Molecular markers in breast carcinoma

A study with focus on molecular phenotypes, angiogenesis and stem cells in an African population

Hawa Nalwoga

Dissertation for the degree of philosophiae doctor (PhD)

The Gade Institute, Section for Pathology
University of Bergen
NORWAY

2010
This thesis is dedicated to my beloved daughters Sophia and Shamim
as well as little Michael Joel
I love you more than you can imagine

May God bless you
CONTENTS

ACKNOWLEDGEMENTS ......................................................................................................................... 6
LIST OF PUBLICATIONS .......................................................................................................................... 8
LIST OF ABBREVIATIONS ........................................................................................................................ 9
INTRODUCTION ........................................................................................................................................ 10

INCIDENCE AND MORTALITY ............................................................................................................... 10
ETIOLOGY AND RISK FACTORS ........................................................................................................... 11
CLINICAL FEATURES .............................................................................................................................. 12
TUMOR BIOLOGY ................................................................................................................................... 13
   Cell cycle regulators and proliferation ................................................................................................. 15
   Genetic factors .................................................................................................................................. 17
   DNA ploidy ....................................................................................................................................... 22
   Steroid receptors ............................................................................................................................... 22
   Apoptosis ......................................................................................................................................... 23
   Angiogenesis .................................................................................................................................... 24
   Invasion and metastasis ..................................................................................................................... 27
   Cancer stem cells ............................................................................................................................. 33
MOLECULAR PHENOTYPES OF BREAST CANCER .............................................................................. 36
   Basal-like subtype ............................................................................................................................ 37
   Other subtypes ................................................................................................................................ 41
PROGNOSTIC AND PREDICTIVE FACTORS ........................................................................................... 42
   Clinical factors ............................................................................................................................... 45
   Histopathologic factors ...................................................................................................................... 45
   Tumor biomarkers and biologic factors ............................................................................................. 49
DIAGNOSIS AND TREATMENT ............................................................................................................... 54
Acknowledgements

I acknowledge my main supervisor Professor Lars A Akslen for tirelessly guiding me through the process, for the professional advice and the ensuring that I have everything that I needed in a timely and friendly manner. I thank my co-supervisor Professor Henry Wabinga for coordinating the collaboration successively and for the scientific nurturing that he provided whenever I approached him. I acknowledge Professor Leif Bostad for giving me guidance when I most needed it.

I would like to extend my gratitude to all my colleges and the staff at the Department of Pathology MUCHS for accepting my long periods of absence and for the endless and warm welcomes that they offered whenever I returned home. They made me feel more determined to achieve my goal.

I acknowledge my colleges Grace Ndeezi, Henry Wamani, Sissy Mitike, Kamaludeen Baba, Wevica Gronberg, Sam Kalungi, Robert Byamugisha, Jolly Kaharuza, Karsten Gravdal and Charles Karamagi with whom I shared a lot. I would to give special thanks to Annet Birungi, Rebecca Babirye, Robina Kusiima, Syllivia Nampebwa, Ssendikadiwa Emmy, Yahya Baweemu, Joseph Kaaya and Martin Kizito for all the social support and encouragement that they provided to my family. I acknowledge my sisters Pastor Gift Murungi and Monica Maririre as well as Betty Namuyiga and Siragi Ahebwa for the wonderful company they offered my family during my absence. I thank Sam Nkata and my brother Swaibu Muhumza Metha for the undying support they always offered whenever need arouse. I wholeheartedly thank my dear Mummy and Daddy for the wonderful blessings and wise encouragement they always offered me throughout the process. I am highly indebted to Lynnette K Tumwine for the unlimited social support which she offered my family, without her my studies would have been next to impossible. I thank her dearly for accepting to parent my daughters. I deeply appreciate the academic advice she offered whenever I felt like the
opposite. In addition, I acknowledge Professor James K Tumwine and Professor George B Kirya for the academic inspiration and advice through the process.

I extend my sincere gratitude to the Norwegian government, which funded this study through NUFU project # 64/2003. In addition, I acknowledge Makerere University for establishing the collaboration with the University of Bergen through which this study was funded. I thank the Department of Pathology, MUCHS and the Kampala Cancer Registry for allowing me to use the archival tissue and the data. I am indebted to Sarah Nambooze, Ruth Nakigudde, Gerd Lillian Hallseth, Bendik Nordanger, Randi Nygaard and Dorothy Lynn Nabbale for excellent technical assistance. I also extend my thanks to all the staff at Centre for International (CIH) and The Gade Institute, University of Bergen and Haukeland University Hospital for the friendly working environment that they offered. At CIH, special thanks go to Solfrid Hornell, Borgny Kvalnes Lavik and Unni Kvernhusvik. I wholly heartily thank Lillian for the motherly love and care that she always offered to me unreservedly. I would in particular like to acknowledge my colleague Jarle Amnes for accepting to help me at a time when I hardly knew anyone at the Institute. To Gyrild, you were a blessing, the speed and efficiency at which you made things happen is beyond what words can express; thank you so much, you made my study at the Institute enjoyable. In addition, I acknowledge the Centre for Disease Control and Prevention (CDC) for computerizing the records at the Department of Pathology, MUCHS which made data retrieval a lot easier.

Special acknowledgement goes to my lovely daughters Sophia and Shamim for honorably accepting my long period of absence and for seeing me through the process. Thank you for the beeping, it kept me going and alert.

Lastly but not least, I thank you all who have made the completion of this thesis possible. May God bless you all.

Bergen, March 2010

Hawa Nalwoga
List of publications

This thesis is based on the following articles referenced to in the text by their respective roman numerals:


### List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AJCC</td>
<td>American Joint Committee on Cancer</td>
</tr>
<tr>
<td>ALDH</td>
<td>Aldehyde dehydrogenase</td>
</tr>
<tr>
<td>BLP</td>
<td>Basal-like phenotype</td>
</tr>
<tr>
<td>CBP</td>
<td>Core basal phenotype</td>
</tr>
<tr>
<td>CK</td>
<td>Cytokeratin</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>ER</td>
<td>Estrogen receptor</td>
</tr>
<tr>
<td>HIF-1α</td>
<td>Hypoxia inducible factor-1α</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemical</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger Ribonucleic acid</td>
</tr>
<tr>
<td>MUCHS</td>
<td>Makerere University College of Health Sciences</td>
</tr>
<tr>
<td>MVD</td>
<td>Microvessel density</td>
</tr>
<tr>
<td>PcG</td>
<td>Poly comb genes</td>
</tr>
<tr>
<td>pMVD</td>
<td>Proliferative microvessel density</td>
</tr>
<tr>
<td>PR</td>
<td>Progesterone receptor</td>
</tr>
<tr>
<td>SI</td>
<td>Staining index</td>
</tr>
<tr>
<td>TNP</td>
<td>Triple negative phenotype</td>
</tr>
<tr>
<td>UICC</td>
<td>International union against cancer</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
<tr>
<td>VPI</td>
<td>Vascular Proliferation Index</td>
</tr>
</tbody>
</table>
INTRODUCTION

INCIDENCE AND MORTALITY

Worldwide, breast cancer is the most prevalent cancer as well as the most common female neoplasm accounting for 23% of all female cancers.\textsuperscript{1,2} According to the world cancer report, more than one million cases occur worldwide each year, and 45% of these are in developing countries.\textsuperscript{1,3} The incidence of breast cancer is increasing in most countries\textsuperscript{1,4-7} but the outcome is now much better in the western world. The five-year survival rates are over 70% in most of them\textsuperscript{1,3} although racial differences still exist.\textsuperscript{3,8,9} This reduction in the morbidity and mortality rates of breast cancer in the developed countries has been due to increasing early detection by way of mass screening as well as improved targeted therapy.\textsuperscript{1,3,8,10}

In spite of this, breast cancer still remains the leading cause of cancer mortality in women worldwide. In 2002, the estimated number of deaths was about 411,000 (14% of female cancer deaths).\textsuperscript{1} Although the risk is still low in sub-Saharan Africa, the incidence of breast cancer is increasing rapidly in most African countries,\textsuperscript{2,11} where breast cancer is more common in the urban population compared to the rural population.\textsuperscript{12} In Uganda it has doubled over three decades from 11/100,000 in 1965 to 22/100,000 in 1995 (Figure 1).\textsuperscript{13} It is now the second most common non-HIV related cancer\textsuperscript{14-16} affecting women in Uganda.\textsuperscript{13} Unfortunately, the outcome is still very poor. Five-year survival rates have been found to be very low 29% and 34% for patient with grade 3 and 2 tumors, respectively, in one study,\textsuperscript{17} and similarly the overall 5-year survival rate was 38%\textsuperscript{18} and 56%\textsuperscript{19} in previous reports.
ETIOLOGY AND RISK FACTORS

The etiology of breast cancer is multi-factorial involving both genetic and environmental influences. Well known factors include genetic, dietary and reproductive factors plus related hormonal imbalances. Numerous studies have shown that most etiological and risk factors for breast cancer are related to the cumulative exposure of the breast to estrogens both endogenous and exogenous and include early menarche, nulliparity, late age at first pregnancy, late menopause (after 55 years) and hormonal replacement therapy. In addition, the other major influences on the risk of breast cancer include genetic susceptibility, body size and obesity, alcohol, physical activity, and possibly diet plus the western lifestyle.1-3, 20

Regarding genetic susceptibility, these factors contributes about 5-10% of breast carcinoma risk.21 The highest risk is due to germline mutations in the high penetrance breast cancer genes which include BRCA1, BRCA2, TP53, CDH1, PTEN and LKB1/STK11. Breast cancer susceptibility genes with low to moderate risk include CHEK2, TGFβ1, CASP8 and ATM.22 Of these, BRCA1 and BRCA2 germline mutations have been extensively studied.
Mutations of the other genes such as TP53, PTEN, STK11 CHEK2 and ATM result in a small proportion of breast cancer syndromes.\textsuperscript{23} Although family history has been reported to be a marker of risk of breast cancer in the African setting, the prevalence of BRCA1 mutations in African populations is not clearly known.\textsuperscript{24}

**CLINICAL FEATURES**

Breast cancer is a highly heterogeneous disease with regards to morphology, hormonal receptor expression, invasive behavior, metastatic potential, as well as clinical behavior including response to treatment. Nevertheless, most primary invasive breast cancers are characterized by a palpable mass or lump, most frequently located in the upper outer quadrant and most often discovered by the patient.\textsuperscript{25} Other symptoms include nipple discharge, nipple lesions, skin edema plus redness and axillary lymphadenopathy. A small proportion will present with skin ulceration and skin retraction of the overlying skin or nipple.\textsuperscript{25}

However, the spectrum of breast cancer clinical presentation has been considerably changed by the introduction of mass screening by use of mammography. As a result of mammography screening, breast cancer tumor size and stage at presentation or detection have decreased.\textsuperscript{26-28} Whereas detection of non-invasive disease\textsuperscript{29} as well as impalpable breast lesions\textsuperscript{30} is more frequent. This has major implications on the management of breast cancer as well as on screening programs because of false positive cases and the lead time bias effects. Further, some racial differences in tumor size at presentation have been reported.\textsuperscript{31}

Regarding clinical presentation in African populations, it has been noted that African and African-American patients with breast cancer present in the late stages of the disease.\textsuperscript{12, 19, 32-34} In Uganda, a recent retrospective study of medical records of breast cancer presenting at the national referral hospital revealed that a majority (77\%) of patients presented in the late
stage according to the AJCC staging.\textsuperscript{35} Stage III was the peak stage at presentation with 51% of all patients, whereas 26% of patients had metastatic disease at presentation (Figure 2).\textsuperscript{19}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{stage.png}
\caption{Stage at presentation of breast cancer patients presenting at the national referral hospital from 1996-2000. Adapted in part from Gakwaya et al. (2008)\textsuperscript{19}}
\end{figure}

**TUMOR BIOLOGY**

The complex processes that characterize the development and progression of malignant tumors, the \textit{hallmarks of cancer}, have been well described. They include self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, unlimited replicative potential, sustained angiogenesis and tissue invasion and metastasis (Figure 3)\textsuperscript{36} plus the ability of the cancer to escape the immune response through several complex processes and events.\textsuperscript{37,38}
Further, the identification of cancer stem cells in breast cancer has led to more elucidation about evolution and progression of breast cancer.\textsuperscript{39} Also, research on morphologic and molecular features of hereditary breast cancer, especially in patients with germline mutations in $BRCA1$,\textsuperscript{23} a candidate stem cell regulator,\textsuperscript{40,41} has increased our understanding of breast cancer biology. Gene expression profiling studies have extended our understanding of the molecular mechanisms involved in tumorigenesis and progression of breast cancer. Basic research on genes involved in signaling pathways modulating proliferation, apoptosis, survival, angiogenesis, invasion, metastasis and drug resistance have provided more answers to the heterogeneity of breast cancer. There is increasing evidence that this heterogeneity finds its source in genetic variability.\textsuperscript{42-45}

\textbf{Figure 3: Hallmarks of cancer (Adapted from Hanahan and Weinberg, 2000).}\textsuperscript{36} Loss of normal growth control as a hallmark of cancer which encompasses four (self-sufficiency in growth signals, insensitive to antigrowth signals, sustained proliferation and evasion of apoptosis) of the six hallmarks of cancer as defined by Hanahan and Weinberg (2000)\textsuperscript{36} involve control over the cell cycle.
Cell cycle regulators and proliferation

The cell cycle is a highly organized and complex process comprised of a series of tightly controlled events that drive the replication of DNA and ensures correct cell division. Cells are normally in the resting phase \( G_0 \), and after appropriate stimuli they enter the proliferative phases of the cell cycle which is made of four phases; \( G_1 \), S, \( G_2 \) and M phase (Figure 4).

**Figure 4: The stages of the cell cycle indicating site of activity of regulatory CDK/cyclin complexes.**

In the \( G_1 \) phase, the cell is in a preparation for the S phase, in which DNA synthesis occurs followed by a second gap phase (\( G_2 \)) in preparation for the phase M in which the cell undergoes mitosis to generate two diploid \( G_0 \) cells which may reenter the cell cycle or persist in the resting phase.\(^{47, 48}\) Cells are stimulated to divide in response to numerous external signals, including growth factors, hormones and cellular adhesion.\(^{47-49}\) During the \( G_1 \) phase of
the cell cycle, cells are responsive to the external stimuli and are dependent on them until they reach the restriction point (R). This is a point of no return beyond which the cell is committed to enter the cell cycle and thereafter the process becomes autonomous.49

The transition through the cell cycle phases is mediated by sequential assembly and activation of a family of serine/threonine proteins, the cyclin dependent kinases (CDK; CDK1, CDK2, CDK4, CDK6 and CDK7) and the CDK inhibitors (CKI; INK4 family: p15, p16, p18, p19; Cip/Kip family: p21, p27). The CKI are regulated by both internal and external signals such as the TP53 tumor suppressor gene and Transforming Growth Factor β (TGF-β). The cell cycle has several check-points48 (Figure 5) to ensure an orderly sequence of events in the cell cycle as well as complete and accurate replication of the cell before division.48 Of these, the DNA damage check points (G1/S and G2/M) are well elucidated. Although it appears that oncogenic defects may target any major check-point, the most frequently

Figure 5: Check-points of the cell cycle. Adapted in part from Gillet et al. (1998).48
involved is the G1/S transition, and it encompasses many of the important cell cycle events that may be specifically altered in breast cancer including actions of the oncogenes (such as cyclin D1 and cyclin E) and tumor suppressors (such as p27).

Control of cell proliferation in the normal mammary gland is steroid hormone dependent, and it involves complex interactions with other hormones, growth factors and cytokines as well as three proto-oncogenes (c-myc, cyclin D1 and cyclin E1). Proliferation is essential for tissue turnover but it exposes the cell to the occurrence of DNA damage. Cell proliferation plays an important role in the clinical behavior of breast carcinoma and it is a significant prognostic factor in breast cancer. Tissue homeostasis results from the balance between cell proliferation, differentiation and death in the form of apoptosis. An imbalance between cell proliferation and apoptosis contributes to tumorigenesis and tumor progression.

Genetic factors

Cancer is considered to be a genetic disease caused by genomic instability at the chromosomal or DNA level, and breast cancer has all the hallmarks of a multistep genetic disease. Studies have shown that the development of human breast cancer is based on the accumulation of various genetic alterations, and almost every chromosome presents at least one site involved in cancer-related genetic alterations (chromosomal losses, DNA amplifications, mutations or altered DNA methylation patterns). The multistep process in breast cancer is driven by both inherited and acquired genetic alterations which result in changed expression of mRNA and various proteins. These abnormalities may be categorized into two; loss-of-function defects of tumor suppressor genes that have been inactivated by DNA mutation and unmasked by deletion or allelic loss, and gain-of-function genetic events that activate oncogenes. Several genetic alterations have been identified; somatic and
germline mutations have been described in tumor suppressor genes whereas oncogenes have been found to be activated.

