

ORIGINAL RESEARCH ARTICLE

Cancer-testis antigens in gastric adenocarcinoma: Integration of differential expression, clinical associations, survival, and co-expression networks

Jade Dias Valente¹, Livia Érika Carlos Marques^{1*}, Ronald Matheus da Silva Mourão¹, Bianca de Fátima dos Reis Rodrigues¹, Louise Sousa de Souza¹, Jéssica Manoelli Costa da Silva¹, Samir Mansour Moraes Casseb¹, André Salim Khayat¹, Samia Demachki¹, Geraldo Ishak¹, Fabiano Cordeiro Moreira¹, Rommel Mario Rodríguez Burbano², and Paulo Pimentel de Assumpção¹

¹Oncology Research Center, Federal University of Pará, Belém, Pará, Brazil

²Molecular Biology Laboratory, Ophir Loyola Hospital, Belém, Pará, Brazil

***Corresponding author:**
Livia Érika Carlos Marques
(liviamarques@ufpa.br)

Citation: Valente JD, Marques LÉC, da Silva RM, *et al.* Cancer-testis antigens in gastric adenocarcinoma: Integration of differential expression, clinical associations, survival, and co-expression networks. *Microbes & Immunity*. 2026;3(3):025510135. doi: 10.36922/MI025510135

Received: December 16, 2025

Revised: March 12, 2026

Accepted: March 17, 2026

Published online: April 24, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Gastric adenocarcinoma (GAC) remains one of the deadliest malignancies worldwide, and the identification of molecular biomarkers is essential to improve prognostic stratification and guide therapeutic strategies. Cancer-testis antigens (CTAs) represent promising targets for immunotherapy, yet their expression landscape in GAC remains limited, especially when considering comprehensive and updated gene lists. In this study, we analyzed 389 samples, including tumor, peritumoral, metaplastic, and normal tissues, using RNA sequencing data to characterize the transcriptional profile of CTAs in GAC. We identified 104 CTAs differentially expressed among groups, with a predominance of upregulated genes in tumors compared with other tissues. The expression of several CTAs, such as *MAGEA2*, *MAGEA3*, *MAGEA6*, *MAGEA12*, *CTCF*, *IGF2BP1*, and *GAST*, showed significant associations with clinically relevant variables, including tumor–node–metastasis staging, lymph node involvement, metastasis, histological subtype, and therapeutic management. Survival analysis revealed 17 CTA genes associated with overall survival, of which 14 correlated with poor prognosis, whereas *PIK3R3*, *SYCE1L*, and *ZNF683* demonstrated a protective effect. Co-expression analysis identified five well-defined functional modules, reflecting coordinated CTA signatures related to germline reprogramming, proliferation, and immune remodeling within the tumor microenvironment. Collectively, our results provide the most comprehensive characterization of CTAs in GAC to date, highlighting genes with clinical relevance, prognostic value, and translational potential as candidates for future personalized immunotherapeutic approaches.

Keywords: Antigens; Cancer-testis antigens; Gastric cancer; Molecular biomarkers; Therapeutic targets; Transcriptomics

1. Introduction

Gastric adenocarcinoma (GAC) remains a persistent global public health challenge, ranking as the fourth leading cause of cancer-related mortality worldwide.¹ Most patients have an unfavorable prognosis, with five-year survival rates ranging from 20% to 30% in most regions of the world,² primarily due to late diagnosis.³

For resectable GAC or gastroesophageal junction tumors, perioperative fluorouracil, leucovorin, oxaliplatin, and docetaxel (FLOT) remains a standard regimen, although recurrence remains common. Recently, the phase III MATTERHORN clinical trial demonstrated that adding durvalumab (anti-programmed death-ligand 1) to perioperative FLOT improved pathologic complete response and event-free survival in patients with potentially curable disease.⁴

Over the past decades, immunotherapy has emerged as an alternative for cancers that do not respond to conventional treatments, representing a promising therapeutic strategy to improve outcomes for patients with GAC. Accordingly, recent research has increasingly focused on selective targeted therapies, immunotherapies, and the identification of predictive biomarkers and therapeutic targets.⁵⁻⁷

Cancer-testis antigens (CTAs) are proteins encoded by genes whose expression is normally restricted to germline and trophoblastic tissues but aberrantly re-expressed in multiple cancers.⁸ CTAs have been implicated in tumorigenesis, therapy resistance, and metastatic progression.^{9,10} They participate in multiple cancer-related cellular pathways and contribute to nearly all hallmarks of cancer, including both the original and emerging ones.¹¹

Moreover, CTAs exhibit unique characteristics that position them as attractive targets for immune-directed cancer therapies.¹² The testis and placenta, the tissues in which they are expressed, are immunoprivileged sites due to the presence of the blood–testis and blood–placental barriers, which limit immune system access and prevent the establishment of long-lasting immune tolerance to these antigens.¹³ Therefore, when CTAs are aberrantly expressed in other tissues, they are recognized as “non-self” and trigger cytotoxic T lymphocyte responses, eliciting an antitumor immune reaction.¹⁴ Consequently, CTAs are currently being evaluated in a wide range of clinical trials aimed at developing new immunotherapy-based strategies, including therapeutic vaccines and chimeric antigen receptor (CAR)-T cell approaches, underscoring their translational potential.¹⁵⁻¹⁷

In 2009, more than 200 CTA-encoding genes were catalogued in a dedicated database ([http://www.cta.lncc.](http://www.cta.lncc.br/)

[br/](http://www.cta.lncc.br/)).¹⁸ Currently, new technologies have led to a substantial expansion in the number of identified CTAs,¹⁹ although these newly described genes have not been incorporated into the database. Because no clear and up-to-date registry of CTA genes is currently available, many studies of CTA expression in cancer still rely on outdated gene lists,^{20,21} limiting interpretation and reproducibility.

Advances in next-generation sequencing technologies have enabled comprehensive transcriptomic analyses, allowing the identification of specific expression patterns and the detection of differentially expressed genes (DEGs), thereby providing valuable insights into the discovery of potential biomarkers and novel therapeutic targets.^{22,23} The integration of newly generated data with existing knowledge may offer additional benefits for gastric tumor immunotherapy in the coming years.²⁴

In this study, we investigated the transcriptional profile of CTAs by analyzing their differential expression in GAC tissues, peritumoral tissue (PTT), metaplastic (MP) tissue, and normal tissue without neoplastic lesions. The resulting DEGs were correlated with clinical variables, as well as their impact on overall survival (OS). Additionally, we constructed co-expression networks to explore the functional interactions between CTAs and partner genes. This study provides comprehensive insights into CTAs in GAC and highlights their relevance as potential biomarkers and therapeutic targets for immunotherapy.

2. Methods

2.1. Ethical considerations and sample characterization

A total of 389 samples from patients in the Brazilian Amazon region (Northern Brazil) were collected and classified into four categories: 156 GAC samples, comprising both Lauren histological types, as well as Epstein–Barr virus-positive and -negative tumors and *Helicobacter pylori*-positive and -negative tumors; 186 samples of PTT adjacent to the tumor; 20 samples of MP, a precursor condition that may progress to gastric cancer (GC); and 27 samples of gastric mucosa without neoplastic lesions (control samples) obtained from patients undergoing upper gastrointestinal endoscopy for other clinical indications at João de Barros Barreto University Hospital (HUIBB) and Ophir Loyola Hospital (HOL).

During gastric resection, 0.5 cm fragments were collected from each tumor, 0.5 cm fragments from MP tissue, and an additional 0.5 cm fragment located at least 5 cm from the macroscopic tumor margins for cases included in the PTT analysis. For patients without neoplastic lesions, 4 mm tissue fragments were obtained

through endoscopic biopsy. All samples were stored in RNAlater and subsequently preserved at -80°C .

The medical records of all patients included in this study were reviewed to obtain clinical data such as sex, age, Lauren histological subtype, tumor-node-metastasis (TNM) staging, tumor location, therapeutic management, and T, N, and M categories (Table 1). Individuals were recruited at HUJBB and HOL. All participants signed an informed consent form, voluntarily agreed to participate, and were informed about the objectives of the study. The conduct of this study was approved by the Research Ethics Committee under protocol number CAAE 47580121.9.0000.5634.

2.2. Total RNA extraction

Total RNA was extracted from tissue samples using TRIzol® (Thermo Fisher Scientific, United States of America [USA]), following the manufacturer's instructions. Initially, approximately 50–100 mg of tissue from each sample was homogenized, and 1 mL of TRIzol was then added to the processed tissue at room temperature. RNA integrity and concentration were assessed using the Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA) and the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA). The ideal criteria for RNA integrity were A260/A280 ratios between 1.8 and 2.2, A260/A230 ratios > 1.8 , and an RNA integrity number ≥ 5 . The extracted total RNA was stored at -80°C until further use.

2.3. Library preparation and next-generation sequencing

Library preparation was performed using 1 μg of total RNA per sample in a total volume of 10 μL , employing the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold (20040529, Illumina, USA) according to the manufacturer's instructions. After library construction, integrity was reassessed using the 2,200 TapeStation System (Agilent, USA), and the final product displayed a band of approximately 260 base pairs. cDNA libraries were sequenced on the Illumina NextSeq platform in paired-end mode (NextSeq 500 System Illumina, USA). Sequencing was carried out using the commercial NextSeq® 500 ID Output V2—150 cycles kit (20024907, Illumina, USA) under the conditions specified by the manufacturer.

2.4. Quality control, alignment, and quantification

RNA sequencing (RNA-Seq) data were processed using the nf-core/rnaseq pipeline²⁵ to ensure greater reproducibility and consistency in the analyses. Reads obtained were converted to FASTQ format using Reporter software, with all sequences and quality scores encoded in ASCII. Read quality was assessed using FastQC (v0.11.9), followed

by removal of adapters and low-quality reads with Trimmomatic,²⁶ applying a minimum Phred quality score threshold of QV15. Reads were filtered and aligned to the human transcriptome reference (hg38), annotated with GENCODE (v42), using STAR (v2.7.10a), generating BAM files. Subsequently, transcript abundance was quantified using Salmon (v1.5.2)²⁷ in two steps: quantification of mRNAs using FASTA sequences of human protein-coding transcripts, and identification of ncRNAs using FASTA sequences of non-coding transcripts. The resulting transcript abundances were imported into R (v4.5.1; <https://www.r-project.org>)²⁸ using the Tximport package.²⁹

2.5. Differential transcript expression analysis

The resulting count matrices were normalized, and transcript abundance at the gene level was estimated using the DESeq2³⁰ package in RStudio. Differential expression analysis was performed considering the four sample groups (GAC, PTT, MP, and normal), with comparisons conducted through specific contrasts between the groups of interest. Transcripts with $|\log_2(\text{fold change})|$ (L2FC) > 2 and an adjusted p -value < 0.05 were considered differentially expressed, and results were visualized using volcano plots. Global expression of CTA genes across the groups was represented with boxplots, and differences were assessed using the Kruskal–Wallis test. For downstream analyses, variance-stabilized and batch-corrected expression values were used to test gene associations with clinical variables, OS analysis, and construction of correlation networks.

