Helena Isabel Rosa Pinto Caiado

Molecular basis for the epigenetic regulation of MGP in cancer

2022

Helena Isabel Rosa Pinto Caiado

Molecular basis for the epigenetic regulation of MGP in cancer

Doutoramento em Mecanismos de Doença e Medicina Regenerativa

Trabalho efetuado sob a orientação de:

Professora Doutora M. Leonor Cancela

Professora Doutora Natércia Conceição

2022

Molecular basis for the epigenetic regulation of MGP in cancer

Declaração de autoria de trabalho

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

____

Copyright© Helena Isabel Rosa Pinto Caiado

A Universidade do Algarve reserva para si o direito, em conformidade com o disposto no Código do Direito de Autor e dos Direitos Conexos, de arquivar, reproduzir e publicar a obra, independentemente do meio utilizado, bem como de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição para fins meramente educacionais ou de investigação e não comerciais, conquanto seja dado o devido crédito ao autor e editor respetivos.

Para a minha Mãe,

Irmão e Irmã

"None of us can know what we are

capable of until we are tested"

Dr. Elizabeth Blackwell

Agradecimentos/Acknowledgments

É com uma enorme gratidão que começo por agradecer às minhas orientadoras, a professora Leonor e Natércia, por toda a ajuda, compreensão, incentivo e motivação, por me terem acompanhado em todas as minhas conquistas, mas também nas minhas maiores fraquezas, pelos "puxões de orelhas" que bem me fizeram crescer, mas sobretudo por me terem acompanhado em todo o meu percurso tanto a nível profissional como pessoal, por acreditarem sempre nas minhas capacidades e por nunca me deixarem desistir.

Agradeço ainda pela oportunidade de ter pertencido ao Bioskel lab, é e será sempre a minha casa, o local onde aprendi tudo o que hoje sei e que me possibilitou conhecer pessoas que irei guardar para a vida.

Agradeço ao Vincent e ao Paulo Gavaia pela ajuda que sempre me foi dada.

Agradeço também aos meus colegas Márcio, Gil, João, Daniela, Sunil, Débora e Tatiana pela entreajuda e companheirismo em ultrapassar as adversidades do dia a dia no laboratório, pelas conversas e pelos conselhos que me foram dados.

Deixo ainda o meu especial ao João cardeira, à Joana, à Cátia, à Vânia, à Íris, ao Nacho, ao Mike, à Cindy, à Andreia, ao Marco, à Kátia e ao Alessio, por terem estado presentes de alguma forma e me terem acompanhado ao longo desta jornada e por me ajudarem sempre em qualquer altura.

Deixo ainda uma nota de agradecimento a todos os colegas do Cebal, em especial à Fátima, à Teresa e à Ivone pelo constante incentivo e apoio e por acreditarem em mim.

Por fim, mas não menos importante, agradeço à minha família pelo amor e força incondicional ao longo da minha formação. Em especial à minha mãe, cuja força e coragem, contra todas as adversidades da vida, me permitiu ser quem sou hoje e que sem ela não teria sido possível chegar onde cheguei. Ao meu pai, por tudo o que me ensinou e que me acompanhará sempre. Ao meu irmão e irmã, os meus segundos pais, por nunca me terem deixado faltar nada, por me terem acompanhado sempre e por acreditarem sempre que serei capaz de ultrapassar todas as adversidades e aos meus tios Chico e Céu por tudo o que me têm transmitido ao longo da minha vida.

A todos, o meu muito obrigado, por de alguma forma terem contribuído para a minha formação enquanto ser humano e me transmitiram importantes valores para a vida.

Funding

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through the doctoral grant PD/BD/128341/2017, to whom I express my thankfulness.
This work was also partially supported by national funds received from the Portuguese Foundation for Science and Technology (FCT) through the project UIDB/04326/2020 (CCMAR).

Abstract

Matrix Gla protein (MGP) is a member of the vitamin K-dependent family of proteins and a known physiological inhibitor of ectopic calcifications. Mutations in its gene cause Keutel syndrome, a rare autosomal recessive disorder characterized by severe soft tissue calcification. Despite its involvement in multiple calcifying pathologies, MGP has been described as playing a potential role in carcinogenesis, thus contributing to trigger interest in this gene towards exploring its potential as a cancer prognostic factor. These findings lead us to the main objective of this work which consisted of the study of the *MGP* gene in different types of tumors, to identify possible epigenetic patterns and transcriptional regulators responsible for controlling *MGP* gene expression in cancer.

In this sense, we analyzed the expression pattern of *MGP* at mRNA and protein levels in a variety of tumors, initially using tissue biopsies from colorectal patients, and then exploring, through bioinformatics analysis, the data available in The Cancer Genome Atlas and The Human Protein Atlas. Results demonstrated a correlation between high levels of *MGP* with advanced stages of cancer progression and poor overall survival outcome, establishing MGP as an independent prognostic factor for the patient's overall survival

We also evaluated the methylation status in four of the six CpG sites, located in the *MGP* promoter region (cg13302154; cg22221831, and cg00431549) and first intron (cg0560958), between healthy and tumoral tissue. Results showed a correlation between *MGP* mRNA expression and the different methylation patterns across all the analyzed tumors, providing evidence for epigenetic regulation of *MGP* transcription.

Moreover, we investigated several putative transcriptional regulators through transient transfections with luciferase reporter assays, demonstrating that YY1, GATA1, and C/EBPα are negative regulators of the *MGP* promoter.

Altogether, our data contribute to provide novel insights on *MGP* regulation, strengthening the hypothesis that *MGP* plays a role during cancer progression/proliferation.

Keywords: Matrix Gla protein, DNA methylation, Epigenetic, Transcriptional regulation, Transcription factors

Resumo

A proteína Gla da Matriz (MGP) pertence à família de proteínas dependentes de vitamina K, sendo um inibidor fisiológico da calcificação. Mutações no seu gene causam a síndrome de Keutel, uma doença autossómica recessiva rara caracterizada por calcificações ectópicas graves. Para alem do seu envolvimento em patologias associadas a mecanismos de calcificação, estudos recentes têm sugerido a intervenção da MGP em processos de carcinogénese, tendo-se afirmado como um potencial fator de prognóstico para o cancro. O principal objetivo deste trabalho consistiu no estudo da expressão do gene *MGP* em diferentes tipos de tumores e na identificação de reguladores transcricionais e possíveis padrões epigenéticos associados ao cancro. Analisamos o padrão de expressão da *MGP* numa variedade de tumores, com recurso a biópsias de tecidos de pacientes com cancro colorretal e bases de dados (*The Human Genome Cancer* e *The Human Protein Atlas)*. Os resultados demonstraram uma correlação entre níveis elevados de expressão do gene *MGP* com estadios mais avançados de progressão do tumor, associados a um pior prognóstico da doença, estabelecendo o gene *MGP* como um fator de prognóstico independente para o tempo de sobrevida dos pacientes.

A analise, entre tecido normal e tumoral, dos níveis de metilação em quatro dos seis locais CpG localizados na região do promotor do gene *MGP* (cg13302154; cg22221831 e cg00431549) e primeiro intrão (cg0560958), demonstrou existir uma correlação entre a expressão do gene *MGP* e os padrões de metilação observados nos tumores analisados.

Investigámos ainda alguns potenciais reguladores transcricionais do gene *MGP* através de ensaios funcionais do seu promotor, demonstrando que YY1, GATA1 e C/EBPα são reguladores negativos da transcrição do gene *MGP*. No seu conjunto, os dados obtidos contribuem para fornecer novas informações sobre a regulação da MGP, fortalecendo a hipótese de que o gene *MGP* poderá estar envolvido na progressão/proliferação de tumores.

Palavras-Chave: Proteína Gla da matriz, Carcinogénese, Metilação do DNA Regulação transcricional, Fatores de transcrição

Table of contents

Agradecimentos/Acknowledgments ... viii
Abstract ... x
Resumo ... xii
Abbreviations and Acronyms ... 1
Preamble ... 3
1. General Introduction ... 6
  1.1 Matrix Gla Protein (MGP) ... 6
  1.2 MGP function ... 7
  1.2.1 Regulation of *MGP* gene ... 7
  1.2.2 *MGP* in Cancer ... 8
 1.3 Cancer Biology ... 9
  1.3.1 Cancer epidemiology ... 11
  1.3.2 Cancer regulation ... 12
 1.4 Concept of Epigenetics ... 13
 1.5. Epigenetic modifications ... 13
  1.5.1 DNA methylation ... 13
  1.5.2 Histone modifications ... 14
  1.5.3 miRNAS ... 14
 1.6 Main objectives ... 14
Chapter 2 ... 17
Assessment of MGP gene expression and ... 17
contribution to cancer prognosis ... 17
 Chapter 2.1 ... 19
 Evaluation of MGP gene expression in Colorectal Cancer ... 19
 2.1.1 Abstract ... 20
 2.1.2 Introduction ... 20
 2.1.3 Materials and Methods ... 22
  2.1.3.1 Patient selection and collection of tissue samples ... 22
  2.1.3.2 RNA Extraction and cDNA synthesis ... 23
  2.1.3.3 Immunohistochemistry ... 23
  2.1.3.4 Statistical analysis ... 24
 2.1.4 Results ... 24
  2.1.4.1 Clinical, demographic and pathological characteristics of patients ... 24
  2.1.4.2 *MGP* gene expression was increased in colorectal cancer patients ... 27
  2.1.4.3 No significant associations were observed between *MGP* gene expression and clinical and pathological features of patients ... 29

2.1.4.4 Transcription factor *RUNX2* is upregulated in patients with colorectal cancer.... 30
2.1.4.5 *MGP* and *RUNX2* are important predictor factors to differentiate the prognosis of colorectal adenocarcinoma amongst patients.. 31
2.1.5 Discussion.. 35
2.1.6 Conclusion.. 38
2.1.7 Supplementary Data 39
Chapter 2.2... 41
Data on the evaluation of *FGF2* gene expression in Colorectal Cancer 41
2.2.1 Abstract.. 42
2.2.3 Value of the Data 43
2.2.4 Data Description 43
2.2.4 Experimental Design, Materials, and Methods 52
2.2.4.1 Clinical, demographic and pathological characteristics of patients 52
2.2.4.2 RT-qPCR.. 52
2.2.4.3 Statistical analysis 53
Chapter 2.3... 55
Assessment of MGP gene expression in cancer and contribution to prognosis 55
2.3.1 Abstract.. 55
2.3.2 Introduction 56
2.3.2 Material and methods 57
2.3.2.1 Sample and data collection 57
2.3.2.2 Gene expression data analysis 57
2.3.2.3 Survival analysis 58
2.3.2.4 Analysis of DNA methylation 58
2.3.2.5 Statistical analysis 59
2.3.3 Results 59
2.3.3.1 *MGP* is differentially expressed between cancer and normal tissue samples....... 59
2.3.3.2 High *MGP* expression is associated with a shorter Overall Survival 65
2.3.3.3 Divergent expression patterns of *MGP* mRNA and MGP protein in cancer and normal tissues.. 69
2.3.3.4 *MGP* CpG sites present different DNA methylation patterns across different cancer types 69
2.3.4 Discussion.. 75
2.3.5 Supplementary data 79
Chapter 3 87
Functional analysis of MGP promoter 87
Chapter 3.1... 89

Transcriptional regulation of the human Matrix GLA Protein promoter: identification of downstream repressors 89

3.1.1 Abstract 89

3.1.2 Introduction 90

3.1.2 Material and Methods 91

3.1.2.1 CpG sites identification in the *MGP* gene 91

3.1.2.2 *In silico* analysis of *MGP* promoter 91

3.1.2.3 Luciferase reporter constructs of *MGP* promoter 91

3.1.2.4 Site-directed mutagenesis 92

3.1.2.5 Transcription factor expressing vectors 93

3.1.2.6 Cell Culture 93

3.1.2.7 Luciferase reporter assays 93

3.1.2.8 TCGA data analysis 94

3.1.2.9 Statistical analysis 94

3.1.3 Results 94

3.1.3.1 *In silico* analysis of transcription binding sites and CpG sites identified in the *MGP* promoter 94

3.1.3.2 Functional analysis of *MGP* promoter activity by the predicted transcription factors 95

3.1.3.2.1 Transcription factor YY1 (Yin Yang 1) represses *MGP* luciferase promoter activity 96

3.1.3.2.2 Transcription factor GATA1 binds to *MGP* promoter repressing its activity ... 98

3.1.3.2.3 C/EBPα has a repressive effect on the *MGP* promoter 98

3.1.3.2.4 Repression in the *MGP* promoter is enhanced by YY1. 99

3.1.3.2.5 RUNX2 enhances *MGP* promoter activity 100

3.1.3.3 Transcription factors YY1, GATA1, and C/EBPα negatively regulate *MGP* expression in cancer 101

3.4 Discussion 103

3.5 Supplementary data 107

Chapter 4 110

Final Conclusions and Future Perspectives 110

4.1 General conclusions and Future perspectives 112

References 116

Abbreviations and Acronyms

AP1: Activating Protein 1

AP2: Activating Protein 2

CRC: Colorectal Cancer

DAB: 3, 3'-diaminobenzidine

DNA: Deoxyribonucleic Acid

FGF2: Basic Fibroblast Growth Factor 2

GAPDH: Glyceraldehyde-3-phosphate

GDC: Genomic Data Commons

GLA: γ-Carboxylated Glutamic Acid

GLU: γ- Carboxyl Glutamic Acid

HR: Hazard Ratio

MGP: Matrix Gla Protein

mRNA: Messenger Ribonucleic Acid

miRNAs: MicroRNAs

OS: Overall Survival

RUNX2: Runt Related Protein 2

RA: Retinoic Acid

RAR: Retinoic Acid Receptor

RNA: Ribonucleic Acid

S.D: Standard Deviation

TCGA: The Cancer Genome Atlas

TF: Transcription Factor

VD: Vitamin D

VDR: Vitamin D Receptor

WHO: World Health Organization

Preamble

This dissertation is divided into four main chapters.

- Chapter one consists of a literature overview of concepts about cancer biology, epigenetic regulation, the different types of epigenetic markers and their role in gene regulation and cancer, followed by the state of the art about Matrix Gla protein and its role in cancer, ending with a brief description of the main objectives of this work.

- Chapter two presents the analysis of *MGP* at mRNA and protein levels in different types of tumors, the correlation between the expression of the *MGP* gene and levels of DNA methylation at specific CpG sites located in the *MGP* gene and its promoter, and a discussion concerning the potential of MGP to be a marker for cancer prognosis. This chapter is divided into three sub-chapters, 2.1, 2.3, and 2.3, each of them organized in the format of a scientific article. Sub-chapter 2.1, focusing on the analysis of the *MGP* gene expression in correlation with specific epigenetic events and through bioinformatic analysis contribute to identify MGP as a novel factor for cancer prognosis. This work is being submitted to the journal "Biochimie".

The sub-chapter 2.2, entails the evaluation of *MGP* expression at mRNA and protein levels and one of its known transcriptional regulators, RUNX2, in normal and tumoral tissue samples biopsies from colorectal patients, as well as the contribution of these genes to cancer prognosis. This sub-chapter was published as a manuscript in the journal "Gene".

Sub-chapter 2.3, comprises the study of FGF2, also known as a transcriptional regulator of the *MGP* gene, in the colorectal patients involved in the study of sub-chapter 2.2 and its contribution to cancer prognosis. This sub-chapter was published as a data manuscript format in the journal "Data in brief"

- Chapter 3 focuses on the functional analysis of one of the identified CpG sites (cg00431549) in the *MGP* promoter, in the presence of several identified transcription factors with the potential to be regulators of *MGP* transcription and explores the correlation of these factors with *MGP* gene expression in different types of tumors. This chapter is being submitted as a manuscript for publication in the journal. "Biochimica et Biophysica Acta – general subjects"

- Chapter 4 includes the main conclusions drawn from the work presented in this thesis, how these findings are relevant for cancer prognosis and some perspectives for future work.

Chapter 1

General introduction

1. General Introduction

1.1 Matrix Gla Protein (MGP)

Matrix Gla Protein (MGP) is a 14kDa protein belonging to the vitamin K-dependent protein family and containing 84 amino acid residues in its mature form. This protein was initially discovered in the demineralized bovine bone matrix (Price, Poser, et al. 1976; Price et al. 1976), being later isolated from cartilage (Hale et al. 1988). Its primary structure comprises a signal peptide, a phosphorylation domain, and a γ-carboxylase recognition site. The human MGP contains five γ-carboxyl glutamic acid (Glu) residues that are converted post-transcriptionally to γ-carboxylated glutamic acid (Gla) residues by γ-glutamyl carboxylase, an ubiquitous enzyme dependent on vitamin K (Figure 1.2) and has been extensively conserved throughout evolution (Laizé et al. 2005).

Since its discovery, MGP has been found in a variety of tissues, such as the heart, lung, and kidney (Fraser and Price, 1988).

```
                          5                      10                       15
 1  Tyr-Leu-Asp-His -Trp-Leu-Gly-Ala-Hyp-Ala -Pro -Tyr-Pro -Asp-Pro -
16  Leu-Gla -Pro -Lys-Arg-Gla -Val-Cys-Gla -Leu-Asn-Pro-Asp-Cys -Asp-
31  Glu-Leu-Ala -Asp-His -Ile  -Gly-Phe-Gln -Glu -Ala -Tyr-Arg -Arg-Phe -
46  Tyr-Gly -Pro -Val
```

Figure 1.1 Primary structure of MGP protein when first isolated from bovine bone. Image adapted from (Price, Poser, and Raman. (1976)

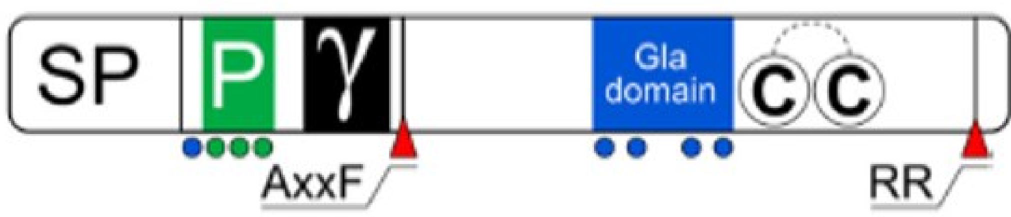


Figure 1.2 Schematic representation of the human MGP protein.
Sp represents the signal peptide, green box and dots represent phosphorylation sites; Blue box and dots represent the Gla residues and domain; black box contains the site for γ glutamyl carboxylase enzyme and red triangles indicate proteolytic cleavage sites. Figure Adapted from Cancela et al. (2014).

1.2 MGP function

Concerning the biological function, MGP has been described as an important inhibitor of calcification in cartilage and vasculature (Shanahan et al. 1993; Yagami et al. 1999). Genetic studies with mice lacking MGP have shown a premature death around two months of age due to excessive and abnormal calcification in the arteries, causing exsanguination by the blood vessel rupture, establishing the role of MGP as a physiological inhibitor of calcification (Luo et al. 1997).

In the course of the MGP studies, it was also shown its involvement in cell differentiation (Boström et al. 2004), as well as in proliferation (Cancela et al. 2014).

Mutations found in the human *MGP* gene were proven to cause the Keutel syndrome, a rare autosomal recessive disease characterized by abnormal cartilage calcification, short stature, multiple peripheral pulmonary stenoses, brachytelephalangia, and inner ear deafness (Munroe et al. 1999).

Several other pathologies have been associated with a deregulation of MGP, such as atherosclerosis, chronic kidney disease, and a wide range of cancers (Gheorghe and Crăciun, 2015; Wu et al. 2017; Tuo and Ye, 2017; Fu et al. 2018; Post et al. 2019; Gong et al. 2019; Huang et al. 2021; Li et al. 2020; Shiomi et al. 2021; Huang et al. 2021; Rong et al. 2022)

1.2.1 Regulation of *MGP* gene

The human *MGP* gene was cloned and localized in chromosome 12 (Cancela et al. 1990) and until a few years ago, it was thought that the genomic structure of the *MGP* gene would give rise to only one single transcript. However, there is some evidence that more than one variant can arise from this gene (E5; accession number: NM_001190839.3) (Figure 1.2).

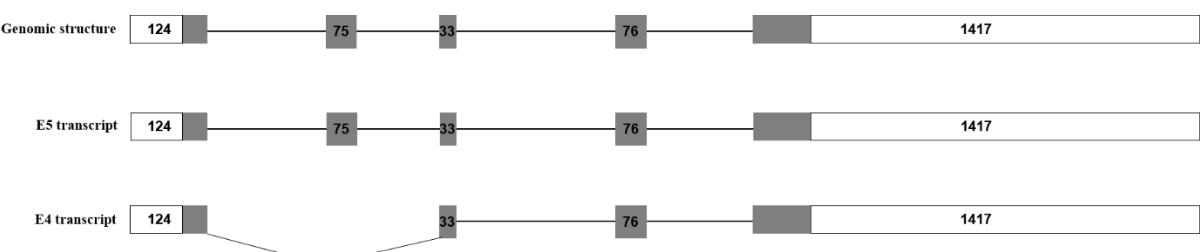


Figure 1.2 Schematic representation of alternative splicing of *MGP* transcripts. Schematic representation of *MGP* splice variants E4 and E5. Solid boxes indicate the coding regions and open boxes indicate untranslated sequences. Lines represent intronic regions of the gene

As shown in figure 1.2, this new transcript contains a new exon of 75bp, however, so far, the function of this new *MGP* transcript remains to be further elucidated (Cancela, et al. 2014).

The raising interest in *MGP* has promoted multiple studies to understand its regulation. In this regard, many putative binding sites for transcriptional regulators have been reported, but only a few were shown to be functional.

In a previous study of human *MGP* gene promoter, it was revealed the existence of several consensus structural motifs important for the putative binding of the typical TATA and CAAT boxes, AP1 and AP2 transcription factors, and metal-responsive elements. Also, it was shown the existence of two polymorphisms, located in the promoter, a region essential for transcription activity. Since the second polymorphism was associated with the AP-1 binding site it was suggested that it would influence AP1 complex binding to the *MGP* promoter, thus altering gene expression (Farzaneh-Far et al. 2001)

In other studies, specific binding sites for retinoic acid (RA) and vitamin D (VD) receptors were identified in the *MGP* promoter, however, the observed effects of these regulatory elements were far from being consensual. Regarding RAR, it was observed that depending on the recipient cell model used, the RAR either repressed or activated *MGP* gene expression (Kirfel et al. 1997). In the case of VDR, it was shown to up-regulate *MGP* expression in some cell types, while in others, it did not promote any significant changes unless it was administered in combination with RA (Kirfel et al. 1997).

In mice, the basic fibroblast growth factor (FGF2) was proven to regulate the *Mgp* promoter activity in proliferative chondrocytes (Stheneur et al. 2003).

Another transcription factor that was shown to regulate *MGP* gene expression was Runt-related transcription factor 2 (RUNX2). It was evidenced that co-transfection of ATF4, RUNX2, and SATB2 enhanced the *MGP* human promoter activity in ATDC5 cells (Roberto et al. 2018).

In mice, parathyroid hormone regulates *Mgp* expression through the transcription factors Sp and Runx2 (Suttamanatwong et al., 2009). In addition, it has been demonstrated that *MGP* is regulated by Runx2 in *Xenopus laevis* (Fazenda et al., 2010). Interestingly, in this model, two functional promoters (proximal and distal) were identified in the *MGP* gene, and both were shown to be regulated by Runx2.

1.2.2 *MGP* in Cancer

MGP was recently implicated in tumorigenic processes such as angiogenesis and shown to be abnormally regulated in several tumors, including cervical, ovarian, urogenital, and breast

(Hough et al. 2001; Sterzyńska et al. 2018; Levedakou et al. 1992) However, little is known about how *MGP* is involved in tumor signaling pathways, how the deregulated expression is related to different kinds of tumors and how its expression levels affect the prognosis of patients.

Interestingly, several studies already have demonstrated the potential of this gene as a possible biomarker in a wide range of cancers. For example, for breast and glioblastomas cancers (Fan et al. 2001; Yoshimura et al. 2009; Mertsch et al. 2009; Kuzontkoski et al. 2010), a relationship between the upregulation of *MGP* and a poor prognosis was shown (Yoshimura et al. 2009; Mertsch et al. 2009). In another study, also in breast cancer, the putative value of *MGP* for the poor prognosis was reported, and the results showed that patients exhibiting a poor prognosis had *MGP* mRNA expression up-regulated, however, it was not possible to establish a correlation between the protein levels and overall survival rate (Yoshimura et al. 2009). More recently it has been shown that the *MGP* gene regulation is promoted by HOCXC8 in triple-negative breast cancer cells (Gong et al. 2019)

MGP overexpression was also found in ovarian cancer (Hough et al. 2001; Sterzyńska et al. 2018), prostate epithelial cells undergoing apoptosis and urogenital malignancies (Levedakou et al. 1992), and more recently in gastric cancer (Wang et al. 2020).

In the case of colorectal cancer, the expression of *MGP* is far from consensual. For example, in the first study of *MGP* in colorectal cancer, a down-regulation of *MGP* expression was observed in the tumoral tissue, but it was not possible to establish a correlation between *MGP* gene expression and the tumor stages (Fan et al. 2001). More recently, down-regulation of *MGP* in the colorectal tumoral tissue was also described (Alon et al. 1999; Notterman et al. 2001; Fan et al. 2001; Sabates-Bellver et al. 2007). However, other studies have shown that *MGP* was overexpressed in colorectal tumor tissue (Ki et al. 2007; Li et al. 2020; Huang et al. 2021).

A down-regulation of MGP was also observed in head and neck cancer (Mishra et al. 2018).

Nevertheless, these studies suggest that additional players must be involved in *MGP* regulation, and more studies need to be developed to understand the role of *MGP* in tumorigenesis.

1.3 Cancer Biology

Cancer consists of a complex multistep process, which ultimately leads to uncontrolled cellular growth.

In normal circumstances, a cell can divide and proliferate when receives external signals to be able to maintain the equilibrium, but if those are altered, death can occur, either promoted by

the normal aging of the cell or in circumstances promoting some sort of tissue or DNA injury (Mendelsohn et al. 2008).

To occur cell division, the cells need to have specialized cellular machinery to be able to synchronize this entire process, which is known as the cell cycle. This machinery comprises a complex and specific network that can be triggered by external signals, molecules (paracrine regulation), or hormones (endocrine regulation) or produced by the own cells (autocrine regulation), thus ensuring the normal functioning of the cells (Mendelsohn et al. 2008)

From the beginning until the end of the cell cycle, the cell must pass orderly through four distinct phases, the G1 phase (cell growth); S phase (DNA synthesis), G2 (pre-division), and M phase (cell division), to give rise to two identical cells derived from the parental cell. In the S stage, DNA replication occurs, and all the genetic information contained in the DNA is copied to be passed to the daughter cells (Mendelsohn et al. 2008).

Sometimes during these events, some errors during DNA replication or cell division might occur leading to the malfunctioning of the cell. To prevent the dissemination of those errors, between each phase of the cell cycle exists the so-called checkpoints that stop the progression of the cell cycle and target the cell for death (apoptosis) (Mendelsohn et al. 2008)

In cancerous cells, all the dynamics of this complex network related to cell regulation are in some way deregulated. In most cases, during DNA replication, some errors, such as insertions or deletions of nucleotides might occur. This affects the proper function of important genes involved in cell cycle regulation or even the proper functioning of genes themselves involved in other molecular pathways, causing the gain of fundamental characteristic traits that prompts these cells to evade apoptosis. Ultimately, this may result in abnormal growth of cells if those errors are not properly corrected (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011; Hanahan, 2022).

Not all the abnormal growth of cells will give rise to malignant tumors, for instance, cells may start to slowly divide uncontrollably, but remain in the primary site location where they arose, originating benign tumors. On the other hand, those cells can acquire other specific features that will allow them to rapidly start to divide and spread to distant tissues from the original tumor site leading to the development of metastasis (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011).

Some of the most important features of cancer cells rely on the ability to create advantageous microenvironments that enable them to proliferate and survive in hostile conditions by i) recruiting other cells (through secretion of cytokines) that will emit important stimulatory signals to induce tumor growth; ii) emit their signals to prevent immune responses; iii) resist to

cell death; iv) gain the ability to become immortal and replicate continuously, or v) producing enzymes that will digest the involving matrix allowing cancer cells to migrate into surrounding tissues (Hanahan and Weinberg, 2000).

Another key characteristic of tumors is their genome instability, which can correspond either to microsatellite or chromosome instability, creating genetic diversity among tumors and promoting the carcinogenic process (Hanahan & Weinberg, 2011).

Since the implementation of the physiological behaviors of cancer cells by Hanahan and Weinberg in 2000, much has been accomplished over the past two decades and how these traits can be caused (Hanahan, 2022).

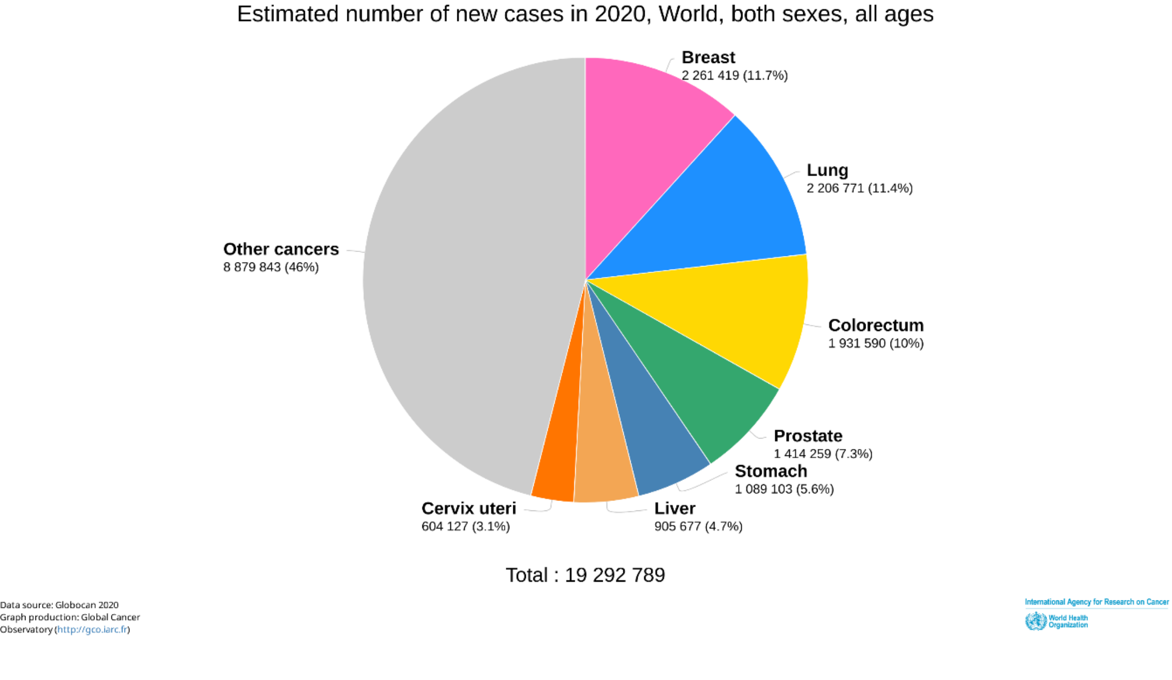
Nevertheless, the vaster and more complex concept related to the biology of cancer not only relies on the deregulation of molecular processes such as the ones involved in the cell cycle but also is related to the crosstalk with other molecular events that can promote abnormal gene regulation (Hanahan, 2022)

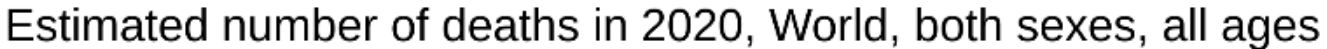
1.3.1 Cancer epidemiology

According to the latest data released by the International Agency for Research on Cancer, cancer is one of the leading causes of death worldwide, accounting for almost 10 million deaths and around 19 million new cases in 2020. The most common cancers were breast, lung, colon and rectal, prostate, skin, and stomach, in case of the incidence of new cases in the world and lung, colon and rectal, liver, stomach, and breast as the most common cause of deaths (Figure 1.3) (Ferlay et al. 2021).

According to the published data, there is a risk of one in five persons developing cancer disease and one in 10 persons perishing from this disease worldwide. Furthermore, most types of cancers can vary depending on the country.

Being cancer one of the major concerns all over the world, that could affect any person, the strategy is now focusing on the prevention (by avoiding external cancer-causing agents, i.e. tobacco, alcohol, etc.) and early detection, which in the case of mortality, it can be reduced when tumors are detected at an early state and treated in a more specific and appropriate manner.





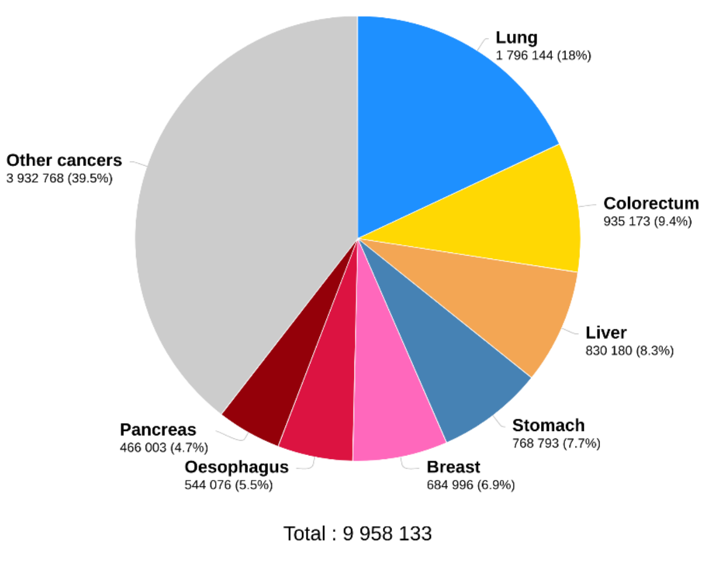


Figure 1.3 Cancer incidence and mortality worldwide by female and male genders and all ages. Estimation of 19 million new cancers and around 10 million deaths in 2020 according to the International Agency for Research on Cancer, World Health Organization (WHO) data.

1.3.2 Cancer regulation

Normal cells depend on the proper functioning of specific key genes that are important in the prevention, development, and progression of cancer.

Normally, in cancer cells, these genes are deregulated, promoting abnormal proliferation. Some of these important genes are known as proto-oncogenes (in normal cells) or oncogenes (in cancer cells), which contribute to the progression and inhibition of cell death; alternatively, deregulation of tumor suppressor genes, which are responsible to prevent abnormal cell division, leading to cell death, is also common in cancer (Mendelsohn et al. 2008).
Numerous genes have been identified as proto-oncogenes, and some of these genes are responsible for the stimulation of cell growth and control of cell death.
Usually in cancer, only one copy of the mutated proto-oncogene (oncogene) is necessary to provide signals for constant cell division and avoid cell death (Mendelsohn et al. 2008)

1.4 Concept of Epigenetics

One of the emerging concepts over the past decades related to this crosstalk between disease and malfunctioning genes is epigenetics. Despite the enormous and complex molecular and genetic processes that regulate the cell cycle, several other molecular events promote gene regulation without altering genetic information.
This concept relies on the regulation of gene functions without changing the DNA sequence, through small changes affecting gene regulation in a reversible manner (Inbar-Feigenberg et al. 2013; Chen et al. 2014)

1.5. Epigenetic modifications

1.5.1 DNA methylation

DNA methylation is one of the most studied epigenetic mechanisms (Inbar-Feigenberg et al. 2013). This mechanism comprises the transference of a methyl group to cytosine on CpG dinucleotides (Jones & Baylin, 2002; Chen et al. 2014) and is typically associated with gene silencing through the binding of methylation-sensitive DNA binding proteins or through interaction with various histone proteins modifications that can modulate chromatin states and consequently the access of gene promoters to transcriptional machinery (Jones and Baylin, 2002)
In the case of CpGs dinucleotides or CpG islands (when CpGs dinucleotides are present in a concentrated manner, ranging from 200bp to several kilobases in size), these can be found throughout the genome. When present in gene promoters, the Methyl groups can modulate gene regulation usually by inhibiting gene transcription by i) preventing the binding of enzymes or proteins necessary to the transcription machinery or gene activation, or ii) recruiting DNA

binding proteins that block gene transcription (Hassler and Egger, 2012). Less frequently, they can also promote activation of transcription by inhibiting the binding of repressor proteins, thus contributing to active genes promoting cancer cell division (Leão et al. 2018).
When a gene is highly methylated (hypermethylation) the silence of gene transcription is more prominent, whereas the lack of methylation groups (hypomethylation) is usually related to a more active state of the gene transcription (Hassler and Egger, 2012).

1.5.2 Histone modifications

The structural DNA is organized in nucleosomes, which are composed of DNA involving histones and forming chromatin. Each nucleosome is an octamer of histone proteins (two of H2A, H2B, H3, and H4). These histones are normally organized in terms of the location of their tails and are available for the binding of proteins or enzymes to ensure the proper opening or closure of chromatin (Kungulovski and Jeltsch, 2016).
It is in these tails that occurs the main post-translation modifications of histones, consequently altering the way how DNA interacts with regulatory proteins, ultimately leading to the activation or silence of genes.
There are different types of histone modifications including: i) histone phosphorylation; ii) histone methylation, iii) histone acetylation, iv) histone sumoylation, vi) histone ADP-ribosylation and many others, each one contributing to mark active or repressive chromatin states, as well as alterations related with gene transcription (Hassler and Egger, 2012; Zhou et al. 2014).

1.5.3 miRNAS

MicroRNAs (miRNAS) are small noncoding RNAs that indirectly control gene regulation and many other mechanisms, by inducing the inhibition of translation in a post-transcriptional manner or in a direct way through the regulation of epigenetic mechanisms (Suzuki et al. 2013). For example, microRNAs can regulate 3D chromatin structure, through the histone methyltransferase EZH2, thus altering gene expression (Suzuki et al. 2013).

1.6 Main objectives

Although MGP was implicated in tumorigenic processes until now the role and clinical implications of MGP in different types of cancer are not well-understood. Regarding its regulation, it is still unclear and much remains to be uncovered. So far, MGP was found to be

regulated in human by only a few transcription factors and, more recently, by epigenetic mechanisms.

This work aimed to analyze the DNA methylation pattern of MGP in different types of tumors, through bioinformatics and an *in vitro* approach, to identify a possible epigenetic regulation and new transcriptional regulators of *MGP* transcription.

