ABCS HEALTH SCIENCES

ORIGINAL ARTICLE

OPEN ACCESS

Identification of novel biomarkers with potential for diagnosis and prognosis of gastric cancer: a Bioinformatics Approach

Marcos Vinicius Rossetto[1], Fernanda Pessi de Abreu[1], Pedro Lenz Casa[1], Ivaine Tais Sauthier Sartor[1,2], Scheila de Avila e Silva[1]

[1]Laboratório de Biologia Computacional e Bioinformática, Instituto de Biotecnologia, Universidade de Caxias do Sul (UCS) – Caxias do Sul (RS), Brazil
[2]Hospital Moinhos de Vento (HMV) - Porto Alegre (RS), Brazil

ABSTRACT

Introduction: Gastric cancer (GC) is the fifth most diagnosed neoplasia and the third leading cause of cancer-related deaths. A substantial number of patients exhibit an advanced GC stage once diagnosed. Therefore, the search for biomarkers contributes to the improvement and development of therapies. **Objective:** This study aimed to identify potential GC biomarkers making use of *in silico* tools. **Methods:** Gastric tissue microarray data available in Gene Expression Omnibus and The Cancer Genome Atlas Program was extracted. We applied statistical tests in the search for differentially expressed genes between tumoral and non-tumoral adjacent tissue samples. The selected genes were submitted to an in-house tool for analyses of functional enrichment, survival rate, histological and molecular classifications, and clinical follow-up data. A decision tree analysis was performed to evaluate the predictive power of the potential biomarkers. **Results:** In total, 39 differentially expressed genes were found, mostly involved in extracellular structure organization, extracellular matrix organization, and angiogenesis. The genes *SLC7A8*, *LY6E,* and *SIDT2* showed potential as diagnostic biomarkers considering the differential expression results coupled with the high predictive power of the decision tree models. Moreover, GC samples showed lower *SLC7A8* and *SIDT2* expression, whereas *LY6E* was higher. *SIDT2* demonstrated a potential prognostic role for the diffuse type of GC, given the higher patient survival rate for lower gene expression. **Conclusion:** Our study outlines novel biomarkers for GC that may have a key role in tumor progression. Nevertheless, complementary *in vitro* analyses are still needed to further support their potential.

Keywords: stomach neoplasms; computational biology; gene expression; biomarkers, tumor; tissue array analysis; decision trees.

How to cite this article: Rossetto et al. Identification of novel biomarkers with potential for diagnosis and prognosis of gastric cancer: a Bioinformatics Approach. ABCS Health Sci. 2023;48:e023227 https://doi.org/10.7322/abcshs.2021108.1836

Received: May 25, 2021
Revised: Nov 13, 2021
Approved: Dec 02, 2021

Corresponding author: Marcos Vinicius Rossetto - Laboratório de Biologia Computacional e Bioinformática, Instituto de Biotecnologia, Universidade de Caxias do Sul - Rua Francisco Getúlio Vargas, 1130 – CEP: 95070-560 - Caxias do Sul (RS), Brazil - E-mail: rossettomarcos@gmail.com

Declaration of interests: nothing to declare
Funding: CAPES, CNPq

INTRODUCTION

Gastric cancer (GC) is recognized as the fifth most commonly diagnosed malignant tumor and the third leading cause of cancer-related deaths. This disease displays a rare incidence in adults under 50 years old, being more frequent in men[1]. Furthermore, due to the aging world population, the absolute number of new cases has been increasing

every year[2]. Other factors such as low socioeconomic status, smoking, and high intake of salt, nitrites, nitrates, and alcohol are also related to GC establishment[3,4]. Additionally, some studies associate infections with the *Helicobacter pylori* bacterium as a GC risk factor[5,6].

Approximately 80% of patients diagnosed with GC exhibit the advanced stage of the disease. This scenario emerges as a result of the majority of the patients highlighting unalarming symptoms, or even appearing asymptomatic[7,8]. The overall survival rate of the disease is poor since early diagnosis tends to be less frequent[9]. Therefore, investigating potential biomarkers proves a crucial step towards the improvement of several medical procedures, including screening, estimating cancer development risk, differential diagnosis, determining prognosis, predicting responses to therapy, and monitoring disease recurrence, among others[10].

Multiple biomarker classes exist, ranging from proteins and nucleic acids to antibodies and peptides[10]. There are distinct methodologies that may be employed to identify potential biomarkers, which can be divided into classic approaches (*e.g.*, tumor biology and metabolism of the pharmaceutical agent) and modern technologies (*e.g.*, high-throughput sequencing and gene expression arrays)[10]. In the same manner, GC is also classified through different strategies, often based on tumor histology and gene expression analysis[11,12].

