalloidin TRITC	1:5000	Sigma
DAPI (CAT#D9564)	1:500	Sigma

Table 12 List of antibodies used for immunofluorescence.

Antibodies for phalloidin TRITC and DAPI were used to investigate the morphology and luminal clearing when certain genes were knocked out.

2.4.9 Quantification of 3D acinar culture

4% paraformaldehyde was added to media and cultures overnight (2% final concentration) to minimise acini loss. Cultures were washed (phosphate buffered saline (1X PBS)), mounted and imaged on the Dotslide at a x10 magnification. VSI files were uploaded to Fiji (ImageJ) and each aciniar structure was scored 'normal' or 'abnormal' when compared to either the MCF10A control or MCF10.puro (empty vector) control [189]. GraphPad Prism was used for graph production and statistics. Statistical analysis used was an unpaired t-test.

2.4.10 Confocal microscopy

3D acinar cultures were imaged using a TCS SP5 confocal microscope (Leica) and the application suite advanced fluorescence lite software (Leica). Images were taken with a x20 glycerol lens using two lasers.

2.4.11 Primary fibroblast cell culture

Primary normal breast fibroblasts (NBF) and matched cancer-associated fibroblasts (CAFs) from associated HER2+ and triple receptor negative breast tumours were cultured at 37◦C and 5% CO_2 in fibroblast growth media: Dulbecco's modified Eagle's/Hams F12 (1:1) media (Gibco-Invitrogen and Lonza-Biowhittaker) supplemented with 10% horse serum (Life technologies), 1X penicillin-streptomycin (Sigma), glutamine (Sigma) and Amphotericin B (2.5µg/ml) (Sigma). A phosphate buffered saline (1X PBS) wash was performed prior to trypsinization using 0.05% trypsin-EDTA/PBS solution (Sigma). The media was changed twice a week with 2ml of conditioned medium left in the flask. Media is described in table 13.

Reagent	Concentration
Dulbecco's modified Eagle's/Hams F12 (1:1) media	1:1
Horse serum	10%
Penicillin-streptomycin	1X
Glutamine	1X
Amphotericin B	2.5μg/ml

Table 13 Primary fibroblast media

Components of primary fibroblast media

2.4.12 Quantitive polymerase chain reaction (qPCR) of stromal markers

100,000 Primary fibroblasts were seeded into a 6 well plate and cultured for 3 days. Cells were harvested, pelleted and RNA extracted following the manufacture's guidelines (Promega).

A high capacity cDNA reverse transcriptase kit (Applied Biosystems) was used to reverse transcribe the extracted RNA on a MJ Research PTC-200 Peltier Thermal Cycler following the manufacture's instructions. Product complementary DNA (cDNA) was used for real-time qPCR using Sybrgreen on a 7500 Real-Time PCR System (Applied Biosystems). The expression of stromal genes *ACTA2*, *COL1A1*, *FN1* and *CTGF* were investigated using the primers in the table below. Expression of *ACTB* was used as a control. Primers are listed in table 14.

Gene	Forward Primer	Reverse Primer	Company
ACTA2	GACAATGGCTCTGGGCTCTGTAA	ATGCCATGTTCTATCGGGTACTT	Sigma
COL1A1	ACGAAGACATCCCACCAATCACCT	AGATCACGTCATCGCACAACACCT	Sigma
FN1	TGTGGTTGCCTTGCACGA	GCTTGTGGGTGTGACCTGAGT	Sigma
CTGF	CCCTCGCGGCTTACCGACTG	GGCGCTCCACTCTGTGGTCT	Sigma
ACTB	TGGCACCCAGCACAATGAA	CTAAGTCATAGTCCGCCTAGAAGCA	Sigma

Table 14 List of primers for Taqman stromal marker analysis

List of primers used to investigate stromal gene expression in primary fibroblasts.

2.4.13 Statistics for stromal expression data

GraphPad Prism was used for graph production and statistics. CAF expression data was normalised to their matched NBF and a second set of analysis involved organising CAFs via their associated breast tumour subtype. CAFs were normalised to the average expression of the triple receptor negative CAFs for that particular gene.

Chapter 3: Morphology and immunohistochemistry

3.1 Tumour morphology review

A full morphology review was described as part of the characterisation of breast tumours with a germline *TP53* mutation. Two histopathologists (Dr Matthew Sommerlad and Dr Guy Martland) reviewed each case independently in which they reported on a total of 45 patients from 136 formalin fixed paraffin embedded (FFPE) breast and associated lymph node blocks. All cases had a variety of pathological features evaluated including tumour type, grade, presence and growth pattern of ductal carcinoma *in situ* (DCIS), sclerosis, vascular invasion and lymphocytic infiltration. Additionally the same histopathologists analysed a second group that were carefully selected from POSH as a control. Samples were selected for HER2+, matched DCIS and availability of germline data ruling out *BRCA1, BRCA2* and *TP53* mutations (n= 55).

The same reporting review was firstly implemented for a subset of cases from the POSH study. Data from *BRCA1* carriers (n= 60), *BRCA2* carriers (n= 61) and young breast cancer (YBC) with no underlying genetic predisposition (n= 98) were made available by the POSH steering group [82]. These groups were selected to see if their genomic background was having any influence over breast disease phenotype. For these subsets receptor status was not investigated as part of this study therefore these groups are not mutually exclusive. For the POSH morphology review, there was not a third histopathologist to review cases where readers 1 and 2 called features differently. As a result of caller discrepancy, a substantial amount of data has had to be recorded as 'missing' particularly for more subtle features such as stroma not routinely reported in the clinic. Graphs presenting the percentages of certain features demonstrate the proportion of cases when missing data is excluded. See chapter 2.1 for further information regarding cohorts.

Using immunohistochemistry (IHC), protein expression of the oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor 2 (HER2), p53, integrin αvβ6, alpha smooth muscle actin (α-SMA) and pSMAD2/3 were investigated in the invasive tumour and DCIS cells. Staining for the expression of these particular proteins will indicate possible metabolic and cellular pathways which are potentially driving these tumours (see section 1.8 for more information). Additionally receptor status data describing the POSH cohort (n=2956) was used as an age matched control [81].

3.1.1 Tumour morphology review: germline *TP53* carriers tumour type and grade

Patients with a germline *TP53* mutation developed tumours that were typically high grade ductal no special type (NST) with associated widespread high grade ductal carcinoma *in situ* (DCIS). 80% (36 patients) of cases were shown to contain tumour whist the remaining 20% (9 patients) had not yet progressed any further from high grade DCIS. From those 80% of tumour cases, 94% were of ductal no special type. Ductal NST tumours are the most common type of invasive breast disease and no significance was found between the early onset breast tumour groups from POSH. The data is described in fig. 17.

a)

Cohort	Tumour Type			
	Ductal (NST)	**Pure special type (90% purity)**	**Mixed tumour type (50-90% special type)**	**Missing data**
TP53	32/36 (88.9%)	2/36 (5.6%)	2/36 (5.6%)	0 (0.0%)
BRCA1	43/60 (71.7%)	1/60 (1.7%)	2/60 (3.3%)	14/60 (23.3%)
BRCA2	45/61 (73.8%)	3/61 (4.9%)	1/61 (1.6%)	12/61 (19.7%)
HER2+	50/55 (90.9%)	3/55 (5.5%)	2/55 (3.6%)	0 (0.0%)
YBC	77/98 (78.6%)	5/98 (5.1%)	0/98 (0.0%)	16/98 (16.3%)

b)

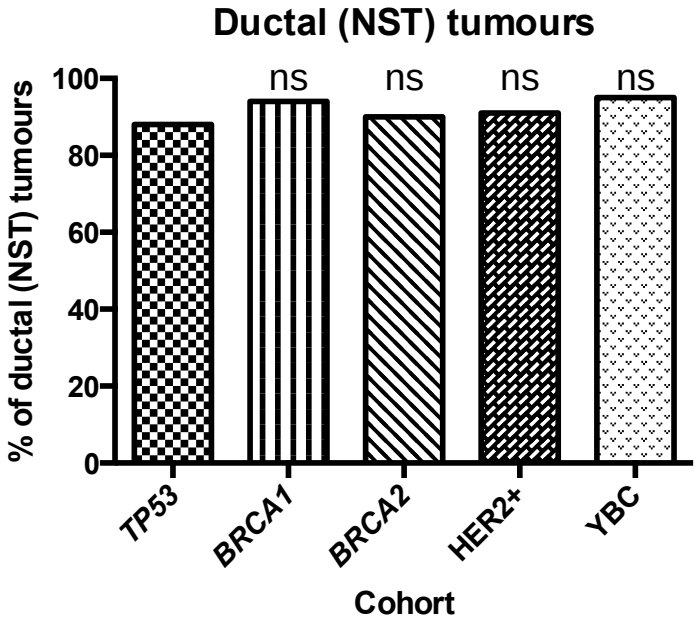


Figure 17 Ductal (NST) tumours across cohorts

Frequency of ductal no special type tumours. a) Table to show the frequency of ductal (NST), pure special type and mixed tumour type across the cohorts. b) Bar chart to show a comparison of ductal (NST) tumours when *TP53* germline carriers are compared to the POSH subgroups. Missing data was excluded from percentages presented in the graph. Ductal (NST) is the most common tumour type with no significant difference across the cohorts. Statistics used Pearson Chi Square.

Tumour grade was investigated across all cohorts in which tumours were typically grade 2 or 3. The data is shown in fig. 18.

a)

Cohort	Tumour Grade (%)			
	1	2	3	Missing data
TP53	2/36 (5.6%)	16/36 (44.4%)	18/36 (50.0%)	0 (0.0%)
BRCA1	4/60 (6.7%)	12/60 (20.0%)	23/60 (38.3%)	21/60 (35.0%)
BRCA2	1/61 (1.6%)	13/61 (21.3%)	22/61 (36.1%)	25/61 (41.0%)
HER2+	1/55 (1.8%)	26/55 (47.3%)	28/55 (47.3%)	0 (0.0%)
YBC	5/98 (5.1%)	18/98 (18.4%)	32/98 (32.7%)	43/98 (43.9%)

b)

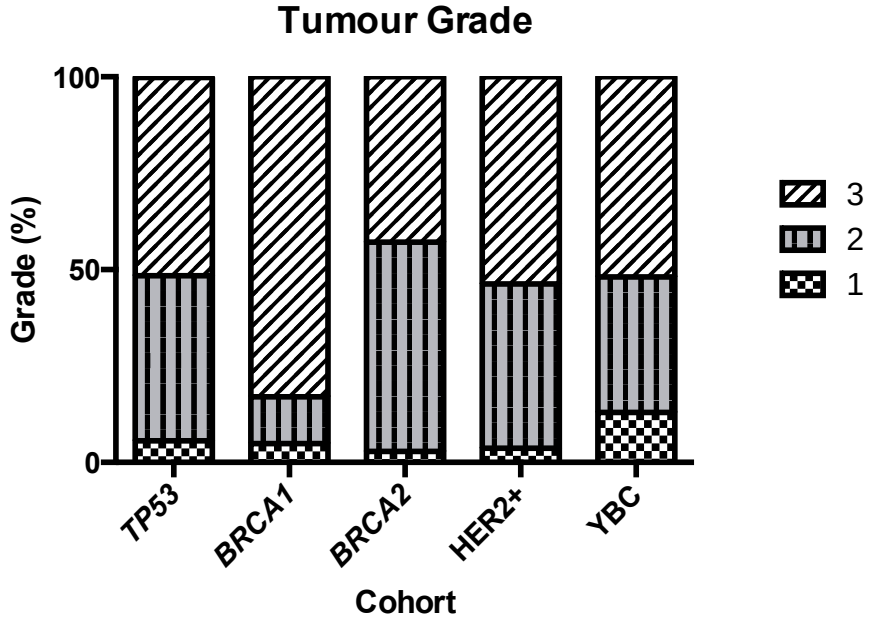


Figure 18 High grade tumours amongst early onset breast cancer cohorts.

Young breast cancer cohorts are typically grade 2 or 3. a) Table to show tumour grading in five cohorts. b) Graph to show the tumour grade distribution between the groups. Missing data was excluded from percentages presented in the graph.

Few grade one tumours were reported in early onset breast cancer cohorts. However, there was a clear difference between grade 3 tumours especially when the *BRCA1* patients were graded. *BRCA1* carriers had a significantly higher incidence of grade 3 tumours compared to the *TP53* carriers ($p=<0.001$, Pearson Chi Square). A prediction would be that a mutation in the tumour

suppressor gene *TP53*, would yield similar findings when it comes to tumour grade to that of the DNA repair gene *BRCA1*. By delving a little deeper into the way the tumours are scored, a possible explanation arose. One of the difficulties with the *TP53* cohort is the age of the samples and poor fixing. One of the many disadvantages of poor fixing is the loss of mitoses. This causes the overall tumour grade to drop from a grade 3 to a grade 2. Therefore, this cohort could have been in some cases, under scored because of this poor fixing and lose of mitoses. The scoring is shown below in fig. 19.

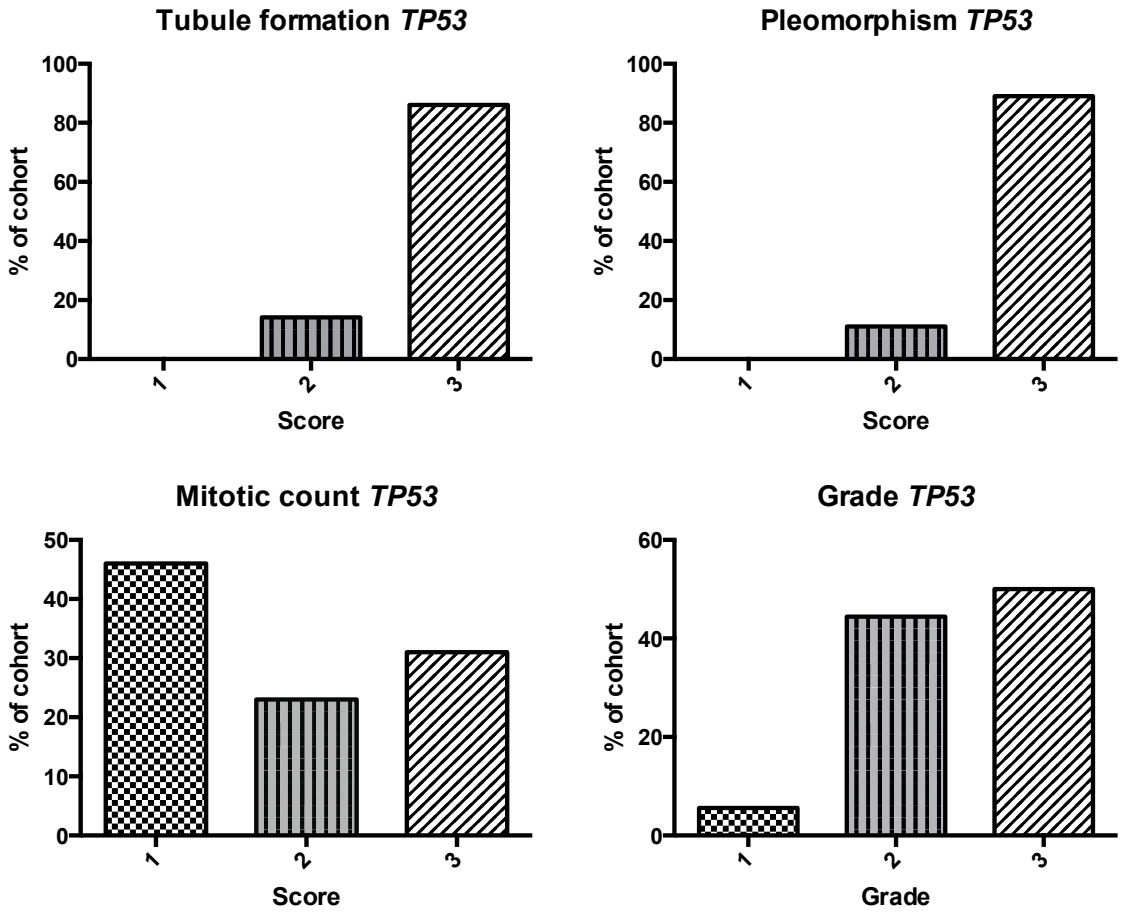


Figure 19 Scoring of the germline *TP53* cohort

The majority of the *TP53* invasive breast cancers typically scored 3 for tubule formation and pleomorphism. The scoring is much more widespread for mitotic count with 46% of the cohort scoring a 1.

3.1.2 Tumour morphology review: High frequency of infiltrative tumour border within *TP53* carriers and HER2+ breast tumours

There was significant variation between these groups when the type of tumour border was investigated. A similar frequency of an infiltrative tumour border was reported in *TP53* carriers (100%) and HER2+ cases from POSH (95%). The comparisons are shown in fig. 20.

a)

Cohort	Tumour Border (%)		
	Pushing	Infiltrative	Missing data
TP53	0/36 (0.0%)	36/36 (100.0%)	0/36 (0.0%)
BRCA1	15/60 (25.0%)	23/60 (38.3%)	22/60 (36.7%)
BRCA2	23/61 (37.7%)	26/61 (42.6%)	12/61 (19.7%)
HER2+	3/55 (5.5%)	52/55 (94.5%)	0/55 (0.0%)
YBC	31/98 (31.6%)	37/98 (37.8%)	30/98 (30.6%)

b)

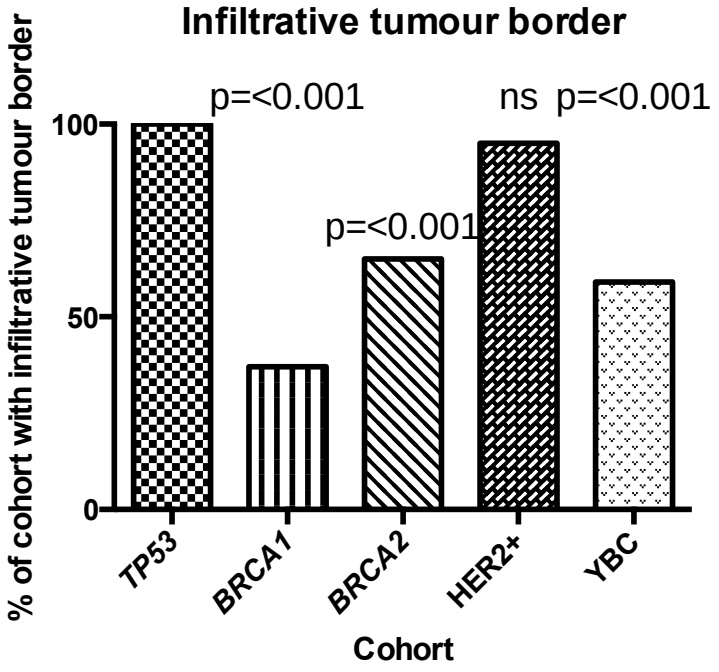


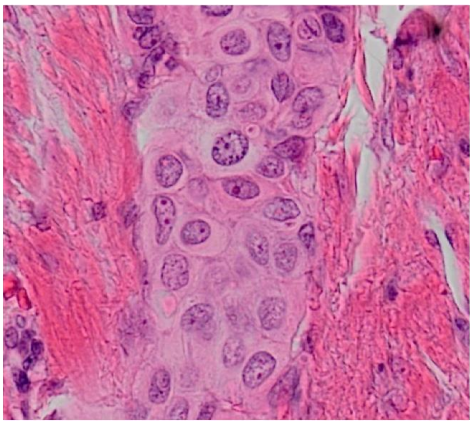
Figure 20 High frequency of infiltrative tumour border in *TP53* carriers and HER2+ tumours

All *TP53* breast cancers were scored as having an infiltrative border. a) Table to show the frequency of different tumour border types. b) Table to show the frequency of an infiltrative tumour border amongst cohorts. All cohorts were compared against the *TP53* cohort. There was no significance between *TP53* carriers and the HER2+ cohort. Statistics used Fisher's Exact test. Missing data was excluded from percentages presented in the graph.

3.1.3 Tumour morphology review: *TP53* carriers have a high frequency of sclerotic tumour stroma

A striking feature of the *TP53* carriers was a high prevalence of sclerotic tumour stroma. 81% of cases had this particular type of stroma which was significantly higher than HER2+, *BRCA1* carriers, *BRCA2* carriers and YBC subgroups from POSH. Cases of sclerotic tumour stroma from COPE are displayed in fig. 21 and statistical analysis is shown in fig. 22.

a)



b)

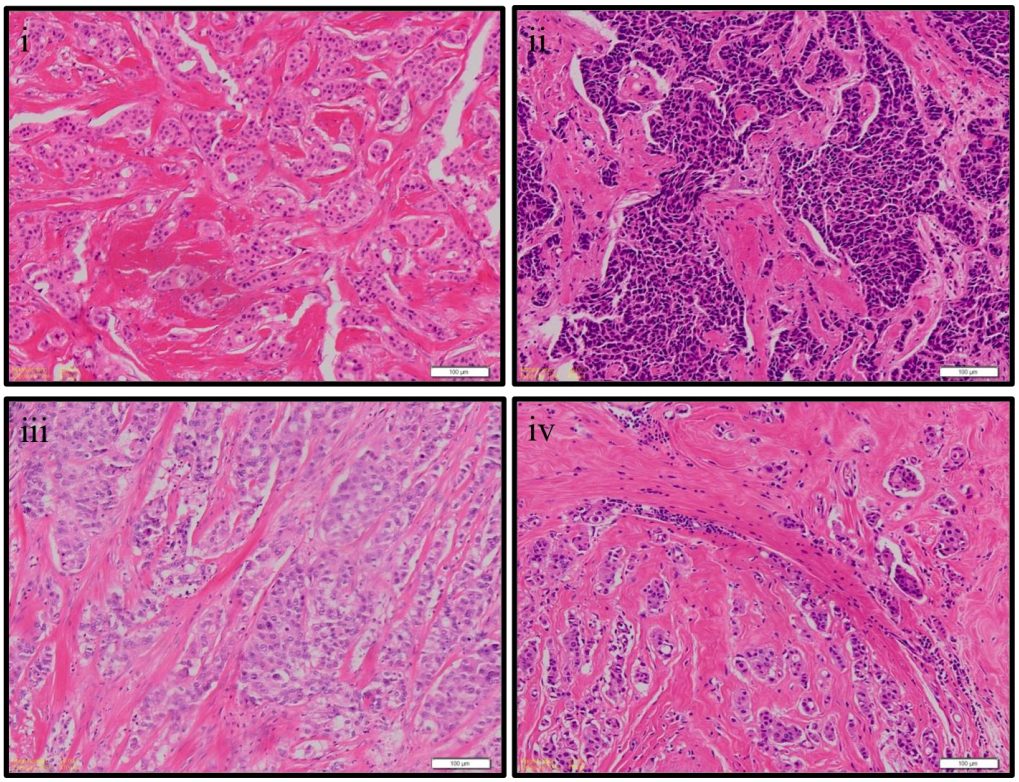


Figure 21 Invasive breast cancer with a surrounding sclerotic stroma in germline *TP53* breast tissue

A high frequency of sclerotic stroma was reported in *TP53* carriers. a) A close up of an area of tumour surrounded by sclerotic stroma. b) **i**, COPE case 30091007 shows a grade 3 ductal carcinoma surrounded by a sclerotic stroma. **ii**, COPE case 30091102 contains a grade 3 ductal carcinoma of basaloid cell type with a sclerotic stroma. **iii**, COPE case 30091003 again shows a grade 3 ductal carcinoma with a sclerotic stroma running through the tissue. **iv**, COPE case 30091123 contains islands of invasive grade 3 ductal carcinoma surrounded by this sclerotic stroma. Images were taken on the Olympus Dotslide at an objective magnification of x20 (a) or x10 (b).

a)

Cohort		Tumour Stroma (%)				
	Cellular	**Sclerotic**	**Desmoplastic**	**Myxoid**	**Other**	**Missing data**
TP53	1/36 (2.8%)	29/36 (80.6%)	6/36 (16.7%)	0/36 (0.0%)	0/36 (0.0%)	0/36 (0.0%)
BRCA1	2/60 (3.3%)	12/60 (20%)	5/60 (8.3%)	0/60 (0.0%)	1/60 (1.7%)	40/60 (66.7%)
BRCA2	2/61 (3.3%)	13/61 (21.3%)	5/61 (8.2%)	1/61 (1.6%)	1/61 (1.6%)	39/61 (63.9%)
HER2+	6/55 (10.9%)	28/55 (50.9%)	20/55 (36.4%)	1/55 (1.8%)	0/55 (0.0%)	0/55 (0.0%)
YBC	4/98 (4.1%)	21/98 (21.4%)	10/98 (10.2%)	0/98 (0.0%)	2/98 (2.0%)	61/98 (62.2%)

b)

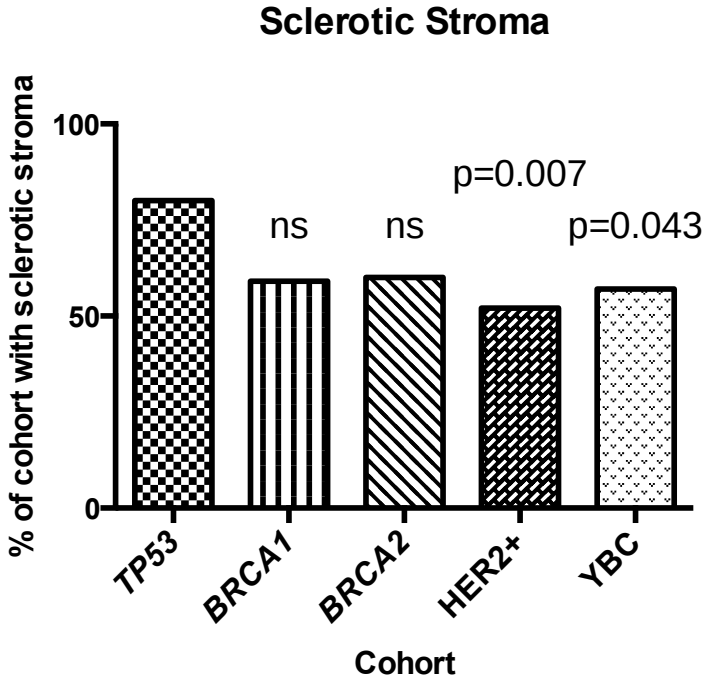


Figure 22 *TP53* carriers had a significantly higher proportion of sclerotic stroma

81% of *TP53* carriers were scored as having a sclerotic stroma a) Table to show the distribution of stromal types in young breast cancer onset cohorts. b) Bar chart showing the frequencies of sclerotic stroma between cohorts. *TP53* carriers had a significantly higher incidence of sclerotic stroma than HER2+ and YBC POSH comparison groups. Statistics were performed on *TP53* carriers against POSH groups using the Fisher's Exact test. Missing data was excluded from percentages presented in graph.

3.1.4 Tumour morphology review: Differences in lymphocytic infiltration in young onset breast cancer cohorts

Very little difference were observed between *TP53* carriers and HER2+ tumours. There was surprisingly, no significant difference between the *TP53* and *BRCA1* carriers. The young breast cancer (YBC) (p=0.048) and *BRCA2* (p=<0.001) groups had a significantly higher level of lymphocytic infiltration. Data is presented in fig. 23 below.

a)

Cohort	Lymphocytic Infiltration (%)		
	Absent/Mild	**Prominent**	**Missing data**
TP53	30/36 (83.3%)	6/36 (16.7%)	0/36 (0.0%)
BRCA1	21/60 (35.0%)	12/60 (20.0%)	27/60 (45.0%)
BRCA2	13/61 (21.3%)	16/61 (26.2%)	32/61 (52.5%)
HER2+	45/55 (81.8%)	10/55 (18.2%)	0/55 (0.0%)
YBC	34/98 (34.7%)	19/98 (19.4%)	45/98 (45.9%)

b)

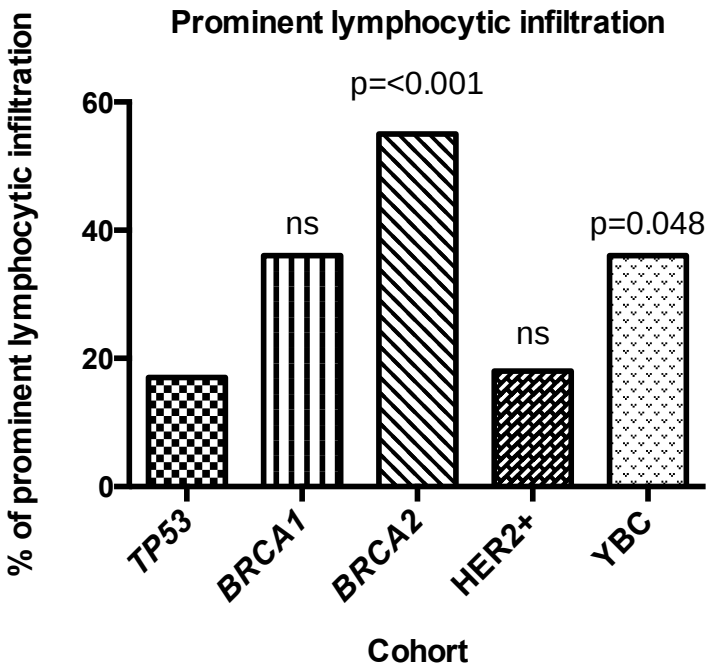


Figure 23 Frequency of prominent lymphocytic infiltration across early onset breast cohorts

TP53 carriers were reported as having the lowest frequency of prominent lymphocytic infiltration. a) Table to show the frequency of lymphocytic infiltration tumour across the cohorts. b) Bar chart to show the frequency of prominent lymphocytic infiltration across cohorts when the *TP53* germline carriers were compared to the POSH subgroups. *TP53* carriers had a significantly lower incidence of prominent lymphocytic infiltration than *BRCA2* carriers and young breast cancer (YBC) cohorts. Statistics used Pearson Chi Square. Missing data was excluded from percentages presented in the graph.

3.1.5 Tumour morphology review: *TP53* carriers have a similar incidence of vascular invasion to HER2+ tumours

When the incidence of vascular invasion amongst *TP53* carriers were compared to the subgroups from the POSH cohort, *BRCA1,* and *BRCA2* were shown to have significantly less vascular invasion. Very similar findings were found in the *TP53* carriers and the HER2+ subgroup from the POSH cohort. Vascular invasion data is presented in fig. 24.

a)

Cohort	Vascular Invasion (%)		
	Absent	**Present**	**Missing data**
TP53	24/36 (66.7%)	12/36 (33.3%)	0/36 (0%)
BRCA1	47/60 (78.3%)	7/60 (11.7%)	6/60 (10.0%)
BRCA2	44/61 (72.1%)	7/61 (11.5%)	10/61 (16.4%)
HER2+	36/55 (65.5%)	19/55 (34.5%)	0/55 (0.0%)
YBC	65/98 (66.3%)	14/98 (14.3%)	19/98 (19.4%)

b)

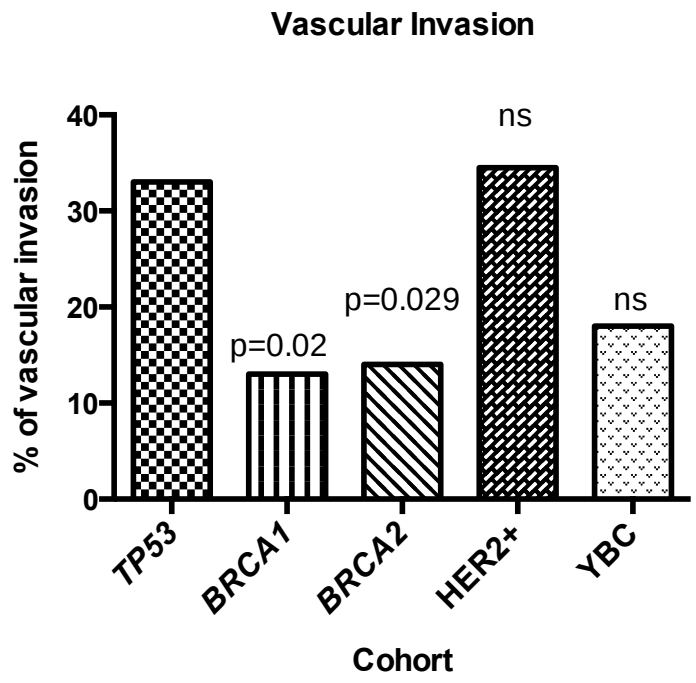


Figure 24 *TP53* carriers and HER2+ breast tumours have a high proportion of vascular invasion

TP53 carriers and HER2+ tumours were reported as having the highest frequency of vascular invasion. a) Table to show the presence of vascular invasion across the cohorts. *BRCA1* carriers had the lowest proportion of cases positive for vascular invasion and *TP53* carriers and HER2+ had the highest. b) Bar chart to show the frequency of vascular invasion across cohorts when the *TP53* germline carriers were compared to the POSH subgroups. *TP53* carriers had a significantly higher incidence of vascular invasion than *BRCA1* and *BRCA2* carrier cohorts. Statistics used Pearson Chi Square. Missing data was excluded from percentages presented in the graph.

3.2 Ductal carcinoma *in situ* (DCIS) morphology review

Ductal carcinoma *in situ* (DCIS) is a neoplastic proliferation of the epithelial cells and a precursor lesion to invasive carcinoma. At this stage of the disease however, the basement membrane and myoepithelial layer are still intact restricting these abnormal cells to the breast ducts. This precursor lesion is graded low, intermediate or high depending on the lack of polarisation and the particular architectural pattern of the proliferating cells of the duct [190]. In addition to grading, DCIS can grow in seven distinct growth patterns including: solid, comedo, cribriform, micropapillary, papillary, apocrine and flat. Patients often present with more than one of these growth patterns and each growth pattern is associated with a different risk of progression. Comedo DCIS is associated with high grade disease with a higher risk of becoming invasive whereas a flat growth pattern is typically associated with lower grade and a reduced risk of becoming invasive [190]. When comparisons were made between DCIS derived in *TP53* carriers and subgroups from POSH, the 9 *TP53* cases of pure DCIS were excluded from analysis. This is because of differences in the eligibility criteria for the 2 cohorts. For POSH, only cases with invasive disease fulfilled recruitment criteria (see chapter 2.1 for further information).

3.2.1 Ductal carcinoma *in situ* (DCIS) morphology review: Incidence of DCIS amongst young onset breast cancer cohorts

As previously stated, those patients with a germline *TP53* mutation typically had associated widespread high grade ductal carcinoma *in situ* (DCIS) (3.1.1). 91% of this cohort were DCIS positive with a 20% subset only containing high grade DCIS which had not yet progressed any further and become invasive. This 20% of pure DCIS cases were excluded from analysis and the incidence of DCIS were compared to other young onset breast cancer cohorts from the POSH study including subsets *BRCA1* carriers, *BRCA2* carriers, HER2+ and YBC.

a)

Cohort	Presence of DCIS (%)		
	Absent	Present	Missing data
TP53	4/36 (11.1%)	32/36 (88.9%)	0/36 (0.0%)
BRCA1	5/60 (8.3%)	7/60 (11.7%)	48/60 (80.0%)
BRCA2	8/61 (13.1%)	4/61 (6.6%)	49/61 (80.3%)
HER2+	6/55 (10.9%)	49/55 (89.1%)	0/55 (0.0%)
YBC	9/98 (9.2%)	7/98 (7.1%)	82/98 (83.7%)

b)

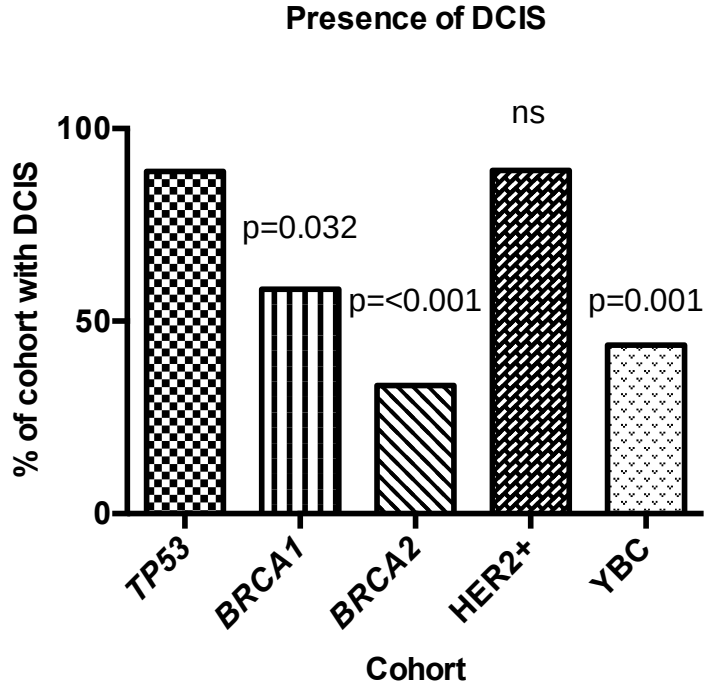


Figure 25 *TP53* carriers and HER2+ breast tumours have a high proportion of DCIS

TP53 carriers and HER2+ tumours were reported as having the highest frequency of matched DCIS. a) Table to show the presence of DCIS across the cohorts. b) Bar chart to show the incidence of DCIS across cohorts when the *TP53* germline carriers are compared to the POSH subgroups. *TP53* carriers had a significantly higher incidence of DCIS than *BRCA1* carriers, *BRCA2* carriers and young breast cancer (YBC) cohorts. Statistics used Fisher's Exact. Missing data was excluded from percentages presented in the graph.

3.2.2 Ductal carcinoma *in situ* (DCIS) morphology review: DCIS grade

After statistical significance was determined for the presence of DCIS between the cohorts, the grade of this precursor lesion was investigated. For those patients with a germline *TP53* mutation and that had matched DCIS present (88.9% of the cohort), 96.9% of those were described as having high grade DCIS. Fig. 26 shows high grade DCIS from four patients with a germline *TP53* mutation.

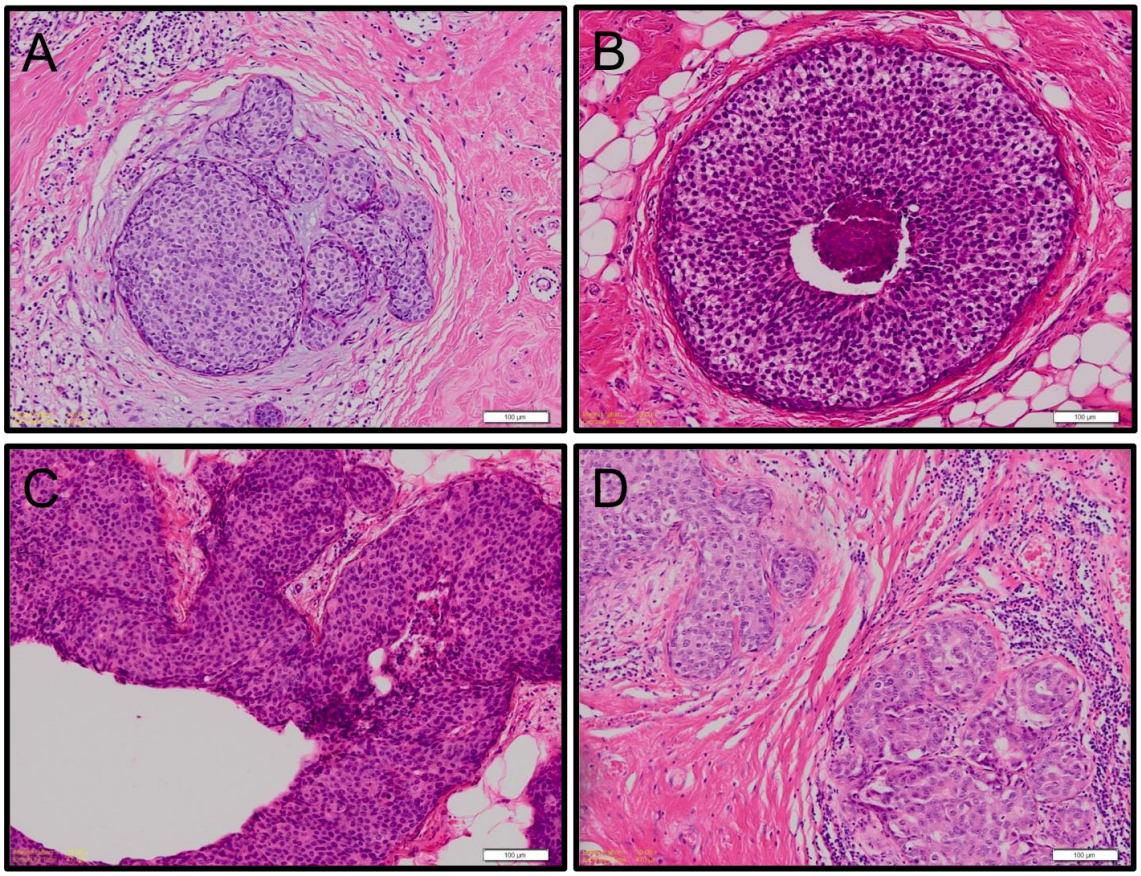


Figure 26 High grade ductal carcinoma *in situ* (DCIS) in germline *TP53* carriers.

A, COPE case 30095001 shows solid DCIS with cancerization of lobules (right). **B**, COPE case 30091122 shows DCIS of a comedo growth pattern with characteristic central necrosis. **C**, COPE case 30091007 contains a large area of solid DCIS. **D**, COPE case 30091003 is made up of two areas of solid DCIS. Images were taken on the Olympus Dotslide at an objective magnification of x10.

This feature of the *TP53* carriers were compared to *BRCA1* carriers, *BRCA2* carriers, HER2+ and YBC subgroups from the POSH cohort. *TP53* carriers were shown to have a significantly higher proportion of high grade DCIS compared to *BRCA2* carriers and YBC cohorts. The statistics for high grade DCIS are described in fig. 27.

a)

Cohort	DCIS Grade (%)			
	Low	Intermediate	High	Missing data
TP53	0/36 (0.0%)	1/36 (2.8%)	31/36 (86.1%)	4/36 (11.1%)
BRCA1	0/60 (0.0%)	0/60 (0.0%)	14/60 (23.3%)	46/60 (76.7%)
BRCA2	0/61 (0.0%)	6/61 (9.8%)	10/61 (16.4%)	44/61 (72.1%)
HER2+	0/55 (0.0%)	6/55 (10.9%)	43/55 (78.2%)	6/55 (10.9%)
YBC	1/98 (1.0%)	9/98 (9.2%)	18/98 (18.4%)	70/98 (71.4%)

b)

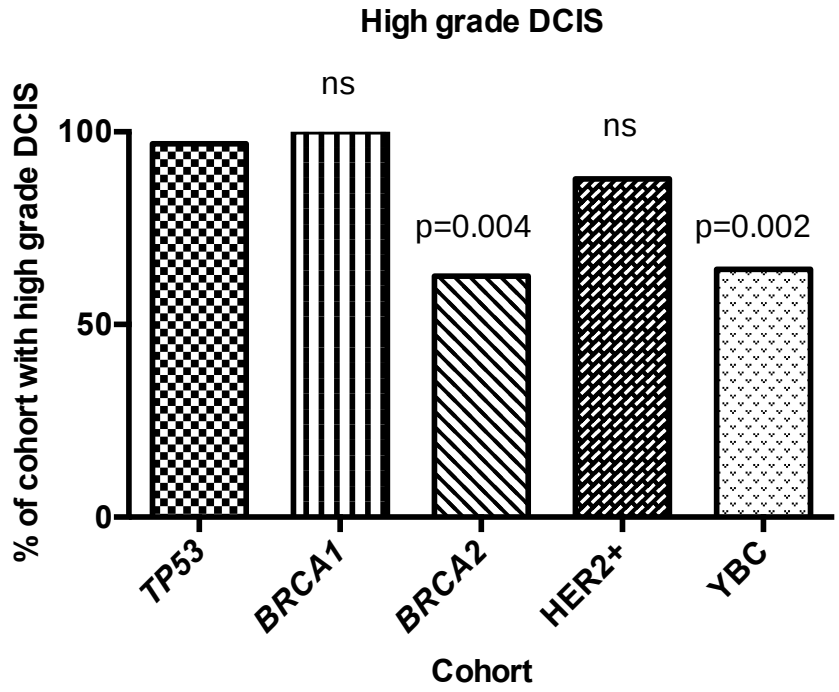


Figure 27 High grade ductal carcinoma *in situ* (DCIS) in early onset breast cancer cohorts.

Incidence of high grade matched DCIS in cohorts from POSH. a) Table to show the grade of matched DCIS across the cohorts. b) Bar chart to show the incidence of high grade DCIS across cohorts when *TP53* carriers were compared to the POSH subgroups. *TP53* carriers had a significantly higher incidence of high grade DCIS than *BRCA2* carriers and young breast cancer (YBC) cohorts. Statistics used Fisher's Exact test. Missing data and DCIS negative cases were excluded from percentages presented in the graph.

3.2.3 Ductal carcinoma *in situ* (DCIS) morphology review: DCIS growth patterns

Of the 41 *TP53* carriers that had developed DCIS, five growth patterns were described: solid, comedo, cribriform, micropapillary and flat DCIS. For each case the two most common growth patterns were reported. Solid and comedo DCIS were the most common growth patterns. This data is described in fig. 28.

a)

Cohort	DCIS growth pattern (%)					
	Solid	**Comedo**	**Cribriform**	**Micropapillary**	**Flat**	**Not applicable**
TP53	31/45 (68.9%)	26/45 (57.8%)	9/45 (20.0%)	5/45 (11.1%)	2/45 (4.4%)	4/45 (8.9%)

b)

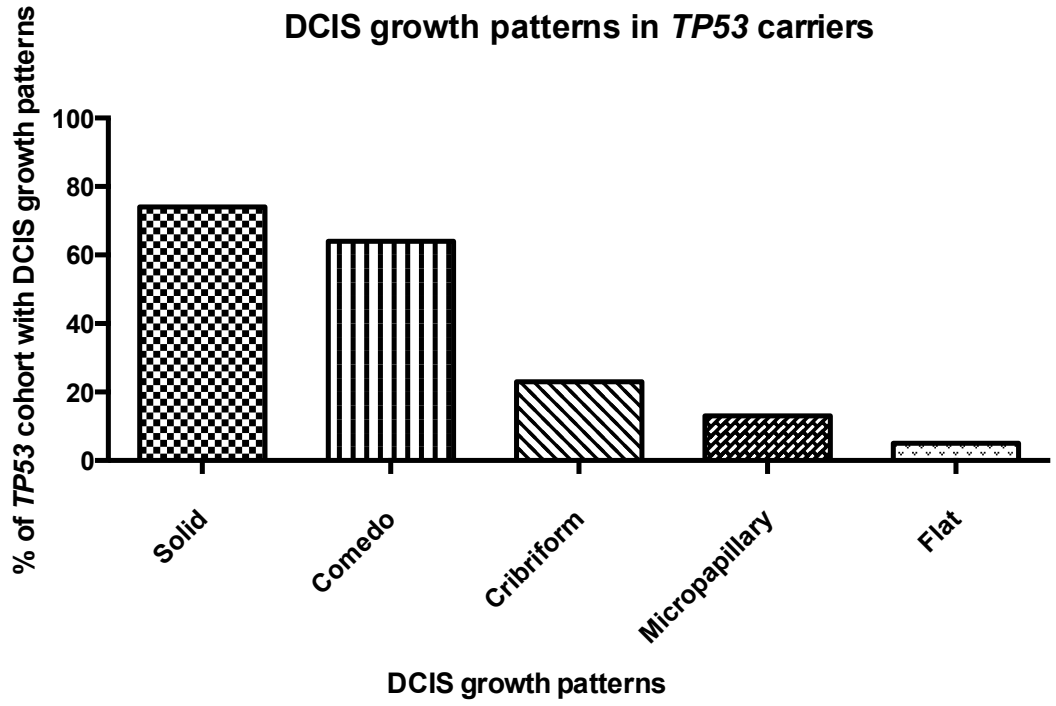


Figure 28 DCIS growth patterns in *TP53* carriers.