Chapter 2

Assessment of MGP gene expression and contribution to cancer prognosis

Chapter 2.1

Evaluation of MGP gene expression in Colorectal Cancer

Helena Caiado[1,2,3]

Natércia Conceição[2,3,4*]

Daniel Tiago[2&]

Ana Marreiros[3,4]

Susana Vicente[5]

Jose Luis Enriquez[5]

Ana Margarida Vaz[6]

Artur Antunes[6]

Horácio Guerreiro[6]

Paulo Caldeira[3,6]

M. Leonor Cancela[2,3,4,7*]

This sub-chapter was published as a research manuscript to *Gene*, 723 (2019): 144120.
doi.org/10.1016/j.gene.2019.144120

[1]ProRegeM PhD Programme in Mechanisms of disease and regenerative medicine, University of Algarve, Faro 8005-139, Portugal
[2]Center of Marine Sciences (CCMAR), University of Algarve, Faro 8005-139, Portugal
[3]Department of Biomedical Sciences and Medicine, University of Algarve, Faro 8005-139, Portugal
[4]Algarve Biomedical Center, University of Algarve, Faro 8005-139, Portugal
[5]Pathology Department, University Hospital of Algarve, Faro 8000-386
[6]Gastroenterology Department, University Hospital of Algarve, Faro 8000-386
[7] Centre for Biomedical Research, University of Algarve, Faro 8005-139, Portugal
[&]Present address: Residency at University Hospital of Algarve, Faro 8000-386
*Correspondence to: M. L. Cancela, Department of Biomedical Sciences and Medicine, University of Algarve, 8005 Faro, Portugal, Email Address: lcancela@ualg.pt
N. Conceição, Department of Biomedical Sciences and Medicine, University of Algarve, 8005 Faro, Portugal, Email Address: nconcei@ualg.pt

Author's roles: Study design: NC; DT and MLC. Study conduct: HC, NC and DT; Data collection: HC, SV, JLE, AMV, AA, HG and PC. Data analysis: HC, NC, DT, AM and MLC. Data interpretation: All authors. Drafting manuscript: HC, NC, DT and MLC. Revising manuscript content: All authors. Approving final version of manuscript: All authors.

2.1.1 Abstract

Purpose Matrix Gla protein (MGP) is a vitamin K-dependent, γ-carboxylated protein that was initially found to be a physiological inhibitor of ectopic calcifications affecting mainly cartilage and the vascular system. Mutations in the *MGP* gene were found to be responsible for a human pathology, the Keutel syndrome, characterized by abnormal calcifications in cartilage, lungs, brain and vascular system. *MGP* was recently implicated in tumorigenic processes such as angiogenesis and shown to be abnormally regulated in several tumors, including cervical, ovarian, urogenital and breast. This fact has triggered our interest in analyzing the expression of *MGP* and of its regulator, the transcription factor runt related transcription factor 2 (*RUNX2*), in colorectal cancer (CRC).

Methods *MGP* and *RUNX2* expression were analyzed in cancer and non-tumor biopsies samples from 33 CRC patients and 9 healthy controls by RT-qPCR. Consequently, statistical analyses were performed to evaluate the clinical-pathological significance of *MGP* and *RUNX2* in CRC. MGP protein was also detected by immunohistochemical analysis.

Results showed an overall overexpression of *MGP* in the tumor mucosa of patients at mRNA level when compared to adjacent normal mucosa and healthy control tissues. In addition, analysis of the expression of *RUNX2* mRNA demonstrated an overexpression in CRC tissue samples and a positive correlation with *MGP* expression (Pearson correlation coefficient 0.636; $p \leq 0.01$) in tumor mucosa. However, correlations between *MGP* gene expression and clinical-pathological characteristics, such as gender, age and pathology classification did not provide relevant information that may shed light towards the differences of *MGP* expression observed between normal and malignant tissue.

Conclusions We were able to associate the high levels of *MGP* mRNA expression with a worse prognosis and survival rate lower than five years.

These results contributed to improve our understanding of the molecular mechanism underlying *MGP* deregulation in cancer.

2.1.2 Introduction

Matrix Gla Protein (MGP) was the second extracellular Gla protein to be discovered from bone following the identification of Bone Gla protein or Osteocalcin, both being dependent on vitamin K to undergo a post-translation modification required to shift to their calcium-binding active forms (Stenflo et al. 1974; Price, Otsuka, et al. 1976; Price, Poser, et al. 1976; Price et al. 1983).

Following its discovery in bone, MGP was identified in cartilage, being secreted into the extracellular matrix by chondrocytes (Hale et al. 1988), and later it was also detected in lung, heart and kidney (Fraser and Price, 1988).

The human MGP is a small, insoluble secreted protein with 84 amino acid residues in its mature form, preceded by a signal peptide. In its N terminal region MGP contains a phosphorylation domain with three phosphorylated serine residues and a target site for the enzyme ɣ-glutamyl carboxylase, required to ɣ -carboxylate the five glutamic acid (Glu) residues into ɣ -carboxyglutamic acid (Gla) residues through a vitamin K dependent posttranslational modification, (Price, Otsuka, et al. 1976; Price, Poser, et al. 1976; Price and Williamson, 1985; Fraser and Price, 1988; Cancela et al. 1990);.

Regarding its biological function, MGP-deficient (KO) mice displayed a premature death, as a result of an abnormal soft tissue calcification including their arteries, thus leading to blood vessel rupture and clarifying the role of MGP as a physiological inhibitor of calcification (Luo et al. 1997). Furthermore, patients harboring mutations in the human *MGP* gene exhibited features with similarities to those found in the MGP KO mice, and therefore its association with Keutel syndrome was established (Munroe et al. 1999). Matrix Gla Protein (MGP) has been also described as having an important role on mineralization inhibition and chondrocytes maturation (Yagami et al. 1999). In addition, MGP has been shown to be expressed with different levels of carboxylation but its significance in terms of functional effect remains unclear (Cranenburg et al. 2008; Schurgers et al. 2010; Mayer et al. 2014; Vermeer et al. 2015). MGP has been described as being abnormally expressed in β-thalassemia (Boraldi et al. 2013), atherosclerosis (Vassalle and Iervasi, 2014) and chronic kidney disease (Karsli Ceppioğlu et al. 2011). In addition to pathologies associated with abnormal calcifications, abnormal *MGP* expression has also been found to occur in different types of tumors including breast cancer (Yoshimura et al. 2009), colon adenocarcinoma (Fan et al. 2001) and glioblastomas (van den Boom et al. 2003; Martini et al. 2013; Fu et al. 2017) but its functional significance remains to be elucidated.

Little is known of how *MGP* acts during tumorigenesis pathways and in what manner this deregulation of *MGP* expression is related to tumorigenesis. Furthermore, contradictory data regarding its gene expression and role in cancer have been published, possibly related to the heterogeneity of tumors and associated pathologies, inherent to each patient. Both in breast cancer and in glioblastomas, a relationship between upregulation of *MGP* and a poor prognosis was identified (Mertsch et al., 2009; Yoshimura et al. 2009), however in another glioblastomas study, a downregulated expression was also observed to be correlated with an unfavorable

clinical status such as a poor differentiated state, a larger tumor size, and lymph node metastasis (Kuzontkoski et al. 2010). In patients with breast cancer the putative value of MGP as a biomarker for poor prognosis was reported. Results showed that in patients exhibiting a poor prognosis and in whom *MGP* mRNA expression was found to be upregulated in the tumor compared to normal tissue, *MGP* had the potential to serve as a prognostic indicator of the disease, but no correlation was established between protein level and overall survival rate (Yoshimura et al. 2009). *MGP* was also found to be overexpressed in other types of tumors, such as urogenital (Levedakou et al. 1992), ovarian (Hough et al. 2001), skin (Micke et al. 2007) and gastric cancer (Guo et al. 2010). Concerning colorectal cancer, the function of *MGP* gene expression is more ambiguous. In one recent study a significant upregulation of *MGP* gene expression was observed in tumor tissue when compared with normal tissue (Galamb et al. 2016), while a previous study showed the inverse situation (Fan et al. 2001), with levels of *MGP* mRNA found to be downregulated in colorectal tumor tissue as compared with their paired normal tissue samples (Fan et al. 2001).

Since colorectal accounts for the second most frequent cancer and one associated with high number of deaths (Ferlay et al. 2018; Malvezzi et al. 2018), we wanted to understand how the involvement of *MGP*, as well as one of its transcriptional regulators, *RUNX2,* could affect tumor progression, by relating their gene expression pattern with tumor differentiation, and also if there was any correlation with the different stages of tumor, age or gender of the patients affected with colorectal cancer.

2.1.3 Materials and Methods

2.1.3.1 Patient selection and collection of tissue samples

Tissue samples, as well as clinical and pathological information were obtained from 33 patients with colon adenocarcinomas when colonoscopic examination was performed at the Gastroenterology Service, University ospital of Algarve (CHUA) in Faro between January 2010 and December 2016. Samples were obtained through forceps biopsy directly from tumor (TM) and from matched normal mucosa collected 5 cm away from the tumor site (NM).

Colorectal tumor biopsies embedded in paraffin after being collected from patients were screened as a complementary diagnosis through direct Sanger sequencing for mutations on exons 2, 3 and 4 of the *RAS* genes family (*KRAS* and *NRAS*).

Colorectal tissue samples from the control group (9 subjects) without a pathology related to colorectal adenocarcinomas were collected during colonoscopic routine exams, to investigate

signs and symptoms of abdominal pain, gastrointestinal bleeding, changes in bowel habits, chronic diarrhea, or to check for macroscopic lesion, without being related with CRC. Samples were preserved in RNAlater (Sigma-Aldrich) and maintained at 4ºC up to 24 hours. Histopathological diagnosis was confirmed by the Pathology department of CHUA.

Clinical staging of the samples was also assessed, according to American Joint Committee on Cancer (AJCC) (Bochner et al. 2017). Samples were obtained according to the current legislation and approved by CHUA ethics commission, with full knowledge of the patients who signed informed consents.

2.1.3.2 RNA Extraction and cDNA synthesis

Total RNA was extracted following Chomczynski and Sacchi protocol (Chomczynski and Sacchi, 1987). For cDNA synthesis, 1µg of the extracted RNA was treated with RQ1 DNase (1U for each 1µg of RNA; Promega) and M-MLV reverse transcriptase (ThermoFisher Scientific) according to the manufacturer's instructions.

Relative gene expression was analyzed by $2^{-\Delta\Delta Ct}$ method and normalized with the expression of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) as reference gene. Primer sequences for *GAPDH*, *MGP* and *RUNX2* were as follows:

GAPDH: Forward: 5'-TCAACGGATTTGGTCGTATTGGGCG-3';
Reverse:5'-CTCGCTCCTGGAAGATGGTGATGGG-3';
MGP: Forward: 5'-TGCTGCTACACAAGACCCTGAGACTGA-3';
Reverse: 5'-GTAGCGTTCGCAAAGTCTGTAGTCATCAC-3';
RUNX2: Forward: 5'-GGAGTGGACGAGGCAAGAGTTTCACC-3';
Reverse: 5'-GCGGGACACCTACTCTCATACTGGG-3'.

Data were presented as the relative quantity of target mRNA normalized with *GAPDH* and relative to the mean expression of the control group.

2.1.3.3 Immunohistochemistry

MGP detection was performed by immunostaining in normal and tumor tissue samples using a primary antibody against MGP (ab192396, Abcam) at 1:200 concentration in the BenchMark Ultra (Ventana) equipment. Hematoxylin and eosin were used as counterstains and all the techniques were performed according to the manufacturer's protocols. Color development was

carried out using ultraView Universal DAB (3,3'-diaminobenzidine) detection kit (Ventana). Negative controls were carried out by the absence of the primary antibody and showed the lack of background staining by secondary antibodies (data not shown). Images were acquired under the light microscope Axio Imager Z2 microscope (Zeiss, Germany) at 50x and 400x magnification.

2.1.3.4 Statistical analysis

Statistical analysis was performed using SPSS software program version 25. Values of gene expression are presented as mean and standard deviation (SD) and two-sided *P* value less than 0.05 was demarcated as statistically significant. Fold changes presented were considered by the ratio of the values from tumor mucosa *versus* normal mucosa. Comparisons between group variables and gene expression were estimated using non parametric statistical tests: Mann–Whitney U and Kruskal–Wallis.

The cutoff value to distinguish the patients with low and high *MGP* and *RUNX2* levels were estimated taking into account the median value of the fold change for both *MGP* and *RUNX2*. A multivariate classification technique of two step clusters (Chiu et al. 2001) was performed to determine possible patient profiles, taking into account the characteristics of categorical and numerical variables, simultaneously. Pearson and Spearman coefficients were considered to analyze the correlation between *MGP* and *RUNX2* fold change values by the interest groups, namely, clusters and tissue samples. Overall survival probability for two groups of patients (clusters 1 and 2) was calculated using the Kaplan–Meier method; intergroup differences were determined using a log-rank test. Logistic regression analysis and χ^2 analysis was used to evaluate the independent influence of factors on the final prognosis.

2.1.4 Results

2.1.4.1 Clinical, demographic and pathological characteristics of patients

Clinical, demographic and histopathological information regarding patients (case group) is depicted in Tables 1 and 2. The mean of age (±SD) for the control group (n=9) and case group (n=33) were 67.78 (±11.14) and 71.70 (±12.00) years, respectively, with age ranging between 51 and 84 years for control group and 36 to 92 years for case group. Most of the case group were male (64%), without familial cancer history (79%), had localized disease (no metastasis,

73%) and presented previous pathologies (82%), most of them not related with colorectal cancer (Table 2.1.1), for example, diabetes, hypertension, Parkinson or psoriasis.

Most of the tumors were localized in the rectum (43%), well differentiated (91%), between stage III and IV (58%), without the occurrence of metastasis (76%) and without the presence of *RAS* genes family mutations (*KRAS* gene 64%) (Table 2.1.2). A *KRAS* mutation was present in codon 12 in 10 tumors and in codon 146 in the other 2. None of the case group analyzed had mutations in the *NRAS* gene.

Table 2.1.1 Demographic features of patients

	MGP (n=33)			*RUNX2* (n=33)		
Characteristics	Number (%)	Mean value of fold change	*p* value	Number (%)	Mean value of fold change	*p* value
Gender			0.113			0.122
Male	21 (64)	5.297		21 (64)	3.970	
Female	12 (36)	6.164		12 (36)	9.418	
Age (median: 71,70 years)			0.957			0.343
<72	15 (45)	5.083		15 (45)	3.072	
≥72	18 (55)	6.176		18 (55)	8.501	
Familial Cancer History			0.476			0.268
Yes	7 (21)	6.267		7 (21)	3.276	
No	26 (79)	3.473		26 (79)	6.733	
Previous Pathologies			0.476			0.268
Yes	27 (82)	5.257		27 (82)	5.149	
No	6 (18)	7.563		6 (18)	9.296	
Metastasis			0.476			0.268
Yes	9 (27)	7.824		9 (27)	6.722	
No	24 (73)	4.806		24 (73)	5.658	

Mann-Whitney U test

Table 2.1.2 Histopathological features of patients

Characteristics	*MGP* (n=33) Number (%)	*MGP* Mean value of fold change	*p* value	*RUNX2* (n=33) Number (%)	*RUNX2* Mean value of fold change	*p* value
Tumor Location			0.691			0.620
Rectum	14 (43)	4.828		14 (43)	4.494	
Rectosigmoid Junction	4 (12)	5.082		4 (12)	3.237	
Ascending Colon	4 (12)	3.787		4 (12)	4.675	
Sigmoid	5 (15)	8.106		5 (15)	4.276	
Cecum	2 (6)	2.793		2 (6)	11.333	
Hepatic Angle	3 (9)	3.385		3 (9)	1.551	
Transverse Colon	1 (3)	2.576		1 (3)	3.317	
Tumor Location grouped			0.686			0.406
Right side	10 (30)			10 (30)		
Left side	23 (70)			23 (70)		
Tumor Histology			0.287			0.091
Well Differentiated	30 (91)	4.419		30 (91)	3.777	
Poorly Differentiated	3 (9)	9.785		3 (9)	11.087	
Tumor Stage			0.382			0.489
I - II	14 (42)	4.352		14 (42)	3.615	
III - IV	19 (58)	5.315		19 (58)	5.051	
T classification			0.826			0.375
pT1	1 (3)	2.576		1 (3)	3.317	
pT2	5 (15)	3.598		5 (15)	7.730	
pT3	25 (76)	5.176		25 (76)	3.959	
pT4	2 (6)	5.967		2 (6)	2.825	
N classification			0.592			0.684
N0	14 (43)	4.352		14 (43)	3.615	
N1	11 (33)	6.156		11 (33)	6.245	
N2	8 (24)	4.159		8 (24)	3.409	
M classification			0.275			0.240
M0	25 (76)	4.120		25 (76)	3.795	
M1	8 (24)	7.365		8 (24)	6.463	
Hepatic Metastasis			0.252			0.333
Yes	7 (21)	8.050		7 (21)	6.913	
No	26 (79)	4.060		26 (79)	3.777	
Pulmonary Metastasis			0.491			0.573
Yes	3 (9)	9.930		3 (9)	9.930	
No	30 (91)	4.404		30 (91)	3.893	
Peritoneal Metastasis			1.000			0.529
Yes	1 (3)	2.576		1 (3)	3.317	
No	32 (97)	4.979		32 (97)	4.477	
KRAS mutations			0.369			0.331
Yes	12 (36)	3.328		12 (36)	5.479	
No	21 (64)	5.809		21 (64)	3.850	

Mann-Whitney U test

2.1.4.2 *MGP* gene expression was increased in colorectal cancer patients

RNA expression levels were measured by RT-qPCR in colorectal cancer using the RNA extracted from samples collected from the case group and control group (Figure 2.1.1). Overall, our results showed a significant increase of *MGP* expression in the majority of colorectal tumor mucosa samples compared with the paired normal mucosa samples (Figure 2.1.1, *p*=0.001). It was also investigated if the results obtained for normal mucosa (collected from tissue adjacent to the tumor (NM)) and the colorectal normal mucosa samples collected from the group control were comparable, to attest if this could validate the expression observed in the paired normal tissue of the patients. Results showed that in the mucosa from healthy individuals, there is no difference in MGP levels compared to those observed for normal mucosa samples, thus validating the levels of *MGP* expression in normal mucosa of CRC patients (Figure 2.1.1, *p*=0.105).

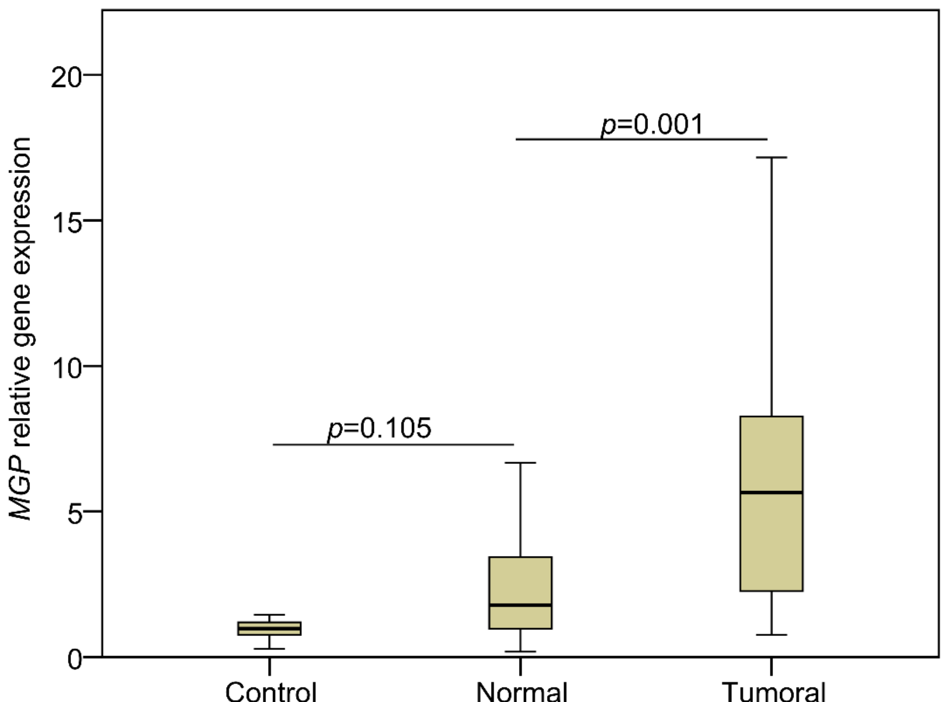


Figure 2.1.1 Relative MGP gene expression in samples from patients with colon adenocarcinoma Relative *MGP* gene expression levels were analyzed by RT-qPCR in a total of 9 control group and 33 colorectal cancer tissue samples (normal and tumor mucosa). Overall overexpression in tumor versus normal mucosa (n=33, *p*=0.001). *MGP* expression in normal mucosa and control group (n=9) without significant differences (*p*=0.105). Values are presented as mean ± SD. The Mann-Whitney and Kruskal Wallis non parametric tests were performed for the statistical analysis

In addition, it was possible to identify, among the 33 patients, a subgroup of five patients where levels of *MGP* were differentially expressed in normal versus tumor mucosa samples; however,

the difference in expression (between 0.4 and 0.8-fold change) was not statistically significant in this subgroup. (Supplementary Figure S2.1.1, panel A, *p*=0.548). We also compared the levels of *MGP* gene expression from the group control and the normal mucosa samples from this subgroup and results showed a significant difference between them (Supplementary Figure S2.1.1, panel A, *p*=0.004).

At protein level, it was possible to confirm the expression of MGP by immunohistochemistry in the patients whose expression was upregulated in tumor mucosa (Figure 2.1.2, panel A) and in the patients that presented an increased *MGP* gene expression in normal mucosa (Figure 2.1.2 panel B).

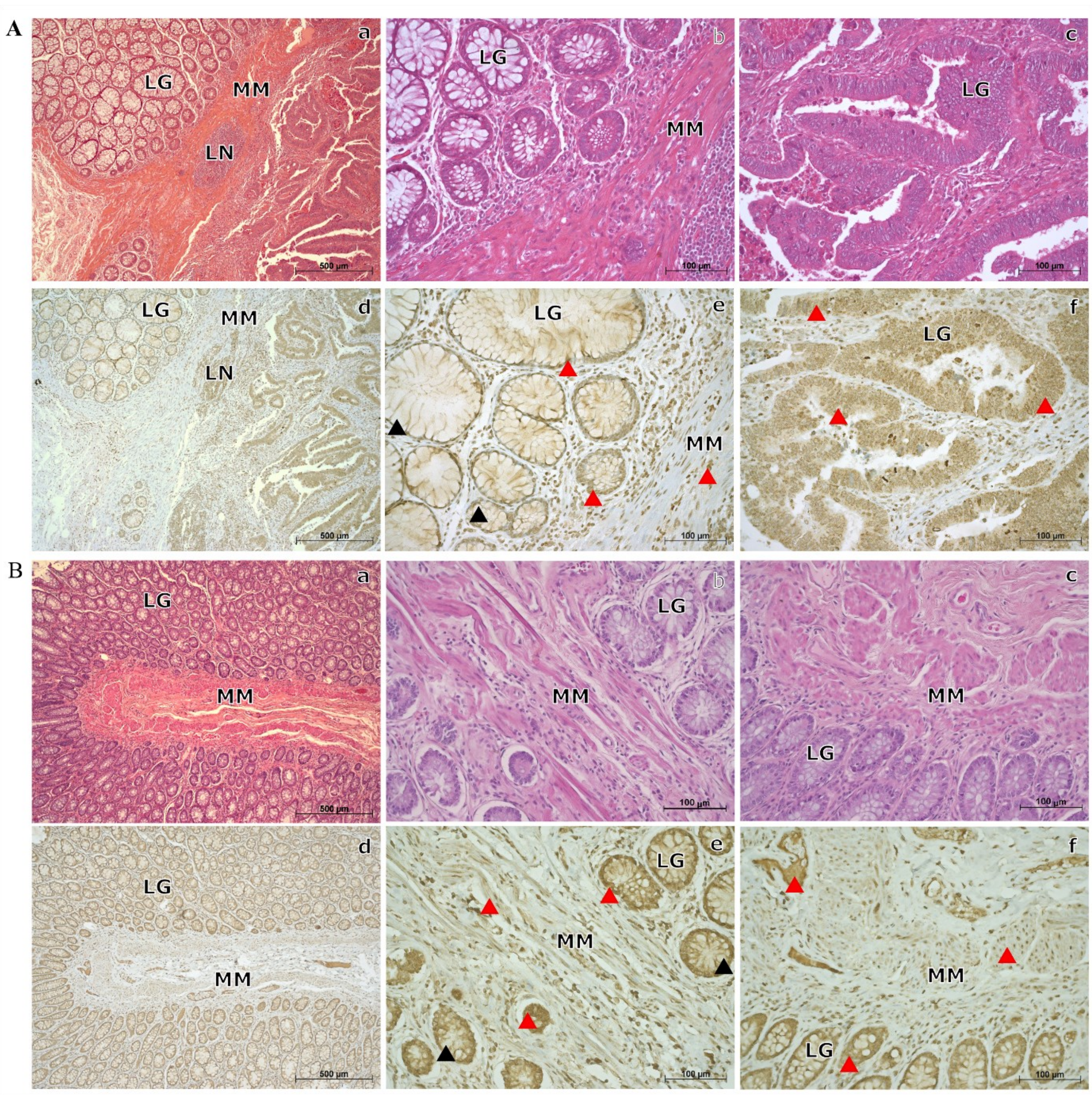


Figure 1.1.2 MGP immunostaining in colorectal cancer tissue samples from two patients. (A) MGP expression at protein level in a CRC patient with higher expression in tumor mucosa. Hematoxylin and eosin staining of (**a)** normal and tumor tissue, magnification 50x; (**b)** detail of normal tissue, magnification 200x; (**c)** detail of tumor tissue, magnification 200x. MGP immunostaining in (**d)** normal and tumor tissue, magnification 50x; (**e)** detail of normal tissue, magnification 200x; (**f)** detail of tumor tissue, magnification 200x.
(B) MGP expression at protein level in a CRC patient with higher expression in normal mucosa. Hematoxylin and eosin staining of (**a**) normal and tumor tissue, magnification 50x; (**b**) detail of normal tissue, magnification 200x; (**c)** detail of tumor tissue, magnification 200x. MGP immunostaining in (**d)** normal and tumor tissue, magnification 50x; (**e)** detail of normal tissue, magnification 200x; (**f)** detail of tumor tissue, magnification 200x.
Lymphatic Nodule (LN), Muscularis Mucosae (MM) and *Lieberkühn* Glands (LG). Black arrows indicate the Goblets cells and the red arrows indicate MGP expression

Based on the verified upregulation of *MGP* in most of the tumor mucosa samples, we were interested in evaluating whether this upregulation could have an impact on the patients overall survival (OS). For this, we established the median value of the fold change (ratio tumor mucosa/normal mucosa) as the cutoff to define the patients who would have a higher or lower *MGP* expression. We found that 35.3% of the patients presented a higher mortality rate when levels of *MGP* expression were upregulated (Supplementary Figure S2.1.2, panel A, log-rank test, *p*=0.026).

2.1.4.3 No significant associations were observed between *MGP* gene expression and clinical and pathological features of patients

To evaluate whether *MGP* expression could be correlated with clinical and pathological features of the patients, we analyzed all variables summarized in Tables 2.1.1 and 2.1.2. We did not identify any statistically significant associations between *MGP* expression with gender (*p*=0.113), age (*p*=0.957), familial cancer history (Yes vs. No; *p*=0.476), previous pathologies (Yes vs. No; *p*=0.476) or the presence of metastasis (Yes vs. No; *p*=0.476) (Table 2.1.1). We also did not found a significant association between *MGP* expression and pathological features of patients, such as tumor localization (*p*=0.691) or tumors grouped into right *versus* left sided (*p*=0.686), tumor histology (well differentiated vs poorly differentiated; *p*=0.287), tumor stage (stage I-II, stage III-IV; *P*=0.382), T classification (pT1, pT2, pT3, pT4; *p*=0.826), N classification (N0, N1, N2; *p*=0.592), M classification (M0, M1; *p*=0.275), hepatic (*p*=0.252), pulmonary (*p*=0.491) and peritoneal (*p*=1.000) metastasis and the presence or absence of *KRAS* mutations (*p*=0.369) (Table 2.1.2).
Moreover, we assessed if the high expression of *MGP* presented in the normal mucosa of the five patients could be related with clinical and pathological features of the patients but we did not find a significant correlation between them (results not shown).

2.1.4.4 Transcription factor *RUNX2* is upregulated in patients with colorectal cancer

To understand this deregulation of *MGP* in tumoral mucosa, we explored the levels of expression of the transcription factor *RUNX2*, since it is essential for gene expression of bone matrix proteins and it has been proposed as a regulator of *MGP* transcription (Stock et al. 2004; Takahashi et al. 2005; Suttamanatwong et al. 2009; Fazenda et al. 2010).

The results demonstrated a pattern of expression for *RUNX2* similar to the one obtained for *MGP*, with significantly increased expression in tumor versus normal mucosa (p=0.030) when compared with control group (p=0.695) (Figure 2.1.3). Since we found a small group of patients whose expression of *MGP* was higher in normal mucosa, we questioned if some of these patients could also present a higher expression of *RUNX2* in normal mucosa. In fact, we identified a group of nine patients whose expression was significantly higher in normal mucosa compared with tumor mucosa, indicating the heterogeneity of tumor populations (Supplementary Figure S2.1.1, panel B, p=0.040). Moreover, in the group of five and nine patients where *MGP* and *RUNX2* expression were higher in the normal mucosa, respectively, only three patients presented a higher expression for both genes in normal mucosa.

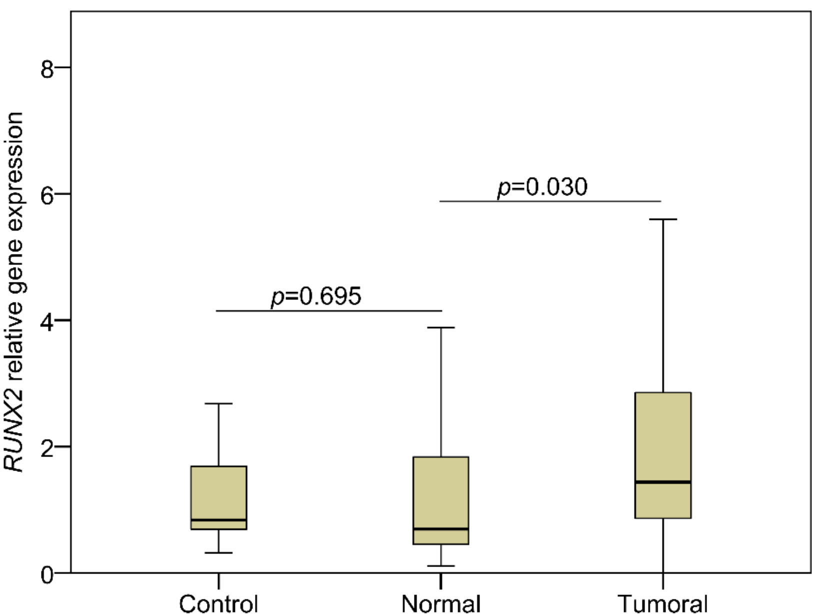


Figure 2.1.3 Relative RUNX2 gene expression in samples from patients with colon adenocarcinoma. Relative *RUNX2* gene expression levels were analyzed by RT-qPCR in a total of 9 control group and 33 colorectal cancer tissue samples (normal and tumor mucosa). Overall overexpression in tumor versus normal mucosa (n=33, p=0.030). *RUNX2* expression in normal mucosa and control group (n=9) without significant differences (p=0.695). Values are presented as mean ± SD. The Mann-Whitney and Kruskal Wallis non parametric tests were performed for the statistical analysis

Since we found an association between high levels of *MGP* and a poor OS rate, we also analyzed the association between *RUNX2* gene expression and OS. We found that the same percentage of patients that passed away in the first two years and presented high levels of *MGP*, also presented high levels of *RUNX2* (Supplementary Figure S2.1.2, panel B, log-rank test, p=0.034).

We then evaluated the correlation between *RUNX2* and *MGP* gene expression in tumor and normal mucosa. The correlation Pearson coefficients, between *MGP* and *RUNX2* gene expression was 0.636 ($p \leq 0.01$) in the tumor mucosa and was 0.204 (p=0.255) in the normal mucosa of patients (Table 4), establishing a significant, positive and strong correlation between *RUNX2* and *MGP* gene expression in tumor mucosa.

Table 2.1.4 Correlation between RUNX2 and MGP gene expression in normal and tumoral tissue

TISSUE SAMPLE			***RUNX2***
NORMAL	***MGP***	Pearson Correlation	0,204
		Bilateral Significance	0,255
		N	33
TUMORAL	***MGP***	Pearson Correlation	0,636**
		Bilateral Significance	0,000068
		N	33

** Correlation in significant at 0.01 level (2 bilateral)

2.1.4.5 *MGP* and *RUNX2* are important predictor factors to differentiate the prognosis of colorectal adenocarcinoma amongst patients

Since there was no association between *MGP* expression with the characteristics previously analyzed, we decided to further investigate which factors could best differentiate the patients by separating (at low and high) groups based on *MGP* and *RUNX2* gene expression, using the median value of fold change as a cut-off point (fold change *MGP* = 2.57, fold change *RUNX2*=1.80).

We then performed a two-step cluster analysis to determine the classification of patients according to the variables with the greatest predictive importance in the separation of the groups of individuals, considering the new variable of the dichotomized fold *MGP*, as well as the fold change of *RUNX*2 gene expression categorized in the same way. The variables considered were: T classification, N classification, tumor staging, gender, deceased, fold change *MGP*

categorized, fold change *RUNX2* categorized, fold change *MGP*, fold change *RUNX2*, tumor histology, *KRAS* mutations, tumor location, survival rate (years), polyposis and stroke.

According to the multivariate analysis of the predictors, patients were divided into clusters 1 and 2 (Table 2.1.3). Patients in cluster 1 presented N1 classification (62.5%), meaning that the majority of these patients had lymph node metastasis, the tumor was either stage III (56.3%) or stage IV (43.8%), mostly female (68.7%), with high *MGP* (75%) and *RUNX2* (68.8%) levels of expression, with a fold change for *MGP* of 6.2 (±5.6) and a fold change for *RUNX2* of 5.9 (±7.7), with a tumor histology moderately (50%) and poorly differentiated (12.5%), without mutation on *KRAS* (56.3%), with a T3 classification (81.3%), with a mean survival time of 2.5 (±2.0) years, with the tumor mostly located in rectum (37.5%) and without the presence of polyposis (93.8) and stroke (93.8%). In cluster 2 patients presented a N0 classification (82.4%), meaning that the majority of these patients did not present lymph node metastasis, the tumor was either stage I (23.5%) or stage II (58.8%), mostly male gender (94.1%), with low *MGP* (70.6%) and *RUNX2* (64.7%) levels of expression, with a fold change for *MGP* of 3.7 (±4.7) and a fold change for *RUNX2* of 3.1 (±4.4), with a tumor histology well differentiated (70.6%), without mutation on *KRAS* (70.6%), with a T3 classification (70.6%), with a mean survival time of 3 (±1.9) years, with the tumor mostly located in rectum (47.1%) and without the presence of polyposis (88.2%) and stroke (94.1%).

Despite the majority of patients in clusters 1 and 2 do not present mutations on *KRAS* gene, among the patients who have mutations in this gene (Table 2.1.5), these are mostly found in cluster 1, where 7 (43.7%) of the 16 patients showed mutations, while in cluster 2 only 5 (29.4%) of the 12 patients had mutations in the *KRAS* gene, these proportions are within the range (7-63%) reported in the literature (Ren et al. 2012).

We can conclude that high levels of *MGP* and *RUNX2*, combined with other factors, such as a higher tumor staging and advanced N and T classifications of tumors, are important predictor factors to profile the patients accordingly to their prognosis and in this case, patients categorized in cluster 1 have the worst prognosis and lowest survival rate associated with high levels of *MGP* and *RUNX2*, when compared with patients in cluster 2 (Table 2.1.3).

Table 2.1.3 Multivariate analysis of predictor factors

Characteristics	Cluster 1 (n=16, %)	Cluster 2 (n=17, %)	*p* value
N Classification			***p*≤0.05**[1]
N0	0 (0)	14 (82.4)	
N1	10 (62.5)	1 (5.9)	
N2	6 (37.5)	2 (11.8)	
Tumor Staging			***p*=0.05**[1]
Stage I	0 (0)	4 (23.5)	
Stage II	0 (0)	10 (58.8)	
Stage III	9 (56.3)	2 (11.8)	
Stge IV	7 (43.8)	1 (5.9)	
Gender			***p*=0.05**[1]
Male	5 (31.3)	16 (94.1)	
Deceased			***p*=0.05**[1]
No	9 (56.3)	17 (100)	
Fold change *MGP* categorized			***p*=0.05**[1]
High *MGP*	12 (75)	5 (29.4)	
Fold change *RUNX2* categorized			*p*=0.055[1]
High *Runx2*	11 (68.8)	6 (35.3)	
Fold Change *MGP*, mean (SD[2])	6.2(±5.6)	3.7(±4.7)	***p*=0.05**[3]
Fold change *RUNX2*, mean (SD[2])	5.9(±7.7)	3.1(±4.4)	*p*=0.110[3]
MGP vs *RUNX2*[5]	r=0.450; P=0.080	r=0.365; P=0.149	
Tumor Histology			*p*=0.108[1]
Well differentiated	4 (25)	12 (70.6)	
Moderately differentiated	8 (50)	3 (17.6)	
Poorly Differentiated	2 (12.5)	1 (5.9)	
Mucinous	1 (6.3)	0 (0)	
Mucinous well differentiated	1 (6.3)	1 (5.9)	
KRAS mutations			*p*=0.392[1]
No	9 (56.3)	12 (70.6)	
T classification			*p*=0.421[1]
T1	1 (6.3)	0 (0)	
T2	1 (6.3)	4 (23.5)	
T3	13 (81.3)	12 (70.6)	
T4	1 (6.3)	1 (5.9)	
Survival Rate (Years), mean (SD[2])	2.5(±2.0)	3.0(±1.9)	***p*=0.05**[4]
Tumor Location			*p*=0.447[1]
Rectum	6 (37.5)	8 (47.1)	
Rectosigmoid junction	3 (18.8)	1 (5.9)	
Ascending colon	1 (6.3)	3 (17.6)	
Sigmoid	2 (12.5)	3 (17.6)	
Cecum	2 (12.5)	0 (0)	
Hepatic angle	1 (6.3)	2 (11.8)	
Transverse Colon	1 (6.3)	0 (0)	
Polyposis			*p*=0.582[1]
No	15 (93.8)	15 (88.2)	
Stroke			*p*=0.965[1]
No	15 (93.8)	16 (94.1)	

[1] Chi Square test
[2] Standard Deviation
[3] Mann-Whitney test
[4] Log Rank test
[5] Spearman correlation coefficient
Boldfaced values - Variables with p ≤ 0.05

Table 2.1.5 Mutations of KRAS gene found in the patients

RAS GENES	EXONS	CODON MUTATION
KRAS	Exon 2	Gly 12 Val Gly 12 Asp Gly 12 Cys
	Exon 4	Ala 146 Thr

Subsequently, a Kaplan-Meier survival analysis was performed to assess if *MGP* and *RUNX2* could be in fact good prognostic factors in terms of overall survival rate and if their high expression might also be related with advanced tumor stages for the two groups of patients found in the two-step clusters analysis. Patients in cluster 1, which presented a worst prognosis, had a higher mortality rate when compared with patients in cluster 2 (log-rank test *p*= 0.003) (Figure 4) and the expression of *MGP* was significantly associated with mortality rate (Mann-Whitney test, *p*=0.019), while *RUNX2* expression seems not to have a significant influence on the mortality rate of these patients (Mann-Whitney test, *p*=0.110) (Table 2.1.3).