Messenger RNA (mRNA) is considered critical in the progression and maintenance of tumoral cells. Furthermore, mRNA displays a high potential of reflecting cellular phenotypes, since it contains a higher quantity of information in the system[13]. Therefore, the use of tools dedicated to analyzing transcripts from high-throughput techniques, such as microarrays and RNA-Seq, provides a means to investigate potential biomarkers related to the diagnosis and prognosis of diverse types of neoplasia[14,15].

The term biomarker, according to the National Cancer Institute (https://www.cancer.gov/), is applied to "a biological molecule that is a sign of a normal or abnormal process, or of a condition or disease." In this sense, multiple biomarker classes exist, ranging from proteins and nucleic acids to antibodies and peptides. There are distinct methodologies that may be employed in identifying potential biomarkers, those being divided into classic approaches (e.g., tumor biology and metabolism of the pharmaceutical agent) and modern technologies (e.g., high-throughput sequencing and gene expression arrays)[10].

Gene expression analysis was proposed in the late 90s as a complementary method to support morphology-based tumoral classification systems, once tumors of the same histotype can reach considerably distinct clinical outcomes[11,12]. One of the main challenges posed by the "omics era" is the pursuit of biologically relevant information, considering that experimental techniques like microarrays and RNA-Seq produce large volumes of data. At present, various open-access repositories collect and store data resulting from those methods, such as the Gene Expression Omnibus (GEO)[16] ArrayExpress[17] and The Cancer Genome Atlas (TCGA)[18].

The information available in databases is essential in the identification of novel biomarkers. Besides this, the use of an *in silico* approach contributes to the class determination as well as genomic architecture characterization of each cancer[19,20]. Several studies have made use of bioinformatics procedures in the search for potential biomarkers. Sartor et al.[21] for instance, identified the *TULP3* gene as a prospective biomarker for pancreatic cancer, verifying that high transcriptional levels of *TULP3* may fulfill a fundamental role in tumor progression. In another work, Xue et al.[22] investigated the potential of the *KIF4A* gene as a prognostic and diagnostic biomarker for breast cancer.

Among current medical research, biomarker studies show high promise for therapy improvement and cost reduction. The establishment of correlations between potential biomarkers and diseases can render new tools, both for diagnosis and treatment adjustment for patients[23].

From this perspective, the objective of the present paper was to identify potential GC biomarkers by using *in silico* techniques and public repository data.

METHODS

Data acquisition

Gene expression data were obtained from the GEO and TCGA repositories. Every selected dataset consists of information from patients diagnosed with gastric adenocarcinoma, including both tumoral tissue (GC) and non-tumoral adjacent tissue (NT). Two datasets were selected from the GEO database: GSE33335[24] (25 GC and 25 NT samples) and GSE54129 (111 GC and 21 NT samples). As for the TCGA database, data from the TCGA-STAD study[18] (415 GC and 35 NT samples) were obtained.

Data normalization

The raw gene expression data from studies GSE54129 (GPL570) and GSE33335 (GPL5175) was normalized by employing the Robust Multi-array Average technique, implemented in affy and oligo packages from the BioConductor repository. For the STAD-TCGA study, RNA-Seq data were preprocessed with the *TCGAbiolinks* package, also obtained from BioConductor. To avoid biased expression values, STAD-TCGA data were normalized and filtered, so that only samples situated in the interquartile range (25-75%) were considered. Afterward, the data from all studies were transformed into a logarithmic scale, for gene expression comparison between GC and non-tumoral (NT) tissue samples. A principal component analysis was then applied to identify variance distribution and filter biased samples.

Microarray probe mapping was required exclusively for GEO-derived data since in this repository every gene carries a probe-specific code. The gene-to-probe-code relation is available as a separate archive, which is standardized according to the platform used by the microarray technique. Consequently, a new probe mapping based on the GPL5175 platform annotation system was created for the GSE33335 study with the corresponding genes, once the data could not be directly imported to R. All data preprocessing and analyses were performed in R *v.*3.4.3 statistical software.

Differential expression and gene selection

For both databases, the limma package was used for the differential expression analysis[25] with a selected value of logFC of 1.5. Benjamini-Hochberg correction was used for multiple comparisons. In the GSE54129 study, JetSet scoring[26] was used to select the probe that best represents a gene. More specifically, given that a single gene can be measured by a probeset, JetSet provides individual gene mapping to the probe that best represents its expression. In the following procedure, the VennDiagram package[27] was employed to overlap the studies and pinpoint the common genes between them. This main intersection of differentially expressed genes then underwent a functional enrichment analysis conducted with the cluster profile package[28].

Analysis of biomarker potential

The selected genes were submitted to an in-house developed tool to conduct analyses of functional enrichment, survival rate, Laurén and World Health Organization (WHO) histological classifications, TCGA molecular classification, and clinical follow-up data. Furthermore, a decision tree algorithm was utilized to perform a complementary analysis of GEO data, conducted with the Orange Data Mining *v*.3.26.0 software[29].

