



# Differential expression of mRNA 3'-end isoforms in cervical and ovarian cancers

Didem Naz Dioken<sup>a</sup>, Ibrahim Ozgul<sup>a</sup>, Gozde Koksal Bicakci<sup>a</sup>, Kemal Gol<sup>b</sup>, Tolga Can<sup>c,1</sup>, Ayse Elif Erson-Bensan<sup>a,\*</sup>

<sup>a</sup> Department of Biological Sciences, Middle East Technical University (METU), Dumlupinar Blv No: 1 Universiteler Mah., Cankaya, Ankara, 06800, Turkiye

<sup>b</sup> Gynecology Clinic, Ugur Mumcu Cad 17/2, Cankaya, Ankara, Turkiye

<sup>c</sup> Department of Computer Engineering, Middle East Technical University (METU), Dumlupinar Blv No: 1, Universiteler Mah., Ankara, 06800, Turkiye

## ABSTRACT

Early diagnosis and therapeutic targeting are continuing challenges for gynecological cancers. Here, we focus on cancer transcriptomes and describe the differential expression of 3'UTR isoforms in patients using an algorithm to detect differential poly(A) site usage. We find primarily 3'UTR shortening cases in cervical cancers compared with the normal cervix. We show differential expression of alternate 3'-end isoforms of *FOXPI*, *VPS4B*, and *OGT* in HPV16-positive patients who develop high-grade cervical lesions compared with the infected but non-progressing group. In contrast, in ovarian cancers, 3'UTR lengthening is more evident compared with normal ovary tissue. Nevertheless, highly malignant ovarian tumors have unique 3'UTR shortening events (e.g., *CHRAC1*, *SLC16A1*, and *TOP2A*), some of which correlate with upregulated protein levels in tumors. Overall, our study shows isoform level deregulation in gynecological cancers and highlights the complexity of the transcriptome. This transcript diversity could help identify novel cancer genes and provide new possibilities for diagnosis and therapy.

## 1. Introduction

Compared to prevalent cancers, cancers of the female reproductive organs are relatively less common, yet all women are at risk with increasing age. Gynecologic cancers consist of multiple diseases with different etiologies and clinicopathological characteristics. Symptoms are usually not disease-specific and often overlap with other conditions, resulting in late detection and advanced disease states. Hence, early diagnosis and effective treatment are still significant challenges for these cancers, contributing considerably to the global cancer burden. Further research is needed to understand the underlying molecular mechanisms, identify new biomarkers for early detection, and develop targeted treatments.

High-throughput expression methods for genomic, transcriptomic, and proteomic changes have greatly benefitted molecular and clinical cancer research by providing insight into disease mechanisms and revealing potential markers for disease subtypes, stages, and drug responsiveness. With the advancement of sequencing techniques, the actual depth of transcriptome complexity is more appreciated. Hence, it is becoming more apparent that the transcriptome complexity generated by alternative RNA processing is not an exception but the norm. With this perspective, the Genotype-Tissue Expression (GTEx) project is an excellent resource for understanding mRNA isoform diversity in normal tissues [1].

On the other hand, deregulated RNA processing in malignancies further complicates the transcriptome [2–5]. As a result, isoforms

\* Corresponding author.

E-mail address: [erson@metu.edu.tr](mailto:erson@metu.edu.tr) (A.E. Erson-Bensan).

<sup>1</sup> Current address: Computer Science Department, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401 USA.



