Journal of
Personalized Medicine

MDPI

Article

Differentially Expressed Genes Identify FIGO Stage II Cervical Cancer Patients with a Higher Risk of Relapse in a Small Cohort

Carolina P. S. Melo [1], Angelo B. Melo [2], Fábio R. Queiroz [1], Álvaro P. Costa [3], Laurence R. Amaral [2], Ramon A. Pereira [1], Izabela F. G. Amorim [1], Jorge G. G. Ferreira [1], Wander J. Jeremias [4], Pedro L. L. Bertarini [2], Matheus S. Gomes [2], Letícia C. Braga [1,*] and Paulo G. O. Salles [1,*]

[1] Laboratory of Translational Research in Oncology, Teaching, Research and Innovation Center, Mario Penna Institute, Belo Horizonte 30380-490, MG, Brazil; carolina.melo@mariopenna.org.br (C.P.S.M.); fabio.queiroz@mariopenna.org.br (F.R.Q.); ramon.pereira@mariopenna.org.br (R.A.P.); izabela.amorim@professor.faminas.edu.br (I.F.G.A.); jorge.ferreira@mariopenna.org.br (J.G.G.F.)
[2] Laboratory of Bioinformatics and Molecular Analysis, Federal University of Uberlandia, Patos de Minas 38702-178, MG, Brazil; angelo.melo@ufu.br (A.B.M.); laurence@ufu.br (L.R.A.); bertarini@ufu.br (P.L.L.B.); matheusgomes@ufu.br (M.S.G.)
[3] Postgraduate Program in Sciences Applied to Surgery and Ophthalmology, Faculty of Medicine, Federal University of Minas Gerais, Belo Horizonte 30130-100, MG, Brazil; alvaro.costa@mariopenna.org.br
[4] Laboratory of Experimental Pharmacology, School of Pharmacy, Federal University of Ouro Preto, Ouro Preto 35402-163, MG, Brazil; wander.jeremias@ufop.edu.br
* Correspondence: leticia.braga@mariopenna.org.br (L.C.B.); paulo.salles@mariopenna.org.br (P.G.O.S.)

Abstract

Background/Objectives: Most studies investigating prognostic biomarkers in cervical cancer (CC) analyze patients irrespective of FIGO stage, potentially masking molecular features that underlie the aggressiveness of some FIGO II tumors. To address this, we investigated differential gene expression in a FIGO II CC cohort to identify a gene signature predictive of progression-free survival (PFS) within five years of treatment initiation. **Methods**: Tumor samples from 15 CC patients were analyzed using RNA sequencing, bioinformatics, and machine learning to identify differentially expressed genes (DEGs) associated with prognosis. Findings were validated in an independent CC cohort (n = 174). **Results**: High expression of B3GALT1 (HR = 5.11), GTF3C2-AS1 (HR = 18.73), and ZKSCAN4 (HR = 5.18) was significantly associated with an increased risk of recurrence in our cohort. Elevated expression of these transcripts is also associated with shorter PFS in the external dataset. Notably, GTF3C2-AS1 expression alone was sufficient to classify all fifteen patients into their respective prognostic groups using a decision tree model, achieving 93.3% accuracy in leave-one-out cross-validation (LOOCV). Additional candidates, including RCAN2-DT, MYH9-DT, IGKC, IGHG1, and IGHG3, were associated with PFS in our cohort but could not be externally validated due to a lack of available data. **Conclusions**: Transcriptomic profiling revealed potential biomarkers that refine prognostic stratification in cervical cancer beyond FIGO staging. Among them, GTF3C2-AS1 consistently emerged as a potential predictor of recurrence risk. Additional candidates, including B3GALT1, ZKSCAN4, and immunoglobulin transcripts, provided complementary insights but require further validation. These preliminary results highlight intra-stage heterogeneity in FIGO II CC and underscore the promise of molecular markers to improve risk assessment.

Keywords: cervical cancer prognosis; gene expression signature; FIGO staging; machine learning model; cervical cancer recurrence; artificial intelligence for oncology; network-based selection method

check for updates

Academic Editors: Daniela Calina, Luca Falzone, Himanshu Arora and Luigi Della Corte

Received: 7 July 2025
Revised: 29 August 2025
Accepted: 13 October 2025
Published: 16 October 2025

Citation: Melo, C.P.S.; Melo, A.B.; Queiroz, F.R.; Costa, Á.P.; Amaral, L.R.; Pereira, R.A.; Amorim, I.F.G.; Ferreira, J.G.G.; Jeremias, W.J.; Bertarini, P.L.L.; et al. Differentially Expressed Genes Identify FIGO Stage II Cervical Cancer Patients with a Higher Risk of Relapse in a Small Cohort. *J. Pers. Med.* **2025**, *15*, 497. https://doi.org/10.3390/jpm15100497

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cervical cancer (CC) ranks as one of the most lethal types of cancer, precisely the fourth leading cause of death among females globally. Significant investments have been made in CC diagnosis and treatment, including the development of vaccines, highlighting economic disparities among countries. In 2022, about 94% of CC-related deaths occurred in less developed countries [1].

Diagnoses have increased with the expansion of CC screening programs, underscoring the urgent need to enhance disease stratification strategies. The International Federation of Gynecology and Obstetrics (FIGO) staging system is the predominant method used worldwide for evaluating prognosis in CC patients. Its latest update in 2018 incorporated imaging techniques and pathological features, such as lymph node involvement and tumor size, to improve prognostic outcomes' differentiation and effectively guide clinical management [2]. A key advantage of the updated FIGO staging system is its flexibility in utilizing clinical, radiological, or pathological findings to determine stage assignment, an approach more inclusive, especially when considering the limited resources of less developed countries [3–6].

FIGO stage II in CC is characterized by the carcinoma invasion beyond the uterus into the surrounding tissues, but without extension to the lower third of the vagina or the pelvic wall [7]. It is subdivided into IIA and IIB, depending on the region where the carcinoma has spread. Although with a relatively good prognosis compared to stages III and IV, the estimated 5-year survival rate is around 66% for IIB patients [2], exposing an intragroup heterogeneity. Considering the inherent complexity of tumors, CC molecular features may be responsible for the observed differences in outcomes, and their identification could improve treatment stratification and clinical management for CC patients.

The concept of using molecular markers for patient stratification in CC is not new. With the increasing prevalence of studies employing next-generation sequencing (NGS), there has been substantial progress in understanding the influence of molecular features on aspects ranging from disease susceptibility to progression and survival, facilitating the identification of potential biomarkers and therapeutic targets [8]. In recent years, multiple prognostic biomarkers for CC have been described; however, their clinical application remains limited due to insufficient validation in large cohorts or difficulty in implementation. Su et al. reported an association between Zac1 expression and poor prognosis in CC [9]. Zhao et al. developed and validated an expression-based prognostic signature comprising SPP1, EFNA1, MMP1, ITM2A, and DSG2 genes as an independent predictor of CC survival [10]. Similarly, Ju et al. identified a five-mRNA (GALNTL6, ARSE, DPAGT1, GANAB, and FURIN) prognostic signature related to post-translational modifications for predicting both disease-free and overall survival in CC [11].

The studies describing CC prognostic biomarkers often group patients regardless of their FIGO stage. One of the few exceptions is the study of Nguyen et al., who proposed a 70-gene signature for predicting treatment outcomes comparing early-stage (stage I–IIA) and advanced-stage (stage IIB–IV) CC [12]. Although often grouped as advanced-stage disease, FIGO stage III and IV CC are characterized by more extensive disease, with an expected poorer prognosis, whereas FIGO stage II is more heterogeneous. Stage IIB in particular is considered locally advanced; however, some patients may still experience prolonged progression-free survival [13]. Thus, a biomarker able to identify, at diagnosis, FIGO II patients who will present shorter progression-free survival (PFS) after treatment could be used as an additional tool in the management of high-risk patients.

To address this hypothesis, this study investigated the differential gene expression in a FIGO II CC cohort between patients with good and poor prognoses. We first profiled the differentially expressed genes (DEGs) of FIGO II CC patients from our cohort, based

on which genes were up- or downregulated in patients with recurrent and non-recurrent disease after treatment. Subsequently, we evaluated the impact of these genes on the hazard ratio (HR) for PFS. Furthermore, a machine learning (ML) analysis was performed using normalized gene expression counts to identify potential biomarker candidates for improved stratification of FIGO II CC patients with distinct prognoses. The prognostic value of these candidate biomarkers was further validated using data from The Cancer Genome Atlas (TCGA) cohort. Our findings provide novel insights into the molecular heterogeneity underlying FIGO stage II CC and may support clinicians in identifying high-risk patients warranting closer clinical surveillance.