Frequency of DCIS growth patterns in *TP53* carriers. a) Table to show the presence of the DCIS growth patterns in *TP53* carriers. b) Bar chart to show the incidence of the particular DCIS growth patterns in *TP53* germline carriers. Solid DCIS was more prevalent than any other growth pattern. Comedo DCIS was the second most common growth pattern.

Many *TP53* carriers presented with multiple DCIS growth patterns. When the two most dominant DCIS growth patterns for each case was investigated, a clear pattern emerged showing that 19/41 (46.3%) of cases presented with a combination of solid and comedo DCIS growth patterns.

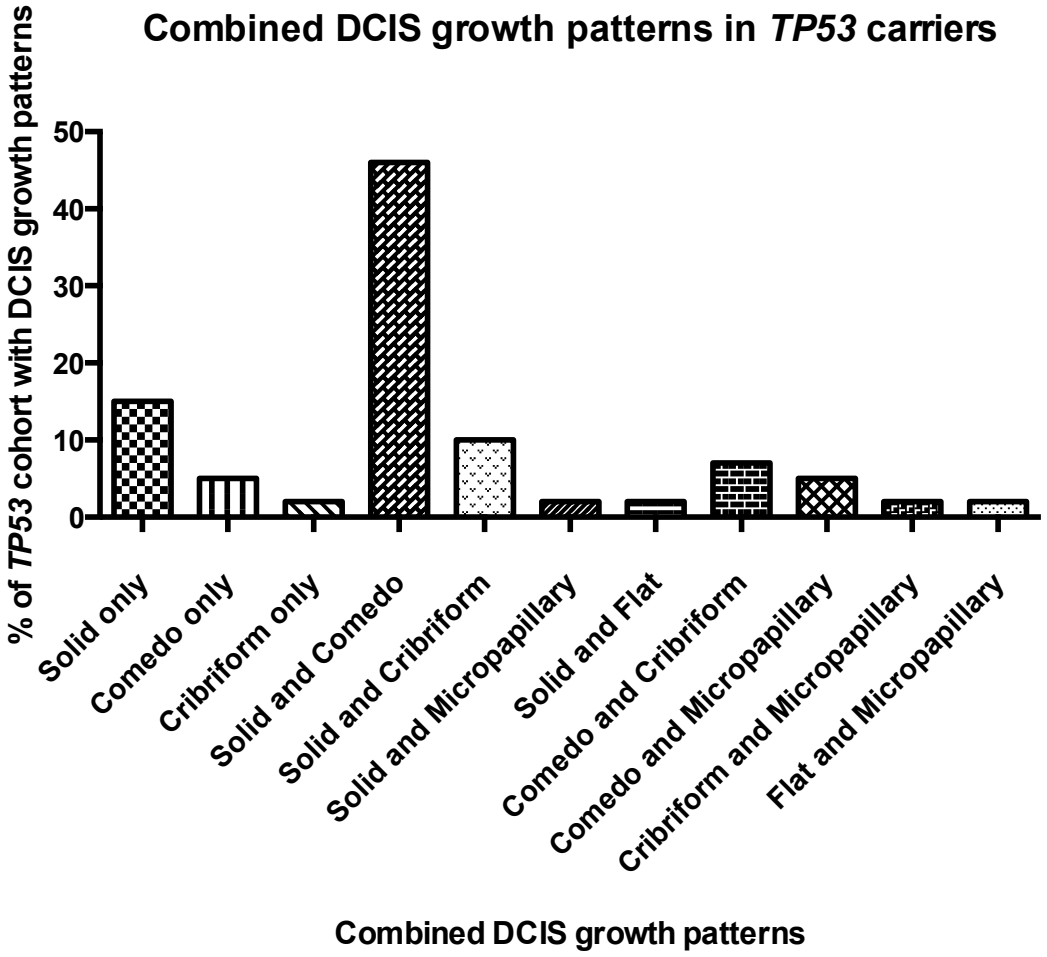


Figure 29 Combined DCIS growth patterns in *TP53* carriers.

Frequency of combined DCIS growth patterns in *TP53* carriers. Bar chart to show the incidence of the particular combined DCIS growth patterns in *TP53* germline carriers. A combined solid and comedo DCIS growth pattern were the most prevalent growth pattern phenotype.

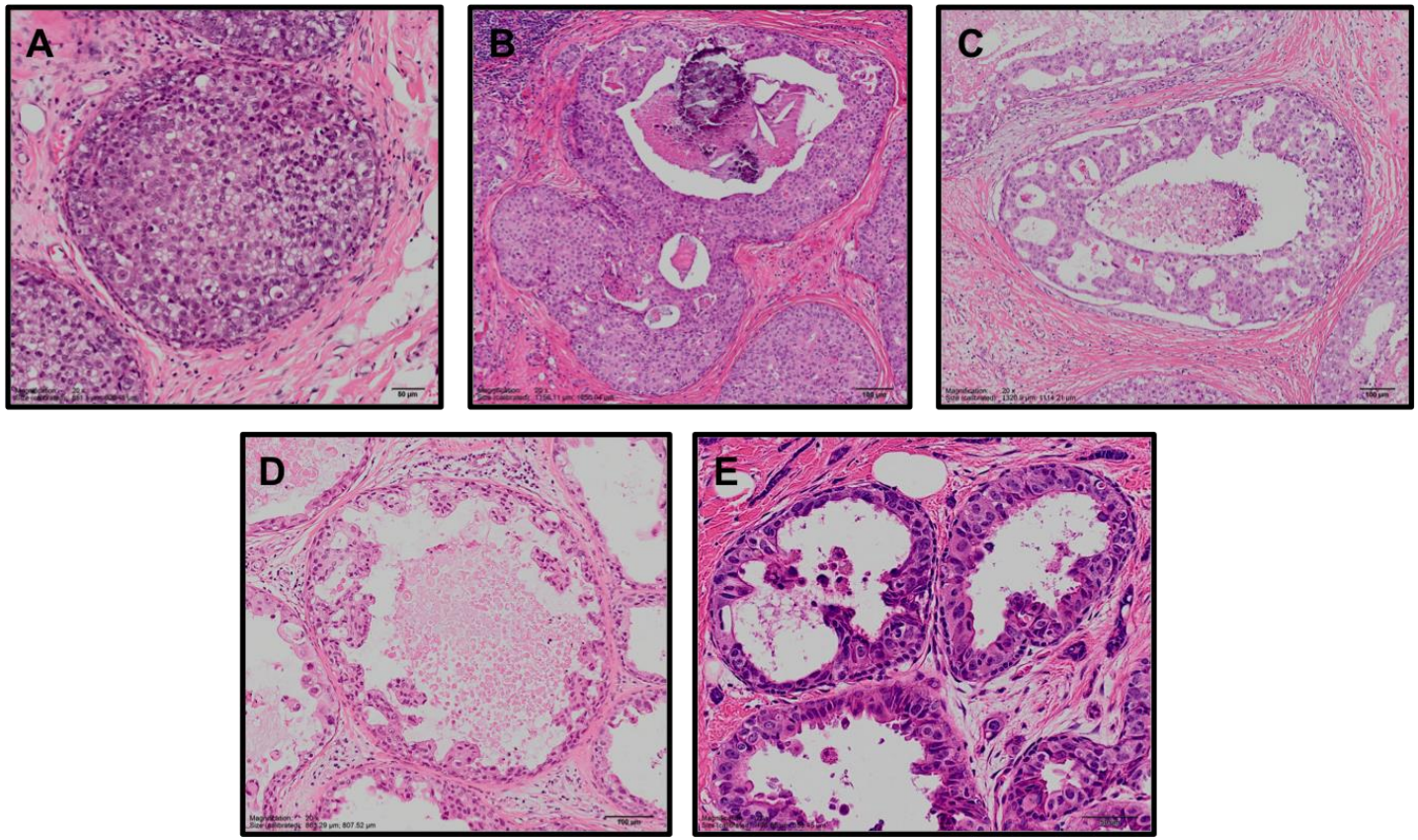


Figure 30 DCIS growth patterns in *TP53* carriers

From the *TP53* carrier cohort, five different growth patterns were described: solid, comedo, cribriform, micropapillary and flat. **A**, Patient 30091001: solid DCIS. B, Patient 30091401: comedo DCIS. C, Patient 30092901: cribriform DCIS. D, Patient 30091006: micropapillary DCIS. E, Patient 30091139: flat DCIS.

The growth patterns that presented in *TP53* carriers were compared to the POSH cohort to examine if particular growth patterns were a feature of a certain group. As part of the POSH morphology review, comedo DCIS was not reported therefore comparative statistics were not performed when describing these features. The types of DCIS growth patterns are described in table 15 and DCIS only cases in the *TP53* carrier group were once again excluded from the table.

Cohort	**DCIS growth pattern (% present in cases)**					
	Solid	**Comedo**	**Cribriform**	**Micropapillary**	**Flat**	**N/A***
TP53	23/36 (63.9%)	19/36 (52.8%)	7/36 (19.4%)	5/36 (13.9%)	2/36 (5.6%)	4/36 (11.1%)
BRCA1	13/60 (21.7%)	N/A	8/60 (13.3%)	2/60 (3.3%)	1/60 (1.7%)	46/60 (76.7%)
BRCA2	16/61 (26.2%)	N/A	8/61 (13.1%)	0/0 (0.0%)	0/0 (0.0%)	42/61 (68.9%)
HER2+	30/55 (54.5%)	37/60 (61.7%)	9/55 (16.4%)	0/60 (0.0%)	1/60 (1.7%)	6/60 (10.0%)
YBC	16/98 (16.3%)	N/A	21/98 (21.4%)	3/98 (3.1%)	1/98 (1.0%)	71/98 (72.4%)

Table 15 DCIS growth patterns

Frequency of DCIS growth patterns in early onset breast cancer cohorts. Table to show the spectrum of DCIS growth patterns in various early onset breast cancer cohorts. *TP53* carriers contained mostly solid and comedo growth patterns. * Includes cases with no DCIS reported and missing data.

3.3 Immunohistochemistry of breast cancers derived in a germline *TP53* background

Tissue microarrays (TMAs) were constructed from *TP53* carriers. Using immunostaining, clinically significant proteins and proteins believed to be implicated in tumourogenesis were

examined in DCIS and invasive tumour cores. These proteins included oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth receptor 2 (HER2), p53 tumour suppressor protein, integrin αvβ6, alpha smooth muscle actin (α-SMA) and phospho-Smad2/3 (pSMAD2/3).

3.3.1 The receptor status of germline *TP53* breast tumours: High frequency of HER2+

TMAs were stained for HER2 across the cohort which revealed 59% of patients were HER2+ (3+). This feature was also described in the POSH cohort with HER2 overexpression confirmed in 717/2956 (24%) of the entire cohort [81]. This receptor status feature was significant when these two groups were compared with a significance of <p=0.001. The data is described in fig. 31.

a)

Cohort	HER2- (% of cohort)	HER2+ (% of cohort)	Borderline	Missing data
COPE (*TP53*)	12/36 (33.3%)	19/36 (52.8%)	1/36 (2.8%)	4/36 (11.1%)
POSH	1839/2956 (62.2%)	717/2956 (24.3%)	45/2956 (1.5%)	355/2956 (12.0%)

b)

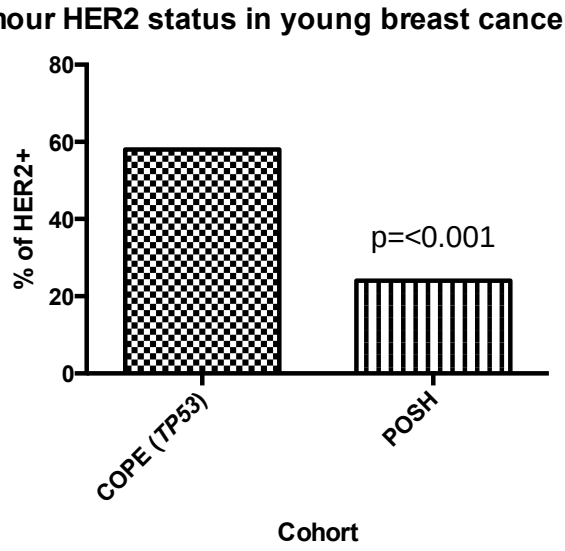


Figure 31 Overexpression of HER2 in *TP53* carriers

HER2 overexpression in early onset breast cancer cohorts. a) Table to show the incidence of HER2 overexpression in *TP53* carriers and all early onset breast cancer subgroups from the POSH cohort. b) Bar chart to show the incidence of HER2 overexpression amongst early onset breast cancer subgroups. Statistics used Pearson Chi Square test. Missing data was excluded from statistics.

There was no significance between ER and PR status between *TP53* carriers and cases from POSH. In general for clinical purposes, at least ER and HER2 status is evaluated. Therefore these receptors will be investigated further and comparisons made against the POSH cohort. PR+ was reported in 58.3% (21/36, missing data 3) of *TP53* carriers and 45% of the POSH cohort (1342/2956, missing data 581) [81]. ER data is described in table 16.

Cohort	ER- (% of cohort)	ER+ (% of cohort)	Missing data
COPE (*TP53*)	9/36 (25.0%)	24/36 (66.7%)	3/36 (8.3%)
POSH	997/2956 (33.7%)	1947/2956 (65.9%)	12/2956 (0.4%)

Table 16 ER status in early breast tumour subtypes

ER expression in *TP53* carriers and young breast cancer. a) Table to show the expression of ER in *TP53* carriers and young breast cancer from the POSH cohort. No significance was found between cohorts. Statistics used Pearson Chi Square test. Missing data was excluded from statistics

3.3.2 The receptor status of germline *TP53* breast tumours: High frequency of HER2+/ER+

Many breast tumours derived from *TP53* carriers were HER2+ and ER+ with very few tumours HER2-/ER-. The combination of HER2 and ER status were compared to the POSH cohort.

a)

Tumour receptor status	COPE, *TP53* carriers (% of cohort)	POSH (% of cohort)	p value*
HER2+/ER+	13/36 (36.1%)	461/2956 (15.6%)	0.002
HER2-/ER+	10/36 (27.8%)	1238/2956 (41.9%)	ns
HER2+/ER-	6/36 (16.7%)	256/2956 (8.7%)	ns
HER2-/ER-	3/36 (8.3%)	643/2956 (21.8%)	ns
Missing data	4/36 (11.1%)	358/2956 (12.1%)	-

b)

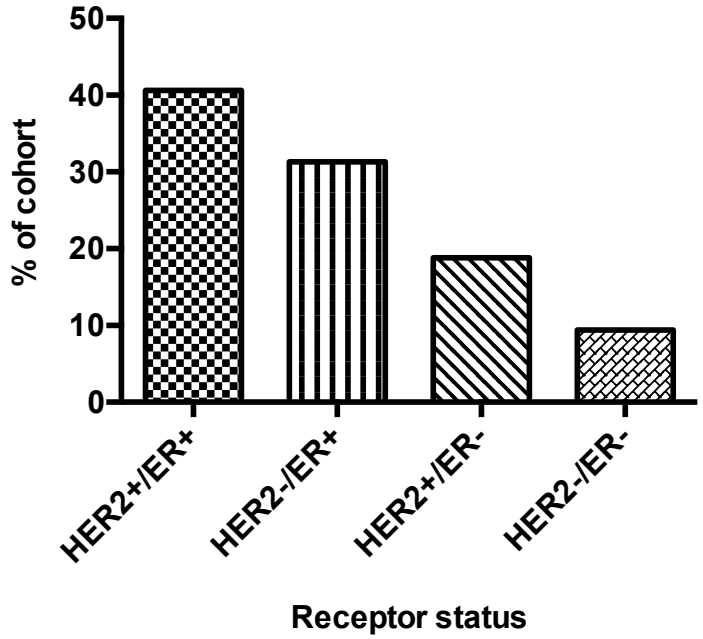
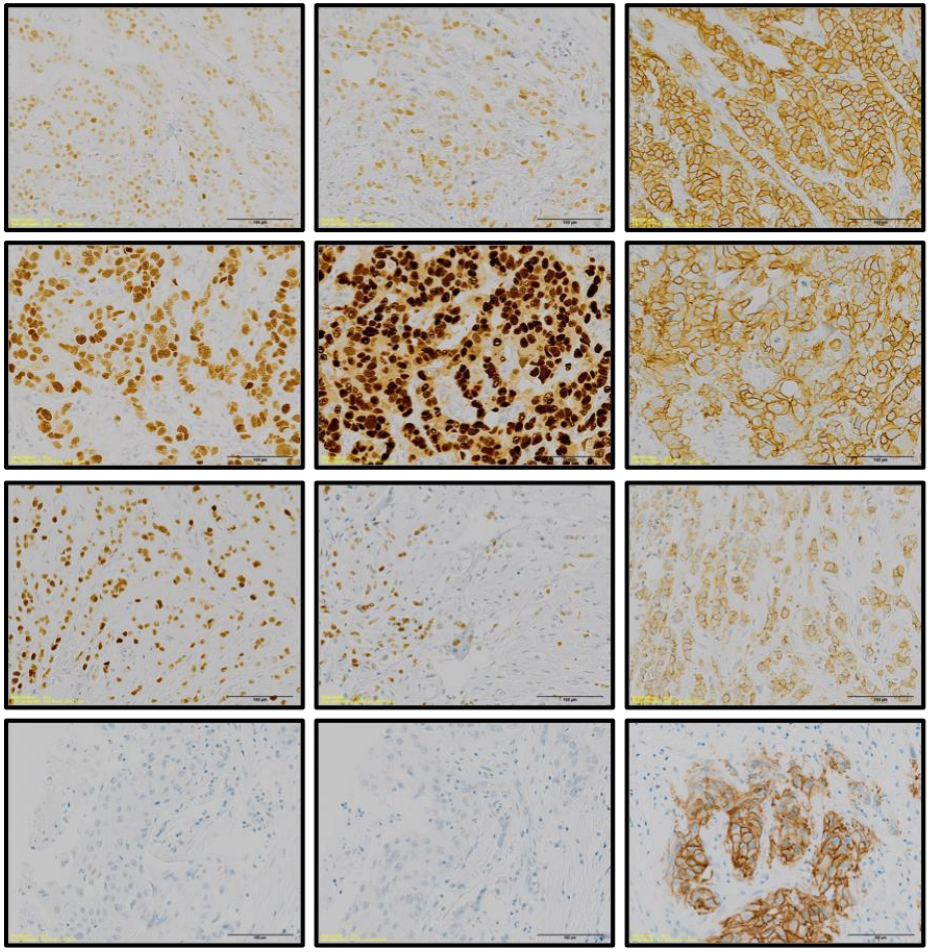


Figure 32 Tumour receptor status in *TP53* carriers

Breast cancer receptor status in early onset breast cancer cohorts. a) Table to show the frequency of ER and HER2 co-expression in *TP53* carriers and the POSH cohort. b) Bar chart to show the incidence of receptor co-expression in *TP53* carriers. The most common receptor status presented in this cohort was HER2+/ER+ positive. HER2 borderline was considered negative. *Statistics used Fisher's Exact.

Statistical significance was only present between *TP53* carriers and the POSH cohort for HER2+/ER+ (p=0.002) tumours. Despite only a small subset of *TP53* carriers developing HER2-/ER- breast tumours, the numbers were not high enough to be deemed statistically significant when compared to the POSH cohort. There was however a difference in the proportion of cases with the p value estimated to be reaching significance (p=0.06). The most common receptor status combinations are presented in fig. 33.



Patient 30091007
TP53 mutation c.659A>G, p.Y220C
(missense)
ER+/PR+/HER2+
(ER: 8, PR: 7, HER2: 3+)

Patient 30124101
TP53 mutation c.993+2T>G, p.?
(splicing)
ER+/PR+/HER2+
(ER: 8, PR: 8, HER2: 3+)

Patient 30091122
TP53 mutation c.743G>A, p.R248Q
(missense)
ER+/PR+/HER2-
(ER: 8, PR: 7, HER2: 2+)

Patient 30091003
TP53 mutation c.112C>T, p.Q38*
(nonsense)
ER-/PR-/HER2+
(ER: 0, PR: 0, HER2: 3+)

Figure 33 Tumour receptor status in *TP53* carriers: the three most common receptor combinations

Common receptor combinations in breast tumours derived from *TP53* carriers.

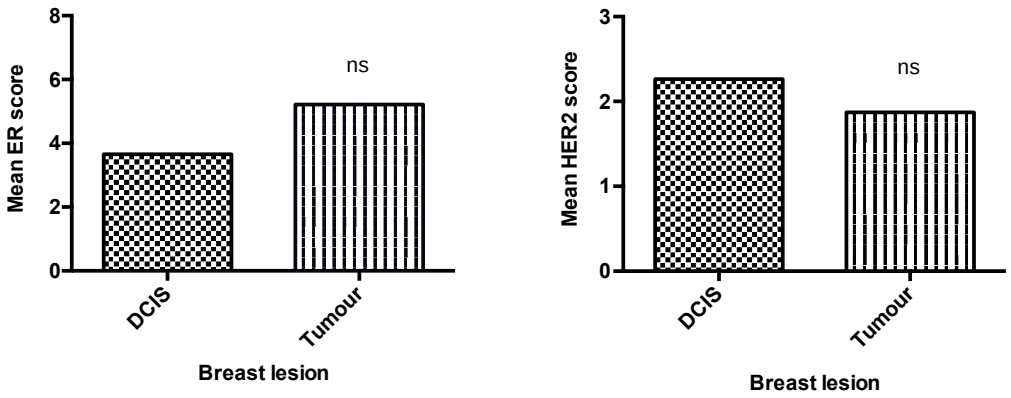
3.3.3 HER2+ status was retained from ductal carcinoma *in situ* (DCIS) to invasive ductal carcinoma in germline *TP53* carriers

Receptor status was compared between the DCIS and tumour tissue. Tumour samples across the cohort had a higher mean score for ER expression and DCIS tissue was shown to have a higher mean score for HER2. No statistical significance was found between the mean scores of HER2 and ER expression between the DCIS and tumour. It appears HER2 and ER expression is maintained in the invasive tumour in *TP53* carriers where HER2 expression is typically lost once disease becomes invasive. This correlation is described in fig. 34.

a)

	Overall cases (Mean score)		Matched cases (Mean score)	
Receptor	**DCIS**	**Invasive tumour**	**DCIS**	**Invasive tumour**
ER+	3.66 (n=36)	5.22 (n=33)	4.15 (n=27)	4.97 (n=27)
HER2+	2.26 (n=39)	1.87 (n=32)	2.14 (n=29)	2.04 (n=29

b)



c)

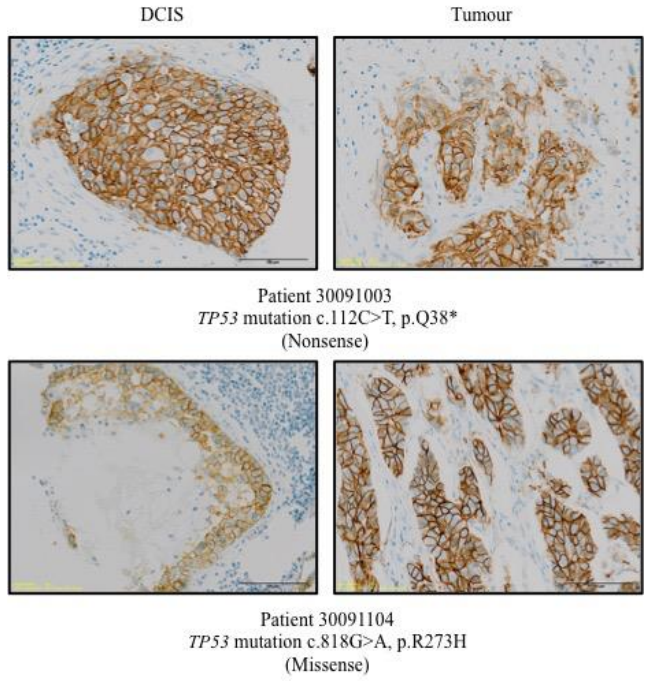


Figure 34 *TP53* carriers maintain HER2 status during tumour progression

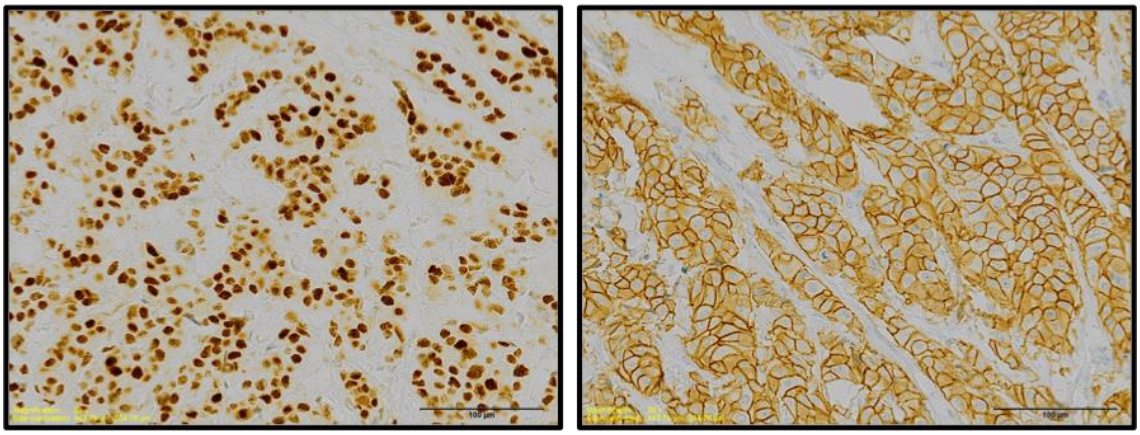
Expression of ER and HER2 in DCIS and tumour. a) Table to show the mean IHC score for ER and HER2 from *TP53* carriers. b) Bar charts to show the mean ER and HER2 expression in DCIS and tumour samples. No significant difference were determined between DCIS and tumour samples. Statistics used Wilcoxon signed rank test. c) HER2 status is retained in the tumour samples. Positive peroxidase staining indicates HER2 overexpression. Images of matched DCIS and tumour samples from TMA cores were taken on the Olympus Dotslide at x20 magnification.

3.3.4 p53 and HER2 expression in germline *TP53* breast tumours

As previously stated, this cohort of breast tumours with a germline *TP53* mutation was typically HER2 amplified and often ER positive. Where data were available, 32/32 (100.0%) of tumour samples were positive for p53 whereas 29/37 (78.4%) of DCIS samples were positive. This positive staining for p53 strongly suggests there is a stabilisation of mutant p53 as in normal tissue; the p53 protein is degraded very quickly resulting in negative IHC. This potentially could be a key step in driving a HER2+ breast tumour phenotype. Matched stained samples are presented in fig. 35.

a)

Patient 30091007, *TP53* mutation c.659 A>G, p.Y220C



p53 staining: 3/4 HER2 status: 3+

b)

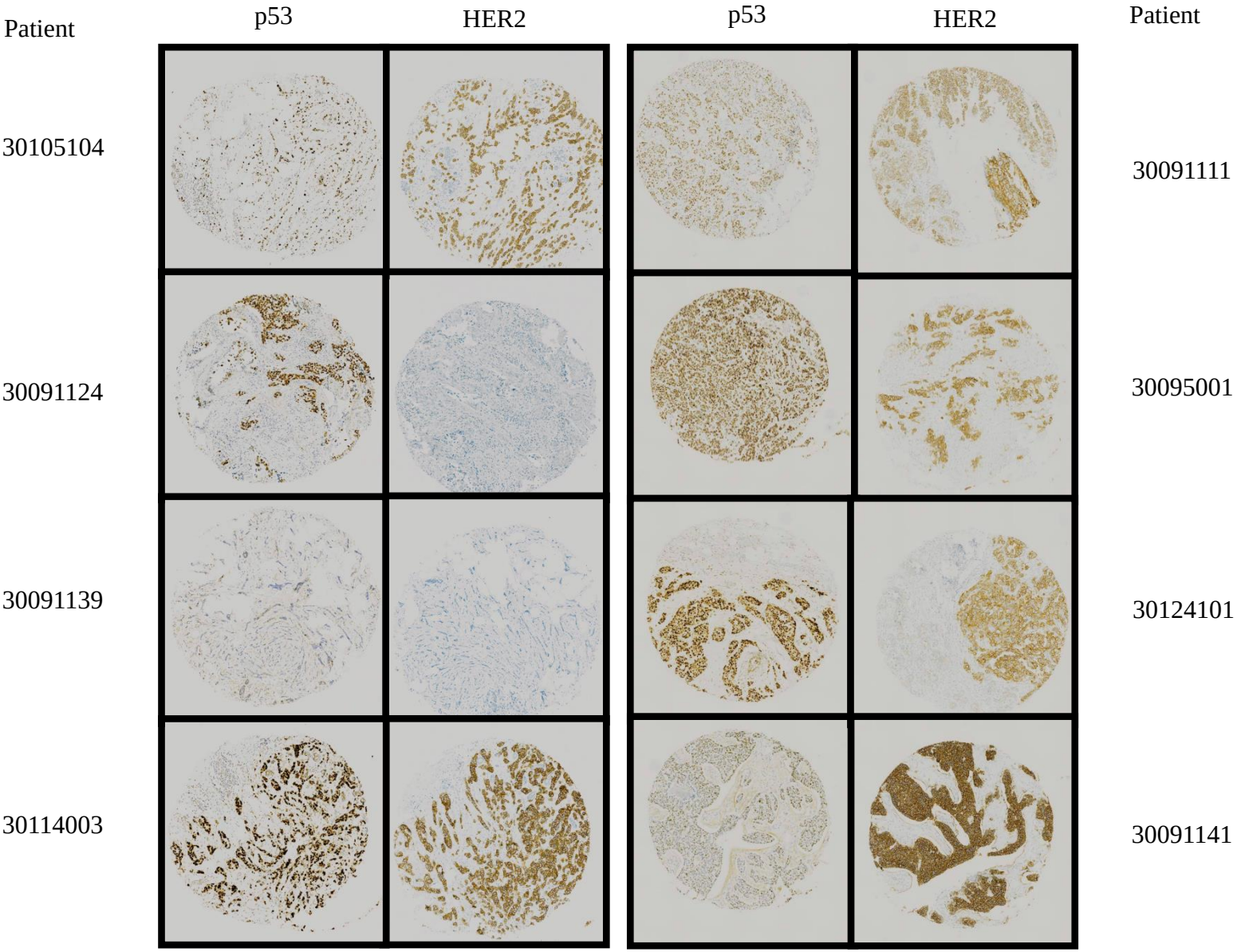

Stain
Patient
p53
HER2
p53
HER2
Patient
30105104
30091111
30091124
30095001
30091139
30124101
30114003
30091141

Figure 35 p53 and HER2 expression in germline *TP53* breast tumours

A high expression of p53 and HER2 staining were reported in breast tumours with an underlying *TP53* mutation. a) Patient 30091007 has a germline *TP53* mutation c.659 A>G, p.Y220C. The tumour showed high expression of p53 (3/4) and HER2 (3+). Images were taken on the Olympus Dotslide, magnification 20x. b) p53 and HER2 immunohistochemistry stains were selected for eight patients. Positive peroxidase staining indicates protein expression. HER2; human epidermal growth factor receptor 2. Images were taken on the Olympus Dotslide

3.3.5 p53 expression in DCIS and breast tumours from *TP53* carriers

The expression of p53 was compared in DCIS and tumour. As previously discussed, 100.0% of tumour samples were positive for p53 whereas 78.4% of DCIS samples were positive. Tumours typically had stronger staining throughout the abnormal cells whilst DCIS was less intense and more patchy throughout the *in situ* disease. Some matched examples are shown in fig. 36.

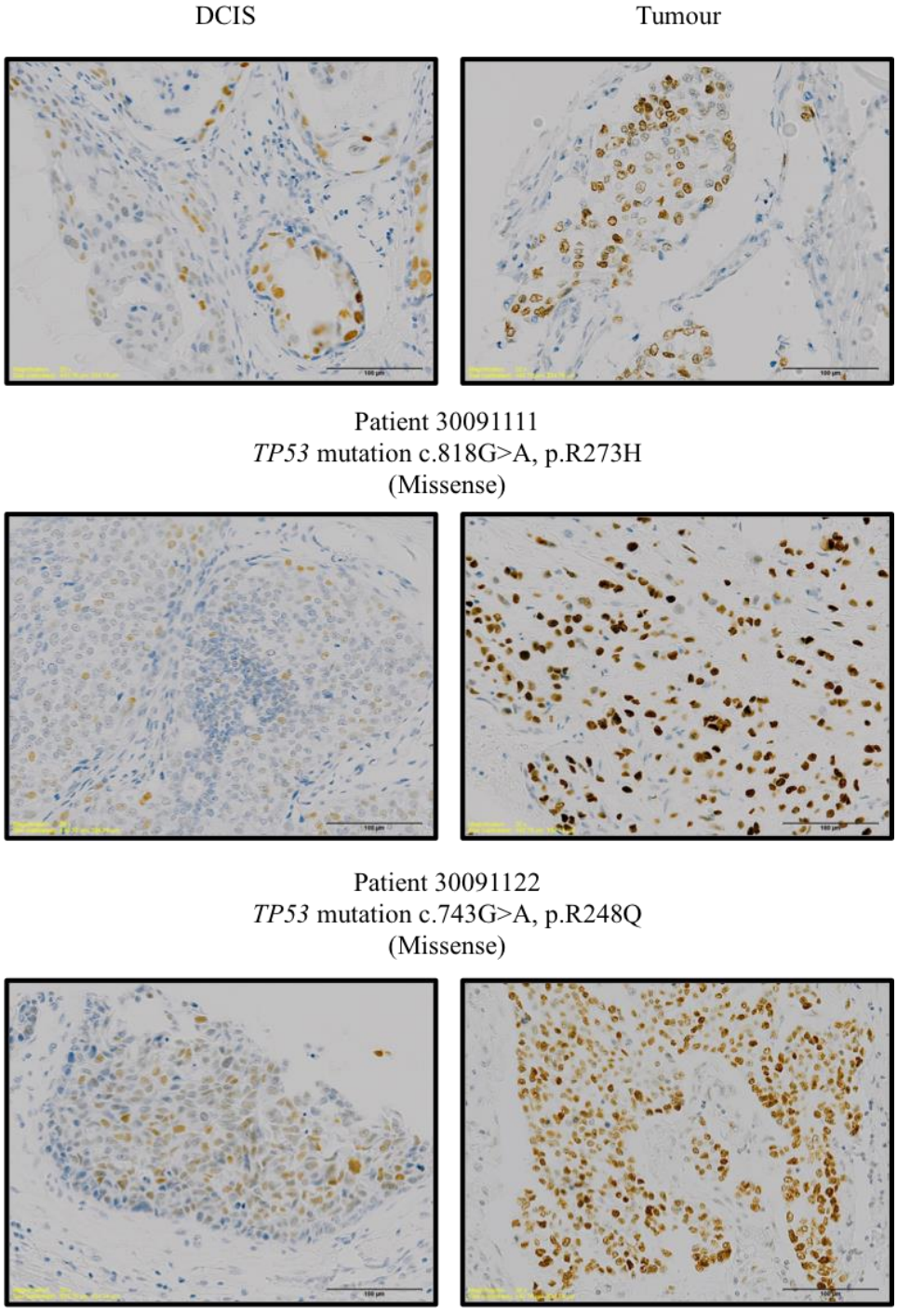


Figure 36 p53 staining was stronger in the tumour than the matched DCIS

Positive staining of p53 in matched DCIS and tumour samples. Higher scores were described in the tumour samples. Positive peroxidase staining indicates p53 expression. Images of spots were taken on the Olympus Dotslide.

TP53 was scored out of 7 (see chapter 2.2.5 for details) and the scores from the DCIS lesions were compared to the matched invasive tissue. Matched p53 data was available for 26 patients in which 74% of those cases expressed higher levels of p53 in the tumour compared to their matched DCIS. Due to the skewed distribution of the scoring, a paired non-parametric t-test: Wilcoxon signed ranks test was used to see if these differences in p53 expression were significant. There was a statistically significant increase in p53 staining in the tumour samples (p=<0.001). The difference in average staining is presented in fig. 37.

a)

	p53 scoring (out of 7)	
Tissue	**DCIS**	**Invasive tumour**
Mean	3.92	5.85
Median	5.0	7.0

b)

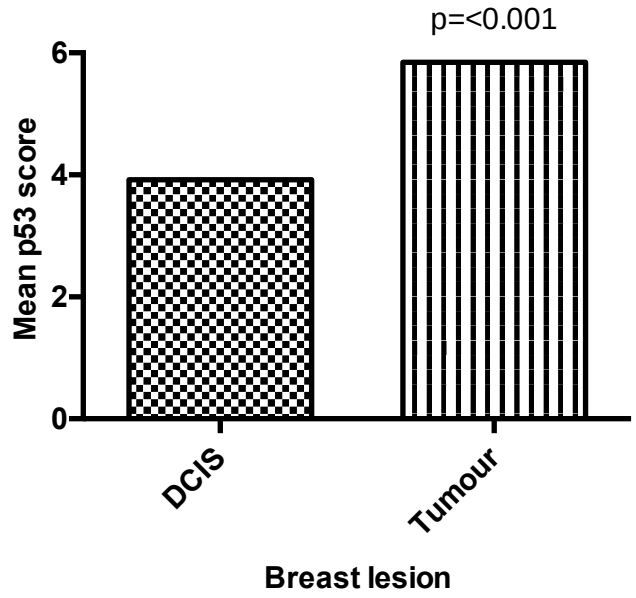


Figure 37 Higher expression of p53 in invasive tumour samples compared to matched DCIS lesions

Positive staining of p53 in matched DCIS and tumour samples. a) Table to show the mean and median p53 score for the DCIS and invasive tumour. b) Bar charts to show the mean p53 expression in matched DCIS and invasive tumour samples. p53 staining was shown to be significantly higher in the invasive tumour compared to the matched DCIS (DCIS median: 5, tumour median: 7, p=<0.001). Statistics used Wilcoxon signed rank test.

This data set suggests that the early stabilisation of the p53 protein is implicated in driving these tumours to become invasive. Potentially a loss of wild type (WT) p53 function could be occurring in this cohort. This would explain the weaker, often patchy p53 staining in the precursor lesion as some of these *in situ* cells still have functional p53.

3.3.6 Type of *TP53* mutation and p53 expression in DCIS and breast tumours from *TP53* carriers

There is some degree of variation within p53 staining, so the next line of enquiry was to examine particular types of *TP53* germline mutations to see if that had an affect on p53 staining. Across the cohort there are missense, splice and nonsense *TP53* mutations and one patient with an entire *TP53* deletion. Typically, those patients with a missense or a splicing *TP53* mutation scored more highly for p53 IHC particularly in the tumour. Cases with germline nonsense mutations had less intense p53 staining in comparison. One patient, whose germline mutation was a whole *TP53* deletion, had no p53 staining in the DCIS and this had not yet progressed to invasive disease. We speculated that these observations were comparable with the hypothesis that a loss of the WT p53 function is a critical step in the progression to invasive disease. The differences in p53 staining depending on the type of mutation are described in fig. 38.

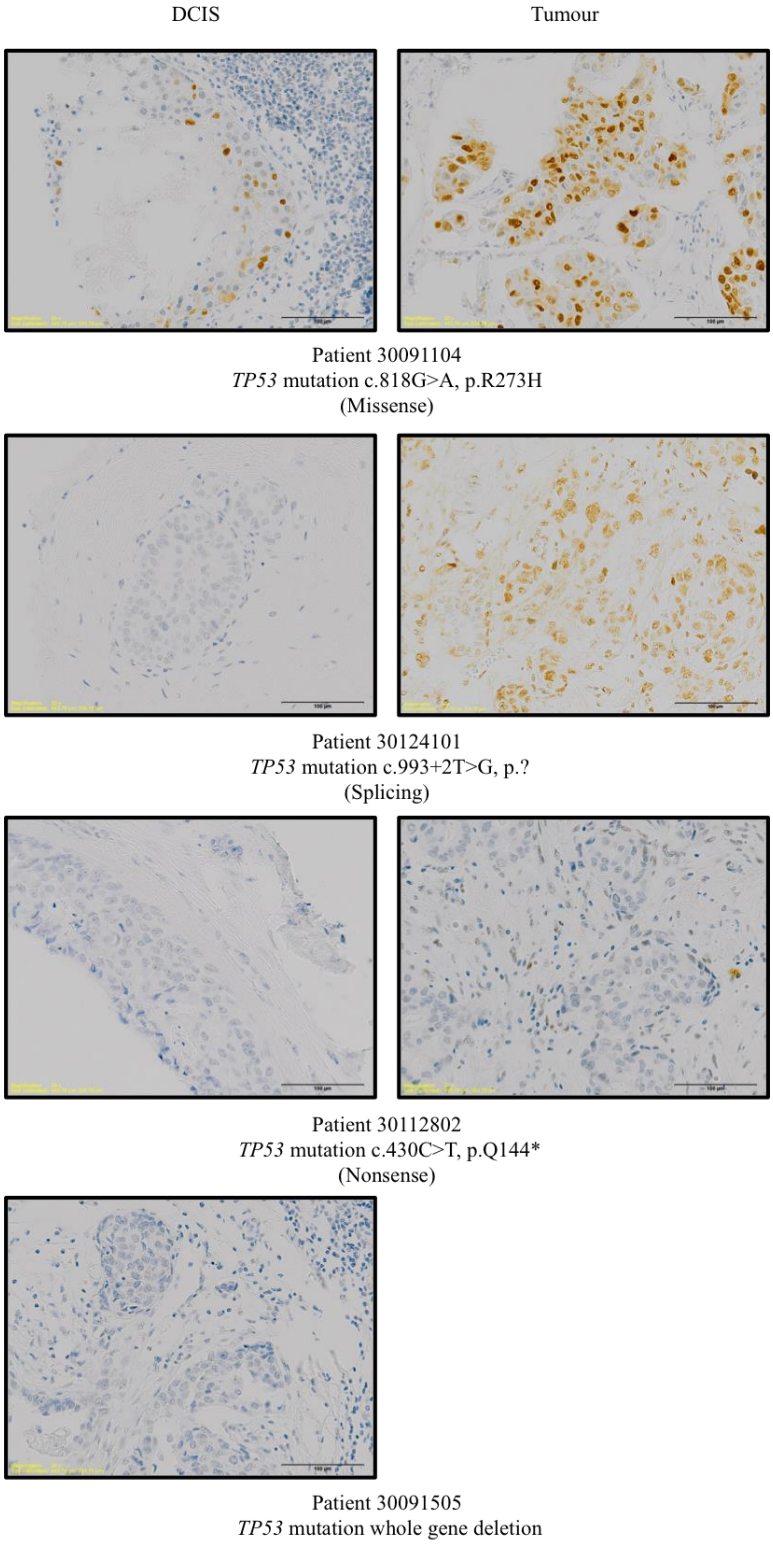


Figure 38 The type of *TP53* mutation leads to different p53 expression

Different types of *TP53* mutation result in different levels of p53 expression using IHC. Patients with a missense or splicing mutation typically had high p53 expression in their tumours. Nonsense mutations resulted in low p53 expression and a patient with a whole *TP53* mutation had no p53 staining. Positive peroxidase staining indicates p53 expression. Images of the cores were taken on the Olympus Dotslide.

3.4 Immunohistochemistry: αvβ6 and α-SMA expression

3.4.1 The effect of the stroma and TGFβ: αvβ6 and α-SMA expression

Following on from the observation of a high frequency of sclerotic tumour stroma in *TP53* carriers (section 3.1.3), the expression of integrin αvβ6 and α-SMA were investigated to try and understand the mechanism behind the development of this type of stroma. The expression of these proteins were investigated because of their role in driving a sclerotic tumour stroma. Integrin αvβ6 is considered a major activator of TGFβ and this cytokine drives myofibroblast differentiation, the cellular component of the microenvironment that deposit the rich collagen layer defining a sclerotic stroma [159-161, 163]. These cells are typically identified through their α-SMA expression [162]. See section 1.8 for further information.

Expression of αvβ6 and α-SMA were scored in the DCIS and the invasive tumour tissue in *TP53* carriers using IHC on TMA cores. Samples were scored as absent, low, moderate or high. Those samples that presented with a sclerotic tumour stroma consistently expressed at least a moderate level of αvβ6 and α-SMA. This would suggest that the stroma, particularly through TGFβ signalling, is playing a key role in promoting tumorigenesis in these cases. Data from patient 30112802 is presented in fig. 39.

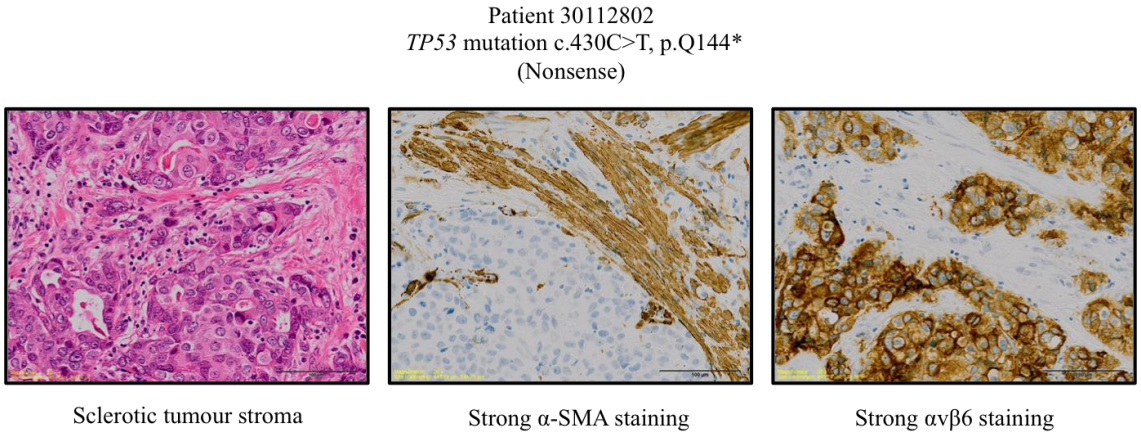


Figure 39 αvβ6 and α-SMA expression in germline *TP53* breast tumours

A sclerotic tumour stroma was consistent with moderate/high α-SMA and αvβ6 immunohistochemistry stains. Positive peroxidase staining indicates protein expression. αvβ6; integrin alpha v beta 6, α-SMA; alpha smooth muscle actin.