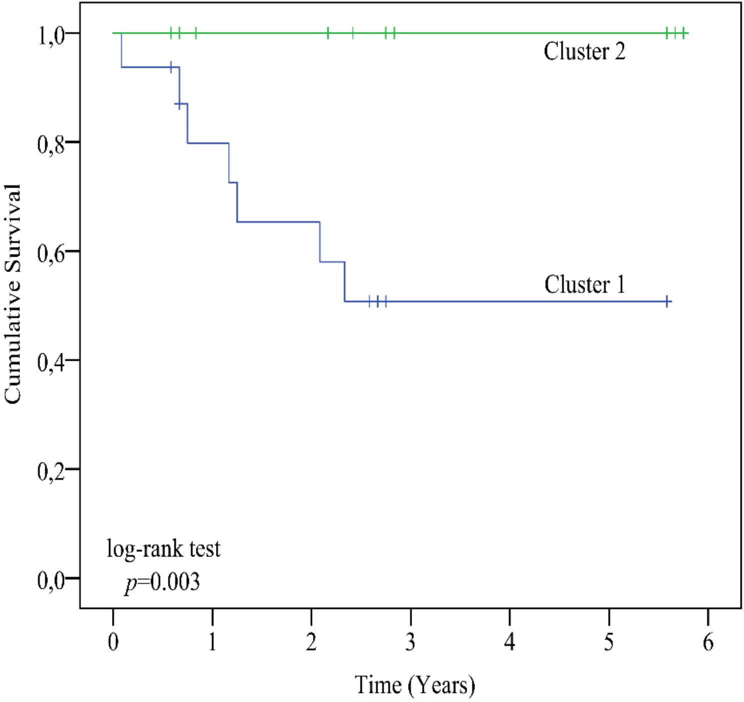


Figure 2.1.4 Overall survival curve for patients categorized by clusters 1 and 2. Patients in Cluster 2 present a better survival rate, when compared with patients in cluster 1 with a lower survival rate and a worse prognosis (*p*=0.003). Small vertical lines indicate the censored cases. Censored cases refer to the number of patients that have not reached the terminal event during this study. *P*-value was calculated by log-rank test

To evaluate the effect of *MGP* and *RUNX2* gene expression in the mortality rate of the patients within five years by clusters, we categorized the patients divided by clusters according to the length of their survival time above and below five years and considered the fold change of *MGP* and *RUNX2* categorized according to their median values (fold change *MGP* = 2.57, fold change *RUNX2*=1.80). We found that high levels of *MGP* (log-rank test, *p*=0.036) and *RUNX2* (log-rank test, *p*=0.041) are significantly correlated with a survival time lower than five years in patients in cluster 1 with a worst prognosis.

2.1.5 Discussion

Colorectal cancer is an exceptional example in the investigation of cancer, since it has relatively slow development that allows researchers to explore specific genes involved in critical steps of its progression. For example, *β-catenin* (Wai et al. 2006), *C-MYC* (Wai et al. 2006) and *KRAS* (Tong et al. 2014) are oncogenes involved in colorectal cancer progression, and are currently used as biomarkers for cancer prognosis and therapeutic decision (C.-W. Fan et al. 2013).
MGP has been found to be associated with and/or play an important role in several types of pathologies (Weaver et al. 2014; Gluba-Brzózka et al. 2014; Gerdes et al. 2014) and was found to be present in different types of tumors, such as breast cancer and glioblastoma, as recently reviewed by Gheorghe and Crăciun (2015). Interestingly, being one of the most common and well-described cancer types, the association of *MGP* to colorectal cancer has been so far poorly investigated (Fan et al. 2001; Gheorghe and Crăciun, 2015).
Although many studies have been previously conducted to try to understand how tumor cells proliferate and migrate to other organs (Mertsch et al. 2009), only few studies have related MGP to those events (Mertsch et al. 2009; Sharma and Albig, 2013), and even fewer studies have actually investigated the influence of *MGP* expression in tumorigenesis (Mertsch et al. 2009; Yoshimura et al. 2009; Tiago et al. 2016).
Our results, in contrast with a previous study where *MGP* was shown to be down-regulated in colorectal tumor tissues (Fan et al. 2001), demonstrate the association between *MGP* gene expression and the poor prognosis of patients with colorectal cancer. According to Fan et al. (2001) approximately 79% (n=63) of patients presented a downregulation of *MGP* gene expression in tumor mucosa samples, while 21% (n=17) of patients presented an increased *MGP* gene expression in tumor mucosa, whereas our results demonstrated that 84.8% (n=28) of patients presented a significant upregulation of *MGP* in tumor mucosa samples, while 15.2% (n=5) of patients presented an increased expression in normal mucosa samples. This

discrepancy between both data could be associated to different populations, since the previous study evaluated *MGP* gene expression in an Asiatic population (80 patients from Taiwan), while our analysis focused on a European Caucasian population (33 patients from South Portugal). Furthermore, tumors are known to present high heterogeneity and the presence of differential gene expression in tumors with similar locations has been shown for other genes (Micke et al. 2007), thus this could account for the differences identified for *MGP* expression. Throughout this analysis of expression, we found a minor population of five CRC patients who presented a non-significant upregulation of *MGP* expression in normal mucosa compared to tumor mucosa (confirmed by immunohistochemistry, Figure 2.1.2) in agreement to what the previous authors have shown (Fan et al. 2001). Yet, it was not possible to establish a significant correlation between these results and clinical and histopathological features of those patients. In a more recent study by Galamb et al. (2016), it was demonstrated that several age-related DNA methylation alterations could affect the expression of genes during the development and progression of colorectal cancer. In this study, the authors analyzed *MGP* gene expression using Affymetrix HGU133 Plus2.0 whole transcriptome data of 49 healthy controls, 49 patients with adenomas, 49 patients with CRC and 6 healthy children as controls (GEO accession numbers: GSE37364, GSE10714, GSE4183, GSE37267). and found that *MGP* gene expression was more similar between healthy children and CRC patients than the other analyzed patients with adenomas or healthy adults. In addition, the levels obtained for CRC patients are in the same range of values as the ones obtained in our study, thus contributing to validate the results obtained in our study. These results also confirm the heterogeneity of CRC tumors and indicate the usefulness of searching for additional tumor predictors to help better characterize patient populations and thus contribute to improve their treatments. Such analysis could also contribute to understand the inherent molecular mechanisms underlying MGP deregulation in cancer, and contribute to find new biomarkers for CRC helping to better characterize the patient populations.

Transcription factor *RUNX2* is a regulator of *MGP* transcription (Stock et al. 2004; Takahashi et al. 2005; Suttamanatwong et al. 2009; Fazenda et al. 2010). Previous experiments have shown that *RUNX2* was associated with tumor progression in colorectal cancer by affecting the normal regulation of metastatic genes in murine colorectal cancer cells (Wai et al. 2006). Furthermore, *RUNX2* was shown to be associated with higher risk of colorectal cancer (Slattery et al. 2011) and has been proposed as a possible prognostic factor (Sase et al. 2012). Therefore, since *RUNX2* is a known regulator of *MGP* transcription in *Xenopus laevis* (Fazenda et al. 2010), that also regulates *MGP* expression through parathyroid hormone in rodents via PKA and

ERK/MAPK signaling pathway and was reported to be also a *MGP* regulator in breast cancer metastasis (Suttamanatwong et al. 2009), we hypothesized whether this transcription factor could also be upregulated in our samples. We found that patients with high levels of *RUNX2* also presented a worst prognosis almost identically of what we found for patients with high levels of *MGP*, suggesting a possible role on *MGP* regulation as previously demonstrated in another biological system (Fazenda et al. 2010). Interestingly, we found a group of nine patients whose *RUNX2* expression was upregulated in the normal mucosa, three of those patients presenting both upregulation of *MGP* and *RUNX2* in normal mucosa. However, by using a Pearson correlation analysis between *RUNX2* and *MGP* expression we found a strong and positive relationship in the tumor mucosa, but not in normal mucosa (Table 2.1.4). This could be explained by the involvement of other signaling pathways in the regulation of *RUNX2* that can indirectly regulate expression of *MGP*, and therefore, in this particular case, it is not possible to affirm that the tendency for a possible upregulation of *RUNX2* is related to the higher levels of *MGP*. Nevertheless, and given the known function of *RUNX2* in cancer (Wai et al. 2006; Slattery et al. 2011; Ferrari et al. 2013), our results bring a new insight towards future experiments aiming to test whether the effect of RUNX2 disturbs or not *MGP* regulation in cancer.

It is known that patients in advanced stages of cancer have multiple factors related with the disease, for example, the presence of lymph node metastasis or even the presence of distant metastasis, which are important factors that should be taken into consideration at the time of treatments (Zhang et al. 2010). In addition, it is not always possible to detect the presence of micrometastasis and in these cases the tumor can be much more aggressive than initially perceived and endorse a worse prognosis (Bork et al. 2014; Lee et al. 2015). When considering the factors that could better categorize the patients according to the variables with the greatest predictive importance in the separation of the groups of individuals in the multivariate analysis, we found that the upregulation of both *MGP* and *RUNX2* combined with other factors, such as T, N, and M status, tumor staging, gender, tumor histology, the presence or absence of *RAS* genes family (*KRAS*) mutations, tumor location, survival rate, presence of polyposis and stroke, formed an important and relevant set of factors capable of differentiating the patient population in two clusters, with patients in cluster 2 having a better survival outcome than patients in cluster 1.Accordingly, and from what is known in the literature, from the several factors associated with a worse overall outcome, a poorly differentiated tumor is more likely to metastasize than a tumor that has a moderate differentiation (Tsigris et al. 2002) a result which

is in agreement with our data since most of the patients in cluster 1 presented a poorly differentiated tumor tissue with the presence of lymph node metastases.

Taking all these factors into account for the prognosis of the patients we could perceive that all together, but not individually, they can characterize the population into two distinct groups with a worst and better prognosis, and present a significant association with *MGP* and/or *RUNX2* gene expression. In addition, this analysis opens a new perspective on how these genes, i.e., *MGP* and its transcription factor, *RUNX2*, may be interacting in the tumor microenvironment of colorectal cancer.

2.1.6 Conclusion

The overall results from this study suggest that high expression levels of *MGP* and *RUNX2* are related with a worst prognosis of CRC and are important predictor factors to differentiate patients with colorectal cancer with a better or worse survival outcome, in particular when analyzed together with a given set of identified predictive factors. However, this study also highlights the existence of different patient populations differentially expressing these genes, indicative of the presence of genetic variability among the patients which likely accounts for the different expression pattern observed for *MGP* and *RUNX2* in tumor mucosa.

Further experiments may provide additional relevant information towards understanding the role of MGP in cancer, in particular aiming to further explore i) MGP association to tumor mucosa, ii) MGP association to specific phases of tumor progression, iii) cellular mechanisms underlying its regulation, and iv) epigenetic factors involved in *MGP* regulation.

2.1.7 Supplementary Data

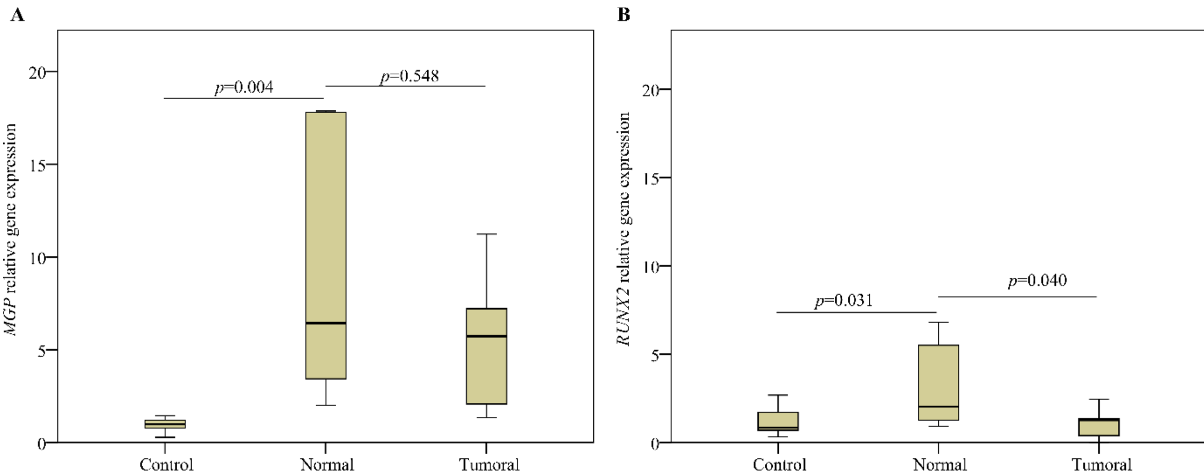


Figure S2.1.1 Increased expression of MGP and RUNX2 gene expression in normal samples from patients with colon adenocarcinoma Differential expression of *MGP* in the normal mucosa samples when compared with tumor mucosa samples without significant differences (panel A, n=5, *p*=0.548) and significant overexpression of *RUNX2* gene expression in normal mucosa when compared with tumor mucosa (panel B, n=9, *p*=0.011). Values are presented as mean ± SD. Mann-Whitney and Kruskal Wallis non parametric tests

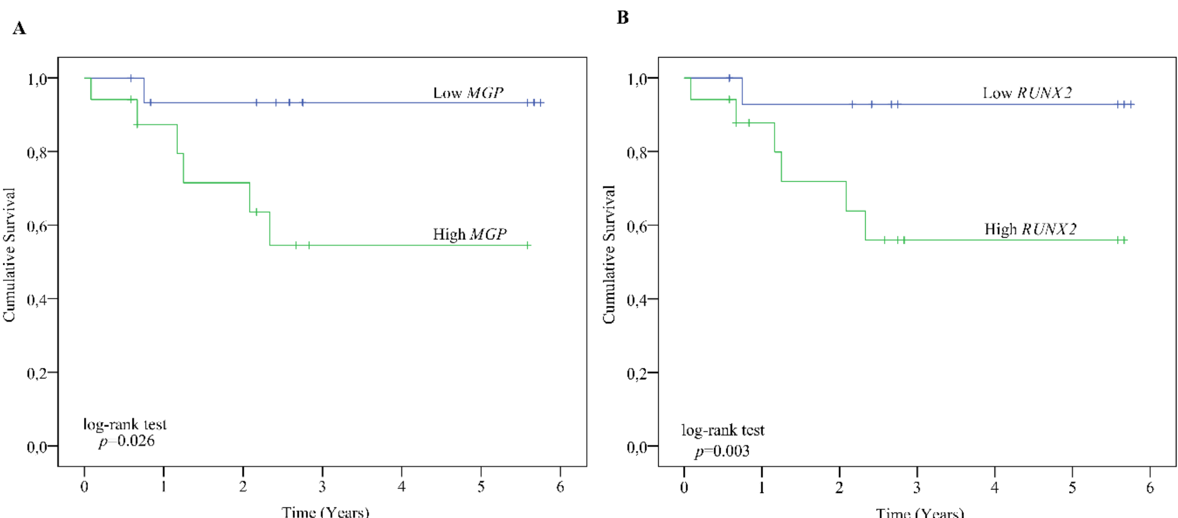


Figure S2.1.2 Overall survival curve of patients with overexpression of MGP and RUNX2 in tumor mucosa. Patients with high levels of *MGP* (panel A, *p*=0.026) and *RUNX2* gene expression (panel B, *p*=0.003) have a worst survival rate associated. Small vertical lines indicate the censored cases. Censored cases refer to the number of patients that have not reached the terminal event during this study. *P*-value was calculated by log-rank test

Chapter 2.2

Data on the evaluation of *FGF2* gene expression in Colorectal Cancer

Helena Caiado[1,2,3]

Natércia Conceição[2,3,4*]

Daniel Tiago[2&]

Ana Marreiros[3,4]

Susana Vicente[5]

Jose Luis Enriquez[5]

Ana Margarida Vaz[6]

Artur Antunes[6]

Horácio Guerreiro[6]

Paulo Caldeira[3,6]

M. Leonor Cancela[2,3,4,7*]

This sub-chapter was published as a research manuscript to Data in Brief, 31 (2020): 105765. doi.org/10.1016/j.dib.2020.105765

[1]ProRegeM PhD Programme in Mechanisms of disease and regenerative medicine, University of Algarve, Faro 8005-139, Portugal
[2]Center of Marine Sciences (CCMAR), University of Algarve, Faro 8005-139, Portugal
[3]Department of Biomedical Sciences and Medicine, University of Algarve, Faro 8005-139, Portugal
[4]Algarve Biomedical Center, University of Algarve, Faro 8005-139, Portugal
[5]Pathology Department, University Hospital of Algarve, Faro 8000-386
[6]Gastroenterology Department, University Hospital of Algarve, Faro 8000-386
[7] Centre for Biomedical Research, University of Algarve, Faro 8005-139, Portugal
[&]Present address: Residency at University Hospital of Algarve, Faro 8000-386
*Correspondence to: M. L. Cancela, Department of Biomedical Sciences and Medicine, University of Algarve, 8005 Faro, Portugal, Email Address: lcancela@ualg.pt
N. Conceição, Department of Biomedical Sciences and Medicine, University of Algarve, 8005 Faro, Portugal, Email Address: nconcei@ualg.pt

Author's roles: Study design: NC; DT and MLC. Study conduct: HC, NC and DT; Data collection: HC, SV, JLE, AMV, AA, HG and PC. Data analysis: HC, NC, DT, AM and MLC. Data interpretation: All authors. Drafting manuscript: HC, NC, DT and MLC. Revising manuscript content: All authors. Approving final version of manuscript: All authors.

2.2.1 Abstract

The data presented in this article is related with the research paper entitled “Evaluation of MGP gene expression in colorectal cancer”, available on Gene journal (Caiado et al. 2020). From all the transcription factors known to regulate *MGP*, FGF2 is the most described in colon adenocarcinoma and colon tumor cell lines, where it was shown to: i) contribute for the invasiveness potential; and ii) promote proliferation and survival of colorectal cancer cells. These *in vitro* studies pose the hypothesis that FGF2 associated signaling pathways could be promoting the regulation of others genes, such as *MGP,* that may lead to tumor progression which ultimately could result in poor prognosis in colon adenocarcinoma.

2.2.2 Specifications Table

SUBJECT	MOLECULAR BIOLOGY
SPECIFIC SUBJECT AREA	Colorectal cancer, Molecular biology
TYPE OF DATA	Table Graph Figure
HOW DATA WERE ACQUIRED	qRT-PCR, SPSS
DATA FORMAT	Raw Analysed
PARAMETERS FOR DATA COLLECTION	FGF2 was shown to be both a regulator of MGP and an inhibitor of cellular differentiation in colorectal cancer organoids, and FGF family proteins were proven to have an important role on the survival and growth of stem cells during embryogenesis, carcinogenesis and tissue regeneration
DESCRIPTION OF DATA COLLECTION	FGF2 gene expression analysis through qRT-PCR and assessment of the correlation with MGP gene expression and clinical and histopathological data analysis using SPSS software in colorectal patients
DATA SOURCE LOCATION	University of Algarve Faro Portugal
DATA ACCESSIBILITY	Data is available with this publication

RELATED RESEARCH ARTICLE	Caiado, H. et al. 2019 Evaluation of MGP gene expression in colorectal cancer Gene doi.org/10.1016/j.gene.2019.144120

2.2.3 Value of the Data

- The data presented here were obtained in order to evaluate *FGF2* gene expression in patients with colorectal cancer. This data may be of great relevance in trying to understand how *MGP* gene expression deregulation may affect patient´s prognosis.
- Beneficiaries of these data are all those who seek knowledge about the molecular mechanisms that could be underlying *MGP* deregulation in tumorigenesis.
- These data report the upregulation of *FGF2* gene expression in tumor tissue and its positive correlation with *MGP* gene expression in CRC. These results could provide future insights for the search of new therapeutic targets associated with *MGP* gene expression and its deregulation in cancer.

2.2.4 Data Description

The fibroblast growth factor (FGF) signaling network has been implicated in several pathways, such as normal cell growth, differentiation, angiogenesis and tumor development (Akl et al. 2016). The transcription factor FGF2 is one of the most studied in terms of its role in carcinogenesis including its role in tumor cell differentiation and proliferation (Akl et al. 2016). Moreover, it is known that FGF2 induces transcription of the *MGP* gene (Stheneur et al. 2003). In this report, we describe data regarding the expression analysis performed by qRT-PCR for *FGF2,* for both normal and tumor tissues, of 23 out of 33 CRC patients (Caiado et al. 2020) whose samples were still available, and 9 samples from the control group (Figure 2.2.1). The data showed that the expression of *FGF2* was significantly up-regulated in CRC tissues compared to matched normal tissues (p=0.002). Our data is in accordance with what was already described in the literature regarding the increase of *FGF2* expression in various tumor tissues, such as lung (Li et al. 2018), colorectal (George et al. 2002), bladder (Gazzaniga et al. 1999) and prostate (Soulitzis et al. 2006).

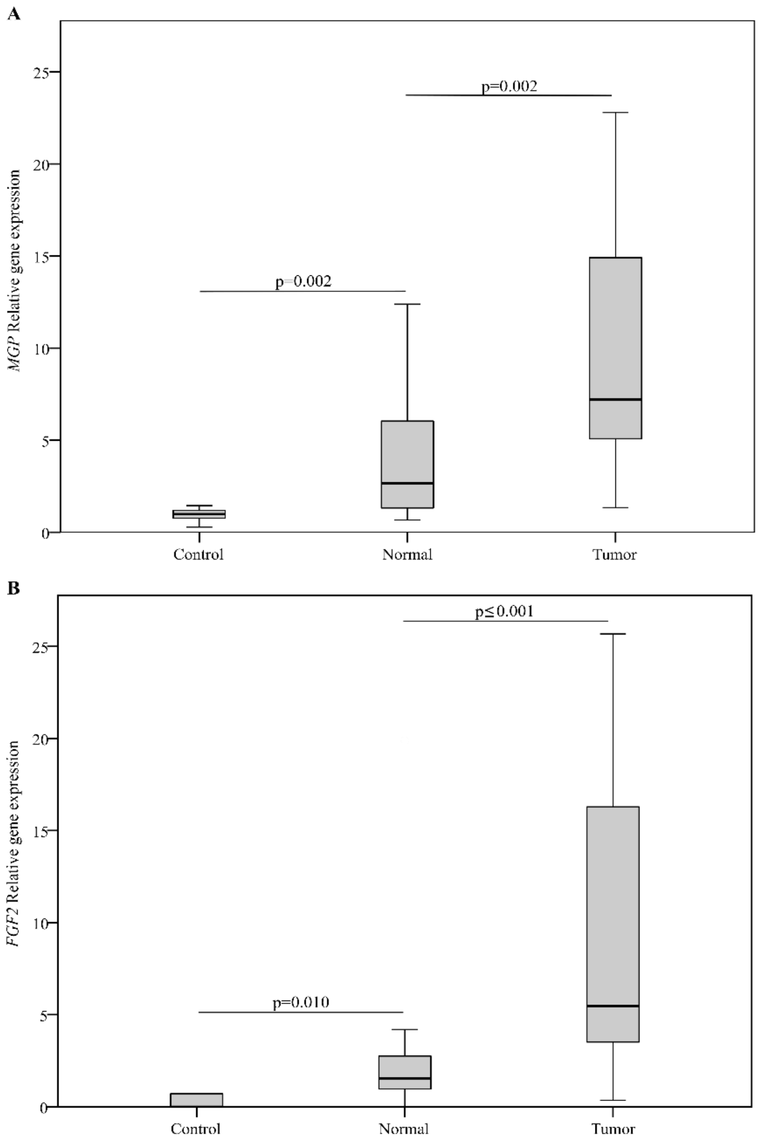


Figure 1 Relative MGP and FGF2 gene expression in samples from patients with colon adenocarcinoma Relative *MGP* (A) and *FGF2* (B) gene expression levels were analyzed by RT-qPCR in a total of 9 samples from control group and 23 samples from colorectal cancer tissue (normal and tumor mucosa). The latter showed higher mRNA levels of *MGP* and *FGF2* than non-tumor tissues (MGP $p=0.002$; FGF2 $p \leq 0.001$). Values are presented as mean ± SD. The Mann-Whitney and Kruskal Wallis non parametric tests were performed for the statistical analysis.

To evaluate if there is a correlation between *FGF2* expression and the clinical-pathological features of the patients, we analyzed all the variables shown in Tables 2.2.1 and 2.2.2. No statistically significant associations were found between *FGF2* expression and the clinical and pathological features of the patients.

Table 2.2.1 Demographic features of colorectal patients

	MGP (n=23)			*FGF2* (n=23)		
Characteristics	Number (%)	Mean value of fold change	*p* value	Number (%)	Mean value of fold change	*p* value
Gender			**0.033**			0.439
Male	14 (61)	3.135		14 (61)	2.000	
Female	9 (39)	6.648		9 (39)	1.000	
Age (median: 71,70 years)			0.548			0.776
<72	8 (35)	2.898		8 (35)	4.437	
≥72	15 (65)	5.369		15 (65)	5.640	
Familial Cancer History			0.671			0.579
Yes	7 (30)	3.034		7 (30)	3.495	
No	16 (70)	5.155		16 (70)	5.977	
Previous Pathologies			0.691			1.000
Yes	18 (78)	3.732		18 (78)	5.460	
No	5 (22)	7.308		5 (22)	4.363	
Metastasis			0.177			0.812
Yes	6 (26)	8.082		6 (26)	5.445	
No	17 (74)	3.249		17 (74)	5.143	

Mann-Whitney U test

Table 2.2.2 Histopathological features of patients

	MGP (n=23)			*FGF2* (n=23)		
Characteristics	Number (%)	Mean value of fold change	*p* value	Number (%)	Mean value of fold change	*p* value
Tumor Location			0.618			0.493
Rectum	12 (52)	4.672		12 (52)	3.967	
Rectosigmoid Junction	3 (13)	6.217		3 (13)	2.479	
Ascending Colon	2 (9)	2.633		2 (9)	10.730	
Sigmoid	1 (4)	8.004		1 (4)	3.653	
Cecum	2 (9)	2.793		2 (9)	14.938	
Hepatic Angle	3 (13)			3 (13)		
Tumor Histology			0.196			0.655
Well Differentiated	10 (44)	4.014		10 (44)	3.400	
Moderately Differentiate	9 (39)	2.164		9 (39)	7.867	
Poorly Differentiated	1 (4)	24.042		1 (4)	5.530	
Mucinous	1 (4)	8.004		1 (4)	3.653	
Mucinous well Differentiate	2 (9)	6.028		2 (9)	3.054	
Tumor Stage			0.201			0.336
I - II	9 (39)	3.155		9 (39)	3.017	
III - IV	14 (61)	5.380		14 (61)	6.639	
T classification			0.815			0.447
pT2	4 (18)	3.983		4 (18)	1.866	
pT3	18 (78)	4.763		18 (78)	5.918	
pT4	1 (4)	2.055		1 (4)	6.109	
N classification			0.372			0.592
N0	9 (39)	3.155		9 (39)	3.017	
N1	8 (35)	5.626		8 (35)	6.717	
N2	6 (26)	5.053		6 (26)	6.536	
M classification			0.227			0.745
M0	18 (78)	3.294		18 (78)	5.505	
M1	5 (22)	8.884		5 (22)	4.201	
Hepatic Metastasis			0.227			0.745
Yes	5 (22)	8.884		5 (22)	4.201	
No	18 (78)	3.294		18 (78)	5.505	
Pulmonary Metastasis			0.158			0.198
Yes	2 (9)	14.057		2 (9)	8.597	
No	21 (91)	3.600		21 (91)	4.900	
KRAS mutations			0.728			0.265
Yes	8 (35)	4.022		8 (35)	7.826	
No	15 (65)	4.770		15 (65)	3.833	

Mann-Whitney U test

We then evaluated the correlation between *FGF2* and *MGP* expression. *FGF2* mRNA expression determined by qRT-PCR was well correlated (r=0.572, p=0.004) with that determined for *MGP* (H. Caiado et al. 2020) (Figure 2.2.2).

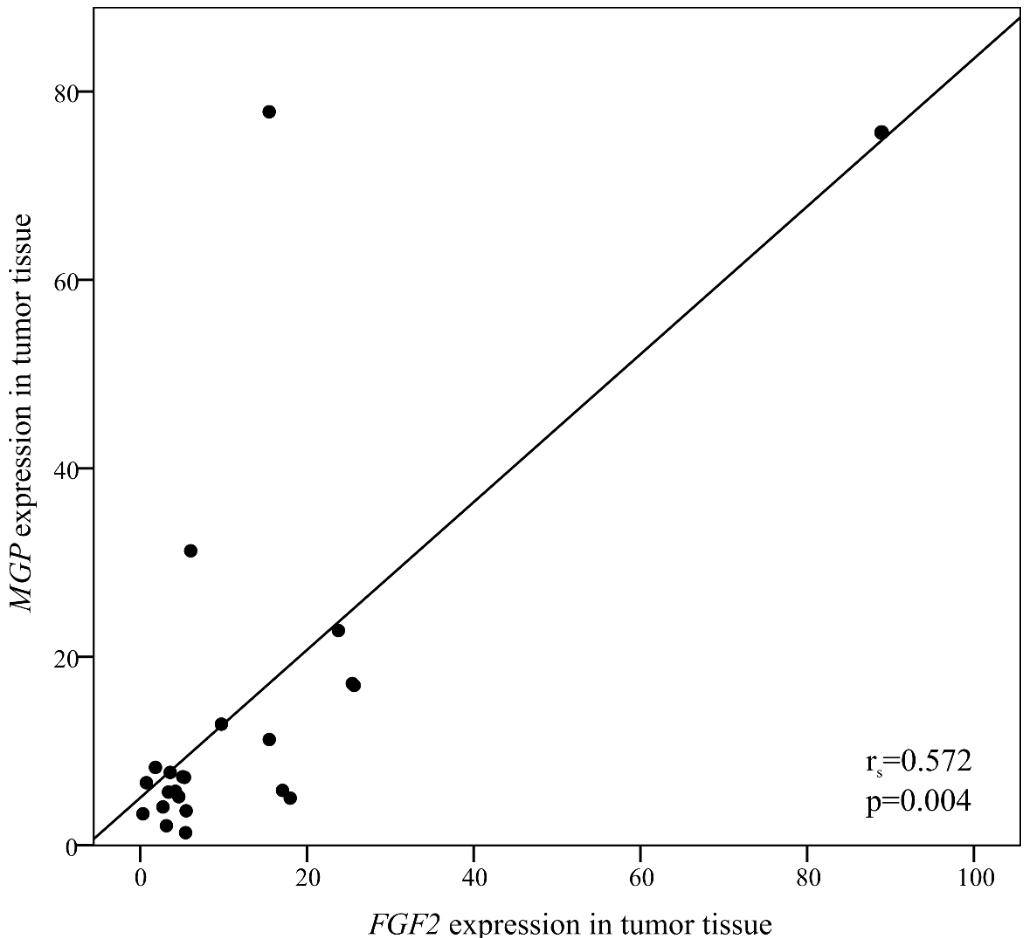


Figure 2.2.2 Correlation between FGF2 and MGP gene expression in tumor tissue
As described in experimental design in materials and methods, the correlation between *MGP* and *FGF2* gene expression was evaluated through the SPSS software, applying the Spearman coefficient correlation test in the tumor tissue and establishing a positive and significant correlation between expression of both genes (r=0.572; p=0.004).

In our previously published study, we found that the two-step cluster analysis of the CRC samples allowed differentiating patients with a better or worse survival outcome (Caiado et al. 2020). Subsequently, we performed a multivariate classification of two step clusters (Chiu et al. 2001) to determine possible patient profiles, taking into account the characteristics of categorical and numerical variables (Table 2.2.3). This type of analysis allows the exploitation of data taking into account each variable independently from each other's, to try to identify

homogeneous groups depending on their characteristics. Since we did not find any correlation between the high expression of *FGF2* and the overall patient survival rate (Figure 2.2.3), we then evaluated the prognostic value of different variables to differentiate patients in different groups according to the influence of these factors.

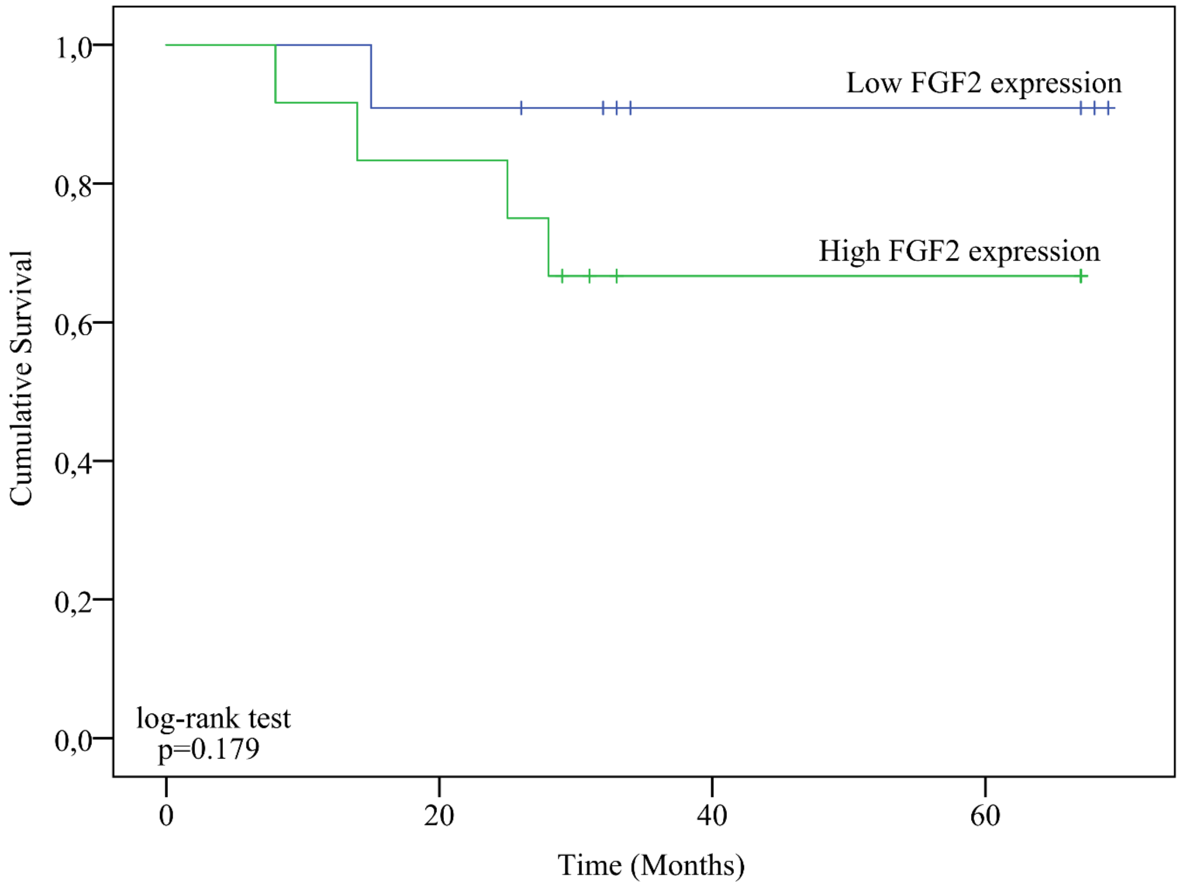


Figure 3 Overall survival curve of patients with overexpression of FGF2
Patients with high *FGF2* gene expression appear to have a lower survival rate although this was not statistically significant (p=0.179). Small vertical lines indicate the censored cases referring to the number of patients that have not reached the terminal event during the data collection. *p*-value was calculated by log-rank test.

The variables considered were: T classification, N classification, tumor staging, gender, deceased, fold change MGP categorized, fold change FGF2 categorized, fold change MGP, fold change FGF2, tumor histology, KRAS mutations, tumor location, survival rate (months), polyposis and stroke. According to this analysis, patients were divided into clusters 1 and 2. Patients in cluster 1 presented a stage N0 of lymph node metastasis (50%), the tumor was either in stage II (33.3%) or stage III (44.4%), mostly male (72.2%), with low MGP (72.2%) and FGF2 (55.6%) levels of expression, with a fold change for MGP of 3.09 (±3.03) and for FGF2 of 4.89 (±6.81), with a tumor histology showing either a moderately (44.4%) or well

differentiated tumor (44.4%), without mutation on KRAS (61.1%), with a T3 classification (72.2%), with a mean survival time of 49.61 (±18.6) months, with the tumor mostly located in rectum (38.8%) and without the presence of polyposis (88.9) and no stroke (88.9%). Patients in cluster 2 presented a stage N1 of lymph node metastasis (60%), the tumor was either in stage III (20%) or stage IV (80%), mostly female (80%), with high MGP (100%) and FGF2 (80%) levels of expression, with a fold change for MGP of 9.61 (±8.4) and for FGF2 of 6.38 (±5.0), with a well differentiated tumor histology (40%), without mutation on KRAS (80%), with a T3 classification (100%), with a mean survival time of 18.00 (±8.2) months, with the tumor located in rectum (100%) and without the presence of polyposis (100%) and no stroke (100%).

Moreover, we performed a Kaplan-Meier survival analysis to assess if MGP and FGF2 could be in fact good prognostic factors in terms of overall survival rate for the two groups of patients found in the two-step cluster analysis. Patients in cluster 2, which presented a worst prognosis, had a higher mortality rate when compared with patients in cluster 1 (log-rank test $p \leq 0.001$) (Figure 2.2.4).

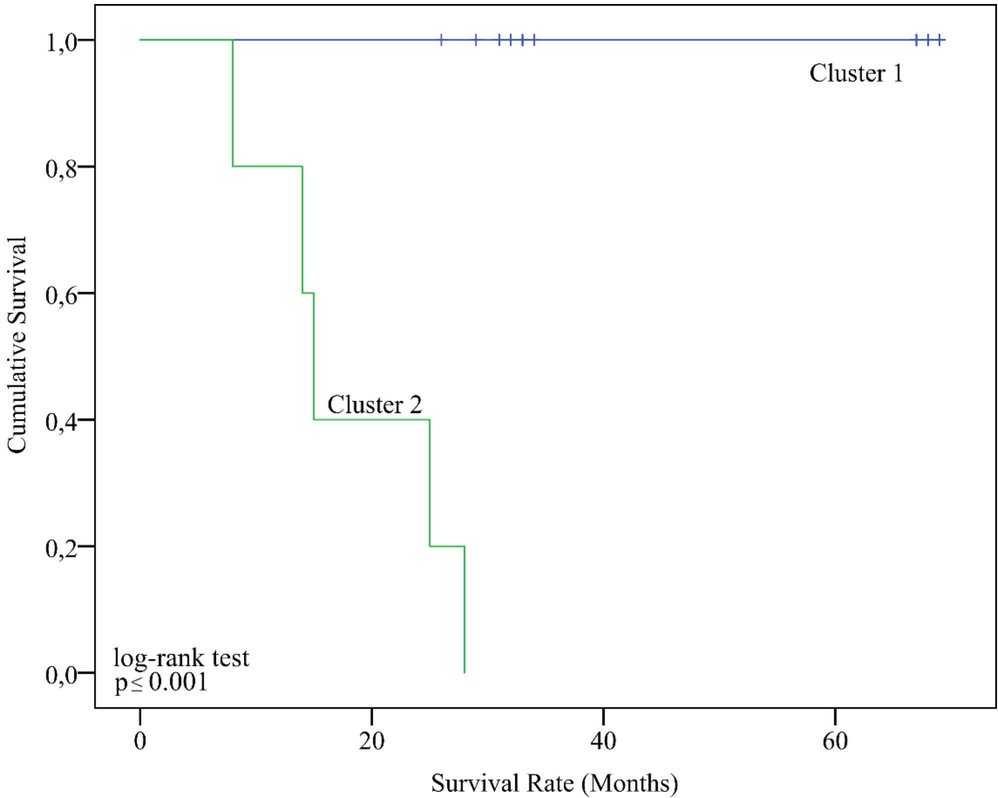


Figure 2.2.4 Overall survival curve for patients categorized by clusters 1 and 2
Patients in Cluster 1 present a better survival rate, when compared with patients in cluster 2, who have a lower survival rate and a worse prognosis.
Small vertical lines indicate the censored cases referring to the number of patients that have not reached the terminal event during data collection. *p*-value was calculated by log-rank test.