RESULTS

Gene selection and potential biomarkers

In total, 39 genes were found to be differentially expressed considering the intersection of GSE33335 and GSE54129 studies, with a defined limit of *p*-value <0.05 and a LogFC cut-off of 1.5 (Figure 1). In addition, we were able to discern which genes exhibited a higher potential to be used as biomarkers, those being: *SLC7A8*, *LY6E,* and *SIDT2*. The complete gene list in conjunction with the statistical results from the expression analysis can be found in Supplemental Material I (Figure 1).

Functional enrichment analysis

To determine the biological function of the 39 differentially expressed genes a functional enrichment analysis was conducted (Figure 2). This approach was able to uncover the cellular pathways involved with the selected genes: (*i*) extracellular structure organization (12 genes); (*ii*) extracellular matrix organization (12 genes); and (*iii*) angiogenesis (10 genes). Additionally, a p-value of < 0.0005 was obtained for these 3 functions.

DISCUSSION

In the conducted analyses, *SLC7A8* expression presented significant statistical differences between tumoral and non-tumoral adjacent tissue samples. While tumor samples displayed lower expression values and a wider distribution range, non-tumoral ones were characterized by a higher *SLC7A8* expression and a narrower distribution range (Figure 3). The same result was observed in the Laurén and WHO histological classifications, as well as for the TCGA molecular classification. On the other hand, no statistical significance was found for the differential expression of this gene in tumoral staging and the survival rate of patients (Supplemental Material II).

The constructed decision tree models support the notion that *SLC7A8* expression values can operate as a classification attribute. The resultant model for the GSE33335 dataset determines that: "If the expression value is lower or equal to 4.86, then the sample will be classified as tumoral"; and "If the expression value is higher than 4.86, the sample will be classified as non-tumoral". Similarly, the model for the GSE54129 dataset found that: "If the expression value is lower or equal to 9.0, then the sample will be classified as tumoral"; and "If the expression value is higher than 9.0, the sample will be classified as non-tumoral". A graphical visualization of the trees, in association with their confusion matrix and statistical metrics used in the evaluation of the models, can be found in Supplemental Material III.

The *SLC7A8* gene (solute carrier family 7 member 8), or LAT2, is an amino acid transporter of the L type and not dependent[30] on Na+. The *SLC7A8* expression level has been related to other neoplasias. In lung and breast cancers, for example, a high gene expression was associated with a higher survival rate of patients and favorable

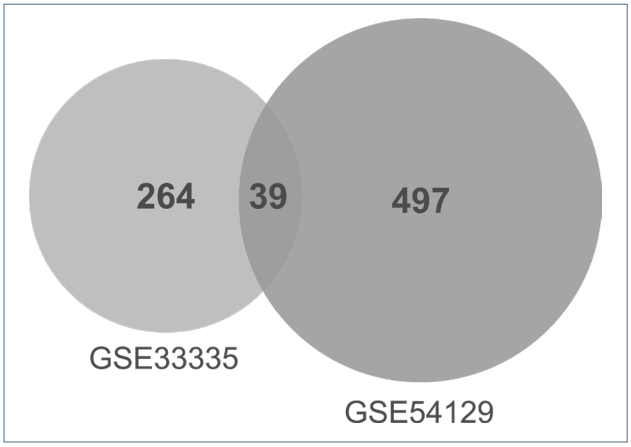

Figure 1: Venn Diagram of differentially expressed genes from studies GSE33335 and GSE54129.