The level of α-SMA and integrin αvβ6 were compared between DCIS and tumour tissue. Scores were grouped either absent/low or moderate/high. Marginal differences were seen in integrin

αvβ6 expression between the DCIS and invasive tumour but this was not statistically significant. Statistical significance was confirmed for α-SMA expression (p=0.047) (fig. 40).

a)

	Tissue	Stain Intensity		
		Absent/low (% of cohort)	Moderate/high (% of cohort)	Missing data
αvβ6	DCIS	14/41 (34.15%)	21/41 (51.22%)	6/41 (14.63%)
	Tumour	11/36 (30.56%)	21/36 (58.33%)	4/36 (11.11%)
α-SMA	DCIS	6/41 (14.63%)	24/41 (58.54%)	11/41 (26.83%)
	Tumour	1/36 (2.78%)	32/36 (88.89%)	3/36 (8.33%)

b)

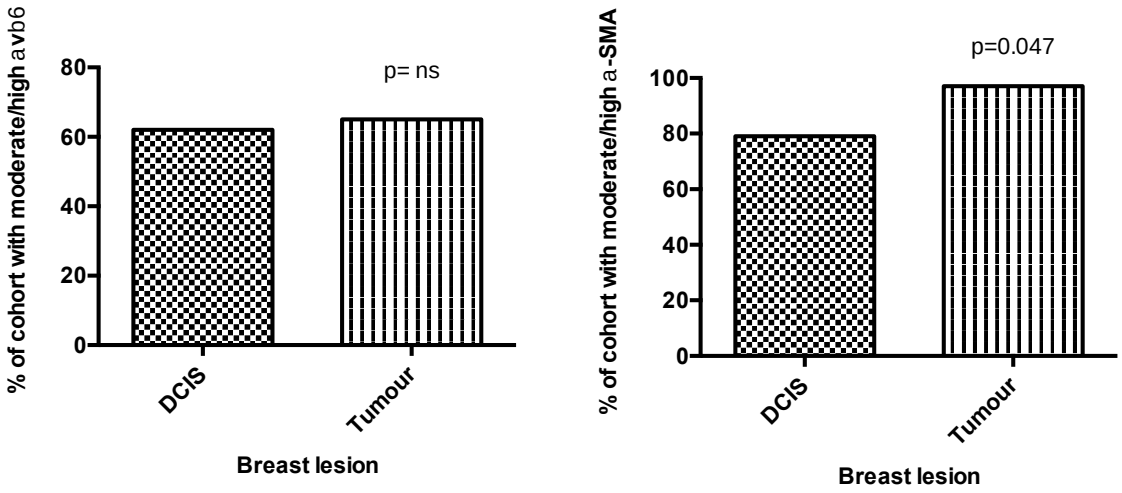


Figure 40 High expression of αvβ6 and α-SMA expression in germline *TP53* breast tumours: a comparison between DCIS and tumour

Invasive tumour and DCIS in *TP53* carriers typically expressed moderate to high expression of αvβ6 and α-SMA. a) Table to show expression of integrin αvβ6 and α-SMA in DCIS and invasive tumour. b) Graphs to show moderate/high expression of α-SMA and αvβ6 in DCIS and invasive tumour. No statistical significance was detected for αvβ6 expression but significance was determined for α-SMA (p=0.047). Statistics used Fisher's Exact test. Missing data was excluded from statistics. αvβ6; integrin αvβ6, α-SMA; alpha smooth muscle actin.

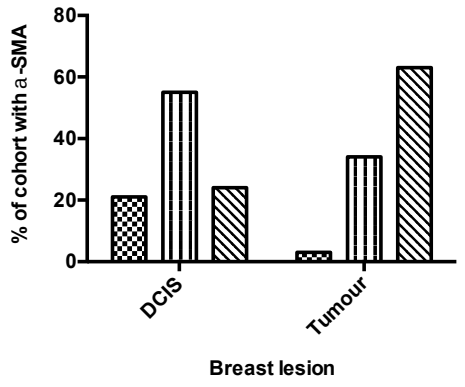
3.4.2 The effect of the stroma and TGFβ: Differences in moderate and high α-SMA in breast lesions

Whilst scoring this cohort, clear differences between the frequency of moderate to high α-SMA staining was observed between the precursor lesion and the invasive tumour. Although moderate staining is considered positive, it is worth noting that a higher incidence of moderate staining was scored in the DCIS compared to the tumour. Furthermore, when only high expression of α-SMA was compared between the DCIS and tumour samples, the significance was larger between the two breast lesions (p=<0.001). Fig. 41 shows these differences.

a)

***TP53* carrier breast lesion**	**α-SMA (% of cohort)**		
	Absent/low	**Moderate**	**High**
DCIS	6/30 (20.0%)	17/30 (56.7%)	7/30 (23.3%)
Invasive tumour	1/33 (3.0%)	11/33 (33.3%)	21/33 (63.6%)

b)



c)

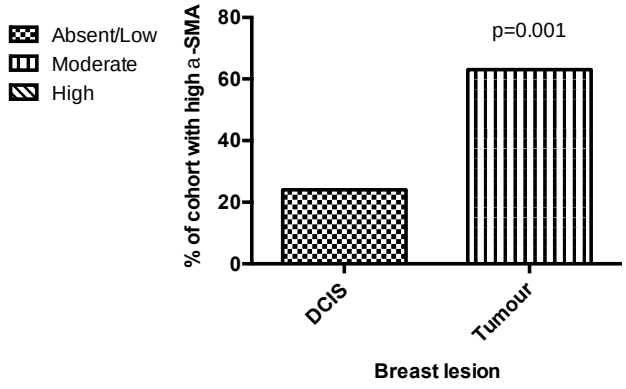


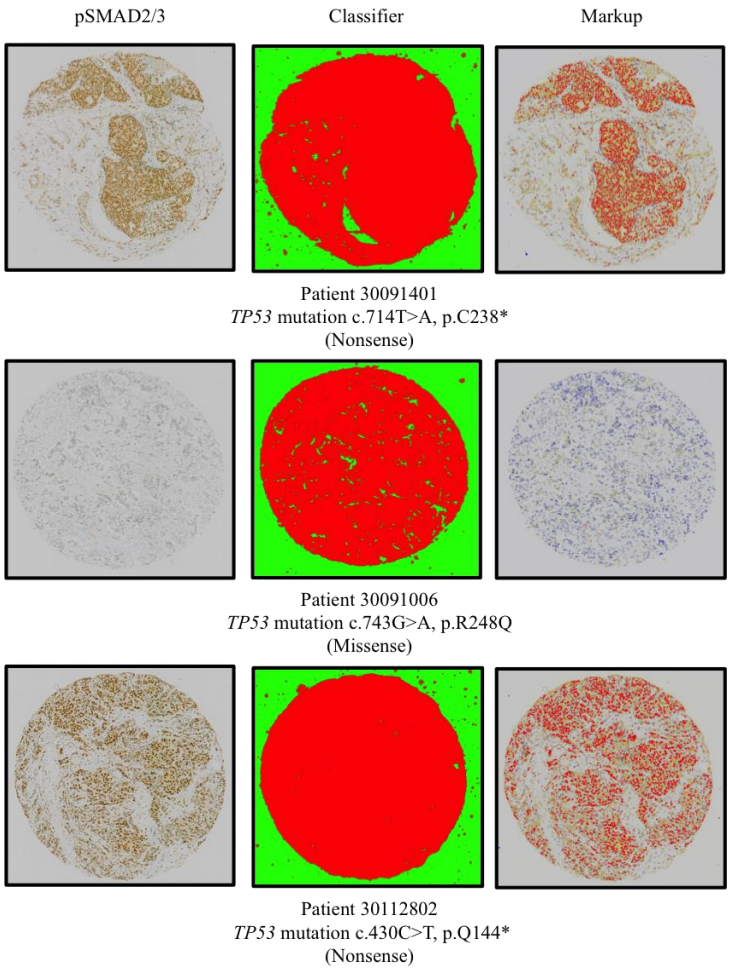
Figure 41 High expression of α-SMA expression in germline *TP53* breast tumours: a comparison between DCIS and tumour

DCIS tissue typically expressed moderate staining of α-SMA rather than high. a) Table to show differences in moderate and high α-SMA staining in DCIS and invasive tumour samples. b) Graph to show the spectrum of staining in DCIS and tumour samples in *TP53* carriers. Tumour samples typically expressed high levels of α-SMA whereas DCIS usually expressed only moderate expression. c) Graph to show significant difference in high α-SMA expression between breast lesions (p=0.001). Statistics used Pearson Chi Square test. α-SMA; alpha smooth muscle actin.

3.5 Immunohistochemistry: pSMAD upregulation of TGFβ signalling

The final line of enquiry into the mechanism for stromal development was to determine if TGFβ signalling was playing an active role in these tumours and potentially driving tumorgenesis. In order to investigate TGFβ signalling, TMA sections were stained for pSMAD2/3. These proteins are activated through a phosphorylation cascade which forms the basis of initiated TGFβ signalling (see chapter 1.8.2 for further information). Once activated, these proteins migrate from the cytoplasm to the nucleus and initiate transcription and the downstream deposition of collagen. TMA sections were scored separately as a proportion of stained cells and the staining intensity. The proportion of cells were scored 1-4 based on the quartile staining and the intensity was scored weak, moderate and strong depending how the majority of the cells in that core were scored. The digital pathology Halo Image Analysis software was trained and used to analyse this staining (fig. 42).

a)



b)

Patient	Tissue	% Positive cells	Proportion score	% Weak staining	% Moderate staining	% Strong staining	Staining intensity score
30091401	DCIS	93.6	4	33.1	28.1	32.4	1
30091006	Tumour	27.5	2	24.2	3.2	0.1	1
30112802	Tumour	95.7	4	21.5	34.4	39.9	3

Figure 42 pSMAD2/3 staining and analysis using Halo Image Analysis

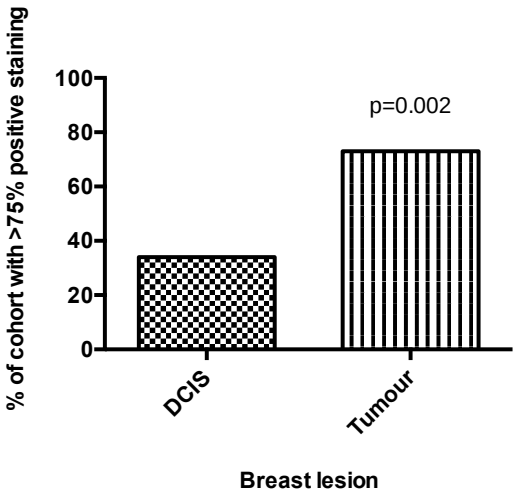
Each core was scanned using an Olympus Dotslide at a x40 magnification and the images were processed through the digital pathology Halo Image Analysis software. The software was trained to score all cell types in each core and produce an output analysis file. a) 3 cores and their analysis classifier and markup files. The classifier shows the area scored (red) and the white space that was excluded from the scoring statistics (green). The markup file shows how the cores were scored with strong stained cells displayed as red, moderate staining shown as orange, weak staining yellow and negative cells as blue. b) Scoring statistics from the above cores. The proportion and staining intensity were scored based on the % of positive cells and the highest frequency of intensity. Typically DCIS and tumour scored more highly than stromal cells.

All cores were scored for pSMAD2/3 expression to check for activation of TGFβ signalling. Each core was checked for tissue content, classifier and markup. Cores that either contained normal tissue, damaged or the software had difficulties scoring were excluded. Data is shown in fig. 43.

a)

	Scoring	DCIS (% of cohort)	Tumour (% of cohort)
Staining proportion	1	2/41 (4.9%)	1/36 (2.8%)
	2	2/41 (4.9%)	0/36 (0.0%)
	3	19/41 (46.3%)	7/36 (19.4%)
	4	13/41 (31.7%)	23/36 (63.9%)
Staining intensity	1	4/41 (9.8%)	2/36 (5.6%)
	2	27/41 (65.9%)	20/36 (55.6%)
	3	5/41 (12.2%)	9/36 (25.0%)
Missing data		5/41 (12.2%)	5/36 (13.9%)

b)



c)

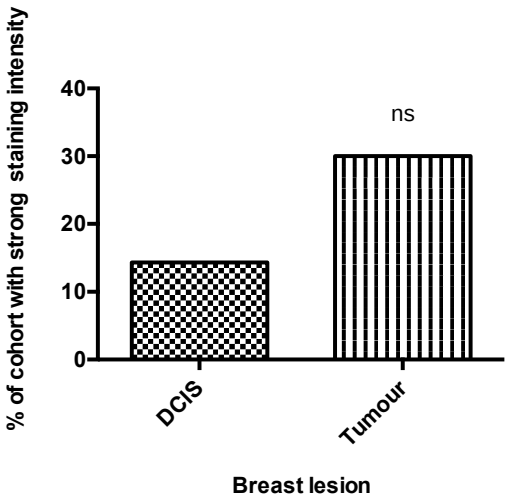


Figure 43 pSMAD2/3 staining in breast and DCIS tissue from *TP53* carriers

Each core was scored following analysis on the Halo Image Software. a) Table to show pSMAD staining: intensity and proportion. b) Tumour cells were shown to have a significantly higher proportion of positive staining (p=0.002). c) Tumour cells often had a higher staining intensity but this was not statistically significant.

3.6 Immunohistochemistry case summary

Protein express using immunohistochemical analysis was carried out on DCIS and invasive tumour tissue from germline *TP53* carriers. Fig. 44 shows an overview from those *TP53* carriers.

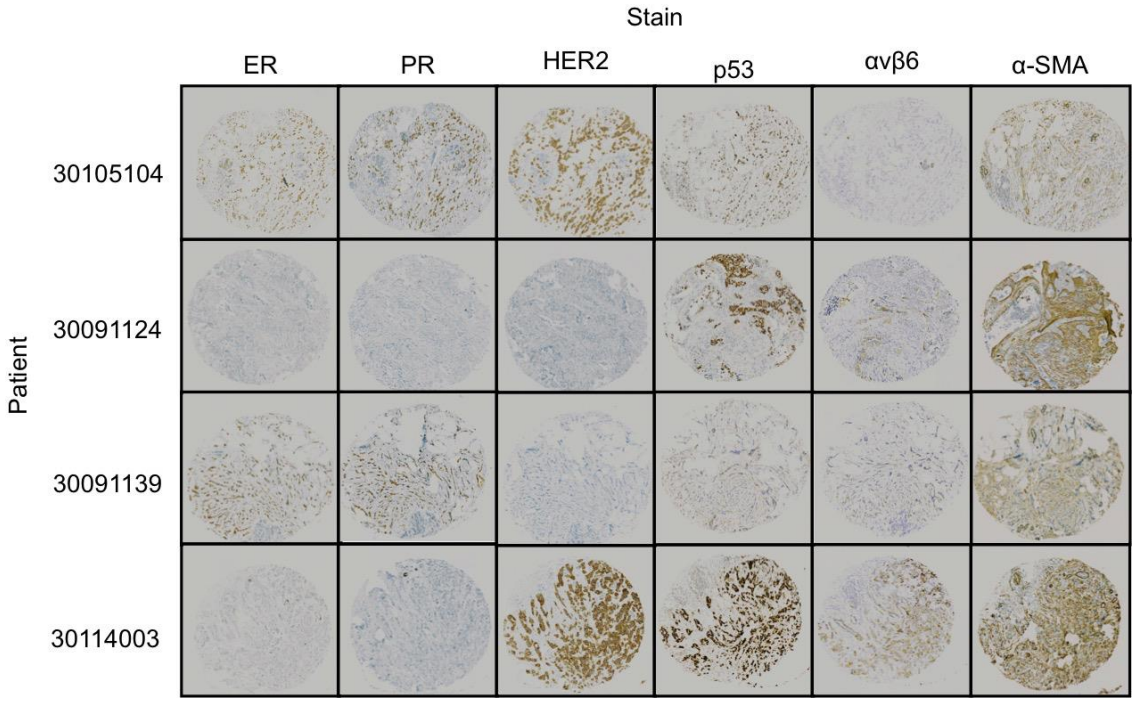
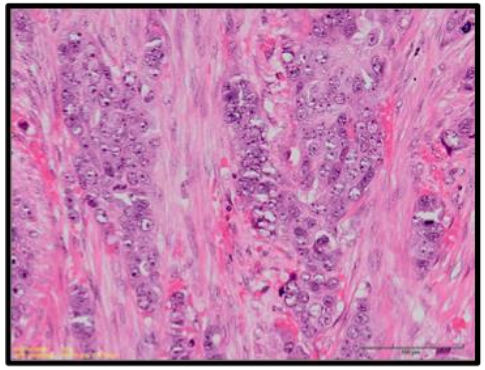


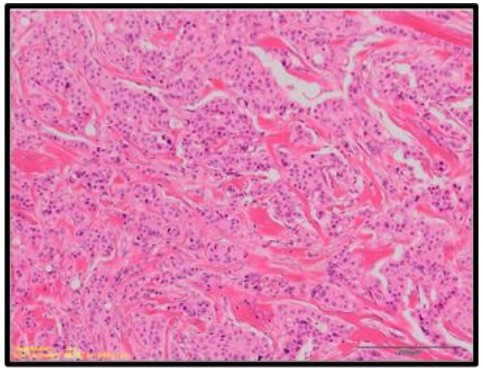
Figure 44 Immunohistochemistry summary: 4 cases

3 tumour cores were taken from patients 30105104, 30091124, 30091139 and 30114003. These were stained for ER, PR, HER2, p53, αvβ6 and α-SMA. ER; oestrogen receptor, PR; progesterone receptor, HER2; human epidermal growth factor receptor 2; αvβ6, αvβ6 integrin; α-SMA, alpha smooth muscle actin.

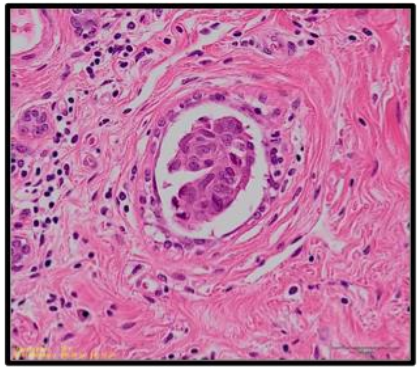
	Tumour (% of cohort)						DCIS (% of cohort)		
	Ductal (NST)	**Grade 3**	**Infiltrative border**	**Sclerotic stroma**	**Prominent lymphocytic infiltration**	**Present vascular invasion**	**Present DCIS**	**High grade DCIS**	**Solid and comedo DCIS**
TP53	32/36 (88.9%)	18/36 (50.0%)	36/36 (100.0%)	29/36 (80.6%)	6/36 (16.7%)	12/36 (33.3%)	41/45 (91.1%)	40/45 (88.9%)	19/41 (46.3%)



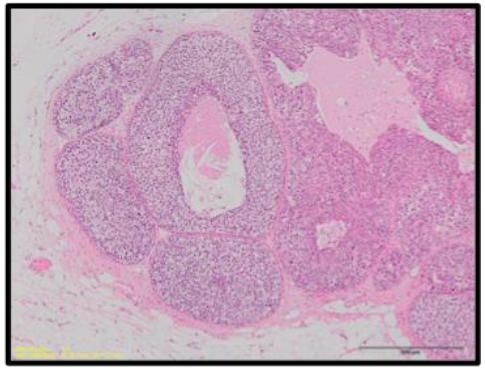
Grade 3 ductal carcinoma



Sclerotic tumour stroma and absent lymphocytic infiltration

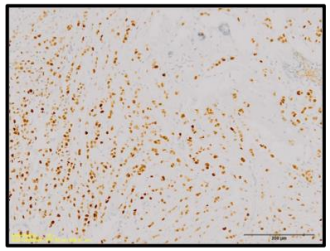


Vascular invasion

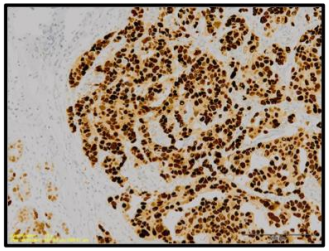


Solid and comedo DCIS

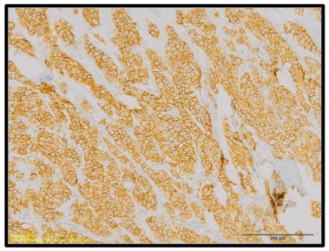
	Stains (% of *TP53* carriers)							
	ER+	**PR+**	**HER2+**	**Mean p53 score**	**Integrin αvβ6 (moderate/high)**	**α-SMA (moderate/high)**	**pSMAD2 proportion >75%**	**pSMAD2/3 intensity 3**
DCIS	21/41 (51.2%)	15/41 (36.6%)	27/41 (65.9%)	4/7 (n=26)	21/41 (51.2%)	24/41 (58.5%)	13/41 (31.7%)	5/41 (12.2%)
Tumour	24/36 (66.7%)	21/36 (58.3%)	19/32 (52.8%)	6/7 (n=26)	21/36 (58.3%)	32/36 (88.9%)	23/36 (63.9%)	9/36 (25.0%)



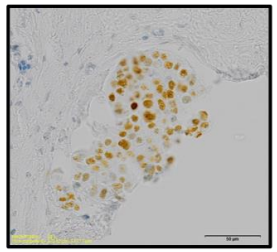
ER positive



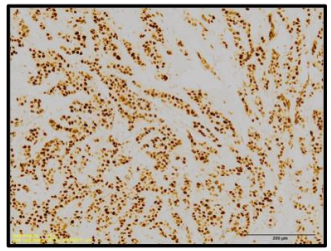
PR positive



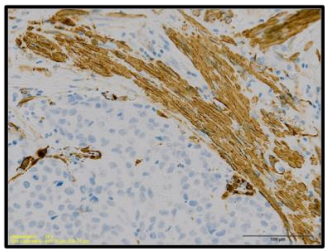
HER2 positive



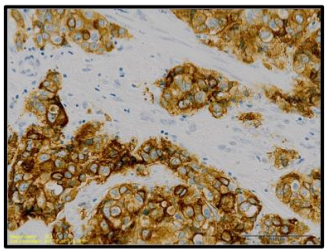
DCIS: moderate p53 staining



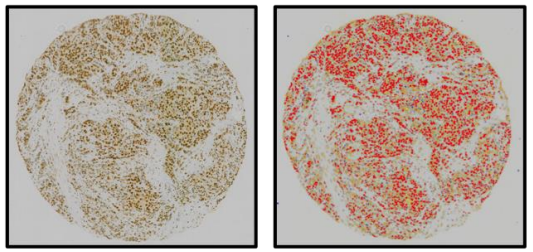
Tumour: strong p53 staining



α-SMA positive



Integrin αvβ6 positive



Activation of TGFβ signalling: pSMAD2/3 positive

Figure 45 Summary of the morphology and immunohistochemistry features of *TP53* carriers

The morphology and immunohistochemistry features were investigated in the DCIS and invasive tumour. ER; oestrogen receptor, PR; progesterone receptor, HER2; human epidermal growth factor receptor 2, αvβ6; αvβ6 integrin, α-SMA; alpha smooth muscle actin, pSMAD2/3; phospho-SMAD2

3.7 Discussion

Breast tumours derived from germline *TP53* carriers were typically HER2+/ER+ high grade ductal no special type tumours associated with widespread high grade DCIS. 34/36 (94.4%) of the cohort were either grade 2 or grade 3 tumours but when this scoring was investigated in more detail, many of the cohort were awarded a 3 for pleomorphism, 3 for tubule formation and only a 1 for mitotic count. A significant proportion of cases that were scored in this manner, were poorly fixed and this had a possible impact on correct grading of breast tumour samples. Poor fixation often results in the underscoring of tumour grade due to very low or absent mitotic figures. Three elements need to be taken into account for correct fixation: the thickness of the tissue, type and volume of the fixative, for example if buffered formalin has been used, and the time in which it has been left in the fixative. If particularly thick tissues are used, tissue at the centre has often stopped proliferating by the time the fixative has reached it. Therefore, for at least a number of germline *TP53* cases, tumours were under scored as a 2 rather than a 3 due to poor fixing. An example of a particularly poorly fixed case from this cohort is shown in fig 46.

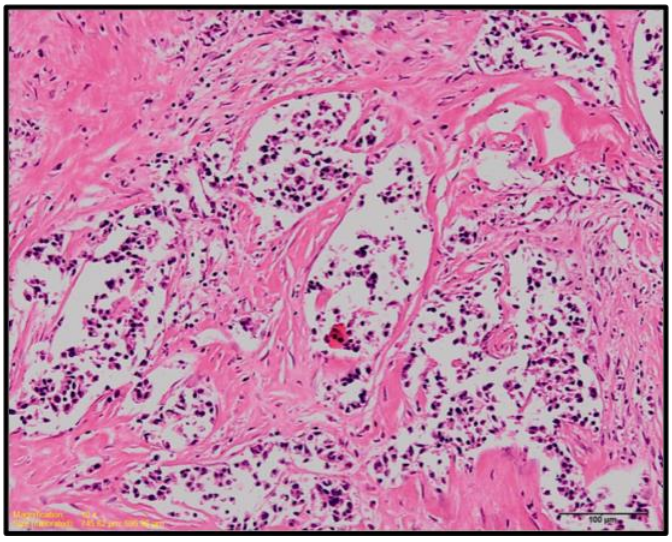


Figure 46 Poor fixing in the tumour samples of *TP53* carriers

Grading of some patient samples was difficult because of poor fixing. 30091006 shown above was an extreme case but prolonged exposure to formalin prevents accurate grading due to the loss of mitoses.

A feature of the cohort was a high frequency of sclerotic tumour stroma identified by a rich collagen layer. A reactive tumour stroma has previously been implicated in driving tumour invasion and vascular invasion [146-148]. Data from this study would support the finds of other groups as a higher incidence of vascular invasion was identified in patients with a germline *TP53* mutation compared to control groups. Additionally, groups have suggested the role a collagen rich stroma plays in preventing entry of TILs by acting as a physical barrier [191]. Only a small proportion of this cohort was scored as having a prominent lymphocytic infiltration so perhaps in breast tumours derived in germline *TP53* carriers, their *TP53* mutation is driving a sclerotic stroma and in turn is preventing the infiltration of immune cells. Together a picture emerges from the morphology review of a very aggressive breast tumour subtype with predicted poor prognosis inline with clinical data from Li-Fraumeni Syndrome survival studies.

Additionally, a striking high grade DCIS field effect is consistently observed in *TP53* carriers. These lesions are noticeably more widespread and larger than any of the control groups. Nearly half of all cases had both solid and comedo DCIS growth patterns associated with high grade and invasion potential [12, 13]. The high incidence of comedo DCIS would suggest that these lesions are highly proliferative due to the size and extent of necrosis in lesions. Despite comedo DCIS not undergoing assessment in the POSH study, one could predict that the frequency of comedo DCIS would be lower due to a lower incidence of high grade DCIS reported [81, 82]. In order to address this proliferative question, further quantification would be required between groups to measure the average size of DCIS lesions that present in these patients. Additionally using IHC and staining for Ki67 would address how proliferative these lesions are. Studies such as this would also allow the

speed in which these lesions develop in young breast cancer subgroups. A cohort with very little DCIS and a ductal (NST) tumour type would suggest that disease progression and invasion has been fast.

IHC was used to investigate p53 expression in DCIS and the invasive tumour. p53 IHC was positive in all tumour samples, which would suggest a loss of wild type p53 function and stabilisation of the mutant. Interestingly, p53 staining was present in many DCIS lesions but at a lower intensity and to a patchy extent across lesions. Interpretation of this data would indicate that a stabilisation of mutant p53 is underway in some clones in the DCIS. Consequently this would implicate this loss of WT p53 function is a key event driving disease progression from *in situ* to invasive ductal carcinoma.

Additionally, the patchy nature of p53 staining in the DCIS would support previous work by Heselmeyer-Haddad and Hernandez that DCIS is a heterozygous lesion containing distinct clones [132, 133]. This heterogeneity of DCIS clones would in addition suggest that an evolutionary bottleneck is underway rather than a convergence model of evolution [130, 131]. See chapter 1.7 for further information.

Following on from this line of enquiry, HER2 status was investigated in parallel. HER2 expression is a common feature of DCIS but during tumour progression, most matched invasive lesions are HER2 negative [118, 126, 127]. A feature of *TP53* carriers is that they retain this HER2 amplification. The fact we are seeing what appears to be the start of a loss of WT p53 function in the DCIS and retained HER2 status in the invasive tumour, would suggest that these HER2+ clones in the DCIS are losing functional p53. This could be giving these clones a selective and invasive advantage. Therefore this evolution bottleneck proposed to describe tumour progression and the patchy p53 staining in the DCIS, would support why patients with a germline *TP53* go on to develop HER2+ invasive ductal carcinoma.

The cohort consists of patients with various types of *TP53* mutations including missense, nonsense, splicing, and frameshift mutations. Interestingly differences were observed in the intensity of p53 staining between the types of mutations. Stronger staining was observed in tumours with a missense mutation and this could be because of a gain of function (GOF) associated with this type of mutation or may be purely because the antibody has a lower binding affinity in truncating mutations.

Lastly the expression of integrin αvβ6, α-SMA and pSMAD2/3 was investigated to see if TGFβ signalling was upregulated in these tumours to give some insight as to what is driving this sclerotic tumour stroma. It is well established the role integrin αvβ6 plays in driving myofibroblast differentiation through TGFβ activation and as a consequence, the production of a sclerotic stroma [159-161, 163]. Through IHC analysis, *TP53* carriers were confirmed to have high α-SMA and αvβ6 expression. Excluding missing data, 65.6% (21/32) of tumour cases were confirmed to have

moderate to strong integrin αvβ6 staining. A study of two cohorts containing over 2000 women found that only 15-16% of invasive ductal carcinomas expressed this integrin suggesting that this could be a *TP53* driven phenomenon [168]. DCIS IHC data was comparable with tumour samples so a possible interpretation of this is that αvβ6 expression is an early event possibly initiated through loss of p53 function. Alternatively, this could be a consequence of catastrophic early copy number events that have previously been reported in triple negative breast cancers [192]. Integrin αvβ6 expression has previously been implicated in DCIS progression so this could be an additional mechanism in which DCIS in *TP53* carriers breaks through the myoepithelial barrier becoming invasive [164, 165].

In conclusion breast tumours derived in a *TP53* background were typically high grade HER2+/ER+ ductal carcinomas with associated high grade DCIS. A high frequency of sclerotic stroma and high expression of integrin αvβ6, α-SMA and pSMAD2/3 suggests that TGFβ signalling is playing a role in tumourogenesis. Positive p53 IHC in the tumour and less intense more patchy staining in the DCIS is suggestive of early loss of WT p53 function.

Chapter 4: The Genomics of breast cancers derived in *TP53* carriers

From the morphology review, *TP53* carriers typically developed HER2+ breast tumours with widespread DCIS. We wanted to determine the mutation spectrum in the COPE tumours and the timing of genomic lesions driving tumour evolution. DNA was extracted from patient FFPE material following the extraction protocol outlined in the methods (chapter 2.3.3 and 2.3.4). We compared COPE data against 9 sporadic HER2 amplified tumours in matched young patients from the POSH cohort. These control cases were selected based on firstly their HER2+ phenotype, secondly their associated DCIS and finally the availability of germline DNA for that case. These particular HER2+ cases with DCIS were selected in order to investigate early molecular events and to test the hypothesis that an early loss of *TP53* was critical in driving a HER2+ invasive breast phenotype. Due to the poor quality of DNA derived from archival FFPE material, the final number of cases available for next generation sequencing was substantially reduced. The cases are summarised in table 17.

	COPE		**POSH**		
	Tumour	**DCIS**	**Tumour**	**DCIS**	**Blood**
Total cases available	24	9	9	8	9
DNA targeted sequence data available	16	4	9	0	9
Cases with multiple sequence data sets to assess evolution	4		0		

Table 17 DNA Summary of cases

Table to show cases and data availability.

The availability of blood DNA for the POSH cases was a great advantage allowing us to exclude germline variants from somatic mutation analysis. Germline material was extracted from surrounding normal breast tissue, skin and lymph nodes (tumour negative) but this was of too poor quality to process for NGS (see appendix, chapter 7.2).

4.1 Challenges of using archival FFPE derived DNA for targeted next-generation sequencing

DNA quality is important in determining the likelihood of successful sequencing particularly when DNA is derived from FFPE. The target enrichment approach required the use of the Nanodrop to gain an insight into DNA purity, Qubit for an accurate reading of pure double stranded DNA (dsDNA) and a PCR based assay to allow an assessment of how amplifiable the template would be.

Bioinformatics assessment of sequence variants had to take into account the high likelihood of false positive calls due to fixation artefacts. Using a higher input of DNA for these poorer samples were shown to optimise for fewer suspect false positives called once these were visualised in IGV. This correlation between the number of variants called, versus the DNA input for samples of a level B fragmentation category and below can be seen in fig. 47.

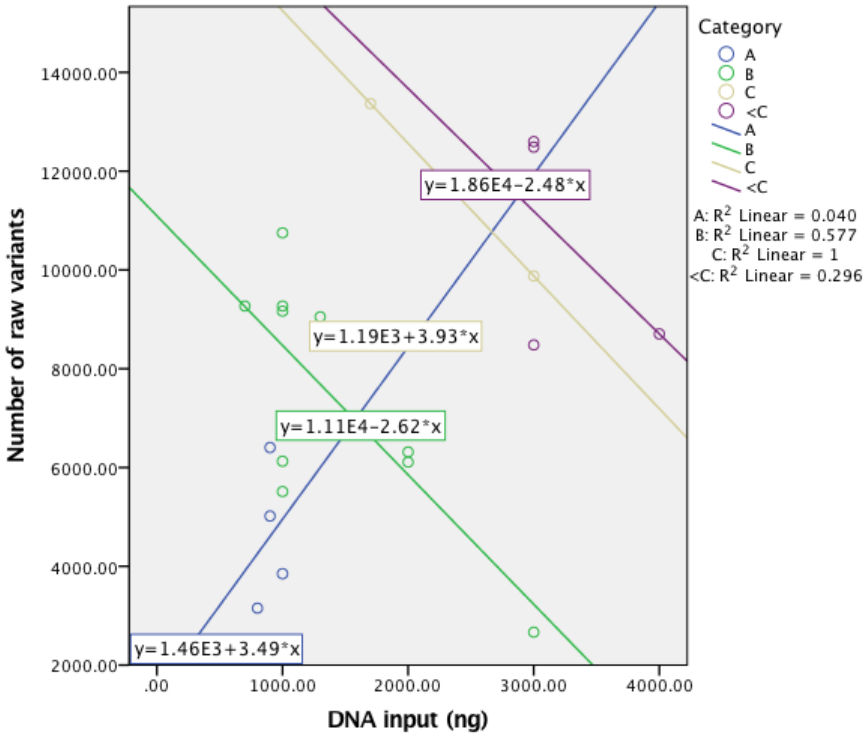


Figure 47 A higher input of DNA reduces the chance of false positive calling

A correlation was observed between the amount of DNA input and the number of suspect false positives in samples that fall into a category B or below for fragmentation. Quality category: blue; A, green; B, yellow; C, purple; <C.

During this project, a number of selection criteria were introduced due to a large number of samples failing quality control. This was especially important as the DNA quantity required was

greater if the DNA integrity was low. An example of a poor and good quality sample can be seen in fig. 48.

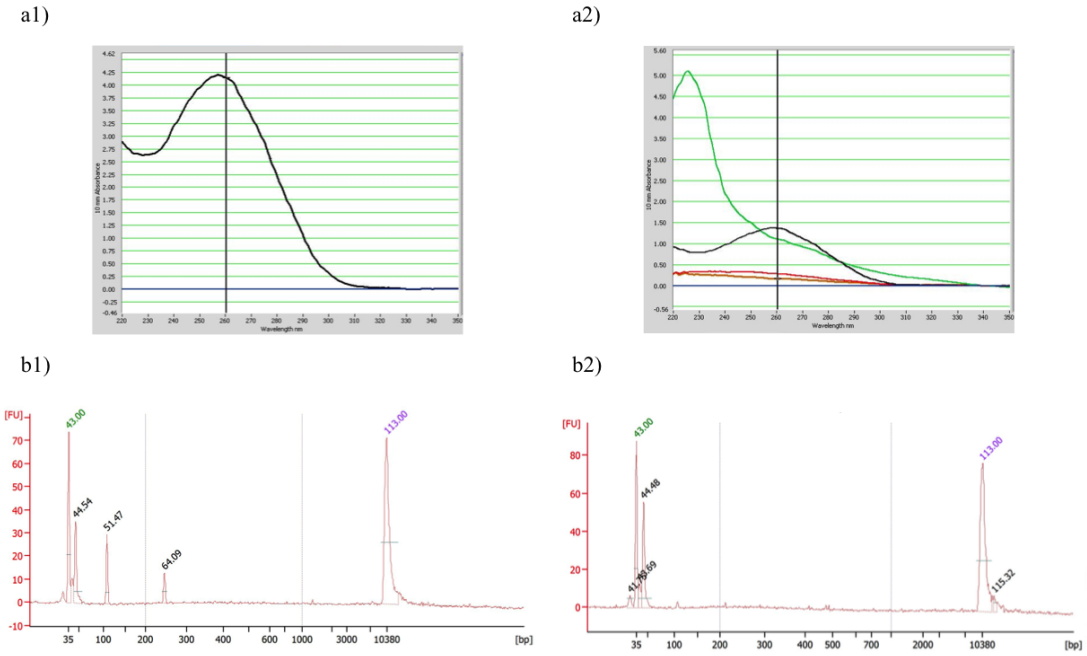


Figure 48 Quality of samples: nanodrop and fragmentation

The COPE cohort yielded varying qualities of DNA from these samples. a1) Nanodrop curve of a good quality sample. a2) Nanodrop curves of poor quality samples that failed quality control. b1) Bioanalyzer electropherogram of a good quality sample (peak 51.47: 107 bp fragments; 64.09: 245 bp fragments. B2) Bioanalyzer electropherogram of a poor heavily fragmented sample (no fragment peaks). Peaks 43.00 (green) and 113.0 (purple) are markers used to identify the size of sample peaks.

Following initial DNA extractions and quality assessments, careful case selection was based on morphology (excluding poorly fixed cases, cases with very limited tumour cells, little material in block and cases with large numbers of lymphocytic tumour infiltration). In addition, we prioritised cases more recently diagnosed. Despite this careful selection criteria, numbers that passed quality control were low. After liaising with the Agilent application specialists, a small fraction of samples that did not meet the selection criteria but yielded a high amount of DNA were tested. This pass criteria included a DNA purity 260/280 ratio between 1.8-2 and following PCR, amplified fragments at 107 bps and 245 bps to score at least a C (see methods chapter 2.3.6).

9 patient samples fell into this criteria with 8/9 successfully producing NGS data. The 8 samples that underwent successful target enrichment fell below the required threshold for fragmentation with amplified fragments at only 107 bps. Additional DNA was used in the target enrichment of

these samples to try and counteract the limited sizes of the DNA fragments. The failed sample (30091506) revealed excellent DNA yields but a category C for fragmentation, and a poor 260/280 ratio score of 1.57. This contamination had significant implications with this sample failing target enrichment despite good DNA yields and larger DNA fragmentation sizes. The consequences of using poor samples can be seen in fig. 49.

a)

Patient	Tissue	Nanodrop 260/280	Input DNA (Qubit dsDNA µg)	Fragmentation category
30091104	DCIS, 97	1.91	3	<C
30091506	Tumour	1.57	2.5	C

b) c)

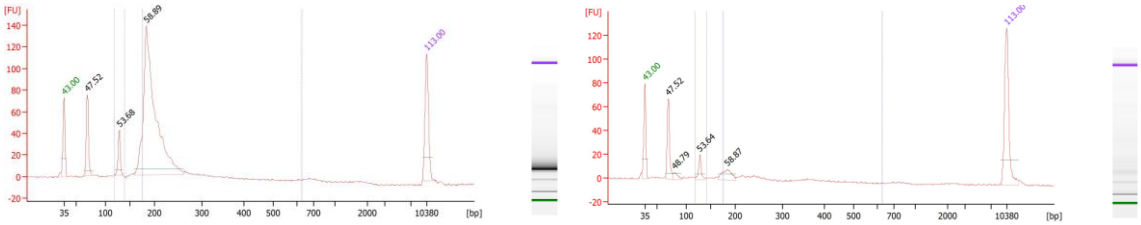


Figure 49 Importance of DNA purity and fragmentation

Samples 30091104 (DCIS 1997 episode) and 30091506 (tumour) did not meet the quality control criteria but were still selected for target enrichment. Samples need a 260/280 ratio between 1.8-2.0 and fall into at least a category C fragmentation score. a) Table to show the quality control data from samples 30091104 (DCIS, 97 episode) and 30091506 (tumour). b) Bioanalyzer electropheragram of sample 30091104 (DCIS, 97 episode) following successful target enrichment. c) Bioanalyzer electropheragram of sample 30091506 (tumour) following failed target enrichment.

Despite successful target enrichment of some fragmented samples that did not meet the selection criteria, these samples unsurprisingly revealed a poorer coverage and depth of genes. Differences in the percentage of bases covered at different depths can be seen alongside the category of fragmentation shown in table 18. The addition of more DNA input can counteract some of these issues but ultimately, one needs to accept the limitations of this archival heavily fragmented material.

Sample	Category	Kit Coverage 20x (%)	Kit Coverage 100x (%)	Coverage *TP53* (%) 100x	Coverage *ERBB2* (%) 100x	Coverage *MUC17* (%) 100x
30091123 (tumour)	A	68.12	53.74	93.1%	93.0%	95.4
30101705 (DCIS)	B	65.01	50.48	92.7%	97.4%	94.9%
30091111 (tumour)	C	38.95	27.61	77.3%	78.0%	40.2%
200510979 (tumour)	<C	50.16	38.19	82.8%	91.7%	58.6%

Table 18 DNA fragmentation and the affect on coverage

Table to show differing fragmentation categories and the affect on coverage at different depths.

Each case from the COPE cohort was evaluated for the predicted DNA quality, i.e. if the sample was poorly fixed or only a small area of tissue was present following morphological analysis. If the sample was selected for DNA extraction the DNA was tested for purity (260/280 ratio), concentration and amplifiable material. The COPE invasive tumour cases are summarised in table 19 and the DCIS cases summarised in table 20.

COPE number	NGS data	Failed QC step	Failed due to morphology	Nanodrop (260/280)	Qubit [dsDNA]	PCR
30091001	✓	-	-	1.96	43.2	A
30091003	×	Qubit	-	1.85	6.26	C
30091006	×	Morphology	Fixing	-	-	-
30091007	✓	-	-	1.91	143.8	<C
30091102	×	PCR	-	1.84	158.4	F
30091104, 2001	✓	-	-	1.83	416.0	B
30091107	×	Qubit	-	1.87	6.9	<C
30091108	×	Morphology	Fixing	-	-	-
30091111	✓	-	-	1.87	32.2	C
30091113	×	Morphology	Material	-	-	-
30091122	×	PCR	-	1.87	14.84	<C
30091123	✓	-	-	1.83	64.0	A
30091124	×	Morphology	Cellular content	-	-	-
30091125	✓	-	-	1.95	46.8	A
30091128	×	PCR	-	1.89	16.1	F
30091139	×	Morphology	Fixing	-	-	-
30091140	×	Qubit	-	1.86	15.5	<C
30091141	✓	-	-	1.93	228.0	<C
30091402	✓	-	-	1.94	256.0	B
30091504	✓	-	-	1.91	216.0	B
30091506	×	HaloPlex	-	1.57	41	C
30092701	×	Qubit	-	2.06	6.46	B
30092702	✓	-	-	2.00	17.8	B
30092703	✓	-	-	1.91	23.00	B
30095001	×	Morphology	Material	-	-	-

30101705	✓	-	-	2.01	21.2	B
30101711	×	Qubit	-	1.97	3.34	-
30105101	×	Morphology	Material	-	-	-
30105104	✓	-	-	2.00	39.6	A
30112802	✓	-	-	1.91	105.33	C
30114001	×	Morphology	Material	-	-	-
30114003	✓	-	-	1.98	48.6	B
30114004	×	Morphology	Fixing	-	-	-
30124101	✓	-	-	1.9	212	B
30132101	×	Morphology	Fixing	-	-	-
2007022398	×	Morphology	Fixing	-	-	-

Table 19 Summary of COPE invasive tumour cases selected for NGS

Table to show a summary of COPE invasive tumour selection for NGS. Morphology failures include the size of the diseased area and the material remaining in the block (material), poor fixing and underscoring (fixing) and a high frequency of tumour invading microenvironment cells which would reduce the tumour DNA purity (cellular content). See chapter 2.3.6 for further information regarding passing PCR scores A to <C. A failed score (F) indicates no amplification of 105 bp fragments.

COPE number	NGS data	Failed QC step	Morphology	Nanodrop (260/280)	Qubit [dsDNA]	PCR
30091001	×	Qubit	-	1.92	65	<C
30091003	×	Morphology	Tumour	-	-	-
30091006	×	Morphology	Fixing	-	-	-
30091007	×	Morphology	Material	-	-	-
30091102	×	Morphology	Tumour	-	-	-
30091104, 1996	×	Qubit	-	1.95	14.26	<C
30091104, 1997	✓	-	-	1.91	55.8	<C
30091104, 2001	×	Morphology	Material	-	-	-
30091107	×	Qubit	-	1.85	5.04	F
30091108	×	Morphology	Fixing	-	-	-
30091111	×	Nanodrop	-	1.59	-	-
30091113	×	Morphology	Material	-	-	-
30091122	×	Morphology	Tumour	-	-	-
30091123	×	Qubit	-	1.72	25.8	B
30091124	×	Morphology	Tumour	-	-	-
30091125	×	Qubit	-	1.81	8.24	A
30091127	×	Morphology	Material	-	-	-
30091139	×	Morphology	Fixing	-	-	-
30091141	×	Qubit	-	1.91	44.4	<C
30091401	✓	-	-	2.00	15.68	B
30091402	×	Morphology	Material	-	-	-
30091504	×	Morphology	Material	-	-	-
30091505	×	Morphology	Material	-	-	-
30091506	×	Morphology	Material	-	-	-
30092701	×	Qubit	-	1.92	18.1	B

30092703	×	Morphology	Material	-	-	-
30092901	×	Morphology	Material	-	-	-
30095001	×	Morphology	Material	-	-	-
30101702	×	Morphology	Material	-	-	-
30101705	✓	-	-	1.98	43.6	B
30101707	×	Morphology	Material	-	-	-
30101710	×	Morphology	Material	-	-	-
30101711	×	Qubit	-	1.96	3.78	-
30105101	×	Morphology	Material	-	-	-
30105104	×	Nanodrop	-	2.61	-	-
30112802	×	Morphology	Material	-	-	-
30114003	×	Morphology	Material	-	-	-
30114004	×	Morphology	Fixing	-	-	-
30124101	✓	-	-	1.9	134	B
30132101	×	Morphology	Fixing	-	-	-
2007022398	×	Morphology	Fixing	-	-	-

Table 20 Summary of COPE ductal carcinoma *in situ* (DCIS) cases selected for NGS

Table to show a summary of COPE DCIS selection for NGS. Morphology failures include the size of the diseased area and the material remaining in the block (material), poor fixing and underscoring (fixing) and if the matched tumour samples had previously failed the selection criteria (tumour). See chapter 2.3.6 for further information regarding passing PCR scores A to <C. A failed score (F) indicates no amplification of 105 bp fragments.

A Flowchart of this process and case dropout can be seen in fig 50.