From the analysis it was perceived that patient in cluster 2 had a worst prognosis, in the way that all of these patients presented a small survival rate, and higher tumor stages when compared with patients in cluster 1. It's also worthy of note, that the variables that significantly contributed to the division of the patients were the tumor staging, the presence of high level of MGP, gender and the survival rate. This means that, per se, the high levels of FGF2 alone are not sufficient for the clustering of patients, but in combination with other multiple variables can profile the patients into groups with a better or worst prognosis.

Despite the presence of some patients in cluster 1 presenting a T staging of 3 or even 4, this does not mean that these patients will actually have an associated worst prognosis. In fact, it was already shown in the literature that patients who presented a tumor stage III could have a better prognosis than those with a tumor stage II. For example, according to the American Joint Committee (AJCC) staging manual (Weiser, 2018), when TNM staging is being evaluated, the clinicians have to take into account the tumor size (T), the number of lymph node metastasis, and the presence of metastasis. The stage is then categorized according to the combination of those three major factors, but the prognosis of the disease is reflected by its combination with other external variables that may also contribute to a worst and better prognosis. The conclusion from this analysis shows that it is the combination of the multiple variables analyzed, together with the high expression of FGF2 in tumor tissue, that can differentiate patients in two groups associated to a better or worst prognosis.

Table 2.2.3 Multivariate analysis of predictor factors

Characteristics	Cluster 1 (n=18, %)	Cluster 2 (n=5, %)	*p* value
N Classification			*p*=0.126[1]
N0	9 (50)	0 (0)	
N1	5 (27.8)	3 (60)	
N2	4 (22.2)	2 (40)	
Tumor Staging			***p*=0.05[1]**
Stage I	3 (16.7)	0 (0)	
Stage II	6 (33.3)	0 (0)	
Stage III	8 (44.4)	1 (20)	
Stage IV	1 (5.6)	4 (80)	
Gender			***p*=0.05[1]**
Male	13 (72.2)	1 (20)	
Female	5 (27.8)	4 (80)	
Deceased			***p*=0.05[1]**
No	18 (100)	0 (0)	
Yes	0 (0)	5 (100)	
Fold change *MGP* categorized			***p*=0.05[1]**
High *MGP*	5 (27.8)	5 (100)	
Fold change *RUNX2* categorized High *RUNX2*	7 (38.9)	5 (100)	*p*=0.016[1]
Fold change *FGF2* categorized High *FGF2*	8 (44.4)	4 (80)	*p*=0.159[1]
Fold Change *MGP*, mean (SD[2])	3.09(±3.03)	9.61(±8.4)	***p*=0.05[3]**
Fold change *FGF2*, mean (SD[2])	4.89(±6.81)	6.38(±5.00)	*p*=0.403[3]
MGP vs *FGF2*[5]	r=0.373; p=0.128	r=-0.200; p=0.747	
Tumor Histology			*p*=0.246[1]
Well differentiated	8 (44.4)	2 (40)	
Moderately differentiated	8 (44.4)	1 (20)	
Poorly Differentiated	0 (0)	1 (20)	
Mucinous	1 (5.6)	0 (0)	
Mucinous well differentiated	1 (5.6)	1 (20)	
KRAS mutations			*p*=0.433[1]
No	11 (61.1)	4 (80)	
T classification			*p*=0.412[1]
T1	0 (0)	0 (0)	
T2	4 (22.2)	0 (0)	
T3	13 (72.2)	5 (100)	
T4	1 (5.6)	0 (0)	
Survival Rate (Months), mean (SD[2])	49.61(±18.6)	18.00(±8.2)	***p*=0.05[4]**
Tumor Location			*p*=0.320[1]
Rectum	7 (38.8)	5 (100)	
Rectosigmoid junction	3 (16.7)	0 (0)	
Ascending colon	2 (11.1)	0 (0)	
Sigmoid	1 (5.6)	0 (0)	
Cecum	2 (11.1)	0 (0)	
Hepatic angle	3 (16.7)	0 (0)	
Polyposis			*p*=0.435[1]
No	16 (88.9)	5 (100)	
Stroke			*p*=0.435[1]
No	16 (88.9)	5 (100)	

[1] Chi Square test
[2] Standard Deviation
[3]Mann-Whitney test

[4]Log Rank test
[5]Spearman coefficient correlation test
Boldfaced values - Variables with $p \leq 0.05$

2.2.4 Experimental Design, Materials, and Methods

In this report we present briefly the materials and methods used to obtain the data here described. To see a more detailed material and methods, please refer to (Caiado et al. 2020).

2.2.4.1 Clinical, demographic and pathological characteristics of patients

Tissue samples, as well as clinical and pathological information, were obtained as described in the research article "Evaluation of MGP Gene Expression in Colorectal Cancer".
Clinical, demographic and histopathological information regarding patients is depicted in tables 2.2.1 and 2.2.2.

2.2.4.2 RT-qPCR

Total RNA was extracted from fresh biopsies stored in RNALater (CRC (n=23) including normal adjacent tissue and healthy colonic tissue (n=9)). After quality and quantity measurements, cDNA synthesis was performed using 1 µg of the extracted RNA treated with RQ1 DNase (1U per µg of RNA; Promega) and M-MLV reverse transcriptase (ThermoFisher Scientific) according to manufacturer's instructions.
The expression of mRNA for *FGF2* was analyzed by $2^{-\Delta\Delta Ct}$ method and normalized with the expression of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) as reference gene. Primer sequences for *GAPDH* and *FGF2* were as follows:
GAPGH: forward: 5'-TCAACGGATTTGGTCGTATTGGGCG-3' and reverse: 5'-CTCGCTCCTGGAAGATGGTGATGGG-3';
FGF2: forward: 5'-CAAAAACGGGGGCTTCTTCCTG-3' and reverse: 5'-CCATCTTCCTTCATAGCCAGGTAACG-3'.
Data were presented as the relative quantity of target mRNA normalized with *GAPDH* and relative to the mean expression of the control group. Please refer to the research article "Evaluation of MGP Gene Expression in Colorectal Cancer" for the analyses of expression of mRNA for *MGP* (Caiado et al., 2020).

2.2.4.3 Statistical analysis

Statistical analysis was performed using SPSS software program version 25. Values for gene expression are presented as mean and standard deviation (SD) and two-sided *P* value less than 0.05 was defined as statistically significant. Fold changes presented correspond to the ratio of the values from tumor mucosa *versus* normal mucosa. Comparisons between group variables and gene expression were estimated using non parametric statistical tests: Mann–Whitney U and Kruskal–Wallis.

The cutoff value to distinguish the patients with low and high *MGP* and *FGF2* levels were estimated taking into account the median value of the fold change for both *MGP* and *FGF2*.

A multivariate classification of two step clusters (Chiu et al. 2001) was performed to determine possible patient profiles, taking into account the characteristics of categorical and numerical variables (Table 2.2.3). This allowed the formation of cluster 1 (n=18) and cluster 2 (n=5). Spearman coefficients were considered to analyze the correlation between *MGP* and *FGF2* fold change values by the interest groups, namely, clusters and tissue samples. Overall survival probability for two groups of patients (clusters 1 and 2) was calculated using the Kaplan–Meier method; intergroup differences were determined using a log-rank test. Logistic regression analysis and χ^2 analysis was used to evaluate the independent influence of factors on the final prognosis.

Chapter 2.3

Assessment of MGP gene expression in cancer and contribution to prognosis

Helena Caiado[1,2,3]

M. Leonor Cancela[2,3,4*]

Natércia Conceição[2,3,4*]

This sub-chapter was submitted as a research manuscript to Biochimie journal

[1]ProRegeM PhD Programme in Mechanisms of disease and regenerative medicine, University of Algarve, Faro 8005-139, Portugal
[2]Center of Marine Sciences (CCMAR), University of Algarve, Faro 8005-139, Portugal
[3]Faculty of Biomedical Sciences and Medicine, University of Algarve, Faro 8005-139, Portugal
*Correspondence to: Natércia Conceição (nconcei@ualg.pt) and M. Leonor Cancela (lcancela@ualg.pt)

Author's roles: Study design: NC and MLC. Study conduct: All authors; Data collection: HC and NC. Data analysis: All authors. Data interpretation: All authors. Drafting manuscript: All authors. Revising manuscript content: All authors. Approving final version of manuscript: All authors.

2.3.1 Abstract

Matrix Gla protein (*MGP*) was first identified as a calcification physiological inhibitor and the causal agent of the Keutel syndrome. *MGP* has been suggested to play a role in development, cell differentiation, and tumorigenesis. This study aimed to compare *MGP* expression and methylation status in different tumors and adjacent tissues, using The Cancer Genome Atlas (TCGA) data repository. We investigated if changes in *MGP* mRNA expression were correlated to cancer progression and whether the correlation coefficients could be used for prognosis. Strong correlations were observed between altered *MGP* levels and disease progression in breast, kidney, liver, and thyroid cancers, suggesting that they could serve as a supplement to current clinical biomarkers, for early cancer diagnosis.

We have also analyzed *MGP* methylation and identified CpG sites in its promoter and first intron with clear differences in methylation status between healthy and tumoral tissue providing evidence for epigenetic regulation of *MGP* transcription. Furthermore, we demonstrate that these alterations correlate with the overall survival of the patients suggesting that its assessment can serve as an independent prognostic indicator of patients' survival.

2.3.2 Introduction

Matrix Gla Protein (MGP) has been mainly recognized as a physiological inhibitor of calcification and the causal gene for the Keutel syndrome, a condition characterized by abnormal calcium deposition, mostly in the cartilage and vascular system (Munroe et al. 1999; Schurgers et al. 2008). However, over the past years, the study of this gene has gained additional relevance related to implications for tumor progression (Gheorghe and Crăciun, 2015).

Regarding its function in carcinogenesis, *MGP* gene expression may be a tumor-type dependent factor since it has been shown to present a negatively correlated with tumor progression and metastasis in some tumors, such as renal and prostate carcinoma (Levedakou et al. 1992), while its up-regulation in breast tumors and glioblastomas is associated with tumor progression and poor prognosis .

Epigenetic events, such as DNA methylation, histone modification, and miRNA have been progressively more associated with cancer, being identified as a possible cause for the activation or silencing of certain genes depending on the context (Jones and Baylin, 2002). In fact, in the literature, there is some evidence that *MGP* gene expression might be regulated by epigenetic events, such as DNA methylation (Tuo and Ye, 2017) and microRNAs (Tiago et al. 2016).

The identification of cancer-specific biomarkers can provide a way to distinguish pathologic from normal samples. Therefore, uncovering their roles in tumorigenesis could contribute to understanding how genetic and epigenetic events affect the incidence of cancer and thus uncover novel therapeutic targets.

In this study, we aimed to evaluate *MGP* expression and its DNA methylation status in tumoral tissue as compared to normal tissue using the cancer genome atlas (TCGA) database. *MGP* expression was found to be downregulated in tumoral tissue compared to the adjacent normal tissue ($p \leq 0.001$). In addition, three of the four CpG sites located in the promoter and first intron of the *MGP* gene showed clear differences in methylation status between healthy and tumor tissue indicating clear deregulation in cancer and strengthening the hypothesis that *MGP* plays a role during cancer progression/proliferation. Altogether, our data provide evidence for possible regulation of the *MGP* gene by epigenetic events and suggest that changes in *MGP* expression could be used as an independent prognostic factor for the overall survival of the patients.

2.3.2 Material and methods

2.3.2.1 Sample and data collection

Cohort studies from eleven distinct tumor types were downloaded from The Cancer Genome Atlas (TCGA) Genomic Data Commons (GDC) portal (http://xena.ucsc.edu/, accessed between 2017 to 2021), including data from cancerous tissue and normal samples (TCGA code in parenthesis), namely, bile duct (CHOL), bladder (BLCA), breast (BRCA), colorectal (colon cancer (COAD) and rectal cancer (READ)), esophageal (ESCA), head and neck (HNSC), kidney (kidney chromophobe (KICH); kidney clear cell carcinoma (KIRC) and kidney papillary cell carcinoma (KIRP)), liver (LIHC), lung (lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)), prostate (PRAD), and thyroid (THCA). For simplicity, hereafter, we will refer to these tumor types according to the tissue of origin (bile duct, bladder, breast, colorectal, esophageal, head and neck, kidney, liver, lung, prostate, and thyroid cancer). Clinical information includes age, gender, pathologic stage, and survival time. The standard clinical and pathological variables considered were the Tumor stage (from I to IV) and the TNM system (T stage, N stage, and M stage). The Tumor stage indicates the extension of cancer and whether it has spread outside the position of origin. Regarding the TNM system, T indicates the size and extent of the tumor, the N stage indicates the presence of lymph node metastases, and the M stage provides the presence or absence of metastasis of primary tumor spread to other parts of the body.

Patients with missing or incomplete data for the inclusion criteria for subsequent analysis were excluded, having resulted in 6002 tumoral samples and 634 adjacent normal tissues in total, distributed from all the eleven cohort studies, unless otherwise stated. Data regarding clinicopathological variables from patients were collected from the selected tumor cohorts and the respective number of patients analyzed per tumor is depicted in their respective figures or figure legend. All data used in this study were downloaded from TCGA following TCGA publication guidelines and data access policies, thus, additional approval by the local Ethics Committee was not needed.

To confirm the expression of MGP in the normal and tumoral tissues, immunohistochemistry was evaluated by the Human Protein Atlas (http://www.proteinatlas.org).

2.3.2.2 Gene expression data analysis

Data including raw HTSeq-counts for *MGP* gene expression were downloaded collectively with clinical information at the same time as the tumor cohorts selection from the UCSC Xena

browser (TCGA) data portal (https://xenabrowser.net/, accessed between 2017 to 2021) to explore the potential gene expression profiles between tumoral tissue and adjacent normal tissue samples across the eleven selected cohorts and evaluate possible correlations with clinicopathological data.

Data processing regarding raw HTSeq-counts gene expression was conducted according to the Genomic Data Commons Data portal (GCD hub hosting TCGA data, https://portal.gdc.cancer.gov/) and is publicly available for consultation at https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/.

2.3.2.3 Survival analysis

To explore the impact of *MGP* gene expression on the Overall Survival (OS) of patients from the eleven cohorts analyzed, *MGP* gene expression cut-offs for each tumor were established through Receiver Operating Characteristic (ROC) Curve (AUC) analysis (Greiner et al. 2000) to explore the potential to discriminate tumor from normal samples. Values above and below the cut-offs were considered as high and low *MGP* gene expression. The Kaplan-Meir, univariate and multivariate Cox regression analyses were conducted to explore the effect of high and low *MGP* expression combined with the clinical factors on survival.

2.3.2.4 Analysis of DNA methylation

To evaluate the impact of DNA methylation in the regulation of *MGP* expression, this information was retrieved collectively for all the patients across the eleven tumor cohorts, between 2017 to 2021.

DNA methylation beta values collected resulted from the readings based on intensities at known CpG sites through the Illumina Infinium HumanMethylation450K array and processed according to the Genomic Data Commons Data portal (GCD hub hosting TCGA data, https://portal.gdc.cancer.gov/) guidelines, publicly available at https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Methylation_Pipeline/.

The location of the six identified CpG probes retrieved from the TCGA for the *MGP* gene, as well as the presence of histone marks and chromatin organization located at the sites of interest were explored using the UCSC genome browser (http://genome.ucsc.edu/, accessed between 2017 to 2021, GRCh38/HG38).

Patients with missing methylation β-values were discarded from the analysis, and the effective number of patients per tumor used for the analysis is depicted in the respective figure and figure legend.
The average of methylation β values for each CpG site was compared for each tumor database and was considered differentially methylated when the *p*-value was equal to or lower than 0.05.

2.3.2.5 Statistical analysis

Comparisons between group variables, gene expression, and DNA methylation were estimated using non-parametric statistical tests: Mann–Whitney U and Kruskal–Wallis. Values for gene expression are presented as the median values of *MGP* gene expression in the tumoral tissues. Correlation analysis between *MGP* expression and DNA methylation was performed by the non-parametric Spearman correlation coefficient test. A two-sided *p*-value less than 0.05 was demarcated as statistically significant.
Overall survival plot was calculated using the Kaplan–Meier method; intergroup differences of high and low *MGP* expression were determined using a log-rank test and the univariate and multivariate Cox regression analysis was used to evaluate the independent influence of factors on the final prognosis.
All the statistical analysis was performed using SPSS software program version 26 (IBM, 2010, Chicago, IL, USA).

2.3.3 Results

2.3.3.1 *MGP* is differentially expressed between cancer and normal tissue samples

Basic information from 632 patients with paired tumors and normal samples, from 11 different types of cancer from TCGA was collected and used for analysis. It contained 9 samples for bile duct, 16 for bladder, 112 for breast, 48 for colorectal, 10 for esophageal, 43 for head and neck, 126 for kidney, 50 for liver, 108 for lung, 52 for prostate, and 58 for thyroid. Compared with the control group, *MGP* was significantly downregulated in the bladder ($P \leq 0.001$), breast ($P \leq 0.001$), colorectal ($P \leq 0.001$), head and neck ($P \leq 0.001$), kidney ($P \leq 0.001$), lung ($P \leq 0.001$) and thyroid ($P \leq 0.001$) cancer tissues. In contrast, the expression of *MGP* was upregulated in the bile duct (P=0.002) and liver (P=0.011) cancer tissues compared with normal tissues. No significantly different *MGP* expression was observed for esophageal and prostate, between the two groups (Figure 2.3.1).

Similar results were obtained using a different approach taking into consideration all samples. Indeed, of the 11 types of cancer collected from TCGA, a total of 6002 tumor samples were available while only 634 normal samples were provided. By integrating all these samples, we observed that compared with the control group, *MGP* was significantly downregulated in the bladder ($P \leq 0.001$), breast ($P \leq 0.001$), colorectal ($P \leq 0.001$), head and neck ($P \leq 0.001$), kidney ($P \leq 0.001$), lung ($P \leq 0.001$) and thyroid ($P \leq 0.001$) cancer tissues. The expression of *MGP* was upregulated in the bile duct ($P \leq 0.01$) and liver ($P \leq 0.01$) cancer tissues compared with normal tissues. No significantly different *MGP* expression was observed for esophageal and prostate, between the two groups (Figure 2.3.2).

Altogether, the data analyzed, irrespective of the strategy used, indicated the mRNA expression of *MGP* was abnormally expressed across different cancer types.

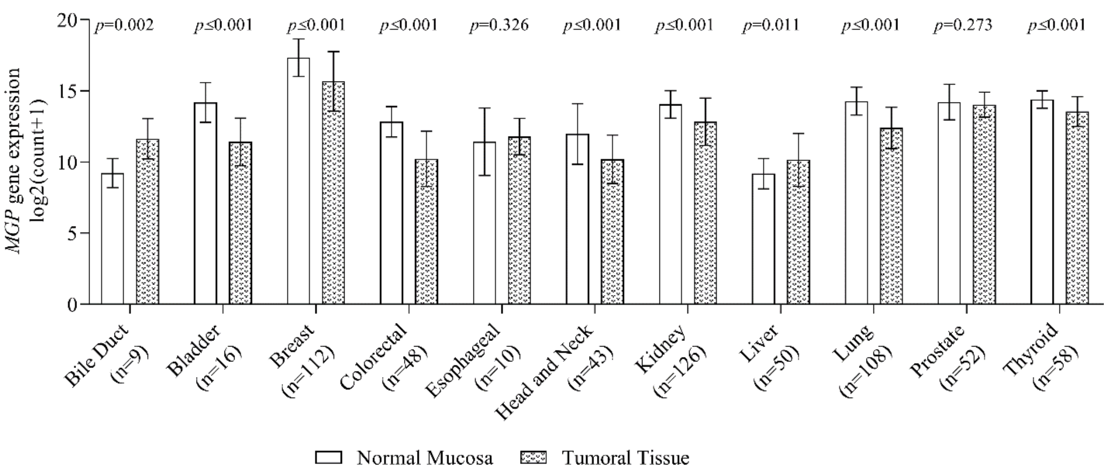


Figure 2.3.1 Analysis of *MGP* gene expression in paired tumor-normal samples from TCGA databases. *MGP* expression was upregulated in the normal tissues compared to the corresponding tumor tissue, except for bile duct (p=0.002) and liver (p=0.011) cancers which presented a significant upregulation in the tissue of the primary tumor. Esophageal (p=0.326) and prostate (p=0.273) cancers did not present significant differences in *MGP* expression between normal and tumoral tissues. Mean ± SD, Mann-Whitney statistic test, the p-value was considered statistically different ($p \leq 0.05$).

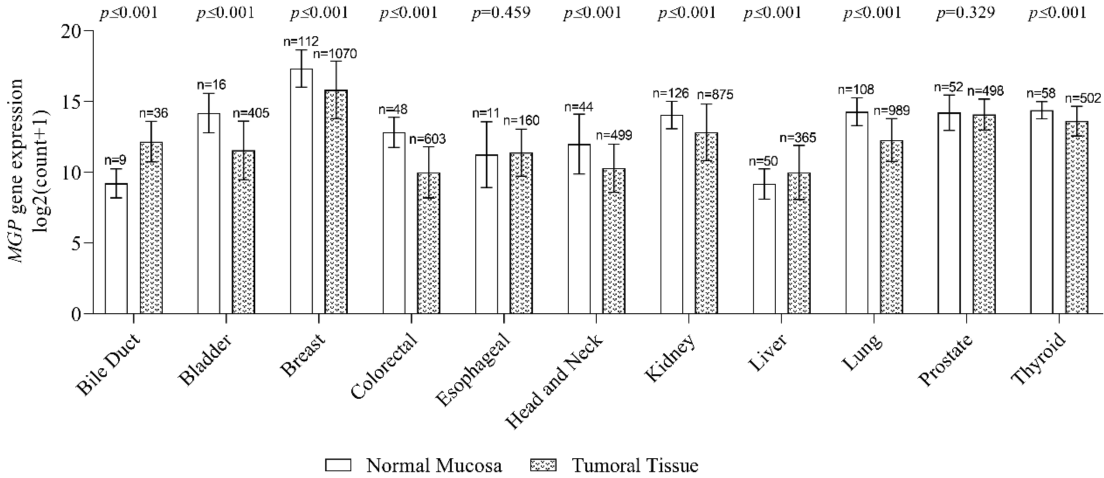


Figure 2.3.2 Analysis of MGP gene expression in normal and tumoral samples from TCGA databases. MGP expression was upregulated in the normal tissues compared to the corresponding tumor tissue, except for bile duct and liver cancers which presented a significant upregulation in the tissue of the primary tumor (p≤0.001). Esophageal (p=0.459), and prostate (p=0.329) cancers did not present significant differences in MGP expression between normal and tumoral tissues. Mean ± SD, Mann-Whitney statistic test, the p-value was considered statistically different (p≤0.05).

We then analyzed the 6002 cancer samples from TCGA to investigate the relationship between the expression level of *MGP* and clinicopathologic variables. The related results are shown in Tables 2.3.1 and 2.3.2. As shown in Table 2.3.1, higher expression levels of *MGP* were observed in female patients with breast (P=0.031) and lung (P≤0.001) cancers and in patients older than 70 years with bladder (P=0.035), breast (P=0.009), colorectal (P=0.008), lung (P=0.042), and thyroid (P=0.038) cancers (Table 2.3.1).

The expression of *MGP* was significantly correlated with the clinical stage (stage I-II versus stage III-IV) in the bladder (P≤0.001), colorectal (P≤0.001), kidney (P≤0.001), and prostate (P≤0.001) cancers. The expression of *MGP* was also significantly correlated with the T stage in the bladder (P≤0.001), breast (P=0.035), colorectal (P≤0.001), kidney (P≤0.001), lung (P≤0.001), and prostate (P≤0.001) cancers; N stage in the bladder (P≤0.001), colorectal (P≤0.001), head and neck (p=0.002), kidney (P=0.002), lung (P≤0.001), and prostate (P=0.006) cancers; and M stage in the bladder (P≤0.001), breast (P=0.035), colorectal (P=0.003), kidney (P≤0.001), and thyroid (P<0.001) cancers (Table 2.3.2). Taken together, these results suggested that the expression of *MGP* increased gradually in the development of the tumors. MGP is correlated with lymph node metastases and high TNM stage (Table 2.3.2), further suggesting that MGP may act as a biomarker of poor prognosis for several tumors.

Table 2.3.1 Demographic features of patients in the different types of tumors

	Gender					Age				
	Male		Female			≤70		>70		
Characteristics	Number (n)	Median value	Number (n)	Median value	*p* value	Number (n)	Median value	Number (n)	Median value	*p* value
Bile Duct	20	12.26	16	12.33	*p*=0.849	23	12.17	13	12.28	*p*=0.633
Bladder	299	11.34	106	11.44	*p*=0.897	239	11.15	182	12.04	***p*=0.035**
Breast	12	14.77	1058	15.99	***p*=0.031**	874	15.90	196	16.55	***p*=0.009**
Colorectal	327	9.71	276	9.75	*p*=0.696	357	9.89	246	9.46	***p*=0.008**
Esophageal	137	11.10	23	11.33	*p*=0.774	114	11.03	46	11.78	*p*=0.52
Head and Neck	366	10.48	133	10.27	*p*=0.352	395	10.39	104	10.48	*p*=0.688
Kidney	591	13.08	284	13.30	*p*=0.092	674	13.24	201	12.99	*p*=0.083
Liver	246	9.96	119	10.17	*p*=0.235	292	10.03	73	9.80	*p*=0.938
Lung	590	12.13	399	12.77	***p*≤0.001**	640	12.24	349	12.60	***p*=0.042**
Prostate	498	14.11	-	-	-	461	14.11	37	14.12	*p*=0.630
Thyroid	135	13.65	367	13.72	*p*=0.285	462	13.72	40	13.30	***p*=0.038**

Mann-Whitney U test
[a]Median value of *MGP* expression in tumoral mucosa

Table 2.3.2 Pathological characteristics of TCGA tumors cohorts

Pathological Characteristics	Bile Duct			Bladder			Breast			Colorectal			Esophageal			Head and Neck		
	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value
Tumor Stage			p=0,447			**p≤0,001**			p=0,247			**p≤0,001**			p=0,135			p=0,154
I - II	28	12,26		135	10,77		801	15,85		333	9,46		90	10,98		103	9,92	
III - IV	8	12,63		286	12,15		269	15,75		270	9,95		70	11,42		396	10,44	
T stage			p=0,100			**p≤0,001**			**p=0,035**			**p≤0,001**			p=0,394			p=0,457
TX				1	8,61					1	13,89		2	10,95		3	11,86	
Tis										1	9,35							
T0				1	8,17													
T1	19	12,28		4	10,21		40	16,13		20	9,26		27	10,92		44	10,53	
T1a							1	16,92										
T1b							16	14,89										
T1c							221	16,26										
T2	6	12,14		58	10,63		619	15,94		106	9,32		39	11,07		149	10,51	
T2a	2	10,18		34	10,59		1	18,83										
T2b	4	12,32		60	11,04		1	14,84										
T2c																		
T3	5	13,4		45	11,73		132	16,16		410	9,78		87	11,4		118	10,25	
T3a							1	17,37										
T3b				83	12,63													
T3c																		
T4				11	12,43		8	14,43		29	9,6		4	11,28		12	9,59	
T4a				45	12,56					26	12,17		1	15,95		169	10,41	
T4b				5	13,14		27	15,77		10	10,01					4	10,55	
T4c																		
T4d							3	12,67										

Mann-Whitney U test
[a]Median value of MGP expression in tumoral tissue

Table 2.3.2 Pathological characteristics of TCGA tumors cohorts (continued)

Pathological Characteristics	Bile Duct			Bladder			Breast			Colorectal			Esophageal			Head and Neck		
	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value
N stage			p=0,701			**p≤0,001**			p=0,195			**p≤0,001**			p=0,336			**p=0,002**
NX	5	12,25		42	10,16		13	16,47		3	12,42		6	10,6		13	10,21	
N0	28	12,19		246	11,25		320	16,11		343	9,45		67	10,94		205	9,84	
N0(i-)							156	15,97										
N0(i+)							28	16,88										
N0(mol+)							1	17,14										
N1	5	13,40		46	11,94		122	15,78		105	9,69		71	11,28		74	10,77	
N1mi							36	16,25										
N1a							168	15,94		8	11,43							
N1b							32	16,54		21	10,95							
N1c							1	16,92										
N2				80	12,64		56	15,27		79	9,78		10	11,38		14	10,37	
N2a							64	15,61		11	10,97					12	11,53	
N2b										20	12,44					120	10,66	
N2c																52	11,05	
N3				7	13,19		26	15,96					6	12,91		9	9,62	
N3a							43	16,18										
N3b							3	16,82										
N3c							1	18,16										
M stage			p=0,406			**p≤0,001**			**p=0,035**			**p=0,003**			p=0,291			p=0,073
MX	3	12,46		204	11,08		156	15,72		69	10,23		17	11,71		48	9,83	
M0	28	12		206	11,02		888	16,05		449	9,64		129	11,22		448	10,48	
M0(i+)							6	15,32										
M1	5	13,1		11	12,56		22	14,74		71	9,78		7	10,68		3	9,38	
M1a										11	11,96		6	10,98				
M1b										3	9,71		1	8,66				
M1c																		

Mann-Whitney U test
[a]Median value of *MGP* expression in tumoral tissue

Table 2.3.2 Pathological characteristics of TCGA tumors cohorts (continued)

Pathological Characteristics	Kidney			Liver			Lung			Prostate			Thyroid		
	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value
Tumor Stage			**p≤0,001**			p=0,475			p=0,174			**p≤0,001**			p=0,723
I - II	582	12,89		271	10,03		791	12,39		192	13,84		335	13,75	
III - IV	293	13,74		94	9,99		198	12,24		306	14,33		167	13,59	
T stage			**p≤0,001**			p=0,502			**p≤0,001**			**p≤0,001**			p=0,069
TX				3	10,47		3	13,94		6	14,60		2	14,04	
Tis															
T0															
T1	47	12,02		180	9,96		112	12,73					44	13,59	
T1a	250	13,26					71	13,01					19	13,91	
T1b	183	12,93					96	12,81					80	13,9	
T1c										1	15,26				
T2	92	12,62		89	10,03		322	12,81					164	13,55	
T2a	19	12,6		1	12,56		169	12,37		13	13,42				
T2b	15	11,50		1	11,38		56	11,59		10	13,79				
T2c										166	13,84				
T3	14	13,62		44	9,97		116	12					170	13,72	
T3a	171	13,75		28	10,16					158	14,09				
T3b	66	13,81		6	8,33					134	14,58				
T3c	3	10,82													
T4	15	14,01		13	9,91		41	12,21		10	14,53		9	13,94	
T4a													14	13,33	
T4b															
T4c															
T4d															

Mann-Whitney U test
[a]Median value of MGP expression in tumoral tissue

Table 2.3.2 Pathological characteristics of TCGA tumors cohorts (continued)

Pathological Characteristics	Kidney			Liver			Lung			Prostate			Thyroid		
	n	Median value[a]	p-value	n	Median value[a]	p-value	n	Median value[a]	p-value	n	Median value[a]	p-value	n	Median value[a]	p-value
N stage			**p=0,002**			p=0,691			**p≤0,001**			**p=0,006**			p=0,419
NX	500	12,98		113	10,03		18	13,57		73	13,77		50	13,67	
N0	326	13,49		248	10,01		638	12,37		347	14,11		229	13,68	
N0(i-)															
N0(i+)															
N0(mol+)															
N1	43	13,31		4	9,77		218	12,24		78	14,5		58	13,6	
N1mi															
N1a													90	13,69	
N1b													75	13,9	
N1c															
N2	6	13,92					108	12,32							
N2a															
N2b															
N2c															
N3							7	13,19							
N3a															
N3b															
N3c															
M stage			**p≤0,001**			p=0,176			p=0,795			p=0,137			**p≤0,001**
MX	208	11,05		99	10,14		227	12,34		41	13,82		211	13,52	
M0	578	13,53		263	9,95		731	12,38		454	14,12		282	13,84	
M0(i+)															
M1	89	13,88		3	13,02		22	12,47					9	13,06	
M1a							3	12,38		1	16,23				
M1b							6	11,64		1	9,7				
M1c										1	13,41				

Mann-Whitney U test
[a]Median value of MGP expression in tumoral tissue

To explore further the observed *MGP* expression associated with a more advanced tumor stage, we performed the same analysis with paired samples from tumoral and normal tissue from the same patient, across all tumors. Due to the limited number of paired samples per tumor, we only performed this analysis in cancers with more than 45 patients. We collected paired tumoral and normal tissue from 112 breast, 48 colorectal, 126 kidney, 50 liver, 108 lung, and 58 thyroid cancer patients (Supplementary Materials Table S2.3.1). Our results showed higher *MGP* expression levels in male patients with colorectal cancer (P=0.034), and in female patients with lung cancer (P=0.031) (Supplementary Materials Table S2.3.1). The expression of *MGP* was significantly correlated with the clinical stage (stage I-II versus stage III-IV) only in prostate cancer (P≤0.03). The expression of *MGP* was also significantly correlated with the N stage in prostate cancer (P=0.043) and the M stage in kidney cancer (P=0.002) (Supplementary Materials Table S2.3.2).

2.3.3.2 High *MGP* expression is associated with a shorter Overall Survival

To explore the relationship between *MGP* expression and the Overall Survival of patients, we first applied the Receiver Operating Characteristic curve (AUC) model to the levels of *MGP* gene expression in the tumors presenting a statistically significant difference between normal and tumoral tissues, to predict the corresponding optimal cutoff, for each tumor. Through this model it is possible to determine the values of the Area Under the Curve (AUC) with sensitivity and specificity, which can be used as a diagnostic test tool to distinguish between samples, where AUC values of 0.5 are non-informative, values between 0.5 and ≤0.7 are less accurate, values between 0.7 and ≥0.8 are considered moderately accurate, and values between 0.9 and 1 are highly accurate. Next, to perform the Kaplan-Meier curves, we selected those tumors that presented an AUC ≥0.7 and characterized as low *MGP* expression for those below the optimal cutoff value, and high *MGP* expression for those above the optimal cutoff value (Supplementary figure S2.3.1 A-I). As shown in Figure 2.3.3 (A-I), the overall survival of patients presenting high levels of *MGP* expression was significantly shorter than for patients presenting low levels, including in breast (p=0.018), bladder (p=0.017); colorectal (p=0.005), lung (p=0.048), head and neck (p=0.036), and kidney (p=0.03), whereas for bile duct (p=0.219), liver (p=0.054) and thyroid (p=0.296), despite presenting AUC values considered to be moderately accurate, it was not possible to differentiate the patients by normal and tumoral mucosa, as well as to establish a correlation between *MGP* expression and overall survival (Figure 2.3.3 A-I and Supplementary Materials Figure S2.3.1 A-I). These results suggest that high *MGP* expression might be a good indicator of poor prognosis in most of the analyzed tumors.

Since we were able to predict a survival association with high *MGP* expression, a univariate cox regression model was performed individually to evaluate if *MGP* expression, as well as age groups (age less or equal than 70, and age above 70 years), tumor stage (I-II/III-IV), T stage, N stage, and M stage, could have the potential to be good independent predictor factors for patient’s prognosis for each cancer.

Results displayed in Table 2.3.3 show the significant potential of the queried factors associated with the overall survival, especially for the categorized *MGP* expression in breast (p=0.019), bladder (p=0.018), colorectal (p=0.006), head and neck (p=0.037) and kidney (p=0.004) cancers, as seen with the Kaplan-Meir curves. In liver (p=0.055) and lung (p=0.060) the level of significance is higher than 0.05 therefore we cannot conclude if *MGP* expression might be a potential prognosis factor for these two cancers. Also, in the bile duct (p=0.236) and thyroid

(p=0.317), *MGP* expression did not show a significant correlation with the overall survival as an individual prognostic factor (Table 2.3.3).