2. Materials and Methods

2.1. Patient Recruitment and Sample Selection

The study was approved by the Institutional Ethics Committee (CAAE: 41114915.5.0000.5121). Tumor samples from fifteen patients with CC were collected at the Mario Penna Institute (Belo Horizonte, Brazil) from August 2017 until May 2019. Patients who met the inclusion criteria, such as no previous history of cancer or any immune diseases, histopathological diagnosis of adenocarcinoma or squamous cell carcinoma (SCC), diagnosis of cervical cancer at stage II according to the FIGO staging system, and who had signed the informed consent form, were included. After the diagnosis and biopsy, the patients underwent chemoradiotherapy, and clinical data were collected and analyzed. Patients were classified as having either a poor prognosis (PP) or a good prognosis (GP), based on PFS being lower than or higher than five years, respectively.

2.2. General View of the Study

This work is part of a previous study of our group, which investigated DEGs in CC stem-like cells (CCSCs) from tumor biopsies at diagnosis between patient responders and non-responders to chemoradiotherapy [14]. In the present study only the non-stem cells fraction was used. The flow diagram of this study is presented in Figure 1.

2.3. Fluorescence-Activated Cell Sorting (FACS)

The protocol for cell sorting has been previously described [14]. Briefly, FACS was used to isolate enriched CCSCs from a complex mixture of tumor cells based on their light scatter and fluorescent staining profiles. CC tissue fragments (5 mm) from patients' biopsies were disaggregated using the BD™ Medimachine System according to the manufacturer's instructions (BD-Biosciences, Franklin Lakes, NJ, USA). The resulting cell suspension was frozen in 20% HES cryoprotective solution (100 mL anhydrous glucose 1.7 g/L; Na (+1) 140 mEq/L; Cl (−1) 98 mEq/L; K (+1) 5 mEq/L; Mg (+2) 3 mEq/L; Gluconate 23 mEq/L; Acetate 27 mEq/L) and stored in liquid nitrogen until use. Two monoclonal antibodies corresponding to the cell surface markers CD45 (APC-H7 Clone 2D1) and CD34 (PE Clone 563) were used in the cell suspension, increasing the cell concentration to 5×10^6 cells/mL. CCSCs-enriched subpopulations were isolated using the FACSAria® flow-sorter (BD-Biosciences, Franklin Lakes, NJ, USA). Yield mode was performed at 45 psi with 85-µm nozzle at a frequency of ~51 kHz. CCSCs were sorted into cytometry tubes using a gate containing the CD45−/CD34+ population to eliminate contamination with hematopoietic stem cells. Non-stem cells were kept for subsequent RNA sequencing.

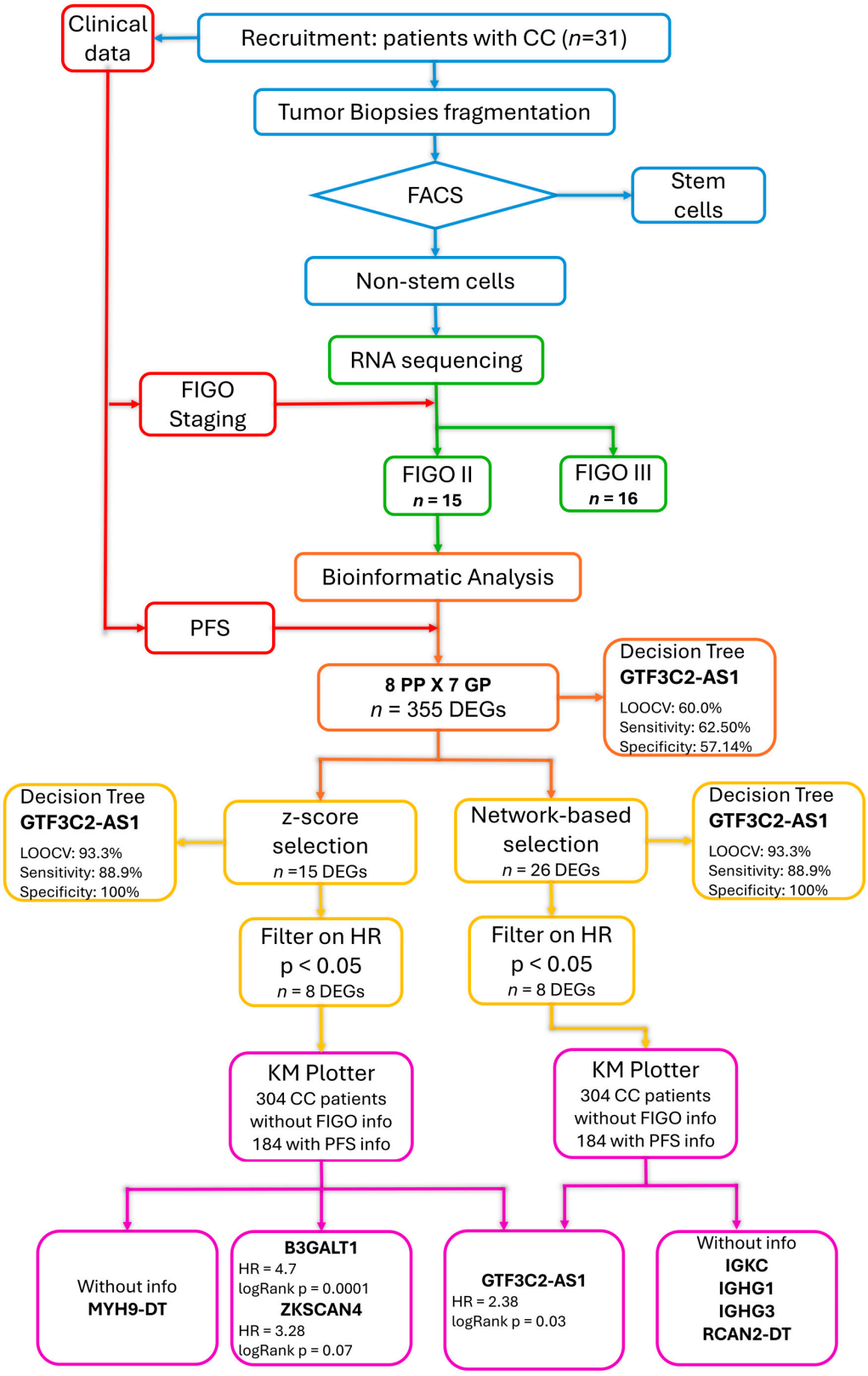


Figure 1. Schematic representation of the study workflow. The colors indicate the different methodological steps: cervical cancer non-stem cell selection (blue), clinical data collection (red), transcriptome sequencing (green), identification of differentially expressed genes (DEGs) between poor (PP) and good (GP) prognosis groups (orange), variable selection considering their association with progression-free survival (PFS) measured as hazard ratio (HR) (yellow), and validation in an external cohort (purple).

2.4. Next Generation Sequencing (NGS)

The non-stem cell pellet was processed with the SMART-Seq v4 Ultra-low Input RNA Sequencing Kit (Takara Bio USA, Mountain View, CA, USA) for cDNA synthesis, according to the manufacturer's instructions. Qualitative and quantitative analysis of cDNA were performed using the Agilent High Sensitivity DNA Chip run on the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) and the Qubit™ dsDNA HS Assay Kit on a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), respectively. cDNA suitable for NGS presented fragments from 400 bp to 10,000 bp, and concentration $\geq$ 0.3 ng/µL.

Sequencing libraries were prepared using the Nextera XT Library Prep Kit (Illumina, San Diego, CA, USA) and the Nextera XT Index Kit V2 Set A (Illumina, San Diego, CA, USA), following the manufacturer's instructions. The libraries' fragment size was measured using the High Sensitivity D1000 kit on the 2200 TapeStation System (Agilent, Santa Clara, CA, USA). Libraries' concentration was determined with the Qubit™ dsDNA HS Assay Kit on a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). NGS was performed using a NextSeq® 500/550 High Output Kit v2 (150 cycles) on a NextSeq® 550 sequencer (Illumina, San Diego, CA, USA).

2.5. Differential Gene Expression Analysis

Raw reads obtained from RNA-Seq underwent quality assessment using FastQC (version 0.11.9) [15]. Adapters and sequences with lengths or Phred scores lower than 35 and 30, respectively, were discarded by Trimmomatic (version 0.39) [16]. Filtered reads were aligned to the Homo Sapiens reference genome, version GRCh38, using the STAR software (version 2.7.10b) [17] and the GTF file from the same genome obtained from Ensembl (version 109) [18]. The gene expression quantification was performed using the "outFilterMultimapNmax 1" and "quantMode GeneCounts" STAR options. The raw counts were filtered using HTSFilter (version 1.48) [19] and then normalized with DESeq2 (version 1.48.1) [20]. Both packages are part of the R software environment (version 4.3.2) [21]. Finally, differential gene expression analysis was performed using DESeq2 (version 1.38.3). Only DEGs with log2FoldChange (log2FC) > 1 or <−1 and *p*.adjusted (*p*adj) value < 0.05 were considered. After removing duplicated genes, hierarchical clustering analysis was performed to group upregulated and downregulated DEGs, using the pheatmap R package (version 1.0.13) [22].