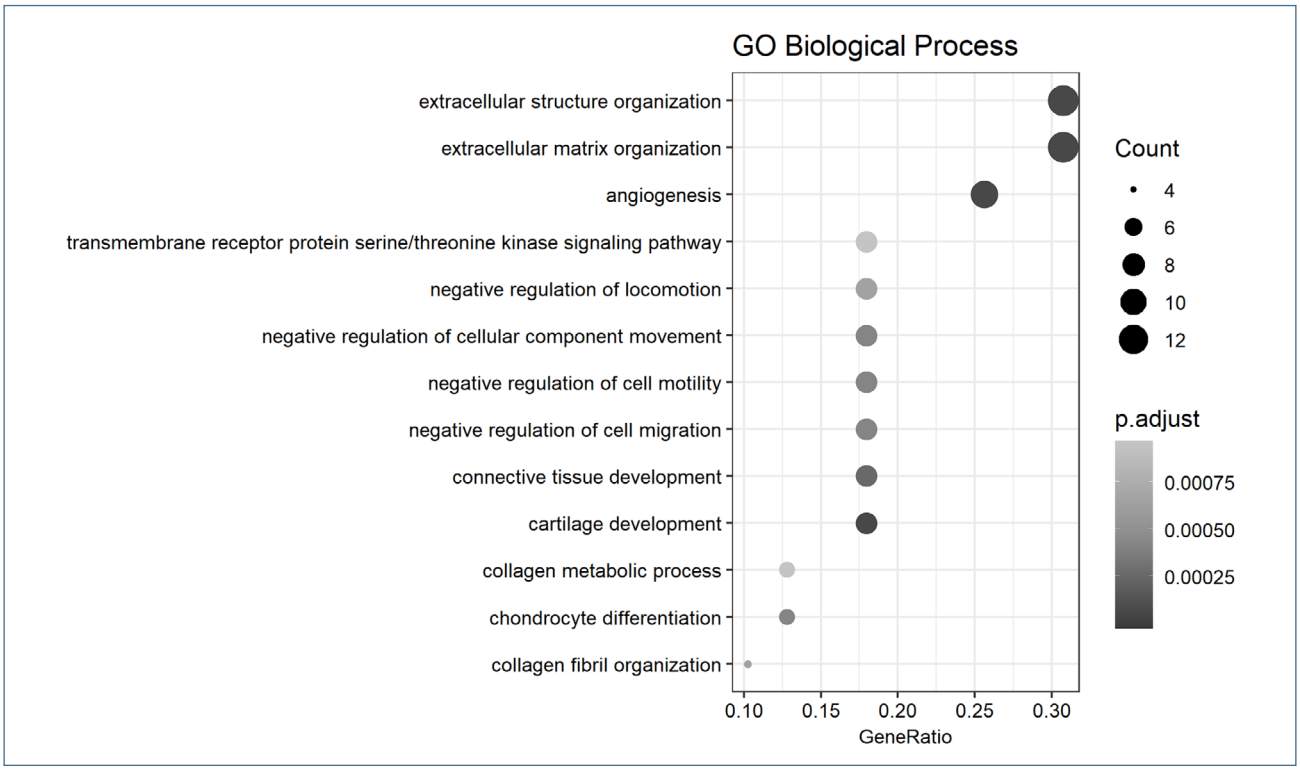

Figure 2: Gene ontology functional enrichment analysis.

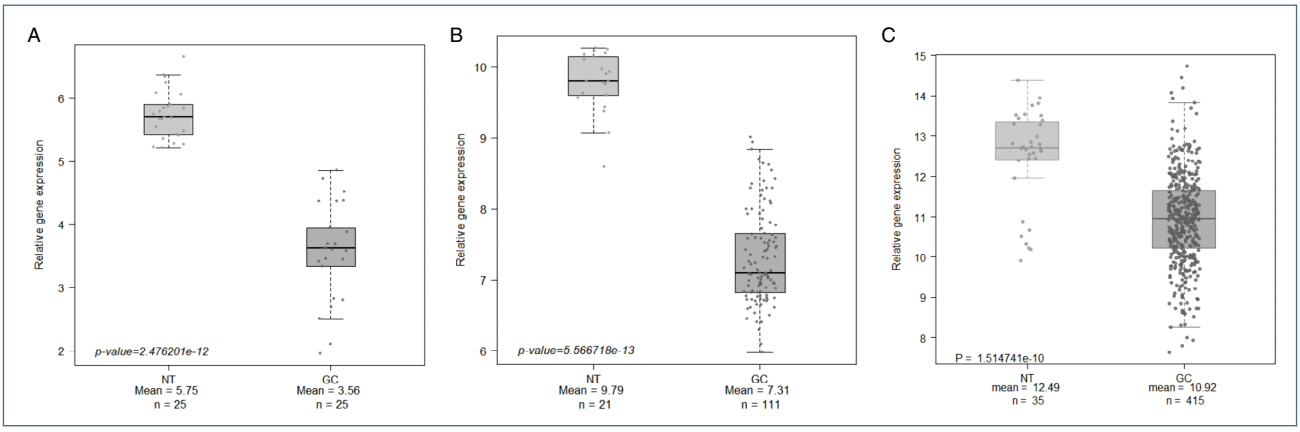

Figure 3: Comparison of the *SLC7A8* gene expression between tumoral (Gastric cancer) and non-tumoral adjacent tissue samples. Relative gene expressions were plotted for a) GSE33335; b) GSE54129; and c) TCGA data.

prognostics[31,32]. On top of that, the overexpression of *SLC7A8* in breast cancer was found to be involved in tumor suppression[32]. However, in contrast with the previously reported results, overexpression of the *SLC7A8* gene has been indicated in basal cell carcinoma, a type of skin cancer, and neuroendocrine tumor tissues[33,34].

Contrary to the previous gene, the *LY6E* gene (Lymphocyte Antigen 6 Family Member E) displayed higher expression values in GC than in NT samples (Figure 4). The same result was observed in the Laurén and WHO histological classifications, as well as the TCGA molecular classification. As for tumor staging and patient survival, there was no statistical significance in the differential expression of the gene (Supplemental Material IV). Furthermore, the models obtained from decision trees (GSE33335 and GSE54129) specify that expression values equal to and below 6.7 return non-tumoral samples, whereas values above this threshold classify as tumor samples. The graphical tree visualization, confusion matrix, and statistical evaluation metrics are available in Supplemental Material V.