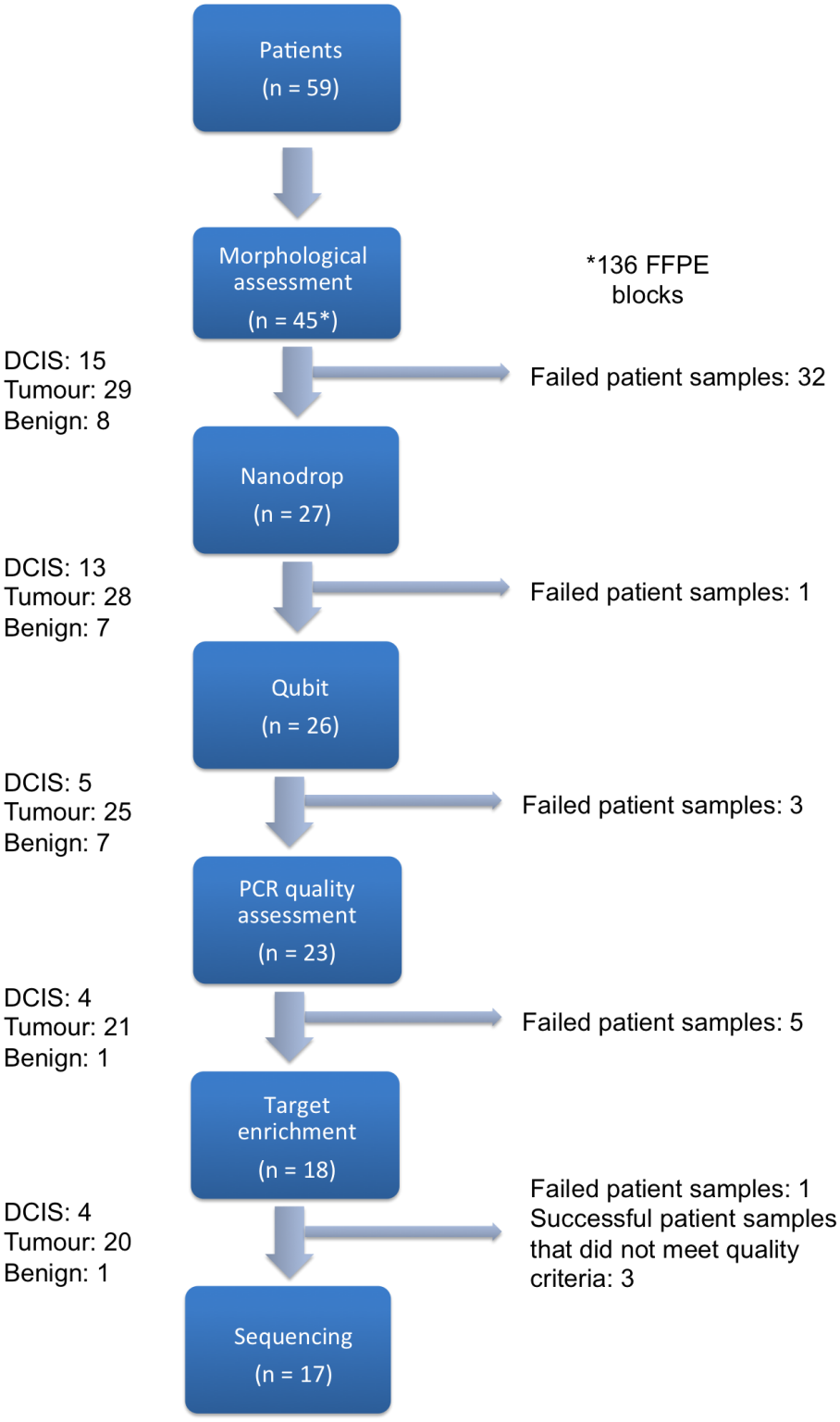


Figure 50 Flowchart of sample dropout for the COPE cohort

A flowchart to show the number of DCIS and invasive tumour samples that failed each quality control (QC) step. Tissue was only available for 45 patients.

Similar findings were found in the POSH HER2+ cases. A smaller proportion of cases failed based on morphological assessment that can in part be explained by the age of the samples (patient recruitment 2000-2008) as well as the stringent selection criteria for these cases (see methods chapter 2.1) [81]. Similar to the COPE cohort, each case was evaluated for the predicted DNA quality, i.e. if the sample was poorly fixed or only a small area of tissue was present following morphological analysis. Additionally POSH cases were selected based on the quality of their matched DCIS to test the hypothesis that an early loss of *TP53* was important in driving a HER2+ breast tumour phenotype. If the sample was selected for DNA extraction the DNA was tested for purity (260/280 ratio), concentration and amplifiable material. No POSH HER2+ DCIS cases underwent NGS but enough DNA was obtained in some cases to valid and test any variants called in *TP53* as part of the future work for this project. The POSH HER2+ invasive tumour cases are summarised in table 21 and the DCIS cases summarised in table 22.

POSH number	NGS data	Failed QC step	Morphology	Nanodrop (260/280)	Qubit [dsDNA]	PCR
2003090261	✓	-	-	2.00	35.2	A
2003120358	✓	-	-	1.95	33.4	<C
2004060642	×	Matched DCIS	-	1.93	208.00	<C
2004060672	×	Qubit	-	1.95	70.69	<C
2004070700	✓	-	-	1.86	782.42	<C
2005010979	✓	-	-	1.87	94.6	<C
2005041139	✓	-	-	1.97	646.77	A
2005041169	✓	-	-	2.00	101.00	<C
2005051253	×	Qubit	-	2.08	7.96	A
2005071332	×	Matched DCIS	-	1.95	262	<C
2005071383	×	Matched DCIS	-	1.96	296	<C
2005111597	×	Morphology	Fixed	-	-	-
2005121653	×	Matched DCIS	-	1.89	220	<C
2006011712	×	Matched DCIS	-	-	-	-
2006031839	×	Nanodrop	-	1.75	-	-
2006051963	×	Morphology	Material	-	-	-
2006102217	✓	-	-	1.92	348	A
2006102270	×	Morphology	Fixing	-	-	-
2007022386	×	Morphology	Material	-	-	-
2007052497	✓	-	-	1.92	31.4	B
2007092704	×	Qubit	-	1.92	41.8	-

2007102743	×	Matched DCIS	-	-	-	-
2008073046	✓	-	-	1.94	73.2	B

Table 21 Summary of POSH HER2+ invasive tumour cases selected for NGS

Table to show a summary of POSH HER2+ invasive tumour selection for NGS. Morphology failures include the size of the diseased area and the material remaining in the block (material), poor fixing and underscoring (fixing) and if the matched DCIS samples had previously failed the selection criteria. See chapter 2.3.6 for further information regarding passing PCR scores A to <C. A failed score (F) indicates no amplification of 105 bp fragments.

POSH number	NGS data	Failed QC step	Morphology	Nanodrop (260/280)	Qubit [dsDNA]	PCR
2003090261	×	Nanodrop	-	1.59	-	-
2003120358	×	Morphology	Material	-	-	-
2004060642	×	Nanodrop	-	2.45	-	-
2004060672	×	Morphology	Material	-	-	-
2004070700	×	Morphology	Material	-	-	-
2005010979	×	Qubit	-	1.89	48.4	<C
2005041139	×	Qubit	-	2.00	25.8	<C
2005041169	×	Qubit	-	1.97	28.6	<C
2005051253	×	Qubit	-	2.07	10.3	-
2005071332	×	Morphology	Material	-	-	-
2005071383	×	PCR	-	1.98	12.1	F
2005111597	×	Morphology	Fixing	-	-	-
2005121653	×	Qubit	-	1.97	4.16	-
2006031839	×	Nanodrop	-	1.72	-	-
2006051963	×	Morphology	Material	-	-	-
2006102217	×	Qubit	-	1.97	11	B
2006102270	×	Morphology	Fixing	-	-	-
2007022386	×	Morphology	Material	-	-	-
2007052497	×	Qubit	-	1.91	4.36	-
2007092704	×	Qubit	-	2.1	9.98	-
2008073046	×	Qubit	-	1.54	6.94	B

Table 22 Summary of POSH HER2+ ductal carcinoma *in situ* (DCIS) cases selected for NGS

Table to show a summary of POSH HER2+ DCIS selection for NGS. Morphology failures include the size of the diseased area and the material remaining in the block (material) and poor fixing and underscoring (fixing). DNA obtained will be used for validation. See chapter 2.3.6 for further information regarding passing PCR scores A to <C. A failed score (F) indicates no amplification of 105 bp fragments.

A Flowchart of this process and case dropout can be visualised in fig 51.

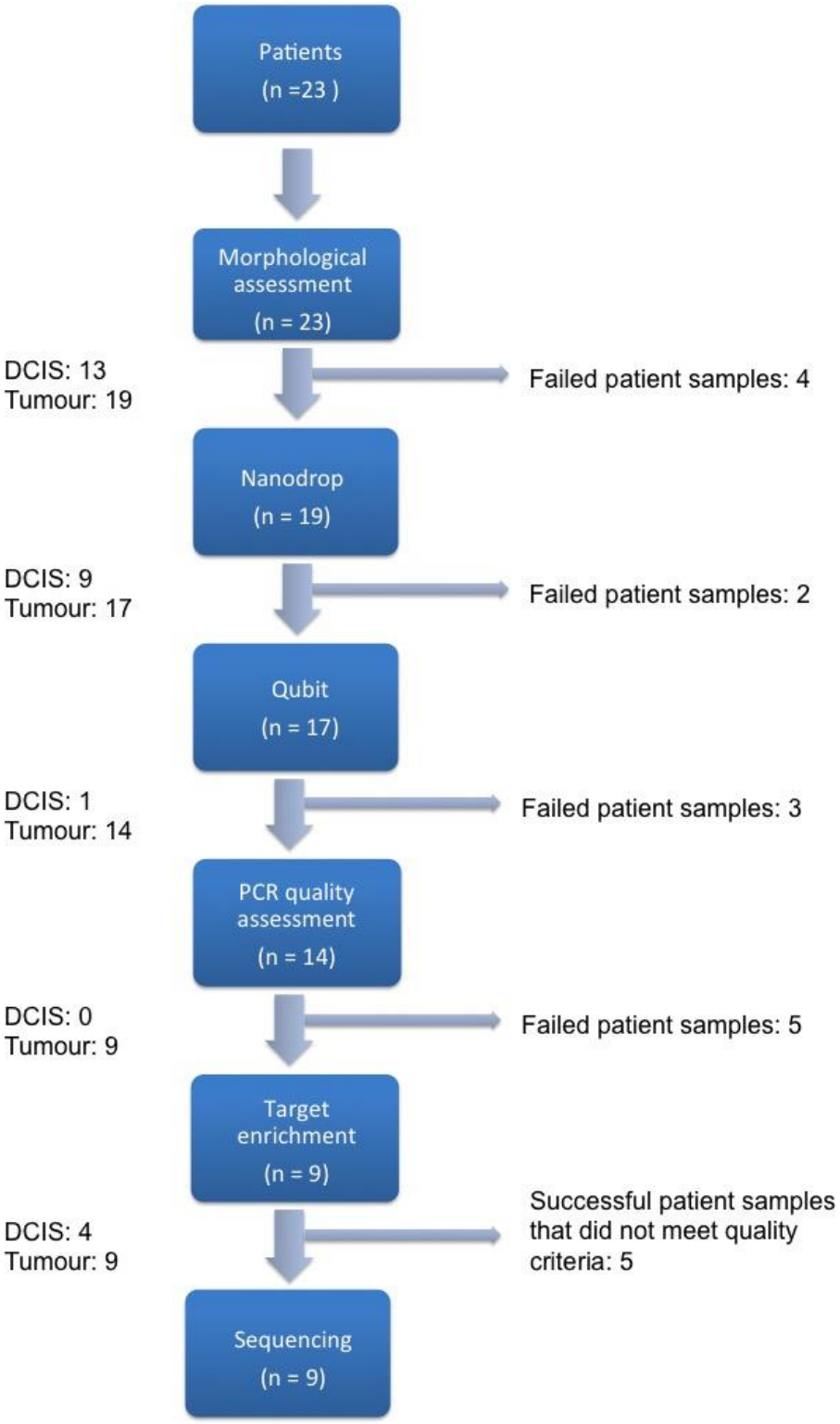


Figure 51 Flowchart of sample dropout for the HER2+ POSH control cohort

A flowchart to show the number of DCIS and invasive tumour samples that failed each quality control (QC) step.

4.2 Germline *TP53* mutations in the COPE cohort

The majority of *TP53* mutations in the COPE patients were in the DNA binding domain and were missense mutations. A summary of germline mutations in COPE patients is shown in figure 52.

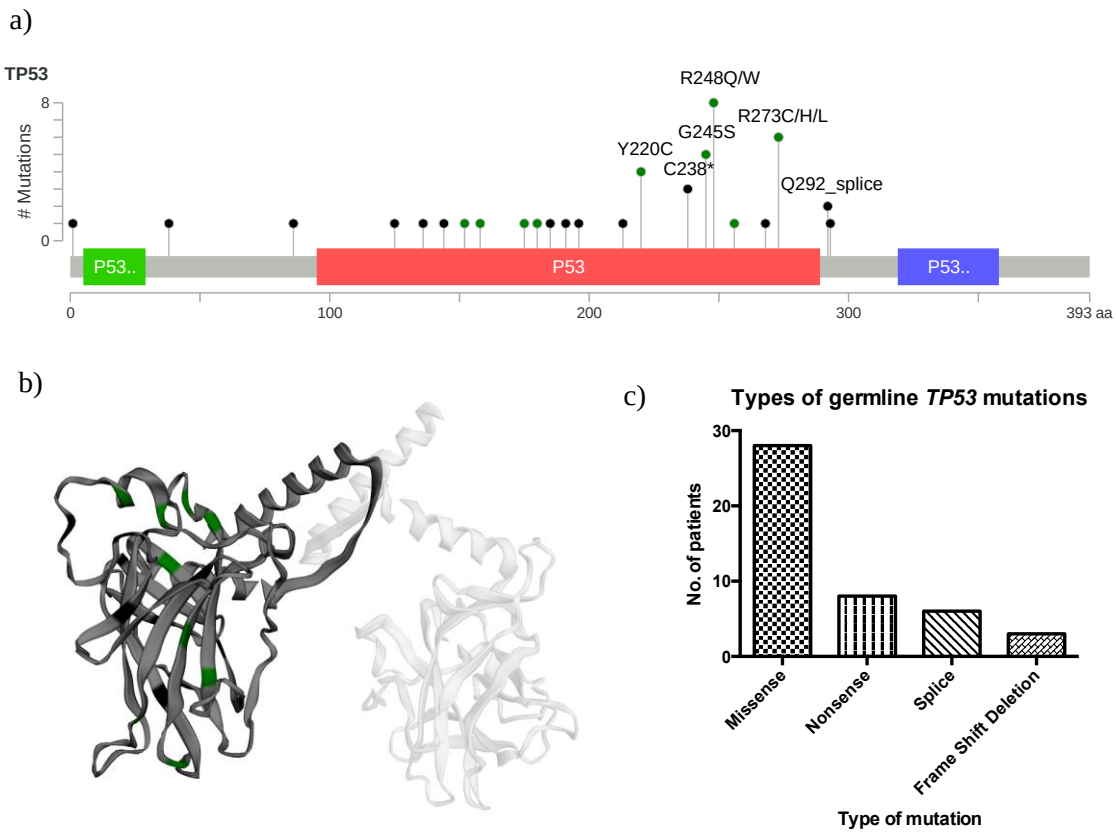


Figure 52 Spectrum of germline *TP53* mutations

The germline *TP53* mutations in the 45 patients recruited to the COPE study. a) MutationMapper (cBioPortal for Cancer Genomics) of the germline *TP53* mutations in the cohort. The domains are the transactivation motif (codons 5-29, green), the DNA-binding domain (codons 95-289, red) and the tetramerisation domain (codons 319-358, blue). Missense mutations are presented in green and truncating mutations (nonsense, frameshift indel and splice) are presented in black. A position containing more than one mutation is marked. b) PyMol three-dimensional representation of *TP53* showing the locations of the missense (green) and truncating (black) mutations and where they sit in the DNA-binding cleft. c) Bar chart to show the frequency of the different mutational effects in the cohort.

The International Agency for Research on Cancer (IARC) *TP53* mutation database was used to explore the frequency and distribution of *TP53* mutations in cancer reported globally. The data is shown in fig 53.

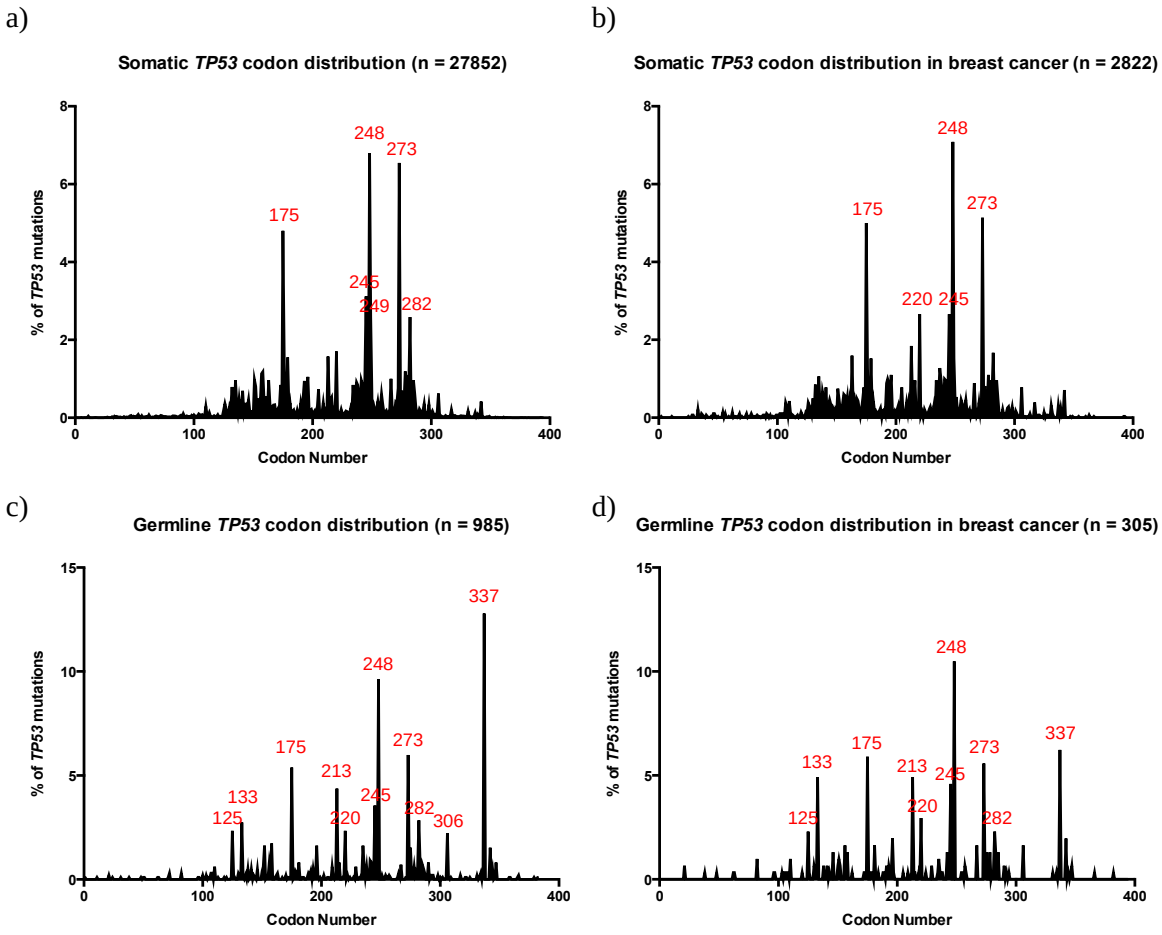


Figure 53 *TP53* variant codon distribution in selected cohorts

Codon distribution across the *TP53* gene in different cohorts. Hotspots are labelled and reported as a frequency of at least 2% of variants called in a cohort. a) Codon distribution of somatic *TP53* mutations. b) Codon distribution of somatic *TP53* mutations detected in breast cancer. c) Codon distribution of germline *TP53* variants. d) Codon distribution of germline *TP53* variants in those patients presenting with breast cancer.

Somatic mutations across all cancers (a) and somatic mutations in breast cancer (b) are similar and cluster around the DNA binding domain (codons 95-289) with some hotspots including codons 175, 245, 248 and 273.

A similar picture was seen in germline mutations in LFS cases (c) and in LFS patients with breast cancer (d). The outstanding difference between germline mutations and the COPE samples was the hotspot at codon 337. This mutation does not follow the usual trend associated with *TP53* germline mutations and instead of sitting within the DNA binding domain, it is located in the tetramerization domain. This is a unique mutation predominantly reported in the South Eastern part of Brazil, one family in Portugal with Brazilian ancestry and a single patient was detected in France with Portuguese ancestry [31, 57, 58]. The drop in incidence of this codon when germline aberrations

were investigated solely in breast cancer is because this R337H missense mutation is often associated with childhood adrenal cortical carcinoma [56, 58, 193]. For more information see chapter 1.2.5. The types of mutational effects can be seen in fig. 54.

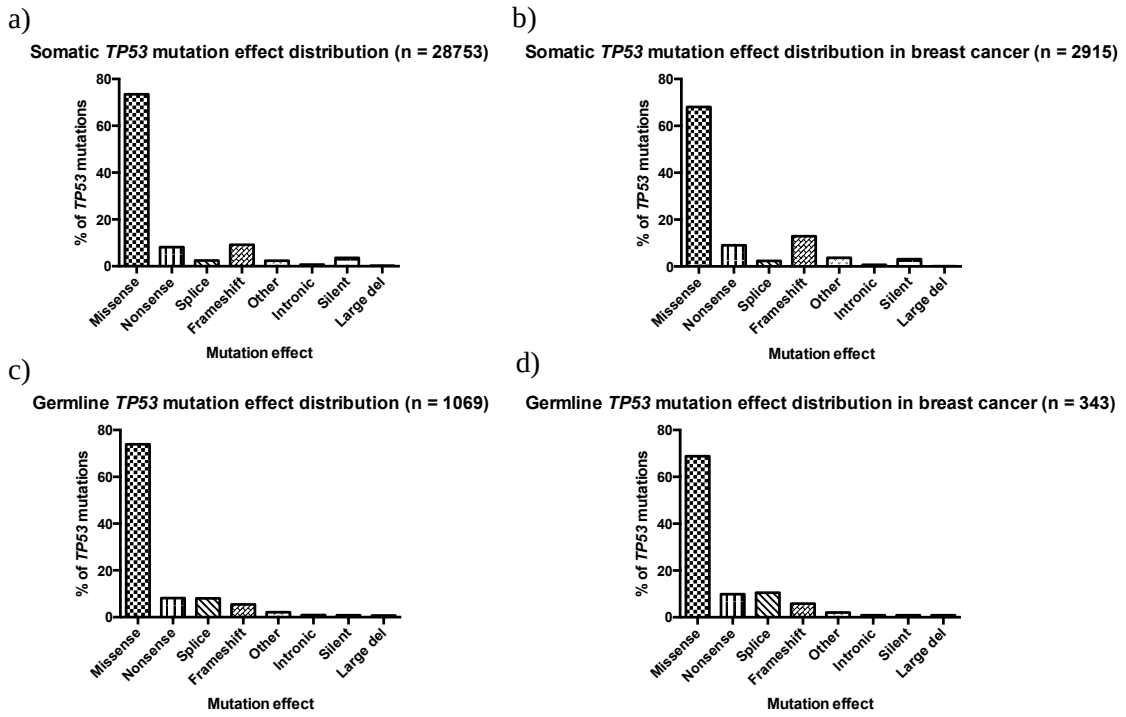


Figure 54 *TP53* variant mutation effect distribution in selected cohorts

Mutational effect distribution across the *TP53* gene in different cohorts. Across all cohorts missense mutations dominate the genomic landscape. a) Mutation effect in somatic *TP53* mutations. b) Mutation effect in somatic *TP53* mutations detected in breast cancer. c) Mutation effect in germline *TP53* variants. d) Mutation effect in germline *TP53* variants in those patients presenting with breast cancer.

Missense mutations dominate the mutational effect landscape in all cohorts investigated including the COPE cohort. Missense mutations were reported to a slightly less extent in the COPE cohort (62%) and there was a higher incidence of nonsense (18%) and splice (13%) variants (see fig. 52). Frameshift mutations were reported to a similar extent in the COPE cohort (7%). These differences may by due to the size of the COPE cohort and the high probability of some of these patients being related.

4.3 Somatic mutations in *TP53* carriers and an age matched HER2+ control cohort

Raw sequence data was processed to a variant cell file by the bioinformatics team (Dr. R. Pengelly). Further selection of variants of interest was based on likely functionality (for example by removing synonymous mutations) and mutations reported in dbSNP were excluded to try and

achieve a list that was unlikely to be germline. Additionally, variants were selected based on their biological relevance reported in the literature and in >5% of reads at a depth of 100 reads.

This still left too many variants for validation and manual inspection in the Integrative Genomics Viewer (IGV) so to focus further, genes of particular biological interest were selected including all *TP53* mutations (see appendix for full list of filtered variants in COPE DCIS samples (chapter 7.4), COPE invasive tumour (chapter 7.5) and POSH HER2+ invasive tumour samples (chapter 7.6)). Genes of particular biological interest included known cancer genes in pathways associated with for example HER2+ breast cancer or TGFβ signalling and sclerotic stroma development identified from the morphology review [194]. The gene lists presented in this chapter have been prioritised.

4.3.1 Somatic mutations in breast tumour tissue from *TP53* carriers

20 DNA samples derived from breast tumour FFPE samples obtained from patients with a germline *TP53* mutation underwent target enrichment. These samples were from 16 patients with 2 of these patients having 3 primary tumour samples from the same time point sequenced in order to address the heterogeneity in this cohort. These 2 cases will be discussed later and have been excluded from the analysis in table 23.

Patient	Germline *TP53* mutation	Number of filtered somatic variants	Effect
30091001	c.672+1G>T	34	Truncating
30091007	c.659A>G (p.Y220C)	92	Missense
30091111	c.818G>A (p.R273H)	657	Missense
30091123	c.743G>A (p.R248Q)	25	Missense
30091125	c.743G>A (p.R248Q)	50	Missense
30091141	c.743G>A (p.R248Q)	190	Missense
30091402	c.714T>A (p.C238*)	37	Truncating
30091504	c.766A>G (p.T256A)	70	Missense
30092702	c.743G>A (p.R248Q)	67	Missense
30092703	c.714T>A (p.C238*)	191	Truncating
30101705	c.733G>A (p.G245S)	194	Missense
30105104	c.733G>A (p.G245S)	39	Missense
30114003	c.659A>G (p.Y220C)	104	Missense
30124101	c.993+2T>G	201	Truncating

Table 23 Tumour filtered somatic variants in *TP53* carriers

Table to show the number of somatic variants called in the tumour samples of 14 patients with a germline *TP53* mutation.

The raw data was filtered using the filtering strategy outlined in the methods (chapter 2.3.10). Based on the number of filtered variants there does not seem to be any pattern regarding the

number of somatic variants called and the type of germline *TP53* mutation in invasive tumours. As previously stated, samples that scored very high for fragmentation typically contained a high number of somatic variants called. However these are likely to be false positives.

Genes that were mutated in more than 90% of cases were manually inspected in IGV to rule out the possibility of poor areas of sequencing either because of the sequence itself (for example if it was GC rich), as a result of poor quality DNA derived from FFPE or because of mismapping. The genes *P4HTM*, *ITPR1* and *ARID1A* were removed. An example of poor sequencing is shown in fig. 55. When this was visualised in IGV alongside sequence from a high quality DNA sample (DNA derived from blood), the poor quality of sequence in this region is clear.

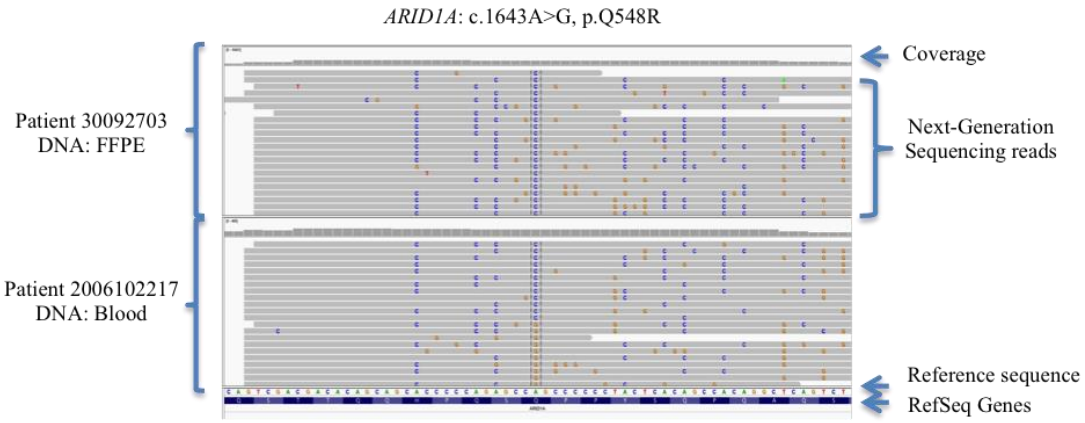


Figure 55 Region of poor sequencing in the *ARID1A* gene

A variant called in this area of the *ARID1A* gene is highly likely to be a false positive when the reads were manually investigated in IGV. Reads that are different to the reference sequence are shown as a coloured base. The grey areas of read match the reference genome.

A heat map showing the filtered clinically relevant genes is presented in fig. 56.

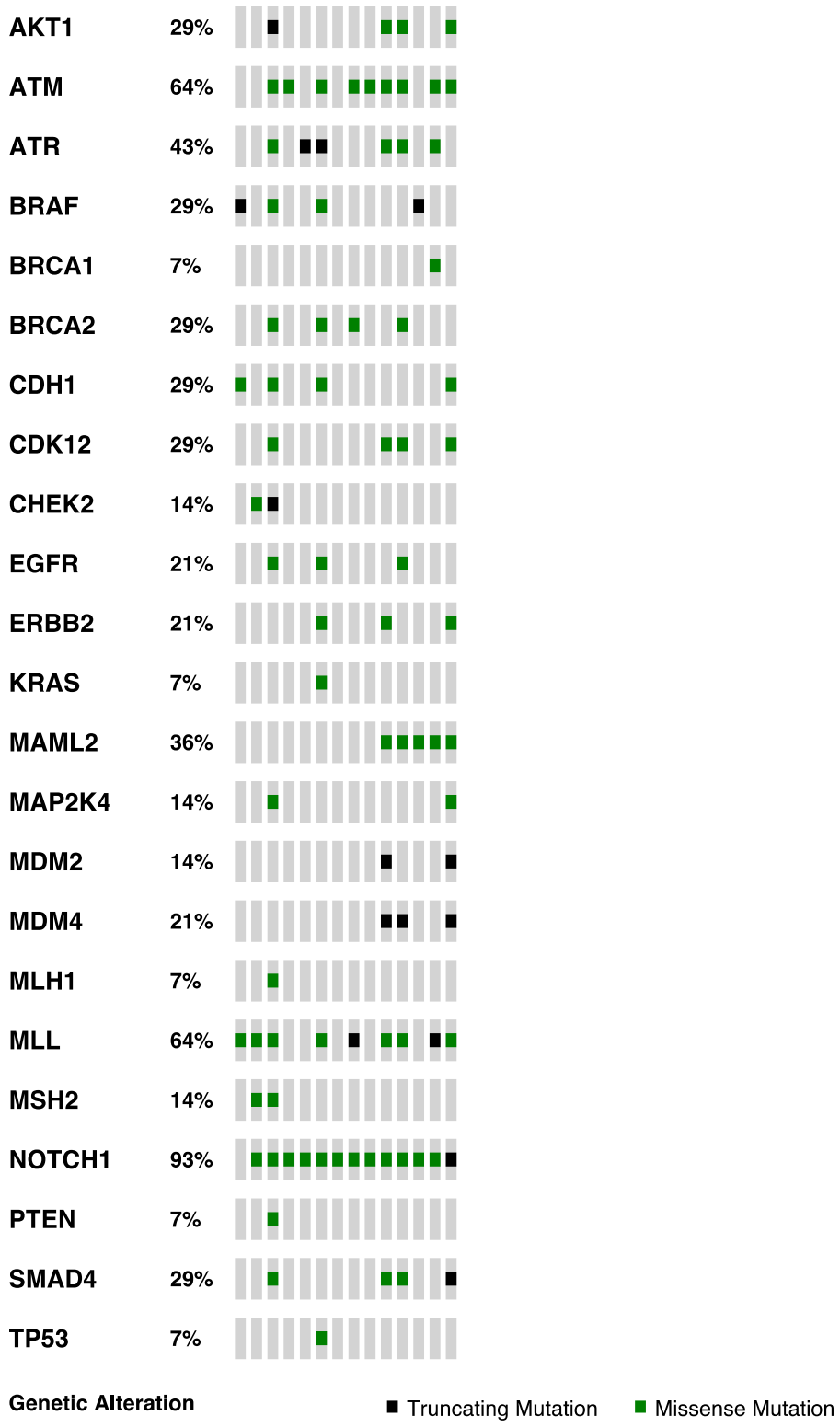


Figure 56 Heat map of filtered somatic tumour variants in *TP53* carriers.

Heat map of somatic variants present in the tumour within *TP53* carriers Genetic alteration: black, truncating mutation; grey, inframe mutation and green, missense mutation. Patient order: 30091001, 30091007, 30091111, 30091123, 30091125, 30091141, 30091402, 30091504, 30092702, 30092703, 30101705, 30105104, 30114003 and 30124101.

A number of the filtered genes that were reported in over 90% of cases were likely false positives including *GATA3*, *EP400* and *MKLN1*. Some of these variants reported did differ between cases but these were typically areas of poor sequence. Some of these variants however once visualised in IGV, did look like they could be true variants (see appendix for full list, chapter 7.5).

The original filtered list (see appendix, chapter 7.5) was processed through the DAVID pathway analysis database to analyse potential deregulated pathways [186]. From this analysis, 14 pathways were significantly affected as a result of the somatic mutations. Pathway analysis in the invasive tumour samples from the COPE cohort is presented in table 24.

Pathway	% of genes	Genes	Ease score
Focal adhesion	14.2	*TLN1, TLN2, ERBB2, PTEN, AKT1, LAMB4, SOS1, COL11A1, FN1, PRKCA, PIK3CG, EGFR, VAV3, BRAF, ROCK1, PIK3CB, FLT4, FLNC, FLNB, VAV1, FLNA, LAMA2, LAMA1,*	$2.39x10^{-12}$
ErbB signalling	8.0	*PRKCA, PIK3CG, EGFR, ERBB4, BRAF, PIK3CB, ERBB3, ERBB2, MAP2K4, AKT1, KRAS, SOS1, MTOR, NRG2*	$2.67x10^{-8}$
Notch signalling	6.3	*DVL2, NOTCH3, NOTCH1, EP300, DLL4, MAML2, CREBBP, NOTCH4, JAG2, JAG1, NCOR2*	$3.61x10^{-8}$
MAPK signalling	11.9	*PRKCA, EGFR, TRAF2, BRAF, NF1, MAP2K4, TP53, TGFB3, MECOM, FLNC, FLNB, FLNA, AKT1, ACVR1B, RPS6KA3, KRAS, RASGRF2, SOS1, MAP3K1, MAPK8IP3, MAP3K11*	$6.55x10^{-7}$
TGF-beta signalling	6.8	*INHBA, ACVR1B, EP300, RBL2, ROCK1, ZFYVE16, CREBBP, RBL1, SMAD4, BMPR2, TGFB3, SMURF2*	$1.89x10^{-6}$
Cell cycle	6.8	*EP300, RBL2, CREBBP, RBL1, SMAD4, TGFB3, TP53, MDM2, RB1, ATR, CHEK2, ATM*	$6.34x10^{-5}$
p53 signalling	5.1	*CASP8, TP53, MDM2, ATR, CHEK2, MDM4, PTEN, ATM, TP73*	$8.75x10^{-5}$
mTOR signalling	4.5	*PIK3CG, AKT1, RPS6KA3, TSC1, BRAF, PIK3CB, STK11, MTOR*	$1.04x10^{-4}$
Apoptosis	5.1	*IRAK4, PIK3CG, AKT1, TRAF2, CASP10, PIK3CB, CASP8, TP53, ATM*	$4.94x10^{-4}$
Regulation of actin	7.4	*PIK3CG, EGFR, VAV3, BRAF, ROCK1, PIK3CB, VAV1, KRAS, TIAM1, SOS1, MYLK, APC, FN1*	$2.00x10^{-3}$
Adherens junction	4.0	*EGFR, ACVR1B, EP300, ERBB2, CREBBP, SMAD4, CDH1*	$5.83x10^{-3}$
ECM-receptor interaction	3.4	*LAMA2, LAMA1, LAMB4, LAMA4, COL11A1, FN1*	$3.34x10^{-2}$
Wnt signalling	4.5	*PRKCA, DVL2, EP300, ROCK1, CREBBP, SMAD4, TP53, APC*	$4.25x10^{-2}$
Endocytosis	5.1	*EGFR, ACVR1B, RET, ERBB4, ERBB3, MDM2, SMURF2, ITCH, KIT*	$4.29x10^{-2}$

Table 24 Pathways affected in the tumour of patients with a germline *TP53* mutation

Table to show the deregulated pathways in the invasive tumour cases from *TP53* carriers as a consequence of somatic mutations acquired.

4.3.2 Somatic mutations in DCIS tissue from *TP53* carriers

4 COPE cases had DNA from the precursor lesion DCIS successfully sequenced. Patient 30091401 presented with only DCIS, patient 30091104 initially presented with DCIS and later developed invasive ductal carcinoma and the remaining 2 patients presented with both DCIS and invasive tumour. The cases are summarised in the table 25.

Patient	Germline *TP53* mutation	Number of filtered	Effect	Clinical presentation
30091104, 1997	c.818G>A (p.R273C)	109	Missense	1996 pure DCIS 1997 pure DCIS 2001 invasive ductal carcinoma with associated DCIS
30091401	c.714T>A (p.C238*)	55	Truncating	2006 pure DCIS
30101705	c.733G>A (p.G245S)	155	Missense	2009 invasive ductal carcinoma with associated DCIS
30124101	c.993+2T>G	157	Truncating	2009 invasive ductal carcinoma with associated DCIS

Table 25 DCIS filtered somatic variants in *TP53* carriers

The number of somatic variants called in the DCIS of 4 patients with a germline *TP53* mutation.

Firstly, based on the number of filtered somatic variants called, the two cases with invasive tumour at initial presentation, revealed the highest number of variants in DCIS. The case that did not progress any further had the lowest count of variants (30091401) and case 30091104 which later progressed into invasive ductal carcinoma, had 109 variants falling into the middle of this spectrum. This may be a reflection of a greater level of genomic instability in the immediate preinvasive DCIS. Genes with the same genetic alteration in all cases were manually inspected in IGV to check for areas of poor sequencing and suspected false positives. The genes in which these variants are located, are shown in the heat map below (fig. 57).

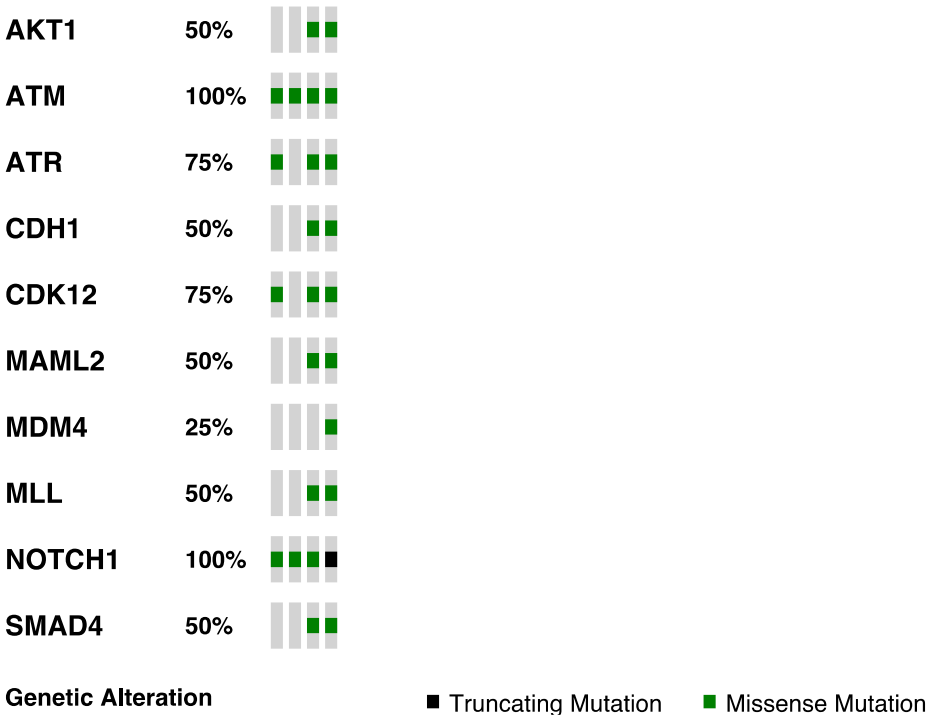


Figure 57 Heat map of filtered variants in DCIS lesions

Heat map of somatic variants present in DCIS within *TP53* carriers. Genetic alteration: black, truncating mutation; grey, inframe mutation and green, missense mutation. Patient order: 30091104; 1997, 30091401, 30101705 and 3012410.

The original filtered list (see appendix, chapter 7.4) was processed through the DAVID pathway analysis database to analyse potential deregulated pathways [186, 187]. Table 26 shows the pathways that are deregulated in DCIS derived from *TP53* carriers.

Pathway	% of genes affected	Genes	EASE score
Notch signalling	12.9	*NOTCH3, NOTCH1, EP300, DLL4, MAML2, CREBBP, NOTCH4, JAG2, NCOR2,*	$2.17x10^{-9}$
Adherens junction	7.1	*ACVR1B, EP300, CREBBP, SMAD4, CDH1*	$3.0x10^{-3}$
MAPK signalling	11.4	*AKT1, ACVR1B, NF1, MAPK8IP3, MECOM, FLNC, FLNA, MAP3K11*	$4.9x10^{-3}$
TGF-beta signalling	7.1	*ACVR1B, EP300, CREBBP, SMAD4, SMURF2*	$5.0x10^{-3}$
Focal adhesion	10	*AKT1, LAMB4, TLN1, TLN2, FLNC, FLNA, FN1*	$5.0x10^{-3}$
p53 signalling	5.7	*ATR, MDM4, ATM, TP73*	$1.7x10^{-2}$
Cell cycle	7.1	*EP300, CREBBP, SMAD4, ATR, ATM*	$1.7x10^{-2}$

Table 26 Pathways affected in the DCIS of patients with a germline *TP53* mutation

Table to show the deregulated pathways in the DCIS of *TP53* carriers as a consequence of the somatic mutations acquired.

When the affected pathways were compared between the DCIS and tumour, a higher frequency of significantly affected pathways were described in the tumour samples. Although a greater number of tumour samples were analysed, the expected greater genomic instability in the tumour samples would predictably lead to more genomic lesions accrued and the deregulation of additional pathways. The overall number of pathways significantly affected in the DCIS was 7 compared to the 14 pathways deregulated in the tumour samples. From the top 5 significantly affected pathways in the DCIS and tumour samples, one pathway stands out and that is ErbB2 signalling. This was the second most significantly affected pathway in the tumour samples ($2.67x10^{-8}$) and was also deregulated in the DCIS tissue. Additional pathways significantly affected in the tumour samples were mTOR, and Wnt signalling, ECM-receptor interactions, regulation of actin, apoptosis and endocytosis (see table 24 for further details).

4.3.3 Somatic mutations present in tumour tissue from a control HER2+ breast cancer subgroup from the POSH cohort.

The advantage of exploring hypotheses in this cohort of breast tumours is that germline information is available through gDNA derived from blood. Therefore, for the nine patients that passed quality control, a sample of their gDNA derived from blood samples also underwent target enrichment. This allowed determination of germline variants as well as poor areas of sequencing independent of the fixing process in the blocks. A good example of a gene with areas of poor sequencing is *NOTCH1*, a gene notorious for difficulties in sequencing regardless of the technology used. *NOTCH1* is regarded as a gene difficult to sequence due to its high GC content (>65%) [195]. This GC content bias has previously been reported which describes the dependence of coverage, based on the regional GC content of the gene [196]. Fig. 58 shows an area of poor sequencing in this gene when comparing a matched blood and FFPE tumour sample. This variant sits in a GC repeat region so it is unsurprising this is an area where accurate sequencing was difficult.

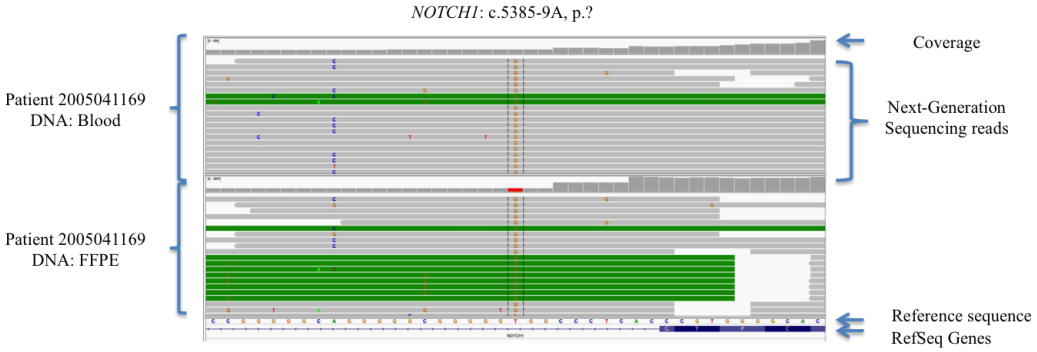


Figure 58 Area of poor sequencing in the *NOTCH1* gene

A splicing variant was called in the *NOTCH1* gene that was visualised in the matched blood sample. Grey reads represent bases that match the reference genome. Coloured bases represented bases that differ from the reference genome.

Following on from the analysis of somatic mutations in those patients with a germline *TP53* mutation, somatic mutations in the control HER2+ cases from the POSH cohort were next investigated. Table 27 shows the cases selected and those patients with a somatic *TP53* mutation.

Patient	Number of filtered somatic variants	Somatic *TP53* mutation	Effect
2005010979	91	-	-
2003090261	33	c.A733C, (p.T245P) c.G614T, (p.R205L)	Missense Missense
2003120358	16	-	-
2004070700	113	c.C437G, (p.P146R)	Missense
2005041139	33	c.A733C, (p.T245P)	Missense
2005041169	248	-	-
2006102217	168	c.T122G, (p.V41G)	Missense
2007052497	96	c.G273A, (p.W91*) c.G217C, (p.V73L)	Truncating Missense
2008073046	120	-	-

Table 27 Filtered somatic variants in HER2+ invasive tumour control cases from POSH

Table to show the number of somatic variants and *TP53* status in HER2+ cases from the POSH cohort.

The individual *TP53* mutations were investigated next to see if the variant allele frequency (VAF) would suggest an early molecular event. This data is shown in table 28. Similar to *TP53* carriers, the VAF scores for the mutation were high suggesting that the mutation is present in a high proportion of the tumour and that based on Darwinian evolution, could be an early event driving tumourogenesis in HER2+ breast tumours.

Patient	Germline *TP53* mutation	Number of reads	% of reads	% tumour cellularity	% reads normalised to cellularity
2003090261	c.A733C, (p.T245P)	30/109	27.52	60	45.87
	c.G614T, (p.R205L	497/2785	17.85	60	29.75
2004070700	c.C437G, (p.P146R)	722/1334	54.12	70	77.31
2005041139	c.A733C, (p.T245P)	73/189	38.62	65-70	57.21
2006102217	c.T122G, (p.V41G)	1192/3086	38.63	70	55.19
2007052497	c.G273A, (p.W91*)	1563/3188	49.03	70	70.04
	c.G217C, (p.V73L)	11/211	5.21	70	7.44

Table 28 *TP53* somatic mutations HER2+ control cases from POSH

Table to show the VAF of *TP53* mutations. Each mutation was normalised to the tumour cellularity.