Two approaches were used to select the transcripts most associated with patients PFS. In the first method the initial DEGs list were filtered to include those that were upregulated (z-score > 0) in at least 50% of the PP group and, simultaneously, downregulated (z-score < 0) in at least 50% of the GP group, or vice versa. The z-score values were calculated with the following equation:

$$z = (x - \text{Mean}(x))/\text{SD}(x), \quad (1)$$

where

- x = data frame with normalized counts obtained from DESeq2;
- Mean(x) = mean expression of each DEG across all patients, calculated using the R basic function *rowMeans*;
- SD(x) = standard deviation across all patients, calculated using the *rowSds* function from the matrixStats R package (version 1.5) [23].

The second approach employed a network-based variable selection strategy using the regnet R package (version 1.0.2) [24]. This method not only incorporates correlations among genomic features but also applies a network-based penalization under the accelerated failure time (AFT) model, combined with Kaplan–Meier weights. By accounting

for correlations in survival outcomes, this approach has been widely applied for variable selection in high-dimensional cancer genomic datasets [24]. Gene expression, as normalized counts, and PFS time were used as input data. The response and penalty parameters were set to "survival" and "network", respectively, and the robust method was chosen. To compute the optimal value of lambda, the number of folds for cross-validation was set to 15.

2.6. Statistical Analysis

Five-year progression-free and overall survival (OS) rates were assessed through Kaplan–Meier method and compared through the Cox–Mantel (log-rank) test using the Kaplan–Meier (KM) plotter web-based platform [25,26]. HR's considering PFS time were obtained through Cox regression using the 'coxph' function from the 'survival' R package (version 4.4.1) [27] for each variable selected. Patients were divided into high and low groups based on gene expression levels above or below the median. High expression levels were set to 1, and low expression levels were set to 0. Log-rank test p-values < 0.05 were considered significant.

Machine learning (ML) analysis using normalized counts from the DEGs was performed via the J48 algorithm in WEKA software (Waikato Environment for Knowledge Analysis, version 3.6.11, University of Waikato, Hamilton, New Zealand) [28]. Briefly, this algorithm builds decision trees based on a set of training data, selecting at each node the attribute that most effectively splits the samples into subsets. To reduce the likelihood of overfitting, the Confidence Factor (C) parameter, used for pruning the decision tree, was set to 0.25, and the Minimum Number of Instances per Leaf (M) parameter, which prevents the creation of nodes based on very few samples, was set to 2. A higher value for this parameter typically results in smaller, more generalizable trees. No hyperparameter optimization was performed outside the validation loop, ensuring that all model adjustments occurred exclusively within the cross-validation process to avoid information leakage.

Leave-one-out cross-validation (LOOCV) was applied to evaluate test performance, which is particularly suitable for small datasets. In this approach, each of the 15 samples was systematically left out once, with the model trained on the remaining 14 samples, and the procedure repeated until every sample had served as the test instance exactly once. The classification accuracy calculated represents the average performance across all folds, reflecting the model's overall generalizability rather than the performance of a single model.

2.7. Validation with an External Cohort

Searches were conducted for larger CC expression datasets across different databases; however, only one dataset included PFS information in the clinical data, while the other two provided only OS (Table 1).

For this reason, the prognostic value of the selected genes was evaluated in an independent cohort using the Kaplan–Meier (KM) plotter web-based platform [25,26]. This tool enables real-time survival analysis based on transcriptomic data from large, publicly available patient cohorts. A cervical squamous cell carcinoma cohort comprising 304 patients was selected under the "Pan Cancer RNA-seq" option. Since FIGO stage information was not available for this dataset, all patients with documented recurrence status were included. Samples were stratified into two groups (high and low expression) using the median expression values for each one of the selected genes that presented significant HR values ($p < 0.05$). Survival analyses were performed using recurrence-free survival (RFS) as the clinical endpoint. The follow-up threshold was set at 60 months, and patients were censored beyond this time point. Group comparisons were assessed using the Cox–Mantel (log-rank) test. Differences in gene expression between high and low groups were

observed using boxplot and compared through Mann–Whitney test in GraphPad Prism 9.0 (GraphPad Software, Boston, MA, USA).

Table 1. Public datasets with larger CC gene expression acquired through RNA sequencing.

	TCGA-CESC	CGCI-HTMCP-CC	KMplotter
Method	RNAseq	RNAseq	RNAseq
Cases	306	212	304
Cases with Gene Expression	303	123	304
FIGO II Stage cases	70	91	NA
Cases with Treatment Outcome	39	34	174
OS time	Yes	Yes	Yes
PFS time	No	No	Yes
Reference	[29]	[30]	[26]

TCGA: The Cancer Genome Atlas; CESC: Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma Collection; CGCI: Cancer Genome Characterization Initiative; HTMCP: HIV+ Tumor Molecular Characterization Project; CC: cervical cancer; KM: Kaplan Meier; RNA: ribonucleic acid; FIGO: International Federation of Gynecology and Obstetrics; NA: not available; OS: overall survival; PFS: progression free survival.

3. Results

In this study, we aimed to identify a gene signature differentially expressed in FIGO II CC patients who are more likely to experience relapse within five years of starting treatment. Tumor biopsies from 15 women with a median age of 48 years (range 24–81) were submitted to RNA sequencing of their non-stem cell fraction. The cohort clinicopathological characteristics are summarized in Table 2 (for details, please see Table S1). Approximately 93.3% of patients (14/15) had SCC, with only one patient having adenocarcinoma. Stage FIGO IIB was predominant (14/15), and one patient was classified as FIGO IIA. The mean tumor size was 6.1 cm, and most patients had parametrial (93.3%) and/or vaginal involvement (93.3%). Most tumors were moderately (53%) or poorly differentiated (40%), and one was undifferentiated (grade IV).

Table 2. Clinicopathological characteristics of cervical cancer patients.

	PP Group $n = 8$	GP Group $n = 7$	Total (%) $n = 15$
Median age (years)	39	64	48
Diagnosis			
SCC	7 (88%)	7 (100%)	14 (93%)
Adenocarcinoma	1 (12%)	-	1 (7%)
Histological grade			
II	4 (50%)	4 (57%)	8 (53%)
III	3 (38%)	3 (43%)	6 (40%)
IV	1 (12%)	-	1 (7%)
FIGO stage			
IIA	1 (12%)	-	1 (7%)
IIB	7 (88%)	7 (100%)	14 (93%)
Tumor size (cm)			
>4 cm	7 (88%)	4 (57%)	11 (73%)
≤4 cm	-	3 (43%)	3 (20%)
NA	1 (12%)	-	1 (7%)
Parametrial Involvement			
Bilaterally	1 (12%)	5 (71%)	6 (40%)
Unilaterally	6 (76%)	2 (29%)	8 (53%)
Free	1 (12%)	-	1 (7%)

Table 2. *Cont.*

	PP Group $n = 8$	GP Group $n = 7$	Total (%) $n = 15$
Vaginal Involvement			
Present	8 (100%)	4 (57%)	12 (80%)
Absent	-	2 (29%)	2 (13%)
NA	-	1 (14%)	1 (7%)
Distant metastasis after treatment			
Yes	8 (100%)	-	8 (53%)
No	-	7 (100%)	7 (47%)

SCC: Squamous Cell Carcinoma. PP: poor prognosis; GP: good prognosis; FIGO: International Federation of Gynecology and Obstetrics; NA: not available.

Patients were categorized into two prognostic groups based on whether they presented or did not present distant metastasis after treatment. The PP group was characterized by a lower median age at diagnosis (39 years, ranging from 24 to 67), tumor size greater than 4 cm (88%), and unilateral parametrial involvement in most cases (76%). On the other hand, the GP group was characterized by a higher median age at diagnosis (64 years, ranging from 32 to 82) and bilateral parametrial involvement in most cases (71%). Patients classified as PP had a median PFS of 9.4 months, with 5-yr OS rate of only 12.5% (HR = 7.44, 95% CI 1.47–37.81, $p < 0.01$) (Figure 2). In contrast, GP patients exhibited a PFS of more than 60 months and a 5-year OS rate of 74.2%.

