TRANSCRIPTOMIC PROFILING AND VALIDATION OF EARLY ONSET ESTROGEN RECEPTOR-POSITIVE BREAST CANCER PATIENTS

ALAA ABDELAZIZ MAHMOUD SIDDIG

UNIVERSITI SAINS MALAYSIA

2023

TRANSCRIPTOMIC PROFILING AND VALIDATION OF EARLY ONSET ESTROGEN RECEPTOR-POSITIVE BREAST CANCER PATIENTS

by

ALAA ABDELAZIZ MAHMOUD SIDDIG

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

December 2023

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to the almighty Allah for giving me the opportunity to continue my education to this level. I would like to extend my sincere appreciation to my main supervisor Associate professor Dr. Wan Faiziah Wan Abdel Rahman, for her continuous support and guidance throughout my PhD journey. Her trust, enthusiasm, and motivation have been invaluable, and this work would not have been possible without her ongoing assistance. I am also grateful to my co-supervisors - Dr Maya Mazuwin, Associate Professor Dr Sarina Sulong, Dr Siti Norasikin Binti Mohd Nafi, and Dr Tengku Ahmad Damitri Al-Astani Tengku Din - for their support and valuable advice during my challenging times. I would like to express my thanks to all the staff and students of the Department of Pathology and Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia (USM). Throughout my PhD journey, I had the privilege of meeting exceptional researchers, and I would like to specifically acknowledge Dr Shazana Hilda Shamsuddin and Dr Ahmad Aizat Bin Abdul Aziz for their immense support in academics and research. I am deeply grateful to the Ministry of Higher Education Malaysia (MOHE) for funding this study through the Fundamental Research Grant Scheme (FRGS), with grant number FRGS/1/2019/SKK08/USM/02/15. I would like to dedicate this work to my beloved mother, Zainab Farah, and my father, Abdelaziz Mahmoud Siddig, as well as my sisters Fatima, Banan, and Dana, and my brothers Khalid and Yassir. Additionally, this work is dedicated to all those who have suffered from breast cancer.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURE

LIST OF ABBREVIATIONS

HER-2	Human epidermal growth factor receptor
ERBB2	Erythroblastic oncogene B
TNBC	Triple-negative breast cancer
LRR	Loco-Regional Recurrence
PI3K	Phosphatidylinositol-3 kinase
mTOR	Mammalian target of rapamycin
KRAS	Kirsten rat sarcoma viral oncogene
ER	Estrogen receptor
PR	Progesterone receptor
DEGs	Differentially expressed genes
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
IHC	Immunohistochemistry
TSR	Tumour stroma ratio
HR	Hazard ratio
CI	Confidence interval
LRFS	Local recurrence-free survival
OS	Overall survival
DFS	Disease free survival
BRCA1	Breast cancer gene 1
BRCA2	Breast cancer gene 2
TP53	Tumour protein 53
TCGA	The Cancer Genome Atlas Database
PAH	Polycyclic aromatic hydrocarbons

mRNA	Messenger RNA
DNA	Deoxyribonucleic acid
HRD	Homologous recombination deficiency
CNAs	Copy number alterations
aCGH	Comparative genomic hybridization arrays
JAK/STAT	Janus kinase/signal transducers and activators of transcription
PPAR	Peroxisome proliferator-activated receptors
ECM	Extracellular matrix
ERR	Extracellular signal-regulated kinase
MAPK	Mitogen-activated protein kinase
TME	Tumour microenvironment
GLYATL-1	Glycine N-acyltransferase-like protein 1
FISH	Fluorescent in situ hybridization
cDNA	Complementary DNA
WT	Whole Transcript
TAC	Transcriptome analysis console software
RMA	Robust Multi-Array Analysis
ORA	Overrepresentation Analysis
FDR	False discovery rate
BH	Benjamini Hochberg
TBE	Tris-Boric Acid- EDTA buffer
RT	Reverse Transcription
dNTPs	Deoxyribonucleotide triphosphate
CT	Cycle threshold
CP	Crossing point
RANBP3L	Ran-binding protein 3 like

RANK	Receptor activator of nuclear factor kappa-B
FFPE	Formalin-fixed paraffin-embedded
KEGG	Kyoto Encyclopedia of Gene and Genome

LIST OF APPENDICES

PEMPROFILAN DAN VALIDASI TRANSKRIPTOMIK PESAKIT KANSER PAYUDARA RESEPTOR ESTROGEN POSITIF PADA AWAL USIA

ABSTRAK

Kanser payudara awal (≤45 tahun) adalah bidang perdebatan disebabkan ciri-ciri klinikopatologi yang merugikan dan hasil klinikal yang tidak menguntungkan yang diperhatikan pada pesakit awal berbanding pesakit lewat (>45 tahun). Kajian ini membandingkan ciri-ciri klinikopatologi pesakit kanser payudara awal dan lewat yang didiagnosis di Hospital Universiti Sains Malaysia pada tempoh 2013 – 2022 (n=570). Oleh itu, profil transkriptom kanser payudara Estrogen Receptor (ER)-positif dibandingkan antara pesakit awal dan lewat menggunakan cip gen Clariom S dari Affymetrix (n=14), pengesahan dalam saiz sampel yang lebih besar dilakukan menggunakan analisis in silico dan reaksi rantai polimerase secara real-time (RT-qPCR). Plot Kaplan-Meier digunakan untuk menilai makna prognostik gen yang menunjukkan perbezaan ekspresi antara tisu kanser payudara awal dan lewat. Analisis imunohistokimia digunakan untuk menilai tahap ekspresi protein RANBP3L dan GLYATL-1; gen yang dihasilkan oleh gen yang berbeza ekspresinya. Akhirnya, distribusi nisbah stroma tumor dan limfosit infiltrasi tumor dibandingkan antara kanser payudara ER-positif awal dan lewat dalam seksyen tisu pewarnaan hematoxylin dan eosin. Hasil menunjukkan bahawa kanser payudara awal cenderung mempunyai saiz besar (>5cm) (28.3% vs 13.5%), dan lebih banyak jenis luminal B (22.8% vs 14.2%). Kejadian kanser payudara bilateral lebih tinggi pada pesakit awal berbanding pesakit lewat (9.4% vs 3.6%). Profil transkriptom kanser payudara ER-positif awal secara berbeza menghasilkan 11 gen (perubahan lipatan <3 atau >3; nilai p terubah suai