The filtered somatic variants in the HER2+ control tumours are shown in figure 59.

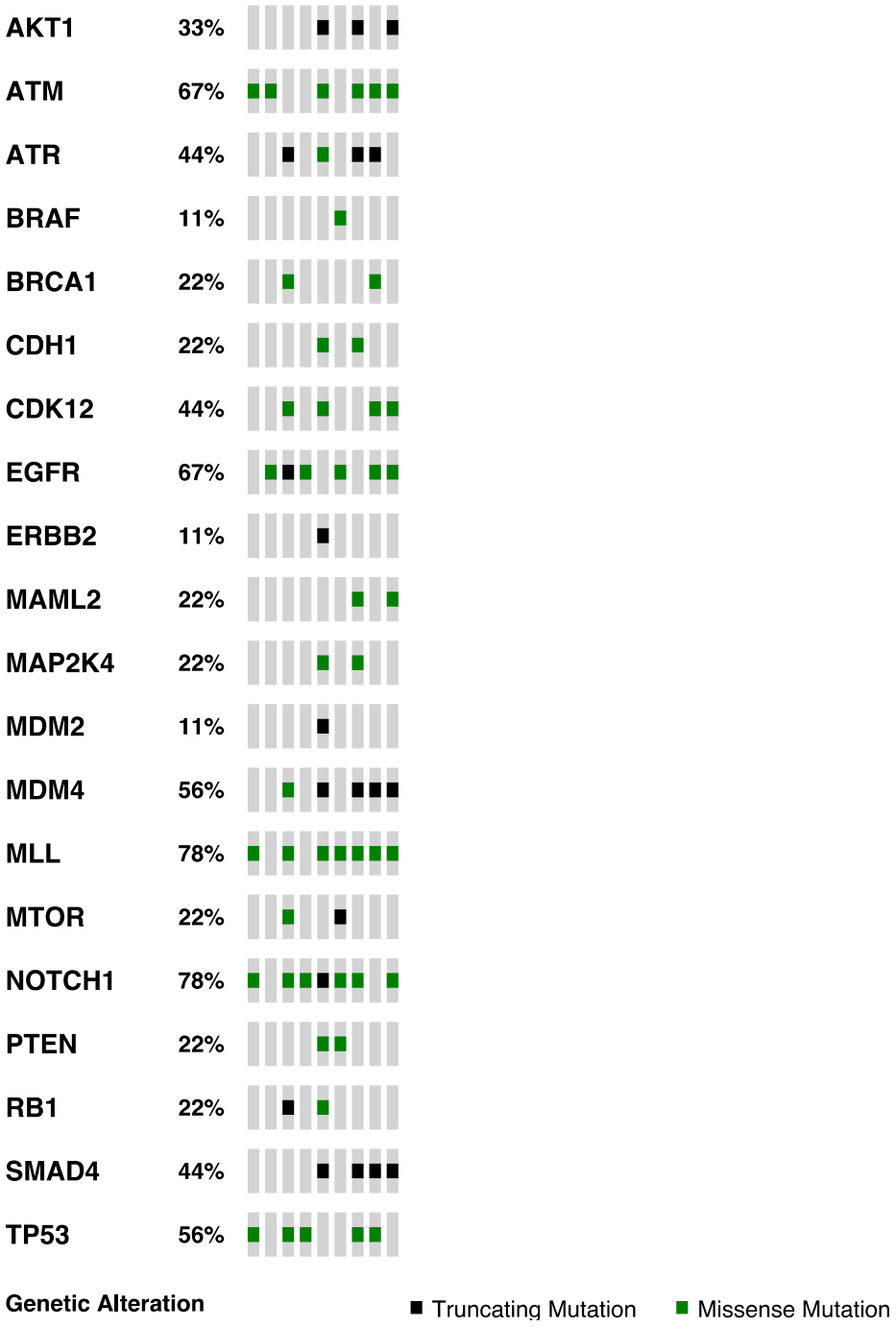


Figure 59 Filtered somatic variants in the control HER2+ POSH cohort

Heat map of somatic variants present in HER2+ cases from the POSH cohort. Genetic alteration: black, truncating mutation; grey, inframe mutation and green, missense mutation. Patient order: 2005010979, 2003090261, 2003120358, 2004070700, 2005041139, 2005041169, 2006102217, 2007052497 and 2008073046.

Pathway	% of genes	Genes	EASE score
Focal adhesion	15.3 (n=21)	*EGFR, PIK3CG, TLN1, ROCK1, BRAF, TLN2, GRB2, ERBB2, FLNC, PTEN, FLNB, FLNA, LAMA2, AKT1, LAMA1, LAMB4, LAMA4, SOS1, COL11A1*	$3.1x10^{-11}$
ErbB signalling	9.5 (n=13)	*PIK3CG, EGFR, BRAF, ERBB4, ERBB3, GRB2, ERBB2, MAP2K4, AKT1, SOS1, MTOR*	$1.0x10^{-8}$
Notch signalling	6.6 (n=9)	*NOTCH1, EP300, DLL4, MAML2, CREBBP, NOTCH4, JAG2*	$6.7x10^{-7}$
MAPK signalling	13.1 (n=18)	*EGFR, TRAF2, BRAF, GRB2, NF1, MAP2K4, TP53, TGFB3, FLNC, FLNB, FLNA, AKT1, ACVR1B, SOS1, MAP3K1*	$9.3x10^{-7}$
TGF-beta signalling	7.3 (n=10)	*EP300, RBL2, ROCK1, ZFYVE16, CREBBP, SMAD4, BMPR2, TGFB3*	$9.7x10^{-6}$
Cell cycle	7.3 (n=10)	*RBL2, CREBBP, SMAD4, TGFB3, TP53, MDM2, RB1, ATR*	$1.7x10^{-4}$
mTOR signalling	5.1 (n=7)	*AKT1, HIF1A, TSC1, BRAF, STK11*	$1.8x10^{-4}$
p53 signalling	5.1 (n=7)	*MDM2, ATR, MDM4, PTEN, ATM*	$7.7x10^{-4}$
Adherens junction	5.1 (n=7)	*ACVR1B, EP300, ERBB2, CREBBP, SMAD4*	$1.5x10^{-3}$
Apoptosis	4.4 (n=6)	*AKT1, TRAF2, CASP10, TP53*	$1.3x10^{-2}$
Wnt signalling	5.1 (n=7)	*EP300, ROCK1, CREBBP, SMAD4, TP53*	$3.6x10^{-2}$
ECM-receptor interaction	3.6 (n=5)	*LAMA1, LAMB4, LAMA4*	$4.8x10^{-2}$

Table 29 Pathway analysis of HER2+ control cases from POSH

Table to show the pathways affected due to somatic mutations in the HER2+ POSH cohort

The same pathways were potentially deregulated in both cohorts (tables 24 and 29). These pathways including MAPK signalling outlined in table 29, have been reported as having a high frequency of somatic variants often reported in HER2+ breast cancer [194]. Deregulation of these pathways is not exclusive to HER2+ breast cancer but is often implicated in disease of this breast tumour subtype [194]. Somatic variants in these genes can cause a loss or gain of function resulting in deregulation of these pathways implicated in driving oncogenesis. Differences were identified in the number and spectrum of filtered variants between the two cohorts. This observation can probably be attributed to the higher fragmentation of DNA analysed from the COPE cohort and the increased number of suspect false numbers. In order to reduce the risk of including many false positives, samples that had a coverage below 50% at a depth of 20x for the panel of genes, were excluded from the analysis [197]. Studies have shown that discrepancies between FFPE and fresh tissue from matched samples were in the majority of cases located in low coverage regions [198, 199].

Initially, the number of somatic mutations was investigated in the COPE cohort to see if as a result of a germline *TP53* mutation, there were fewer downstream somatic mutations. A second line of enquiry was to see if the type of *TP53* mutation was playing a role in the acquisition of somatic mutations. Differences were observed in the number of somatic mutations when those patients with a missense and truncating germline *TP53* mutation were compared. A Mann-Whitney statistical test revelled no statistical significance. This is primarily an observation and in order to decipher any relevant information, more patient samples will need to be investigated ideally from fresh tissue to rule out differences because of fixing. The data is revealed in table 30.

Missense mutation		Truncating mutation	
Germline mutation	**No. of somatic mutations**	**Germline mutation**	**No. of somatic mutations**
C.743G>A (p.R248Q)	25	c.672+1G>T	34
C.743G>A (p.R248Q)	50	C.714T>A (p.C238*)	37
c.766A>G (p.T256A)	70	C.714T>A (p.C238*)	191
C.743G>A (p.R248Q)	67	c.993+2T>G	201
c.659A>G (p.Y220C)	104		
c.733G>A (p.G245S)	39		
c.733G>A (p.G245S)	194		
Mean average somatic variants	**78**	**Mean average somatic variants**	**116**
Median	**67**	**Median**	**114**

Table 30 Number of somatic mutations in patients with a different germline *TP53* mutation

Table to show the number of somatic variants in invasive breast cancers in missense and truncating germline *TP53* carriers. No stastical signifacnce was found. Statistics used Mann-Whitney test.

Next, the number of somatic variants were investigated in the POSH HER2+ cases. This cohort was split between those cases with wild type (WT) *TP53* and those following filtering, with a somatic mutation in *TP53*. A correlation was once again observed between non-functional *TP53* and fewer additional somatic mutations however, a Mann-Whitney statistical test showed no statistical significance. This can be compared to the number of filtered mutations present in germline *TP53* cases shown in table 23.

TP53 WT		*TP53* mutant	
Somatic mutation	**No. of somatic mutations**	**Somatic *TP53* mutation**	**No. of somatic mutations**
WT	248	c.A733C, (p.T245P) c.G614T, (p.R205L)	33
WT	91	c.A733C, (p.T245P)	33
WT	120	c.T122G, (p.V41G)	168
		c.G273A, (p.W91*)	96
Mean average somatic variants	**78**	**Mean average somatic variants**	**83**
Median	**120**	**Median**	**64.50**

Table 31 Number of somatic mutations in HER2+ from the POSH cohort

Table to show the number of somatic variants and *TP53* status in HER2+ cases from the POSH cohort.

4.4 Loss of Heterozygosity (LOH) in germline *TP53* carriers

One of the leading hypotheses at the start of the project was that an early loss of *TP53* predisposes to a HER2+ breast tumour subtype. From the morphology review, this cohort of patients typically developed HER2+ breast cancer with widespread high grade DCIS. As a consequence of the high frequency of high grade precursor lesion DCIS, this cohort allowed a unique opportunity to investigate the order of molecular lesions and *TP53* status in cancer evolution.

Once the tumour samples had been sequenced and the data had been processed through the cancer VarScan 2 pipeline, the data was checked for the germline *TP53* mutations [182]. This pipeline allows identification of germline variants in individual samples, somatic mutations, copy number aberrations, LOH events, and *de nova* mutations and Mendelian inheritance in family trios [182, 200]. Varscan 2 has been modified from the previous release of Varscan to include tools to detect somatic mutations and copy number aberrations (CNA) in matched tumour normal pairs [200].

During this quality control stage, the ratio of mutant reads compared to wild type reads was observed to differ in the majority of cases from the heterozygous 50:50 ratio. This increased frequency of mutant bases would therefore suggest that there has been a loss of the wild type allele. No differences were observed in LOH status and the type of germline *TP53* mutation. Those cases with a lower % of mutant reads can in most cases be explained because of a decreased purity of

tumour DNA, increasing the proportion of germline DNA shown when proportions were normalised to cellularity. The exceptions to this are patients, 30091111 which performed particularly poorly, and 30091141 where the VarScan 2 pipeline failed to call the germline variant. Manually inspecting this position in IGV, showed the germline mutation was present but at a very low frequency. This would suggest loss of heterozygosity of the mutant allele. This patient did acquire a further *TP53* mutation that was present at an expected clonal level. The frequency of mutant *TP53* reads is described in table 32.

Patient	Germline *TP53* mutation	Number of reads	% of reads	% tumour cellularity	% reads normalised to cellularity
30091001	c.672+1G>T	269/360	74.72	70	100.00
30091007	c.659A>G (p.Y220C)	708/903	78.41	80	98.01
30091111	c.818G>A (p.R273H)	81/229	35.37	70	50.53
30091123	c.743G>A (p.R248Q)	1605/2531	63.41	70	90.59
30091125	c.743G>A (p.R248Q)	2001/2824	70.86	65-70	100.00
30091141	c.743G>A (p.R248Q)	-	-	-	-
30091402	c.714T>A (p.C238*)	2897/4121	70.30	70	100.00
30091504	c.766A>G (p.T256A)	5207/6988	74.51	75-80	99.35
30092702	c.743G>A (p.R248Q)	1756/2375	73.94	70	100.00
30092703	c.714T>A (p.C238*)	1926/3281	58.70	65-70	90.31
30101705	c.733G>A (p.G245S)	1496/2452	61.01	65-70	93.86
30105104	c.733G>A (p.G245S)	2114/3374	62.66	65-70	96.40
30114003	c.659A>G (p.Y220C)	1513/2608	58.01	65-70	89.25
30124101	c.993+2T>G	1155/1562	73.94	65-70	100.00

Table 32 Breast tumour *TP53* status in germline *TP53* carriers

Table to show the frequency of mutant *TP53* reads for germline *TP53* mutations.

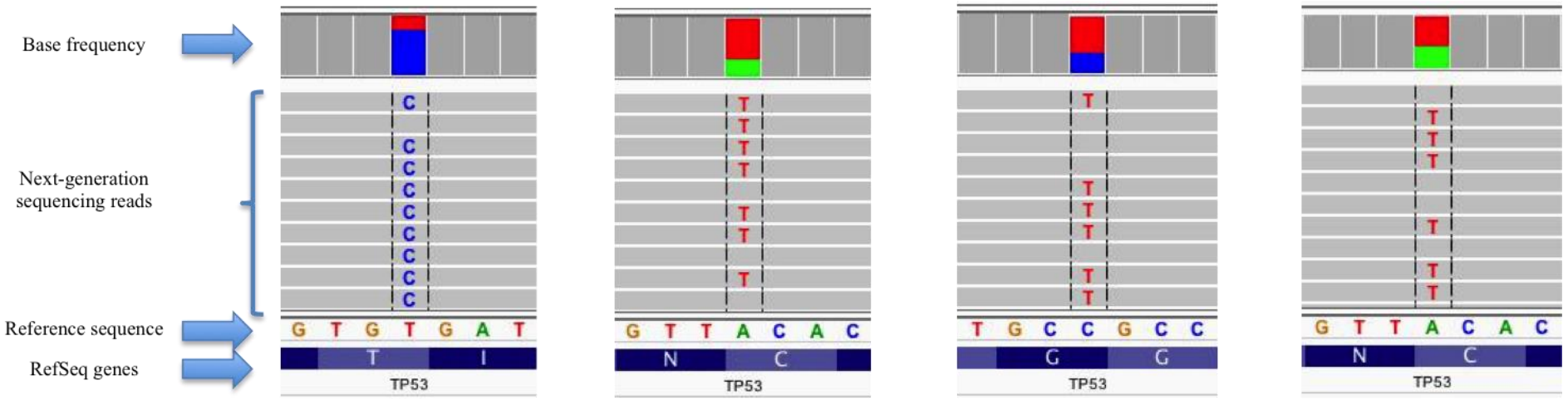
TP53 status was investigated in DCIS where NGS data was available. At the time of fixing, patients 30091104 and 30091401 had only presented with DCIS so the data from these cases was invaluable in interpretation of *TP53* status. These cases revealed a higher proportion of mutant

reads suggesting LOH of the WT *TP53* allele. As this LOH was also present in the DCIS, this would be suggestive of LOH of WT *TP53* being an early event driving tumorigenesis. The LOH DCIS data is shown in table 33.

Patient	**Germline *TP53* mutation**	**Number of reads**	**% of reads**	**% tumour cellularity**
30091104, 1997	c.818G>A (p.R273C)	168/267	62.92	90+
30091401	c.714T>A (p.C238*)	2025/2942	68.83	90+
30101705	c.733G>A (p.G245S)	1851/2570	72.02	90+
30124101	c.993+2T>G	926/1183	78.28	90+

Table 33 Breast DCIS *TP53* status in germline *TP53* carriers

Table to show the frequency of mutant *TP53* reads for DCIS cases derived from germline *TP53* mutations.



Patient	30091504
***TP53* mutation**	c.766A>G, p.T256A
Type of mutation	Missense
No. of reads	5207/6988
% of reads	74.51%
% tumour cellularity	75-80%
% of reads normalised	99.35%

Patient	30091402
***TP53* mutation**	c.714T>A, p.C238*
Type of mutation	Truncating
No. of reads	2897/4121
% of reads	70.30%
% tumour cellularity	70%
% of reads normalised	100.00%

Patient	30105104
***TP53* mutation**	c.733G>A, p.G245S
Type of mutation	Missense
No. of reads	2114/3374
% of reads	62.66%
% tumour cellularity	65-70%
% of reads normalised	96.40%

Patient	30092703
***TP53* mutation**	c.714T>A, p.C238*
Type of mutation	Truncating
No. of reads	1926/3281
% of reads	58.70%
% tumour cellularity	65-70%
% of reads normalised	90.31%

Figure 60 Tumour *TP53* loss of heterozygosity in *TP53* carriers

IGV traces of the germline *TP53* mutations in invasive breast cancer samples. Normalised reads have been normalised to the tumour cellularity.

4.5 Early genomic events in *TP53* carriers: matched DCIS and invasive tumour samples

Matched targeted sequencing data was available during the same disease episode for two cases: 30101705 and 30112401. The DCIS was laser captured and the invasive tumour was partly laser captured and partly macrodissected. Due to the nature of this sample and the substantial number of tumour microenvironment cells, it was not possible to gain a 100% tumour purity DNA sample. When selecting cases for NGS, the cut off was 60% for tumour cellularity in those diseased areas to maximise the chance of detecting somatic variation. Taking this into account, *TP53* status was investigated.

The germline *TP53* mutations in the COPE cohort are heterozygous so the VAF score should be at a frequency of 50%. An increase in the mutant *TP53* allele was observed in the DCIS precursor lesion. This would indicate that LOH of the WT *TP53* allele has already taken place in the DCIS suggestive of an early event in tumourogenesis. When the number of mutant *TP53* reads were normalised to the tumour cellularity of the sample, this frequency of reads increased further suggestive of widespread LOH of the WT allele. This observation supports the original hypothesis that an early loss of functional *TP53* drives a HER2+ breast tumour subtype. The matched data is shown in table 34.

Patient	Tissue	Germline *TP53* mutation	Number of reads	% of reads	% DCIS/tumour cellularity	% reads normalised to cellularity
30101705	DCIS	c.733G>A (p.G245S)	1851/2570	72.02	90+	-
30101705	Invasive tumour	c.733G>A (p.G245S)	1496/2452	61.01	65-70	93.86
30124101	DCIS	c.993+2T>G	926/1183	78.28	90+	-
30124101	Invasive tumour	c.993+2T>G	1155/1562	73.94	65-70	100.00

Table 34 LOH of wild-type *TP53* in matched DCIS and invasive breast tumours

Table to show *TP53* status in matched DCIS and invasive tumour samples. DCIS was laser captured therefore minimal contaminating microenvironment cells should be present in DCIS DNA samples. Invasive tumour samples were normalised to the tumour cellularity and due to the high purity, DCIS was not normalised to cellularity.

In addition, the somatic mutations were studied in matched DCIS and invasive tumour cases. As expected the number of filtered somatic variants were higher in the tumour samples compared to the matched DCIS. Pathways that were significantly affected in all samples were Notch and MAPK signalling, the cell cycle and the adherens junction pathway. The suggested deregulation of these pathways is unremarkable as these are often associated with HER2+ breast tumours. In addition, these pathways were also significantly affected in the control HER2+ samples (see table 29). DAVID Bioinformatics Resources 6.7 was used for pathways analysis [186, 187]

Patient	Tissue	Germline *TP53* mutation	Number of somatic mutations	Pathways affected by somatic variants	EASE score
30101705	DCIS	c.733G>A (p.G245S)	155	Notch signalling	$6.3x10^{-9}$
				Cell cycle	$5.8x10^{-3}$
				Adherens junction	$1.0x10^{-2}$
				MAPK signalling	$1.9x10^{-2}$
				Focal adhesion	$2.9x10^{-2}$
30101705	Invasive tumour	c.733G>A (p.G245S)	194	Notch signalling	$1.1x10^{-9}$
				Cell cycle	$2.2x10^{-3}$
				MAPK signalling	$3.2x10^{-3}$
				TGFβ signalling	$3.8x10^{-3}$
				p53 signalling	$1.4x10^{-2}$
				Focal adhesion	$1.6x10^{-2}$
				Adherens junction	$1.9x10^{-2}$
30124101	DCIS	c.993+2T>G	157	Notch signalling	$2.8x10^{-9}$
				Cell cycle	$3.9x10^{-3}$
				p53 signalling	$5.3x10^{-3}$
				Adherens junction	$7.5x10^{-3}$
				TGFβ signalling	$1.1x10^{-2}$
				MAPK signalling	$1.2x10^{-2}$
30124101	Invasive tumour	c.993+2T>G	201	Notch signalling	$9.0x10^{-10}$
				Adherens junction	$2.2x10^{-3}$
				MAPK signalling	$1.2x10^{-2}$
				Cell cycle	$1.2x10^{-2}$
				p53 signalling	$1.3x10^{-2}$
				Focal adhesion	$1.4x10^{-2}$
				TGFβ signalling	$2.5x10^{-2}$

Table 35 Affected pathways in matched DCIS and breast tumours from *TP53* carriers

Table to show the pathways significantly affected in matched DCIS and invasive tumours. Pathways that are not significantly affected in the matched DCIS and tumour samples are shown in bold.

4.6 Clonality and tumour evolution in breast tumour samples in *TP53* carriers

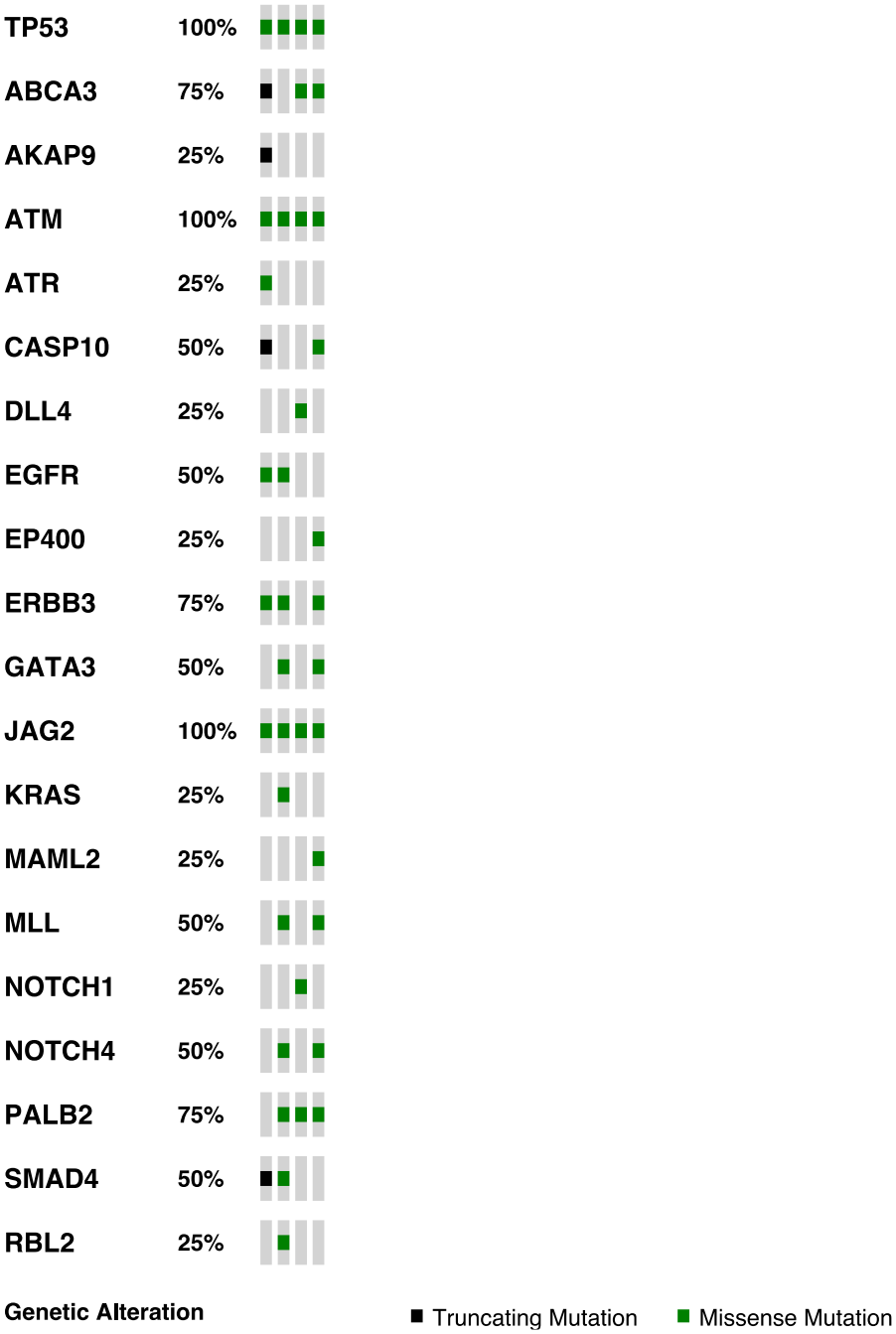
An additional hypothesis for this project was that an early loss in *TP53* leads to a dominant clone thus resulting in a low level of clonality. In order to assess the clonality of these unique tumours, three FFPE blocks from the same primary tumour were dissecting and processed for target enrichment. Furthermore, this allowed the LOH of WT *TP53* to be investigated and whether this pattern is present throughout the tumour. Clonality was investigated in two patients: 30091104 (missense mutation) and 30112802 (truncating mutation).

Across both patients there were few genomic lesions that seemed to be shared across primary breast tumour sites suggestive of a branching tumour evolution (see chapter 1.7 for additional information). Unfortunately due to a limited availability of clinical data, the exact location in which these sites derived from was unavailable. However due to the differences observed in between tumour samples, it would be fair to assume these were distant sites. Other than a mutation in *ATM* and *JAG2*, the only other lesion which was consistent across sites in the tumours and additional time points in patient 30091104, is this WT LOH of *TP53*. In patient 30091104 the only variant that was detected across all 3 tumour samples was a missense mutation in *PALB2*. Additional variants seem to be at low levels across the samples suggesting that it is this WT LOH of *TP53*, that is critical in driving tumourogenesis and the breast subtype observed in *TP53* carriers. This evolution is described in fig. 61.

a)

30091104 (c.818G>A, p.R273H)		**Mutant germline *TP53* status**				**Somatic status**
Year	**Tissue**	**Number of reads**	**% of reads**	**% DCIS/tumour cellularity**	**% reads normalised to cellularity**	**Number of somatic mutations**
1997	DCIS	168/267	62.92	90+	-	109
2001	Invasive tumour block 01	210/411	51.09	65-70	78.60	25
2001	Invasive tumour block 02	343/612	56.05	70	80.07	34
2001	Invasive tumour block 03	310/629	49.28	65-70	75.82	24

b)


TP53 100%
ABCA3 75%
AKAP9 25%
ATM 100%
ATR 25%
CASP10 50%
DLL4 25%
EGFR 50%
EP400 25%
ERBB3 75%
GATA3 50%
JAG2 100%
KRAS 25%
MAML2 25%
MLL 50%
NOTCH1 25%
NOTCH4 50%
PALB2 75%
SMAD4 50%
RBL2 25%
Genetic Alteration
Truncating Mutation
Missense Mutation

c)

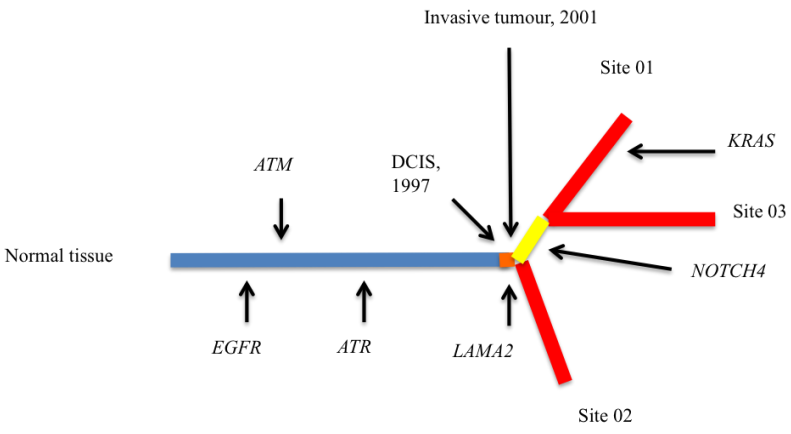

Invasive tumour, 2001
Site 01
KRAS
ATM
DCIS,
1997
Site 03
Normal tissue
NOTCH4
EGFR
ATR
LAMA2
Site 02

Figure 61 Tumour evolution in patient 30091104

Tumour samples showed a high level of heterogeneity within samples. *TP53* WT LOH is consistent throughout samples. a) Table to show number of genomic events and *TP53* status across samples. b) Heatmap to show shared genomic lesions. C) Phylogenetic tree showing disease evolution and acquisition of additional somatic variants. Lines are to scale based on the number of mutations.

Similar clonality findings were detected in patient 30112802. Very few genomic mutations were identified across all tumour areas apart from similar levels of WT *TP53* LOH. Details of this case are described in table 36.

30112802 (c.430C>T, p.Q144*)	**Mutant germline *TP53* status cellularity**				**Somatic status mutations**
Tissue	**Number of reads**	**% of reads**	**% DCIS/tumour cellularity**	**% reads normalised to cellularity**	**Number of somatic mutations**
Invasive tumour block 07	57/105	54.29	65-70	83.52	32
Invasive tumour block 08	2030/3081	65.89	70	94.13	25
Invasive tumour block 09	2613/4194	62.30	65-70	95.85	48

Table 36 Tumour heterogenity in patient 30112802

Table to show the number of genomic lesions and *TP53* status across tumour sites.

4.7 Discussion

From this study, breast tumours derived in patients with a germline *TP53* mutation typically developed HER2+ invasive breast tumour with associated widespread high grade DCIS. Our hypothesis was that an early loss of *TP53* was driving a HER2+ breast tumour phenotype. Evaluating *TP53* status from the VAF scores of the NGS data in this cohort, revealed LOH of the WT allele in 14/16 (87.5%) of invasive tumour samples. This seemed to be independent of the type of germline *TP53* mutation. Additionally the VAF score at the particular site of a patient's germline mutation would suggest that tumours are losing p53 activation through a LOH of the WT allele rather than through a gain of p53 function (GOF).

NGS data in patients 30091111 and 30091141 did not suggest that these tumours had undergone WT LOH. In fact, patient 30091141 seemed to have LOH of the mutant allele with a frequency of 1.1% (103/9469) of the confirmed germline mutation. In patient sample 30091111 the mutant allele frequency was 35.37% (81/229) and once normalised to the tumour cellularity, this rises to 50.53%. This could be because similar to patient 30091411, this patient has *TP53* LOH of the mutant allele and something else is driving tumourgenesis, or this patient's germline *TP53* mutation is in fact a *de novo* mosaic germline variant. Unfortunately we are limited with clinical, family history and additional germline data so investigating this further without this information is difficult. This kind of genomic behaviour has previously been reported in LFS in a young girl that presented with adrenocortical adenoma at 1 year and osteosarcoma at 5 years [201]. She was found to have allelic imbalance with only 1/3 of her gDNA sample tested to be heterozygous for the mutation [201]. A similar case was reported by Behjati et al. describing a young patient who presented with 3 separate primary cancers and a R248Q mutation in *TP53* was identified in 3-20% of sequencing reads [202].

Following on from the invasive tumour samples, *TP53* status was evaluated in the precursor lesion DCIS. The VAF scores in these samples also favoured the mutant allele suggestive of WT LOH. Patients 30091104 and 30091401 were of particular interest as these patients had only presented with DCIS at the time. Interestingly, these frequencies are different if the malignant cellularity is considered. DCIS samples were sequenced with little contaminating germline material through laser capture microdissection whereas tumour samples had contaminating microenvironment cells present. This contamination was unavoidable in the tumour FFPE samples as laser capture microdissection of individual contaminating immune cells for example, was just not feasible. Therefore considering malignant cellularity, tumour samples had a higher VAF score of mutant *TP53*. This would indicate that LOH is an on going process in DCIS and required for progression to invasive ductal carcinoma.

Clonality and the evolution of cancer in these patients were also investigated. Patient 30091104 was invaluable in underpinning this work as blocks were available for her 3 disease presentations: DCIS 1996, DCIS 1997 and invasive ductal carcinoma in 2001. When this patient first came to our

attention it was our intention to sequence samples from all 3 disease episodes however, the quality of DNA from 1996 was of too poor quality. DNA from the DCIS in 1997 did not fulfil the selection criteria so caution needs to be applied when investigating the number of somatic mutations and an acceptance of the limitations of this work. Additional work and extensive validation needs to be applied to those somatic mutations as it is highly likely that a proportion of those are false positives.

Despite the poor quality and limitations underpinning the DCIS sample from patient 30091104, this is not a concern for investigation into LOH of *TP53*. This is because the NGS reads were able to cover the known region of the germline variant so an assessment of WT LOH can be made. Tumour samples had a high level of contaminating DNA from surrounding microenvironment cells but what is striking, is if this contamination is taken into consideration, there is a consistent *TP53* VAF score. It appears that WT *TP53* LOH is not a clonal event and is consistent across the tumour originally described earlier in the precursor lesion. Additionally, the number of somatic mutations in the tumour appear to be low with these variants observed at frequencies just over 5%. However, these variants could be being masked by the contaminating germline DNA.

Similar observations were made in patient 30112802. Few somatic mutations were called with a small number of variants shared between sites. Some of these may be FFPE artefact due to fixation but this would be confirmed in the validation. It is possible that this cohort is displaying a field affect of low-level passenger mutations as a result of genomic instability through loss of p53 function. Therefore, a feature of *TP53* carriers from the two cases investigated in this study, is that patients have a small number of low level variants and that it is the loss of WT *TP53* that is critical to driving oncogenesis.

Somatic *TP53* mutations were detected in 5/9 (55.6%) of HER2+ control invasive tumours from the POSH cohort. These were detected at higher VAF frequencies than what would be expected for a clonal mutation suggesting that these *TP53* mutations are present throughout the tumour. If that is correct and these mutations are present in the majority of tumour cells, this would suggest that this genomic lesion was acquired early on in disease. DNA derived from the DCIS has been obtained in 4/5 of these patients and the presence of this mutation in their DCIS will be critical in determining if an early loss of *TP53* is driving a HER2+ breast tumour subtype.

Work by Silwal-Pandit and colleagues investigated the frequency of somatic *TP53* mutations in breast tumour subtypes finding that HER2+ (53%) and basal (65%) tumours have the highest frequency of mutations in this gene [70]. What is not known is why there is such a high frequency of mutations in *TP53* in two very biologically distinct subtypes. During the course of this project we wanted to understand why this was the case and if an early loss of *TP53* was driving a HER2+ subtype whereas a later loss of this gene was involved in basal breast tumours. Data would suggest

that a loss of p53 function is driving a HER2 phenotype so the next step for this work would be genomic investigation into basal tumours.

Ideally fresh tissue would be a better model to study this work in removing the technical difficulties of using FFPE and allowing single cell sequencing. Single cell analysis would allow heterogeneity to be investigated and also the removal of contaminating germline material using fluorescence-activated cell sorting (FACS) to separate and identify a pure population. Using single cell analysis would also allow investigation at the single cell level to identify if *ERBB2* amplification and *TP53* LOH are indeed occurring in the same cell. However, basal tumours do not typically present with DCIS causing difficulties in tracking tumour evolution. Work by Zhang et al. in non-small cell lung cancer investigated heterogeneity of *TP53* mutations and found that somatic and LOH events in this gene can be clonal [203]. From this they found that a loss of functional p53 despite its role as well established driver of tumorigenesis, was not always necessarily an early event [203].

To conclude, NGS data from *TP53* carriers and matched HER2+ control breast tumour cases have suggested that an early loss of *TP53*, could be driving a HER2+ breast tumour phenotype. In order to confirm this, data needs to be validated in tumour samples and matched DCIS. Clonality and evolutionary analysis for *TP53* carriers has suggested that the genomic landscape is littered with low-level genomic lesions which possibly could be because of a loss of p53 function field effect. *TP53* WT LOH seems to be the primary genomic lesion consistent across patients and areas of the tumour in clonality assessment.

Chapter 5: Cell biology models of breast cancers derived in *TP53* carriers

An additional approach to investigate this cohort of patients was using cellular biology methods to gain clarity on the mechanisms involved. The main clinical question which this work investigated was 'why do *TP53* carriers typically develop HER2+ breast cancer?' The morphology review confirmed that patients with a germline *TP53* mutation typically develop HER2+ breast cancer and it is well established in the literature that patients with a germline *BRCA1* mutation, go on to develop triple receptor negative breast cancer [204-206]. Using the immortalized nontransformed mammary epithelial MCF10A cell line, the effects of these oncogenes on cellular processes and their complementary status were investigated.

Furthermore, cellular biology assays were utilised to study this striking sclerotic stroma feature. From the morphology review, it appears that Li-Fraumeni associated breast carcinoma also develop this collagen rich sclerotic tumour stroma. This was significantly higher in this cohort compared to HER2+ invasive tumours (reported by Dr Matthew Sommerlad and Dr Guy Martland) and the various cohort subgroups in POSH that were reported as part of an investigation to test the robustness of using virtual slides for pathology review [82]. Primary matched normal surrounding fibroblasts (NBFs) and cancer-associated fibroblasts (CAFs) were obtained from patients with associated HER2+ and triple receptor negative invasive breast tumours. This area of work was undertaken to see if the receptor status of the tumour was playing a role in the expression of stromal genes in these cancer-associated fibroblasts. This would give some indication as to the role HER2 has if any, into the acquisition of a sclerotic stroma.

5.1 Expression of oncogenes: knockdown of p53 and BRCA1, and overexpression of ErbB2

In order to investigate the roles these oncogenes play in tumorigenesis, the expression of p53, BRCA1 and ErbB2 was tested and confirmed. Western blotting was utilised to confirm successful knockdown of p53 (fig. 63) and expression of ErbB2 (fig. 62) with quantitative PCR (qPCR) utilised to confirm knockdown of BRCA1 (fig. 64).

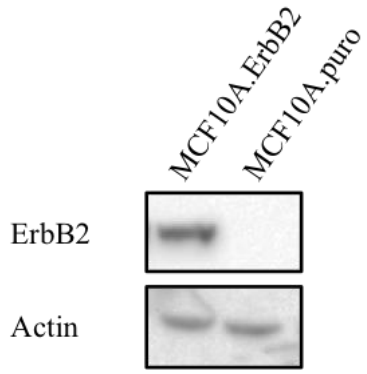


Figure 62 ErbB2 expression in the MCF10A cell line

Stable ErbB2 expression was confirmed by western blotting in the virally transduced MCF10A.ErbB2 cell line. The empty vector MCF10A.puro cell line was negative for ErbB2.

Transient knockdown of p53 was maintained and confirmed up to at least day 8. This was deemed a long enough knockdown for the cellular assays planned. Therefore, stable viral knockdown were not required and RNAi was utilised for all experiments (fig. 63).

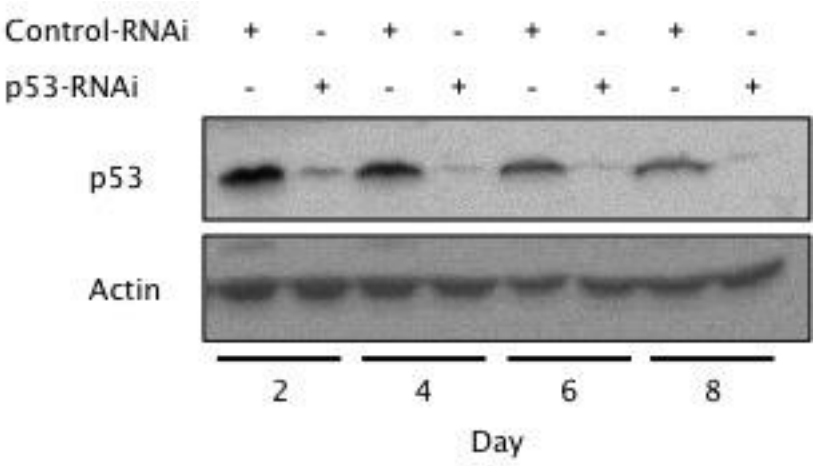


Figure 63 Transient knockdown of p53 in the MCF10A cell line

Transient knockdown of p53 was confirmed by western blotting in the MCF10A cell line.

BRCA1 knockdown was confirmed using qPCR. This approach was used instead of western blotting because of a poor BRCA1 antibody. BRCA1 knockdown is demonstrated in fig. 64.

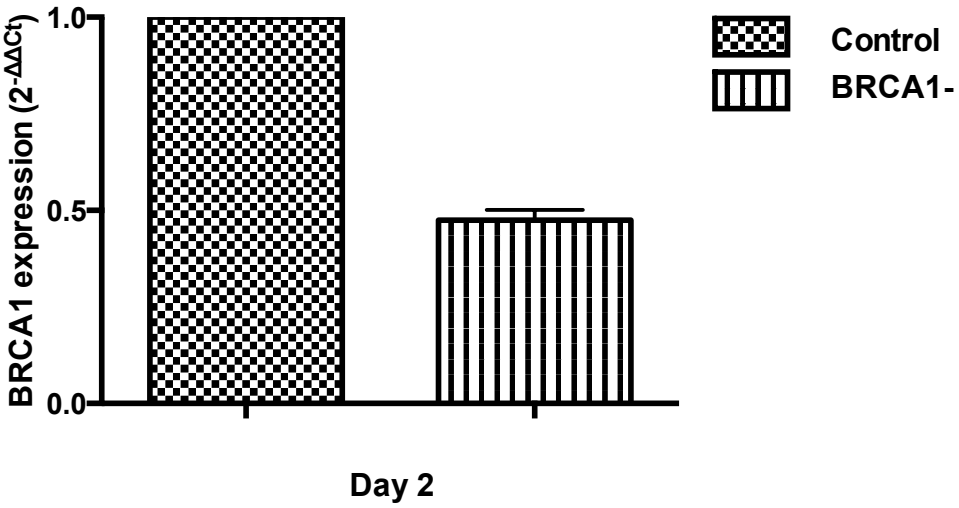


Figure 64 Transient knockdown of BRCA1 in the MCF10A cell line

Transient knockdown of BRCA1 was confirmed by qPCR in the MCF10A cell line.

5.2 The effect of oncogenes p53, ErbB2 and BRCA1 on proliferation

In order to address the clinical question 'why do *TP53* carriers typically develop HER2+ breast cancer?' proliferation was investigated in order to establish cooperation of oncogenes.

5.2.1 p53 and ErbB2 induces continued proliferation through contact inhibition

p53 and ErbB2 were first analysed individually against the MCF10A.puro control samples. Both of these oncogenes exhibited similar features in two-dimensional culture with significant differences in cell number consistently observed at day 8. Additionally, significant differences were observed at day 10 when ErbB2 alone was analysed. During proliferation, both oncogenes were observed to function similarly and this can be seen in fig. 65 with p53 seemingly affecting proliferation earlier. Very little differences were observed between MCF10A.ErbB2 control and MCF10A.ErbB2 –p53 with slight significance observed consistently at day 4 (n=3) (fig. 65.d-e). This data would suggest that ErbB2 expression and loss of p53 complement each other and would give cells a growth and survival advantage.

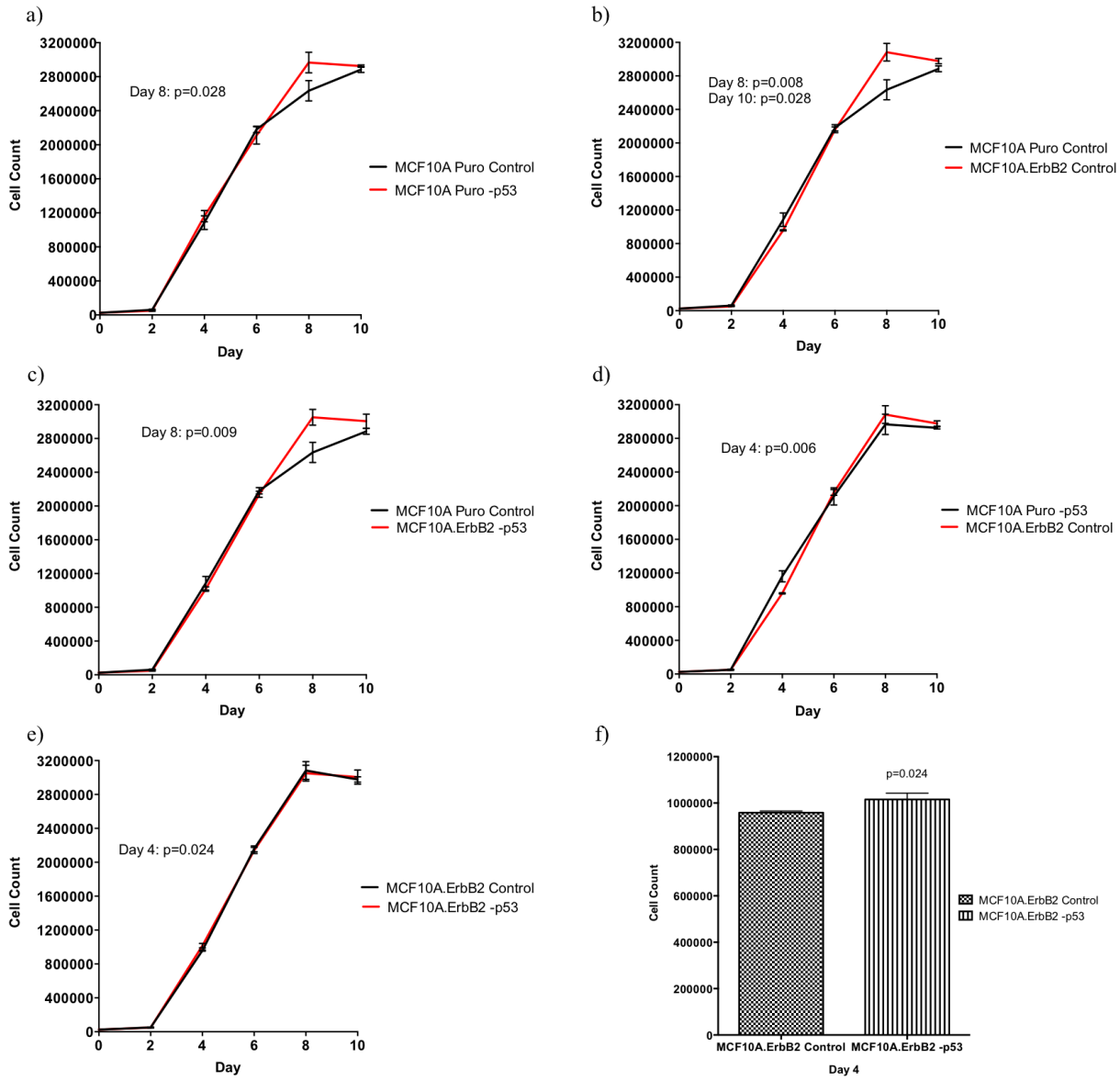


Figure 65 Cooperation of ErbB2 and p53 induces proliferation through contact inhibition

The affect of ErbB2 expression and a loss of p53 on proliferation. The figure shows 1 of 3 proliferation assays. The pattern of proliferation was consistent across assays. a) Comparison of MCF10A.puro control against MCF10A.puro –p53. b) Comparison of MCF10A.puro control against MCF10A.ErbB2 control. c) Comparison of MCF10A.puro control against MCF10A.ErbB2 –p53. d) Comparison of MCF10A.puro –p53 against MCF10A.ErbB2 control. e) Comparison of MCF10A.ErbB2 control against MCF10A.ErbB2 –p53. f) Bar chart to show significant day 4 comparison of MCF10A.ErbB2 control against MCF10A.ErbB2 –p53. Multiple t tests were used for statistics.