The high expression levels for the *LY6E* gene in tumor samples encountered in this paper are in accord with results obtained in a work by Lv et al.[35]. In their research, the authors verified, through microarray assays, that the respective gene was overexpressed in GC samples when compared to non-tumoral cases. In addition, their study

evaluated the immunohistochemical profile of *LY6E*, revealing that 78.7% of GC samples presented proteic overexpression. Even more, the overexpression was found to be correlated to tumor grading and staging[35]. Much like *LY6E*, other *LY6* family genes are positively regulated in tumoral tissue, unlike non-tumoral tissue samples. In this background, the elevated expression of the LY6 gene family has been related to an unfavorable prognosis in distinct neoplasias[36].

The last potential biomarker here highlighted is the *SIDT2* gene, which presented lower expression in GC samples than in the NT group (Figure 5). The same result carries over when considering Laurén and WHO histological classes and TCGA molecular classification. As for tumor staging, no statistical significance was found in the differential expression of this gene (Supplemental Material VI).

Regarding the survival rate of patients, in a general manner, no significant difference was verified in the STAD-TCGA study (Supplemental Material VI). However, when considering the Laurén histological classification, *SIDT2* expression for the diffuse type of GC had statistical significance in the survival rate of patients. In this particular case, the high gene expression in GC is associated with a lower patient survival rate (Figure 6A). The ROC curve (Figure 6B) for survival rate shows an AUC value of 65,5% in the cases of high *SIDT2* expression. Nguyen et al.[37] have described the role of *Sidt2* in tumor progression for lung and intestinal adenocarcinoma, using animal models. In his work, the author observed that mice without *Sidt2* expression would show a reduction in tumor progression together with an increase in survival rate[37].

The decision tree analysis demonstrates that different expression levels of *SIDT2* can be used as a classifying feature. For the GSE33335 dataset: "If the expression value is lower or equal to 5.34, then the sample will be classified as tumoral"; and "If the expression value is higher than 5.34, the sample will be classified as non-tumoral". Whereas for the GSE54129 dataset: "If the expression value is lower or equal to 10.2, then the sample will be classified as tumoral"; and "If the expression value is higher than 10.2, the sample will be classified as non-tumoral". A graphical rendering of the tree models, in addition to their confusion matrices, and statistical evaluation metrics are available in Supplemental Material VII.

The *SIDT2* protein mediates RNA transport to lysosomes, promoting a degradation process known as RNAutophagy[37,38]. In agreement with the NT and GC expression results found in our paper, Beck et al.[39]

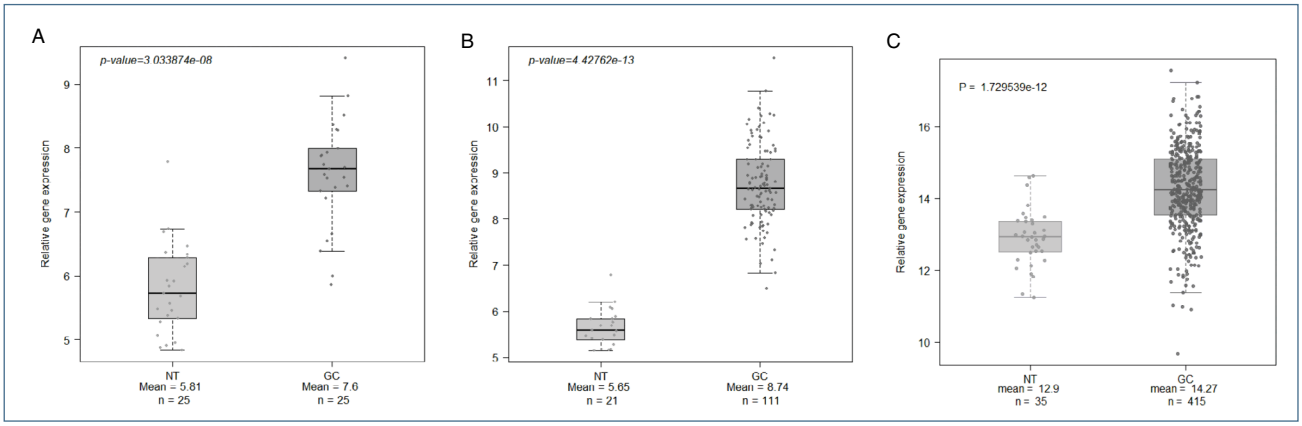

Figure 4: Comparison of the *LY6E* gene expression between tumoral (Gastric cancer) and non-tumoral adjacent tissue samples. Relative gene expressions were plotted for a) GSE33335; b) GSE54129; and c) TCGA data.

