U.PORTO

MESTRADO EM ONCOLOGIA

ESPECIALIZAÇÃO EM ONCOLOGIA MOLECULAR

Metastatic breast cancer: finding molecular biomarkers in liquid biopsies

Helena Estevão Pereira

M

2018

U.PORTO

Helena Estevão Pereira

Metastatic breast cancer: finding molecular biomarkers in liquid biopsies

M.ICBAS 2018

Metastatic breast cancer: finding molecular biomarkers in liquid biopsies

Helena Estevão Pereira

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Helena Sofia Casanova Estevão Pereira

Metastatic Breast Cancer: Finding Molecular Biomarkers in Liquid Biopsies

Dissertação de Candidatura ao grau de **Mestre em Oncologia** – Especialização em Oncologia Molecular submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto

Orientadora
Professora Doutora Carmen de Lurdes Fonseca Jerónimo
Professora Associada convidada com Agregação
Departamento de Patologia e Imunologia Molecular
Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto
Investigadora Auxiliar e Coordenadora do Grupo de Epigenética e Biologia do Cancro
Centro de Investigação
Instituto Português de Oncologia do Porto Francisco Gentil

Coorientadora
Professora Doutora Meriem Lamghari Moubarrad
Professora Afiliada
Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto
Investigadora Auxiliar e Coordenadora do Grupo de Circuitos Neuro-Esqueléticos
Instituto de Investigação e Inovação em Saúde

“Permanence, perseverance and persistence in spite of all

obstacles, discouragements, and impossibilities:

It is this, that in all things distinguishes the strong soul from the weak.”

Thomas Carlyle

Financial Support

This study was funded by a grant of the Research Centre of Portuguese Oncology Institute of Porto (PI 74-CI-IPOP-19-2016).

ACKNOWLEDGEMENTS

A concretização desta Dissertação de Mestrado significou o término de uma fase de imensurável valor no meu percurso académico. Em 2016, o meu desejo focava-se em enveredar por uma área tão multidisciplinar e dinâmica como a Oncologia. No desenrolar destes dois anos, o Mestrado em Oncologia não só me ofereceu o privilégio de aprofundar os meus conhecimentos, me equipou com ferramentas de reflexão fundamentais para uma análise crítica, como também me proporcionou um contacto direto com o quotidiano de um hospital especializado. Naturalmente, seria impensável não expressar o meu singelo agradecimento a todos os que contribuíram e estiveram presentes nesta etapa.

Em primeiro lugar, agradeço à Professora Carmen Jerónimo, na qualidade de Diretora do Mestrado em Oncologia, por ter aceite e selecionado a minha candidatura, permitindo-me iniciar este novo desafio. Como minha Orientadora, os motivos de agradecimento são infindáveis. Para além do voto de confiança depositado ao permitir a minha integração no Grupo de Epigenética e Biologia do Cancro, sem dúvida serei eternamente grata por cada palavra de incentivo no decorrer de todos os contratempos, cada minuto das incontáveis reuniões em que foram partilhados conselhos e críticas construtivas, cada ensinamento transmitido e cada sincera palavra de encorajamento para perseguir o meu sonho e nunca desistir. Serei sempre verdadeiramente grata por tudo.

À Professora Meriem Lamghari, na qualidade de coorientadora, por ter possibilitado esta colaboração com o INEB que potenciou a partilha de diversos pontos de vista e enriquecimento científico.

Ao Professor Rui Henrique, por todas as valiosas e pertinentes críticas, conselhos e exímio rigor científico partilhados em cada reunião de grupo, que despoletaram uma reflexão crítica e foram essenciais para o desenvolvimento deste trabalho.

Ao Professor Manuel Teixeira, na qualidade de Diretor do Centro de Investigação, por me receber e possibilitar a realização desta Dissertação neste Centro de Investigação.

Ao serviço de Anatomia Patológica, em particular à Técnica Paula Lopes, pela realização da imunohistoquímica necessária para a completa caracterização das séries incluídas neste trabalho e à Técnica Mariana Cantante pelos inumeráveis cortes, sem os quais não teria iniciado este trabalho, e extrema compreensão e disponibilidade.

Aos membros da Clínica da Mama, em especial à Doutora Susana Sousa por me ter recebido sempre que necessário e disponibilizado o seu tempo para me ajudar na seleção das doentes incluídas neste estudo.

A toda a equipa que diariamente trabalha na Central de Colheitas, em particular à Enfermeira Berta, pela amabilidade e disponibilidade com que me receberam. A vossa colaboração foi fulcral para a concretização deste trabalho.

A todas as Pacientes que contribuíram com o seu consentimento, às quais apresento a minha mais profunda consideração e agradecimento.

Ao Serviço de Epidemiologia, em especial ao Engenheiro Luís Antunes, pelo seu contributo na execução da análise estatística.

À Daniela Sousa e à Catarina Lourenço, membros do grupo de Circuitos Neuro-Esqueléticos do INEB. À Daniela, por estar sempre disponível para me receber, por toda a paciência, ajuda e partilha de opiniões, ideias e novas perspetivas. À Catarina, mini Kate, por me teres acompanhado nesta longa caminhada. Foi a realização deste Mestrado que transformou duas colegas de curso em duas verdadeiras amigas. Tenho a certeza que esta fase, em que tanto partilhámos, foi o início de uma longa amizade.

À Francisca Dias, membro do grupo de Oncologia Molecular e Patologia Viral do CI-IPO-Porto, pela amabilidade com que partilhou o seu conhecimento, por toda a disponibilidade e paciência com que sempre me recebeu.

Aos meus colegas do Grupo de Epigenética e Biologia do Cancro, que me acolheram e em diversos níveis contribuíram para a concretização deste trabalho. A vossa boa disposição proporcionou uma admirável leveza a cada dia de trabalho, muito obrigada.

À Maria, são tantas as razões para agradecer. Mesmo tendo eu chegado ao grupo na altura mais decisiva e atarefada do teu percurso, acolheste-me de uma forma impressionante. A maneira como respondeste às minhas mil perguntas, me ensinaste cada técnica, me motivaste e te disponibilizaste para me auxiliar em todos os passos demonstrou não só a excelente profissional, mas a genuína e fantástica pessoa que és. Muito obrigada.

Aos mais velhos, Vera, David, Ângela, Dani, Ana Laura, Bárbara e Catarina, por todos os ensinamentos indispensáveis, sugestões pertinentes e momentos partilhados ao longo deste ano.

À Catarina, Sandra e Lameirinhas, porque compartilhámos todos os dias deste percurso, desde as aulas do Mestrado às incontáveis horas de trabalho no laboratório resguardadas por muitos risos, brincadeiras e companheirismo. À Catarina, por me

permitires ocupar cada pedacinho desocupado da nossa mesa, mesmo quando ultrapassava aquele limite por outros considerado aceitável. À Sandra, pela companhia numa longa jornada, desde a Licenciatura em Bioquímica até ao término desta etapa tão enriquecedora. Foram incontáveis os momentos que partilhamos, de vitória e de frustração, os pequenos almoços antes dos exames, que acabaram por se tornar o nosso amuleto da sorte, os debates e partilha de ideias, a entreajuda e as brincadeiras. À Lameirinhas, porque tudo começou com espirituosos entendimentos e desentendimentos por espaço que culminaram numa amizade com tantos momentos partilhados. Lala, foram muitas horas de trabalho, muitas conversas, algumas preenchidas de sentido e outras sem sentido algum, muitas trocas de ideias e "conspirações", muitos risos incontrolados, que alimentaram o nosso companheirismo e possibilitaram um livro repleto de histórias nossas.

À Joana e à Inês, apesar de temporária, a vossa presença no laboratório também deixou uma marca. Não só por toda a companhia e ajuda, mas por todo um variadíssimo leque de momentos divertidos e caricatos.

Às mais novinhas, Verita, Ritinha, Cláudia, Mariana, Isabel e Catarina, porque a vossa chegada foi uma lufada de ar fresco. O facto de ter partilhado alguns valiosos ensinamentos que no início deste percurso me foram incutidos, e tentado transmitir alguma calma e motivação no início desta vossa etapa, permitiu uma realização pessoal imensurável.

À D. Marta, agradeço pela constante alegria e disponibilidade, o carinho com que todos os dias me cumprimentava, pelas nossas conversas matinais, pequenas brincadeiras e apoio.

Quanto à pequena família que encontrei no laboratório… simplesmente não há palavras suficientes para vos agradecer por tudo.

À Sara, por partilhares comigo aqueles intervalos no mundo da distração que resultaram em momentos de muita risada, os "baldinhos" de café à hora do almoço, a necessidade de bolo de chocolate, as tardes de fim de semana repletas de frustração e desespero por resultados positivos, mas enriquecidas com conversas leves e risos genuínos, e claro, por me teres presenteado com o mais recente e inovador *housekeeping*.

Ao Dr. João Lobo, interno no Serviço de Anatomia Patológica, pelo contributo na seleção de todos casos e esclarecimento de todas as dúvidas. Mas especialmente ao João, porque ao longo desta jornada desempenhaste o papel de irmão mais velho nesta pequena família. Obrigada por toda a disponibilidade, paciência, conselhos, conversas,

brincadeiras, momentos partilhados em que ouvíamos e "sentíamos mentalmente" os raspanetes da Sofs por irrevogavelmente sermos adoráveis e angelicais.

Ao Diogo, insubstituível técnico Didi, por todos os preciosos momentos. Porque no início de uma tempestade, as tuas brincadeiras, a tua capacidade de me atribuíres inovadoras e carinhosas alcunhas que ligeiramente me provocam, mas alegravam, os teus abraços sinceros, foram um porto seguro. Talvez tenham sido as nossas semelhanças, ou simplesmente a genuína pessoa que és, que permitiram que um entendimento e amizade florescesse de forma tão inata, formando um laço que indubitavelmente permanecerá intemporal.

À Sofs, porque no meio de tantas razões, difícil é escolher as mais marcantes. Curioso como o universo funciona… relembrar que tudo surgiu de um ligeiro *bullying* ocasional. Ao longo deste curto, mas longo percurso, entre as nossas finitas semelhanças e incontáveis diferenças, alcançamos um equilíbrio que despoletou uma verdadeira amizade e uma sensação familiar. Dia após dia, aprendemos a tolerar os nossos dias maus, respeitamos os nossos defeitos, partilhamos as nossas qualidades e concluímos brilhantemente que em muitos casos o melhor era concordar em discordar. Foram tantos os momentos, tamanha a partilha e infinito o apoio, que alcançamos uma sólida sintonia que perpetuará no tempo e no espaço. Sem dúvida não são suficientes as palavras para expressar o valor do teu papel nesta jornada e importância da tua amizade.

À tia Carol, que apesar de não ser parte integrante deste grupo, é um membro indispensável desta pequena família. Obrigada pelas visitas, pelas brincadeiras, pelas reuniões de família e por sempre me relembrares que sou a bebé.

Ao meu pequeno eterno grupo de amigas, porque sem o vosso constante e ilimitado apoio, tudo teria sido mais difícil. À Nádia, à Águeda e à Kelly, por todos os jantares repletos de brincadeiras de criança, conversas sem sentido e gargalhadas genuínas. À Catarina, por todo o apoio ao longo de todos estes anos e sempre teres acreditado que seria capaz de alcançar os meus objetivos. À Bruna, a minha parceira de crime, por todas as conversas profundas e sinceras, por todas as palavras de incentivo, todos os momentos de diversão e conspiração, que por vezes testaram a paciência do André e o assustaram ligeiramente. À Nessie, por um apoio incondicional e amizade genuína, com a qual sei que poderei permanentemente contar. Incontestavelmente, desde o secundário desenvolvemos uma amizade que nos tornou família. À Rita, a minha eterna companheira de guerra, porque nesta nossa etapa compartilhamos conquistas e frustrações, suspiros de cansaço e alívio, momentos de descanso e diversão. Mesmo nos momentos de desalento, descobrimos atalhos para nos apoiarmos mutuamente.

Indubitavelmente, a faculdade despoletou uma amizade que perdurará ao longo das nossas vidas.

Por último, à minha família. Porque até os mais pequenos momentos se tornaram grandiosas motivações. À minha prima Joana, porque após uma semana exaustiva de trabalho, a tua boa disposição, as nossas infinitas brincadeiras infantis, o nosso verdadeiro e inexplicável entendimento proporcionaram a melhor distração e harmonizaram cada fim de semana. Ao meu avô, por amorosamente me ter apoiado e demonstrado, com toda a sua simplicidade, humildade e experiência de vida, que desistir não é opção. A vida é uma jornada repleta de caminhos árduos, mas o laço que nos une é mais forte do que qualquer obstáculo. E finalmente, aos meus pais, porque incontestavelmente são a primordial razão daquilo que fui, sou e serei. O vosso apoio incondicional, a maneira infatigável como me procuraram dar tudo o que necessitava para as minhas tentativas em alcançar as minhas ambições e ensinamentos incutidos foram a pedra basilar para cada caminho escolhido.

Obrigada.

RESUMO

Introdução: O cancro da mama permanece a principal causa de morte por cancro nas mulheres em todo o mundo, principalmente devido ao desenvolvimento de recorrência e/ou doença metastática. De fato, ao momento do diagnóstico, cerca de 5% das pacientes apresentam metastização à distância e aproximadamente 15% das pacientes desenvolvem metastização à distância durante os primeiros 3 anos. Consequentemente, houve um impulso em desenvolver novos biomarcadores minimamente invasivos para doença avançada. Os microRNAs possuem potencial como biomarcadores para o cancro, principalmente devido à sua estabilidade em amostras de tecidos e fluidos corporais. Nos últimos anos, inúmeras evidências indicam que microRNAs específicos desempenham um papel funcional em diversas etapas do processo metastático e comportam-se como mediadores de sinalização, permitindo a colonização de um órgão específico.

Objetivo: O principal objetivo desta Dissertação de Mestrado foi avaliar o desempenho do *miR-30b-5p* como biomarcador de predição da progressão do cancro de mama e avaliar a sua viabilidade como biomarcador de doença avançada em biópsias líquidas.

Materiais e Métodos: Inicialmente, os níveis de expressão de *miR-30b-5p* foram validados numa grande série de amostras de tecidos de tumores primários e respetivas lesões metastáticas. Posteriormente, os níveis de expressão de *miR-30b-5p* foram avaliados em amostras de plasma de uma coorte de pacientes com cancro de mama localizado ou avançado. Seguidamente, o seu potencial como biomarcador de prognóstico foi avaliado através da construção da curva de ROC. Por fim, os níveis de expressão do *miR-30b-5p* foram avaliados em linhas celulares de cancro da mama, nomeadamente BT-474, MDA-MB-231 e Bo-1833 e nos seus respetivos meios condicionados.

Resultados e Discussão: O *miR-30b-5p* foi diferencialmente expresso em tumores primários comparativamente com correspondentes lesões metastáticas, tendo-se observados níveis mais elevados de *miR-30b-5p* nas metástases ósseas. O mesmo se verificou nos correspondentes tumores primários, sugerindo um papel importante deste microRNA na disseminação tumoral e na modulação do tropismo para um órgão específico. Adicionalmente, verificou-se que pacientes com doença avançada apresentaram elevados níveis de expressão de *miR-30b-5p* plasmático comparativamente a doentes com cancro de mama localizado. De facto, a expressão de *miR-30b-5p* discriminou pacientes com estadios avançados de pacientes com doença localizada com sensibilidade de 88,9%, especificidade de 66,7% e acuidade de 75,6%. Relativamente aos estudos *in vitro*, a linha celular proveniente de um tumor primário apresentou níveis significativamente mais elevados de expressão intracelular de *miR-30b-5p* comparativamente às linhas celulares

de cancro de mama metastático. Adicionalmente apenas o meio condicionado da primeira linha celular apresentou expressão de *miR-30b-5p*. Além do mais, os seus fenótipos epiteliais e mesenquimais parecem associar-se com a expressão do *miR-30b-5p*, sugerindo um papel funcional tanto na transição epitélio-mesenquima como na transição mesenquima-epitélio.

Conclusões e Perspetivas Futuras: Os resultados sugerem que a expressão do *miR-30b-5p* pode identificar doentes com cancro da mama que apresentam maior risco de progressão de doença, podendo ser uma abordagem clinicamente útil para a monitorização dos doentes, possibilitando um tratamento atempado e mais eficaz. No entanto, a validação em coortes multicêntricas é necessária para confirmar estes nossos achados. Como principal perspetiva futura, pretendemos avaliar a expressão do *miR-30b-5p* em amostras de seguimento, a fim de avaliar o seu potencial como biomarcador de monitorização para a deteção precoce de metástases do cancro da mama.

ABSTRACT

Introduction: Breast cancer (BrCa) remains the leading cause of cancer-related death in women worldwide, mainly due to development of recurrent and/or metastatic disease. Indeed, at time of diagnosis around 5% of patients present distant metastases and approximately 15% of patients develop distant metastases within the first 3 years. Consequently, there is an urge to bring out novel minimally invasive biomarkers for advanced disease. MicroRNAs hold promise as cancer biomarkers due to their stability in tissues and bodily fluids. In the last years, increasing evidence strongly indicates that specific microRNAs play a functional role in several steps of the metastatic cascade, behaving as signaling mediators to enable the colonization of a specific organ.

Aims: The main objective of this Master Dissertation was to evaluate the biomarker performance of *miR-30b-5p* expression for predicting BrCa progression and to assess its feasibility as a biomarker of advanced disease in liquid biopsies.

Material and Methods: Firstly, *miR-30b-5p* expression level was validated using quantitative reverse transcription polymerase chain reaction in a large set of formalin-fixed paraffin-embedded primary and metastatic tumors tissue samples. Then, *miR-30b-5p* expression level was assessed in a plasma BrCa patients' cohort composed by patients with localized or advanced BrCa. A ROC curve was constructed to evaluate *miR-30b-5p* prognostic performance. Finally*, miR-30b-5p* expression levels were evaluated in BrCa cell lines, namely BT-474, MDA-MB-231 and Bo-1833, and in the respective conditioned mediums.

Results and Discussion: *MiR-30b-5p* was differentially expressed in primary tumors and paired metastatic lesions, with bone metastases displaying significantly higher *miR-30b-5p* expression levels, paralleling the corresponding primary tumors, suggesting an important role in tumor dissemination and a potential role in modulation of metastatic organ tropism. Interestingly, patients with advanced disease disclosed increased plasma *miR-30b-5p* expression compared to patients with localized BrCa. In fact, *miR-30b-5p* expression discriminated advanced from localized BrCa patients with 88.9% sensitivity, 66.7% specificity and 75.6 accuracy. Regarding *in vitro* studies, primary BrCa cell line displayed significantly higher intracellularly *miR-30b-5p* expression when compared to metastatic BrCa cell lines. Remarkably, only conditioned medium from the primary BrCa cell line showed *miR-30b-5p* expression. Moreover, their epithelial and mesenchymal phenotypes might be correlated with *miR-30b-5p* expression, suggesting a functional role on the plastic process of epithelial-to-mesenchymal transition and mesenchymal-to-epithelial transition.