<0.05). Enam daripada sebelas gen telah mengesahkan ekspresi. Analisis pengayaan menunjukkan gen yang berbeza ekspresinya antara kanser payudara ER-positif awal dan lewat dengan nilai p tidak diselaraskan <0.01 yang terlibat dalam laluan yang berkaitan dengan pelekatan fokal, panduan akson, dan kaskad komplemen dan koagulasi (nilai p FDR <0.05). Analisis kemandirian menunjukkan bahawa kehilangan ekspresi SFXN2, GLYATL-1, RANBP3L, dan ESR-1 yang merupakan ciri profil transkriptom kanser payudara ER-positif awal secara signifikan berkaitan dengan kelangsungan hidup bebas kambuh yang lebih singkat. Analisis IHC menunjukkan bahawa ekspresi protein GLYATL-1 dan RANBP3L memperlihatkan hubungan dengan positiviti nod limfa dan gred histopatologi rendah (P<0.005 dan 0.038) masing-masing. Nisbah stroma tumor yang rendah secara signifikan berkaitan dengan pesakit kanser payudara ER-positif pada usia (50-59 tahun). Bukti yang dikumpulkan menunjukkan bahawa awal kanser payudara (≤45 tahun) menunjukkan ciri-ciri klinikopatologi yang merugikan berbanding lewat kanser payudara (>45 tahun). Profil transkriptom kanser payudara ER-positif awal secara berbeza menghasilkan gen yang terlibat dalam perkembangan kanser dan metastasis. Kanser payudara yang berkembang pada usia tua dicirikan oleh nisbah stroma tumor yang rendah yang mungkin meramalkan prognosis yang baik.

TRANSCRIPTOMIC PROFILING AND VALIDATION OF EARLY ONSET ESTROGEN RECEPTOR-POSITIVE BREAST CANCER PATIENTS

ABSTRACT

Early-onset breast cancer (≤45 years) is an area of debate due to the adverse clinicopathological features and the unfavourable clinical outcomes observed in early-onset patients compared to late-onset patients (> 45 years). The present study compared the clinicopathological features of early and late-onset breast cancer patients diagnosed at Hospital Universiti Sains Malaysia during the period 2013 – 2022 (n=570). Consequently, the transcriptomic profile of Estrogen Receptor (ER)- positive breast cancer was compared between early and late-onset patients using the Clariom S gene chip from Affymetrix (n=14), validation in larger sample size conducted using in silico analysis and real-time polymerase chain reaction (RT-qPCR). The Kaplan-Meier plot used to assess the prognostic significance of the genes that exhibit differential expression between early and late-onset breast cancer tissues. Immunohistochemical analysis utilized to assess the protein expression level of RANBP3L and GLYATL-1; genes encoded by the differentially expressed genes. Finally, the distribution of the tumour stroma ratio and the tumour infiltrating lymphocytes were compared between early and late-onset ER-positive breast cancer haematoxylin and eosin-stained tissue sections. Results showed that early-onset breast cancer tend to have large size (> 5cm) (28.3% vs 13.5%), and a higher proportion of luminal B subtype (22.8% vs 14.2%). The bilateral breast cancer incidence is higher in early than in late-onset patients (9.4% vs 3.6%). The transcriptomic profile of ER-positive early-onset breast cancer differentially expressed 11 genes (fold change < 3

or >3; adjusted p-value <0.05). Six genes out of eleven have validated expression. Enrichment analysis revealed that differentially expressed genes between early and late-onset ER-positive breast cancer with unadjusted p-value < 0.01 implicated in pathways related to focal adhesion, axon guidance and complement and coagulation cascades (FDR p-value <0.05). Survival analysis showed that loss of expression of *SFXN2, GLYATL-1, RANBP3L* and *ESR-1* which is characteristic of the transcriptomic profile of early-onset ER-positive breast cancer was significantly associated with shorter relapse-free survival. IHC analysis showed that the expression of *GLYATL-1* and *RANBP3L* proteins exhibit an association with lymph node positivity and low histopathological grade (P<0.005 and 0.038) respectively. A low tumour stroma ratio was significantly associated with ER-positive breast cancer patients with age (50-59 years). The collected evidence indicates that the early onset of breast cancer (≤45 years) exhibits adverse clinicopathological features compared to the late onset of breast cancer (>45 years). The transcriptomic profile of the early-onset ER-positive breast cancer differentially expressed genes that are implicated in cancer progression and metastasis. Breast cancer developed in old age is characterized by a low tumour stromal ratio which may predict a good prognosis.

CHAPTER 1

INTRODUCTION

1.1 What is cancer?

Cancer is defined as a multistep transformation of a normal cell to a malignant cell through acquiring several functional capabilities including; induction of proliferation signalling, escaping growth suppressors regulating mechanisms, evading cell death signalling, maintaining replicative immortality, creation of new blood vessels, promoting invasion and metastasis, control cellular metabolism program, and evading immune system (Hanahan, 2022).

1.2 Breast cancer

The global cancer burden report for the year 2020; produced by the International Agency for Research on Cancer (IARC), reported 19.3 million new cancer cases, and around 10 million cancer-related death documented during the year 2020. Of this huge number, female breast cancer was the most commonly diagnosed cancer counting around 2.3 million cases (11.7%). Furthermore, breast cancer ranks fifth as a cause of cancer-related death. Even though breast cancer has a higher incidence rate in developed countries in comparison to developing countries (55.9 vs 29.7 per 100,000, respectively), the death rate due to breast cancer is high in developing countries in comparison to developed countries (15.0 vs 12.8 per 100,000). By the year 2040, it is expected that cancer incidence will increase by 47% compared to 2020 incidence. (Sung et al., 2021).