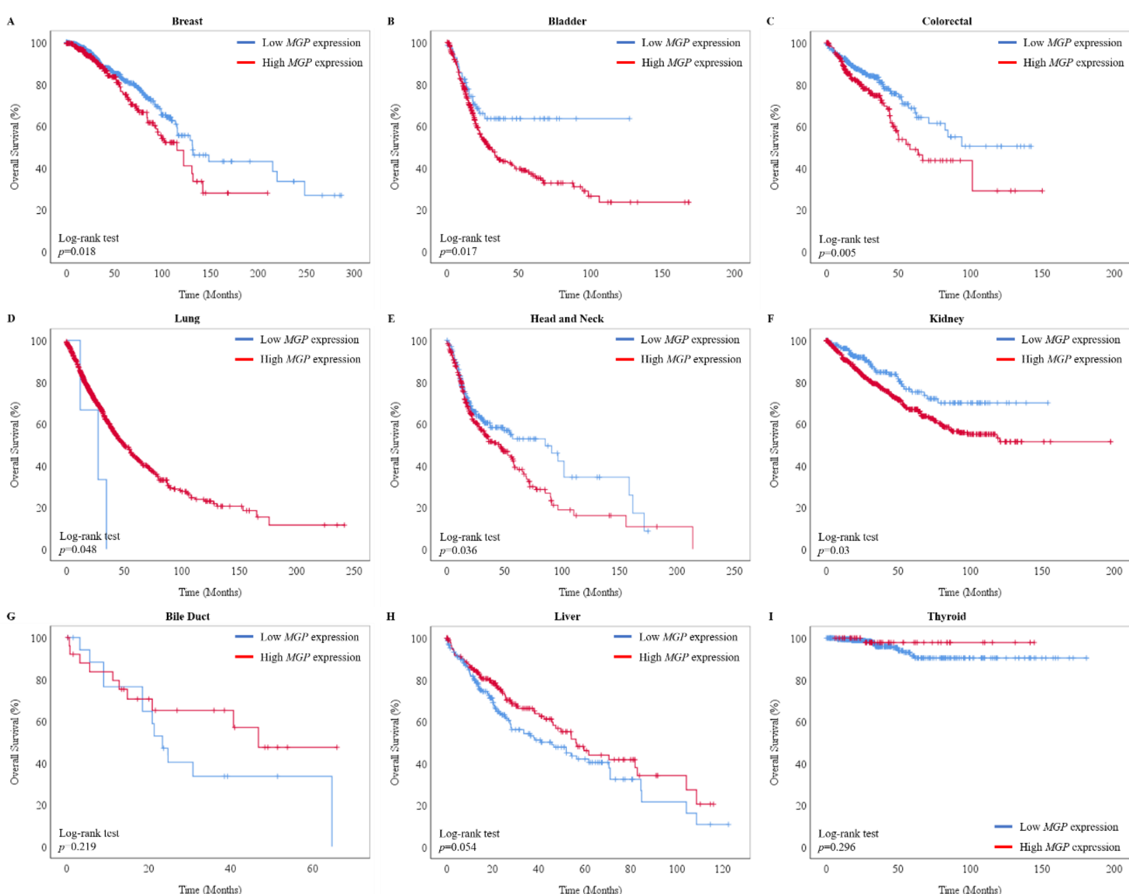


Figure 2.2.3 Overall Survival (OS) of *MGP* gene expression in the different types of TCGA tumor cohorts. Most tumors presented a poor overall survival rate when *MGP* expression was high. Optimal cutoff values into low *MGP* expression (blue line) and high *MGP* expression (red line) were establish by ROC curve analysis. Log-rank test considered statistically significant when $p \leq 0.005$

Table 2.3.3 Univariate Cox regression analysis of clinical factors and overall survival

Clinical factors	Breast				Bladder				Colorectal			
		95% CI				95% CI				95% CI		
	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value
Age (≤70/>70)	2,615	1,916	3,569	**0,001**	1,473	1,107	1,959	**0,008**	1,904	1,359	2,667	**0,001**
Gender	0,580	0,081	4,143	0,587	0,871	0,636	1,193	0,390	1,048	0,749	1,466	0,785
Tumor Stage (I-II/III-IV)	2,189	1,636	2,929	**0,001**	2,161	1,513	3,087	**0,001**	3,261	2,271	4,682	**0,001**
T stage	1,145	1,073	1,222	**0,001**	1,170	1,095	1,251	**0,001**	1,666	1,389	1,999	**0,001**
N stage	1,097	1,056	1,139	**0,001**	1,193	1,084	1,314	**0,001**	1,241	1,161	1,327	**0,001**
M stage	1,854	1,460	2,354	**0,001**	1,148	0,994	1,326	0,610	1,847	1,419	2,404	**0,001**
MGP (High/Low)	1,414	1,059	1,887	**0,019**	1,710	1,096	2,669	**0,018**	1,603	1,148	2,239	**0,006**
Clinical factors	**Lung**				**Head and Neck**				**Kidney**			
		95% CI				**95% CI**				**95% CI**		
	HR	**Lower**	**Higher**	**p-value**	**HR**	**Lower**	**Higher**	**p-value**	**HR**	**Lower**	**Higher**	**p-value**
Age (≤70/>70)	1,377	1,141	1,662	**0,001**	1,784	1,352	2,253	**0,001**	1,989	1,535	2,577	**0,001**
Gender	0,795	0,656	0,965	**0,020**	0,747	0,572	0,975	**0,032**	1,096	0,847	1,416	0,486
Tumor Stage (I-II/III-IV)	1,892	1,539	2,326	**0,001**	1,662	1,204	2,293	**0,002**	4,706	3,623	6,112	**0,001**
T stage	1,13	1,075	1,187	**0,001**	1,131	1,038	1,232	**0,005**	1,259	1,207	1,314	**0,001**
N stage	1,356	1,205	1,525	**0,001**	1,157	1,089	1,230	**0,001**	1,817	1,517	2,175	**0,001**
M stage	1,197	0,986	1,454	0,069	1,271	0,760	2,123	0,361	3,944	3,120	4,987	**0,001**
MGP (High/Low)	0,336	0,108	1,047	0,060	1,315	1,017	1,701	**0,037**	1,672	1,180	2,369	**0,004**
Clinical factors	**Bile Duct**				**Liver**				**Thyroid**			
		95% CI				**95% CI**				**95% CI**		
	HR	**Lower**	**Higher**	**p-value**	**HR**	**Lower**	**Higher**	**p-value**	**HR**	**Lower**	**Higher**	**p-value**
Age (≤70/<70)	1,437	0,600	3,439	0,416	1,444	1,039	2,005	**0,029**	10,862	4,491	26,269	**0,001**
Gender	1,306	0,569	2,994	0,529	0,959	0,700	1,314	0,796	2,087	0,85	5,121	0,108
Tumor Stage (I-II/III-IV)	1,701	0,657	4,406	0,274	2,246	1,645	3,068	**0,001**	5,824	2,232	15,192	**0,001**
T stage	1,149	0,881	1,499	0,306	1,217	1,139	1,301	**0,001**	1,515	1,039	2,209	**0,031**
N stage	3,160	1,848	5,404	**0,001**	0,719	0,523	0,990	**0,043**	0,968	0,658	1,424	0,868
M stage	2,177	1,198	3,958	**0,011**	0,689	0,495	0,958	**0,027**	2,672	1,15	6,206	**0,022**
MGP (High/Low)	0,604	0,262	1,391	0,236	0,741	0,544	1,007	0,055	0,358	0,048	2,679	0,317

HR – Hazard Ratio; CI – Confidence Interval; $p \leq 0,05$ was considered statistically significant and values are presented in bold

We further included all the possible predictor factors in the multivariate cox regression model to establish the prognostic value of *MGP* expression combined with the clinical factors. This analysis further showed that high *MGP* expression was an independent factor for the overall survival in breast (p=0.010, HR=1.479, 95%CI: 1.096-1.995), colorectal (p=0.018, HR=1.519, 95% CI: 1.073-2.149), while the low *MGP* expression was an independent prognostic factor in

liver (p=0.048, HR=0.732, 95% CI= 0.538-0.997) and lung (p=0.019, HR=0.250, 95% CI=0.078-0.797) (Table 2.3.4). Altogether, data suggest that *MGP* expression could be a good prognostic factor for the overall survival of the patients.

Table 2.3.4 Multivariate Cox regression analysis of clinical factors and overall survival

Clinical factors	Breast				Bladder				Colorectal			
		95% CI				95% CI				95% CI		
	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value
Age (≤70/>70/)	2,570	1,879	3,515	**0,001**	1,474	1,105	1,967	**0,008**	2,539	1,784	3,614	**0,001**
Gender	0,581	0,081	4,180	0,589	0,802	0,584	1,103	0,175	1,057	0,752	1,486	0,750
Tumor Stage (I-II/III-IV)	1,160	0,709	1,900	0,554	1,395	0,855	2,274	0,182	2,507	1,538	4,088	**0,001**
T stage	1,067	0,987	1,153	0,103	1,134	1,034	1,245	**0,008**	1,458	1,177	1,806	**0,001**
N stage	1,061	1,006	1,120	**0,029**	1,284	1,153	1,430	**0,001**	1,025	0,930	1,130	0,618
M stage	1,663	1,306	2,117	**0,001**	1,079	0,932	1,250	0,310	1,419	1,101	1,829	**0,007**
MGP (High/Low)	1,479	1,096	1,995	**0,010**	1,489	0,945	2,346	0,086	1,519	1,073	2,149	**0,018**
Clinical factors	Lung				Head and Neck				Kidney			
		95% CI				95% CI				95% CI		
	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value
Age (≤70/>70/)	1,427	1,181	1,724	**0,001**	1,898	1,424	2,530	**0,001**	2,195	1,686	2,858	**0,001**
Gender	0,829	0,68	1,01	0,062	0,761	0,578	1,004	0,053	0,970	0,746	1,261	0,819
Tumor Stage (I-II/III-IV)	1,309	0,96	1,785	0,089	1,162	0,753	1,795	0,498	2,702	1,563	4,671	**0,001**
T stage	1,081	1,025	1,14	**0,004**	1,075	0,968	1,194	0,176	1,010	0,925	1,102	0,824
N stage	1,204	1,025	1,415	**0,024**	1,146	1,068	1,231	**0,001**	1,464	1,220	1,756	**0,001**
M stage	1,061	0,864	1,302	0,573	1,186	0,712	1,978	0,512	2,564	1,971	3,336	**0,001**
MGP (High/Low)	0,250	0,078	0,797	**0,019**	1,181	0,907	0,907	0,217	0,931	0,650	1,334	0,698
Clinical factors	Bile Duct				Liver				Thyroid			
		95% CI				95% CI				95% CI		
	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value	HR	Lower	Higher	p-value
Age (≤70/>70/)	6,566	1,770	24,361	**0,005**	1,366	0,971	1,920	0,073	7,306	2,709	19,703	**0,001**
Gender	1,524	0,571	4,068	0,400	0,970	0,704	1,338	0,855	1,86	0,704	4,91	0,210
Tumor Stage (I-II/III-IV)	1,357	0,170	10,827	0,773	1,069	0,469	2,438	0,874	3,754	1,093	12,892	**0,036**
T stage	0,990	0,551	1,778	0,974	1,202	1,011	1,431	**0,038**	1,006	0,695	1,455	0,974
N stage	4,868	2,171	10,914	**0,001**	0,950	0,608	1,482	0,820	0,806	0,517	1,255	0,339
M stage	2,499	0,983	6,357	0,054	0,698	0,450	1,084	0,110	2,508	1,166	5,396	**0,019**
MGP (High/Low)	0,635	0,238	1,691	0,364	0,732	0,538	0,997	**0,048**	0,283	0,037	2,158	0,223

HR – Hazard Ratio; CI – Confidence Interval; $p \leq$ 0,05 was considered statistically significant and values are presented in bold

2.3.3.3 Divergent expression patterns of *MGP* mRNA and MGP protein in cancer and normal tissues

The results of stains on normal cancer tissues demonstrated that the MGP was only highly expressed in tumoral breast tissues (Figure S2.3.2). Concerning normal samples, the staining was not detected, and the intensity was negative. Among 49 tumor tissue specimens tested, the MGP protein (tumor tissue vs. normal tissue) was detected in 17 (34.7 %) tumors (Figure S2.3.2).

2.3.3.4 *MGP* CpG sites present different DNA methylation patterns across different cancer types

Since DNA methylation levels in promoter CpG sites are known to be closely correlated with gene expression, we examined if *MGP* had promoter CpG sites and whether they were hypo- or hyper-methylated in the 11 cancers under study. First, we compared the methylation values of the samples of adjacent normal tissue and the tumor samples and we found that *MGP* was significantly hypomethylated in the bladder ($P=0.003$), colorectal ($P\leq 0.001$), liver ($P\leq 0.001$), and thyroid ($P\leq0.001$) tumoral tissues compared with non-tumor tissues and it was significantly hypermethylated in breast ($P\leq0.005$), kidney ($P\leq0.001$), lung ($P\leq0.001$), and prostate ($P\leq0.001$) tumoral tissues compared with non-tumor tissues (Table 2.3.5).

The *MGP* gene was significantly less expressed and hypermethylated in breast, kidney, and lung tumor tissues (Figure 2.3.2 and Table 2.3.5).

Furthermore, we identified the location of the six CpG sites along the *MGP* gene, four of them located in the promoter region (cg16617257, cg13302154, cg22221831, and cg00431549), one located in the 5’ UTR, at the beginning of the transcription start site according with NCBI sequence (NG_023331.2) (cg06601891) and one located in the first intron (cg05360958) (Figure 2.3.4).

Table2.3.5 Overall methylation levels of *MGP* CpG probes across TCGA cohorts

	Mean β-value			
Tumors	**Normal Tissue**	**Tumoral Tissue**	**|Δβ|**	**p-value**
Bile Duct	,6150	,5232	-0,092	0,182
Bladder	,2838	,2523	-0,032	**0,003**
Breast	,1671	0,223367	0,056	**0,005**
Colorectal	,4957	,3954	-0,100	**≤0,001**
Esophageal	,2787	,3312	0,053	0,203
Head and Neck	,3420	,3489	0,007	0,579
Kidney	,4749	,6343	0,159	**≤0,001**
Liver	,6013	,4192	-0,182	**≤0,001**
Lung	,3190	0,398651	0,080	**≤0,001**
Prostate	,1878	,3421	0,154	**≤0,001**
Thyroid	,3024	,2422	-0,060	**≤0,001**

Mann-Whitney U test

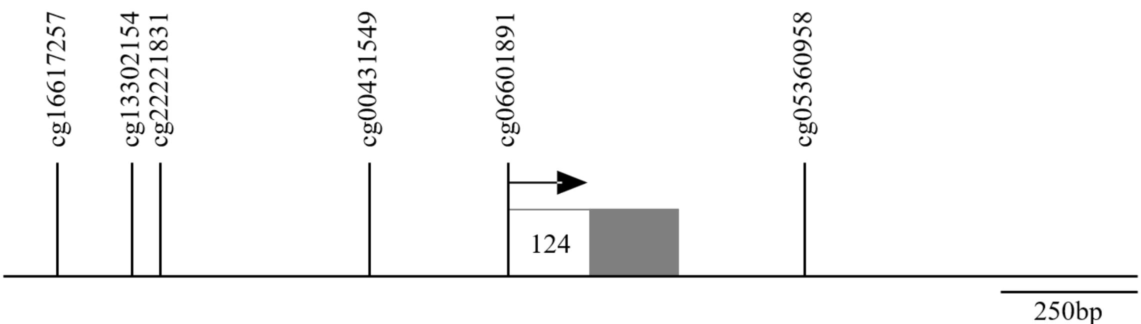


Figure 2.3.4 CpG methylation sites of human MGP gene.
Lines represent introns and solid boxes represent exons. The transcription start site is indicated by an arrow. Gray and white areas indicate exonic coding and non-coding regions, respectively. The localization of the CpGs methylation sites located in the promoter region and the first intron on the *MGP* gene was identified through the TCGA, UCSC Genome Browser, and NCBI databases.

We further explored the methylation levels for the identified CpG sites in the selected cancers databases, except for the CpG's cg16617257 and cg06601891 since the methylation β- values were not available from the TCGA databases.

Results showed that, depending on the type of cancer, the four CpG sites presented different methylation status (hypomethylation vs hypermethylation).

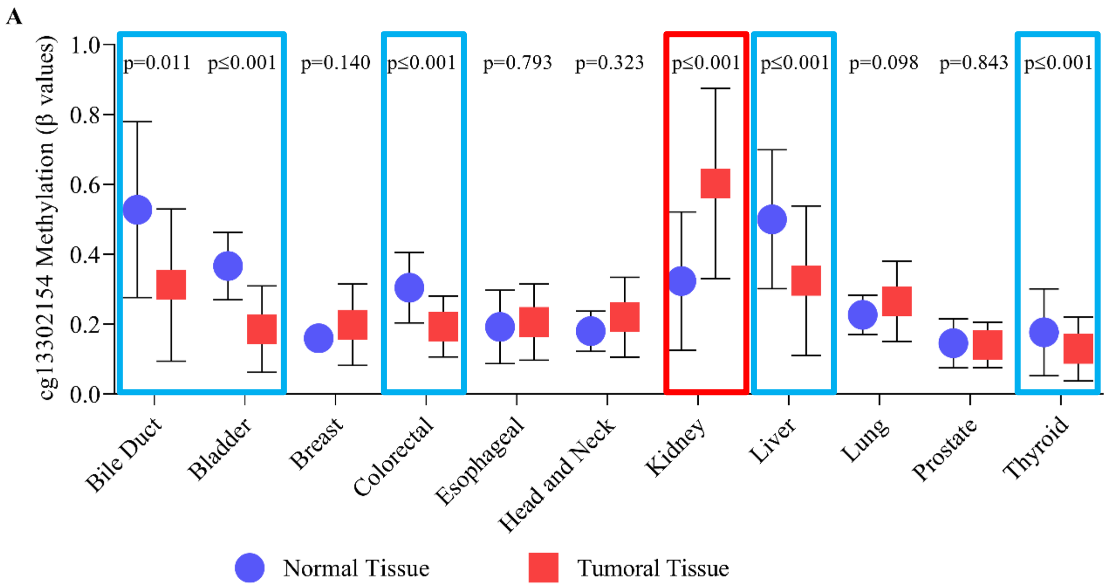
The CpG cg13302154 site was significantly hypomethylated in the bile (p=0.011), bladder (*p*≤0.001), colorectal (p≤0.001), liver (*p*≤0.001), and thyroid (*p*≤0.001) cancers (highlighted in blue), while in the kidney this CpG site was shown to be hypermethylated (*p*≤0.001) (highlighted in red) in the tumoral tissues. In the breast, esophageal, head and neck, lung and prostate there were no significant differences in the methylation levels for this CpG site between normal and tumoral tissue (Figure 2.3.5A).

The CpG site cg2221831 presented a significantly hypomethylated state in the tumoral tissues of the bladder (*p*≤0.001), colorectal (*p*≤0.001), liver (*p*=0.002), and thyroid (*p*=0.007) cancers (highlighted in blue), whereas in the breast (p=0.01), head and neck (*p*=0.003), kidney (*p*≤0.001), and lung (*p*=0.030) cancers it displayed a hypermethylated state (figure 2.3.5B) (highlighted in red) in the tumoral tissues. In the bile duct, esophageal, and prostate there were no significant differences in the methylation levels for this CpG site between normal and tumoral tissues (Figure 2.3.5B).

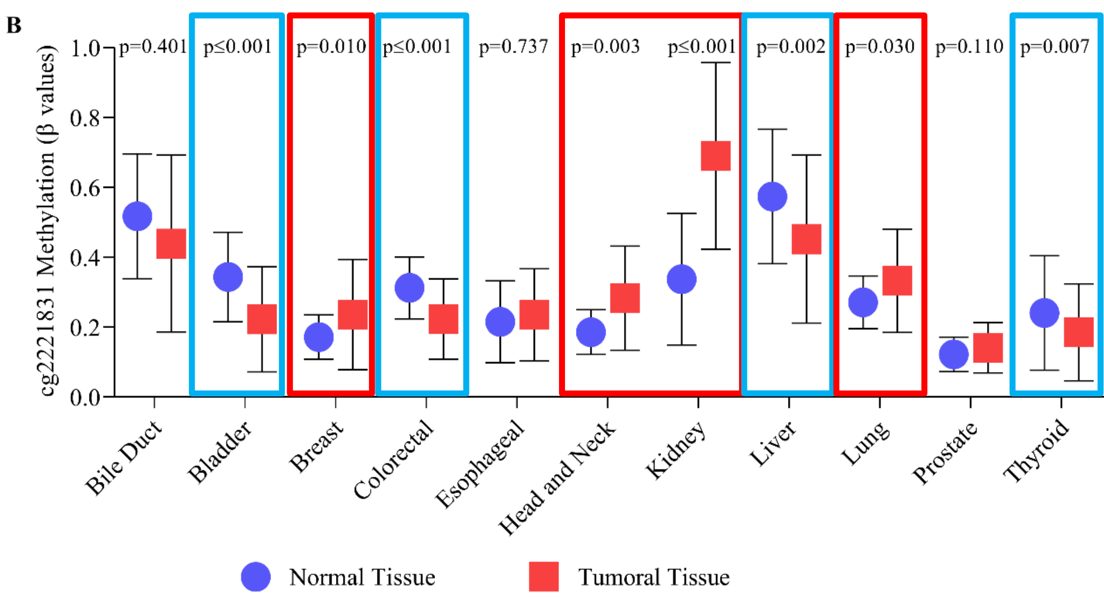
The CpG cg00431549 showed a significant hypomethylation status in the liver (*p*=0.002), thyroid (p=0.041) tumoral tissues (highlighted in blue), whereas in the breast (*p*≤0.001), esophageal (p=0.042), lung (*p*≤0.001) and prostate (*p*≤0.001) cancers this CpG site was found to be hypermethylated (highlighted in red) in the tumoral tissue. For the remaining cancers, no significant differences were observed (figure 2.3.5C).

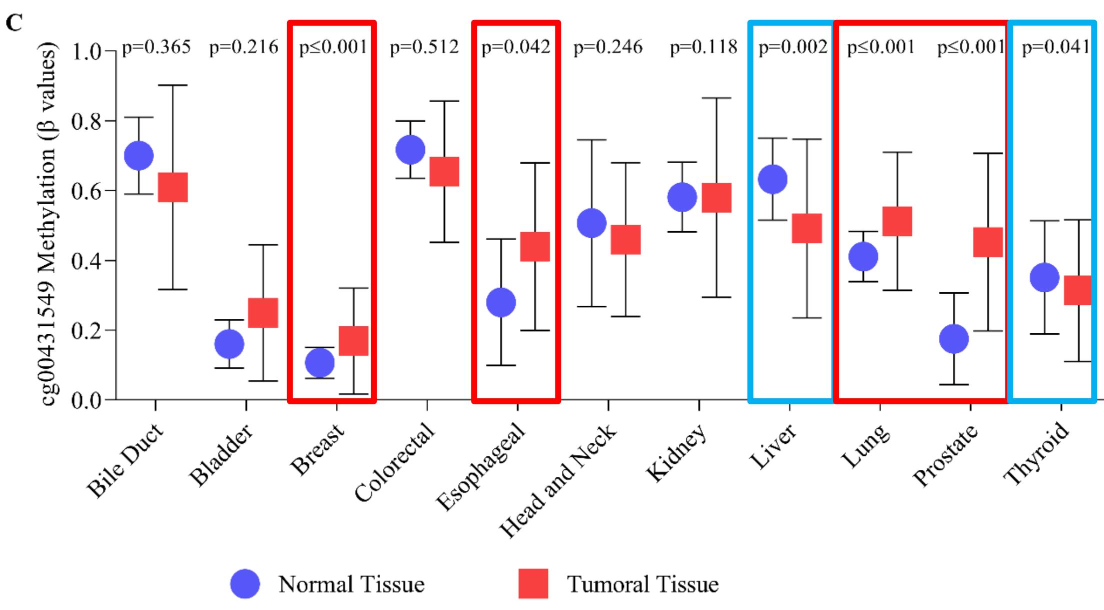
The CpG cg05360958 site, located in the first intron of the *MGP* gene, was found to be hypomethylated in colorectal (*p*=0.002), liver (*p*≤0.001), and thyroid (*p*≤0.001) tumoral tissues (highlighted in blue), while it was hypermethylated in the kidney (*p*=0.018), lung (*p*≤0.001) and prostate (*p*≤0.001) tumoral tissues (highlighted in red) (Figure 2.3.5D). For the remaining tissues, no significant differences were observed (figure 2.3.5D).

A



B





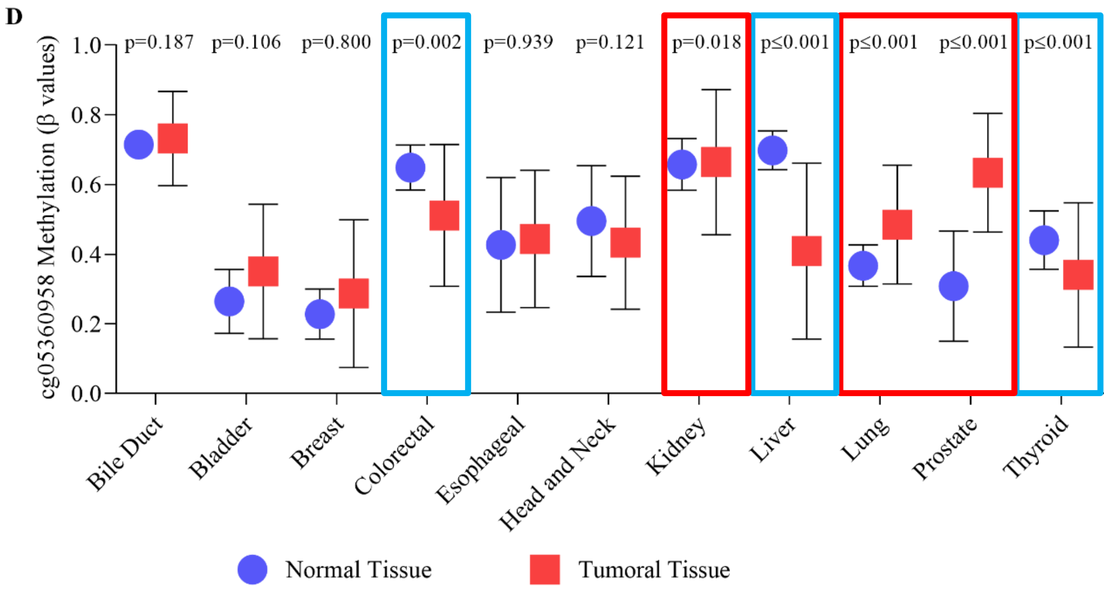


Figure 2.3.5 **Differentially methylated *MGP* CpG sites in the TCGA cancer cohorts. A**. Probe cg13302154; **B**. Probe cg2221831; **C**. Probe cg00431549; **D**. Probe cg05360958. Beta values (β, methylation values) were compared between normal and tumor samples. CpG sites were considered as differentially methylated whenever the p-value≤0.05 (Mann–Whitney test). Blue lines highlight the cancers that are significantly hypomethylated in tumoral tissue, while red lines demark the cancers that are significantly hypermethylated in the tumoral tissue.

Using the same approach in paired normal-tissue samples we could validate the methylation status of the four probes that was consistent with our previous results (Supplementary Figure S2.3.3).

To uncover if those alterations in DNA methylation could affect gene expression, we performed a correlation analysis between DNA methylation and gene expression for the probes that presented a statistically significant difference between the methylation status across the tumors that were identified to have different *MGP* expression levels (Figure 2.3.6 A-O and Supplementary Figure 2.3.4 A-H). The DNA methylation values of the four probes were significantly correlated with the expression of the *MGP* gene in the different types of tumors (Figure 2.3.6 A-O). Among the four probes, those hypermethylated were found to be significantly negatively correlated with *MGP* expression, except for cg2221831 in head and neck, and lung cancers. These results indicate that the down-regulation of *MGP* in several cancer tissues is caused by hypermethylation of important DNA methylation sites. The hypomethylation of the cg00431549 and the cg05360958 also have an inhibiting effect on gene expression in thyroid cancer.

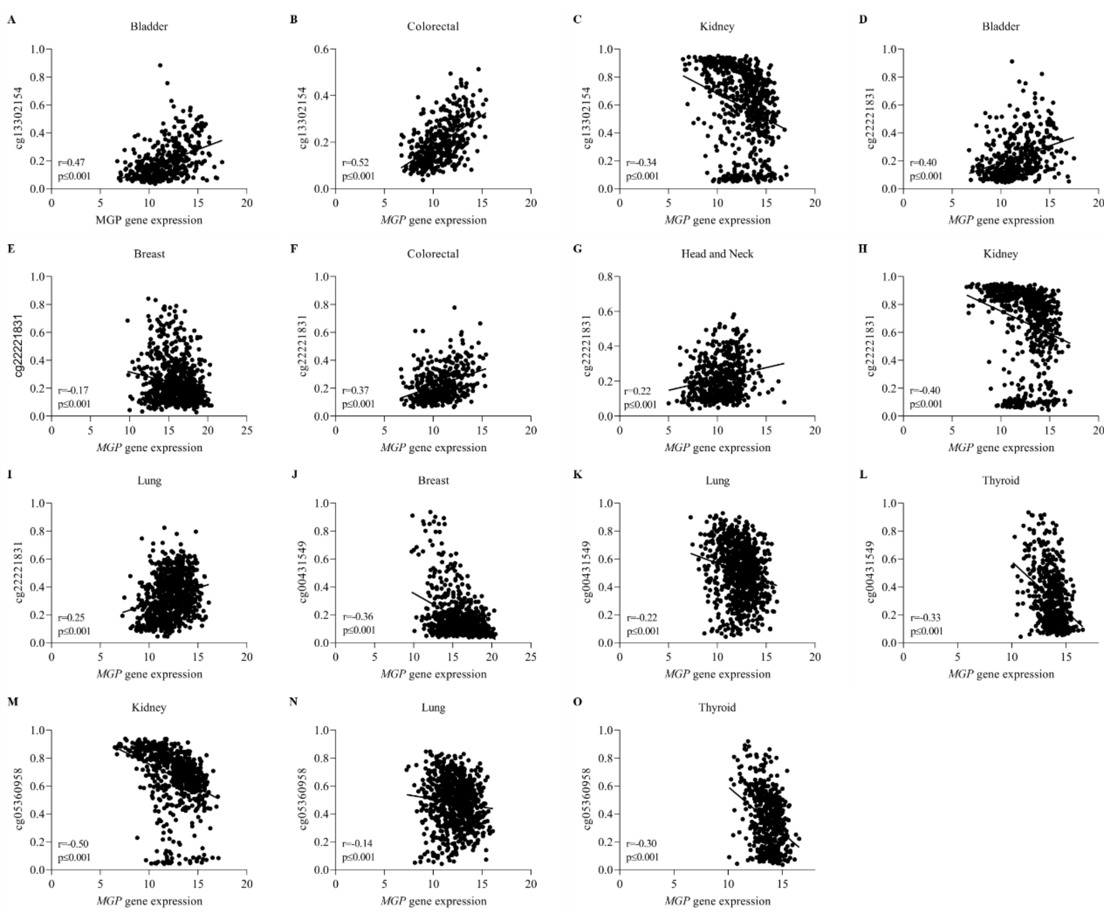


Figure 2.3.6 Correlations between *MGP* expression and CpG sites. Different correlation patterns between *MGP* expression and methylation levels were found depending on the hypomethylation or hypermethylation state of CpG sites. The results are presented as Spearman's correlation coefficient (r) and p-value. Results were considered significant with a p-value ≤ 0·05.

Additionally, we further explored the presence of other epigenetic markers and chromatin states in the region of the CpG sites, through the UCSC Genome Browser. The analysis revealed the presence of histone marks H3K4Me1, H3K4Me3, and H3K27Ac epigenetic modifications that are usually found near the regulatory regions of genes, as well as an open chromatin state, suggesting that the locations of the CpG sites in the promoter region might be associated to binding sites of transcription factors that could regulate *MGP* transcription.
The significant correlation between *MGP* hypermethylation and expression levels in some tumors combined with the presence of histone marks indicators of transcriptional regulation further suggested that these CpG sites might play a role in *MGP* gene regulation.

2.3.4 Discussion

MGP has been proposed as an important factor involved in the development and progression of various human tumors. However, its role in cancer pathogenesis has yet not been fully elucidated. According to data available in the literature, *MGP* may have different expression patterns depending on the location and type of tumor. As revised in Gheorghe and Crăciun (2015), depending on the type of tumor, *MGP* expression could be associated with a better or worst prognosis. There has been an increase in research papers reporting the controversial effect of *MGP* gene expression in tumorigenesis. For example, in head and neck cancer a decrease in *MGP* expression in the tumor tissue has been shown (Mishra et al. 2018), while in breast cancer (Gong et al. 2019), ovarian cancer (Sterzyńska et al. 2018), gastric cancer (Wang et al. 2020), glioblastoma (Mertsch et al. 2009) and colorectal cancer (Caiado et al. 2020), *MGP* was found to be overexpressed and associated with unfavorable prognosis. Although many studies have shown that *MGP* was overexpressed in colorectal tumor tissue (Ki et al., 2007; Li et al. 2020; Huang et al. 2021) other studies demonstrated that *MGP* was downregulated in CRC cancerous tissue (Alon et al. 1999; Notterman et al. 2001; Fan et al. 2001; Sabates-Bellver et al. 2007).. Loss of *MGP* expression in metastatic renal cell and prostatic carcinomas compared with primary tumors has been associated with tumor progression and metastasis (Levedakou et al. 1992). In glioblastomas, Kuzontkoski et al. (2010) have shown that *MGP* was downregulated

and correlated with unfavorable clinical status such as a poorly differentiated state, a larger tumor size, and lymph node metastasis.

In this work, the association between the expression of the *MGP* gene and eleven distinct cancers was analyzed based on the TCGA database. The results revealed that *MGP* was overexpressed in the bile duct and liver tumoral tissues and downregulated in bladder, breast, colorectal, head and neck, kidney, lung, and thyroid tumoral tissues. We also noted that older patients with bladder, breast, colorectal, lung, or thyroid cancers, presented high levels of *MGP* in the tumoral tissue. The high levels of *MGP* at a more advanced age might be related to the fact that, in most cases, cancer is an age-related disease and this age-related molecular alteration in signaling pathways, might contribute to deregulating *MGP* expression in cancer (Galamb et al. 2016).

Although *MGP* expression was found to be downregulated in most tumoral tissues, we were able to observe that in bladder, breast, colorectal, kidney, and lung cancers, patients with high levels of *MGP* presented the worst prognosis in terms of tumor stage (stage III-IV vs I-II), T, N and M stages. Moreover, we were able to differentiate patients into low and high *MGP* expression and correlate high *MGP* expression with a poor overall survival outcome. Also, the expression of *MGP* was predicted to be an independent prognostic factor related to overall survival in univariate and multivariate cox regression analysis for bladder, breast, colorectal, head and neck, kidney, liver, and lung cancers, as well as, other variables, such as age, T, N and M stages, and tumor stage. Tuo and Ye (2017), predicted that high levels of *MGP* are associated with better relapse-free survival in patients with lumina type A breast cancer patients.

Our study has demonstrated a lower *MGP* mRNA content in tumor tissues of bladder, breast, colorectal, head and neck, kidney, lung, and thyroid cancer patients than in the normal mucosa, while MGP protein contents in those tissues taken from the Human Protein Atlas did not validate the staining and intensity of MGP for the majority of those tissues. Furthermore, MGP expression in breast cancer patients was higher than in the breast samples of healthy controls. It appears that the high level of *MGP* transcripts expression in normal breast samples of healthy subjects was accompanied by relatively low MGP protein content. Thus, our observations suggest that in normal tissue, posttranscriptional events limit the *MGP* expression, and this regulation becomes disturbed in breast carcinoma (Tiago et al. 2016). Tuo and Ye, 2017 have also analyzed the Human Protein Atlas database and showed the widely MGP expression among the 11 cases of breast cancer tissues. The same authors also analyzed the breast cancer cohort from the TCGA database and showed that *MGP* downregulation was related to DNA hypermethylation in both luminal A and luminal B subtypes of breast cancer.

To evaluate if *MGP* DNA methylation is associated with the differential expression in cancer patients we proceeded to investigate the methylation levels of *MGP* taking advantage of the TCGA database which allowed us to correlate DNA methylation levels with gene expression of the patients. We found that the CpG sites located in the promoter region and the first intron of the *MGP* gene were differentially methylated across all tumors. We were able to observe, from the Spearman correlation coefficient analysis, that the hypermethylation state of CpG probes cg13302154 (in the kidney), cg 22221831 (in breast and kidney), cg 00431549 (in breast, lung, and thyroid) and cg 05360958 (in kidney, lung, and thyroid) in the tumoral tissue was inversely correlated with *MGP* gene expression, whereas for the CpG sites cg13302154 (in bile, bladder, colorectal, liver, and thyroid); cg 22221831 (in the bladder, colorectal, liver, and thyroid); cg00431549 (in liver and thyroid) and cg05360958 (in colorectal, liver, and thyroid) a hypomethylation state in the tumoral tissue was observed.

From what is known in the literature so far concerning the effect of methylation on *MGP* expression, the results regarding the hypomethylation status of cg13302154 that we observed are in agreement with those published by Galamb et al. 2016. These authors reported that the upregulation of *MGP* expression in the tumoral tissue is related to the hypomethylation status of cg13302154 and cg00431549 (Galamb et al. 2016). However, in our results, we could not observe the increase of *MGP* expression in the tumoral tissue of colorectal cancer we observed a significant downregulation of *MGP* and a positive correlation with the hypomethylation state of cg13302154, along with a hypomethylation status of cg22221831 and cg05360958 in the tumoral tissue of colorectal cancer. One of the possible reasons to explain the downregulation of *MGP* expression in the tumoral tissue could be the binding of possible transcriptional repressors to this CpG site, thus deregulating *MGP* expression, but additional experiments are required to validate this hypothesis.

In breast cancer, the authors Tuo and Ye (2017), published that the downregulation of *MGP* could be due to the methylation sites in the *MGP* promoter. However, the authors do not refer to which CpGs sites were analyzed and only refer to the state of hypermethylation of two analyzed CpGs located in the promoter. Nevertheless, in terms of expression, our results are in agreement with those observed in that study, since we observed a downregulation of *MGP* expression in breast cancer patients, and the *MGP* gene expression is negatively correlated with the hypermethylation state.

It is known that epigenetic events, such as DNA methylation, can inhibit gene transcription when present mostly in gene regulatory regions (Jones, 2012). The fact that three of the four identified CpG sites in the *MGP* gene are located in the promoter region, combined with the

presence of histone marks (Javaid and Choi, 2017) and an open chromatin state (Calo and Wysocka, 2013), all usually found in the vicinity of regulatory regions of a gene, supports the hypothesis that these CpG sites might be important for *MGP* gene regulation through the binding of possible negative transcriptional regulators, thus explaining the downregulation of *MGP* expression in the tumoral tissue.

