UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

DEPARTAMENTO DE BIOLOGIA CELULAR E MOLECULAR E BIOAGENTES PATOGÊNICOS

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

**Estudo do gene EMC2 em câncer de mama:
abordagens de bioinformática e funcionais**

Orientadora: Profa. Dra. Enilza Maria Espreadico
Aluna: Marcela Motta de Castro

Versão corrigida após sessão de defesa realizada no dia 15 de Junho de 2018

Ribeirão Preto - SP
2018

MARCELA MOTTA DE CASTRO

Estudo do gene EMC2 em câncer de mama: abordagens de bioinformática e funcionais

Relatório final, apresentado a Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de mestre.

Ribeirão Preto, 15 de Junho de 2018.

BANCA EXAMINADORA

Prof. Leticia Fröhlich Archangelo
Afiliações

Prof. Andréia Machado Leopoldino
Afiliações

Prof. Robson Francisco Carvalho
Afiliações

DEDICATÓRIAS

Ao meu amigo Lucas Chiaretti da Cruz, que passou por um câncer de testículo e me despertou o desejo de estudar o câncer. Dedico também a minha querida avó Maria Faria de Castro, que passou por um câncer de mama. E dedico também a todas as pessoas que tem ou tiveram essa doença.

AGRADECIMENTOS

Há muita gratidão envolvida neste trabalho!

Agradeço a Deus, pelo dom da vida, pelo equilíbrio espiritual e emocional, pelos ensinamentos de amor que guiam meus passos!

Aos meus pais, Nei e Carminha, sou infinitamente grata, por serem minha base, acreditarem sempre em mim, pelo amor incondicional e por terem me ensinado os caminhos coerentes da vida. Agradeço a minha irmã Leticia, por ser minha companheira e alicerce, assim como meu querido cunhado Gustavo.

Agradeço a toda minha família, pelo abundante amor, exemplo, educação e carinho!!!

Ao meu namorado Alan, por me contagiar com sua alegria, e pelo companheirismo nos momentos da vida incluindo me acompanhar nos experimentos a noite e finais de semana no laboratório.

Agradeço a professora Enilza Maria Espreafico, pelo amor a ciência que irradia e atinge a todos ao seu redor! Sou grata pelo sim que ela me deu, acreditando que uma física médica pudesse aprender um pouco de biologia rs. Agradeço por tudo que me ensinou e colaborou para a execução deste trabalho.

Aos meus amigos de laboratório, que são fortemente responsáveis pelo resultado deste trabalho. Agradeço ao Lucas pelo apoio inicial, e todos ensinamentos sobre bioinformática, ao Rui pela vigilante disponibilidade de ajuda, ao Johnny pelas dicas de protocolo e ao Luis e Jackeline pelos auxílios diários nos experimentos e pela parceria em experimentos cruciais de meu mestrado. Agradeço aos integrantes que passaram pelo laboratório, Carlos e Maiaro que encaminharam meus primeiros experimentos, assim como a Mariane e a Fernanda. Agradeço a amizade de Talita, Iuly, Raquel, Gabriel e Camila!!! E aos novos integrantes do laboratório, Barbara, Santiago e Luis, sou grata pelas conversas no intervalo e pelas ajudas na escrita desta dissertação. Não poderia faltar uma peça fundamental do nosso laboratório: a técnica Joana, que além de todo suporte técnico, coloca-se como “mãe” dos alunos na ajuda por quaisquer tipos de problema. Agradeço também a técnica Silmara, que se aposentou no início de meu mestrado e sempre

contribuiu muito para o nosso laboratório. E as técnicas dos laboratórios multiusuários Bete, Roberta, Vani e Tânia por toda ajuda e contribuição nos resultados.

Agradeço também ao professor Alfredo Ribeiro da Silva pela paciência nos experimentos de imunohistoquímica e a sua técnica Laura pelo capricho nos experimentos.

Agradeço aos meus amigos da graduação, que mesmo alguns estando distantes, sabemos que a amizade está viva e que torcemos muito pela felicidade uns dos outros. Agradeço a Luciana, Enock, Sarah e Débora pelo apoio de sempre.

E aos meus amigos de infância, Fernanda, Juliana, Stephania e Caroline, que são como partes da minha família.

Gostaria de agradecer as agências de fomento CNPQ, Capes, Fapesp e FAEPA pelo apoio financeiro que viabilizou a execução desse trabalho.

THE
AWKWARD YETI

THEAWKWARDYETI.COM

©2014 THE AWKWARD YETI

099

@THEAWKWARDYETI

FACEBOOK.COM/AWKWARDYETI

RESUMO

DE CASTRO, M. M., **Estudo do gene EMC2 em câncer de mama: abordagens de bioinformática e funcionais**. Dissertação (Mestrado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo – USP, Ribeirão Preto, 2018.

Em mulheres, o câncer de mama é o tipo mais incidente depois do tumor de pele não melanoma e é uma das principais causas de morte por câncer. Apesar dos avanços já alcançados na caracterização da doença, a busca por novos marcadores moleculares para diagnóstico, tratamento e entendimento molecular da doença é de extrema importância. Estudos em nosso laboratório apontaram a proteína EMC1 (do inglês, Endoplasmic Reticulum Complex 1) como relacionada a propriedades malignas em linhagens celulares de câncer de mama e melanoma, incluindo aumento no crescimento tumoral em ensaios *in vivo*. Como essa proteína faz parte do complexo EMC, formado por 10 proteínas, despertou-se em nosso laboratório o interesse no estudo das demais proteínas do complexo. Assim, o presente estudo tem início com a análise de dados em larga escala do banco The Cancer Genome Atlas (TCGA) em um painel de 32 tipos de câncer. Os resultados obtidos indicaram uma significativa associação de expressão elevada de diversos genes EMCs com pior sobrevida dos pacientes. O gene EMC2, que se localiza na região cromossômica altamente amplificada em diversos tumores (8q23.1), se destaca pela alta frequência (40%) de pacientes com superexpressão em câncer de mama entre todos os casos do TCGA (n=960). Em linhagens de câncer de mama, o *knockdown* de EMC2 causa redução na taxa proliferativa, assim como um atraso na progressão do ciclo celular através da fase M, após a remoção da droga despolimerizante de microtúbulos, nocodazol, utilizada para sincronização das células em mitose. Além disso, o *knockdown* de EMC2 causou mudanças na morfologia celular sugestivas de transição epitélio-mesenquima, o que pode justificar a menor proporção de pacientes com marcação de EMC2 média e alta em pacientes com câncer de mama em estágios mais avançados, feitos por imuno-histoquímica. Em conjunto, nossos dados sugerem que o complexo EMC pode favorecer o desenvolvimento de tumores e influenciar a sua malignidade. Além disso, a desregulação nos níveis de EMC2 pode afetar diversas características celulares, incluindo proliferação, ciclo celular e transição epitélio-mesenquima.

Palavras-chaves: Complexo EMC, bioinformática, EMC2, câncer de mama, proliferação celular, ciclo celular, transição epitélio-mesenquima.

ABSTRACT

DE CASTRO, M. M., **A study of the EMC2 gene in breast cancer: bioinformatics and functional approaches**. Dissertation (Master) – Ribeirão Preto Medical School, University of São Paulo – USP, Ribeirão Preto, 2018.

In women, breast cancer is the most common type of cancer, excluding non-melanoma skin cancers, and it is one of the leading cause of death from cancer. Despite the advances already achieved in the disease characterization, the search for novel molecular biomarkers of diagnostic, treatment and disease investigation is extremely important. Studies performed in our laboratory pointed Endoplasmic Reticulum Complex 1 (EMC1) involvement in malignant properties in breast cancer and melanoma cell lines, including the promotion of tumor growth in *in vivo* assays. As this protein is part of the EMC complex, consisting of 10 proteins, these findings prompted us to investigate other members of the EMC complex. Thus, in the present study, we began by analyzing a large-scale database from The Cancer Genome Atlas (TCGA) in a range of 32 cancer type. The results indicated a significant association of high expression of EMC genes with poor survival outcome, in many instances. The EMC2 gene, located in a commonly amplified chromosomal region in cancer (8q23.1), is upregulated in 40% of all breast cancer cases (n=960) from TCGA. In breast cancer cells, EMC2 knockdown led to reduced proliferation rates and delayed progression through M phase of the cell cycle, after wash out of microtubule depolymerizing drug Nocodazole used for synchronization of the cells in mitosis. In addition, EMC2 knockdown caused changes in cell morphology suggestive of epithelial-mesenchymal transition, which may account for the lower proportion of patients with medium and high EMC2 labeling in patients with advanced stage breast cancer, made by immunohistochemistry. Together, the our data suggest that the EMC complex favors tumor growth and may influence their malignance. Additionally, a deregulation in EMC2 levels may affect several cellular characteristics, including proliferation, cell cycle and epithelial-mesenchymal transition.

Key words: EMC complex, bioinformatics, EMC2, breast cancer, cell proliferation, cell cycle, epithelial-mesenchymal transition.

LISTA DE FIGURAS

Figura 1 – Estrutura da glândula mamária. (Adaptado de Dimri et al., 2005)

Figura 2 – Modelo de progressão do câncer de mama. (Adaptado de Polyak et al., 2007)

Figura 3 – Modelo esquemático das proteínas do complexo EMC de mamíferos de acordo com as estruturas topológicas preditas. (Molina et al., 2014)

Figura 4 – Isoformas de EMC2. Adaptado de http://atlasgeneticsoncology.org/Genes/GC_EMC2

Figura 5- Comparação entre os níveis de expressão gênica do tecido canceroso e do tecido normal adjacente ao tumor dos genes EMC1-10, utilizando os dados do TCGA. Os dados foram obtidos através do software cBioPortal. Os valores são apresentados como diagrama de caixa Tukey e analisados estatisticamente pelo teste t de duas caudas paramétrico de amostras não pareadas, onde o valor de p é representado por * <0.05, **<0.01, ***<0.001 e ****<0.0001.

Figura 6- Alterações genéticas de EMC1-10 no painel de câncer do TCGA. São analisadas as alterações: amplificação, homodeleção e mutações com troca de sentido, sem sentido, e fora de fase. O download dos dados foi feito através do software cBioPortal e a quantidade de pacientes de cada tipo de câncer está especificada na seção 3.1 (Métodos).

Figura 7– Alteração de expressão dos genes EMC1-10, no painel de câncer do TCGA. A classificação em subexpresso e superexpresso considera o valor de z-score <-2 e >2, respectivamente. O download dos dados foi feito através do software cBioPortal e a quantidade de pacientes de cada tipo de câncer está especificada na seção 3.1 (Métodos).

Figura 8- Curvas de sobrevida comparando os pacientes com alta e baixa expressão gênica. O grupo “Alta” é composto por 25% dos pacientes com maior expressão gênica e o grupo “Baixa”, 25% dos pacientes com menor expressão gênica. Todos os tipos de câncer foram analisados com relação aos 10 genes do complexo EMC. Nessa figura são apresentados os casos em que $p\text{-value} < 0.05$. Os gráficos foram obtidos diretamente do site da OncoLnc.

Figura 9- Alterações genéticas de EMC2 nos pacientes com câncer de mama do TCGA. A) Cada retângulo representa um paciente, e estes foram dispostos lado a lado, totalizando 960 pacientes. A fita superior apresenta as alterações gênicas: amplificação, superexpressão de mRNA, homodeleção, subexpressão de mRNA e mutações troca de sentido e sem sentido; e a fita inferior mostra o *heatmap* da expressão gênica. As fitas foram ampliadas duas vezes em regiões convenientes para serem visualização com mais detalhe e nitidez. **B)** Localização das alterações na proteína EMC2 que foram consequentes das mutações gênicas encontradas em 5 dos 960 pacientes. Os diagramas foram obtidos diretamente do software cBioPortal.

Figura 10 – Amplificação da região cromossômica 8q em pacientes com câncer de mama do TCGA. O retângulo da esquerda mostra toda extensão do braço q do cromossomo 8 e o retângulo da direita focaliza na região cromossômica onde se localiza o gene EMC2. Em vermelho é marcada a região com amplificação e em azul deleção. A figura foi obtida diretamente do software UCSC Cancer Browser.

Figura 11 – A) Esquema do cromossomo 8 e as localizações dos genes EMC2 e MYC, mostrando a proximidade entre eles. **B)** Perfil de alterações de EMC2 e MYC nos pacientes com câncer de mama do TCGA. Os pacientes foram dispostos na mesma sequência na régua EMC2 e MYC. **C)** Curva de sobrevida de Kaplan-Meier comparando o grupo com superexpressão de EMC2 e com expressão normal de MYC, contra o restante dos pacientes. Os gráficos B e C foram obtidos diretamente do software cBioPortal.

Figura 12 – Expressão de EMC2 com relação aos biomarcadores moleculares para o câncer de mama: receptor de estrogênio, receptor de progesterone e HER2. Os valores são apresentados como diagrama de caixa Tukey e analisados estatisticamente pelo teste t de duas caudas, onde o valor de p é representado por * <0.05, **<0.01, ***<0.001 e ****<0.0001.

Figura 13 – **A)** Correlação entre o conjunto total de genes e EMC2, calculados através dos coeficientes de Pearson e Spearman. Em verde, genes selecionados com maior correlação com EMC2 (coeficiente de Pearson ≥ 0.3). **B)** Heat map representando a expressão gênica dos genes selecionados no grupo de pacientes com menor e maior expressão de EMC2. **C)** Predição de vias metabólicas utilizando os genes selecionados

Figura 14 – **Superexpressão de EMC2 na linhagem HEK293-T.** **A)** Células transfectadas com o plasmídeo controle (mCherryCTRL) e **B)** o plasmídeo mCherryEMC2. Em vermelho está a fluorescência da proteína mCherry e em azul a marcação do núcleo com DAPI. Nessa figura também aparecem os campos claros e as respectivas sobreposições dos canais.

Figura 15 – **Silenciamento de EMC2 e ensaio de proliferação.** **A)** Western blot e PCR em tempo real comprovando silenciamento de EMC2 na linhagem celular 4T1. **B)** Ensaio de proliferação utilizando cristal violeta que mostra diminuição na proliferação das células silenciadas para EMC2 comparadas ao controle. Os valores são apresentados como média $\pm$ desvio padrão e foram analisados estatisticamente pelo teste t de duas caudas, onde o valor de p é representado por * <0.05, **<0.01, ***<0.001 e ****<0.0001. Os testes feitos comparam cada grupo ao shCTRL, e em B cada tempo foi analisado separadamente.

Figura 16 – Células 4T1 silenciadas para EMC2 e controle, sincronizadas com nocodazol na progressão do ciclo celular. O ajuste dos picos foi feito automaticamente utilizando o software ModFit.

Figura 17– Fuso mitótico de células 4T1 knockdown de EMC2 (sh260) e controle (shCTRL). Em vermelho e verde estão as imunomarcações de EMC2, $\alpha\beta$ tubulina, respectivamente e em azul a marcação do DNA com DAPI. Na esquerda, estão as reconstruções 3D e na direita, as projeções máximas (barra corresponde a 10 μ m).

Figura 18 – Mudança na morfologia celular após o silenciamento de EMC2, na célula de câncer de mama MCF7.

Figura 19 – Imunomarcação de EMC2 (em verde) na linhagem celular MCF7. Em azul está a marcação do núcleo através do DAPI.

Figura 20 – Imunomarcação de EMC2 (em vermelho) na linhagem celular 4T1. Em verde está a marcação de β -tubulina e em azul, a marcação do núcleo com DAPI. Ainda está apresentado a imagem obtida no campo claro e a sobreposição dos canais.

Figura 21 – Controle da reação de imunohistoquímica. Tecido de rim cujas reações foram feitas na **A)** ausência e **C)** presença da proteína anti-EMC2, sendo portanto, os controles negativo e positivo, respectivamente. **B)** e **D)** são ampliações das figuras **A)** e **C)**, nessa ordem.

Figura 22 – Parâmetro interno estabelecido para classificação das amostras de câncer de mama e mama normal adjacente ao câncer dos pacientes do HC-Ribeirão Preto de acordo com intensidade de marcação da proteína EMC2 por imunohistoquímica. As amostras são de pacientes diagnosticados com carcinoma invasivo mamário.

Figura 23 – Quantificação das amostras analisadas classificadas quanto a intensidade de marcação de EMC2 (negativo, fraco, médio e forte).

Figura complementar 1 – Western blot da linhagem celular MCF7 mostrando a alta especificidade do anticorpo anti-EMC2 marcando uma única banda (**D**) na altura

aproximada de 35kDa. Em **A** e **C** encontra-se o marcador de peso molecular e em **B** a raia do lisado proteico corado com Ponceau (solução descrita em Materiais e Métodos).

Figura Suplementar 2- Colocalização da marcação de EMC2 através do anticorpo anti-EMC2 (verde) e da proteína EMC2 fusionada a mCherry (vermelha) mostrando que o anticorpo reconhece a proteína induzida pelo plasmídeo transfectado mCherry-EMC2.

LISTA DE TABELAS

Tabela 1 – Características dos subtipos moleculares de acordo com padrões de imunomarcacão. Retirado de Cinqueira, et al. 2011.