2.6. Cancer-testis antigen gene selection

In this study, a comprehensive list of 1,103 putative CTA genes, previously described by Silva *et al.*,¹⁹ was used. In that work, the authors identified genes with preferential or exclusive expression in the testis based on RNA-Seq data from the Human Body Map, Genotype-Tissue Expression Project, and Human Protein Atlas databases. As a selection criterion, a proportional score ≥ 0.9 was considered, indicating that at least 90% of the total gene expression across the evaluated tissues was derived from the testis. Therefore, for subsequent analyses, we directly adopted this previously established list of CTAs ($n = 1,103$ genes).

2.7. Functional enrichment analysis

Gene Ontology (GO) enrichment analysis was performed using an overrepresentation approach implemented in the clusterProfiler package, based on the top 15 DEGs and upregulated genes identified in the comparison between GAC and PTT, presented in Table 2. These genes were used as input, and the background universe was defined as all genes present in the tumor expression matrix normalized by variance-stabilizing transformation. Gene symbols were

converted to Entrez IDs using org.Hs.eg.db. Biological process terms were tested with enrichGO(), applying the Benjamini–Hochberg correction, and terms with an adjusted p -value ≤ 0.05 were considered significant. The main enriched categories were visualized using scatter plots.

2.8. Association of cancer-testis antigens with clinical variables

Normalized data were used to assess the relationship between CTA gene expression in GAC and (i) pathological variables: Lauren histological subtype, therapeutic management, TNM staging, and T, N, and M categories; and (ii) epidemiological characteristics: sex and age. Statistical associations were evaluated using DESeq2 in RStudio (adjusted p -value < 0.05), with multiple testing controlled by the Benjamini–Hochberg false discovery rate (FDR) procedure, and univariate designs of the form $\sim$ clinical_variable. The Wilcoxon test ($p < 0.05$) was applied between pairs of clinical variables as a complementary approach. Results were visualized using boxplots.

2.9. Survival analysis

For OS analysis, univariate Cox regression models with 95% confidence intervals were used via the Survival package in R. The p -values were obtained from both the log-rank test and the Cox model, and the corresponding hazard ratios (HRs) were reported. p -values were adjusted for multiple comparisons using FDR correction (Benjamini–Hochberg), and genes with adjusted p -values < 0.05 were considered significant. Kaplan–Meier curves were generated using the Survminer package to visualize differences in survival between high and low gene expression groups.

2.10. Construction of co-expression networks

Using the variance-stabilized expression matrix, Spearman correlations (ρ) were calculated exclusively across tumor samples. CTAs previously identified as DEGs between tumor and non-tumor samples were defined as target genes. For each CTA, ρ was calculated between its expression and that of all other genes, selecting the five partners with the highest $|\rho|$ for each target. This strategy allowed the analysis to focus on the most informative local network context of each CTA, reducing noise and multiple testing issues, while generating a biologically interpretable subnetwork. Subsequently, an undirected graph weighted by $|\rho|$ was constructed, preserving the correlation sign as an edge attribute, and a $|\rho| \geq 0.30$ filter was applied to retain only relevant associations.

Cluster detection was performed using the Leiden algorithm under the Constant Potts Model, prioritizing

the identification of co-expression macro modules surrounding CTAs, which are useful for generating hypotheses of gene co-regulation and involvement in shared pathways. The layout for visualization was obtained using the layout_with_graphopt function from the igraph package (version 1.3.5), and networks were rendered with the ggraph package (version 2.2.2).

3. Results

3.1. Differential expression of cancer-testis antigens

Among patients with available clinical data, the cohort of 156 individuals with GAC was primarily composed of patients aged 60–69 years (20.5%), and male patients constituted the largest reported sex subgroup (47.4%). Most patients presented with stage III disease according to the clinical TNM (cTNM) system (28.2%). When evaluated by T, N, and M categories, the majority were classified as T4 (33.3%), N3 (26.9%), and M0 (51.9%). Regarding Lauren's classification, 48.1% were intestinal type and 22.4% were diffuse type. Most tumors were located in non-cardia regions (60.9%; Table 1).

The flowchart in Figure 1 illustrates the steps of the analysis. The study included 389 samples divided among GAC tissues (156), PTT (186), MP (20), and normal tissues (27). Comparison of gene expression profiles between these groups led to the identification of 104 DEGs belonging to the previously defined list of CTAs, according to the cohort criteria. The differential expression analysis performed using the DESeq2 package in R is presented in Figure 2.

Comparison between cancer and PTT identified 14 CTA DEGs, of which 71.43% (10/14) were upregulated, and 28.57% (4/14) were downregulated (Figure 2A). Comparing cancer and metaplasia revealed 47 DEGs, with 91.49% (43/47) upregulated and only 8.51% (4/47) downregulated (Figure 2B). Analysis between tumor tissue and normal tissue identified 81 DEGs, with 58.02% (47/81) upregulated and 41.98% (34/81) downregulated (Figure 2C). CTA gene expression was also compared between PTT and metaplasia, resulting in 39 DEGs, comprising 92.31% (36/39) upregulated genes and 7.69% (3/39) downregulated genes (Figure 2D).

The distribution of global CTA gene expression levels across different tissue types is shown in Figure 3. Highlighted genes were selected based on statistical and/or biological relevance. Most evaluated genes exhibited significant differences in expression between groups (Kruskal–Wallis, $p < 0.05$). Several CTAs, including members of the MAGE and DAZ families, showed significantly higher expression in GAC compared with normal tissue. In PTT, expression levels were often intermediate, suggesting a tumor

influence on the peritumoral microenvironment. Some genes showed no significant variation, indicating non-specific expression.

Table 1. Clinicopathological characteristics of patients (N = 156) with gastric adenocarcinoma

Characteristics	Category	n (%)
Age	30–39	8 (5.1)
	40–49	18 (11.5)
	50–59	29 (18.6)
	60–69	32 (20.5)
	70–79	23 (14.7)
	≥80	8 (5.1)
	Not available	38 (24.4)
Gender	Female	44 (28.2)
	Male	74 (47.4)
	Not available	38 (24.4)
TNM stage	I	5 (3.2)
	II	30 (19.2)
	III	44 (28.2)
	IV	30 (19.2)
	Not available	47 (30.1)
T stage	T1	7 (4.5)
	T2	13 (8.3)
	T3	42 (26.9)
	T4	52 (33.3)
	Not available	42 (26.9)
N stage	N0	23 (14.7)
	N1	25 (16.0)
	N2	26 (16.7)
	N3	42 (26.9)
	Not available	40 (25.6)
M stage	M0	81 (51.9)
	M1	30 (19.2)
	Not available	45 (28.9)
Lauren type	Diffuse	35 (22.4)
	Intestinal	75 (48.1)
	Not available	46 (29.5)
Tumor location	Cardia	21 (13.5)
	Non-cardia	95 (60.9)
	Not available	40 (25.6)

Abbreviations: n: Number of cases; N: Total number of samples in the gastric adenocarcinoma group.

Additionally, pronounced intra-group heterogeneity was observed, reflected by the wide dispersion of individual values across all tissues. This variability is consistent with the biological diversity of GC, including differences among patients, molecular subtypes, and heterogeneous cellular composition. These findings highlight the need for further validation in stratified cohorts and reinforce the superiority of multigene panels over individual markers for more robust characterization.

Among the 486 upregulated DEGs in the comparison between GAC and PTT, 4 CTAs were identified among the 15 most highly expressed genes in tumors (Table 2). Notably, *MAGEA3* ranked 6th with an L2FC > 8, followed by *MAGEA6* in 8th place (L2FC > 7). *GOLGA6L1* and *MAGEA2* were ranked 11th and 13th, respectively, with L2FC values >7 and >6.

Table 2. Top 15 upregulated differentially expressed genes identified between gastric adenocarcinoma and peritumoral tissue

Gene	L2FC	Adjusted p-value	Category
<i>ZIC1</i>	11.9	$p < 0.001$	Others
<i>STRA6</i>	10.5	$p < 0.001$	Others
<i>FOXC2</i>	10.5	$p < 0.001$	Others
<i>ZIC2</i>	10.3	$p < 0.001$	Others
<i>ITIH2</i>	9.2	$p < 0.001$	Others
<i>MAGEA3</i>	8.3	8.45×10^{-9}	CTA
<i>SERPINB3</i>	8	$p < 0.001$	Others
<i>MAGEA6</i>	7.9	7.33×10^{-7}	CTA
<i>CRABP1</i>	7.8	$p < 0.001$	Others
<i>FIBIN</i>	7.6	$p < 0.001$	Others
<i>GOLGA6L1</i>	7.2	0.03	CTA
<i>FBN3</i>	6.9	$p < 0.001$	Others
<i>MAGEA2</i>	6.6	5.99×10^{-7}	CTA
<i>SFRP2</i>	6.4	$p < 0.001$	Others
<i>GAP43</i>	6.3	$p < 0.001$	Others

Abbreviations: CTA: Cancer-testis antigen; L2FC: Log2 fold change.

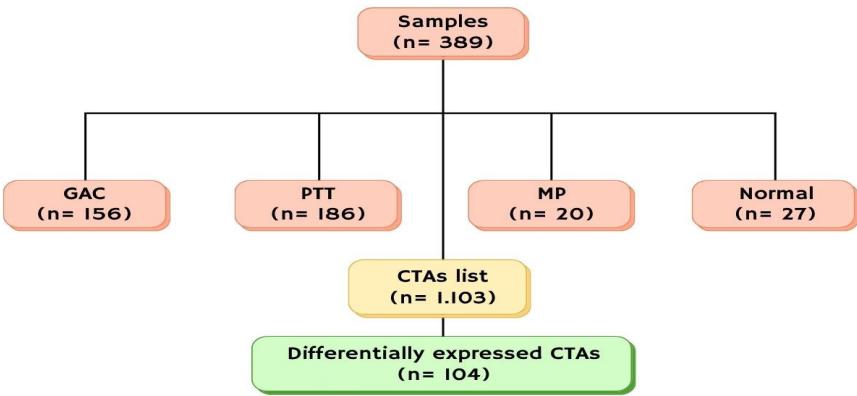


Figure 1. Analysis workflow. Distribution of samples ($n = 389$) across the gastric adenocarcinoma (GAC), peritumoral tissue (PTT), metaplastic (MP), and normal groups, identification of differentially expressed genes, and selection of cancer-testis antigens (CTAs). n represents the number of samples.

