UNIVERSIDADE DO ALGARVE

Departamento de Ciências Biomédicas e Medicina

Expression of matrix gla protein (MGP) in breast cancer: a molecular and cellular approach

KARINE ROSÁRIO DA SILVA TEIXEIRA

Dissertação para obtenção do Grau de

Mestre em Oncobiologia

Trabalho efetuado sob a orientação científica da Doutora Natércia Conceição e Doutora Sofia Braga

2016

UNIVERSIDADE DO ALGARVE

Departamento de Ciências Biomédicas e Medicina

Expression of matrix gla protein (MGP) in breast cancer: a molecular and cellular approach

KARINE ROSÁRIO DA SILVA TEIXEIRA

Dissertação para obtenção do Grau de

Mestre em Oncobiologia

Trabalho efetuado sob a orientação científica da Doutora Natércia Conceição e Doutora Sofia Braga

2016

Expression of matrix gla protein (MGP) in breast cancer: a molecular and cellular approach

Declaro ser o(a) autor(a) deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Declaração Copyright

A Universidade do Algarve tem o direito, perpétuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

____

Karine Rosário da Silva Teixeira

Abstract

Every year around 8.2 million people die due to cancer, in Portugal, breast cancer is the leading cause of women death (in cancer patients). Several risk factors could be pointed being age (over 50 years old), gender (females), reproductive and hormonal factors and genetic predisposition (presence of mutations in BRCA 1 and BRCA 2) the most commons. Staging is done regarding several aspects, e.g.: histology, histopathologic localization, hormone receptors, mutations or expression of cluster genes.

Research from last decades is focused into targeted therapies. MGP, matrix Gla protein, was pointed as a possible target. MGP primary structure consists in a signal peptide, a phosphorylation domain and a γ-carboxylase recognition site. MGP has also five Gla residues. For a long time, only an isoform with four exons was known, but recently was discovered a second isoform with five exons, with three additional putative binding sites, conferring, possible, a higher binding capability as well as a presumably higher calcification inhibitory power.

Although MGP role in cancer is not fully understood, being upregulated in some types of tumors and downregulated others, in breast cancer, was established, as a prognostic factor, being overexpressed in poor diagnose cases.

Being MGP role in tumorigenesis to be defined, the aims of this work were: clarify MGP expression in normal vs tumoral tissue (using *in silica* method – TCGA database); as well as clarify regulatory mechanisms of MGP (TCGA database); analyze expression of each isoform in breast cancer cell lines (results were obtained performing qRT-PCR followed by electrophoresis and sequencing of the extracted bands) and assess the result of overexpression of each isoform (using XTT and wound healing assay).

We could conclude that both variants could be expressed in simultaneous. Every cell line presented E4 isoform but just a few presented E5 isoform. Hound healing assay revealed that an overexpression of E4 and E5 variants slow down the rate of the area invaded by migrating cells, specially an overexpression of E4 isoform. After *in silica* analyze we could conclude that MGP is more expressed in tumoral tissue vs normal tissue, existing a negative correlation between MGP expression and mir155 expression and methylation, that means that: in tumoral tissue mir155 is less expressed than in normal tissue and occurs an hypomethylation in tumoral tissue. Knowing that MGP

inhibits calcification and is correlated with angiogenesis, our hypothesis is that overexpression of MGP in tumoral tissue is a mechanism of cancer to facilitate angiogenesis – a hallmark of cancer.

Keywords: Breast Cancer; MGP, mir155; prognostic factor, angiogenesis, calcification inhibition.

Resumo

Introdução: o cancro é caracterizado por um crescimento anormal e incontrolado de células que podem adquirir características malignas por acumulação de mutações. Hanahan and Weinberg (2000) descreveram que o desenvolvimento tumoral baseado em vários passos está assente na aquisição, por parte do tumor, de seis princípios biológicos: **i**) manter sinalização proliferativa; **ii)** desenvolver mecanismos de defesa contra supressores de crescimento; **iii)** resistir á morte celular; **iv**) ativar imortalidade replicativa, **v)** induzir angiogénese e **vi)** ativar invasão e metástase. Se uma visão mais tradicional atribuía estas alterações ao resultado de mutações genéticas, uma visão mais atual do cancro engloba também a epigenética como fator crucial na repressão e ativação de genes capazes de conferir as características malignas aos tumores; a metilação genómica é um exemplo de um mecanismo já descrito com um papel importante na atividade tumoral, sendo que está associado a hipometilação do genoma com o cancro.

Determinar os mecanismos moleculares desta doença aproximam a população científica de terapias alvo que permitam ajudar a controlar a disseminação desta doença que é considerada pela organização mundial de saúde uma doença epidémica em ascensão. Todos os anos morrem no mundo 8.2 milhões de pessoas devido a cancro (cerca de 13% de todas as mortes do mundo), sendo a segunda causa de morte em Portugal, EUA e Europa. Em Portugal, o cancro do pulmão e brônquios é a primeira causa de morte por cancro nos homens, enquanto que a primeira causa de morte por cancro nas mulheres é o Cancro da Mama (CM). Existem vários fatores de risco para o CM, sendo os principais: **i)** idade acima dos 50 anos; **ii)** ser do sexo feminino (apenas 1% dos casos corresponde a homens); **iii)** história pessoas ou familiar de CM; **iv)** fatores reprodutivos e hormonais; **v)** exposição endógena a estrogénios (como p.e. uso da pílula) e **vi)** predisposição genética (comumente mutação nos genes BRCA1 e BRCA2). O estadiamento da doença é feito usando TNM (avalia tamanho do Tumor, Nódulos e Metástases). Contudo, outras características são usadas em cancro para definir o tratamento, nomeadamente: localização histopatológica (divide os carcinomas *in situ* dos carcinomas invasivos); expressão de determinados recetores hormonais e expressão de genes em clusters/aglomerados (conferindo a divisão em Luminal A, Luminal B, HER2 e tipo basal), sendo que nas últimas décadas tem se procurado

encontrar terapias direcionas, sendo que cada vez mais se procuram novos alvos terapêuticos.

Um possível alvo terapêutico é a MGP (proteína Gla da matriz). Está localizada no braço pequeno do cromossoma 12 (12p12.3) e é uma proteína dependente da vitamina K. A sua estrutura primária consiste num péptido de sinal, um domínio de fosforilação e um local de reconhecimento da γ-carboxilase; tem também cinco resíduos Gla que após conversão pela enzima γ-glutamil carboxilase passam a ter uma maior afinidade para minerais e iões minerais (como o cálcio, fosfato e cristais de hidroxiapatite – componentes minerais do esqueleto). Durante muito tempo apenas uma variante era conhecida E4 (quatro exões) contudo foi identificada uma nova isoforma: E5 (cinco exões), que apresenta três locais putativos adicionais de ligação a γ-carboxilação; esta característica pode provocar um aumento da sua capacidade de ligação, levando a um poder inibitório mais forte. MGP está associada à inibição de calcificação (devido à sua afinidade de ligação mineral e iónica), e mutações neste gene estão associadas a aterosclerose (calcificação vascular), síndrome de Keutel e cancro. No caso do cancro estudos revelam um papel na diferenciação, proliferação e migração celular, no início da angiogénese e na tumorigénese. Contudo o seu papel na oncogénese e a correlação entre a sua expressão e o tipo de tumores está ainda por determinar; sendo que para alguns tipos de tumores, apresenta maior expressão no tecido tumoral e em outros mais em tecido normal. No cancro da mama contudo, tem sido apontado como fator de prognóstico, estando a sua sobreexpressão associada a um pior prognóstico.

Objectivos: Estando o papel da MGP no cancro ainda por definir, o principal objetivo deste trabalho passa por tentar correlacionar a sua expressão em tecido normal e tumoral assim como em subgrupos de pacientes; clarificar a expressão das variantes em linhas celulares do cancro da mama, assim como qual o resultado a sobreexpressão das variantes na proliferação e migração celular e, identificar mecanismos regulatórios da MGP e o sua expressão em tecido normal vs tumoral.

Materiais e métodos: Para cada linha celular foi averiguado qual a(s) variante(s) expressa(s), recorrendo para isso a qRT-PCR e posteriormente a uma confirmação por sequenciação das bandas obtidas por electroforese, das amostras do qRT-PCR. O efeito da sobreexpressão das isoformas na proliferação e migração celular foi avaliado em duas linhas celulares da mama (linhas não comerciais) previamente transfetadas com o plasmídeo de interesse (E4, E5 e um vector vazio usado como

controlo) que foram sujeitadas a ensaios de XTT e do risco. A relação entre a expressão da MGP em tecidos tumorais, assim como a análise da expressão do mir155 foi averiguada recorrendo a um estudo *in silico* a partir da base de dados *The Cancer Genome Atlas* (TCGA). Usou-se esta ferramenta também para avaliação da metilação de quatro CpGs da MGP. Contudo, fez-se também uma avaliação de uma CpGs na linha celular de cancro da mama T47D, fazendo inicialmente um tratamento com 5-azacitidina para desmetilar o genoma. A análise da metilação foi feita recorrendo a um tratamento com bissulfito e posterior análise por PCR.

Conclusão: As duas variantes podem ser expressas em simultâneo, sendo que em todas as linhas celulares em estudo foi comprovada a expressão da variante E4, já a variante E5 não foi detetada em várias linhas celulares. Ensaio do risco mostrou que uma maior expressão da MGP (tanto variante E4 como E5, sendo que na variante E4 os resultados foram mais significativos) parece promover uma diminuição da taxa de migração celular, sendo que após as 72h de ensaio as células não transfectadas (ou transfetadas com vetor vazio) tinham já ocupado 100% da área, enquanto que as células transfetadas com E5 só foi possível verificar os 100% às 96h, e nas células transfetadas com E4 ao fim das 96h ainda não tinha ocorrido 100% de área invadida.

Os ensaios *in silico* demostraram que a MGP está mais expressa no tecido tumoral, e que há uma correlação negativa entre expressão da MGP e metilação (ocorrendo uma hipometilação no tecido tumoral) assim como com a expressão de mir155. Sabendo que MGP inibe calcificação e, está relacionada com a angiogénese, parece que a maior expressão da MGP em tecido tumoral seja um mecanismo do tumor para promover a angiogénese – uma marca do cancro (*Hallmark of Cancer*).

Palavras-chave: Cancro da mama; MPG; mir155; factor de prognóstico, angiogénse, inibição da calcificação

Abbreviations and acronyms

5-Aza – 5-Azacytidine

A - adenine

Aa – amino acid

AG – astrocytic gliomas

AJCC – American joint committee on cancer

ALND – axillary lymph node

BC – breast cancer

BGP – bone gla protein

bp – base pairs

BRCA – Breast invasive Carcinoma (as well as a gene name)

C – cytosine

cDNA – complementar DNA

CH3 – methyl group

CO2 – carbon dioxide

CTL – control

DMEM – Dulbecco's Modified Eagle Medium

DNA – deoxyribonucleic acid

DNM – DNA methyl transferse

ECM – Extracellular matrix

ER – estrogen receptor

ESR1 – estrogen receptor 1

F – forward

FBS – fetal bovine serum

FS – first strand

G - guanine

GAPDH – glycealdehyde-3-phosphate Dehydrogenese

Gla – γ carboxylated glutamic acid

HER2 – epidermal growth factor receptor 2

HRT – hormonal replacement therapy

IHM – immunohistochemistry

kD – kilo Dalton

LN – lymph node

MGP – Matrix Gla Protein

miRNA – micro RNA

M-MLV – moloney-murine leukemia virus

mRNA – messenger RNA

NHGRI – National Human Genome Research Institute

NIH – The National Cancer Institute

NT – normal tissue

ºC – Celsius degrees

P/S – penicillin/ streptomycin

PBS - phosphate buffered saline

PCR - polymerase chain reaction

PR – progesterone receptor

pre-mRNA – pre-messenger RNA

pT – pathologic stage

qRT-PCR – quantitative reverse transcriptase polymerase chain reaction

R – reverse

RNA - ribonucleic acid

RNases – RNA nucleases

RPMI – Roswell Park Memorial Institute

RT-PCR – reverse transcriptase polymerase chain reaction

SLND – sentinel lymph node

T – timine

TAE – tris acetate edta

TCGA – The Cancer Genome Atlas

TNBC – triple negative breast cancer (all tumors with negative receptors and a cell line)

TNM – tumor, node and metastasis

TP53 – tumor protein 53

TT – tumoral tissue

U – units

UK – United Kingdom

USA – United States of America

V – volts

VKD – vitamin K dependent

WT – wild typ

XTT – tetrazolium salt

Index

Abstract iv
Resumo vi
Abbreviations and acronyms ix
Chapter I – Introduction 1
The genetic material of life 2
Constitution and Structure of DNA 2
Cancer Genetics 3
DNA Methylation 5
Epidemiology 6
Breast Cancer 8
Epidemiology 8
Risk factors 9
Staging 11
Histopathologic Type 19
Types of tumor 20
MGP – Matrix Gla Protein 23
MGP variants 23
Characterization of the protein and its functions 24
MGP and cancer 25
Cell culture 27
Aims 28
Chapter II - Materials and Methods 29
Cell lines and cultures 30
RNA extraction and purification 30
Synthesis of cDNA by Reverse Transcriptase reaction 32
qRT-PCR – Quantitative Real Time – Polymerase Chain Reaction 32

Confirming results of qRT-PCR by Electrophoresis and Sequencing 33
Transient transfection 34
Cell Viability Assays 34
Wound healing assay – motility assay 35
Demethylating treatment with 5 – Azacytidine 35
Genomic DNA extraction from culture cells 36
Methylation analysis by bisulfite conversion 36
Bioinformatic database – The Cancer Genome Atlas (TCGA) 39
Patients selection 39
Chapter III – Results and Discussion 40
MGP expression in different cell lines 41
Proliferation and Migration Assays 43
XTT – CFI-TERT and TNBC 43
Wound healing assay 45
Methylation analysis 46
Analysis breast cancer genomics data using the cBioPortal 46
MGP mRNA expression is higher in tumor tissue 47
Methylation pattern and mir155 expression are correlated with MGP expression 48
MGP in tumoral tissue is hypomethylated in comparison with matched normal tissue 48
mir155 is downregulated in tumoral tissue, comparing to matched normal tissue 51
Final Remarks and Future Perspectives 52
Final Remarks and Future Perspectives 53
Bibliography 55
Annexes .... a

Index of Figures

Figure 1 | Scheme of semiconservative replication.. 3

Figure 2 | Hallmarks of cancer. 4

Figure 3 | Emerging Hallmarks and Enabling Characteristics 5

Figure 4 | Cytosine methylation. 6

Figure 5 | Trends in cancer incidence and death rates by sex, in USA 7

Figure 6 | Ten leading cancer types by sex in American citizens. 7

Figure 7 |A) Mortality by gender in Portugal (2000-2012); B) Incidence by gender in Portugal 8

Figure 8 | Distribution of breast cancers cases by sporadic vs hereditary cancer . 10

Figure 9 | Female and breast anatomy.. 12

Figure 10 | Diagram of breast and regional lymph nodes. 13

Figure 11 | Schematic representation of matrix Gla protein (MGP), pointing out the primary structure and Gla domains.. 23

Figure 12 | MGP schematic representation. 24

Figure 13 | Wound healing assay draft.. 35

Figure 14 | Primers localization in MGP gene. 41

Figure 15 | Relative expression of E4 and E5 MGP isoforms in several breast cancer cell lines 41

Figure 16 | Electrophoresis assay 42

Figure 17 | Chromatograms from sequencing 42

Figure 18 | XTT assay to determine optimal incubation time of the reagent........ 43

Figure 19 | XTT assay in A) CFI-TERT and B) TNBC cell lines. 44

Figure 20 | Wound healing assay over 96h in CFI-TERT cells 45

Figure 21 | Invaded área after 24h, 48, 72h and 96 hours in diferente conditions. 45

Figure 22 | Electrophoretic assay of PCR samples after bisulfite treatment......... 46

Figure 23 | Heatmap of MGP expression in BRCA samples 46

Figure 24 | MGP expression is higher in tumoral tissue than in normal tissue (BRCA samples)... 47

Figure 25 | Localization of methylation probes in MGP gene 48

Figure 26 | Comparison of the methylation status, in four probes localizated in MGP gene 49

Figure 27 | Correlation between MGP expression and methylation in normal and tumoral tissue.. 50
Figure 28 | mir155 expression in NT vs TT... 51

Index of tables

Table 1 | TMN stage grouping for breast cancer 19

Table 2 | Histopathologic classification of breast cancers 19

Table 3 | MGP protein expression in tumoral vs normal tissue staining 26

Table 4 | Cell lines characterization and culture medium conditions 30

Table 5 | Cycling protocol for bisulfite conversion reaction. 37

Table 6 | PCR conditions using two different Taq polymerase 38

Chapter I – Introduction

The genetic material of life

The mathematician and philosopher Norbert Wiener based on Schrodinger's descriptions said that "*just as the amount of information in a system is a measure of its degree of organization, so the entropy of a system is a measure of its degree of disorganization*" [1], this theoretical concept applies to everything known in this universe even when we are not able to understand the connection between the information.[2] This mathematician put the focus on information, and its importance. The seek for information is as old as life, and due to last decades technology improvement, the pursuit for information had a big revolution and was rich in a wide-ranging fields, this revolution allowed, for instance, the birth of genetic as a science, contributing to answer big questions like:[3]

- How information is regulated and transmitted cell to cell in a multicellular organism, holding the information almost unaffected;
- What molecule is able to replicate almost limitless in such precise way;
- How the information is organized in order to allow storing so much information in a space as small as a cell.

Slowly these questions were answered and some 1940s studies were very important, because they showed that: [3]

- Proteins work as "*building blocks"* in cells and they have a play in almost all cell's functions;
- Deoxyribonucleic acid (DNA) was likely the carrier of genetic information. [3];

One of the biggest finding of 20th century was "the structure of DNA and the mechanisms by which information coded in DNA is translated into the amino acid (aa) sequence of proteins",[4] without these findings scientific perspective of inheritance probably would still be almost none and based on empirical data, turning biological disciplines like modern cancer research a "descriptive science that cataloged diverse biological phenomena without being able to explain the mechanics of how they occur." [5]

Constitution and Structure of DNA

Deoxyribonucleic acid is the combination of two poly-nucleotide chains (strands), arranged in a double helical structure, with a sugar-phosphate backbone

attached to a set of pyrimidine (thymine (T) and cytosine(C)) and purine (adenine (A) and guanine (G)) bases. Sugar molecules – deoxyribose – are connected by a phosphodiester bond from the 5' position of one deoxyribose to the 3' position of the next. The helical structure is held by hydrogen bonds between bases, which have a specific way to bound between them, thymines only bond to adenines and cytosines only bond to guanines. To a set of one base (A, T, C or G) and a sugar is given the name of **nucleoside** if the molecule has phosphate group(s) is called **nucleotide**. [3,4]

In most multicellular organisms genetic information carried in each cell is the same;[6] to make that happen cells undergo semiconservative replication of DNA – using both strands of a DNA molecule to produce two complete double helices, is called semiconservative because each original (parental) strand serve as template for new (daughter) strand (Figure 1). Replication have error associated (one error per 10^7 nucleotides copied) as well as correction machinery that correct 99%, still some errors are not corrected - mutations. [3]

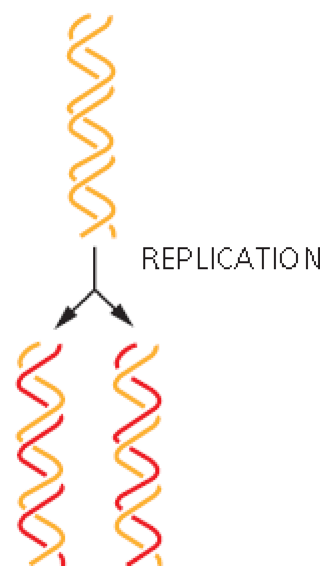


Figure 1 | **Scheme of semiconservative replication,** each strand is used as a template for the new double helix. Replication is semiconservative because each daughter DNA double helix is composed by a conserved strand and newly synthesized strand. Originals strands remain intact for many cell generations. Adapted from Alberts *et al* (2010)[3]

Cancer Genetics

Cancer is usually characterized by an abnormal growth and uncontrolled proliferation of cells. Cells undergo several dynamic processes in which they can acquire genetic changes, that provoke, for instance, loss of response for several signals controlling cellular growth and death, therefore, an accumulation of genetic alteration

can occur in somatic cells, resulting sporadic cancers, or in germline cells causing hereditary predisposition to cancer.[7–10] Although mutations occur aleatory, cancer is a result of mutations in three types of genes: oncogenes (cells promoting cell growth and survival – with gain-of-function mutations), tumor suppressor genes (inhibitors of cell growth and survival – with recessive loss-of-function)[5,11] and stability genes (caretakers – that maintain DNA mechanisms and repair errors). [11,12]

Hanahan and Weinberg (2000) described that tumors acquire six biological principles during multistep development (Figure 2):[10] sustaining proliferative signaling; evading growth suppressors; resisting cell death; enabling replicative immortality;

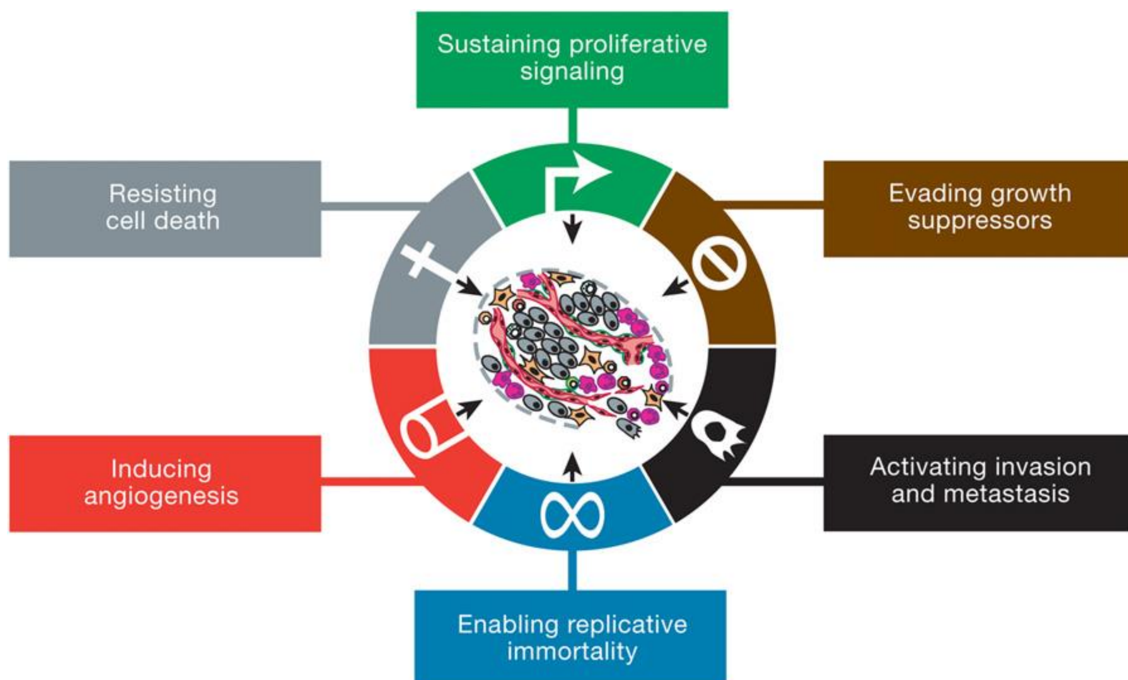


Figure 2 | Hallmarks of cancer. Adapted from *Hanahan. and Weinberg (2000)*

inducing angiogenesis and activating invasion and metastasis.[10,13]

After a decade of conceptual progress The Hallmarks of Cancer "continue to provide a solid foundation for understanding the biology of cancer"; even though they reevaluated new findings and added two emerging and two enabling characteristics to hallmarks (traits) list, revealing that tumorigenesis cannot be just defined by self traits but also regarding tumor microenvironment (Figure 3).[10]

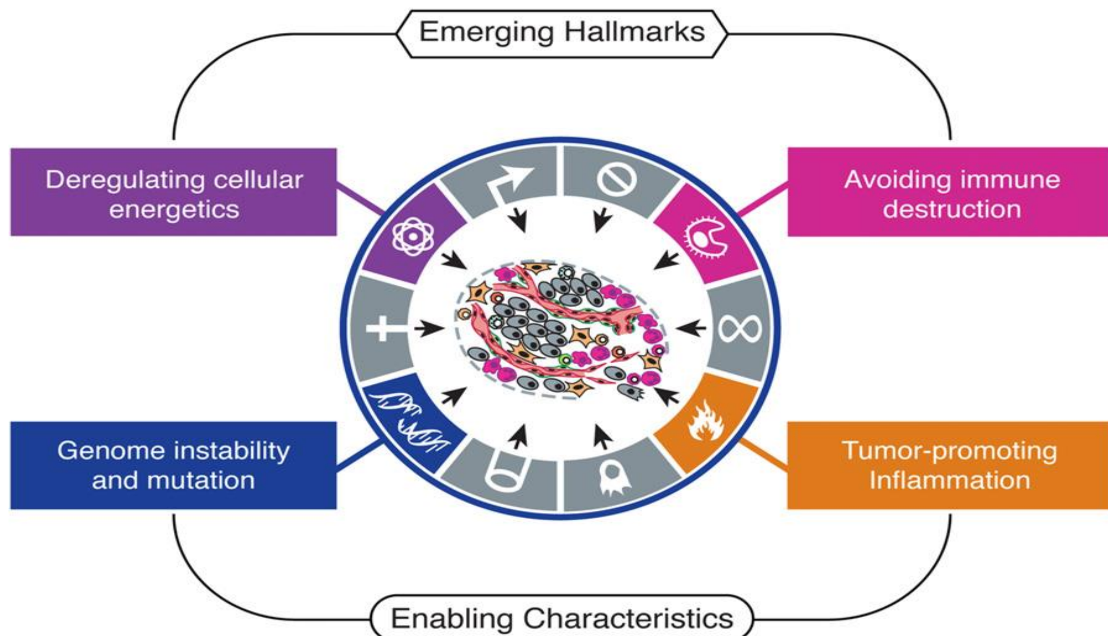


Figure 3 | Emerging Hallmarks and Enabling Characteristics. Adapted from *Hanahan and Weinberg (2011)*

Traditional view of cancer describes it as an accumulation of genetic alteration/mutation, but several studies regarding epigenetic[14] (study of heritable changes in gene function that not entail a change in DNA sequence[15,16]) showed that alterations in the epigenome have an important role in gene activation and repression regulation related to cancer. These heritable changes are disseminated as "covalent chemical changes to the cytosine bases and are referred to as DNA methylation".[16] Epigenetic regulation and chromatin compaction also understand histone tail modifications, ATP-dependent chromatin remodeling or non-coding RNA (ribonucleic acid) play, yet heritability is less clear. [16]

DNA Methylation

Involved in several epigenetic processes like "gene expression, imprinting, X chromosome inactivation, silencing of retroviral and transposable DNA elements, and chromatin organization,"[16] assuring proper regulation of gene expression and stable gene silencing.[17] In cancer DNA is usually hipomethylated.[18]

Methylation is a covalent addition of a methyl group (-CH_3) that occurs "exclusively at the 5 position of the cytosine moiety" catalyzed by DNMT (DNA methyltransferase), this occurs within CpG islands[a] dinucleotides (Figure 4).[17] An alteration in gene expression caused by methylation could lead to an oncogene activation or inactivation of a tumor suppressor gene, increasing cancer risk.

[a] CpG islands are short interspersed DNA sequences rich in guanidine and cytosine (GC).[103]

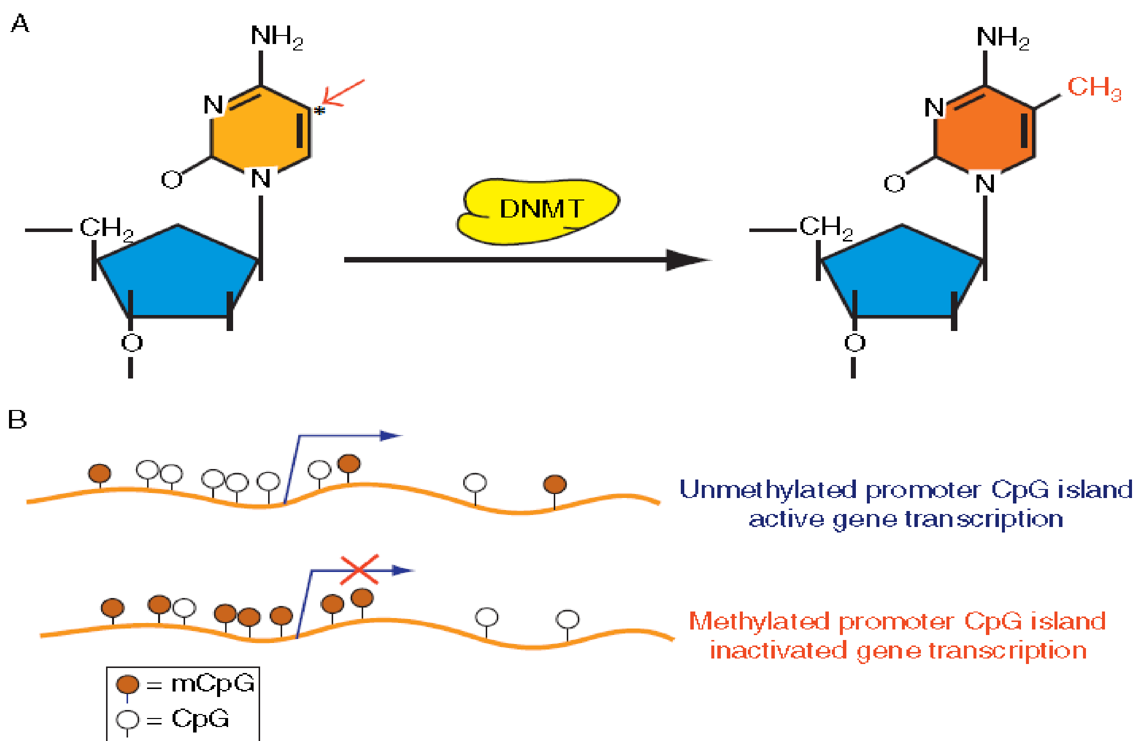


Figure 4 | **Cytosine methylation. A)** Methylation occurs at position 5 of cytosine moiety (marked with an [*], the process is catalyzed by DNMT. **B)** Within the promoter regulation region, unmethylated CpGs (dark brown) allow gene expression; when methylation occurs gene becomes silenced. Adapted from Kulis, M. and Esteller, M. (2010).[17]

Epidemiology

Cancer is considered a rising epidemic disease[19]. Each year, worldwide appears approximately 14 million new cases of cancer and 8.2 million people die from cancer; this represents 13% of all death worldwide.[19,20] In United States as in Portugal cancer is the second leading cause of death, but is expected to overcame heart diseases in the next few years.[20–22] In 2016 around 595 690 Americans are expected to die due cancer, this represents an average of 1 630 death per day.[21]

Using trends, in cancer death rates to measure the improvement of the battle against cancer, is possible to see that in the 20th century, the total cancer death rate was high, reaching a peak in 1991 (Figure 5), this peak is often explained by the tobacco epidemic.[21] Tobacco intake is associated as a risk factor for numerous tumors like "*lung, larynx, oesophagus, oral cavity and pharynx, bladder, pancreas, kidney, liver, stomach, bowel, cervix, leukemia, and ovarian cancers*".[21] So was expected that a decrease of its consumption – happened in the 70's in United Kingdom (UK)[23] and United States of America (USA)[24], e.g. – would lead to a decrease of cancer death, but

that just happened between 1991 and 2012 (rate lowed in 23%).[20] This is explained by the fact that usually is necessary many years, or decades, for the damages in DNA caused by smoking lead to cancer. The improvement in early detection and treatment is pointed as well as a cause of the reduction of cancer death.[20]

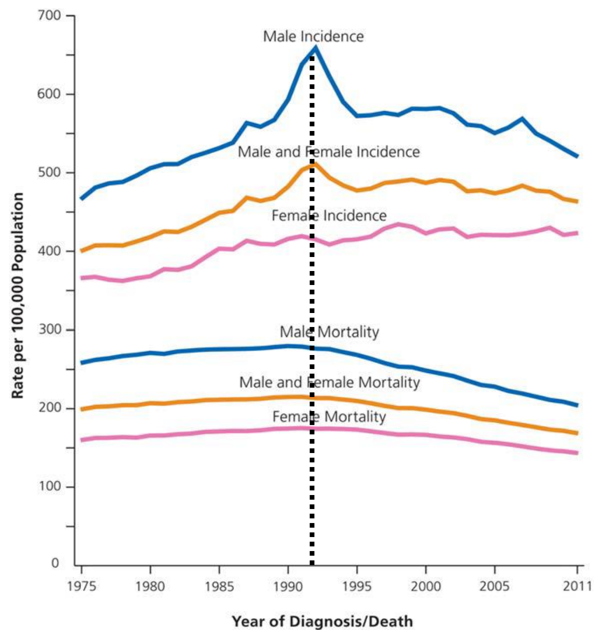


Figure 5 | Trends in cancer incidence and death rates by sex, in USA between 1975 and 2011 – adapted from Siegel et al

In USA, in spite of in both males and females the leading cause of death by cancer is lung and bronchus cancer, incidence is gender specific – prostate cancer in men and breast cancer (BC) in women (Figure 6).[20]

Estimated New Cases

Males			Females		
Prostate	220,800	26%	Breast	231,840	29%
Lung & bronchus	115,610	14%	Lung & bronchus	105,590	13%
Colon & rectum	69,090	8%	Colon & rectum	63,610	8%
Urinary bladder	56,320	7%	Uterine corpus	54,870	7%
Melanoma of the skin	42,670	5%	Thyroid	47,230	6%
Non-Hodgkin lymphoma	39,850	5%	Non-Hodgkin lymphoma	32,000	4%
Kidney & renal pelvis	38,270	5%	Melanoma of the skin	31,200	4%
Oral cavity & pharynx	32,670	4%	Pancreas	24,120	3%
Leukemia	30,900	4%	Leukemia	23,370	3%
Liver & intrahepatic bile duct	25,510	3%	Kidney & renal pelvis	23,290	3%
All Sites	**848,200**	**100%**	**All Sites**	**810,170**	**100%**

Estimated Deaths

Males			Females		
Lung & bronchus	86,380	28%	Lung & bronchus	71,660	26%
Prostate	27,540	9%	Breast	40,290	15%
Colon & rectum	26,100	8%	Colon & rectum	23,600	9%
Pancreas	20,710	7%	Pancreas	19,850	7%
Liver & intrahepatic bile duct	17,030	5%	Ovary	14,180	5%
Leukemia	14,210	5%	Leukemia	10,240	4%
Esophagus	12,600	4%	Uterine corpus	10,170	4%
Urinary bladder	11,510	4%	Non-Hodgkin lymphoma	8,310	3%
Non-Hodgkin lymphoma	11,480	4%	Liver & intrahepatic bile duct	7,520	3%
Kidney & renal pelvis	9,070	3%	Brain & other nervous system	6,380	2%
All Sites	**312,150**	**100%**	**All Sites**	**277,280**	**100%**

Figure 6 | Ten leading cancer types by sex in American citizens. **A)** Estimated new cases; **B)** Estimated deaths. Estimates are rounded to the nearest 10 and excludes cases of basal cell and squamous cell skin cancers and in situ carcinoma except urinary bladder – adapted from Cancer Statistics, Siegel (2015)

In Portugal mortality rates are quite different (Figure 7), in males lung and bronchus tumors are still the leading cause of death, but in females the first position is taken by breast cancer (lung and bronchus are in fifth place). Colorectal cancer is the second major cause of cancer in Portugal, in both genders. Although, cancer incidence in Portugal as is USA is gender specific. [25]

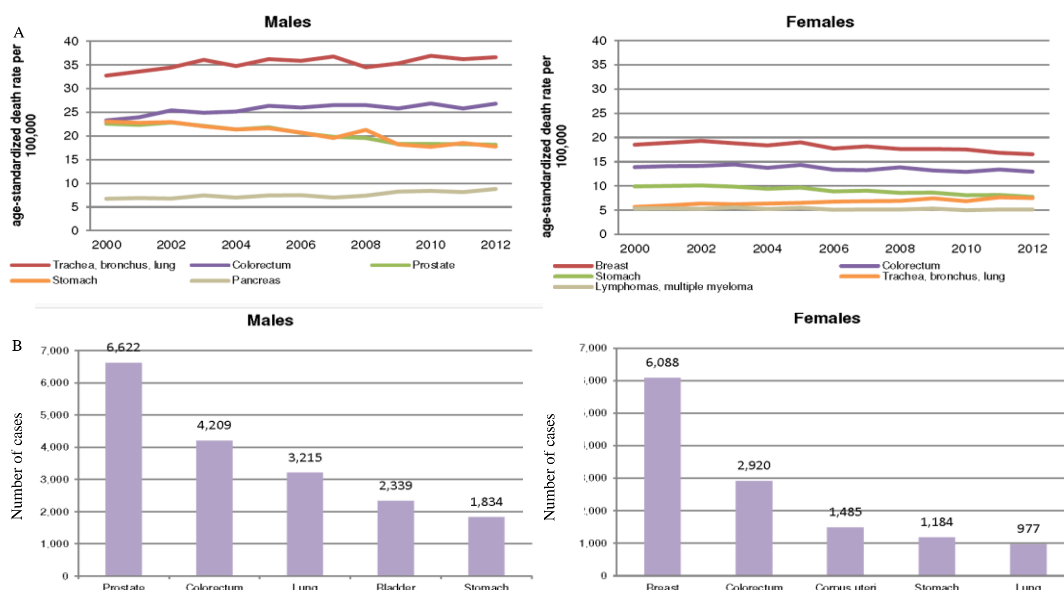


Figure 7 |**A)** Mortality by gender in Portugal (2000-2012); **B)** Incidence by gender in Portugal. Adapted from World Health Organization (2014)

Breast Cancer

Epidemiology

Breast cancer is a disease that affects many people worldwide, in specially women. In United States, 29% of the new cases of cancer in women are breast cancer, it is the second leading cause of death in women with cancer, representing 15% of the deaths[20], in Portugal that number rises to 17%.[25] Cancer statistics 2015 show that new cases and deaths in men represents only 1% of all cases,[20] the same was assessed to Portuguese population.[26] Comparatively to worldwide the incidence of BC in Western Europe is higher, although, Portugal have the smallest rate of mortality (in comparison with other European state members). In part this could be a result of early detection, promoted by mammographies screenings, that in 2016 is expected to reach 60% of coverage of women (between 45 and 69 years old).[27]

Risk factors

There are some risk factors associated with breast cancer and they are commonly seen in about 50% of woman diagnosed. Those risk factors could be:

– Age – above 50 years old[28,29];

– Gender – occurs about 150 times more in females than in males;

– Lifestyle, diet and environment factors;

– Personal or family history of breast cancer or benign disease;

– Hormonal and reproductive factors – females with no pregnancies – nulliparity – or pregnancies in an age greater than 30 years old, early menarche and late menopause;

– Endogenous estrogen exposure/reproductive factors;

– Genetic predisposition.[28]

Hormonal risk

As we saw above, hormones play a tremendous role in breast cancer, and it have been demonstrated that an increased estrogen exposure is correlated with an increasing risk of cancer, these suggests a hormone dependence.[28,30] Although the cancer risk, estrogen is essential for breast development and for the reproductive system. The number of menstrual cycles a woman goes throw in her life will be determinant to define her cancer risk because natural estrogen (produced in self organism) is "*released from ovaries during every menstrual cycle*"[30]. This way: women with earlier menarche and late menopause are exposed to more natural estrogen[28,30]; for each year earlier from the average menarche (12 years old) the risk increases by 5% ,[30,31] and for late menopauses the risk increases 3% per year.[30] In other hand having children is benefic to women due to a lack of periods during the pregnancy, this could be a risk factor for western societies where women delayed giving birth and have less children; each pregnancy is thought to lower the risk in 7%. [30] The risk could be lowered even further by breastfeeding [4.3% for every 12 months], studies suggest that breast cells could be changed in the process, making them less prone to develop cancer. [30,32,33]

Women use hormones commonly as a contraceptive (the pill) or as a post-menopausal Hormone Replacement Therapy (HRT), this increases slightly breast cancer risk[30,34–36]. Notwithstanding, the risk vanishes slowly after quitting, being null after 10 years quitting the pill [30,37] and after 5 years after stopping HRT.[28,30] The risk, is higher using combined therapy of estrogen and progestin compared with using estrogen

alone.[28] In spite of the risk, big number of menopausal patients use HRT to relieve biological changes like hot flashes, vaginal dryness or as a way of protection against osteoporosis or other diseases of bone.[28,38]

Genetic risk

Between 70-80% (Figure 8) of breast cancer cases are due to sporadic cases (mix of genetic damages obtained over a lifetime) and a small percentage due to hereditary causes (genetic alterations passed by one generation to another) and family clusters (bigger percentage of breast cancer in a same family considering rate of sporadic cancer; putting in evidence a combination of risk factors like inherited susceptibility and environmental factors).[39]

Breast cancer 1 and 2 genes (BRCA1 and BRCA2) are the cause of most hereditary breast cancer cases, but mutations can occur also in PTEN, p53, MLH1, MLH2 and STK11.[28] BRCAs are autosomal dominant genes that encode tumor suppressor protein, which help repair damaged DNA, playing an essential role in cell stability. When either of one is mutated or altered causing a malfunction of the protein, DNA may not be repaired properly and additional genetic alteration could lead to cancer. Specific mutations in these genes not only increase female risk of breast cancer but also ovarian cancers. Is known that 12% of woman in general population will develop breast cancer during their lives; recent studies show that 55 to 65% women with harmful mutations in BRCA1 and 45% with harmful mutations in BRCA2 will

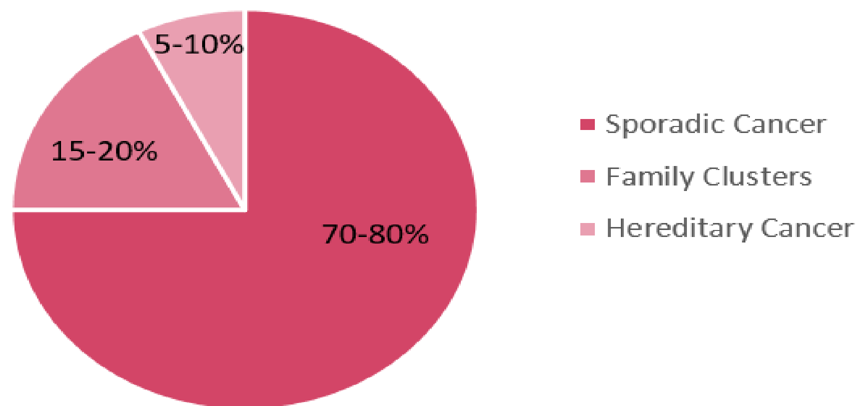


Figure 8 / Distribution of breast cancers cases by sporadic vs hereditary cancer. Adapted from Staten Island University Hospital (2014)[39]

develop breast cancer by the age of 70 years old. [40]

Staging

For staging it is mandatory doing a microscopic confirmation where the histologic type and grade should be recorded. Staging for carcinoma of the breast applies to infiltrating (including microinvasive) and *in situ* carcinomas. It is usually used TNM (Tumor, Node and Metastasis) classification of malignant tumors, published by the American Joint Committee on Cancer (AJCC) as staging system.[28,41]

Physicians use results from scans and diagnostic tests to answer questions like:

“**T** (Tumor) – How large is the primary tumor? Where is it located?

N (Node) – Has the tumor spread to the lymph nodes? If so, where and how many?

M (Metastasis) – Has the cancer metastasized to other parts of the body? If so where and how much?”[43]

Anatomy

Female breast is mostly composed by adipose tissue. Each breast is divided in lobes, lobules and milk ducts. A healthy breast have between 12 to 20 lobes (which one is formed by many smaller lobules[28] – gland that produces milk in nursing woman). Lobules and lobes are connected by milk ducts that when necessary carry milk to the nipple. In adipose tissue a network of ligaments, fibrous connective tissue, nerves, lymph vessels, lymph nodes and blood vessels (Figure 9).[42]

Lymph system

The lymph is part of the immune system and work in a similar way of the circulatory system, where lymph vessels and nodes are running throughout the entire body to transport disease-fighting cells and fluids. Bean-shaped lymph nodes clusters “are fixed in areas throughout the lymph system act as filters by carrying abnormal cells away from healthy tissue”.

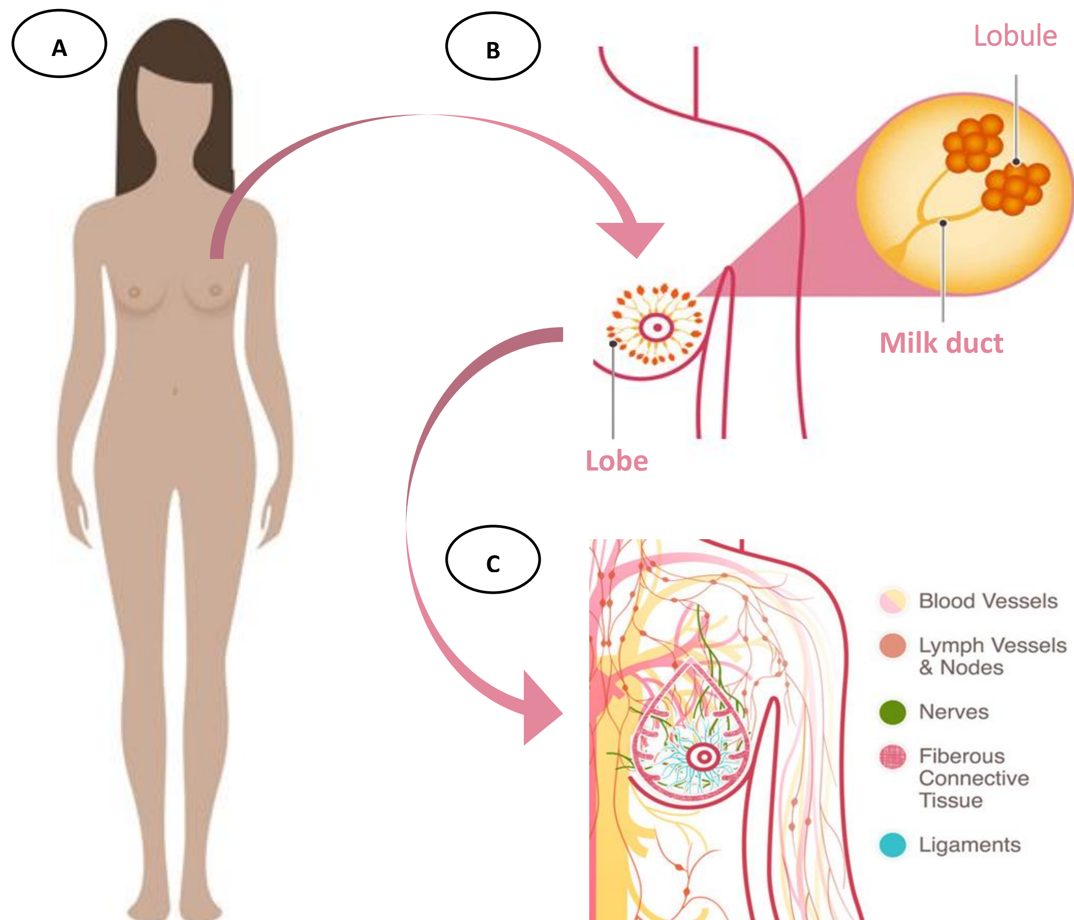


Figure 9 / Female and breast anatomy. **A)** Female anatomy; **B)** Localization of lobules, lobes and milk ducts in breasts; **C)** **S**cheme of intrinsic systems in breast tissue and neighborhood. Adapted from National Breast Cancer Foundation (2015).

Usually the type of breast cancer is determined by the origin of the growth of cancer cells (usually in the lobes, lobules and ducts). Cancer found in nearby lymph nodes (LN) can help physicians determine if the tumor has spread; some additional distal nodes can be tested to understand how far the cancer has spread.

Regional lymph nodes

Axillary, transpectoral and internal mammary nodes are drained by the breast lymphatic vessels. For staging purposes intrammamary lymph nodes are coded as axillary, and supraclavicular lymph nodes are coded as regional lymph nodes. Metastasis to other lymph nodes (as cervical or contralateral internal mammary lymph nodes) are classified as distant (M1) (Figure 10)[28]

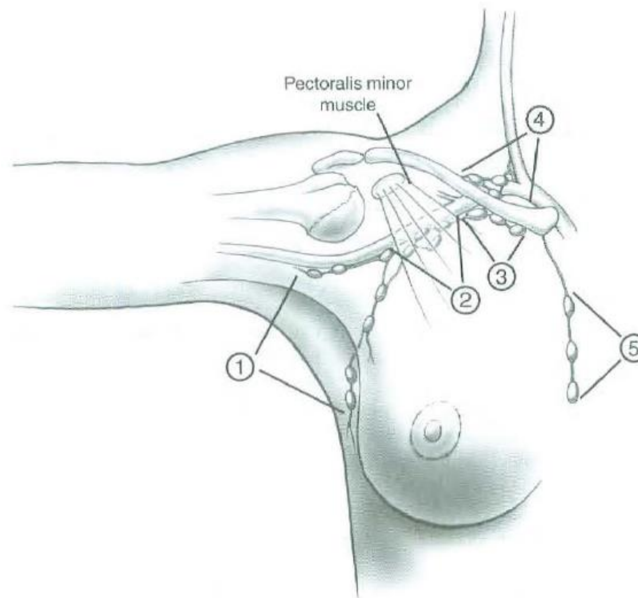


Figure 10 | **Diagram of breast and regional lymph nodes**.
1 – Low axillary, Level I; 2 – Mid-axillary, Level II; 3 – High-axillary, apical, Level III; 4 – Supraclavicular; 5 – Internal mammary nodes.

1. **Axillary (ipsilateral)**: interpectoral (Rotter's) nodes and lymph nodes along axillary vein and its tributaries that may (but not required) divided into:
    a. Level I – low axilla
    b. Level II – mid-axilla
    c. Level III – apical axilla (more anatomical information in Appendix table 3)
2. **Internal mammary (ipsilateral)**: LN in the intercostal spaces along the edge of the sternum in the endothoracic fascia.
3. **Supraclavicular**: LN in the supraclavicular fossa, a triangle defined by the omohyod muscle and tendon, the internal jugular vein, and the clavicle and subclavicle vein. Adjacent LN outside of this triangle are considered to be lower cervical nodes (M1).

Classification

1. Clinical staging

The first step is a physical examination of the area, including observation and palpation of the skin, mammary gland and lymph nodes (axillary, supraclavicular and cervical), imaging and an examination of breast tissues and/or other tissues. [28,41]

2. Pathologic staging

Includes all previous information and data from "surgical exploration and resection as well as pathologic examination of the primary carcinoma, regional lymph nodes and, metastatic sites (if applicable) including not less than excision of the primary carcinoma with no macroscopic tumor in any margin of resection by pathologic examination."[41]

The pathologic stage (pT) only can be assessed if the marge involvement is microscopic and not macroscopic. If there is a macroscopic involvement the cancer is coded as pTX since the total extent of the primary tumor could not be measured. If the primary tumor is invasive and not just microinvasive it is recommended at least the resection of the low axillary lymph nodes (Level 1) (Figure 10) this is used for pathologic classification (pN). If the surgery occurs after "neoadjuvant therapy, hormonal therapy, immunotherapy, or radiation therapy, the prefix "y" should be used with the TNM classification" (e.g. ypTNM)[41]

TNM Classification

A. Primary Tumor (T)

For each case, it is used a specific kind of measurement to classify the primary tumor; it could be obtained by physical examination or mammographies and ultrasounds. Only the invasive component is measured, after that it is removed tissue to prosecute some specific studies, for instance, to evaluate estrogen and progesterone receptors. Some patients have to do multiple core biopsies, this could lead to an under classifying of the T component, in that cases tumor size should be reconstructed based on imaging and histological findings. [41]

Applying TNM for staging the letter "T" is grouped to a number or letter to define the size and location. Some stages have some smaller groups – sub stages – to describe the tumor in more detail[43]

TX: As said previously, is used when the primary tumor cannot be evaluated.[28,41,43]

TO: There is no evidence of breast cancer. [28,41,43]

Tis – Carcinoma *in situ*, with no evidence of an invasive component[41], its restricted within ducts or lobules of the breast tissue. [41,43] Tis is sub divided in three types:

Tis (DCIS) – Intraductal carcinoma *in situ*, is a noninvasive cancer, that can evolve to an invasive breast cancer if not removed. [28,41,43]

Tis (LCIS) – Lobular carcinoma *in situ*, called to abnormal cells found in the lobules or glands of the breast – in spite of not being cancer increases the risk of progress to invasive breast cancer. [28,41,43]

Tis (Paget's) – Paget's disease of the nipple[28,41,43], "rare form of early, noninvasive cancer"[43] that only affect skin cells of the nipple. Simultaneously can occur an invasive breast cancer, staging in this case is only given by the invasive tumor.[28,43]

T1 – Tumor is no bigger than 20 millimeters (mm) in its widest area. It is subdivided usually in substages:

T1mic – Microinvasion no larger than 1mm in the widest area.[28,41]

T1a – Tumor is larger than 1mm but smaller than 5mm in the widest area.[28,41,43]

T1b – Tumor larger than 5mm but not more than 10mm in the widest area.[28,41,43]

T1c – Tumor larger than 10mm but not more than 20mm in the widest area. [28,41,43]

T2 – Tumor between 20mm and 50mm in the widest area. [28,41,43]

T3 – Tumor greater than 50mm. [28,41,43]

T4 – Tumor of any size with direct extension to: [28,41,43]

T4a – Chest wall, not including pectoralis muscle. [28,41,43]

T4b – Skin (Edema "[including peau d'orange] or ulceration of the skin of the breast, or satellite skin nodules confined to the same breast)". [41]

T4c – Chest wall and skin. [28,41,43]

T4d – Inflammatory carcinoma. [28,41,43]

Notes:

i. In case of bilateral breast carcinoma, each breast is considered an independent organ, this way both will have an independent stage.
ii. Inflammatory carcinoma is characterized by "diffuse erythema and edema (peau d'orange) of the breast, often without an underlying palpable mass."[41] The symptoms are common to mastitis (inflammation of the breast). In this type of cancer, cancer cells block lymph vessels, causing the visual symptoms.[44]

iii. Carcinoma *in situ* – cancer located in epithelial cells is designated carcinoma. It occurs in glands and ducts, that is why most breast cancers are carcinomas. Benign (non-cancerous) cells do not invade beyond epithelial tissue, in carcinoma *in situ* abnormal cells look similar to invasive carcinoma cells (when analyzed under microscope). It was assumed that these cells could become invasive if not treated, but recent studies showed that the transition to invasive carcinoma is more "complex and subtle" than the previous idea based on microscopic resemblance. Long term follow-up of patients with carcinomas *in situ* established that not all of them progress to invasive cancer. [45]

B. Regional Lymph Nodes (N)

Definitions for classifying the regional lymph nodes (N) are different for *Clinical* and for *Pathologic* classification. [28,41,43,46]

B.1. Clinical Classification[28,41]

NX – Regional lymph node cannot be assessed (it could be removed previously, or not removed, e.g.).

NO - No regional lymph node metastasis.

N1 – Metastasis to movable ipsilateral axillary lymph nodes

N2 - Metastasis in ipsilateral axillary lymph nodes fixed or matted (axillary lymph nodes that are fixed to each other), or in clinically apparent ipsilateral internal mammary nodes in the absence of clinically evident axillary node metastasis.

N2a – Metastasis in ipsilateral axillary lymph nodes fixed to one another or to other structures

N2b – Metastasis only in clinically apparent ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastasis.

N3 – Metastasis to ipsilateral infraclavicular lymph node(s) with or without clinically evident axillary lymph nodes, or in clinically apparent ipsilateral internal mammary lymph node(s) and in the presence of clinically evident axillary lymph node metastasis, or metastasis in ipsilateral supraclavicular lymph nodes with or without axillary or internal mammary nodal involvement.

N3a – Metastasis to ipsilateral infraclavicular lymph node(s)

N3b – Metastasis to ipsilateral internal mammary lymph node(s) and clinically apparent axillary lymph nodes

N3c – Metastasis in ipsilateral supraclavicular lymph nodes with or without axillary or internal mammary nodal involvement

B.2. Pathologic Classification [28,41,46]

Classification based on axillary lymph node dissection (ALND) with or without sentinel lymph node dissection (SLND). Classification based merely in SLND should be designated ***sn***.

NX – Regional lymph node cannot be assessed (they could be removed previously, or not removed, e.g.).

NO - No regional lymph node metastasis.

N1 – Metastasis to movable ipsilateral axillary lymph nodes.

pNX – Regional lymph nodes cannot be assessed (e.g., previously removed, or not removed for pathologic study).

pNO - No regional lymph node metastasis; no additional examination for isolated tumor cells (ITCs, defined as single tumor cells or small lusters not greater than 0.2 mm, usually detected only by immunohistochemistry (IHC) or molecular methods but which may be verified on hematoxylin and eosin stains. ITCs do not usually show evidence of malignant activity [e.g., proliferation or stromal reaction])

pNO (i-) – No histological nodal metastasis, and negative by IHC

pNO (i+) – No histological nodal metastasis but positive by IHC, with no cluster greater than 0.2 mm in diameter

pNO (mol-) – No histological nodal metastasis and negative molecular findings (by reverse transcriptase polymerase chain reaction, RT-PCR)

pNO (mol+) – No histological nodal metastasis, but positive molecular findings (by RT-PCR)

pN1 – Metastasis in 1-3 ipsilateral axillary lymph node(s) and/or in internal mammary nodes with microscopic disease detected by SLND but not clinically apparent

pN1mi – Metastasis (greater than 0.2 mm, none greater than 2.0mm)

pN1a – Metastasis in 1-3 axillary lymph nodes

pN1b – Metastasis to internal mammary lymph nodes with microscopic disease detected by SLND but not clinically apparent

pN1c – Metastasis in 1-3 ipsilateral axillary lymph node(s) and in internal mammary nodes with microscopic disease detected by SLND but not clinically apparent. If associated with more than three positive axillary nodes, the internal mammary nodes are classified as N3b to reflect increased tumor burden.

pN2 – Metastasis in 4-9 axillary lymph nodes or in clinically apparent internal mammary lump nodes in the absence of axillary lymph nodes

pN2a – Metastasis in 4-9 axillary lymph nodes (at least one tumor deposit >2 mm)

pN2b – Metastasis in clinically apparent internal mammary lymph nodes in the absence of axillary nodes

pN3 – Metastasis in 10 or more axillary lymph nodes, or in infraclavicular lymph nodes, or in clinically apparent ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary nodes; or in more than three axillary lymph nodes with clinically negative microscopic metastasis in internal mammary lymph nodes; or in ipsilateral supraclavicular lymph node(s)

pN3a – Metastasis in 10 or more axillary lymph nodes (at least one tumor deposit greater than 2.0 mm), or metastasis to the infraclavicular lymph nodes

pN3b – Metastasis in clinically apparent ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary nodes: or in more than three axillary lymph nodes detected by SLND but not clinically apparent

pN3c – Metastasis in ipsilateral supraclavicular lymph node(s).

C. Metastasis

MX - Distant metastasis cannot be assessed.

M0 – No distant metastasis

M1 - Distant metastasis

D. TNM stage grouping for breast cancer

Cancer stage is a combination of T, N and M. Staging is usually confirmed after a surgery, but for neoadjuvant therapy[b] staging is chiefly determined clinically. Physicians may use stage I to stage IIA to define early stages, and stages IIB to stage II

[b] Neoadjuvant therapy – treatment given before surgery.[43]

for locally advanced. Furthermore, it is imperative to retain the idea that tumor biology, like specific markers (*e.g.* estrogen receptors) have a huge impact on the advised treatment as well as in the prognosis. [43]

Table 1 | *TMN stage grouping*[c] *for breast cancer, as in the 7thEdition of AJCC Cancer Staging Manual (2010).*[43,46] *(* –* T1 includes T1mi)

Stage	TNM	N	M
Stage 0	Tis	NO	MO
Stage IA	T1*	NO	MO
Stage IB	TO-T1*	N1mi	MO
Stage II A	TO	N1	MO
	T1*	N1	MO
	T2	N0	MO
Stage II B	T2	N1	MO
	T3	NO	MO
Stage III A	TO	N2	MO
	T1*	N2	MO
	T2	N2	MO
	T3	N1	MO
	T3	N2	MO
Stage III B	T4	N0	MO
	T4	N1	MO
	T4	N2	MO
Stage III C	Any T	N3	MO
Stage IV	Any T	Any N	M1

Histopathologic Type

Histological classification has been a valuable tool in past decades, and divide *in situ* carcinomas and invasive carcinomas;[41,47] Relies on histology and put apart molecular markers that have a significant weight on prognosis.[47]

Table 2 | Histopathologic classification of breast cancers - as in the 7thEdition of AJCC Cancer Staging Manual (2010)[48]

Tumor Location	Histologic subtype
Carcinoma, NOS (Not otherwise specified)	
Ductal	Intraductal (*in situ*)
	Invasive with predominant component

[c] TMN stage grouping used in TCGA (The Cancer Genome Atlas, a database used in the thesis) is from the7th Edition of AJCC Cancer Staging Manual, in order to even out the results it was decided to place the 7th edition classification here; 6th edition can be found in Appendix table 1.

	Invasive, NOS
	Comedo
	Inflammatory
	Medullary with lymphocytic infiltrate
	Mucinous (colloid)
	Papillary
	Scirrhous
	Tubular
	Other
Lobular	Invasive with predominant *in situ* component
	Invasive
Nipple	Paget disease, NOS
	Paget disease with intraductal carcinoma
	Paget disease with invasive ductal carcinoma
Other	Undifferentiated carcinoma
	Metaplastic

The following tumor subtypes occur in the breast but are not considered typical breast cancers:

- Phyllodes tumor.
- Angiosarcoma.
- Primary lymphoma

Types of tumor

Commonly breast cancer is organized in subtypes, aiming the division in some (1) special receptors or in a (2) specific cluster of genes.

1. **Hormone receptors (routinely assays are performed by pathologists on tumor samples)** [28]
   a. **Hormone receptor positive** – Tumors express estrogen receptors (ER), progesterone receptors (PR),[44] ER responsive genes and other genes that encode typical proteins of luminal epithelial cells.[49] They may depend on hormones (estrogen and progesterone) to grow.[44]
   b. **HER2 positive** – Represents 20% - to 25% of breast cancers, these tumors are dependent of the gene HER2 (epidermal growth factor receptor 2) to grow and have a high abnormal number of HER2 receptors or copies of the gene. [44]

c. **Triple negative** – if the tumor do not express any of the previous hormone receptors (ER, PR and/or HER2). This type may grow faster than the hormone positive and be more sensitive to chemotherapy. [44]

2. **Molecular classification of breast cancer subtypes**

Molecular classification was described first by Perou *et* al in 2000[50], and gradually emerged as an answer to predict a response to targeted therapies and represents a high success in the design of individualized therapies in breast cancer, promoting an improvement in disease-specific survival.[51,52]Altough, studies show that combining both histological and molecular classifications result in significantly better predictive value. [47,51] After the first classification given by Perou *et* al (2000) several adjustments were made and now tumors are commonly classified into four major subtypes: luminal A, luminal B, $HER2^+$ and basal like.[50–54] This classification is a result of gene expression profiling given by "microarray datasets and progressed to a PCR-based test with a curated list of 50 genes (the PAM50 gene signature)."[54] Each subtype has a particular set of "risk factors for incidence, response to treatment, risk of disease progression and preferential organ sites of metastases."[51]

Luminal A – most common (50 to 60% of all BC); is characterized by an increase in estrogen receptor 1 (ESR1) and/or PR^+/$Her2^-$ status[47,53,55] and lower levels of proliferation related genes, as Ki 67.[d] [49,56] It is also associated to low-grade tumors and good prognosis; [47,53,55] they will likely have a good response to hormonal treatments; with relapse rate lower than other subtypes. Recurrence is common in bone, but in tissues like liver, lung and central nervous system occur in a little less than 10% of patients. Include special histological types (tubular, invasive cribriform, mucinous and lobular).[51]

Luminal B – Less common than luminal A (about 20% of all BC) although approximately 30% of HER2-positive tumors defined by IHC are assigned to the luminal B subtype. Demonstrate increase expression of growth receptor signaling genes and have a more aggressive[49] phenotype with bad prognosis.[47,53]

As an IHC point of view luminal B is defined by:

- ER positive; HER2 negative; high Ki 67 or;
- ER positive; HER2 positive.

[d] Ki 67 protein expression is strictly associated with cell proliferation,[104] high levels are associated with worse outcomes[56]

But in fact, this definition does not englobe a small percentage (6%) of luminal B that are ER and HER negative. [49,56]

Note:

Main difference between both luminal subgroups is increased expression of proliferation-related genes. Ki67 index is pointed as a potential marker to differentiate luminal A and B in clinical practice.[49] Cheang *et al* (2009) stablished Ki67 index of 14% or more Ki67-positive tumor nuclei as the best cut point.

To summarize: Ki67 < 14% (low Ki67) – Luminal A

Ki67 ≥ 14% (high Ki67 – Luminal B

However, Ki67 immunohistochemistry is not optimized and standardized as well the Ki67 cut off. [49,56]

HER 2$^+$ – About 15-20% of BC. Characterized by high expression of HER2 gene and other HER2 pathway related genes.[49]

HER2 is a member of four membrane tyrosine kinases. "*Upon ligand binding to their extracellular domains, HER proteins undergo dimerization and transphosphorylation of their extracellular domains.*" Due to HER2 lack of ligand, HER2 relies on "*heterodimerization with another family member or homodimerization with itself to be activated*" at high levels. These "*phosphorylated tyrosine residues interact with numerous intracellular signaling molecules*" provoking downstream activation of second messenger pathways and crosstalk with other membrane signaling pathways, activating many genes involved in cell proliferation, differentiation, survival, angiogenesis, invasion and metastasis, [49,57–60] conferring more aggressive biological and clinical behavior. In the absence of treatment, HER2$^+$ have a poor diagnosis and a susceptibility to metastasize to brain and visceral organs.[49]

Basal-like – Low frequency (3-15% of all BC); frequently lack hormone receptors (*e.g*. ER and PR) and HER2, hence are called triple negative breast cancer (TNBC), but not all basal like are TNBC and not all TNBC are basal-like.[49] Either characterized for the expression of keratin 5, 6 and 17, integrin beta, fatty acids and laminin.[50,51,53,61,62] Currently there is "*no molecular-based targeted therapies*"[51] and only 20% have a good response to standard chemotherapy, nevertheless numerous agents are in clinical trials that unfortunately seem to fail treating TNBC (as a result

have a poor outcome) further studies point to a subdivision of TNBC in five or six classes, with specific molecular features, that can explain the low response to drugs; [51,63–65]

Basal like cancers undergo frequent mutations in TP53[e] (tumor protein 53) gene; approximately ¾ of basal-like tumors are BRCA1 (breast cancer 1).

MGP – Matrix Gla Protein

Matrix Gla Protein (MGP) was first isolated from urea extracts of demineralized bovine bone, being the second vitamin K-dependent (VKD)[f] to be discovered in bone, after osteocalcin (also known BGP or bone Gla protein).[66,67] Different tissues (*e.g.* heart, lung, kidney and breast) express MGP.[66,68,69]

Human MGP has a single copy gene[66] located in the small arm of chromosome 12 (12p12.3) and encodes a 10-kD (kilo Dalton) skeletal extracellular matrix (ECM) protein.[70]

MGP primary structure consists in a signal peptide, a phosphorylation domain and a γ-carboxylase recognition site. MGP has also five Gla residues (γ-carboxylated glutamic acid) (Figure 11), which are converted by γ-glutamyl carboxylase enzyme (dependent of vitamin K[66,71,72] – using it as a cofactor)[70] through a posttranslational modification; this alteration gives high affinity for mineral and mineral ions[70,73,74] like calcium, phosphate and hydroxyapatite crystals (mineral components of skeletal ECM).[70]

Figure 11 | Schematic representation of matrix Gla protein (MGP), pointing out the primary structure and Gla domains.
Primary structure: Signal peptide – SP; Phosphorylation domain – P, in a green box (phosphoserine residues are represented by green dots; γ-carboxylase recognition site – γ, in a blue box.
Gla residues: blue dot. Adapted from Cancela *et al* (2014)

MGP variants

For several years MGP was known as a four exon gene (with a total of 1398 base pairs [bp]) but recently a specific MGP variant was identified in human fetal tissue, with an insertion of a new protein domain, resulting in the synthesis of a longer protein

[e] TP53 – evidence of genomic instability and inactivation of the retinoblastoma (Rb) pathway.[49]
[f] VKD proteins have several functions, including: hemostasis, apoptosis and growth control.[74]

isoform (with 1473bp). This novel isoform was labeled E5 (contain five exons) while the previous isoform was labeled E4 (contain four exons) (Figure 12). Isoform E5 have three putative additional sites of gamma carboxylation,[71,72] this feature could increase MGP binding capacity, resulting in a more powerful inhibitory function.

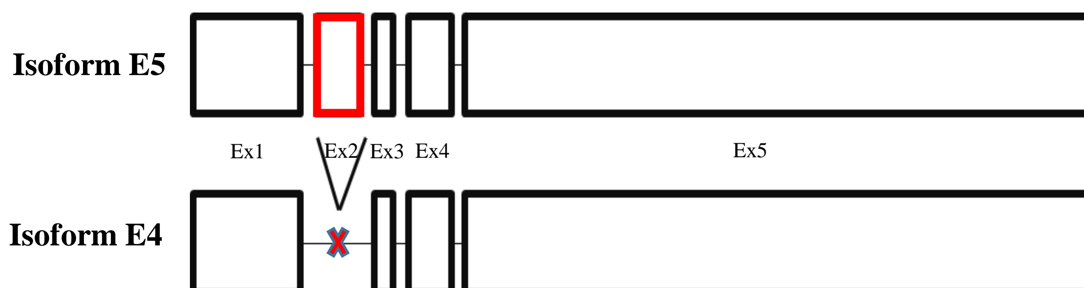


Figure 12 | **MGP schematic representation**. Variant E5 (5 exons) and variant E4 (4exons) – novel isoform (E5) have one extra exon (75bp) due to alternative splicing in MGP gene. Exon at red (Ex2) is only present in isoform E5.

Alternative splicing

Humans produce around 90 000 different proteins, and scientific community was expecting that human genome project would reveal a similar number of genes ("one gene to one protein") so it was a huge surprise that the real number of genes is almost four times smaller than expected – about 25000 genes. Insights in this paradigm resulted in the concept of alternative splicing – a mechanism where a single gene is the origin of multiple proteins, tallying protein diversity.[75] In average, one human gene has eight exons and seven introns, that gives rise in average to three or more alternatively spliced mRNA (messenger RNA). Recent high-throughput sequencing studies point that 100% of human genes produce at least two or three isoforms.[76]

The first transcript of the gene is pre-messenger RNA (pre-mRNA) – copy of the gene including both introns (intended to be removed during pre-mRNA processing) and exons. In RNA splicing exons are retained and exons targeted for removal in order to create different mRNA from the same primordial information (pre-mRNA). The same nucleotide sequence can be an exon or an intron, depending on the splicing target.[77]

Characterization of the protein and its functions

Fraser and Price in 1988[69] described MGP as the first well characterized substrate of VKD found in discrete tissue-specific cells (kidney, lung, heart and spleen

cells), showing that MGP function is beyond specific connective tissues, thus, is unlikely that MGP acts just for accumulation in an extracellular matrix.

In 1997, Luo *et al*[73] developed homozygous mice without Mgp expression (MGP^{m1}/ MGP^{m1}). Those mice until the second week were similar to control mice but then they became shorter and with a faster heart beat; within two months they died due hemorrhage triggered by thoracic or abdominal aortic rupture, showing signals of prominent calcification. Findings suggested:

- Mgp inhibits calcification of certain ECMs, probably as a result of its capacity of mineral-ion-binding.
- In early development other gene products play an important role to prevent earlier calcification, otherwise newborn or one-week-old MGPm^{1}/ MGP^{m1} mice must have showed defects. [73]

Studies show also the involvement of MGP in cell differentiation[71,78] and proliferation.[71,72]

Mutations (with loss of function) in this gene were related with: atherosclerosis (vascular calcification), cancer[68,79,80] and were proven to lead Keutel syndrome whom is characterized by abnormal formation of cartilage calcifications (in auricles, nose, larynxes, trachea and ribs), multiple peripheral pulmonary stenosis, neural hearing loss and short terminal phalanges. [70]

MGP and cancer

As stated above, MGP was showed a role in cell differentiation and proliferation, and was suggested a role also in cell migration, differentiation, onset of angiogenesis[81] and tumorigenesis.[72,82,83] However, MGP role in oncogenesis is unclear as well as the correlation between its expression and tumor type.[68] Some studies were performed and in some tumor types expression was higher and in other types lower than in normal tissue, giving in some cases prognostic value.

Regarding MGP mRNA expression analyses we can see:

Ovarian cancer – MGP mRNA expression in tumor was higher compared to normal tissue. [68,84]

Renal cell carcinoma – higher expression in tumor tissue with a significant inverse correlation between mRNA expression and tumor size, lymph node metastasis and grade.[68,85]

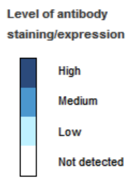
Colorectal cancer – Fan *et al* (2001) [86] found a down-regulation of MGP mRNA in human adenocarcinomas but results obtained in our laboratory comparing cancer tissue and paired adjacent normal tissue showed an up-regulation.[71]

Astrocytic gliomas (AG) – comparing high and low grade AG, mRNA levels were significantly increased in high-grade tumors. [68,83]

Regarding protein expression, The Human Protein Atlas[87] stained twenty different types of cancers and controls, the results (Table 3) in brief, show that malignant cells exhibited moderate to strong cytoplasmic and nuclear staining. Samples vary from 4 to 12, and the most exciting result regard MGP protein expression is in gliomas and liver cancers, where it could be observed a small upregulation of MGP relatively to normal tissue. Analyses of more samples could give more reliable results.

Table 3 | MGP protein expression in tumoral vs normal tissue staining. Colors represent different levels of antibody/staining expression per sample. Adapted from The Human Protein Atlas.

Tissue	Cancer staining	Normal tissue staining	Tissue	Cancer staining	Normal tissue staining
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		



MGP and Breast Cancer

A study using cDNA hybridization revealed that MGP expression in breast cancer is 20-fold higher in metastatic cancer (cell line 600PEI) than in normal epithelium, [68,88] and Yoshimura *et* al (2009) presented MGP mRNA expression as a potential prognostic factor in BC patients, being overexpressed in poor prognostic patients (using microarray analyses). However, immunohistostaining of breast tissue microarrays did not revealed a correlation between overall survival or ER status metastasis with protein expression,[68] this information is also supported by results in Table 3, where MGP is highly to medium expressed in cancer staining and normal tissue is also highly expressed.

Cell culture

Is unethical starting research in humans or other animals before testing it before, for this reason a variety of models were and are being developed. For instance, cancer cell lines which are used currently in several biomedical research laboratories.[89,90]

Cell lines are obtained by primary cultures (growing cells from tissue taken directly from an individual growing in a flask or petri dish. They provide an almost unlimited supply of cells (with similar genotypes and phenotypes). There are various observations that reveals cell lines as a good model for cancer research, as a first approach, for instance:[90]

Histopathology: most human cancer cell lines transplanted to immunedeficient mice lead to tumors; histopathologic analysis of 127 human cell lines that formed tumors in mice reveled that they are all correlated with tumor origin. [90]

Molecular genetics and receptor expression: comparison between lung cancer and breast cancer cell line with origin cancers revealed that phenotypic (as ER receptors expression) and genotypic properties are retained for long periods of time. [90]

Aims

More and more cancer research is focused on target therapies and biomarkers. This kind of approach is imperative to refine strategies of screening and diagnosis leading to early detection, that is correlated with better prognosis. Furthermore, target therapies are known as less toxic and with reduce side effects, improving the wellbeing of cancer patients. To improve this approach is necessary to enlighten the role of genes related with tumorigenesis. Knowing that MGP role in breast cancer (and cancer in general) is unclear, our biggest aim is to try to clarify it, searching for correlations between its expression in normal vs tumoral samples of breast cancer, as well as in particular subsets of patients. Alongside our main aims were either:

Clarify isoforms expression in breast cancer cell lines, as well as the result of isoforms overexpression in cellular proliferation and migration.

Identify some regulatory mechanisms of MGP expression, in particular epigenetic (methylation) and posttranscriptional regulation (mir155 analyze, even as, its correlation in normal and tumoral tissue.

Chapter II - Materials and Methods

Cell lines and cultures

It was used eight cell lines from human breast cancer tissue; those, five were commercial lines[91] and three were primary cultures obtained from patient samples from CUF. Cells were maintained in the appropriate cell culture medium according to their specific requirements, as seen Table 4.

Table 4 | Cell lines characterization and culture medium conditions; RPMI (Roswell Park Memorial Institute; Gibco, Life Technologies) 1640 Medium GlutaMAX; DMEM (Dulbecco's Modified Eagle Medium; Gibco, Life Technologies); P/S – Pen Strep (Penicillin/Steptomycin mixture, Life Technologies); FBS (fetal bovine serum, Sigma Aldrich)

	Cell line	Gender	Age	Disease	Cell type	Culture Medium	FBS (%)	Glutamine (%)	P/S (%)
Commercial	ZR-75-01	Female	63	Ductal carninoma	Epithelial	RPMI 1640 GlutaMAX	10	0	1
	MDA-MB-231	Female	51	Adenocarcinoma	Epithelial	DMEM	10	1	1
	MCF-7	Female	69	Adenocarcinoma	Epithelial	DMEM	10	1	1
	T47D	Female	54	Ductal carninoma	Epithelial	DMEM	10	1	1
	HCC1428	Female	49	TNM stage IV, grade 4, adenocarcinoma	Epithelial	DMEM	10	1	1
Primary culture	CFI-TERT	Female				DMEM	20	1	1
	TNBC-1	Female	39			DMEM	20	1	1
	BC	Female	64	Cell pellet					

All cell lines were maintained at 37ºC (Celsius degrees) and at 5% CO_2 (carbon dioxide) and subdivided 1:2 three times a week, except TNBC-1 that was subdivided one to two times a week. BC cells were obtained already in a pellet for downstream applications.

RNA extraction and purification

Extraction and purification of total RNA was performed with GeneJET RNA Purfication Kit (ThermoScientific)[92], according to manufacturer's instructions.

Suspension and adherent cells have different needs in sample preparation for pelleting, thus:

- Up to $1x10^7$ **cells in suspension** were collected in culture medium and centrifuged for 5 minutes at 250 x g. After discarding the supernatant, the cells were rinsed with PBS and centrifuged again to remove residual growth medium.
- Culture medium from **adherent cells** was removed and cells were rinsed with PBS to assure total removal of residual medium. Washing PBS was discarded and an additional 1mL of PBS was added to the petri plate, cells were detached using a scraper and transferred into a 1.5mL microcentrifuge tube and pellet by centrifugation (5 minutes at 250 x g).

In both cases after discarding the supernatant the pellet was used do extract RNA or stored at -70ºC until use.

Pellets were resuspended in 600 µL of Lysis Buffer supplemented with β-mercaptoethanol, (the mixture presents a detergent basis use to break lipid barriers in cells[93], exposing genetic information; solution contains also guanidine thiocyanate – (chaotropic) salt capable of cell lysis and protecting RNA from endogenous RNases[RNA nucleases]) and vortexed for 10 seconds to mix. The lysates were homogenized passing the mixture into a blunt 20-gauge needle fitted to an RNase-free syringe several times. This step is important because an incomplete homogenization results in a significant reduction of RNA yields.

After adding 360 µL off ethanol (96%-100%) (to precipitate RNA[94]) and mixing by pipetting, up to 700 µL of the lysate/mixture was transferred to a GeneJET RNA Purification Column[g] (inserted in a collection tube) and centrifuged for 1 minute at ≥12000 x g. The flow-through was discarded. This process was repeated till all the lysate was processed in the column and centrifuged; at the end the collection tube was changed. Then 700 µL of Wash Buffer 1 was added to the Column and centrifuged for 1minute at ≥12000 × g, the flow-through was discarded and 600 µL of Wash Buffer 2 was added and centrifuge and discard step were repeated. An additional volume of 250 µL of Wash Buffer 2 was added and centrifuged for 2 minutes at ≥12000 × g. After discarding the flow –through one plus step of centrifugation were performed at maximum speed for 1 minute to guarantee that all the residual solution was discarded. The final step – RNA elution was performed placing the column in a 1.5mL RNase-free microcentrifuge tube and adding 50 µL of sterile water (Sigma) in the center of the

[g] GeneJET RNA Purification Column – with a sílica-based membrane technology that presents acid-bindings properties in the presence of guanidine thiocyanate.

column, after one final centrifugation at ≥12000 × g for 1 minute the column was discarded and purified RNA was used or stored at -20ºC or -70ºC until use.

Quality and quantity of RNA was assessed by spectrophotometry (Nanodrop ND-1000, Thermo Scientific), through reading the absorbance of each sample at 230, 260 and 280 nm, and determining the ratios at 260/230 nm and 260/280 nm.

Synthesis of cDNA by Reverse Transcriptase reaction

Total RNA (1 µg) was treated with RQ1 DNase (1U for each 1 µg of RNA: Promega), 1 µL of 1x RQ1 DNase Buffer (10x, Promega) and water (Sigma) till the final volume of 10 µL, according to manufacturer's instructions. After incubation (30 minutes at 37ºC) the reaction was stopped by adding 1 µL RQ1 DNase Stop Solution (1U for each 1 µg of RNA, Promega) and incubating for 10 minutes at 65ºC. This step uses a endonuclease (DNase I) to degrade double and single-stranded DNA assuring the integrity of RNA.

RNA samples after the DNase treatment were reverse transcribed into cDNA using Moloney-murine leukemia virus (M-MLV) reverse transcriptase (200U/µL). For that, it was added to RNA samples:

- 4 µL (1x) First Strand (FS) Buffer (5x FS Buffer, Invitrogen);
- 1 µL DTT, (0.1M, Invitrogen);
- 1 µL oligo dT primer (10µM, Invitrogen);
- 1 µL dNTPs (10mM each [dATP, dTTP, dCTP, dGTP]).

Incubated at 65ºC, 5 minutes

Then the mixture stayed for 5 minutes at 4ºC and were supplemented with

- 1 µL of RNase Out (40U/µL, Invitrogen) – incubated at 37ºC, 2 minutes.
- 1 µL M-MLV – incubates at 37ºC for 50 minutes and to inactivate the reaction incubated at 70ºC for 15 minutes.

The cDNA can at this point be used as a template for PCR or qRT-PCR or stored at -20ºC.

qRT-PCR – Quantitative Real Time – Polymerase Chain Reaction

cDNA samples obtained by reverse transcriptase reaction were used as templates for real-time quantitative polymerase chain reaction (qRT-PCR) for analyze the expression of MGP mRNA, using specific set of primers (Appendix table 4).

PCR amplification was placed in a 96-well PCR micro-plate (Thermo Scientific); where 2 µL of cDNA (diluted 1/10) was mixed with 10µL of SsoFast™ Eva Green Mix (BioRad), 0.6 µL of each forward and reverse primers (10 µM) and 6.8 µL (to a final volume of 20 µL). Micro-plates were sealed with adhesive PCR film (Thermo Scientific) and amplifications were performed in a CFX96™ Real-Time PCR Detection System (BioRad) under the following conditions: denaturation and polymerase activation at 95ºC for 30 seconds; 40 cycles of denaturation (5 seconds at 95ºC) and annealing/elongation (30 seconds at 57ºC); melt curve (65ºC to 95ºC, incrementing 0.5ºC) for determination of melting point.

As a negative control PCR amplifications were performed with sterile water; a relative expression was determined using $2^{-\Delta\Delta Ct}$ method[95]. Threshold cycles obtained for each variant were normalized using a housekeeping gene[h] GAPDH (Glyceraldehyde-3-phosphate Dehydrogenase) and the expression of MCF-7 cell line which previous work in the lab confirmed the presence of the two isoforms.

Confirming results of qRT-PCR by Electrophoresis and Sequencing

Electrophoresis is the standard protocol to separate DNA by size, and was performed in the products of qRT-PCR, to confirm the presence of different isoforms (with different DNA fragments). Samples were loaded on a 1.5% agarose gel containing Green Safe Premium (NZY Tech) intercalating dye and run with 1x TAE (Tris-Acetate-EDTA) buffer at 120 Volts (V) for 15-20 minutes. The fragments were visualized under UV light.

PCR products with the expected size were removed with a sterile blade and DNA was purified using GeneJET Gel Extraction Kit (Thermo Scientific) according manufacturer's instructions; the technique underlie under the principle that in high temperatures (50ºC to 60ºC) the agarose gel melt; using binding properties of DNA to the silica based column and washing buffers proteins and contaminants are removed and DNA is purified and diluted in sterile water. A sample of the purified DNA and specific primers was sent to sequencing evaluation at CCMAR facilities.

[h] Housekeeping genes – they encode "proteins that are required universally to maintain viability of all cell types throughout the body or to carry out certain biological functions common to all cell types."[5]

Transient transfection

Cells were transfected with EzWay™ Transfection Reagent (Koma Biotech inc.) – a cationic lipid with DNA binding affinity.

The day before the transfection cells were plated to rich 80 – 90% of confluency in the next day – previous tests were made with a differential number of cells to assess best transfection conditions; transient transfections were performed in TNBC cells and CFI-TERT. Per each well the transfection solution was added drop by drop and was prepared according to the manufacturing instructions. For 24 well plate ($3x10^4$ cells/well): 0.8 µg of DNA was diluted in 50 µL of Opti-Mem (Gibco, Life technologies, ThermoFisher); 1 µL of reagent diluted in 50 µL of Opti-Mem; both solutions were combined gently and incubated for 30 minutes at room temperature, to allow liposome-DNA complexes to form; and then 100 µL of the complex was added per well. In 96 well plate the protocol was the same, but were placed a smaller number of cells ($1.5x10^4$ cells/well) and DNA (3.5 µg). As a negative control were used wells without transfection reagents and as a positive control an empty plasmid (pcDNA). Plasmids tested contained E4 and E5 isoforms (work done previously in the lab).

Cell Viability Assays

XTT (tetrazolium salt, AppliChem) was performed to assay viability of wild type (WT) cells in comparison to transfected cells with E4 and E5 variants.

Live cells can reduce tetrazolium salts into colored (orange) formazan compounds through mitochondrial enzymes, which are inactivated after death. The dye (water soluble) intensity can be measured (wavelength) in a spectrophotometer (microplate reader Synergy 4). It was performed a XTT optimization protocol to evaluate the optimal incubation time of the reagent, and hourly measurements were performed at 450nm and 690 nm (for background corrections) after adding XTT to cells as described by the manufacturer – it was determined an optimal incubation time of 4h. Right after transfection it was added 50 µL of XTT mixture to each well (ratio for XTT preparation was 1 µL of activation solution to 50 µL of XTT reagent). XTT measurements were performed at time 0h (zero) – 4h after transient transfection; 24h, 48h and 72h. Blank measurements (triplicates) were performed in wells containing only growing medium and XTT and the average absorbance at 450nm was subtracted to specifics results.

Specific absorbance was defined by subtracting results of absorbance at 690nm for results of 450nm absorbance.

Wound healing assay – motility assay

Wound healing assay is used to assess cell migration. $3x10^4$ of CFI-TERT cells were plated in 24 well plate and transfection assay were performed as described before. When cells reached about 90% of confluency the bottom of each well was scratched using a sterile tip (1000 µL) (Figure 13), the scratch is used to mimic a wound, and should be possible visualize the two edges using a magnification of 100 times. At least 3 photographs (Canon PowerShot G12) were taken of each condition (with replicates) at 0h, 24h, 48h, 72h and 96h. After the scratch and before taken pictures the wells were rinsed two times with PBS. At the end was calculated the percentage of invaded area.

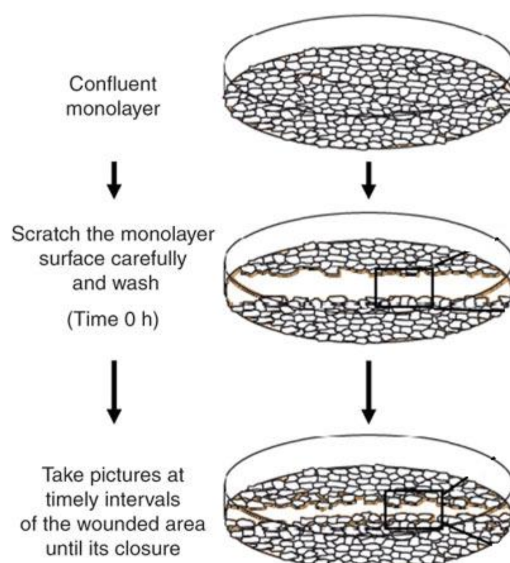


Figure 13 | Wound healing assay draft

Demethylating treatment with 5 – Azacytidine

5-Azacytidine (5Aza) (Sigma Aldrich) is a pyrimidine ring analog of cytidine, that inhibit enzymes that methylate cytosine residues in eukaryotic DNA, altering methylation status and activating genes that other way were inactivated by methylation. Because of its cytotoxic effect additional safety procedures were taken.

5Aza was dissolved in PBS (1x – phosphate buffered saline) to a 5mM concentration and then it was added to T47D cells in a 6 well plate (plated with $6x10^5$ cells in the previous day). The treatment was repeated after 24h and 48h; and after 72h RNA and DNA extraction was performed. As a negative control it was added only PBS (without treatment) to the cells. 5Aza is very unstable in aqueous solutions, so fresh solutions were made every treatment day.

Genomic DNA extraction from culture cells

Genomic DNA was extracted using DNeasy Blood & Tissue kit (Qiagen). The kit is based in a buffer system optimized to allow direct cell lysis (by proteinase K) followed by selective binding of DNA to the DNeasy membrane (due to high concentration of chaotropic salt, that removes water from hydrated molecules in solution). Centrifugation process and buffer condition are used for completely remove contaminants and enzymes inhibitors like proteins.

Culture medium was removed and cells were rinsed with PBS to assure total removal of residual medium. Washing PBS was discarded and an additional volume of appropriate PBS was added to each well, cells were detached using a scraper and transferred into a 1.5 mL microcentrifuge tube and pellet by centrifugation (5minutes at 300 x g); pellet was then ressuspended in 200 µL PBS and 20 µL of proteinase K was added; as well 200 µL of Buffer AL (mixed by vortex), the homogeneous mixture was then incubated at 56ºC for 10 minutes; afterward was added 200 µL of ethanol (96-100%) and the mixture was mix through vortex; at this point solution was pipetted to a DNeasy Mini spin column and was centrifuged at ≥6000 x g per 1 minute; the flow-through was discarded as well the 2 mL collection tube, that was replaced for a new one; 500 µL of buffer AW1 was added and centrifuged again for 1 minute at ≥6000 x g; the flow-through was discarded as well the collection tube and an additional volume of 500 µL of buffer AW2 was added and centrifuged for 3 minutes at 20000 x g, to dry the DNeasy membrane, to guarantee that residual ethanol was not present, once it could interfere with succeeding reactions; the spin column was placed in a 1.5 mL microcentifuge tube and 100 µL sterile water (Sigma Aldrich) was added; an incubation time of 1 minute at room temperature was performed before the last centrifugation for 1 minute at ≥6000 x g.

Quality and quantity of DNA was assessed by spectrophotometry (Nanodrop ND-1000, Thermo Scientific), through reading the absorbance of each sample at 230, 260 and 280 nm, and determining the ratios at 260/230 nm and 260/280 nm.

Methylation analysis by bisulfite conversion

Sodium bisulfite conversion is the standard method for methylation analyses. The DNA methylation, that usually occur in CpG dinucleotides, modify cytosines into

5-methylcytosines (5-mC), and the outcome of incubation of target DNA with sodium bisulfite is the conversion of unmodified cytosines into uracils (5-mC stay intact).

Sodium bisulfite conversion was performed on genomic DNA, extracted from cells as described above, using the EpiMark Bisulfite Conversion kit (New England Biolabs Inc). Briefly, in a 0.2 mL PCR tube it was gently mixed by pipetting 1 µg of genomic DNA and sterile water (Sigma Aldrich) – up to 10 µL – and 130 µL of bisulfite mix; mixture was then transferred to a thermocycler (Biometra) under the following conditions on Table 5.

Table 5 | Cycling protocol for bisulfite conversion reaction.

Cycle Step	Temperature	Time (minutes)
Denaturation	95ºC	5
Incubation	65 ºC	30
Denaturation	95 ºC	5
Incubation	65 ºC	60
Denaturation	95 ºC	5
Incubation	65 ºC	90
Hold	18 – 20 ºC	Up to 12 hours

The next step was the desulphonation and clean-up of the converted DNA that was accomplished by transferring the individual reactions into 1.5 mL microcentrifuge tubes and adding 550 µL of DNA Binding Buffer. After mix it briefly, the sample was them loaded in an EpiMark spin column with a 2 mL collection tube and centrifuged for 1 minute at 15000 x g. The flow-through was discarded and 500 µL of Wash Buffer was added, the centrifugation step was repeated, the flow-through discarded, and 500 µL of Desulphonation Reaction Buffer was added and the mixture was left incubating at room temperature for 15 minutes. Afterwards the centrifugation step was repeated; two wash steps were performed adding 500 µL of Wash Buffer followed by the centrifugation step; the spin column was putted into a sterile 1.5 mL microcentrifuge tube and 20 µL of sterile water (Sigma) was incubated for one minute at room temperature and a last centrifugation step was performed. DNA eluted was then analyzed by PCR.

CpG islands upstream of the *MGP* gene were identified based on information in the UCSC genome browser (https://genome.ucsc.edu). Primers to amplify an upstream CpG island were designed with MethPrimer software (http://www.urogene.org/methprimer) which amplified a region containing one CpG. PCR analyze was performed with two different Taq DNA polymerase (Kapa

Biosystems polymerase kit and Invitrogen kit), according manufacturer's instructions, as described in Table 6. The PCR products were run on a 1.5% agarose gel containing Green Safe Premium (NZY Tech) dye at 120V V for 15 – 20 minutes. Resulting DNA fragments of the expected size were purified using the GeneJET Gel Extraction Kit (Thermo Scientific) according manufacturer's instructions and sequenced at CCMAR facilities and methylation status of the CpG was assessed.

Table 6 | PCR conditions using two different Taq polymerase, in particular Kapa Biosystems kit and Invitrogen Kit.

PCR component – Kapa Biosystems		PCR Run			
Reagents	25 µL reaction	Step	Temperature (ºC)	Duration	Cycles
5x KAPA HiFi buffer	5 µL	Initial denaturation	95	30 s	1
		Denaturation	95	15s	40
10 mM dNTPs	0.5 µL	Annealing	55	30 s	
10 µM Forward Primer	0.5 µL	Extension	68	1 min.	
10 µM Reverse Primer	0.5 µL	Final extension	68	5 min	1
Genomic DNA	6 µL	Hold	4	Indefinitely	
KAPA HiFi HotStart DNA Polymerase	1 µL				
Sterile water	11.5 µL				
PCR component – Invitrogen		**PCR run**			
Reagents	50 µL reaction	Step	Temperature (ºC)	Duration	Cycles
10x PCR bufffer	5µL	Initial denaturation	94	3 min.	1
50mM $MgCl_2$	1,5 µL	Denaturation	94	45 s	40
10 mM dNTPs	1 µL	Annealing	55	30 s	
10 µM Forward primer	2.5 µL	Extension	72	90 s	
10 µM Reverse	2.5 µL	Final extension	72	10 min	1

Primer				
Genomic DNA	6 µL	Hold	4	Indefinitely
KAPA HiFi HotStart DNA Polymerase	0.25 µL			
Sterile water	31.25 µL			

Bioinformatic database – The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) is the outcome of a collaboration between The National Cancer Institute's (NIH) and the National Human Genome Research Institute (NHGRI). Is a platform of sharing data that includes multi-dimensional maps of the key genomic changes in 33 types of cancer. The data is collected by tumor tissue and matched normal tissue. In particular, breast cancer is referred as BRCA (Breast invasive Carcinoma), and more than one thousand samples are available for analyze.

Patients selection

Using Cancer Genomics Browser (University of California in Santa Crus, USA) and cBioPortal (Memorial Sloan Kettering – Cancer Center) we were able to download selected information for every sample; then we select patients with matched samples for tumoral and normal tissue[i]. Selected samples are from women – only one man matched sample was available, and was excluded because it would not be statistical relevant); for the same reason one sample from one Asian woman was excluded. Information was organized in an Excel sheet (Microsoft Office, 2016) and analyzed with SPSS (IBM Analytics) and GraphPad Prism 6 (GraphPad software, Inc.). Selected information can be seen in Appendix table 5.

[i] Normal Tissue – samples were collected from adjacent tissue of the tumor.

Chapter III – Results and Discussion

MGP expression in different cell lines

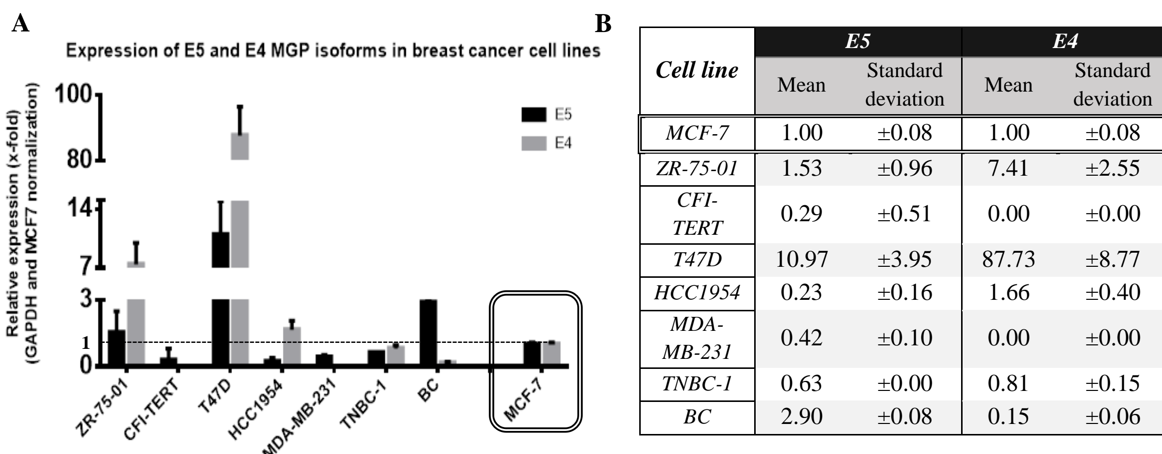
Several breast cancer cell lines were analyzed to assess the presence of the two known variants of MGP – E4 and E5. First, it was performed a qRT-PCR, and then the results were confirmed by electrophoresis and sequencing. The set of primers used to assess expression of:

– **E5 isoform** was F1 (drawn in exon 1) and R4 (drawn in exon 2 – exclusively of E5 variant).

– **E4 isoform** was F1 and F5 (drawn in exon 5) (Figure 14).

Figure 14 | Primers localization in *MGP* gene. The red band represents Forward 1 (F1) primer, in exon 1; orange band represents Reverse 5 (R5) primer, in exon 5 (used to identify isoform E4 with F1); and blue band represents Reverse 4 (R4) primer, in exon 2 (used to identify isoform E5).

As shown before, qRT-PCR results were normalized using $2^{-\Delta\Delta Ct}$. Expression results obtained are a comparison between the several cell lines tested and MCF-7, which expression values are equal to one (Figure 15).



Cell line	*E5* Mean	*E5* Standard deviation	*E4* Mean	*E4* Standard deviation
MCF-7	1.00	±0.08	1.00	±0.08
ZR-75-01	1.53	±0.96	7.41	±2.55
CFI-TERT	0.29	±0.51	0.00	±0.00
T47D	10.97	±3.95	87.73	±8.77
HCC1954	0.23	±0.16	1.66	±0.40
MDA-MB-231	0.42	±0.10	0.00	±0.00
TNBC-1	0.63	±0.00	0.81	±0.15
BC	2.90	±0.08	0.15	±0.06

Figure 15 | **Relative expression of E4 and E5** ***MGP*** **isoforms in several breast cancer cell lines.** Expression was obtained from qRT-PCR – normalized using $2^{-\Delta\Delta Ct}$ method relative to *GAPDH* (housekeeping gene) and MCF-7.

At first sight E5 seem to be expressed in every cell line and E4 seem be expressed in low quantity or absent in CFI-TERT; MDA-MB-231 and in BC. To confirm results from qRT-PCR an electrophoresis assay was performed and the bands were extracted and undergo sequencing was possible to verify the presence or absence of the two isoforms in each cell line.

Electrophoresis showed very different results than qRT-PCR. Isoform E4 is present in every cell line and E5 appears to be present in ZR-75-01 and T47D and maybe in CFI-TERT (Figure 16).

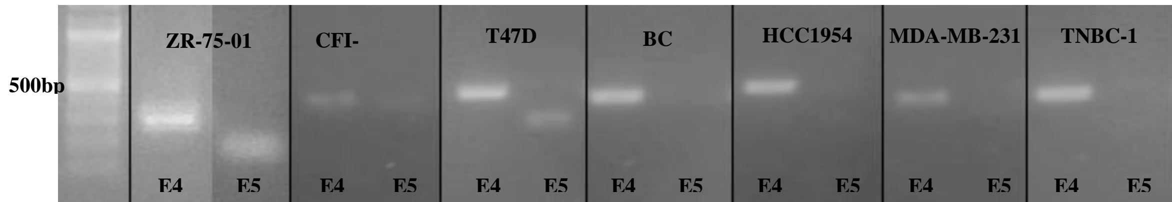


Figure 16 | Electrophoresis assay suggesting the presence of E4 in all cell lines and the presence of E5 only in ZR-75-01, CFI-TERT, T47D and HCC195 (a light band appeared, although is not visible in the picture).

Due to the disparity in qRT-PCR and electrophoresis assay sequencing was essential to confirm the results (Figure 17). We could conclude that on fact E4 is expressed in every cell line and E5 seems to be expressed only in ZR-75-01; T47D and HCC1954 cell lines.

Figure 17 | Chromatograms from sequencing revealed that E4 is present in every cell line and confirmed the presence oE5 only in ZR-75-01, T47D and HCC1954 cell lines. (Images obtained using Geneious, Biomatters Limited)

Even though preliminary results derived from qRT-PCR suggested the E5 expression in BC cell line it was not possible to detect a band in electrophorese neither sequencing.

From the three techniques, sequencing is the more reliable, thus, although the disparity of the results, is secure to say that E4 expression is certain in all cell lines and E5 expression is certain in only three cell lines. qRT-PCR were performed in machines that we later found out were not working well, in a certain way the disparity of results could be explained because of that.

Specificity of primers could be an explanation either. However, several set of primers were used to assess the isoforms expression and the set F1+R4 and F1+R5 showed to be the more reliable. In spite of the fact that F1+R5 in theory is not specific for E4 isoform, sequencing results showed only the expression of E4, either in this analysis as in others performed previous in our lab.

Proliferation and Migration Assays

XTT – CFI-TERT and TNBC

An evaluation of the optimal incubation time of the reagent was determined by hourly measurements and was established in 4h for both cell lines (CFI-TERT and TNBC).

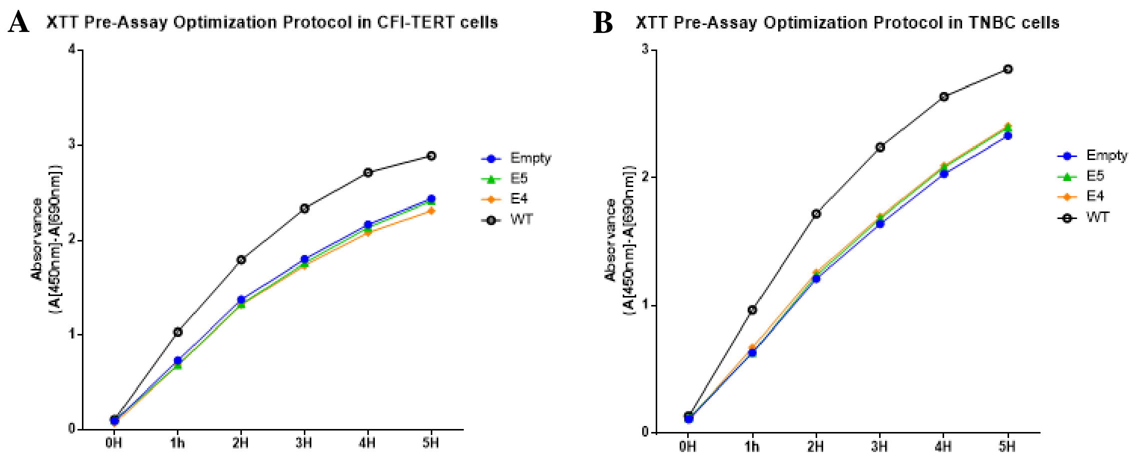


Figure 18 | XTT assay to determine optimal incubation time of the reagent in: A) CFI-TERT and **B)** TNBC cell lines.

As a control we used WT (wild type – cells not transfected) and cells transfected with an empty vector, to exclude faux variations derived from the transfection assay. We expected to see that both WT and cells transfected with the empty vector proliferate similarly.

In CFI-TERT cell line (Figure 19**A**), transfected cells present a similar behavior, very different from the WT, suggesting that the variation observed is due to the effect of transfection. In TNBC (Figure 19**B**) cells, at initial time zero (four hours after the transfection and XTT protocol) we see a big variation, and then a similar behavior is seen; to evaluate if it is a result of the transfection, or human error, it is necessary to

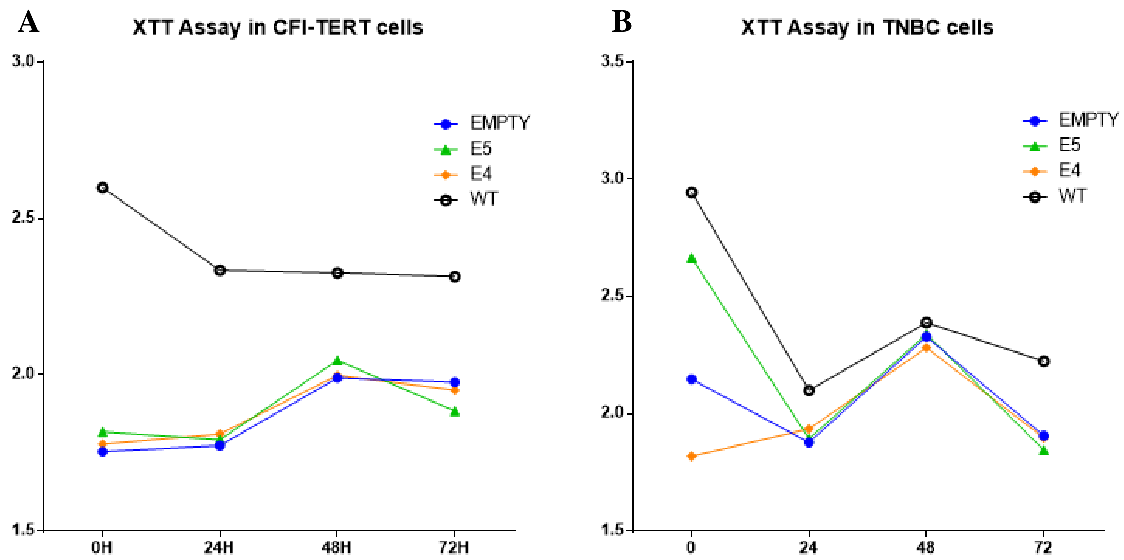


Figure 19 | XTT assay in **A)** CFI-TERT and **B)** TNBC cell lines.

repeat the treatment.

Both cell lines are not commercial, they are more sensitive in specially TNBC cells that have a low growth rate and are easily contaminated. In the second assay TNBC cells got contaminated. XTT treatment was performed in the same plate, for that reason a cross contamination of CFI-TERT could had occur although no evidence was seen, and could alter the results. Additional repetitions were not performed because of time – to rich the necessary number of cells to perform this assay cells must be in culture in average three to four weeks and then XTT protocol is performed (requiring one additional week) – if a contamination occur is necessary return to point zero and defrost a new batch, impairing the task.

Wound healing assay

This assay was performed in CFI-TERT cells, to evaluate cell migration under overexpression of each isoform during 96h. Percentage of invaded area was calculated using the average value of replicates for each condition, at time zero. Results are shown in Figure 20 and Figure 21.

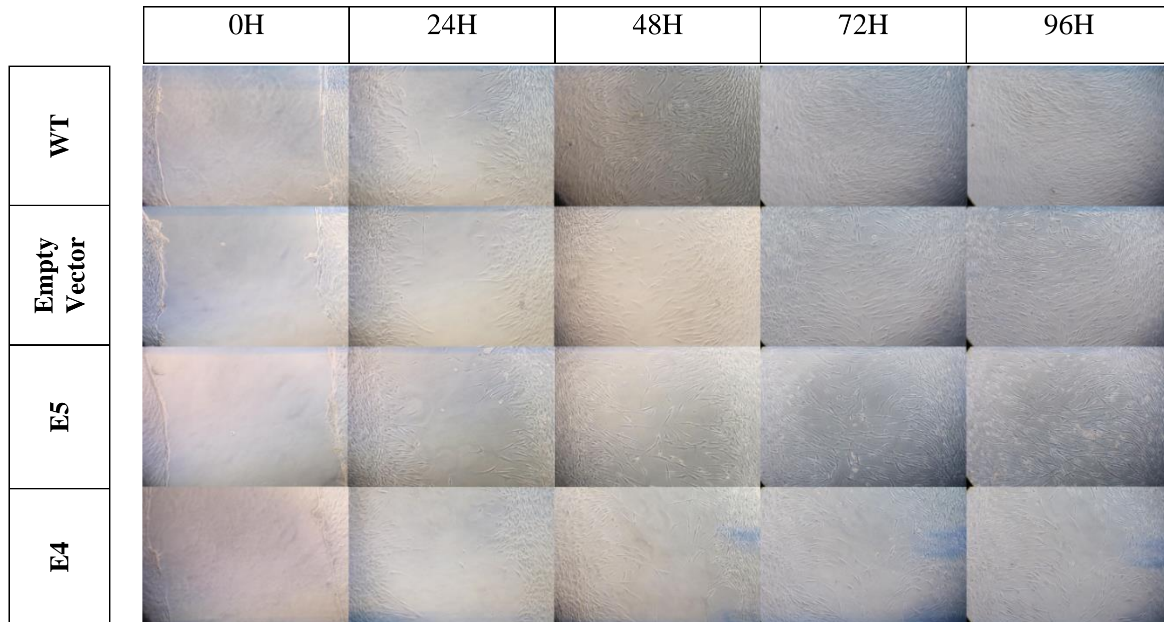


Figure 20 | Wound healing assay over 96h in CFI-TERT cells – WT, transfected with an empty vector or E5, or E4.

Comparing WT and cells transfected with an empty vector, results are similar, as expected. When comparing cells with overexpression of E5 or E4 to WT was possible to observe a reduction of area invaded, revealing an already described paper in cell migration of MGP)

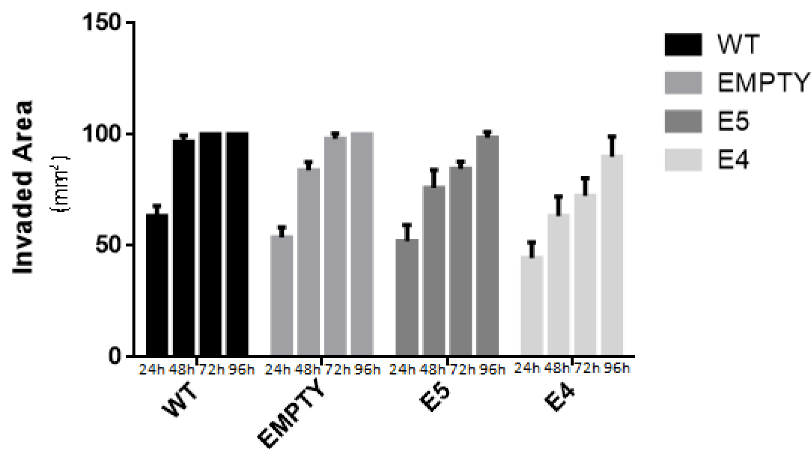


Figure 21 | Invaded área after 24h, 48, 72h and 96 hours in diferente conditions (WT, empty vector, E5 and E4).

Methylation analysis

After 5-Aza treatment was performed a bisulfite conversion, where occurs the conversion of unmodified cytosines into uracils (5-mC stay intact), with in mind this, two set of primers were made (F1/R1 and F2/R2) to identify the absence or presence of methylation status. As a control it was used the T47D cells without 5-aza treatment.

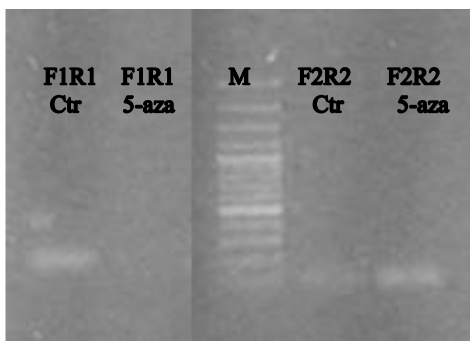


Figure 22 | Electrophoretic assay of PCR samples after bisulfite treatment. Control samples are refered as (Ctr) and samples treated with 5-aza as (5-aza)

After the PCR, an electrophoretic assay was performed and only in F1/R1 control (CTL) did not appear one band (Figure 22). Bands were extracted and undergo sequencing but without no results. Even after altering *Taq* polymerase and conditions of PCR, it was not possible to obtain sequencing results. The expected size of the band was different from the ones obtained, this made us wondering about primer specificity. Furthermore, in one bisulfite conversion of gDNA, occurred a formation of a precipitate, leading us doubting the integrity of the kit.

In summary, the experiment was not conclusive and need more optimization.

Analysis breast cancer genomics data using the cBioPortal

In the absence of envisaged fresh samples of breast cancer tissue to evaluate the levels of *MGP* mRNA expression, in cancer *vs* normal tissue, an *in silico* approach was used, using international TCGA database (breast cancer samples are labeled as BRCA). At a first sight, using a heatmap (https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/ – Figure 23) of MGP expression was possible to observe that there is a difference in gene expression between normal and tumoral tissue, so we further analyzed this data.

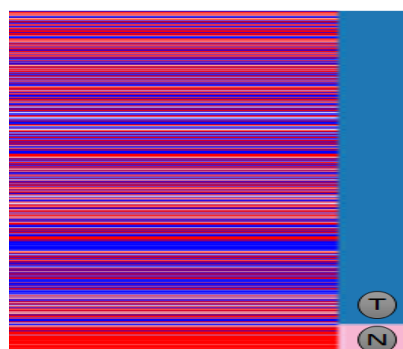


Figure 23 | Heatmap of MGP expression in BRCA samples, presenting a possibility of differential expression in normal vs tumoral tissue. (T) represents data from tumoral tissue (blue) and (N) represents normal tissue (pink). Data was normalized subtracting the mean of gene expression from each sample (data values in blue represent gene expression less than zero; in red more than zero and in white equal to zero – mean). Adapted from Genome Browser Cancer.

At this point, it was important to select a population to assess statistical information from the dataset; using only matched samples of the same patient for normal (tissue collected from an adjacent

position relative to the tumor) and tumoral tissue, to guarantee more reliable results. Adjacent tissue is largely used as a control. However, it is still under debate if it should be treated as "normal" tissue, due specially to these theories: **1)** tumor cell contamination – cells extending beyond the invasive tumor margin; **2)** tumor microenvironment – suggest that normal tissue environment is modified for signals that either could promote or suppress tumor development; **3)** field cancerization – suggests that adjacent tissue is "in an intermediate state between normal and tumor". [96] Until a final confirmation of the reliability of these samples, the following results have clearly some limitations.

MGP mRNA expression is higher in tumor tissue

Comparing the mRNA expression of MGP in matched normal tissue (NT) and tumoral tissue (TT) it was possible to determine that MGP expression is increased by 1.13 (±0.09) times in tumoral tissue (Figure 24). This result is in agreement with previous studies that showed a higher expression (20-fold) in metastatic cancer (compared to normal epithelium)[68,88] as well as a correlation between higher MGP expression and poor prognosis.[68] Although an increase of 1.13-fold could appear not so expressive, comparing with the 20-fold that occur in metastatic tissue, it is important to refer that metastatic tumors are an evolution of primary tumors, with poor prognosis, so it can be expected a smaller increase in tumoral tissue.

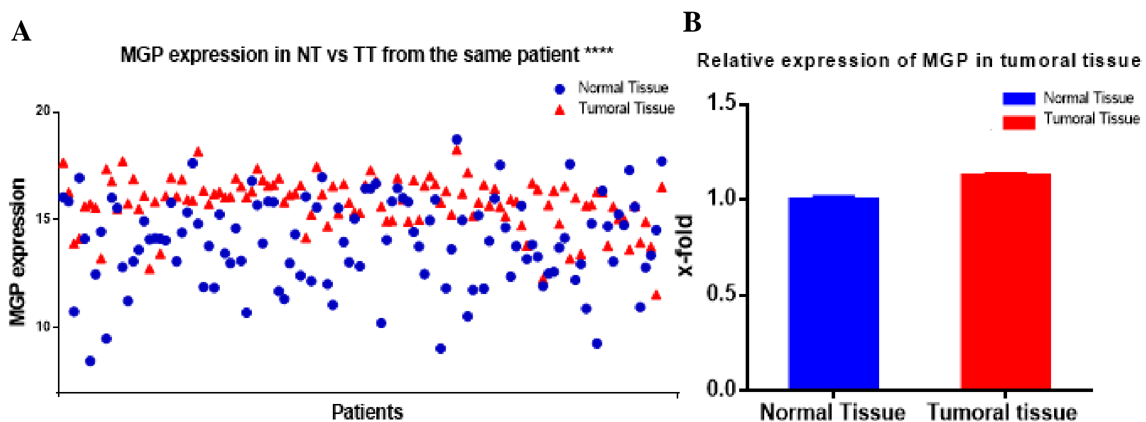


Figure 24 | **MGP expression is higher in tumoral tissue than in normal tissue (BRCA samples). A)** Comparison between MGP expression in NT and TT, per patient – each row represents a patient (n=112) – show a statistical different expression of MGP. Results were obtained by two-way ANOVA with p-value < 0.0001 (****). **B)** MGP is increased 1.13 (±0.09) times in TT.

The absence of fresh samples and results in TCGA database it was not possible to try clarify a correlation between mRNA levels and protein levels, however, previous

studies did not find a correlation between an increase of mRNA and protein expression.[68,87,88]

Methylation pattern and mir155 expression are correlated with MGP expression

Once the implication of mRNA levels as prognosis factor were not demonstrate at a protein level, in our lab, Tiago *et al* postulated the involvement of a post transcriptional regulatory mechanism, and it was showed that MGP (protein) is significantly repressed by miR-155 in BC cell lines, and in parallel it exists a stimulation of cell proliferation and cell invasiveness. [98] mi-RNAs are a subclass of small noncoding RNA with approximately 22 nucleotides with a tremendous role in gene regulation by targeting mRNA for cleavage or translational repression; after binding to the target mRNA (by complementary matching bases) miRNA will prevent translation or promote mRNA degradation.[98,99] Mattiske *et* al had already validated 147 targets of miRNA-155 with several roles in apoptosis, differentiation, angiogenesis, proliferation, and epithelial–mesenchymal transition.[100] Epigenetic (*e.g.* methylation) also play an important role in gene expression, so to assess this hypothesis it was performed a methylation evaluation in the gene as well as in mir155.

MGP in tumoral tissue is hypomethylated in comparison with matched normal tissue

Using TCGA database in Cancer Genomic Brower, methylation data was downloaded and was performed an analyze in matched samples to verify the presence of specific methylation in tumor vs normal tissue. For *MGP* gene there are four probes available along the gene (Figure 25):

CpG1 – cg13302154	CpG2 – cg22221831	CpG3 – cg00431549	CpG4 – cg5360958

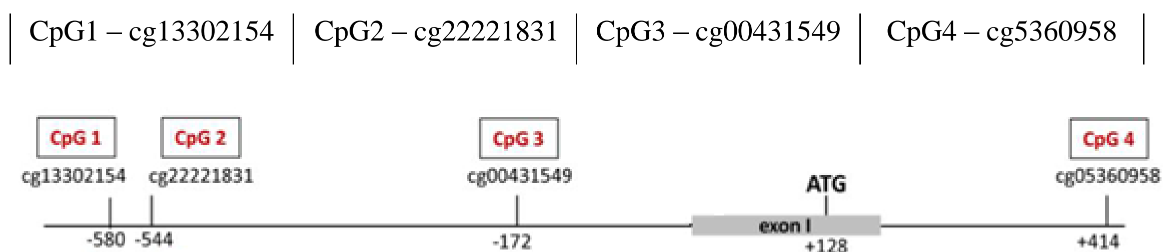


Figure 25 | Localization of methylation probes in *MGP* gene. ATG is the first codon.

We observed that from the four probes, only one site (CpG4) did not shown statistical different methylation status in the matched samples (p(v)=0.0822). The remaining probes suggest an hypomethylation in tumoral tissue vs normal tissue. CpG2 and CpG3 have the most significant difference with p(v)=0.0004 and p(v)=0.0003,

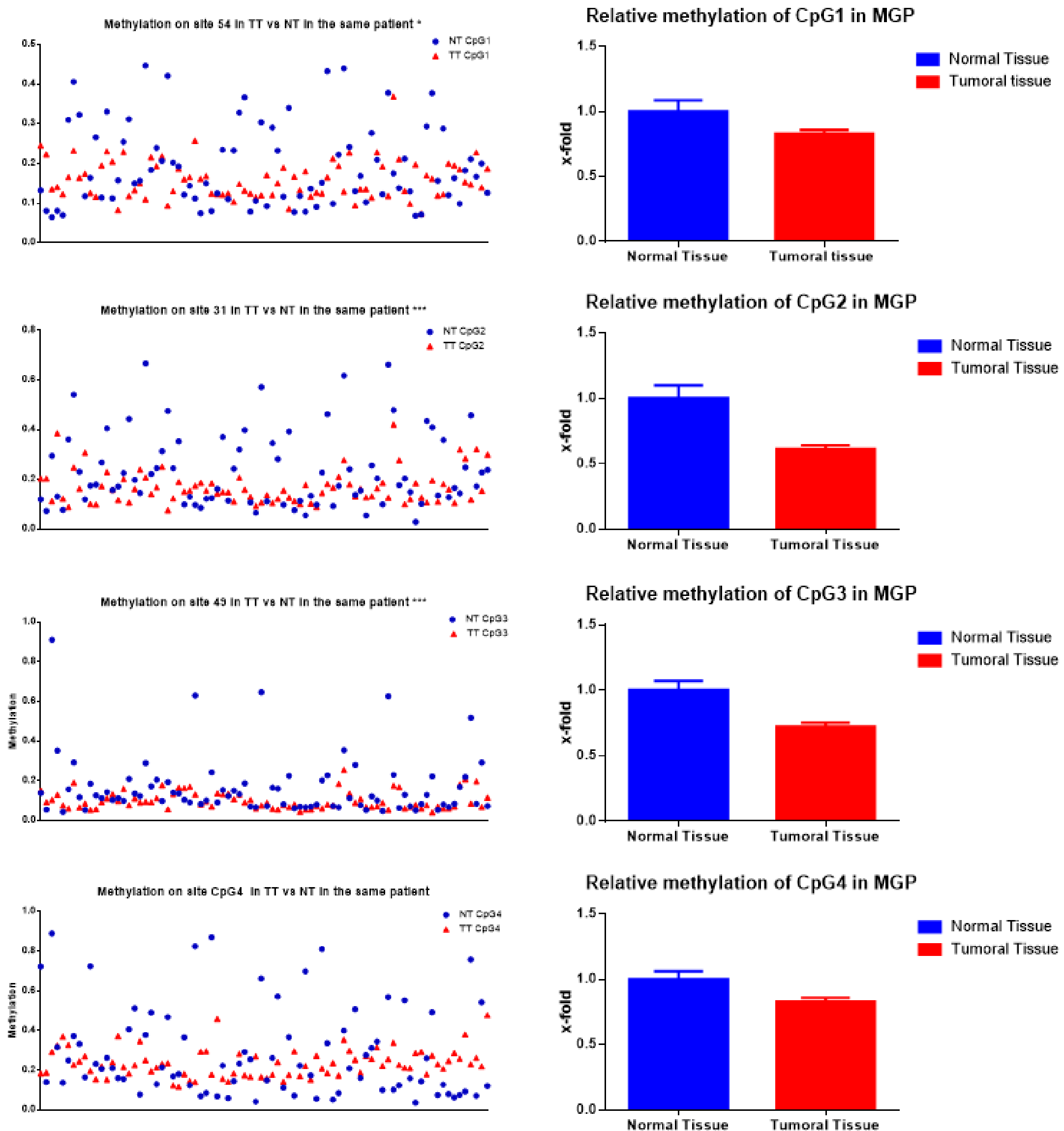


Figure 26 | Comparison of the methylation status, in four probes localizated in MGP gene, in matched NT and TT samples. CpG1, CpG2 and CpG3 are statistically more methylated in normal samples, occuring na hypomethylation on tumoral tissue.

respectively, and CpG1 with p(v)=0.014 (Figure 26).

An hypomethylation in cancer is not a novelty, in fact, was the first epigenetic abnormality recognized in human tumors.[18] The hypomethylation in *MGP* gene suggest that some regulatory factors are silenced in the absence of tumors.

If we go further, and compare *MGP* expression and methylation status (Figure 27) we can observe that it exists in some CpGs a negative correlation between them, meaning that a hypomethylation lead to more expression of *MGP*. Finding the

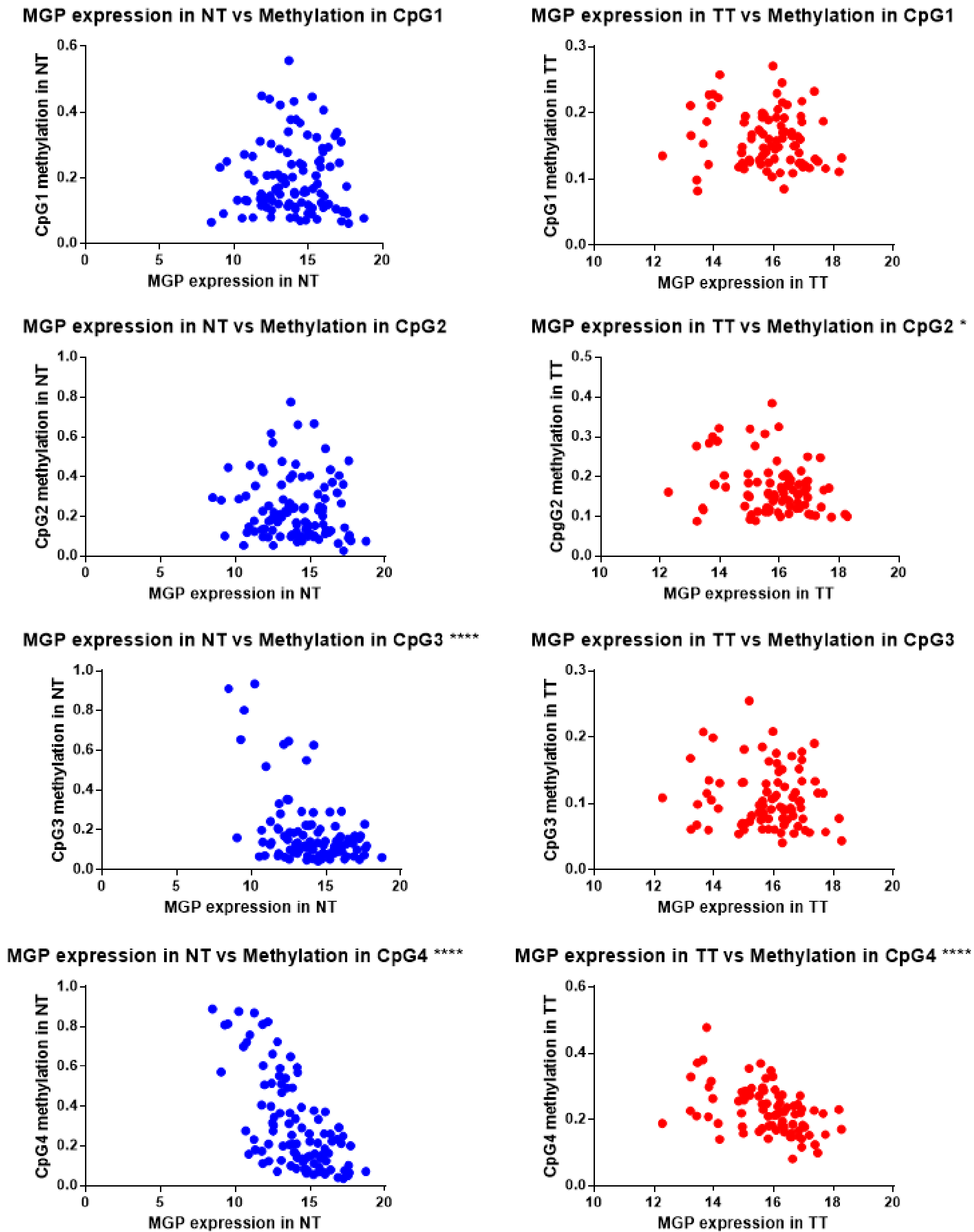


Figure 27 | Correlation between *MGP* expression and methylation in normal and tumoral tissue.

regulatory factors that bind in the sequence that corresponds to these CpGs, would be very important. It is important to refer that only CpG4 (the one that showed no statistical difference in methylation status between NT and TT) have consistent negative correlation in normal and tumoral tissue, so next steps should be cautious.

mir155 is downregulated in tumoral tissue, comparing to matched normal tissue

Comparing the mir155 expression in matched normal and tumoral tissue it was possible to determine that it is more expressed in normal tissue ($p(v)<0.0001$). In fact the expression of mir155 in TT is 0.85 (±0.099s.d) folds the expression in NT (Figure 28).

Mattiske (2012) in the first review of the role of mir155 in BC progression revealed that high levels of mir155 are associated with "clinicopathologic markers, tumor subtype, and poor survival rates", with all previous work showing a higher expression of mir155 in TT rather NT.[100]

Tiago *et al* found that an overexpression of mir155 leads to a decrease around 30% of MGP protein expression (in a breast cell line (MCF7)).[98] Despite variations in mRNA expression did not have always proved to lead to a respective variation in protein expression, if we assume that mir155 is decreasing MGP protein by reducing the quantity of functional *MGP* mRNA, our results appear to be consistent with the mir155 inhibitor role.

One of the Hallmarks of Cancer is the ability of cancer cells to induce angiogenesis. Blood is where oxygen and nutrients needed to feed cells are; vessel formation occurs during organogenesis and then is carefully regulated, this way a growing cancer is further from the feeding source; inducing angiogenesis is the best hypothesis for tumor survival.[13] MGP was already related with angiogenesis.[101,102] In Kuzontkoski (2010) work of xenografts in immunosuppressed mice, better vascularization was found in MGP rich xenografts, in comparison to control tumors; MGP inhibition resulted in smaller and less vascularized xenografts.[101] This information reveals a huge role of MGP in angiogenesis, and therefore on tumor progression. Regarding mir155 role in breast cancer, it appears to have a protective effect.

The contradictory information of this findings open space to more discuss and research.

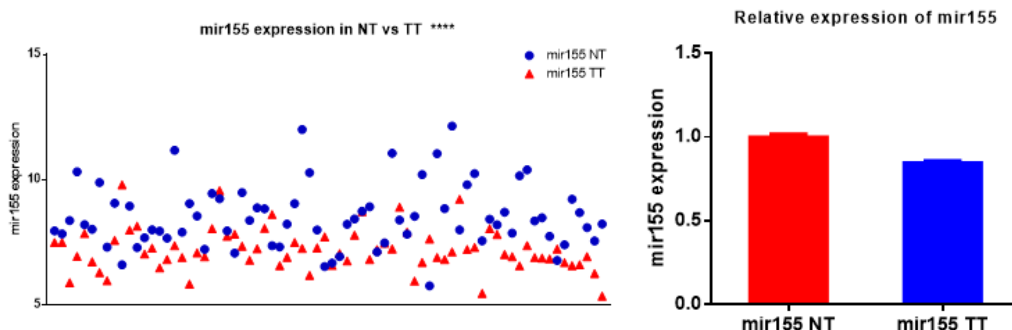


Figure 28 | mir155 expression in NT vs TT. mir155 is more expressed in normal tissue ($p(v)<0.0001$).

Final Remarks and Future Perspectives

Final Remarks and Future Perspectives

MGP expression seems to be different according to the type of tumor, but in several cases is overexpressed in tumoral tissue (as seen in ovarian cancer, renal cell carcinoma and astrocytic gliomas); two independent works in our lab show a higher expression of MGP in tumoral tissue in breast cancer as well as in colorectal cancer (although in this case a previous report from other group showed a downregulation).

According to our data the expression of MGP in tumoral and normal tissue is associated with epigenetic regulation and posttranscriptional regulation. Appearing to exist a negative correlation between:

- Hypermethylation (of four CpGs islands in MGP gene) and MGP expression. Occurs an hypomethylation in cancer tissue.
- mir155 expression and MGP expression – suggesting that mir155 inhibits the expression of MGP; mir155 is higher expressed in normal tissue.

Adding to this results the fact that MGP inhibits calcification, and that is involved in angiogenesis, an hallmark of cancer, it is a good starting point to new research, in a way to develop new treatments or tools to make MGP an easy biomarker.

Our work present new and exciting insights into the expression of MGP variants, especially because there is almost none information about it in the literature. Till now, results presented as MGP are, presumably, the result of total MGP (both isoforms). Data obtained could be the starting line to evaluate MGP isoforms in breast cancer progression and to assess the feasibility of using MGP as a prognostic factor or a biomarker.

We confirmed that some cell lines express both variants simultaneous as well that the expression of E4 isoform occurred in all samples, but E5 expression was more discreet. An overexpression of each variant in CFI-TERT cell line revealed a lower migration ratio, comparing to WT cells. Although cell lines are a good research model, new studies in fresh samples and/or 3D cultures could give a better insight of the real variation of expression of both isoforms as well as the effect of an overexpression of each. Invasiveness should also be tested. Knowing that:

- MGP is overexpressed in tumoral tissue and assuming that the main reason is for facilitate angiogenesis, by inhibiting vessels calcification and that;

- E5 variant have three putative additional sites of gamma carboxylation, and that this feature could increase MGP binding capacity, resulting in a more powerful inhibition of calcification;

We could expect that tumoral tissue present more expression of E5 variant than normal tissue. Notwithstanding, this theory could not be tested due to a lack of fresh samples and databases didn't distinguish E4 and E5 expression in MGP.

Bibliography

1. Wiener N. Cybernetics or control and communication in the animal and the machine. *Psychol Bull*. 1949;46(3):236-237. doi:10.1037/h0051026.

2. Reading A. Meaningful Information: The Bridge Between Biology, Brain, and Behavior. In: Vol 1. Springer-Verlag New York; 2011:158. doi:10.1007/978-1-4614-0158-2.

3. Alberts B, Bray D, Hopkin K, et al. *Essential Cell Biology*. Third. Garland Science; 2010. doi:citeulike-article-id:4505949.

4. Korf B. Human Genetics and Genomics. In: Vol 53. Third. Blackwell; 2007:2-18. doi:10.1017/CBO9781107415324.004.

5. Weinberg RA. *The Biology of Cancer*. Vol 44. 2nd ed. Garland Science; 2014. doi:10.1016/j.mpmed.2015.10.005.

6. Nature Education. Ribosomes, Transcription, and Translation. *Nat Cell Biol*. 2014;6(6):483-486. http://www.nature.com/scitable/topicpage/ribosomes-transcription-and-translation-14120660. Accessed November 15, 2015.

7. Strachan T, Read A. *Human Molecular Genetics*. 4th ed.; 2014. http://hmg.oxfordjournals.org/content/10/7/747/F1.expansion.

8. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell*. 2013;153(1):17-37. doi:10.1016/j.cell.2013.03.002.

9. Neve RM, Chin K, Fridlyand J, et al. A collection of breast cancer cell lines for the study of functionally. *Cancer Cell*. 2009;10(6):515-527. doi:10.1016/j.ccr.2006.10.008.A.

10. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013.

11. Ashworth A, Lord CJ, Reis-Filho JS. Genetic interactions in cancer progression and treatment. *Cell*. 2011;145(1):30-38. doi:10.1016/j.cell.2011.03.020.

12. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med*. 2004;10(8):789-799. doi:10.1038/nm1087.

13. Hanahan D, Weinberg RA, Aplin A., et al. The Hallmarks of Cancer. *Cell*. 2000;100(1):57-70. doi:10.1016/S0092-8674(00)81683-9.

14. You JS, Jones PA. Cancer Genetics and Epigenetics: Two Sides of the Same Coin? doi:10.1016/j.ccr.2012.06.008.

15. Dupont C, Armant DR, Brenner CA. Epigenetics: definition, mechanisms and clinical perspective. *Semin Reprod Med*. 2009;27(5):351-357. doi:10.1055/s-0029-1237423.

16. Sadikovic B, Al-Romaih K, Squire JA, Zielenska M. Cause and consequences of genetic and epigenetic alterations in human cancer. *Curr Genomics*. 2008;9(6):394-408. doi:10.2174/138920208785699580.

17. Kulis M, Esteller M. *DNA Methylation and Cancer*. Vol 70.; 2010:27-56. doi:10.1016/B978-0-12-380866-0.60002-2.

18. Ehrlich M. DNA hypomethylation in cancer cells. *Epigenomics*. 2009;1(2):239-

259. doi:10.2217/epi.09.33.

19. WHO - World Health Organization. *WHO Report on the Global Tobacco Epidemic*. Vol 5.; 2013. doi:10.1002/aehe.3640230702.

20. Siegel R, Miller K, Jemal A. Cancer statistics , 2015 . *CA Cancer J Clin*. 2015;65(1):29. doi:10.3322/caac.21254.

21. American Cancer Society. *Cancer Facts & Figures 2016*.; 2016. doi:cancer.

22. Cancer G, Profiles C, Profile E, et al. Explanatory Notes Background Cancer mortality Cancer incidence Risk factors. WHO. http://www.who.int/cancer/country-profiles/explanatory_notes.pdf?ua=1. Published 2012. Accessed January 15, 2016.

23. Cancer Research UK. Tobacco statistics. http://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/tobacco#collapseZero. Accessed January 15, 2016.

24. Jemal A, Chu KC, Tarone RE. Recent trends in lung cancer mortality in the United States. *J NatlCancer Inst*. 2001;93(4):277-283. http://jnci.oxfordjournals.org/content/93/4/277.full.pdf+html.

25. WHO - World Health Organization. *Country Profile - Portugal*.; 2014. http://www.who.int/cancer/country-profiles/prt_en.pdf?ua=1.

26. Liga Portuguesa Contra o Cancro. Cancro da Mama. https://www.ligacontracancro.pt/cancro-da-mama/. Published 2015. Accessed September 19, 2016.

27. Altobelli E, Lattanzi A. Breast cancer in European Union: an update of screening programmes as of March 2014 (review). *Int J Oncol*. 2014;45(5):1785-1792. doi:10.3892/ijo.2014.2632.

28. Chabner Bruce, Lynch Jr. T, Longo Dan. Harrison's Manual of Oncology. In: Chabner BA, Lynch TJ, Longo DL, eds. *Harrison's Manual of Oncology*. First. The McGraw-Hill Companies; 2008:511-515.

29. Armstrong K. EA. WB, Armstrong K, Eisen A, Weber B. Assessing the risk of breast cancer. *N Engl J Med*. 2000;342(8):564-571. http://www.jacobimed.org/public/Ambulatory_files/mlove/CurriculumWomenandGeri/Breast Care/Breast Care articles/assessing the risk of breast ca.1.pdf.

30. The Health & Environment Alliance. Factors influencing the risk of breast cancer – established and emerging. *Factors Influ risk breast cancer - Establ Emerg*. 2008. http://www.chemtrust.org.uk/wp-content/uploads/factorsinfluencingriskofbreastcancer.pdf.

31. Hunter DJ, Spiegelman D, Adami HO, et al. Non-dietary factors as risk factors for breast cancer, and as effect modifiers of the association of fat intake and risk of breast cancer. *Cancer Causes Control*. 1997;8(1):49-56. doi:10.1023/A:1018431104786.

32. Beral V, Bull D, Doll R, Peto R, Reeves G. Breast cancer and breastfeeding: Collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women

without the disease. *Lancet*. 2002;360(9328):187-195. doi:10.1016/S0140-6736(02)09454-0.

33. Schack-nielsen L, Larnkjær A, Michaelsen KF. Long Term Effects of Breastfeeding on the Infant and Mother. :16-23. http://link.springer.com/chapter/10.1007/1-4020-3535-7_3.

34. Million Women Study Collaborators. Breast cancer and hormone replacement therapy in the Million Women Study. *Lancet*. 2003;362:419-427. http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(03)14065-2.pdf.

35. Greiser CM, Greiser EM, Dören M. Menopausal hormone therapy and risk of breast cancer: A meta-analysis of epidemiological studies and randomized controlled trials. *Hum Reprod Update*. 2005;11(6):561-573. doi:10.1093/humupd/dmi031.

36. Trial RC. Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women. 2009;288(3):321-333. http://jama.jamanetwork.com/article.aspx?articleid=195120.

37. Collaboratative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet*. 1996;347(9017):1713-1727. doi:10.1016/S0140-6736(96)90806-5.

38. Keating NL, Cleary PD, Rossi AS, Zaslavsky AM, Ayanian JZ. Hormone Replacement Therapy. Annals of Internal Medicine. https://medlineplus.gov/hormonereplacementtherapy.html. Published 1999.

39. Staten Island University Hospital. The Hereditary Cancer Genetics Program - Sporadic vs. Hereditary Cancer. http://www.siuh.edu/Our-Services/Clinical-Services/Cancer-Services/The-Hereditary-Cancer-Genetics-Program/Sporadic-vs-Hereditary-Cancer.aspx. Published 2014. Accessed March 14, 2016.

40. National Cancer Institute. BRCA1 and BRCA2: Cancer risk and genetic testing. http://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet#q1. Accessed March 13, 2016.

41. American Joint Committee on Cancer. *AJCC Cancer Staging Handbook*. 6th ed. (Springer, ed.). Greene, Frederick L. Page, David L. Fleming, Irvin D. Fritz, April G. Balch, Charles M. Haller, Daniel G. Morrow, Monica; 2002. doi:10.1007/978-0-387-88443-1.

42. National Breast Cancer Foundation. Breast Anatomy. 2015. http://www.nationalbreastcancer.org/breast-anatomy. Accessed April 25, 2016.

43. American Society of Clinical Oncology. Breast Cancer - Stages. http://www.cancer.net/cancer-types/breast-cancer/stages. Accessed March 11, 2016.

44. American Society of Clinical Oncology. Breast Cancer - Inflammatory: Overview. http://www.cancer.net/cancer-types/breast-cancer-inflammatory/overview. Published 2015. Accessed February 12, 2016.

45. American Cancer Society. Cancer Facts & Figures 2015. *Cancer Facts Fig 2015*.

2015:1-9. doi:10.1097/01.NNR.0000289503.22414.79.

46. Haydaroglu A, Ozyigit G. Principles and practice of modern radiotherapy techniques in breast cancer. In: *Principles and Practice of Modern Radiotherapy Techniques in Breast Cancer*. ; 2013:13-19. doi:10.1007/978-1-4614-5116-7.

47. Malhotra GK, Zhao X, Band H, Band V. Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther*. 2010;10(10):955-960. doi:10.4161/cbt.10.10.13879.

48. Institute NC. Breast Cancer Treatment - Health Professional Version. doi:10.1017/CBO9781107415324.004.

49. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol*. 2014;5(3):412-424. doi:10.5306/wjco.v5.i3.412.

50. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-752. doi:10.1038/35021093.

51. Polyak K. Review series introduction Heterogeneity in breast cancer. *JClinInvest*. 2011;121(10):2011-2013. doi:10.1172/JCI60534.3786.

52. Perez E a. Breast cancer management: opportunities and barriers to an individualized approach. *Oncologist*. 2011;16 Suppl 1(suppl 1):20-22. doi:10.1634/theoncologist.2011-S1-20.

53. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869-10874. doi:10.1073/pnas.191367098.

54. Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*. 2015;8:54. doi:10.1186/s12920-015-0129-6.

55. Eroles P, Bosch A, Alejandro P??rez-Fidalgo J, Lluch A. Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. *Cancer Treat Rev*. 2012;38(6):698-707. doi:10.1016/j.ctrv.2011.11.005.

56. Cheang MCU, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst*. 2009;101(10):736-750. doi:10.1093/jnci/djp082.

57. Barnes CJ, Kumar R. Biology of the epidermal growth factor receptor family. *Cancer Treat Res*. 2004;119:1-13. http://www.ncbi.nlm.nih.gov/pubmed/15164870.

58. Bazley LA, Gullick WJ. The epidermal growth factor receptor family. *Endocr Relat Cancer*. 2005;12(SUPPL. 1):17-28. doi:10.1677/erc.1.01032.

59. Gutierrez C, Schiff R. Biology, Detection, and Clinical Implications of HER". *Arch Pathol*. 2011;135:12-14. doi:10.1043/2010-0454-rar.1.

60. Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*. 2007;26(45):6469-6487. doi:10.1038/sj.onc.1210477.

61. Badve S, Dabbs DJ, Schnitt SJ, et al. Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod Pathol*. 2011;24(2):157-167. doi:10.1038/modpathol.2010.200.

62. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100(14):8418-8423. doi:10.1073/pnas.0932692100.

63. Lehmann BDB, Bauer J a J, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*. 2011;121(7):2750-2767. doi:10.1172/JCI45014DS1.

64. Prat A, Parker JS, Karginova O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res*. 2010;12(5):R68. doi:10.1186/bcr2635.

65. Russnes HG, Kristian H, Vollan M, Lingjærde OC. Europe PMC Funders Group Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. 2014;2(38). doi:10.1126/scitranslmed.3000611.Genomic.

66. Cancela L, Hsieh CL, Francke U, Price PA. Molecular structure, chromosome assignment, and promoter organization of the human matrix Gla protein gene. *J Biol Chem*. 1990;265(25):15040-15048.

67. Price PA, Williamson MK. Primary structure of bovine matrix Gla protein, a new vitamin K-dependent bone protein. *J Biol Chem*. 1985;260(28):14971-14975.

68. Yoshimura K, Takeuchi K, Nagasaki K, et al. Prognostic value of matrix Gla protein in breast cancer. *Mol Med Rep*. 2009;2(4):549-553. doi:10.3892/mmr-00000135.

69. Fraser J, Price P a. Lung, heart, and kidney express high levels of mRNA for the vitamin K-dependent matrix Gla protein. *J Biol Chem*. 1988;263(23):11033-11036.

70. Munroe PB, Olgunturk RO, Fryns JP, et al. Mutations in the gene encoding the human matrix Gla protein cause Keutel syndrome. *NatGenet*. 1999;21(1061-4036 (Print)):142-144. doi:10.1038/5102.

71. Caiado H. Functional analysis of a transcriptional variant of Matrix Gla Protein (MGP). 2014.

72. Cancela ML, Laizé V, Conceição N. Matrix Gla protein and osteocalcin: From gene duplication to neofunctionalization. *Arch Biochem Biophys*. 2014;561:56-63. doi:10.1016/j.abb.2014.07.020.

73. Luo GB, Ducy P, McKee MD, et al. Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. *Nature*. 1997;386(6620):78-81. doi:10.1038/386078a0.

74. Hallgren KW, Zhang D, Kinter M, Willard B, Berkner KL. Methylation of γ-carboxylated Glu (Gla) allows detection by liquid chromatography-mass spectrometry and the identification of Gla residues in the γ-glutamyl carboxylase. *J Proteome Res*. 2013;12(6):2365-2374. doi:10.1021/pr3003722.

75. Roy B, Haupt LM, Griffiths LR. Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity. *Curr Genomics*. 2013;14(3):182-194. doi:10.2174/1389202911314030004.

76. Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*. 2015;(March):1-33. doi:10.1146/annurev-biochem-060614-034316.

77. Diaxonhit. Alternative RNA splicing.

78. Conceição N, Laizé V, Simões B, Pombinho AR, Cancela ML. Retinoic acid is a negative regulator of matrix Gla protein gene expression in teleost fish Sparus aurata. *Biochim Biophys Acta*. 2008;1779(1):28-39. doi:10.1016/j.bbagrm.2007.11.003.

79. Zhou H, Hammonds RG, Findlay DM, Fuller PJ, Martin TJ, Ng KW. Retinoic acid modulation of mRNA levels in malignant, nontransformed, and immortalized osteoblasts. *J Bone Miner Res*. 1991;6(7):767-777. doi:10.1002/jbmr.5650060715.

80. Kirfel J, Kelter M, Cancela LM, Price PA, Schüle R. Identification of a novel negative retinoic acid responsive element in the promoter of the human matrix Gla protein gene. *Proc Natl Acad Sci U S A*. 1997;94(6):2227-2232. http://www.ncbi.nlm.nih.gov/pubmed/9122176. Accessed September 8, 2016.

81. Correia E, Conceição N, Cancela ML, Belo JA. Matrix Gla Protein expression pattern in the early avian embryo. *Int J Dev Biol*. 2016;60(1-3):71-76. doi:10.1387/ijdb.150365jb.

82. Mertsch S, Schurgers LJ, Weber K, Paulus W, Senner V. Matrix gla protein (MGP): an overexpressed and migration-promoting mesenchymal component in glioblastoma. *BMC Cancer*. 2009;9:302. doi:10.1186/1471-2407-9-302.

83. van den Boom J, Wolter M, Kuick R, et al. Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. *Am J Pathol*. 2003;163(3):1033-1043. doi:10.1016/S0002-9440(10)63463-3.

84. Hough CD, Cho KR, Zonderman a B, Schwartz DR, Morin PJ. Coordinately up-regulated genes in ovarian cancer. *Cancer Res*. 2001;61(10):3869-3876. http://www.ncbi.nlm.nih.gov/pubmed/11358798. Accessed September 8, 2016.

85. Levedakou EN, Strohmeyer TG, Effert PJ, Liu ET. Expression of the matrix Gla protein in urogenital malignancies. *Int J cancer*. 1992;52(4):534-537. http://www.ncbi.nlm.nih.gov/pubmed/1399132. Accessed September 8, 2016.

86. Fan C, Sheu D, Fan H, Hsu K, Allen Chang C, Chan E. Down-regulation of matrix Gla protein messenger RNA in human colorectal adenocarcinomas. *Cancer Lett*. 2001;165(1):63-69. http://www.ncbi.nlm.nih.gov/pubmed/11248420.

87. The Human Protein Atlas. Expression of MGP in cancer - Summary. http://www.proteinatlas.org/ENSG00000111341-MGP/cancer. Accessed May 23, 2016.

88. Chen L, O'Bryan JP, Smith HS, Liu E. Overexpression of matrix Gla protein mRNA in malignant human breast cells: isolation by differential cDNA hybridization. *Oncogene*. 1990;5(9):1391-1395. http://www.ncbi.nlm.nih.gov/pubmed/2216462. Accessed September 8, 2016.

89. Festing S, Wilkinson R. The ethics of animal research Talking Point on the use of animals in scientific research.

90. Masters JR. Human cancer cell lines: fact and fantasy. *Nat Rev Mol Cell Biol*. 2000;1(3):233-236. doi:10.1038/35043102.

91. ATCC. Breast cancer and normal cell lines. ATCC. doi:CB-0513-.

92. Thermo Scientific. Product information:GeneJET RNA Purification Kit, #K0731, #K0732. 2014:1-19. https://tools.thermofisher.com/content/sfs/manuals/MAN0012664_GeneJET_RNA_Purification_UG.pdf. Accessed September 12, 2016.

93. Thermo Fisher. Cell Lysis Solutions. https://www.thermofisher.com/pt/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/cell-lysis-solutions.html. Published 2015. Accessed July 12, 2016.

94. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*. 1987;162(1):156-159. doi:10.1016/0003-2697(87)90021-2.

95. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc*. 2008;3(6):1101-1108. doi:10.1038/nprot.2008.73.

96. Huang X, Stern DF, Zhao H. Transcriptional Profiles from Paired Normal Samples Offer Complementary Information on Cancer Patient Survival – Evidence from TCGA Pan-Cancer Data. *Sci Rep*. 2016;6(February):20567. doi:10.1038/srep20567.

97. Clare O'Connor. Essentials of Cell Biology. In: Clare O'Connor, ed. Cambridge: NPG Education; 2010. http://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010.

98. Tiago DM, Conceição N, Caiado H, Laizé V, Cancela ML. Matrix Gla protein repression by MIR-155 promotes oncogenic signals in breast cancer MCF-7 cells. *FEBS Lett*. 2016;590(8):1234-1241. doi:10.1002/1873-3468.12155.

99. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*. 2004;116:281-297. http://www.cell.com/cell/pdf/S0092-8674(04)00045-5.pdf. Accessed June 10, 2016.

100. Mattiske S, Suetani RJ, Neilsen PM, Callen DF. The oncogenic role of miR-155 in breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2012;21(8):1236-1243. doi:10.1158/1055-9965.EPI-12-0173.

101. Kuzontkoski PM, Mulligan-Kehoe MJ, Harris BT, Israel M a. Inhibitor of DNA binding-4 promotes angiogenesis and growth of glioblastoma multiforme by elevating matrix GLA levels. *Oncogene*. 2010;29(26):3793-3802. doi:10.1038/onc.2010.147.

102. Sharma B, Albig AR. Matrix Gla Protein Reinforces Angiogenic Resolution. *Elsevier*. 2014:24-33. doi:10.1016/j.mvr.2012.10.005.

103. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011;25(10):1010-1022. doi:10.1101/gad.2037511.

104. Scholzen T, Gerdes J. The Ki-67 protein: From the known and the unknown. *J Cell Physiol*. 2000;182(3):311-322. doi:10.1002/(SICI)1097-4652(200003)182:3<311::AID-JCP1>3.0.CO;2-9.

Annexes

Appendix table 1 | TMN stage grouping for breast cancer, as in the 6th Edition of AJCC Cancer Staging Manual (2002)[28,41]

Stage	TNM	
Stage 0	Tis NO MO	Cancer cells are confined in ducts and lobules. Not spread to surrounding tissue. Noninvasive cancer (Tis, NO, MO)
Stage I	T1 NO MO (including T1mic)	
Stage II A	T0 N1 MO T1 N1 MO (including T1mic) T2 NO MO	
Stage II B	T2 N1 MO T3 NO MO	
Stage III A	TO N2 MO T1 N2 MO (including T1mic) T2 N2 MO T3 N1 MO T3 N2 MO	
Stage III B	T4 Any N MO	
Stage III C	Any T N3 MO	
Stage IV	Any T Any N M1	

Appendix table 2 | | Histopathologic classification of Breast Cancers - as in the 6th Edition of AJCC Cancer Staging Manual (2002) [41]

In situ Carcinomas	Invasive Carcinomas
NOS	NOS
Intraductal	Ductal
Paget's disease and intraductal	Inflammatory
	Medullary, NOS
	Medullary with lymphoid stroma
	Mucinous

	Papillary (predominantly micropapillary pattern)
	Tubular
	Lobular
	Paget’s disease and infiltrating
	Undifferentiated
	Squamous cell
	Adenoid cystic
	Secretory
	cribriform

Appendix table 3 | Regional lymph nodes anatomy

1. **Axillary (ipsilateral)**: interpectoral (Rotter’s) nodes and lymph nodes along axillary vein and its tributaries that may (but not required) divided into:
    a. Level I (low axilla) – LN lateral to the lateral border or pectoralis minor muscle.
    b. Level II (mid-axilla) – LN between the medial and lateral borders of the pectoralis minor muscle and the interpectoral (Rotter’s) lymph nodes.
    c. Level III (apical axilla) – LN medial to the medial margin of the pectoralis minor muscle, including those designated as apical.
2. **Internal mammary (ipsilateral)**: LN in the intercostal spaces along the edge of the sternum in the endothoracic fascia.
3. **Supraclavicular**: LN in the supraclavicular fossa, a triangle defined by the omohyod muscle and tendon, the internal jugular vein, and the clavicle and subclavicle vein. Adjacent LN outside of this triangle are considered to be lower cercical nodes (M!)

Appendix table 4 | Primers used in qRT-PCR.

Primer	Sequence
hGADPH Forward	5'-TCAACGGATTTGGTCGTATTGGGCG-3'
hGADPH Reverse	5'-CTCGCTCCTGGAAGATGGTGATGGG-3'
hMGP Forward 1 (MGP F1)	5'- TGCTGCTACACAAGACCCTGAGACTGA-3'
hMGP Reverse 4 (MGP Rev4)	5'- CCGAAGTTTTCTTCTTTCTGCCACTCTCC-3'
hMGP Reverse 5 (MGP Rev5)	5'-GTAGCGTTCGCAAAGTCTGTAGTCATCAC-3'
Set used in MGP: E4 variant – F1 and Rev5 E5 variant – F1 and Rev4	

Appendix table 5 | Legend of characteristics analyzed in the study population

Legend of the characteristics analyzed in the study population

Population: 136 patients with paired samples of tumoral tissue and adjacent tissue, all patients are female.

(A) 1 Column: **Tumoral type – histology**
- **1** – Breast Invasive ductal carcinoma (DC)
- **2** – Breast invasive lobular carcinoma (LC)
- **3** – Others (Mixed [LD+LC], Metaplastic cancer and Paget´s disease of the nipple)

(B) 2 Column: **Race**
- **1** – White
- **2 –** Black or African American

(C) 3 Column: **Age at diagnosis**

(D) 4 Column: Tumor Stage
- **1 –** T1
- **2 –** T2
- **3 –** T3

(E) 5 Column: Lymph nodes stage
- 0 – N0
- 1 – N1
- 2 – N2
- 3 – N3

(F) 6 Column: **ER (Estrogen Receptor) status**
- 1 – Positive
- 2 – Negative

(G) 7 Column: **PR (Progesterone Receptor) status**
- 1 – Positive
- 2 – Negative

(H) 8 Column: **HER status**
- 1 – Positive
- 2 – Negative

(I) 9 Column: **AJCC staging**

1 – Stage I, IA, IB
2 – Stage II, IIA, IIB
3 – Stage III, IIIA IIB, IIIC
4 – Stage IV

(J) **10 Column: PAM 50 staging in adjacent tissue**
1 – Normal
2 – Luminal A
3 – Luminal B
4 – Basal-Like
5 – Her+

(K) **11 Column: PAM 50 staging in tumoral tissue**
1 – Normal
2 – Luminal A
3 – Luminal B
4 – Basal-Like
5 – Her+

(L) 12 Column: **Undergo radiation therapy:**
1 – Positive
2 – Negative

(M) 13 Column**: Menopause status**
1 – Peri (6-12 months since last menstrual period (LMP))
2 – Pre (<6 months since LMP and no prior bilateral ovariectomy and not on estrogen replacement)
3 – Post (prior bilateral ovariectomy OR >12 months since LMP with no prior hysterectomy)

(N) 14 Column: **Anatomic subdivision of breast neoplasia**
1 – Right breast
2 – Left breast

(O) 15 Column: **Neoadjuvant history**
1 – Yes
2 – No

(P) 16 Column: Post operative rx and tx
1 – Yes
2 – No

(Q) 17 Column: Overall Survival in months

(R) 18 Column: MGP expression in adjacent tumor tissue

(S) 19 Column: MGP expression in tumoral tissue

(T) 20 Column: Mir155 expression in adjacent tumor tissue

(U) 21 Column: Mir155 expression in tumoral tissue

Methylation of cg in MGP

(V) 22 Column: Methylation of cg cg05360958 adjacent tumor tissue

(W)23 Column: Methylation of cg cg05360958

(X) 24 Column: Methylation of cg00431549 adjacent tumor tissue

(Y) 25 Column: Methylation of cg00431549 in tumoral tissue

(Z) 26 Column: Methylation of cg22221831 adjacent tumor tissue

(AA) 27 Column: Methylation of cg22221831 in tumoral tissue

(AB) 28 Column: Methylation of cg13302154 adjacent tumor tissue

(AC) 29 Column: Methylation of cg13302154 in tumoral tissue