Tabela 2 – Sequências de primers utilizados.

Tabela suplementar 1 – Regressão de Cox utilizando dados de sobrevida dos pacientes do banco de dados TCGA para as integrantes do complexo EMC calculadas através do software OncoLnc.

Tabela suplementar 2 – Genes com alta correlacão de Pearson com EMC2 (Corr. de Pearson ≥ 3.0) selecionados para a análise de predicão de vias. Dados obtidos através do cBioPortal.

LISTA DE SIGLAS

ACC	<i>Adrenocortical Carcinoma</i> (Carcinoma Adrenocortical)
AML	<i>Acute Myeloid Leukemia</i> (Leucemia Mieloide Aguda)
BLCA	<i>Bladder Urothelial Carcinoma</i> (Carcinoma Urotelial de Bexiga)
BRCA	<i>Breast Invasive Carcinoma</i> (Carcinoma de Mama Invasivo)
CESC	<i>Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma</i> (Carcinoma Cervical de Células Escamosas)
CHOL	<i>Cholangiocarcinoma</i> (Colangiocarcinoma)
COAD	<i>Colon adenocarcinoma</i> (Adenocarcinoma de Cólon)
DLBC	<i>Lymphoid Neoplasm Diffuse Large B-cell Lymphoma</i> (Linfoma Neoplásico Linfoide Difuso de Células B Grandes)
EMC	<i>Endoplasmic Reticulum Membrane Protein Complex</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático)
EMC 1	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 1</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 1)
EMC 2	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 2</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 2)
EMC 3	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 3</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 3)
EMC 4	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 4</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 4)
EMC 5	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 5</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 5)
EMC 6	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 6</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 6)

EMC 7	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 7</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 7)
EMC 8	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 8</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 8)
EMC 9	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 9</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 9)
EMC 10	<i>Endoplasmic Reticulum Membrane Protein Complex Subunit 10</i> (Complexo de Proteínas da Membrana do Retículo Endoplasmático Subunidade 10)
ERAD	Endoplasmic Reticulum Associated Protein Degradation (Degradação de Proteínas Associado ao Retículo Endoplasmático)
ESCA	<i>Esophageal Carcinoma</i> (Carcinoma Esofágico)
GBM	<i>Glioblastoma Multiforme</i> (Glioblastoma Multiforme)
HC	Hospital das clínicas
HER2	<i>Human Epidermal Growth Factor Receptor 2</i> (Receptor do Fator de Crescimento Epidermal Humano 2)
HNSC	<i>Head and Neck Squamous Cell Carcinoma</i> (Carcinoma de Células Escamosas de Cabeça e Pescoço)
KICH	<i>Kidney Chromophobe</i> (Cromofobia de Rim)
KIRC	<i>Kidney Renal Clear Cell Carcinoma</i> (Carcinoma Renal de Células Claras de Rim)
KIRP	<i>Kidney Renal Papillary Cell Carcinoma</i> (Carcinoma Renal de Células Papilares de Rim)
LGG	<i>Brain Lower Grade Glioma</i> (Glioma de Baixo Grau de Cérebro)
LIHC	<i>Liver Hepatocellular Carcinoma</i> (Carcinoma Hepatocelular de Fígado)
LUAD	<i>Lung Adenocarcinoma</i> (Adenocarcinoma de Pulmão)
LUSC	<i>Lung Squamous Cell Carcinoma</i> (Carcinoma de Células Escamosas de Pulmão)
MESO	<i>Mesothelioma</i> (Mesotelioma)

OV	<i>Ovarian Serous Cystadenocarcinoma</i> (Cistadenocarcinoma Seroso de Ovário)
PAAD	<i>Pancreatic Adenocarcinoma</i> (Adenocarcinoma Pancreático)
PCPG	<i>Pheochromocytoma and Paranglioma</i> (Feocromocitoma e Paranglioma)
PRAD	<i>Prostate Adenocarcinoma</i> (Adenocarcinoma de Prostata)
RE	Retículo Endoplasmático
SARC	<i>Sarcoma</i> (Sarcoma)
SKCM	<i>Skin Cutaneous Melanoma</i> (Melanoma Cutâneo de Pele)
STAD	<i>Stomach Adenocarcinoma</i> (Adenocarcinoma de Estômago)
TGCT	<i>Testicular Germ Cell Tumors</i> (Tumor de Células Germinativa Testicular)
THCA	<i>Thyroid Carcinoma</i> (Carcinoma de Tireóide)
THYM	<i>Thymoma</i> (Timoma)
UCS	<i>Uterine Carcinosarcoma</i> (Carcinosarcoma Uterino)
UCEC	<i>Uterine Corpus Endometrial Carcinoma</i> (Carcinoma Endometrial de Corpos Uterinos)
UPR	<i>Unfolded Protein Response</i> (Resposta a Proteínas Mal Enoveladas)
UVM	<i>Uveal Melanoma</i> (Melanoma Uveal)

SUMÁRIO

RESUMO	v
ABSTRACT	vi
1.INTRODUÇÃO	1
1.1 Aspectos moleculares do câncer de mama.....	2
1.2 Complexo EMC.....	5
1.2.1 Envolvimento de membros do complexo EMC em câncer.....	7
1.3 Proteína EMC2.....	7
2. OBJETIVOS	8
3. MÉTODOS	10
3.1 Bioinformática – Análise dos dados do TCGA	11
3.1.1 Estudo das proteínas do complexo EMC	11
3.1.1.1 Download da expressão do tecido normal	11
3.1.1.2 Download da expressão do tecido tumoral	11
3.1.1.3 Análise das alterações genéticas	12
3.1.1.4 Análises de superexpressão e subexpressão	12
3.1.1.5 Análise de sobrevida	12
3.1.1.6 Análise estatística dos dados do TCGA.....	13
3.1.2 Estudo da proteína EMC2 em câncer de mama.....	13
3.1.2.1 Genes com alta correção de Pearson.....	13
3.1.2.2 Heat map.....	13
3.1.2.3 Predição de vias.....	13
3.2 Análises <i>in vitro</i>	14
3.2.1 Clonagem da proteína EMC2	14
3.2.2 Cultivo da linhagem celular	14
3.2.3 Midiprecipitação plasmidial	15
3.2.4 Estabelecimento da linhagem knockdown para EMC2	15

3.2.5 Transfecção por lipofectamine	15
3.2.6 Dosagem proteica	15
3.2.7 SDS-PAGE	16
3.2.8 Western blot	16
3.2.9 Extração de RNA	17
3.2.10 Síntese de cDNA	17
3.2.11 PCR em Tempo Real	17
3.2.12 Ensaio de proliferação celular	18
3.2.13 Ensaio de ciclo celular	18
3.2.14 Imunocitoquímica	19
3.2.15 Anticorpos e marcadores fluorescentes utilizados	19
3.3 Imunohistoquímica	20
4. RESULTADOS	21
4.1 Análise do complexo EMC usando o banco de dados TCGA.....	22
4.2 Análise de EMC2 em câncer de mama.....	27
4.3 Superexpressão e silenciamento de EMC2.....	32
4.4 Análise do silenciamento de EMC2 no ciclo celular	35
4.5 Análise por imunocitoquímica do fuso mitótico	38
4.6 Mudança na morfologia em células silenciadas para EMC2	39
4.7 Localização de EMC2 nas células de câncer de mama	40
4.8 Imunohistoquímica de amostras de pacientes com câncer de mama	43
5. DISCUSSÃO	46
6. CONCLUSÕES	51
7. BIBLIOGRAFIA.....	53
8. ANEXO.....	59

INTRODUÇÃO

1. Introdução

Câncer é uma doença proveniente de células mutantes, cujas alterações causam descontrole em características celulares como proliferação, migração, morte celular, e assim compromete-se a harmonia tecidual (Hanahan & Weinberg, 2001).

Devido ao aumento na expectativa de vida e ao crescente acesso ao diagnóstico do câncer, aumenta-se anualmente a incidência dessa doença. Na mulher, o câncer de mama é o tipo com maior ocorrência, depois do de pele não melanoma, e contabiliza 28% dos novos casos de câncer por ano. No homem a incidência é mais rara, com apenas 1% dos casos da doença (INCA, 2018).

Alguns aspectos são tidos como fatores de risco para o desenvolvimento do câncer de mama. Dentre eles está a idade, fatores relacionados a comportamento ou ambiente, como exposição à radiação ionizante, fatores endócrinos ou relacionados a vida reprodutiva, como menarca precoce, menopausa tardia e nuliparidade, além de fatores genéticos como mutações nos genes BRCA1 e BRCA2. A elucidação dos mecanismos responsáveis pelo desenvolvimento do câncer de mama é de difícil compressão, o que é agravado pela alta complexidade molecular e heterogeneidade celular desse tipo tumoral (Turashvili et al., 2017).

1.1 Aspectos moleculares do câncer de mama

O epitélio glandular mamário é composto por lóbulos e ductos, os quais apresentam células mioepiteliais e luminais (Figura 1). Quando ocorre a hiperplasia, células crescem desarmoniosamente ocupando o espaço luminal. Como proposto por Polyak e colaboradores (2007), no estágio *in situ*, estágio mais avançado que a hiperplasia, as células mioepiteliais são alteradas fenotipicamente e epigeneticamente e ocorre o aumento de fibroblastos no estroma assim como miofibroblastos, linfócitos e células endoteliais. Quando há a perda células mioepiteliais juntamente com o rompimento da membrana basal caracteriza-se a fase de carcinoma invasivo, cujo nome é dado devido a capacidade das células tumorais conseguirem invadir o tecido adjacente. Se as células atingirem a corrente sanguínea ou endócrina e alcançarem outros tecidos distantes, é denominado metástase (Figura 2).

Figura 1 – Estrutura da glândula mamária. (Adaptado de Dimri, et al. 2005)

Figura 2 – Modelo de progressão tumoral do câncer de mama. (Adaptado de Polyak et al., 2007)

Em vista da progressão tumoral, o tumor mamário pode ser classificado como lesão precursora, lesão de baixo potencial maligno, proliferação do epitélio benígno, neoplasias fibroepiteliais, mioepiteliais ou mesenquimais, carcinoma in situ e invasivo (Sinn et al., 2013). Há ainda a classificação de acordo com aspectos moleculares. Nesse contexto, o câncer de mama é dividido em seis principais subtipos: luminal A, luminal B, superexpressão de HER2, basalóide, *normal-like*

(Rorou et al., 2000) e o mais recentemente nomeado *claudin-low* (Herschkowitz et al. 2007).

Cada subtipo molecular apresenta aspectos específicos, que além das características histológicas, também são considerados aspectos moleculares como níveis proteicos de receptores HER2, ER, PR e marcação de células proliferativas Ki-67, na maioria das vezes avaliados por imunohistoquímica. Além desses, outros marcadores podem ser utilizados, como é o caso do EGFR e citoceratinas (Cinqueira et al., 2011). A tabela 1 mostra algumas características dos subtipos moleculares.

Subtipo molecular	Classificação com o índice de Ki-67 de 14%
	Padrão de Imunomarcação
Luminal A	RE+ e/ou RP+, HER2- e Ki-67 <14%
Luminal B	RE+ e/ou RP+, HER2- e Ki-67 ≥14% RE+ e/ou RP+, HER2+ (luminal HER2)
Superexpressão de HER2	RE-, RP- e HER2+
Basaloide	RE-, RP-, HER2-, CK5+ e/ou EGFR+
Triplo-negativo não basaloide	RE-, RP-, HER2-, CK5- e EGFR-

HER2: receptor tipo 2 do fator de crescimento epidérmico humano; EGFR: receptor tipo 1 do fator de crescimento epidérmico; RE: receptor de estrogênio; RP: receptor de progesterona; CK5: citoceratina 5.

Tabela 1 – Características dos subtipos moleculares de acordo com padrões de imunomarcação. Retirado de Cinqueira (2011).

O subtipo *claudin-low*, por sua vez, apresenta baixa expressão de genes envolvidos em junções ocludentes e glicoproteínas de adesão célula-célula, incluindo as claudinas 3, 4 e 7, as ocludinas e a E-caderina (Herschkowitz et al., 2007). Além disso, esse subtipo molecular apresenta aumento na expressão de proteínas relacionadas a transição epitélio-mesenquima e marcadores associados a células tronco tumorais (Prat et al., 2010).

Um dos primeiros genes cuja perda ou mutação foi associada com câncer de mama foram BRCA1 e BRCA2. Hoje, se sabe que esses genes também podem se encontrar inativos por alterações epigenéticas. Eles são considerados genes supressores de tumor e atuam em reparo do DNA (Claus et al., 1991; Romagnolo et al., 2015).

Outros genes que aparecem frequentemente mutados ou com expressões aberrantes em câncer de mama são PTEN, TP53, PIK3CA e RB. Esses genes, assim como outros já mencionados anteriormente, podem desregular vias como a de

EGFR/PI3K/PTEN/Akt/mTORC1 e com isso causar transformações malignas, prevenção de apoptose, além de metástase (Davis et al., 2014).

Apesar dos estudos desenvolvidos até o momento, e dos interessantes achados moleculares, em especial, no câncer de mama, ainda há muito o que ser elucidado dentro da biologia celular e biologia do câncer. O advento de novas ferramentas em biologia celular traz, para o atual momento da ciência, um cenário propício para esse avanço, fazendo com que o entendimento da doença seja um desafio contemporâneo.

Em nosso laboratório, através de um estudo de expressão diferencial usando fragmentos ORESTES (*Open Reading Frames ESTs*) gerados pelo projeto Genoma Humano do Câncer, foi identificada uma proteína (KIAA0090) diferencialmente expressa em linhagens de melanoma metastático (Sousa et al., 2002). Essa proteína foi posteriormente descrita como participante de um complexo, chamado EMC, e por esse motivo foi renomeada como EMC1 (Jokinas et al., 2009). Estudos *in vitro* de nosso laboratório, mostraram que a proteína apresenta relação com características tumorais como proliferação, migração e formação de clones, assim como crescimento tumoral em modelos *in vivo* (Silva, 2014; Molina, 2014; Cincos, 2017). Devido aos resultados interessantes encontrados no estudo de EMC1 em câncer, o nosso grupo procurou ampliar a pesquisa para as outras proteínas do complexo. O estudo atual busca possíveis relações das proteínas do complexo EMC e dados de diversos tipos de câncer, e posteriormente, busca aprofundar a investigação de possíveis relações entre EMC2 e câncer de mama.

1.2 Complexo EMC

O retículo endoplasmático (RE) é uma organela essencial, responsável por funções como enovelamento e secreção de proteínas, biossíntese de lipídeos e colesterol e homeostase de cálcio. Além disso, o RE assegura a qualidade do enovelamento proteico num processo chamado degradação de proteínas associado ao retículo endoplasmático (ERAD). Em situações de estresse, pode ocorrer o

acúmulo de proteínas mal enoveladas no RE, e com isso, ativar uma outra via chamada resposta a proteínas mal enoveladas (UPR) (Tsai & Weissman, 2010).

Jonikas e colaboradores (2009) em um estudo em leveduras, em 2009, fez uma busca por novas proteínas relacionadas a via de UPR. Através de silenciamentos duplos e utilizando um sensor endógeno Ire1 para verificar o estado de enovelamento de proteínas no RE, o estudo mostrou que um grupo de proteínas, até então pouco caracterizadas, pareciam fazer parte da via de UPR, assim como se organizarem em um complexo. Esse grupo, composto por 6 proteínas, foi denominado *Endoplasmatic Reticulum Membrane Complex* (EMC). Em 2011, Bircham e colaboradores corroboram os achados anteriores, mostrando que duplos mutantes para os membros do complexo EMC mostram-se hipersensíveis a droga tunicamicina, indutora de estresse de retículo.

Em 2011, Christianson e colaboradores identificaram o complexo em células de mamífero, composta por 10 membros (EMC1-10). O estudo também verificou que o complexo interage com as proteínas UBAC2 e DERLIN-2 pertencentes a via de ERAD, e dessa forma mostrando relação entre o complexo e a via de translocação e degradação de proteínas.

Lihiri e colaboradores (2014) mostraram que o complexo EMC é importante na transferência de PS do RE para a mitocôndria, reduz o “tethering” ER-mitocôndria e provoca danos nas funções mitocondriais.

Recentemente, foi visto que a infecções de vírus como WNV, zika, dengue e YFV é fortemente dependente do complexo EMC para sua replicação e nos estágios iniciais da infecção (Ma et al., 2016; Savitis et al., 2016).

Figura 3 – Modelo esquemático das proteínas do complexo EMC de mamíferos de acordo com as estruturas topológicas preditas. (Molina et al., 2014)

1.2.1 Envolvimento de membros do complexo EMC em câncer

Já foram relatados na literatura alguns estudos que mostram proteínas do complexo EMC envolvidas com o câncer. Em 2011, Valsesia e colaboradores apontaram EMC1 como um gene com expressão aumentada em linhagens de melanoma metastático humano comparado a amostras de melanócito. Em células de glioma, duas proteínas do complexo apareceram como supressoras na proliferação celular: EMC10 (Junes-Gill, et al. 2010) e EMC6 modulada por autofagia (Shen, et al. 2016).