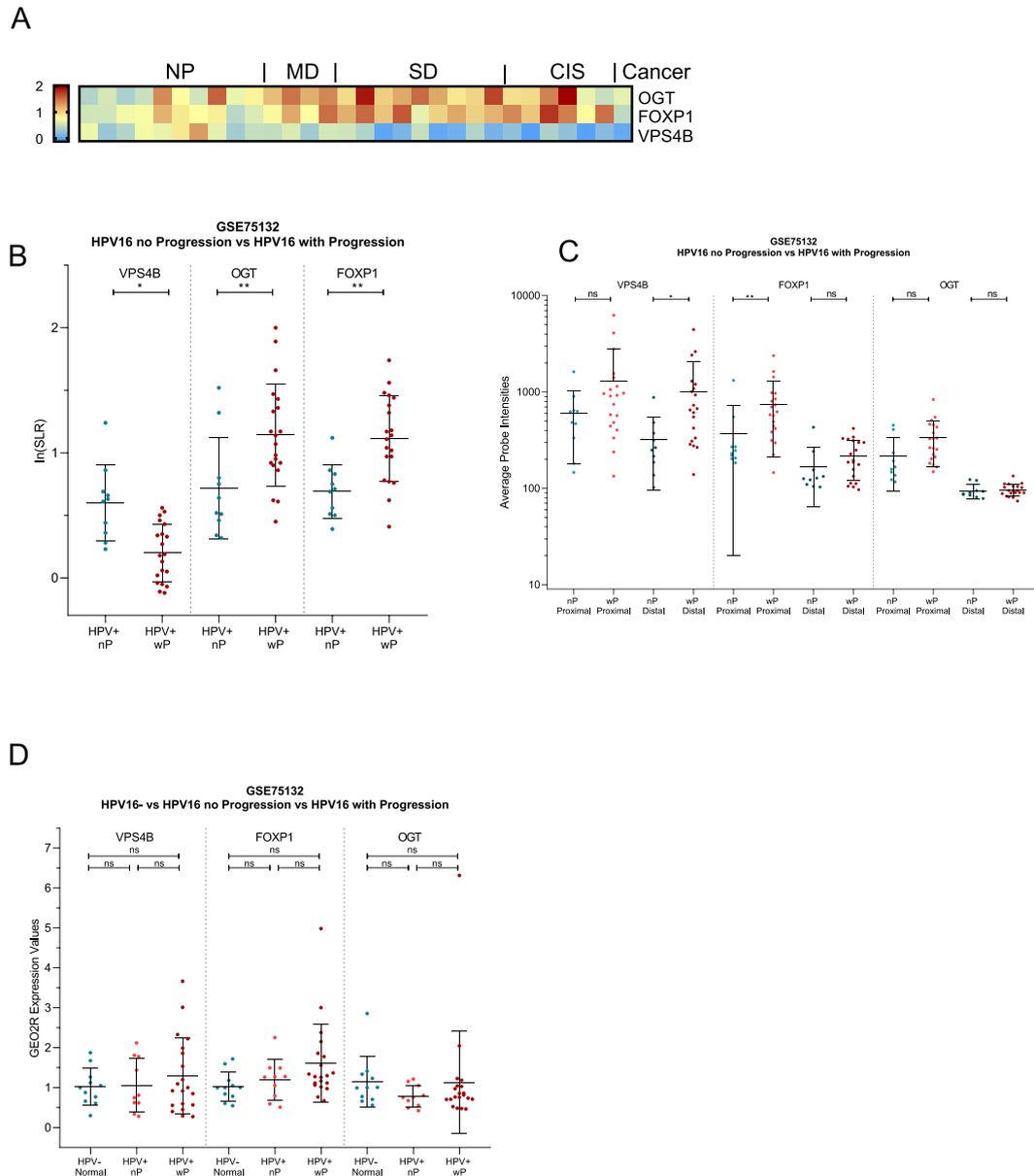
Finally, significant changes in the signal intensities are reported as SLR ((Short + Long)/Long ratio) values.

These results may have future implications in basic research, biomarker discovery, and precision medicine approaches for these malignancies.

## 2. Results

We analyzed transcript expression data for common gynecological cancers (cervix and ovary) to discover cancer-specific mRNA 3'-end isoforms using APADetect.

To investigate whether we can detect cancer-specific deregulation of 3'-end isoforms, first, we compiled expression data for cervical



**Fig. 2. SLR changes in the HPV16 infected patients (GSE75132).** **A.** Heatmap of tumor-specific SLR changes (BFL) in moderate to severe dysplasia and cancer cases ( $n = 20$ ) compared to persistently HPV16-infected patients who have not progressed (nP) ( $n = 10$ ) in GSE75132, MD: moderate dysplasia, SD: severe dysplasia, CIS: Carcinoma *in situ*, **B.** SLR values of individual tumors of progressed patients (wP: with progression) compared with the no-progression group, **C.** Average probe intensity graphs for isoform level expression values. Proximal probe sets recognize total transcripts, distal probes recognize longer 3'UTR isoforms for *VPS4B*, *FOXP1* and *OGT*, **D.** Total mRNA expressions of *VPS4B*, *FOXP1* and *OGT* in normal cervix tissue, no progression and progression groups (Unpaired *t*-test, Kolmogorov-Smirnov test, ns: not significant, \*\*\*\* ( $p < 0.0001$ ), \*\*\* ( $0.0001 < p < 0.001$ ), \*\* ( $0.001 < p < 0.01$ )).