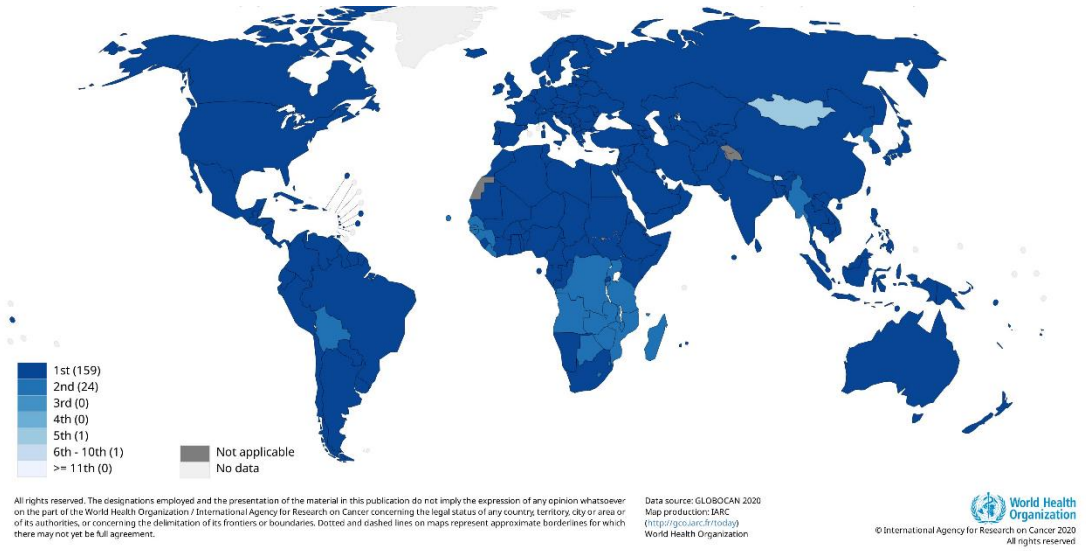

Figure 1.1 Ranking of breast cancer globally; based on the number of incidences in 2020 (World Health Organization 2020) (Sung *et al.*, 2021)

1.3 Breast cancer subtypes

Breast cancer encompasses a heterogeneous group of diseases, which are characterized by different biological natures, pathological characteristics, clinical manifestations, therapeutic responses, clinical behaviour, and outcomes (Al-thoubaity, 2020).

Based on the receptors expressed on the surface of the tumour cells, breast cancer is classified into several subtypes those receptors are either nuclear or cytoplasmic and are detected routinely by immunohistochemical staining.

Based on immunohistochemical analysis breast cancer is classified into; hormone receptor-positive, human epidermal growth factor receptor-2 (HER-2) overexpressed, and Triple-negative breast cancer. Hormones receptor-positive tumours exhibit immunoreactivity for antibodies against estrogen and progesterone nuclear receptors, while HER-2 overexpressed show positivity for antibodies directed against the membranous protein HER-2 and finally, triple-negative tumours characterized by a

lack of immunoreactivity to estrogen, progesterone and HER2 (Zaha, 2014). Immunohistochemical classification is the most commonly used in clinical practice, due to it is specificity, cost-effectiveness and simple methodology (Al-thoubaity, 2020).

On a molecular level earlier gene expression study carried out by Perou and his colleagues classified breast cancer into; estrogen-positive tumours which express genes of mammary luminal cells (Keratin 8, Keratin 18, GATA-binding protein 3, X-box binding protein 1, and hepatocyte nuclear factor 3a), estrogen negative tumours which express genes of mammary basal cells (Keratin 5, Keratin 17, Integrin-β4 and laminin), overexpression of erythroblastic oncogene B (ERBB2) tumours which express a high level of ERBB2 oncogene and low levels of estrogen, finally normal like tumours which express genes of basal epithelial cells, adipose cells and lower expression of genes of luminal epithelial cells. This gene expression-based classification was a gold discovery in breast cancer research, it tremendously improved the understanding of the molecular biology of breast cancer and the management approach was improved accordingly (Perou et al., 2000). In terms of prognosis estrogen receptor-positive tumours have favourable outcomes compared to HER-2 and Triple-negative breast cancer (TNBC) which both have unfavourable prognoses (Prat et al., 2015).

1.4 Breast Cancer Status in Malaysia

Based on the Malaysian National Cancer Registry's (MNCR) last report for the years (2012-2016) issued in June 2019. Breast cancer was the most commonly diagnosed

type of cancer; counting around 21,925 (19%) of all cancer incidences 115,238 (100%) during the 4 years.

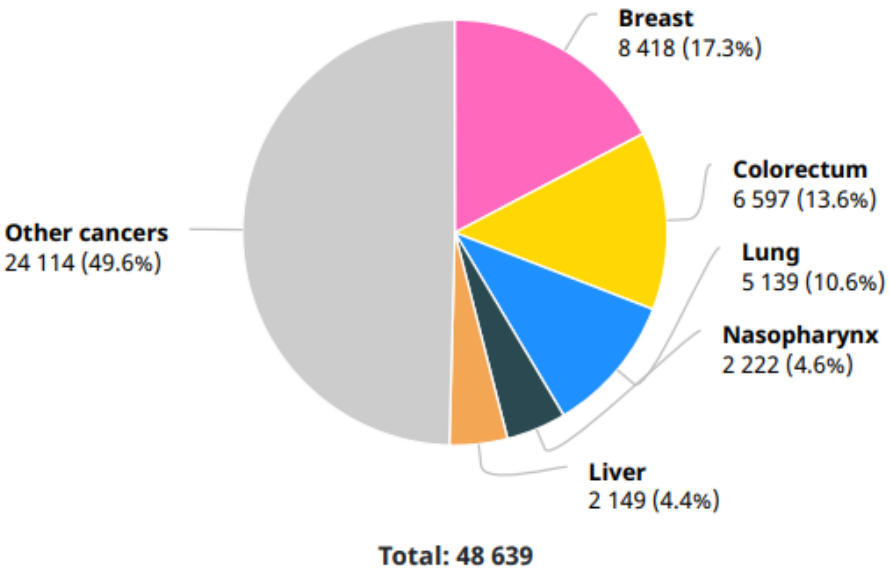

Figure 1.2 Cancer incidence in Malaysia for the year 2020. Breast cancer was the most commonly diagnosed type of cancer among the Malaysian population accounting for around 32.9% of all cancer incidences. Adopted from the GLOBOCAN cancer report for the year 2020 (Sung *et al.*, 2021).