1.3 Proteína EMC2

Inicialmente nomeada como KIAA0103, EMC2 apresenta domínios *tetratricopeptide repeats*, o que lhe rendeu outro nome: TTC35 (Nagase et al., 1995). Esse domínio é encontrado em proteínas como membros do complexo APC, cdc16, OGT, entre outras. É um domínio conhecido por servir de estrutura para interação proteína-proteína e ajuda na formação de complexos multiproteicos (Krachler, 2010).

Localizada na região cromossômica 8q23.1, o gene EMC2 contém 43.318 bases orientados na fita positiva. São produzidos 5 variantes de transcritos, dos quais 4 correspondentes a mRNA e 1 RNA não codificante (Figura 4). A proteína EMC2 contém 297 aminoácidos, com massa molecular equivalente a 34,8kDa (<https://www.ncbi.nlm.nih.gov/>, acessado 17 de Outubro de 2017).

A proteína EMC2 é altamente distribuída na célula, sendo encontrada no RE (Hoja et al. 2000; Jonikas et al. 2011), envelope nuclear (Greger et al. 2001), citosol, mitocôndria e peroxissomos (GeneCards).

Figura 4 – Isoformas de EMC2. Adaptado de http://atlasgeneticsoncology.org/Genes/GC_EMC2

OBJETIVOS

2. Objetivos

2.1 Objetivo geral

O objetivo geral deste trabalho consiste em analisar possíveis relações entre as proteínas do complexo EMC e o desenvolvimento de cânceres, em especial, a proteína EMC2 em câncer de mama.

2.2 Objetivos específicos

- Análise em bancos de dados da expressão gênica, alteração genética e sobrevida dos pacientes, em um painel de cânceres;
- Geração de linhagens de câncer de mama com superexpressão e silenciamento de RNA de EMC2;
- Análises funcionais dos efeitos da superexpressão na linhagem HEK293T e silenciamento de EMC2 nas linhagens MCF7 e 4T1;
- Análises da marcação de EMC2 por imunohistoquímica de amostras do tecido tumoral de pacientes com câncer de mama, em diferentes estágios da doença.

MÉTODOS

3. Métodos

3.1 Bioinformática – Análise dos dados do TCGA

Todos os dados analisados nessa seção (3.1) pertencem ao banco de dados The Cancer Genome Atlas (TCGA), grupo este que reúne dados de pacientes diagnosticados com câncer de centros cadastrados dos EUA.

3.1.1 Estudo das proteínas do complexo EMC

3.1.1.1 Download da expressão do tecido normal

Os dados de expressão normalizada do tecido normal foram obtidos através do FireBrowse (<http://firebrowse.org/>), acessado em 10 de Julho de 2016. BLCA (n=19), BRCA (n=112), CESC (n=3), CHOL (n=9), ESCA (n=11), GBM (n=5), HNSC (n=44), KICH (n=25), KIRC (n=72), KIRP (n=32), LIHC (n=50), LUAD (n=59), LUSC (n=51), PAAD (n=4), PCPG (n=3), PRAD (n=52), SARC (n=2), SKCM (n=1), STAD (n=35), THCA (n=59), THYM (n=2) e UCEC (n=35).

3.1.1.2 Download da expressão do tecido tumoral

Os dados de expressão normalizada de tumores foram adquiridos através do cBioPortal (Gao et al. 2013 & Cerami et al. 2012). AML (n=173), ACC (n=79), BLCA (n=408), BRCA (n=1100), CESC (n=306), CHOL (n=36), COAD (n=382), DLBC (n=48), ESCA (n=185), GBM (n=166), HNSC (n=522), KICH (n=66), KIRC (n=534), KIRP (n=291), LGG (n=530), LIHC (n=373), LUAD (n=517), LUSC (n=501), MESO (n=87), OV (n=307), PAAD (n=179), PCPG (n=184), PRAD (n=498), SARC (n=263), SKCM (n=472), STAD (n=415), TGCT (n=156), THCA (n=509), THYM (n=120), UCS (n=57), UCEC (n=177) e UVM (n=80).

3.1.1.3 Análise das alterações genéticas

As alterações no número de cópias putativo (GISTIC) e os dados de mutação somática foram obtidos através do cBioPortal (Gao et al. 2013 & Cerami et al. 2012), acessado em 10 de Julho de 2016. Apenas deleções homozigóticas e amplificação de alto nível foram consideradas. O número de amostras com dados disponíveis das alterações no número de cópias (CNA) e mutações são dispostos como CNA/Mutações: AML (n=188/188), ACC (n=90/90), BLCA (n=408/130), LGG (n=513/286), BRCA (n=1080/977), CESC (n=295/194), CHOL (n=36/35), COAD (n=616/223), DLBC (n=48/48), ESCA (n=184/185), GBM (n=577/290), HNSC (n=522/510), KICH (n=66/66), KIRC (n=528/451), KIRP (n=288/282), LIHC (n=370/373), LUAD (n=516/230), LUSC (n=501/177), MESO (n=87/Na), OV (n=579/316), PAAD (n=184/150), PCPG (n=162/179), PRAD (n=492/498), SARC (n=257/247), SKCM (n=367/366), STAD (n=441/395), TGCT (n=150/149), THCA (n=499/402), THYM (n=123/123), UCS (n=56/57), UCEC (n=539/248) and UVM (n=80/80).

3.1.1.4 Análises de superexpressão e subexpressão

A expressão relativa z-score é calculada pela equação $z - score = (x - \mu) / \sigma$, onde x equivale a expressão do gene do paciente em questão, μ e σ representam, respectivamente, a média e o desvio padrão do gene de um paciente dentro do câncer que ele foi diagnosticado. Valores de z-score ≥ 2.0 são considerados superexpressos e z-score ≤ -2.0 subexpressos, de acordo com o cBioPortal.

3.1.1.5 Análise de sobrevida

Foram feitos os downloads dos dados de sobrevida pre-computados pelo banco de dados OncoLnc, onde a regressão de Cox foi realizada considerando expressão gênica, sexo, idade, grau e histologia como multivariáveis. Os cânceres com ausência de classificação por graus ou que a maioria dos pacientes eram do mesmo estágio, foram excluídos esses parâmetros como multivariáveis (e.g. COAD,

GBM, KIRP, LAML, LUAD, LUSC, SARC e SKCM). As análises de sobrevida Kaplan-Meier foram usadas para plotar as sobrevidas dos paciente com Alta e Baixa expressão do gene definido divididos pelo menor e maior quartil de expressão desse gene.

3.1.1.6 Análise estatística dos dados do TCGA

Os dados de expressão (Figura 5) foram testados para normalidade (distribuição Gaussiana) com o teste de normalidade D'Agostino-Pearson omnibus, que determinou o uso dos testes paramétricos (teste T não pareado) e não paramétricos (teste Mann Whitney) para a comparação de colunas individuais.

3.1.2 Estudo da proteína EMC2 em câncer de mama

3.1.2.1 Genes com alta correção de Pearson

Os dados da correção de Pearson de EMC2 foram obtidos através do cBioPortal. Selecionamos para análises posteriores de heat map e predição de vias, genes com correlação de Pearson ≥ 0.3 .

3.1.2.2 Heat map

Para a construção do gráfico, foram utilizadas os dados de expressões gênicas dos genes com correlação de Pearson ≥ 0.3 , de 10% dos pacientes com maior e menor expressão de EMC2. Utilizamos o software Morpheus (<https://software.broadinstitute.org/morpheus/>).

3.1.2.3 Predição de vias

A predição de vias foi realizada utilizando o software DAVID (<https://david.ncifcrf.gov/summary.jsp>).

3.2 Análises *in vitro*

3.2.1 Clonagem do cDNA EMC2

Para amplificar o gene EMC2, utilizamos como template cDNAs da linhagem celular RPE e o kit de PCR Go Taq G2 Flex DNA. O primer foi desenhado utilizando a sequência de cDNA de EMC2 depositada no NCBI (<https://www.ncbi.nlm.nih.gov/>) (Tabela 3) que corresponde ao transcrito NM_014673.4. A amplificação gênica por PCR foi feita utilizando um *pull* de células da linhagem RPE. O produto de PCR foi submetido à eletrofose no gel de agarose 1% e a banda próxima ao tamanho esperado foi purificada utilizando o kit Wizard SV Gel and PCR Clean-Up System. O fragmento foi, então, subclonado no vetor PGEM-T Easy, da Promega utilizando o kit do fabricante e posteriormente foi feita a transformação das bactérias cálcio competentes. Após a midiprecipitação plasmidial das bactérias transformadas, fizemos a reação de sequenciamento utilizando o kit BigDye Terminator v3.1 Cycle Sequencing. O sequenciamento encontra-se em anexo. Após a subclonagem, o plasmídeo foi digerido com as enzimas HindIII e BamHI (Fermentas) seguindo o protocolo do fabricante e em seguida foi feita a eletrofose em gel de acrilamida. Purificamos novamente a banda na região do número de bases esperada e ligamos o fragmento ao vetor comercial MCherry-Parkin (Plasmid #23956) que havíamos, previamente, retirado a proteína Parkina. O novo plasmídeo, nomeado por nosso grupo como MCherry-EMC2, foi novamente transformado nas bactérias cálcio competentes, e os plasmídeos foram extraídos utilizando o kit de midiprecipitação plasmidial. Foi contruído um vetor de MCherry sem a ligação de EMC2 para ser utilizado como controle. Para isso, foi feita a digestão dupla do plasmídeo MCherry-Parkin utilizando as enzimas BglII e BclI, e em seguida, foi feita a ligação das extremidades do fragmento de interesse. Chamamos o vetor de MCherry-controle. Os sequenciamentos foram feitos e estão disponíveis no tópico Anexo.

3.2.2 Cultivo da linhagem celular

As células de câncer de mama murino 4T1 e as células embriogênicas de rim HEK293T foram cultivadas em meio DMEM (alta glicose, 110mg/L de piruvato de sódio) suplementado com 10% SFB em atmosfera úmida contendo 5% de CO₂ a 37°C.

3.2.3 Midiprecipitação plasmidial

Para extração e purificação dos plasmídeos das bactérias DH5α, utilizamos o kit PureYield™ Plasmid Midiprep System e seguimos as recomendações sugeridas pelo fabricante.

3.2.4 Estabelecimento da linhagem knockdown para EMC2

Utilizamos três plasmídeos shRNA-pLKO.1 com as sequências sh260, sh885 e sh1007, específicas para o mRNA de EMC2 camundongo, que foram adquiridos pela empresa Sigma-Aldrich (MISSION®). Como controle, utilizamos um shRNA sem alvo no genoma de mamíferos (shCTRL). Para produção de partículas lentivirais, células HEK293T foram transfectadas utilizando reagente PEI, com os plasmídeos shRNA-pLKO.1, Δ8.9 contendo o gene gag, pol e rev, e VSV-G que expressa o envelope viral.

3.2.5 Transfecção por lipofectamina

As células foram cultivadas em placas de 6 poços (área = 9.6 cm²) até chegar a confluência de 80%. Em um microtubo de 1.5mL, foi adicionado 250μL de DMEM (sem adição de soro) e 2.5μg de DNA plasmidial. Em outro microtubo, foi adicionado 250μL de DMEM (sem adição de soro) e 5μL de Lipofectamine™ 1000 (Thermo Fisher Scientific). Após 5 minutos de incubação à temperatura ambiente, juntou-se as duas soluções, homogeneizamos cuidadosamente e deixamos incubando por 20 minutos. Então, o meio das células foi trocado por DMEM sem soro (1.5mL) e foi adicionado 500μL da solução prepada em cada poço. Após 3 horas de incubação à

37°C em atmosfera úmida contendo 5% de CO₂, o meio foi trocado por meio de cultura de costume.

3.2.6 Dosagem proteica

Para dosagem de concentração proteica, utilizamos o kit Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) e seguimos as recomendações do fabricante.

3.2.7 SDS-PAGE

As células foram lavadas duas vezes com PBS e foi adicionado 100µL de tampão Ripa. O perfil eletroforético protéico foi analisado por SDS-PAGE em mini-géis contituídos de um gel de empilhamento (5%) (Tris 0.125mM pH6.8; acrilamida/Bis 4.8%; SDS 0.1%; persulfato de ammonia 0.06%; azul de bromofenol 1%; TEMED 0.1% e água MiliQ) e um gel de separação de poliacrilamida 12% (Tris 0.37M pH8.8; acrilamida/Bis 12%; SDS 0.1%; persulfato de ammonia 0.06%; TEMED 0.1% e água MiliQ). Para a corrida da eletroforese foi utilizado o tampão de eletrodo (Tris 25mM; glicina 187 mM e SDS 0.1%), com amperagem inicial de 15mA, passando-se para 20mM após a entrada das amostras no gel de separação.

3.2.8 Western blot

As proteínas separadas por SDS-PAGE foram tranferidas por eletroforese por 16 horas com uma diferença de potencial de 37 volts a 4°C para uma membrana de nitrocellulose (BioRad), utilizando-se o tampão de transferência (Tris 25mM; glicina 0.2mM e methanol 20%). Após a transferência, a membrana foi corada com Ponceau 0.5% em ácido tricloroacético (TCA) 3% por 15 minutos e lavada com água para remover o excesso de corante. Antes da imunomarcção, os sítios inespecíficos foram neutralizados com solução de bloqueio (5% de leite emp ó desnatado em TBS/T (Tris 20mM, NaCl 137mM pH7.5 e Tween-20 0.1% (V/V))) sob agitação continua por 1 hora à temperature ambiente. Em seguida, o anticorpo primário foi adicionado a membrana, diluído em 5% de BSA em TBS-T, que ficou incubada *overnight* a 4°C. A membrana foi Lavada 4 vezes por 5 minutos em TCS/T, sob agitação continua à temperature ambiente para retirada do anticorpo primário

não ligado. Foi acrescentado, então, o anticorpo secundário polyclonal anti-IgG de coelho ou camundongo, conjugado a peroxidase (Promega) diluído 1:10000 em TBS/T, e a membrana foi incubada por 1 hora à temperatura ambiente, sob agitação. Para a retirada do anticorpo secundário não ligado, a membrana foi lavada 4 vezes de 5 minutos em TBS/T, sob agitação contínua à temperatura ambiente. A revelação do anticorpo ocorreu em câmara escura. A membrana foi banhada com soluções caseiras ECL (Enhanced chemiluminescence) solução I (Tris/HCl 0.011M pH8.5; Luminol (5-Amino-2,3-dihydro-1,4-phythalazinedione free acid-Sigma-Aldrich) 2.78mM; 0.44mM ácido p-cumárico (Sigma-Aldrich); água MiliQ) e solução II (0.02% H₂O₂; 0.1M Tris/HCl pH8.5; água MiliQ) na proporção 1:1 por 1 minuto. A revelação foi realizada em Sistema automatizado ImageQuant LAS 4000 (GE).

3.2.9 Extração de RNA

A extração de RNA total foi realizada utilizando o reagent Trizol® (Invitrogen, Life Technologies). Utilizando placas de cultivo com poços de 10cm² de diâmetro, retiramos o meio de cultura, lavado com PBS e adicionamos 1mL de Trizol® por poço. Homogeneizamos a solução para que as células desgrudassem totalmente do fundo do poço, e transferimos para um tubo de 1,5mL, ao qual foi adicionado posteriormente 200µL de clorofórmio e agitado vigorosamente por 15 segundos. Após incubação à temperatura ambiente por 5 minutos, centrifugamos por 15 minutos a 4°C, 13200rpm. O sobrenadante incolor foi transferido para outro tubo de 1,5mL, onde foi acrescentado 500µL de isopropanol. Inverteu-se os tubos para misturar a solução e foi incubado por 5 minutos a temperatura ambiente. Em seguida, foi feita a centrifugação por 10 minutos a 4°C, 13200rpm. Cuidadosamente, removemos o sobrenadante, adicionamos 1mL de etanol 75% gelado e foi feita a centrifugação por 5 minutos, a 4°C, 13200rpm. Removemos o sobrenadante e deixamos o *pellet* secando por aproximadamente 10 minutos. A seguir, ressuspendemos o *pellet* em 20µL de água (*Nuclease-Free water* - Promega) e quantificamos utilizando o espectrofotômetro NanoVue Plus (GE HealthCare).

3.2.10 Síntese de cDNA

Para síntese de cDNA, utilizamos o kit High-Capacity cDNA Reverse Transcription (Applied Biosystems) e seguimos as recomendações do fabricante.

3.2.11 PCR em Tempo Real

Os experimentos para análise de expressão gênica foram realizados no equipamento Applied Biosystems 7500 Real-Time PCR System, utilizando o kit GoTaq® qPCR Master Mix (Promega). Os primers utilizados com suas respectivas sequências estão listados na tabela 2. Utilizamos como controle endógeno o gene TBP (*TATA Box Binding Protein*), um fator de transcrição geral que age no centro de um fator multiproteico de ligação ao DNA. Os dados obtidos foram analisados pelos valores de Ct (*Cycle Threshold*), que foi convertido em expressão relativa de acordo com o método $2^{-\Delta\Delta Ct}$ (Livak et al., 2001). Utilizamos os valores de Ct dos transcritos do gene TBP para a normalização.