and ovarian cancers from the GEO database (Table S1). Following APADetect analysis for differentially expressed isoforms in tumors compared to normal tissue, we used the WEKA machine learning platform [24] to find significant SLR changes to discriminate cancer samples from normal tissue. Correlation-based feature selection subset evaluation (CfsSubsetEval) was used. These transcripts were further grouped and reported as the BFL (Best First List).

### 2.1. Cervical cancer

We investigated GSE9750, GSE6791 and GSE63514 datasets to find cervix cancer-specific SLR changes. GSE9750 [25] has expression data from the normal cervical epithelium ( $n = 20$ ) and primary tumors (squamous cell carcinoma) ( $n = 33$ ). Using APADetect, we identified 67 cases with higher SLR values and 10 with lower SLR in tumors (Table S2). After SAM analysis, the total number of significant events was 48 (Table S3). The cancer-specific isoforms (12) in the BFL are presented in Fig. 1A, Table S4. SLR changes in individual tumor samples compared to normal tissue are shown in Fig. 1B.

In GSE6791 [26], there are eight normal cervix tissue and 20 cervical tumors. Using APADetect, we identified 235 significant SLR events (166 events with high SLR and 69 with low SLR) (Table S2). After SAM analysis, the total number of significant transcripts was 163 (Table S3). The CfsSubsetEval method in WEKA revealed BFL with 12 events (8 events with  $SLR > 1.5$ , 4 events with  $SLR < 0.6$ ) (Fig. 1C, Table S4). SLR changes in individual tumor samples compared to normal tissue are shown in Fig. 1D.

In a third dataset, GSE63514, we analyzed 24 normal tissue and 21 cervix tumors. We identified 80 significant SLR events (59 events with high SLR, 21 with low SLR) (Table S2). After SAM analysis, the total number of significant SLR events was 54 (Table S3). The BFL list had 17 events (Fig. 1E, Table S4). SLR changes in individual tumor samples compared to normal tissue are shown in Fig. 1F.

Based on the three independent datasets for 74 cervix cancer patients, the general pattern of increased SLR values indicated proximal poly(A) preference. *TCF3* caught our attention with increased proximal poly(A) site usage in all datasets (in BFL or significant SLR events lists) (Table S3, Table S4). Of note, *TCF3* has been proposed as a potential oncogene in cervical cancers. Hence, the 3'UTR shortening we detected here could contribute to the upregulation of *TCF3* [27,28]. Other than *TCF3*, the three independent datasets did not share many common significant SLR changes (Fig. S1, Table S3). We reasoned that the tumors in independent cohorts were heterogeneous due to disease stage, age of patients, or positivity with different human papillomavirus (HPV) types. Indeed, patients in all three datasets, were infected with different HPV types (GSE9750: 16, 45, 18, 31; GSE6791: 31, 33, 35, 58, 66, 16, 18, and in GSE63514: 16, 18 or other unspecified types).

