



**CONFORME SOLICITAÇÃO DO AUTOR, ESTA
PRODUÇÃO INTELECTUAL POSSUI
RESTRIÇÃO DE ACESSO**

**CAXIAS DO SUL
2018**

UNIVERSIDADE DE CAXIAS DO SUL
ÁREA DO CONHECIMENTO DE CIÊNCIAS DA
VIDA
INSTITUTO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

**ANÁLISE DE EXPRESSÃO DIFERENCIAL DE GENES: UMA
SOLUÇÃO COMPUTACIONAL PARA IDENTIFICAÇÃO DE
BIOMARCADORES DE TUMORES GÁSTRICOS EM HUMANOS**

MARCOS VINÍCIUS ROSSETTO

CAXIAS DO SUL

2019

MARCOS VINÍCIUS ROSSETTO

**ANÁLISE DE EXPRESSÃO DIFERENCIAL DE GENES: UMA
SOLUÇÃO COMPUTACIONAL PARA IDENTIFICAÇÃO DE
BIOMARCADORES DE TUMORES GÁSTRICOS EM HUMANOS**

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia da Universidade de Caxias do Sul, visando a obtenção de grau de Mestre em Biotecnologia
Orientadora: Prof^a. Dr^a Scheila de Avila e Silva.
Co-orientadora: Dr^a. Ivaine Taís Sauthier Sartor.

CAXIAS DO SUL

2019

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade de Caxias do Sul
Sistema de Bibliotecas UCS - Processamento Técnico

R829a Rossetto, Marcos Vinícius

Análise de expressão diferencial de genes [recurso eletrônico] : uma solução computacional para identificação de biomarcadores de tumores gástricos em humanos / Marcos Vinícius Rossetto. – 2019.

Dados eletrônicos.

Dissertação (Mestrado) - Universidade de Caxias do Sul, Programa de Pós-Graduação em Biotecnologia, 2019.

Orientação: Scheila de Avila e Silva.

Coorientação: Ivaine Taís Sauthier Sartor.

Modo de acesso: World Wide Web

Disponível em: <https://repositorio.ucs.br>

1. Neoplasias gástricas. 2. Estômago - Tumores. 3. Marcadores bioquímicos. 4. Biotecnologia. I. Silva, Scheila de Avila e, orient. II. Sartor, Ivaine Taís Sauthier, coorient. III. Título.

CDU 2. ed.: 616.33-006

Catalogação na fonte elaborada pela(o) bibliotecária(o)
Márcia Servi Gonçalves - CRB 10/1500

MARCOS VINICIUS ROSSETTO

ANÁLISE DE EXPRESSÃO DIFERENCIAL DE GENES: UMA SOLUÇÃO COMPUTACIONAL PARA IDENTIFICAÇÃO DE BIOMARCADORES DE TUMORES GÁSTRICOS EM HUMANOS

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia da Universidade de Caxias do Sul, visando à obtenção do título de Mestra em Biotecnologia.

Orientador: Prof^a. Dr^a. Scheila de Avila e Silva.

Co-orientadora: Dr^a. Ivaine Taís Sauthier Sartor

DISSERTAÇÃO APROVADA EM 17 DE DEZEMBRO DE 2019.

Orientadora: Prof^a. Dr^a. Scheila de Avila e Silva

Co-orientadora: Dr^a. Ivaine Taís Sauthier Sartor

Prof^a. Dr^a. Mariana Roesch Ely

Prof. Dr. Clodis Boscaroli

Prof. Dr. Rafael Andrade Caceres

ÍNDICE

1 INTRODUÇÃO	7
1.1 OBJETIVOS	11
1.1.1 Objetivo geral	11
1.1.2 Objetivos específicos.....	11
1.2 ESTRUTURA DO TRABALHO.....	11
2 FUNDAMENTAÇÃO TEÓRICA	13
2.1 NEOPLASIA MALIGNA	13
2.2 BANCO DE DADOS DE EXPRESSÃO GÊNICA	24
2.3 TRABALHOS RELACIONADOS	26
3 METODOLOGIA	30
3.1 DADOS BIOLÓGICOS	31
3.2 AQUISIÇÃO E PROCESSAMENTO DOS DADOS	31
3.3 ANÁLISE DA EXPRESSÃO DIFERENCIAL.....	32
3.4 DESENVOLVIMENTO DO C-GEMIS	33
4 RESULTADOS	34
4.1 CAPÍTULO I - <i>C-Gemis: A computational tool for gastric cancer data analysis</i> ...	35
4.2 CAPÍTULO II - Busca de potenciais biomarcadores em câncer gástrico utilizando ferramentas in silico	40
5 CONSIDERAÇÕES FINAIS	67