Conclusions and Future Perspectives: Our findings suggest that *miR-30b-5p* might identify breast cancer patients at higher risk of disease progression and may constitute a useful clinical tool for patient monitoring, entailing earlier and more effective treatment. Nonetheless, additional validation in larger multicentric cohorts are needed to confirm our findings. As a future perspective, we intend to assess *miR-30b-5p* levels in additional follow-up samples to evaluate its potential as monitoring biomarker for early detection of BrCa metastases.

TABLE OF CONTENTS

FIGURES INDEX

TABLES INDEX

LIST OF ABBREVIATIONS

Ago – Argonaute Protein

AJCC-UICC – American Joint Committee on Cancer - Union for International Cancer Control

ATCC – American Type Culture Collection

AUC – Area Under the Curve

BrCa – Breast Cancer

BRCA1 – BRCA1, DNA Repair Associated

BRCA2 – BRCA2, DNA Repair Associated

cDNA – Complementary DNA

ChT – Chemotherapy

DCIS – Ductal Carcinoma *In Situ*

DGCR8 – Di-George Syndrome Critical Region Gene 8 Protein

DMEM – Dulbecco's Modified Eagle Medium

DNA – Deoxyribonucleic Acid

EDTA – Ethylenediamine Tetraacetic Acid

EMT – Epithelial-to-Mesenchymal Transition

ER – Estrogen Receptor

ERBB2 – Erb-b2 Receptor Tyrosine Kinase 2

ESMO – European Society for Medical Oncology

ET – Endocrine Therapy

FBS – Fetal Bovine Serum

FFPE – Formalin-Fixed Paraffin-Embedded

H&E – Hematoxylin and Eosin

HER2 – Human Epidermal Growth Factor 2 Receptor

HRT – Hormone Replacement Therapy

IDC – Invasive Ductal Carcinoma

IHC – Immunohistochemistry

ILC – Invasive Lobular Carcinoma

LCIS – Lobular Carcinoma *In Situ*

LncRNAs – Long non-coding RNAs

MET – Mesenchymal-to-Epithelial Transition

MiRNAs – MicroRNAs

MRI – Magnetic Resonance Imaging

mRNA – Messenger RNA

NcRNAs – Non-coding RNAs

NPV – Negative Predictive Value

NST – No Special Type

Nt – Nucleotides

ORF – Open Reading Frames

Pen/Strep – Penicillin-Streptomycin

PPV – Positive Predictive Value

PR – Progesterone Receptor

PTEN – Phosphatase and Tensin Homolog

RISC – RNA-induced Silencing Complex

RNA – Ribonucleic Acid

ROC – Receiver Operating Characteristic

rRNAs – Ribosomal RNAs

RT – Radiotherapy

RT-qPCR – Quantitative Real-Time Polymerase Chain Reaction

SiRNAs – Small interfering RNAs

SLNB – Sentinel Lymph Node Biopsy

snoRNAs – Small Nucleolar RNAs

SSC – Special Subtypes Carcinomas

TP53 – Tumor Protein p53

TRBP – Transactivation Response RNA-binding Protein

tRNAs – Transfer RNAs

UTR – Untranslated Regions

WHO – World Health Organization

XPO5 – Exportin 5

INTRODUCTION

Breast Cancer

Epidemiology

Breast cancer (BrCa) is the second most common cancer worldwide, corresponding to 11.9% of all cancer incidence. It is the most frequently diagnosed cancer in women, accounting for 1.67 million estimated new cases, approximately 25% of all new cancer diagnoses in 2012 (1). The estimated age-standardized BrCa incidence rates distribution varied nearly four-fold across the world in 2012, with the highest incidence rates observed in more developed regions, whereas the lowest rates were observed in less developed regions **(Figure 1)** (2).

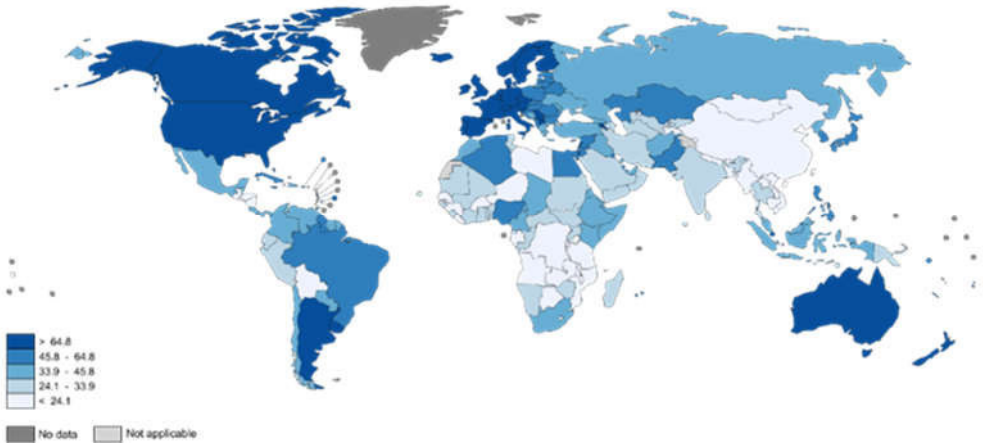

Figure 1 - Estimated Age-Standardized Breast Cancer Incidence Worldwide in 2012. From (3).

The variations in BrCa incidence rates are mostly due to the introduction of mammography screening programs, which allows early detection, and the population ageing in the developed regions. Moreover, advances in cancer detection and treatments have improved cancer survival rates and life expectancy, increasing the prevalence of BrCa in most Western countries (1, 4). Worldwide in 2012, BrCa was the fifth cause of cancer-related death, accounting with 522 thousand deaths. Indeed, it was the most frequent cause of cancer-related death in females in less developed regions and the second cause in more developed regions **(Figure 2)** (1, 2).

In Europe, BrCa was the leading cancer and the main cause of cancer-related death in women in 2012. In Portugal, BrCa was the foremost cancer with 67.6 per 100,000 new cases in 2012 and the main cause of cancer death in female with approximately 13.1 per 100,000 deaths **(Figure 3)** (5).

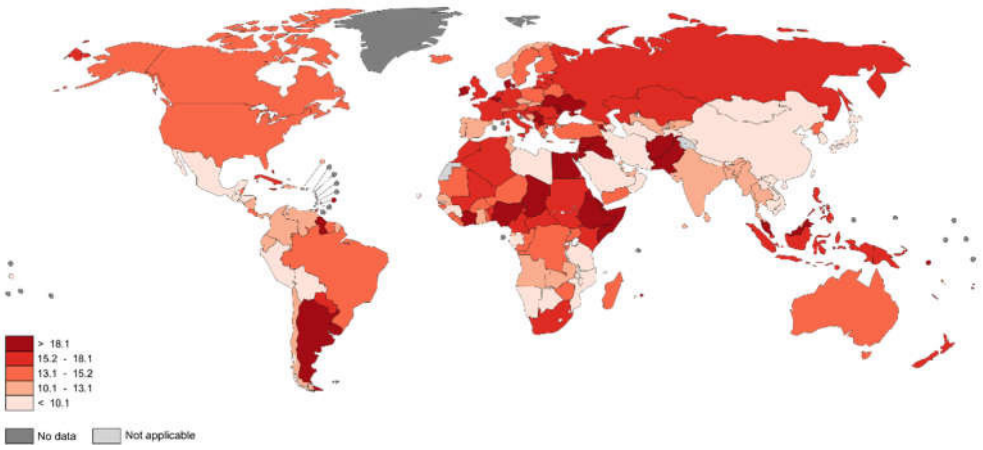

Figure 2 - Estimated Age-Standardized Breast Cancer Mortality Worldwide in 2012. From (3).

Despite the improvement in cancer early detection and treatment, BrCa remains the leading cause of death from cancer in women, mainly due to development of recurrent and/or metastatic disease (4). In fact, at the time of diagnosis approximately 5% of patients present distant metastases and up to 15% of patients develop distant metastases within the first 3 years (6). Undeniably, metastatic BrCa is in most cases incurable (7).

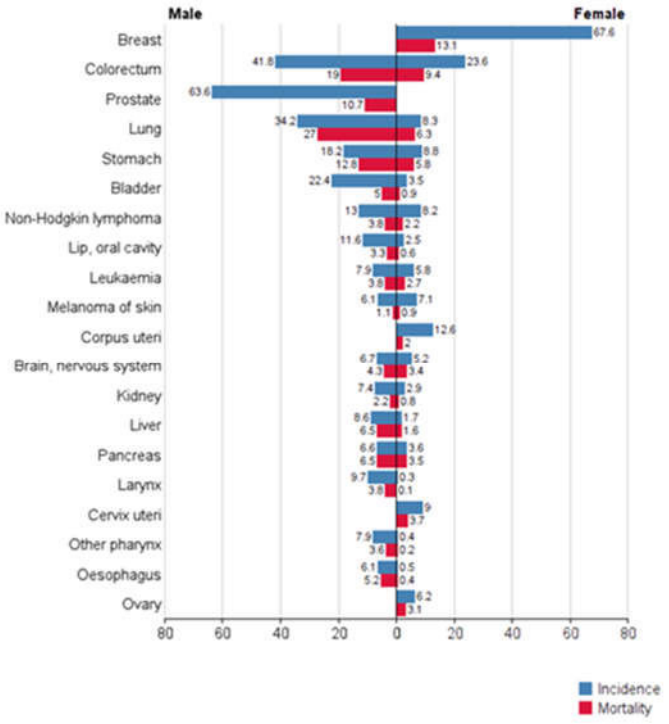

Figure 3 - Estimated Age-Standardized Incidence and Mortality Rates (per 100,000) in Portugal in 2012. From (3).

Risk Factors

Several factors were associated with BrCa risk, including demographic factors, family history, hormone exposure and lifestyle factors, among others (**Table 1**). However, most of these factors are associated with a minor to moderate increase in risk, being estimated that about 50% of female who develop BrCa have no recognizable risk factors elsewhere gender and increased age (8, 9). Undeniably, BrCa incidence sharply increases with age, being usually diagnosed in women among 45 and 74 years (10). BrCa incidence is higher in African-American women younger than 40 years and in Caucasian women older than 40 years which reveals an increased risk of BrCa development associated with ethnicity and race (11).

Table 1 - Magnitude of risk of the main Breast Cancer risk factors. Adapted from (8, 9).

Relative Risk < 2	Relative Risk 2 – 4	Relative Risk > 4
• Early age at menarche; • Late age at menopause; • HRT; • Alcohol consumption; • Cigarrete smoking; • Postmenopausal obesity;	• BrCa history in first-degree relatives; • Late age at first delivery;	• Atypical hyperplasia; • *BRCA1* and *BRCA2* mutations; • Exposure to ionizing radiation;

Abbreviations: BrCa – Breast Cancer; *BRCA1 – BRCA1, DNA repair associated*; *BRCA2 – BRCA2, DNA repair associated*; HRT – hormone replacement therapy

Furthermore, women with atypical epithelial hyperplasia have an increased risk of developing subsequent invasive BrCa, mostly in premenopausal women (12, 13). Moreover, women with history of BrCa in family members, particularly first degree relative or a relative diagnosed before 40 years, have an increased BrCa risk (8, 9, 14). Although familial BrCa accounts only to 5-10% of all BrCa cases, mutations in *BRCA1, DNA repair associated (BRCA1) and BRCA2, DNA repair associated (BRCA2)* are strongly related with higher lifetime risk of BrCa that differs from 26 to 85%, mainly in younger premenopausal women (14, 15). *Tumor protein p53 (TP53)* and p*hosphatase and tensin homolog (PTEN)* are also shown to be involved in familial BrCa, although with a minor role (8).

Additionally, women's reproductive history is also associated with BrCa risk. Women with early age at menarche, late age at first delivery and last full-term pregnancy or late age at menopause have an increased BrCa risk, while parity, premenopausal oophorectomy and breastfeeding contribute as a protective effect on the risk of developing BrCa (9, 16). Some of these risk factors might be explained by their association with estrogen levels exposition since elevated levels of endogenous estrogen are related with normal and malignant breast cells proliferation (9). Likewise, obesity and hormone replacement therapy (HRT) in

postmenopausal women contribute to a minor rise of BrCa risk, while in premenopausal women, a high body mass index seems to be a protector effect (9, 17).

Also, a connection between environmental factors such as exposure to ionizing radiation and BrCa is well-recognized. Indeed, an increased risk of developing BrCa has been described in women who performed radiation treatments at younger ages (8). Although some associations are controversial, several studies tried to establish a relation between lifestyle factors and BrCa risk. A healthy diet with high consumption of vegetables and physical activity seems to contribute as a protective effect, whereas alcohol consumption and cigarrete smoking are associated with an increased BrCa risk (9, 18, 19).

Screening and Diagnosis

BrCa detecting at a pre-clinical stage, before it acquires the potential to spread, is the major goal of implemented population-based mammography screening programs. In females between 50 and 69 years, mammography screening every two years has revealed the most effective mortality reduction benefit. In women with familial BrCa, an annual screening with magnetic resonance imaging (MRI) along with or alternating with mammography every six months is recommended and should start ten years younger than the youngest case known in the family (20).

BrCa diagnosis is based on physical examination, comprising bimanual palpation of the breasts and locoregional lymph nodes in concomitance with imaging and pathological confirmation (20, 21). Nowadays, mammography is the standard imaging technique for the detection of BrCa. However, its sensitivity is influenced by the breast's density, being necessary to performed MRI in specific situations (20, 22). Apart from imaging techniques, the confirmation of malignant involvement can only be appropriately determined by tissue sampling. The gold standard technique for palpable and impalpable breast abnormalities' diagnosis remains the biopsy. Fine-needle biopsy and needle-core biopsy are performed on palpable lesions. Although both techniques have a good sensitivity, for patients who will receive preoperative systemic therapy, needle-core biopsy is required to guarantee a histopathological diagnosis and assess immunostaining markers. Pathological diagnosis for non-palpable lesions is based on an image-guided core needle biopsy (8, 20).

Histological Subtypes

BrCa is extremely heterogeneous both morphologically and clinically. Currently, World Health Organization (WHO) Classification of Tumors of the Breast distinguishes more than twenty different histological subtypes. The majority of BrCa arises from epithelial cells and might be divided into two main categories: *in situ* and invasive carcinomas (22).

In situ carcinomas are characterized as pre-invasive lesions in which malignant cells still restricted to the ductal or lobular tree of the breast without invading the basement membrane of the surrounding stroma (23). These pre-invasive lesions might be further subdivided into lobular carcinoma *in situ* (LCIS) and ductal carcinoma *in situ* (DCIS) which are distinguished by cytological and architectural features and not by the microanatomical site of origin (23, 24). With the implementation and improvement of the screening programs, DCIS accounts approximately to 20-25% of newly diagnosed BrCa (22, 23).

Nevertheless, invasive carcinomas comprise 70 to 80% of malignant mammary carcinomas. This group can be subdivided into invasive carcinoma of no special type (NST), also recognized as invasive ductal carcinoma (IDC), and special subtypes carcinomas (SSC)(22). The IDC category represents 75% of the invasive carcinomas and comprises all the tumors which lack histologic features for being categorized as one of the SSC (24). In SSC group are included more than ten histological types, being the invasive lobular carcinoma (ILC) the most frequent type, representing 5 to 15% of all invasive BrCa (22, 23) However, tumors exhibiting combined morphology, such as SSC and NST patterns are classified as mixed (22).

Staging, Prognostic and Predictive Biomarkers

The Biomarkers Definition Working Group defined a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (25).

A prognostic factor is a measurement related to patients’ outcome that might be applied to estimate the chance of recovery or recurrence, whereas a predictive factor is a measurement that predicts the responsiveness to a determined treatment. Some biomarkers might have prognostic and predictive value (8).

Histological grade is a measure of how close a malignancy remains of its original tissue. The method of grading is based on three parameters: the grade of architectural differentiation, nuclear pleomorphism and mitotic index. Nowadays, the Nottingham

Combined Histologic Grade (**Appendix I**) is considered a significant prognostic factor, providing useful information for clinical management (22).

Currently, BrCa staging is based on the TNM staging system reported by the American Joint Committee on Cancer - Union for International Cancer Control (AJCC-UICC) (22) (**Appendix II**). This staging system specifically combines extension of the primary tumor (T), involvement of the regional lymph nodes (N) and development of distant metastases (M). In BrCa patients both clinical and pathological staging can be applied. The clinically staging (c) is defined considering information prior to surgery or any primary treatment, including physical examination, imaging techniques and pathologic report based on fine-needle biopsy. On the other hand, the pathological staging (p) comprises clinical stage information as well as all the information from surgery and pathologic examination of the primary tumor, lymph nodes and metastatic lesions, if applicable (26). Based in TNM classification, five stages are established (**Appendix III**), allowing to evaluate the spread of the disease and patients' prognosis (26).

Despite the prognostic value provide by the grade system and TNM staging, patients with same stage displayed different outcomes. Thus, it was essential to introduced others biomarkers of patients' outcome and, specially, predictive therapy response (8). Currently, assessment of progesterone receptor (PR) and estrogen receptor (ER) *status* by immunohistochemistry (IHC) and human epidermal growth factor 2 receptor (HER2) *status* by IHC or in situ hybridization when necessary is part of the currently clinical practice (20).

ER is a nuclear transcription factor activated by the hormone estrogen to stimulate the development and differentiation of normal breast cells. ER-positive tumors, by IHC correspond up to 75% of invasive BrCa. In fact, ER-positivity is related with less aggressive and well-differentiated tumors, so a better outcome in comparison to ER-negative tumors (27, 28). Furthermore, ER-positive patients generally have a better response to anti-estrogen or aromatase inhibitors (28). PR is also a nuclear transcription factor activated by the hormone progesterone to stimulate cell proliferation. ER regulates the expression of PR, so PR expression suggests an active ER signaling pathway (27). Approximately 75% of BrCa are PR-positive by IHC, being PR expression associated to a better endocrine therapy response (ET) (28). Besides, tumors ER-positive and PR-negative are less responsive to hormone therapy in comparison to tumors positive for both receptors (22). Although both PR and ER *status* display weak prognostic value, they play a major role in determining the responsiveness to hormone therapy (28).

Erb-b2 receptor tyrosine kinase 2 (ERBB2) gene, predominantly known as *HER2* gene, is an oncogene localized in chromosome 17 that encodes a transmembrane protein for a

growth factor receptor present in breast epithelial cell surface (27, 29). About 15% to 20% of all BrCa demonstrate *HER2* gene amplification or protein overexpression. Moreover, more than 55% of these tumors do not express ER or PR (28). HER2 expression is related with poor prognostic and at same time to a positive response to HER2-targeted therapy for instance trastuzumab (22, 28). Moreover, HER2-positivity is predictive of favorable response to anthracycline and taxane-based regimens (30) .

The IHC evaluation of proliferation-related markers such as Ki67 index has been applied in clinical practice as it supplies additional valuable information as prognostic factor (31). Moreover, age is also a noteworthy prognostic factor since BrCa patients younger than 35 years have worst prognosis than older patients (8).