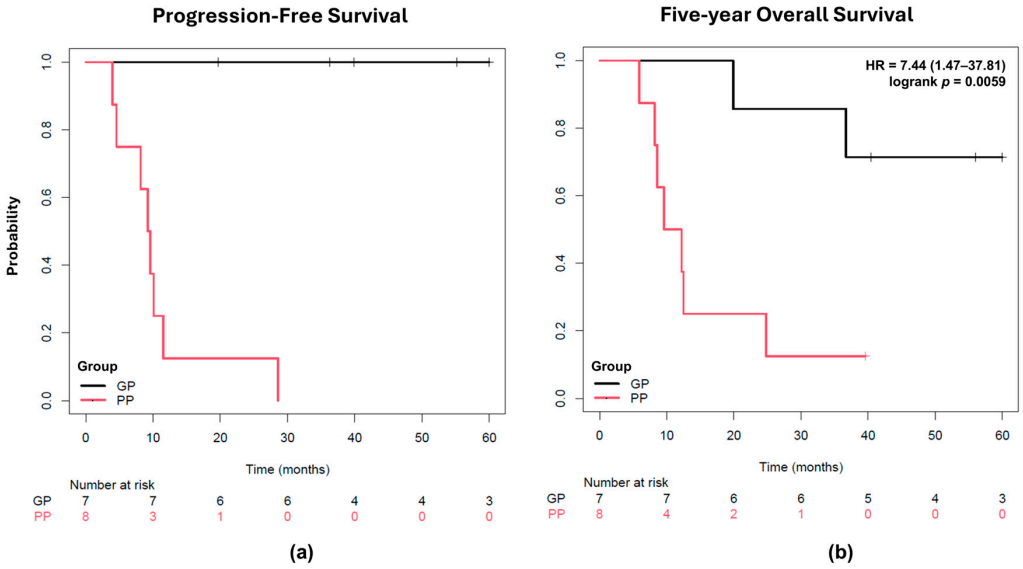


Figure 2. Kaplan–Meier plots for cervical cancer patients after 5 years of follow-up. (**a**) Progression-free survival was considered for a period of five years after diagnosis. (**b**) Overall survival in five years was compared using the Cox–Mantel (log-rank) test. The occurrence of metastasis after treatment was used as a criterion for patient assignment in the PP group and was significantly associated with shorter 5-yr overall survival. PP: poor prognosis; GP: good prognosis; HR: hazard ratio.

The comparison between gene expression data of both groups revealed 355 DEGs associated with patient outcome (Tables S2 and S3), most of them were upregulated in only two GP patients (Figure 3).

To identify the transcripts most strongly associated with patient prognosis, two approaches were applied. In the first analysis, we filtered the dataset for genes that were upregulated in at least 50% of the PP group and simultaneously downregulated in at least 50% of the GP group, or vice versa, based on z-score values (Table S4). Fifteen transcripts

were selected based on these criteria (Tables S5–S7), and the resulting heatmap is depicted in Figure 4a. The second analysis used a network-based variable selection strategy, which incorporates correlations among genomic features and accounts for correlations in survival outcomes. Twenty-six transcripts were selected through this method (Tables S8 and S9) and are shown as a heatmap in Figure 4b. Only the GTF3C2-AS1 transcript was selected by the two methods.

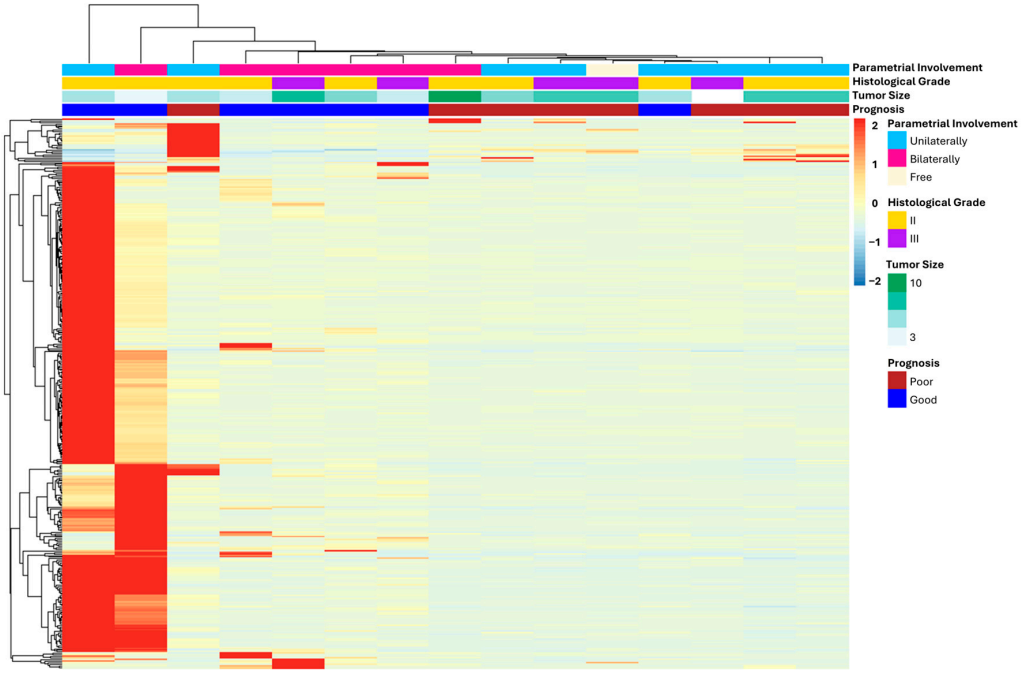


Figure 3. Heatmap of differentially expressed genes between FIGO II CC patients with poor and good prognosis. Only DEGs with log2FC > 1 or <−1 and padj value < 0.05 were considered, totaling 355 genes. Most of them are upregulated in only two GP patients (left side of the heatmap).

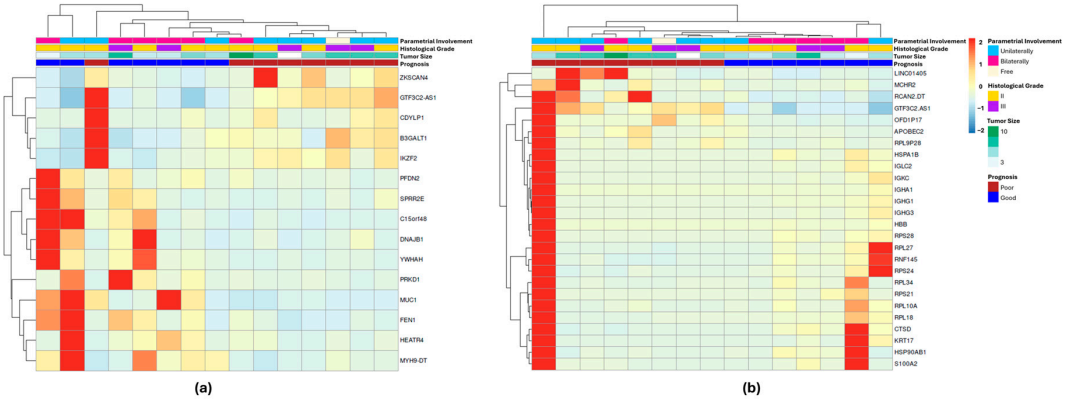


Figure 4. Differentially expressed genes associated with FIGO II CC patients' prognosis selected by two different approaches. (**a**) Heatmap for the 15 DEGs selected by the z-score method: upregulated in at least 50% of the PP group and, at the same time, downregulated in at least 50% of the GP group, or vice versa. (**b**) Heatmap for the 26 DEGs selected by network-based method. GTF3C2-AS1 was the only transcript in common between the two lists.

Given the small sample size of our cohort, we sought larger CC gene expression datasets in public repositories to validate the identified prognostic signature. Among the three available RNA-seq datasets, only one included PFS information. Therefore, we utilized the KMplotter web-based platform to assess the relationship between gene expression and PFS in an independent CC cohort. This dataset comprised RNA-seq profiles

from 304 CC samples, of which 184 had available PFS data. As FIGO stage information was not provided, analyses could not be restricted to stage II patients.

Univariable Cox regression identified eight (60%) of the 15 DEGs from the z-score selection method as significantly associated with PFS (Table S10), three of which (B3GALT1, GTF3C2-AS1, ZKSCAN4) were validated in an external CC cohort (n = 174) (Table 3). Higher expression of these transcripts correlated with increased relapse risk and shorter PFS (Figure 5). MYH9-DT showed a protective effect in our cohort (HR = 0.1, 95% CI: 0.01–0.85, p = 0.011), though external validation was not possible due to lack of data. Other transcripts (IKZF2, MUC1, PRKD1, YWHAH) were significant internally but not externally.

Table 3. Hazard ratios for transcripts selected by z-score method significantly associated with PFS in the internal cohort.