### 2.2. High risk patients

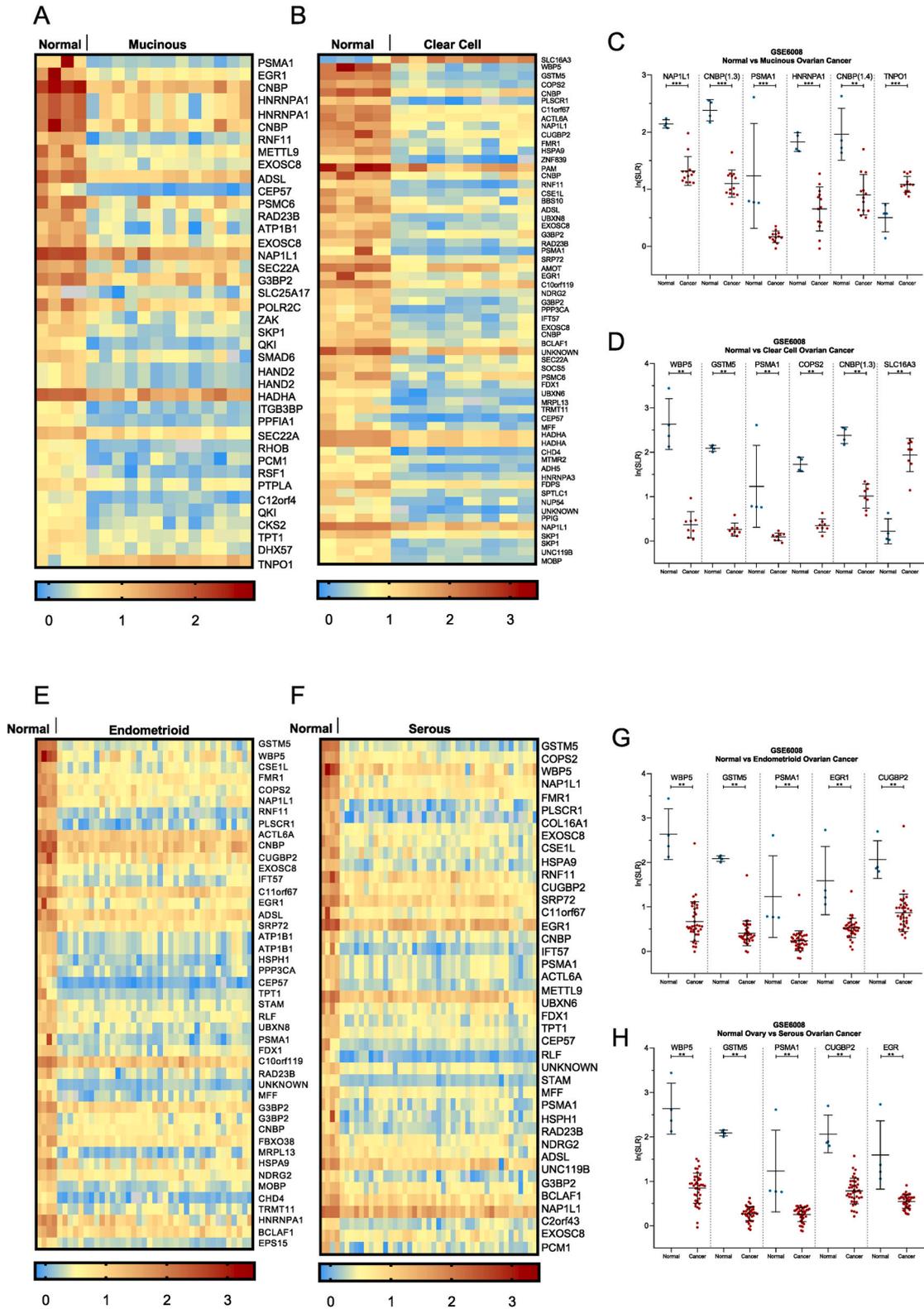
To focus on high-risk HPVs (e.g., 16, 18) as a major risk factor for cervical cancer [29,30], we looked into a fourth dataset (GSE75132) consisting of only HPV16-positive patients [31]. Curiously, only some women persistently infected with HPV16 are at high risk for progressing to CIN3+/cancer, whereas other persistently infected patients are not. *TMEM45A*, *p16INK4a*, and *SERPINB5* have been linked to progression to malignancy; however, the molecular mechanisms underlying progression are not completely known [31]. Hence, to address whether mRNA 3'-end isoforms are differentially expressed in the high-risk group, we divided the persistently HPV16-positive patients into two groups, those who have not progressed for up to 19 years ( $n = 10$ ) and those who developed moderate to severe dysplasia or cancer ( $n = 20$ ). APADetect revealed 61 events significantly altered in the infected and progressing patients (Table S2). The most significant 28 events were filtered using SAM (Table S3). The CfsSubsetEval method in WEKA revealed three events that differed the most between the no-progression and progression groups (Fig. 2A, Table S4). In the HPV16+ persistent group who progressed, *OGT* (O-linked *N*-acetylglucosamine (GlcNAc) transferase) and *FOXPI* (Forkhead Box P1) had  $SLR > 1.5$ . On the other hand, *VPS4B* (Vacuolar Protein Sorting 4 Homolog B) SLR was low (SLR, 0.66) in the HPV16+ patients who developed cancer.