For Malaysian women, the lifetime risk of developing breast cancer is 1 in every 27 women. Nearly half of Malaysian patients (47.9%) seek medical attention at a late stage of the disease (III & IV). According to the cancer registry report among young age Malaysian patients (25 – 59 years), breast cancer is the most encountered cancer followed by colorectal and cervical cancers, whereas among old age patients (≥ 70 years) colorectal cancer was slightly more prevalent than breast cancer (Azizah *et al.*, 2019).

Over the past 20 years, the global research community has been particularly interested in breast cancer cases affecting women under the age of 45 years. This is because those women are at an age of highest productivity in terms of family planning or career life.

Suffering from such disease at this age will adversely affect their quality of life. In addition to the emotional and psychological distress associated with dealing with this illness, there are also social implications to consider.

1.5 Definition of young age in breast cancer

Oncology literature defined early-onset breast cancer as breast cancer that developed before the age of 40 years (Poggio et al., 2018; Paluch-Shimon et al., 2020), however, detailed stratification was adopted by several groups of researchers, where they divided the young age group into two, very young which indicates breast cancer arising below the age of 35 years (Eiriz et al., 2021), and young which donates patients developed breast cancer between 40- 49 years (Azim et al., 2012). However, the general practice by breast cancer researchers is to deal with age factors as a discriminator between breast cancers developed pre-or post-menopause period (<50 years as premenopausal and >50 years indicating post-menopause patients) (Pfitzner et al., 2014).

1.6 Breast cancer at a young age at a glance

Breast cancer arises at a young age tends to be large, in advanced stage and grade, with a higher proportion of aggressive molecular subtypes (HER-2 overexpressed and basal subtypes), and a higher risk of locoregional recurrence and contralateral breast cancer, all these factors resulted in inferior prognosis compared to breast cancer that developed late in life (between 40 – 60 years) (Kim et al., 2022).

1.7 Risk factors for developing breast cancer at a young age

A substantial number of studies claimed that the distribution of risk factors of breast cancer differs among different age groups and that patients' age must be taken into consideration when assessing the effect of risk factors (Bissell *et al.*, 2020). However several recent studies failed to find the claimed heterogeneity. Hence, further investigation in this area is warranted. In this section, the previous literature that assessed the distribution of several risk factors of breast cancer among different age groups was summarized in the following sections.

1.7.1 Reproductive factors

An earlier study reported that a longer breastfeeding period and younger age at first childbirth decreases the risk of breast cancer in young age patients (<40 years) compared to older age patients (>40 years). Moreover, the study found that the number of childbirths does not appear to reduce the risk of breast cancer at a young age. In addition, the study reported that late first childbirth (after the age of 30 years) increases the risk of breast cancer more prominently in women diagnosed in age below 40 and had given birth to one child only (OR = 7.06; 95% CI = 2.16 – 23.01) (Tryggvadóttir et al., 2002). In contrast, Warner et al. failed to find heterogeneity between young (<40 years) and old breast cancer patients (>40 years) in terms of reproductive factors such as the age of menarche, parity, age at first childbirth, and breastfeeding (Warner et al., 2013). A case-control study investigates breast cancer risk factors in young breast cancer patients (younger than 50 years at the time of diagnosis) matched with healthy participants. The study concludes that the recent trend of decreasing parity and short duration of breastfeeding with the increased use of oral contraceptives may explain the increase in breast cancer incidence among young patients (Ghiasvand et al., 2011).

Another case-control study included breast cancer patients who were diagnosed at the age of 50 years or younger revealed that younger age at menarche and late age at first-term pregnancy were significantly associated with an increased risk of early-onset breast cancer (O'Brien et al., 2015). A recent study assessed the effect of early pregnancy (< 21 years) on breast cancer development at an early age of life (<50 years) in BRCA1 and BRCA2 carriers. The findings showed that early pregnancy was associated with a lower incidence of early onset breast cancer specifically in women aged between 30 – 39 years, OR 0.73 for BRCA2 ($p = 0.002$) and 0.78 for BRCA1 ($p = 0.005$) (Evans et al., 2018). From the above, we may conclude that regarding reproductive factors, the findings are controversial and the variation in the cut-off age may contribute to this heterogeneity.

1.7. 2 Obesity

In general obesity (Body mass index ≥ 30 kg/m2) increases the risk of several types of cancer. A person with a BMI over 40% has a 70% increased risk of developing cancer. In contrast to the norm, several studies reported an inverse relationship between obesity and breast cancer at a young age, where obesity appears to prevent the development of breast cancer at an early age but the exact mechanism has not yet been elucidated (Magnusson et al., 2005; O'Brien et al., 2015; Micaily et al., 2021). However, the protective effect of obesity differs among different ethnicities, In a study conducted among American women, higher BMI was associated with a lower risk of breast cancer development in premenopausal women (OR −0.46, 95%, CI: 0.26, 0.80) though, this observation was seen in Non-Hispanic White women, but not in African American women (Micaily et al., 2021).