Primer

Gene	Sequência	Propósito
EMC2 (humano)	F – AAGCTTATGCGAAGGTCTCAGAGCTT R – GGATCCGTAAATCATCATGGTTTATTT	Clonagem
EMC2 (humano)	F – TCGTGGAACAGTACGAACGC R – GCAAGTAAACCTCTACAAATGTGG	Sequenciamento do plasmídeo mCherry
EMC2 (murino)	F – ATCGCTTGACTCTGTGACTGC R – GAGGATGCTGTCACTTTCTGG	PCR em tempo real
TBP (murino)	F – GAGCTGTGATGTGAAGTTTCC R – TCTGGGTTTGATCATTCTGTAG	PCR em tempo real

Tabela 2 – Sequências de primers utilizados.

3.2.12 Ensaio de proliferação celular

As células 4T1 silenciadas para EMC2 e shCTRL foram plaqueadas em placa de 96 poços na densidade 5000 células/poço. Em diferentes tempos de incubação (0, 24 e 48 horas) as células foram fixadas com etanol 70% e coradas com cristal violeta. Após a lavagem dos poços com água miliQ por 4 vezes (ou até a água sair transparente – sinal que todo cristal violeta em excesso foi removido), as células foram ressuspensas em 100µL de ácido acético 10%. Deixamos incubando por 30

minutos à temperatura ambiente e fizemos a leitura da absorbância dos poços em 540nm.

3.2.13 Ensaio de ciclo celular

200.000 células foram plaqueadas por poço em placa de 6 poços. Após 24 horas, foi adicionado Nocodazol a concentração final de 200nM. A droga ficou agindo por 16 horas, e após duas lavagens com PBS, as células foram fixadas com etanol 70% por pelo menos 2 horas a 4°C. Em seguida, as células foram marcadas com Iodeto de Propídeo, e passaram pela citometria de fluxo BD FACSCanto.

3.2.14 Imunocitoquímica

As células foram cultivadas em placas de 100cm² em meio de cultivo apropriado, conforme descrito previamente, até atingirem a confluência. Posteriormente, as células foram transferidas para placa de 24 poços sobre lamínulas (13mm de diâmetro) e cultivadas em atmosfera úmida com 5% de CO₂ à 37°C. Após 48h as células foram lavadas com PBS (Na₂HPO₄ 4.3mM; KH₂PO₄ 1.4mM pH=7.3; NaCl 137mM; KCl 2.7mM), fixadas com paraformaldeído 2% (diluído em PBS) por 20 minutos à 37°C. Após levagem das lamínulas por 2 vezes com PBS, utilizamos Triton-X100 0.003% para permeabilizar a membrana das células, por 10 minutos. Novamente as células foram lavadas com PBS por 3 vezes e em seguida, fez-se o bloqueio dos aldeídos livres com Glicina 100mM, durante 5 minutos. Lavou-se com PBS e utilizamos BSA/PBS 2% para o bloqueio de sítios inespecíficos. Após o bloqueio, as células foram incubadas com o anticorpo primário diluído na solução de bloqueio por 1 hora à temperatura ambiente. Em seguida, as lamínulas foram lavadas com PBS (5 vezes de 5 minutos) e incubadas com o anticorpo secundário conjugado com um radical fluorescente, durante 1 hora à temperatura ambiente. As células foram, então, lavadas com PBS (5 vezes de 5 minutos), montadas sobre lâminas utilizando o meio de montagem Prolong Diamond (Thermo Fisher Scientific) e seladas com esmalte cosmético.

3.2.15 Anticorpos e marcadores fluorescentes utilizados

- Marcador nuclear DAPI (dihidrocloreto de 4',6' diamino-2-fenilindol – Molecular Probes) utilizado na concentração de 2µg/ml.
- Anticorpo primário policlonal anti-EMC2 (Protein Tech, 25443-1-AP) produzido em coelho. Para o ensaio de Western Blot e imunocitoquímica foi utilizada a diluição 1:200, e na imunohistoquímica 1:50.
- Anticorpo primário monoclonal anti-γ-tubulina Clone GTU-88 (Sigma T6557) utilizado na diluição.
- Anticorpo primário monoclonal anti-β-tubulina Clone TUB 2.1 (Sigma T4026) utilizado na diluição 1:200.
- Anticorpo secundário IgG de cabra anti-coelho Alexa Fluor 546 dye (Life Technologies A11035) na diluição 1:400.
- Anticorpo secundário IgG de jumento anti-camundongo Alexa Fluor 488 dye (Invitrogen A21202) na diluição 1:400.
- Anticorpo anti-mouse IgG conjugado com HRP (Horseradish Peroxidase) (Promega W4028) na diluição 1:2500.
- Anticorpo produzido em jumento anti-coelho conjugado com HRP (Horseradish Peroxidase) (NA934V GE Healthcare UK Limited) na diluição 1:10000.

3.3 Imunohistoquímica

As amostras utilizadas foram gentilmente cedidas pelo prof. Dr. Alfredo Ribeiro da Silva, do departamento de Patologia e Medicina Legal da Faculdade de Medicina de Ribeirão Preto. São amostras provenientes de pacientes do Hospital das Clínicas de Ribeirão Preto diagnosticados com fibroadenoma mamário (n=33), carcinoma ductal mamário in situ (n=43) e carcinoma mamário invasivo (n=92), além de amostras de tecido normal de mama adjacente ao tumor (n=8). As amostras, parafinizadas e dispostas na lamina para a técnica de microarranjo tecidual (TMA, do inglês *tissue microarray*), foram imunomarcadas com o anti-EMC2 na diluição 1:50. Foi utilizado o kit REVEAL Sistema de detecção Livre de biotina e seguido o protocolo do fabricante.

RESULTADOS

4. Resultados

4.1 Análise do complexo EMC usando o banco de dados TCGA

O banco de dados *The Cancer Genome Atlas* (TCGA) é uma plataforma online pública que reúne dados de pacientes diagnosticados com câncer de centros cadastrados dos EUA. Para download dos dados do TCGA, utilizamos a plataforma cBioPortal for Cancer Genomics (www.cbioportal.org), que apresenta também ferramentas de análises dos dados online e interativas para o público.

Inicialmente, analisamos os dados de expressão gênica dos pacientes em um painel de 22 tipos de câncer, comparando os níveis transcricionais dos tecidos tumorais e normais adjacentes ao tumor, para todas as proteínas do complexo EMC (Figura 5). Numa visão geral, os genes apresentaram tendência de estarem superexpressas no tecido tumoral, com exceção de EMC3 e EMC5. O gene EMC8 aparece com aumento em 13 tipos de câncer, seguido de EMC10 que aparece em 11 (50% dos tumores analisados).

Dos tipos de câncer pesquisados, o carcinoma pulmonar de células escamosas (LUSC) foi o que mais apresentou proteínas alteradas, totalizando 9. O carcinoma mamário invasivo (BRCA), objeto de análises posteriores *in vitro* neste trabalho, aparece com aumento em 7 membros do complexo, sendo que os 3 genes restantes também apresentam aumento nas medianas, porém com valor de p menor que 0,05.

Figura 5- Comparação entre os níveis de expressão gênica do tecido canceroso e do tecido normal adjacente ao tumor dos genes EMC1-10, utilizando os dados do TCGA. Os dados foram obtidos através do software cBioPortal. Os valores são apresentados como diagrama de caixa Tukey e analisados estatisticamente pelo teste t de duas caudas paramétrico de amostras não pareadas, onde o valor de p é representado por * <0.05, **<0.01, ***<0.001 e ****<0.0001.

Os mesmos grupos de pacientes foram analisados quanto à existência de alterações genéticas: amplificação, homodeleção e/ou mutações com troca de

sentido, sem sentido, e fora de fase (Figura 6). O gene é considerado amplificado quando o número de cópias gênicas é maior do que dois e homodeletado quando não há cópias do gene. Já nas mutações, há alterações em um ou mais nucleotídeos que resultam em troca de aminoácido causando parada precoce na tradução (mutação sem sentido), mudança na fase traducional (mutação fora de fase) ou mudança pontual em um único aminoácido (mutação troca de sentido).

Em destaque, aparece EMC2 com altos níveis de amplificação em um padrão uniforme ao longo dos tumores. A maior ocorrência de alterações genéticas é vista em câncer de ovário (OV), no qual cerca de 23% dos pacientes apresentam amplificação. Em segundo lugar aparece BRCA com 16% de amplificação. A região cromossômica de EMC2 (8q23) é relatada em muitos trabalhos como sendo uma região amplificada em cânceres (Kang et al. 2014).

Genetic Alteration

Figura 6- Alterações genéticas de EMC1-10 no painel de câncer do TCGA. São analisadas as alterações: amplificação, homodeleção e mutações com troca de sentido, sem sentido, e fora de fase. O download dos dados foi feito através do software cBioPortal e a quantidade de pacientes de cada tipo de câncer está especificada na seção 3.1 (Métodos).

Na análise de expressão dos genes EMCs, levamos em consideração o parâmetro *z-score*, de modo que os pacientes com *z-score*>2 são considerados superexpressos e *z-score*<-2, subexpressos. Nessa classificação, EMC2 aparece

com impressionantes proporções de pacientes superexpressos, chegando a 43% em tumor epitelial de ovário (OV), 40% em carcinoma invasivo mamário e 37% em carcinoma hepatocelular de fígado (LIHC) (Figura 7).

Figura 7– Alteração de expressão dos genes EMC1-10, no painel de câncer do TCGA. A classificação em subexpresso e superexpresso considera o valor de z-score <-2 e >2 , respectivamente. O download dos dados foi feito através do software cBioPortal e a quantidade de pacientes de cada tipo de câncer está especificada na seção 3.1 (Métodos).

Outra análise realizada foi quanto a sobrevida dos pacientes, comparando os que apresentam baixa expressão (25º percentil) com os com alta expressão (75º percentil) dos genes do complexo EMC. Levamos em consideração a coeficiente de regressão COX e o valor de p. O gene EMC1 foi o que apareceu em mais tipos cancer (n=6), no qual, em todos eles, a superexpressão associa-se com diminuição na sobrevida (Tabela suplementar 1). Além de EMC1, também EMC6 e EMC9 aparecem com coeficiente de regressão de COX positivo em todos os cânceres que foram encontrados com relação com sobrevida.

Figura 8- Curvas de sobrevida comparando os pacientes com alta e baixa expressão gênica. O grupo “Alta” é composto por 25% dos pacientes com maior expressão gênica e o grupo “Baixa”, 25% dos pacientes com menor expressão gênica. Todos os tipos de câncer foram analisados com relação aos 10 genes do complexo EMC. Nessa figura são apresentados os casos em que p-value < 0.05. Os gráficos foram obtidos diretamente do site da OncoLnc.

4.2 Análise de EMC2 em câncer de mama

Devido aos interessantes resultados encontrados com EMC2, como alto índice de amplificação e superexpressão em diversos tumores, assim como diferença na sobrevida de alguns tipos tumorais, buscamos aprofundar o estudo desse gene. Escolhemos o câncer de mama como modelo de estudo pois além de ter sido um dos tipos tumorais com resultados mais relevantes encontrados nas análises feitas até agora, esse modelo tumoral também já é utilizado em nosso laboratório, incluindo no estudo feito sobre EMC1. Além disso, no banco de dados do TCGA, o câncer de mama é um dos tipos tumorais com maiores números de pacientes, o que é desejável para uma melhor acurácia nos resultados.

A plataforma cBioPortal (Gao et al., 2003; Cerami et al. 2012) agrupa dados de diversos bancos, incluindo o TCGA, e apresenta ferramentas para visualização dos dados de forma organizada e didática. De um total de 960 pacientes com carcinoma invasivo mamário disponíveis, 410 (43%) apresentam alguma alteração em EMC2. Dentre as alterações, 16% corresponde a pacientes com amplificação no gene e 40% são pacientes classificados como superexpressos ($\text{score-z} \geq 2.0$). Há ainda 5 pacientes com subexpressão, sendo que 1 deles apresenta homodeleção; e outros 5 pacientes apresentam mutações, nas quais 4 são consideradas com troca de sentido e 1 apresenta codon de parada precoce (Figura 9).

Figura 9- Alterações genéticas de EMC2 nos pacientes com câncer de mama do TCGA. **A)** Cada retângulo representa um paciente, e estes foram dispostos lado a lado, totalizando 960 pacientes. A fita superior apresenta as alterações gênicas: amplificação, superexpressão de mRNA, homodeleção, subexpressão de mRNA e mutações troca de sentido e sem sentido; e a fita inferior mostra o *heatmap* da expressão gênica. As fitas foram ampliadas duas vezes em regiões convenientes para serem visualização com mais detalhe e nitidez. **B)** Localização das alterações na proteína EMC2 que foram consequentes das mutações gênicas encontradas em 5 dos 960 pacientes. Os diagramas foram obtidos diretamente do software cBioPortal.

Como visto nas análises feitas até agora, EMC2 aparece com amplificação em muitos tipos de tumores. Isso se dá pois a região cromossômica em que o gene se encontra é altamente amplificada em câncer (Kang JU, Bosch T et al., Garnis C et al.). Pelos dados do TCGA, também é possível verificar tais amplificações em pacientes com câncer de mama (Figura 10).

Figura 10 – Amplificação da região cromossômica 8q em pacientes com câncer de mama do TCGA. O retângulo da esquerda mostra toda extensão do braço q do cromossomo 8 e o retângulo da direita focaliza na região cromossômica onde se localiza o gene EMC2. Em vermelho é marcada a região com amplificação e em azul deleção. A figura foi obtida diretamente do software UCSC Cancer Browser.

Verificamos na seção anterior, que pacientes com câncer de mama do TCGA com maior expressão de EMC2 apresentam pior sobrevida comparado aos pacientes com menor expressão (Figura 8). Entretanto, um importante oncogene, MYC, localiza-se na região cromossômica 8q24, ou seja, próximo ao locus de EMC2, e portanto em muitos casos encontra-se também amplificado (Figura 11 A e B). Pensamos na possibilidade de que o pior prognóstico encontrado nos pacientes com maior expressão de EMC2 tenha fortes influências da amplificação de MYC, e desse modo, não seria EMC2 o responsável pela intensificação na malignidade, mas sim ele seria apenas um gene passageiro da região frequentemente amplificada. Para testarmos essa hipótese, comparamos a sobrevida dos pacientes com

superexpressão de EMC2 porém MYC com níveis de mRNA normais contra o restante dos pacientes. Com isso, conseguimos verificar que mesmo sem a influência da superexpressão de MYC, pacientes com altos índices de expressão de EMC2 e outras alterações no gene apresentam pior sobrevida (Figura 11C).

Figura 11 – A) Esquema do cromossomo 8 e as localizações dos genes EMC2 e MYC, mostrando a proximidade entre eles. **B)** Perfil de alterações de EMC2 e MYC nos pacientes com câncer de mama do TCGA. Os pacientes foram dispostos na mesma sequência na régua EMC2 e MYC. **C)** Curva de sobrevida de Kaplan-Meier comparando o grupo com superexpressão de EMC2 e com expressão normal de MYC, contra o restante dos pacientes. Os gráficos B e C foram obtidos diretamente do software cBioPortal.

Verificamos ainda o nível de expressão de EMC2 considerando os principais biomarcadores moleculares para o câncer de mama: receptor de estrogênio, receptor de progesterona e HER2. Verificamos que pacientes negativos para os receptores de estrogênio de progesterona apresentam expressão maior de EMC2, assim como pacientes triplo negativos (Figura 12).

Figura 12 – Expressão de EMC2 com relação aos biomarcadores moleculares para o câncer de mama: receptor de estrogênio, receptor de progesterone e HER2. Os valores são apresentados como diagrama de caixa Tukey e analisados estatisticamente pelo teste t de duas caudas, onde o valor de p é representado por * <0.05, **<0.01, ***<0.001 e ****<0.0001.

Utilizando as mostras de câncer de mama do banco TCGA, selecionamos os genes com maior correlação com a expressão de EMC2 (coeficiente de Pearson ≥ 0.3) (Figuras 13A e B). Fizemos então uma análise de predição de vias através do software DAVID Bioinformatics Resources 6.8 usando *KEGG Pathway*. Foram encontradas 7 categorias com alta correlação com EMC2 (Figura C). Em primeiro lugar, aparece a via de ciclo celular, no qual 10 genes dessa categoria correlacionados com a expressão de EMC2. Em seguida, aparece a via de proteólise mediado por ubiquitina.

Figura 13 – A) Correlação entre o conjunto total de genes e EMC2, calculados através dos coeficientes de Pearson e Spearman. Em verde estão representados os genes selecionados com maior correlação com EMC2 (coeficiente de Pearson ≥ 0.3). **B)** *Heat map* da expressão gênica dos genes selecionados no grupo de pacientes com menor e maior expressão de EMC2. **C)** Predição de vias metabólicas utilizando os genes selecionados

4.3 Superexpressão e silenciamento de EMC2

Com o intuito de analisar *in vitro* o papel de EMC2, fizemos a superexpressão e o silenciamento do gene em linhagens de câncer de mama.