LISTA DE QUADROS

Tabela 1 - Descrição das características da Classificação de Laurén	16
Tabela 2 - Correlação entre as classificações de WHO e Laurén	18
Tabela 3 - Bibliotecas utilizadas no R	31

LISTA DE FIGURAS

Figura 1 - Capacidades tumorais adquiridas.....	15
Figura 2 - Principais características dos subtipos moleculares do câncer gástrico ...	19
Figura 3 - Exemplo de processo da técnica de microarranjo	22
Figura 4 - Processos do RNA-Seq	23
Figura 5 - Desenho Metodológico	30

LISTA DE ABREVIATURAS E SIGLAS

DNAc	DNA complementar
CG	Câncer Gástrico
CIN	<i>Chromosomal instability</i> / Instabilidade cromossômica
EBV	<i>Epstein-Barr virus</i> / Vírus de Epstein Barr
FDR	<i>False Discovery Fate</i>
GEO	<i>Gene Expression Omnibus</i>
GS	<i>Genomically stable</i> / Genomicamente estáveis
RNA	<i>Ribonucleic acid</i> / Ácido ribonucleico
RNAm	RNA mensageiro
MSI	<i>Microsatellite instability</i> / Instabilidade de microssatélites
PCA	<i>Principal Component Analysis</i> / Análise de componentes principais
TCGA	<i>The Cancer Genome Atlas</i> / O Atlas do Genoma do Câncer
WHO	<i>World Organization Health</i> / Organização Mundial da Saúde

RESUMO

O Câncer é a segunda maior doença causadora de mortes mundialmente, sendo o câncer gástrico o terceiro mais fatal. Neste contexto, o presente trabalho tem por objetivo a identificação de possíveis biomarcadores em tumores de câncer gástrico humano, por meio do desenvolvimento de uma ferramenta computacional dedicada a exploração e visualização da expressão gênica diferencial e da análise de sobrevivência de pacientes clínicos. Na realização desta pesquisa, utilizou-se dados de microarranjos disponibilizados no banco de dados público GEO com os seguintes códigos: GSE33335, GSE54129. Utilizando os conjuntos de dados, foram aplicados testes estatísticos na busca de genes estatisticamente diferencialmente expressos. Foram encontrados 39 genes com expressão diferencial entre o grupo de tecido tumoral e tecido não tumoral adjacente. Dentro desse grupo de genes diferencialmente expressos, foram selecionados os 3 genes com o menor p-value. Estes genes foram submetidos a ferramenta C-Gemis (desenvolvimento próprio) para o enriquecimento dos dados utilizando os dados do repositório TCGA. Essa análise indicou que o gene *SIDT2* apresentou potencial como biomarcador para câncer gástrico do subtipo difuso da classificação de Laurén.

Palavras-chave: Câncer gástrico. Análise de expressão diferencial. Sobrevida. Biomarcador.

ABSTRACT

Cancer is the second leading cause of death worldwide, with gastric cancer the third most fatal form of this disease. In this way, the present work aims to identify possible biomarkers in human gastric cancer tumors through the development of a computational tool dedicated to explore and display differential gene expression and survival analyses of clinical patients. In carrying out this research, the use microarray data available in the public GEO database was made with the following codes: GSE33335, GSE54129. Using these data sets, statistical tests were applied in the search for statistically significant differentially expressed genes. In total, 39 genes were found with differential expression between the kind of tumor tissue and adjacent non-tumor tissue. Within this group of differentially expressed genes, the three with lowest p-values were selected. These genes were then submitted to the C-Gemis tool (own development) for data improvement, by crossing data from the TCGA repository. The results indicate the SIDT2 gene as a potential biomarker for gastric cancer of the diffuse subtype of the Laurén classification.

Keywords: *Gastric cancer. Differential expression analysis. Survival. Biomarker.*