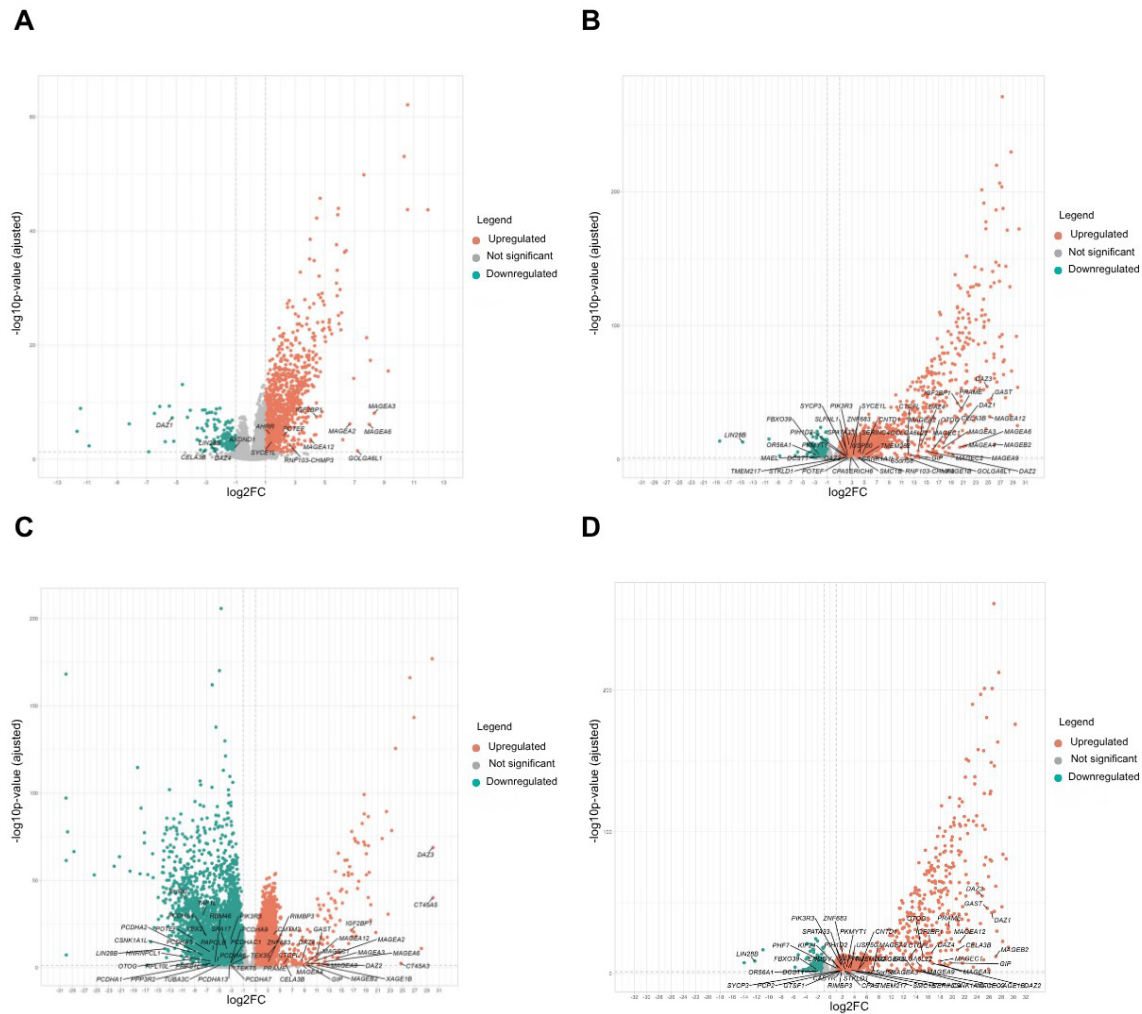


Figure 2. Volcano plot of differentially expressed genes between groups. (A) GAC vs. PTT. (B) GAC vs. MP. (C) GAC vs. Normal. (D) PTT vs. MP. The x-axis represents the L2FC, and the y-axis corresponds to the $-\log_{10}$ adjusted p -value. Genes in red are upregulated, those in green are downregulated, and those in gray are not significant. Abbreviations: GAC: Gastric adenocarcinoma; L2FC: Log2 fold change; MP: Metaplastic; PTT: Peritumoral tissue.

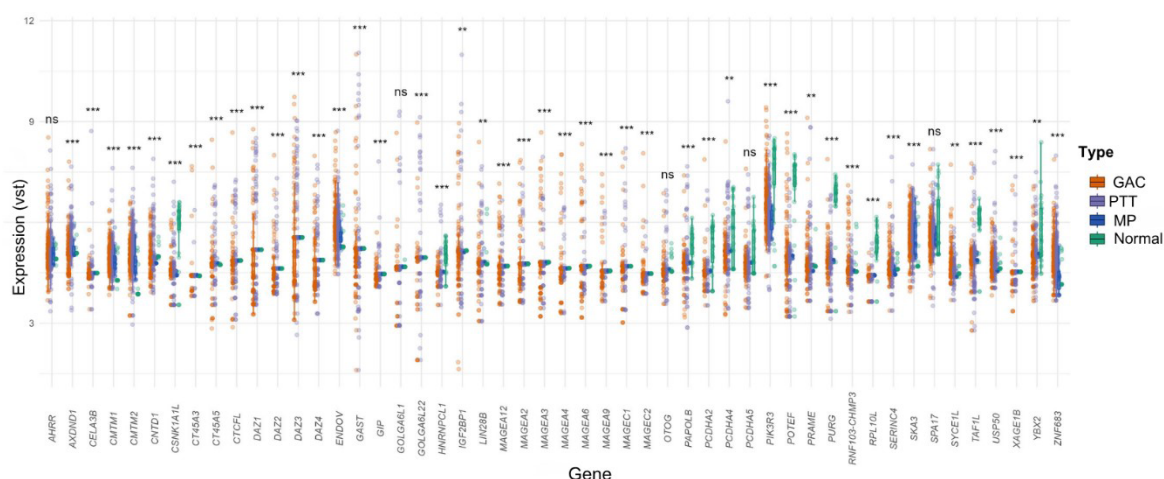


Figure 3. Global expression of CTA genes across different tissue types. The x-axis represents the selected CTAs, and the y-axis displays gene expression values transformed by VST. Each point corresponds to one sample from the GAC, PTT, and MP or normal group. Significance levels are indicated above each gene. Note: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Abbreviations: CTA: Cancer-testis antigen; GAC: Gastric adenocarcinoma; MP: Metaplastic; ns: Not significant; PTT: Peritumoral tissue; VST: Variance-stabilizing transformation.

3.2. Functional enrichment analysis

A GO analysis was performed using the top 15 differentially expressed and upregulated genes between GAC and PTT to identify the biological processes involved (Figure 4). The main enriched biological processes indicate the involvement of genes in embryonic morphogenesis and development, cellular differentiation, cell adhesion, and mesenchymal signatures, suggesting that genes upregulated in the tumor participate in processes associated with cellular reprogramming and the reactivation of embryonic-like programs.

3.3. Association of cancer-testis antigens with clinical variables

Gene expression analysis revealed distinct associations with clinically relevant variables in GAC. Initially, normalized gene expression was used to assess the relationship between DEGs and variables of interest using the DESeq2 package in R. This approach identified 29 genes with significant associations across different clinical parameters, including cTNM staging, T, N, and M categories, Lauren classification, therapeutic management, and sex. Additionally, pairwise contrast and analyses between clinical groups were performed (Figure S1).

Significant associations were observed for *POTEF* (FDR = 4.41×10^{-6} –0.00816), *MAGEA6* (FDR = 0.000395–0.00138), *MAGEA3* (FDR = 0.00158–0.00461), and *CTCF* (FDR = 0.0335) with the cTNM system. Additional candidates, including *PIK3R3*, *CMTM1*, *TAF1L*, *YBX2*, *AHRR*, *CMTM2*, *PAPOLB*, *PURG*, *PCDHA2/4/5*, and *AXDND1*,

also showed significant differences in expression across tumor stages. Pairwise comparisons revealed higher *CTCF* expression in stage IV, both between stages II and IV ($p = 0.0027$) and stages III and IV ($p = 0.0089$), suggesting a potential role for this gene in tumor progression.

Regarding tumor invasion depth (T category), genes from the MAGE family (*MAGEA2*, *MAGEA3*, *MAGEA6*, *MAGEA12*), *POTEF*, *PRAME*, *IGF2BP1*, *DAZ1/3/4*, *PCDHA4*, *GAST*, and *ZNF683* were significantly associated with this clinical variable (FDR < 0.05). Pairwise evaluation revealed increased *IGF2BP1* expression in patients classified as T2 when comparing T1 vs. T2 and T2 vs. T3, both with $p = 0.045$. For the ypN category of staging, *MAGEA3*, *MAGEA12*, *TAF1L*, *GAST*, *AHRR*, *GOLGA6L1*, *GOLGA6L22*, *PURG*, *AXDND1*, and *ENDOV* were identified as significant candidates associated with lymph node involvement after treatment.

In the evaluation of the ypM category, greater expression variability was observed in non-metastatic tumors (M0), whereas metastatic tumors (M1) exhibited more homogeneous expression. The analysis revealed strong associations for *GAST* (FDR = 3.16×10^{-8}), *POTEF* (FDR = 1.42×10^{-7}), *PIK3R3* (FDR = 0.000557), *YBX2* (FDR = 0.00123), *PCDHA4* (FDR = 0.00248), and MAGE genes (*MAGEA6*, *MAGEA12*), focusing on the role of these antigens in tumor dissemination. *PURG*, *TAF1L*, *CMTM2*, *RNF103-CHMP3*, and *DAZ3* were also significant.