Para a superexpressão, subclonamos a sequência gênica no plasmídeo pGEM, e em seguida, transferimos para o plasmídeo mCherry-Parkin modificado (materiais e métodos). Vinte e quatro horas (24h) após a transfecção, as células contendo o plasmídeo controle ficaram com a marcação difusa na célula, ao passo que as células transfectadas com mCherry-EMC2 estavam arredondadas e com marcação perinuclear (Figura 14). Mantendo as células em cultura, percebemos que

o número de células transfectadas, ou seja, aquelas que possuíam marcação vermelha da mCherry, rapidamente diminuía. Desse modo, antes mesmo de conseguirmos selecionar as células transfectadas com a droga de seleção G418, as células vermelhas já não eram mais encontradas na cultura. Nas células com MCherry-controle, também notamos redução no número de células vermelhas com o tempo. Acreditamos que a elevação na expressão de EMC2 possa ter potencializado a morte celular possivelmente por estresse, mas como a diminuição das células transfectadas também aconteceu com a cultura mCherry-controle, o ideal seria testarmos outro plasmídeo.

Figura 14 – Superexpressão de EMC2 na linhagem HEK293-T. A) Células transfectadas com o plasmídeo controle (mCherryCTRL) e **B)** o plasmídeo mCherryEMC2. Em vermelho está a fluorescência da proteína mCherry e em azul a marcação do núcleo com DAPI. Nessa figura também aparecem os campos claros e as respectivas sobreposições dos canais.

Realizamos o silenciamento de EMC2 na linhagem de câncer de camundongo 4T1 utilizando três sequências distintas de sh (Figura 15A).

Sabe-se que uma importante característica das células tumorais é a capacidade proliferativa aumentada. Realizamos então o ensaio de proliferação e verificamos que as células silenciadas para EMC2 proliferam menos comparadas ao controle (Figura 15B).

Figura 15 – Silenciamento de EMC2 e ensaio de proliferação. A) Western blot e PCR em tempo real comprovando silenciamento de EMC2 na linhagem celular 4T1. B) Ensaio de proliferação utilizando cristal violeta que mostra diminuição na proliferação das células silenciadas para EMC2 comparadas ao controle. Os valores são apresentados como média $\pm$ desvio padrão e foram analisados estatisticamente pelo teste t de duas caudas, onde o valor de p é representado por * <0.05, **<0.01, ***<0.001 e ****<0.0001. Os testes feitos comparam cada grupo ao shCTRL, e em B cada tempo foi analisado separadamente.

4.4 Análise do silenciamento de EMC2 no ciclo celular

Em vista da diminuição na capacidade de proliferação de células silenciadas para EMC2, e em conjunto com a predição de vias que relaciona a expressão de EMC2 a genes associados ao ciclo celular, decidimos investigar uma possível relação entre EMC2 e essa via de ciclo celular. Para isso, sincronizamos as células com nocodazol e vimos que a saída da mitose em células silenciadas é ligeiramente mais devagar comparado a controle (Figura 16).

Figura 16 – Células 4T1 silenciadas para EMC2 e controle, sincronizadas com nocodazol na progressão do ciclo celular. O ajuste dos picos foi feito automaticamente utilizando o software ModFit.

4.5 Análise por imunocitoquímica do fuso mitótico

Dado que a progressão do ciclo celular, quando houve a sincronização na fase M pela droga nocodazol, mostrou-se mais devagar nas células silenciadas comparando ao controle, buscamos avaliar se havia problemas na formação do fuso mitótico, o que justificaria os resultados. Dessa forma, realizamos a imunomarcagem do fuso mitótico através da marcação de β -tubulina e DNA, além da marcação da proteína EMC2 com o uso do anticorpo anti-EMC2 (Figura 17). Contudo, não notamos diferença ou possíveis problemas na formação do fuso mitótico em células silenciadas para EMC2 já que foi possível observar os dois fusos simétricos e o DNA condensado ligado a ele.

Figura 17– Fuso mitótico de células 4T1 knockdown de EMC2 (sh260) e controle (shCTRL). Em vermelho e verde estão as imunomarcações de EMC2, β -tubulina, respectivamente e em azul a marcação do DNA com DAPI. Na esquerda, estão as reconstruções 3D e na direita, as projeções máximas (barra corresponde a 10 μ m).

4.6 Mudança na morfologia em células silenciadas para EMC2

Ao silenciarmos as células de câncer de mama MCF7, notamos mudança na morfologia celular. A célula epitelial, após o silenciamento, passou a ficar fusiforme e com prolongamentos maiores, característica de células mesenquimais (Figura 18). O mesmo fenótipo foi visto quando silenciarmos a 4T1 utilizando o sh260, sh885 e sh1007 (dado não mostrado), porém a mudança na morfologia foi mais evidente no sh260.

Figura 18 – A) PCR em tempo real comprovando silenciamento de EMC2 na linhagem celular de câncer de mama humano MCF7. **B)** Mudança na morfologia celular após o silenciamento de EMC2.

4.7 Localização de EMC2 nas células de câncer de mama

Imunomarcamos a proteína EMC2 nas linhagens de câncer de mama MCF7 (Figura 19) e 4T1 (Figura 20). É possível notar que a marcação de EMC2 é encontrada por todo o citoplasma, e é possível notar uma marcação reticular, sugestiva de associado ao retículo endoplasmático. Também observamos uma marcação mais fraca de EMC2 no núcleo.

Figura 19 – Imunomarcacão de EMC2 (em vermelho) na linhagem celular MCF7. Em azul está a marcação do núcleo através do DAPI.

Figura 20 – Imunomarcação de EMC2 (em vermelho) na linhagem celular 4T1. Em verde está a marcação de β -tubulina e em azul, a marcação do núcleo com DAPI. Ainda está apresentado a imagem obtida no campo claro e a sobreposição dos canais.

O anticorpo utilizado para a marcação de EMC2 foi testado por nós em ensaios de western blot (Figura suplementar 1) e imunocitoquímica nas células transfectadas com o plasmídeo mCherry-EMC2 (Figura suplementar 2). Nota-se uma alta especificidade do anticorpo no ensaio de western blot já que marca apenas um banda e na altura espera para EMC2 (~35kDa). Na figura suplementar 2, é possível notar que o anti-EMC2 marca intensamente as células transfectadas, sendo que essa marcação sobressai a marcação da proteína endógena das células não transfectadas devido a quantidade de proteína em cada célula. Dessa forma, a marcação de anti-EMC2 na célula transfectada representar alta especificidade do

anticorpo, e além disso, também mostra que a proteína produzida a partir do plasmídeo transfectado mCherry-EMC2 parece estar similar à endógena, já que é reconhecida pelo anticorpo.

4.8 Imunohistoquímica de amostras de pacientes com câncer de mama

Sabemos a importância da análise de amostras de pacientes para o estudo de genes com potencial associação com câncer, em especial, análises a nível proteico. Por isso, imunomarcamos EMC2 em amostras de pacientes do Hospital das Clínicas de Ribeirão Preto diagnosticados com fibroadenoma de mama, carcinoma in situ e carcinoma invasivo mamário, além de tecidos adjacentes ao tumor de margem cirúrgica.

Como controle da reação de imunohistoquímica, utilizamos tecido de rim (Figura 21). É possível notar que a intensidade da marcação de EMC2 varia de acordo com o tipo celular. Isso mostra tanto alta especificidade do anticorpo quanto variação da expressão de EMC2 de forma tecido específico. Além disso, nota-se que a marcação de EMC2 é citoplasmática, e em algumas células, marca-se intensamente o núcleo. Nos tecidos de mama de margem cirúrgica e no tecido canceroso de mama, a marcação é somente citoplasmática (Figura 19).

Figura 21 – Controle da reação de imunohistoquímica. Tecido de rim cujas reações foram feitas na **A)** ausência e **C)** presença da proteína anti-EMC2, sendo portanto, os controles negativo e positivo, respectivamente. **B)** e **D)** são ampliações das figuras **A)** e **C)**, nessa ordem.

As amostras de mama foram classificadas de acordo com a intensidade de marcação de EMC2 e foram divididas em quatro categorias: negativo, fraco, médio e forte. (Figura 22). A categoria forte corresponde à maior intensidade encontrada nas amostras de mama analisadas neste trabalho, e a negativa corresponde à ausência de marcação.

Figura 22 – Parâmetro interno estabelecido para classificação das amostras de câncer de mama e mama normal adjacente ao câncer dos pacientes do HC-Ribeirão Preto de acordo com intensidade de marcação da proteína EMC2 por imunohistoquímica. As amostras são de pacientes diagnosticados com carcinoma invasivo mamário.

Após a análise das amostras, verificamos que no tecido de mama normal, a marcação de EMC2 é, em sua maioria, médio e forte, mostrando que essa proteína pode apresentar um papel importante nesse tipo de tecido (Figura 23). Curiosamente, foi visto também que em células de fibroadenoma, a marcação é mais fraca do que no tecido normal, assim como nos carcinomas in situ e invasivo. Portanto, a partir destes dados, vemos que os baixos níveis de EMC2 parecem se relacionar a estágios mais avançados da doença.

Figura 23 – Quantificação das amostras analisadas classificadas quanto a intensidade de marcação de EMC2 (negativo, fraco, médio e forte).

DISCUSSÃO

5. Discussão

O complexo EMC localiza-se na membrana do retículo endoplasmático e está associado a vias de degradação de proteínas mal enoveladas (ERAD), resposta ao estresse de RE via UPR, e transferência de PS do RE para a mitocôndria (Jonikas et al., 2011; Christianson et al., 2012; Lahiri et al., 2014). Em células de câncer, devido a diversos fatores causadores de estresse como hipóxia e privação de nutrientes, vias relacionadas ao RE tais como UPR, encontram-se frequentemente ativadas (Kato, et al. 2015). Dessa forma, o complexo EMC pode estar relacionado a processos envolvidos ao câncer.

Na busca de uma investigação ampla sobre os membros do complexo EMC em câncer, utilizamos os dados do banco TCGA contendo expressão gênica, mutações e sobrevida dos pacientes. Comparando a expressão gênica de amostras tumorais em relação aos tecidos normais adjacentes ao tumor, vimos que EMC1 aparece aumentado ($p < 0.05$) em 9 dos 22 tipos de câncer analisados, incluindo carcinoma urotelial de bexiga (BLCA) e carcinomas de pulmão (LUAD e LUSC). Em 2002, estudos em nosso laboratório haviam apontado EMC1 como um gene diferencialmente expresso em linhagens de melanoma metastático (Sousa, 2002), o que também foi visto por Valsesia e colaboradores em 2011. Em nosso estudo, não foi possível analisar esses dados de melanoma pois só havia 1 paciente com amostra normal (melanócito) disponível no TCGA.

Ainda sobre EMC1, este foi o gene do complexo que apareceu em mais cânceres com diferença na curva de sobrevida ($n=6$). E todos os coeficientes de regressão de Cox são positivos, ou seja, pacientes com maiores níveis de expressão de EMC1 apresentam pior sobrevida do que os pacientes com menor expressão.

Os demais genes do complexo EMC também apresentam tendência de estarem com expressão aumentada no câncer em comparação aos tecidos normais, porém não ocorre com todos os genes. Por exemplo, EMC5 aparece com expressão diminuída em 8 dos 13 tipos de tumor que ele se encontra diferencialmente expresso. Em glioblastoma (GBM), EMC10 apresenta subexpressão e homodeleção em alguns pacientes, o que corrobora com o trabalho de Junes-Gill e colaboradores, que em 2010, descreveram a proteína como supressora de proliferação celular em células de glioma.

Porém, em um aspecto geral, os genes EMCs encontram-se superexpressos em um grupo significativo de amostras, de acordo com a classificação através do *z-score*. Além disso, muitos tipos tumorais apresentam amplificação de genes do complexo, o que justifica alguns valores de superexpressão. Ainda foi visto, que em muitos tipos tumorais, membros do complexo EMC apresentam coeficiente de regressão de Cox positivos, o que significa diminuição da sobrevida em pacientes com expressão mais elevada. Dessa forma, estes dados trazem evidências de que o aumento de função do complexo parece favorecer o desenvolvimento do tumor.

Dos membros do complexo EMC, o gene EMC2 se destacou nas análises dos dados do TCGA pela frequente amplificação e superexpressão encontrada em diversos tumores. Em câncer de mama, os pacientes com alterações em EMC2 apresentam sobrevida pior que o restante dos pacientes, até mesmo quando excluímos os pacientes com superexpressão do importante oncogene MYC, que por se localizar também na região amplificada 8q, em alguns casos também aparecem superexpressos.

Para delinear mecanismos em que EMC2 pode estar atuando na célula, fizemos a análise de predição de vias de genes altamente correlacionados a EMC2, e encontramos as vias do ciclo celular e ubiquitinação. Em vista disso, realizamos experimentos *in vitro* em ensaios de ciclo celular, e vimos que o knockdown de EMC2 parece retardar a progressão da fase M. Além disso, verificamos que o silenciamento da proteína diminui a taxa proliferativa da célula, o que pode ser explicado pelo retardo do ciclo celular. Questionamos também se o atraso na progressão da fase M era devido a má formações no fuso mitótico. Mas após imunomarcarmos o fuso através da β -tubulina, vimos que aparentemente não há defeitos na formação do fuso.

Nos ensaios de imunohistoquímica dos pacientes com câncer de mama do HC-Ribeirão Preto, vimos que em pacientes com estágios mais avançados da doença há menores proporções de pacientes com marcação de EMC2 médio e forte. A princípio, esse resultado foi oposto ao esperado já que os resultados até o momento indicaram que o aumento da função da proteína parece favorecer a progressão do câncer. Porém, quando silenciemos EMC2 notamos uma diferença na morfologia celular, dando indícios de que a célula passa a apresentar um fenótipo mais mesenquimal em contraposição ao fenótipo epitelial. Essa transformação

epitélio-mesenquima também foi visto no estudo do gene EMC1 e bem caracterizado pelos alunos Mariani Cincos e Roberto Molina. Em conjunto, os dados dão evidências de que EMC2 pode atuar de duas formas: a alta expressão contribui com a maior capacidade de proliferação, mas o silenciamento facilitaria outras etapas da progressão tumoral, como por exemplo, a transição epitélio-mesenquima.

É válido ressaltar que apesar da participação efetiva da proteína EMC2 no complexo EMC, acreditamos que, devido a alta distribuição da proteína na célula, que é encontrada em todo o citoplasma e também no envelope nuclear (Greger et al. 2001) e por ela ser uma proteína sem domínio transmembrana, é provável que EMC2 possa atuar em outras funções da célula independentes do complexo EMC. Dados que favorecem essa hipótese são as interações proteína-proteína encontradas de EMC2 com outras proteínas como a RASSF1C (Dmitriev et al., 2007), CCDC33 e IKZF3 (Rolland et al., 2014) através de ensaios de dois híbridos.

Apesar dos esforços realizados neste trabalho na busca de entender funções da EMC2, há muito ainda o que ser estudado sobre essa proteína ainda pouco caracterizada e que o início do estudo é recente em nosso laboratório. Neste trabalho, trouxemos indícios da participação de EMC2 na proliferação celular, ciclo celular e transição epitélio-mesenquima, o que pode estar influenciando a progressão tumoral. Porém, a partir desse ponto, é necessário aprofundar os experimentos para elucidarmos os mecanismos de forma completa.

CONCLUSÃO

6. Conclusão

Os dados do TCGA mostraram que em muitos tipos de câncer os membros do complexo EMC estão com expressão gênica aumentada em comparação ao tecido normal e, em geral, pacientes que possuem maior expressão apresentam pior sobrevida. O gene EMC2, em especial, se localiza na região cromossômica 8q23.1 altamente amplificada em diversos tumores, e possui alta proporção de pacientes superexpressos. Em câncer de mama, pacientes com alterações em EMC2 apresentam pior sobrevida comparado aos demais pacientes, mesmo quando excluímos os pacientes com superexpressão de MYC, importante oncogene localizado próximo ao EMC2 e também amplificado em muitos casos.

Ainda utilizando os dados dos pacientes com câncer de mama do TCGA, a predição de vias dos genes altamente correlacionados com EMC2 apontam possível relação do gene com a via de ciclo celular. Ensaios de sincronização da célula utilizando a droga despolimerizante de microtúbulos, nocodazol, mostraram que as células 4T1 silenciadas para EMC2 apresentam atraso na progressão do ciclo celular na fase M. Além disso, o silenciamento das células causaram diminuição na taxa proliferativa das células, o que pode ter relação com o atraso no ciclo celular.

Ensaios de imunocitoquímica indicam que a proteína EMC2 localiza-se no citoplasma de modo sugestivo de associação com RE. A localização citoplasmática também é vista nos ensaios de imunohistoquímica nas células epiteliais do tecido mamário e renal. Além disso, também é visto em algumas células de rim uma interessante marcação nuclear intensa.