1.7.3 Physical activity

Moderate physical activity significantly reduces the risk of breast cancer, which is explained by the reduction of ovarian steroids in the circulation that is known to promote breast cancer tumorigenesis (Spei *et al.*, 2019). Physical activity may also play a crucial role in enhancing immune system function and promoting anti-carcinogenic pathways through enhancing insulin sensitivity, lowering levels of adipokines, and inflammatory and oxidative stress markers (Friedenreich, Ryder-Burbidge and McNeil, 2021). However, in a study that included 1560 premenopausal breast cancer patients (<36 years) and 1548 controls, no relation was found between sport participation and the risk of breast cancer (Magnusson et al., 2005). In line with this finding, a meta-analysis study reviewed ten studies that assess the relationship between physical activity and the risk of breast cancer development in young women (40 – 49 years). All ten studies found no relation between physical activity and the reduction of risk of breast cancer in premenopausal women (Nelson et al., 2012).

1.7.4 Hereditary factors

Inheritance of pathogenic mutations in the high and low penetrance breast cancer susceptibility genes is an unmodifiable risk factor for breast cancer, this factor explains around 5-10% of all breast cancer cases (Siddig et al., 2021). By comparing the replication behaviour and replication stress of blood lymphocytes which is used to assess the genomic stability; Dikomey et al., found that patients who developed breast cancer at a young age (≤45 years) have significantly higher germline mutations compared to old-age breast cancer patients (>60 years) (Dikomey, 2021).

In contrast, Andrikopoulou et al. compared the genomic profile of early-onset breast cancer (<40 years) with late onset of breast cancer (>40 years) using the powerful

technology the next-generation sequencing. The study showed that the most common pathogenic germline mutations encountered in the young age group were in (CHEK2, BRCA1, BRCA2, and TP53), whereas (BRCA1, ATM and CHEK2) were the most common in old age patients. However the observed difference didn't reach the statistical significance level, and thus researchers concluded that young and old age breast cancer patients, share the same genomic profile (Andrikopoulou et al., 2022).

1.7.5 Modifiable Risk Factors

A previous study investigated the modifiable risk factors of breast cancer in light of the patient's age and found that 12% of breast cancer incidences in premenopausal women were explained by alcohol consumption (CI = 4.3–20.2%) and 7% by current use of oral contraceptives for ≥5 years (CI = 0.3–13.5%), Whereas for postmenopausal women, 12.8% was explained by overweight or obesity (CI = 7.8–17.5%), 6.9 current use of hormone replacement therapy and 6.6% due to regular alcohol consumption (CI = 4.8–8.9%) (Arriaga *et al.*, 2019).

One of the most important modifiable risk factors for breast cancer is the choice of diet, a prospective study found that significant intake of red meat during adolescence is linked to an increased likelihood of premenopausal breast cancer. Conversely, substituting red meat with alternative sources of dietary protein during adolescence appeared to lower the potential risk of premenopausal breast cancer (Taha and Eltom, 2018).

Interestingly the dietary pattern risk was found to be associated with menopausal status, the western dietary pattern was found to be related to 20% elevation in breast cancer risk among postmenopausal women, but not among premenopausal women.

On the other hand, the prudent dietary pattern was found to reduce breast cancer risk in premenopausal women by 23% however no significant reduction was observed among postmenopausal women (Xiao *et al.*, 2019).

Furthermore, previous evidence showed that diets rich in carotenoids and vitamin A result in a reduction in the risk of breast cancer mainly for premenopausal women who may have a family history that makes them predisposed to breast cancer (Forshee, Storey and Ritenbaugh, 2003).

These findings highlight a significant difference in the distribution of the modifiable risk factors of breast cancer between young and old age women, conducting meta-analyses and systematic reviews with larger sample sizes can accurately reflect the true situation and offer valuable insights into the variations in the distribution of risk factors of breast cancer.

1.8 Young age as an independent risk factor for poor clinical outcomes

Despite the enormous number of studies that aimed to elucidate the prognostic significance of young age at the time of breast cancer diagnosis, up-to-date findings are controversial on whether young age is an independent factor for poor prognosis in breast cancer or not. This section sheds light on recent studies that investigate this area.

Analyzing the clinicopathological characteristics and follow-up data of 25,898 breast cancer patients from the Japanese Breast Cancer Registry showed that after adjusting potential clinicopathological features that considered cofounders such as TNM classification, neo-adjuvant/adjuvant therapy and breast cancer subtype; young age

patients (<35 years) predict poorer disease-free survival and overall survival than middle-aged (35–50 years) and old aged patients (>50 years) (Kataoka et al., 2016). In contrast, a retrospective single institutional-based study assessed the clinicopathological features and calculated the overall and disease-free survival for 1611 breast cancer patients of which 281 (17.4%) patients were 40 years or younger, the study reported that although young age patients present with larger tumour size at time of diagnosis and frequent TNBC in comparison to old age patients, young age at diagnosis was not an independent risk factor for worse prognosis. The study showed that the overall and disease-free survival of young age patients exhibited a drop however it did not reach the statistical significance threshold (Abdulla et al., 2022).

Clinicopathological data of 243,012 breast cancer patients diagnosed during the period 1988-2003 were extracted from the Surveillance, Epidemiology, and End Results Program data (SEER), data analysis demonstrated that after controlling cofounders, young age breast cancer patients (<40 years) were more likely to die due to breast cancer in comparison with their older counterparts (≥40 years) only if they were diagnosed at an early stage of the disease (stage I and II), (adjusted HR = 1.44; CI, 1.27 to 1.64) (adjusted HR = 1.09; CI, 1.03 to 1.15) respectively (Gnerlich et al., 2009). Partridge et al. found that the poor clinical outcome of breast cancer at a young age is prominent in hormone receptor-positive cases, whereas in triple-negative and HER-2 subtypes young age doesn't appear to worsen the risk of breast cancer mortality (Partridge et al., 2013). Suggesting that age alone is not considered an independent predictor for clinical outcomes in HER-2 and TNBC (Partridge et al., 2016).

Based on the information provided above, it can be inferred that under certain conditions, young age may serve as an independent predictor.

1.9 Factors that may contribute to the poor clinical outcomes seen in early-onset breast cancer patients

Several researchers suggest that poor clinical outcomes seen in young age breast cancer patients can be related to several factors, including but not limited to non-compliance to hormonal therapy, resistance to endocrine therapy, fast restoration of ovarian function post-chemotherapy, and a distinct biology, however, the exact reason is still not clear. (Kim et al., 2022). In this section previous literature that interrogates the suggested factors was summarized and presented.