Next, we looked further into the expression levels of these transcripts detected by proximal and distal probe sets. Probe sets recognizing different isoforms are shown in Fig. S2. Fig. 2B shows the SLR changes, and Fig. 2C shows increased average intensities for proximal and distal probes recognizing the total and long 3'UTR isoforms of *VPS4B*, *OGT*, and *FOXPI* in the progressed group. Higher proximal signal intensity due to proximal poly(A) site usage in the high-risk patients results in increased SLR values. For *VPS4B*, expression levels of isoforms detected by proximal and distal probes were both high in the progressed patients. Still, the longer 3'UTR isoform was more upregulated due to increased distal poly(A) site usage in the progressed patients. The isoform-specific regulation of *OGT*, *FOXPI*, and *VPS4B* could be important in defining high-risk patients and needs to be tested in larger patient cohorts to evaluate their biomarker potential to predict the development of high-grade cervical lesions in HPV16+ patients.

It is worth noting that such changes in isoform levels can be under-detected when differential expression of isoforms is not considered in the overall mRNA quantification methods. Indeed, the overall mRNA levels (calculated by taking the mean signal intensity of all probes) of *OGT*, *FOXPI*, and *VPS4B* are not different in the two groups (Fig. 2D). A similar quantification of overall mRNA levels by RNA-sequencing and other quantification methods disregarding isoforms could mask the discovery of potentially significant cancer-specific changes. These results highlight the importance of isoform-level quantification.

### 2.3. Ovarian cancers

To find subtype-specific differences, we grouped ovarian cancer patients into mucinous, clear cell carcinomas, endometrioid and serous. The GSE6008 dataset has normal ovary tissue ( $n = 4$ ) and mucinous ovarian tumors ( $n = 13$ ) (Table S1). APADetect revealed 174 events with  $SLR < 0.6$  and 31 with  $SLR > 1.5$  in mucinous ovarian cancers (Table S2). Sixty-nine events were selected as significant



**Fig. 3.** SLR changes in mucinous, clear cell carcinomas, endometrioid, and serous ovarian cancers. **A.** Heatmap of SLR fold changes (BFL) in mucinous ovarian cancer (n = 13) compared to normal tissue (n = 4) in GSE6008, **B.** Heatmap of SLR fold changes (BFL) in clear cell ovarian cancers (n = 8) compared to normal tissue (n = 4) in GSE6008, **C, D.** Representative SLR events that best distinguish mucinous ovarian cancer or clear cell ovarian cancer compared to normal tissue, respectively. **E.** Heatmap of SLR fold changes (BFL) in endometrioid ovarian cancers (n = 4) compared to

normal tissue (n = 37) in GSE6008, F. Heatmap of SLR fold changes (BFL) in serous ovarian cancers (n = 41) compared to normal tissue (n = 4) in GSE6008. G, H. Representative SLR events in BFL that best distinguish normal tissue and tumors of endometrioid or serous ovarian cancers, respectively (Unpaired *t*-test, Kolmogorov-Smirnov test, \*\*\* (0.0001 < p < 0.001, \*\* 0.001 < p < 0.01).

by SAM (Table S3). The CfsSubsetEval method in WEKA analysis identified 40 events distinguishing normal tissue from mucinous cancers (Fig. 3A, Table S4). Representative SLR values are given in Fig. 3C. The overall pattern of low SLR values indicates a shift towards distal poly(A) site usage in mucinous ovarian tumors. Next, eight clear cell carcinomas (CCC), a rare form of ovarian cancer with poor prognosis, were compared to normal ovarian tissue. We identified 170 events with SLR < 0.6 and 41 with SLR > 1.5 in tumors (Table S2). After SAM analysis, the number of significant events was 107 (Table S3). Sixty-one of these were within the BFL (Fig. 3B, Table S4). Representative SLR values are given in Fig. 3D.