Nevertheless, due to the limitations of the current biomarkers, gene expression profiles which recognizes genes that might be used as a molecular signature in predicting prognosis and identifying patients who are most likely to benefit from specific therapies have been developed. Oncotype DX, MammaPrint and PAM50 are some of these molecular prognostic profiles, but the high costs of molecular signatures limited their use in clinical practice (32, 33).

Molecular Subtypes

BrCa is a heterogeneous disease. Patients with tumors with identical histological type and stage might present diverse outcomes and treatment responses (34). Gene expression profiling and hierarchical clustering have enabled to categorized BrCa into four intrinsic molecular subtypes (luminal A, luminal B, HER2-enriched and basal-like) that are associated with diverse clinical outcomes and responsiveness (35, 36).

Luminal BrCa is mainly distinguished by the expression of high levels of ER and luminal epithelial cytokeratins. The luminal subtype represents approximately 70 to 80% of BrCa and can be subdivided into Luminal A and Luminal B (34, 37, 38). Luminal BrCa exhibiting higher expression of ER-regulated genes, no amplification of HER2 and low expression of proliferation-related genes are classified as luminal A. Contrarily, luminal B tumors are known to have a lower expression of ER-related genes, a higher expression of proliferation-related genes and a variable amplification of HER2, being consequently associated with worse prognosis in comparison to luminal A tumors (39, 40).

The non-luminal or ER-negative BrCa includes two intrinsic subtypes: the HER2-enriched subtype and the basal-like subtype. The HER2-enriched subtype exhibits high expression of several genes in the HER2 amplicon at 17q22.24, including *HER2*. Although these

cancers display an aggressive clinical outcome because of their poor differentiation and high proliferation, they present a good response to HER2-targeted therapy (37). Low expression of luminal epithelial genes and high expression of basal epithelial genes characterize the basal-like subtype. Even if there is around 80% overlap between triple-negative (negative for ER, PR and HER2) and basal-like subtype, triple-negative subtype also includes special histological types like adenoid cystic and medullary carcinoma with low risks of distant relapse (41).

Therefore, the molecular subtypes of BrCa have a valuable role in evaluating prognosis and determining the responsiveness to therapy, providing a personalized treatment (38). Nowadays, the assessment of the intrinsic molecular subtypes is based on a cost-effective IHC assays (**Table 2**), but can also be defined by gene expression profiling using multiparameter molecular tests such as PAM-50 (37, 42).

Table 2 - Characterization of Breast Cancer molecular subtypes according to European Society for Medical Oncology (ESMO). Adapted from (20).

Intrinsic Subtype	Clinicopathological Surrogate Markers
Luminal A[1]	**"Luminal A-like"** ER-positive HER2- negative Ki67 low[2] PR high[2]
Luminal B[1]	**"Luminal B-like" (HER2-negative)** ER-positive HER2- negative And either Ki67 high[2] or PR low[2] **"Luminal B-like" (HER2-positive)** ER-positive HER2- positive Any Ki67 and any PR
HER2-Enriched	**"HER2-positive (non-luminal)"** HER2- positive ER and PR absent
Basal-like	**"Triple-negative"**[3] HER2- negative ER and PR absent

[1] If molecular signature is available, Luminal A BrCa are associated with a low-risk signature, whereas Luminal B BrCa with a high-risk signature.

[2] Scores should be interpreted in the light of local laboratory values.

[3] There is around 80% overlap between 'triple-negative' and intrinsic 'basal-like' subtype, however 'triple-negative' also comprises special histological types such as medullary and adenoid cystic carcinoma with low risks of distant relapse.

Abbreviations: ER – estrogen receptor; HER2- human epidermal growth factor 2 receptor; PR- progesterone receptor

Advanced Breast Cancer

Advanced BrCa includes both locally advanced and metastatic BrCa (43). Locally advanced BrCa comprises the most advanced BrCa without metastases at distant organs which end up being related to higher risk of locoregional and systemic relapse. Although definition of locally advanced BrCa is still controversial, recent guidelines describes it as a AJCC stage III BrCa (44).

Metastatic BrCa arises following the aggressive proliferation of cancer from its primary location to distant organs (45). The metastatic process involves multiple sequential steps. Initially, malignant cells undergo an epithelial-to-mesenchymal transition (EMT), acquiring mesenchymal characteristics such as higher motility and invasiveness. These features allow malignant cells to disseminate from primary tumor, invade through adjacent tissues and basement membranes and enter into circulation by the lymphatics and/or blood vascular system. Subsequently, the survival malignant cells arrest and extravasate into the foreign microenvironment where they pass through a mesenchymal-to-epithelial transition (MET) to revert to an epithelial phenotype, crucial for the colonization and establishment of a secondary tumor (**Figure 4**) (45, 46).

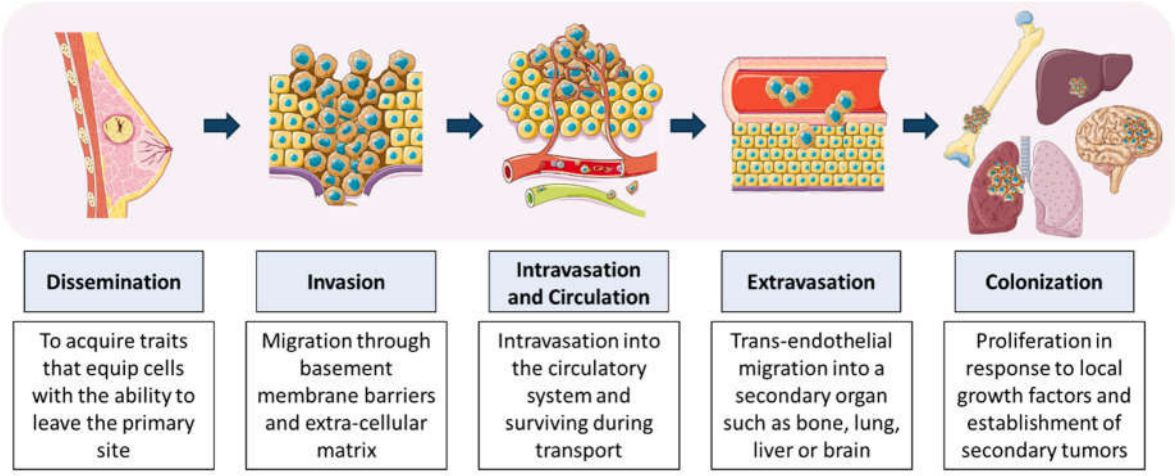

Figure 4 - Multiple sequential steps of the metastatic process. Metastasis is a multistep process that starts with the dissemination of malignant cells from primary tumor. The epithelial-to-mesenchymal transition (EMT) allows the acquisition of features essential for migration through surrounding tissues and basement membranes. After escaping from the primary tumor, malignant cells might intravasate into circulation until they arrest and extravasate in a secondary organ. At this final phase, tumor cells undergo a mesenchymal-to-epithelial transition (MET) and proliferate, finally establishing a secondary tumor. Estevão-Pereira H. *unpublished*.

The metastases can show an organ-specific pattern of spread. Indeed, the most common site of BrCa metastases is the bone, with lungs, liver and brain as the second, third and fourth most common metastatic sites, respectively (6, 47). Besides, the molecular subtype of the primary tumor is also associated to the metastatic spread and to distant metastases sites (48). Luminal subtypes are related to a slower metastatic proliferation, low relapse

rates and better outcomes in comparison to basal-like subtype. Patients with luminal tumors display a higher predisposition to develop bone metastasis, whereas basal-like cancers metastasize preferentially to lung and brain. HER2-positive BrCa show the highest rate of recurrence and have higher potential to develop brain metastasis (6, 48-51).

Treatment

Due to the added value of providing a personalized practical approach, BrCa treatment should be determined by a multidisciplinary clinical team (20). Tumor site, extension and biology, the proliferation of disease and its metastatic potential, prognostic biomarkers *status* as well as patient's age, general health *status*, menopausal *status* and preferences are some of the main factors that should be consider (52). According to these, BrCa treatment might embrace one or more strategies such as surgery, chemotherapy (ChT), radiotherapy (RT), ET and target therapies (20).

Concerning local treatment, approximately 60 to 80% of early-stage BrCa patients are amenable for conservative surgery proceed by RT. Nonetheless, in some cases due to tumor size and multicentricity, incapability to accomplish negative surgical margins after several resections, contraindications to RT or patient's preference, mastectomy is still performed (20, 53). Owing to its association with reduced morbidity, sentinel lymph node biopsy (SLNB) is the standard approach for axillary staging in early, clinically node-negative BrCa. Though, in patients with sentinel node metastasis, axillary lymph node clearance is obligatory (20, 54).

BrCa recurrence might be prevented by systemic adjuvant treatment. According to the predictive response to therapy, the overall benefit and the risk of relapse, the treatment chosen might be RT, ChT, ET and/or target therapies (20, 55).

As previously mentioned, postoperative RT is strongly recommended after a conservative surgery and in patients that carried out a mastectomy and had positive lymph nodes (55).

On the other hand, the decision of treatment with ChT is complex and usually based on the molecular subtypes (56). The advantage of adjuvant ChT is higher in ER-negative tumors. However, ChT is also recommended in HER2-positive, "triple-negative" tumors and luminal B BrCa with HER2 amplification or high recurrence risk (20, 56). In cases of doubt, gene expression profiles like PAM50 might be used to determine the risk of relapse and predict the benefit of ChT (20).

BrCa with positive hormone receptors benefit of ET. The prescription of the agent is principally defined by patients' menopausal *status*. In premenopausal women tamoxifen for

5 to 10 years is the standard treatment, while in postmenopausal, aromatase inhibitors like letrozole are recommended (20, 56).

The most known target therapy for BrCa is the HER2-directed therapy. Trastuzumab is a monoclonal antibody against HER2 and several studies demonstrated that when combined with ChT in patients with HER2 amplification halves the risk of relapse in comparison with ChT alone (20).

For the treatment of advanced BrCa, RT, ChT, ET and/or target therapies might be applied in the neoadjuvant setting before breast surgery and axillary node clearance (when conceivable) (43). Indeed, a multimodality treatment seems to improve significantly the patients' outcomes (44). Moreover, neoadjuvant treatment might also be performed in patients with multifocal BrCa and/or with large operable tumors that would otherwise require mastectomy because of the tumor size (20).

The management of metastatic BrCa patients also involves treatment of the related symptoms to improve the patients' quality of life (43). Indeed, metastatic BrCa remains in most cases an incurable disease (7).

Epigenetics

In 1942, epigenetic was defined by Conrad Waddington as *"the causal interaction between genes and their products, which bring the phenotype into being"* (57). Nevertheless, due to the increased amount of knowledge in this area, the concept has evolved over the years, being currently defined as heritable alterations in gene function and regulation that are not owed to any change in the nucleotide sequence (57-59).

Epigenetic processes are crucial to guarantee the normal development and homeostasis of the organism. In fact, epigenetic deregulation has been verified on early steps of the tumorigenesis process (60). Four major mechanisms are involved in epigenetic regulation: non-coding RNAs (ncRNAs), DNA methylation, post-translational modifications of histones and histone variants (58, 60) (**Figure 5**). Notwithstanding their crucial role as epigenetic mechanisms, the last three mechanisms above mentioned will not be the focus of this dissertation, so the concepts will not be discussed.

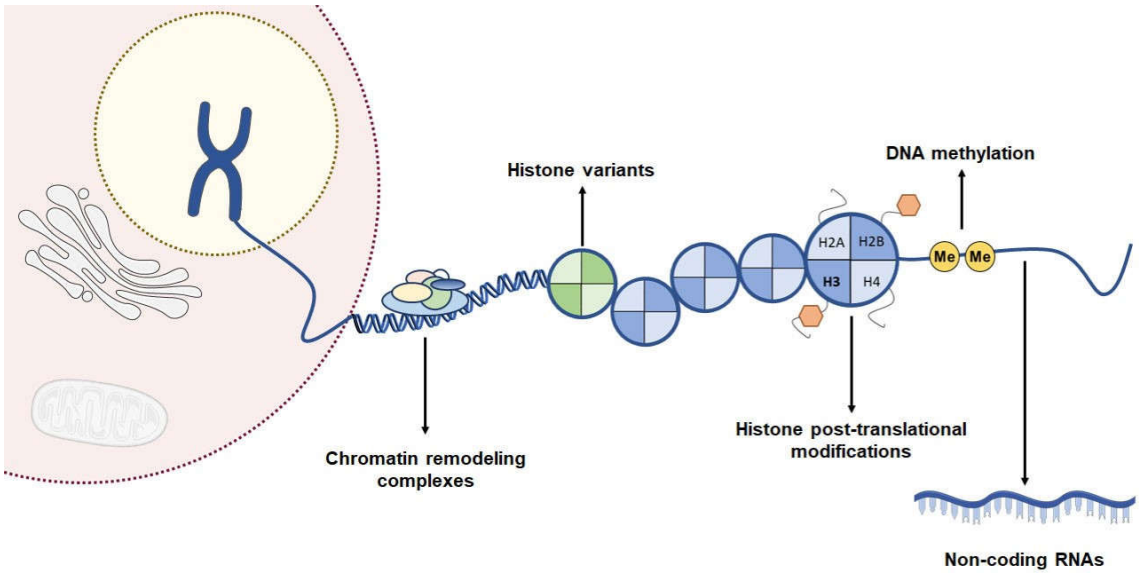

Figure 5 – Schematic representation of the epigenetic mechanisms. The major mechanisms involved in epigenetic regulation are non-coding RNAs, DNA methylation, post-translational modifications of histones and histone variants. (Kindly provided by Lameirinhas A.unpublished)

Non-coding RNAs

In the last years, recent evidence has given emphasis to the crucial role of the transcribed genes that do not encode proteins, particularly ncRNAs (61, 62). According to their size, ncRNAs can be divided into long non-coding RNAs (lncRNAs) (more than 200 base pair) and small non-coding RNAs (less than 200 nucleotides (nt)). Moreover, ncRNAs can also be classified depending on their function, specifically molecules that are generally constitutively expressed such as transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs) and ribosomal RNAs (rRNAs), and molecules which play a regulatory role like small interfering RNAs (siRNAs) and microRNAs (miRNAs) (63). Indeed, ncRNAs play an important role at several levels of gene expression, being its deregulation involved in the development of many different disorders (61).

MicroRNAs

The miRNAs are endogenous, highly conserved small ncRNAs of approximately 22 nt in length, originally discovered in *Caenorhabditis elegans* (64, 65). MiRNAs represent an emerging class of molecules that play important roles at posttranscriptional regulation of gene expression in several cellular processes. Moreover, the association between their deregulation and cancer development has being studied during the last years (61). In fact, it is thought that around 50% of the human transcriptome is conditioned to miRNA regulation (66).

Biogenesis and Mechanisms of Action

The majority of miRNAs genes are found in intergenic regions, but they can also be located in exonic or intronic regions, both in sense or antisense orientation, being the furthest denominated “mirtrons” (67). Moreover, miRNAs might be found in gene clusters or as single genes (68).

The biogenesis of miRNAs is a multistep process that starts in the nucleus and finishes in the cytoplasm (**Figure 6**). In the canonical miRNA biogenesis pathway, miRNAs are mainly transcribed by RNA polymerase II that synthetizes a large primary precursor with a hairpin structure (pri-miRNA). The endonuclease Drosha and its cofactor Di-George syndrome critical region gene 8 (DGCR8) protein, the microprocessor complex, recognize and cleave pri-miRNA to a RNA hairpin intermediate (pre-miRNA) with two nt 3’ overhang (65, 67). Alternatively, the mirtrons bypass the Drosha processing and follow a splicing pathway to originate debranched introns that mimic the pre-miRNA structure (69).

Both RNA hairpin intermediate from canonical and non-canonical pathway are recognized by exportin 5 (XPO5) and are actively transported into the cytoplasm, where they are processed by the endonuclease Dicer and transactivation response RNA-binding protein (TRBP), forming a double-stranded miRNA duplex (67). This molecule is loaded into the Argonaute protein (AGO) and the mature miRNA guide is incorporated into RNA-induced silencing complex (RISC) (67, 69).

RISC directs the regulation of mRNA by identifying miRNAs targets through base-paring interactions between the “seed sequence” and the targeted mRNA which contains a partially or fully complementary sequence generally located in 3’ untranslated regions (UTR). MiRNAs repress gene expression at the posttranscriptional level according to the complementary sequence of the target mRNA: mRNA is cleavage when the complementarity is nearly perfect and mRNA translation is inhibited when the complementarity is partial (66). However, recent reports have shown that 5’UTR and open reading frames (ORF) also contain target sequences for miRNAs (70). The complementarity between the “seed sequences” and these regions is associated to an upregulation of target mRNA translation (71).

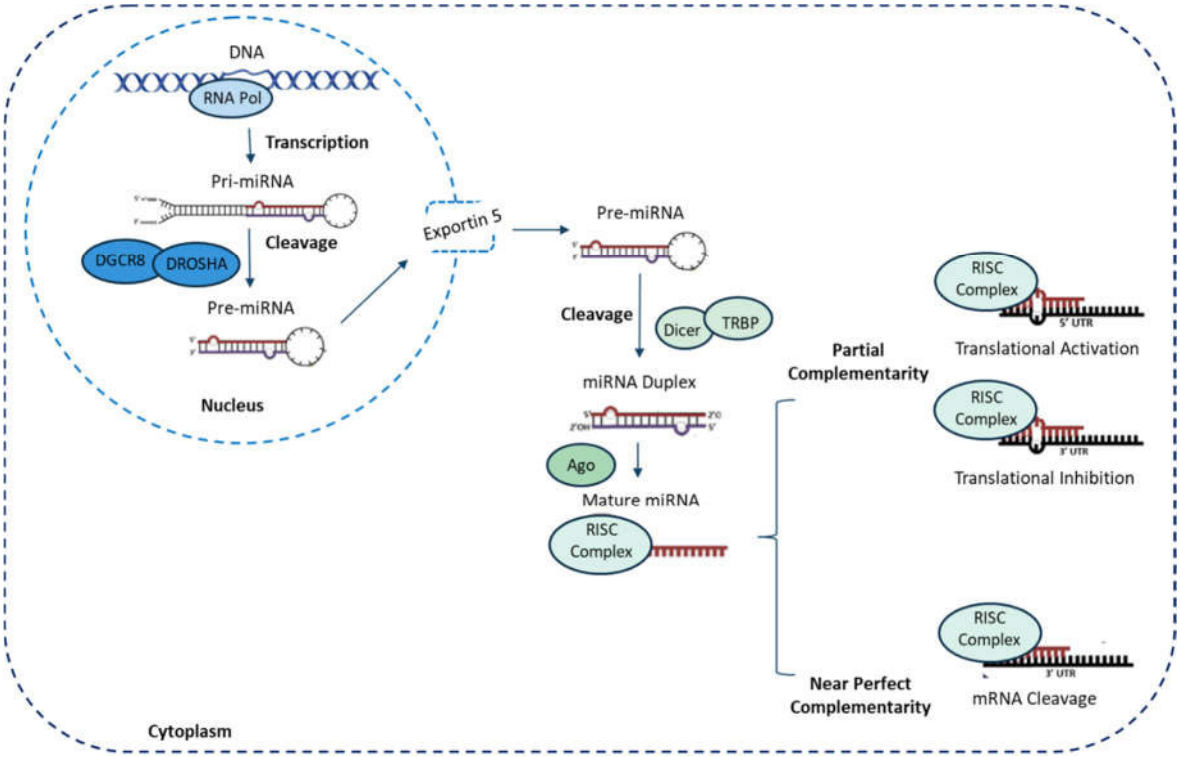

Figure 6 - Canonical pathway of microRNA biogenesis and mechanisms of action. miRNAs are canonically transcribed by RNA polymerase II that synthetize a primary precursor with a hairpin structure, the pri-miRNA. These molecules are processed by Drosha and Di-George syndrome critical region gene 8 (DGCR8) protein into pre-miRNAs. RNA hairpin intermediates from canonical and non-canonical pathway (not represented in the figure) are sequestered into the cytoplasm by exportin 5 (XPO5) where they are cleaved by endonuclease Dicer and transactivation response RNA-binding protein (TRBP), forming a double-stranded miRNA duplex. The double-stranded miRNA duplex is loaded into the Argonaute protein (AGO) and the mature miRNA guide is incorporated into RNA-induced silencing complex (RISC). RISC recognizes the target mRNA by identifying base-paring interactions. MiRNAs might regulate gene expression by mRNA cleavage, translational repression and translational activation. Estevão-Pereira H. *unpublished*.