1.9.1 Dense breast tissue and reduced mammography sensitivity

The breast is composed of ductal epithelium, fibrous stroma and fat tissues. Young women usually have high-density breast tissue in comparison to old age women, as with age fat constituent increased and the fibro-glandular tissue decreased. The majority of premenopausal women have dense breasts. In mammograms dense breast tissue appears radiopaque comparable to breast cancers, reducing visual contrast. Thus, the sensitivity of mammography screening was reduced to 62% when examining dense breasts, whereas the percentage increased to 88% with fatty breasts (Chelmow et al., 2020).

1.9.2 Non-adherence to endocrine therapy

The majority of breast cancer is estrogen-receptor-positive tumours, which are treated by endocrine therapy (Lei et al., 2019). An earlier study aimed to determine oral endocrine therapy non-persistence and non-initiation among early-onset breast cancer patients (≤ 40 years) found that younger age was significantly associated with higher odds of non-persistence. Moreover, after adjusting age and other clinical features, non-

persistence was found to be significantly associated with lower odds of recurrence (Rosenberg et al., 2022).

1.9.3 Locoregional recurrence and young age at diagnosis

The definition of Loco-Regional Recurrence (LRR) is any recurrence in the ipsilateral breast, chest wall or axilla (Holleczek et al., 2019). In a recently published study young age was on top of several factors that were attributed to the locoregional recurrence of breast cancer including tumour stage, molecular subtype, surgical margins, and lack of adjuvant radiotherapy (Merino et al., 2018). A substantial number of researchers examined the relationship between age and the risk of breast cancer LRR, a recent study reported that with each 1-year decrease in patient age; there is a 7% increase in the relative risk of breast cancer loco-regional recurrence (Buchholz, Ali and Hunt, 2020; Li et al., 2021).

1.9.4 Resistance to hormonal therapy

More than 70% of all breast cancer cases are classified as estrogen receptor-positive tumours based on immunohistochemical analyses (Scabia et al., 2022), this type of tumour is treated through hormonal therapy. An earlier study aimed to determine the overall and breast cancer-specific survival for 1444 breast cancer patients aged < 35 years and 8,441 patients with age between 35 and 50 years, results showed that tamoxifen and adjuvant chemotherapy combination did not improve the survival in young age patients (< 35 years) (Ahn et al., 2007).

Goldhirsch et al. attributed the resistance to endocrine therapy seen in very young age patients to several factors including therapy duration, menopausal symptoms related

to the therapy, concern regarding sexual functioning and family planning (Goldhirsch et al., 2001).

1.9.5 The biology of breast cancer developed at a young age

There is a huge debate regarding whether breast cancer developed at a young age has distinct biology from that developed at old age. An earlier study assessed the transcriptomic profile of young (≤45 years) and old age (≥65 years) breast cancer and the results revealed that 350 gene sets show differential expression. Later, the researchers reanalysed the exact data and amended their findings, proposing that young age patients have a high proportion of the aggressive molecular subtypes of breast cancer such as HER2 overexpressed and TNBC, but the biology of both age groups appears homogenous (Carey K. Anders *et al.*, 2008; Anders *et al.*, 2011). A recent study focused on transcriptional changes in breast cancer related to age, reported that breast cancer at a young age (< 40 years) exhibited an increase in oncogenic signalling (PI3K /mTOR, MYC and KRAS), suggesting that the high estrogen level in young age patients may promote and increase the proliferation activity and decreases the apoptotic activity resulting in aggressive tumour phenotype (Ingebriktsen et al., 2022).

1.10 Problem statement

The incidence of breast cancer at a young age varies according to the geographical distribution ranging between 6–20% of all breast cancer cases (Gómez-Flores-Ramos *et al.*, 2017). Asian patients have a high proportion of pre-menopausal breast cancer compared to their Western counterparts (Lin *et al.*, 2019). Breast cancer that develops at an age below 40 years exhibits aggressive clinicopathological characteristics and

poor clinical outcomes compared to that developed late in life (Hu *et al.*, 2021). Several factors were claimed to be attributed to the aggressive behaviour of breast cancer seen in young age patients (Kim *et al.*, 2022). However, the historical hypothesis suggests that early-onset breast cancer has a unique biology that differs from late-onset breast cancer. Numerous studies assessed the biology of early-onset breast cancer at genomic, transcriptomic and epigenetic levels, furthermore, several studies also compared the distribution of risk factors and the clinicopathological characteristics between young and old age patients. However, up to date, the findings are controversial and no specific diagnostic or prognostic markers for this group of tumours.

1.11 Objective of the study

General objective

To explore the clinicopathological features, transcriptomic profile and the tumour microenvironment components of breast cancer that developed in young age women (≤45 years) in comparison to that developed in old age women (> 45 years).

Specific objectives

1- To determine the association between the clinicopathological characteristics of breast cancer and patient age at diagnosis.

2- To compare the transcriptomic profile of young and old age estrogen receptor (ER)-positive breast cancer via DNA microarray analysis.

3- To validate the expression of the differentially expressed genes (DEGs) by young age ER-positive breast cancer using real-time polymerase chain reaction (RT-qPCR).

4- To compare the protein expression level encoded by the DEGs between young and old age ER-positive breast cancer using immunohistochemical (IHC) analysis, and to associate the protein expression level with the various clinicopathological parameters.

5- To determine the distribution of the tumour stroma ratio (TSR) and the tumour infiltrating lymphocytes (TILs) level between young (≤45 years) and old age (> 45 years) ER-positive breast cancer.

CHAPTER 2

LITERATURE REVIEW

This chapter will provide a summary of existing research investigating the difference between early and late-onset breast cancer. Focusing on several aspects, including the clinicopathological features, the molecular profile (such as genomic, transcriptomic, and epigenomics), as well as the difference in the components of the tumour microenvironment.