**Tumor suppressor genes**

Several tumor suppressor genes have been implicated in breast carcinogenesis; mutations in genes such as *BRCA1, BRCA2, TP53, PTEN* or *ATM* or epigenetic functional inactivation of genes such as *SYK* and *NESI* play important early roles in formation of some breast cancers.\(^{54}\) Of particular significance are the germline mutations in the *BRCA1/BRCA2* genes and somatic alterations in the *TP53* gene. Other genes of interest include the retinoblastoma gene (*pRb*), *p16, NM23* and *MASPIN*.\(^{55}\)

**BRCA1 and BRCA2 genes**

Studies have indicated that *BRCA1* and *BRCA2* are tumor suppressor genes which are essential for cellular development and are involved in repair of double-stranded breaks (DSB) and the maintenance of genome integrity as well as cell cycle control.\(^{56-59}\) *BRCA1* has also been suggested to represent a stem cell regulator.\(^{40, 41}\) Mutations in these genes contribute to about 25-30% of hereditary breast cancer among young patients. In addition, it has been suggested that hypermethylation of *BRCA1* and *BRCA1* with inactivation may have a potential role in the carcinogenesis and aggressive phenotype of sporadic breast cancer.\(^{60-64}\)

Further, breast carcinomas occurring in women with *BRCA1* mutations are more likely to occur at an earlier age and are frequently high grade, aneuploid, estrogen receptor (ER) and progesterone receptor (PR) negative, p53 positive, have abundant lymphocyte infiltration and pushing margins.\(^{21, 22}\) They have also been shown to be associated with the basal-like phenotype,\(^{65}\) and are enriched with CD44+/CD24− stem cells.\(^{66}\) Decreased BRCA1 expression has been associated with acquisition of metastatic capacity, the solid-tubular
phenotype, poor tubular formation, high tumor grade and overexpression of HER2 in sporadic tumors.  

TP53 gene

The TP53 gene is the most frequently mutated gene in breast cancer and other human cancers. About 25% of breast cancers have somatic TP53 mutations, and 30-50% of breast tumors have overexpression of p53 protein. p53 is a nuclear transcription factor that is involved in control of gene transcription in the cell cycle (check-points) and promotes chromosomal stability maintaining the integrity of the genome. It regulates cell proliferation and apoptosis by preventing replication of damaged DNA and division of genetically altered cells. p53 protein binds to damaged DNA and regulates transcription of a number of genes. Some of these genes, such as GADD45, p21 and MDM2, are transcriptionally activated by p53 whereas genes such as c-myc and c-fos are repressed by p53. The transcriptional activation of p21 during the G1 phase leads to cell cycle arrest and prevents cells with damaged DNA from entering the cell cycle phases of DNA synthesis and replication. In addition, the p53 gene transcriptionally activates bax, a pro-apoptotic gene and down regulates transcription of bcl-2 which is a powerful antiapoptotic proto-oncogene. Consequently, inducing apoptosis through the bcl-2/bax pathway in susceptible cells in which the damage is beyond repair thereby protects the tissue against transmission of DNA abnormalities.

Mutations in p53 adversely affect its ability to bind regulatory DNA sequences of these genes and thus to inhibit their transcriptional regulation resulting into a cascade of downstream effects. Mutation of the p53 gene increases the risk of developing breast cancer and affects the biology of cancer cells and their response to therapy. Mutations are more common in ductal carcinomas than in lobular carcinomas and are commonly associated
with BRCA1 and BRCA2 germline mutations.\textsuperscript{55} Also, TP53 mutations have been associated with more aggressive disease.\textsuperscript{74, 75}

Oncogenes

Oncogenes refer to those genes whose alterations cause \textit{gain-of-function} effects that lead to activation and can contribute to the development of cancer.\textsuperscript{76} Activation of oncogenes can occur through various ways; gene amplification, point mutation and chromosomal translocation.\textsuperscript{76} Oncogenes may also act cooperatively with other genetic or epigenetic changes.\textsuperscript{76} Numerous oncogenes have been characterized in human cancers but only few oncogenes are crucial in the development of breast cancer.\textsuperscript{44, 76} Amplification and overexpression of these oncogene and oncogene products are the major mechanisms through which they contribute to carcinogenesis.\textsuperscript{76}

In breast cancer, oncogene amplification is a common mechanism,\textsuperscript{44} and is an important mechanism for oncogene overexpression.\textsuperscript{77} The HER2, EGFR, c-myc, CCND1, FGFR1, ESR1 and MDM2 are among the frequently amplified oncogenes. Coamplifications (HER2/c-myc or CCND1/FGFR1) have also been reported.\textsuperscript{45, 77}

\textit{HER2/neu} gene

The \textit{HER2/neu} proto-oncogene is amplified in 15-30% of breast cancer.\textsuperscript{78} \textit{HER2} (also known as \textit{neu}, c-erbB-2 or human epidermal growth factor 2) is a transmembrane protein with tyrosine kinase activity. \textit{HER2} has been implicated in breast carcinogenesis and plays an important role in development and progression of cancer.\textsuperscript{78} \textit{HER2} overexpression has been reported in 10-44% of human breast cancers.\textsuperscript{79-81} Overexpression in breast carcinoma occurs through either amplification of the gene or mRNA overexpression. This results in increased
cell proliferation, inhibition of apoptosis, and angiogenesis leading to poor prognosis in breast cancer.\textsuperscript{78, 82-85}

**EGFR gene**

EGFR is another member of the tyrosine kinase family of receptors which are transmembrane proteins regulating major cellular events such as cell proliferation, differentiation, apoptosis, adhesion and cell migration.\textsuperscript{86-89} Several studies have established that \textit{EGFR} gene acts as a cellular oncogene. \textit{EGFR} gene amplification has been identified in 0.8-14\% of breast cancers.\textsuperscript{77, 90, 91} Epidermal growth factor influences the proliferation and differentiation of a wide variety of cancer cells, and plays a role in the pathogenesis of breast cancer.\textsuperscript{92, 93} In addition, it influences cell proliferation and a number of other processes in tumor progression such as cell survival, cell adhesion, cell motility, angiogenesis and tissue invasion.\textsuperscript{92} EGFR expression has been reported in about 45\% (range 14-91\%) of all breast cancers.\textsuperscript{94-96}

Amplification of \textit{EGFR} in breast cancer indicates a more aggressive tumor behavior and a poor patient outcome.\textsuperscript{91, 97-100} Similarly, EGFR expression in breast cancers has been associated with features of poor prognosis including high tumor grade, elevated growth fraction, ER negativity and poor response to endocrine therapy and reduced survival.\textsuperscript{96, 101-104}

**C-myc gene**

\textit{C-myc} amplification is relatively common (8-37\%) in breast cancer and may provide independent prognostic information.\textsuperscript{105} It encodes for a helix-loop/leucine zipper protein and myc responsive genes include those whose protein products regulate cell proliferation and apoptosis.\textsuperscript{105} The \textit{HER2/c-myc} coamplified tumors have worse prognosis than tumors with only one of these amplified.\textsuperscript{45, 77}
CCND1 gene

The CCND1 gene located on chromosome 11q3 and coding for the G1-cyclin protein (cyclin D1) involved in regulation of the cell cycle has been found amplified in 10-27% of breast cancers. CCND1 amplifications are associated with ER and PR positivity, but studies on prognostic significance are still controversial.

MDM2 gene

The MDM2 gene protein product down regulates the TP53 tumor suppressor gene and is amplified in 4-7.7% of breast cancers and has been associated with poor prognosis in some studies.

DNA ploidy

DNA aneuploidy is a manifestation of chromosomal instability which is recognized as an early feature of malignant transformation and found to be an indicator of prognosis in breast cancer. The mechanisms responsible for the frequent instability of genomes of breast cancer cells have been poorly understood although recent functional findings on oncogene and tumor suppressor genes have provided more information about this matter. Studies have suggested that the DNA content of breast cancer cells reflects biologic properties associated with malignant behavior of the tumors.

Steroid receptors

Ovarian steroid hormones estrogen and progesterone are necessary for normal mammary development and growth. The estrogen (ER) and progesterone receptors (PR) belong to the steroid hormone receptor family of inducible transcription factors that play a role in the development and progression of breast cancer. Studies have shown that
estrogen directly increases the growth of breast cancer cells in culture by increasing the number of cells entering the cell cycle (Figure 4). ER directly regulates several key G1 phase cell-cycle regulators (such as cyclin D1, Myc, cyclin E-CDK2 complex, CDK4 and CDK inhibitors) and those required for S phase entry.\textsuperscript{113-115} In addition, studies have suggested that PR induce cell cycle progression via activation of mitogen activated protein kinases in breast cancer cell lines.\textsuperscript{116} Steroid hormone receptors are directly involved in the development, progression and therapeutic responsiveness of breast cancer.

ER is expressed in about 50-95\% of breast carcinomas, while PR is expressed in 60\% to 70\% of the cases.\textsuperscript{117, 118} However, previous independent studies have shown a low prevalence of ER (23-33\%) in women of African and African-American populations,\textsuperscript{32, 119-122} although some studies reported higher frequencies (64-65\%).\textsuperscript{123, 124} ER/PR positive tumors are more common in postmenopausal women and are more likely to be diploid, well differentiated, to have lower proliferative rates, and to be less aggressive than the receptor negative tumors.\textsuperscript{42}

Furthermore, estrogens in mammary epithelial cells and ER positive breast cancer cell lines have been shown to regulate,\textsuperscript{125} the expression of \textit{bcl-2}, a powerful antiapoptotic proto-oncogene.

**Apoptosis**

Apoptosis is a highly complex and tightly regulated process of cell death which deprives the proliferating cellular pool and allows the elimination of genetically damaged cells after their division. It is also a cellular protective mechanism against malignant transformation. Apoptosis regulation is ensured by various genes often associated with breast carcinogenesis, mostly pro-apoptotic (\textit{c-myc}, \textit{p53} and \textit{ras}) and rarely antiapoptotic (\textit{bcl-2}). The \textit{bcl-2} gene was the first antiapoptotic gene to be described and is able to antagonize
apoptosis induced by several stimuli. Bcl-2 is one of the important regulators of apoptosis,\textsuperscript{126} and it delays the induction of apoptosis in mammary glands.\textsuperscript{126, 127} The expression of the \textit{bcl-2} gene is regulated by estrogens\textsuperscript{125} as well as down regulated by p53 in breast cancer cell lines.\textsuperscript{128} Several independent studies have shown that Bcl-2 overexpression in breast cancer correlates with biologic features of a differentiated phenotype (low proliferative rate, high levels of steroid receptor, weak or absent p53 expression and absence of HER2 expression).\textsuperscript{129}

The ability of tumor cells to evade apoptosis, as a \textit{hallmark of cancer},\textsuperscript{36} leads to continued proliferation of tumor cells and ultimate tumor expansion. Thus, dysregulation of apoptosis plays an important role in the pathogenesis and progression of breast cancer. The development and continued growth of cancers involve an interaction between cell proliferation and apoptosis.\textsuperscript{130-133} It has been shown that apoptosis is increased in invasive breast cancer\textsuperscript{134} and is positively correlated with Ki-67 expression.\textsuperscript{135} Breast tumors with increased apoptosis are more likely to be of high histologic grade and to be ER negative. Further, studies have shown that the rate of tumor growth depends in part on the excess of proliferation over apoptosis,\textsuperscript{130-133} and partly on angiogenesis.\textsuperscript{136}

\textbf{Angiogenesis}

In 1971, Judah Folkman suggested that the growth and spread of malignant tumors were dependent on the process of angiogenesis, and that tumors could be treated by attacking their blood supply.\textsuperscript{136-139} Tumor-associated angiogenesis is now considered one of the key elements which contribute to tumorigenesis.\textsuperscript{140-143} Sustained angiogenesis, another \textit{hallmark of cancer},\textsuperscript{36} is a tumor micro-environmental process that is necessary for tumor cell survival, tumor growth, invasiveness, progression and development of metastasis, and beyond a critical volume a tumor can not expand further in absence of neovascularization.\textsuperscript{139, 142, 144-146} Angiogenesis is a complex multi-step process, consisting of coordinated, sequential and
interdependent steps leading to formation of new blood vessels from pre-existing vascular networks.\textsuperscript{146, 147} It is a highly restricted process in normal human adult tissues, and in order to initiate it, a tumor must switch to the angiogenic phenotype. This occurs early in tumor development and limits or determines the rate of tumor progression.\textsuperscript{139, 141, 142, 146-148} The \textit{angiogenic switch} is induced by the secretion of specific endothelial cell growth factors like VEGF (vascular endothelial growth factor) produced by the tumor cells plus other non-malignant host cells recruited by the tumor.\textsuperscript{142, 146}

\textbf{Vascular Endothelial Growth Factor}

The \textit{angiogenic switch} of a tumor is related to a balance between positive and negative regulators (Figure 6). Several pro-angiogenic factors have been identified, and the vascular endothelial growth factor (VEGF) family plays a key role in this process as the major mediator of breast cancer angiogenesis.\textsuperscript{139, 141, 146, 149} VEGF is the most active, specific and potent mitogen for vascular endothelium among the endothelial cell growth factors,\textsuperscript{146, 150, 151} and is a potent inducer of angiogenesis.\textsuperscript{139, 141, 149, 150, 152} It is secreted in response to environmental stimuli like hypoxia which is the main stimuli, certain cytokines and estradiol.\textsuperscript{146, 149} It plays crucial roles in cancer biology including endothelial cell proliferation and migration, promotion of tumor angiogenesis and metastasis.\textsuperscript{152, 153} Studies in breast cell lines showed that down regulation of the \textit{VEGF} gene expression inhibited breast cancer cell-induced angiogenesis and suppressed breast tumor metastasis in mice.\textsuperscript{154}
Figure 6: The levels of the angiogenic inhibitors and activators factors control the angiogenic switch as well as the angiogenic activity of a tumors including breast cancer. Adapted from Hanahan (1996).

Furthermore, in vitro studies have shown that angiogenesis is also related to other molecular mechanisms involved in tumor growth and metastasis. Certain oncogenes such as HER2 signaling pathways promote angiogenesis by up-regulating cancer-cell production of angiogenic factor like VEGF.\textsuperscript{155, 156} In contrast, the p53 transcription factor has been reported to have a role in suppressing angiogenesis through enhancing the expression of Thrombospondin-1, an angiogenic inhibitor,\textsuperscript{157} as well as down regulating VEGF expression\textsuperscript{158} (Figure 6). p53 contributes to the angiogenic switch during tumorigenesis. It inhibits Hypoxia-inducible factor-1 (HIF-1) activity by targeting the HIF-1α for ubiquitination and proteasomal degradation.\textsuperscript{159} Thus, loss of p53 function leads to an amplification of normal HIF-1-dependent response to hypoxia,\textsuperscript{159} which is a key signal for induction of angiogenesis.\textsuperscript{160} Indeed, hypoxia is one of the most potent inducers of VEGF mRNA synthesis, a function achieved through inducing HIF-1α.
Studies have shown that tumor growth, invasion and metastasis of breast carcinoma depend partly on angiogenesis.\textsuperscript{136, 154} Thus increased tumor angiogenesis has been associated with increased incidence of metastasis.\textsuperscript{161, 162}

**Invasion and metastasis**

Most deaths from cancer result from progressive growth of metastases that are resistant to conventional therapies, and in a significant number of patients metastases occur before diagnosis of the primary tumor.\textsuperscript{154} Tissue invasion and metastasis are exceedingly complex processes whose mechanisms are closely related but are poorly understood and are some of the acquired capabilities of cancer.\textsuperscript{36} The existence of an invading cancer does not necessarily imply metastasis, but invasive growth is a prerequisite for metastasis. Cancer cell invasion involves the breaching of tissue barriers by the cancer cells, and subsequent infiltration of these cells throughout the surrounding tissue.\textsuperscript{163} Several gene families are involved in this process. Acquired genetic alterations conferring growth advantage to the cells in addition to loss of cell-cell adhesion or cell-matrix adhesion and matrix remodeling all interplay to confer a migratory plasticity to the cancer cells.\textsuperscript{163, 164} Studies have indicated that the motility machinery of the cells is extremely important; and acquisition of a motile phenotype is essential for the tumor cells to become invasive.\textsuperscript{164, 165} Tumor cell motility is the hallmark of invasion and an essential step in metastasis, and evidence shows that tumor microenvironment might initiate the expression of genes that induce cell motility, invasion and metastasis.\textsuperscript{165-168}

Single epithelial cells can migrate through two predominant mechanisms.\textsuperscript{169} The mesenchymal migration which requires an epithelial-mesenchymal transition (EMT) is the predominant mechanism and requires matrix degrading enzymes. The second type, the amoeboid migration, enables cells to squeeze their way through the matrix without need for
the proteases and requires a mesenchymal-amoeboid transition (MAT). This migratory method has implications for the treatment of breast cancer since it is used as a compensatory mechanism when the predominant one has been blocked. Further, factors from the tumor microenvironment such as cytokines, growth factors, proteases and angiogenic factors secreted from multiple cell types plays a major role in determining the potential invasion and later metastasis in cancer.

Cell adhesion molecules play major roles in the invasion-metastasis cascade. Whereas activation of integrin αvβ3 initiates calcium-dependent signaling pathway leading to increased cell motility and proteolysis, loss of E-cadherin expression facilitates tumor cell detachment enabling invasion and metastases. During tumor progression E-cadherin can be functionally inactivated or silenced by different mechanisms and loss of E-cadherin expression and/transcriptional repression of its mRNA are hallmarks of epithelial-mesenchymal transition (EMT).