MicroRNAs and Breast Cancer

The fundamental role of miRNAs in the development of several disorders, particularly the miRNA deregulation in human malignancies is well-recognized. Interestingly, several miRNAs are located in fragile regions of the genome that are susceptible to genetic abnormalities such as translocation, deletion or amplification (72). Moreover, miRNAs might also be deregulated by epigenetic mechanisms such as abnormal DNA methylation of their promoter regions (73, 74). In different steps of the tumorigenic process, miRNAs can act as tumor suppressor or oncogenes (oncomiRs). Tumor suppressor miRNAs act by negatively regulating the expression of oncogenes, being usually downregulated in cancer, while oncomiRs act by targeting tumor suppressor genes, being often upregulated in cancer. Besides, according to tumor type and cellular context, miRNAs might present a dual function (75).

The increased amount of evidence has brought forward the role of miRNAs at different steps of the BrCa development (76). Currently, since each malignancy type seems to have a distinct miRNA profile that allows to differentiate it from normal tissue and other tumors type, miRNAs might be used as diagnostic tools for BrCa (64). Iorio and colleagues recognized a 13-miRNA signature that could discriminate BrCa from normal breast tissues with 100% accuracy (77). Among these miRNAs, the most constantly dysregulated in BrCa were *miR-10b*, *miR-145, miR-125b* (downregulated), *miR-155* and *miR-21* (upregulated), suggesting their potential role as tumor suppressor genes or oncogenes, respectively. Additionally, expression profiles able to distinguish BrCa patients from healthy individuals in bodily fluids have also been investigated. Heneghan and coworkers identified higher levels of *miR-195* and *let-7a* in BrCa patients' blood (78), whereas Zhang *et al.* recognized a 3-miRNA signature (*miR-29c*, *miR-424* and *miR-199a*) as a diagnostic signature for non-invasive early detection of BrCa (79). Though, perhaps due to the clinicopathological variables and heterogeneity in BrCa, disparities between the miRNA signatures continue to be verified, leading to the investigation of miRNA profiles that might reflect different histopathological characteristics for instance ER, PR and HER2 *status* (80-83).

Furthermore, a correlation between several miRNAs and clinicopathological features related to different outcomes has been proposed, prompting the identification of miRNAs with prognostic value (64, 76, 84). In recent years, several lines of evidence imply multiple functions of miRNAs in BrCa metastases. Indeed, miRNAs might function either as promotors or suppressors of metastases by targeting multiple signaling pathways and important proteins that are major players in different steps of the metastatic process (85, 86). Besides, miRNAs seem to be involved in the phenotypic alterations correlated with metastases' development, acting as regulators of the EMT/MET processes (86). Currently, *miR-9* (87), *miR-10b* (88), *miR-21* (89), *miR-29a* (90), *miR-155* (91) are known as metastases promoters, whereas *miR-126*, *miR-335* (92), *miR-30* family (93, 94), *miR-200* family and *miR-205* (95) are predominantly describe as metastases suppressors. As previously mentioned, it is important to emphasize that due to their versatile role, some of these miRNAs, for instance *miR-200* family might act both as metastases promoter and metastases suppressor (96, 97).

Lastly, the role of miRNAs as predictive biomarkers has also been described. Mailot *et al.* found higher levels of *miR-21*, *miR-23b* and *miR-181b*, which were shown differentially expressed in tamoxifen-resistant cell lines, upon ET (98). Moreover, higher expression levels of *miR-210* was associated with increased risk of relapse in patients treated with tamoxifen (99) and resistance to trastuzumab (100), while *miR-100* was related with sensitivity to ChT using paclitaxel (101).

Therefore, miRNAs have been emerging as promising diagnostic, prognostic and predictive biomarkers for BrCa.

PRELIMINARY RESULTS

Preliminary Data

The project presented in this Master Dissertation arises from a previous unpublished work developed at the Cancer Epigenetics and Biology Group (GBEC), whose main goal was to evaluate the diagnostic and prognostic potential of some miRNAs in tissues samples from BrCa patients.

MiR-30b-5p expression level was assessed in 176 fresh-frozen BrCa samples and 26 normal breast tissue samples with no evidence of preneoplastic or neoplastic alterations (**Table 3**). Age distribution significantly differed between patients and controls (p=0.003).

No statistically difference was depicted between BrCa tissues and normal breast tissues invalidating *miR-30b-5p* potential as diagnostic biomarkers (**Figure 7A**). When the association between *miR-30b-5p* expression and clinicopathological features was evaluated, significantly higher *miR-30b-5p* expression was observed in N3 patients when compared to N0 and N1 patients (p=0.018 and p=0.0025, respectively, **Figure 7B**), suggesting an invasive and metastasis promoter function.

MiR-30b-5p expression level was further analyzed in a formalin-fixed paraffin-embedded (FFPE) primary and metastatic tumors available from BrCa patients' cohort of 16 BrCa patients comprising 38 tumor samples (16 primary BrCa and 22 paired metastases, **Table 3** and **Table 4**). Significantly higher *miR-30b-5p* expression was observed in metastatic lesions compared to the corresponding primary breast tumors (p=0.0066, **Figure 8A**). Specifically, in 10 of 16 patients, *miR-30b-5p* expression level was significantly increased in metastatic lesions *versus* primary tumors with a fold variation higher than 1 (**Figure 8B**).

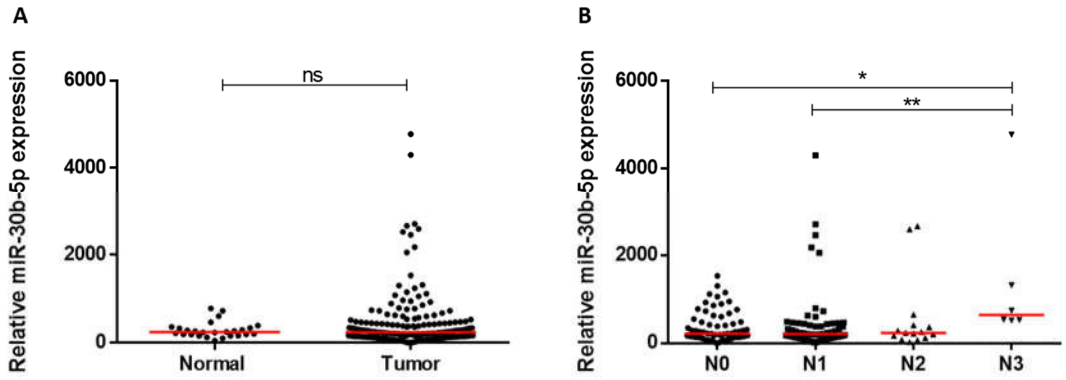

Figure 7 – (A) Scatter-plot of *miR-30b-5p* relative expression in normal breast tissues and Breast Cancer tissues. A ns denotes p-value>0.05 by non-parametric Mann-Whitney U test. **(B)** Scatter-plot of *miR-30b-5p* relative expression according to N stage. * p-value <0.05 and ** p-value <0.01 by non-parametric Kruskal-Wallis test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

Table 3 - Clinicopathological data of Breast Cancer and normal breast samples included in the study.

	Fresh-Frozen Series		FFPE Series
Clinicopathological Features	**BrCa**	**NBr**	**BrCa**
Patients (n)	176	26	16
Age median (range)	61 (41-75)	54 (40-70)	58 (35-78)
Molecular subtype[1] (%)			
Luminal A	56 (31.8)		4 (25.0)
Luminal B	80 (45.4)	n.a.	12 (75.0)
HER2-enriched	12 (6.8)		-
Basal-like	28 (15.9)		-
Histological type (%)			
Invasive carcinoma of NST	155 (88.1)		13 (81.3)
Invasive lobular carcinoma	11 (6.3)	n.a.	3 (18.8)
Other special subtype carcinoma	2 (1.1)		-
Mixed type carcinoma	8 (4.5)		-
Grade (%)			
G1	21 (11.9)		2 (12.5)
G2	76 (43.2)	n.a.	4 (25.0)
G3	67 (38.1)		8 (50.0)
Gx	12 (6.8)		2 (12.5)
ER receptor *status* (%)			
Positive	136 (77.3)	n.a.	16 (100.0)
Negative	40 (22.7)		-
PR receptor *status* (%)			
Positive	110 (62.5)	n.a.	10 (62.5)
Negative	66 (37.5)		6 (37.5)
HER2 receptor *status* (%)			
Positive	28 (15.9)	n.a.	3 (18.8)
Negative	148 (84.1)		13 (81.3)
T Stage (%)			
T1	47 (29.9)		4 (25.0)
T2	97 (55.1)		9 (56.3)
T3	6 (3.4)	n.a.	1 (6.3)
T4	7 (4.0)		2 (12.5)
Tx	19 (10.8)		-
N Stage (%)			
N0	65 (36.9)		5 (31.3)
N1	63 (35.8)		7 (43.8)
N2	15 (8.5)	n.a.	2 (12.5)
N3	6 (3.4)		2 (12.5)
Nx	27 (15.3)		-
Stage (%)			
I	26 (14.8)		3 (18.8)
II	100 (56.8)	n.a.	8 (50.0)
III	30 (17.0)		5 (31.5)
Not determined	20 (11.4)		-

[1]Assessed by immunohistochemistry. Abbreviations: BrCa – breast cancer; ER – estrogen receptor; FFPE – formalin-fixed paraffin-embedded; G – grade, HER2 – human epidermal growth factor receptor 2; n.a. – not applicable; NBr – normal breast; NST – no special type; PR – progesterone receptor

Table 4 – Detail information about primary tumors and the matched metastases per each patient included in FFPE Breast Cancer patients' cohort.

Patient number	Age at diagnosis	Molecular subtype of primary tumor[1]	Metastasis localization	Time interval after primary tumor (years)
1	39	Luminal B	Lung	20.43
2	60	Luminal A	Axillary lymph node	16.07
3	36	Luminal B	Bone marrow	3.45
4	35	Luminal B	Liver	11.05
5	74	Luminal B	Pleural	11.75
6	64	Luminal B	Liver	3.54
7	78	Luminal B	Breast Skin	2.73
8	61	Luminal B	Bone	2.76
9	43	Luminal A	Axillary lymph node	11.68
10	55	Luminal B	Breast Skin	6.55
11	51	Luminal A	Lung	6.43
12	63	Luminal B	Pleural	2.90
13	56	Luminal B	Breast skin	3.48
			Axillary lymph node	4.59
14	66	Luminal A	Mediastinum	8.53
			Esophagus	8.93
15	51	Luminal B	Contralateral breast	6.44
			Axillary lymph node	6.52
			Pleural	11.02
			Contralateral breast skin	11.39
16	60	Luminal B	Bone marrow	1.51
			Skin	3.38

[1]Assessed by immunohistochemistry.

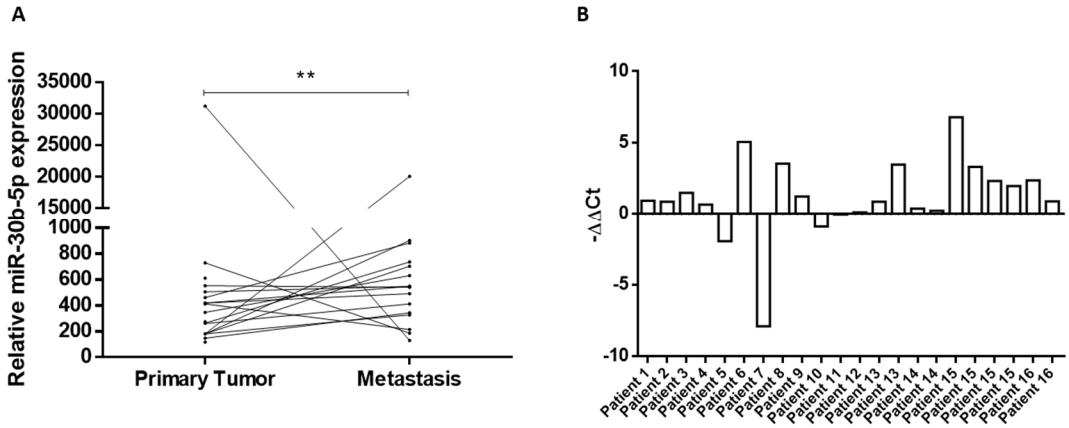

Figure 8 - (A) *MiR-30b-5p* relative expression levels in primary tumors and the corresponding paired metastasis. ** *p*-value <0.01 by non-parametric Wilcoxon paired sample test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000. **(B)** Comparison of *miR-30b-5p* expression in primary breast tumors *versus* corresponding metastasis. X-axis represents each patient. Y-axis denotes -ΔΔCt values, corresponding positive values to higher expression in the distant metastasis compared to the corresponding primary tumor.

AIMS

Aims

BrCa is the second most common cancer worldwide and, by far, the most frequently diagnosed cancer in women. Notwithstanding, BrCa remains the leading cause of cancer-related death in women, mainly due to development of recurrent and/or metastatic disease. Consequently, there is an urge to bring out novel minimally invasive biomarkers for advanced BrCa.

The hypothesis to be tested in this Master Dissertation is that *miR-30b-5p* might be involved in the metastatic cascade and might be biologically and clinically relevant for advanced BrCa patients. Thus, the major goal is to test *miR-30b-5p* expression levels as biomarker for prediction of progression and prognosis of BrCa and to assess the feasibility of using *miR-30b-5p* as a biomarker of advanced disease in liquid biopsies. This might provide a tool required to plan the treatment to maximize efficacy and improve personalized advanced BrCa treatments at the time of diagnosis.

Hence, the specific tasks were:

- Validate *miR-30b-5p* expression level in a large series of FFPE metastatic breast tissues and the paired primary tumors to analyze its potential value as prognostic biomarker;
- Assess *miR-30b-5p* value as a non-invasive biomarker to discriminate advanced BrCa from localized disease;
- Evaluate if *miR-30b-5p* might be involved in modulation of metastatic organ tropism.

MATERIAL AND METHODS

Patients and Samples Collection

FFPE primary and metastatic tumors from BrCa patients available at Portuguese Oncology Institute of Porto (IPO-Porto) were included in this study (validation cohort #1). Relevant clinical data was retrieved from patients' charts. All cases were revised by an experienced pathologist and graded according to Bloom and Richardson's Modified system and staged according to the AJCC system (22, 26). Paraffin-embedded histological sections (4 μm of thickness) were cut from each tissue block and stained with hematoxylin and eosin (H&E) staining, followed by a pathologist examination to select the most representative tumor lesion. Tumor areas identified were then macrodissecte in 6 consecutive 8 μm sections for tumor cells enrichment (>80%).

Additionally, peripheral blood samples from 20 patients with localized BrCa and 25 patients with advanced BrCa were collected at IPO-Porto after informed patients' consent (validation cohort #2). Briefly, peripheral blood was collected into EDTA-containing tubes and centrifuged at 2000 rpm for 10 minutes at 4ºC. Plasma was immediately separated, aliquoted into 1.5 mL tubes and properly stored at -80ºC until further use.

This study was approved by institutional ethical committee (CES 120/015). Patients' blood samples collection was approved by the institutional review board of IPO-Porto (CES-IPOFG-EPE 019/08) and was performed in accordance with the Declaration of Helsinki.

Cell Line Characterization

Three BrCa cell lines were used in the present study: BT-474 (isolated from a solid, invasive ductal mammary carcinoma from a 60 years old female), and metastatic cell lines MDA-MB-231 (isolated from pleural effusion from a 51 years old woman with a metastatic adenocarcinoma of the breast) and Bo-1833 (a MDA-MB-231 subpopulation that preferentially metastasize to the bone) (102). All the cell lines were grown in Dulbecco's Modified Eagle Medium (DMEM) high glucose supplemented with 10% fetal bovine serum (FBS) and 10 μg/mL Penicillin-Streptomycin (Pen/Strep). The characterization of breast cell lines and the culture medium are summarized **Table 5**.

Biological triplicates of each cell pellets and cells' conditioned mediums were kindly provided by Professor Meriem Lamghari from the Neuro-Skeletal Circuits Group of INEB (Portugal). All samples were stored at -80ºC until further RNA extraction.

Table 5 – Characterization of Breast Cancer cell lines selected.

Cell Lines	Derived From	Molecular Subtype	Immunoprofile (103)	Culture medium	ATCC Reference
BT-474	Solid, invasive ductal carcinoma of the breast	Luminal B	ER^{+}, $PR^{+/-}$, $HER2^{+}$	DMEM High Glucose	HTB20™
MDA-MB-231	Metastatic site: pleural effusion	Claudin-low	ER^{-}, PR^{-}, $HER2^{-}$		HTB26™
Bo-1833	–	–	–		–

Abbreviations: ATCC- American Type Culture Collection; DMEM- Dulbecco's Modified Eagle's medium; ER- estrogen receptor; HER2- human epidermal growth factor receptor 2; PR- progesterone receptor

RNA Extraction

RNA was extracted from paraffin-embedded histological sections (12 µm of thickness) from FFPE tissue blocks, using a commercially extraction kit (FFPE RNA/DNA Purification Plus Kit, Norgen Biotek, Thorold, Canada) following manufacturer's instructions. Briefly, FFPE samples were deparaffinized and digested with proteinase K [20mg/mL (NZYTECH, Portugal)] and digestion buffer provided by the extraction kit for 15 minutes at 55°C. Then, samples were centrifuged, the RNA-containing supernatant was transferred to a new RNase-free tube and the DNA-containing pellet was stored at -20ºC. The provided buffer and absolute ethanol were added to the RNA-containing solution which was loaded into an RNA Purification Micro Column to proceed the RNA binding. Finally, the RNA bonded to the column was washed with the provided wash solution and eluted in 15 or 30 µL of elution solution according to the initial sample amount.