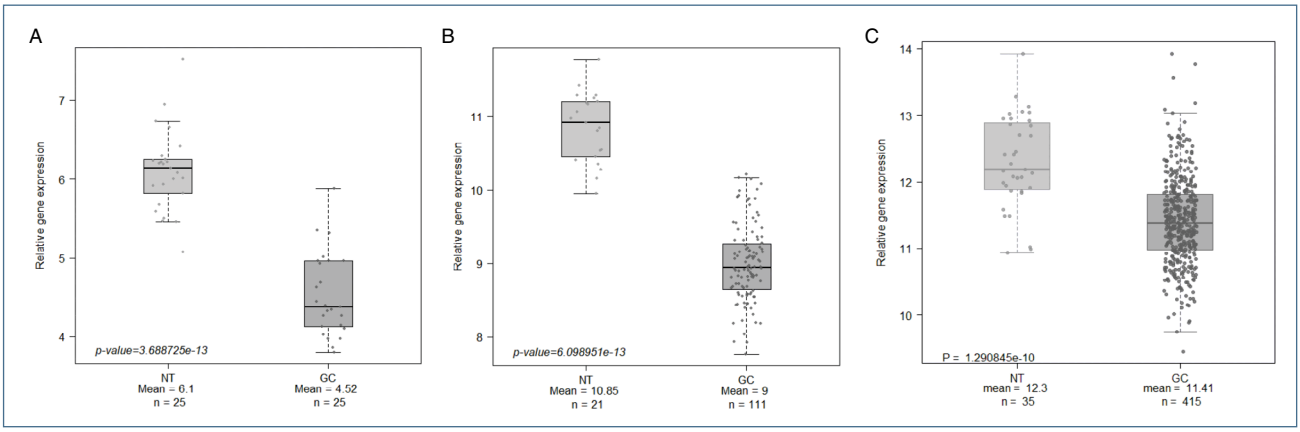

Figure 5: Comparison of the *SIDT2* gene expression between tumoral (Gastric cancer) and non-tumoral adjacent tissue samples. Relative gene expressions were plotted for a) GSE33335; b) GSE54129; and c) TCGA data.

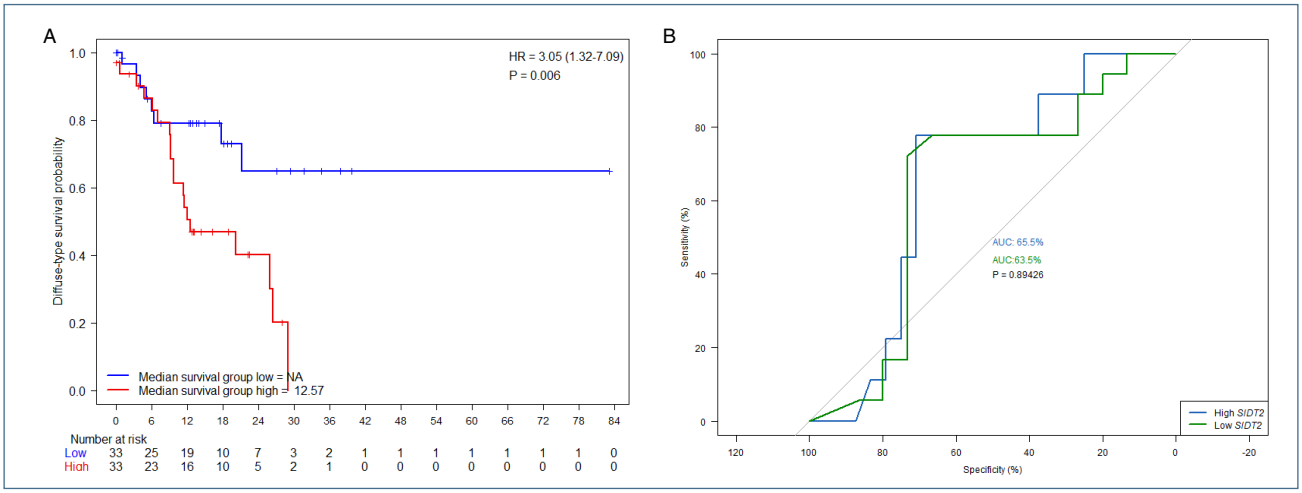

Figure 6: Comparison of survival rates based on *SIDT2* expression for the diffuse type of GC: a) Patient survival rate, b) ROC curve.

identified high expression of *SIDT2* transcripts in healthy human tissue including the stomach, pancreas, spinal cord, prostate, testicles, and placenta. Moreover, they detail that *SIDT2* displays a negative regulation level in tumor tissue in comparison to the corresponding healthy tissue. Similarly, Brady et al.[40] report that the *SIDT2* gene is found underexpressed in a variety of mice and human tumors. Even more, the authors delineate its action as a *TP53*-dependent tumor suppressor. In this regard, Nguyen et al.[37] when investigating the role of *Sidt2* on mice lung adenocarcinoma tumorigenesis, reported that animals with *Sidt2* deficiency developed significantly fewer tumors and showed a substantial reduction in total tumor yield. These results evidence the tumor suppressive action of *Sidt2* and can explain the survival rate of our findings.

The use of *in silico* tools enabled the identification of novel biomarkers for GC that may have a role in the disease progression. Our study outlines three possible diagnostic biomarkers, the genes *SLC7A8, LY6E,* and *SIDT2*, given that they displayed a statistically significant differential expression between tumoral and non-tumoral adjacent tissue samples. Furthermore, the *SIDT2* gene exhibited a potential role as a prognostic GC biomarker for the diffuse type of cancer, considering the association between the high gene expression and the lower survival rate.

Considering the diverse types of GC, studies that identify potential diagnostic or prognostic biomarkers histology-specific are important to the contribution of the improvement of the knowledge of the GC. These three genes also appear related to other kinds of neoplasia in the literature. However, complementary *in vitro* analyses are still needed to provide further support to these genes as potential biomarkers for gastric cancer.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. https://doi.org/10.3322/caac.21492
2. Luo G, Zhang Y, Guo P, Wang L, Huang Y, Li K. Global patterns, and trends in stomach cancer incidence: Age, period and birth cohort analysis. Int J Cancer. 2017;141(7):1333-44. https://doi.org/10.1002/ijc.30835
3. Gao J, Lindsay J, Watt S, Bahceci I, Lukasse P, Abeshouse A, et al. The cBioPortal for cancer genomics and its application in precision oncology. Cancer Res. 2016;76(14 Suppl):5277. https://doi.org/10.1158/1538-7445.AM2016-5277
4. Kelley JR, Duggan JM. Gastric cancer epidemiology and risk factors. J Clin Epidemiol. 2003;56(1):1-9. https://doi.org/10.1016/S0895-4356(02)00534-6
5. Choli-Papadopoulou T, Kottakis F, Papadopoulos G, Pendas S. Helicobacter pylori neutrophil activating protein as target for new drugs against H. pylori inflammation. World J Gastroenterol. 2011;17(21):2585-91. https://doi.org/10.3748/wjg.v17.i21.2585
6. Flora S, La Maestra S. Epidemiology of cancers of infectious origin and prevention strategies. J Prev Med Hyg. 2015;56(1):E15-20.
7. Bornschein J, Kandulski A, Selgrad M, Malfertheiner P. From gastric inflammation to gastric cancer. Dig Dis. 2010;28(4-5):609-14. https://doi.org/10.1159/000320061
8. Schmidt N, Peitz U, Lippert H, Malfertheiner P. Missing gastric cancer in dyspepsia. Aliment Pharmacol Ther. 2005;21(7):813-20. https://doi.org/10.1111/j.1365-2036.2005.02425.x
9. Lee JY, Kim HI, Kim YN, Hong JH, Alshomimi S, An JY, et al. Clinical significance of the prognostic nutritional index for predicting short- and long-term surgical outcomes after gastrectomy: a retrospective analysis of 7781 gastric cancer patients. Medicine (Baltimore). 2016;95(18):e3539. https://doi.org/10.1097/MD.0000000000003539
10. Henry NL, Hayes DF. Cancer biomarkers. Mol Oncol. 2012;6(2):140-6. https://doi.org/10.1016/j.molonc.2012.01.010

11. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286(5439):531-7. https://doi.org/10.1126/science.286.5439.531

12. Wan C, Li J. Synthesis of well-dispersed magnetic CoFe2O4 nanoparticles in cellulose aerogels via a facile oxidative co-precipitation method. Carbohydr Polym. 2015;134:144-50. https://doi.org/10.1016/j.carbpol.2015.07.083

13. Cristescu R, Lee J, Nebozhyn M, Kim K-M, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21(5):449-56. https://doi.org/10.1038/nm.3850

14. Broadhead ML, Clark JCM, Dass CR, Choong PFM. Microarray: an instrument for cancer surgeons of the future? ANZ J Surg. 2010;80(7-8):531-6. https://doi.org/10.1111/j.1445-2197.2010.05379.x

15. D'Angelo G, Di Rienzo T, Ojetti V. Microarray analysis in gastric cancer: a review. World J Gastroenterol. 2014;20(34):11972-6. https://doi.org/10.3748/wjg.v20.i34.11972

16. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10. https://doi.org/10.1093/nar/30.1.207

17. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update - from bulk to single-cell expression data. Nucleic Acids Res. 2019;47(D1):D711-5. https://doi.org/10.1093/nar/gky964

18. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513(7517):202-9. https://doi.org/10.1038/nature13480

19. Zheng H, Zhang G, Zhang L, Wang Q, Li H, Han Y, et al. Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis. Front Oncol. 2020;10:68. https://doi.org/10.3389/fonc.2020.00068

20. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19(1A):A68-77. https://doi.org/10.5114/wo.2014.47136

21. Sartor ITS, Zeidán-Chuliá F, Albanus RD, Dalmolin RJS, Moreira JCF. Computational analyses reveal a prognostic impact of TULP3 as a transcriptional expert regulator in pancreatic ductal adenocarcinoma. Mol Biosyst. 2014;10(6):1461-8. https://doi.org/10.1039/c3mb70590k

22. Xue D, Cheng P, Han M, Liu X, Xue L, Ye C, et al. An integrated bioinformatical analysis to evaluate the role of KIF4A as a prognostic biomarker for breast cancer. Onco Targets Ther. 2018;11:4755-68. https://doi.org/10.2147/OTT.S164730

23. Poste G. Bring on the biomarkers. Nature. 2011;469:156-7. https://doi.org/10.1038/469156a

24. Cheng L, Yang S, Yang Y, Zhang W, Xiao H, Gao H, et al. Global gene expression and functional network analysis of gastric cancer identify extended pathway maps and GPRC5A as a potential biomarker. Cancer Lett 2012;326(1):105-13. https://doi.org/10.1016/j.canlet.2012.07.031

25. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007

26. Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene. BMC Bioinform. 2011;12:474. https://doi.org/10.1186/1471-2105-12-474

27. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinform. 2011;12:35. https://doi.org/10.1186/1471-2105-12-35

28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7. https://doi.org/10.1089/omi.2011.0118

29. Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, et al. Orange: data mining toolbox in python. J Mach Learn Res. 2013;14(35):2349-53.

30. Fotiadis D, Kanai Y, Palacín M. The SLC3 and SLC7 families of amino acid transporters. Mol Aspects Med. 2013;34(2-3):139-58. https://doi.org/10.1016/j.mam.2012.10.007

31. Asada K, Kobayashi K, Joutard S, Tubaki M, Takahashi S, Takasawa K, et al. Uncovering Prognosis-Related Genes and Pathways by Multi-Omics Analysis in Lung Cancer. Biomolecules. 2020;10(4):524. https://doi.org/10.3390/biom10040524

32. El Ansari R, Alfarsi L, Craze ML, Masisi BK, Ellis IO, Rakha EA, et al. The solute carrier SLC7A8 is a marker of favourable prognosis in ER-positive low proliferative invasive breast cancer. Breast Cancer Res Treat. 2020;181(1):1-12. https://doi.org/10.1007/s10549-020-05586-6

33. Tina E, Prosén S, Lennholm S, Gasparyan G, Lindberg M, Göthlin Eremo A. Expression profile of the amino acid transporters SLC7A5, SLC7A7, SLC7A8 and the enzyme TDO2 in basal cell carcinoma. Br J Dermatol. 2019;180(1):130-40. https://doi.org/10.1111/bjd.16905

34. Barollo S, Bertazza L, Watutantrige-Fernando S, Censi S, Cavedon E, Galuppini F, et al. Overexpression of L-Type Amino Acid Transporter 1 (LAT1) and 2 (LAT2): Novel Markers of Neuroendocrine Tumors. PLoS One. 2016;11(5):e0156044. https://doi.org/10.1371/journal.pone.0156044

35. Lv Y, Song Y, Ni C, Wang S, Chen Z, Shi X, et al. Overexpression of Lymphocyte Antigen 6 Complex, Locus E in Gastric Cancer Promotes Cancer Cell Growth and Metastasis. Cell Physiol Biochem. 2018;45(3):1219-29. https://doi.org/10.1159/000487453

36. Upadhyay G. Emerging Role of Lymphocyte Antigen-6 Family of Genes in Cancer and Immune Cells. Front Immunol. 2019;10:819. https://doi.org/10.3389/fimmu.2019.00819

37. Nguyen TA, Bieging-Rolett KT, Putoczki TL, Wicks IP, Attardi LD, Pang KC. SIDT2 RNA Transporter Promotes Lung and Gastrointestinal Tumor Development. iScience. 2019;20:14-24. https://doi.org/10.1016/j.isci.2019.09.009

38. Aizawa S, Contu VR, Fujiwara Y, Hase K, Kikuchi H, Kabuta C, et al. Lysosomal membrane protein SIDT2 mediates the direct uptake of DNA by lysosomes. Autophagy. 2017;13(1):218-22. https://doi.org/10.1080/15548627.2016.1248019

39. Beck A, Fecher-Trost C, Wolske K, Philipp SE, Flockerzi V, Wissenbach U. Identification of Sidt2 as a lysosomal cation-conducting protein. FEBS Lett. 2017;591(1):76-87. https://doi.org/10.1002/1873-3468.12528

40. Brady CA, Jiang D, Mello SS, Johnson TM, Jarvis LA, Kozak MM, et al. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. Cell 2011;145(4):571-83. https://doi.org/10.1016/j.cell.2011.03.035