**Epithelial-mesenchymal transition (EMT)**

By this process, polarized epithelial cells are converted into motile mesenchymal cells. The initial step of metastasis is epithelial-mesenchymal transition (EMT) which involves disruption of the adhesive interactions with surrounding cells and the acquisition of a motile phenotype. EMT is characterized by loss of polarity and down regulation of epithelial proteins, mostly E-cadherin, but also occludin, claudins, cytokeratins or catenin proteins in addition to inducing mesenchymal proteins like N-cadherin, vimentin and others. Multiple signaling pathways and effectors induce or contribute to the EMT and the key players include Receptor Tyrosine Kinases, the Transforming Growth Factor β superfamily, NF-κB, WNT signaling, Notch signaling and Hedgehog signaling. In addition, the EMT transcriptome program is controlled by several transcription factors outlined in Table 1.
Transcription factor

**Snail family**
- SNAI1 (Snail)
- SNAI2 (Slug)

**ZEB family**
- SIP1/ZEB-2
- βEF-1/ZEB-1
- TWIST1
- TWIST2
- E12/E47 (E2A gene product)
- FOXC2
- Goosecoid

**Table 1:** *Transcription factors involved in the epithelial-mesenchymal transition.*

The current model proposes that EMT is a two way process and EMT occurs at the invasion front of tumors whereas mesenchymal-epithelial transition (MET) occur at the secondary site (Figure 7).

**Figure 7:** A reversible EMT model in tumor metastasis, with deregulation of cell proliferation and eventual acquisition of a motile phenotype; tumor cells breach the
basement membrane and enter the blood or lymphatic vessels. At the distant organ, the
cancer cells exit the vessels and undergo a reverse mesenchymal-epithelial transition
(MET) and regain their ability to proliferate. Adapted from Thiery et al. (2002)

EMT can promote metastasis in several ways and some of the EMT mediators also
inhibit apoptosis (snail and twist families) which promotes tumor growth and expansion and
mediate tumor immunosuppression (snail) potentially facilitating metastasis.

Metastasis is a complex process including primary tumor growth, local invasion
through basement membrane and extracellular matrix, angiogenesis and lymphangiogenesis,
dissemination to lymphatic and/or blood circulation, transport to distant organs and
colonization at the secondary site. Recent evidence indicates that metastatic capacity is an
early and inherent feature of breast tumors and not a late event. In breast cancer, metastases
occur most commonly in the bone, lung and liver (Figure 8). Other relatively frequent sites
include adrenal glands, pleura, gastrointestinal tract, brain and the peritoneum. Studies have
shown that gene expression signatures can predict the likelihood of distant metastases with
90% accuracy as well as the site of breast cancer metastases.

In addition, gene expression studies identified a poor-prognosis signature which
included genes involved in the cell cycle, signal transduction, angiogenesis, invasion and
metastasis. These also included genes almost exclusively expressed by stromal cells such as
MMP1 and MMP9 which are required for extracellular matrix (ECM) degradation and tumor
invasion.
Further, gene expression profiles have demonstrated that the tumor microenvironment plays a significant role in tumorigenesis. Current evidence shows that tumor microenvironment initiates the expression of genes that induce cell motility, invasion and metastasis. Many of the EMT-inducing pathways play prominent roles in development and stem cell self-renewal. There is rapidly accumulating evidence which suggest that a link exists between stem cells and EMT. Mani et al. (2008) demonstrated that EMT induced by twist or snail endows breast epithelial cells with stem cell-like properties. Conversely, normal and neoplastic stem cells isolated from breast tissues show several features of EMT, and several signaling pathways that mediate stem cell self-renewal also induce EMT. Further, it has been proposed that the biological and molecular heterogeneity as well as the risk of distant metastasis corresponds with the amount of breast cancer stem cells (see Figure 8: Common metastasis sites of breast cancer as seen at autopsy. Adapted from Weigelt, (2005).}
next section) in the tumor. Consequently, a new integrated model of breast cancer metastasis which is illustrated in Figure 9 has been proposed by Weigelt et al. (2005).

Figure 9: The integrated model of breast cancer metastasis, adapted from Weigelt et al. (2005). Oncogenic mutations occurring in the breast stem cells (red) and the differentiated progenitor cells (yellow) generate metastatic ‘poor prognosis’ (orange) and non-metastatic ‘good prognosis’ breast cancers (pink), respectively. In the metastatic tumors, under the influence of stromal fibroblasts, a small population of breast cancer stem cells has the ability to metastasize. There might be variants of cancer stem cells that differ in their tissue selectivity for metastasis, expressing additional tissue-specific profile (such as green; bone, purple; lung).

Interestingly, studies in brain tumor cell lines have shown that cancer stem cells (CSC) contributes to the angiogenic drive in tumors by generating VEGF and other factors to induce angiogenesis. The CSC-mediated VEGF production led to amplified endothelial cell migration and tube formation in vitro suggesting that cancer stem cells may be a crucial source of key proangiogenic factors in cancers. At the same time, tumor vasculature aids in maintaining CSC self-renewal and maintenance. Cancer stem cells depend on CSC
maintenance signals created by the vasculature similar to what has been observed in normal stem cells (Figure 10).^187,188

Figure 10: CSC generate pro-angiogenic factors to stimulate angiogenesis while the tumor vasculature aids in maintaining CSC self renewal and maintenance. Adapted in part from Eyler and Rich (2008).^188

Cancer stem cells

The term ‘cancer stem cell’ is an operational term defined as a cancer cell that has the ability to self-renew giving rise to another cancer stem cell as well as undergo differentiation to give rise to phenotypically diverse mixed populations of non-tumorigenic tumor cell populations in the tumor.^190^-192 The cancer stem cell hypothesis has fundamental implications for cancer biology and clinical management of patients.^[193,194 It implies that breast cancers arise in mammary stem or progenitor cells through dysregulation of the normally tightly regulated process of self renewal.^[195 These “cancer stem cells” are thought to drive the growth and spread of tumors.^[190,191 Therefore, failure to target them would set the stage for
recurrences and treatment failures.\textsuperscript{190, 191, 194} Studies in mouse models and established breast cancers have suggested that breast cancer behavior may be programmed in the precancer stem cells,\textsuperscript{196} and the amount of cancer stem cells within breast tumors may correspond to the biologic and molecular heterogeneity of the tumors\textsuperscript{184} as well as to risk of distant metastases.\textsuperscript{185, 186} Pece et al. (2010),\textsuperscript{184} characterized the transcription signature of human normal mammary stem cells (hNMSC signature) and by using markers of this signature isolated stem cells from both the normal gland and breast tumors. In xenografts, the hNMSC signature was able to predict the biologic and molecular features of breast cancers.\textsuperscript{184} The ability to identify these tumorigenic cancer cells has facilitated the elucidation of pathways that regulate their growth and survival,\textsuperscript{190, 191} and might lead to development of novel CSC-targeted therapies\textsuperscript{197-199} which will eliminate breast CSCs.\textsuperscript{198, 199}

Importantly, exploration of cellular and molecular mechanism involved in the relationship between CSC and tumor angiogenesis that has been established in brain tumors\textsuperscript{187, 189} will provide opportunities for the development of novel CSC-targeted antiangiogenic therapies with advantage over currently available therapies.\textsuperscript{188, 197}

**Stem cell markers**

The existence of stem cells in rodent mammary glands was first demonstrated by Kordon and Smith (1998).\textsuperscript{195, 200} Consequently, human mammary stem cells have been identified and purified based on their surface antigen expression.\textsuperscript{191, 201} Human breast cancers are reported to contain a subpopulation of cancer cells similar to epithelial stem cells, the “cancer stem cells”.\textsuperscript{185, 190, 192} Studies have shown that human breast cancers and cell lines contain a subpopulation of cells characterized by CD44\textsuperscript{+}/CD24\textsuperscript{−/low}/Lin\textsuperscript{−} cell surface markers, and these cells have stem cell properties.\textsuperscript{190, 202} Breast cancer stem cells which expressed a combined CD44\textsuperscript{+}/CD24\textsuperscript{−/low}/ALDH1\textsuperscript{+} phenotype showed an especially high tumorigenic
capacity. In a recent study of 33 breast cell lines derived from human breast cancers and normal breast tissue, the results indicated that 23 of the cell lines contained functional cancer stem cells with metastatic capacity. In addition to increased aldehyde dehydrogenase 1 (ALDH1) expression, BMI-1 expression has been reported as stem cell marker.

**BMI-1 expression**

BMI-1 expression, a putative stem cell marker, is one of the several polycomb genes (PcG) which have been identified as oncogenes. It was first identified as an oncogene that co-operates with c-myc in the generation of mouse pre B-cell lymphomas. It is a transcriptional repressor which acts as a key regulator of self-renewal activity in both normal and tumorigenic human mammary stem cells. The PcG play a role in maintenance of cellular identity and contribute to regulation of the cell cycle by preserving gene silencing after cell division. Thus, dysregulation of this gene silencing machinery can lead to cancer, and BMI-1 has been implicated in breast cancer carcinogenesis, tumor progression and metastasis.

**Aldehyde dehydrogenase 1**

ALDH1 is another stem cell marker which is considered to be an indicator of both normal and malignant stem and progenitor cells in the breast. ALDH is a family of cytosolic isoenzymes responsible for oxidizing intracellular aldehydes, leading to oxidation of retinol to retinoic acid in early stem cell differentiation, which is important for proliferation, differentiation and survival. ALDH1 (also known as ALDH1A1) is the predominant ALDH isoform in mammals, and it is highly expressed in the hematopoietic progenitors and in intestinal crypt cells as well as in breast tumor cells. In breast cancer, ALDH1 expression has been associated with poor clinical outcome, resistance to chemotherapy and
the basal-like phenotype of breast cancer.\textsuperscript{193, 194, 205} Also, in a recent study, in both \textit{in vitro} and xenografts, the results showed that invasion and metastasis in inflammatory breast cancer are mediated by a CSC component that displays ALDH enzymatic activity,\textsuperscript{223} and ALDEFLUOR-positive cells were found to be responsible for mediating metastasis in a study involving 33 cell lines derived from breast tissues.\textsuperscript{203}

**MOLECULAR PHENOTYPES OF BREAST CANCER**

Gene clustering analyses have indicated that breast cancer can be divided into two broad categories; ER+ and ER− groups which can further be subdivided into additional biologically different and clinically significant subgroups. Thus, five different sets of intrinsic gene clusters were recognized (luminal A, luminal B, the HER2+ subtype, the basal-like and the normal breast-like category) with different prognosis in multiple independent studies.\textsuperscript{224-228} Although gene expression profiling is the gold standard for molecular classification of breast cancer, its large scale clinical use or use in retrospective studies is limited by the strict tissue requirements (fresh and frozen tissue) and by issues of cost, complexity and technical feasibility.\textsuperscript{229, 230} Consequently, in an attempt to develop a molecular classification that is clinically significant, technically simple, reproducible, and readily available, investigators have proposed an immunohistochemical-based classification.\textsuperscript{229} These biomarkers can define the molecular subgroups in the routine and readily available formalin-fixed, paraffin embedded tissues by way of immunohistochemical staining. Although some of the proposed IHC markers have been validated using a 930-case tissue microarray,\textsuperscript{231, 232} there is, however, still no consensus on these definitions,\textsuperscript{233} and overlapping categories exist. By using the immunohistochemical classification, four similar major subgroups have emerged as well as the unclassified tumors (Figure 11) which encompass the normal breast-like class of breast cancer that is still poorly characterized immunohistochemically.\textsuperscript{233} Of these, the basal-like
breast cancer (also known as basal-like phenotype or basal-like subtype) and the HER2+ subtype are of particular interest since they have a poor prognosis.\textsuperscript{226}

**Figure 11:** *The immunohistochemical subclassification of breast cancer (simplified)*

**Basal-like subtype**

There is no consensus on how to define this subgroup. Basically, these tumors might be defined on the basis of expression of various basal markers. Alternatively, negativity for ER, HER2 and eventually PR might be added to obtain more composite basal-like profiles. According to the latter, basal-like breast carcinomas usually lack ER and HER2 and express genes characteristic of basal or myoepithelial cells such as basal cytokeratins (CK5, CK14, CK17) and other genes characteristic of basal-like cells of the breast.\textsuperscript{233-237} In addition to
structural roles, many of the basal-like gene products have been implicated in cellular proliferation, suppression of apoptosis, cell migration and invasion, all hallmarks of cancer.\textsuperscript{36, 228, 233} Indeed, gene expression studies have further shown that a majority of basal-like tumors express the activated wound-response signature,\textsuperscript{238} which represents important processes likely to contribute to cancer invasion and metastasis such as matrix remodeling, cell motility and angiogenesis.

Further, it has been suggested that different subtypes of breast cancer might originate from breast stem or progenitor cells at distinct stages of lineage differentiation, with basal-like tumors arising from the most-primitive ER-negative stem cells.\textsuperscript{239-241} Gene expression profile studies of basal-like tumors have suggested a less differentiated breast stem cell or progenitor cell of origin for these tumors\textsuperscript{241} and several gene products in the basal cluster are also expressed in stem cells of various tissue types.\textsuperscript{241, 242} Given the central role of BRCA1 in normal mammary development,\textsuperscript{243} Foulkes (2004)\textsuperscript{40} proposed that BRCA1 regulates differentiation of breast stem cells, and defects in the BRCA1 pathway might arrest further differentiation of these cells leading to cancer. Subsequent studies have provided some evidence that basal-like breast cancers originate in stem cells with maturation defects and genomic instability caused by BRCA1 mutations.\textsuperscript{41, 194}

The basal-like breast carcinomas contribute about 8-25\% of all breast cancers as defined using gene expression or IHC surrogate criteria.\textsuperscript{233, 241, 244} They express basal markers such as basal cytokeratins in addition to other makers like EGFR, P-cadherin, p63 and c-kit.\textsuperscript{231, 232, 236, 245-247} However, unexpectedly, basal-like tumors might also co-express luminal cytokeratins CK8 and CK18.\textsuperscript{241} The basal-like subgroup partially overlaps with the so called triple negative tumors defined as being ER\textsuperscript{−}/PR\textsuperscript{−}/HER2\textsuperscript{−}, as well as the BRCA1 associated breast cancers (Figure 12).\textsuperscript{233, 248} A majority (82\%) of basal-like breast tumor were found to contain p53 mutations.\textsuperscript{226}
Clinically, the basal-like tumors have been associated with younger age (< 40 years) and are more likely premenopausal African-American women in some studies. Studies have shown that the basal-like subtype seems to differ by race and age, whereas other major subtypes do not seem to show a clear difference. Also, previous reports have indicated that the hormonal receptor negative tumors as well as the basal-like subtype are overrepresented in women from African population.

![Diagram of interrelationships of basal-like (BP), triple negative (TNBCs), and BRCA1 associated breast carcinomas.](image)

**Figure 12: The interrelationships of the basal like (BP), the triple negative (TNBCs) and the BRCA1 associated breast carcinomas. Adapted from Diaz et al. (2007)**

Morphologically, a majority of basal-like breast cancers are usually of high histologic grade and invasive ductal carcinomas. The basal-like tumors are seen as sheets of cells with minimal tubule formation which are more likely to have a pushing non-infiltrative tumor border, higher degree of stromal lymphoplasmacytic infiltration and larger zones of geographic necrosis than the non basal-like tumors. These characteristics represent medullary features. Morphologic characteristics of basal-like breast cancers that have been confirmed in a number of independent studies, although not in all, are listed in Table 2, but however they may also be found in other grade 3 non basal-like tumors. The high proliferative rate of basal-like tumors which has been reported in some reports may explain their
overrepresentation among so-called interval breast cancers. Further, Foulkes et al. (2004) observed that glomeruloid microvascular proliferation (GMP), a histologic marker of an aggressive angiogenic phenotype in human cancer, was significantly more frequent in the basal-like subgroup of breast cancer.

### Table 2: Morphologic characteristics of basal like breast cancer

<table>
<thead>
<tr>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pushing invasion border</td>
</tr>
<tr>
<td>Central scar or sclerosis</td>
</tr>
<tr>
<td>Geographic tumor necrosis</td>
</tr>
<tr>
<td>Marked cellular pleomorphism</td>
</tr>
<tr>
<td>High nuclear grade</td>
</tr>
<tr>
<td>High mitotic count (average 45 mitoses per 10 high power fields)</td>
</tr>
<tr>
<td>High nuclear-cytoplasmic ratio</td>
</tr>
<tr>
<td>Vesicular chromatin</td>
</tr>
<tr>
<td>Prominent nucleoli</td>
</tr>
<tr>
<td>Lack of tubule formation</td>
</tr>
<tr>
<td>Frequent apoptotic cells</td>
</tr>
<tr>
<td>Spindled tumor cells</td>
</tr>
<tr>
<td>Metaplastic features such as squamous cell metaplasia</td>
</tr>
<tr>
<td>Scant stromal content</td>
</tr>
<tr>
<td>Exaggerated stromal lymphoplasmacytic response</td>
</tr>
</tbody>
</table>

Interestingly, the basal-like subgroup is reported to have a specific pattern of metastatic spread with reduced lymphatic metastases and increased hematogenic spread to sites associated with poor prognosis. They show relatively increased propensity for the
lungs and brain metastases whereas they have a decreased propensity for bone and liver metastases. Thus it has been proposed that basal-like breast cancer posses a distinct mechanism of metastatic spread.

Regarding the therapeutic implications of the molecular subtypes, a recent report has suggested that some chemotherapeutic agents might have different mechanisms of action in different subtype of breast cancer. The basal-like breast cancers are resistant to currently available therapeutic targets for breast cancer, although they may be responsive to EGFR targeted therapy. Also, studies have shown that ER negative tumors benefit twice as much from chemotherapy than the ER positive tumors. A number of studies, although not all, have indicated that basal-like tumors have a higher response rate to chemotherapy both as adjuvant and neoadjuvant regimens compared to the luminal subtype.