Avaliando EMC2 a nível proteico através da marcação por imunohistoquímica de amostras de pacientes com câncer de mama do HC-RP, incluindo fibroadenoma, carcinoma in situ, carcinoma invasivo e amostras de mama normal, vimos que em estágios mais avançados da doença, a proporção de pacientes com marcação média e forte é curiosamente menor comparado ao tecido normal. Esse resultado pode estar relacionado à possível transição epitélio-mesênquima de células com baixa expressão de EMC2. Essa transição foi hipotetizada devido a mudança morfológica das células causada pelo silenciamento de EMC2 nas linhagens MCF7 e 4T1. Porém, este é um dado preliminar que precisa de mais investigações para se ter uma conclusão correta.

BIBLIOGRAFIA

7. Bibliografia

CERAMI, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*. 2012 2; 401.

CHRISTIANSON JC, OLZMANN JA, SHALER TA, et al. Defining human ERAD network through an integrative mapping strategy. *Nat Cell Biol* 14: 93-105. 2012.

CINCOS, MLM. EMC1, uma proteína transmembrana do RE, associada com fenótipo mesenquimal e propriedades malignas de células de câncer de mama. Tese de doutorado. Faculdade de Medicina de Ribeirão Preto. Universidade de São Paulo. São Paulo. 2017.

CINQUEIRA MB, MOREIRA MAR, SOARES LR e Freitas-Junior R. Subtipos moleculares do câncer de mama. Revisão. <http://files.bvs.br/upload/S/0100-7254/2011/v39n10/a2965.pdf>

CLAUS EB, RISCH N, THOMPSON WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. *AmJHum Genet*, v.48. 1991.

DAVIS NM, SOKOLOSKY M, STADELMAN K, et al. Deregulation of the EGFR/PI3K/PTEN/Akt/mTORC1 pathway in breast cancer: possibilities for therapeutic intervention. *Oncotarget*. 2014 Jul 14;5(13):4603-50.

DIMRI G, BAND H, BAND V. Mammary epithelial cell transformation: insights from cell culture and mouse models. *Breast Cancer Res* 7: 171-179. 2005.

DMITRIEV, et al. Characterization of hampin/MSL1 as a node in the nuclear interactome. *BBRC*. 2007.

DREGER, et al. Nuclear envelope proteomics: Novel integral membrane proteins of the inner nuclear membrane. *PNAS* 2001.

ETHIER SP. Identifying and validating causal genetic alterations in human breast cancer. *Breast Cancer Res.*, 62, 6240-6245. 2003

GAO et al. Integrative analysis of the complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, 2013, 6 pl1.

GARNIS C, et al. Novel regions of amplification on 8q distinct from the MYC locus and frequently altered in oral dysplasia and cancer. *Genes Chromosomes Cncl.* 39(1):93-8. 2004.

HANAHAN D & WEINBERG RA. The hallmarks of cancer. *Cell* 144: 646-674. 2011.

HERSCHKOWITZ JI, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* 2007; 8(5):R76.

HOJA, et al. A Visual Intracellular Classification Strategy for Uncharacterized Human Proteins. *Experimental Cell Research* 259, 239–246. 2000.

JONIKAS MC, COLLINS SR, DENIC V, et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* 323: 1693-1697. 2009.

JUNES-GILL KS, GALLANHER TK, GLUZMAN-POLTORAK Z, et al. hHSS1: a novel secreted factor and suppressor of glioma growth located at chromosome 19q13.33. *J Neurooncol.* 2011. 102:197-211.

KANG JU. Chromosome 8q as the most frequent target for amplification in early gastric carcinoma. *Oncol Lett*, 7(4):1139-1143. 2014.

KRACHLER AM, SHARMA A, KLWANTHOUS. Self-association of TPR domains: Lessons learned from a designed, consensus-based TPR oligomer. *Proteins*. **78** (9): 2131–43. 2010.

LAHIRI S, et al. A Conserved Endoplasmic Reticulum Membrane Protein Complex (EMC) Facilitates Phospholipid Transfer from the ER to Mitochondria. Plos Biology. 2014.

LOUIE RJ, et al. A yeast phenomic model for the gene interaction network modulating CFTR- Δ F508 protein biogenesis. Genome Medicine, 4:103,2012.

NARENDRA D, TANAKA A, SUEN DF, YOULE RJ. Parkin is recruited to impaired mitochondria and promotes their autophagy. Cell Biol. 183(5):795-803.

MOLINA RAS. EMC1 contribui para a malignidade em células de câncer de mama. Tese de doutorado. Faculdade de Medicina de Ribeirão Preto. Universidade de São Paulo. São Paulo. 2014.

PEROU CM, et al. Molecular portraits of human breast tumours. Nature 2000. Aug 17; 406(6797):747-52.

PRAT A, PARKER J, KARGINOVA O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68.

REEVES, et al. Ras-association domain family 1C protein promotes breast cancer cell migration and attenuates apoptosis. BMC Cancer 2010.

RICHARD M, BOULIN T, ROBERT VJ, RICHMOND JE, BESSEREAU JL. Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. Proc Natl Acad Sci U S A, 2013.

ROLLAND, et al. A Proteome-Scale Map of the Human Interactome Network. Cell 2014.

ROMAGNOLO AP, ROMAGNOLO DF, SELMIN OI. BRCA1 as target for breast cancer prevention and therapy. *Anticancer Agents Med Chem*. 2015; 15(1):4-14.

SHEN X, KAN S, HU J, et al. EMC6/TMEM93 suppresses glioblastoma proliferation by modulating autophagy. *Cell Death and Disease*. 2016. 7, e2043.

SILVA RR. O gene KIAA0090 é ativado em lesão pré-neoplásica e seu silenciamento por siRNA causa morte celular em linhagem de melanoma. Dissertação de mestrado. Faculdade de Medicina de Ribeirão Preto. Universidade de São Paulo. 2002.

SILVA RR. Silenciamento de EMC1 reduz tumorigenicidade em modelo experimental de melanoma. Tese de doutorado. Faculdade de Medicina de Ribeirão Preto. Universidade de São Paulo. 2014.

SINN HP, KREIPE H. A Brief Overview of the WHO Classification of Breast, 4th Edition, Focusing on Issues and Updates from 3rd Edition. *Breast Care (Basel)*. 2013 Maio; 8(2): 149-154.

SOUSA JF. Estudo da expressão gênica em linhagens de melanoma humano de diferentes estágios da progressão tumoral. Tese de doutorado. Faculdade de Medicina de Ribeirão Preto. Universidade de São Paulo. São Paulo. 2002.

TCGA Research Network: <http://cancergenome.nih.gov/>

TSAI YC e WEISSMAN AM. The Unfolded Protein Response, Degradation from the Endoplasmic Reticulum, and Cancer. *Genes Cancer*, v.1(7):764-778,2010.

TURASHVILI G, BROGI E. Tumor Heterogeneity in Breast Cancer. *Front Med (Lausanne)*. 2017; 4:227.

UCSC Xena: <http://xena.ucsc.edu>

VALSESIA A, RIMOLDI D, MARTINET D, et al. Network-Guided Analysis of Genes with Altered Somatic Copy Number and Gene Expression Reveals Pathways Commonly Perturbed in Metastatic Melanoma. Plos One. 2011.

van der Bosch T, van Beek JG, Vaarwater J, Verdijk RM, Naus NC, Paridaens D, de Klein A, Kiliç E. 2012.

ANEXO

Gene	Cancer	Regressão de Cox	p-value	Correlação FDR
EMC1	LGG	0.38	2.40e-05	1.59e-04
	BLCA	0.324	1.80e-05	7.37e-03
	SARC	0.254	1.00e-02	1.02e-01
	LIHC	0.226	1.30e-02	1.03e-01
	CESC	0.332	1.10e-02	1.81e-01
	HNSC	0.156	3.00e-02	2.67e-01
EMC2	PAAD	0.225	3.50e-02	1.48e-01
	SKCM	-0.158	4.10e-02	1.52e-01
	BLCA	0.18	2.60e-02	1.65e-01
EMC3	COAD	-0.227	2.90e-02	3.44e-01
EMC4	LGG	-0.363	1.10e-04	5.50e-04
	HNSC	0.226	1.40e-03	8.39e-02
	LUAD	0.171	3.00e-02	1.53e-01
	LIHC	-0.189	4.50e-02	2.08e-01
EMC5	KIRC	-0.288	4.80e-04	2.90e-03
	SKCM	-0.164	1.90e-02	9.47e-02
	HNSC	0.201	2.60e-03	1.05e-01
	BLCA	0.167	2.50e-02	1.62e-01
	BRCA	0.243	4.40e-03	1.87e-01
EMC6	KIRC	0.283	4.40e-03	1.65e-02
	SKCM	0.155	3.10e-02	1.28e-01
	LUAD	0.168	2.20e-02	1.30e-01
EMC7	SARC	-0.285	5.20e-03	7.39e-02
	HNSC	0.222	2.00e-03	9.71e-02
	BLCA	0.173	2.60e-02	1.65e-01
EMC8	KIRP	-0.424	9.40e-03	4.75e-02
	HNSC	0.183	9.10e-03	1.73e-01
	SKCM	0.178	1.10e-02	6.66e-02
	ESCA	0.306	2.40e-02	9.07e-01
	LUAD	0.157	3.90e-02	1.75e-01
EMC9	KIRC	0.188	2.10e-02	5.57e-02
EMC10	KIRP	-0.435	4.90e-03	3.05e-02
	PAAD	-0.344	1.50e-03	3.06e-02
	LGG	0.235	2.00e-02	4.32e-02

Tabela suplementar 1 – Regressão de Cox utilizando dados de sobrevida dos pacientes do banco de dados TCGA para as integrantes do complexo EMC calculadas através do software OncoLnc.

	Gene	Correlação de Pearson	Gene	Correlação de Pearson	Gene	Correlação de Pearson	Gene	Correlação de Pearson
1	EBAG9	0.78	72	ZNF250	0.54	143	ZC3H3	0.39
2	DCAF13	0.77	73	GRINA	0.53	144	ZNF707	0.39
3	MTDH	0.74	74	ZNF34	0.53	145	VPS13B	0.39
4	NUDCD1	0.73	75	MRPL15	0.53	146	MCM4	0.38
5	DERL1	0.72	76	E2F5	0.52	147	MYBL1	0.38
6	ATP6V1C1	0.71	77	C8ORF59	0.52	148	RPL30	0.38
7	YWHAZ	0.7	78	DSCC1	0.52	149	TIMM17A	0.38
8	AZIN1	0.7	79	CHRA1	0.52	150	PAICS	0.38
9	STK3	0.69	80	OXR1	0.52	151	BOP1	0.38
10	SLC25A32	0.69	81	MED30	0.52	152	C8ORF44	0.38
11	MTERF3	0.69	82	MTBP	0.52	153	C8ORF37	0.38
12	RMDN1	0.69	83	TGS1	0.52	154	RBM12B	0.38
13	POLR2K	0.68	84	TPD52	0.51	155	FNTA	0.37
14	C8ORF76	0.68	85	UQCRB	0.51	156	ABCE1	0.37
15	EIF3E	0.67	86	ZNF696	0.51	157	SPAG1	0.37
16	PTK2	0.66	87	VCPIP1	0.51	158	VDAC1	0.37
17	ZNF706	0.66	88	TRAPPC9	0.51	159	LRRC14	0.37
18	UTP23	0.66	89	DPY19L4	0.51	160	TOMM70	0.37
19	ESRP1	0.66	90	YTHDF3	0.51	161	ZNF517	0.37
20	C8ORF33	0.66	91	VIRMA	0.51	162	SHARPIN	0.37
21	TERF1	0.65	92	TMEM65	0.51	163	SDAD1	0.37
22	ZNF7	0.65	93	TBC1D31	0.51	164	ARL6IP1	0.37
23	TMEM70	0.65	94	RHPN1-AS1	0.5	165	PYCR3	0.37
24	ARMC1	0.65	95	CSPP1	0.5	166	CHCHD4	0.36
25	NSMCE2	0.65	96	HGH1	0.5	167	CENPE	0.36
26	PTDSS1	0.64	97	EFR3A	0.5	168	UBE2K	0.36
27	ZFAND1	0.64	98	ATAD2	0.5	169	HSPA9	0.36
28	LACTB2	0.64	99	LYPLA1	0.49	170	PPAT	0.36
29	FAM91A1	0.64	100	ZHX1	0.49	171	EIF3J	0.36
30	RAD21	0.63	101	GRHL2	0.49	172	DGAT1	0.36
31	CPNE3	0.63	102	MRPS28	0.49	173	PSME3	0.36
32	INTS8	0.63	103	CYC1	0.48	174	ANAPC10	0.36
33	TRAM1	0.63	104	ZNF572	0.48	175	XPOT	0.36
34	ENY2	0.63	105	RRS1	0.48	176	PLEKHF2	0.36
35	TMEM68	0.62	106	HSF1	0.47	177	NAA15	0.36
36	ELOC	0.62	107	NDUFAF6	0.47	178	PAIP2	0.36
37	UBE2V2	0.62	108	UBXN2B	0.47	179	METAP1	0.36
38	MTRF1	0.62	109	UQCRBP1	0.47	180	JPH1	0.36
39	TATDN1	0.62	110	ZNF251	0.47	181	GNL3	0.36
40	OTUD6B	0.62	111	DECR1	0.46	182	EIF4E	0.35
41	TCEA1	0.61	112	FZD6	0.46	183	TONSL	0.35
42	WASHC5	0.61	113	NSMAF	0.46	184	RAD51	0.35
43	COPS5	0.61	114	WWP1	0.46	185	MROH1	0.35
44	IMPAD1	0.61	115	FAM83H	0.46	186	VDAC3	0.35
45	ZNF252P	0.61	116	MCMD2	0.45	187	PCMTD1	0.35
						214	EEF1E1	0.33
						215	NCOA2	0.33
						216	CHCHD7	0.33
						217	NIPAL2	0.33
						218	PIGW	0.33
						219	FBXL6	0.33
						220	NCAPG	0.33
						221	HSP90AB1	0.32
						222	MAD2L1	0.32
						223	TSNARE1	0.32
						224	TXNRD1	0.32
						225	STAM	0.32
						226	SRXN1	0.32
						227	ASAP1	0.32
						228	NAA20	0.32
						229	THAP1	0.32
						230	TMEM33	0.32
						231	BRIX1	0.32
						232	RBM12B-AS1	0.32
						233	LARP4B	0.32
						234	TMEM97	0.32
						235	LRP12	0.32
						236	CCNB1	0.31
						237	GLO1	0.31
						238	PPID	0.31
						239	PVT1	0.31
						240	SLBP	0.31
						241	JRK	0.31
						242	FAM72A	0.31
						243	GGH	0.31
						244	PSMD14	0.31
						245	CCT8	0.31
						246	TMED2	0.31
						247	GRPEL1	0.31
						248	TOP1MT	0.31
						249	UCHL5	0.31
						250	PDP1	0.31
						251	SPDL1	0.31
						252	MTPAP	0.31
						253	NEIL3	0.31
						254	HSPA4L	0.31
						255	WDR45B	0.31
						256	DPH6	0.31
						257	KNSTRN	0.31
						258	C16ORF87	0.31

46	RAB2A	0.6	117	PABPC1	0.45	188	SLC39A4	0.35	259	RACGAP1	0.31
47	TAF2	0.6	118	NDUFB9	0.44	189	CACYBP	0.35	260	FAM72B	0.31
48	UBR5	0.6	119	PABPC3	0.44	190	MLX	0.34	261	CCNA2	0.3
49	FAM49B	0.6	120	SAMD12	0.44	191	CCNE2	0.34	262	DLAT	0.3
50	WDYHV1	0.6	121	TMEM67	0.44	192	PTGES3	0.34	263	GRSF1	0.3
51	UBE2W	0.6	122	PRKDC	0.43	193	NRBP2	0.34	264	GRPEL2	0.3
52	MRPL13	0.6	123	ST3GAL1	0.42	194	SLC10A5	0.34	265	KPNB1	0.3
53	RB1CC1	0.59	124	POP1	0.42	195	LRRCC1	0.34	266	NEK2	0.3
54	RIDA	0.58	125	PUF60	0.42	196	PNPT1	0.34	267	SRSF1	0.3
55	ATP6V1H	0.58	126	FAM84B	0.42	197	LAPTM4B	0.34	268	UBE2E1	0.3
56	PEX2	0.57	127	COMMD5	0.42	198	PIP4P2	0.34	269	NOLC1	0.3
57	RAD54B	0.57	128	SNX16	0.42	199	SCRIB	0.34	270	RECQL4	0.3
58	CHMP4C	0.57	129	RPL7	0.41	200	GTPBP4	0.34	271	MTSS1	0.3
59	STAU2	0.57	130	SLC52A2	0.41	201	RNF19A	0.34	272	STIP1	0.3
60	OSGIN2	0.56	131	THEM6	0.41	202	CCDC43	0.34	273	RNF13	0.3
61	ZNF623	0.56	132	MAF1	0.41	203	AGO2	0.34	274	RNF170	0.3
62	ANKRD46	0.56	133	SPIDR	0.41	204	UBE2T	0.34	275	PRELID3B	0.3
63	NBN	0.55	134	TMED10P1	0.41	205	XKR9	0.34	276	ZC2HC1A	0.3
64	SQLE	0.55	135	TRIQQ	0.41	206	FAM216A	0.34	277	PPIL1	0.3
65	ZNF16	0.55	136	RIPK2	0.4	207	ZBTB10	0.33	278	ARL8B	0.3
66	ARFGEF1	0.55	137	MAL2	0.4	208	LIN54	0.33	279	MAPRE1	0.3
67	RNF139	0.55	138	ZFP41	0.4	209	CISD2	0.33	280	CISD1	0.3
68	RRM2B	0.55	139	ENOPH1	0.4	210	NARS	0.33	281	ZNF252P-AS1	0.3
69	TRMT12	0.55	140	EIF3H	0.39	211	PRDX1	0.33	282	PNO1	0.3
70	IMPA1	0.54	141	ARHGAP39	0.39	212	SMS	0.33	283	NKIRAS1	0.3
71	PHF20L1	0.54	142	CHD7	0.39	213	GPAA1	0.33			

Tabela suplementar 2 – Genes com alta correlação de Pearson com EMC2 (Corr. de Pearson ≥ 3.0) selecionados para a análise de predição de vias. Dados obtidos através do cBioPortal.