2.3.5 Supplementary data

Table S2.3.1 Demographic features of paired patients in the different types of tumors

	Gender					Age				
	Male		Female			≤70		>70		
Characteristics	Number (n)	Median value[a]	Number (n)	Median value[a]	*p* value	Number (n)	Median value[a]	Number (n)	Median value[a]	*p* value
Breast	1	16.36	111	15.64	*p*=0.786	91	15.46	21	16.92	*p*=0.144
Colorectal	23	10.79	25	9.51	***p*=0.034**	22	10.57	26	9.72	*p*=0.230
Kidney	84	13.06	42	13.17	*p*=0.879	96	13.17	30	12.69	*p*=0.299
Liver	28	9.83	22	10.73	*p*=0.245	30	10.03	20	11.09	*p*=0.332
Lung	59	12.02	49	12.47	***P*=0.031**	66	12.24	42	12.50	*p*=0.601
Thyroid	17	13.41	41	13.78	*p*=0.260	52	13.70	6	13.28	*p*=0.402

[a]Median value of *MGP* expression in tumoral tissue
Mann-Whitney U test

Table S2.3.2 Pathological characteristics of paired TCGA tumors cohorts

Pathological characteristics	Breast			Colorectal			Kidney			Liver			Lung			Prostate			Thyroid		
	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value	n	Median value[a]	*p*-value
Tumor Stage			p=0,378			p=0,146			p=0,086			p=0,782			p=0,313			**p≤0,030**			p=0,979
I - II	85	15,75		31	9,61		68	12,88		33	10,41		87	12,39		28	13,59		42	13,67	
III - IV	27	15,27		17	11,09		58	13,77		17	9,91		21	12,07		24	14,25		16	13,6	
T stage			p=0,793			p=0,885			p=0,266			p=0,660			p≤0,311			p=0,169			p=0,561
TX																6	14,60				
T1	4	16,52		2	11		4	12,91		20	9,61		16	12,68					7	13,52	
T1a	1	16,92					23	13,09					3	11,63					2	14,46	
T1b	1	15,46					22	12,97					8	14,36					2	14,27	
T1c	24	15,47																			
T2	63	15,75		7	10,63		22	12,69		14	10,79		44	12,24					22	13,65	
T2a							2	9,93					20	12,77		1	13,19				
T2b	1	14,84					2	11,99					8	11,48							
T2c																28	13,68				
T3	11	14,88		33	9,61		1	13,91		8	10,07		7	11,9					21	13,78	
T3a							28	13,87		4	10,88					15	14,01				
T3b							18	13,71		1	7,81					6	14,66				
T3c							4	13,93													
T4				5	10,79					3	9,91		2	12,85		2	14,69		3	13,41	
T4a																			1	14,26	
T4b	6	16,71		1	9,89																
T4d	1	13,8																			
N stage			p=0,369			p=0,231			p=0,0796			p=0,513			p≤0,152			**p=0,043**			p=0,419
NX	2	16,29					56	13,07		18	10,74		5	14,12		5	13,35		5	13,42	
N0	34	15,56		33	9,51		58	13,06		31	10,14		62	12,29		46	14,02		29	13,88	
N0(i-)	12	15,72																			
N0(i+)		16,87																			
N0(mol+)		17,14																			
N1	13	15,11		7	11,56		11	13,85		1	9,15		25	12,27		1	15,4		14	13,69	
N1mi	5	14,64																			
N1a	16	15,09		1	11,37														7	13,24	
N1b	13	16,91																	3	12,81	
N1c	1	16,92																			
N2	7	14,51		5	10,73		1	11,49					16	12,21							
N2a	5	15,92																			
N2b				2	12,47																
N3	2	18,64																			
N3a	2	14,04																			
M stage			p=0,295			p=0,848			**p=0,002**			p=0,250			p=0,435			p=0,194			p=0,109
MX	7	16,35		7	11,37		15	11		16	10,37		68	13,29		4	14,56		15	13,41	
M0	103	15,48		33	9,87		87	13,09		33	10,14		142	13,64		48	13,99		41	13,89	
M0(i+)																					
M1	2	17,81		8	10,34		24	13,96		1	13,02		6	12,47					2	12,31	

[a]Median value of *MGP* expression in tumoral tissue
Mann-Whitney U test

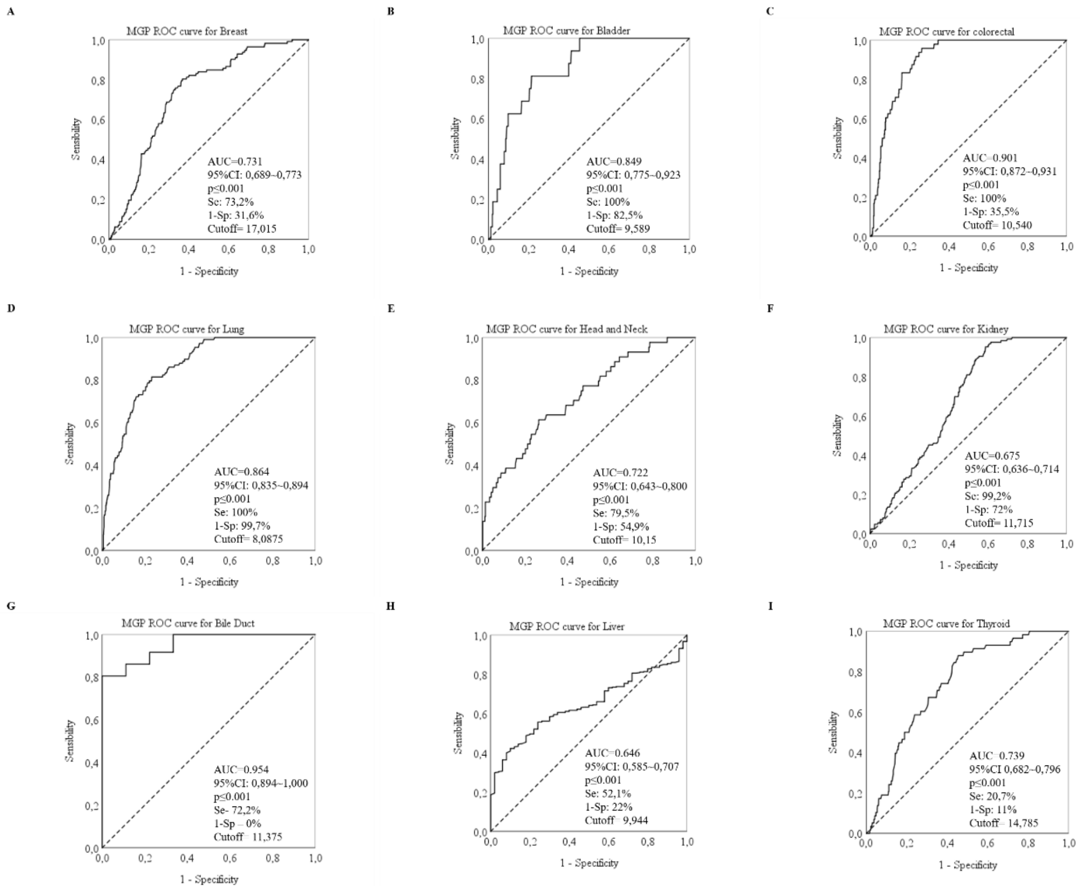
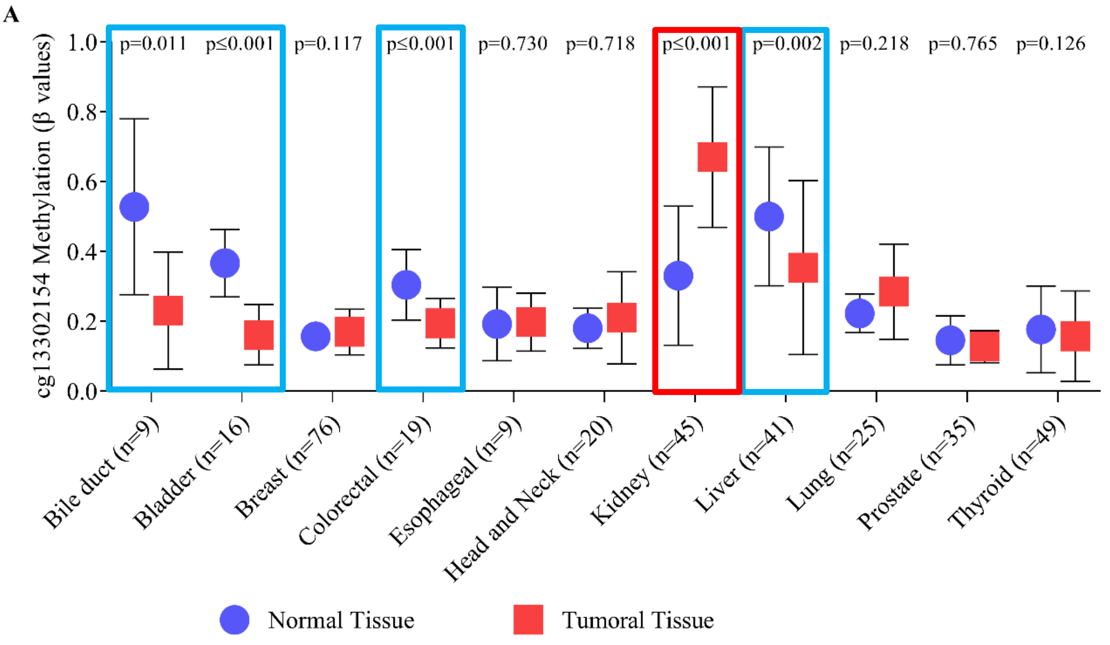
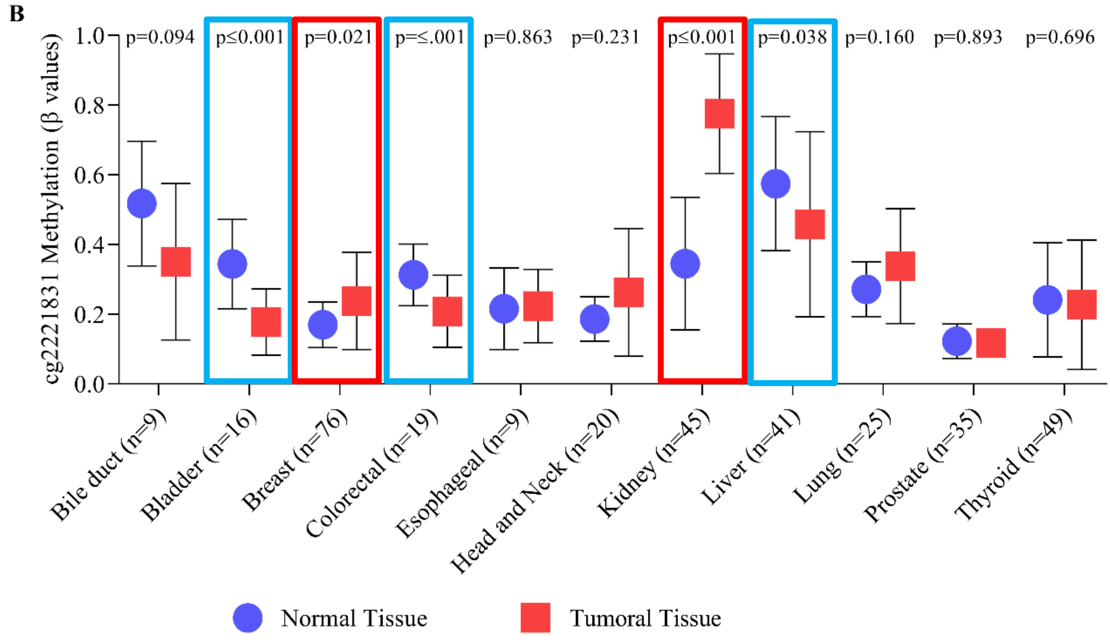


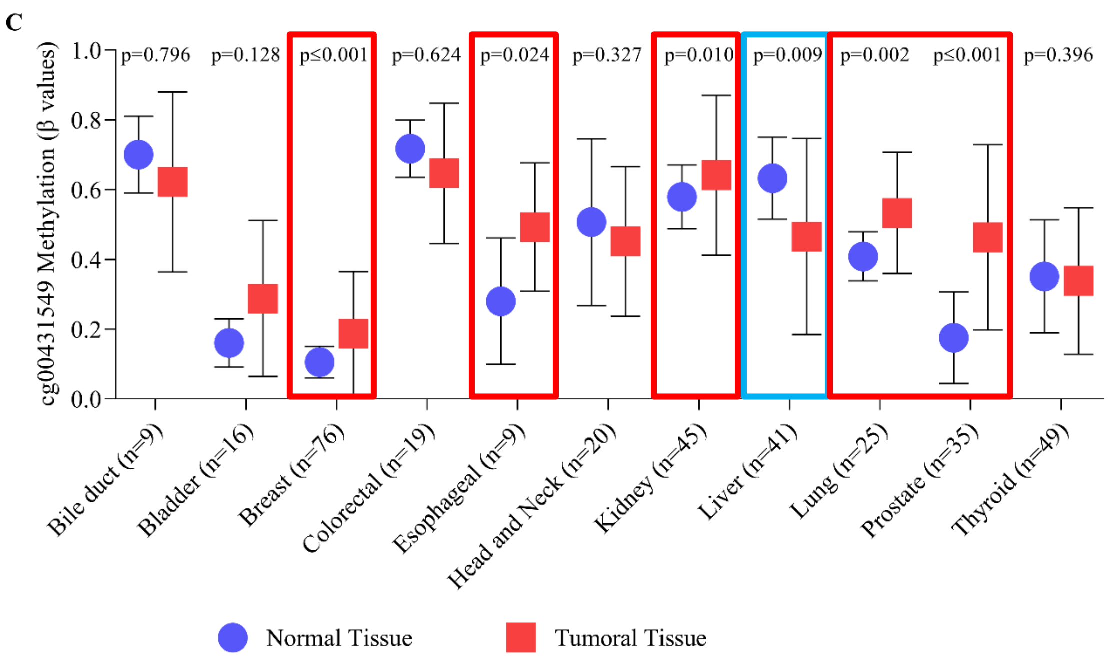
Figure S2.3.1 Receiver operating characteristic curves of *MGP* gene expression. The optimal cutoff value was established by AUC-ROC curve analysis to distinguish between normal and tumoral tissue samples.

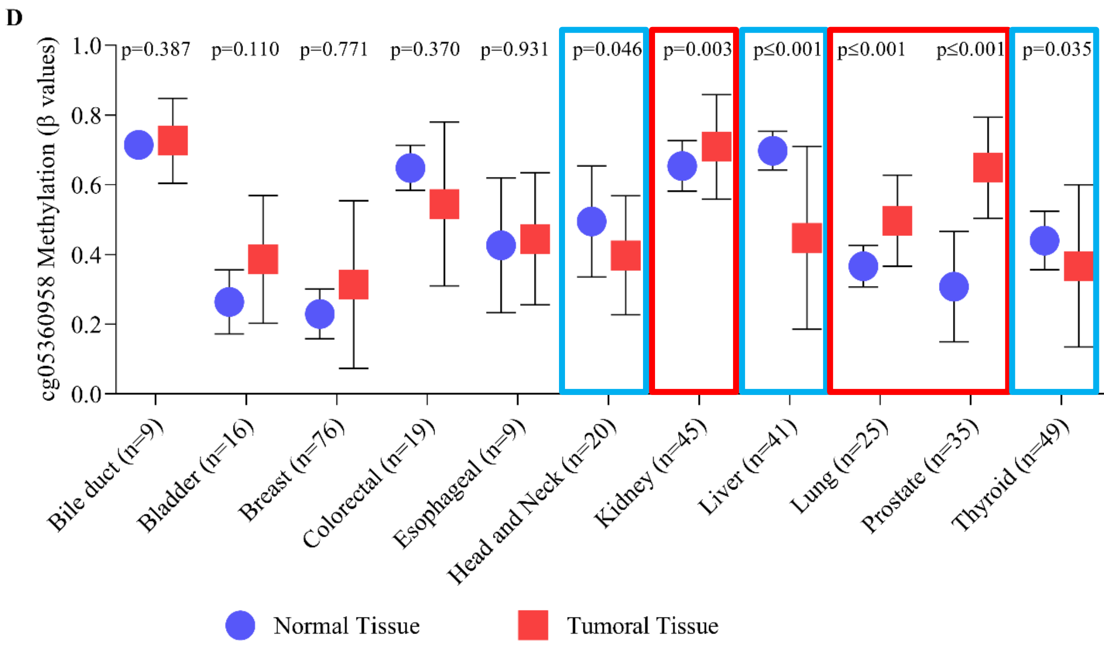
	Normal	Tumoral
Bladder	Male, age 76 Patient id: 4790	Female, age 73 Urinary bladder Urothelial carcinoma, High grade Patient id: 5463 Staining Low (11 of 12 not detected)
Breast	Female, age 43 Patient id: 2104 Staining low in glandular cells	Female, age 66 Breast Lobular carcinoma Patient id: 4229 Staining High (3 of 11) 3 medium, 3 low and 2 not detected)
Colorectal	Male, age 65 Patient id: 4910	Male, age 62 Rectum Adenocarcinoma Patient id: 5554 Not detecded (12 of 12)
Esophageal	Female, age 64 Patient id: 5446	N.D.
Head and neck	N.D.	Male, age 69 Head-Neck Squamous cell carcinoma Patient id: 4117 Not detected (4 of 4)
Kidney	Female, age 41 Patient id: 2530	Female, age 89 Kidney Adenocarcinoma Patient id: 3474 Not detected (12 of 12)
Liver	Female, age 50 Patient id: 2251	Female, age 25 Liver Carcinoma, Hepatocellular Patient id: 4823 (not detected 10 of 10)
Lung	Female, age 43 Patient id: 4840	Male, age 73 Lung Adenocarcinoma Patient id: 1907 Medium (1 of 11)
Prostate	Male, age 51 Patient id: 2053 Staining Medium	Male, age 70 Prostate Adenocarcinoma, High grade Patient id: 5418 Low (6 of 12)
Thyroid	Male, age 68 Patient id: 1922	Male, age 61 Thyroid gland Papillary adenocarcinoma Patient id: 2072 Not detected (4 of 4)

Figure S2.3.2 MGP protein expression analysis in the different type of TCGA tumors. Data was collected from The Human Protein Atlas.









Supplementary Figure S2.3.3 Differentially methylated *MGP* CpG sites in the paired tissue samples from the TCGA cancer cohorts. **A**. Probe cg13302154; **B**. Probe cg2221831; **C**. Probe cg00431549; **D**. Probe cg05360958. Beta values (β, methylation values) were compared between paired normal and tumor samples. CpG sites were considered as differentially methylated whenever the p-value < 0.05 (Mann–Whitney test). Blue lines highlight the cancers that are significantly hypomethylated in tumoral tissue, while red lines demark the cancers that are significantly hypermethylated in the tumoral tissue.

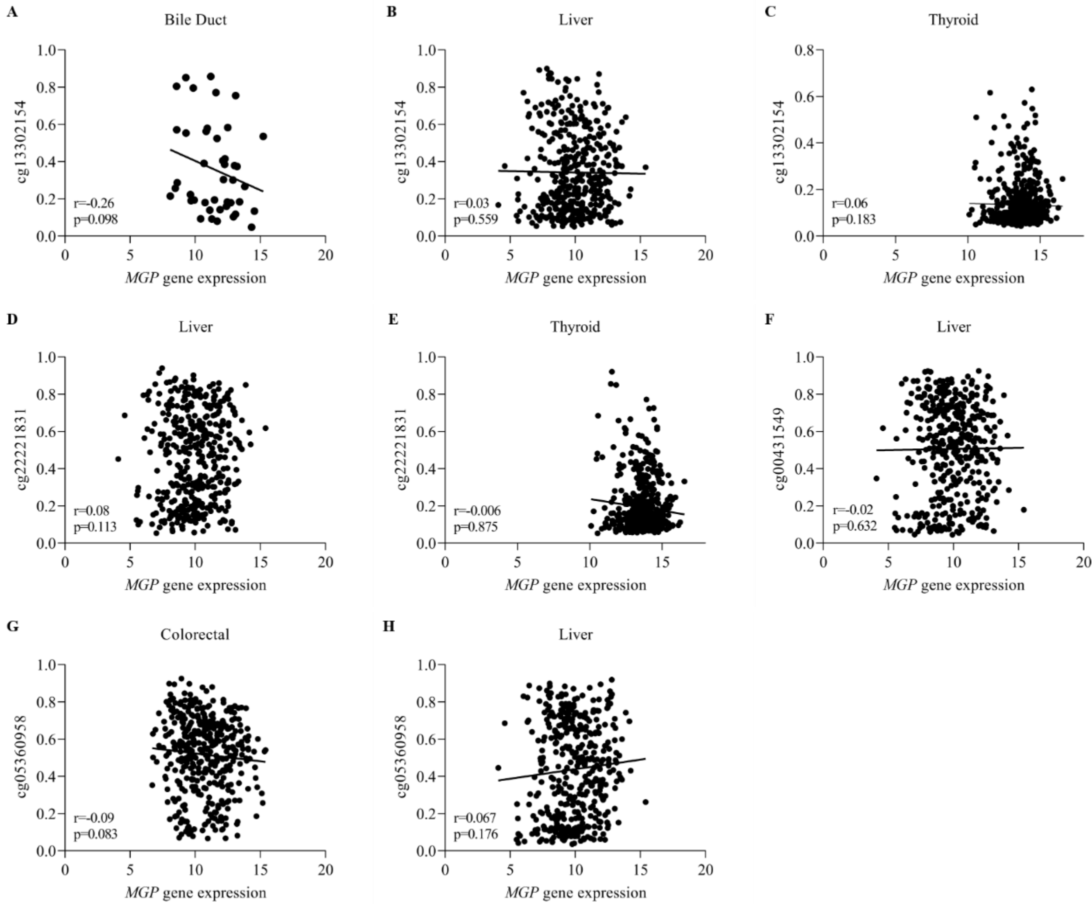


Figure S2.3.4 Correlations between *MGP* expression and CpG sites. No significant differences were observed between *MGP* expression and methylation levels found depending on the hypomethylation or hypermethylation state of CpG sites. Spearman correlation coefficient test, *p*-value was considered statistically significant when $P \leq 0.05$.

Chapter 3

Functional analysis of MGP promoter

Chapter 3.1

Transcriptional regulation of the human Matrix GLA Protein promoter: identification of downstream repressors

Helena Caiado[1,2,3]

M. Leonor Cancela[2,3,4*]

Natércia Conceição[2,3,4*]

This chapter was submitted as a research manuscript to Biochimica et Biophisica Acta - General Subjects

[1]ProRegeM PhD Programme in Mechanisms of disease and regenerative medicine, University of Algarve, Faro 8005-139, Portugal
[2]Center of Marine Sciences (CCMAR), University of Algarve, Faro 8005-139, Portugal
[3]Faculty of Biomedical Sciences and Medicine, University of Algarve, Faro 8005-139, Portugal
*Correspondence to: Natércia Conceição (nconcei@ualg.pt) and M. Leonor Cancela (lcancela@ualg.pt)

Author's roles: Study design: NC and MLC. Study conduct: All authors; Data collection: HC and NC. Data analysis: All authors. Data interpretation: All authors. Drafting manuscript: All authors. Revising manuscript content: All authors. Approving final version of manuscript: All authors.

3.1.1 Abstract

Background: Matrix Gla protein (MGP) is a vitamin K-dependent, γ-carboxylated protein that was initially found to be a physiological inhibitor of ectopic calcifications affecting mainly cartilage and the vascular system. Mutations in the *MGP* gene were found to be responsible for the Keutel syndrome, characterized by abnormal calcifications in cartilage, lungs, brain, and vascular system. MGP was recently shown to be abnormally regulated in several tumors, including cervical, ovarian, urogenital, and breast.
Methods: Using bioinformatic approaches, transcription factor binding sites (TFBSs) containing CpG dinucleotides were identified in the *MGP* promoter, including those for YY1, GATA1, and C/EBPα. We carried out functional tests using transient transfections with a luciferase reporter assay, primarily for the transcription factors YY1, GATA1, C/EBPα, and RUNX2. Results: By co-transfection analysis, we found that these TFs specifically repressed the *MGP* promoter. Furthermore, the co-transfection with RUNX2 activated the *MGP*

promoter. In addition, in several cancer types, *MGP* expression is either negatively or positively correlated with the studied TFs expression levels.

In conclusion: This study provides novel insights into *MGP* regulation by demonstrating that YY1, GATA1, and C/EBPα are negative regulators of the *MGP* promoter, and DNA methylation may influence their activity. The dysregulation of these mechanisms in cancer should be further elucidated.

3.1.2 Introduction

Matrix Gla protein (MGP) is a member of the family of vitamin K-dependent Gla proteins mainly associated with the extracellular matrix. The *MGP* gene located on chromosome 12p12.3 (Cancela et al. 1990) is highly expressed by vascular smooth muscle cells (VSMCs) (Shanahan et al., 1993) and chondrocytes (Yagami et al. 1999). The MGP function was clarified through the identification of mutations in *MGP* that lead to the Keutel syndrome (Munroe et al. 1999). Consistent with the functional importance of MGP in physiology, de-regulated expression of *MGP* has been reported to correlate with the development of various pathologies including colorectal, ovarian, lung, and breast cancer (Gheorghe and Crăciun, 2015).

Despite the knowledge of the involvement of MGP in diverse cellular events, the mechanisms regulating the expression of the *MGP* gene remain largely unknown. The *MGP* promoter has been investigated in previous studies with some putative transcription factor binding sites (TFBSs) reported, such as retinoic acid receptor (RAR) (Kirfel et al. 1997), vitamin D receptor (VDR), cAMP receptor (Farzaneh-Far et al. 2000) and parathyroid hormone (Gopalakrishnan et al. 2001; Suttamanatwong et al. 2007) but only a few were shown to be functional AP1 (Farzaneh-Far et al. 2001), FGF2 (Stheneur et al. 2003), HOXC8 (Gong et al. 2019), and Egr-1 (Dong et al. 2020).

It is known that cytosine methylation, in CpG sites, might change the spatial structure of DNA including at TFBSs, and therefore may regulate transcriptional regulation by changes in the affinity of TFs binding to DNA (Medvedeva et al. 2014).

Since the precise mechanisms involved in *MGP* regulation are unknown, the main objective of this work was to obtain novel knowledge regarding *MGP* transcriptional and epigenetic regulation, and to understand the molecular mechanisms involved. For that, we characterized the putative TFBSs in the *MGP* promoter and analyzed the functionality of those binding sites overlapping CpG sites. To better understand the functional consequences of TF binding, we analyzed the results of transient transfection promoter activity assays carried out in HEK293

cell line. In each assay, we compared the activity of the MGP wild-type promoter construct with that of a mutant promoter construct in which the predicted TF binding site was abolished (see Materials and methods section). The mutations were chosen to mutate only the CG nucleotides or to abolish TF binding by mutating as many as six nucleotides in the most informative positions, that is, those making the greatest contribution to the TF-DNA binding free energy. Our results show that the transcription factors YY1, GATA1, C/EBPα, and RUNX2 contribute to the regulation of the *MGP* gene at the transcriptional level.

3.1.2 Material and Methods

3.1.2.1 CpG sites identification in the *MGP* gene

The identification of CpG nucleotides located in the *MGP* gene was perform through the analysis of USCS genome browser data. This public database is available online and allows the visualization of CpG ID probes for specific CpG sites within a given gene, generated by the Illumina Infinium Human Methylation 450 Bead Array platform.

3.1.2.2 *In silico* analysis of *MGP* promoter

Human *MGP* promoter sequence spanning from -527 to +48 (from now on referred to as MGP560), related to the beginning of the transcriptional start site, was retrieved from the NCBI database (www.ncbi.nlm.nih.gov, NCBI Reference Sequence: NG_023331.1).
Evaluation for putative transcription factors binding sites (TFBSs) was carried out through the analysis of three online bioinformatics tools i) TF PROMO; ii) TF BIND; and iii) LASAGNA – search 2.0. The transcription factors predicted at least in two of those software analyzed were selected for further *in vitro* validation.

3.1.2.3 Luciferase reporter constructs of *MGP* promoter

The *MGP* promoter fragment (MGP560) inserted into pGL2 described previously by Kirfel et al. 1997 was removed and inserted into *Bgl*II – *Hind*III restriction sites of the reporter plasmid pGL3–Basic (Promega) and sequenced in both strands to confirm the correct orientation and insertion into the plasmid.

3.1.2.4 Site-directed mutagenesis

For the evaluation of the effect of the selected TFs and their binding to the CpG sites in the *MGP* promoter, the MGP560 luciferase construct was used as DNA template to generate seven luciferase constructs containing the following site-directed mutations for the CpG site cg00431549 (MGP560mut1), a new mutation abolishing the entire core sequence for YY1 (MGP560mut2); for the second putative binding site for YY1 transcription factor (MGP560mut3); for the third putative binding site of YY1 located in the transcription start site (TSS) (MGP560mut4); with both mutations for the CpG site and the second putative binding site for YY1 (MGP560mut5) and both mutations abolishing the entire YY1 core sequence and the third putative binding site to YY1 (MGP560mut6). These constructs were generated using the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies) following the manufacturer instructions and the specific primers described in Table 1 (the inserted mutations are depicted in bold and underlined).

Table 1 Primers used for PCR-based site-directed mutagenesis. Bold uppercase and underlined bases represent mutations with respect to the wild-type sequence, in the consensus binding sites for transcription factors.

Name	Orientation	Sequence (5' – 3')
MGP560mut1	sense:	5'- CCCAGGTCTGTCCCCAAGCATA**GT**ATGGCCAAAACTT-3'
	antisense:	5'-AAGTTTTGGCCAT**AC**TATGCTTGGGGACAGACCTGGG-3'
MGP560mut2	sense:	5'-TTTGCCCAGGTCTGTCCCCAAGCATA**GTCCCT**CCAAAACTTCTGCACCAGAGC-3'
	antisense:	5'-GCTCTGGTGCAGAAGTTTTGG**AGGGAC**TATGCTTGGGGACAGACCTGGGCAAA-3'
MGP560mut3	sense:	5'-GCCCACTCAGAGTAGATAATATC**CCC**AAGGAATGACTGTTTGGGAAAAG-3'
	antisense:	5'-CTTTTCCCAAACAGTCATTCCTT**GGG**GATATTATCTACTCTGAGTGGGC-3'
MGP560mut4	sense:	5'-TATAAAAACCTCACAGCCTTCCACTAACAT**AATGC**AGGAGCCTCTCTCCCTACTGC-3'
	antisense:	5'-GCAGTAGGGAGAGAGGCTCCT**GCATT**ATGTTAGTGGAAGGCTGTGAGGTTTTTATA-3'

3.1.2.5 Transcription factor expressing vectors

The expression plasmid containing the human YY1 transcription factor was a kind gift from Dr. Yang Shi (Harvard Medical School, pCMV-hYY1). The expression plasmid containing the human GATA1 transcription factor was purchased from Addgene (catalog number #118352) (Hietakangas et al. 2006). The expression plasmid containing C/EBPα transcription factor was a kind gift from Dr. Pierre Fafournoux (French National Institute for Agriculture, Food, and Environment INRAE, Unité de Nutrition Humaine, France, pcDNA-C/EBPα). The expression plasmid containing mouse full-length Runx2 type II isoform was a kind gift of Dr. Gerard Karsenty (Baylor College of Medicine, Houston, TX, USA, pCMV- Osf2/Cbfa1).

3.1.2.6 Cell Culture

Human embryonic kidney cells (HEK-293) were cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% (v/v) of fetal bovine serum (FBS, Invitrogen), 1% (v/v) of L-glutamine (200mM, Invitrogen) and 1% of penicillin-streptomycin (10.000 U/ml). The cells were maintained at 37ºC in a 5% CO_2 humidified atmosphere, and subdivided every 3 days.

3.1.2.7 Luciferase reporter assays

One day before the transfections, HEK-293 cells were seeded in a 24-well plate at the density of 5×10^4 cells per well. After reaching confluence of 50%-60%, 250ng of *MGP* promoter pGL3-luciferase construct (MGP560) or the *MGP* promoter fragments containing the site-directed mutations as described in the section 2.4 were co-transfected with 25ng of each expression plasmids containing the transcription factors or empty expression plasmids with 0.5 µl of XtremeGENE HP DNA transfection reagent (Roche) and 25ng of pRL – TK vector as the internal control per well.

The pGL3-Basic and pGL3-control vectors were used as negative and positive controls to monitor transfections efficiency. After 48h post-transfection, cells were lysed and firefly and renilla activities were measured with Dual-Luciferase Reporter Assay Kit (Promega) according to the manufacturer's instructions, using a multiplate reader (BioTek Synergy 4). The relative luciferase activity was calculated regarding based on the ratio from firefly and renilla luciferase activities and independent experiments were performed at least five times with duplicates per well.

3.1.2.8 TCGA data analysis

Cohort studies from different types of cancer were accessed thought The Cancer Genome Browser TCGA database (http://xena.ucsc.edu/, accessed between 2017 to 2021), data originally published by the National Cancer Institute. Data including gene expression was downloaded from the UCSC Xena browser (TCGA) data portal (https://xenabrowser.net/). All the downloaded data involved in this study were downloaded from TCGA in accordance with TCGA publication guidelines and data access policies, thus, additional approval by the local Ethics Committee was not needed. mRNA expression data from the TCGA database was used to assess correlations between expression of *MGP* and of the transcription factors analyzed (*YY1*, *GATA1*, *C/EBPα*, and *RUNX2*) expression.

3.1.2.9 Statistical analysis

The results are presented as mean ± Standard Deviation (S.D) and the statistical analysis was performed using Prism 8 (GraphPad Software). Significant differences were determined by a one-way ANOVA statistical test, with a Tukey's post-test correction, where a two-sided *p* value less than 0.05 was demarcated for statistically significant ($p \leq 0.05$).
The correlation between expression of MGP and the TFs expression was verified by Spearman rank correlation analysis.

3.1.3 Results

3.1.3.1 *In silico* analysis of transcription binding sites and CpG sites identified in the *MGP* promoter

We started by identifying possible CpG islands or CpG sites located in the human *MGP* promoter, through the analysis of the USCS genome browser. This public database available online, allow us to visualize CpG ID probes for specific CpG sites within a given gene, generated by the Illumina Infinium Human Methylation 450 Bead Array platform.
Results indicated the presence of six CpG sites located in or around the *MGP* gene, with four located in the promoter region, one in the transcription start site according to the sequence from NCBI (Reference Sequence: NG_023331.1), and one located in the first intron (Figure 3.1). Since DNA methylation is one of the factors that may influence gene expression and interfere with the binding of transcription factors, the promoter region of the *MGP* gene (MGP560) spanning -527 to +48 relative to the transcriptional start site, containing the CpG sites

cg00431549 and cg06601891 (Figure 3.1A), was used to search for possible transcription factor binding sites (TFBSs) overlapping the CpG, through an *in silico* approach using three different online prediction databases.

Since *MGP* is dysregulated in several types of cancers, in our analysis we selected transcription factors known from the literature to be involved in carcinogenesis processes. The results showed the presence of three putative binding sites for YY1, GATA1, and C/EBPα overlapping the CpG site cg00431549 and one binding sites for YY1 overlapping the CpG site cg06601891 (Figure 3.1A).

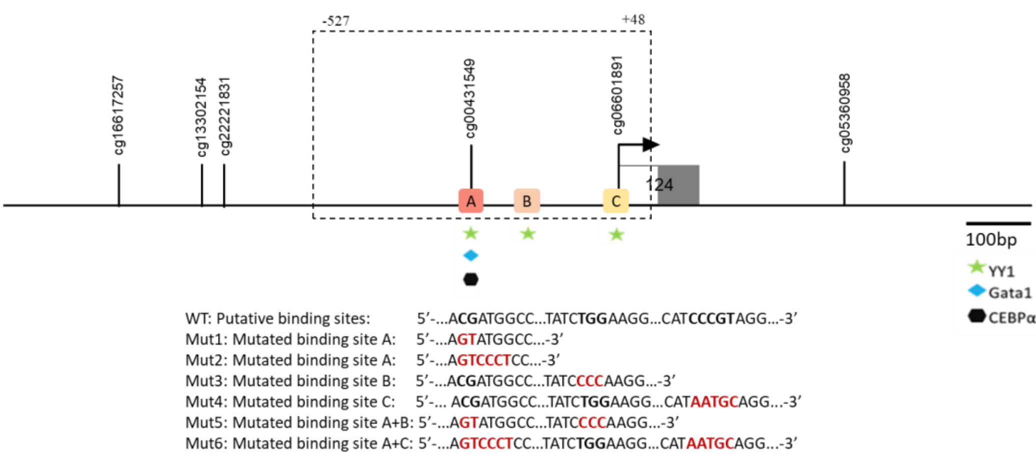


Figure 3.1 Putative transcription factor binding sites in human *MGP* promoter region. (A) Regulatory elements in the promoter region were searched using several bioinformatics tools. Geometric shapes represent the transcription factor binding sites for YY1 (star), GATA1 (diamond), C/EBPα (hexagon). The smaller, closed arrow represent the transcription start site. Each small, vertical line represents a single CpG site. Exons and promoter regions are in scale.
(**B**) Summary of the various mutations introduced in the identified binding sites. The different mutants produced are shown below the wild type (WT) sequence for the three putative binding sites. Mutated nucleotides are shown in red. In construct Mut1, for site A, YY1, GATA1 and C/EBPα sites were mutated in the nucleotides CG. In construct Mut2, for site A, YY1, GATA1 and C/EBPα sites additional nucleotides were mutated. In construct mut3, the binding for YY1 in site B was altered as indicated. In construct mut4, for site C, the YY1 site was mutated as shown. In construct mut5, both sites A and B were mutated as shown, and in construct mut6, both sites A and C were mutated as shown

3.1.3.2 Functional analysis of *MGP* promoter activity by the predicted transcription factors

To assess the promoter activity of the *MGP* gene by the selected transcription factors, we generated a luciferase reporter construct containing the *MGP* promoter (MGP560), as well as

constructs containing site-directed mutations for the respective putative binding sites for YY1, GATA1 and C/EBPα (Figure 3.1B). These fragments were then co-transfected with YY1, GATA1, and C/EBPα expression plasmids and their respective controls.

3.1.3.2.1 Transcription factor YY1 (Yin Yang 1) represses *MGP* luciferase promoter activity

Co-transfection of *MGP* promoter fragment (MGP560) with the YY1 expressing vector resulted in a significant 2.9-fold reduction of the luciferase activity (Figure 3.2A) over the control (the MGP560 promoter fragment co-transfected with the pCMV empty vector), suggesting a possible regulation by this transcription factor. To determine whether the CpG site is important for the binding of the transcription factor YY1, we performed a site directed mutation targeting the CG nucleotides containing the putative binding site for this transcription factor. However, when we co-transfected the expressing vector containing YY1 with the MGP560mut1 construct, we observed a similar repression of 2.6-fold in the promoter activity, suggesting that the mutation of the CpG site was not sufficient to restore the luciferase promoter activity (Figure 3.2B).

Since the repression was visible after mutating the CG (MGP560mut1), we performed a new site directed mutation of six nucleotides targeting the entire core sequence of YY1 (MGP560mut2). Results showed a repression of 1.7-fold over the control with this mutation (Figure 3.2C), suggesting that the repressive effect on *MGP* promoter by YY1 could be due to the binding of this transcription factor in other binding site(s) within this promoter fragment.

To explore the hypothesis of other putative binding site(s) for the YY1 transcription factor in the *MGP* promoter region under analysis, we searched through an *in silico* approach for new possible binding sites for YY1. Results showed another putative binding site for YY1, located at the position -144 to -141 in addition to the one located in the transcription start site (Figure 3.1). First, we explored the effect of this YY1 binding site in the *MGP* promoter luciferase activity by mutating it (MGP560mut3). However, a similar decrease of 2.5-fold over the control in the luciferase activity of this construct fragment MGP560mut3 was observed upon co-transfection with the expression vector carrying YY1 (Figure 3.2D).

Next, we evaluated if the effect of YY1 in the *MGP* luciferase activity was due to binding at the transcription start site by mutating it (MGP560mut4). Results showed again a decrease in the luciferase activity of 1.9-fold over the control, suggesting that this site might not be a binding site for YY1 transcription factor (Figure 3.2E).

Next, we evaluated a possible negative synergistic effect due to the YY1 binding in the identified binding sites. For that, we generated two constructs, the MGP560mut5 (containing mutations in both sites A and B) and MGP560mut6 (containing mutations in both sites A and C) (Figure 3.1B). Results demonstrate a repressive effect on the *MGP* promoter activity, in both constructs when co-transfected with YY1 expressing vector (Figure 3.2F and G), suggesting that the YY1 does not bind to any of these sites. Although we could not identify its binding site, these data still support a role for YY1 in basal *MGP* transcription.

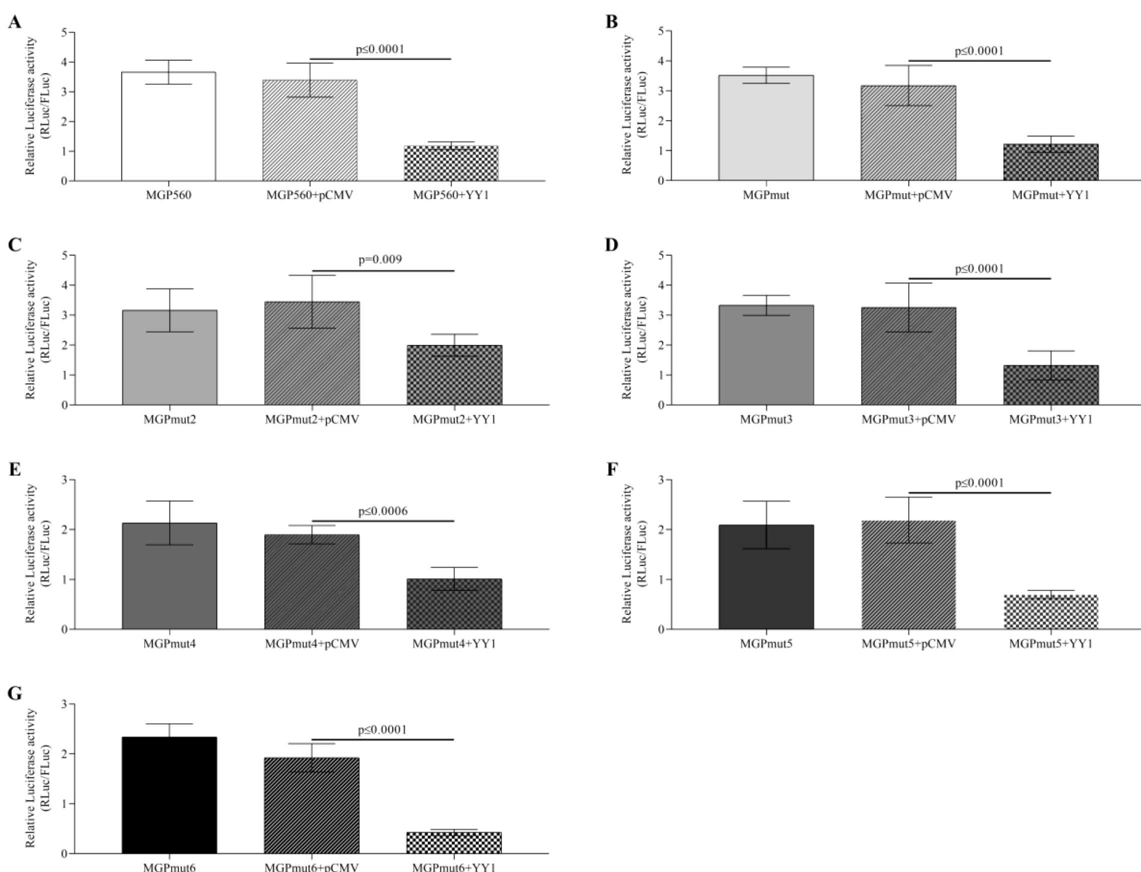


Figure 3.2 Effect of YY1 transcription factor in the *MGP* promoter activity. The *MGP* promoter fragment (MGP560) (**A**), the construct MGP560mut1 (**B**), the construct MGP560Mut2 (**C**), the construct MGP560Mut3 (**D**), the construct MGP560Mut4 (**E**), the construct MGP560Mut5 (**F**) and the construct MGP560Mut6 (**G**) were shown to be repressed in the presence of the transcription factor YY1. Data is presented as Mean±S.D of at least five independent experiments. Statistical significance was determined by One-way ANOVA statistical test with Tukey correction, p-value was considered significant when $p \leq 0.05$.