The current challenge is to identify novel target molecules and pathways for the basal-like subtype which is frequently triple negative. Some possible targets that have been proposed include EGFR and VEGF. C-kit which is expressed in a high proportion of basal-like tumors might also be a suitable target. However, c-kit positive breast tumors have been shown to lack activating c-kit mutations which conveys sensitivity to imatinib, a c-kit inhibitor. The biologic similarities between BRCA1 associated and basal-like tumors have suggested that strategies like PARP inhibitors targeting DNA-repair defects of the BRCA1 pathway dysfunction in basal-like tumors might be effective.

Other subtypes

In general, the HER2+ subtype has been defined as ER−/PR− and HER2 positive tumors. The HER2+ tumors express high levels of genes located in the HER2 amplicon including HER2 and the GATA4 transcription factor. They lack expression of ER and GATA3. Current literature shows that it contributes about 8-12% of the breast cancers.
This is an aggressive subtype which has been associated with high histologic grade and reduced survival. Fortunately, the clinical outcome of patients with HER2 positive tumors has been greatly improved by development of HER2 targeted therapy like trastuzumab which is now routine treatment for breast cancer. Thus, HER2 expression is a predictive factor currently in use.

Luminal tumors are ER positive tumors that express ER responsive genes and other genes that encode characteristic proteins of luminal epithelial cells such as PR, GATA3, BCL-2 and the luminal cytokeratins 8 and 18. They contribute about 50-70% of breast cancers. Luminal tumors are usually associated with increasing age, low histologic grade, they are less aggressive and have a good prognosis and will respond to hormonal therapy. The luminal A subtype is most frequent and has a better prognosis than luminal B tumors which are more frequently ER+/PR− and have a higher tumor cell proliferation.

The normal breast-like subtype is also a predominantly ER negative group. It has relatively high expression of many genes known to be expressed by adipose tissue and other non epithelial cell types as well as strong expression of genes in the basal cluster but low expression of luminal epithelial genes. However, some reports have suggested that it may potentially be due to normal tissue contamination. This group is still poorly characterized, most IHC studies have not included this subtype because of its complex expression patterns which can not be summarized into a simple 5-marker panel.

**PROGNOSTIC AND PREDICTIVE FACTORS**

Currently, histopathologic evaluation of breast cancer includes a detailed description of morphologic patterns and biologic parameters of the tumor, including prognostic and predictive factors. A prognostic marker might be related to molecular mechanisms involved in tumor growth, progression, invasion and metastasis and gives significant
information on clinical outcome for groups of patients. A predictive factor is a clinical, pathologic or biologic feature that is used to estimate the likelihood of response to a particular type of adjuvant therapy. \cite{42,273} Hence, the use of prognostic and predictive factors has mainly three reasons, \cite{274} to identify patients:

- who may not require adjuvant therapy after local surgery
- whose prognosis is poor enough to warrant a more aggressive adjuvant therapy
- whose tumors are more likely to be responsive or resistant to particular types of therapy

Several potentially useful prognostic and predictive factors have been suggested and can broadly be categorized into clinico-pathologic factors and biologic factors including tumor biomarkers as shown in Table 3. The College of American Pathologists\cite{275} has categorized such factors into 3 groups. Category I, are factors with prognostic importance being useful in clinical management of patients; Category II includes factors that have been extensively studied, but whose importance remains to be validated in statistically robust studies; Category III includes all other markers not sufficiently studied to demonstrate their prognostic value.
Table 3: *Useful and potential prognostic and predictive parameters in breast cancer.*\(^{42}\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient related factors</strong></td>
<td></td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td></td>
</tr>
<tr>
<td>Ethnicity/race</td>
<td></td>
</tr>
<tr>
<td><strong>Histopathologic features</strong></td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td></td>
</tr>
<tr>
<td>Tumor differentiation</td>
<td></td>
</tr>
<tr>
<td>Histologic type</td>
<td></td>
</tr>
<tr>
<td>Histologic grade</td>
<td></td>
</tr>
<tr>
<td>Lymph nodes status (stage)</td>
<td></td>
</tr>
<tr>
<td>Vascular invasion</td>
<td></td>
</tr>
<tr>
<td><strong>Cell cycle and proliferation</strong></td>
<td></td>
</tr>
<tr>
<td>Mitotic count/Mitotic index</td>
<td></td>
</tr>
<tr>
<td>Ki-67/MIB-1</td>
<td></td>
</tr>
<tr>
<td>DNA S-phase fraction (flow cytometry)</td>
<td></td>
</tr>
<tr>
<td>DNA/ploidy (flow cytometry)</td>
<td></td>
</tr>
<tr>
<td>Cyclin E</td>
<td></td>
</tr>
<tr>
<td><strong>Steroid Receptors</strong></td>
<td></td>
</tr>
<tr>
<td>ER/PR</td>
<td></td>
</tr>
<tr>
<td><strong>Growth factors and receptors</strong></td>
<td></td>
</tr>
<tr>
<td>HER2</td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td></td>
</tr>
<tr>
<td><strong>Tumor suppressor genes</strong></td>
<td></td>
</tr>
<tr>
<td>TP53</td>
<td></td>
</tr>
<tr>
<td><strong>Measures of invasiveness</strong></td>
<td></td>
</tr>
<tr>
<td>Cathepsin D</td>
<td></td>
</tr>
<tr>
<td>Plasminogen activator inhibitor-1 (PAI-1)</td>
<td></td>
</tr>
<tr>
<td>Urokinase plasminogen activator (uPA)</td>
<td></td>
</tr>
<tr>
<td>Laminin receptors</td>
<td></td>
</tr>
<tr>
<td><strong>Angiogenesis</strong></td>
<td></td>
</tr>
<tr>
<td>MVD</td>
<td></td>
</tr>
<tr>
<td>VEGF</td>
<td></td>
</tr>
<tr>
<td><strong>Multiparameter gene expression analysis</strong></td>
<td></td>
</tr>
<tr>
<td>Oncotype DX assay</td>
<td></td>
</tr>
<tr>
<td>MammaPrint</td>
<td></td>
</tr>
<tr>
<td>Rotterdam signature</td>
<td></td>
</tr>
<tr>
<td>Breast Cancer Gene Expression ratio</td>
<td></td>
</tr>
<tr>
<td><strong>Composite prognostic factors</strong></td>
<td></td>
</tr>
<tr>
<td>Nottingham Prognostic Index</td>
<td></td>
</tr>
<tr>
<td>TNM and pTNM classification</td>
<td></td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
</tr>
<tr>
<td>Tumor necrosis</td>
<td></td>
</tr>
<tr>
<td>Stromal fibrosis /elastosis</td>
<td></td>
</tr>
<tr>
<td>Basal-like phenotype</td>
<td></td>
</tr>
<tr>
<td>Triple negativity</td>
<td></td>
</tr>
<tr>
<td>Stem cell markers</td>
<td></td>
</tr>
</tbody>
</table>
**Clinical factors**

Age at diagnosis is one of the useful prognostic indicator in breast cancer.\textsuperscript{42} Several independent studies have shown that young breast cancer patients (≤35 years) have more aggressive biologic characteristics and poorer prognosis.\textsuperscript{276-278} Consequently, age (<35 years) is one of the parameters which was recommended by the St Gallen 2007 conference, used to determine the risk category of patients.\textsuperscript{279,280} On the other hand, the older patients (>70 years) also exhibit poor survival or higher mortality due to other factors.\textsuperscript{278} Interestingly, breast cancer in African and African-American women is diagnosed about 10-15 years earlier than in women from Caucasians populations.\textsuperscript{281,282}

Related to this, race and ethnicity is another patient-related factor that has been proposed as a prognostic marker although it is still a matter of debate, and numerous independent studies have shown that breast cancer in Africans and African-American has poorer prognosis than in Caucasians.\textsuperscript{33,282-286} Indeed, compared with Caucasian women, African-American women, regardless of age presented with higher histologic grade for each stage of breast cancer and tumor size above 1 cm in a study by Henson et al in 2003.\textsuperscript{284}

**Histopathologic factors**

Histologic grade is one of the most widely used prognostic factors. Using traditional morphologic features (tumor glands, nuclear pleomorphism and mitotic frequency), by careful examining of breast cancer specimens, can provide significant prognostic information required for therapeutic stratification. For accurate evaluation, good fixation and specimen preparation are very important in assessing these features. The traditional factors which are the most widely used prognostic markers and have the greatest value in clinical management of patients include; histologic type, histologic grade, tumor diameter, lymph node status, and vascular invasion,\textsuperscript{42,278,287,288} as well as distance to resection margins.
Assessment of tumor differentiation (histologic type, histologic grade) gives an indication of the underlying biology within a given tumor. The prognostic value of certain histologic types of invasive carcinoma has been well-established and may be grouped into four categories ranging from excellent to very poor prognosis.\textsuperscript{278, 288} However, in multivariate analysis, histologic grade is a more powerful prognostic factor than histologic type.

Several studies have shown significant correlations between histologic grade and survival of breast cancer patients,\textsuperscript{271} although a significant concern has been the reproducibility of grading. Currently, two grading systems are widely used, the Nottingham method (modified Scarff-Bloom-Richardson) and the Fisher nuclear grading method. The Nottingham system with its more objective criteria has good to excellent reproducibility when used by experienced pathologists,\textsuperscript{289, 290} hence, it is the most widely used and is currently recommended. It evaluates glandular differentiation, nuclear pleomorphism and mitotic counts ultimately generating three tumor grades.\textsuperscript{42, 291}

Tumor diameter is one of the strongest prognostic indicators even after 20 years of follow-up.\textsuperscript{278} Consequently, it has become an important quality assurance measure for breast screening programs. However, for its prognostic correlation it should be assessed on pathologic specimens (pathologic tumor size), and the greatest diameter is considered as the final tumor size.\textsuperscript{288}

Multiple studies have shown that histologically determined axillary lymph node status is one of the strongest independent prognostic factors in breast cancer.\textsuperscript{42, 275, 291} However, there is still some debate about the use of axillary clearance or sentinel lymph node biopsy\textsuperscript{291} although the latter is frequently used. At a recent St Gallen meeting (2009), the use of sentinel node biopsy was considered as standard care for patients with clinically negative axilla.\textsuperscript{292} Nevertheless, it is generally recommended that, in order to obtain accurate histologic evaluation of lymph node stage, several blocks from each node submitted for examination
should be examined.\textsuperscript{275, 291, 293} The St Gallen conference 2005 identified nodal status including sentinel node status as the most important feature for defining risk category in patient with breast cancer which was reaffirmed in a subsequent meeting.\textsuperscript{280, 294} The absolute number of nodes involved is useful in determining the thresholds for treatment modalities in the same group of patients.\textsuperscript{292}

Closely related to lymph node status is vascular invasion. Presence of vascular invasion correlates closely with lymph node involvement, and it has been suggested as a surrogate for lymph node status in cases where nodes have not been removed for examination.\textsuperscript{291} Vascular invasion is a powerful predictor of local recurrence following surgery and a prognostic factor for reduced overall survival. It has been recommended that vascular invasion should be assessed in routinely processed tissue with extra care to avoid artifacts of retraction spaces.\textsuperscript{275, 291} Interobserver variability about the topographical patterns of vascular invasion still exists\textsuperscript{275} and is a matter of debate.

Regarding the composite prognostic factors, the UICC TNM classification which evaluates the primary tumor size (T), regional lymph node status (N) and presence of distant metastasis (M) is commonly used for breast cancer patients at the time of diagnosis (Figure 13). The tumor stage at the time of diagnosis is one of the strongest prognostic factors in breast cancer. The pTNM classification requires examination of the primary cancer with no gross tumor at the margins of resection and is similarly categorized as pT (corresponds to T category), pN and pM (corresponds to M category). The pN classification requires the resection and examination of at least the lower axillary (level 1) lymph nodes. However, sentinel nodes may be used, but even then the grading should be designated (sn) for sentinel node, for example pN1(sn).\textsuperscript{2}
Figure 13: TNM classification. The TNM classification is an anatomically based system that records the primary and regional nodal extent of the tumor and the absence or presence of distant metastases. T category describes the primary tumor site, N category describes the regional lymph node involvement, and M category describes the presence or absence of distant metastases.295

The Nottingham Prognostic Index (NPI), another composite parameter which includes nodal status, tumor diameter and histologic grade, is a strong prognostic assessment method, although inclusion of tumor biomarkers like ER status and HER2 expression to the NPI offers additional information about selection of patients for systemic adjuvant therapy.296

In 2007, prognostic factors such as histologic tumor grade, lymph node status, peritumoral vascular invasion (PVI), pathologic tumor size and patients age in addition to biomarkers were the criteria used to determine the risk category of patients by the St Gallen’s conference.294 Addition of proliferation assessed either by Ki-67 or mitotic count to the pathologic factors was later used to determine the algorithm for threshold of treatment modalities by the recent St Gallen’ conference 2009.292

48
Tumor biomarkers and biologic factors

Over the years, researchers have continued to identify and propose several biomarkers as putative prognostic and predictive factors in breast cancer (Table 3) that might help to better stratify patients to various treatment regimens as well as targeted therapies. These novel biomarkers reflect alterations in genes that regulate development and proliferation of tumors. However, three biomarkers (ER/PR and HER2 expression) have become standard measurements in the management of breast cancer patients. In addition, some factors have recently been recommended for clinical use. The uPA and PAI-1 which should be measured by ELISA may be used to determine prognosis in node negative breast cancer, although the St Gallen conference did not accept uPA/PAI-1 as a useful prognostic factor. In addition, the Oncotype Dx multiparameter gene expression analysis may be used to predict the risk of recurrence in patients with ER positive breast cancer who are treated with tamoxifen if readily available.

Estrogen receptor status is a widely applied factor that is used to predict response to hormonal therapy in both early and metastatic disease. Prediction of response can be refined further by combing ER and PR assays. Also, it has been suggested that absence of PR may indicate increased signaling of HER2 and may help clinicians decide between using aromatase inhibitor or tamoxifen. In addition, ER expression has been shown to predict the long-term outcome of hormonal therapy, and is associated with improved overall survival although its use as a prognostic factor is limited.

HER2 expression has been associated with poor prognosis in breast carcinoma including poor response to both chemotherapy and hormonal therapy. Studies have shown that HER2 status may be used to predict resistance to tamoxifen or cyclophosphamide-based therapy and enhanced response to anthracycline-based therapy in early breast cancer. However, its current clinical use is limited to predicting the response to HER2 targeted
therapy and selection of patients for treatment with trastuzumab. Other utilities for HER2 status are still undergoing further evaluation.

Gene expression studies have shown that proliferation is the most important component in many prognostic signatures. Cell proliferation plays a major role in the behavior of breast cancer, and increased proliferation correlates strongly with prognosis irrespective of the methodology used. Several methods of assessing proliferation have been studied (Table 3), but mitosis counting provides the most reproducible and independent prognostic information. Currently, it is recommended by the College of American Pathologists that assessment of cell proliferation should be performed routinely in evaluation of breast cancers, and mitotic figure counting might be sufficient enough for this purpose. It was found to be the most important prognostic component of the Nottingham grading system, and the mitotic activity index (MAI) was validated as the strongest independent and well reproducible prognosticator in lymph node negative patients. Assessment of other proliferation markers such as Ki-67 is currently optional.

Ki-67 is a nuclear antigen that was identified by Gerdes et al. (1991) and is expressed only in the proliferative phases of the cell cycle G1, S, G2 and M but absent in G0. It can be used to stratify patients into good or poor prognostic groups. However, results from different groups are still conflicting and therefore its use in routine clinical management of breast cancer is still undetermined. The Norwegian Breast Cancer Group has recently (February 2010) recommended its use in subgroups of breast cancer. The St Gallen conference in 2009 considered Ki-67-labelling index a useful factor that could be used to indicate the potential value of adding adjuvant chemotherapy in patients with receptor positive disease.

For the TP53 gene, mutation status and gene expression profiles have been suggested as powerful prognostic markers in breast cancer. In addition, p53 expression has
been associated with poor prognostic factors and poor survival,\textsuperscript{308, 309} and may be a prognostic marker in nodal negative breast cancer patients.\textsuperscript{42, 275, 310} In addition, it can help to identify patients likely to response to chemotherapy or radiotherapy,\textsuperscript{42, 275, 310} but its use as a prognostic or predictive factor is still controversial. Present data are insufficient to recommend clinical use of p53 in breast cancer patients. Moreover, the IHC detection of p53 expression is variable and, does not detect all \textit{TP53} mutations. A consensus on how to assess the staining has not yet occurred.\textsuperscript{42, 272, 275}

EGFR expression has been reported as one of the biomarkers which may be a candidate for clinical application in the near future.\textsuperscript{94} Several independent studies have shown that EGFR expression in breast cancer is associated with features of aggressive tumors and poor response to tamoxifen.\textsuperscript{42, 311} Present data suggest that some patient groups with breast cancer could benefit from EGFR-targeted therapy.\textsuperscript{94, 258, 312} EGFR status might possibly have a predictive role for response to such therapy,\textsuperscript{313} although detection and interpretation of EGFR is controversial and still needs to be standardized.\textsuperscript{94, 311} Indeed, results from a trial in which patients were treated on the basis of EGFR expression are promising.\textsuperscript{94, 314} On the contrary, different studies have provided conflicting results concerning the prognostic and predictive significance of EGFR, and its routine value in clinical management of breast cancer patients is still undetermined.\textsuperscript{42, 311, 315}

Growth, invasion and metastasis of breast carcinoma depend on angiogenesis, and thus tumor-associated angiogenesis has attracted much attention and has been extensively studied as a possible prognostic or even predictive factor in breast cancer.\textsuperscript{271, 316, 317} Several independent studies, although not all, have proposed that tumor angiogenesis is an independent prognostic factor and is associated with the risk of distant metastases and poor survival.\textsuperscript{316, 318-320} However, the prognostic significance of angiogenesis remains somewhat controversial. This is mainly due to the variability in measurement of angiogenesis by
assessing the microvessel density (MVD) as a surrogate marker of the degree of angiogenesis.\textsuperscript{139, 275} The assessment of MVD within a selected tumor area (hot spot) is too variable to be clinically useful.\textsuperscript{317, 319, 321}

Another surrogate marker of angiogenesis with prognostic significance is VEGF expression, which was also found to correlate with MVD in some studies.\textsuperscript{318} VEGF has been reported as an independent marker of poor prognosis in some studies,\textsuperscript{42, 308, 322} and high expression of VEGF can identify a subgroup of patients who may benefit from selective anti-angiogenic therapy.\textsuperscript{42, 153} However, the clinical use of VEGF is still undetermined.