Figura suplementar 1 – Western blot da linhagem celular MCF7 mostrando a alta especificidade do anticorpo anti-EMC2 marcando uma única banda (**D**) na altura aproximada de 35kDa. Em **A** e **C** encontra-se o marcador de peso molecular e em **B** a raia do lisado proteico corado com Ponceau (solução descrita em Materiais e Métodos).

Figura Suplementar 2- Colocalização da marcação de EMC2 através do anticorpo anti-EMC2 (verde) e da proteína EMC2 fusionada a mCherry (vermelha) mostrando que o anticorpo reconhece a proteína induzida pelo plasmídeo transfectado mCherry-EMC2.

					Início	Final
NM_001329493.1	Transcrito 1	3587 bp	NP_001316422.1	306 aa	MAKVS	QITQS
NM_014673.4	Transcrito 2	3560 bp	NP_055488.1	297 aa	MAKVS	QITQS
NM_001329494.1	Transcrito 3	895 bp	NP_001316423.1	212 aa	MAKVS	SIYIM
NM_001329495.1	Transcrito 4	3700 bp	NP_001316424.1	298 aa	MKYSS	QITQS
NR_138033.1	Transcrito 5	3502 bp	*	*	*	*

Tabela Suplementar 3- Informações sobre os transcritos de EMC2 (rosa) e as variantes da proteína EMC2 (azul). Os dados foram obtidos pelo site <https://www.ncbi.nlm.nih.gov/>. * representa ausência de informação uma vez que se trata de um transcrito não codificante. bp: pares de base. aa: aminoácidos.

Figura suplementar 3- Esquema que mostra a construção do plasmídeo mCherry-EMC2. Iniciou-se com o plasmídeo comercial mCherry-Parkin (<https://www.addgene.org/23956/>), e após a digestão dupla utilizando as enzimas de restrição BglII e HindIII foi feita a inserção da “Sequência inserto-EMC2”.

Sequência inserto-EMC2:

AGCTTATGCGAAGGTCTCAGAGCTTTACGATGTCACCTGGGAAGAAATGAGAGA
TAAATGAGAAAATGGAGAGAAGAAACTCAAGAAATAGTGAGCAAATTGTGGA
AGTTGGAGAAGAATTAATTAATGAATATGCTTCTAAGCTGGGAGATGATATTTGG

ATCATATATGAACAGGTGATGATTGCAGCACTAGACTATGGTCGGGATGACTTG
GCATTGTTTTGTCTTCAAGAGCTGAGAAGACAGTTCCTGGCAGTCACAGAGTC
AAGCGATTAACAGGCATGAGATTTGAAGCCATGGAAAGATATGATGATGCTATA
CAGCTATATGATAGGATTTTACAAGAAGATCCAACCTAACACTGCTGCAAGAAAGC
GTAAGATTGCCATTCGAAAAGCCCAGGGGAAAAATGTGGAGGCCATTCGGGAG
CTGAATGAGTATCTGGAACAATTTGTTGGAGACCAAGAAGCCTGGCATGAACTT
GCAGAACTTTACATCAATGAACATGACTATGCAAAAGCAGCCTTTTGTITAGAGG
AACTAATGATGACTAATCCACACAACCACTTATACTGTCAGCAGTATGCTGAAGT
TAAGTATACCCAAGGTGGACTTGAAAACCTCGAACTTTCAAGAAAGTATTTTGCA
CAGGCATTGAACTGAACAACAGAAATATGAGAGCTTTGTTTGGACTTTATATGT
CGGCAAGTCATATTGCTTCTAATCCAAAAGCAAGTGCAAAAACGAAAAAGGACA
ACATGAAATATGCTAGTTGGGCAGCTAGTCAAATAAACAGAGCTTATCAGTTTGC
AGGTCGAAGTAAGAAGGAAACCAATATTCTCTTAAGGCTGTCGAAGACATGTT
GGAAACATTGCAGATCACCCAGTCTTAAGGTTTCAAAAACCTTTGACATTAGAT
TTCACAACCTGCACAATTGAACTTATTGGCCTGTAACCTATTTACTAAATGCTCAGT
GCTATTTATATACTACAGTAATTTTCTGTTAAGAAGGCAGTTGTAAAGAATGTGTT
TATATAAACCTAAAAATGCCTTTTACTGCTAAGTGGGGAGATGGGGGAAATCCAT
GGAAGAGAGATTTAAGACTTATTGATTGTACATCAGTCTCTTCATATCACATATAC
ATGTATATATATAAACTCTAATGTAGTATAACCTTGTTAAATAAACCATGATGAT
TTACG

Legenda:

Letra vermelha: bases adicionadas para colocar sequência codificante de EMC2 em fase.

Cor de realce amarelo: Sequência do primer utilizado para amplificar o gene EMC2 a partir do *pull* de cDNA.

Sublinhado: Sequência do transcrito de EMC2 (NM_014673.4)

Em negrito: Sequência dos sítios das enzimas de restrição

AAGCTT - HindIII
GGATCC - BamHI

Sequência do plasmídeo mCherry-EMC2:

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAG
TTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGAC
CCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG
ACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCA
GTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGT
AAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACT
TGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGC
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCC
ACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAATCAACGGGACTTTC
CAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTA
CGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTA
GCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCA
TCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGC
CACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCC
AGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGA
CATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCG
CCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAG
CGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCT
CCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTC
CCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTC
CGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGG
CTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAA
GGCCAAGAAGCCCGTGACAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGG
ACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCC
GAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTCCGGACTCA
GATCTCGAGCTCAAGCTTATGCGAAGGTCTCAGAGCTTACGATGTCACTTGGG
AAGAAATGAGAGATAAAATGAGAAAATGGAGAGAAGAAAATCAAGAAATAGTG
AGCAAATTGTGGAAGTTGGAGAAGAATTAATTAATGAATATGCTTCTAAGCTGGG
AGATGATATTTGGATCATATATGAACAGGTGATGATTGCAGCACTAGACTATGGT
CGGGATGACTTGGCATTGTTTTGTCTTCAAGAGCTGAGAAGACAGTTCCCTGGC
AGTCACAGAGTCAAGCGATTAACAGGCATGAGATTTGAAGCCATGGAAAGATAT
GATGATGCTATACAGCTATATGATAGGATTTTACAAGAAGATCCAACCTAAGT

CTGCAAGAAAGCGTAAGATTGCCATTCGAAAAGCCCAGGGGAAAAATGTGGAG
GCCATTCGGGAGCTGAATGAGTATCTGGAACAATTTGTTGGAGACCAAGAAGCC
TGGCATGAACTTGCAGAACTTTACATCAATGAACATGACTATGCAAAAGCAGCCT
TTTGTTTAGAGGAACTAATGATGACTAATCCACACAACCACTTATACTGTCAGCA
GTATGCTGAAGTTAAGTATACCCAAGGTGGACTTGAAAACCTCGAACTTTCAAGA
AAGTATTTTGCACAGGCATTGAACTGAACAACAGAAATATGAGAGCTTTGTTTG
GACTTTATATGTCGGCAAGTCATATTGCTTCTAATCCAAAAGCAAGTGCAAAAAC
GAAAAAGGACAACATGAAATATGCTAGTTGGGCAGCTAGTCAAATAAACAGAGC
TTATCAGTTTGCAGGTCTGAAGTAAGAAGGAAACCAATATTCTCTTAAGGCTGTC
GAAGACATGTTGGAAACATTGCAGATCACCCAGTCTTAAGGTTTCAAAAACCTTT
TGACATTAGATTTCACAACTGCACAATTGAACTTATTGGCCTGTAACCTTATTTACT
AAATGCTCAGTGCTATTTATATACTACAGTAATTTTCTGTTAAGAAGGCAGTTGTA
AAGAATGTGTTTATATAAACCTAAAAATGCCTTTTACTGCTAAGTGGGGAGATGG
GGGAAATCCATGGAAGAGAGATTTAAGACTTATTGATTGTACATCAGTCTCTTCA
TATCACATATACATGTATATATATAAACTCTAATGTAGTATAACCTTGTTAAATAA
ACCATGATGATTTACGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATA
CCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAA
CCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATA
ATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCA
CTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTAACGCGTAAATT
GTAAGCGTTAATATTTTGTAAATTCGCGTTAAATTTTGTAAATCAGCTCATTT
TTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAGAATAGACCG
AGATAGGGTTGAGTGTTGTTCCAGTTTGAACAAGAGTCCACTATTAAAGAACG
TGGA CTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTA
CGTGAACCATCACCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTA
AATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGC
GAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCG
CTGGCAAGTGTAGCGGTACGCTGCGCGTAACCACCACACCCGCCGCGCTTAA
TGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
ACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACA
ATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAA
AGAACCAGCTGTGGAATGTGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCC
CCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGT

GGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAAGTATGCAAAGCATGCATCTCAA
TTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCATCCCGCCCCCTAACTCC
GCCCAGTTCCGCCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCA
GAGGCCGAGGCCGCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTT
TTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGT
TTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTG
GAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGC
CGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCG
ACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGG
CTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGC
GGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCAT
CTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGG
CTGCATACGCTTGATCCGGCTACCTGCCCATTGACCACCAAGCGAAACATCGC
ATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCT
GGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCCGCCAGGCTCAAG
GCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTT
GCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCG
GCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTG
CTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATC
GCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTC
TGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCA
TCACGAGATTTGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATC
GTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGA
GTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATAC
CGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGT
GTTGGGTCGTTTGTTTCATAAACGCGGGGTTCCGTCCCAGGGGCTGGCACTCTGT
CGATACCCACCGAGACCCCATTTGGGGCCAATACGCCCGCGTTTCTTCCTTTTC
CCCACCCACCCCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACGTC
GGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGAT
TAAAACCTTCATTTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTTGATAATCT
CATGACCAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGT
AGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGC
TTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGGCGGATCAAGAG

CTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAAT
ACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA
CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCG
CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAA
CGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACG
CTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAA
CAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGT
CCTGTCGGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCA
GGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT
GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT
GTGGATAACCGTATTACCGCCATGCAT

Legenda:

Letra vermelha: bases adicionadas para colocar sequência codificante de EMC2 em fase.

Cor de realce amarelo: Sequência do primer utilizado para amplificar o gene EMC2 a partir do *pull* de cDNA

Cor de realce vermelho: Sequência da mCherry

Sublinhado: Sequência do transcrito de EMC2 (NM_014673.4)

Em negrito: Sequência dos sítios das enzimas de restrição

AAGCTT	-	HindIII
GGATCC	-	BamHI

			Fim mCherry		HindIII	→	EMC2		
Esperada	1		ACAAGTCCGGACTCAGATCTCGAGCTC		AAGCTT		TGCGAAGGTCTCAGAGCTTTACGATG	60	
Sequenciada	30		ACAAGTCCGGACTCAGATCTCGAGCTC		AAGCTT		TGCGAAGGTCTCAGAGCTTTACGATG	89	
Esperada	61		TCACCTGGGAAGAAATGAGAGATAAAATGAGAAAATGGAGAGAAGAAAACCTCAAGAAATA					120	
Sequenciada	90		TCACCTGGGAAGAAATGAGAGATAAAATGAGAAAATGGAGAGAAGAAAACCTCAAGAAATA					149	
Esperada	121		GTGAGCAAATTGTGGAAGTTGGAGAAGAATTAATTAATGAATATGCTTCTAAGCTGGGAG					180	
Sequenciada	150		GTGAGCAAATTGTGGAAGTTGGAGAAGAATTAATTAATGAATATGCTTCTAAGCTGGGAG					209	
Esperada	181		ATGATATTTGGATCATATATGAACAGGTGATGATTGCAGCACTAGACTATGGTCGGGATG					240	
Sequenciada	210		ATGATATTTGGATCATATATGAACAGGTGATGATTGCAGCACTAGACTATGGTCGGGATG					269	
Esperada	241		ACTTGGCATTGTTTTGTCTTCAAGAGCTGAGAAGACAGTTCCTGGCAGTCACAGAGTCA					300	
Sequenciada	270		ACTTGGCATTGTTTTGTCTTCAAGAGCTGAGAAGACAGTTCCTGGCAGTCACAGAGTCA					329	
Esperada	301		AGCGATTAAACAGGCATGAGATTTGAAGCCATGGAAAGATATGATGATGCTATACAGCTAT					360	
Sequenciada	330		AGCGATTAAACAGGCATGAGATTTGAAGCCATGGAAAGATATGATGATGCTATACAGCTAT					389	
Esperada	361		ATGATAGGATTTTACAAGAAGATCCAACCTAACACTGCTGCAAGAAAGCGTAAGATTGCCA					420	
Sequenciada	390		ATGATAGGATTTTACAAGAAGATCCAACCTAACACTGCTGCAAGAAAGCGTAAGATTGCCA					449	
Esperada	421		TTTCAAAAAGCCAGGGGAAAAATGTGGAGGCCATTGCGGAGCTGAATGAGTATCTGGAAC					480	
Sequenciada	450		TTTCAAAAAGCCAGGGGAAAAATGTGGAGGCCATTGCGGAGCTGAATGAGTATCTGGAAC					509	
Esperada	481		AATTTGTTGGAGACCAAGAAGCCTGGCATGAACTTGCAAGAACTTTACATCAATGAACATG					540	
Sequenciada	510		AATTTGTTGGAGACCAAGAAGCCTGGCATGAACTTGCAAGAACTTTACATCAATGAACATG					569	
Esperada	541		ACTATGCAAAAAGCAGCCTTTTGTGTTAGAGGAACTAATGATGACTAATCCACACAACCCT					600	
Sequenciada	570		ACTATGCAAAAAGCAGCCTTTTGTGTTAGAGGAACTAATGATGACTAATCCACACAACCCT					629	
Esperada	601		TATACTGTGAGCAGTATGCTGAAGTTAAGTATACCCAAGGTGGACTTGAAAACTCGAAC					660	
Sequenciada	630		TATACTGTGAGCAGTATGCTGAAGTTAAGTATACCCAAGGTGGACTTGAAAACTCGAAC					689	
Esperada	661		TTTCAAGAAAGTATTTTGCACAGGCATTGAAACTGAACAACAGAAATATGAGAGCTTTGT					720	
Sequenciada	690		TTTCAAG - AAGTATTTTGCACAGGCATTGAAACTGAACAACAGAAATATGAGAGCTTTGT					748	

Figura Suplementar 4 – Alinhamento entre a sequência esperada da inserção e a sequenciada utilizando o Primer Foward para o sequenciamento do plasmídeo. Nota-se que do início até a base 660 da sequência esperada há similaridade em 100% comparado a parte sequenciada.