Circulating RNA extraction from plasma samples was performed using miRNeasy Serum/Plasma Kit (Qiagen, Hilden, Germany). Briefly, to 200 µL of plasma were added 1 mL of QIAzol Lysis Reagent (Qiagen) to denature proteins and 200 µL of chloroform (Merck, Darmstadt, Germany). Then, the samples were centrifuged to perform phase separation. Absolute ethanol was added to 600 µL of RNA-containing upper aqueous phase which was loaded into a RNeasy MinEluate Spin Column to proceed the RNA binding. The RNA bonded to the column was washed with the provided buffers and 80% ethanol. Finally, RNeasy MinEluate Spin Column were centrifuged at full speed with opened lids to dry the membrane and RNA was eluted in 14 µL of provided RNase-free water.

Total RNA from cells was extracted by suspended cell pellet samples in 500 µL of TRIzol® reagent (Invitrogen, USA) and 100 µL of chloroform (Merck, Darmstadt, Germany). After an incubation time, samples were centrifuged to perform phase separation and 250 µL of isopropanol were added to RNA-containing upper aqueous phase. The mixture was shaken

vigorously and incubated for RNA precipitation. Then, the samples were centrifuged, supernatants discharged, and pellets washed twice with 75% ethanol. Finally, air dried RNA pellets were eluted in 15 µL sterile distilled water (B. Braun, Melsungen, Germany).

Extraction of miRNA from cells' conditioned mediums was also performed using GRS microRNA kit (GRiSP, Porto, Portugal) with a protocol optimized by Francisca Dias from Molecular Oncology and Viral Pathology Group of CI-IPO-Porto (Portugal).

All RNA concentrations and purity ratios were posteriorly measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and RNA samples were stored at -80°C until use.

MicroRNA cDNA Synthesis

The cDNA synthesis from FFEP and cells RNA was performed using miRCURY LNA™ Universal RT microRNA PCR (Exiqon, Vedbaek, Denmark) in accordance with the manufacturer's instructions. Briefly, RNA samples concentration was adjusted to a 5 ng/µL concentration, using sterile distilled water (B.Braun, Melsungen, Germany). Reverse transcription reaction working-solution was prepared by adding 5 µL of nuclease-free water, 2 µL of 5x Reaction Buffer and 1 µL of enzyme mix for each reaction. On ice, per each RNase-free PCR tube, it was added 8 µL of reverse transcription reaction working-solution and 2 µL of previously concentration-adjusted RNA. Then, RNase-free PCR tubes were gently vortexed and incubated on Veriti® Thermal Cycler (Applied Biosystems, Foster City, CA, USA) for 60 minutes at 42°C, followed by 5 minutes at 95°C to inactivate the reverse transcriptase. Finally, cDNA samples were diluted 20x in sterile distilled water and stored at -20ºC.

Circulating RNA and cells' conditioned medium RNA were reverse transcribed to cDNA using Taqman® Advanced miRNA cDNA Synthesis Kit (Applied Biosystems, Foster City, CA, USA) in accordance with the manufacturer's instructions. The protocol consists in four major steps that improved sensitivity for low-abundant miRNA targets as follows: first it is performed a polyadenylation of the miRNA at the 3' end, followed by the adaptor ligation at the 5' end, that acts as the forward primer binding site for miRNA amplification reaction; then, it is performed reverse transcription reaction using universal reverse transcription primers which bind to the 3' poly(A) tail and finally, miRNA amplification reaction using universal forward and reverse primers to increase the number of cDNA molecules. Briefly, RNA samples concentration was adjusted to a 30 ng/µL concentration, using sterile distilled water (B.Braun, Melsungen, Germany). Poly(A) reaction working-solution was prepared

and on ice, to each RNase-free PCR tube, it was added 3 µL of poly(A) reaction working-solution and 2 µL of previously concentration-adjusted RNA. PCR tubes were incubated on Veriti® Thermal Cycler (Applied Biosystems, Foster City, CA, USA) for 45 minutes at 37°C to performed polyadenylation, followed by 10 minutes at 65°C to stop reaction. Then, to each PCR tube containing the poly(A) tailing reaction product, it was added 10 µL of adaptor ligation reaction working-solution and PCR tubes were incubated for 60 minutes at 16°C. Afterward, to each PCR tube containing the adaptor ligation reaction product, it was added 15 µL of reverse transcription reaction working-solution and PCR tubes were incubated for 15 minutes at 42°C, followed by 5 min at 85°C. Finally, per each 5 µL of reverse transcription reaction product was added 45 µL of miRNA amplification reaction working-solution and PCR tubes were incubated for 5 minutes at 95°C, followed by 3 s at 95°C and 30 s at 60°C replicated for 14 cycles and 10 minutes at 99°C. MiR-Amp reaction products were diluted 10x in sterile distilled water and stored at -20ºC.

MicroRNA Expression Assay

For the detection of cDNA derived from tissue samples and cells, per each well was added: 5 µL of Xpert Fast SYBR (2X) (GRiSP, Porto, Portugal), 1 µL of miRNA specific primer mix (microRNA LNA™ PCR primer set, Exiqon) and 4 µL of previously diluted cDNA. The forward and reverse primes are miRNA specific and optimized with LNA™, allowing a higher sensitivity and specificity as well as low background enabling accurate quantification of very low miRNAs levels.

For detection of cDNA derived from circulating miRNAs and cells' conditioned medium, per each well, it was added: 5 µL of Xpert Fast Probe (2X) (GRiSP, Porto, Portugal), 0.5 µL of TaqMan® Advanced miRNA Assay (20X) and 4.5 µL of diluted cDNA.

Quantitative real-time PCR (RT-qPCR) reactions were performed in 384-well plates. Each amplification reaction was performed in triplicate on a LightCycler 480 instrument (Roche Diagnostics, Manheim, Germany) and each plate contained 2 negative template controls.

For the intercalating green dye chemistry, RT-qPCR protocol consisted in a denaturation step at 95ºC for 2 minutes, followed by 40 amplification cycles at 95ºC for 5 seconds and 60ºC for 20 seconds. Melting curve analysis was performed according to instrument's manufacturer recommendations. For the probe-detection technology, RT-qPCR protocol consisted in a denaturation step at 95ºC for 3 minutes, followed by 45 amplification cycles at 95ºC for 10 seconds and 60ºC for 25 seconds.

SNORD38B was used as a reference gene for normalization. In **Table 6** are present the target sequences of reference gene and miRNA analyzed.

Table 6 - Specific target sequence of reference gene and target microRNA.

Gene	Target sequence
SNORD38B (hsa)	UCUCAGUGAUGAAAACUUUGUCCAGUUCUGCUACUGACAGUAAGUGAAGAUA AAGUGUGUCUGAGGAGA
hsa-miR-30b-5p	UGUAAACAUCCUACACUCAGCU

The relative miRNA expression in each tissue RNA sample and cells samples was calculated by the $2^{-\Delta CT}$ method, using the formula:

$$\text{Relative miRNA expression} = 2^{-\Delta Ct}, \text{ in which } \Delta Ct = Ct_{\text{target microRNA}} - Ct_{\text{reference}}$$

For plasma samples, the relative miRNA expression was calculated by the formula:

$$\text{Relative miRNA expression} = (\text{Mean quantity}_{\text{target microRNA}} / \text{Mean quantity}_{\text{reference}}) \times 1000$$

Herein, a five serial 10x dilutions of a positive control was run in each plate to generate a standard curve.

The relative miRNA expression in cells' conditioned medium samples was calculated by Livak method ($2^{-\Delta\Delta CT}$) (104).

Statistical Analysis

Non-parametric Mann-Whitney U test and Kruskal-Wallis test, followed by Mann-Whitney U tests when appropriate, were used to ascertain the statistical significance of differences in continuous variables among two or more independent datasets, respectively. Bonferroni correction was applied to pairwise comparisons. Differences between paired samples were analyzed using non-parametric Wilcoxon paired sample test. Fold changes for miRNA were calculated using the $2^{-\Delta\Delta CT}$ method (104). Spearman nonparametric correlation test was performed to assess the association between continuous variables.

Receiver Operating Characteristic (ROC) curve was constructed and biomarker performance parameters (sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy) were calculated. Cut-off was established based on the highest value obtained in ROC curve analysis based on Youden's J index (105, 106).

In cell lines, statistical significance for continuous variables comparisons between more than two independent samples was assessed by One-Way Analysis of Variance (one-Way ANOVA) test, followed by Bonferroni correction.

Statistical analysis of obtained data was performed using SPSS software (SPSS Version 20.0, Chicago, IL) and two-tailed *p*-values were considered statistically significant when $p<0.05$. Graphics were built using GraphPad 6 Prism (GraphPad Software, USA).

RESULTS

Validation cohort #1 (FFPE)

Characterization of validation cohort #1

The validation cohort #1 was composed of a larger set of 175 tumor samples (82 primary BrCa and 93 paired metastases) from 82 BrCa patients (**Table 7** and **Appendix IV**).

Table 7 - Clinicopathological data of Breast Cancer patients of the validation cohort #1.

Clinicopathological Features	Validation Cohort #1
Patients (n)	82
Age median (range)	49 (28-76)
Molecular subtype[1] (%)	
Luminal A	19 (23.2)
Luminal B	58 (70.7)
HER2-enriched	1 (1.2)
Basal-like	4 (4.9)
Histological type (%)	
Invasive carcinoma of NST	73 (89.0)
Invasive lobular carcinoma	-
Other special subtype carcinoma	-
Mixed type carcinoma	9 (11.0)
Grade (%)	
G1	4 (4.9)
G2	43 (52.4)
G3	35 (42.7)
Gx	-
ER receptor *status* (%)	
Positive	77 (93.9)
Negative	5 (6.1)
PR receptor *status* (%)	
Positive	65 (79.3)
Negative	17 (20.7)
HER2 receptor *status* (%)	
Positive	15 (18.3)
Negative	67 (81.7)
T Stage (%)	
T1	20 (24.4)
T2	52 (63.4)
T3	5 (6.1)
T4	3 (3.7)
Tx	2 (2.4)
N Stage (%)	
N0	18 (22.0)
N1	33 (40.2)
N2	16 (19.5)
N3	13 (15.9)
Nx	2 (2.4)
Stage (%)	
I	10 (12.2)
II	34 (41.5)
III	24 (29.3)
IV	12 (14.6)
Not determined	2 (2.4)

[1]Assessed by immunohistochemistry. Abbreviations: ER – estrogen receptor; G – grade; HER2 – human epidermal growth factor receptor 2; NST – no special type; PR – progesterone receptor

Among the 93 available paired metastases, 63 were from bone, 17 from lung, 4 from brain and 9 were locoregional or from contralateral breast. It should be noted that 10 patients had multiple metastases with different locations (**Appendix IV**). Overall, the time elapsed between diagnosis of the primary tumor and of the metastasis varied from 0.15 to 18.98 years (median 6.63 years).

Evaluation of *miR-30b-5p* expression levels in validation cohort #1

Once *miR-30b-5p* was differentially expressed between primary tumors and matched metastases in a small set of tumor tissue samples from BrCa patients, its expression level was evaluated in the validation cohort #1 (**Table 7** and **Appendix IV**) to analyze its potential value as prognostic biomarker.

MiR-30b-5p expression levels were significantly higher in metastases than in primary tumors (p<0.0001,**Figure 9**), confirming findings in the small set of tumor tissue samples.

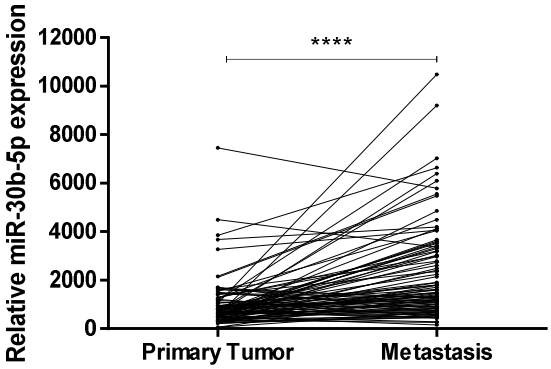

Figure 9 - *MiR-30b-5p* relative expression levels in primary tumors and the corresponding matched metastases. **** p-value <0.0001 by non-parametric Wilcoxon paired sample test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

Interestingly, primary tumors that metastasized to bone disclosed significantly higher *miR-30b-5p* expression levels compared to all other primary tumors (p=0.002, **Figure 10A**). Moreover, bone metastases displayed significantly higher *miR-30b-5p* expression levels than all samples from other metastatic sites (p<0.0001, **Figure 10B**).

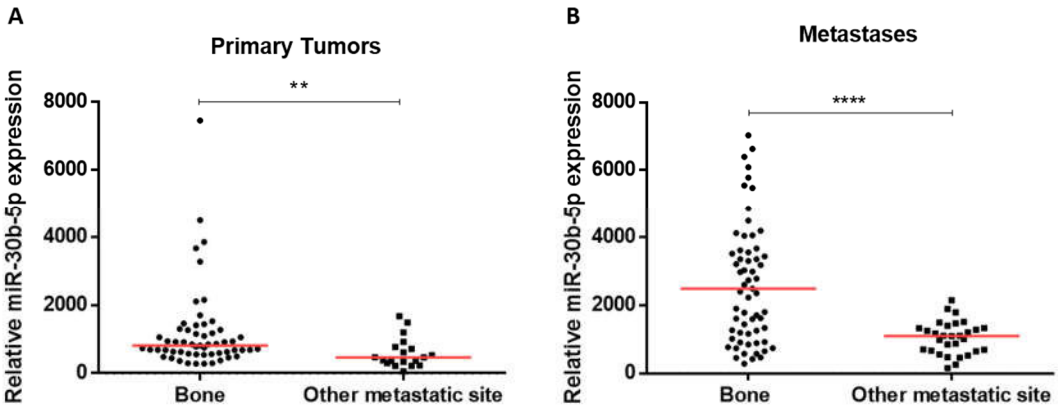

Figure 10 - Scatter-plots of *miR-30b-5p* relative expression in primary tumors **(A)** and metastases **(B)**. ** *p*-value <0.01 and **** *p*-value <0.0001 by non-parametric Mann-Whitney U test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

MiR-30b-5p expression levels: association with clinicopathological features

Except for HER2 *status*, no statistically significant associations were found between *miR-30b-5p* expression levels and any of the clinicopathological parameters (age, histological type, grade, TNM staging, molecular subtype assessed by IHC, ER and PR *status*). Indeed, HER2-negative tumors depicted significantly higher *miR-30b-5p* expression levels compared to HER2-positive BrCa (*p*=0.041, **Figure 11**).

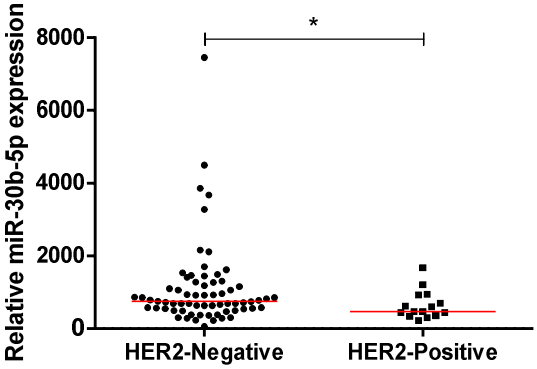

Figure 11 - Scatter-plot of *miR-30b-5p* relative expression according to the HER2 receptor *status*. * *p*-value <0.05 by non-parametric Mann-Whitney U test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

Validation cohort #2 (Plasma)

Characterization of validation cohort #2

The validation cohort #2 was composed of 20 patients with localized BrCa (stage I) and 25 patients with advanced BrCa, comprising both locally advanced (n=12) and metastatic BrCa (n=13) (**Table 8**). No significant differences were found for patients' age between localized and advanced BrCa (*p*=0.417).

Table 8 - Clinicopathological data of Breast Cancer patients of the validation cohort #2.

Clinicopathological Features	Localized BrCa	Advanced BrCa
Patients (n)	20	25
Age median (range)	61 (39-71)	53 (35-82)
Molecular subtype[1] (%)		
Luminal A	7 (35.0)	2 (8.0)
Luminal B	13 (65.0)	16 (64.0)
HER2-enriched	-	5 (20.0)
Basal-like	-	2 (8.0)
Histological type (%)		
Invasive carcinoma of NST	17 (85.0)	18 (72.0)
Invasive lobular carcinoma	2 (10.0)	4 (16.0)
Other special subtype carcinoma	1 (5.0)	2 (8.0)
Mixed type carcinoma	-	1 (4.0)
Grade (%)		
G1	4 (20.0)	-
G2	9 (45.0)	18 (72.0)
G3	7 (35.0)	6 (24.0)
Gx	-	1 (4.0)
ER receptor *status* (%)		
Positive	20 (100.0)	18 (72.0)
Negative	-	7 (28.0)
PR receptor *status* (%)		
Positive	20 (100.0)	14 (56.0)
Negative	-	11 (44.0)
HER2 receptor *status* (%)		
Positive	4 (20.0)	11 (44.0)
Negative	16 (80.0)	14 (56.0)
T Stage (%)		
T1	20 (100.0)	3 (12.0)
T2	-	10 (40.0)
T3	-	5 (20.0)
T4	-	5 (20.0)
Tx	-	2 (8.0)
N Stage (%)		
N0	20 (100.0)	-
N1	-	3 (12.0)
N2	-	-
N3	-	20 (80.0)
Nx	-	2 (8.0)
Stage (%)		
IA	20 (100.0)	n.a.
IIIC	n.a.	12 (48.0)
IV	n.a.	13 (52.0)

[1]Assessed by immunohistochemistry. Abbreviations: BrCa – breast cancer; ER – estrogen receptor; G – grade; HER2 – human epidermal growth factor receptor 2; n.a – not applicable; NST – no special type; PR – progesterone receptor

Assessment of *miR-30b-5p* expression levels as prognostic biomarker in liquid biopsies

A plasma BrCa patient cohort was used for validation in liquid biopsies (**Table 8**). Remarkably, patients with advanced BrCa displayed higher plasma *miR-30b-5p* expression levels (p<0.0001) than patients with localized disease (**Figure 12A**). Moreover, when advanced BrCa group was stratified in locally advanced and metastatic BrCa, *miR-30b-5p* expression levels were significantly higher in both groups (p=0.0002 and p=0.021, respectively) compared to localized BrCa (**Figure 12B**).