Scientists investigating the biological aspects of breast cancer occurring at a young age may face difficulties in comprehending this specific category of tumours due to contradictory discoveries presented in previous research. Nonetheless, it is crucial to acknowledge that certain inconsistencies might be attributed to the influence of race and ethnicity. Nevertheless, earlier studies generally agree on the adverse clinical features and the unfavourable prognosis associated with breast cancer diagnosed in young individuals.

2.1 The pattern of early-onset breast cancer depends on the geographical distribution

According to a large international collaboration study, East Asian women tend to develop breast cancer at a younger age than American women, East Asian patients aged 40-49 years at the time of diagnosis have a significant likelihood of developing ER-positive breast cancer. In contrast, the probability of developing ER-positive breast cancer among American women increases with age (Lin et al., 2019). In concordance with this, an earlier study reported that TNBC was the most common molecular subtype in African-American women who were diagnosed with breast

cancer at a young age (under the age of 35 years) (Ihemelandu et al., 2007). Comparing the early onset of breast cancer among African (Cameron and Congo) and Western cohorts indicated that the mean age at breast cancer diagnosis is 46.5 ± 12.9 and 58.5 ± 13.2 years for the African and Western cohorts, respectively. In addition, the incidence of early-onset breast cancer was 51.7% among African patients versus 15.6% among Western patients (Tonouo et al., 2022).

These findings illustrate variations in incidence proportion and the distribution of the molecular subtypes of early-onset breast cancer cases across diverse populations.

2.2 Breast cancer molecular subtypes and clinical outcomes among young age patients

It is well established that in invasive breast cancer the luminal A-like tumours (ER and/or PR +, HER2 -, Ki-67 <14%) have the best clinical outcomes and that TNBC (ER-, PR-, HER-2 -) have the poorest outcome (Hennigs et al., 2016). However, in early-onset breast cancer, the scenario is different. Fu et al reported that patients diagnosed with breast cancer at the age of 40 years or younger had lower cancer-specific survival rates after 5 and 10 years of primary diagnosis compared to those diagnosed at an older age, and that age is a significant independent prognostic factor in breast cancer-specific survival ($P<0.001$). In addition, their analysis showed that young age patients with ER-positive breast cancer had twice the risk of cancer-specific death compared to older patients (Fu et al., 2018). In line with this observation, a retrospective study found that in very young age breast cancer patients ($\leq$ 35 years); luminal B and pregnancy-associated breast cancer negatively impact overall survival (Tsai et al., 2021).

Partridge and colleagues found that after controlling the sociodemographic, clinicopathological and treatment characteristics of 17,575 breast cancer patients, being diagnosed at age ≤40 years increases the risk of dying from breast cancer (HR1.4; 95% CI 1.2-1.7). The risk of breast cancer death significantly increases in those patients who were diagnosed with luminal A subtype (hazard ratio of 2.1; 95% confidence interval of 1.4 to 3.2), followed by those with luminal B and TNBC. However among HER-2 overexpressed subtype no significant relation between age and risk of breast cancer death. Concluding that young age is associated with the worst clinical outcome among luminal A tumours (Partridge et al., 2016).

Comparing the clinical and molecular features of TNBC in three different age groups: young (≤39 years), intermediate (40-64 years), and old age (≥65 years), showed that patients in the young and old age groups had a higher risk of relapse within 3 years compared to those in the intermediate age group ($P < 0.001$). However, all other clinicopathological features were distributed evenly among the different age groups (tumour size, histopathological grade and stage), except for ki-67 expression and fibrosis. The old age group had significantly lower Ki-67 expression ($P < 0.001$) and higher levels of tissue fibrosis ($P < 0.005$) (Ma et al., 2020).

To conclude the previous literature highlighted the poor clinical outcomes associated with breast cancer developed at a young age stressing the poor survival seen in ER-positive tumours.

2.3 The distribution of the clinicopathological features of breast cancer between young and old age patients

The heterogeneity between young and old age breast cancer patients is not limited to the clinical outcomes however it's also observed in the clinicopathological characteristics of the tumour at the time of presentation. Several studies reported that breast cancer at a young age exhibits adverse clinicopathological characteristics at the time of presentation when compared with breast cancer that arises in old age. A retrospective multi-centre study compared the clinicopathological characteristics of breast cancer of 295 young patients (≤34 years of age) with 2,119 old age patients (35 to 49 years), results showed that breast tumours in young age patients maintained a higher percentage of lymph node metastasis, lower expression of progesterone receptors and have advance stage than the control group ($P<0.05$) (Zhang et al., 2021). Another study with a larger sample size, classified patients into three groups; <30 years (n=564), 30 – 60 years (n=1,4519), and > 60 years (n=670), the study found that the younger age patients (<30 years old) suffers from breast cancer with higher vascular and lymphatic invasion ($p = 0.002$), HER2-expression ($p = 0.001$), ER-negative, and higher frequency of grade III tumours ($p = 0.001$). Furthermore, younger patients (<30 years) have higher rates of regional metastasis and higher rates of brain metastasis ($p < 0.001$) (Akrami et al., 2018). Two retrospective studies reported that patients who were 40 years or below at the time of diagnosis presented with advanced nodal disease (N2-N3) ($p= 0.046$), higher frequency of stage III cancers ($p= 0.011$), more metastasis events ($p<0.001$) than those who were diagnosed at age above 40 years. Multivariate analysis revealed that N2-3 disease, age below 40 years, grade 3 negatively affects overall survival (Yavuz, Aktan and Kanyilmaz, 2020).

Comparing the OS, local recurrence-free survival (LRFS) and disease-free survival (DFS) between young (≤35 years) and old (>35 years) age breast cancer patients (n=11,671). Young women had the lowest 5-year LRFS and DFS ($P < 0.001$ for all). When stratified by molecular subtype (ER, PR positive and Her-2 negative subtype) still young women appear to have the worst outcomes including LRFS, DFS, and OS ($P < 0.05$ for all) (Yang et al., 2022).