We detected 176 SLR changes (135 with SLR < 0.6 and 41 with SLR > 1.5) for endometrioid ovarian tumors (n = 37) (Table S2). The SAM analysis filtered 74 of these events as significant (Table S3), and 45 of these events (SLR < 0.6) were listed as BFL (Fig. 3E, Table S4). Representative SLR values are given in Fig. 3G.

We then looked into serous ovarian cancers, the most malignant and prevalent type [32]. We analyzed 41 tumors compared to normal ovary tissue (n = 4) (GSE6008). A total of 134 events with SLR < 0.6 and 31 with SLR > 1.5 were identified in patients (Table S2). Sixty-six transcripts from these events were selected using SAM (Table S3), and the CfsSubsetEval method in WEKA revealed the best identifiers (BFL) of both groups using 43 events (Fig. 3F, Table S4). Representative SLR values are given in Fig. 3H.

These subtypes of ovarian cancers are histologically different; hence, shared SLR events were rare (Fig. S3, Table S4). While specific isoform ratios could be unique to subtypes, common isoforms shared by different subtypes could indicate functional importance. For example, decreased SLR, common in clear cell carcinomas, endometrioid, and serous ovarian cancers, indicated 3'UTR lengthening of *WBP5* (WW domain-binding protein 5, a.k.a. *TCEAL9* (transcription elongation factor A like 9)). Notably, *WBP5* expression is most abundant in normal ovaries (Fig. S4, Fig. 3D, G, 3H). Hence, it will be essential to understand the functional consequence of the shift in isoform levels of *WBP5* in ovarian cancers. Unique SLR changes detected for rare subtypes could also be biologically relevant.

#### 2.4. Malignancy-related SLRs

Next, we turned our focus to malignancy-related SLR changes in serous ovarian tumors. We compared tumors with low malignancy potential (LMP) (n = 18) to highly malignant tumors (n = 225) (GSE9891). We identified ten events with SLR > 1.5 and five events with SLR < 0.6 in the malignant group compared with the LMP tumors (Table S2). 12 of these transcripts were significant according to SAM (Table S3), and the CfsSubsetEval method in WEKA narrowed the SLR events to nine (Fig. 4A, Table S4). *CYP37A*, *C4A*, *RSPH4A*, and *PPIL6* had decreased SLRs, whereas *CHRAC1*, *KLHL24*, *RPL13*, *SLC16A1*, and *TOP2A* had increased SLRs in the highly malignant tumors compared with the LMP tumors (Fig. 4B). Probe intensities for different isoforms are given in Fig. 4C. Of note, for *RPL13*, while SLR was more than 1.5, the decrease of the long isoform was more than the proximal transcripts. Hence, despite having an SLR > 1.5, *RPL13* did not have upregulated short 3'UTR isoform (Fig. 4C).

#### 2.5. Protein levels

To understand whether the differential expression of 3'-end isoforms may affect protein levels in highly malignant ovarian tumors, we took advantage of the UALCAN database [33]. Protein data were available only for *CHRAC1*, *RPL13*, *TOP2A*, and *SLC16A1*, which had increased SLRs in tumors. The protein levels of *CHRAC1*, *RPL13*, and *TOP2A* are high in primary ovarian tumors (n = 100) compared to normal ovary tissue (n = 25) (Fig. 5A, B, C). *SLC16A1* protein levels were not significantly different in normal tissue and tumors (Fig. 5D). Of note, post-translational modifications, protein stability, and/or activity are other variables to be considered at the proteome level. Nevertheless, it would be interesting to experimentally test whether isoform level changes explain protein over-expression and how this shift in isoform ratios may be linked to malignancy in ovarian tumors.