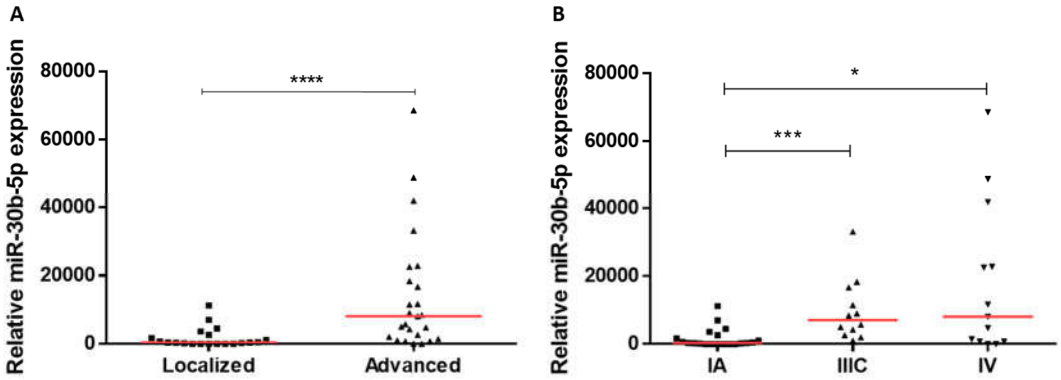

Figure 12 - (A) Scatter-plots of plasmatic *miR-30b-5p* relative expression in localized and advanced Breast Cancer. **** p-value <0.0001 by non-parametric Mann-Whitney U test. **(B)** Scatter-plots of plasmatic *miR-30b-5p* relative expression according to stage. * p-value <0.05 and *** p-value <0.001 by non-parametric Kruskal-Wallis test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

ROC analysis revealed that plasma *miR-30b-5p* expression levels could discriminate advanced from localized BrCa patients with an area under the curve (AUC) of 0.831 (95% CI = 0.721-0.950). Using a cut-off value of 4611, plasma *miR-30b-5p* expression identified advanced disease with 88.9% sensitivity, 66.7% specificity and 75.6 accuracy (**Figure 13** and **Table 9**).

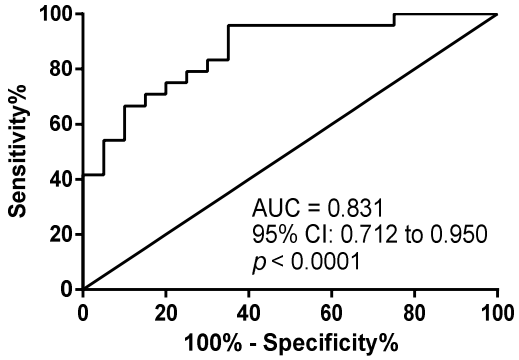

Figure 13 - ROC curve analysis to evaluate the potential of *miR-30b-5p* as a biomarker for discriminate patients with advanced Breast Cancer from patients with localized Breast Cancer.

Table 9 - Performance of *miR-30b-5p* as biomarker for discriminate advanced Breast Cancer from localized Breast Cancer.

Sensitivity%	Specificity %	PPV %	NPV %	Accuracy %
88.9	66.7	64.0	90.0	75.6

Abbreviations: PPV – positive predictive value; NPV – negative predictive value

Association between *miR-30b-5p* expression levels and clinicopathological features

MiR-30b-5p expression levels were significantly higher in plasma samples from patients with T2 tumors and with positive axillary lymph node (p=0.012, **Figure 14A** and p<0.0001, **Figure 14B** respectively).

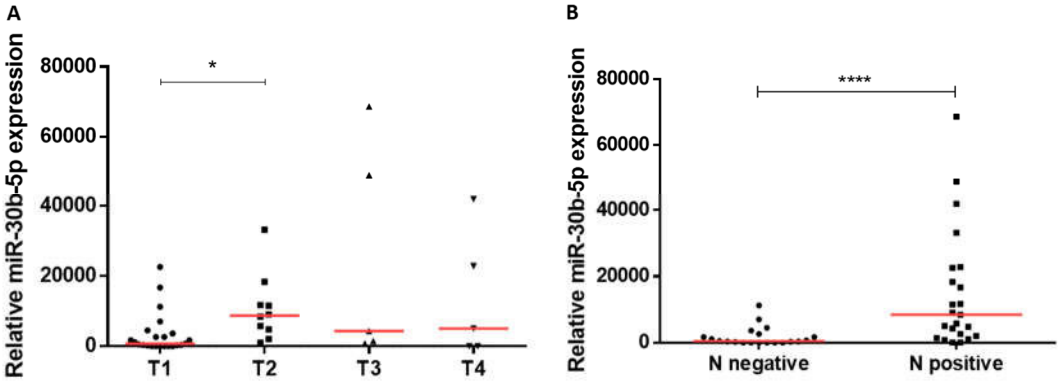

Figure 14 - Scatter-plots of *miR-30b-5p* relative expression according to T stage **(A)** and N stage **(B)**. * p-value <0.05 and **** p-value <0.0001 by non-parametric Mann-Whitney U test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

Patients with distant metastases at diagnosis displayed higher *miR-30b-5p* expression levels, although without statistically significance (p=0.073, **Figure 15**).

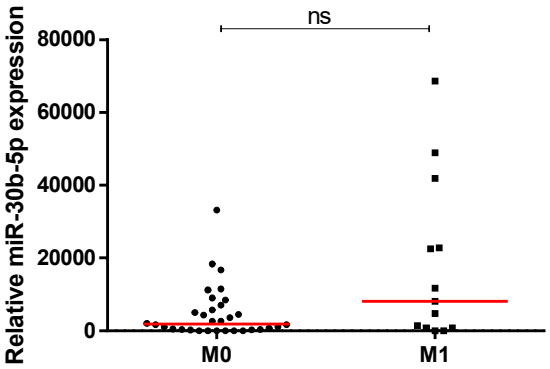

Figure 15 - Scatter-plot of *miR-30b-5p* relative expression according to the presence or absence of distant metastases at diagnosis. A ns denotes p-value >0.05 by non-parametric Mann-Whitney U test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

No significant differences were observed for *miR-30b-5p* expression levels and any of clinicopathological parameters (age, histological type, grade, molecular subtype assessed by IHC, HER2, ER and PR *status*).

Assessment of *miR-30b-5p* expression levels in cell lines and cells' conditioned mediums

Considering the previous results in patients' samples, *miR-30b-5p* expression levels were evaluated in *in vitro* BrCa models.

In order to investigate the possible involvement of *miR-30b-5p* on modulation of metastatic organ tropism, three breast cell lines were selected: BT-474 cells, isolated from a primary carcinoma of the breast, MDA-MB-231 cells, a metastatic cell line and Bo-1833 cells, a subclone with tropism to bone **(Figure 16).**

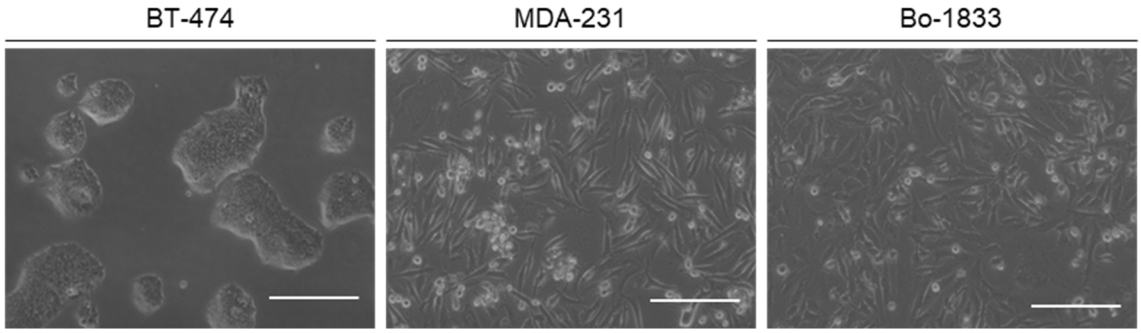

Figure 16 - Morphological phenotype of Breast Cancer cell lines. Photographs taken in microscope Olympus CKX41 (100x magnification). Scale bar denotes 200 µm. Photographs kindly provided by Catarina Lourenço from the Neuro-Skeletal Circuits Group of INEB (Portugal).

Firstly, intracellularly *miR-30b-5p* expression levels were assessed in each cell line. BT-474 cells displayed significantly higher *miR-30b-5p* expression levels when compared to MDA-MB-213 and Bo-1833 (p=0.0028 and p=0.0012, respectively, **Figure 17**). Regarding metastatic cell lines, no significant differences were depicted between *miR-30b-5p* expression levels.

To evaluate the hypothesis that BrCa cells might release *miR-30b-5p*, its extracellular expression was analyzed. Once it was already reported that the medium composition might influence the results of this analysis (107), *miR-30b-5p* levels were also evaluated in unconditioned culture medium, to ensure that the data observed result from cell secretion rather than medium components.

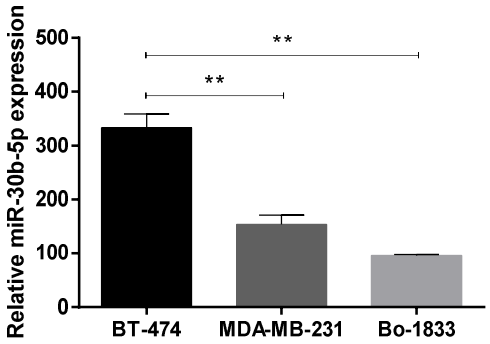

Figure 17 - Intracellularly *miR-30b-5p* expression levels in BT-474, MDA-MB-231 and Bo-1833 cells. ** *p*-value <0.01 by one-Way ANOVA test. Y-axis denotes $2^{-\Delta CT}$ values multiplied by 1000.

Interestingly, BT-474 cells conditioned medium showed *miR-30b-5p* expression, while no expression was detected in conditioned medium of MDA-MB-231 and Bo-1833 cells (**Figure 18**).

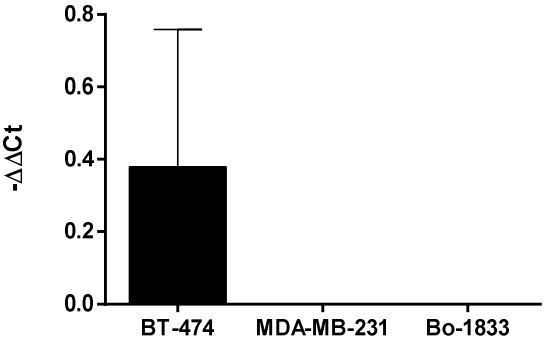

Figure 18 - Comparison of *miR-30b-5p* expression in cells' conditioned mediums *versus* cells' culture medium. X-axis represents each cells' conditioned medium. Y-axis denotes -ΔΔCt values, corresponding positive values to higher expression in the cells' conditioned medium compared to the cells' culture medium.

DISCUSSION

Discussion

BrCa remains the most common malignancy in women and a major cause of morbidity and mortality (1). Although biological features are routinely used for BrCa diagnosis and prognosis assessment, patients with similar clinicopathological features often show different clinical outcome (20). Therefore, identification of biomarkers providing more accurate prognostic information for BrCa patients, complementing currently used parameters, will have a major impact. Hence, assessment of specific miRNAs expression deregulation, which has been associated with several mechanisms underlying BrCa aggressiveness, might be a potential source for biomarkers (64, 76). Most studies addressing miRNAs expression and miRNA-target validation have been performed in cancer cell lines and display several limitations, including absence of epithelial-stromal and tumor-host interactions, that could modulate prognosis *in vivo* (64, 108). Thus, tissue analysis might allow for broader insight into biologically and clinically relevant miRNAs which may serve as prognostic biomarkers. In a previous study, *miR-30b-5p* expression levels were reported to be significantly higher in BrCa tissues from patients with advanced disease (N stage: N3). Moreover, in FFPE primary and metastatic tumors available from 16 BrCa patients, *miR-30b-5p* expression was significantly higher in metastatic lesions compared to matched primary BrCa tissues, suggesting a role in promoting metastasis development and therefore a value as prognostic/progression biomarker (*unpublished* observations).

Hence, we aimed to validate these results in a larger set of tumor tissue samples from BrCa patients (validation cohort #1) in order to evaluate the biomarker potential of *miR-30b-5p* for predicting advanced disease. It should be recalled that stability of miRNAs in FFPE tissues holds an enormous potential (109), especially in BrCa patients in which late relapses frequently occur, as demonstrated in validation cohort #1. However, due to the limited availability of metastatic tissue samples, only a few studies compared miRNA expression levels between primary and correspondent distant metastases (97, 110, 111). Although downregulation of *miR-30* family (*miR-30f*) members and its role as tumor suppressor during BrCa local invasion and metastization have been previously described (93, 94), to the best of our knowledge, *miR-30b-5p* upregulation in BrCa metastases has not been reported thus far. The role of *miR-30b-5p* remains controversial. On the one hand, expression of *miR-30b-5p* and *miR-30c-5p* has been associated with increased cell viability and resistance to apoptosis (112) and *miR-30b-5p* was found to be upregulated in bladder cancer (113), medulloblastoma (114), advanced oral squamous cell carcinoma (115) and associated with metastasis in melanoma (116). On the other hand, *miR-30b-5p* was associated with decreased migration and invasiveness in colorectal cancer (117, 118), and *miR-30a* was reported to be downregulated in primary nasopharyngeal carcinoma tissues

but overexpressed in the corresponding metastases (119), although *miR-30f* members were shown to inhibit early steps of the metastatic process (120). Our data clearly support an oncogenic role for *miR-30b-5p* in BrCa.

Progression of solid malignancies is the result of a multistep cascade in which tumor cells undergo widespread modifications to successfully migrate and colonize other organs. EMT is key for the initial escape of tumor cells by enabling increased cell migration and invasion. Once circulating, tumor cells extravasate from the blood vessels and recover their epithelial properties by undergoing MET (121). In support of the MET hypothesis, several studies have shown that metastatic lesions and the corresponding primary breast tumor have a similar epithelial nature (122, 123). Thus, the dynamic ability to first undergo EMT and subsequently MET is an important feature of metastatic cells. *MiR-30b-5p* modulation might be important in this plastic process. In fact, a recent study showed that decreased *miR-30f* members in BrCa patients without evidence of distant metastases was associated with poor relapse-free survival, which might be associated with the ability of decreased *miR-30f* levels to speed EMT initiation (124). Downregulation of *miR-30f* members might lead to EMT initiation enabling cells to metastasize, while subsequent upregulation might be associated with MET, facilitating re-adaptation of the epithelial phenotype and colonization, crucial to develop macroscopic metastases. Indeed, a similar context-depending role in metastasis has been described for *miR-200* family members (96, 97). Therefore, additional studies are needed to ascertain *miR-30b-5p* functional role in BrCa.

Knowledge of determining patterns of metastatic organ tropism might provide useful information for clinical evaluation of disease stage and to monitor progression. Hence, comparative analyses of *miR-30b-5p* expression according to metastatic site were performed. Interestingly, bone metastases disclosed significantly higher *miR-30b-5p* expression levels compared to other metastases and, remarkably, primary BrCa cases that metastasized to bone also displayed increased levels compared to those that did not. These results strongly suggest that not only *miR-30b-5p* play a role in metastization, but it also predisposes tumor cells to homing at specific organ sites, especially promoting bone colonization by tumor cells. Nevertheless, it should be recalled that miRNAs expression is highly context- and tissue-dependent, and thus, ideally, miRNA expression in normal tissues more prone to receive metastatic cells should also be assessed. Moreover, the mechanisms underlying tumor cell tropism to bone and the extent to which metastatic cells miRNA's profile differ according to their location might add valuable insights into disease development and clinical management.

Circulating miRNAs are stable in body fluids and their assessment might provide valuable diagnostic, prognostic and therapeutic prediction information, allowing for non-invasive

testing and potential individual treatment optimization (64). Recently, *miR-30b-5p* expression levels were shown to distinguish BrCa patients from healthy controls in liquid biopsies (125), although these levels have been associated with aged (126). Remarkably, we found that *miR-30b-5p* could discriminate patients with advanced BrCa from those with localized BrCa with high sensitivity, but modest specificity and overall accuracy, and no association with age was disclosed. Our results suggest that *miR-30b-5p* might identify, at diagnosis, patients which are more likely to endure disease progression.

Several other miRNAs have been implicated in BrCa invasion and metastasis (64). *MiR-10b* was found highly expressed in tissue samples from patients with metastatic BrCa (88) and, more recently, found to be significantly more expressed in tissues from patients with stage III and IV BrCa compared to early stage disease (127). Moreover, circulating *miR-10b* combined with *miR-373* might identify BrCa lymph node metastasis with 72% sensitivity and 94.3% specificity (128) and *miR-21* overexpression was significantly correlated with lymph node metastasis, advanced clinical stage and poor prognosis (129). Notwithstanding the tissue series size (n=113), no stage IV patients were included in this study. Similarly, circulating *miR-21* discriminated stage IV BrCa patients with visceral metastasis from those with stage I, II and III disease with 86% specificity and 70% sensitivity (130). Nonetheless, as far as we know, these metastasis-related miRNAs were only evaluated in primary BrCa tissue and were not assessed in a larger tissue series of primary tumors and the corresponding metastatic lesions. Furthermore, in our study, plasma *miR-30b-5p* expression levels identified advanced disease with higher sensitivity, although with limited specificity.

Importantly, a careful analysis is mandatory since the origin of tumor-associated miRNAs in circulation is not fully elucidated. Thus, it should be considered that they might be released by the primary tumor, circulating cells or metastatic lesions. Besides, due to the low abundance of miRNAs in circulation, their quantification might be a challenge.

Finally, considering the previous results in patients' tissue samples, particularly the higher *miR-30b-5p* expression levels observed in bone metastases and primary tumors that metastasized to the bone, the contribution of *miR-30b-5p* on modulation of metastatic bone tropism was evaluated using *in vitro* models of BrCa. BT-474 cells, derived from a primary mammary carcinoma, MDA-MB-231 cells, a metastatic cell line derived from a pleural effusion and Bo-1833 cells, a MDA-MB-231 subpopulation that metastasized preferentially to bone were selected.

Even though metastatic samples displayed increased *miR-30b-5p* expression levels when compared to the corresponding primary tumors, regarding *in vitro* study, primary BrCa cell

line showed significantly higher *miR-30b-5p* expression levels when compared to metastatic cell lines. Importantly, a cautious comparison is required once the origin of miRNA was different. It must be considered that primary tumor and metastatic lesions samples were representative of the tumor bulk, where malignant cells had already colonized, whereas MDA-MB-231 cell line derived from a pleural effusion. Concerning the morphologies of BrCa lines, BT-474 cells have formed closely associated colonies, while MDA-MB-231 and Bo-1833 cells were characterized by elongated cell bodies usually associated with motility and invasive features. Previously, Kenny *et al*. classified BrCa cell lines into four morphological categories: Mass, Round, Stellate and Grape-like (131). BT-474 cell line was included in Mass category characterized by tightly cohesive colonies with strong cell-cell adhesion, whereas MDA-MB-231 was allocated to Stellate category characterized by limited cell-cell interactions and lack of E-cadherin expression, which are characteristic of EMT and mesenchymal phenotype (103, 131). Thus, these different phenotypes were in agreement with the previous hypothesis that decreased *miR-30b-5p* expression levels might lead to EMT initiation enabling cells with motility and invasive features. This might explain lower *miR-30b-5p* expression in MDA-MB-231 cells, and subsequent upregulation associated with MET, enabling re-adaptation of the epithelial phenotype, which was observed in BT-474 cells.