Even though the age cut-off was not well established in the above-summarized studies, however, all findings provide strong evidence that early-onset breast cancer has aggressive features compared to late-onset breast cancer.

2.4 The molecular profile of early and late-onset breast cancer

2.4.1 Germline and somatic mutations

Cancer genomic profiling refers to the study of any form of genetic alterations that can be utilized in the diagnosis, classification and prediction of outcomes of a specific type of cancer (Chakravarty and Solit, 2021). Germline mutations are mutations that result in the presence of disease, exist in an individual's inherited genetic material and are not developed during life like somatic mutations. These mutations can interrupt the reading frame of protein synthesis or have been reported previously as disease-causing variants in ClinVar (Kan et al., 2018).

Women who carry germline mutations in *BRCA1* almost certainly will develop breast cancer at an early age, but not all early-onset breast cancer patients necessarily carry this genomic aberration. Less than 10% of breast cancer that arises at a young age was attributed to *BRCA1*/*BRCA2* mutation. However, the proportion of *BRCA1*/*BRCA2* germline mutations differs between early and late-onset breast cancer, a previous study

showed that 13.7% of the young-age Asian patients (≤40 years) harboured germline pathogenic mutations in *BRCA1*/*BRCA2* genes compared to 4.8% in old-age Caucasian patients (>40 years) (Kan et al., 2018). A prospective study assessed whether *BRCA1* and *BRCA2* mutations affect the survival of young age (<40 years) breast cancer patients (n= 2733), results showed that 12% of the patients harboured *BRCA1/BRCA2* mutations. No difference was detected between those patients who were harbouring the mutations and those who were non-harbouring them (Copson et al., 2018).

A single institute retrospective study reported that 26.9% of very young breast cancer patients (≤ 35 years) diagnosed within the period of 10 years (2007 – 2017) were harbouring high-risk pathogenic variants, however, those variants did not appear to affect the overall survival or the relapse-free survival of those patients (Tsai et al., 2021).

In a study conducted among Taiwanese breast cancer patients, whole genome and exome sequencing revealed that 40% and 37% of the young age patients (<41 years) were carriers for mutations in the tumour suppressor gene *TP53* and the oncogene *PIK3CA*, respectively. Furthermore, there was a high rate of extracellular structural protein-coding gene mutations *MUC17* (19%), *TTN* (17%), and *FLG* (16%), and the higher mutation prevalence of those genes remained distinct even after comparing the data with non-Taiwanese pooled young age breast cancer patients and pooled old age cohorts (Midha et al., 2020). Mining previous literature showed that *MUC17* expression was identified by Al Amri et al., as a chemotherapy prognostic marker in breast cancer, where in vitro suppression of *MUC17* was related to increased

chemotherapy sensitivity. Moreover, there was an association between a low level of *MUC17* and prolonged survival after receiving chemotherapy (Al Amri et al., 2020).

Comparing the genomic profile of breast cancer among three age groups (≤45, 46-69, and ≥70 years); using data obtained from The Cancer Genome Atlas Database (TCGA), researchers found that when adjusting the clinicopathological characteristics of breast cancer such as tumour size, nodal status, and the tumour histopathological subtype, the highest number of somatic mutations was observed in elderly patient's ≥70 years, and several mutations show association with patient's age. However, only mutations in *GATA3* were significantly associated with the youngest age group (≤45 years) (P= 0.003; 15.2 % Vs 8.2 % Vs 9%) (Azim Jr et al., 2015).

Mealey et al., support Azim et al., findings, reporting that comparing the mutational pattern of 89 early-onset breast cancers (<40 years) with 949 late-onset breast cancers (> 40 years) revealed that early-onset breast cancer harbour a high rate of mutations in *GATA3* and *CTNNB1* whereas late-onset breast cancer shows a high rate of mutations in *PIK3CA*, *CDH1*, and *MAP3K1* genes (Mealey et al., 2020).

In contrast, Andrikopoulou et al. observed that breast cancer developed at a young age (<40 years) characterized by a higher frequency of pathogenic somatic mutations in *PIK3CA* and *TP53*, and germline mutation in *CHEK2*. However, the mutational landscape of breast cancer in young and old age patients appears homogenous. However, the researchers suggested that they may fail to find the claimed heterogeneity between the two age groups due to the small sample size of subjects allocated to the young age group (Andrikopoulou et al., 2022).

Yau and colleagues also compared the genomic profile of early (≤ 45 years) and late-onset (≥ 70 years) ER-positive breast cancer in sporadic Caucasian patients. Several oncogenes such as *MYC*, *CCND1*, *ZNF217*, *AIB1*, *MDM2*, *ESR1*, *ERBB2*, and *TOPO2A* were amplified in the older age group. However, there were no statistically significant changes in the chromosome copy number in terms of gain, loss, amplification or whole chromosome changes (Yau et al., 2007).

Another study investigated the mutational profile of Latin American young age breast cancer patients (20-45 years) supported the above-mentioned findings, where higher mutational loads were observed in *TP53* and *PIK3CA* genes, another interesting finding is that there is a higher percentage of single-base substitution-transversion mutation (G:C>T:A) in *TP53* compared to old age patients (27% vs 14%). (Olivier et al., 2019). This mutational pattern observed in the *TP53* gene reflects exposure to environmental carcinogens in lung cancer patients, mainly polycyclic aromatic hydrocarbons (PAH) (DeMarini et al., 2001). Interestingly, investigations in occupational and environmental medicine provide evidence of an elevated risk of breast cancer associated with PAH exposure, mostly in cases with premenopausal status (Petralia et al., 1999; Lee et al., 2019).

In a recent study, ER-positive breast cancer of very young age Korean patients (≤35 years) was classified into three groups based on the genomic profile which was characterized through RNA and whole exome sequencing analyses: the first group was enriched with luminal A tumours, the second group was enriched with luminal A and B tumours, whereas third group was enriched with HER-2 and luminal B tumours. The most frequent mutations observed in the cohort were in *TP53*, *PIK3CA* and *GATA3*, where *TP53* mutations were more prevalent among the third group and