5.2.2 No proliferative differences were observed between loss of BRCA1 and BRCA1/p53

The affect of knocking down BRCA1 and BRCA1/p53 on proliferation was next investigated to see if loss of these oncogenes had a complementary affect. The original hypothesis for this was that the double BRCA1/p53 knockdown would result in cell lethality giving some insight as to why *BRCA1* carriers go on to develop triple receptor negative breast cancer, and why *TP53* carriers go on to develop HER2+ breast cancer. Surprisingly, little difference was observed between these conditions with both conditions experiencing an initial lag phase during the earlier time points.

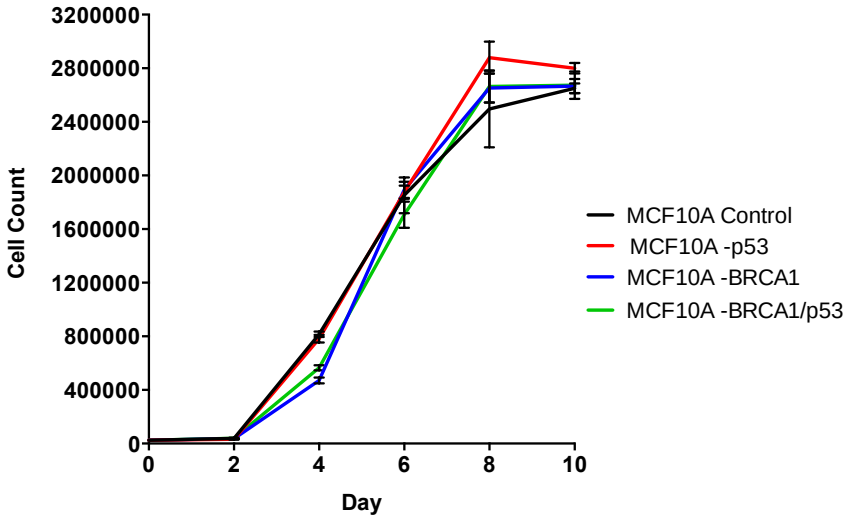


Figure 66 The effect of oncogenes on proliferation

Knockdown of BRCA1 and both BRCA1/p53 showed a similar trend in proliferation. . The figure shows 1 of 3 proliferation assays. The pattern of proliferation was consistent across assays. Statistical significance was only present at day 4 (p=0.005). Multiple t tests were used for statistics.

5.3 The effect of oncogenes p53, ErbB2 and BRCA1 on morphology in three-dimensional acinar culture

Taking this cell work further and the inconclusive evidence obtained from the two-dimensional proliferative assays for BRCA1 and BRCA1/p53 knockdown, these cells were next grown in three-dimensional acinar culture. The MCF10A cell line when cultured in modified assay media can be induced to grow in a way that mimics the ductal normal architecture of the breast. This model of the breast ductal system forms these hollow polarized, growth arrested acini-like spheroids. The model allows investigation into the precursor lesion ductal carcinoma *in situ* (DCIS) and its progression into invasive carcinoma. The transformed morphological phenotype of these cells were

analysed to determine cooperation of oncogenes and aggressiveness (phenotype differing from control).

5.3.1 Aggressive morphological phenotype of MCF10A.ErbB2 –p53 suggests cooperation of oncogenes

Supporting evidence obtained from the proliferation assay, the MCF10A.ErbB2 –p53 cell condition revealed the most intense invasive-like phenotype (fig. 67D). The cooperation of –p53 and ErbB2 in this culture resulted in a loss of polarity and these invasive fringes with culture over coming the growth arrest phenotype of the control. The control culture showed the expected aciniar spheroid growth structure with luminal clearing evident through the presence of apoptotic bodies (fig. 67A). The loss of p53 results in luminal filling and more of a disorganised structure (fig 67.B). Expression of ErbB2 supported previous phenotypic changes reported in the literature (fig 67.C) [188]. The phenotypic changes described can be seen in fig. 67.

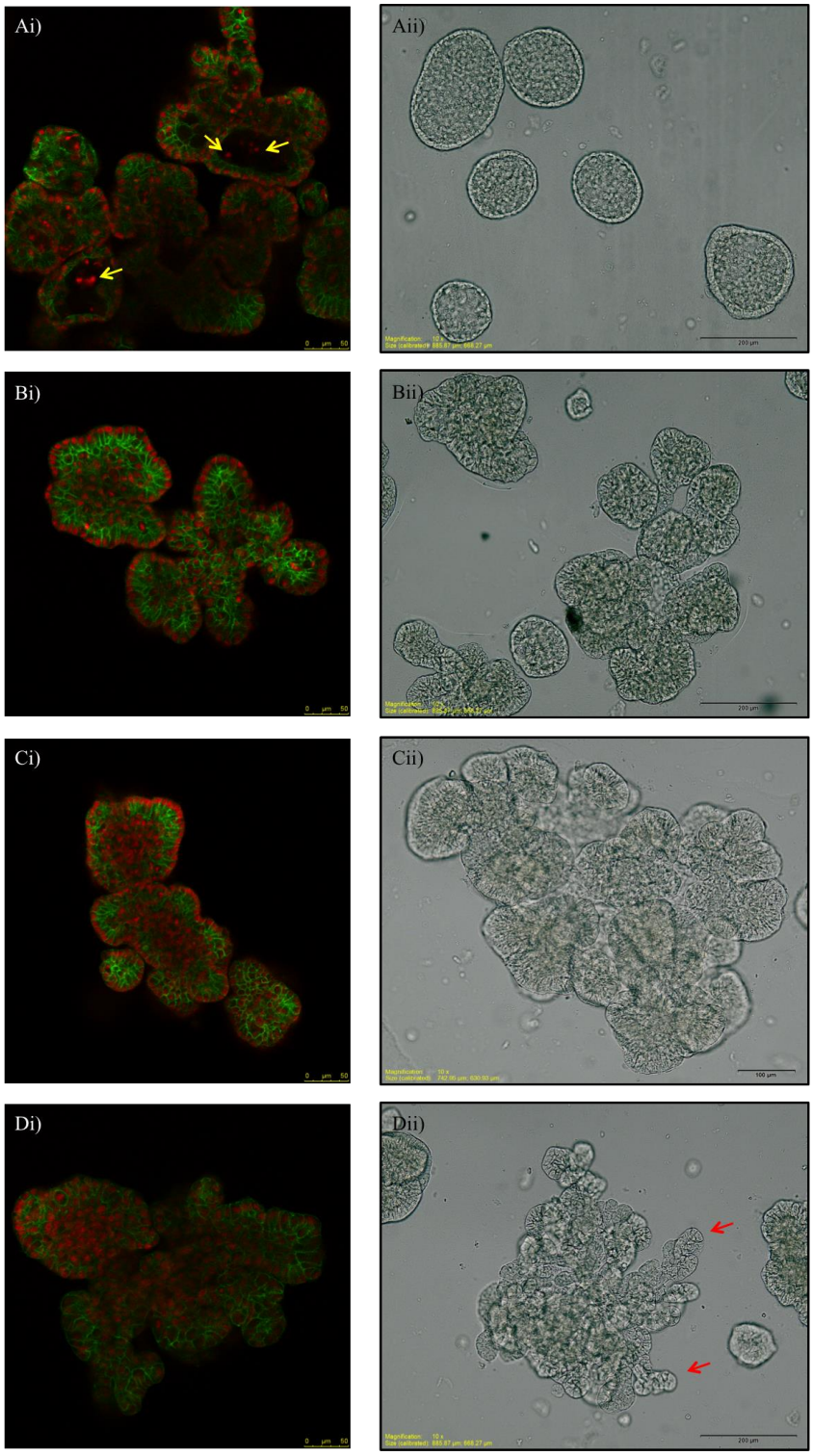


Figure 67 The affect of p53 and ErbB2 on three-dimensional phenotype

Confocal and light microscopy images of MCF10A culture when ErbB2 is expressed and p53 is knocked down. Ai) Confocal image of control MCF10.puro culture. A yellow arrow indicates an apoptotic body. Aii) Light microscopy image of control MCF10A.puro culture. Bi) Confocal microscopy image of MCF10A.puro –p53 culture. Bii) Light microscopy image of MCF10A.puro –p53 culture. Ci) Confocal microscopy image of MCF10A.ErbB2 control culture. Bii) Light microscopy image of MCF10A.ErbB2 control culture. Di) Confocal microscopy image of MCF10A.ErbB2 –p53 culture. Dii) Light microscopy image of MCF10A.ErbB2 –p53 culture. A red arrow indicates an invasive front. Confocal images were taken on a x20 magnification oil lens. Light microscopy images were taken at a 10x magnification on an Olympus DotSlide.

These cultures were quantified by scoring acini three-dimensional culture either ‘normal’ or ‘abnormal’. Each condition was scored against the puro control cases (fig. 68).

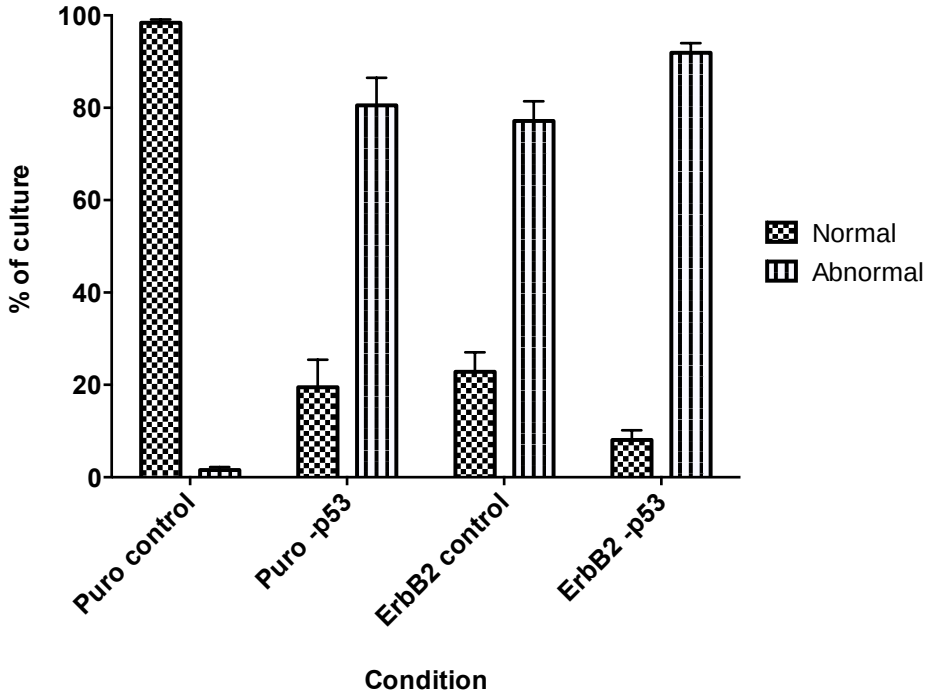


Figure 68 Quantification of three-dimensional phenotype in ErbB2 positive and p53 knockouts

Three-dimensional cultures were scored either normal or abnormal when compared to the Puro control culture (n=3). The software Fiji was used for scoring.

5.3.2 BRCA1 and p53 do not cooperate: a more aggressive morphological phenotype observed in knockdown of BRCA1 compared to BRCA1/p53

Despite no noticeable differences between knockdowns of BRCA1 and BRCA1/p53 in two-dimensional proliferative culture, clear differences in phenotype were observed in three-dimensional culture. Both cultures lacked luminal clearing but those with a single BRCA1 knockdown formed a more aggressive phenotype with protruding invasive fringes and larger three-dimensional cultures. The double knockdown of BRCA1/p53 seemed to revert back to a phenotype similar to cultures with a knockdown of p53 alone. However these double knockdown cultures appeared to be smaller in size. Three-dimensional cultures are presented in fig. 69.

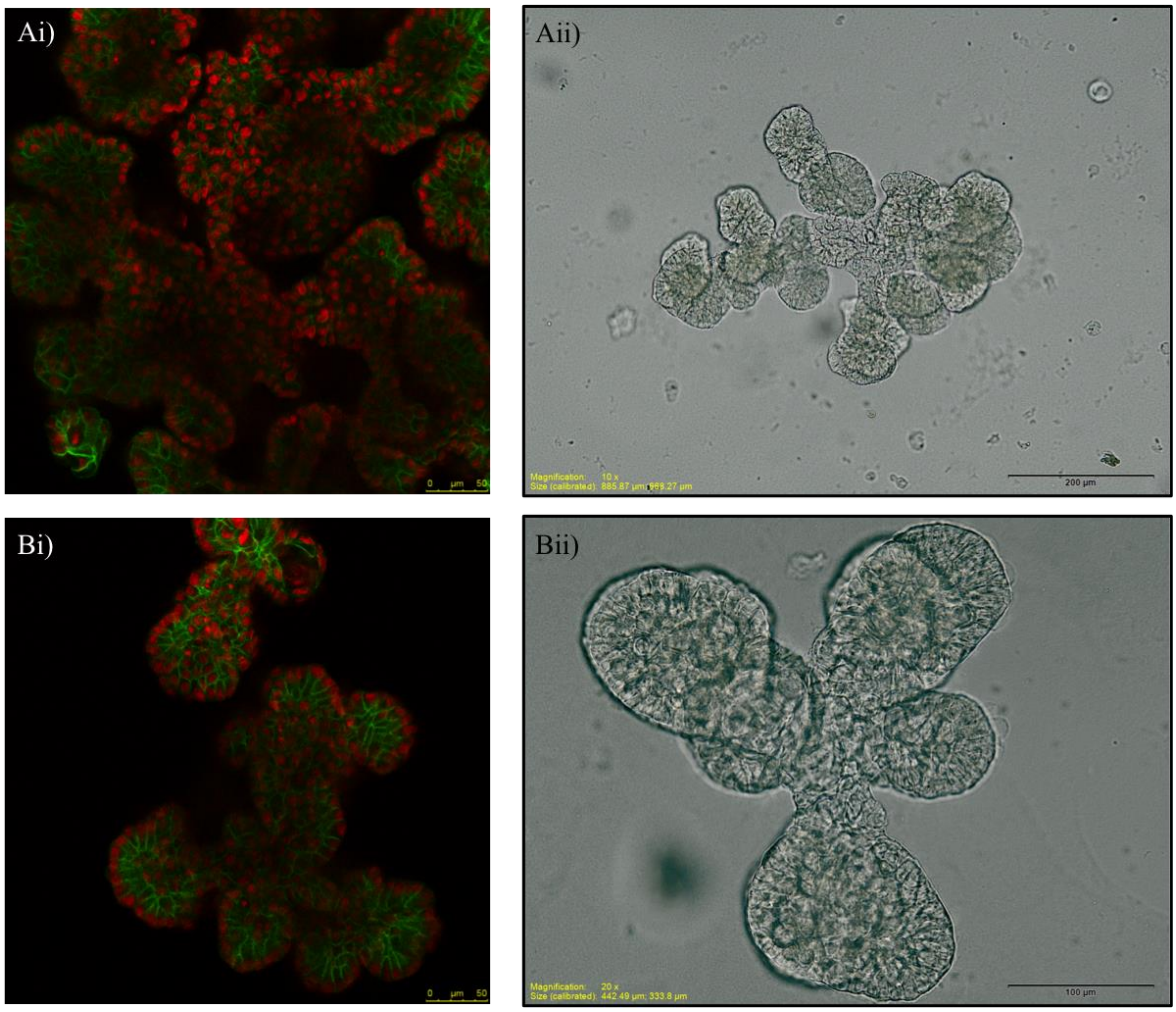
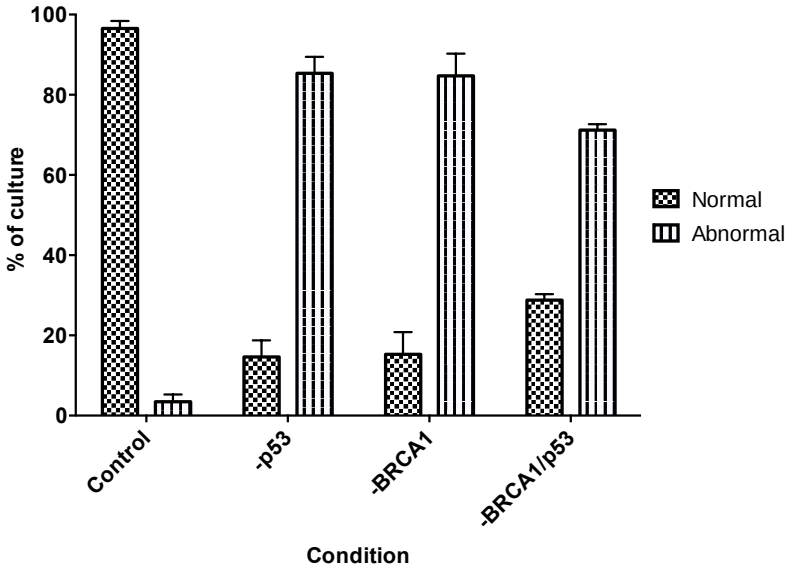


Figure 69 A more aggressive phenotype is observed in three-dimensional culture when BRCA1 only is lost

Confocal and light microscopy images of MCF10A culture when BRCA1 and BRCA1/p53 were knocked down. Ai) Confocal image of MCF10A -BRCA1 culture. Aii) Light microscopy image of control MCF10A -BRCA1 culture. Bi) Confocal microscopy image of MCF10A -BRCA1/p53 culture. Bii) Light microscopy image of MCF10A -BRCA1/p53 culture. Confocal images were taken on a x20 magnification oil lens. Light microscopy images were taken at a 10x (Aii) or 20x (Bii) magnification on an Olympus DotSlide.

Three-dimensional cultures were quantified. Additionally, the -BRCA1/p53 knockdown revealed a lower number of acini structures. This was significantly different compared to the -BRCA1 with a value of p=0.004 (n=3). The quantification is displayed in fig. 70.

a)



b)

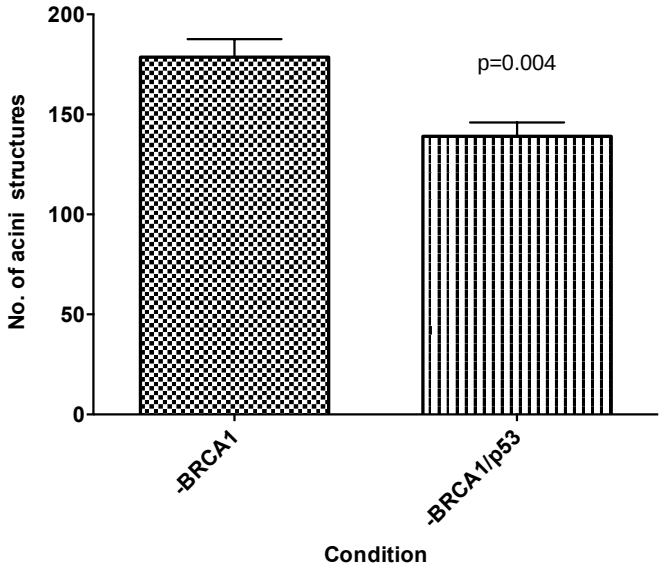


Figure 70 Quantification of three-dimensional phenotype in -BRCA1 and -BRCA1/p53 knockdowns

a) Three-dimensional cultures were scored either 'normal' or 'abnormal' when compared to the control culture (n=3). The software Fiji was used for scoring. b) The number of acini three-dimensional culture was significantly higher in the -BRCA1 culture compared to –BRCA1/p53 (n=3).

5.4 Variation in expression of stromal markers in HER2+ and triple receptor negative associated primary fibroblasts

The expression of stromal genes *ACTA2*, *COL1A1*, *FN1* and *CTGF* were investigated in matched primary normal breast fibroblasts (NBFs) and cancer-associated fibroblasts (CAFs). Expression data was analysed in two ways. Firstly the CAF expression data was normalised to their matched NBF. It is well established in the literature that CAFs typically have an increased coexpression of stromal markers and in particular *ACTA2*, a marker of CAFs, to their matched NBF [146-148, 207-210]. Contradictory to the literature, expression of *ACTA2* varied between NBFs and their matched CAFs which has also been found previously in our lab with head and neck primary fibroblast samples. The same was found across stromal genes showing significant variation between matched CAFs and NBFs (fig. 71).

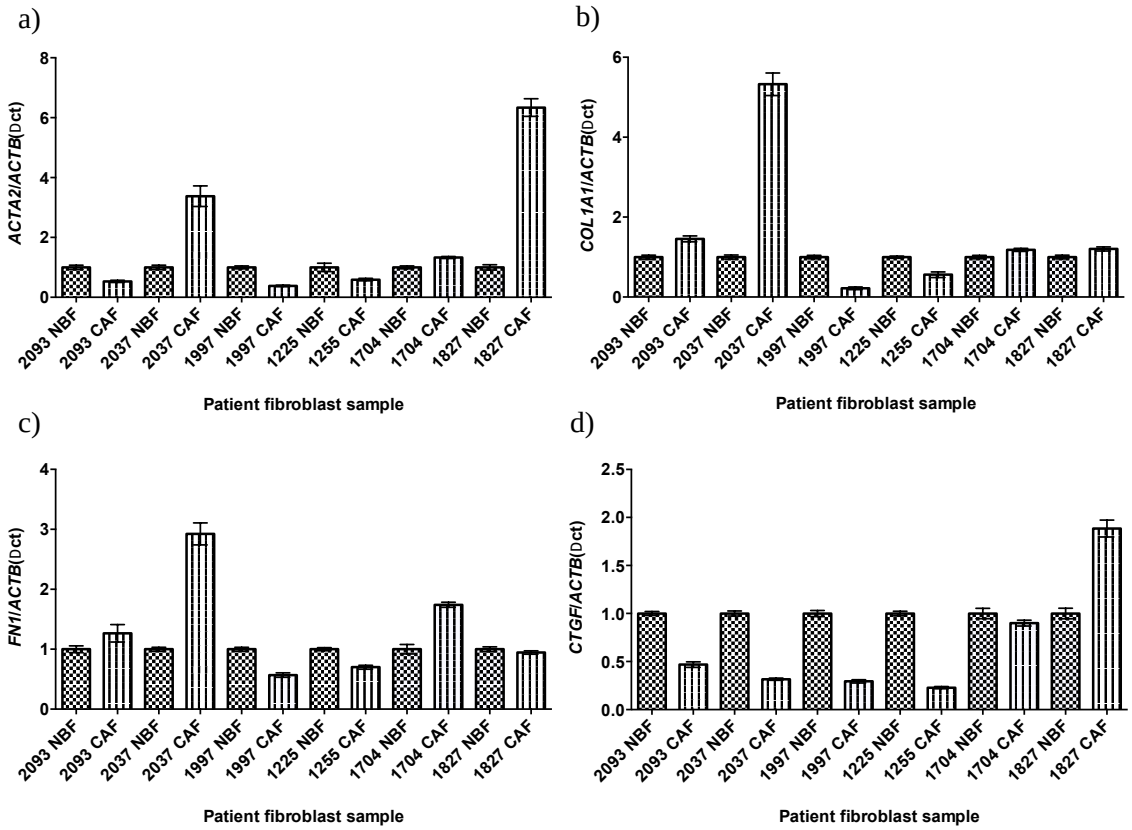


Figure 71 Expression of stromal markers in matched normal breast fibroblasts (NBFs) and cancer associated fibroblasts (CAFs)

Variation in the expression of *ACTA2*, *COL1A1*, *FN1* and *CTGF* in matched NBFs and CAFs. CAFs were normalised to their matched NBF. a) Expression of *ACTA2* across matched NBF and CAF pairs. b) Expression of *COL1A1* across matched NBF and CAF pairs. c) Expression of *FN1* across matched NBF and CAF pairs.) Expression of *CTGF* across matched NBF and CAF pairs.

Following on from this matched analysis, CAFs derived from associated HER2+ and triple negative breast tumours were analysed to see if there were significant differences observed between the two subgroups. CAFs were normalised to the average expression in the triple receptor negative CAFs for each gene. A similar pattern to the matched case analysis was observed with variation observed across genes (fig. 72).

a)

Patient sample	Associated breast tumour subtype
2093	Triple receptor negative
2037	Triple receptor negative
1997	Triple receptor negative
1255	HER2+
1704	HER2+
1827	HER2+

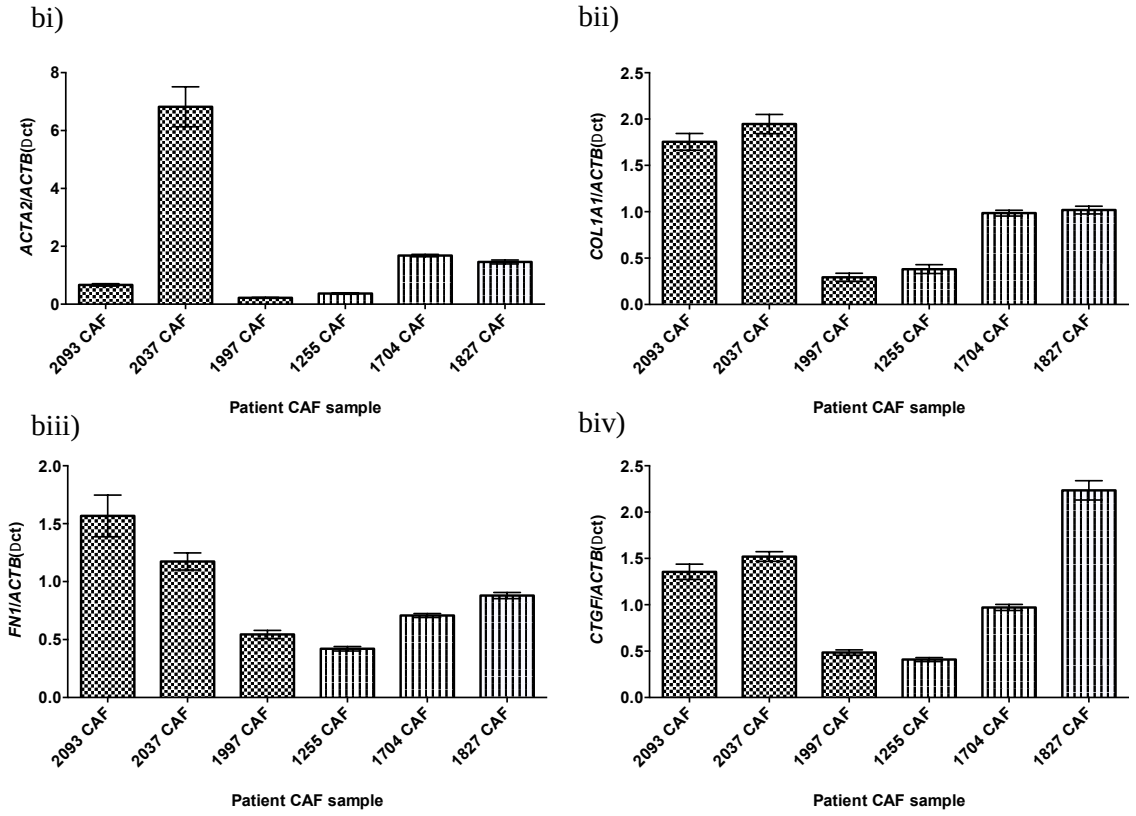


Figure 72 Variation of stromal marker expression in triple receptor negative and HER2+ associated CAFs

Variation in expression was observed across CAFs. a) Table to show patient CAF samples and associated breast tumour subgroups. bi) Expression of *ACTA2* across HER2+ and triple receptor positive associated CAFS. bii) Expression of *COL1A1* across HER2+ and triple receptor positive associated CAFS. biii) Expression of *FN1* across HER2+ and triple receptor positive associated CAFS. biv) Expression of *CTGF* across HER2+ and triple receptor positive associated CAFS.

5.5 Discussion

The work was undertaken as a different approach to try and answer the question 'why do patients with a germline *TP53* typically go on to develop HER2+ breast cancer?' It is well established in the literature that patients that have a *BRCA1* germline mutation typically go on to develop triple receptor negative breast cancer but less is currently known regarding the role *TP53* plays in driving a HER2+ breast tumour phenotype [204-206]. A retro-virally modified MCF10A cell line with ErbB2 overexpressed and successful knockdowns of p53 and BRCA1, allowed this clinical question to be investigated as well as allowing insight as to how DCIS progresses to invasive carcinoma.

These modified MCF10A cell lines were firstly grown in two-dimension to test the proliferative affects of these oncogenes. The original hypothesis was that ErbB2 and p53 complement each other and provides cells in culture a selective and survival advantage. ErbB2 and p53 were shown to have a similar affect resulting in contact inhibition whilst cells in culture became more confluent. Very little differences were observed when ErbB2 was expressed and p53 function lost in culture. A slight difference can be observed at day 4 (n=3) with proliferation slightly higher when p53 is lost as well as ErbB2 expressed. However the similar pattern of proliferation between the two suggests cooperation of these oncogenes.

This data supports the original hypothesis that these two lesions cooperate and complement each other. In two-dimensional culture these cells lose their contact inhibition and continue to proliferate. This was further demonstrated in three-dimensional culture with the apparent aggressive invasive-like phenotype of acinar structures. These structures now unrecognisable from the original puro control, lack polarity and form this disorganised organisation or protruding infiltrative invasive front. This lack of contact inhibition defined in two-dimension, allows an understanding as to why in three-dimensional culture, these ErbB2 positive p53 negative clones go on to form such large protruding structures. This work suggests that a key step in the evolution of DCIS to invasive carcinoma in *TP53* carriers, involves the loss of p53 function and the addition of ErbB2 expression. This would give abnormal cells in the DCIS a growth and invasive advantage allowing migration through the myoepithelial layer and invasion initiation.

Leading on from the cellular work investigating *TP53* carriers, attention was turned to why *BRCA1* carriers go on to develop triple receptor negative breast tumours. It was hypothesised that losing two key tumour suppressor genes in culture would result in cell lethality. Contradictory to this, in two-dimensional culture these double knockout cultures grew and behaved very similarly to the single BRCA1 knockdowns. There seemed to be an initial lag phase in growth when cells loss BRCA1 but proliferation increased post day 4.

From two-dimensional culture, little insight was gained as to why *BRCA1* carriers develop triple receptor negative tumours so three-dimensional culture was incorporated into the project. From three-dimensional culture, differences became obvious with the loss of BRCA1 function alone driving a more aggressive phenotype. Fewer aciniar cultures were able to grow when BRCA1 and p53 were knocked down simultaneously and their morphology also represented a similar phenotype to those cultures with a loss of p53. In conclusion from three-dimensional culture, a loss of BRCA1 and p53 do not cooperate and give cells a selective advantage.

There is a possible explanation as to this three-dimensional phenotype observed in the double BRCA1 and p53 knockout. This less aggressive phenotype observed in the double knockdown and similarity to a p53 alone culture, could be as a result of double knockout cellular death. These cultures could either be undergoing apoptosis as a result of losing two important tumour suppressor genes, or there could be a subclonal population in culture that unsuccessfully had BRCA1 knocked down and it is this population which has gone on to outcompete the double knockdown. Either of these explanations suggests that a loss of both of these genes do not give a selective advantage and biologically suggests why *BRCA1* carriers go on to develop triple receptor negative breast cancer and *TP53* carriers go on to develop HER2+ tumours.

When two and three-dimensional data is combined when investigating the BRCA1 story, a longer lag in growth can be seen in the double knockdown proliferation assay. This could explain as to why in three-dimension, these cultures are typically smaller than p53 alone knockdowns. If what we are seeing is a dominant p53 negative clone driving these assays, the starting cell count would be lower in the double knockdown due to this being a subclonal affect. This would support why three-dimensional acinar structures are smaller and also support genomic studies that have been presented in the literature. Tumours with somatic *TP53* mutations tend to play a dominant role and are typically reported in one of the larger clone populations [211].

Through breast genomic studies, groups have reported a high incidence of somatic *TP53* mutations in HER2+ (53%) and basal tumours (65%) [70]. Other groups have reported that basal tumours typically have extensive low-level chromosomal loss at an early stage. Potentially, this high genomic instability leads to disruption and convergence on similar pathways including the PI3K-AKT pathway and loss of *TP53* is a consequence of this. The absence of complementary status observed between BRCA1 and p53 in this cellular biology work, would suggest that this loss of *TP53* could be a later event rather than the suggested earlier event in HER2+ cases.

This work has given some insight as to why *TP53* carriers go on to develop HER2+ breast tumours and *BRCA1* carriers develop triple receptor negative tumours. Patients with these breast tumour subtypes often have the worst outcome and this cellular biology work supports what is well established in the clinic. Additionally, this complements data described in the morphology review. However, additional assays need to be performed to support these initial findings.

To continue this branch of the project, the pathways involved in breast tumour subtype could be investigated including the expression of pAkt and PTEN in the PI3K-AKT pathway. The heterogeneity of breast tumours is well documented and the underlying biology differs between subgroups. By investigating the expression of key regulators in pathways often deregulated in breast cancer, this model system would give some insight as to the order of molecular lesions acquired by these tumours.

Particularly in the -BRCA1 and -BRCA1/p53 system, there are still many unanswered questions. In order to address this potential *TP53* dominant clone theory, firstly the BRCA1 status needs to be confirmed in three-dimension once cultures are fixed. Following on from this an apoptosis assay could be utilized to test cell death in this double knockdown. This would allow confirmation as to whether this double knockdown causes cell lethality.

Additionally, ErbB2 expression and p53 loss in three-dimensional culture demonstrated an aggressive invasive-like phenotype. This work has suggested that an early loss of *TP53* with ErbB2 expression in the DCIS could be driving an invasive phenotype with cells breaking through the myoepithelial layer of the duct. DCIS is typically HER2+ so an addition loss of *TP53* could be a key event driving tumourogenesis. By taking this work into migration and invasion assays this mechanism and key driver events can be addressed further.

A key feature from the morphology review was that breast tumours derived in *TP53* carriers typically develop a sclerotic tumour stroma. The mechanisms involved in the development of this stroma were investigated using primary fibroblasts derived from associated HER2+ and triple receptor negative invasive tumours. Significant differences were observed in the morphology review when the type of tumour stroma were investigated between subgroups. Triple receptor negative tumours did not typically have this collagen rich stroma as a prominent feature. Therefore, by examining the expression of stromal markers in cancer-associated fibroblasts (CAFs) from these two subgroups, it was possible to address the hypothesis that HER2+ associated CAFs typically expressed a higher level of collagen and stromal genes compared to triple receptor negative associated CAFs.

On examination of *ACTA2*, *COL1A1*, *FN1* and *CTGF* expression, large variation was observed between matched normal breast fibroblasts (NBF) and CAFs in addition to variation observed between the two breast tumour subtypes. Unsurprisingly, there was significant patient variation and in order to investigate this from a subgroup stand, much higher patient numbers are required. Other groups have suggested that specific breast cancer subgroups have specific populations of CAFs with differing expression of stromal genes [212-214].

However, what has been observed in a tiny set of cases might actually be a reflection of the heterogeneity of CAFs which would support previous work published by our group [215]. Particularly as this high frequency of sclerotic tumour stroma in *TP53* carriers is significantly

higher than HER2+ tumours (p=0.007). This sclerotic stroma could be so prominent in this cohort because of the germline *TP53* mutation in these patients producing a specific CAF subtype. Work by Bao and colleagues have suggested that integrin αv is involved in the inactivation of wild type p53 and from the morphology review breast tumours derived in patients with a *TP53* mutation were shown to be typically integrin αvβ6 positive [216]. Additionally, work undertaken in cardiac fibrosis has suggested that a loss of p53 induces expression of collagen [217]. To gain further clarity of this mechanism further work and patient fibroblasts are required.

The next stage in this investigation would be to study these stromal markers in additional patients. Ideally, investigating stromal markers in CAFs derived from germline *TP53* carriers would yield the answers to these questions but due to the rarity of this genomic lesion at the germline level, realistically this is not feasible. A different approach to addressing this question of the role, if any, *TP53* is having on tumour stroma would be the transient knockdown of *TP53* in fibroblast cell line such as the HFFF2 line which is positive for wild type *TP53*.

To conclude, supporting cellular work has reinforced this hypothesis that an early loss of *TP53* and HER2 amplification complement one another. This work suggests that these lesions drive a cell type that has lost its contact inhibition producing an aggressive infiltrative three-dimensional phenotype. Additionally losing BRCA1 and p53 at an early stage offers no selective advantage whereas p53 loss and HER2 overexpression is advantageous for growth and invasion. The expression of stromal markers in CAFs demonstrated heterogeneity independent of associated tumour subtype.

Chapter 6: Final Discussion

This project has identified a specific breast tumour subtype which has been driven by a germline *TP53* mutation. Patients typically developed an ER+/HER2+ high grade infiltrative ductal carcinoma with widespread ductal carcinoma *in situ* (DCIS). The ER+/HER2+ receptor status, which was a feature of *TP53* carriers was significantly higher in this patient cohort compared to the matched POSH control group (p=0.002) suggesting this a p53 driven event.

The key finding during the course of the project was that tumourogenesis was being driven consistently through an early loss of heterozygosity (LOH) of the wild type *TP53* allele. This was firstly suggested by the p53 immunohistochemistry. All tumour cases exhibiting abnormal p53 staining with less intense more patchy staining identified in the DCIS precursor lesion suggestive of early event in the p53 pathway and stabilisation of the mutant p53 protein. This was additionally confirmed in the next generation sequencing data of invasive tumour and DCIS when *TP53* status was investigated. The LOH of the WT *TP53* allele was observed in 14/16 (87.5%) of invasive tumour cases and 4/4 (100.0%) of DCIS cases. Retention of the mutant germline allele indicates that loss of function of the *TP53* gene starts in the DCIS enabling the malignant cells to overcome this myoepithelial barrier and invade surrounding tissue.

The significance of this *TP53* loss of function as a key driver event in early tumourogenesis is supported by the lack of additional dominant driver variants throughout patient cases. Additionally, clonality and evolutionary data in patients 30091104 and 30112802 show a discrepancy of somatic variants across different areas of the tumour and time periods. This could be a sign of a field tumourogenesis effect of genomic instability and accumulation of somatic mutations because of a loss of functional p53. The WT LOH of *TP53* was consistent across each patient sample with the resulting VAF scores very similar between tumour sites. This would indicate further evidence supporting this argument that WT LOH is an early event.

In addition to WT LOH of *TP53*, tumours derived in a germline *TP53* background seemed to retain their HER2+ status. HER2 amplification in DCIS is not always overexpressed when disease evolves to become invasive [118, 126, 127]. A combination of WT LOH of *TP53* and HER2+ in certain DCIS clones would suggest that this is giving these cells a selective and invasive advantage. This was further demonstrated in the POSH control cases that were HER2+. Again taking contaminating germline DNA from microenvironment cells into consideration, all those HER2+ cases with a somatic *TP53* mutation had a high VAF score. This would suggest that these mutations are found throughout the tumour suggesting an early event. DCIS derived DNA was available for 4/5 of patients where a *TP53* mutation was detected so in order to answer this hypothesis, these mutations will need to be validated in the tumour and checked in the DCIS.

The cell biology aspect of this project has supported what was observed in the morphology and genomics regarding p53 loss of function and HER2 overexpression. HER2+ and a loss of p53 function were shown to complement each other driving a cell type that had lost its contact inhibition and thus producing an aggressive infiltrative three-dimensional phenotype. It is well documented in the literature that basal breast tumours and HER2+ tumours have the highest frequency of *TP53* mutations [70]. The *in vitro* experiments were designed to explore why two very distinct breast tumour subtypes have a mutation in the same tumour suppressor gene. The three-dimensional cell biology assays showed that losing either p53 in a HER2+ expressing environment, or BRCA1 alone produced the most aggressive invasive-like lesions as one might expect from the clinical data.

The lack of cooperation between a loss of BRCA1 and p53 became clear when grown in the three-dimensional culture model. Knocking down both genes produced fewer acinar structures but where structures were produced, their morphology was similar to cultures with *TP53* knockdown only. A possible explanation for this is that cultures could be undergoing cell death as a consequence of losing two tumour suppressor genes. Alternatively, proliferation in the double knockdown could be from a subclonal population of cells in which BRCA1 was not successfully knocked down and it is this population which has gone on to outcompete the double knockdown. Either of these explanations would suggest that loosing BRCA1 and p53 at an early stage offers no selective advantage whereas p53 loss and HER2 overexpression is advantageous for growth and invasion. Sequence data and *in vitro* cultures therefore suggest that where the function of both *TP53* alleles is lost through mutation and loss of heterozygosity in DCIS tumour cells, aggressive, invasive HER2 amplified breast cancer is the next step in tumour evolution. Migration and invasion cellular assays would be a potentially informative approach to further explore tumour invasion.

A sclerotic tumour stroma was a striking feature that we observed in the breast cancers of patients with a germline *TP53* mutation. This feature was significantly more common in COPE cases than the control group with HER2+ breast cancer selected from the POSH cohort ($p=0.007$) suggesting that this also relates to early loss of p53 function. The development of this sclerotic tumour stroma is initiated through the activation of the cytokine TGFβ. One of the major mechanisms in which TGFβ is activated is through the expression of integrin αvβ6 on the cell surface of tumour cells [159]. Once activated, TGFβ drives fibroblast differentiation and these differentiated myofibroblasts secrete the pro-tumourgenic components to the ECM that produce this sclerotic tumour stroma [160, 161, 163]. These myofibroblasts are identified through their high expression of α-SMA [162]. For more information see chapter 1.8. Using immunohistochemistry, expression of integrin αvβ6 was confirmed in 58.3% (21/36) of invasive COPE cases compared to only 15-16% noted in a published 2000 patient cohort suggesting this could be a p53 driven event [168]. Further evidence of TGFβ activation can be seen in 88.9% (32/36) of invasive tumour cases that

have high α-SMA and 63.9% (23/36) of tumour samples that had >75% of their TMA cores stained positive for pSMAD2/3.

Tumours with a strong stromal response have been linked with immune infiltration and vascular invasion through the barrier properties of collagen [146-148]. Breast tumours from the COPE cohort were shown to have a high incidence of vascular invasion and low level of lymphocytic infiltration which are both associated with a poor prognosis.

CAFs were either from associated HER2+ or triple receptor negative invasive breast tumours to see if HER2 status was affecting the expression of stromal genes such as *COL1A1*, one of the genes upregulated in a sclerotic stroma. No differences were observed between CAF groups revealing heterogeneity within invasive tumour subgroups supporting previous work by our group [215].

In order to explore further whether the type of *TP53* mutation may be relevant in driving a particular CAF phenotype, data from the morphology review was investigated further to see if there was a correlation between a sclerotic tumour stroma and the type of germline *TP53* mutation. This correlation is displayed in table 37.

Type of germline *TP53* mutation	**Tumour stroma**		
	Sclerotic	**Other**	**Total**
Missense	21/29 (72.4%)	3/7 (42.9%)	24
Truncating	8/29 (27.6%)	4/7 (57.1%)	12
Total	29	7	36

Table 37 Type of *TP53* mutation affects the type of tumour stroma

Patients with a missense mutation appear to be more susceptible to developing a sclerotic tumour stroma

Based on the low frequency of those patients with truncating *TP53* mutation developing a sclerotic tumour stroma, it appears the type of *TP53* mutation is driving the stromal response. Work by other groups has implicated p53 in the upregulation of collagen previously [169]. A loss of p53 was shown to upregulate TGFβ signaling and as a consequence the transcriptional activity of COL1A2 and collagen synthesis [169]. Work in murine models using a system that was heterozygous for $p53^{+/-}$, tumours developed an extensive proliferative stromal reaction that was positive for α-SMA and S100A4, a specific marker for fibroblasts [170]. Lastly, multiple groups have previously

suggested that through indirect cellular contact, tumour cells can inhibit WT p53 activation and stimulate immunosuppression [171, 172].

Taking all of this into account, integrin αvβ6 expressed on the DCIS and tumour cells could be activating TGFβ which in turn is driving myofibroblast differentiation identified through positive α-SMA expression. The missense mutation present in the majority of sclerotic stroma cases, could then be demonstrating a gain of function (GOF) effect and consequently silencing the wild type *TP53* allele. p53 Immunohistochemistry analysis of the stroma was negative unlike the tumour suggesting that the wild type allele is still present.

Bringing all aspects of the project together, the proposed mechanism in which these tumours evolve is illustrated in fig. 73.

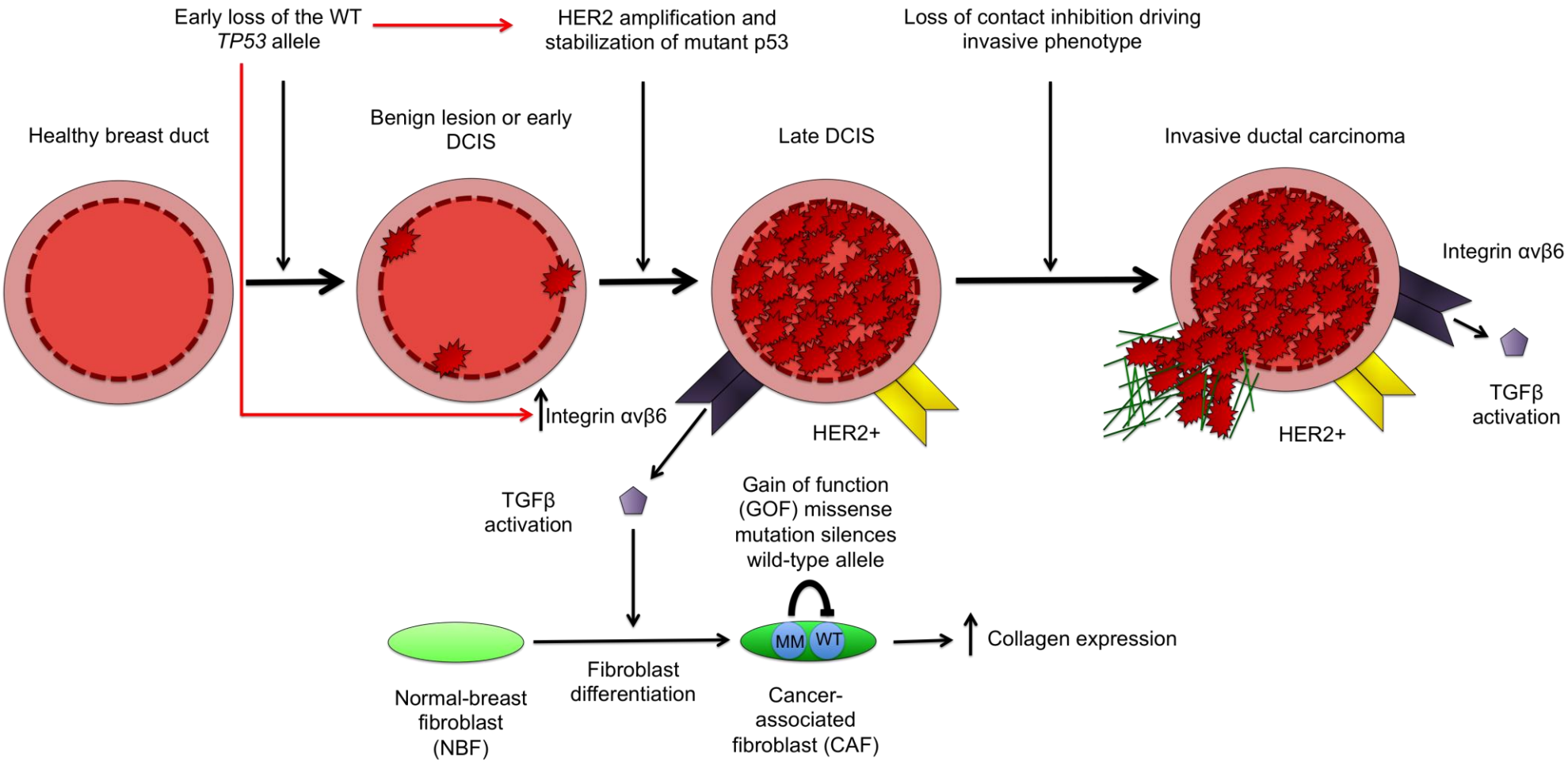


Figure 73 Proposed mechanism in the breast tumour evolution of *TP53* carriers

Proposed breast cancer evolution mechanism in germline *TP53* carriers. MM, missense mutation; WT, wild type.