In addition to the above factors, other parameters which might become of clinical importance in breast cancer include but are not limited to the triple negative phenotype, the basal-like subtype and cancer stem cell markers.\textsuperscript{292, 294} The triple negative phenotype (TNP) is characterized by lack of ER/PR/HER2 and has recently been recognized as a group with therapeutic implications. It lacks targeted therapy whereas it is frequently resistant to standard chemotherapeutic regimens.\textsuperscript{248} Multiple studies have reported the poor prognosis associated with TNP,\textsuperscript{323} although the use of TNP as a prognostic factor is still not well studied.\textsuperscript{248} Further, molecular predictive signatures will enable characterization of triple negative breast cancers better and design of optimal treatment modalities.\textsuperscript{324} Indeed, the recent St Gallen’s conference (2009)\textsuperscript{292} recommended the use of triple negativity as a parameter that can be used to select some patients for chemotherapy.

Similarly, the basal-like phenotype which partially overlaps with the TNP and \textit{BRCA1} associated breast cancer,\textsuperscript{65, 248} has a poor prognosis with clinical importance.\textsuperscript{248, 325} It is associated with the shortest relapse-free and overall survival,\textsuperscript{226, 230} although its use as a prognostic factor is still not well studied.\textsuperscript{305} It has been suggested that pathologists should routinely identify the BLP in breast cancer,\textsuperscript{325} and incorporate the specific morphological features associated with the basal-like tumors (Table 2) with standard biomarkers such as
ER/PR and HER2 which might aid clinicians in developing optimal therapeutic strategies for this group.248

The presence of cancer stem cells might have prognostic and therapeutic implications in breast cancer.191, 193, 194 Studies in cell lines have indicated that cancer stem cells have a drug-resistant phenotype,326, 327 and express drug-resistance proteins such as ABCG2 (breast cancer resistance protein). In addition, breast cancer stem cells displayed resistance to both radiotherapy and tamoxifen treatment at clinically relevant doses.191, 328, 329 Further, it has been found that breast cancer stem cells metastasize to the bone marrow in early-stage breast cancer.330 Hence, it might be difficult to eradicate such drug-resistant stem-like cells from the bone marrow of patients using traditional chemotherapy only. The ability to identify cancer stem cells has facilitated the elucidation of pathways that regulate their growth and survival,190, 191 and has provided a deeper understanding of the natural evolution of cancer as well as clinical behavior and response to treatment.

Studies in vitro and in xenografts have indicated that breast cancer stem cells display distinct molecular signatures; the 413-gene CSC signature203 and the hNMSC signature,184 which might be potential prognostic parameters. The hNMSC signature could be used to predict biologic and molecular features of breast cancer.184 Stem cell markers like ALDH1 expression have been associated with poor prognosis,193, 194 and studies have also suggested that levels of ALDH1 expression in primary breast cancer can be used to predict the response to chemotherapy.193, 222, 331 Also, in the inflammatory breast cancer ALDH1 expression was found to be an independent prognostic marker that can predict metastasis and poor patient outcome,223 whereas in pancreatic adenocarcinoma ALDH1 expression was associated with a worse survival.332 However, the clinical significance of breast cancer stem cells and stem cell markers is still undetermined.
DIAGNOSIS AND TREATMENT

Optimal breast cancer management requires a multidisciplinary or interdisciplinary team approach involving surgeons, oncologists, radiologists, pathologists, geneticists, and possibly psychosocial specialists. Accurate diagnosis is a necessary step in the management of breast cancer and ideally every patient should have a pathologic diagnosis of breast cancer before definitive treatment can be given. Accurate diagnosis confirms the presence or absence of breast cancer and thus, avoids unnecessary treatment in patients with benign conditions, and in addition provides prognostic and predictive features of the cancer, which help in planning treatment and counseling of the patient with breast cancer.

A fundamental principle in evaluation of breast cancer patients is the triple-test diagnosis of breast masses which has been identified as a critical practice in diagnosing breast cancer. The triple test entails a correlation of clinical, pathologic and imaging findings. Regarding pathologic diagnosis, fine needle aspiration biopsy (FNAB) has been recognized as the most cost-effective procedure with short turnaround time, although core needle biopsy and standard surgical biopsy might be used. However, the choice among these three is influenced by availability of the tools and expertise in a limited-resource setting.

In addition to diagnosis, staging of breast cancer, to determine the extent of a disease is necessary for proper breast cancer management as well as providing useful information about the current status of cancer detection and management, and the success of breast health programs. The TNM classification of breast cancer is widely used and has been recommended by a number of regulatory bodies including the UICC among others. Unfortunately, in limited-resource countries, breast cancer is commonly diagnosed at late stages and is therefore characterized by high mortality. Therefore, breast cancer staging in such countries could provide revealing epidemiological information about opportunities for initiating or improving breast health care programs.
Regarding treatment of breast cancer, conservative surgery is currently being promoted. According to the St. Gallen conference in 2009, the use of surgical procedures developed to allow a wide excision with satisfactory results (oncoplastic surgery) were considered standard. The sentinel node biopsy was identified as standard of care for patients with clinically negative axilla; axillary node dissection could be avoided in all patients with a negative sentinel node, and in selected patients with micrometastasis in the sentinel node. However, the definition of adequate surgical margin remains controversial and no detailed specific recommendation on this matter was given.

Other treatment modalities for breast cancer include use of radiation therapy, endocrine therapy, anti-HER2 therapy and the cytotoxic chemotherapy. The St Gallen conference 2009 recommended radiation therapy after local excision of ductal carcinoma in situ (DCIS) as standard and postmastectomy radiation in invasive cancer, for women with four or more axillary lymph node involved. It could also be used in particularly young patients with one to three nodes and in those with poor prognostic features. Generally, it should be avoided in elderly patients and those with low-grade DCIS and clearly negative margins. In addition, the recommendations of the American Society for Clinical Oncology or the European Society of Mastology may be used to guide radiation treatment choice.

For endocrine therapy, anti-HER2 therapy and the cytotoxic chemotherapy, the St Gallen 2009 conference gave a detailed algorithm for the thresholds for these treatment modalities, although adherence to the therapeutic guidelines is greatly affected by the resources available in various geographic settings. Briefly, it recommended endocrine therapy for all patients whose tumors show any presence of ER, anti-HER2 therapy for patients with HER2 positive disease, and chemotherapy for patients receiving anti-HER2 therapy and as the mainstay of adjuvant treatment of most patients with triple negative...
tumors. However, the threshold for use of cytotoxic chemotherapy for some tumor groups was recognized as the most difficult to define.\(^2\)

Briefly, the Breast Cancer Guidelines for Uganda\(^3\) recommended breast self-examination as a way of early detection, the triple assessment approach for diagnosis, the TNM classification for staging and surgery as the mainstay of treatment for breast cancer except in metastatic disease. Also, tumor-free margin should not be less than 10 mm “at surgery”, adjuvant radiation therapy after surgery for most patients groups and adjuvant systemic treatment (chemotherapy and hormonal) for all patients in Uganda, with few exceptions (DCIS, lobular carcinoma \textit{in situ} (LCIS) and Paget’s disease) were recommended. Hormonal therapy alone was recommended for LCIS. Neoadjuvant chemotherapy could be given to down-stage the tumor before local treatment is offered.\(^3\)
BACKGROUND AND AIMS OF THE STUDY

Background

Over the years, numerous biomarkers have been proposed as putative prognostic and predictive factors in breast cancer that might help to stratify patients to various treatment regimens and targeted therapies.\textsuperscript{42, 272, 275, 305} Previous reports have described the prognostic biomarker profiles of breast cancer in African and African-American women,\textsuperscript{119, 282, 343} and striking similarities in breast cancer biology between the two groups have been reported.\textsuperscript{281} In Uganda, there is a paucity of reports about prognostic and molecular biomarkers in breast cancer, whereas a previous study analyzed HER2 oncoprotein expression in breast cancer.\textsuperscript{344} In this population, where the incidence of breast cancer is increasing, early diagnosis remains a challenge and the clinical outcome continues to be poor.\textsuperscript{19}

Consequently, in 2000, the Uganda Breast Cancer Working Group launched Breast Cancer Guidelines for the management of breast cancer.\textsuperscript{345} Its goal was to improve the quality of life of breast cancer patients and their families. Specifically, it aimed at standardizing and harmonizing diagnosis and treatment of breast cancer. In addition, one aim was to enable early detection of the disease with an ultimate goal of improving survival of the patients. The success of such a program needs to be augmented by studies specifically designed to elucidate the nature, behavior, basic processes and prognosis of breast cancer in this setting. Further, there was a need for a study to identify significant clinico-pathological parameters that might assist to achieve some of the objectives of the guidelines.
Aims of the study

General aim

On this background, the aim of this study was to explore the molecular markers in breast cancer with special focus on molecular subtypes, angiogenesis and stem cells in an African population.

Specific aims

Accordingly, the specific aims of the study included:

1. To explore the expression of selected basal-like markers in a series of breast cancers from native Ugandan women in the Kyadondo County, and to determine their frequency and relationship to other prognostic indicators.

2. To evaluate the expression of EGFR and c-kit in relation to the basal-like phenotype and other prognostic factors in breast cancers from an African population.

3. To explore the expression of candidate stem cell markers ALDH1 and BMI-1 in breast cancers from an African population and their associations with the basal-like phenotype (BLP) and other molecular markers.

4. To explore tumor-associated angiogenesis in relation to the basal-like phenotype, the triple negative phenotype and other tumor characteristics in an African population as well as a non-African population.
MATERIALS AND METHODS

STUDY SITES AND STUDY POPULATIONS

This study was carried out as part of the collaboration between the Department of Pathology at Makerere University College of Health Sciences (MUCHS) in Uganda and Section for Pathology, The Gade Institute, University of Bergen in Norway. Data collection was done at Makerere University while laboratory analysis and writing were done at The Gade Institute. The Department of Pathology at MUCHS is a research and teaching centre, for both undergraduate and postgraduate (average 2; range 1-5 per year) students. In addition, it offers histopathologic diagnostic services for the national referral hospital and the other main hospitals from all over the country which had an estimated population of about 32,4 million in 2009 (Figure 14).\textsuperscript{346}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{population.png}
\caption{The population of Uganda by age group as of 2009}
\end{figure}

Furthermore, it also houses the Kampala Cancer Registry (KCR), a population-based registry that was established in 1951 with an aim of determining incidence of cancers in the population of Kyadondo County.\textsuperscript{13} The Registry covers an area of about 1914 km\textsuperscript{2}, which
comprises of Kampala, the capital city of Uganda, and the neighboring urban and semi-urban areas,\textsuperscript{18} with an estimated population of 1.7 million (2002);\textsuperscript{347} the female population above 15 years old is about 530,000. The annual incidence of breast cancer is 22/100,000 in this population.\textsuperscript{13} The Baganda are the largest ethnic group in the county, but all the other ethnic groups are represented. Analysis of the breast cancer cases recorded at the Department of Pathology from 1990-2000 showed that cases from Kyadondo County contributed about 28% of the histologically confirmed breast cancers in females in Uganda. No significant differences were observed in distributions of female breast cancer patients by ethnic groups represented by regions as well as the age structure of cancer patients from Kyadondo County (mean age = 46 years) and those from other counties (mean age = 46 years) (P = 0.697).

The region is served with one 900 bed national referral hospital with attached oncology and radiotherapy units plus three other 100-bed missionary hospitals and hospice. Mammography is available at the national referral hospital but is limited to diagnostic purposes, and no routine screening is available.\textsuperscript{342} The study population mainly included female patients with breast cancer who were registered at the Kampala Cancer Registry or presented at the three missionary hospitals in Kyadondo County. The registry methods of collecting data and results have been previously described.\textsuperscript{348} A small number of study cases were drawn from the general population (39/192) (Figure 15).
In addition to the above population, a second study population included female Ashkenazi Jewish women in North America with breast cancer who were registered in the medical records at the Sir Mortimer B. Davis Jewish General Hospital, Montreal, Quebec, Canada.

PATIENT SERIES AND TISSUES

Series 1

This was the Series on which Paper I and Paper II were based. Cases of female patients with histologically confirmed breast carcinoma were consecutively compiled from the records of the Kampala Cancer Registry. Altogether, 120 cases from the period 1993-2002 were included whereby a total of 65 cases were identified from which suitable paraffin blocks were available and retrieved from the departmental archives and analyzed. Duplicated cases due to repeated biopsies and subsequent mastectomies were included only once. These blocks were originally obtained from mastectomy, incisional, excisional and core needle biopsy specimens that were submitted to the department. Twenty other cases with inadequate tissue available plus 35 cases which were untraceable were excluded.
Series 2

This series was the basis for Paper III and part of Paper IV (referred to as Series I in Paper IV). Cases of female patients with histologically confirmed breast carcinoma were consecutively compiled from the records at the Department of Pathology in addition to cases registered at the Kampala Cancer Registry. Altogether, 314 cases from the period 1990-2002 were included whereby a total of 192 cases (Table 4) with suitable retrieved paraffin blocks were eventually identified. Altogether, 122 (39%) other cases were excluded. These included cases with inadequate tissue available; 24 metastases, 11 ductal carcinoma in situ, 7 benign, 26 with insufficient material, 1 poorly preserved sample, 18 cases where no tumor tissue was identified and 35 cases where the tissue blocks could not be located.

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>37</td>
<td>19.3</td>
</tr>
<tr>
<td>1991</td>
<td>37</td>
<td>19.3</td>
</tr>
<tr>
<td>1992</td>
<td>56</td>
<td>29.2</td>
</tr>
<tr>
<td>1993</td>
<td>8</td>
<td>4.2</td>
</tr>
<tr>
<td>1994</td>
<td>5</td>
<td>2.6</td>
</tr>
<tr>
<td>1995</td>
<td>14</td>
<td>7.3</td>
</tr>
<tr>
<td>1996</td>
<td>6</td>
<td>3.1</td>
</tr>
<tr>
<td>1998</td>
<td>5</td>
<td>2.6</td>
</tr>
<tr>
<td>1999</td>
<td>7</td>
<td>3.6</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>2.1</td>
</tr>
<tr>
<td>2001</td>
<td>7</td>
<td>3.6</td>
</tr>
<tr>
<td>2002</td>
<td>6</td>
<td>3.1</td>
</tr>
<tr>
<td>Total</td>
<td>192</td>
<td>100</td>
</tr>
</tbody>
</table>

Clinical information on the included cases was obtained from the histology request forms. The mean age was 46.2 years (range 18-80 years) for Series 2 (n=192 cases) and 49.8 years (range 27-89 years) for the Series 1 (n=65 cases). Duration of clinical symptoms as reported by a total of 127 patients at the time of presentation ranged from 0.5-108 months with an average of 17.1 months. The stage of the disease at the time of presentation was
available in only 22 patients and the majority 12 (54.5%) were in stage 4, 8 (36.4%) were in stage 3. Tumor size was recorded in 31 cases and average size was 5 cm (range 1-12 cm).

The histologic type and grade of tumor were available in 181/192 (94%) and 107/192 (56%) cases respectively, however these were not according to the recommended criteria. Therefore, all cases were histologically re-typed according to the World Health Organization and re-graded in accordance with the Nottingham criteria. Nuclear grade and mitotic count was also recorded as separate variables according to the same criteria.

**Series 3**

This was part of the series on which Paper IV was based. For Paper IV, a second independent patient series (referred to as Series II in Paper IV) was included. Cases from the ethnically restricted single hospital-based retrospective cohort study as previously described were used in this series. The study was approved by the hospital’s Institutional Review Board. Briefly, the patients were consecutive cases of Ashkenazi Jewish women aged ≤ 65 years diagnosed with a primary, non-metastatic, invasive breast cancer during 1980-1995 at the Sir Mortimer B Davis-Jewish General Hospital, Montreal, Quebec, Canada. In total, 239 cases were included for the analysis and 70 other cases were excluded due to unavailable or unsuitable material, repeated unsatisfactory staining, or inability to amplify DNA after several attempts. Among the 239 cases, 24 BRCA1 and 6 BRCA2 mutation carriers were included. All cases were re-examined histologically and typed according to World Health Organization and grading was performed in accordance with the Nottingham criteria. In all, 181 cases (76%) were treated by breast conserving therapy (BCT), while 58 cases (24%) received mastectomy. Adjuvant chemotherapy was given to 45% of the BCT cases and to 54% of those treated by mastectomy. Radiotherapy was given to 85% of the cases treated by BCT and 7% of those receiving mastectomy. Hormone therapy was given to 56% of hormone receptor
positive patients and 22% of receptor negative cases. The median follow-up time of those who did not die of breast cancer was 9.3 years (n=168), and there were 69 breast cancer related deaths during the follow-up period.