Esperada	546	GCAAAAGCAGCCTTTTGTAGAGGAACTAATGATGACTAATCCACACAACCCTTATAC	605
Sequenciada	301	GCAAAAGCAGCCTTTTGTAGAGGAACTAATGATGACTAATCCACACAACCCTTATAC	360
Esperada	606	TGTCAGCAGTATGCTGAAGTTAAGTATACCCAAGGTGGACTTGAAAACCTCGAACTTTCA	665
Sequenciada	361	TGTCAGCAGTATGCTGAAGTTAAGTATACCCAAGGTGGACTTGAAAACCTCGAACTTTCA	420
Esperada	666	AGAAAGTATTTTGCACAGGCATTGAACTGAACAACAGAAATATGAGAGCTTTGTTTGGGA	725
Sequenciada	421	AGAAAGTATTTTGCACAGGCATTGAACTGAACAACAGAAATATGAGAGCTTTGTTTGGGA	480
Esperada	726	CTTTATATGTCGGCAAGTCATATTGCTTCTAATCCAAAAGCAAGTGCAAAAACGAAAAAG	785
Sequenciada	481	CTTTATATGTCGGCAAGTCATATTGCTTCTAATCCAAAAGCAAGTGCAAAAACGAAAAAG	540
Esperada	786	GACAACATGAAATATGCTAGTTGGGCAGCTAGTCAAATAAACAGAGCTTATCAGTTTGCA	845
Sequenciada	541	GACAACATGAAATATGCTAGTTGGGCAGCTAGTCAAATAAACAGAGCTTATCAGTTTGCA	600
Esperada	846	GGTCGAAGTAAGAAGGAAACCAAATATTCTCTTAAGGCTGTCGAAGACATGTTGGAAACA	905
Sequenciada	601	GGTCGAAGTAAGAAGGAAACCAAATATTCTCTTAAGGCTGTCGAAGACATGTTGGAAACA	660
Esperada	906	TTGCAGATCACCCAGTCTTAAGGTTTCAAAAACCTTTGACATTAGATTTCACTGCA	965
Sequenciada	661	TTGCAGATCACCCAGTCTTAAGGTTTCAAAAACCTTTGACATTAGATTTCACTGCA	720
Esperada	966	CAATTGAACCTATTGGCCTGTAACCTATTTACTAAATGCTCAGTGCTATTTATATACTAC	1025
Sequenciada	721	CAATTGAACCTATTGGCCTGTAACCTATTTACTAAATGCTCAGTGCTATTTATATACTAC	780
Esperada	1026	AGTAATTTCTGTTAAGAAGGCAGTTGTAAGAATGTGTTTATATAAACCTAAAAATGCC	1085
Sequenciada	781	AGTAATTTCTGTTAAGAAGGCAGTTGTAAGAATGTGTTTATATAAACCTAAAAATGCC	840
Esperada	1086	TTTTACTGCTAAGTGGGGAGATGGGGGAAATCCATGGAAGAGAGATTTAAGACTTATTGA	1145
Sequenciada	841	TTTTACTGCTAAGTGGGGAGATGGGGGAAATCCATGGAAGAGAGATTTAAGACTTATTGA	900
Esperada	1146	TTGTACATCAGTCTTTCATATCACATATACATGTATATATATAAACTCTAATGTAGTA	1205
Sequenciada	901	TTGTACATCAGTCTTTCATATCACATATACATGTATATATATAAACTCTAATGTAGTA	960
Esperada	1206	TAACCTTGTTAAATAAACCATGATGATTTAAGGATCCA	1243
Sequenciada	961	TAACCTTGTTAAATAAACCATGATGATTTAAGGATCCA	998

EMC2 ← BamHI

Figura Suplementar 5 – Alinhamento entre a sequência esperada da inserção e a sequenciada utilizando o Primer Reverso para o sequenciamento do plasmídeo. Para fazermos a comparação, alteramos a sequência proveniente do sequenciamento para a respectiva sequência reversa complementar, utilizando o site https://www.bioinformatics.org/sms/rev_comp.html. Nota-se que a partir da base 546 da sequência esperada há similaridade em 100% comparado a parte sequenciada.

Sequenciamento clonagem PGEM EMC2 – Prime Forward

GGCCCATCACGGCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAG
CTCAAGCTTATGCGAAGGTCTCAGAGCTTTACGATGTCACTTGGGAAGAAATGA
GAGATAAAATGAGAAAATGGAGAGAAGAAAACCTCAAGAAATAGTGAGCAAATTG
TGGAAGTTGGAGAAGAATTAATTAATGAATATGCTTCTAAGCTGGGAGATGATAT
TTGGATCATATATGAACAGGTGATGATTGCAGCACTAGACTATGGTCGGGATGA
CTTGGCATTGTTTTGTCTTCAAGAGCTGAGAAGACAGTTCCCTGGCAGTCACAG
AGTCAAGCGATTAAACAGGCATGAGATTTGAAGCCATGGAAAGATATGATGATGC
TATACAGCTATATGATAGGATTTTACAAGAAGATCCAACTAACACTGCTGCAAGA

AAGCGTAAGATTGCCATTGCGAAAAGCCCAGGGGAAAAATGTGGAGGCCATTGCG
GGAGCTGAATGAGTATCTGGAACAATTTGTTGGAGACCAAGAAGCCTGGCATGA
ACTTGCAGAACTTTACATCAATGAACATGACTATGCAAAAGCAGCCTTTTGTTTA
GAGGAACTAATGATGACTAATCCACACAACCACTTATACTGTCAGCAGTATGCTG
AAGTTAAGTATACCCAAGGTGGACTTGAAAACCTCGAACTTTCAAGAAGTATTTT
GCACAAGCATTGAACTGAACAACAGAAATATGAGAGCTTTGTTTGGACTTTATA
TGTCGGCAAGTCATATTGCTTCTAATCCAAAAGCAAGTGCAAAAACGAAAAAGG
ACAACATGAAATATGCTAGTTGGGCAGCTAGTCAAATAAACAGAGCTTATCAGTT
TGCAAGGTCGAAGTAAGAAGGAAACCAAATATTCTCTTAAAGGCTGTCGAAGAC
ATGTTGGGAAAACATTGCAGAATCACCCCAGGTCTTTAAAGTTTCAAAAACCTTCT
TTTGAACATTTAGATTTACAAACTGCACAAATTTGAACTACTGCTGTAACCTTAT
TACTTAATGCTCAGTGCTATTTATATACTACAGTAAATTCTGATTAAGAAAGTAA
GGTGTTTCAGAC

Sequenciamento clonagem PGEM EMC2 – Prime Reverse

TGGCTGATGGCTCGAGATGTATCGACTCACTATAGGGCGAATTGGGCCCCGACG
TCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTAAGCTTATGCG
AAGGTCTCAAAGCTTTACGATGTCACCTTGGGAAGAAATGAGAGATAAAATGAGA
AAATGGAGAGAAGAAAACCTCAAGAAATAGTGAGCAAATTGTGGAAGTTGGAGAA
GAATTAATTAATGAATATGCTTCTAAGCTGGGAGATGATATTTGGATCATATATGA
ACAGGTGATGATTGCAGCACTAGACTATGGTCGGGATGACTTGGCATTGTTTTG
TCTTCAAGAGCTGAAAAACAGTTCCCTGGCAGTCACAGAGTCAAGCGATTAAC
AGGCATGAGATTTGAAGCCATGGAAAGATATGATGATGCTATACAGCTATATGAT
AGGATTTTACAAGAAGATCCAATAACACTGCTGCAAGAAAGCGTAAGATTGCC
ATTCGAAAAGCCCAGGGGAAAAATGTGGAGGCCATTCTGGGAGCTGAATGAGTA
TCTGGAACAATTTGTTGGAGACCAAGAAGCCTGGCATGAACTTGCAAACTTTA
CATCAATGAACATGACTATGCAAAAGCAGCCTTTTGTTTAGAGGAACTAATGATG
ACTAATCCACACAACCACTTATACTGTCAGCAGTATGCTGAAGTTAAGTATACCC
AAGGTGGACTTGAAAACCTCGAACTTTCAAGAAAGTATTTTGCACAAGCATTGAA
ACTGAACAACAGAAATATGAGAGCTTTGTTTGGACTTTATATGTCGGCAAGTCAT
ATTGCTTCTAATCCAAAAGCAAGTGCAAAAACGAAAAAGGGACAACATGAAATAT
GCTAGTTGGGCAGCTAGTCAAATAAAACAAGAGCTTATCAGTTTGCAGGTCGAA
GTAAGAGGAAACCAAATATTCTTCTTGGGCTGTCGAAGACATGTGAAACATGCA

GATCACCCAGTCTAGTTCAAAACTCTTGGACATAGATTCACACCTGCACATGAAC
CTATGGCTGTACTATACTAATGCTCATGCTATATACTACGTAGTACTGTAGAGCA
GTGTAAGAAAG

Sequenciamento clonagem MCherry EMC2 – Primer Foward

GGCCCATCACGGCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAG
CTCAAGCTTATGCGAAGGTCTCAGAGCTTTACGATGTCACTTGGGAAGAAATGA
GAGATAAAATGAGAAAATGGAGAGAAGAAAACCTCAAGAAATAGTGAGCAAATTG
TGGAAGTTGGAGAAGAATTAATTAATGAATATGCTTCTAAGCTGGGAGATGATAT
TTGGATCATATATGAACAGGTGATGATTGCAGCACTAGACTATGGTCGGGATGA
CTTGGCATTGTTTTGTCTTCAAGAGCTGAGAAGACAGTTCCCTGGCAGTCACAG
AGTCAAGCGATTAACAGGCATGAGATTTGAAGCCATGGAAAGATATGATGATGC
TATACAGCTATATGATAGGATTTTACAAGAAGATCCAACTAACACTGCTGCAAGA
AAGCGTAAGATTGCCATTGCGAAAAGCCCAGGGGAAAAATGTGGAGGCCATTGCG
GGAGCTGAATGAGTATCTGGAACAATTTGTTGGAGACCAAGAAGCCTGGCATGA
ACTTGCAGAACTTTACATCAATGAACATGACTATGCAAAAGCAGCCTTTTTGTTTA
GAGGAACTAATGATGACTAATCCACACAACCACTTATACTGTCAGCAGTATGCTG
AAGTTAAGTATACCCAAGGTGGACTTGAAAACCTCGAACTTTCAAGAAGTATTTT
GCACAAGCATTGAACTGAACAACAGAAATATGAGAGCTTTGTTTGGACTTTATA
TGTCGGCAAGTCATATTGCTTCTAATCCAAAAGCAAGTGCAAAAACGAAAAAGG
ACAACATGAAATATGCTAGTTGGGCAGCTAGTCAAATAAACAGAGCTTATCAGTT
TGCAAGGTCGAAGTAAGAAGGAAACCAAATATTCTCTTAAAGGCTGTCGAAGAC
ATGTTGGGAAAACATTGCAGAATCACCCCAGGTCTTTAAAGTTTCAAAAACCTTCT
TTTGAACATTTAGATTTTCAAACTGCACAAATTTGAACTACTGCTGTAACCTTAT
TACTTAATGCTCAGTGCTATTTATATACTACAGTAAATTCTGATTAAGAAAGTAA
GGTGTTTCAGAC

Sequenciamento clonagem MCherry EMC2 – Primer Reverse

CATAAATGATTACGCCAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCA
TCCAACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATT
CACTAGTGATTGGATCCGTAAATCATCATGGTTTATTTAACAAGGTTATACTACAT
TAGAGTTTTATATATACATGTATATGTGATATGAAGAGACTGATGTACAATCAA
TAAGTCTTAAATCTCTCTTCCATGGATTTCCCCCATCTCCCCACTTAGCAGTAAA
AGGCATTTTTAGGTTTATATAAACACATTCTTTACAACCTGCCTTCTTAACAGAAAA

TTACTGTAGTATATAAATAGCACTGAGCATTTAGTAAATAAGTTACAGGCCAATAA
GTTCAATTGTGCAGTTGTGAAATCTAATGTCAAAGAGTTTTTGAAACCTTAAGAC
TGGGTGATCTGCAATGTTTCCAACATGTCTTCGACAGCCTTAAGAGAATATTTGG
TTTCCTTCTTACTTCGACCTGCAAAGCTGATAAGCTCTGTTTATTTGACTAGCTGC
CCAAGTAGCATATTTTCATGTTGTCCTTTTTTCGTTTTTGCACTTGCTTTTGGATTAG
AAGCAATATGACTTGCCGACATATAAAGTCCAAACAAAGCTCTCATATTTCTGTT
GTTCAAGTTTCAATGCTTGTGCAAAATACTTTCTTGAAAGTTTCGAGGTTTTCAAGT
CCACCTTGGGTATACTTAACTTCAGCATACTGCTGACAGTATAAGTGGTTGTGTG
GATTAGTCATCATTAGTTCCTCTAAACAAAAGGCTGCTTTTGCATAGTCATGTTT
ATTGATGTAAAGTTCTGCAAGTTCATGCCAGGCTTCTTGGTCTCCAACAAATTGT
TCCAGATACTCATTAGCTCCCGAATGGCCCTCCACATTTTTTTCCCCTGGGCTTT
TCGAATGGCAAATCTTACGCTTTCTTGACAGCAGTGTTTAGTTTGGATTCTTTCTT
GTTAAAATCCTATCATATAGCTGTATAGCATCAATCATATCTTCCATGGGCTTTCA
ATCTCAATGCTATATCGCTTGAAGTCTGTGAAGTGGCAGGAAGTATCTTCCTCA
GCTTCTTTGGAG

Sequenciamento clonagem MCherry Controle – Primer Foward

TAAAAGAAGGCGCGAGATGTATCGACTCACTATAGGGCGAATTGGGCCCCGACG
TCGCATGCTCCCGGGCCGCATGGCGGGCGCGGAATTCGATTAAGCTTATGCG
AAGGTCTCAAAGCTTTACGATGTCACTTGGGAAGAAATGAGAGATAAAATGAGA
AAATGGAGAGAAGAAAAGTCAAGAAATAGTGAGCAAATTGTGGAAGTTGGAGAA
AAATTAATTAATGAATATGCTTCTAAGCTGGGAGATGATATTTGGATCATATATGA
ACAGGTGATGATTGCAGCACTAGACTATGGTCGGGATGACTTGGCATTGTTTTG
TCTTCAAGAGCTGAAAAACAGTTCCTGGCAGTCACAGAGTCAAGCGATTAAC
AGGCATGAGATTTGAAGCCATGGAAAGATATGATGATGCTATACAGCTATATGAT
AGGATTTTACAAGAAGATCCAAGTCAAGTCTGCAAGAAAGCGTAAGATTGCC
ATTCGAAAAGCCCAGGGGAAAAATGTGGAGGCCATTCTGGGAGCTGAATGAGTA
TCTGGAACAATTTGTTGGAGACCAAGAAGCCTGGCATGAACTTGCAAAACTTTA
CATCAATGAACATGACTATGCAAAAGCAGCCTTTTGTGTTAGAGGAACTAATGATG
ACTAATCCACACAACCACTTATACTGTCAGCAGTATGCTGAAGTTAAGTATACCC
AAGGTGGACTTGAAAACCTCGAACTTTCAAGAAAGTATTTTGCACAAGCATTGAA
ACTGAACAACAGAAATATGAGAGCTTTGTTTGGACTTTATATGTCGGCAAGTCAT
ATTGCTTCTAATCCAAAAGCAAGTGCAAAAACGAAAAAGGACAACATGAAATATG

CTAGTTTGGGCAGCTAGTCAAATAAACAGAGCTTATCAGTTTGCAGGTCTGAAGT
AAGAAGGAAACCAAATATTCTCTTCGCTGTCTGAAGACATGTGAAACATTGCAGAT
CACTCAGTCTAGTTCCAAAACCTCTTTGGACTAGATTCACACTGCACATGAACCTA
TGCTGTACTATTACTAAATGCTCAATGCATTATACTACAGTATTTTCGTAGAGTCAG
TGTTGCG

Sequenciamento clonagem MCherry Controle – Primer Reverse

TGTTATATGATCAGTTATCTAGATCCGGTGGATCTGAGTCCGGACTTGTACAGCT
CGTCCATGCCGCCGGTGGAGTGGCGGCCCTCGGCGCGTTTCGTACTGTTCCAC
GATGGTGTAGTCCTCGTTGTGGGAGGTGATGTCCAACCTTGATGTTGACGTTGTA
GGCGCCGGGCAGCTGCACGGGCTTCTTGGCCTTGTAGGTGGTCTTGACCTCAG
CGTCGTAGTGGCCGCCGTCTTCAGCTTCAGCCTCTGCTTGATCTCGCCCTTCA
GGGCGCCGTCTCGGGGTACATCCGCTCGGAGGAGGCCTCCCAGCCCATGGT
CTTCTTCTGCATTACGGGGCCGTCTCGGAGGGGAAGTTGGTGCCGCGCAGCTTCA
CCTTGTAGATGAACTCGCCGTCTTGCAGGGAGGAGTCCTGGGTCACGGTCACC
ACGCCGCCGTCTCGAAGTTCATCACGCGCTCCCACTTGAAGCCCTCGGGGAA
GGACAGCTTCAAGTAGTCGGGGATGTCTGGCGGGGTGCTTCACGTAGGCCTTGG
AGCCGTACATGAACTGAGGGGACAGGATGTCCCAGGCGAAGGGCAGGGGGCC
ACCCTTGGTCACCTTCAGCTTGGCGGTCTGGGTGCCCTCGTAGGGGCGGGCCCT
CGCCCTCGCCCTCGATCTCGAACTCGTGGCCGTTACGGAGCCCTCCATGTGC
ACCTTGAAGCGCATGAACTCCTTGATGATGGCCATGTTATCCTCCTCGCCCTTG
CTCACCATGGTGGCGACCGGTAGCGCTAGCGGATCTGACGGTTCACTAAACCA
GCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCCATTTGCGTCGAT
GGGGGCGGAGTTGTTACGACATTTTTGGAAAGTCCCGTTTGATTTTGGGTGCCA
AAACATACTCCTCATTGACGTCGATGGGGGTGGAGTACTTTGGTAAATCCCCCG
ATGAGTCAAGCTGCTATCCACGCCTCAATTGGATGGTACCTGGCTAAAACCGGC
ATCATCGATGTAGTAGCGATGACTATACGTAGATGTACCTGACCAAG