	Progression-Free Survival HR (95% CI), *p*	
mRNA	**Internal Cohort (*n* = 15)**	**External Cohort (*n* = 174)**
B3GALT1	5.11 (1.02–25.55) **p = 0.027**	4.7 (1.96–11.26) **p = 0.0001**
GTF3C2-AS1	18.73 (2.22–157.61) **p = 0.0003**	2.38 (1.05–5.39) **p = 0.033**
IKZF2	6.74 (1.33–34.24) **p = 0.0087**	1.24 (0.56–2.74) p = 0.59
MUC1	0.07 (0.01–0.58) **p = 0.002**	1.84 (0.81–4.17) p = 0.14
MYH9-DT	0.1 (0.01–0.85) **p = 0.011**	NA
PRKD1	0.09 (0.01–0.74) **p = 0.0052**	2.14 (0.94–4.85) p = 0.063
YWHAH	0 (0–inf) **p = 0.0001**	0.85 (0.39–1.86) p = 0.68
ZKSCAN4	5.18 (1.04–25.95) **p = 0.026**	3.28 (1.31–8.22) **p = 0.0072**

p-values in bold were considered significant ($p < 0.05$). HR: hazard ratio; CI: confidence interval; mRNA: messenger ribonucleic acid; B3GALT1: Beta-1,3-Galactosyltransferase 1; GTF3C2-AS1: General Transcription Factor IIIC Subunit 2 Antisense RNA 1; IKZF2: IKAROS Family Zinc Finger 2; MUC1: Mucin 1; MYH9-DT: Myosin Heavy Chain 9 Divergent Transcript; PRKD1: Protein Kinase D1; YWHAH: Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Eta; ZKSCAN4: Zinc Finger with KRAB and SCAN Domains 4; NA: not available.

From the network-based selection method, eight of 26 transcripts were associated with PFS internally (Table S10), but only GTF3C2-AS1 (also identified by the z-score method) was validated externally (Table 4). Four transcripts (IGHG1, IGHG3, IGKC, RCAN2-DT) were unavailable in KMplotter; in our dataset, immunoglobulin genes suggested a protective role (HR < 1), while RCAN2-DT correlated with poor prognosis.

Decision tree analysis reinforced GTF3C2-AS1 as a key discriminator between patient groups. While full training achieved perfect classification using all 355 DEGs, leave-one-out cross-validation (LOOCV) resulted in modest performance, with 60% accuracy, 62.5% sensitivity (correctly classifying PP patients), and 57.1% specificity (correctly classifying GP patients). In contrast, restricting the input to transcripts selected by the z-score or network-based methods markedly improved model performance, achieving 93.3% accuracy in LOOCV, with 88.9% sensitivity and 100% specificity. These parameters are summarized in Table 5. These results highlight the importance of dimensionality reduction for enhancing predictive power.

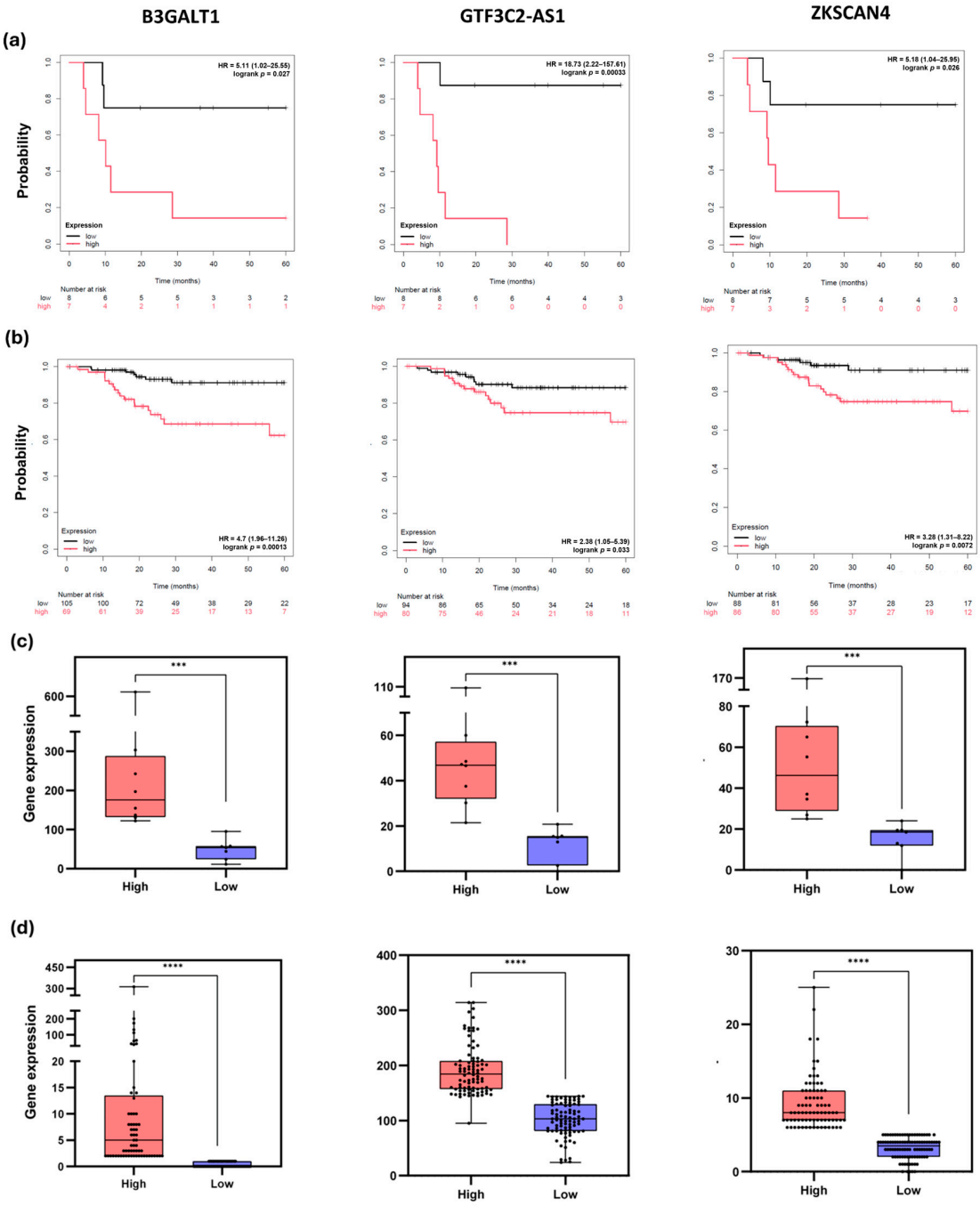


Figure 5. DEGs significantly associated with PFS in the internal and external cohorts. (**a**,**b**) Kaplan–Meier plots of PFS for the (**a**) internal (n = 15) and (**b**) the external (n = 174) cohorts. Curves were generated by KMplotter platform using the expression values for the genes associated with PFS in both cohorts. The results suggest that high expression of B3GALT1, GTF3C2-AS1, or ZKSCAN4 significantly increases the risk of disease recurrence within five years. (**c**,**d**) Boxplot of patients' distribution into high (red) and low (blue) groups based on gene expression level above or below the median. Gene expressions were significantly different between both groups for the three DEGs, either in (**c**) the internal or (**d**) the external cohort. The box extends from the 25th to the 75th percentiles. The line in the middle of the box is plotted at the median. Whiskers were plotted from the minimum to the maximum value. Each dot corresponds to one different patient. Comparison was performed using the Mann–Whitney test. *** $p < 0.001$; **** $p < 0.0001$.

Table 4. Hazard ratios for transcripts selected by the network-based method significantly associated with PFS in the internal cohort.

mRNA	Progression-Free Survival HR (95% CI), *p*	
	Internal Cohort (*n* = 15)	External Cohort (*n* = 174)
GTF3C2_AS1	18.73 (2.22–157.61) **p = 0.0003**	2.38 (1.05–5.39) **p = 0.033**
HSPA1B	0.22 (0.04–1.13) **p = 0.049**	1.23 (0.56–2.72) *p* = 0.6
IGHG1	0.07 (0.01–0.58) **p = 0.002**	NA
IGHG3	0.07 (0.01–0.58) **p = 0.002**	NA
IGKC	0.19 (0.04–0.99) **p = 0.03**	NA
KRT17	0.07 (0.01–0.58) **p = 0.002**	1.31 (0.59–2.89) *p* = 0.5
RCAN2-DT	5.64 (1.12–28.49) **p = 0.019**	NA
RNF145	0.19 (0.04–0.99) **p = 0.03**	1.01 (0.46–2.22) *p* = 0.98

p-values in bold were considered significant ($p < 0.05$). HR: hazard ratio; CI: confidence interval; mRNA: messenger ribonucleic acid; GTF3C2-AS1: General Transcription Factor IIIC Subunit 2 Antisense RNA 1; HSPA1B: Heat Shock Protein Family A (Hsp70) Member 1B; IGHG1: Immunoglobulin Heavy Constant Gamma 1; IGHG3: Immunoglobulin Heavy Constant Gamma 3; IGKC: Immunoglobulin Kappa Constant; KRT17: Keratin 17; RCAN2-DT: Regulator Regulator of Calcineurin 2 Divergent Transcript; RNF145: Calcineuri Ring Finger Protein 145; NA: not available.

Table 5. Decision tree analyses performance parameters in LOOCV.

Parameter	Input Data		
	All DEGs (*n* = 355)	z-Score Selection (*n* = 15)	Network-Based Selection (*n* = 26)
Accuracy	60.0%	93.3%	93.3%
Specificity	57.1%	100%	100%
Sensitivity	62.5%	88.9%	88.9%

DEGs: Differentially Expressed Genes.

4. Discussion

Cervical cancer stands out as the leading cause of cancer-related deaths among women, with approximately 660,000 new cases in 2022, according to the World Health Organization. The FIGO staging system is crucial for classifying patient risk and guiding treatment decisions. However, it still does not fully capture the complexity of the molecular tumor's response to chemotherapy and its inherent heterogeneity. In this context, molecular tests could provide additional information to staging systems, thereby refining the segregation of patients into risk groups. For instance, the updated FIGO 2023 staging for endometrial cancer incorporates, among other parameters, molecular classification, influencing stages I and II [31]. In this study, we identified three DEGs with the potential to subclassify the FIGO II stage CC into good and poor prognosis. Patients with higher expression of B3GALT1, GTF3C2-AS1, and ZKSCAN4 presented an increased risk of disease recurrence in our cohort. Interestingly, the same pattern was observed in CC patients with shorter PFS time using a larger external dataset.

The B3GALT1 is a member of the beta-1,3-galactosyltransferase (beta3GalT) gene family. Its role in cancer is controversial. High levels of B3GALT1 have been shown to enhance CD8+ T cell infiltration and reduce immune evasion, suppressing breast cancer lung metastasis [32]. On the other hand, higher expression of B3GALT1 results in higher

levels of MUC1-associated Sialyl Lewis antigen, which is linked to the metastasis of prostate cancer cells [33]. In this study, we observed higher levels of B3GALT1 in CC patients who present metastasis after treatment, corroborating the association observed in prostate cancer. ZKSCAN4, also referred to as ZNF307, belongs to the Krüppel-associated box zinc finger (KRAB-ZFPs) family, which is the largest group of zinc finger transcription factors [34]. Researchers have shown that ZKSCAN4 reduces cell proliferation, migration, and invasion in a hepatocellular carcinoma model. Coherently, the knockdown of ZKSCAN4 restored the proliferative phenotype [35]. Interestingly, our study revealed that ZKSCAN4 was upregulated in CC patients with poor prognosis. This finding contrasts with its role in hepatocellular carcinoma, suggesting a context-specific function of ZKSCAN4 in different cancer types. GTF3C2-AS1 codes for a lncRNA for which there is limited information regarding its mechanism of action. Nonetheless, it has been associated with poor OS in endometrial carcinoma [36], corroborating our findings. In our study, PP patients exhibited higher expression levels of GTF3C2-AS1. Moreover, a decision tree model successfully classified all patients based solely on GTF3C2-AS1 expression, achieving 93.33% accuracy in predicting prognosis. This result underscores the potential of GTF3C2-AS1 as a promising candidate for a prognostic biomarker in CC.

Gene expression data for five of the transcripts identified as associated with PFS in our cohort were not available in the KMplotter CC patient database. Two of them are divergent transcripts, MYH9-DT and RCAN2-DT. According to HGNC, divergent transcripts are a subgroup of long noncoding RNA (lncRNA) located within 300–500 nucleotides of a protein-coding gene on the opposite strand. It has been suggested that these lncRNA positively regulate the transcription of nearby genes [37]. Thus, we could assume that the upregulation of MYH9-DT would lead to increased levels of MYH9 transcripts. The MYH9 gene encodes non-muscle myosin II-A (NMM-IIA), a protein essential for various cellular functions, including cell division, adhesion, and migration. Its role in cancer is controversial, being implicated as an oncogene or a tumor suppressor, depending on the specific cancer type [38]. One study has shown that using direct RNA interference (RNAi) to silence Myh9 promotes the formation of invasive SCC in tumor-susceptible models, likely due to NMM-IIA's role in regulating post-transcriptional p53 stabilization [39]. In this context, the association between the elevated expression of MYH9-DT and patients' good prognosis observed here could be a consequence of higher p53 stabilization resulting from increased MYH9 transcript levels, ultimately improving their response to chemotherapy. In the same way, the upregulation of the RCAN2-DT would lead to increased levels of RCAN2 transcripts. RCAN2 is one of the regulators of the calcineurin protein family and participates in phosphorylation by inhibiting the activation of calcineurin. Its role in cancer is still unclear. It was described as involved in tumor progression in gastric cancer [40], whereas its overexpression has been demonstrated to enhance apoptosis in vitro and in vivo [41]. In our cohort, RCAN2-DT higher levels were associated with shorter PFS, in consonance with the findings in gastric cancer.

The other three transcripts absent from the validation dataset were immunoglobulins significantly associated with better prognosis in our cohort. IGKC (immunoglobulin kappa constant) has previously been linked to metastasis-free survival in breast cancer and identified as a prognostic marker in non-small-cell lung cancer and colorectal cancer [42]. More recently, IGKC expression was associated with favorable response to immune checkpoint blockade in melanoma [43]. IGHG1 and IGHG3, which encode the immunoglobulin heavy constant gamma 1 and 3 chains, respectively, have been reported with heterogeneous prognostic roles. Several studies have demonstrated that high IGHG1 expression promotes tumor proliferation, migration, and invasion, and correlates with poor prognosis [44–46]. Conversely, other reports describe IGHG1 and IGHG3 as suppressor genes in breast cancer

recurrence [47]. In addition, elevated IGHG3 expression was observed in advanced non-small-cell lung cancer patients who responded to carboplatin plus paclitaxel chemotherapy compared with non-responders [48]. The strong association of IgG with cellular immunity, together with the high correlation of IGHG transcripts with metagenes of adaptive immune cells, suggests that elevated expression of these genes may reflect an active antitumor immune response, which may be responsible for the protective association observed in our patients [49].

Despite current limitations, the application of artificial intelligence (AI) to predict CC outcomes shows considerable promise for improving patient care by supporting clinical decision-making through the integration of genomic and clinical data [50]. In our study, decision tree analysis using the complete set of 355 DEGs was less accurate than models built after variable reduction. Decision trees based on selected transcripts achieved markedly higher sensitivity and specificity, underscoring the importance of rigorous variable selection in ML-based approaches. Using the z-score method, we identified transcripts that most effectively distinguished the two prognostic groups; 60% (8/15) were significantly associated with PFS in our cohort, and three (20%) were also validated in an external cohort. By contrast, the network-based method identified a lower proportion of transcripts that were significantly associated with PFS internally, only 30% (8/26). External validation of these findings was limited, as four of the eight transcripts—three of them immunoglobulin genes—were not available in the external dataset. Given the established role of immunoglobulin transcripts in tumor metastasis and prognosis, it is likely that their relevance will also extend to other CC cohorts, although further validation is required. Importantly, the two selection methods provided complementary insights into the molecular mechanisms underlying CC recurrence. Notably, GTF3C2-AS1 was the only transcript identified by both methods and was also prioritized by the decision tree analyses, where its expression alone was sufficient to stratify patients into the correct risk groups.

It is important to emphasize that this is an exploratory study and the findings should be considered preliminary. A key limitation was the absence of a large independent CC cohort that matched our specific criteria—patients with FIGO stage II disease and PFS data from a 5-year follow-up. Given the small size of our cohort, such validation would have been particularly valuable. As a feasible alternative, we used the KMplotter platform CC dataset to assess associations between gene expression and PFS in an external cohort, although FIGO stage information was unavailable for these patients. Despite these limitations, our results are intriguing and warrant further validation. Moreover, this study addresses a relatively underexplored question: the heterogeneity within patients classified under the same clinical stage.

5. Conclusions

This study highlights the potential of transcriptomic profiling to refine prognostic stratification in cervical cancer (CC) beyond the current FIGO staging system. We identified three transcripts—B3GALT1, ZKSCAN4, and GTF3C2-AS1—consistently associated with poor prognosis in FIGO stage II CC, with GTF3C2-AS1 emerging as a potential biomarker capable of classifying patients by risk with high accuracy. Additional candidates, including divergent transcripts (MYH9-DT, RCAN2-DT) and immunoglobulin genes (IGKC, IGHG1, IGHG3), revealed prognostic associations in our cohort, further suggesting their involvement in CC outcomes, although external validation remains limited.

Importantly, this exploratory study underscores the biological heterogeneity within patients assigned to the same FIGO stage, raising the value of molecular markers for more precise risk classification. Despite the limitations imposed by cohort size and incomplete external datasets, our findings provide preliminary but intriguing evidence that transcript-

based signatures, particularly GTF3C2-AS1, may serve as clinically relevant prognostic biomarkers in CC. Further validation in larger, well-annotated cohorts are warranted.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jpm15100497/s1, Table S1: Clinical Data; Table S2: DESeq2 Results 355 DEGs; Table S3: 355 DEGs; Table S4: zscore 355 DEGs; Table S5: DESeq2 results 15 DEGs; Table S6: 15 DEGs; Table S7: zscore 15 DEGs; Table S8: 26_DEGs; Table S9: zscore_26_DEGs; Table S10: Hazard_Ratios_for_selected_DEGs.

Author Contributions: Conceptualization, C.P.S.M., Á.P.C., R.A.P., I.F.G.A., W.J.J. and F.R.Q.; methodology, C.P.S.M. and A.B.M.; software, A.B.M., L.R.A., M.S.G. and P.L.L.B.; validation, C.P.S.M., Á.P.C. and A.B.M.; formal analysis, C.P.S.M. and A.B.M.; investigation, C.P.S.M. and A.B.M.; resources, L.C.B., P.G.O.S. and M.S.G.; data curation, C.P.S.M. and A.B.M.; writing—original draft preparation, C.P.S.M.; writing—review and editing, L.C.B., R.A.P., I.F.G.A., W.J.J. and P.G.O.S.; visualization, F.R.Q. and J.G.G.F.; supervision, C.P.S.M. and L.C.B.; project administration, L.C.B. and P.G.O.S.; funding acquisition, L.C.B. and P.G.O.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by MINISTRY OF HEALTH OF BRAZIL (PRONON), grant numbers NUP 25000.159953/2014-18 and NUP 25000.079266/2015-09; FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE MINAS GERAIS, grant numbers APQ-02255-22, APQ-02564-22 and RED 00059-23 (Rede Mineira de Pesquisa Translacional em Oncologia). The APC was funded by MINISTRY OF HEALTH OF BRAZIL (PRONON), grant number NUP 25000.079266/2015-09 and FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE MINAS GERAIS, grant number RED 00059-23 (Rede Mineira de Pesquisa Translacional em Oncologia).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Mario Penna Institute (CAAE: 41114915.5.0000.5121 on 9 June 2016).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in SRA database, reference number PRJNA812529.

Acknowledgments: We thank the patients and their families who enrolled in this trial; and Flavia Ivair Santiago de Oliveira from Mario Penna Institute for assistance during the submission process of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAAE	Certificate of Presentation of Ethical Appreciation
CC	Cervical Cancer
DEGs	Differentially Expressed Genes
FACS	Fluorescence-activated cell sorting
FIGO	The International Federation of Gynecology and Obstetrics
GP	Good Prognosis
HR	Hazard Ratio
lncRNA	Long noncoding RNA
log2FC	log2 Fold Change
LOOCV	Leave-One-Out Cross-Validation
NGS	Next Generation Sequencing
OS	Overall Survival
padj	p adjusted

PP	Poor Prognosis
PFS	Progression-free Survival
SCC	Squamous Cell Carcinoma
TCGA	The Cancer Genome Atlas

References

1. Cervical Cancer. Available online: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer?gad_source=1&gclid=Cj0KCQjwwae1BhC_ARIsAK4JfryI8eJ5VN114LbjZ2NK4S9zZJzoNwEi5jCMEZd-o-TKtNjlsyB5wNIaAgLsEALw_wcB (accessed on 1 July 2025).
2. Mohamud, A.; Høgdall, C.; Schnack, T. Prognostic Value of the 2018 FIGO Staging System for Cervical Cancer. *Gynecol. Oncol.* **2022**, *165*, 506–513. [CrossRef] [PubMed]
3. Bhatla, N.; Singhal, S.; Dhamija, E.; Mathur, S.; Natarajan, J.; Maheshwari, A. Implications of the Revised Cervical Cancer FIGO Staging System. *Indian J. Med. Res.* **2021**, *154*, 273–283. [CrossRef]
4. Salib, M.Y.; Russell, J.H.B.; Stewart, V.R.; Sudderuddin, S.A.; Barwick, T.D.; Rockall, A.G.; Bharwani, N. 2018 FIGO Staging Classification for Cervical Cancer: Added Benefits of Imaging. *Radiographics* **2020**, *40*, 1807–1822. [CrossRef]
5. Grigsby, P.W.; Massad, L.S.; Mutch, D.G.; Powell, M.A.; Thaker, P.H.; McCourt, C.; Hagemann, A.; Fuh, K.; Kuroki, L.; Schwarz, J.K.; et al. FIGO 2018 Staging Criteria for Cervical Cancer: Impact on Stage Migration and Survival. *Gynecol. Oncol.* **2020**, *157*, 639–643. [CrossRef]
6. Tang, X.; Guo, C.; Liu, S.; Guo, J.; Hua, K.; Qiu, J. A Novel Prognostic Nomogram Utilizing the 2018 FIGO Staging System for Cervical Cancer: A Large Multicenter Study. *Int. J. Gynecol. Obstet.* **2021**, *155*, 86–94. [CrossRef]
7. Berek, J.S.; Matsuo, K.; Grubbs, B.H.; Gaffney, D.K.; Lee, S.I.; Kilcoyne, A.; Cheon, G.J.; Yoo, C.W.; Li, L.; Shao, Y.; et al. Multidisciplinary Perspectives on Newly Revised 2018 FIGO Staging of Cancer of the Cervix Uteri. *J. Gynecol. Oncol.* **2018**, *30*, e40. [CrossRef]
8. Vallejo-Ruiz, V.; Gutiérrez-Xicotencatl, L.; Medina-Contreras, O.; Lizano, M. Molecular Aspects of Cervical Cancer: A Pathogenesis Update. *Front. Oncol.* **2024**, *14*, 1356581. [CrossRef] [PubMed]
9. Su, H.-C.; Wu, S.-C.; Yen, L.-C.; Chiao, L.-K.; Wang, J.-K.; Chiu, Y.-L.; Ho, C.-L.; Huang, S.-M. Gene Expression Profiling Identifies the Role of Zac1 in Cervical Cancer Metastasis. *Sci. Rep.* **2020**, *10*, 11837. [CrossRef] [PubMed]
10. Zhao, M.; Huang, W.; Zou, S.; Shen, Q.; Zhu, X. A Five-Genes-Based Prognostic Signature for Cervical Cancer Overall Survival Prediction. *Int. J. Genom.* **2020**, *2020*, 8347639. [CrossRef]
11. Ju, M.; Qi, A.; Bi, J.; Zhao, L.; Jiang, L.; Zhang, Q.; Wei, Q.; Guan, Q.; Li, X.; Wang, L.; et al. A Five-mRNA Signature Associated with Post-translational Modifications Can Better Predict Recurrence and Survival in Cervical Cancer. *J. Cell. Mol. Med.* **2020**, *24*, 6283–6297. [CrossRef]
12. Nguyen, N.N.Y.; Choi, T.G.; Kim, J.; Jung, M.H.; Ko, S.H.; Shin, Y.; Kang, I.; Ha, J.; Kim, S.S.; Jo, Y.H. A 70-Gene Signature for Predicting Treatment Outcome in Advanced-Stage Cervical Cancer. *Mol. Ther. Oncol.* **2020**, *19*, 47–56. [CrossRef]
13. Changede, P. Staging of Cervical Cancer: What Has Changed? *J. Obstet. Gynecol. India* **2024**, *74*, 378–381. [CrossRef] [PubMed]
14. Zuccherato, L.W.; Machado, C.M.T.; Magalhães, W.C.S.; Martins, P.R.; Campos, L.S.; Braga, L.C.; Teixeira-Carvalho, A.; Martins-Filho, O.A.; Franco, T.M.R.F.; Paula, S.O.C.; et al. Cervical Cancer Stem-Like Cell Transcriptome Profiles Predict Response to Chemoradiotherapy. *Front. Oncol.* **2021**, *11*, 639339. [CrossRef]
15. Babraham Bioinformatics-FastQC A Quality Control Tool for High Throughput Sequence Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 1 July 2025).
16. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
17. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef] [PubMed]
18. Dyer, S.C.; Austine-Orimoloye, O.; Azov, A.G.; Barba, M.; Barnes, I.; Barrera-Enriquez, V.P.; Becker, A.; Bennett, R.; Beracochea, M.; Berry, A.; et al. Ensembl 2025. *Nucleic Acids Res.* **2025**, *53*, D948–D957. [CrossRef] [PubMed]
19. Rau, A.; Gallopin, M.; Celeux, G.; Jaffrézic, F. Data-Based Filtering for Replicated High-Throughput Transcriptome Sequencing Experiments. *Bioinformatics* **2013**, *29*, 2146–2152. [CrossRef]
20. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef]
21. R Core Team R: A Language and Environment for Statistical Computing. 2024. Available online: https://www.R-project.org/ (accessed on 25 February 2025).
22. Kolde, R. Pheatmap: Pretty Heatmaps. Available online: https://raivokolde.r-universe.dev/pheatmap (accessed on 1 July 2025).
23. Bengtsson, H. *MatrixStats: Functions That Apply to Rows and Columns of Matrices (and to Vectors)*, R Package Version 0.52.2. 2017. Available online: https://github.com/HenrikBengtsson/matrixStats/ (accessed on 15 March 2025).

24. Ren, J.; Du, Y.; Li, S.; Ma, S.; Jiang, Y.; Wu, C. Robust Network-based Regularization and Variable Selection for High-dimensional Genomic Data in Cancer Prognosis. *Genet. Epidemiol.* **2019**, *43*, 276–291. [CrossRef]
25. Lánczky, A.; Győrffy, B. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. *J. Med. Internet Res.* **2021**, *23*, e27633. [CrossRef]
26. Győrffy, B. Integrated Analysis of Public Datasets for the Discovery and Validation of Survival-Associated Genes in Solid Tumors. *Innovation* **2024**, *5*, 100625. [CrossRef]
27. Therneau, T. *A Package for Survival Analysis in R*, R Package Version 3.8-4. 2024. Available online: https://CRAN.R-project.org/package=survival/ (accessed on 15 June 2025).
28. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*, 2nd ed.; Kaufmann, M., Ed.; Elsevier Science & Technology Books, Sony Electronics [Distributor]: San Francisco, CA, USA, 2005; ISBN 0120884070.
29. Cancer Genome Atlas Research Network. Integrated Genomic and Molecular Characterization of Cervical Cancer. *Nature* **2017**, *543*, 378–384. [CrossRef]
30. Gagliardi, A.; Porter, V.L.; Zong, Z.; Bowlby, R.; Titmuss, E.; Namirembe, C.; Griner, N.B.; Petrello, H.; Bowen, J.; Chan, S.K.; et al. Analysis of Ugandan Cervical Carcinomas Identifies Human Papillomavirus Clade–Specific Epigenome and Transcriptome Landscapes. *Nat. Genet.* **2020**, *52*, 800–810. [CrossRef] [PubMed]
31. Occhiali, T.; Poli, A.; Arcieri, M.; Driul, L.; Ditto, A.; Bogani, G.; Mariani, A.; Scambia, G.; Restaino, S.; Vizzielli, G. The Exciting Journey of Progress: Exploring FIGO 2023 Staging for Endometrial Cancer at a Leading ESGO Institution. *Eur. J. Surg. Oncol.* **2025**, *51*, 109695. [CrossRef] [PubMed]
32. Li, Y.; Guo, T.; He, J.; Liu, D.; Peng, S.; Xu, A. SLC35A2-Mediated Bisected GlcNAc-Modified Extracellular Vesicles Enhance Immune Regulation in Breast Cancer Lung Metastasis. *Int. Immunopharmacol.* **2025**, *154*, 114505. [CrossRef]
33. Chachadi, V.B.; Ali, M.F.; Cheng, P.-W. Prostatic Cell-Specific Regulation of the Synthesis of MUC1-Associated Sialyl Lewis a. *PLoS ONE* **2013**, *8*, e57416. [CrossRef]
34. Sobocińska, J.; Molenda, S.; Machnik, M.; Oleksiewicz, U. KRAB-ZFP Transcriptional Regulators Acting as Oncogenes and Tumor Suppressors: An Overview. *Int. J. Mol. Sci.* **2021**, *22*, 2212. [CrossRef] [PubMed]
35. Liang, Y.; Li, Q.; Chen, K.; Ni, W.; Zhan, Z.; Ye, F.; Li, Y.; Fang, Y.; Zhang, F.J.; Chen, L.; et al. Zinc Finger Protein 307 Functions as a Tumor-Suppressor and Inhibits Cell Proliferation by Inducing Apoptosis in Hepatocellular Carcinoma. *Oncol. Rep.* **2017**, *38*, 2229–2236. [CrossRef]
36. Zhou, M.; Zhang, Z.; Zhao, H.; Bao, S.; Sun, J. A Novel LncRNA-Focus Expression Signature for Survival Prediction in Endometrial Carcinoma. *BMC Cancer* **2018**, *18*, 39. [CrossRef]
37. Luo, S.; Lu, J.Y.; Liu, L.; Yin, Y.; Chen, C.; Han, X.; Wu, B.; Xu, R.; Liu, W.; Yan, P.; et al. Divergent LncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. *Cell Stem Cell* **2016**, *18*, 637–652. [CrossRef]
38. Brito, C.; Sousa, S. Non-Muscle Myosin 2A (NM2A): Structure, Regulation and Function. *Cells* **2020**, *9*, 1590. [CrossRef]
39. Schramek, D.; Sendoel, A.; Segal, J.P.; Beronja, S.; Heller, E.; Oristian, D.; Reva, B.; Fuchs, E. Direct in Vivo RNAi Screen Unveils Myosin IIa as a Tumor Suppressor of Squamous Cell Carcinomas. *Science* **2014**, *343*, 309–313. [CrossRef] [PubMed]
40. Hattori, Y.; Sentani, K.; Shinmei, S.; Oo, H.Z.; Hattori, T.; Imai, T.; Sekino, Y.; Sakamoto, N.; Oue, N.; Niitsu, H.; et al. Clinicopathological Significance of RCAN2 Production in Gastric Carcinoma. *Histopathology* **2019**, *74*, 430–442. [CrossRef]
41. Lee, S.-K.; Ahnn, J. Regulator of Calcineurin (RCAN): Beyond Down Syndrome Critical Region. *Mol. Cells* **2020**, *43*, 671–685. [CrossRef] [PubMed]
42. Schmidt, M.; Hellwig, B.; Hammad, S.; Othman, A.; Lohr, M.; Chen, Z.; Boehm, D.; Gebhard, S.; Petry, I.; Lebrecht, A.; et al. A Comprehensive Analysis of Human Gene Expression Profiles Identifies Stromal Immunoglobulin κ C as a Compatible Prognostic Marker in Human Solid Tumors. *Clin. Cancer Res.* **2012**, *18*, 2695–2703. [CrossRef]
43. Onieva, J.L.; Xiao, Q.; Berciano-Guerrero, M.-Á.; Laborda-Illanes, A.; de Andrea, C.; Chaves, P.; Piñeiro, P.; Garrido-Aranda, A.; Gallego, E.; Sojo, B.; et al. High IGKC-Expressing Intratumoral Plasma Cells Predict Response to Immune Checkpoint Blockade. *Int. J. Mol. Sci.* **2022**, *23*, 9124. [CrossRef]
44. Li, Y.; Wang, P.; Ye, D.; Bai, X.; Zeng, X.; Zhao, Q.; Zhang, Z. IGHG1 Induces EMT in Gastric Cancer Cells by Regulating TGF-β/SMAD3 Signaling Pathway. *J. Cancer* **2021**, *12*, 3458–3467. [CrossRef]
45. Zhao, J.; Peng, H.; Gao, J.; Nong, A.; Hua, H.; Yang, S.; Chen, L.; Wu, X.; Zhang, H.; Wang, J. Current Insights into the Expression and Functions of Tumor-Derived Immunoglobulins. *Cell Death Discov.* **2021**, *7*, 148. [CrossRef]
46. Wang, G.; Li, H.; Pan, J.; Yan, T.; Zhou, H.; Han, X.; Su, L.; Hou, L.; Xue, X. Upregulated Expression of Cancer-Derived Immunoglobulin G Is Associated With Progression in Glioma. *Front. Oncol.* **2021**, *11*, 758856. [CrossRef]
47. Hsu, H.-M.; Chu, C.-M.; Chang, Y.-J.; Yu, J.-C.; Chen, C.-T.; Jian, C.-E.; Lee, C.-Y.; Chiang, Y.-T.; Chang, C.-W.; Chang, Y.-T. Six Novel Immunoglobulin Genes as Biomarkers for Better Prognosis in Triple-Negative Breast Cancer by Gene Co-Expression Network Analysis. *Sci. Rep.* **2019**, *9*, 4484. [CrossRef]

48. Mon, M.M.; Srisomsap, C.; Chokchaichamnankit, D.; Watcharatanyatip, K.; Weeraphan, C.; Svasti, J.; Maneechai, K.; Thongsuksai, P.; Raungrut, P. Serum Proteomic Profiling Reveals Differentially Expressed IGHG3 and A1AG1 as Potential Predictors of Chemotherapeutic Response in Advanced Non-Small Cell Lung Cancer. *Anticancer Res.* **2021**, *41*, 1871–1882. [CrossRef] [PubMed]
49. Larsson, C.; Ehinger, A.; Winslow, S.; Leandersson, K.; Klintman, M.; Dahl, L.; Vallon-Christersson, J.; Häkkinen, J.; Hegardt, C.; Manjer, J.; et al. Prognostic Implications of the Expression Levels of Different Immunoglobulin Heavy Chain-Encoding RNAs in Early Breast Cancer. *npj Breast Cancer* **2020**, *6*, 28. [CrossRef] [PubMed]
50. Saini, S.K.; Sharma, D.N.; Chauhan, S.; Srivastava, S.; Gopishankar, N.; Subramani, V. Precision Prediction of Cervical Cancer Outcomes: A Machine Learning Approach to Recurrence and Survival Analysis. *J. Cancer Res. Ther.* **2025**, *21*, 538–546. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.