TISSUE MICROARRAY

Altogether, 192 archival tissues (Series 2) were assembled on tissue microarray (TMA) blocks according to Kononen et al. (1998). Representative tumor areas with the highest histologic tumor grade preferably at the periphery of the tumor were identified on H&E-stained slides, and tissue cylinders with a diameter of 1 mm were punched from selected areas of the donor blocks and mounted into the receipt paraffin blocks using a custom made precision instrument (Beecher Instruments, Silver Spring, MD). To account for intratumoral heterogeneity and to reduce the problem of drop-outs, a minimum of 3 tissue cores were punched from the selected areas. Five μm thick sections of the resulting TMA blocks were made by standard technique. Serial sections were stained with antibodies as shown in Table 5. Also, a total of 230 cases from Series 3 were available on tissue microarray constructed as previously described, and 4 μm pre-cut slides were made available for staining with EGFR antibody as described. Of these, 223 cases with previously registered results were used for further analysis.

IMMUNOHISTOCHEMICAL METHODS

For series 1 and 2, immunohistochemical staining was performed on 5 μm thick sections of both the conventional and TMA slides. After sections were deparaffinized in xylene and alcohols, heat induced epitope retrieval methods were used for all antibodies except EGFR where proteinase kinase pre-digestion was used. As shown in Table 4, the antigen retrieval time, the antibody dilution and incubation were optimized for each antibody.
used. The staining was performed in a Dako Autostainer for all antibodies except the BMI-1 antibody, a non-commercial monoclonal antibody that was kindly provided by Dr Arie P. Otte (The University of Amsterdam, The Netherlands). The main detection system was Dako EnVision+ enzyme labeled polymer with 3,3’-diaminobenzidine (DAB+) as chromogen. For BMI-1, the Catalyzed Signal Amplification II (CSA II) kit (Dako, K1497) which is a biotin-free tyramide signal amplification system was used. Hematoxylin was used as a counter stain. All incubations were done at room temperature. Cases of breast carcinoma (ER, PR CK5/6, P-cadherin, ALDH1 and BMI-1), colonic carcinoma (Ki-67 p53 and EGFR), prostate carcinoma (Factor VIII) and gastrointestinal stromal tumor (GIST) (c-kit) with known immunoreactivity for the respective markers were used as positive controls. Normal breast tissue was used as a control for p63. Replacing the primary antibody with buffer solution served as the negative control.

For evaluation of microvessel density and vascular proliferation, 5 μm conventional sections (Paper IV: Series 2; 192 cases; Series 3; 239 cases) were used. We employed the dual staining procedure with Factor VIII and Ki-67 for endothelial cell proliferation.350 Sections were incubated with a cocktail of polyclonal rabbit anti-human Factor VIII (A0082) and monoclonal mouse antihuman Ki-67 antigen (Table 5). A secondary goat anti-mouse antibody (Dako E0433) with streptavidin alkaline phosphatase (Lab Vision) and Ferangi Blue chromogen kit (Biocare Medical) (Series 2) or StreptABComplex/AP (Dako K0391) and Fast Blue (Series 3) was used for visualization of Ki-67. For, Factor VIII staining, visualization was achieved by using EnVision+ and AEC+ (3-amino-9-ethylcarbazole) in both series. No contrast staining was applied in this protocol.
<table>
<thead>
<tr>
<th>Antibody</th>
<th>Provider</th>
<th>Dilution</th>
<th>Antigen Retrieval</th>
<th>Incubation (minutes)</th>
<th>Detection system</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>Dako/1D5</td>
<td>1:50</td>
<td>MW: 15 min in TE9 buffer</td>
<td>30</td>
<td>EnVision</td>
</tr>
<tr>
<td>PR</td>
<td>Dako/ PgR 636</td>
<td>1:150</td>
<td>MW: 15 min in TE9 buffer</td>
<td>30</td>
<td>EnVision</td>
</tr>
<tr>
<td>HER2</td>
<td>Dako/ Polyclonal</td>
<td>1:500</td>
<td>MW: 15 min in TE9 buffer</td>
<td>60</td>
<td>EnVision</td>
</tr>
<tr>
<td>CK5/6</td>
<td>Dako/D5/16B4</td>
<td>1:200</td>
<td>MW: 15 min in TE9 buffer</td>
<td>30</td>
<td>EnVision</td>
</tr>
<tr>
<td>P-cadherin</td>
<td>BD/56</td>
<td>1:400</td>
<td>MW: 15 min in TE9 buffer</td>
<td>60</td>
<td>EnVision</td>
</tr>
<tr>
<td>Ki-67/MIB-1</td>
<td>Dako/ MIB-1</td>
<td>1:50</td>
<td>MW: 30 min in TE9 buffer</td>
<td>60</td>
<td>EnVision</td>
</tr>
<tr>
<td>EGFR</td>
<td>Zymed/31G7</td>
<td>1:30</td>
<td>Proteinase K</td>
<td>30</td>
<td>EnVision</td>
</tr>
<tr>
<td>c-kit</td>
<td>Dako/ Polyclonal</td>
<td>1:200</td>
<td>MW: 15 min in TE9 buffer</td>
<td>30</td>
<td>EnVision</td>
</tr>
<tr>
<td>p63</td>
<td>Dako/4A4</td>
<td>1:300</td>
<td>MW: 15 min in TE9 buffer</td>
<td>30</td>
<td>EnVision</td>
</tr>
<tr>
<td>p53</td>
<td>Dako/DO-7</td>
<td>1:1000</td>
<td>MW: 15 min in TE9 buffer</td>
<td>60</td>
<td>EnVision</td>
</tr>
<tr>
<td>BMI-1</td>
<td>Dr Otte/6C9</td>
<td>1:1</td>
<td>MW: 25 min in TE9 buffer</td>
<td>60</td>
<td>CSA-kit</td>
</tr>
<tr>
<td>ALDH1</td>
<td>BD/44</td>
<td>1:250</td>
<td>MW: Citrate buffer pH 6</td>
<td>60</td>
<td>EnVision</td>
</tr>
<tr>
<td>Factor VIII and Ki-67 (Series 2)</td>
<td>Dako/ Polyclonal and MIB-1</td>
<td>1:800 and 1:50</td>
<td>MW: 30 min in TE9 buffer</td>
<td>60</td>
<td>EnVision and AP/HRP</td>
</tr>
<tr>
<td>Factor VIII and Ki-67 (Series 3)</td>
<td>Dako/ Polyclonal and MIB-1</td>
<td>1:400 and 1:200</td>
<td>MW: 20 min in TE9 buffer</td>
<td>60</td>
<td>EnVision and AP/HRP</td>
</tr>
</tbody>
</table>

BD=BD Transduction, MW=microwave, TE9=Tri EDTA buffer pH=9, AP=Alkaline phosphatase, HRP=Horseradish peroxidase

**Evaluation of staining**

In Series 2, tumors (2.6%-4.7%) without interpretable cores because of insufficient tumor tissue were omitted from the analysis. In total, 183-187 could be evaluated for the various markers (Table 6). For all biomarkers, evaluation was done by qualitative and quantitative visual assessment, and criteria for evaluation is given in Table 6. For ALDH1, nuclear staining alone was considered non specific and was not included in the analysis. Tumors with any mild to strong staining in at least 10% of cells were considered as positive.
staining for ER, PR and c-kit. Regarding EGFR, tumors with weak to strong cell membrane staining, whether complete or incomplete, and observed in more than 1% of the tumor cells were considered positive in accordance to the Dako criteria.\textsuperscript{101,351}

**Staining index**

To evaluate p53, p63, CK 5/6, P-cadherin, ALDH1 and BMI-1 expressions, a staining index (SI) (values=0-9) was determined by multiplying the score for intensity of staining (none=0, weak=1, moderate=2, and strong=3) with the score for proportion of immunoreactive cells (<10%=1, 10%-50%=2, >50%=3).\textsuperscript{65,246,252,352,353} Cut-off points for the various markers, determined based on median or upper quartile SI for Series 2 in consideration of the frequency distribution curve, the size of the subgroups and the number of events in each subgroup are shown in Table 6.

**Evaluation of Ki-67 expression**

Ki-67 proliferative rate was determined as a proportion (%) of positively stained tumor cell nuclei out of 500 tumor cells examined at high power (x400) using an eyepiece grid. In total, 7 cases with fewer than 500 cells were counted (small tumors). The cut-off point for high tumor cell proliferation by Ki-67 expression was set at 15.4% (Series 1) and 20.0% (Series 2) based on the median values (Table 6).
<table>
<thead>
<tr>
<th>Marker</th>
<th>Interpretable cores</th>
<th>Staining pattern</th>
<th>Cut-off point</th>
<th>Positivity rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>187</td>
<td>Nuclear</td>
<td>10% positive</td>
<td>39</td>
</tr>
<tr>
<td>PR</td>
<td>187</td>
<td>Nuclear</td>
<td>10% positive</td>
<td>28</td>
</tr>
<tr>
<td>HER2</td>
<td>187</td>
<td>Cell membrane</td>
<td>HercepTest criteria\textsuperscript{354}</td>
<td>17</td>
</tr>
<tr>
<td>Ki-67</td>
<td>189</td>
<td>Nuclear</td>
<td>Median; (\geq 15.4%*) or (\geq 20.0%** = high)</td>
<td>51</td>
</tr>
<tr>
<td>p53</td>
<td>187</td>
<td>Nuclear</td>
<td>upper quartile (&gt;4 = \text{positive})\textsuperscript{353}</td>
<td>29</td>
</tr>
<tr>
<td>Cytokeratin 5/6</td>
<td>186</td>
<td>Cell membrane and cytoplasm</td>
<td>Median; 1-9 = positive</td>
<td>15</td>
</tr>
<tr>
<td>P-cadherin</td>
<td>187</td>
<td>Cell membrane and cytoplasm</td>
<td>Median; (&gt;3 = \text{positive})</td>
<td>27</td>
</tr>
<tr>
<td>EGFR</td>
<td>185</td>
<td>Cell membrane</td>
<td>(&gt;1%\text{ stained} = \text{positive}) (Dako criteria)\textsuperscript{351}</td>
<td>20</td>
</tr>
<tr>
<td>c-kit</td>
<td>186</td>
<td>Cytoplasm and/or cell membrane</td>
<td>Median; 2-9= positive</td>
<td>4</td>
</tr>
<tr>
<td>p63</td>
<td>187</td>
<td>Nuclear</td>
<td>Median; 3-9 = positive</td>
<td>17</td>
</tr>
<tr>
<td>ALDH1</td>
<td>183</td>
<td>Cytoplasm</td>
<td>Median; 1-9 = positive</td>
<td>48</td>
</tr>
<tr>
<td>BMI-1</td>
<td>186</td>
<td>Nuclear</td>
<td>Median; 1-9 = positive</td>
<td>25</td>
</tr>
</tbody>
</table>

Regarding Series 3, detailed information about staining methods of the various markers (ER, PR, HER2, p53 CK5/6, P-cadherin, and EGFR) and evaluation of the markers was available from previous publications.\textsuperscript{65, 101, 246, 356, 357} For EGFR, a total of 201 were evaluated for the EGFR staining in accordance to the Dako criteria (EGFR-DA)\textsuperscript{351} as previously described,\textsuperscript{101} 22 tumors (9.9\%) with uninterpretable cores on TMA were omitted.

**Evaluation of Microvessel Density (MVD)**

The average microvessel density (MVD) in a selected tumor area was assessed in accordance to previous studies.\textsuperscript{319, 350, 358, 359} Sections were first scanned at low magnifications.
(x50 and x100) to identify the most vascularized area (hot spot) of the tumor. Then 10 consecutive high power fields at x250 (field size: Series 2; 0.45 mm$^2$, Series 3; 0.42 mm$^2$) from the hot spot were examined, except in a few cases with small tumors where less than 10 fields were examined in Series 2 (n=18). All positively stained vessels (red) were counted including vessels without microlumina according to Weidner’s approach and clusters of endothelial red cells that were clearly separate from the adjacent microvessels were also counted. In cases with vascular nests (glomeruloid proliferations) or long winding vessels and branching vessels, individual lumina or segments were counted to account for increased angiogenic response. When no clear hot spot was identified, vessels were counted in the most cellular area of the tumor periphery. Areas close to necrosis were excluded in the counting. In Series 2, 11 (5.7%) poorly stained (weak staining or excessive background staining) tissues and 4 (2%) cases with insufficient tumor tissue were excluded leaving 177 cases for analysis, whereas 239 cases were fit for analysis in Series 3. The MVD was then determined as the average number of microvessels counted in the 10 fields expressed as microvessels per mm$^2$.

**Vascular proliferation**

Similarly, within the same fields as used for the MVD, the number of microvessels containing positive proliferating endothelial cells were counted. The dividing endothelial cells were recognized by showing distinct Factor VIII/Ki-67 co-expression; red cells with blue nuclei (Figure 16). Positive nuclei outside the endothelial cell layer or within the vessel lumen were not counted. The average number of vessels with proliferating endothelial cells per mm$^2$ (pMVD) was determined and the Vascular Proliferation Index (VPI) was determined as the ratio of pMVD (mm$^2$) to the MVD (mm$^2$) given as a percentage.

The cut-off points for high MVD, high pMVD and high VPI were determined as 80.4/mm$^2$, 1.7/mm$^2$ and 3.1% for Series 2 and as 92.3/mm$^2$, 1.9/mm$^2$ and 3.3% for Series 3,
respectively based on the upper quartile values for the particular series. Thus, tumors that had a pMVD $\geq 1.7\text{mm}^2$ (Series 2) or $\geq 1.9\text{mm}^2$ (Series 3) were considered to have a high pMVD, and those with VPI $\geq 3.1\%$ (Series 2) or $\geq 3.3\%$ (Series 3) were considered to have a high VPI.

Figure 16: Microvessel containing positive proliferating endothelial cells as seen on dual staining x400. The dividing endothelial cells are recognized by distinct Factor VIII/Ki-67 co-expression; red cells with blue nuclei were counted at x250 HPF in 10 consecutive fields in a selected hot spot

Molecular phenotype sub-classification

There is no consensus on how to define different molecular subtypes of breast cancer by immunohistochemical markers, and overlapping categories exist. We used criteria based on current literature for sub-classification into molecular subtypes. In accordance with Carey et al. (2006), we defined the luminal A (ER+ and/or PR+, HER2−), luminal B (ER+ and/or PR+, HER2+), HER2+ subtype (ER−, PR−, HER2+) and the basal-like subtype (ER−, PR−, HER2−, CK 5/6+ and/or EGFR+) subgroups (Table 7). Tumors negative for all the 5 markers (ER, PR, HER2, CK 5/6 and EGFR) were considered as unclassified. This
definition for luminal B tumors does not identify all luminal B tumors because only 30% to 50% are HER2+ and the rest are classified with the luminal A. We therefore merged luminal A and luminal B into the luminal subtype. Further, in accordance with our previous studies, we included P-cadherin staining in some of the definitions of the basal-like phenotype.²⁴⁶,²⁵²

By using the Arnes et al. (2008)²⁰⁴ criteria, we defined the BLP profiles (Table 7) as follows: BLP1: concurrent ER−, HER2− and CK 5/6+; BLP2: concurrent ER−, HER2− and P-cadherin+; BLP3: concurrent ER−, HER2− and EGFR+; BLP4: concurrent ER−, HER2− and CK 5/6+ and/or EGFR+; BLP5: concurrent ER−, HER2− and positivity for one or more basal markers (CK 5/6, P-cadherin and EGFR). BLP4 is identical to the core basal phenotype (CBP) as defined by Nielsen et al. (2004)²³² and Tischkowitz et al. (2007).³⁶⁰

Regarding Papers I and II, tumors that expressed CK 5/6 and/or P-cadherin were considered to have a basal-like phenotype (BLP) in accordance to previous studies.²⁴⁶,²⁵²

**Table 7: Criteria for molecular subgroup classification of breast cancer and basal-like phenotypic definitions as used in this study**

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Biomarker-criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminal A</td>
<td>ER+/and/or PR+/HER−</td>
</tr>
<tr>
<td>Luminal B</td>
<td>ER+/and/or PR+/HER+</td>
</tr>
<tr>
<td>HER2 subtype</td>
<td>ER−/PR−/HER+</td>
</tr>
<tr>
<td>Basal-like subtype</td>
<td>ER−/PR−/HER−/CK5/6+ and/or EGFR+</td>
</tr>
<tr>
<td>Unclassified</td>
<td>ER−/PR−/HER−/CK5/6−/EGFR−</td>
</tr>
<tr>
<td>BLP1</td>
<td>ER−/HER−/CK5/6+</td>
</tr>
<tr>
<td>BLP2</td>
<td>ER−/HER−/P-cadherin+</td>
</tr>
<tr>
<td>BLP3</td>
<td>ER−/HER−/EGFR+</td>
</tr>
<tr>
<td>BLP4</td>
<td>ER−/HER−/CK5/6+ and/or EGFR+</td>
</tr>
<tr>
<td>BLP5</td>
<td>ER−/HER−/CK5/6+ and/or P-cadherin+ and/or EGFR+</td>
</tr>
</tbody>
</table>

**ETHICAL CONSIDERATIONS**

This study was approved by the Institutional Review Board at Makerere University College of Health Sciences.
STATISTICAL METHODS

Statistical analysis was performed using the SPSS 15.0 statistical package. We used the t-test or Mann-Whitney U test to compare continuous variables between different groups. We evaluated associations between categorical variables using the Pearson’s $\chi^2$ test or Fisher’s exact test. Differences between variables were considered statistically significant when the p-value for any statistical test used was <0.05.
MAIN FINDINGS

In Paper I, we found that the basal-like markers were expressed in 34% of the series of breast cancer from an African population, and they were significantly associated with features of aggressive tumors including high histologic grade, high nuclear grade, high mitotic count, and ER/PR negativity.