There have been few studies that have investigated the genomic landscape in the tumours of patients with Li-Fraumeni Syndrome. These have all been individual case reports so this study, is to the best of our knowledge, the largest of its kind that has investigated the genomic landscape and morphologically reviewed a large cohort of these patients. Individual case reports have found a similar range of findings including additional known somatic cancer variants and ErbB2 signalling deregulation [218-220]. In addition, large chromosomal rearrangements have been detected in these patients but with the limitations of the patient material in this study, data such as this was not obtainable.

One of the largest challenges to overcome during this project was to extract DNA of an acceptable quality for next generation sequencing. Germline mutations in the *TP53* gene are extremely rare in the general population, hence it was only possible to recruit 45 patients presenting with breast disease. Many of these cases were archival with some blocks being in excess of 40 years old, prior to any genomic studies using FFPE material. Studies have revealed that there are many aspects of tissue processing which have a derogatory effect on genomic integrity including cold and warm ischemia, tissue size, duration of block storage and possibly one of the most significant factors, whether the formalin has been buffered [221]. Formalin has been used routinely as a fixative to retain tissue pathology for morphological assessment and to preserve cellular proteins for immunohistochemistry analysis. Unfortunately, through its role in the cross linking of proteins and DNA, nucleic acids are often limited, degraded and fragmented with genomic analysis options restricted [222-227]. In an ideal world, whole exome sequencing (WES) would be utilised in the COPE cohort to investigate somatic variation across all coding regions of the genome. Unfortunately, NGS to this scale requires a high input of good quality DNA and with the limited archival material obtained during recruitment, WES quite simply would not be feasible. One of the main advantages of targeted enrichment is the lower input of DNA required to achieve the necessary depth to investigate clonal somatic events. On the other hand, a disadvantage is that potential key genes implicated in tumourigenesis may be absent from the list of targets.

This revolutionary fast paced era of genomic medicine is allowing the release of exciting new technologies optimised for fragmented, poor quality, DNA extracted from FFPE. Through programmes such as Genomics England, histopathology laboratories are better prepared and educated in the preparation of human tissues in order to maintain genomic integrity. With tissue fixing becoming more standardised across the National Health Service (NHS) and an understanding of the implications each stage has on genomic integrity, the future looks promising for this area of research. Investigating the transcriptome is now a possibility in FFPE samples with whole or targeted approaches available. Illumina has recently released the TruSight® Tumor 170 panel which integrates DNA and RNA enrichment from FFPE cancer samples. For the COPE cohort, investigating the transcriptome could possibly have been an interesting route to take in order to investigate the expression profile of this sclerotic stroma phenotype. Potentially using a targeted

approach, the expression of genes implicated in TGFβ signalling and the ECM could have been investigated in order to decipher the mechanism [194]. However there is a limit to what is experimentally possible with archival samples that were degraded to the extent of this cohort. As previously stated, one of the major challenges of this project was to extract DNA of an acceptable quality for NGS. RNA is less stable and more severely affected by the fixing process than DNA and despite RNA extractions and transcriptomic studies having not been attempted in this cohort, the likelihood of successful data collection is doubtful [221, 228, 229].

This work highlights the likely order of events as breast cells progress to *in situ* and then invasive disease when loss of *TP53* function is an early event. It is reasonable to suppose that this is also an explanation for the development of sporadic HER2+ breast cancer and the sequencing data from the POSH study support this hypothesis. Loss of function of both copies of *TP53* will lead to genomic instability, perhaps this causes HER2 amplification and together this drives aggressive invasion. It is rational then to hypothesis that this same set of events underlie sporadic HER2+ invasive breast cancer where both loss of *TP53* and HER2 amplification lead to aggressive invasive disease and may explain why this is a relatively less common subtype of breast cancer since three specific molecular events may be required. In order to explore this further, fresh tumour material with matched DNA derived from blood would be easier and the opportunity to sample multiple areas across the tumour including DCIS in HER2 amplified and triple negative cancers. Even more powerful would be fresh tumour from patients with either germline *TP53* mutations or germline *BRCA1* mutations to compare.

Chapter 7: Appendices

7.1 Morphological assessment

Tumour type	Tumour grade	Tumour border	Stroma	Lymphocytic infiltration	DCIS grade	DCIS growth pattern
Ductal (NST)	1	Pushing	Cellular	Absent	Low	Solid
Pure special type (90% purity)	2	Infiltrative	Sclerotic	Mild	Intermediate	Comedo
Mixed tumour type (50-90% special type)	3		Desmoplastic	Prominent	High	Cribriform
	N/A		Myxoid			Micropapillary
			Other			Flat

Vascular invasion	DCIS
Absent	Absent
Present	Present

7.2 DNA extractions COPE: Germline FFPE material

Germline material was not available for the COPE cohort. DNA extractions were performed on healthy tissues from surrounding normal breast tissue (NB), lymph nodes (LN) negative for tumour and skin FFPE blocks. DNA taken from these areas was extracted to rule out germline variants

identified in the tumour. DNA was consistently extremely poor in these blocks and the data is summarised below. See chapter 2.3.6 for further information regarding passing PCR scores A to <C. A failed score (F) indicates no amplification of 105 bp fragments.

Patient	Tissue type	Nanodrop (260/280)	Qubit (ng/µl)	PCR based assay
30091104	LN 1	1.79	644	F
30091104	LN 2	1.83	586	F
30091104	LN 3	1.85	448	F
30091104	LN 4	1.75	756	F
30091104	Skin 1	1.86	86	F
30091104	Skin 2	1.87	137	F
30091104	Skin 3	1.87	107	F
30091141	LN	1.89	219.4	<C
30091128	NB	1.86	3.58	-
30092701	LN	2.06	6.46	-

7.3 HaloPlex design summary

Design Name: COPE

Design ID: 24212-1393950790

Species: H. sapiens (H. sapiens, hg19, GRCh37, February 2009)

Platform: Illumina

Read Length: 100 bp

Probegroup Summary

Number of Probegroups: 1

Probegroup 1 : COPE_1

Target Summary

Target Region Size: 1.498 Mbp

Amplicon Summary

Total Amplicons: 195187

Total Target Bases Analyzable: 1.476 Mbp

Total Sequenceable Design Size: 4.008 Mbp

Target Coverage: 98.54 %

Recommended Minimum Sequencing per Sample: 801.610 Mbp

Pricing: Illumina Tier 2 (Probe Region Size = 0.501 - 2.599 Mbp; up to 200K probes)

P/N:

Target and Probe Details

TargetID: The identifier entered in the Targets list.

Interval: The genomic interval of the target.

Regions: The number of regions within this target.

Size: The total size (in base pairs) of the regions.

Database(s): The databases in which this target was found.

Coverage: Bases overlapped by probes (extended +/- 100 bp) to represent likely capture.

High Coverage: Number of regions where analyzable amplicon overlap >= 90%.

Low Coverage: Number of regions where analyzable amplicon overlap < 90

TargetID	Interval	Regions	Size	Databases	Coverage (C.)	HighC.	LowC.
ABCA12	chr2:215797348-216002941	54	9009	Gencode, RefSeq, VEGA	100	54	0
ABCA3	chr16:2326665-2378623	31	5921	Gencode, RefSeq, VEGA	99.83	31	0

ABCG2	chr4:89013376-89080401	17	2457	Gencode, RefSeq, VEGA	97.56	16	1
ACTA2	chr10:90694970-90708697	8	1294	Gencode, RefSeq, VEGA	99.23	8	0
ACVR1B	chr12:52345518-52387904	10	2083	Gencode, RefSeq	100	10	0
AIM1	chr6:106808774-107016451	23	6934	Gencode, RefSeq, VEGA	99.91	23	0
AKAP9	chr7:91570404-91739483	52	13366	Gencode, RefSeq, VEGA	99.3	51	1
AKT1	chr14:105236668-105258990	13	1703	Gencode, RefSeq	100	13	0
ALK	chr2:29416080-30143535	31	5738	Gencode, RefSeq, VEGA	99.95	31	0
APC	chr5:112043405-112198243	19	9166	Gencode, RefSeq, VEGA	99.71	19	0
APOB	chr2:21224592-21266827	30	14375	Gencode, RefSeq, VEGA	99.86	30	0
AR	chrX:66764979-66943693	11	3116	Gencode, RefSeq, VEGA	99.78	11	0
ARID1A	chr1:27022885-27107257	20	7260	Gencode, RefSeq, VEGA	98.82	20	0
ARNT	chr1:150784487-150849053	23	2910	Gencode, RefSeq, VEGA	99.24	21	2
ATM	chr11:108098342-108236245	62	10411	Gencode, RefSeq	98.68	59	3
ATP8B1	chr18:55315710-55399049	27	4296	Gencode, RefSeq, VEGA	99.95	27	0
ATR	chr3:142168261-142297556	49	9120	Gencode, RefSeq, VEGA	99.4	48	1
AXL	chr19:41725288-41765819	20	3085	Gencode, RefSeq	100	20	0
BARD1	chr2:215593390-215674303	12	2590	Gencode, RefSeq, VEGA	100	12	0
BMPR2	chr2:203242188-203424679	13	3377	Gencode, RefSeq, VEGA	99.44	13	0
BRAF	chr7:140426284-140624513	21	2799	Gencode, RefSeq, VEGA	99.39	20	1
BRCA1	chr17:41197685-41276123	24	6184	Gencode, RefSeq, VEGA	96.54	23	1
BRCA2	chr13:32890588-32972917	26	10777	Gencode, RefSeq, VEGA	98.49	25	1
CACHD1	chr1:64936418-65157254	27	4365	Gencode, RefSeq, VEGA	99.34	26	1
CASP10	chr2:202050491-202093816	11	2000	Gencode, RefSeq, VEGA	100	11	0
CASP6	chr4:110610476-110624561	7	1023	Gencode, RefSeq, VEGA	100	7	0
CASP8	chr2:202122945-202152207	13	2149	Gencode, RefSeq, VEGA	99.77	13	0

CAV3	chr3:8775553-8787563	2	496	Gencode, RefSeq, VEGA	100	2	0
CCNH	chr5:86690253-86708621	9	1209	Gencode, RefSeq, VEGA	100	9	0
CDH1	chr16:68771309-68868194	17	3135	Gencode, RefSeq, VEGA	99.78	17	0
CDK12	chr17:37618315-37687579	14	4753	Gencode, RefSeq	100	14	0
CELSR1	chr22:46759065-46933077	37	9807	Gencode, RefSeq, VEGA	99.49	36	1
CHEK2	chr22:29083875-29133281	20	2326	Gencode, RefSeq, VEGA	89.68	15	5
CNTNAP1	chr17:40834838-40850938	24	4635	Gencode, RefSeq	99.96	24	0
COL11A1	chr1:103343565-103573744	69	7012	Gencode, RefSeq, VEGA	98.35	67	2
COL12A1	chr6:75794896-75912518	66	10728	Gencode, RefSeq, VEGA	99.66	66	0
COL15A1	chr9:101706334-101832953	43	5090	Gencode, RefSeq, VEGA	99.65	42	1
COL7A1	chr3:48601829-48632602	119	11278	Gencode, RefSeq, VEGA	99.83	118	1
CREBBP	chr16:3777709-3930131	31	8153	Gencode, RefSeq, VEGA	99.61	30	1
CSMD1	chr8:2796097-4851948	73	12258	Gencode, RefSeq, VEGA	99.6	72	1
CUBN	chr10:16866964-17171774	70	12409	Gencode, RefSeq, VEGA	99.69	70	0
DAPK1	chr9:90113983-90322289	26	4888	Gencode, RefSeq, VEGA	99.92	26	0
DIP2C	chr10:323255-735528	41	5983	Gencode, RefSeq, VEGA	100	41	0
DLL4	chr15:41221857-41230242	11	2278	Gencode, RefSeq	100	11	0
DMD	chrX:31132798-33357392	88	13194	Gencode, RefSeq, VEGA	99.83	88	0
DNAH11	chr7:21582854-21940882	84	15338	Gencode, RefSeq, VEGA	98.3	81	3
DNAH5	chr5:13692083-13944557	79	15455	Gencode, RefSeq, VEGA	99.94	79	0
DVL2	chr17:7129174-7137591	15	2511	Gencode, RefSeq, VEGA	98.53	14	1
ECT2	chr3:172472311-172538037	25	3467	Gencode, RefSeq, VEGA	99.88	25	0
EFNA1	chr1:155100444-155106553	5	718	Gencode, RefSeq, VEGA	100	5	0
EGFLAM	chr5:38258847-38464098	25	3641	Gencode, RefSeq, VEGA	99.75	25	0
EGFR	chr7:55086961-55273320	32	4734	Gencode, RefSeq, VEGA	99.77	32	0

EP300	chr22:41488999-41574970	31	7865	Gencode, RefSeq, VEGA	99.16	30	1
EP400	chr12:132445155-132562228	53	10566	Gencode, RefSeq	97.38	52	1
EPHB1	chr3:134514464-134977972	17	3336	Gencode, RefSeq, VEGA	100	17	0
ERBB2	chr17:37855803-37884307	28	4360	Gencode, RefSeq	99.77	28	0
ERBB3	chr12:56474075-56495849	29	4745	Gencode, RefSeq	100	29	0
ERBB4	chr2:212248330-213403264	29	4552	Gencode, RefSeq, VEGA	100	29	0
FAT3	chr11:92085269-92624289	28	14488	Gencode, RefSeq, VEGA	99.3	27	1
FBXO32	chr8:124515603-124553264	9	1248	Gencode, RefSeq, VEGA	99.2	9	0
FLG	chr1:152275166-152287942	2	12226	Gencode, RefSeq, VEGA	84.96	1	1
FLNA	chrX:153576964-153599623	48	9059	Gencode, RefSeq, VEGA	100	48	0
FLNB	chr3:57994282-58156499	49	8990	Gencode, RefSeq, VEGA	99.78	49	0
FLNC	chr7:128470682-128498587	48	9138	Gencode, RefSeq, VEGA	98.99	46	2
FLT4	chr5:180030182-180076555	32	4975	Gencode, RefSeq, VEGA	99.54	31	1
FN1	chr2:216226268-216300535	46	8387	Gencode, RefSeq, VEGA	99.45	45	1
GATA3	chr10:8097609-8115996	5	1437	Gencode, RefSeq, VEGA	100	5	0
GOLGA6L2	chr15:23684680-23692338	9	2148	Gencode	83.01	8	1
GRB2	chr17:73316439-73389719	5	754	Gencode, RefSeq, VEGA	100	5	0
GRIN2D	chr19:48901640-48947204	12	4251	Gencode, RefSeq	99.88	12	0
GRLF1	chr19:47421923-47504642	8	4767	Gencode	99.71	8	0
HERC1	chr15:63901270-64067832	78	16157	Gencode, RefSeq	99.91	78	0
HIF1A	chr14:62162513-62213813	16	2908	Gencode, RefSeq, VEGA	99.93	16	0
HIF3A	chr19:46800324-46842889	18	2833	Gencode, RefSeq, VEGA	100	18	0
HMCN1	chr1:185703902-186159020	107	19048	Gencode, RefSeq, VEGA	99.79	106	1
HRNR	chr1:152185542-152195739	2	8593	Gencode, RefSeq, VEGA	68.74	1	1
INHBA	chr7:41729238-41739982	2	1321	Gencode, RefSeq, VEGA	100	2	0

IRAK4	chr12:44161905-44180528	11	1603	Gencode, RefSeq	100	11	0
ITCH	chr20:32981608-33095609	24	3192	Gencode, RefSeq, VEGA	99.81	24	0
ITPR1	chr3:4558166-4887919	60	9484	Gencode, RefSeq, VEGA	99.89	60	0
JAG1	chr20:10620136-10654188	26	4177	Gencode, RefSeq, VEGA	99.57	26	0
JAG2	chr14:105609022-105634767	26	4237	Gencode, RefSeq, VEGA	100	26	0
KIT	chr4:55524172-55604733	22	3379	Gencode, RefSeq, VEGA	98.96	21	1
KNDC1	chr10:134973962-135038404	30	5898	Gencode, RefSeq, VEGA	99.49	30	0
KRAS	chr12:25362719-25398328	5	787	Gencode, RefSeq	100	5	0
LAMA1	chr18:6942068-7117729	63	10692	Gencode, RefSeq, VEGA	99.46	62	1
LAMA2	chr6:129204381-129837502	67	10736	Gencode, RefSeq, VEGA	99.66	65	2
LAMA4	chr6:112430630-112575362	44	6790	Gencode, RefSeq, VEGA	100	44	0
LAMA5	chr20:60884382-60942311	83	12964	Gencode, RefSeq, VEGA	99.51	82	1
LAMB4	chr7:107664474-107763619	35	6207	Gencode, RefSeq, VEGA	100	35	0
LRP2	chr2:169985163-170218919	79	15592	Gencode, RefSeq, VEGA	99.9	79	0
LRRC7	chr1:70034313-70587580	30	5342	Gencode, RefSeq, VEGA	99.76	30	0
MACF1	chr1:39549829-39951476	107	26205	Gencode, RefSeq, VEGA	99.9	107	0
MAD2L1	chr4:120981263-120987899	5	718	Gencode, RefSeq, VEGA	99.03	5	0
MAGEC1	chrX:140992849-140996629	2	3470	Gencode, RefSeq, VEGA	84.61	1	1
MAML2	chr11:95712102-96075069	5	3571	Gencode, RefSeq	99.1	5	0
MAP2K4	chr17:11924194-12044587	13	1502	Gencode, RefSeq	100	13	0
MAP3K1	chr5:56111391-56189517	20	4939	Gencode, RefSeq, VEGA	100	20	0
MAP3K11	chr11:65365752-65381237	10	2744	Gencode, RefSeq	99.89	10	0
MAP3K14	chr17:43341993-43368121	15	3275	Gencode, RefSeq	99.73	15	0
MAPK8IP3	chr16:1756331-1818835	33	4683	Gencode, RefSeq, VEGA	100	33	0
MDM2	chr12:69202248-69233639	14	1945	Gencode, RefSeq, VEGA	100	14	0

MDM4	chr1:204494637-204527819	14	1925	Gencode, RefSeq, VEGA	100	14	0
MECOM	chr3:168802687-169381170	20	4216	Gencode, RefSeq, VEGA	99.5	20	0
MKLN1	chr7:130827662-131172497	19	2620	Gencode, RefSeq, VEGA	100	19	0
MLH1	chr3:37035029-37107120	20	2711	Gencode, RefSeq, VEGA	98.45	19	1
MLL	chr11:118307218-118392897	38	12846	Gencode	99.79	38	0
MLL3	chr7:151833907-152132881	62	16276	Gencode, VEGA	96.98	55	7
MLLT6	chr17:36861876-36881861	21	3861	Gencode, RefSeq, VEGA	99.27	20	1
MRE11A	chr11:94153281-94225977	21	2583	Gencode, RefSeq	100	21	0
MSH2	chr2:47630321-47739583	18	3344	Gencode, RefSeq, VEGA	98.62	17	1
MTOR	chr1:11166652-11319476	60	9217	Gencode, RefSeq, VEGA	99.54	58	2
MUC16	chr19:8959598-9091824	84	45208	Gencode, RefSeq	99.58	82	2
MUC17	chr7:100663407-100701335	13	13742	Gencode, RefSeq, VEGA	96.91	12	1
MUC5AC	chr11:1151617-1162370	15	2170	Gencode	99.95	15	0
MUC5B	chr11:1244335-1284412	54	19002	Gencode, RefSeq	82.98	53	1
MYLK	chr3:123332942-123512698	31	6365	Gencode, RefSeq, VEGA	99.42	30	1
NCOA3	chr20:46250982-46282171	21	4725	Gencode, RefSeq, VEGA	99.79	21	0
NCOR1	chr17:15935600-16097893	47	8523	Gencode, RefSeq, VEGA	99.27	46	1
NCOR2	chr12:124809938-124979807	50	8713	Gencode, RefSeq, VEGA	99.51	48	2
NEB	chr2:152342264-152589680	183	29442	Gencode, RefSeq, VEGA	88.1	159	24
NF1	chr17:29422318-29705959	60	9902	Gencode, RefSeq, VEGA	99.04	58	2
NOTCH1	chr9:139390513-139440248	34	8348	Gencode, RefSeq, VEGA	100	34	0
NOTCH3	chr19:15271463-15311732	33	7632	Gencode, RefSeq	100	33	0
NOTCH4	chr6:32163204-32191715	30	6612	Gencode, RefSeq, VEGA	98.85	30	0
NOX4	chr11:89059914-89224424	21	2367	Gencode, RefSeq	92.14	18	3
NRG2	chr5:139227492-139422664	15	3079	Gencode, RefSeq, VEGA	100	15	0

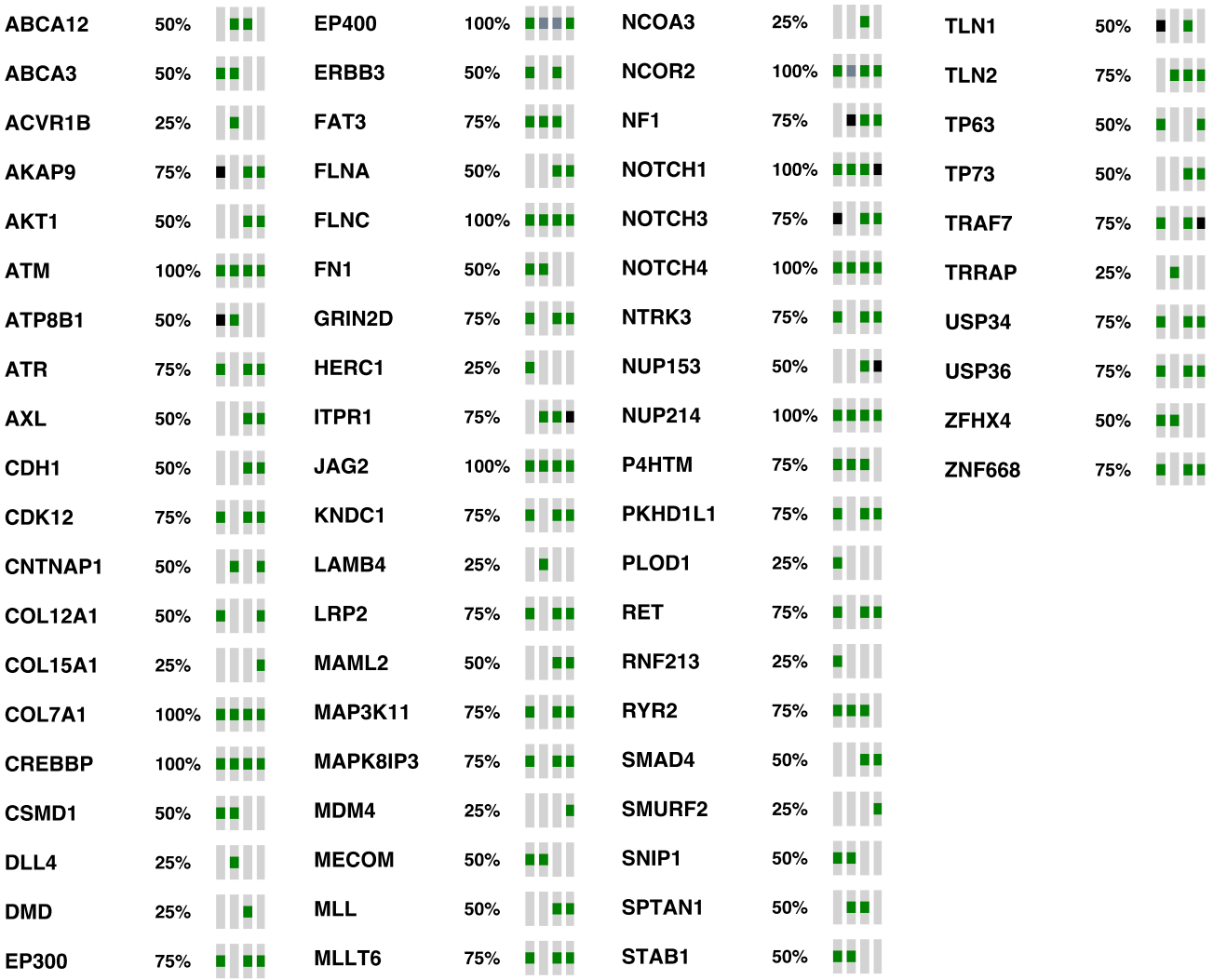
NTRK3	chr15:88420156-88799394	24	3342	Gencode, RefSeq, VEGA	100	24	0
NUP153	chr6:17616318-17706694	23	5047	Gencode, RefSeq, VEGA	100	23	0
NUP160	chr11:47800647-47869982	37	5335	Gencode, RefSeq	99.19	36	1
NUP214	chr9:134000983-134108884	38	7350	Gencode, RefSeq, VEGA	99.7	38	0
NUP98	chr11:3697379-3803357	34	6179	Gencode, RefSeq, VEGA	99.53	34	0
OBSCN	chr1:228399475-228566506	117	30353	Gencode, RefSeq, VEGA	99.82	116	1
OR9K2	chr12:55523543-55524570	1	1028	Gencode, RefSeq	100	1	0
P4HA1	chr10:74767970-74834651	15	1980	Gencode, RefSeq, VEGA	100	15	0
P4HA2	chr5:131528693-131630958	16	2160	Gencode, RefSeq, VEGA	100	16	0
P4HTM	chr3:49027680-49044350	9	1875	Gencode, RefSeq, VEGA	100	9	0
PALB2	chr16:23614770-23652488	13	3821	Gencode, RefSeq	100	13	0
PALLD	chr4:169432646-169847546	23	4778	Gencode, RefSeq, VEGA	99.67	23	0
PARP14	chr3:122399696-122447454	18	5996	Gencode, RefSeq, VEGA	100	18	0
PCM1	chr8:17793110-17885181	39	7036	Gencode, RefSeq, VEGA	99.72	39	0
PIK3CA	chr3:178916604-178952162	20	3609	Gencode, RefSeq, VEGA	98.5	18	2
PIK3CB	chr3:138374221-138478195	22	3669	Gencode, RefSeq, VEGA	99.84	22	0
PIK3CG	chr7:106507997-106545842	10	3509	Gencode, RefSeq, VEGA	99.15	9	1
PIK3R4	chr3:130398149-130464072	19	4457	Gencode, RefSeq, VEGA	99.89	19	0
PKHD1L1	chr8:110374800-110542329	78	14292	Gencode, RefSeq, VEGA	99.72	77	1
PLOD1	chr1:11994827-12034875	22	2891	Gencode, RefSeq, VEGA	98.79	21	1
PLOD2	chr3:145788494-145881415	22	2910	Gencode, RefSeq, VEGA	99.79	22	0
PLOD3	chr7:100849552-100861260	20	2747	Gencode, RefSeq, VEGA	99.82	20	0
POSTN	chr13:38137460-38172873	23	2971	Gencode, RefSeq, VEGA	99.29	23	0
PPM1L	chr3:160474087-160786955	6	1232	Gencode, RefSeq, VEGA	99.27	6	0
PRKCA	chr17:64298960-64800165	18	2436	Gencode, RefSeq	100	18	0

PTEN	chr10:89624217-89725239	9	1392	Gencode, RefSeq, VEGA	99.07	9	0
PTGFR	chr1:78958419-79002382	3	1211	Gencode, RefSeq, VEGA	100	3	0
RAD50	chr5:131893007-131978791	27	4751	Gencode, RefSeq, VEGA	99.49	27	0
RASGRF2	chr5:80256548-80521599	28	4296	Gencode, RefSeq, VEGA	100	28	0
RB1	chr13:48878039-49054217	27	3327	Gencode, RefSeq, VEGA	99.91	27	0
RBL1	chr20:35627152-35724341	22	3656	Gencode, RefSeq, VEGA	99.37	22	0
RBL2	chr16:53468459-53524222	24	4040	Gencode, RefSeq, VEGA	99.8	24	0
RET	chr10:43572697-43623727	20	3777	Gencode, RefSeq, VEGA	98.89	19	1
RNF213	chr17:78237471-78367308	69	17473	Gencode, RefSeq, VEGA	99.76	68	1
ROCK1	chr18:18531335-18690881	33	4725	Gencode, RefSeq, VEGA	92.42	29	4
ROR1	chr1:64240013-64644548	10	3043	Gencode, RefSeq, VEGA	99.8	10	0
ROR2	chr9:94456634-94712255	12	3267	Gencode, RefSeq, VEGA	100	12	0
RPS6KA3	chrX:20173506-20284760	23	2722	Gencode, RefSeq, VEGA	100	23	0
RYR2	chr1:237205812-237995957	111	17285	Gencode, RefSeq, VEGA	99.54	110	1
RYR3	chr15:33603237-34157437	107	16830	Gencode, RefSeq	99.83	106	1
SCN3A	chr2:165946650-166032914	27	6635	Gencode, RefSeq, VEGA	98.45	27	0
SEMA3A	chr7:83590677-83823912	17	2656	Gencode, RefSeq, VEGA	99.36	16	1
SEPT9	chr17:75277608-75495589	17	3199	Gencode, RefSeq	99.84	17	0
SH3RF2	chr5:145317482-145442274	9	2370	Gencode, RefSeq, VEGA	99.2	9	0
SLC22A3	chr6:160769442-160872098	11	1894	Gencode, RefSeq, VEGA	98.42	11	0
SMAD4	chr18:48573407-48604847	13	1999	Gencode, RefSeq, VEGA	100	13	0
SMC6	chr2:17846756-17927223	28	3966	Gencode, RefSeq, VEGA	96.55	26	2
SMURF2	chr17:62541956-62658008	19	2627	Gencode, RefSeq	99.85	19	0
SNIP1	chr1:38003339-38019840	4	1271	Gencode, RefSeq, VEGA	100	4	0
SOS1	chr2:39212955-39347573	24	4524	Gencode, RefSeq, VEGA	99.65	23	1

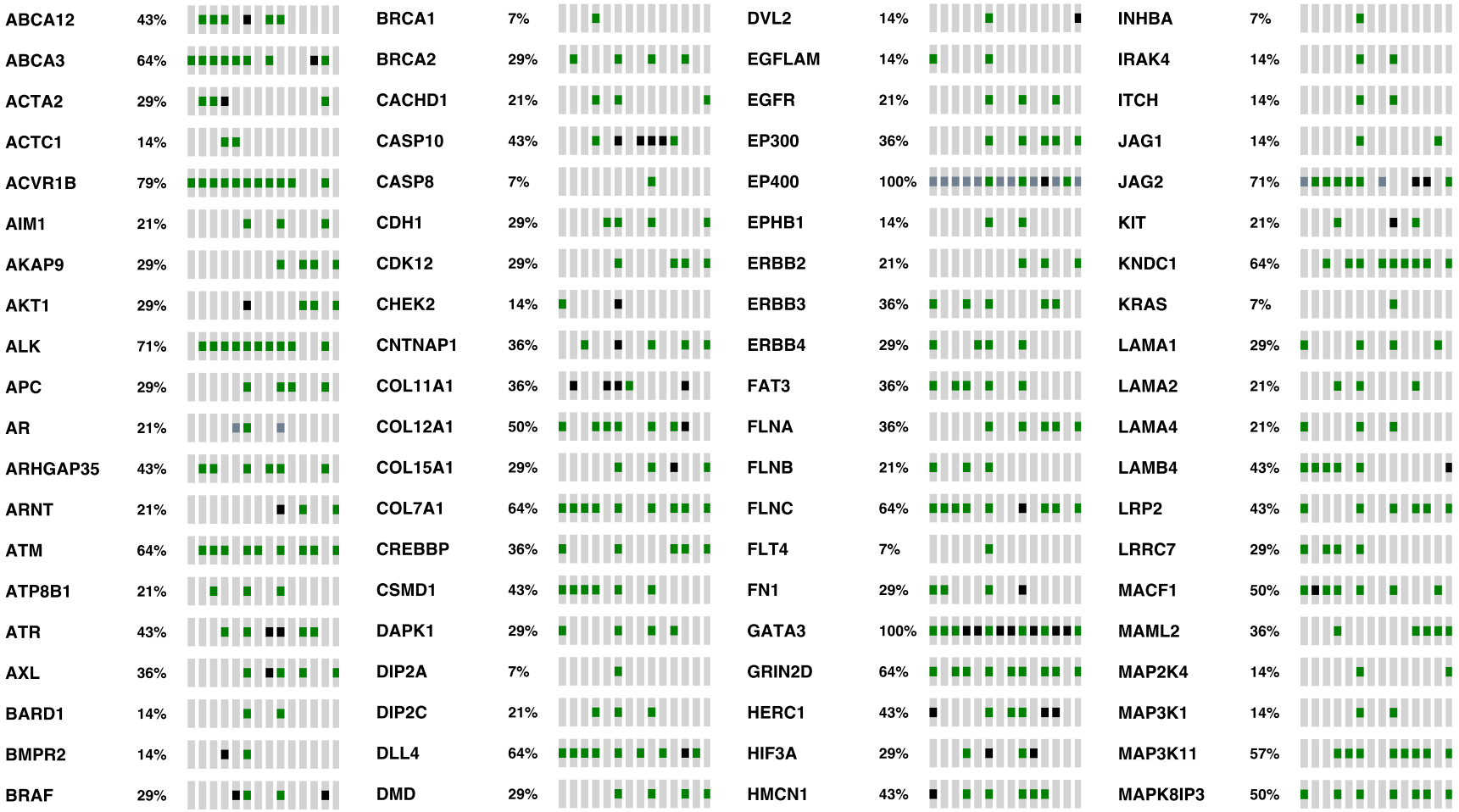
SPTA1	chr1:158581044-158656317	52	8300	Gencode, RefSeq, VEGA	99.89	52	0
SPTAN1	chr9:131329010-131395623	58	8647	Gencode, RefSeq, VEGA	99.6	57	1
STAB1	chr3:52529420-52558382	69	9093	Gencode, RefSeq, VEGA	100	69	0
STK11	chr19:1206903-1226656	9	1482	Gencode, RefSeq	100	9	0
SYNE1	chr6:152443561-152949476	151	30352	Gencode, RefSeq, VEGA	99.64	150	1
SYNE2	chr14:64375857-64692292	124	24169	Gencode, RefSeq, VEGA	99.84	124	0
TAF1	chrX:70586155-70749583	45	7017	Gencode, RefSeq, VEGA	98.93	43	2
TCF3	chr19:1611696-1650257	18	2717	Gencode, RefSeq	100	18	0
TEX14	chr17:56634362-56729372	32	5134	Gencode, RefSeq	100	32	0
TG	chr8:133879236-134147048	50	9450	Gencode, RefSeq, VEGA	100	50	0
TGFB3	chr14:76425520-76447246	7	1379	Gencode, RefSeq	99.85	7	0
TIAM1	chr21:32361891-32639298	27	5452	Gencode, RefSeq, VEGA	99.52	26	1
TLN1	chr9:35697778-35725701	56	8746	Gencode, RefSeq, VEGA	99.94	56	0
TLN2	chr15:62939500-63132819	56	8749	Gencode, RefSeq, VEGA	99.93	56	0
TLR7	chrX:12885688-12906787	2	3190	Gencode, RefSeq, VEGA	100	2	0
TP53	chr17:7565247-7579922	14	1697	Gencode, RefSeq, VEGA	97.53	13	1
TP63	chr3:189349295-189612301	17	2756	Gencode, RefSeq, VEGA	98.19	16	1
TP73	chr1:3598920-3649653	14	2230	Gencode, RefSeq, VEGA	99.87	14	0
TRAF2	chr9:139780940-139820363	11	1954	Gencode, RefSeq, VEGA	99.8	11	0
TRAF5	chr1:211526572-211546054	10	1874	Gencode, RefSeq, VEGA	100	10	0
TRAF7	chr16:2213912-2226585	20	2413	Gencode, RefSeq, VEGA	100	20	0
TRRAP	chr7:98478764-98609988	72	13098	Gencode, RefSeq, VEGA	99.92	72	0
TSC1	chr9:135771612-135804269	21	3987	Gencode, RefSeq, VEGA	99.92	21	0
TTN	chr2:179391729-179682294	364	121767	Gencode, RefSeq, VEGA	98.25	341	23
USH2A	chr1:215799113-216595688	72	17141	Gencode, RefSeq, VEGA	99.89	72	0

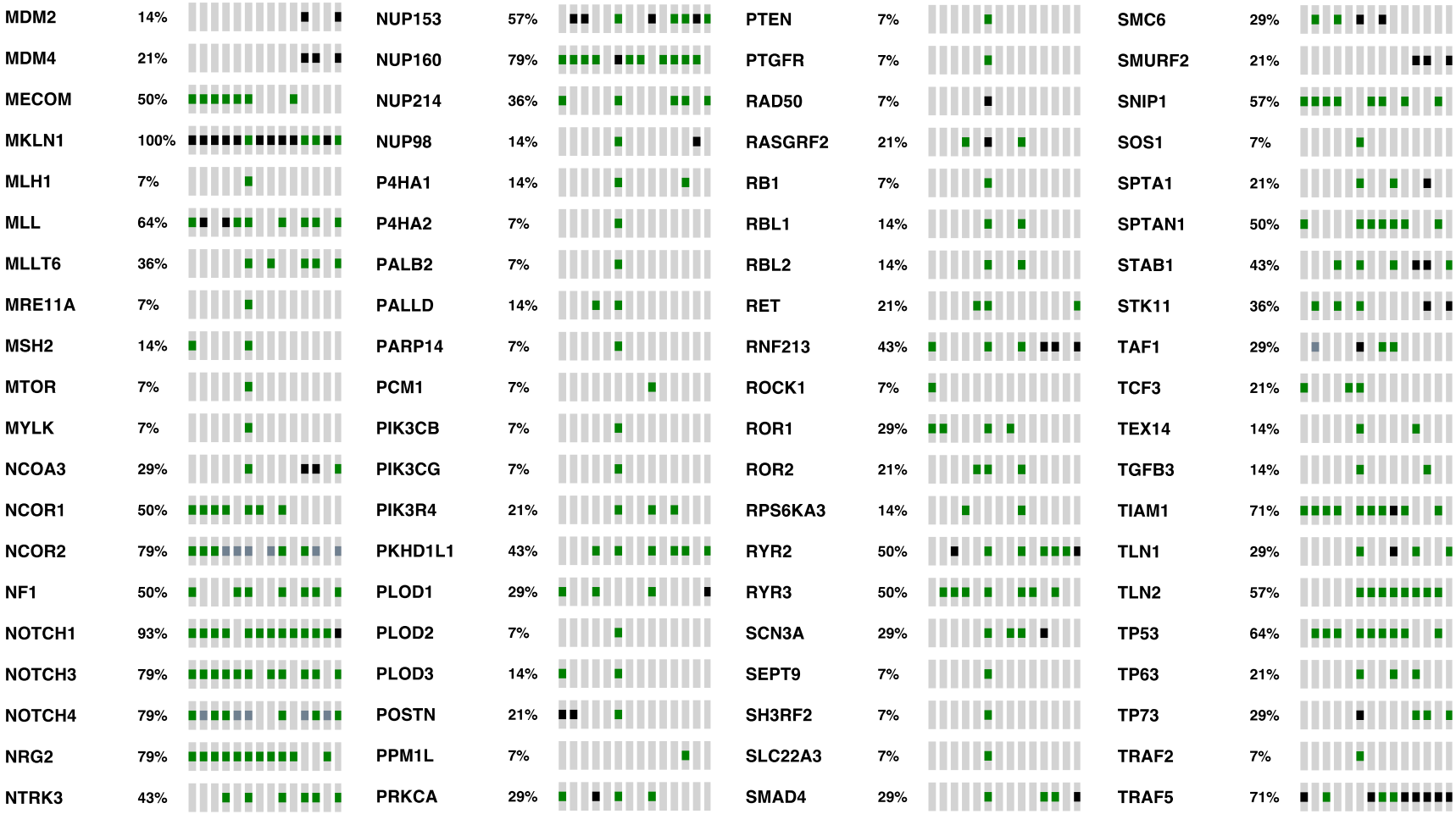
USP24	chr1:55534708-55680796	68	9223	Gencode, RefSeq, VEGA	99.8	67	1
USP32	chr17:58256606-58469310	35	5552	Gencode, RefSeq	87.99	29	6
USP34	chr2:61415227-61697837	80	12245	Gencode, RefSeq, VEGA	99.84	79	1
USP36	chr17:76794492-76832455	18	3732	Gencode, RefSeq	99.92	18	0
VAV1	chr19:6772809-6857128	27	3078	Gencode, RefSeq	100	27	0
VAV3	chr1:108115943-108507501	30	3381	Gencode, RefSeq, VEGA	100	30	0
VCPIP1	chr8:67543708-67579203	4	3765	Gencode, RefSeq, VEGA	100	4	0
VHL	chr3:10183522-10191659	3	702	Gencode, RefSeq, VEGA	99.15	3	0
ZEB2	chr2:145121053-145274927	13	4125	Gencode, RefSeq, VEGA	99.73	12	1
ZFHX4	chr8:77616314-77776811	13	11216	Gencode, RefSeq, VEGA	99.26	12	1
ZFYVE16	chr5:79729957-79773206	19	5060	Gencode, RefSeq, VEGA	99.94	19	0
ZNF423	chr16:49525176-49856606	8	4015	Gencode, RefSeq	100	8	0
ZNF668	chr16:31072379-31075949	3	1989	Gencode, RefSeq, VEGA	100	3	0
ZNF77	chr19:2896744-2944848	5	2072	Gencode, RefSeq	98.02	4	1

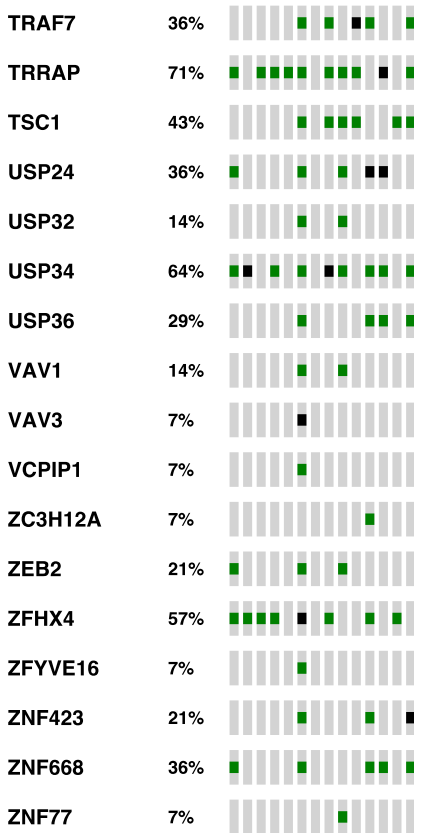
7.4 Full filtered gene list: COPE DCIS


ABCA12 50%
ABCA3 50%
ACVR1B 25%
AKAP9 75%
AKT1 50%
ATM 100%
ATP8B1 50%
ATR 75%
AXL 50%
CDH1 50%
CDK12 75%
CNTNAP1 50%
COL12A1 50%
COL15A1 25%
COL7A1 100%
CREBBP 100%
CSMD1 50%
DLL4 25%
DMD 25%
EP300 75%
EP400 100%
ERBB3 50%
FAT3 75%
FLNA 50%
FLNC 100%
FN1 50%
GRIN2D 75%
HERC1 25%
ITPR1 75%
JAG2 100%
KNDC1 75%
LAMB4 25%
LRP2 75%
MAML2 50%
MAP3K11 75%
MAPK8IP3 75%
MDM4 25%
MECOM 50%
MLL 50%
MLLT6 75%
NCOA3 25%
NCOR2 100%
NF1 75%
NOTCH1 100%
NOTCH3 75%
NOTCH4 100%
NTRK3 75%
NUP153 50%
NUP214 100%
P4HTM 75%
PKHD1L1 75%
PLOD1 25%
RET 75%
RNF213 25%
RYR2 75%
SMAD4 50%
SMURF2 25%
SNIP1 50%
SPTAN1 50%
STAB1 50%
TLN1 50%
TLN2 75%
TP63 50%
TP73 50%
TRAF7 75%
TRRAP 25%
USP34 75%
USP36 75%
ZFHX4 50%
ZNF668 75%