Regarding Lauren classification, *PURG* (FDR = 0.000634), *POTEF* (FDR = 0.00824), *GOLGA6L22* (FDR =

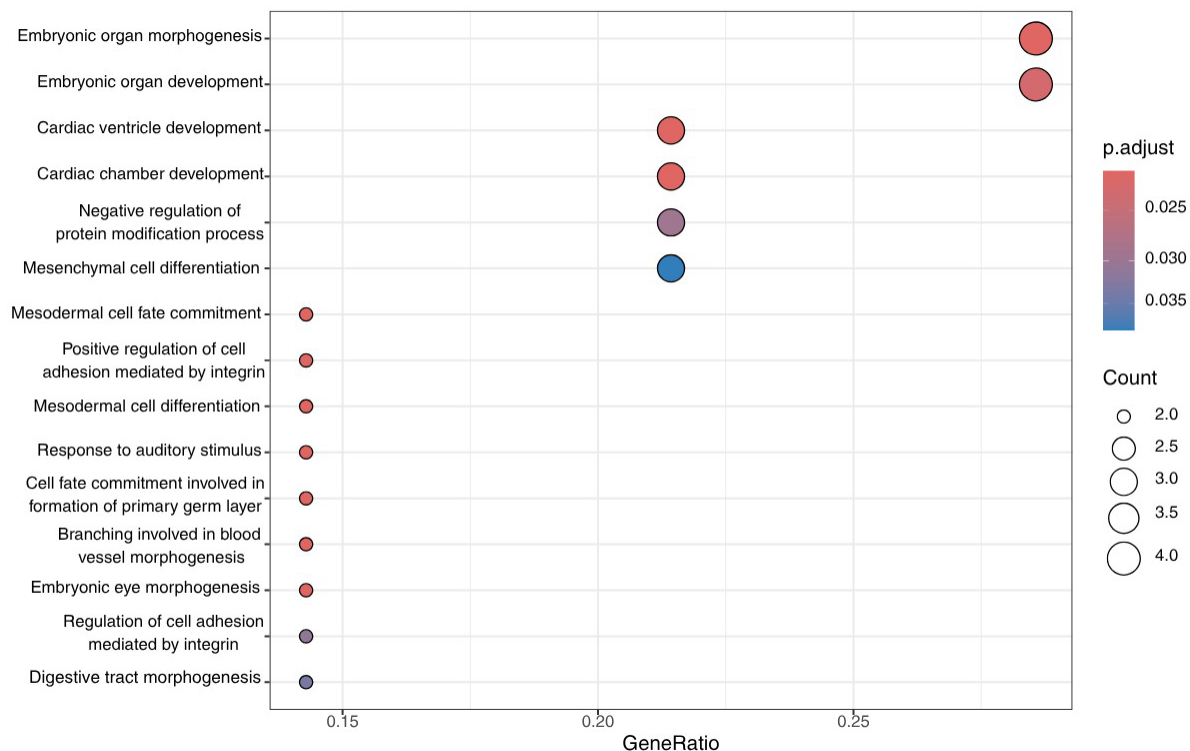


Figure 4. GO enrichment bubble plot of the top 15 upregulated differentially expressed genes identified between gastric adenocarcinoma and peritumoral tissue (FDR-adjusted $p < 0.05$). The x-axis represents the GeneRatio, the y-axis represents the GO terms, bubble size represents the number of associated genes (count), and color indicates the adjusted p -value.

Abbreviations: FDR: False discovery rate; GO: Gene Ontology; p .adjust: p -value adjust.

0.0189), and *PCDHA5* (FDR = 0.0466) showed significant differences in expression between diffuse and intestinal subtypes, indicating distinct molecular profiles between histological subtypes. In terms of therapeutic management (surgery vs. neoadjuvant therapy), *TAF1L* was the only gene showing a significant association (FDR = 0.00602). Finally, for sex related differences, strong associations were detected for members of the *DAZ* family, including *DAZ1* (FDR = 2.92^{-13}), *DAZ3* (FDR = 0.0452), and *DAZ4* (FDR = 0.00138), reflecting sex-linked expression patterns.

3.4. Survival analysis

In the survival analysis, 17 CTAs showed statistically significant differences in OS between high and low expression groups (log-rank test, $p < 0.05$) (Figure S2). Among these, only the gene *OTOG* did not reach statistical significance in the estimated risk of death using the Cox model, suggesting that its effect on risk is not consistent over time despite the differences observed between the curves. Of the 17 genes, high expression of 14 CTAs (82.4%) was associated with poorer outcomes in GAC patients, while only 3 CTAs (17.6%) showed high expression associated

with better prognosis in GAC.

The genes *DAZ3*, *GAST*, *CT45A5*, *MAGEA2*, *IGF2BP1*, *DAZ1*, *POTEF*, *LIN28B*, *MAGEC1*, and *CTCFL* stand out with HR above 2, indicating a significant increase in the risk of death in patients with high expression. Notably, high expression of *IGF2BP1* (HR = 2.86; $p = 0.004$), *CTCFL* (HR = 2.77; $p = 0.0039$), *CT45A5* (HR = 2.81; $p = 0.0033$), and *LIN28B* (HR = 2.72; $p = 0.0045$) showed the highest relative risks, reinforcing their role as potential markers of tumor aggressiveness. In contrast, *SYCE1L* (HR = 0.56; $p = 0.040$), *PIK3R3* (HR = 0.50; $p = 0.012$), and *ZNF683* (HR = 0.56; $p = 0.045$) were associated with improved OS.

3.5. Co-expression network analysis

The co-expression network constructed from differentially expressed CTAs revealed five main clusters, pointing out the presence of gene modules with similar expression patterns (Figure 5). Cluster 1 (blue) grouped genes from the *DAZ* family (*DAZ1*–*DAZ4*), which were strongly correlated with each other ($|r| = 0.64$ to 0.81), with *DAZ1*, *DAZ3*, and *DAZ4* as the main target genes. This cluster

also showed moderate positive correlations with the genes *AADAC* ($|\rho| = 0.48$ to 0.60), *FBP2* ($|\rho| = 0.36$ to 0.58), and *OVCH2* ($|\rho| = 0.43$ to 0.58).

Cluster 2 (purple) grouped genes from the *MAGEA* family, exhibiting strong correlations ($|\rho| = 0.30$ to 0.73), along with additional connections to *OPRPN* ($|\rho| = 0.40$ to 0.71), *OR2M7* ($|\rho| = 0.43$ to 0.70), and *CSAG3* ($|\rho| = 0.39$ to 0.79), forming a highly connected group. Additionally, a small number of genes showed a more dispersed distribution within the cluster or exhibited negative correlations. Among the target genes, *MAGEA2*, *MAGEA3*, *MAGEA6*, *MAGEA9*, *MAGEA12*, *IGF2BP1* ($|\rho| = 0.33$ to 0.48), and *CTCF* ($|\rho| = 0.30$ to 0.46) were prominent within this module. *GOLGA6L22* also showed

retained associations within the filtered network, with correlations reaching $|\rho| = 0.50$.

Cluster 3 (pink) involved a larger number of genes compared to the other clusters. Among the target genes, *YBX2*, *POTEF*, members of the *PCDHA* family (*PCDHA2*, *PCDHA4*, *PCDHA5*), *TAF1L*, *PAPOLB*, and *PURG* were consistently integrated into the filtered network and remained strongly interconnected, with retained correlations reaching $|\rho| = 0.57$. Other co-expressed genes, such as *CACNG7*, *NKAIN1*, *GRIN2B*, *KCNA4*, *ROBO2*, *TMEFF2*, and *SLC7A10*, also showed strong retained correlations within the cluster, reaching $|\rho| = 0.66$.

Cluster 4 (orange) was distributed across different subgroups. One set stood out in the network for exhibiting

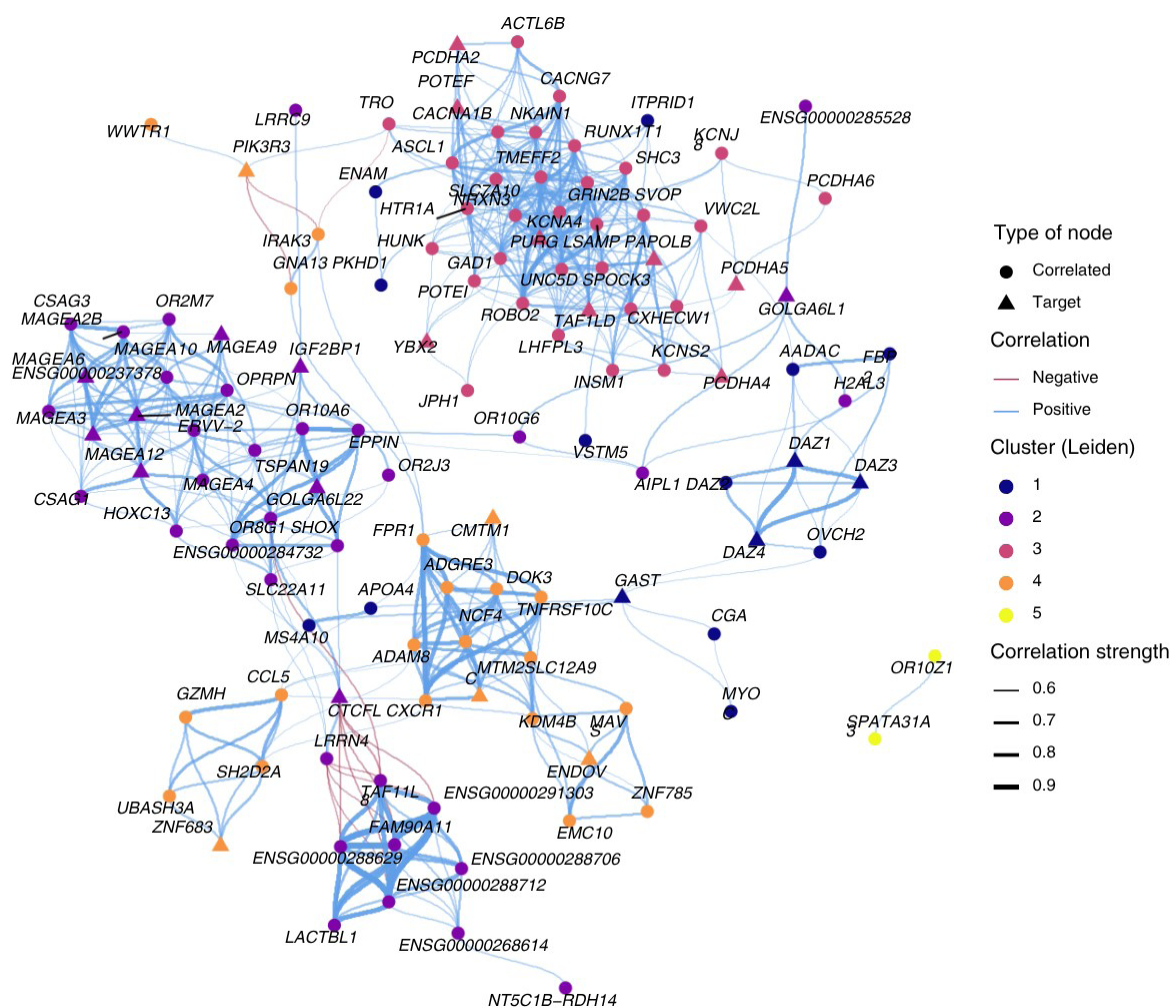


Figure 5. Gene correlation network constructed from differentially expressed cancer-testis antigens (target genes) and their correlated partners based on Spearman's correlation coefficient. Each node represents a gene, and edges indicate positive (blue) or negative (red) correlations, with thickness proportional to the correlation strength. Co-expression modules were detected using the Leiden algorithm and are highlighted in different colors (clusters 1 to 5).