In Paper II, we focused on the expression of tyrosine kinase growth factors (EGFR and c-kit) in relation to basal-like breast carcinoma. We found a strong and significant association between EGFR and/or c-kit expression and the basal-like phenotype. In addition, EGFR and/or c-kit expression was significantly associated with poor prognostic features; histologic grade, nuclear grade, mitotic count, and ER−/PR−/HER2− (triple negativity).

In Paper III, we explored the expression of candidate breast cancer stem cell markers, ALDH1 and BMI-1 and their associations with the basal-like phenotype and other molecular characteristics. We found a high prevalence of ALDH1 expression in the series of breast carcinoma from an African population as well as a more extensive ALDH1 staining in cases that were positive, compared to Caucasian and Asian populations from the literature. Expression of ALDH1 was significantly associated with the basal-like phenotype and basal markers as well as features of aggressive tumors (high histologic and nuclear grades, high mitotic count, ER/PR negativity and p53 expression). On the other hand, BMI-1 expression was associated with good prognostic features, low histologic grade and ER positivity whereas it was inversely associated with ALDH1 staining.

In Paper IV, we determined vascular proliferation as a marker of tumor angiogenesis, and found that the basal-like subtype had increased tumor vascular proliferation compared to the luminal subtype in two independent breast cancer series. Also, we found that increased angiogenesis was associated with TNP, EGFR, p53 and p63 expression.
DISCUSSION

METHODOLOGICAL CONSIDERATIONS

Patient series

Series 1 and 2

This was a retrospective study with some limitations. The quality of data obtained depends on the accuracy of the records at the Kampala Cancer Registry. The registry methods of collecting data and results have been previously described.\textsuperscript{348} However, completeness and accuracy of data in studies done by cancer registries is of major concern particularly in Africa.\textsuperscript{18}

Nevertheless, the Kampala Cancer Registry as a population-based registry was established in 1951 with an aim of determining the incidence of cancers in the population of Kyadondo County,\textsuperscript{13} and it is one of the longest standing cancer registries in the African continent.\textsuperscript{18} Some of the cases included in Series 1 as well as Series 2 were part of the 174 breast cancer patients enrolled for the survival study at the KCR,\textsuperscript{18} which included incident cases diagnosed and registered between 1993-1997. Of these, 109 cases (63\%) were histologically or cytologically confirmed breast cancers. However, as indicated by Gondos et al. (2005),\textsuperscript{18} collection of follow-up data was particularly challenging, and a large number of patients could not be included in their study. For our study, reduced availability of archival tissue presented another limitation since a number of eligible cases were not evaluated; the 35 tissue blocks on record for the year 1997 were not available.

We eventually identified histologically confirmed breast carcinoma cases and were able to include 65 of 120 cases in Series 1 (1993-2002) whereas 192 of 314 cases were included for Series 2 (1990-2002) upon incorporating cases recorded at the departmental records. Altogether, 122 cases including the 35 tissue blocks on record for the year 1997 were not available and were not analyzed. Therefore, the unavailability of some archival tissues
could cause a selection bias. In addition, 31% (87 of 279) of the retrieved samples were not analyzed due to inappropriate tumor tissue.

A majority of samples used in Series 1 and 2 were from patients who presented to the national referral hospital, and this might also represent a selection bias. However, analysis of the records from 1990-2000 at the department revealed that the patients from Kyadondo County contributed about 28% of the breast cancer patients in the whole country. Therefore, our results might be an indication of the general population. In addition, analysis of breast cancer patients from the population of Kyadondo County compared with patients from the other counties revealed no significant difference in the age at diagnosis.

**Series 3**

This series is a single hospital and retrospective cohort series from a major teaching clinic in Montreal, Canada. The main objective in recruiting patients was to study the impact of \textit{BRCA}-mutations on breast cancer. Given that the frequency of \textit{BRCA}-mutations are rare in the general population, a restricted population was considered with regard to the feasibility of conducting large-scale genetic analysis as well as the clinical impact of the study. Ashkenazi Jewish women in North America have a well-known high frequency of \textit{BRCA}-mutations, and these can be attributed to a few dominant founder-mutations, highly simplifying the detection of mutations. For that reason, the study recruited only Ashkenazi women with an age restriction of <65 years based on the observation that breast cancers in older Ashkenazi women had similar frequency of \textit{BRCA}-mutations as the general population.

The study originally included 309 patients, of which 17 patients (5.5%) were immediately excluded due to the inability to locate tissue blocks, lack of invasive carcinoma in the available blocks or repeated inability to amplify DNA for mutation analysis, leaving 292 cases. Eventually, a total of 239 cases where available after excluding 4 cases that lacked
follow-up information in our data base and another 49 which lacked tissue blocks or had inadequate staining. Of these 230 cases were available on TMA blocks. Although this reduction in numbers may be considerable compared to the original series we consider the available cases as a random selection. No significant differences in the means of the included cases and the complete series were detected concerning tumor size, histologic grade, axillary nodal status or patient age at diagnosis (data not shown).

**Clinical-pathological variables**

We obtained much of the clinical information by carefully evaluating the histology request forms. The Department of Pathology at MUCHS has a vast number of archival tissue as well as accompanying histology request forms and reports. In collaboration with Centre for Disease Control and Prevention (CDC), all records available at the department up to 2000 were computerized. This made it easy to identify duplicate cases as well as identify more cases for Series 2 in comparison with the records at the KCR.

However, relative lack of clinical information was a limitation of this study. Accurate information about prognostic variables like tumor size was available in only a few cases (31/192; 16%), while the stage of the disease at time of diagnosis was also available in a limited number of cases (22/192; 12%). In addition, lack of outcome or follow-up information for most of the patients on record caused limitations in assessing the actual prognostic significance of the studied variables.

**Use of archival tissue**

The main advantage in using archival tissue for research lies in the availability of large tissue archives in pathology laboratories as well as long follow-up of a large patient series. This presents an invaluable tool for research and it has been demonstrated that most proteins
retain their antigenicity for more than 60 years. This study is based on material from the University teaching hospital in Uganda and a major Canadian research clinic originating from 1980. However, variations in tissue handling including fixation time and methods, sampling techniques and storage conditions are some of the major disadvantages as these would reduce immunoreactivity of some antigens. The large archival repositories have been supplemented by the use of tissue microarray.

**Tissue microarray (TMA)**

The use of tissue microarrays is an efficient approach in studies of most biomarkers and significantly reduces costs and time as well as conserving research tissues has facilitated the use of large archives. The tissue microarray technique has been validated in several studies since its introduction in 1998 and has been well established in our research group at The Gade Institute since 2000. After a few practice sessions, precision and accuracy was achieved and we used this technique to assemble the 192 cases (Series 2) for analysis. Similarly, 230 cases from Series 3 were assembled and made available to our group for staining. However, regarding evaluation of angiogenesis (Paper IV), conventional 5 μm thick tissue sections were used as well as in Paper I and II with 65 cases.

Further, we used pre-cut sections for both conventional slides and the TMA slides which were stored at −20°C, and recent studies have indicated that stored slides can still be valuable for research purposes in spite of the aging effects.

**Use of immunohistochemical methods**

Immunohistochemistry is widely used to study expression of specific proteins in human malignancies and is an invaluable tool in assessing prognostic and predicative biomarkers in breast cancer. Although formalin fixation retains tissue morphology, antigens
may be masked or even lost during fixation. Thus, the fixed form of proteins must be adequately retrieved to be recognized by antibodies. Antigen retrieval is one of the most important factors for achieving accurate and consistent results for biomarker studies by IHC. Consequently, different methods of antigen retrieval have evolved as an approach to standardize immunohistochemistry protocols for formalin fixed archival tissue, with enzymatic digestion and heat-induced retrieval being the best described.\textsuperscript{368} The requirements for optimal staining results vary with the choice of antibody as well as the antigen of interest, in terms of the pH of retrieval buffer, the heating time and the temperature. We used a test battery approach in staining some of the study cases as well as control tissues including different retrieval buffers with different boiling schemes, different antibody dilutions and different incubation periods.

However, antigen retrieval of the archival tissue presented one of the biggest challenges of this study. Whereas prolonged formalin fixation is rarely a problem in some institutes,\textsuperscript{369} it might be a major concern at our department at MUCHS. There is a considerable variation in the fixation times in formalin at our department (MUCHS) which might induce a bias towards prolonged fixation considering the years that were included. Moreover, the actual fixation time is difficult to determine. Studies have shown that prolonged formalin fixation might lead to decreased antigenicity of some antigens.\textsuperscript{362, 363} These factors might have contributed to the difficulty in antigen retrieval, although, this was overcome by determining the appropriate antigen retrieval times for each antibody under consideration\textsuperscript{362} as described above.

The choice of antibody clones used was done in consideration of the current literature for the particular antigen in question and considering widely used antibodies. We generally preferred antibodies that had been previously used by our group for consistency and adjusted for differences in reactivity between various tissues by modifying the staining protocols.
Accordingly, for the simultaneous dual staining we used Factor VIII which has been used in breast cancer and other cancers by multiple studies in our laboratory (The Gade Institute) with good results and is now well established,\textsuperscript{254, 350, 370-372} although CD31 and CD34 has been considered to be good alternatives.\textsuperscript{373} Previous reports have also shown that the staining results using the three antibodies (Factor VIII, CD31 and CD34) are comparable.\textsuperscript{374} In addition to these three widely used pan-endothelial cell marker, another antibody CD105 has been used in breast cancer and other tumors.\textsuperscript{375-378} CD105 is reported as a highly a specific marker for endothelial cells and has been shown to bind only to activated endothelial cell,\textsuperscript{375, 378} and may be better marker of angiogenesis in tumors\textsuperscript{378-380} although a number of studies in breast cancer have mainly used fresh or frozen tissue.\textsuperscript{375, 378, 381} Also, the non-commercial BMI-1 antibody (donated by Dr Arie Otte, The University of Amsterdam, The Netherlands) used in this study, has been used in various previous studies and was validated in our laboratory;\textsuperscript{382} Western blot analysis supported specificity and correlations with the commercial BMI-1 antibody. For candidate antibodies, the supplied data sheets from the manufacturers and literature studies provided information on their reactivities and were tested accordingly before the main staining.

Appropriate positive controls were selected from well-known sections that expressed the antigen of interest and were included for every round of staining. Negative control was achieved by omitting the primary antibody.

Regarding the detection method, whenever possible we used the EnVision system which on top of reducing the cost of staining, reduces the assay time and the workload.\textsuperscript{383} Also, we employed the Dako Autostainer which ensures equal staining conditions for all samples in every run, for all antibodies except the BMI-1 antibody.
Evaluation of staining: staining index

We employed the staining index (SI) (values=0-9) as the method of assessing immunohistochemical staining for most markers in the study. This is a semi-quantitative and subjective grading system which considers both the intensity of staining and proportion of cells stained, obtained as described in the methods section. This method of evaluation is now a well established robust and reproducible parameter and has been validated by several studies carried out in our team at The Gade Institute. 252, 382, 384-387

Selection of cut-off points used to create dichotomous variables was based mainly on median values and quartiles of the staining index, considering the frequency distribution curve, the size of the subgroups and the number of events in each subgroup. A selected cut-off point, once chosen, was then used consistently in all analyses involving that marker.

Estimation of microvessel density

Angiogenic activity is heterogeneous within a given tumor, and MVD assessment within the most vascularized selected tumor area (hot spot) has been the most widely used technique to quantify intratumoral angiogenesis in breast cancer, since its introduction in 1991. 319, 321 The major drawback, however, has been lack of standardization of the method used to assess MVD, 321 although the method used by Weidner et al. (1991) 319 has been generally accepted with some modifications. 350, 358, 359 The variability is mainly due to issues like choice of antibody and “hot spot” based counts versus “global” counts as well as automated versus manual counts. It is now generally agreed that the vessel counts should be in the hot spots preferably at the tumor periphery. 317 In accordance to previous studies at our institute, we used a double staining with Factor VIII/Ki-67 for endothelial cell proliferation, 350, 370 and we evaluated tumor angiogenesis in hot spots counting vessels in 10 fields at x250 magnification and eventually determined MVD per mm$^2$. MVD estimation by
Weidner’s method is fairly rapid, but is influenced by the training and experience of the investigator. Thus, after initial discussion of the criteria and some basic training, a training set (n=25) was scored twice for intra-observer variability with good results (κ=0.83). The inter-observer variability (HN, JBA) using the same set was also good (κ=0.64).

In addition, availability of a representative tissue block\textsuperscript{359} for estimation of MVD in the archival tissue depends on the initial tumor sampling technique that was done by both the clinician and at the department since it is not uncommon for the clinicians to submit just a small part of the tumor of a mastectomy specimen (personal observation from Uganda). However, previous studies have reported a concordance rate of 71-78% between different blocks sampled from the same tumor;\textsuperscript{388} thus, use of archival tissue did probably not greatly affect the estimated microvessel count.

**Vascular proliferation index**

Vascular proliferation determined by the number of vessels with evidence of dividing endothelial cells given by dual immunohistochemical staining may be a better indicator of angiogenesis than MVD which is more of a marker for the metabolic demand of tissue rather than angiogenesis.\textsuperscript{138} Studies from our laboratory have shown that counting vessels with actively dividing endothelial cells revealed significantly stronger associations with clinicopathologic phenotype and patient prognosis in endometrial and prostate cancers.\textsuperscript{350, 389} In this present study we accordingly determined vascular proliferation as a marker of angiogenesis.
**Molecular sub classification**

Determining the molecular subtype of breast cancer by immunohistochemical markers has some limitations.\(^{233}\) There is no consensus on how to define basal-like breast cancer and overlapping categories exist. Although a majority of the basal-like breast cancers are triple negative, ER or HER2 expression has been reported in about 15%-45% of the basal-like cancers.\(^{226, 232, 233}\) Therefore, the obvious limitation of using IHC to define subtypes is that it might result into misclassification bias of some tumors. In this study, we used criteria based on current literature.\(^{231, 244, 267}\) In this classification, the definition of basal-like breast cancer as ER−/PR−/HER−/CK5/6+ and/or EGFR+ encompasses the criteria used by Nielsen et al. (2004)\(^{232}\) which showed a sensitivity of 76% and specificity of 100% to identify the basal-like phenotype as defined by expression profiling analysis. In addition, we used individual basal markers for this subgroup such as cytokeratin 5/6, P-cadherin and EGFR as well as composite definitions such as basal-like subtype, various basal-like phenotypic profiles (BLP1-5) and the triple negative phenotype which overlaps significantly with the basal-like category (see Figure 12).

**Comparison with previous studies**

There is no uniformly accepted method for the registration of IHC staining that can be applied for most biomarkers. This presents a challenge in interpretation and comparison of results from different research groups, and this was a major source of concern according to the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK).\(^{390}\) Hence, the report encouraged transparency and complete reporting to facilitate usefulness and reproducibility of scientific data in the field of tumor biomarker research.
DISCUSSION OF MAIN FINDINGS

Molecular subtypes of breast cancer

Microarray studies have indicated that breast cancers may be divided into five major subtypes.\textsuperscript{225, 227} Defined by IHC staining, the luminal A subtype is the most prevalent contributing 47-69\% of all breast cancers. The other subtypes contribute 8-25\% (basal-like), 6-17\% (luminal B), 6-10\% (HER2 subtype) while the unclassified accounts for 1-7\%.\textsuperscript{231, 241, 244, 267} The normal breast-like subtype originally identified by gene profiling studies\textsuperscript{226} is still poorly characterized immunohistologically.\textsuperscript{229, 233} The basal-like and the HER2 subtypes are of particular interest because of their prognostic implications.\textsuperscript{226} The basal-like phenotype partially overlaps with the triple negative phenotype (ER\textsuperscript{−}/PR\textsuperscript{−}/HER2\textsuperscript{−}),\textsuperscript{360} which is another category of breast cancer that has attracted attention as an easily recognizable prognostic group with therapeutic implications.\textsuperscript{391} It contributes about 10-24\% of all cases.\textsuperscript{391, 392}

Studies have reported that both the basal-like and the TNP are overrepresented in young African-American and African women,\textsuperscript{119, 120, 231, 393, 394} and striking similarities have been reported in features of breast cancer in the two groups.\textsuperscript{281} Therefore, an important aspect of our studies was to determine the prevalence of the basal-like phenotype and its associations with other tumor characteristics with focus on breast cancer in an African population.