3.1.3.2.2 Transcription factor GATA1 binds to *MGP* promoter repressing its activity

Luciferase activity was significantly decreased when GATA1 expression vector was co-transfected with *MGP* promoter fragment MGP560, suggesting that GATA1 might be a negative regulator of *MGP* promoter activity (Figure 3.3A).

Co-transfection of GATA1 transcription factor with the MGPmut2 promotor construct resulted in a loss of *MGP* promoter activity (Figure 3.3B), suggesting that this predicted binding site could be functional.

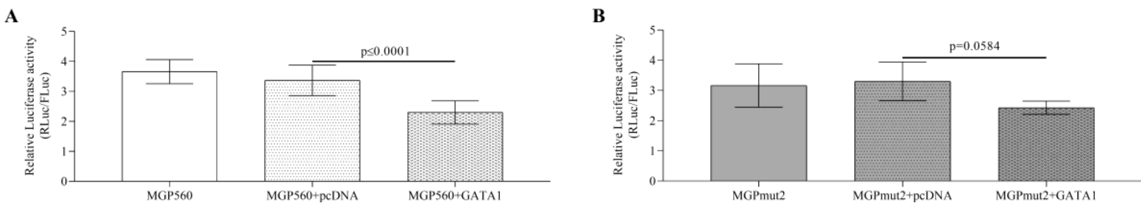


Figure 3.3 Effect of GATA1 transcription factor in *MGP* promoter activity. The *MGP* promoter fragment MGP560 (**A**), and the construct MGP560Mut2 (**B**), were co-transfected with the transcription factor GATA1. Data is presented as Mean±S.D of at least five independent experiments. Statistical significance was determined by One-way ANOVA statistical test with Tukey's correction, p-value was considered significant when $p \leq 0.05$.

3.1.3.2.3 C/EBPα has a repressive effect on the *MGP* promoter

Co-transfection of the expressing vector carrying the C/EBPα and the MGP560 construct resulted in a significant decrease in the luciferase activity (Figure 3.4A). To evaluate if this TF could bind in the CpG site, we transiently co-transfected it with MGP560mut1 construct containing the mutation for the CpG site. Results shown the loss of the repressive effect seen in the *MGP* promoter fragment, suggesting that C/EBPα might bind to this CpG site and could be a possible regulator for *MGP* gene (Figure 3.4B).

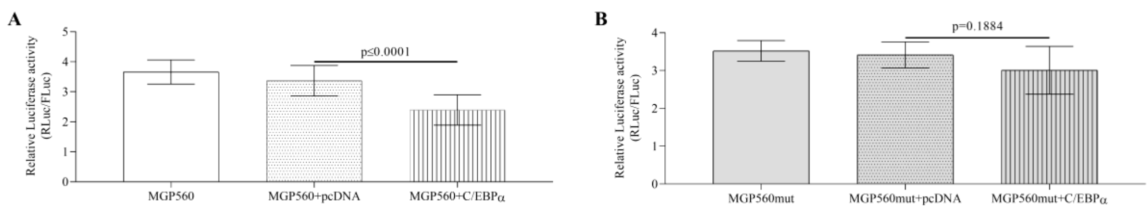


Figure 3.4 Effect of C/EBPα transcription factor in *MGP* promoter activity. The *MGP* promoter fragment MGP560 (**A**), and the construct MGP560Mut1 (**B**), were co-transfected with the transcription factor C/EBPα. Data is presented as Mean±S.D of at least five independent experiments. Statistical significance was determined by One-way ANOVA statistical test with Tukey's correction, p-value was considered significant when $p \leq 0.05$.

3.1.3.2.4 Repression in the *MGP* promoter is enhanced by YY1.

We employed a series of co-transfection assays with the *MGP* promoter alone or combined with the expression vectors for YY1, GATA1 and C/EBPα TFs to test the possible combined involvement of these factors in the observed response in *MGP* promoter. Co-transfection of either YY1, GATA1 or C/EBPα along with MGP560 construct resulted in significant repression in the promoter activity (Figure 3.5A). However, the level of repression was dramatically affected when YY1 and GATA1 or YY1 and C/EBPα or YY1, GATA1 and C/EBPα were transfected together with MGP560 construct, causing 3.9, 2.2 and 2.4-fold repression of luciferase activity, respectively, when compared to the effect observed due to these TFs in the absence of YY1, as shown in Figure 3.5A.

To further verify our finding about the synergistic action of these TFs, we performed co-transfections with MGP650mut2 construct. The results showed a repression of the promoter activity only in the presence of YY1 (Figure 3.5B) suggesting that the repression due to YY1 is independently of the binding site identified in the *MGP* promoter and that YY1 may interact with the two TFs to repress *MGP* promoter.

These data suggest that YY1 specifically targets GATA1 and C/EBPα-dependent transcription, although they do not rule out other possibilities (see Discussion).

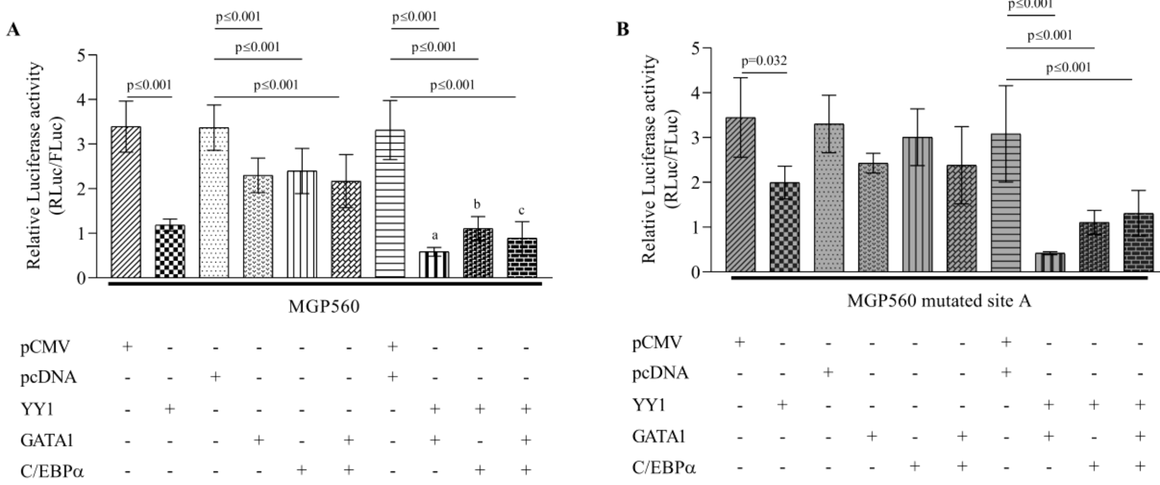


Figure 3.5 Combinatory interactions of YY1, GATA1 and C/EBPα in the repression of *MGP* gene. (A) Luciferase reporter assays after co-transfection of expression vectors YY1, GATA1 and/or C/EBPα along with reporter construct MGP560. Repression of promoter activity was observed when either YY1, GATA1 or C/EBPα were overexpressed in the cell culture system. However, simultaneous overexpression of YY1 and GATA1, YY1 and C/EBPα, and YY1, GATA1 and C/EBPα dramatically enhanced the intensity of repression of luciferase activity. Small letters denote the statistical difference between: (a) MGP560 with GATA1 and YY1 vs MGP560 with GATA1 ($p \leq 0.001$); (b) MGP560 with C/EBPα and YY1 vs MGP560 with C/EBPα ($p \leq 0.001$) and (c) MGP560 with C/EBPα, GATA1 and YY1 vs MGP560 with C/EBPα and GATA1 ($p \leq 0.001$) **(B)** Mutation analysis of *MGP* promoter by luciferase reporter assays. The binding site for YY1, GATA1 and C/EBPα was mutated in the MGP560 reporter construct and then the construct was co-transfected along with the expression vectors YY1, GATA1 and C/EBPα. The repression in the luciferase activity observed with the MGP560mut2 construct was not compromised with the mutation in the binding site when YY1 was present. Error bars represent ± SD of at least five independent experiments. Statistical significance was determined by One-way ANOVA statistical test with Tukey's correction, p-value was considered significant when $p \leq 0.05$.

3.1.3.2.5 RUNX2 enhances *MGP* promoter activity

In our previous work we showed a positive correlation between the expression of *RUNX2* mRNA in CRC tissue samples and *MGP* expression, therefore we wanted to evaluate if RUNX2 had an effect in *MGP* transcription regulation. Co-transfection of the expression vector carrying the RUNX2 triggered a significant increase of 1.5-fold over control in the luciferase activity of MGP560 construct (Figure 3.6A). Co-transfection of RUNX2 with the MGP560mut1 construct still promoted an enhancement in *MGP* promoter activity of 1.8-fold over the control (Figure 3.6B).

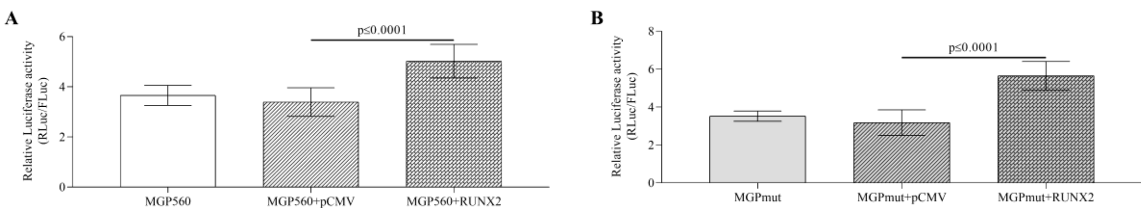


Figure 3.6 Effect of RUNX2 transcription factor in *MGP* promoter activity. The *MGP* promoter fragment MGP560 (**A**), and the construct MGP560Mut1 (**B**), were co-transfected with the transcription factor RUNX2 expressing vector. Data is presented as Mean±S.D of at least five independent experiments. Statistical significance was determined by One-way ANOVA statistical test with Tukey correction, p-value was considered significant when $p \leq 0.05$.

3.1.3.3 Transcription factors YY1, GATA1, and C/EBPα negatively regulate *MGP* expression in cancer

We next asked whether the pattern of *MGP* regulation by YY1, GATA1, C/EBPα and RUNX2 from our *in vitro* analysis may also be observed in primary clinical data. These factors were significantly up- or downregulated in different cancer tissues relative to noncancerous tissue (Supplementary Figure S3.1A, B, C, D).

Data from the TCGA database (http://xena.ucsc.edu/), using paired samples of cancer and noncancerous tissues, indicated that *YY1* mRNA levels are significantly upregulated in various types of malignancies, including bladder, breast, colorectal, esophageal, head and neck, lung, and prostate cancers (Supplementary Figure S3.1A). Next, we performed an *in silico* analysis and, by doing so, found that *YY1* and *MGP* expression exhibited a significant inverse correlation in bile duct, bladder, colorectal, esophageal, lung and thyroid cancers (Figure 3.7A-F). Therefore, decreased expression of the negative transcription factor *YY1* may be one of the key reasons for *MGP* overexpression in these cancers. Hence, these data indicate that YY1 may serve as a negative regulator of *MGP* and that it may play oncogenic roles in different cancers.

GATA1 expression was significantly higher in noncancerous tissue than in bladder, breast, liver, and lung cancers (Supplementary Figure S3.1B). *MGP* expression was positively closely related with *GATA1* in bladder, colorectal, esophageal, head and neck, lung, and prostate cancers (Figure 3.7G-M).

C/EBPα expression was significantly higher in noncancerous tissue than in bile duct, breast, lung, and thyroid cancers, and it was significantly upregulated in kidney and liver cancers (Supplementary Figure S3.1C). *C/EBPα* and *MGP* expression exhibited a significant inverse correlation in bile duct, bladder, esophageal, and kidney cancers and a positive correlation in liver, lung, and thyroid (Figure 3.7 N-T).

RUNX2 was overexpressed in breast, esophageal, kidney, lung, and thyroid cancerous tissues and significantly downregulated in liver and prostate cancerous tissues (Supplementary Figure S3.1D). *RUNX2* and *MGP* expression exhibited a significant inverse correlation in bladder and lung cancers and a positive correlation in bile duct and head and neck cancers (Figure 3.7 U-Z).

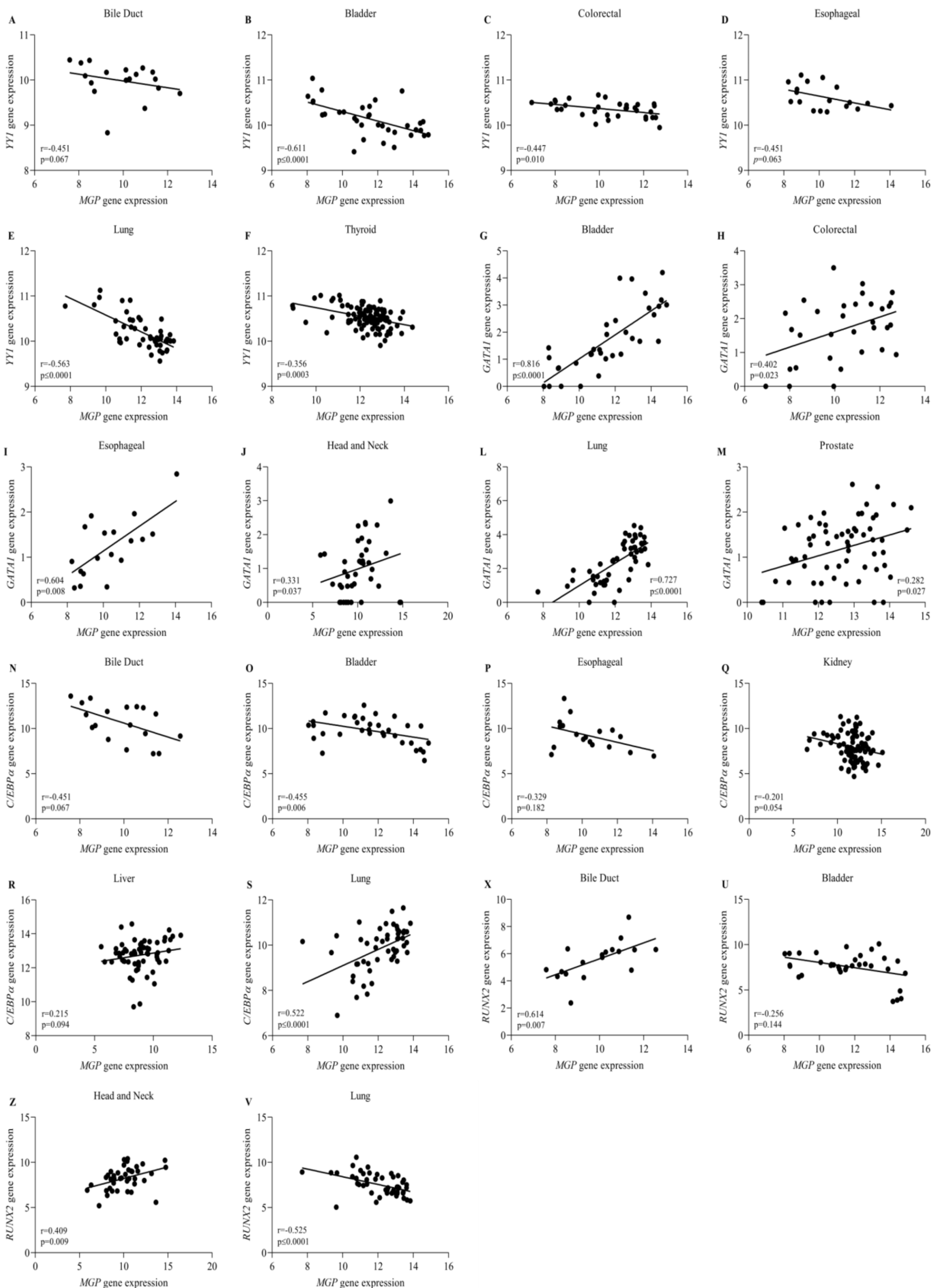


Figure 3.7 Correlation between expression of *MGP* and TFs in cancer. (**A-F**) *MGP* expression levels was negatively closely related with *YY1* in bile duct, bladder, colorectal, esophageal, lung and thyroid cancers. (**G-M**) *MGP* expression was positively closely related with *GATA1* in bladder, colorectal, esophageal, head and neck, lung, and prostate cancers. (**N-T**) *MGP* expression was negatively closely related with *C/EBPα* in bile duct, bladder, esophageal, and kidney cancers, and positively closely related with *C/EBPα* in liver, lung, and thyroid

cancers. (**U-Z**) *MGP* expression was negatively closely related with *RUNX2* in bladder and lung cancers, and positively closely related with *C/EBPα* in bile duct and head and neck cancers.

3.4 Discussion

Epigenetic events have been progressively more associated with cancer, being identified as a possible cause for the activation or silencing of certain genes depending on their biological context (Jones and Baylin, 2002), More recently, it has been published that the downregulation of *MGP* in breast cancer could be due to one methylation site in the proximal region of *MGP* promotor and higher levels of *MGP* could predict a better survival outcome (Tuo and Ye, 2017). Given our previous analysis of the methylation levels in the different types of tumors and their association with *MGP* expression (Caiado, Cancela and Conceição, submitted), our interest was to explore how the CpG site found in the promoter region could affect *MGP* transcription.

Transcription factors regulate gene expression to exert its specified function. This study utilized an *in silico* approach to search for TFs that putatively bind to the CpG sites identified in the *MGP* promoter region. There were three such TFs, YY1, GATA1 and C/EBPα, overlapping the cg00431549 located at -238bp from the TSS. Our studies suggest that these TFs and RUNX2 may be involved in the regulation of *MGP* expression; of these GATA1 and C/EBPα appear to be capable of interacting with the CpG site within the *MGP* promoter.

The regulation of *MGP* expression is likely multifactorial involving many transcription factors with activator or repressor functions responding to distinct signaling pathways. Transcription factor YY1 (Ying-Yang 1) belongs to the family of "Zinc-Fingers" proteins called GLI-Kruppel (Hosler et al. 1989; Ruppert et al. 1988). This transcription factor can interact in promoter regions of different genes acting either as an inhibitor or activator of gene expression, thus performing contrary and versatile functions, depending on external factors (Wang, Chen and Yang, 2006; Gordon et al. 2006; Fu et al. 2022). YY1 can exert transcriptional changes via direct DNA binding or through protein–protein interaction. To determine if *MGP* repression seen in the reporter assays is the result of YY1's direct binding to the *MGP* promoter, electrophoretic mobility shift assays (EMSA) or ChIP should be performed.

Our results are in concordance with results recently published (Fu et al. 2022) that showed, using TCGA and GEO datasets, that *YY1* was expressed at high levels in most malignancies, and its level of expression was statistically associated with the prognosis of tumor patients.

GATA1 has two highly conserved zinc fingers domains, being the one located at the C-terminal responsible for the binding to typical elements (recognize (A/T)GATA(A/G) motifs) in target gene promoters (Lamonica et al., 2006). GATA1 first reports indicated a critical function on

the formation of early eosinophil precursors and on the differentiation of committed erythroid precursors and megakaryocytes (Ferreira et al. 2005). Emerging evidence indicates GATA1 is activated in several tumors and is involved in cell growth, apoptosis, tumorigenesis, and aggressiveness of solid tumors. In pancreatic ductal adenocarcinoma tissues, GATA1 was found to be highly expressed and an independent predictor of prognosis and response to gemcitabine therapy through the anti-apoptotic pathway (Chang et al. 2019). In CRC, GATA1 was upregulated and associated with a predicted poor clinical outcome (Yu et al. 2019). In breast cancer, GATA1 was found overexpressed, and it was demonstrated to promote survivin expression (Boidot et al. 2010). Furthermore, it was shown that GATA1 promotes breast cancer growth and metastasis through regulating VEGF expression (Zhang et al. 2016). In glioblastomas, it was demonstrated that the interaction of GATA1 and MMP-2 enhanced glioblastoma invasion and migration (Kesanakurti et al. 2013). It was also suggested that in ovarian cancer GATA1 regulated JAG1 plays a key regulatory role in cancer cell proliferation and metastasis. On the other hand, it was shown that decreased mRNA expression of GATA1 was associated with tumor aggressiveness and poor outcome in clear cell renal cell carcinoma (Peters et al. 2015), implying that GATA1 may be associated with the progression and aggressiveness of renal clear cell carcinoma.

Our results are in concordance with results recently published (Xu et al. 2021) that compared the mRNA expression levels of the GATA family genes in different types of cancers and normal tissue samples by the Oncomine database and showed that GATA1 was expressed at relatively lower levels in most kinds of cancers than in normal tissues (Xu et al. 2021). However, the data in TCGA suggested that the *GATA1* expression was positively correlated with the *MGP* expression in cancer tissues. Further *in vivo* and *in vitro* studies are needed to better understand the association of GATA1 and MGP and their role in the different tumors.

We report hereby that YY1 and GATA1 act synergistically coordinating the repression of *MGP*. It appears that this synergistic action only requires the presence of the corresponding binding site for GATA1 in *MGP* regulatory region. Synergistic or cooperative regulation of gene expression is a widely occurring phenomenon. In most cases, TFs responsible for synergistic or cooperative regulation of gene expression co-occupy regulatory regions of common target genes. The TFs that are involved in synergistic or cooperative gene regulation are often observed to be interacting directly with each other (Raich et al. 1995; Gordon et al. 2006). Whether YY1 and GATA1 directly interact after their binding to the *MGP* gene promoter remains to be determined.

The CCAAT/Enhancer binding protein (C/EBPα) is a member of the leucine zipper family of transcription factors (Landschulz et al. 1988). The association between C/EBPα and cancer has been well documented in acute myeloid leukaemia (AML) (Tenen, 2001; Schuster and Porse, 2006). C/EBPα is important for myeloid differentiation, and dysregulation of its expression is observed in AML and other haemato-lymphoid malignancies. Reduced expression of *C/EBPα* has been observed in lung, breast, and head and neck cancers through epigenetic mechanisms including loss of heterozygosity and DNA methylation (Landschulz et al. 1988; Halmos et al. 2002; Costa et al. 2007; Gery et al. 2005; Bennett et al. 2007). In gastric cancer *C/EBPα* expression was shown to be downregulated (Regalo et al. 2010). In ovarian cancer (Konopka et al. 2016) and hepatocellular carcinomas (Lu et al. 2010), an up-regulation of *C/EBPα* mRNA expression was associated with worsening outcome. Also, in clear cell renal cell carcinoma (ccRCC) *C/EBP*α mRNA expression was up-regulated in TCGA data (Park et al. 2018). Altogether, these results show that, in agreement with our data obtained using TCGA database, *C/EBPα* expression was significantly up-regulated in some cancer types and down-regulated in others. This dysregulation is likely to be associated with poor clinical outcomes in cancer. Our *in vitro* results show that *C/EBPα* seems to bind *MGP* promoter in a specific binding site repressing its activity, and our *in silico* analysis using TCGA data show that in some tumors *C/EBPα* and *MGP* expression exhibited a significant inverse correlation, while in others there was a positive correlation.

The transcriptional and post-transcriptional regulation of *MGP* expression are complex (Farzaneh-Far et al. 2001; Cancela et al. 2014), and it is therefore possible that downstream of YY1, GATA1 and C/EBPα's downregulation of *MGP* transcription other factors are significantly involved in determining the ultimate expression of *MGP* and the corresponding clinical sequelae. It is also important to note that while our work demonstrates robust *MGP* promoter activity reduction via reporter assay, the extent to which these TFs-mediated reductions of *MGP* expression result in increased apoptosis, alterations in cell cycle progression, or modulation of other hallmarks of cancer progression is currently under investigation in our laboratory.

Another transcription factor that was shown to be able to regulate *MGP* gene expression was Runt-related transcription factor 2 (RUNX2). In mice, parathyroid hormone regulates *Mgp* expression through the transcription factors Sp and Runx2 (Suttamanatwong et al. 2009). In addition, it has been demonstrated that *MGP* is regulated by Runx2 in *Xenopus laevis* (Fazenda et al. 2010). Interestingly, in this model, two functional promoters (proximal and distal) were identified in the *MGP* gene, and both were shown to be regulated by Runx2. It was evidenced

that co-transfection of ATF4, RUNX2 and SATB2 enhanced the *MGP* human promoter activity in ATDC5 cells (Roberto et al. 2018). Recently, we showed that *MGP* and *RUNX2* were overexpressed in CRC tissue samples and there was a positive correlation between the two expressions in tumor mucosa (Caiado et al. 2020). In our present results, we showed that RUNX2 binds to *MGP* promoter and induces its transcription. Also, our *in silico* analysis showed that *RUNX2* was overexpressed in breast, esophageal, kidney, lung, and thyroid cancerous tissues and significantly downregulated in liver and prostate cancerous tissues, and that *RUNX2* and *MGP* expression exhibited a significant inverse correlation in bladder, lung cancer and a positive correlation in bile duct, and head and neck cancers.

Whether the alterations in the TFs expression and/or disturbance in *MGP* methylation are implicated in cancer should be further investigated by comparing the expression of these TFs and the levels of CpG dinucleotide methylation in the *MGP* gene of individuals with different type of cancers compared to the controls.

Our discovery of novel transcriptional repressors of *MGP* may provide new ways of understanding altered *MGP* expression in a significant subset of *cancers*. We also provide evidence for a possible role of YY1, GATA1 and C/EBPα in *MGP* transcription regulation and in human cancer. The clinical significance of this finding across different cancer types has yet to be determined.

3.5 Supplementary data

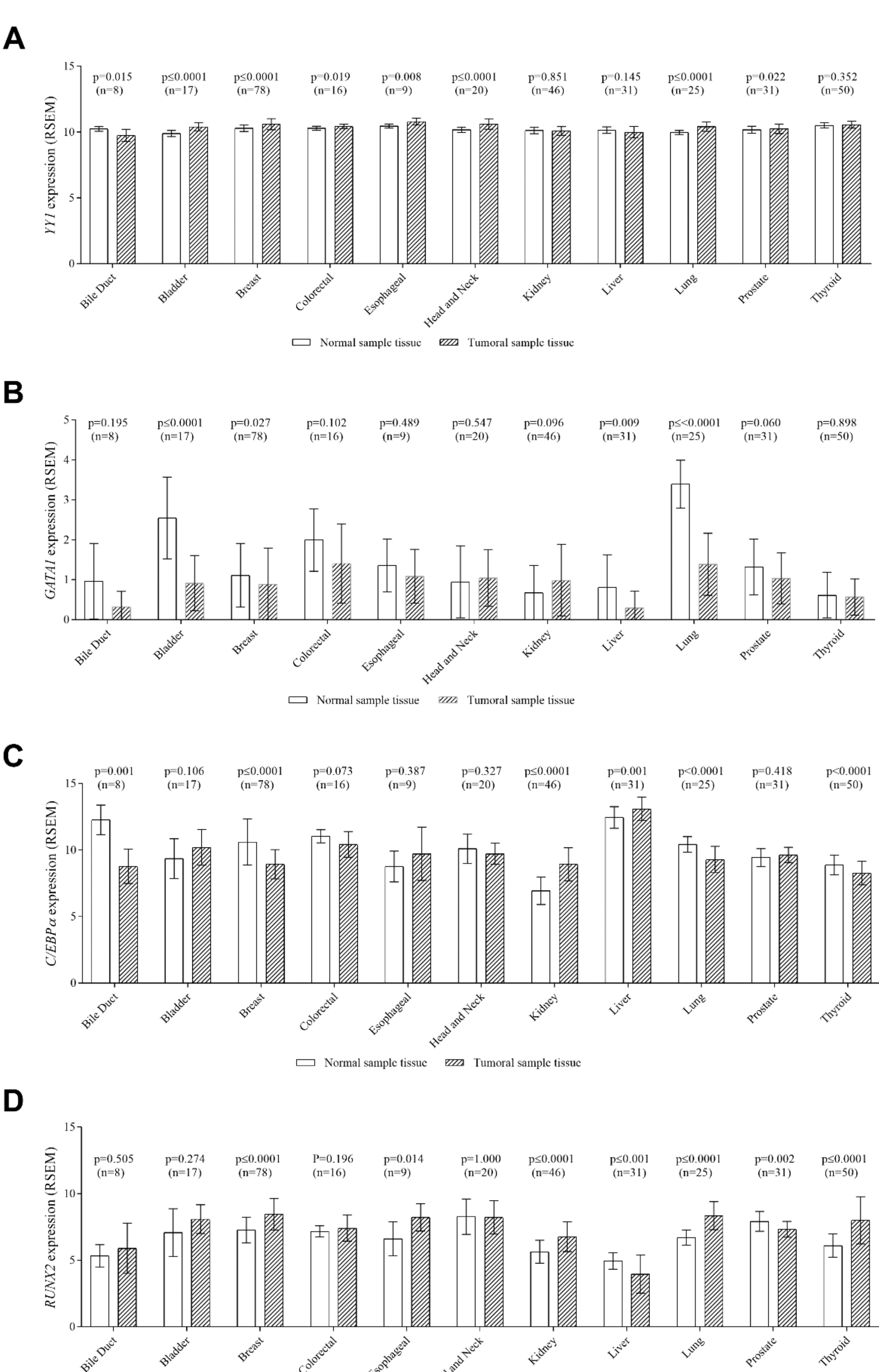


Figure S3.1 Analysis of *YY1*, *GATA1*, and *C/EBPα* gene expression in paired tumor-normal samples from TCGA databases. Mean ± SD, Mann-Whitney statistic test, the *p*-value was considered statistically different ($p \leq 0.05$).

Chapter 4

Final Conclusions and Future Perspectives

4.1 General conclusions and Future perspectives

Cancer comprises multi-step and complex processes that involve multiple deregulated signaling pathways to thrive and proliferate in "normal conditions".

Being cancer one of the major causes of disease and a major concern worldwide, it is of great interest to discover new molecular biomarkers for early-stage cancer detection, to help characterize patient populations, and to contribute to improving a more personalized therapy for the patients.

Recently, *MGP* has been shown to be deregulated in several types of tumors, Nonetheless, there is still too much that remains to be answered regarding the role of the *MGP* gene in cancer. Therefore, in this work, we proposed to study *MGP* gene expression in different types of tumors, to identify possible epigenetic patterns and new possible transcriptional regulators that could be involved in the regulation of *MGP* gene in cancer.

We started by evaluating expression of the *MGP* gene and genes of two of its known transcriptional regulators, *RUNX* and *FGF2*, in normal and tumoral tissues in a group of 33 colorectal patients and in 9 non-tumoral samples. In these studies, we found that *MGP*, *RUNX2,* and *FGF2* were mostly upregulated in the tumoral tissue when compared with tissue from paired normal mucosa and from non-tumoral samples. Also, we found that high levels of *MGP* in combination with high levels of the transcriptional regulators *RUNX2* or *FGF2* were important predictive factors to differentiate the patients according to their prognosis. However, it was not possible to observe a correlation between *MGP* gene expression and clinical and histopathological features of those patients.

Given the small number of samples from patients with colorectal cancer and to better understand if the previous expression pattern observed was maintained in other types of tumors, we analyzed the data available in the TCGA data repository for 11 types of tumors to accomplish this task. Interestingly, we found that the majority of the analyzed tumors presented a downregulation of *MGP* gene expression in the tumoral tissue, except for bile duct and liver cancers. However, when we analyzed the correlations between *MGP* expression in the tumoral tissue and the clinical and pathological features of the patients, we found that high *MGP* expression was correlated with more advanced stages of the disease in some tumors. Notably, in terms of survival and prognosis, we were able to distinguish the patients according to *MGP* expression in normal and tumoral tissues and correlate this expression with the overall survival of the patients. We observed that patients presenting high levels of expression of this gene had a poor overall survival. Moreover, it was possible to establish the high levels of *MGP*

expression as an independent prognostic factor for the survival in some of the analyzed tumors, reinforcing the hypothesis that *MGP* expression could be a good prognostic factor for patients overall cancer survival.

Given the discrepancy observed between tumors, one possible explanation for the differences observed between our previous findings and the results obtained in the present study with colorectal cancer patients might be related to the heterogeneity of the tumor or the individual genetic variability amongst patients in the different types of tumors.

To better understand the *MGP* expression pattern across tumors, it would be interesting to explore another publicly available data repository to evaluate *MGP* gene expression in the different types of tumors and extend this analysis to non-solid tumors, as well as to explore the expression levels of the different *MGP* transcripts within the tumors.

One of the multiple and complex molecular processes involved in gene regulation is known as epigenetic regulation. These epigenetic alterations, when localized in regulatory DNA regions, are capable of silencing the transcriptional activity of a gene.

In that sense, we hypothesized if the observed *MGP* deregulation could be related to epigenetic alterations present in regulatory regions of the *MGP* gene. To explore this hypothesis, we start by scanning for possible CpG sites or islands located within the *MGP* gene. From this analysis, we found the presence of six CpG sites located in the MGP gene, four of them located in the promoter region, one at the beginning of the transcription start site, and one in the first intron.

However, we only were able to analyze the methylation pattern of four of the six probes, due to the absence of methylation values for the probe located at the beginning of the transcription start site (cg06601891) and the probe located more downstream in the promoter region (cg16617257). Given the continuous amount of data generated and publicly available in the data repository, it would be of great interest to explore those databases to better understand how DNA methylation, histone tail modifications, and chromatin conformation are involved in *MGP* regulation in tumoral and normal tissues.

Concerning the methylation pattern in the analyzed tumors, we found a different methylation state depending on the type of tumor. Furthermore, we found that these differentially methylated patterns were either negatively or positively correlated with *MGP* gene expression for some of the analyzed tumors. Moreover, we found the presence of histone marks and an open chromatin state, normally found near regulatory regions. We can conclude from these results, that the methylation of these probes appears to have an impact on *MGP* regulation.

Regarding epigenetic regulation, several studies have shown that epigenetic factors, including DNA methylation, chromatin structure, and histone modifications, can interact with each other

and form a complex regulatory network, affecting regulatory processes, such as alternative splicing. Accordingly, we suggest in a future work to analyze the effect of epigenetic factors in the alternative splicing of *MGP* isoforms, to better understand if there is a possible correlation between the different isoforms in the normal and tumoral tissues, not distinguishable at the gene expression level.

One of the most common implications associated with the presence of these epigenetic modifications affecting the regulatory elements of a gene is the repression of transcriptional activity. However, it is worth noticing that these events are a dynamic process and can occur due to external factors, consequently affecting gene expression through the binding of transcriptional regulators that may be activators or repressors of gene transcription.

Under this hypothesis, we explored if there was any transcription factor that could bind to these CpG sites and regulate transcription of the *MGP* gene, which could in part explain the deregulation observed in the different types of cancers.

Through the functional analysis of the promoter region of the *MGP* gene by site directed mutagenesis in one of the CpG sites (cg00431549), we found three novel transcriptional repressors of the *MGP* gene, which in part could explain the different and dynamic *MGP* gene expression in cancer.

To validate these results, we propose to perform ChIP assays using cultured tumor cells to confirm if the identified repressors effectively bind to the CpG site cg00431549 and possibly, identify new putative transcriptional regulators of the *MGP* gene that could bind to the other CpG sites.

Although much remains to be answered, the novel results obtained in the scope of this work, reinforce the hypothesis that the *MGP* gene might be a good prognosis factor for cancer and the overall survival prognosis of the patients, providing new findings to better understand how *MGP* deregulation might be modulated by epigenetic modifications and transcriptional regulators in cancer.

References

Akl, M. R., Nagpal, P., Ayoub, N. M., Tai, B., Prabhu, S. A., Capac, C. M., Gliksman, M., Goy, A., and Suh, K. S. (2016). Molecular and clinical significance of fibroblast growth factor 2 (FGF2/bFGF) in malignancies of solid and hematological cancers for personalized therapies. Oncotarget, 7(28), 44735–44762. https://doi.org/10.18632/oncotarget.8203

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., and Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences, 96(12), 6745–6750. https://doi.org/10.1073/pnas.96.12.6745

Bennett, K. L., Hackanson, B., Smith, L. T., Morrison, C. D., Lang, J. C., Schuller, D. E., Weber, F., Eng, C., and Plass, C. (2007). Tumor Suppressor Activity of CCAAT/Enhancer Binding Protein α Is Epigenetically Down-regulated in Head and Neck Squamous Cell Carcinoma. Cancer Research, 67(10), 4657–4664. https://doi.org/10.1158/0008-5472.CAN-06-4793

Boidot, R., Végran, F., Jacob, D., Chevrier, S., Cadouot, M., Feron, O., Solary, E., and Lizard-Nacol, S. (2010). The transcription factor GATA-1 is overexpressed in breast carcinomas and contributes to survivin upregulation via a promoter polymorphism. Oncogene, 29(17), 2577–2584. https://doi.org/10.1038/onc.2009.525

Boraldi, F., Garcia-Fernandez, M., Paolinelli-deVincenzi, C., Annovi, G., Schurgers, L., Vermeer, C., Cianciulli, P., Ronchetti, I., and Quaglino, D. (2013). Ectopic calcification in β-thalassemia patients is associated with increased oxidative stress and lower MGP carboxylation. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1832(12), 2077–2084. https://doi.org/10.1016/j.bbadis.2013.07.017

Bork, U., Grützmann, R., Rahbari, N. N., Schölch, S., Distler, M., Reissfelder, C., Koch, M., Weitz, J. (2014). Prognostic relevance of minimal residual disease in colorectal cancer. 20(30), 10296–10304. https://doi.org/10.3748/wjg.v20.i30.10296

Boström, K., Zebboudj, A. F., Yao, Y., Lin, T. S., and Torres, A. (2004). Matrix GLA protein stimulates VEGF expression through increased transforming growth factor-beta1 activity in endothelial cells. The Journal of Biological Chemistry, 279(51), 52904–52913. https://doi.org/10.1074/jbc.M406868200

Caiado, h., Cancela, M. L., and Conceição, N. (2022). Assessment of MGP gene expression and contribuition to cancer prognosis. Submitted to Biochimie

Caiado, H., Conceição, N., Tiago, D., Marreiros, A., Vicente, S., Enriquez, J. L., Vaz, A. M., Antunes, A., Guerreiro, H., Caldeira, P., and Cancela, M. L. (2020). Evaluation of MGP gene expression in colorectal cancer. Gene, 723(October 2019), 144120. https://doi.org/10.1016/j.gene.2019.144120

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Molecular Cell, 49(5), 825–837. https://doi.org/10.1016/j.molcel.2013.01.038

Cancela, L., Hsieh, C. L., Francke, U., and Price, P. A. (1990). Molecular structure, chromosome assignment, and promoter organization of the human matrix Gla protein

gene. Journal of Biological Chemistry, 265(25), 15040–15048. https://doi.org/10.1016/s0021-9258(18)77221-9

Cancela, M. L., Laizé, V., and Conceição, N. (2014). Matrix Gla protein and osteocalcin: From gene duplication to neofunctionalization. Archives of Biochemistry and Biophysics. https://doi.org/10.1016/j.abb.2014.07.020

Chang, Z., Zhang, Y., Liu, J., Guan, C., Gu, X., Yang, Z., Ye, Q., Ding, L., and Liu, R. (2019). GATA1 Promotes Gemcitabine Resistance in Pancreatic Cancer through Antiapoptotic Pathway. Journal of Oncology, 2019. https://doi.org/10.1155/2019/9474273

Chen, Q. W., Zhu, X. Y., Li, Y. Y., and Meng, Z. Q. (2014). Epigenetic regulation and cancer (review). Oncology Reports, 31(2), 523–532. https://doi.org/10.3892/or.2013.2913

Chiu, T., Fang, D., Chen, J., Wang, Y., and Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '01, 263–268. https://doi.org/10.1145/502512.502549

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry, 162(1), 156–159. https://doi.org/10.1016/0003-2697(87)90021-2

Costa, D. B., Li, S., Kocher, O., Feins, R. H., Keller, S. M., Schiller, J. H., Johnson, D. H., Tenen, D. G., and Halmos, B. (2007). Immunohistochemical analysis of C/EBPα in non-small cell lung cancer reveals frequent down-regulation in stage II and IIIA tumors: A correlative study of E3590. Lung Cancer, 56(1), 97–103. https://doi.org/10.1016/j.lungcan.2006.11.023

Cranenburg, E. C. M., Vermeer, C., Koos, R., Boumans, M.-L., Hackeng, T. M., Bouwman, F. G., Kwaijtaal, M., Brandenburg, V. M., Ketteler, M., and Schurgers, L. J. (2008). The circulating inactive form of matrix Gla Protein (ucMGP) as a biomarker for cardiovascular calcification. Journal of Vascular Research, 45(5), 427–436. https://doi.org/10.1159/000124863

Dong, X. Y., Wu, M. X., Zhang, H. M., Lyu, H., Qian, J. M., and Yang, H. (2020). Association between matrix Gla protein and ulcerative colitis according to DNA microarray data. Gastroenterology Report, 8(1), 66–75. https://doi.org/10.1093/gastro/goz038

Fan, C.-W., Chen, T., Shang, Y.-N., Gu, Y.-Z., Zhang, S.-L., Lu, R., OuYang, S.-R., Zhou, X., Li, Y., Meng, W.-T., Hu, J.-K., Lu, Y., Sun, X.-F., Bu, H., Zhou, Z.-G., and Mo, X.-M. (2013). Cancer-initiating cells derived from human rectal adenocarcinoma tissues carry mesenchymal phenotypes and resist drug therapies. Cell Death & Disease, 4, e828. https://doi.org/10.1038/cddis.2013.337

Fan, C., Sheu, D., Fan, H., Hsu, K., Allen Chang, C., and Chan, E. (2001). Down-regulation of matrix Gla protein messenger RNA in human colorectal adenocarcinomas. Cancer Letters, 165(1), 63–69. https://doi.org/10.1016/s0304-3835(01)00416-5

Farzaneh-Far, a, Proudfoot, D., Weissberg, P. L., & Shanahan, C. M. (2000). Matrix gla protein is regulated by a mechanism functionally related to the calcium-sensing receptor.

Biochemical and Biophysical Research Communications, 277(3), 736–740. https://doi.org/10.1006/bbrc.2000.3747

Farzaneh-Far, A., Davies, J. D., Braam, L. A., Spronk, H. M., Proudfoot, D., Chan, S. W., O'Shaughnessy, K. M., Weissberg, P. L., Vermeer, C., and Shanahan, C. M. (2001). A Polymorphism of the Human Matrix γ-Carboxyglutamic Acid Protein Promoter Alters Binding of an Activating Protein-1 Complex and Is Associated with Altered Transcription and Serum Levels. Journal of Biological Chemistry, 276(35), 32466–32473. https://doi.org/10.1074/jbc.M104909200

Farzaneh-Far, A., Weissberg, P. L., Proudfoot, D., and Shanahan, C. M. (2001). Transcriptional regulation of matrix gla protein. Zeitschrift Fur Kardiologie, 90(SUPPL. 3), 38–42. https://doi.org/10.1007/s003920170040

Fazenda, C., Simões, B., Kelsh, R. N., Cancela, M. L., and Conceição, N. (2010). Dual transcriptional regulation by runx2 of matrix Gla protein in Xenopus laevis. Gene, 450(1–2), 94–102. https://doi.org/10.1016/j.gene.2009.10.007

Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., Gavin, A., Visser, O., and Bray, F. (2018). Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. European Journal of Cancer. https://doi.org/10.1016/j.ejca.2018.07.005

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., and Bray, F. (2021). Cancer statistics for the year 2020: An overview. In International Journal of Cancer (Vol. 149, Issue 4, pp. 778–789). https://doi.org/10.1002/ijc.33588

Ferrari, N., McDonald, L., Morris, J. S., Cameron, E. R., and Blyth, K. (2013). RUNX2 in mammary gland development and breast cancer. Journal of Cellular Physiology, 228(6), 1137–1142. https://doi.org/10.1002/jcp.24285

Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 Function, a Paradigm for Transcription Factors in Hematopoiesis. In Molecular and Cellular Biology (Vol. 25, Issue 4, pp. 1215–1227). https://doi.org/10.1128/mcb.25.4.1215-1227.2005

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

Fu, M.-H., Wang, C.-Y., Hsieh, Y.-T., Fang, K.-M., and Tzeng, S.-F. (2017). Functional Role of Matrix gla Protein in Glioma Cell Migration. Molecular Neurobiology. https://doi.org/10.1007/s12035-017-0677-1

Fu, M., Wang, C., Hsieh, Y., and Fang, K. (2018). Functional Role of Matrix gla Protein in Glioma Cell Migration. 4624–4636. https://doi.org/10.1007/s12035-017-0677-1

Fu, X., Ji, F., He, Q., and Qiu, X. (2022). A Systematic Pan-Cancer Analysis of YY1 Aberrations and their Relationship with Clinical Outcome, Tumor Microenvironment, and Therapeutic Targets. In Journal of Immunology Research (Vol. 2022, pp. 1–23). https://doi.org/10.1155/2022/5826741

Galamb, O., Kalmár, A., Barták, B. K., Patai, Á. V., Leiszter, K., Péterfia, B., Wichmann, B., Valcz, G., Veres, G., Tulassay, Z., and Molnár, B. (2016). Aging related methylation influences the gene expression of key control genes in colorectal cancer and adenoma. World Journal of Gastroenterology, 22(47), 10325–10340.

https://doi.org/10.3748/wjg.v22.i47.10325

Gazzaniga, P., Gandini, O., Gradilone, A., Silvestri, I., Giuliani, L., Magnanti, M., Gallucci, M., Saccani, G., Frati, L., and Aglianò, A. M. (1999). Detection of basic Fibroblast Growth Factor mRNA in urinary bladder cancer: Correlation with local relapses. International Journal of Oncology, 14(6), 1123–1127. https://doi.org/10.3892/ijo.14.6.1123

George, M. L., Tutton, M. G., Abulafi, A. M., Eccles, S. A., and Swift, R. I. (2002). Plasma basic fibroblast growth factor levels in colorectal cancer: A clinically useful assay? Clinical and Experimental Metastasis, 19(8), 735–738. https://doi.org/10.1023/A:1021322201816

Gerdes, S., Osadtschy, S., Buhles, N., Baurecht, H., and Mrowietz, U. (2014). Cardiovascular biomarkers in patients with psoriasis. Experimental Dermatology, 23(5), 322–325. https://doi.org/10.1111/exd.12381

Gery, S., Tanosaki, S., Bose, S., Bose, N., Vadgama, J., and Koeffler, H. P. (2005). Down-regulation and growth inhibitory role of C/EBPα in breast cancer. Clinical Cancer Research, 11(9), 3184–3190. https://doi.org/10.1158/1078-0432.CCR-04-2625

Gheorghe, S. R., and Crăciun, A. M. (2015). Matrix Gla protein in tumoral pathology. Clujul Medical, 0(0), 319–321. https://doi.org/10.15386/cjmed-579

Gluba-Brzózka, A., Michalska, M., Franczyk-Skóra, B., Nocun, M., Banach, M., and Rysz, J. (2014). Markers of increased cardiovascular risk in patients with chronic kidney disease. Lipids in Health and Disease, 13(1), 135. https://doi.org/10.1186/1476-511X-13-135

Gong, C., Zou, J., Zhang, M., Zhang, J., Xu, S., Zhu, S., Yang, M., Li, D., Wang, Y., Shi, J., and Li, Y. (2019). Upregulation of MGP by HOXC8 promotes the proliferation, migration, and EMT processes of triple-negative breast cancer. Molecular Carcinogenesis, 58(10), 1863–1875. https://doi.org/10.1002/mc.23079

Gopalakrishnan, R., Ouyang, H., Somerman, M. J., McCauley, L. K., and Franceschi, R. T. (2001). Matrix γ-carboxyglutamic acid protein is a key regulator of pth-mediated inhibition of mineralization in MC3T3-E1 osteoblast-like cells. Endocrinology, 142(10), 4379–4388. https://doi.org/10.1210/endo.142.10.8413

Gordon, S., Akopyan, G., Garban, H., and Bonavida, B. (2006). Transcription factor YY1: Structure, function, and therapeutic implications in cancer biology. Oncogene, 25(8), 1125–1142. https://doi.org/10.1038/sj.onc.1209080

Greiner, M., Pfeiffer, D., and Smith, R. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive Veterinary Medicine, 45(1–2), 23–41. https://doi.org/10.1016/S0167-5877(00)00115-X

Guo, L., Guo, X. B., Jiang, J. L., Zhang, J. N., Ji, J., Liu, B. Y., Zhu, Z. G., and Yu, Y. Y. (2010). Discovery and verification of matrix gla protein, a TNM staging and prognosis-related biomarker for gastric cancer. In Chinese Journal of Pathology (Vol. 39, Issue 7, pp. 436–441). https://doi.org/10.3760/cma.j.issn.0529-5807.2010.07.002

Hale, J. E., Fraser, J. D., and Price, P. A. (1988a). The Identification of Matrix Gla Protein in Cartilage *. Journal of Biological Chemistry, 263(12), 5820–5824. https://doi.org/10.1016/S0021-9258(18)60639-8

Halmos, B., Karp, D. D., Kocher, O., Huettner, C. S., Tenen, D. G., and Ferenczi, K. (2002). Down-regulation and antiproliferative role of C/EBPα in lung cancer. Cancer Research, 62(2), 528–534.

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery, 12(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan, D., and Weinberg, R. A. (2000). The Hallmarks of Cancer. Cell, 100(1), 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. Cell, 144(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hassler, M. R., and Egger, G. (2012). Epigenomics of cancer - Emerging new concepts. Biochimie, 94(11), 2219–2230. https://doi.org/10.1016/j.biochi.2012.05.007

Hietakangas, V., Anckar, J., Blomster, H. A., Fujimoto, M., Palvimo, J. J., Nakai, A., and Sistonen, L. (2006). PDSM, a motif for phosphorylation-dependent SUMO modification. Proceedings of the National Academy of Sciences of the United States of America, 103(1), 45–50. https://doi.org/10.1073/pnas.0503698102

Hosler, B. A., Larosa, G. J., Grippo, J. F., and Gudas, L. J. (1989). Expression of REX-I, a Gene Containing Zinc Finger Motifs, Is Rapidly Reduced by Retinoic Acid in F9 Teratocarcinoma Cells. Molecular and Cellular Biology, 9(12), 5623–5629. https://doi.org/10.1080/0020739840150214 _3

Hough, C. D., Cho, K. R., Zonderman, A. B., Schwartz, D. R., and Morin, P. J. (2001). Coordinately up-regulated genes in ovarian cancer. Cancer Research, 61(10), 3869–3876.

Huang, C., Wang, M., Wang, J., Wu, D., Gao, Y., Huang, K., and Yao, X. (2021). Suppression MGP inhibits tumor proliferation and reverses oxaliplatin resistance in colorectal cancer. Biochemical Pharmacology, 189(December 2020), 114390. https://doi.org/10.1016/j.bcp.2020.114390

Inbar-Feigenberg, M., Choufani, S., Butcher, D. T., Roifman, M., and Weksberg, R. (2013). Basic concepts of epigenetics. Fertility and Sterility, 99(3), 607–615. https://doi.org/10.1016/j.fertnstert.2013.01.117

Javaid, N., and Choi, S. (2017). Acetylation- and methylation-related epigenetic proteins in the context of their targets. Genes, 8(8). https://doi.org/10.3390/genes8080196

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics, 13(7), 484–492. https://doi.org/10.1038/nrg3230

Jones, P. A, and Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. Nature Reviews. Genetics, 3(6), 415–428. https://doi.org/10.1038/nrg816

Karsli Ceppioğlu, S., Yurdun, T., and Canbakan, M. (2011). Assessment of matrix Gla protein, Klotho gene polymorphisms, and oxidative stress in chronic kidney disease. Renal Failure, 33(9), 866–874. https://doi.org/10.3109/0886022X.2011.605534

Kesanakurti, D., Maddirela, D., Gujrati, M., Rao, J. S. (2013). Critical role of MMP-2 and GATA-1 nuclear interaction in the regulation of ionizing radiation (IR)-induced IL-

10/STAT3-mediated invasion and migration in glioblastoma. Cancer Research, 73, 420–420. https://doi.org/10.1158/1538-7445.AM2013-420

Ki, D. H., Jeung, H., Park, C. H., Kang, S. H., Lee, G. Y., Lee, W. S., Kim, N. K., Chung, H. C., and Rha, S. Y. (2007). Whole genome analysis for liver metastasis gene signatures in colorectal cancer. International Journal of Cancer, 121(9), 2005–2012. https://doi.org/10.1002/ijc.22975

Kirfel, J., Kelter, M., Cancela, L. M., Price, P. A., and Schüle, R. (1997a). Identification of a novel negative retinoic acid responsive element in the promoter of the human matrix Gla protein gene. Proceedings of the National Academy of Sciences, 94(6), 2227–2232. https://doi.org/10.1073/pnas.94.6.2227

Konopka, B., Szafron, L. M., Kwiatkowska, E., Podgorska, A., Zolocinska, A., Pienkowska-Grela, B., Dansonka-Mieszkowska, A., Balcerak, A., Lukasik, M., Stachurska, A., Timorek, A., Spiewankiewicz, B., El-Bahrawy, M., and Kupryjanczyk, J. (2016). The significance of c.690G>T polymorphism (rs34529039) and expression of the CEBPA gene in ovarian cancer outcome. Oncotarget, 7(41), 67412–67424. https://doi.org/10.18632/oncotarget.11822

Kungulovski, G., and Jeltsch, A. (2016). Epigenome Editing: State of the Art, Concepts, and Perspectives. Trends in Genetics, 32(2), 101–113. https://doi.org/10.1016/j.tig.2015.12.001

Kuzontkoski, P. M., Mulligan-Kehoe, M. J., Harris, B. T., and Israel, M. A. (2010). Inhibitor of DNA binding-4 promotes angiogenesis and growth of glioblastoma multiforme by elevating matrix GLA levels. Oncogene, 29(26), 3793–3802. https://doi.org/10.1038/onc.2010.147

Laizé, V., Martel, P., Viegas, C. S. B., Price, P. A., and Cancela, M. L. (2005). Evolution of matrix and bone γ-carboxyglutamic acid proteins in vertebrates. Journal of Biological Chemistry, 280(29), 26659–26668. https://doi.org/10.1074/jbc.M500257200

Lamonica, J. M., Vakoc, C. R., and Blobel, G. A. (2006). Acetylation of GATA-1 is required for chromatin occupancy. In Blood (Vol. 108, Issue 12, pp. 3736–3738). https://doi.org/10.1182/blood-2006-07-032847

Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J., and McKnight, S. L. (1988). Isolation of a recombinant copy of the gene encoding C/EBP. Genes & Development, 2(7), 786–800. https://doi.org/10.1101/gad.2.7.786

Leão, R., Apolónio, J. D., Lee, D., Figueiredo, A., Tabori, U., and Castelo-Branco, P. (2018). Mechanisms of human telomerase reverse transcriptase (hTERT) regulation: clinical impacts in cancer. Journal of Biomedical Science, 25(1), 22. https://doi.org/10.1186/s12929-018-0422-8

Lee, C. M., Cho, J. M., Jang, Y. J., Park, S. S., Park, S. H., Kim, S. J., Mok, Y. J., Kim, C. S., and Kim, J. H. (2015). Should Lymph Node Micrometastasis be Considered in Node Staging For Gastric Cancer?: The Significance of Lymph Node Micrometastasis in Gastric Cancer. Annals of Surgical Oncology, 22(3), 765–771. https://doi.org/10.1245/s10434-014-4073-z

Levedakou, E. N., Strohmeyer, T. G., Effert, P. J., and Liu, E. T. (1992). Expression of the

matrix Gla protein in urogenital malignancies. International Journal of Cancer. Journal International Du Cancer, 52(4), 534–537.

Li, L., Zhang, S., Wei, L., Wang, Z., Ma, W., Liu, F., and Qian, Y. (2018). FGF2 and FGFR2 in patients with idiopathic pulmonary fibrosis and lung cancer. Oncology Letters, 16(2), 2490–2494. https://doi.org/10.3892/ol.2018.8903

Li, X., Wei, R., Wang, M., Ma, L., Zhang, Z., Chen, L., Guo, Q., Guo, S., Zhu, S., Zhang, S., and Min, L. (2020). MGP Promotes Colon Cancer Proliferation by Activating the NF- k B Pathway through Upregulation of the Calcium Signaling Pathway. 17(June), 371–383. https://doi.org/10.1016/j.omto.2020.04.005

Lu, G., Leung, C. H., Yan, B., Tan, C. M., Low, S. Y., Aung, M. O., Salto–Tellez, M., Lim, S. G., and Hooi, S. C. (2010). C/EBPα Is Up-regulated in a Subset of Hepatocellular Carcinomas and Plays a Role in Cell Growth and Proliferation. Gastroenterology, 139(2), 632-643.e4. https://doi.org/10.1053/j.gastro.2010.03.051

Luo, G., Ducy, P., Mckee, M. D., Pinero, G. J., Loyer, E., Behringer, R. R., and Karsenty, G. (1997). Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. Nature, 386(6).

Malvezzi, M., Carioli, G., Bertuccio, P., Boffetta, P., Levi, F., Vecchia, C. La, and Negri, E. (2018). European cancer mortality predictions for the year 2018 with focus on colorectal cancer. Annals of Oncology, 29(4), 1016–1022. https://doi.org/10.1093/annonc/mdy033

Martini, M., Cenci, T., D'Alessandris, G. Q., Cesarini, V., Cocomazzi, A., Ricci-Vitiani, L., De Maria, R., Pallini, R., and Larocca, L. M. (2013). Epigenetic silencing of Id4 identifies a glioblastoma subgroup with a better prognosis as a consequence of an inhibition of angiogenesis. Cancer, 119(5), 1004–1012. https://doi.org/10.1002/cncr.27821

Mayer, O., Seidlerová, J., Bruthans, J., Filipovský, J., Timoracká, K., Vaněk, J., Cerná, L., Wohlfahrt, P., Cífková, R., Theuwissen, E., and Vermeer, C. (2014). Desphospho-uncarboxylated matrix Gla-protein is associated with mortality risk in patients with chronic stable vascular disease. Atherosclerosis, 235(1), 162–168. https://doi.org/10.1016/j.atherosclerosis.2014.04.027

Medvedeva, Y. A., Khamis, A. M., Kulakovskiy, I. V, Ba-Alawi, W., Bhuyan, M. S. I., Kawaji, H., Lassmann, T., Harbers, M., Forrest, A. R., and Bajic, V. B. (2014). Effects of cytosine methylation on transcription factor binding sites. BMC Genomics, 15(1), 119. https://doi.org/10.1186/1471-2164-15-119

Mendelsohn, A. C., Howley, A., Israel, S., Gray, J. E., and Thompson, T. (2008). The molecular Basis of Cancer (3rd ed.). Elsivier.

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

Micke, P., Kappert, K., Ohshima, M., Sundquist, C., Scheidl, S., Lindahl, P., Heldin, C.-H., Botling, J., Ponten, F., and Ostman, A. (2007). In situ identification of genes regulated specifically in fibroblasts of human basal cell carcinoma. The Journal of Investigative Dermatology, 127(6), 1516–1523. https://doi.org/10.1038/sj.jid.5700714

Mishra, A., Sriram, H., Chandarana, P., Tanavde, V., Kumar, R. V., Gopinath, A., Govindarajan, R., Ramaswamy, S., and Sadasivam, S. (2018). Decreased expression of cell adhesion genes in cancer stem-like cells isolated from primary oral squamous cell carcinomas. Tumor Biology, 40(6), 1–15. https://doi.org/10.1177/1010428318780859

Munroe, P. B., Olgunturk, R. O., Fryns, J.-P., Maldergem, L. Van, Ziereisen, F., Yuksel, B., Gardiner, R. M., and Chung, E. (1999). Mutations in the gene encoding the human matrix Gla protein cause Keutel syndrome. Nature Genetics, 21(1), 142–144. https://doi.org/10.1038/5102

Notterman, D. A., Alon, U., Sierk, A. J., and Levine, A. J. (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Research, 61(7), 3124–3130.

Park, J. H., Jung, M., and Moon, K. C. (2018). The prognostic significance of nuclear expression of PHF2 and C/EBPα in clear cell renal cell carcinoma with consideration of adipogenic metabolic evolution. Oncotarget, 9(1), 142–151. https://doi.org/10.18632/oncotarget.19949

Peters, I., Dubrowinskaja, N., Tezval, H., Kramer, M. W., von Klot, C. A., Hennenlotter, J., Stenzl, A., Scherer, R., Kuczyk, M. A., and Serth, J. (2015). Decreased mRNA expression of GATA1 and GATA2 is associated with tumor aggressiveness and poor outcome in clear cell renal cell carcinoma. Targeted Oncology, 10(2), 267–275. https://doi.org/10.1007/s11523-014-0335-8

Post, A. E. M., Bussink, J., Smid, M., Sweep, F. C. G. J., and Span, P. N. (2019). Downregulation of matrix Gla protein is a biomarker for tamoxifen-resistant and radioresistant breast cancer. 13, 841–850.

Price, P. A., Otsuka, A. A., Poser, J. W., Kristaponis, J., and Raman, N. (1976). Characterization of a gamma-carboxyglutamic acid-containing protein from bone. Proceedings of the National Academy of Sciences of the United States of America, 73(5), 1447–1451. https://doi.org/10.1073/pnas.73.5.1447

Price, P. A., Poser, J. W., and Raman, N. (1976). Primary structure of the -y-carboxyglutamic acid-containing protein from bovine bone Biochemistry : 73(10), 3374–3375.

Price, P. A., Urist, M. R., and Otawara, Y. (1983). Matrix Gla protein, a new γ-carboxyglutamic acid-containing protein which is associated with the organic matrix of bone. Biochemical and Biophysical Research Communications, 117(3), 765–771. https://doi.org/10.1016/0006-291X(83)91663-7

Price, P. A., and Williamson, M. K. (1985). Primary structure of bovine matrix Gla protein, a new vitamin K-dependent bone protein. Journal of Biological Chemistry, 260(28), 14971–14975. https://doi.org/10.1016/s0021-9258(18)95688-7

Raich, N., Clegg, C. H., Grofti, J., Roméo, P. H., and Stamatoyannopoulos, G. (1995). GATA1 and YY1 are developmental repressors of the human ε-globin gene. EMBO Journal, 14(4), 801–809. https://doi.org/10.1002/j.1460-2075.1995.tb07058.x

Regalo, G., Resende, C., Wen, X., Gomes, B., Durães, C., Seruca, R., Carneiro, F., and Machado, J. C. (2010). C/EBPα expression is associated with homeostasis of the gastric epithelium and with gastric carcinogenesis. Laboratory Investigation, 90(8), 1132–1139.

https://doi.org/10.1038/labinvest.2010.79

Ren, J., Li, G., Ge, J., Li, X., and Zhao, Y. (2012). Is K-ras Gene Mutation a Prognostic Factor for Colorectal Cancer. Diseases of the Colon & Rectum, 55(8), 913–923. https://doi.org/10.1097/DCR.0b013e318251d8d9

Roberto, V. P., Gavaia, P., Nunes, M. J., Rodrigues, E., Cancela, M. L., and Tiago, D. M. (2018). Evidences for a New Role of miR-214 in Chondrogenesis. Scientific Reports, 8(1), 1–16. https://doi.org/10.1038/s41598-018-21735-w

Rong, D., Sun, G., Zheng, Z., Liu, L., Chen, X., Wu, F., Gu, Y., Dai, Y., Zhong, W., Hao, X., Zhang, C., Pan, X., Tang, J., Tang, W., and Wang, X. (2022). MGP promotes CD8+ T cell exhaustion by activating the NF-κB pathway leading to liver metastasis of colorectal cancer. International Journal of Biological Sciences, 18(6), 2345–2361. https://doi.org/10.7150/ijbs.70137

Ruppert, J. M., Kinzler, K. W., Wong, A. J., Bigner, S. H., Kao, F.-T., Law, M. L., Seuanez, H. N., O 'brien, S. J., and Vogelstein1, B. (1988). The GLI-Kruppel Family of Human Genes. Molecular and Cellular Biology, 8(8), 3104–3113. https://doi.org/10.1128/MCB.8.8.3104.Updated

Sabates-Bellver, J., Van Der Flier, L. G., De Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M. A., Bujnicki, J. M., Menigatti, M., Luz, J., Ranalli, T. V., Gomes, V., Pastorelli, A., Faggiani, R., Anti, M., Jiricny, J., Clevers, H., and Marra, G. (2007). Transcriptome profile of human colorectal adenomas. Molecular Cancer Research, 5(12), 1263–1275. https://doi.org/10.1158/1541-7786.MCR-07-0267

Sase, T., Suzuki, T., Miura, K., Shiiba, K., Sato, I., Nakamura, Y., Takagi, K., Onodera, Y., Miki, Y., Watanabe, M., Ishida, K., Ohnuma, S., Sasaki, H., Sato, R., Karasawa, H., Shibata, C., Unno, M., Sasaki, I., and Sasano, H. (2012). Runt-related transcription factor 2 in human colon carcinoma: a potent prognostic factor associated with estrogen receptor. International Journal of Cancer. Journal International Du Cancer, 131(10), 2284–2293. https://doi.org/10.1002/ijc.27525

Schurgers, L. J., Barreto, D. V, Barreto, F. C., Liabeuf, S., Renard, C., Magdeleyns, E. J., Vermeer, C., Choukroun, G., and Massy, Z. A. (2010). The circulating inactive form of matrix gla protein is a surrogate marker for vascular calcification in chronic kidney disease: a preliminary report. Clinical Journal of the American Society of Nephrology : CJASN, 5(4), 568–575. https://doi.org/10.2215/CJN.07081009

Schurgers, L. J., Cranenburg, E. C. M., and Vermeer, C. (2008). Matrix Gla-protein: The calcification inhibitor in need of vitamin K. Thrombosis and Haemostasis. https://doi.org/10.1160/TH08-02-0087

Schuster, M. B., and Porse, B. T. (2006). C/EBPα: A tumour suppressor in multiple tissues? Biochimica et Biophysica Acta - Reviews on Cancer, 1766(1), 88–103. https://doi.org/10.1016/j.bbcan.2006.02.003

Shanahan, C. M., Weissberg, P. L., and Metcalfe, J. C. (1993). Isolation of gene markers of differentiated and proliferating vascular smooth muscle cells. Circulation Research, 73(1), 193–204. https://doi.org/10.1161/01.RES.73.1.193

Sharma, B., and Albig, A. R. (2013). Matrix Gla protein reinforces angiogenic resolution.

Microvascular Research, 85, 24–33. https://doi.org/10.1016/j.mvr.2012.10.005

Shiomi, A., Kusuhara, M., Sugino, T., and Sugiura, T. (2021). Comprehensive genomic analysis contrasting primary colorectal cancer and matched liver metastases. 1–13. https://doi.org/10.3892/ol.2021.12727

Slattery, M. L., Lundgreen, A., Herrick, J. S., Caan, B. J., Potter, J. D., and Wolff, R. K. (2011). Associations between genetic variation in RUNX1, RUNX2, RUNX3, MAPK1, and eIF4E and riskof colon and rectal cancer: additional support for a TGF-β-signaling pathway. Carcinogenesis, 32(3), 318–326. https://doi.org/10.1093/carcin/bgq245

Soulitzis, N., Karyotis, I., Delakas, D., and Spandidos, D. A. (2006). Expression analysis of peptide growth factors VEGF, FGF2, TGFB1, EGF and IGF1 in prostate cancer and benign prostatic hyperplasia. International Journal of Oncology, 29(2), 305–314. https://doi.org/10.3892/ijo.29.2.305

Stenflo, J., Fernlund, P., Egan, W., and Roepstorff, P. (1974). Vitamin K dependent modifications of glutamic acid residues in prothrombin. Proceedings of the National Academy of Sciences of the United States of America, 71(7), 2730–2733. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=388542&tool=pmcentrez&rendertype=abstract

Sterzyńska, K., Klejewski, A., Wojtowicz, K., Świerczewska, M., Andrzejewska, M., Rusek, D., Sobkowski, M., Kędzia, W., Brazert, J., Nowicki, M., and Januchowski, R. (2018). The role of matrix gla protein (MGP) expression in paclitaxel and topotecan resistant ovarian cancer cell lines. International Journal of Molecular Sciences, 19(10). https://doi.org/10.3390/ijms19102901

Stheneur, C., Dumontier, M. F., Guedes, C., Fulchignoni-Lataud, M. C., Tahiri, K., Karsenty, G., and Corvol, M. T. (2003). Basic fibroblast growth factor as a selective inducer of matrix Gla protein gene expression in proliferative chondrocytes. Biochemical Journal, 369(1), 63–70. https://doi.org/10.1042/BJ20020549

Stock, M., Schäfer, H., Fliegauf, M., and Otto, F. (2004). Identification of Novel Target Genes of the Bone-Specific Transcription Factor Runx2. Journal of Bone and Mineral Research, 19(6), 959–972. https://doi.org/10.1359/jbmr.2004.19.6.959

Suttamanatwong, S., Franceschi, R. T., Carlson, A. E., and Gopalakrishnan, R. (2007). Regulation of matrix Gla protein by parathyroid hormone in MC3T3-E1 osteoblast-like cells involves protein kinase A and extracellular signal-regulated kinase pathways. Journal of Cellular Biochemistry, 102(2), 496–505. https://doi.org/10.1002/jcb.21314

Suttamanatwong, S., Jensen, E. D., Schilling, J., Franceschi, R. T., Carlson, A. E., Mansky, K. C., and Gopalakrishnan, R. (2009). Sp proteins and Runx2 mediate regulation of matrix gla protein (MGP) expression by parathyroid hormone. Journal of Cellular Biochemistry, 107(2), 284–292. https://doi.org/10.1002/jcb.22124

Suzuki, H., Maruyama, R., Yamamoto, E., and Kai, M. (2013). Epigenetic alteration and microRNA dysregulation in cancer. Frontiers in Genetics, 4(DEC). https://doi.org/10.3389/fgene.2013.00258

Takahashi, T., Kato, S., Suzuki, N., Kawabata, N., and Takagi, M. (2005). Autoregulatory mechanism of Runx2 through the expression of transcription factors and bone matrix

proteins in multipotential mesenchymal cell line, ROB-C26. Journal of Oral Science, 47(4), 199–207. http://www.ncbi.nlm.nih.gov/pubmed/16415564

Tenen, D. G. (2001). Abnormalities of the CEBP alpha transcription factor: A major target in acute myeloid leukemia. Leukemia, 15(4), 688–689. https://doi.org/10.1038/sj.leu.2402088

Tiago, D. M., Conceição, N., Caiado, H., Laizé, V., and Cancela, M. L. (2016). Matrix Gla protein repression by MIR-155 promotes oncogenic signals in breast cancer MCF-7 cells. FEBS Letters, 590(8), 1234–1241. https://doi.org/10.1002/1873-3468.12155

Tong, J. H. M., Lung, R. W. M., Sin, F. M. C., Law, P. P. Y., Kang, W., Chan, A. W. H., Ma, B. B. Y., Mak, T. W. C., Ng, S. S. M., and To, K. F. (2014). Characterization of rare transforming KRAS mutations in sporadic colorectal cancer. Cancer Biology & Therapy, 15(6), 768–776. https://doi.org/10.4161/cbt.28550

Tsigris, C., Karayiannakis, A. J., Zbar, A., Syrigos, K. N., Baibas, N., Diamantis, T., and Alexiou, D. (2002). Clinical significance of serum and urinary c-erbB-2 levels in colorectal cancer. Cancer Letters, 184(2), 215–222. https://doi.org/10.1016/S0304-3835(02)00205-7

Tuo, Y., and Ye, Y. (2017). MGP is downregulated due to promoter methylation in chemoresistant ER + breast cancer and high MGP expression predicts better survival outcomes. 3871–3878.

Van den Boom, J., Wolter, M., Kuick, R., Misek, D. E., Youkilis, A. S., Wechsler, D. S., Sommer, C., Reifenberger, G., and Hanash, S. M. (2003). Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. American Journal of Pathology, 163(3), 1033–1043. https://doi.org/10.1016/S0002-9440(10)63463-3

Vassalle, C., and Iervasi, G. (2014). New insights for matrix Gla protein, vascular calcification and cardiovascular risk and outcome. Atherosclerosis, 235(1), 236–238. https://doi.org/10.1016/j.atherosclerosis.2014.04.037

Vermeer, C., Drummen, N. E. A., Knapen, M. H. J., and Zandbergen, F. J. (2015). Uncarboxylated Matrix Gla Protein as a Biomarker in Cardiovascular Disease: Applications for Research and for Routine Diagnostics. In Biomarkers in Cardiovascular Disease (pp. 1–17). Springer Netherlands. https://doi.org/10.1007/978-94-007-7741-5_14-1

Wai, P. Y., Mi, Z., Gao, C., Guo, H., Marroquin, C., and Kuo, P. C. (2006). Ets-1 and runx2 regulate transcription of a metastatic gene, osteopontin, in murine colorectal cancer cells. The Journal of Biological Chemistry, 281(28), 18973–18982. https://doi.org/10.1074/jbc.M511962200

Wang, C. C., Chen, J. J. W., and Yang, P. C. (2006). Multifunctional transcription factor YY1: A therapeutic target in human cancer? Expert Opinion on Therapeutic Targets, 10(2), 253–266. https://doi.org/10.1517/14728222.10.2.253

Wang, M., Chen, L., Chen, Y., Wei, R., Guo, Q., Zhu, S., and Guo, S. (2020). Intracellular matrix Gla protein promotes tumor progression by activating JAK2 / STAT5 signaling in

gastric cancer. 14, 1045–1058. https://doi.org/10.1002/1878-0261.12652

Weaver, K. N., El Hallek, M., Hopkin, R. J., Sund, K. L., Henrickson, M., Del Gaudio, D., Yuksel, A., Acar, G. O., Bober, M. B., Kim, J., and Boyadjiev, S. A. (2014). Keutel syndrome: report of two novel MGP mutations and discussion of clinical overlap with arylsulfatase E deficiency and relapsing polychondritis. American Journal of Medical Genetics. Part A, 164A(4), 1062–1068. https://doi.org/10.1002/ajmg.a.36390

Weiser, M. R. (2018). AJCC 8th Edition: Colorectal Cancer. Annals of Surgical Oncology, 25(6), 1454–1455. https://doi.org/10.1245/s10434-018-6462-1

Wu, X., Ruan, L., Yang, Y., and Mei, Q. (2017). Analysis of gene expression changes associated with human carcinoma-associated fibroblasts in non-small cell lung carcinoma. Biological Research, 50(1), 1–8. https://doi.org/10.1186/s40659-017-0108-9

Xu, J., Cheng, K., Lin, H., Han, W., He, T., Nie, X., Sun, Y., Qiuman, S., Reheman, Y., and Chen, Q. (2021). Characterization of the GATA Transcription Factor Family and Exploration of Their Relevance to Immune Infiltration and Tumor Microenvironment in Pancreatic Cancer. International Journal of General Medicine, Volume 14(November), 9083–9101. https://doi.org/10.2147/IJGM.S342741

Yagami, K., Suh, J. Y., Enomoto-Iwamoto, M., Koyama, E., Abrams, W. R., Shapiro, I. M., Pacifici, M., and Iwamoto, M. (1999). Matrix GLA protein is a developmental regulator of chondrocyte mineralization and, when constitutively expressed, blocks endochondral and intramembranous ossification in the limb. Journal of Cell Biology, 147(5), 1097–1108. https://doi.org/10.1083/jcb.147.5.1097

Yoshimura, K., Takeuchi, K., Nagasaki, K., Ogishima, S., Tanaka, H., Iwase, T., Akiyama, F., Kuroda, Y., and Miki, Y. (2009). Prognostic value of matrix Gla protein in breast cancer. Molecular Medicine Reports, 2(4), 549–553. https://doi.org/10.3892/mmr_00000135

Yu, J., Liu, M., Liu, H., and Zhou, L. (2019). GATA1 promotes colorectal cancer cell proliferation, migration and invasion via activating AKT signaling pathway. In Molecular and Cellular Biochemistry (Vol. 457, Issues 1–2, pp. 191–199). https://doi.org/10.1007/s11010-019-03523-w

Zhang, S., Gao, F., Luo, J., and Yang, J. (2010). Prognostic factors in survival of colorectal cancer patients with synchronous liver metastasis. Colorectal Disease, 12(8), 754–761. https://doi.org/10.1111/j.1463-1318.2009.01911.x

Zhang, Y., Liu, J., Lin, J., Zhou, L., Song, Y., Wei, B., Luo, X., Chen, Z., Chen, Y., Xiong, J., Xu, X., Ding, L., and Ye, Q. (2016). The transcription factor GATA1 and the histone methyltransferase SET7 interact to promote VEGF-mediated angiogenesis and tumor growth and predict clinical outcome of breast cancer. Oncotarget, 7(9), 9859–9875. https://doi.org/10.18632/oncotarget.7126

Zhou, H. L., Luo, G., Wise, J. A., and Lou, H. (2014). Regulation of alternative splicing by local histone modifications: Potential roles for RNA-guided mechanisms. Nucleic Acids Research, 42(2), 701–713. https://doi.org/10.1093/nar/gkt875