strong positive correlations, linking *FPRI*, *ADGRE3*, *DOK3*, *TNFRSF10C*, *NCF4*, *ADAM8*, *CXCR1*, and *SLC12A9* as correlated genes, with *CMTM2* and *CMTM1* as target genes ($|p| = 0.48$ to 0.67). In another subgroup of the cluster, weak negative correlations were also observed among genes such as *IRAK3*, *GNA13*, and *PIK3R3* ($|p| = 0.56$ to 0.61). Finally, Cluster 5 (yellow) showed low connectivity, comprising only *SPATA31A3* and *OR10Z1* ($|p| = 0.55$).

4. Discussion

Gastric adenocarcinoma remains a major public health concern, and recent estimates indicate that global mortality from the disease will remain significant in the coming years.³¹ In this context, the identification of new therapeutic targets and biomarkers is essential to improve risk stratification and available treatment options.

The ideal antigen for cancer therapies should be tumor-specific, highly immunogenic, broadly expressed in tumor cells, shared among patients, and essential for tumor survival.³² CTAs satisfy most of these criteria, making them attractive candidates for immunotherapy.³³ However, for this approach to be translationally useful, it is necessary to comprehensively understand their expression patterns across different stages and clinical contexts of GAC.

Although recent studies have investigated the expression of these genes in GC, the use of restricted CTA lists for analysis is a limiting factor. Additionally, analyses comparing tumor tissue with PTT and precursor tissues can provide complementary information. Furthermore, evaluating expression in association with clinically relevant prognostic variables may offer new insights.

In this study, we identified 104 DEGs of CTAs across the analyzed contrasts. Between GAC and PTT, only 14 DEGs were observed, while the comparison between GAC and metaplasia revealed 47 DEGs. Additionally, a large proportion of genes upregulated in the tumor compared to MP tissue (23) were also upregulated in PTT relative to metaplasia. These findings suggest that PTT is more similar to tumor tissue than MP tissue, which can be explained by the field cancerization process, in which regions adjacent to the tumor already carry genetic and/or epigenetic alterations that make them biologically closer to the tumor and prone to developing new lesions or local recurrence.³⁴ The contrast between PTT and MP tissue resulted in 39 DEGs, highlighting a high divergence between these two tissue types.

The highest number of DEGs (81) was found in the comparison between tumor tissue and normal tissue, reflecting the characteristic molecular differences of each

condition. Most CTAs were upregulated in cancer, in agreement with previous literature. Additionally, a large proportion of genes downregulated in cancer are not classical CTAs and/or have been previously reported to be expressed in normal tissues.

In the comparison between GAC and PTT, among the 486 upregulated DEGs, four CTAs stood out among the 15 most abundant: *MAGEA3*, *MAGEA6*, *GOLGA6L1*, and *MAGEA2*. In addition, *IGF2BP1* and *MAGEA12* were also upregulated in the tumor. Except for *GOLGA6L1*, all other CTAs showed upregulation in GAC compared to all other tissues. Moreover, these genes were also elevated in PTT relative to MP, suggesting progressive epigenetic reactivation during gastric carcinogenesis, as previously described, and potentially serving as early biomarkers.

The GO enrichment analysis of the main upregulated genes in GAC highlights biological processes related to embryonic morphogenesis and development, mesenchymal differentiation, and integrin-mediated regulation of cell adhesion. These findings suggest the reactivation of mechanisms normally silenced in adult tissues, a pattern consistent with the expression of CTA genes and frequently observed in tumors.⁹ This profile may reflect increased cellular plasticity, remodeling of the tumor microenvironment, and the acquisition of features associated with invasion and progression of GC and other malignancies.³⁵⁻³⁸ In the context of CTAs, these results further support their association with mechanisms involved in tumor progression.

These results are consistent with the broadly available literature on the MAGE family, classified as classical CTAs. High expression of *MAGEA3* and *MAGEA6* has been associated with increased tumor proliferation and invasion, and they have also been proposed as potential prognostic biomarkers for poor outcomes in bladder cancer.³⁹

In our data, findings indicate that *MAGEA6* expression varies according to tumor stage and the presence of metastasis, suggesting potential involvement of this gene in cancer progression. However, the lack of clear visual differentiation between groups points to biological heterogeneity and a limited sample size, which precludes conclusions regarding specific expression patterns at each stage. Similar results were observed with other CTA genes, emphasizing the complexity of antigen regulation in tumors.

Analysis of *MAGEA3* also showed relevant variation in expression across cTNM stages, suggesting a potential association with clinical cancer progression and positioning it as a candidate of clinical relevance. Conversely, when focusing on individual components of staging, significant

differences were identified among T category groups, indicating an association between *MAGEA3* expression and the extent of the primary tumor. Similarly, differences were detected among N category groups, suggesting a relationship between gene expression and lymph node dissemination. These findings reinforce that *MAGEA3* showed significant variation across cTNM stages and was also associated with T and N categories, suggesting links with local invasion and nodal spread.

Supporting our findings, another study identified that *MAGEA3* is associated with lymph node metastasis and correlates with immune infiltration levels in GC.⁴⁰ Recently, *MAGEA3* was targeted by an mRNA vaccine using lipid nanoparticles against colorectal cancer, with preclinical data indicating prophylactic and therapeutic potential.⁴¹ *MAGEA2* has also emerged as a promising candidate for dendritic cell-based immunotherapy, promoting cytotoxic T lymphocyte activation against prostate cancer.⁴²

A study showed that the genes *MAGEA3* and *MAGEA12* were highly expressed in a subset of breast cancer cell lines. Depletion of *MAGEA12* resulted in the downregulation of other target genes, whereas its overexpression promoted increased cell migration and invasion.⁴³ Similarly, *MAGEA12* suppression reduced proliferation, migration, invasion, and tumor formation in cutaneous squamous cell carcinoma cells, and high expression correlated with poor prognosis.⁴⁴ In this context, future studies using similar approaches in GC cell lines could help elucidate how MAGE family genes influence tumor invasion and progression in GAC.

In our results, the analysis of *MAGEA2* indicated an association between gene expression and primary tumor staging, as well as a negative impact on OS, suggesting that higher *MAGEA2* levels may be linked to poorer prognosis. This finding is consistent with the hypothesis that *MAGEA2* contributes to tumor progression processes, possibly influencing tumor aggressiveness. Studies in renal cancer and glioma, for example, have reported an association between high *MAGEA2* expression and unfavorable prognosis.^{45,46}

MAGEA12 also showed significant variation across the T, N, and M staging components, with a notable association with distant metastasis (M1), indicating that its expression may be related to the tumor's capacity for dissemination. These findings reinforce the idea that different members of the MAGEA family may exert complementary roles in cancer progression, with specific impacts on local invasiveness, lymph node involvement, and metastatic potential. The disparity observed among the groups suggests that the expression of these genes is influenced by diverse biological factors and highlights the

need for additional studies to elucidate their functional roles in tumor progression.

Using single-cell RNA-seq, a study revealed that *MAGEA2*, *MAGEA3*, *MAGEA6*, and *MAGEA12* constitute a co-expressed subgroup, corroborating our results in which these genes exhibited similar expression patterns across contrasts, suggesting shared regulatory mechanisms.⁴⁷ Nevertheless, the study identified that within the same tumor, a subset of tumor cellular populations may not express any of these genes, indicating the intratumoral heterogeneity characteristic of CTAs. From a translational perspective, this co-expression reinforces the potential of therapeutic strategies that target multiple antigens to increase the fraction of tumor cells reached.

Similarly, the involvement of *IGF2BP1* has been described in several cancer types, including neuroblastoma, non-small cell lung cancer, and esophageal cancer.⁴⁸⁻⁵⁰ *IGF2BP1* functions as a post-transcriptional regulator in cancer cells, stabilizing pro-oncogenic mRNAs and contributing to cell proliferation, migration, invasion, and, in some cases, even metastasis. In addition, there is evidence that *IGF2BP1* influences the regulation of tumor immunogenicity, suggesting that this gene may represent a potential immunotherapeutic target in cancer.⁵¹

In the present analysis, we observed that *IGF2BP1* expression showed a significant association with clinically relevant variables, particularly the T category of the cTNM system, suggesting a possible relationship with tumor invasiveness. Consistently, elevated levels of *IGF2BP1* correlated with poorer OS, indicating that patients with high expression have an almost three-fold increased risk of death compared with those with low expression. These findings align with the known function of *IGF2BP1*, reinforcing the hypothesis that its overexpression contributes to a more aggressive phenotype. However, future studies employing approaches such as RNA immunoprecipitation sequencing could help elucidate which oncogenic transcripts are stabilized by *IGF2BP1* in GAC.

Conversely, *FBXO39* and *OR56A1* showed downregulation in GAC tissue and in PTT tissue compared with MP tissue. This pattern suggests that these genes are more active in tissues with premalignant alterations and become progressively silenced as tumor transformation occurs. Future validation studies, such as real-time quantitative polymerase chain reaction or immunohistochemistry, that confirm this gene expression profile may position them as potential biomarkers of progression, indicating the transition from metaplasia to tumor. These genes may also be useful for the early detection of tissues nearing malignant transformation during gastric carcinogenesis.

Although metaplasia is recognized as a precursor condition to cancer, to date, we have not found direct evidence of *FBXO39* expression in MP tissues. Most research has focused on *FBXO39* expression in tumor tissues, particularly in cancers such as breast, colon, and gliomas.⁵²⁻⁵⁴ However, some studies have examined this gene's expression by comparing tumor and adjacent normal tissues, though without a detailed analysis of the adjacent tissue itself.^{52,55} Therefore, further research is needed to fully understand the role of *FBXO39* in these contexts.

In the case of *OR56A1*, the available data are scarce. The current literature on this gene in the context of cancer is extremely limited, with no evidence supporting its expression in tumors, let alone in MP tissues. Most studies on olfactory receptors in cancer focus on other members of the family, such as *OR2B6* and *OR2W3*, which have been associated with tumor progression and may serve as biomarkers or therapeutic targets in various cancer types.⁵⁶

The analysis revealed that *POTEF* is upregulated in tumor tissue compared with PTT and MP tissues, and its expression was significantly associated with several clinically relevant variables, including cTNM stage, T category, M category, and Lauren histological subtype. Additionally, high *POTEF* expression was associated with poorer prognosis and reduced OS, indicating a potentially relevant role in the biology of GAC. These findings suggest that elevated *POTEF* expression within tumors may be linked to more aggressive phenotypes.

Supporting this notion, studies in prostate cancer have shown that *POTEF*-AS1, a long non-coding RNA derived from the *POTEF* locus, promotes cell survival.⁵⁷ Moreover, research on members of the *POTE* family has demonstrated that their high expression in ovarian cancer is associated with worse prognosis and increased tumor aggressiveness.⁵⁸ Future studies are needed to investigate the role of the *POTEF* gene in cancer, including the validation of its expression in tumor tissues through immunohistochemistry and functional assays in GC cell lines, in order to clarify whether this gene plays an active role in tumor aggressiveness and progression.

The genes *CTCF*, *GAST*, and *MAGEC1* exhibited the same expression pattern, with upregulation in GAC compared with MP tissue and normal tissue. Additionally, they were upregulated in PTT relative to MP tissue. This pattern indicates that the activation of these genes may contribute to the maintenance of the tumor phenotype and reflect molecular alterations within the field of cancerization. In the OS analysis, high expression of these genes was also associated with poorer prognosis in patients with GAC, conferring a two-fold increased risk of unfavorable outcomes.

CTCF showed heterogeneous expression across the cTNM stage groups; however, pairwise comparisons revealed higher expression in stage IV than in stages II and III, suggesting a role in tumor progression. Previous studies indicate that *CTCF* overexpression may contribute to carcinogenesis through modulation of chromatin architecture.⁵⁹ Taken together, our results suggest that *CTCF* may serve as a biomarker of advanced stage and prognosis in GC, as well as a potential therapeutic target for limiting tumor progression. Chromatin immunoprecipitation sequencing approaches may help clarify its regulatory role in GC.

In the literature, *MAGEC1* expression has been associated with a significantly higher number of lymph node metastases in melanoma, suggesting its role in metastatic progression.⁶⁰ Regarding *GAST*, a study reported that elevated serum gastrin levels correlate with more aggressive tumors, higher rates of lymph node metastasis, advanced clinical stages, and poorer OS in GC.⁶¹ Consistently, in our analysis, *GAST* gene expression was associated with multiple clinical characteristics, suggesting that its activation may accompany tumor advancement, lymph node dissemination, and metastasis occurrence, a hypothesis that could be further explored through functional assays in GC cell lines. Thus, at both the molecular and clinical levels, *GAST* has potential as a marker of tumor aggressiveness and prognosis in GC.

In the present study, most genes showed that high expression was associated with poorer prognosis. In contrast, *PIK3R3*, *SYCE1L*, and *ZNF683* were associated with a protective effect, suggesting the presence of a potentially more favorable immune microenvironment in these cases. These findings indicate that the expression profile of CTAs may hold prognostic value and aid in risk stratification. Moreover, several CTAs associated with worse outcomes have previously been implicated in oncogenic mechanisms. These mechanisms strengthen the hypothesis that the elevated expression of these genes is not merely a passive marker but may functionally contribute to tumor progression.

Gene expression of CTAs is often considered a consequence of epigenetic dysregulation during tumorigenesis and has traditionally been interpreted as a passenger event in cancer. However, growing evidence indicates that several CTAs may function as drivers, actively contributing to tumor progression through mechanisms such as the promotion of epithelial-mesenchymal transition, maintenance of pluripotency, genomic instability, and therapeutic resistance.^{62,63}

Despite the negative association observed in our cohort, independent data demonstrate that the prognostic

impact of CTAs is dependent on the biological and therapeutic context. In The Cancer Genome Atlas cohort, patients positive for *MAGEA1*, *MAGEA3*, *MAGEA4*, *MAGEA12*, and *CTAG1B* exhibited improved survival when treated with immune checkpoint inhibitors (ICIs), with similar trends reported for *GAGE13* and *SSX2*.⁶⁴ Although our analysis focused on prognostic associations, these findings suggest that CTAs may also have predictive value for therapeutic response, possibly due to their role in increasing tumor immunogenicity and promoting better responses to immunotherapy. However, in the absence of treatment, high expression of these genes may also be associated with more aggressive tumors. Thus, the dual nature of these genes is highlighted, reinforcing the need for studies integrating treatment response data to confirm their predictive potential in cancer, as well as interpretations that consider the molecular and therapeutic context in clinical practice.

The co-expression network analysis constructed from the differentially expressed CTAs revealed the existence of well-defined functional modules, suggesting that specific groups of these antigens share regulatory mechanisms and may participate in common biological processes within the tumor microenvironment. This organization is consistent with recent evidence showing that CTAs are often expressed as coordinated signatures, reflecting both epigenetic plasticity and the reactivation of germline programs in solid tumors.^{12,64,65}

The correlation analysis revealed five distinct clusters among the differentially expressed CTAs, well organized around central target genes. Cluster 1, composed of members of the DAZ family, indicates reactivation of germline programs in cancer.⁶⁶ Cluster 2, formed by the *MAGEA* genes, shows strong correlation with other regulatory CTAs, such as *IGF2BP1* and *CTCF*, suggesting a central role in proliferation and immune evasion.⁶⁷⁻⁶⁹ Cluster 3 includes genes associated with cell adhesion and transcriptional regulation, such as *YBX2*, *POTEF*, and members of the *PCDHA* family, which also act as target genes.^{70,71} Cluster 4 groups genes related to immune response, including *PIK3R3* and *ZNF683*, showing negative correlations with proliferative CTAs, suggesting a possible antagonism between tumor-promoting and inflammatory pathways.⁷²⁻⁷⁴ Finally, Cluster 5 displayed low connectivity, with *SPATA31A3* and *OR10Z1*, genes possibly associated with accessory functions. These findings highlight the functional organization of CTAs into correlation clusters, with potential implications for tumor regulation and immune response.

Taken together, these results reinforce the notion that CTAs are organized into co-expression signatures that

influence key processes such as germline reprogramming, immune evasion, cellular plasticity, and remodeling of the tumor microenvironment. Understanding these clusters not only aids in elucidating mechanisms of tumor progression but also contributes to the development of biomarkers and the identification of therapeutic targets based on CTA-directed immunotherapy.^{20,64,65}

Despite providing a comprehensive characterization of CTAs in GAC, several limitations should be acknowledged. The retrospective nature of the clinical data analysis and the lack of experimental validation restrict the generalizability of the findings, and the lack of available clinical information limits the robustness of the observed associations. Moreover, the absence of consensus regarding the classification of certain genes as CTAs may introduce bias. Another consideration is that co-expression analyses do not establish causal relationships, requiring additional functional validation. Survival analyses were performed using univariate models, and therefore the identified associations cannot be interpreted as independent prognostic effects. In addition, the study did not include immunogenomic analyses that could further contextualize CTA expression within the tumor immune microenvironment. Finally, the study assessed expression only at the transcriptional level, without accounting for protein abundance. These limitations highlight the need for future studies integrating experimental validation and multi-omics approaches to strengthen and expand the clinical relevance of the identified biomarkers.

5. Conclusion

In this study, we conducted the most comprehensive investigation to date of CTA expression profiles in GAC, integrating differential expression analyses, clinical associations, prognostic impact, and gene co-expression structure. Using a cohort of 389 samples, including tumor, PTT, metaplasia, and normal tissue, we identified 104 differentially expressed CTAs, highlighting a set of genes with potential relevance for tumor biology and clinical stratification.

We demonstrated that CTA expression in GAC is characterized by high heterogeneity yet organized into functional modules that reflect processes associated with tumor progression, invasion, metastasis, and aggressiveness. Consistently, 17 CTAs showed significant associations with overall survival: 14 were markers of poorer prognosis, whereas three (*PIK3R3*, *SYCE1L*, and *ZNF683*) exhibited a protective effect, possibly reflecting more responsive immune microenvironments. Additionally, external evidence indicates that some CTAs traditionally associated with aggressiveness may also predispose tumors

to better immunotherapy responses, reinforcing their dual nature and their potential as biomarkers and therapeutic targets, particularly in strategies such as tumor vaccines and CAR-T cell therapies.

The network analysis revealed highly connected co-expression modules, indicating that specific groups of CTAs may act in a coordinated manner during gastric carcinogenesis. These findings strengthen the notion that multigene panels, rather than isolated markers, offer greater precision for the biological and clinical characterization of the disease.

Taken together, our results substantially expand the current understanding of the CTA expression landscape in GC, identifying genes with strong prognostic value, robust clinical associations, and translational potential as immunotherapeutic targets. Future studies, including validation in independent cohorts, functional experiments, and assessment of the immunogenicity of these antigens, will be essential to advance the development of personalized therapies for GAC.

Acknowledgments

The authors thank the Núcleo de Pesquisas em Oncologia at the Universidade Federal do Pará (UFPA) for the technical support, infrastructure, and resources provided for the development of this study.

Funding

This research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil (Funding Code 001), and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (385189/2025-7), through the awarding of scholarships.

Conflict of interest

The authors declare that this research was conducted without any commercial or financial relationships that could be interpreted as potential conflicts of interest.

Author contributions

Conceptualization: Jade Dias Valente, Livia Érika Carlos Marques, André Salim Khayat, Paulo Pimentel de Assumpção, Rommel Mario Rodríguez Burbano

Data curation: Jade Dias Valente, Ronald Matheus da Silva Mourão, Fabiano Cordeiro Moreira, Bianca de Fátima dos Reis Rodrigues, Jéssica Manoelli Costa da Silva

Formal analysis: Jade Dias Valente, Ronald Matheus da Silva Mourão, Fabiano Cordeiro Moreira

Funding acquisition: Livia Érika Carlos Marques, Samia Demachki, Samir Mansour Moraes Casseb, Geraldo

Ishak, Paulo Pimentel de Assumpção, Rommel Mario Rodríguez Burbano

Investigation: Jade Dias Valente, Livia Érika Carlos Marques, Ronald Matheus da Silva Mourão, Fabiano Cordeiro Moreira, Bianca de Fátima dos Reis Rodrigues, Louise Sousa de Souza, Jéssica Manoelli Costa da Silva, Samir Mansour Moraes Moraes Casseb, Geraldo Ishak

Methodology: Jade Dias Valente, Livia Érika Carlos Marques, Ronald Matheus da Silva Mourão, Fabiano Cordeiro Moreira, Jéssica Manoelli Costa da Silva, Samir Mansour Moraes Casseb

Project administration: Livia Érika Carlos Marques, Paulo Pimentel de Assumpção, Rommel Mario Rodríguez Burbano

Resources: Livia Érika Carlos Marques, Samir Mansour Moraes Moraes Casseb, André Salim Khayat, Samia Demachki, Geraldo Ishak, Fabiano Cordeiro Moreira, Mario Rodríguez Burbano, Paulo Pimentel de Assumpção

Supervision: Livia Érika Carlos Marques, Paulo Pimentel de Assumpção, Rommel Mario Rodríguez Burbano

Visualization: Jade Dias Valente, Livia Érika Carlos Marques, Ronald Matheus da Silva Mourão, Fabiano Cordeiro Moreira

Writing—original draft: Jade Dias Valente, Bianca de Fátima dos Reis Rodrigues, Louise Sousa de Souza

Writing—review & editing: All authors

Ethics approval and consent to participate

This study was approved by the Research Ethics Committee of the Núcleo de Pesquisas em Oncologia (Approval No. 47580121.9.0000.5634), which authorized the analyses and procedures described. The committee specifically approved the collection of biological samples and the review of medical records. As the study involved human subjects, all relevant ethical procedures were followed. All participants provided written informed consent, agreeing to take part in the research in accordance with current ethical guidelines.

Consent for publication

All participants included in this study provided written informed consent for the publication of their research data. The study did not involve the publication of participant images. In addition, no individually identifiable information was disclosed, and all sensitive data were appropriately withheld to ensure confidentiality.

Availability of data

The data generated and analyzed in this study include sensitive clinical information derived from patient medical

records and biological samples, which are subject to restrictions imposed by the Research Ethics Committee. Therefore, these data cannot be made publicly available. Anonymized data may be provided upon reasonable request to the corresponding author.

Further disclosure

This manuscript was previously deposited as a preprint on the bioRxiv server under the doi: 10.1101/2025.09.17.676427. The content of this submission corresponds to the revised version of the material originally made available in the preprint.

References

1. Song H, Ferlay J, Siegel RL, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249.
doi: 10.3322/caac.21660
2. Brenner H. Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. *Lancet.* 2002;360(9340):1131-1135.
doi: 10.1016/S0140-6736(02)11199-8
3. Conti CB, Agnesi S, Scaravaglio M, *et al.* Early gastric cancer: update on prevention, diagnosis and treatment. *Int J Environ Res Public Health.* 2023;20:2149.
doi: 10.3390/ijerph20032149
4. Janjigian YY, Al-Batran S, Wainberg ZA, *et al.* Perioperative durvalumab in gastric and gastroesophageal junction cancer. *N Engl J Med.* 2025;393:217-230.
doi: 10.1056/NEJMoa2503701
5. Liu B, Zhou HY, Tan L, Siu PH, Guan XY. Exploring treatment options in cancer: tumor treatment strategies. *Signal Transduct Target Ther.* 2024;9:175.
doi: 10.1038/s41392-024-01856-7
6. Jiang L, Lau HC, Zeng R, Yu J. Diet, Gastric Microbiota, and Metabolites in Gastric Tumorigenesis. *Research.* 2025;8(0693).
doi: 10.34133/research.0693
7. Wu Y, Jiang XY, Qing H, *et al.* Targeting the ANXA8-SP1-PPA1 Axis to Modulate TCA Cycle and Matrix Deposition in Diffuse-Type Gastric Cancer. *Research.* 2025;8(0838).
doi: 10.34133/research.0838
8. Shim K, Jo H, Jeoung D. Cancer/testis antigens as targets for RNA-based anticancer therapy. *Int J Mol Sci.* 2023;24(14679).
doi: 10.3390/ijms241914679
9. Gibbs Z, Whitehurst AW. Emerging contributions of cancer/testis antigens to neoplastic behaviors. *Trends Cancer.* 2018;4(10):701-712.
doi: 10.1016/j.trecan.2018.08.005
10. Wu Q, Zhang W, Wang Y, *et al.* MAGE-C3 promotes cancer metastasis by inducing epithelial-mesenchymal transition and immunosuppression in esophageal squamous cell carcinoma. *Cancer Commun.* 2021;41:1354-1372.
doi: 10.1002/cac2.12203
11. Maxfield KE, Taus PJ, Corcoran K, *et al.* Comprehensive functional characterization of cancer-testis antigens defines obligate participation in multiple hallmarks of cancer. *Nat Commun.* 2015;6:8840.
doi: 10.1038/ncomms9840
12. Ai H, Yang H, Li L, Ma J, Liu K, Li Z. Cancer testis antigens: promising immunotherapy target for digestive tract cancers. *Front Immunol.* 2023;14:1190883.
doi: 10.3389/fimmu.2023.1190883
13. Fijak M, Meinhardt A. The testis in immune privilege. *Immunol Rev.* 2006;213:66-81.
doi: 10.1111/j.1600-065X.2006.00438.x
14. Fratta E, Coral S, Covre A, *et al.* The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Mol Oncol.* 2011;5(2):164-182.
doi: 10.1016/j.molonc.2011.02.001
15. Hong DS, Van Tine BA, Olszanski AJ, *et al.* Phase I dose escalation and expansion trial to assess the safety and efficacy of ADP-A2M4 SPEAR T cells in advanced solid tumors. *J Clin Oncol.* 2020;38(15):102.
doi: 10.1200/JCO.2020.38.15_suppl.102
16. Liu Q, Qian H, Shao J, *et al.* Abstract CT237: A phase I study to assess the safety, tolerability, and immunogenicity of personalized neoantigen/cancer testis antigen nanovaccine (PVAC) for high-risk resected gastric cancer. *Cancer Res.* 2020;80(16).
doi: 10.1158/1538-7445.AM2020-CT237
17. Hong DS, Van Tine BA, Biswas S, *et al.* Autologous T-cell therapy for MAGE-A4+ solid cancers in HLA-A*02+ patients: a phase 1 trial. *Nat Med.* 2023;29:104-114.
doi: 10.1038/s41591-022-02128-z
18. Almeida LG, Sakabe NJ, Oliveira AR, *et al.* CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res.* 2009;37:D816-D819.
doi: 10.1093/nar/gkn673
19. Silva VL, Fonseca AF, Fonseca M, *et al.* Genome-wide identification of cancer/testis genes and their association with prognosis in a pan-cancer analysis. *Oncotarget.* 2017;8(54):92966-92977.

- doi: 10.18632/oncotarget.21715
20. Ansari S, Nikpour P. Identification of cancer/testis antigens related to gastric cancer prognosis based on co-expression network and integrated transcriptome analyses. *Adv Biomed Res.* 2023;12:52.
doi: 10.4103/abr.abr_400_21
 21. Bian G, Cao J, Li W, *et al.* Identification and validation of a cancer-testis antigen-related signature to predict the prognosis in stomach adenocarcinoma. *J Cancer.* 2024;15(11).
doi: 10.7150/jca.91842
 22. Menyhárt O, Györfy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis and diagnosis. *Comput Struct Biotechnol J.* 2021;19:949-960.
doi: 10.1016/j.csbj.2021.01.009
 23. Yang L, Wang J, Altreuter J, *et al.* Tutorial: integrative computational analogy of bulk RNA-sequencing data to characterize tumor immunity using RIMA. *Nat Protoc.* 2023;18:2404-2414.
doi: 10.1038/s41596-023-00841-8
 24. Hashimoto T, Nakamura Y, Mishima S, *et al.* Whole-transcriptome sequencing in advanced gastric or gastroesophageal cancer: a deep dive into its clinical potential. *Cancer Sci.* 2024;115:1622-1633.
doi: 10.1111/cas.16109
 25. Ewels PA, Peltzer A, Fillinger S, *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol.* 2020;38(3):276-278.
doi: 10.1038/s41587-020-0439-x
 26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120.
doi: 10.1093/bioinformatics/btu170
 27. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: a method for quantifying transcript abundance from RNA-seq reads. *Nat Methods.* 2017;14(4):417-419.
doi: 10.1038/nmeth.4197
 28. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: <https://www.r-project.org/> [Last accessed on September 12, 2025].
 29. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 2015;4:1521.
doi: 10.12688/f1000research.7563.2
 30. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
doi: 10.1186/s13059-014-0550-8
 31. Lin JL, Lin JX, Lin GT, *et al.* Global incidence and mortality trends of gastric cancer and predicted mortality of gastric cancer by 2035. *BMC Public Health.* 2024;24:1763.
doi: 10.1186/s12889-024-19104-6
 32. Cheever MA, Allison JP, Ferris AS, *et al.* The prioritization of cancer antigens: a National Cancer Institute pilot project for the acceleration of translational research. *Clin Cancer Res.* 2009;15(17):5323-5337.
doi: 10.1158/1078-0432.CCR-09-0737
 33. Li XF, Ren P, Shen WZ, Jin X, Zhang J. The expression, modulation and use of cancer-testis antigens as potential biomarkers for cancer immunotherapy. *Am J Transl Res.* 2020;12(11):7002-7019.
 34. Braakhuis BJ, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res.* 2003;63(8):1727-1730.
 35. Cheng W, Li HL, Xi SY, *et al.* Growth differentiation factor 1-induced tumour plasticity provides a therapeutic window for immunotherapy in hepatocellular carcinoma. *Nat Commun.* 2021;12:7142.
doi: 10.1038/s41467-021-27525-9
 36. Fazilaty H, Basler K. Reactivation of embryonic genetic programs in tissue regeneration and disease. *Nat Genet.* 2023;55:1792-1806.
doi: 10.1038/s41588-023-01526-4
 37. Hamidi H, Ivaska J. Every step of the way: integrins in cancer progression and metastasis. *Nat Rev Cancer.* 2018;18:533-548.
doi: 10.1038/s41568-018-0038-z
 38. Ullah A, Xing H, Wang X, *et al.* Andrographolide Promotes Ferroptosis in Pancreatic Cancer via Targeting and Activating HSP90/GPX4 Ubiquitination. *Biofactors.* 2025;51(6).
doi: 10.1002/biof.70063
 39. Verma S, Swain D, Kushwaha PP, *et al.* Melanoma antigen family A (MAGE-A) as promising biomarkers and therapeutic targets in bladder cancer. *Cancers (Basel).* 2024;16(2):246.
doi: 10.3390/cancers16020246
 40. Jin J, Tu J, Ren J, *et al.* Comprehensive analysis to identify MAGEA3 expression correlated with immune infiltrates and lymph node metastasis in gastric cancer. *Front Oncol.* 2021;11:784925.
doi: 10.3389/fonc.2021.784925
 41. Choi K, Jeong H, Lee JW, *et al.* Innovative cancer immunotherapy with MAGE-A3 mRNA cancer vaccines. *Cancers (Basel).* 2024;16(19):3428.

- doi: 10.3390/cancers16193428
42. Bakhshi P, Nourizadeh M, Sharifi L, Farajollahi MM, Mohsenzadegan M. Development of dendritic cell-loaded MAGE-A2 long peptide: a potential target for tumor-specific T cell-mediated prostate cancer immunotherapy. *Cancer Cell Int.* 2023;23(1).
doi: 10.1186/s12935-023-03108-0
 43. Oh C, Kim HR, Oh S, *et al.* Epigenetic upregulation of MAGE-A isoforms promotes breast cancer cell aggressiveness. *Cancers.* 2021;13(13):3176.
doi: 10.3390/cancers13133176
 44. Zhao G, Bae JY, Zheng Z, *et al.* Overexpression and implications of melanoma-associated antigen A12 in pathogenesis of human cutaneous squamous cell carcinoma. *Anticancer Res.* 2019;39(4):1849-1857.
doi: 10.21873/anticancer.13292
 45. Mohsenzadegan M, Fattahi F, Kalantari E, Abolhasani M, Madjd Z, Zanjani LS. Overexpression of MAGE-A2 is related to the malignant degree and progression of disease in patients with clear cell renal cell carcinoma. *Appl Immunohistochem Mol Morphol.* 2025;33(3):188-194.
doi: 10.1097/PAI.0000000000001243
 46. Meng Q, Luo G, Liu B, Sun Y. Melanoma-associated antigen A2 is overexpressed in glioma and associated with poor prognosis in glioma patients. *Neoplasma.* 2018;65(4):604-609.
doi: 10.4149/neo_2018_170625N440
 47. Traynor S, Jakobsen MK, Green TM, *et al.* Single-cell sequencing unveils extensive intratumoral heterogeneity of cancer/testis antigen expression in melanoma and lung cancer. *J Immunother Cancer.* 2024;12(6).
doi: 10.1136/jitc-2023-008759
 48. Sun M, Wang L, Ge L, Xu D, Zhang R. IGF2BP1 facilitates non-small cell lung cancer progression by regulating the KIF2A-mediated Wnt/ β -catenin pathway. *Funct Integr Genomics.* 2024;24(1):4.
doi: 10.1007/s10142-023-01275-x
 49. Hagemann S, Misiak D, Bell JL, *et al.* IGF2BP1 induces neuroblastoma via a druggable feedforward loop with MYCN promoting 17q oncogene expression. *Mol Cancer.* 2023;22:88.
doi: 10.1186/s12943-023-01792-0
 50. Wang JJ, Chen DX, Zhang Y, *et al.* Elevated expression of the RNA-binding protein IGF2BP1 enhances the mRNA stability of INHBA to promote invasion and migration of esophageal squamous cancer cells. *Exp Hematol Oncol.* 2023;12:75.
doi: 10.1186/s40164-023-00429-8
 51. Elcheva IA, Gowda CP, Bogush D, *et al.* IGF2BP family of RNA-binding proteins regulate innate and adaptive immune responses in cancer cells and tumor microenvironment. *Front Immunol.* 2023;14:1224516.
doi: 10.3389/fimmu.2023.1224516
 52. Seifi-Alan M, Shamsi R, Ghafouri-Fard S, Mirfakhraie R, Zare-Abdollahi D, Movafagh A. Expression analysis of two cancer-testis genes, FBXO39 and TDRD4, in breast cancer tissues and cell lines. *Asian Pac J Cancer Prev.* 2014;14(11):6625-6629.
doi: 10.7314/APJCP.2013.14.11.6625
 53. Ramadan RA, Morsi H, Samir M, Shamseya M, Shalaby M, El-Attar E. Significance of cancer testis-associated antigens (SPAG9 and FBXO39) in colon cancer. *Indian J Cancer.* 2022;59(3):394-401.
doi: 10.4103/ijc.IJC_873_19
 54. Wu J, Yao F, Li Y, *et al.* The cancer-testis antigen FBXO39 predicts poor prognosis and is associated with stemness and aggressiveness in glioma. *Pathol Res Pract.* 2022;239:154168.
doi: 10.1016/j.prp.2022.154168
 55. Motalebzadeh J, Shabani S, Rezayati S, *et al.* Prognostic value of FBXO39 and ETS-1 but not BMI-1 in Iranian colorectal cancer patients. *Asian Pac J Cancer Prev.* 2018;19(5):1357-1362.
doi: 10.22034/APJCP.2018.19.5.1357
 56. Masjedi S, Zwiebel LJ, Giorgio TD. Olfactory receptor gene abundance in invasive breast carcinoma. *Sci Rep.* 2019;9:13736.
doi: 10.1038/s41598-019-50085-4
 57. Misawa A, Takayama K, Fujimura T, Homma Y, Suzuki Y, Inoue S. Androgen-induced lncRNA POTE-AS1 regulates apoptosis-related pathway to facilitate cell survival in prostate cancer cells. *Cancer Sci.* 2017;108(3):373-379.
doi: 10.1111/cas.13151
 58. Barger CJ, Zhang W, Sharma A, *et al.* Expression of the POTE gene family in human ovarian cancer. *Sci Rep.* 2018;8:17136.
doi: 10.1038/s41598-018-35567-1
 59. Tong X, Gao Y, Su Z. Interaction of CTCF and CTCFL in genome regulation through chromatin architecture during spermatogenesis and carcinogenesis. *PeerJ.* 2024;12:18240.
doi: 10.7717/peerj.18240
 60. Lazăr AD, Dinescu S, Sleiman L, Dumitru AV, Costache M, Costache M. Comparative expression profiling reveals molecular markers involved in the progression of cutaneous melanoma towards metastasis. *Int J Mol Sci.* 2023;24(7):6565.
doi: 10.3390/ijms24076565
 61. Huang JW, Lin C, Lu CJ, Wang HS, Zou DD. Serum gastrin

- and *Helicobacter pylori* infection correlate with tumor aggressiveness and prognosis in gastric cancer. *World J Gastrointest Oncol.* 2025;17(9):110981.
doi: 10.4251/wjgo.v17.i9.110981
62. Yang P, Huo Z, Liao H, Zhou Q. Cancer/testis antigens trigger epithelial-mesenchymal transition and genesis of cancer stem-like cells. *Curr Pharm Des.* 2015;21(10):1292-1300.
doi: 10.2174/1381612821666141211154707
 63. Vaziri C, Forker K, Zhang X, Wu D, Zhou P, Bowser JL. Pathological modulation of genome maintenance by cancer/testes antigens (CTAs). *DNA Repair (Amst).* 2025;147:103818.
doi: 10.1016/j.dnarep.2025.103818
 64. Kalvapudi S, Pachimatla AG, Seager RJ, Conroy J, Pabla S, Mukherjee S. Cancer/testis antigen expression and co-expression patterns in gastroesophageal adenocarcinoma. *Med Oncol.* 2024;41(9).
doi: 10.1007/s12032-024-02475-6
 65. Seager RJ, Senosain MF, Roey EV, *et al.* Cancer testis antigen burden (CTAB): a novel biomarker of tumor-associated antigens in lung cancer. *J Transl Med.* 2024;22:141.
doi: 10.1186/s12967-024-04918-0
 66. Fu XF, Cheng SF, Wang LQ, Yin S, Felici M, Shen W. DAZ family proteins: key players for germ cell development. *Int J Biol Sci.* 2015;11(10):1226-1235.
doi: 10.7150/ijbs.11536
 67. Peng W, Ye L, Xue Q, *et al.* Silencing of circCRIM1 drives IGF2BP1-mediated immune evasion in non-small cell lung cancer. *Cells.* 2023;12(2):273.
doi: 10.3390/cells12020273
 68. Haase J, Misiak D, Bauer M, *et al.* IGF2BP1 is the first positive marker for the diagnosis of anaplastic thyroid carcinoma. *Mod Pathol.* 2021;34(1):32-41.
doi: 10.1038/s41379-020-0630-0
 69. Ye X, Lin J, Chen Y, Wang X. IGF2BP1 accelerates aerobic glycolysis to enhance its immune evasion in the hepatocellular carcinoma microenvironment. *Front Immunol.* 2024;15:1480834.
doi: 10.3389/fimmu.2024.1480834
 70. Du J, Wu K. PCDHA1 high expression is associated with poor prognosis and correlated with immune cell infiltration in breast cancer. *Clin Breast Cancer.* 2023;23(4):397-407.
doi: 10.1016/j.clbc.2023.02.001
 71. Yuan Z, Li B, Liao W, *et al.* A comprehensive pan-cancer analysis of the YBX family reveals YBX2 as a potential biomarker in liver cancer. *Front Immunol.* 2024;15:1382520.
doi: 10.3389/fimmu.2024.1382520
 72. Kitakaze M, Uemura M, Hara T, *et al.* Cancer-specific tissue-resident memory T cells express ZNF683 in colorectal cancer. *Br J Cancer.* 2023;128(10):1828-1837.
doi: 10.1038/s41416-023-02202-4
 73. Chen Q, Sun X, Luo X, Wang J, Hu J, Feng Y. PIK3R3 inhibits cellular senescence through p53/p21 signaling. *Cell Death Dis.* 2020;11:798.
doi: 10.1038/s41419-020-02921-z
 74. Zhu S, Ma AH, Zhu Z, *et al.* Synergistic antitumor activity of pan-PI3K inhibition and immune checkpoint blockade in bladder cancer. *J Immunother Cancer.* 2021;9(11):e002917.
doi: 10.1136/jitc-2021-002917