In agreement with these previous reports, we found a high prevalence (34\%) of the basal-like differentiation in the series of tumors from an African population (Paper I) using two individual basal markers CK5/6 and P-cadherin. This is comparable to what Carey et al. (2006)\textsuperscript{231} found (39\%) in premenopausal African-American women. Further, in Paper III, with more cases included, the prevalence of the basal-like subtype (22\%) as defined according to Carey et al’s definition, using composite criteria, was again comparable to the overall prevalence of BLP in African-American women (26\%) reported in that study. Studies in Caucasians have reported a prevalence of 8-16\%.\textsuperscript{231, 244, 267} In the same vein, we found a
higher prevalence of triple negative tumors (40-41%) (Papers II and III) compared to previous reports (10-21%) among Caucasians.\textsuperscript{392, 393, 395, 396} The high frequency of both the BLP and TNP in patients of an African population compared to Caucasians provides further support that breast cancer in women of African ancestry might be, to some extent biologically different from that in Caucasians.\textsuperscript{282, 343, 397-399} Indeed, a recent gene expression profile study has indicated that differences beyond the knowledge of current markers might exist in tumor biology of African-American compared to Caucasians.\textsuperscript{400} Moreover, breast cancer in the black and African-American women occurs at a much younger age (10-15 years younger on average) than in Caucasians.\textsuperscript{281, 282}

However, the higher frequency (40-41%) of TNP (Papers II and IV) in our study compared to some studies in African-American (21-25%)\textsuperscript{393, 394, 396} might indicate that other factors besides ethnical and racial background are important in breast cancer, although others reported comparable frequency (47%).\textsuperscript{395} Further, more comparative studies are required to answer this.

Base-like subtype and expression of EGFR and c-kit

We found a significant association (Paper II) between EGFR or EGFR and/or c-kit expression and the base-like category\textsuperscript{232, 233, 245, 360} as well with the triple negative phenotype\textsuperscript{401} in accordance with other studies. This provides more support to the stipulation that the base-like\textsuperscript{258} and triple negative tumors\textsuperscript{312} represent a group of breast cancers that might potentially benefit from EGFR-targeted therapies in addition to chemotherapy. Whether these tumors will actually respond to such therapies is yet to be known\textsuperscript{392, 402} from the results of on-going phase II clinical trials in the base-like subtype as well as triple negative tumors.\textsuperscript{403, 404} Studies in cell lines have shown that base-like cells are sensitive to EGFR targeted therapy in combination with chemotherapy\textsuperscript{405} and the two act synergistically. Basal-
like cells are more sensitive to growth inhibition by dasatinib, a multi-targeted kinase inhibitor, compared to the luminal cell lines.\textsuperscript{264}

Moreover, we (\textbf{Paper II}) and others\textsuperscript{232, 244} found that 23-31\% of basal-like breast cancers express c-kit, whereas about 50\% of BLP express either EGFR or c-kit, or both (\textbf{Paper II}). However, results from phase II trials with a potent inhibitor of tyrosine kinases including c-kit are still negative.\textsuperscript{406}

\textbf{Basal-like subtype and tumor-associated angiogenesis}

We found that tumor angiogenesis was increased in the basal-like tumors compared to the luminal subtype (\textbf{Paper IV}) in two independent breast cancer series. This is in agreement with a previous report which found that glomeruloid microvascular proliferation was significantly more frequent in the basal-like phenotype.\textsuperscript{253} In the present study, vascular proliferation was significantly associated with multiple basal markers (such as CK 5/6, P-cadherin and EGFR). To support this, gene expression studies have indicated that the majority of basal-like tumors express the activated wound-response signature, which represents important processes associated with angiogenesis among others.\textsuperscript{238}

Anti-angiogenesis treatment has now been approved for breast cancer and other tumor types,\textsuperscript{407} although it is presently not clear whether this treatment should be given to certain subgroups of malignant tumors.\textsuperscript{160} Our findings therefore suggest that anti-angiogenic therapy might be a possible target in the basal-like subgroup. Further clinical trials targeting different pathways will give more insight on this.\textsuperscript{160}

Similarly, the significant association between TNP and increased tumor angiogenesis which we found in our study (\textbf{Paper IV}), in accordance with previous reports,\textsuperscript{52, 320} supports the rational of phase II clinical trials that are currently accessing the potential benefit of triple negative tumors from a combination of anti-angiogenic therapy and chemotherapy.\textsuperscript{233} Current
studies should also address whether markers of BLP or TNP might be used to predict the response to anti-angiogenic therapy in breast cancer.

**Basal-like subtype and poor prognosis**

Basal-like breast cancer is an aggressive phenotype, but the underlying biology is still poorly understood. It has been reported to have a poor prognosis with a specific pattern of metastatic spread to sites associated with decreased survival, like hematogenous metastases. Several independent studies have established that angiogenesis plays a central role in tumor development and subsequent metastases. To speculate, increased angiogenesis in basal-like breast cancer might partly explain the frequent metastasis reported in this subgroup. In general, the prognostic significance of increased angiogenesis as determined by MVD has been confirmed in several independent studies.

Additionally, we found (Paper I) that the basal-like phenotype was associated with features of aggressive tumors in accordance with previous reports as well as EGFR expression (Paper II). Given that, both EGFR expression and increased angiogenesis are associated with poor prognosis, and our findings provide further evidence that the basal-like breast cancer exhibit multiple features of aggressive tumors which, in part, explains the poor prognosis.

**Cancer stem cells and the basal-like subtype**

Our results (Paper III) showed that the basal-like subtype was significantly associated with ALDH1 expression which has been associated with poor clinical outcome in breast cancer. Also ALDH1 expression was associated with the different BLP profiles, as well as with individual basal markers CK 5/6, P-cadherin and EGFR (Paper III), similar to what others have reported. In breast cancer mouse models, breast cancer stem cells which
expressed a combined CD44+/CD24-^low/ALDH1+ phenotype showed an especially high tumorigenic capacity.\(^{194}\)

Further, others have suggested that the proportion of cancer stem cells within breast tumors may correspond to the risk of distant metastases\(^ {185, 186}\) and the heterogeneous phenotypic and molecular traits of breast cancers are a function of their CSC content.\(^ {184}\) In the integrative model of breast cancer metastasis it has been proposed that oncogenic mutations which occur in the breast stem cells might generate ‘poor prognosis’ metastatic breast cancers.\(^ {181}\) In these cancers, the resulting breast cancer stem cells under the influence of stromal fibroblasts have the ability to metastasize, and variants of the cancer stem cells which express different tissue-specific profiles determine the tissue selectivity for metastasis.\(^ {181}\) The frequent hematogenous metastases as well as specific pattern of metastatic spread that have been reported in the basal-like subtype might relate to the significant association found between ALDH1 expression and the BLP. Moreover, we also found more extensive ALDH1 staining in cases that were positive, compared to reports among Caucasians and Asians populations.\(^ {194, 409}\) More studies are required to understand the biology of cancer stem cells in the development and clinical behavior of basal-like breast cancer.

ALDH1 expression has been associated with resistance to chemotherapy\(^ {193, 222}\) in breast cancers. Also, there is some evidence that the limitation of chemotherapy and radiation treatment may be associated with the inability to target breast cancer stem cells.\(^ {191, 202, 328, 329, 410}\) To speculate, our findings might be related to the aggressive behavior and therapy resistant features of the basal-like breast cancer subtype.\(^ {202, 226, 410, 411}\)
Cancer stem cell markers in African breast cancer

Our result indicate a higher prevalence of ALDH1 expression (48%) in a series of breast cancer from an African population (Paper III) compared to 19% and 30% in two different Caucasian populations\textsuperscript{194} and 10% and 19% in Asian populations.\textsuperscript{193, 409} We also found more extensive staining in positive cases.\textsuperscript{194, 409} Further, in comparison with data derived from breast tumors in Caucasian and Asian populations\textsuperscript{193, 194, 409} regarding ALDH1 positivity rate in tumors with similar characteristics (histologic grade, ER, HER2, Ki-67), we observed that tumors from our study (Paper III) stained in a higher percentage of cases in poor prognosis categories (such as high histologic grade, ER negative cases, and tumors with high Ki-67 expression). Thus, apart from methodological discrepancies, biologic differences might be present when comparing breast cancers from African and Caucasian populations.\textsuperscript{282, 284, 343, 397, 398, 400}

In line with this, a difference in the spectrum of tumor characteristics and prognostic features such as the presence of tumor necrosis, low ER positivity rate, high HER2 positive rate, p53 expression, overexpression (p16 and cyclin E) as well as low expression (cyclin D) of cell-cycle regulatory proteins and a high frequency of basal-like features have been reported in African and African-American patients when compared with breast cancers among Caucasians.\textsuperscript{32, 119, 231, 282, 396, 398} To speculate, our findings might indicate that poor prognosis of breast cancer in Africans and African-American is a preordained event. In support of this, a high prevalence of TNP in young premenopausal African-American was a contributory factor to the poor outcome which was reported in that group,\textsuperscript{394} whereas other independent studies have reported poor prognosis in African and African-American patients.\textsuperscript{19, 281} Indeed, a poorer outcome of breast cancer has been observed in the two populations compared to Caucasians.\textsuperscript{282, 285, 412}

Additionally, we found that ALDH1 expression was significantly associated with features of poor prognosis including high histologic grade, high nuclear grade, high mitotic
count, p53 expression and ER/PR negativity (Paper III). Also, ALDH1 expression was associated with a short duration of symptoms. Moreover, we found a significant association between ALDH1 expression and the triple negative tumors, a group whose poor prognosis has been widely reported.\textsuperscript{323} Our results provide further support that ALDH1 status might be an indicator of poor prognosis in breast cancer.\textsuperscript{194, 409}

To support this, BMI-1 expression, a candidate stem cell marker,\textsuperscript{204} which has been associated with features of good prognosis such as low grade, low mitotic count, ER positivity and absence of TNP in breast cancer in our study and in previous reports,\textsuperscript{204, 214, 413} was inversely associated with ALDH1 expression (Paper III). As expected, the frequency of BMI-1 was lower (25\%) in the series of breast cancers from an African population compared to breast cancer from Caucasian and Asian populations (43-62\%).\textsuperscript{204, 214} Others have found different results, BMI-1 being associated with poor prognosis.\textsuperscript{186, 214, 414} In addition, Glinsky et al. (2005)\textsuperscript{186} found that the expression of a BMI-1 driven 11 gene signature was associated with the risk of metastases in breast carcinoma. The reason of this inverse relationship is not known. To speculate, given that, BMI-1 has been implicated in breast carcinogenesis and is involved in stem cell activation and self-renewal,\textsuperscript{210} whereas ALDH1 is reported to have a role in early stem cell differentiation, proliferation and survival,\textsuperscript{215-217} this might suggest a regulatory relationship between ALDH1 and BMI-1 during carcinogenesis. Hence more studies are required to understand the regulatory pathways and biology of cancer stem cells in breast cancer carcinogenesis.

Finally, to speculate, the significant association between a basal-like phenotype and increased angiogenesis, EGFR expression or ALDH1 expression might indicate that patients with basal-like breast cancer may benefit from combined therapies targeting pathways involved in cell proliferation, angiogenesis as well as cancer stem cell activation and proliferation. These hypotheses still require further exploration.
CONCLUSIONS

1. The basal-like phenotype was frequent in the series of African breast cancer from the Kyadondo County in Uganda and is strongly associated with features of poor prognosis (Paper I).

2. There was a high frequency of tyrosine kinase growth factor (EGFR and c-kit) expression in basal-like breast carcinoma in the series of breast cancer from Uganda, and their expression was associated with features of aggressive tumors. (Paper II).

3. There was a high frequency of ALDH1 expression in the series of invasive breast carcinomas from Uganda which was significantly associated with a basal-like phenotype and with features of aggressive tumors (Paper III).

4. Tumor-associated angiogenesis was increased in basal-like breast cancer in two independent series of breast cancer and was associated with the triple negative phenotype, EGFR expression, p53, p63 and with features of aggressive tumors (Paper IV).

CONCLUDING REMARKS AND FUTURE PROSPECTS

Breast cancer is a highly heterogeneous disease in terms of tumor biology, clinical behavior, prognosis and response to treatment.\textsuperscript{45} It is believed that this is due to molecular differences even within histologically similar tumors.\textsuperscript{178, 224, 261} Interestingly, race appears to be one of the contributory factors to the final outcome.\textsuperscript{9, 33, 281, 286, 399, 415-417}

In line with the aim of this study, we have explored novel molecular markers that could be relevant for the understanding and management of breast cancers in African patients. We observed that breast cancer is occurring at a young age (mean 46 years) during the most productive period of the women. This might have implications on the socio-economic status of the families affected.\textsuperscript{342}
Additionally, we found a high frequency of poor prognosis features in the series of breast cancers from Ugandan women including a high prevalence of basal-like breast cancer, the TNP and high ALDH1 expression. These factors might influence the final clinical outcome of breast cancer in this population.\textsuperscript{18, 19} The prognostic significance of the high and extensive ALDH1 expression deserves further investigations.

Our results have indicated another potential therapy target\textsuperscript{258} for the basal-like subtype, since tumor angiogenesis was found to be increased in this category. Thus, well-controlled clinical trials of targeted therapy combining multiply pathways are urgently needed especially in a population where early diagnosis and effective treatment is especially challenging.\textsuperscript{19} However, standardization of the method used to assess tumor angiogenesis is still required.\textsuperscript{321}

Regarding immunohistochemistry, optimal antigen retrieval is a major concern in both research and the routine setting. The continuing need for routine assessment of hormonal receptors and HER2 to select high-risk patients and provide valuable information on treatment options for breast cancer patients in Uganda cannot be overemphasized. Fortunately, the establishment of routine IHC staining for breast cancer is under way, and personnel are available and have been trained. However, lack of adequate infrastructure is still a problem in some areas.

On the basis of the literature and the present study, we suggest the routine assessment of hormonal receptors and HER2 for treatment of breast cancer patients in African populations. The histologic evaluation of breast carcinoma specimens, including the fixation, should be standardized according to guidelines and specified in the pathology report. This will ensure an improved quality as a basis for better treatment. Regarding research, more clinico-pathologic studies are required to determine the significance of specific entities such as basal-like breast cancer and triple negative tumors in Uganda and how to best stratify patients
towards traditional treatment as well as novel targeted therapies. Other molecular markers in translational breast cancer research also need to be explored and validated in African populations. Ultimately, clinical studies should also be performed in this setting.

In a country like Uganda, with limited resources, and where breast cancer incidence is increasing rapidly, health policies have to be designed to promote early diagnosis in order to improve clinical outcome. National diagnosis and treatment guidelines for breast cancer in line with the WHO recommendations, considering the limited resources, will help promote equity of health care delivery. At the policy level, our results underscore the need to establish routine assessment of ER/PR and HER2 in breast cancer patients in addition to the currently used prognostic factors. This will improve the quality of medical care offered to patients by the health care system.
ERRATA

**Introduction:** Page 46, paragraph 2, line 6: “…experienced pathologists, 289, 290…” should read “…experienced pathologists, 289, 290, 291…”.

Page 47, paragraph 1, line 1: “…be examined. 275, 291, 293…” should read “…be examined. 271, 275, 293…”.

Page 47, paragraph 2, line 4: “…examination. 291…” should read “…examination. 271…”.

Page 47, paragraph 2, line 7: “…retraction spaces. 275, 291…” should read “…retraction spaces. 271, 275…” and paragraph 3, line 6-7: “…categorized as pT (corresponds to T category), pN and pM (corresponds to M category)…” should read “…categorized as pT; primary tumor (the pT categories correspond to the T categories), pN; regional lymph nodes and pM; distant metastasis (the pM categories correspond to the M categories)…”.

Page 55, paragraph 2, line 10: “…choice. 335…” should read “…choice. 294…”.

**Materials and methods:** Page 64, paragraph 3, line 4: “..Table 4..” should read “..Table 5..”.

**Discussion:** Page 77, paragraph 2, line 2: “…research tissues has facilitated…” should read “…research tissues and has facilitated…”.

Page 87, paragraph 2, line 11: “…BLP” should read “…BLP in our study (Paper III).”.

Page 88, paragraph 1, line 14: “…(cyclin D) of cell-cycle regulatory proteins…” should read “…of cell-cycle regulatory proteins (cyclin D)…”.

**Paper II:** Page 517, Statistical methods, paragraph 1, line 3: “…G^2…” should read “…X^2…”.

**Paper III:** Page 370, Patient series, paragraph 1, line 3: “…Cancer Registry at the…” should read “…Cancer Registry and at the…” and Page 371, Molecular subtypes, paragraph 1, line 8-9: “…(ER−, HER2− and CK 5/6 and/or EGFR+)…” should read “…(ER−, PR−, HER2− and CK 5/6 and/or EGFR+)…”.

**Paper IV:** Page 6, Patient series, paragraph 1, line 3: “…Cancer Registry at the…” should read “…Cancer Registry and at the…” and Page 16, Discussion, paragraph 1, line 3: “…Our previous study…” should read “…A previous study…”. 
REFERENCES


56. Frankish H. BRCA1 has a pivotal role in repairing DNA. *Lancet*. 2001;357:1678-1678.


Goldstein NS, Decker D, Severson D, et al. Molecular classification system identifies invasive breast carcinoma patients who are most likely and those who are least likely to achieve a complete pathologic response after neoadjuvant chemotherapy. *Cancer*. 2007;110:1687-1696.


393. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative