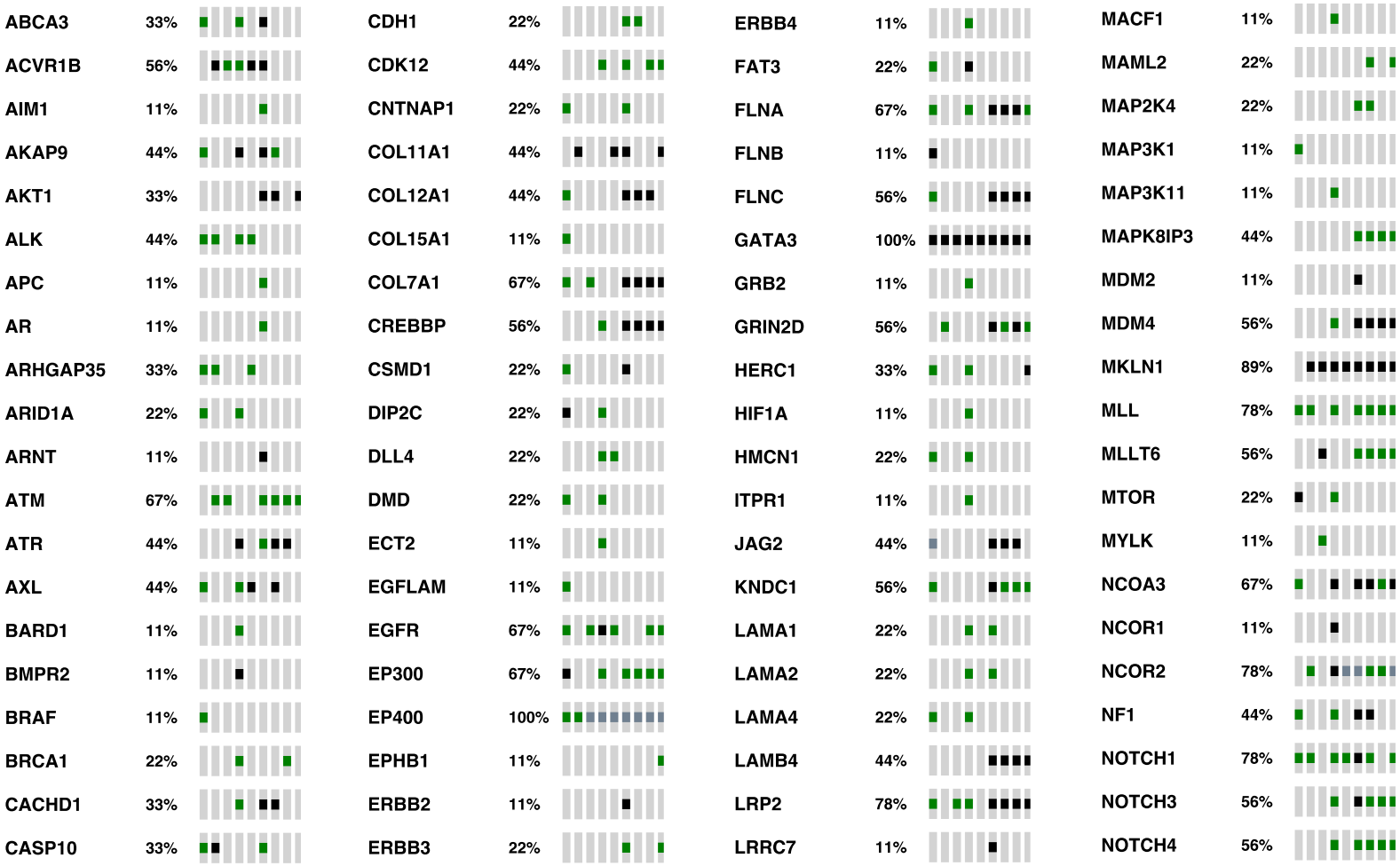
7.5 Full filtered gene list: COPE Invasive

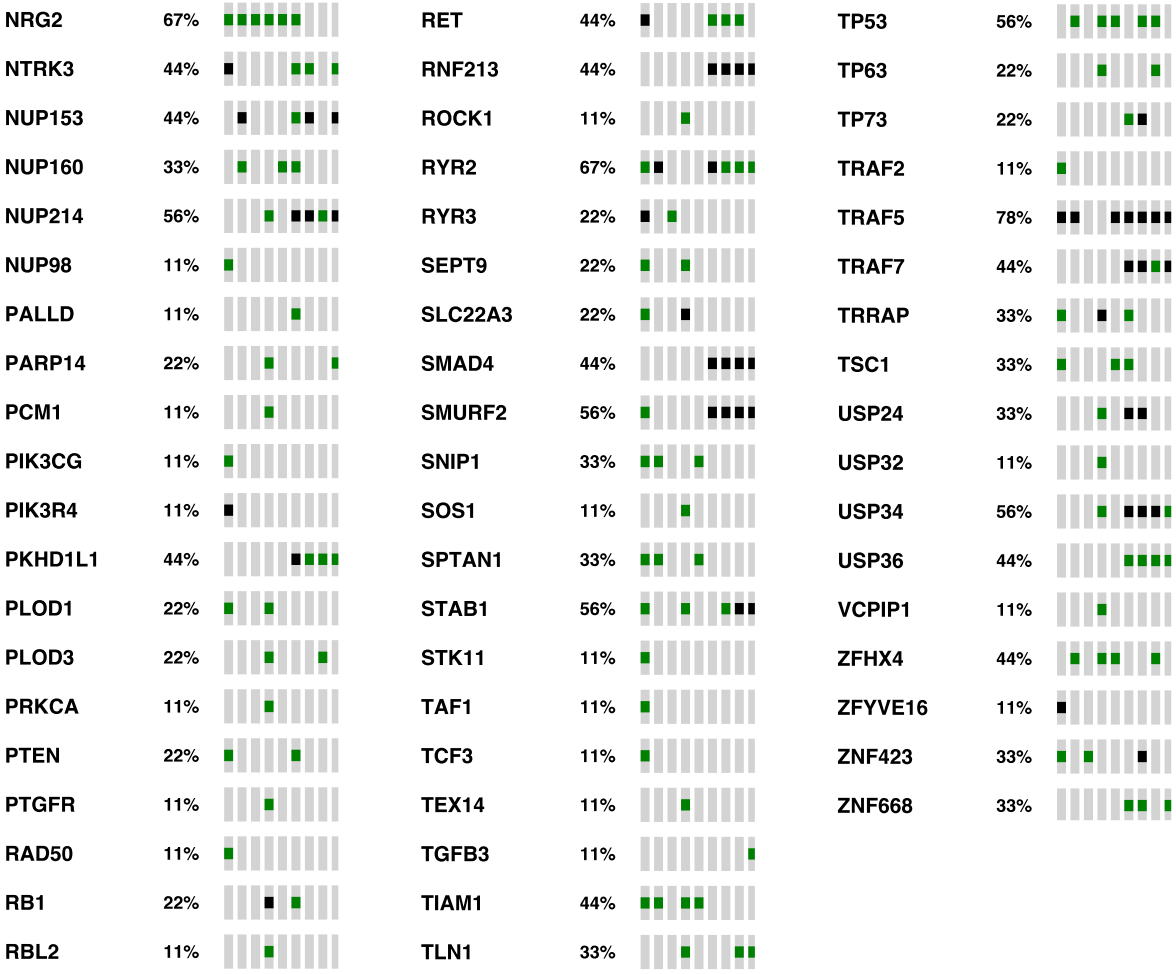



MDM2 14%
MDM4 21%
MECOM 50%
MKLN1 100%
MLH1 7%
MLL 64%
MLLT6 36%
MRE11A 7%
MSH2 14%
MTOR 7%
MYLK 7%
NCOA3 29%
NCOR1 50%
NCOR2 79%
NF1 50%
NOTCH1 93%
NOTCH3 79%
NOTCH4 79%
NRG2 79%
NTRK3 43%
NUP153 57%
NUP160 79%
NUP214 36%
NUP98 14%
P4HA1 14%
P4HA2 7%
PALB2 7%
PALLD 14%
PARP14 7%
PCM1 7%
PIK3CB 7%
PIK3CG 7%
PIK3R4 21%
PKHD1L1 43%
PLOD1 29%
PLOD2 7%
PLOD3 14%
POSTN 21%
PPM1L 7%
PRKCA 29%
PTEN 7%
PTGFR 7%
RAD50 7%
RASGRF2 21%
RB1 7%
RBL1 14%
RBL2 14%
RET 21%
RNF213 43%
ROCK1 7%
ROR1 29%
ROR2 21%
RPS6KA3 14%
RYR2 50%
RYR3 50%
SCN3A 29%
SEPT9 7%
SH3RF2 7%
SLC22A3 7%
SMAD4 29%
SMC6 29%
SMURF2 21%
SNIP1 57%
SOS1 7%
SPTA1 21%
SPTAN1 50%
STAB1 43%
STK11 36%
TAF1 29%
TCF3 21%
TEX14 14%
TGFB3 14%
TIAM1 71%
TLN1 29%
TLN2 57%
TP53 64%
TP63 21%
TP73 29%
TRAF2 7%
TRAF5 71%



7.6 Full filtered gene list: POSH HER2+ invasive


ABCA3 33%
ACVR1B 56%
AIM1 11%
AKAP9 44%
AKT1 33%
ALK 44%
APC 11%
AR 11%
ARHGAP35 33%
ARID1A 22%
ARNT 11%
ATM 67%
ATR 44%
AXL 44%
BARD1 11%
BMPR2 11%
BRAF 11%
BRCA1 22%
CACHD1 33%
CASP10 33%
CDH1 22%
CDK12 44%
CNTNAP1 22%
COL11A1 44%
COL12A1 44%
COL15A1 11%
COL7A1 67%
CREBBP 56%
CSMD1 22%
DIP2C 22%
DLL4 22%
DMD 22%
ECT2 11%
EGFLAM 11%
EGFR 67%
EP300 67%
EP400 100%
EPHB1 11%
ERBB2 11%
ERBB3 22%
ERBB4 11%
FAT3 22%
FLNA 67%
FLNB 11%
FLNC 56%
GATA3 100%
GRB2 11%
GRIN2D 56%
HERC1 33%
HIF1A 11%
HMCN1 22%
ITPR1 11%
JAG2 44%
KNDC1 56%
LAMA1 22%
LAMA2 22%
LAMA4 22%
LAMB4 44%
LRP2 78%
LRRC7 11%
MACF1 11%
MAML2 22%
MAP2K4 22%
MAP3K1 11%
MAP3K11 11%
MAPK8IP3 44%
MDM2 11%
MDM4 56%
MKLN1 89%
MLL 78%
MLLT6 56%
MTOR 22%
MYLK 11%
NCOA3 67%
NCOR1 11%
NCOR2 78%
NF1 44%
NOTCH1 78%
NOTCH3 56%
NOTCH4 56%


NRG2 67%
NTRK3 44%
NUP153 44%
NUP160 33%
NUP214 56%
NUP98 11%
PALLD 11%
PARP14 22%
PCM1 11%
PIK3CG 11%
PIK3R4 11%
PKHD1L1 44%
PLOD1 22%
PLOD3 22%
PRKCA 11%
PTEN 22%
PTGFR 11%
RAD50 11%
RB1 22%
RBL2 11%
RET 44%
RNF213 44%
ROCK1 11%
RYR2 67%
RYR3 22%
SEPT9 22%
SLC22A3 22%
SMAD4 44%
SMURF2 56%
SNIP1 33%
SOS1 11%
SPTAN1 33%
STAB1 56%
STK11 11%
TAF1 11%
TCF3 11%
TEX14 11%
TGFB3 11%
TIAM1 44%
TLN1 33%
TP53 56%
TP63 22%
TP73 22%
TRAF2 11%
TRAF5 78%
TRAF7 44%
TRRAP 33%
TSC1 33%
USP24 33%
USP32 11%
USP34 56%
USP36 44%
VCPIP1 11%
ZFHX4 44%
ZFYVE16 11%
ZNF423 33%
ZNF668 33%

Bibliography

1. https://www.breastcancercare.org.uk/news/media-centre/facts-statistics. 2013.
2. Maxmen, A., *The hard facts.* Nature, 2012. **485**(7400): p. S50-S51.
3. http://www.ons.gov.uk/ons/rel/vsob1/cancer-statistics-registrations--england--series-mb1-/index.html, 2013.
4. http://www.qub.ac.uk/research-centres/nicr/CancerData/OnlineStatistics/, 2013.
5. http://www.wcisu.wales.nhs.uk/home, 2013.
6. http://www.isdscotland.org/Health-Topics/Cancer/Publications/index.asp - 605, 2013.
7. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/incidence/ - source1, 2014.
8. http://www.cancerresearchuk.org/about-cancer/type/breast-cancer/treatment/statistics-and-outlook-for-breast-cancer, 2014.
9. Goldhirsch, A., et al., *Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011.* Ann Oncol, 2011. **22**(8): p. 1736-47.
10. http://www.cancerresearchuk.org/about-cancer/type/breast-cancer/treatment/which-treatment-for-breast-cancer - hormchem, 2014.
11. Vargo-Gogola, T. and J.M. Rosen, *Modelling breast cancer: one size does not fit all.* Nat Rev Cancer, 2007. **7**(9): p. 659-72.
12. Groen, E.J., et al., *Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS).* Breast, 2017. **31**: p. 274-283.
13. Hughes, L.L., et al., *Local excision alone without irradiation for ductal carcinoma in situ of the breast: a trial of the Eastern Cooperative Oncology Group.* J Clin Oncol, 2009. **27**(32): p. 5319-24.
14. Li, F.P. and J.F. Fraumeni, Jr., *Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome.* J Natl Cancer Inst, 1969. **43**(6): p. 1365-73.
15. Li, F.P., et al., *A cancer family syndrome in twenty-four kindreds.* Cancer Res, 1988. **48**(18): p. 5358-62.
16. Lynch, H.T., et al., *Genetic and pathologic findings in a kindred with hereditary sarcoma, breast cancer, brain tumors, leukemia, lung, laryngeal, and adrenal cortical carcinoma.* Cancer, 1978. **41**(5): p. 2055-64.
17. Strong, L.C., M. Stine, and T.L. Norsted, *Cancer in survivors of childhood soft tissue sarcoma and their relatives.* J Natl Cancer Inst, 1987. **79**(6): p. 1213-20.
18. Olivier, M., et al., *Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype.* Cancer Res, 2003. **63**(20): p. 6643-50.

19. Gonzalez, K.D., et al., *Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations.* J Clin Oncol, 2009. **27**(8): p. 1250-6.
20. Nichols, K.E., et al., *Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(2): p. 83-7.
21. Birch, J.M., et al., *Prevalence and diversity of constitutional mutations in the p53 gene among 21 Li-Fraumeni families.* Cancer Res, 1994. **54**(5): p. 1298-304.
22. Eeles, R.A., *Germline mutations in the TP53 gene.* Cancer Surv, 1995. **25**: p. 101-24.
23. Birch, J.M., et al., *Excess risk of breast cancer in the mothers of children with soft tissue sarcomas.* Br J Cancer, 1984. **49**(3): p. 325-31.
24. Brugieres, L., et al., *Screening for germ line p53 mutations in children with malignant tumors and a family history of cancer.* Cancer Res, 1993. **53**(3): p. 452-5.
25. Chompret, A., et al., *Sensitivity and predictive value of criteria for p53 germline mutation screening.* J Med Genet, 2001. **38**(1): p. 43-7.
26. Bougeard, G., et al., *Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families.* J Med Genet, 2008. **45**(8): p. 535-8.
27. Tinat, J., et al., *2009 version of the Chompret criteria for Li Fraumeni syndrome.* J Clin Oncol, 2009. **27**(26): p. e108-9; author reply e110.
28. Kamihara, J., H.Q. Rana, and J.E. Garber, *Germline TP53 mutations and the changing landscape of Li-Fraumeni syndrome.* Hum Mutat, 2014. **35**(6): p. 654-62.
29. Malkin, D., et al., *Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms.* Science, 1990. **250**(4985): p. 1233-8.
30. Ruijs, M.W., et al., *TP53 germline mutation testing in 180 families suspected of Li-Fraumeni syndrome: mutation detection rate and relative frequency of cancers in different familial phenotypes.* J Med Genet, 2010. **47**(6): p. 421-8.
31. Chompret, A., et al., *P53 germline mutations in childhood cancers and cancer risk for carrier individuals.* Br J Cancer, 2000. **82**(12): p. 1932-7.
32. Frebourg, T., et al., *Germ-line p53 mutations in 15 families with Li-Fraumeni syndrome.* Am J Hum Genet, 1995. **56**(3): p. 608-15.
33. MacGeoch, C., et al., *Heterogeneity in Li-Fraumeni families: p53 mutation analysis and immunohistochemical staining.* J Med Genet, 1995. **32**(3): p. 186-90.
34. Varley, J.M., et al., *Germ-line mutations of TP53 in Li-Fraumeni families: an extended study of 39 families.* Cancer Res, 1997. **57**(15): p. 3245-52.
35. Kleihues, P., et al., *Tumors associated with p53 germline mutations: a synopsis of 91 families.* Am J Pathol, 1997. **150**(1): p. 1-13.
36. Gonzalez, K.D., et al., *High frequency of de novo mutations in Li-Fraumeni syndrome.* J Med Genet, 2009. **46**(10): p. 689-93.
37. Wu, C.C., et al., *Joint effects of germ-line p53 mutation and sex on cancer risk in Li-Fraumeni syndrome.* Cancer Res, 2006. **66**(16): p. 8287-92.

38. Hwang, S.J., et al., *Germline p53 mutations in a cohort with childhood sarcoma: sex differences in cancer risk.* Am J Hum Genet, 2003. **72**(4): p. 975-83.
39. Malkin, D., *Predictive genetic testing for childhood cancer: taking the road less traveled by.* J Pediatr Hematol Oncol, 2004. **26**(9): p. 546-8.
40. Bond, G.L., et al., *A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans.* Cell, 2004. **119**(5): p. 591-602.
41. Bond, G.L., W. Hu, and A. Levine, *A single nucleotide polymorphism in the MDM2 gene: from a molecular and cellular explanation to clinical effect.* Cancer Res, 2005. **65**(13): p. 5481-4.
42. Bougeard, G., et al., *Impact of the MDM2 SNP309 and p53 Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome.* J Med Genet, 2006. **43**(6): p. 531-3.
43. Dumont, P., et al., *The codon 72 polymorphic variants of p53 have markedly different apoptotic potential.* Nat Genet, 2003. **33**(3): p. 357-65.
44. Ruijs, M.W., et al., *The single-nucleotide polymorphism 309 in the MDM2 gene contributes to the Li-Fraumeni syndrome and related phenotypes.* Eur J Hum Genet, 2007. **15**(1): p. 110-4.
45. Bond, G.L., et al., *MDM2 SNP309 accelerates tumor formation in a gender-specific and hormone-dependent manner.* Cancer Res, 2006. **66**(10): p. 5104-10.
46. Bueso-Ramos, C.E., et al., *Abnormal expression of MDM-2 in breast carcinomas.* Breast Cancer Res Treat, 1996. **37**(2): p. 179-88.
47. Gudas, J.M., et al., *Differential expression of multiple MDM2 messenger RNAs and proteins in normal and tumorigenic breast epithelial cells.* Clin Cancer Res, 1995. **1**(1): p. 71-80.
48. Marchetti, A., et al., *mdm2 gene alterations and mdm2 protein expression in breast carcinomas.* J Pathol, 1995. **175**(1): p. 31-8.
49. Okumura, N., et al., *Distinct promoter usage of mdm2 gene in human breast cancer.* Oncol Rep, 2002. **9**(3): p. 557-63.
50. Sheikh, M.S., et al., *The p53-binding protein MDM2 gene is differentially expressed in human breast carcinoma.* Cancer Res, 1993. **53**(14): p. 3226-8.
51. Kinyamu, H.K. and T.K. Archer, *Estrogen receptor-dependent proteasomal degradation of the glucocorticoid receptor is coupled to an increase in mdm2 protein expression.* Mol Cell Biol, 2003. **23**(16): p. 5867-81.
52. Khan, S., et al., *Estrogen receptor/Sp1 complexes are required for induction of cad gene expression by 17beta-estradiol in breast cancer cells.* Endocrinology, 2003. **144**(6): p. 2325-35.

53. Petz, L.N., et al., *Differential regulation of the human progesterone receptor gene through an estrogen response element half site and Sp1 sites.* J Steroid Biochem Mol Biol, 2004. **88**(2): p. 113-22.
54. Shlien, A., et al., *Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome.* Proc Natl Acad Sci U S A, 2008. **105**(32): p. 11264-9.
55. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.
56. Ribeiro, R.C., et al., *An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma.* Proc Natl Acad Sci U S A, 2001. **98**(16): p. 9330-5.
57. Pinto, C., et al., *TP53 germline mutations in Portugal and genetic modifiers of age at cancer onset.* Fam Cancer, 2009. **8**(4): p. 383-90.
58. Figueiredo, B.C., et al., *Penetrance of adrenocortical tumours associated with the germline TP53 R337H mutation.* J Med Genet, 2006. **43**(1): p. 91-6.
59. Garritano, S., et al., *Detailed haplotype analysis at the TP53 locus in p.R337H mutation carriers in the population of Southern Brazil: evidence for a founder effect.* Hum Mutat, 2010. **31**(2): p. 143-50.
60. Pinto, E.M., et al., *Founder effect for the highly prevalent R337H mutation of tumor suppressor p53 in Brazilian patients with adrenocortical tumors.* Arq Bras Endocrinol Metabol, 2004. **48**(5): p. 647-50.
61. Palmero, E.I., et al., *Detection of R337H, a germline TP53 mutation predisposing to multiple cancers, in asymptomatic women participating in a breast cancer screening program in Southern Brazil.* Cancer Lett, 2008. **261**(1): p. 21-5.
62. Achatz, M.I., et al., *The TP53 mutation, R337H, is associated with Li-Fraumeni and Li-Fraumeni-like syndromes in Brazilian families.* Cancer Lett, 2007. **245**(1-2): p. 96-102.
63. Custodio, G., et al., *Impact of neonatal screening and surveillance for the TP53 R337H mutation on early detection of childhood adrenocortical tumors.* J Clin Oncol, 2013. **31**(20): p. 2619-26.
64. Petitjean, A., et al., *Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database.* Hum Mutat, 2007. **28**(6): p. 622-9.
65. DiGiammarino, E.L., et al., *A novel mechanism of tumorigenesis involving pH-dependent destabilization of a mutant p53 tetramer.* Nat Struct Biol, 2002. **9**(1): p. 12-6.
66. Mesiano, S. and R.B. Jaffe, *Developmental and functional biology of the primate fetal adrenal cortex.* Endocr Rev, 1997. **18**(3): p. 378-403.
67. Jaffe, R.B., et al., *The regulation and role of fetal adrenal development in human pregnancy.* Endocr Res, 1998. **24**(3-4): p. 919-26.

68. Spencer, S.J., et al., *Proliferation and apoptosis in the human adrenal cortex during the fetal and perinatal periods: implications for growth and remodeling.* J Clin Endocrinol Metab, 1999. **84**(3): p. 1110-5.
69. Lane, D.P., *Cancer. p53, guardian of the genome.* Nature, 1992. **358**(6381): p. 15-6.
70. Silwal-Pandit, L., et al., *TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance.* Clin Cancer Res, 2014. **20**(13): p. 3569-80.
71. Toledo, F. and G.M. Wahl, *Regulating the p53 pathway: in vitro hypotheses, in vivo veritas.* Nat Rev Cancer, 2006. **6**(12): p. 909-23.
72. Hollstein, M., et al., *p53 mutations in human cancers.* Science, 1991. **253**(5015): p. 49-53.
73. Bieging, K.T., S.S. Mello, and L.D. Attardi, *Unravelling mechanisms of p53-mediated tumour suppression.* Nat Rev Cancer, 2014. **14**(5): p. 359-70.
74. Manoukian, S., et al., *Two new CHEK2 germ-line variants detected in breast cancer/sarcoma families negative for BRCA1, BRCA2, and TP53 gene mutations.* Breast Cancer Res Treat, 2011. **130**(1): p. 207-15.
75. Sodha, N., et al., *Increasing evidence that germline mutations in CHEK2 do not cause Li-Fraumeni syndrome.* Hum Mutat, 2002. **20**(6): p. 460-2.
76. Evans, D.G., J.M. Birch, and S.A. Narod, *Is CHEK2 a cause of the Li-Fraumeni syndrome?* J Med Genet, 2008. **45**(1): p. 63-4.
77. Kussie, P.H., et al., *Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain.* Science, 1996. **274**(5289): p. 948-53.
78. Meek, D.W. and C.W. Anderson, *Posttranslational modification of p53: cooperative integrators of function.* Cold Spring Harb Perspect Biol, 2009. **1**(6): p. a000950.
79. Wilson, J.R., et al., *A novel HER2-positive breast cancer phenotype arising from germline TP53 mutations.* J Med Genet, 2010. **47**(11): p. 771-4.
80. Eccles, D., et al., *Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol.* BMC Cancer, 2007. **7**: p. 160.
81. Copson, E., et al., *Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study.* J Natl Cancer Inst, 2013. **105**(13): p. 978-88.
82. Shaw, E.C., et al., *Observer agreement comparing the use of virtual slides with glass slides in the pathology review component of the POSH breast cancer cohort study.* J Clin Pathol, 2012. **65**(5): p. 403-8.
83. Eccles, D.M., et al., *Genetic testing in a cohort of young patients with HER2-amplified breast cancer.* Ann Oncol, 2016. **27**(3): p. 467-73.
84. Copson, E., et al., *Ethnicity and outcome of young breast cancer patients in the United Kingdom: the POSH study.* Br J Cancer, 2014. **110**(1): p. 230-41.
85. Copson, E.R., et al., *Obesity and the outcome of young breast cancer patients in the UK: the POSH study.* Ann Oncol, 2015. **26**(1): p. 101-12.

86. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
87. Nadji, M., et al., *Immunohistochemistry of estrogen and progesterone receptors reconsidered: experience with 5,993 breast cancers.* Am J Clin Pathol, 2005. **123**(1): p. 21-7.
88. Slamon, D.J., et al., *Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.* Science, 1989. **244**(4905): p. 707-12.
89. Foulkes, W.D., et al., *Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer.* J Natl Cancer Inst, 2003. **95**(19): p. 1482-5.
90. Couch, F.J., et al., *Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer.* J Clin Oncol, 2015. **33**(4): p. 304-11.
91. Fostira, F., et al., *Prevalence of BRCA1 mutations among 403 women with triple-negative breast cancer: implications for genetic screening selection criteria: a Hellenic Cooperative Oncology Group Study.* Breast Cancer Res Treat, 2012. **134**(1): p. 353-62.
92. *Familial Breast Cancer: Classification and Care of People at Risk of Familial Breast Cancer and Management of Breast Cancer and Related Risks in People with a Family History of Breast Cancer*. 2013, Cardiff UK: National Collaborating Centre for Cancer.
93. http://publications.cancerresearchuk.org/cancerstats/statsbreast/kfbreast.html.
94. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2.* Nature, 1995. **378**(6559): p. 789-92.
95. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**(5182): p. 66-71.
96. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21.* Science, 1990. **250**(4988): p. 1684-9.
97. Wooster, R., et al., *Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13.* Science, 1994. **265**(5181): p. 2088-90.
98. Sims, A.H., et al., *Origins of breast cancer subtypes and therapeutic implications.* Nat Clin Pract Oncol, 2007. **4**(9): p. 516-25.
99. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
100. Curtis, C., et al., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* Nature, 2012. **486**(7403): p. 346-52.
101. Leary, R.J., et al., *Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers.* Proc Natl Acad Sci U S A, 2008. **105**(42): p. 16224-9.

102. Bignell, G.R., et al., *Signatures of mutation and selection in the cancer genome.* Nature, 2010. **463**(7283): p. 893-8.
103. Lukusa, T. and J.P. Fryns, *Human chromosome fragility.* Biochim Biophys Acta, 2008. **1779**(1): p. 3-16.
104. *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
105. Ciriello, G., et al., *Mutual exclusivity analysis identifies oncogenic network modules.* Genome Res, 2012. **22**(2): p. 398-406.
106. Hanel, W. and U.M. Moll, *Links between mutant p53 and genomic instability.* J Cell Biochem, 2012. **113**(2): p. 433-9.
107. Stephens, P.J., et al., *The landscape of cancer genes and mutational processes in breast cancer.* Nature, 2012. **486**(7403): p. 400-4.
108. Garraway, L.A. and E.S. Lander, *Lessons from the cancer genome.* Cell, 2013. **153**(1): p. 17-37.
109. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.
110. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. **481**(7381): p. 306-13.
111. Anderson, K., et al., *Genetic variegation of clonal architecture and propagating cells in leukaemia.* Nature, 2011. **469**(7330): p. 356-61.
112. Ortmann, C.A., et al., *Effect of mutation order on myeloproliferative neoplasms.* N Engl J Med, 2015. **372**(7): p. 601-12.
113. Yates, L.R., et al., *Subclonal diversification of primary breast cancer revealed by multiregion sequencing.* Nat Med, 2015. **21**(7): p. 751-9.
114. Cowell, C.F., et al., *Progression from ductal carcinoma in situ to invasive breast cancer: revisited.* Mol Oncol, 2013. **7**(5): p. 859-69.
115. Ma, X.J., et al., *Gene expression profiles of human breast cancer progression.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5974-9.
116. Robanus-Maandag, E.C., et al., *Association of C-MYC amplification with progression from the in situ to the invasive stage in C-MYC-amplified breast carcinomas.* J Pathol, 2003. **201**(1): p. 75-82.
117. Jang, M., et al., *FGFR1 is amplified during the progression of in situ to invasive breast carcinoma.* Breast Cancer Res, 2012. **14**(4): p. R115.
118. Burkhardt, L., et al., *Gene amplification in ductal carcinoma in situ of the breast.* Breast Cancer Res Treat, 2010. **123**(3): p. 757-65.
119. Yao, J., et al., *Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression.* Cancer Res, 2006. **66**(8): p. 4065-78.

120. Gao, Y., et al., *Genetic changes at specific stages of breast cancer progression detected by comparative genomic hybridization.* J Mol Med (Berl), 2009. **87**(2): p. 145-52.
121. Liao, S., et al., *Differential copy number aberrations in novel candidate genes associated with progression from in situ to invasive ductal carcinoma of the breast.* Genes Chromosomes Cancer, 2012. **51**(12): p. 1067-78.
122. Porter, D., et al., *Molecular markers in ductal carcinoma in situ of the breast.* Mol Cancer Res, 2003. **1**(5): p. 362-75.
123. Johnson, C.E., et al., *Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma.* Breast Cancer Res Treat, 2012. **133**(3): p. 889-98.
124. Lee, S., et al., *Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer.* Cancer Res, 2012. **72**(17): p. 4574-86.
125. Moelans, C.B., et al., *Molecular differences between ductal carcinoma in situ and adjacent invasive breast carcinoma: a multiplex ligation-dependent probe amplification study.* Cell Oncol (Dordr), 2011. **34**(5): p. 475-82.
126. Latta, E.K., et al., *The role of HER2/neu overexpression/amplification in the progression of ductal carcinoma in situ to invasive carcinoma of the breast.* Mod Pathol, 2002. **15**(12): p. 1318-25.
127. Park, K., et al., *HER2 status in pure ductal carcinoma in situ and in the intraductal and invasive components of invasive ductal carcinoma determined by fluorescence in situ hybridization and immunohistochemistry.* Histopathology, 2006. **48**(6): p. 702-7.
128. Tamimi, R.M., et al., *Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer.* Breast Cancer Res, 2008. **10**(4): p. R67.
129. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-92.
130. Turner, N.C. and J.S. Reis-Filho, *Genetic heterogeneity and cancer drug resistance.* Lancet Oncol, 2012. **13**(4): p. e178-85.
131. Yap, T.A., et al., *Intratumor heterogeneity: seeing the wood for the trees.* Sci Transl Med, 2012. **4**(127): p. 127ps10.
132. Hernandez, L., et al., *Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection.* J Pathol, 2012. **227**(1): p. 42-52.
133. Heselmeyer-Haddad, K., et al., *Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of MYC during progression.* Am J Pathol, 2012. **181**(5): p. 1807-22.
134. Fisher, R., et al., *Development of synchronous VHL syndrome tumors reveals contingencies and constraints to tumor evolution.* Genome Biol, 2014. **15**(8): p. 433.

135. Salgado, R., et al., *Tumor-Infiltrating Lymphocytes and Associations With Pathological Complete Response and Event-Free Survival in HER2-Positive Early-Stage Breast Cancer Treated With Lapatinib and Trastuzumab: A Secondary Analysis of the NeoALTTO Trial.* JAMA Oncol, 2015. **1**(4): p. 448-54.
136. Dieci, M.V., et al., *Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials.* Ann Oncol, 2015. **26**(8): p. 1698-704.
137. Hida, A.I., et al., *Prognostic and predictive impacts of tumor-infiltrating lymphocytes differ between Triple-negative and HER2-positive breast cancers treated with standard systemic therapies.* Breast Cancer Res Treat, 2016. **158**(1): p. 1-9.
138. Schalper, K.A., et al., *Objective measurement and clinical significance of TILs in non-small cell lung cancer.* J Natl Cancer Inst, 2015. **107**(3).
139. Liu, S., et al., *CD8+ lymphocyte infiltration is an independent favorable prognostic indicator in basal-like breast cancer.* Breast Cancer Res, 2012. **14**(2): p. R48.
140. Diamantopoulos, P. and H. Gogas, *Melanoma immunotherapy dominates the field.* Ann Transl Med, 2016. **4**(14): p. 269.
141. Underwood, T.J., et al., *Cancer-associated fibroblasts predict poor outcome and promote periostin-dependent invasion in oesophageal adenocarcinoma.* J Pathol, 2015. **235**(3): p. 466-77.
142. De Monte, L., et al., *Intratumor T helper type 2 cell infiltrate correlates with cancer-associated fibroblast thymic stromal lymphopoietin production and reduced survival in pancreatic cancer.* J Exp Med, 2011. **208**(3): p. 469-78.
143. Marsh, D., et al., *Stromal features are predictive of disease mortality in oral cancer patients.* J Pathol, 2011. **223**(4): p. 470-81.
144. Surowiak, P., et al., *Occurence of stromal myofibroblasts in the invasive ductal breast cancer tissue is an unfavourable prognostic factor.* Anticancer Res, 2007. **27**(4c): p. 2917-24.
145. Tsujino, T., et al., *Stromal myofibroblasts predict disease recurrence for colorectal cancer.* Clin Cancer Res, 2007. **13**(7): p. 2082-90.
146. Massarelli, G., et al., *Myofibroblasts in the epithelial-stromal junction of basal cell carcinoma.* Appl Pathol, 1983. **1**(1): p. 25-30.
147. Lewis, M.P., et al., *Tumour-derived TGF-beta1 modulates myofibroblast differentiation and promotes HGF/SF-dependent invasion of squamous carcinoma cells.* Br J Cancer, 2004. **90**(4): p. 822-32.
148. Vong, S. and R. Kalluri, *The role of stromal myofibroblast and extracellular matrix in tumor angiogenesis.* Genes Cancer, 2011. **2**(12): p. 1139-45.
149. Inman, G.J., *Switching TGFbeta from a tumor suppressor to a tumor promoter.* Curr Opin Genet Dev, 2011. **21**(1): p. 93-9.

150. Yang, L. and H.L. Moses, *Transforming growth factor beta: tumor suppressor or promoter? Are host immune cells the answer?* Cancer Res, 2008. **68**(22): p. 9107-11.
151. Roberts, A.B. and L.M. Wakefield, *The two faces of transforming growth factor beta in carcinogenesis.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8621-3.
152. Khan, Z. and J.F. Marshall, *The role of integrins in TGFbeta activation in the tumour stroma.* Cell Tissue Res, 2016. **365**(3): p. 657-73.
153. Robertson, I.B., et al., *Latent TGF-beta-binding proteins.* Matrix Biol, 2015. **47**: p. 44-53.
154. Wipff, P.J. and B. Hinz, *Integrins and the activation of latent transforming growth factor beta1 - an intimate relationship.* Eur J Cell Biol, 2008. **87**(8-9): p. 601-15.
155. Pickup, M., S. Novitskiy, and H.L. Moses, *The roles of TGFbeta in the tumour microenvironment.* Nat Rev Cancer, 2013. **13**(11): p. 788-99.
156. Derynck, R. and Y.E. Zhang, *Smad-dependent and Smad-independent pathways in TGF-beta family signalling.* Nature, 2003. **425**(6958): p. 577-84.
157. Feng, X.H. and R. Derynck, *Specificity and versatility in tgf-beta signaling through Smads.* Annu Rev Cell Dev Biol, 2005. **21**: p. 659-93.
158. Schmierer, B. and C.S. Hill, *Kinetic analysis of Smad nucleocytoplasmic shuttling reveals a mechanism for transforming growth factor beta-dependent nuclear accumulation of Smads.* Mol Cell Biol, 2005. **25**(22): p. 9845-58.
159. Munger, J.S., et al., *The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis.* Cell, 1999. **96**(3): p. 319-28.
160. Lygoe, K.A., et al., *AlphaV integrins play an important role in myofibroblast differentiation.* Wound Repair Regen, 2004. **12**(4): p. 461-70.
161. Serini, G. and G. Gabbiani, *Mechanisms of myofibroblast activity and phenotypic modulation.* Exp Cell Res, 1999. **250**(2): p. 273-83.
162. Desmouliere, A., et al., *Transforming growth factor-beta 1 induces alpha-smooth muscle actin expression in granulation tissue myofibroblasts and in quiescent and growing cultured fibroblasts.* J Cell Biol, 1993. **122**(1): p. 103-11.
163. Ohtani, H. and N. Sasano, *Stromal cell changes in human colorectal adenomas and carcinomas. An ultrastructural study of fibroblasts, myofibroblasts, and smooth muscle cells.* Virchows Arch A Pathol Anat Histopathol, 1983. **401**(2): p. 209-22.
164. Allen, M.D., et al., *Altered microenvironment promotes progression of preinvasive breast cancer: myoepithelial expression of alphavbeta6 integrin in DCIS identifies high-risk patients and predicts recurrence.* Clin Cancer Res, 2014. **20**(2): p. 344-57.
165. Allen, M.D., J.F. Marshall, and J.L. Jones, *alphavbeta6 Expression in myoepithelial cells: a novel marker for predicting DCIS progression with therapeutic potential.* Cancer Res, 2014. **74**(21): p. 5942-7.

166. Allegra, C.J., et al., *National Institutes of Health State-of-the-Science Conference statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22-24, 2009.* J Natl Cancer Inst, 2010. **102**(3): p. 161-9.
167. Sanders, M.E., et al., *The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up.* Cancer, 2005. **103**(12): p. 2481-4.
168. Moore, K.M., et al., *Therapeutic targeting of integrin alphavbeta6 in breast cancer.* J Natl Cancer Inst, 2014. **106**(8).
169. Ghosh, A.K., S. Bhattacharyya, and J. Varga, *The tumor suppressor p53 abrogates Smad-dependent collagen gene induction in mesenchymal cells.* J Biol Chem, 2004. **279**(46): p. 47455-63.
170. Hill, R., et al., *Selective evolution of stromal mesenchyme with p53 loss in response to epithelial tumorigenesis.* Cell, 2005. **123**(6): p. 1001-11.
171. Bar, J., et al., *Cancer cells suppress p53 in adjacent fibroblasts.* Oncogene, 2009. **28**(6): p. 933-6.
172. Guo, G., et al., *Trp53 inactivation in the tumor microenvironment promotes tumor progression by expanding the immunosuppressive lymphoid-like stromal network.* Cancer Res, 2013. **73**(6): p. 1668-75.
173. Kodama, T., et al., *Increases in p53 expression induce CTGF synthesis by mouse and human hepatocytes and result in liver fibrosis in mice.* J Clin Invest, 2011. **121**(8): p. 3343-56.
174. Samarakoon, R., et al., *Loss of tumour suppressor PTEN expression in renal injury initiates SMAD3- and p53-dependent fibrotic responses.* J Pathol, 2015. **236**(4): p. 421-32.
175. Lyons, T.R., et al., *Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2.* Nat Med, 2011. **17**(9): p. 1109-15.
176. Eid, J., et al., *Real-time DNA sequencing from single polymerase molecules.* Science, 2009. **323**(5910): p. 133-8.
177. Barker, H.E., et al., *LOXL2-mediated matrix remodeling in metastasis and mammary gland involution.* Cancer Res, 2011. **71**(5): p. 1561-72.
178. Levental, K.R., et al., *Matrix crosslinking forces tumor progression by enhancing integrin signaling.* Cell, 2009. **139**(5): p. 891-906.
179. Harvey, J.M., et al., *Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer.* J Clin Oncol, 1999. **17**(5): p. 1474-81.
180. Lawson, J., et al., *Crowdsourcing for translational research: analysis of biomarker expression using cancer microarrays.* Br J Cancer, 2017. **116**(2): p. 237-245.

181. McCarty, K.S., Jr., et al., *Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors.* Cancer Res, 1986. **46**(8 Suppl): p. 4244s-4248s.
182. Koboldt, D.C., D.E. Larson, and R.K. Wilson, *Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection.* Curr Protoc Bioinformatics, 2013. **44**: p. 15.4.1-17.
183. Fuentes Fajardo, K.V., et al., *Detecting false-positive signals in exome sequencing.* Hum Mutat, 2012. **33**(4): p. 609-13.
184. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.
185. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.
186. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.
187. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
188. Debnath, J., S.K. Muthuswamy, and J.S. Brugge, *Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures.* Methods, 2003. **30**(3): p. 256-68.
189. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
190. Pinder, S.E., et al., *A new pathological system for grading DCIS with improved prediction of local recurrence: results from the UKCCCR/ANZ DCIS trial.* Br J Cancer, 2010. **103**(1): p. 94-100.
191. Hartmann, N., et al., *Prevailing role of contact guidance in intrastromal T-cell trapping in human pancreatic cancer.* Clin Cancer Res, 2014. **20**(13): p. 3422-33.
192. Gao, R., et al., *Punctuated copy number evolution and clonal stasis in triple-negative breast cancer.* Nat Genet, 2016. **48**(10): p. 1119-30.
193. Latronico, A.C., et al., *An inherited mutation outside the highly conserved DNA-binding domain of the p53 tumor suppressor protein in children and adults with sporadic adrenocortical tumors.* J Clin Endocrinol Metab, 2001. **86**(10): p. 4970-3.
194. Forbes, S.A., et al., *COSMIC: somatic cancer genetics at high-resolution.* Nucleic Acids Res, 2017. **45**(D1): p. D777-d783.
195. Wooderchak-Donahue, W.L., et al., *A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective.* BMC Med Genomics, 2012. **5**: p. 50.

196. Benjamini, Y. and T.P. Speed, *Summarizing and correcting the GC content bias in high-throughput sequencing.* Nucleic Acids Res, 2012. **40**(10): p. e72.
197. Carrick, D.M., et al., *Robustness of Next Generation Sequencing on Older Formalin-Fixed Paraffin-Embedded Tissue.* PLoS One, 2015. **10**(7): p. e0127353.
198. Munchel, S., et al., *Targeted or whole genome sequencing of formalin fixed tissue samples: potential applications in cancer genomics.* Oncotarget, 2015. **6**(28): p. 25943-61.
199. Astolfi, A., et al., *Whole exome sequencing (WES) on formalin-fixed, paraffin-embedded (FFPE) tumor tissue in gastrointestinal stromal tumors (GIST).* BMC Genomics, 2015. **16**: p. 892.
200. Koboldt, D.C., et al., *VarScan: variant detection in massively parallel sequencing of individual and pooled samples.* Bioinformatics, 2009. **25**(17): p. 2283-5.
201. Prochazkova, K., et al., *Somatic TP53 mutation mosaicism in a patient with Li-Fraumeni syndrome.* Am J Med Genet A, 2009. **149a**(2): p. 206-11.
202. Behjati, S., et al., *A pathogenic mosaic TP53 mutation in two germ layers detected by next generation sequencing.* PLoS One, 2014. **9**(5): p. e96531.
203. Zhang, L.L., et al., *Multiregion sequencing reveals the intratumor heterogeneity of driver mutations in TP53-driven non-small cell lung cancer.* Int J Cancer, 2016.
204. Atchley, D.P., et al., *Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer.* J Clin Oncol, 2008. **26**(26): p. 4282-8.
205. Rakha, E.A., J.S. Reis-Filho, and I.O. Ellis, *Basal-like breast cancer: a critical review.* J Clin Oncol, 2008. **26**(15): p. 2568-81.
206. Lee, E., et al., *Characteristics of triple-negative breast cancer in patients with a BRCA1 mutation: results from a population-based study of young women.* J Clin Oncol, 2011. **29**(33): p. 4373-80.
207. Calon, A., D.V. Tauriello, and E. Batlle, *TGF-beta in CAF-mediated tumor growth and metastasis.* Semin Cancer Biol, 2014. **25**: p. 15-22.
208. Wang, R.F., et al., *Effects of the fibroblast activation protein on the invasion and migration of gastric cancer.* Exp Mol Pathol, 2013.
209. Valcz, G., et al., *Importance of carcinoma-associated fibroblast-derived proteins in clinical oncology.* J Clin Pathol, 2014. **67**(12): p. 1026-31.
210. Casbas-Hernandez, P., J.M. Fleming, and M.A. Troester, *Gene expression analysis of in vitro cocultures to study interactions between breast epithelium and stroma.* J Biomed Biotechnol, 2011. **2011**: p. 520987.
211. Shah, S.P., et al., *The clonal and mutational evolution spectrum of primary triple-negative breast cancers.* Nature, 2012. **486**(7403): p. 395-9.
212. Niemiec, J.A., et al., *Triple-negative, basal marker-expressing, and high-grade breast carcinomas are characterized by high lymphatic vessel density and the expression of*

podoplanin in stromal fibroblasts. Appl Immunohistochem Mol Morphol, 2014. **22**(1): p. 10-6.

213. Tchou, J., et al., *Human breast cancer associated fibroblasts exhibit subtype specific gene expression profiles.* BMC Med Genomics, 2012. **5**: p. 39.
214. Park, S.Y., H.M. Kim, and J.S. Koo, *Differential expression of cancer-associated fibroblast-related proteins according to molecular subtype and stromal histology in breast cancer.* Breast Cancer Res Treat, 2015. **149**(3): p. 727-41.
215. Hanley, C.J., et al., *A subset of myofibroblastic cancer-associated fibroblasts regulate collagen fiber elongation, which is prognostic in multiple cancers.* Oncotarget, 2016. **7**(5): p. 6159-74.
216. Bao, W. and S. Stromblad, *Integrin alphav-mediated inactivation of p53 controls a MEK1-dependent melanoma cell survival pathway in three-dimensional collagen.* J Cell Biol, 2004. **167**(4): p. 745-56.
217. Tamaki, Y., et al., *Metastasis-associated protein, S100A4 mediates cardiac fibrosis potentially through the modulation of p53 in cardiac fibroblasts.* J Mol Cell Cardiol, 2013. **57**: p. 72-81.
218. Chang, V.Y., et al., *Whole exome sequencing of pediatric gastric adenocarcinoma reveals an atypical presentation of Li-Fraumeni syndrome.* Pediatr Blood Cancer, 2013. **60**(4): p. 570-4.
219. Chao, A., et al., *Molecular characteristics of endometrial cancer coexisting with peritoneal malignant mesothelioma in Li-Fraumeni-like syndrome.* BMC Cancer, 2015. **15**: p. 8.
220. Rausch, T., et al., *Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations.* Cell, 2012. **148**(1-2): p. 59-71.
221. Bass, B.P., et al., *A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen?* Arch Pathol Lab Med, 2014. **138**(11): p. 1520-30.
222. Gilbert, M.T., et al., *The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when?* PLoS One, 2007. **2**(6): p. e537.
223. Daugaard, I., et al., *The influence of DNA degradation in formalin-fixed, paraffin-embedded (FFPE) tissue on locus-specific methylation assessment by MS-HRM.* Exp Mol Pathol, 2015. **99**(3): p. 632-40.
224. Auerbach, C., M. Moutschen-Dahmen, and J. Moutschen, *Genetic and cytogenetical effects of formaldehyde and related compounds.* Mutat Res, 1977. **39**(3-4): p. 317-61.
225. Feldman, M.Y., *Reactions of nucleic acids and nucleoproteins with formaldehyde.* Prog Nucleic Acid Res Mol Biol, 1973. **13**: p. 1-49.
226. Karlsen, F., et al., *Modifications of human and viral deoxyribonucleic acid by formaldehyde fixation.* Lab Invest, 1994. **71**(4): p. 604-11.

227. Srinivasan, M., D. Sedmak, and S. Jewell, *Effect of fixatives and tissue processing on the content and integrity of nucleic acids.* Am J Pathol, 2002. **161**(6): p. 1961-71.
228. Chung, J.Y., T. Braunschweig, and S.M. Hewitt, *Optimization of recovery of RNA from formalin-fixed, paraffin-embedded tissue.* Diagn Mol Pathol, 2006. **15**(4): p. 229-36.
229. Zhang, P., et al., *The Utilization of Formalin Fixed-Paraffin-Embedded Specimens in High Throughput Genomic Studies.* Int J Genomics, 2017. **2017**: p. 1926304.