Contrary to what was expected, no significant differences were found between MDA-MB-231 and Bo-1833 cells. Bo-1833 cells are a subpopulation of MDA-MB-231 cell line, a triple negative BrCa cell line. It is well-recognized that BrCa patients prone to develop bone metastasis are mainly those which display luminal features (48, 50). Thus, Bo-1833 cell line is a limited experimental model to study bone metastatic process. A recent study verified increased *miR-30f* members expression levels in luminal BrCa cell lines compared to MDA-MB-231. Furthermore, *miR-30f* members were also found to inhibit BrCa bone metastases in an experimental model (124). Nevertheless, these results were only derived from triple negative BrCa cell lines, which represent a (very) limited subset of BrCa patients who do not commonly develop bone metastases, a limitation in our work. Therefore, this experimental model can hardly be considered representative of the clinically apparent heterogeneity.

Additionally, as previously mentioned, *in vitro* studies are characterized by several limitations, including absence of epithelial-stromal, tumor-host interactions and signaling from extracellular matrix (64, 108). Although malignant cells need to acquire features that allow them to proliferate to distant organs, the particular organ microenvironment and its state before metastases development can moderate the metastatic process, namely by modifying gene expression signatures of tumor cells (46, 132, 133). Indeed, in bone,

malignant cells colonization is strongly modulated by the physical microenvironment (132). The majority of BrCa bone metastases lead to osteolytic lesions owing to exceeding bone loss and deficient bone replacement as a result of aberrant induction of osteoclasts activity and inhibition of osteoblasts differentiation (134). BrCa cells cooperate with resident cells by releasing molecules that enhance osteoclasts activity, perturbing the dynamic among osteoclasts and osteoblasts, which might lead to excessive bone degradation (45, 134). Subsequently, growth factors secreted by bone matrix exacerbate malignant cells proliferation and growth factors production, the so called “vicious” cycle of osteolytic bone metastasis (133-135).

Nevertheless, cells’ conditioned medium analysis revealed that contrarily to MDA-MB-231 and Bo-1833 cells, BT-474 cells might release *miR-30b-5p*. This result might be due to the variability of confluence among cell culture flasks. Moreover, several studies have reported that tumor cells might communicate with the microenvironment by releasing small vesicles (30 to 100 nm in diameter), including exosomes (45, 136). These extracellular vesicles comprise several functional biomolecules such as miRNAs already implicated in tumor invasion and metastization by favoring a pro-metastatic environment, a key factor to organ-specific metastases (136, 137). Indeed, release of *miR-21* and *miR-10b* in extracellular vesicles to tumor microenvironment correlated with enhancement of cell viability, growth, and ability to form colonies (138). In our study, the lack of *miR-30b-5p* in conditioned medium of MDA-MB-231 and Bo-1833 cells might be due to the fact that we did not isolate exosomes from the cells’ conditioned mediums, but instead extraction of miRNA not associated to exosomes was performed.

CONCLUSIONS AND FUTURE PERSPECTIVES

Conclusions and Future Perspectives

In this Master Dissertation, we found that *miR-30b-5p* was overexpressed in metastatic BrCa, suggesting an important role in tumor dissemination. Thus, additional studies are required to better understand *miR-30b-5p* role on the plastic process of EMT/MET, which are decisive to invasion and development of macroscopic metastases.

Interestingly, bone metastases and their paired primary tumors displayed higher *miR-30b-5p* expression levels, implicating its involvement in modulation of metastatic organ tropism. As a future perspective, we intend to assess *miR-30b-5p* expression in normal metastasis-host tissues to ascertain whether differential expression of miRNAs in the primary tumors *versus* metastatic tissues might be a consequence of their modulation in the metastatic microenvironment.

Importantly, advanced BrCa patients displayed significantly higher plasmatic *miR-30b-5p* expression levels than patients with localized BrCa, highlighting its potential as non-invasive biomarker to identify BrCa patients at higher risk of disease progression in liquid biopsies. Furthermore, studies in larger multicentric cohorts are needed to further validate the value of *miR-30b-5p* in BrCa management and prognostic. Moreover, it would be interesting to evaluate *miR-30b-5p* expression in additional follow-up analyses to assess whether *miR-30b-5p* expression levels monitoring in plasma might provide a useful tool for early recurrence/metastases detection.

Regarding *in vitro* studies, primary BrCa cell line displayed increased intracellular *miR-30b-5p* levels when compared to metastatic cell lines, whereas no significant difference was verified among MDA-MB-231 and Bo-1833 cells. Since it is already acknowledged that luminal BrCa patients preferentially develop bone metastasis, we intent to evaluate *miR-30b-5p* levels in luminal metastatic cell lines. Once more, it must be considered that *in vitro* models lack clinical heterogeneity found in primary tumors.

Moreover, we plan to evaluate intracellular *miR-30b-5p* expression levels of human osteoblasts and osteoclasts, since several reports have suggested that these cells might share features with tumoral cells, and the other way around was described bone-related genes that allow cancer cells to preferentially metastasized to bone (134).

Additionally, to understand the effect of *miR-30b-5p* released by BrCa cell lines in osteoclasts and osteoblasts dynamic, we intend to establish co-cultures of human osteoblasts/osteoclasts and expose them to BrCa cells' conditioned medium.

Overall, our results support a prognostic value of *miR-30b-5p* expression levels in BrCa. If proven, this marker would provide a useful clinical tool for patient monitoring, entailing

earlier and more effective treatment. This highlight the requirement to standardize experimental conditions, before its clinical application in daily BrCa patients' management.

REFERENCES

1. Ferlay J*, et al.* (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136:** E359-386.
2. DeSantis CE*, et al.* (2015) International Variation in Female Breast Cancer Incidence and Mortality Rates. *Cancer Epidemiol Biomarkers Prev* **24:** 1495-1506.
3. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. (2013) GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide. In: *IARC CancerBase No. 11 [Internet].* Lyon, France: International Agency for Research on Cancer, globocan.iarc.fr.
4. Torre LA*, et al.* (2015) Global cancer statistics, 2012. *CA Cancer J Clin* **65:** 87-108.
5. Ferlay J*, et al.* (2013) Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* **49:** 1374-1403.
6. McGuire A, Brown JA, Kerin MJ. (2015) Metastatic breast cancer: the potential of miRNA for diagnosis and treatment monitoring. *Cancer Metastasis Rev* **34:** 145-155.
7. Redig AJ, McAllister SS. (2013) Breast cancer as a systemic disease: a view of metastasis. *J Intern Med* **274:** 113-126.
8. Burstein HJ, Harris JR, Morrow M. (2011) Malignant tumors of the breast. In: *DeVita, Hellman, and Rosenberg's cancer : principles & practice of oncology.* DeVita VT, Lawrence TS, Rosenberg SA (eds.) Lippincott Williams & Wilkins, pp. 1401-1446.
9. Vogel VG. (2012) Epidemiology of Breast Cancer. In: *Breast Pathology.* Dabbs DJ (ed.) Elsevier Health Sciences, pp. 44-56.
10. *SEER cancer statistics review, 1975-2015. Based on November 2017 SEER data submission, posted to the SEER web site April 2018* [Internet].
11. Anderson WF, Rosenberg PS, Menashe I, Mitani A, Pfeiffer RM. (2008) Age-related crossover in breast cancer incidence rates between black and white ethnic groups. *J Natl Cancer Inst* **100:** 1804-1814.
12. Marshall LM*, et al.* (1997) Risk of breast cancer associated with atypical hyperplasia of lobular and ductal types. *Cancer Epidemiol Biomarkers Prev* **6:** 297-301.
13. McPherson K, Steel CM, Dixon JM. (2000) ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ* **321:** 624-628.
14. Loman N, Johannsson O, Kristoffersson U, Olsson H, Borg A. (2001) Family history of breast and ovarian cancers and BRCA1 and BRCA2 mutations in a population-based series of early-onset breast cancer. *J Natl Cancer Inst* **93:** 1215-1223.
15. Fackenthal JD, Olopade OI. (2007) Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer* **7:** 937-948.
16. Kelsey JL, Gammon MD, John EM. (1993) Reproductive factors and breast cancer. *Epidemiol Rev* **15:** 36-47.

17. van den Brandt PA*, et al.* (2000) Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am J Epidemiol* **152:** 514-527.
18. Singletary KW, Gapstur SM. (2001) Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. *JAMA* **286:** 2143-2151.
19. Wu Y, Zhang D, Kang S. (2013) Physical activity and risk of breast cancer: a meta-analysis of prospective studies. *Breast Cancer Res Treat* **137:** 869-882.
20. Senkus E*, et al.* (2015) Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **26 Suppl 5:** v8-30.
21. Perry N*, et al.* (2008) European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document. *Ann Oncol* **19:** 614-622.
22. Lakhani SR. (2012) *WHO classification of tumours of the breast.* International Agency for Research on Cancer.
23. Tuzlali S. (2016) Pathology of Breast Cancer. In: *Breast Disease: Diagnosis and Pathology.* Aydiner A, İğci A, Soran A (eds.) Springer International Publishing, Cham, pp. 241-266.
24. Schnitt SJ, Lakhani SR. (2014) Breast Cancer. In: *World Cancer Report 2014.* Stewart B, Wild CP (eds.) International Agency for Research on Cancer, Lyon, pp. 362-373.
25. Biomarkers Definitions Working G. (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* **69:** 89-95.
26. Edge SB*, et al.* (2010) AJCC cancer staging manual. In*.* Springer, New York, pp. 364-375.
27. Wu Y, Sahin AA. (2016) Prognostic and Predictive Factors of Invasive Breast Cancer. In: *Breast Disease: Diagnosis and Pathology.* Aydiner A, İğci A, Soran A (eds.) Springer International Publishing, Cham, pp. 187-206.
28. Dai X, Xiang L, Li T, Bai Z. (2016) Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes. *J Cancer* **7:** 1281-1294.
29. Dabbs DJ, Allred CD, Bhargava R. (2012) Predictive and Prognostic Markers Testing in Breast Pathology: Immunophenotypic Subclasses of Disease. In: *Breast Pathology.* Dabbs DJ (ed.) Elsevier Health Sciences, pp. 151-172.
30. Pritchard KI*, et al.* (2006) HER2 and responsiveness of breast cancer to adjuvant chemotherapy. *N Engl J Med* **354:** 2103-2111.
31. Dowsett M*, et al.* (2011) Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst* **103:** 1656-1664.

32. Győrffy B*, et al.* (2015) Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res* **17:** 11.
33. Verma A, Kaur J, Mehta K. (2015) Molecular oncology update: Breast cancer gene expression profiling. *Asian J Oncol* **1:** 65.
34. Perou CM*, et al.* (2000) Molecular portraits of human breast tumours. *Nature* **406:** 747-752.
35. Sørlie T. (2004) Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *Eur J Cancer* **40:** 2667-2675.
36. Parker JS*, et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27:** 1160-1167.
37. Sorlie T*, et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98:** 10869-10874.
38. Sorlie T*, et al.* (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100:** 8418-8423.
39. Kittaneh M, Montero AJ, Gluck S. (2013) Molecular profiling for breast cancer: a comprehensive review. *Biomark Cancer* **5:** 61-70.
40. van 't Veer LJ*, et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530-536.
41. Prat A*, et al.* (2013) Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *Oncologist* **18:** 123-133.
42. Bastien RR*, et al.* (2012) PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Med Genomics* **5:** 44.
43. Cardoso F*, et al.* (2017) 3rd ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 3). *Ann Oncol* **28:** 3111.
44. Garg PK, Prakash G. (2015) Current definition of locally advanced breast cancer. *Curr Oncol* **22:** e409-410.
45. Lambert AW, Pattabiraman DR, Weinberg RA. (2017) Emerging Biological Principles of Metastasis. *Cell* **168:** 670-691.
46. Chambers AF, Groom AC, MacDonald IC. (2002) Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* **2:** 563-572.
47. Minn AJ, Massagué J. (2011) Invasion and Metastasis. In: *DeVita, Hellman, and Rosenberg's cancer : principles & practice of oncology.* DeVita VT, Lawrence TS, Rosenberg SA (eds.) Lippincott Williams & Wilkins, pp. 113-127.
48. Kennecke H*, et al.* (2010) Metastatic behavior of breast cancer subtypes. *J Clin Oncol* **28:** 3271-3277.

49. Wang H*, et al.* (2017) The prognosis analysis of different metastasis pattern in patients with different breast cancer subtypes: a SEER based study. *Oncotarget* **8:** 26368-26379.
50. Sihto H*, et al.* (2011) Breast cancer biological subtypes and protein expression predict for the preferential distant metastasis sites: a nationwide cohort study. *Breast Cancer Res* **13:** R87.
51. Buonomo OC*, et al.* (2017) New insights into the metastatic behavior after breast cancer surgery, according to well-established clinicopathological variables and molecular subtypes. *PLoS One* **12:** e0184680.
52. Coates AS*, et al.* (2015) Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol* **26:** 1533-1546.
53. Association of Breast Surgery at B. (2009) Surgical guidelines for the management of breast cancer. *Eur J Surg Oncol* **35 Suppl 1:** 1-22.
54. Lyman GH*, et al.* (2005) American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol* **23:** 7703-7720.
55. *NCCN Clinical Practice Guidelines–Breast Cancer. V. 2.2016. National Comprehensive Cancer Network Web site* [Internet].
56. Nounou MI*, et al.* (2015) Breast Cancer: Conventional Diagnosis and Treatment Modalities and Recent Patents and Technologies. *Breast Cancer (Auckl)* **9:** 17-34.
57. García-Giménez JL, Ushijima T, Tollefsbol TO. (2016) Chapter 1 - Epigenetic Biomarkers: New Findings, Perspectives, and Future Directions in Diagnostics. In: *Epigenetic Biomarkers and Diagnostics.* Academic Press, Boston, pp. 1-18.
58. Esteller M. (2008) Epigenetics in cancer. *N Engl J Med* **358:** 1148-1159.
59. Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. (2018) Cancer epigenetics: Moving forward. *PLoS Genet* **14:** e1007362.
60. Sandoval J, Esteller M. (2012) Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev* **22:** 50-55.
61. Esteller M. (2011) Non-coding RNAs in human disease. *Nat Rev Genet* **12:** 861-874.
62. Costa FF. (2008) Non-coding RNAs, epigenetics and complexity. *Gene* **410:** 9-17.
63. Pasculli B, Barbano R, Parrella P. (2018) Epigenetics of breast cancer: Biology and clinical implication in the era of precision medicine. *Semin Cancer Biol* **51:** 22-35.
64. Bertoli G, Cava C, Castiglioni I. (2015) MicroRNAs: New Biomarkers for Diagnosis, Prognosis, Therapy Prediction and Therapeutic Tools for Breast Cancer. *Theranostics* **5:** 1122-1143.

65. Bartel DP. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116:** 281-297.
66. Huntzinger E, Izaurralde E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* **12:** 99-110.
67. Iorio MV, Croce CM. (2012) MicroRNA involvement in human cancer. *Carcinogenesis* **33:** 1126-1133.
68. Wiemer EA. (2007) The role of microRNAs in cancer: no small matter. *Eur J Cancer* **43:** 1529-1544.
69. Li Z, Rana TM. (2014) Therapeutic targeting of microRNAs: current status and future challenges. *Nat Rev Drug Discov* **13:** 622-638.
70. Iorio MV, Croce CM. (2012) MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Mol Med* **4:** 143-159.
71. Vasudevan S, Tong Y, Steitz JA. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318:** 1931-1934.
72. Nelson KM, Weiss GJ. (2008) MicroRNAs and cancer: past, present, and potential future. *Mol Cancer Ther* **7:** 3655-3660.
73. Davalos V, Esteller M. (2010) MicroRNAs and cancer epigenetics: a macrorevolution. *Curr Opin Oncol* **22:** 35-45.
74. Melo SA, Esteller M. (2011) Dysregulation of microRNAs in cancer: playing with fire. *FEBS Lett* **585:** 2087-2099.
75. Zhang B, Pan X, Cobb GP, Anderson TA. (2007) MicroRNAs as oncogenes and tumor suppressors. *Dev Biol* **302:** 1-12.
76. Amorim M, Salta S, Henrique R, Jeronimo C. (2016) Decoding the usefulness of non-coding RNAs as breast cancer markers. *J Transl Med* **14:** 265.
77. Iorio MV*, et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* **65:** 7065-7070.
78. Heneghan HM*, et al.* (2010) Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. *Ann Surg* **251:** 499-505.
79. Zhang L*, et al.* (2015) A circulating miRNA signature as a diagnostic biomarker for non-invasive early detection of breast cancer. *Breast Cancer Res Treat* **154:** 423-434.
80. Wang PY*, et al.* (2013) Higher expression of circulating miR-182 as a novel biomarker for breast cancer. *Oncol Lett* **6:** 1681-1686.
81. Lowery AJ*, et al.* (2009) MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Res* **11:** R27.

82. Cizeron-Clairac G*, et al.* (2015) MiR-190b, the highest up-regulated miRNA in ERalpha-positive compared to ERalpha-negative breast tumors, a new biomarker in breast cancers? *BMC Cancer* **15:** 499.
83. Blenkiron C*, et al.* (2007) MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* **8:** R214.
84. Iorio MV, Casalini P, Tagliabue E, Menard S, Croce CM. (2008) MicroRNA profiling as a tool to understand prognosis, therapy response and resistance in breast cancer. *Eur J Cancer* **44:** 2753-2759.
85. Shi M, Liu D, Duan H, Shen B, Guo N. (2010) Metastasis-related miRNAs, active players in breast cancer invasion, and metastasis. *Cancer Metastasis Rev* **29:** 785-799.
86. Wang L, Wang J. (2012) MicroRNA-mediated breast cancer metastasis: from primary site to distant organs. *Oncogene* **31:** 2499-2511.
87. Ma L*, et al.* (2010) MiR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat Cell Biol* **12:** 247-256.
88. Ma L, Teruya-Feldstein J, Weinberg RA. (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449:** 682-688.
89. Song B*, et al.* (2010) MicroRNA-21 regulates breast cancer invasion partly by targeting tissue inhibitor of metalloproteinase 3 expression. *J Exp Clin Cancer Res* **29:** 29.
90. Gebeshuber CA, Zatloukal K, Martinez J. (2009) MiR-29a suppresses tristetraprolin, which is a regulator of epithelial polarity and metastasis. *EMBO Rep* **10:** 400-405.
91. Petrovic N*, et al.* (2016) MiR-155 expression level changes might be associated with initial phases of breast cancer pathogenesis and lymph-node metastasis. *Cancer Biomark* **16:** 385-394.
92. Tavazoie SF*, et al.* (2008) Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451:** 147-152.
93. Zhang N*, et al.* (2014) MicroRNA-30a suppresses breast tumor growth and metastasis by targeting metadherin. *Oncogene* **33:** 3119-3128.
94. Bockhorn J*, et al.* (2013) MicroRNA-30c targets cytoskeleton genes involved in breast cancer cell invasion. *Breast Cancer Res Treat* **137:** 373-382.
95. Gregory PA*, et al.* (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10:** 593-601.
96. Hilmarsdottir B, Briem E, Bergthorsson JT, Magnusson MK, Gudjonsson T. (2014) Functional Role of the microRNA-200 Family in Breast Morphogenesis and Neoplasia. *Genes (Basel)* **5:** 804-820.

97. Gravgaard KH*, et al.* (2012) The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer. *Breast Cancer Res Treat* **134:** 207-217.
98. Maillot G*, et al.* (2009) Widespread estrogen-dependent repression of micrornas involved in breast tumor cell growth. *Cancer Res* **69:** 8332-8340.
99. Rothe F*, et al.* (2011) Global microRNA expression profiling identifies MiR-210 associated with tumor proliferation, invasion and poor clinical outcome in breast cancer. *PLoS One* **6:** e20980.
100. Jung EJ*, et al.* (2012) Plasma microRNA 210 levels correlate with sensitivity to trastuzumab and tumor presence in breast cancer patients. *Cancer* **118:** 2603-2614.
101. Zhang B*, et al.* (2016) MicroRNA 100 sensitizes luminal A breast cancer cells to paclitaxel treatment in part by targeting mTOR. *Oncotarget* **7:** 5702-5714.
102. Kang Y*, et al.* (2003) A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3:** 537-549.
103. Holliday DL, Speirs V. (2011) Choosing the right cell line for breast cancer research. *Breast Cancer Res* **13:** 215.
104. Livak KJ, Schmittgen TD. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods* **25:** 402-408.
105. Schisterman EF, Perkins NJ, Liu A, Bondell H. (2005) Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* **16:** 73-81.
106. Youden WJ. (1950) Index for rating diagnostic tests. *Cancer* **3:** 32-35.
107. Wei Z, Batagov AO, Carter DR, Krichevsky AM. (2016) Fetal Bovine Serum RNA Interferes with the Cell Culture derived Extracellular RNA. *Sci Rep* **6:** 31175.
108. Shekhar MP, Pauley R, Heppner G. (2003) Host microenvironment in breast cancer development: extracellular matrix-stromal cell contribution to neoplastic phenotype of epithelial cells in the breast. *Breast Cancer Res* **5:** 130-135.
109. Xi Y*, et al.* (2007) Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA* **13:** 1668-1674.
110. Smeets A*, et al.* (2011) Prediction of lymph node involvement in breast cancer from primary tumor tissue using gene expression profiling and miRNAs. *Breast Cancer Res Treat* **129:** 767-776.
111. Schrijver WA, van Diest PJ, Dutch Distant Breast Cancer Metastases C, Moelans CB. (2017) Unravelling site-specific breast cancer metastasis: a microRNA expression profiling study. *Oncotarget* **8:** 3111-3123.

112. Jiang L*, et al.* (2016) A microRNA-mediated decrease in eukaryotic initiation factor 2alpha promotes cell survival during PS-341 treatment. *Sci Rep* **6:** 21565.
113. Mahdavinezhad A*, et al.* (2015) Evaluation of miR-141, miR-200c, miR-30b Expression and Clinicopathological Features of Bladder Cancer. *Int J Mol Cell Med* **4:** 32-39.
114. Lu Y*, et al.* (2009) Amplification and overexpression of Hsa-miR-30b, Hsa-miR-30d and KHDRBS3 at 8q24.22-q24.23 in medulloblastoma. *PLoS One* **4:** e6159.
115. Shao C*, et al.* (2012) Amplification and up-regulation of microRNA-30b in oral squamous cell cancers. *Arch Oral Biol* **57:** 1012-1017.
116. Gaziel-Sovran A*, et al.* (2011) MiR-30b/30d regulation of GalNAc transferases enhances invasion and immunosuppression during metastasis. *Cancer Cell* **20:** 104-118.
117. Zhao H, Xu Z, Qin H, Gao Z, Gao L. (2014) MiR-30b regulates migration and invasion of human colorectal cancer via SIX1. *Biochem J* **460:** 117-125.
118. Liao WT*, et al.* (2014) MicroRNA-30b functions as a tumour suppressor in human colorectal cancer by targeting KRAS, PIK3CD and BCL2. *J Pathol* **232:** 415-427.
119. Wang HY*, et al.* (2014) MicroRNA-30a promotes invasiveness and metastasis in vitro and in vivo through epithelial-mesenchymal transition and results in poor survival of nasopharyngeal carcinoma patients. *Exp Biol Med (Maywood)* **239:** 891-898.
120. Yang SJ*, et al.* (2017) The miR-30 family: Versatile players in breast cancer. *Tumour Biol* **39:** 1010428317692204.
121. Yao D, Dai C, Peng S. (2011) Mechanism of the mesenchymal–epithelial transition and its relationship with metastatic tumor formation. *Mol Cancer Res* **9:** 1608-1620.
122. Chao YL, Shepard CR, Wells A. (2010) Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition. *Mol Cancer* **9:** 179.
123. Kowalski PJ, Rubin MA, Kleer CG. (2003) E-cadherin expression in primary carcinomas of the breast and its distant metastases. *Breast Cancer Res* **5:** R217-222.
124. Croset M*, et al.* (2018) MiRNA-30 Family Members Inhibit Breast Cancer Invasion, Osteomimicry, and Bone Destruction by Directly Targeting Multiple Bone Metastasis-Associated Genes. *Cancer Res* **78:** 5259-5273.
125. Zhang K*, et al.* (2017) Identification of microRNA biomarkers in the blood of breast cancer patients based on microRNA profiling. *Gene* **619:** 10-20.
126. Hatse S*, et al.* (2014) Circulating MicroRNAs as easy-to-measure aging biomarkers in older breast cancer patients: correlation with chronological age but not with fitness/frailty status. *PLoS One* **9:** e110644.

127. Zhang J*, et al.* (2018) MicroRNA-10b expression in breast cancer and its clinical association. *PLoS One* **13:** e0192509.
128. Chen W, Cai F, Zhang B, Barekati Z, Zhong XY. (2013) The level of circulating miRNA-10b and miRNA-373 in detecting lymph node metastasis of breast cancer: potential biomarkers. *Tumour Biol* **34:** 455-462.
129. Yan LX*, et al.* (2008) MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA* **14:** 2348-2360.
130. Asaga S*, et al.* (2011) Direct serum assay for microRNA-21 concentrations in early and advanced breast cancer. *Clin Chem* **57:** 84-91.
131. Kenny PA*, et al.* (2007) The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol Oncol* **1:** 84-96.
132. Croucher PI, McDonald MM, Martin TJ. (2016) Bone metastasis: the importance of the neighbourhood. *Nat Rev Cancer* **16:** 373-386.
133. Weilbaecher KN, Guise TA, McCauley LK. (2011) Cancer to bone: a fatal attraction. *Nat Rev Cancer* **11:** 411-425.
134. Chen YC, Sosnoski DM, Mastro AM. (2010) Breast cancer metastasis to the bone: mechanisms of bone loss. *Breast Cancer Res* **12:** 215.
135. Zheng Y, Zhou H, Dunstan CR, Sutherland RL, Seibel MJ. (2013) The role of the bone microenvironment in skeletal metastasis. *J Bone Oncol* **2:** 47-57.
136. Azmi AS, Bao B, Sarkar FH. (2013) Exosomes in cancer development, metastasis, and drug resistance: a comprehensive review. *Cancer Metastasis Rev* **32:** 623-642.
137. Peinado H*, et al.* (2017) Pre-metastatic niches: organ-specific homes for metastases. *Nat Rev Cancer* **17:** 302-317.
138. Melo SA*, et al.* (2014) Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. *Cancer Cell* **26:** 707-721.

SUPPLEMENTARY MATERIAL

Appendix I - Nottingham Combined Histologic Grade. Breast Cancer grade scoring adapted from (27)

Parameters	Score		
	1	2	3
Architectural differentiation (glandular/tubular differentiation)	>75% of tumor area	10–75 % of tumor area	<10 % of tumor area
Nuclear pleomorphism	Nuclei small and uniform with little increase in size when compared to normal breast epithelium	Cells larger in comparison with normal with open vesicular nuclei, visible nucleoli, and moderate variability in size and shape	Vesicular nuclei, frequently with prominent nucleoli, exhibiting obvious variation in shape and size, sporadically with very large and bizarre forms
Mitotic Index	≤4 HPF	5–9 HPF	≥10 HPF
Total Score	3-5	6-7	8-9
Grade	1 – Well differentiated	2 – Moderately differentiated	3 – Poorly differentiated

Abbreviations: HPF – High-power field

Appendix II - TNM staging system reported by the American Joint Committee on Cancer-Union for International Cancer Control (AJCC–UICC). Adapted from (26)

Primary tumor (T)[1]	
TX	Primary tumor cannot be evaluated
T0	No evidence of primary tumor
Tis	Carcinoma *in situ*
DCIS	Ductal carcinoma *in situ*
LCIS	Lobular carcinoma *in situ*
Paget's	Paget's disease of the nipple not associated with invasive carcinoma and/or DCIS or LCIS in the underlying breast parenchyma. Carcinomas in the breast parenchyma associated with Paget's disease are categorized based on the size and characteristics of the parenchymal disease, although the presence of Paget's disease should still be noted.
T1	Tumor ≤20 mm in greatest dimension
T1mi	Tumor ≤1 mm in greatest dimension
T1a	Tumor >1 mm but ≤5 mm in greatest dimension
T1b	Tumor >5 mm but ≤10 mm in greatest dimension
T1c	Tumor >10 mm but ≤20 mm in greatest dimension
T2	Tumor >20 mm but ≤50 mm in greatest dimension
T3	Tumor >50 mm in greatest dimension
T4	Tumor of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules)
T4a	Extension to the chest wall, not including only pectoralis muscle adherence/invasion
T4b	Ulceration and/or ipsilateral satellite nodules and/or edema (including peau d'orange) of the skin, which do not meet the criteria for inflammatory carcinoma
T4c	Both T4a and T4b
T4d	Inflammatory carcinoma

Regional lymph nodes – Clinical (cN)	
cNX	Regional lymph nodes cannot be assessed
cN0	No regional lymph node metastases
cN1	Metastases to movable ipsilateral level I, II axillary lymph node(s)
cN2	Metastases in ipsilateral level I, II axillary lymph nodes that are clinically fixed or matted; or in clinically detected[2] ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastases
cN2a	Metastases in ipsilateral level I, II axillary lymph nodes fixed to one another (matted) or to other structures
cN2b	Metastases only in clinically detected[2] ipsilateral internal mammary nodes and in the absence of clinically evident level I, II axillary lymph node metastases
cN3	Metastases in ipsilateral infraclavicular lymph node(s) with or without level I, II axillary lymph node involvement; or in clinically detected[2] ipsilateral internal mammary lymph node(s) with clinically evident level I, II axillary lymph node metastases; or metastases in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement
cN3a	Metastases in ipsilateral infraclavicular lymph node(s)
cN3b	Metastases in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
cN3c	Metastases in ipsilateral supraclavicular lymph node(s)

Regional lymph nodes – Pathological (pN)	
pNX	Regional lymph nodes cannot be assessed
pN0	No regional lymph node metastases identified histologically
pN0(i−)	No regional lymph node metastases histologically, negative IHC
pN0(i+)	Malignant cells in regional lymph node(s) not >0.2 mm detected by hematoxylin and eosin staining or IHC including isolated tumor cell clusters
pN0(mol−)	No regional lymph node metastases histologically, negative molecular findings (RT-PCR)

pN0(mol+)	Positive molecular findings (RT-PCR), but no regional lymph node metastases detected by histology or IHC
pN1	Micrometastases or metastases in 1-3 axillary lymph nodes; and/or in internal mammary nodes with metastases detected by SLNB but not clinically detected[2]
pN1mi	Micrometastases (>0.2 mm and/or >200 cells, but none >2.0 mm)
pN1a	Metastases in 1-3 axillary lymph nodes, at least one metastasis >2.0 mm
pN1b	Metastases in internal mammary nodes with micrometastases or macrometastases detected by SLNB but not clinically detected[2]
pN1c	Metastases in 1-3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by SLNB but not clinically detected[2]
pN2	Metastases in 4-9 axillary lymph nodes; or in clinically detected[2] internal mammary lymph nodes in the absence of axillary lymph node metastases
pN2a	Metastases in 4-9 axillary lymph nodes (at least one tumor deposit >2.0 mm)
pN2b	Metastases in clinically detected[2] internal mammary lymph nodes in the absence of axillary lymph node metastases
pN3	Metastases in ≥10 axillary lymph nodes; or in infraclavicular lymph nodes; or in clinically detected[2] ipsilateral internal mammary lymph nodes in the presence of ≥1 positive level I, II axillary lymph nodes; or in ≥3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by SLNB but not clinically detected[2]; or in ipsilateral supraclavicular lymph nodes
pN3a	Metastases in ≥10 axillary lymph nodes (at least one tumor deposit >2.0 mm); or metastases to the infraclavicular nodes
pN3b	Metastases in clinically detected[2] ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in ≥3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by SLNB but not clinically detected[2]
pN3c	Metastases in ipsilateral supraclavicular lymph nodes

Distant metastases (M)[3]	
M0	No clinical or radiographic evidence of distant metastases
cM0(i+)	No clinical or radiographic evidence of distant metastases, but deposits of molecularly or microscopically detected tumor cells in circulating blood, bone marrow or other non-regional nodal tissue that are not >0.2 mm in a patient without symptoms or signs of metastases
M1	Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven >0.2 mm

[1] Definition for classifying the primary tumor (T) is the same regardless of it is based on clinical or for pathologic parameters.

[2] Clinically detected refers to detection by clinical examination or imaging techniques and having characteristics highly suspicious for malignancy or a presumed pathological macrometastases based on fine-needle biopsy.

[3] Definition for classifying the distant metastases (M) is the same regardless of it is based on clinical or for pathologic parameters.

Abbreviations: DCIS - ductal carcinoma *in situ*; IHC – immunohistochemistry; LCIS - lobular carcinoma *in situ*; RT-PCR - real-time polymerase chain reaction; SLNB - sentinel lymph node biopsy.

Appendix III – Anatomic Stage/Prognostic Groups. From (26)

Anatomic Stage	Prognostic Groups		
	T	N	M
0	Tis	0	0
IA	1	0	
IB	0	1mi	
	1	1mi	
IIA	0	1	
	1	1	
	2	0	
IIB	2	1	
	3	0	
IIIA	0	2	
	1	2	
	2	2	
	3	1	
	3	2	
IIIB	4	0	
	4	1	
	4	2	
IIIC	Any T	3	
IV	Any T	Any N	1

Appendix IV – Detail information about primary tumor and the matched metastases per each patient included in FFPE Breast Cancer patients' validation cohort #1.

Patient number	Age at diagnosis	Molecular subtype of primary tumor[1]	Metastasis localization	Time Interval after primary tumor (years)
1	30	Luminal B-like	Brain	10.07
2	37	Basal-like	Brain	1.2
3	36	Luminal B-like	Brain	9.82
4	37	Luminal B-like	Brain	10.45
			Lung	10.4
5	28	Luminal B-like	Bone	7.7
6	39	Luminal B-like	Bone	9.95
			Lung	11.26
7	32	Luminal A-like	Bone	3.64
8	49	Luminal A-like	Bone	13.08
9	51	Luminal A-like	Bone	5.66
10	65	Luminal B-like	Bone	8.8
11	64	Luminal B-like	Bone	11.78
12	57	Luminal B-like	Bone	2.47
13	44	Luminal A-like	Bone	4.94
14	58	Basal-like	Bone	2.46
15	31	Luminal B-like	Bone	0.15
16	57	Luminal B-like	Bone	-
17	44	Luminal B-like	Bone	9.52
18	76	Luminal B-like	Bone	1.93
19	41	Luminal B-like	Bone	3.41
20	71	Luminal B-like	Bone	6.12
21	56	Luminal B-like	Bone	6.66
22	46	Luminal A-like	Bone	11.49
23	42	Luminal B-like	Bone	1.94
24	56	Luminal B-like	Bone	-
25	46	Luminal A-like	Bone	7.67
26	36	Luminal B-like	Bone	2.42
27	62	Luminal A-like	Bone	6.47
28	46	Luminal B-like	Bone	8.31
29	38	Luminal B-like	Bone	3.84
30	33	Luminal B-like	Bone	4.47
31	49	Luminal B-like	Bone	6.61
			Locoregional	7.05
			Contralateral Breast	7.21
32	71	Luminal B-like	Bone	-
33	43	Luminal A-like	Bone	5.98
34	49	Luminal A-like	Bone	12.61
			Contralateral Breast	8.17
35	58	Luminal B-like	Bone	4.99
36	40	Luminal A-like	Bone	4.21
37	73	Luminal B-like	Bone	-
38	58	Luminal A-like	Bone	14.39
			Locoregional	-
39	43	Luminal A-like	Bone	3.42
40	64	Luminal B-like	Bone	3.69
41	42	Luminal A-like	Bone	4.83
42	47	Luminal B-like	Bone	7.76

43	71	Luminal A-like	Bone	3.44
44	51	Luminal B-like	Bone	3.61
45	58	Luminal B-like	Bone	5.5
46	40	Luminal B-like	Bone	2.04
47	73	Luminal A-like	Bone	4.71
48	61	Luminal B-like	Bone	10.5
			Contralateral Breast	3.78
49	47	Luminal B-like	Bone	17
50	59	Luminal B-like	Bone	-
51	43	Luminal A-like	Bone	4.3
			Locoregional	3.96
52	45	Luminal B-like	Bone	-
53	33	Luminal B-like	Bone	6.28
54	37	Luminal A-like	Bone	7.07
55	53	Luminal B-like	Bone	-
56	46	Luminal B-like	Bone	-
57	69	Luminal B-like	Bone	3.92
58	63	Luminal B-like	Bone	11.72
59	65	Luminal B-like	Bone	-
60	45	Luminal B-like	Bone	3.18
61	61	Luminal A-like	Bone	6.9
62	46	Luminal B-like	Bone	7.93
63	53	Luminal B-like	Bone	8.93
64	61	Luminal B-like	Bone	1.92
65	32	Luminal B-like	Bone	3.01
66	45	Luminal B-like	Bone	4.81
67	56	Basal-like	Bone	8.38
68	43	Luminal B-like	Lung	4.73
69	54	HER2-enriched	Lung	17.72
			Contralateral Breast	13.19
70	70	Luminal B-like	Lung	8.29
			Contralateral Breast	9.7
71	75	Luminal B -like	Lung	-
72	56	Luminal B-like	Lung	18.98
73	35	Luminal B-like	Lung	4.83
74	67	Luminal B-like	Lung	-
75	51	Luminal B-like	Lung	7.95
			Locoregional	7.93
76	41	Luminal B-like	Lung	9.43
77	48	Luminal B-like	Lung	10.65
78	58	Luminal B-like	Lung	6.15
79	74	Luminal B-like	Lung	9.01
80	64	Basal-like	Lung	1.03
81	40	Luminal A-like	Lung	12.13
82	54	Luminal B-like	Lung	-

[1]Assessed by immunohistochemistry.

- Patients diagnosed with stage IV Breast Cancer