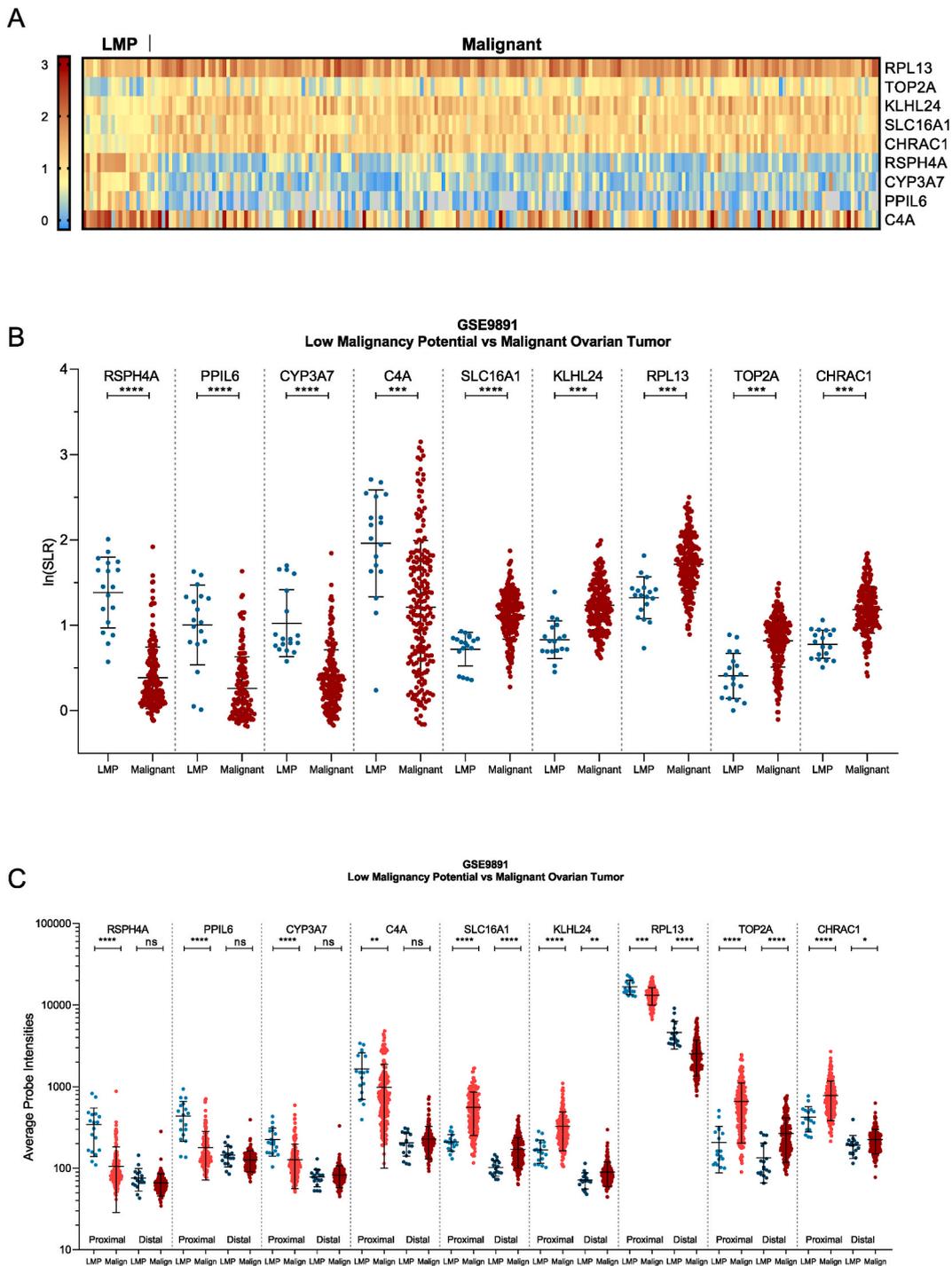
Overall, our results show malignancy-associated isoform-level deregulation of a group of transcripts in cervical and ovarian cancers. Altered isoform ratios could be critical for defining novel genes for understanding malignancy in ovarian cancers.

### 3. Discussion

mRNAs harbor protein-coding information flanked by 5'UTRs and 3'UTRs, targeted by *trans*-factors such as RNA-binding proteins and microRNAs. These *trans* factors alter mRNAs' translation rate, stability, and localization. Despite these critical features and mounting evidence showing alterations in 3'UTR lengths in cancer transcriptomes [34], 3'-end isoforms are generally under-detected due to technical limitations [14,35]. Here, we utilized an isoform-specific approach and determined differentially regulated isoforms with alternate 3'-ends in common gynecological cancers.

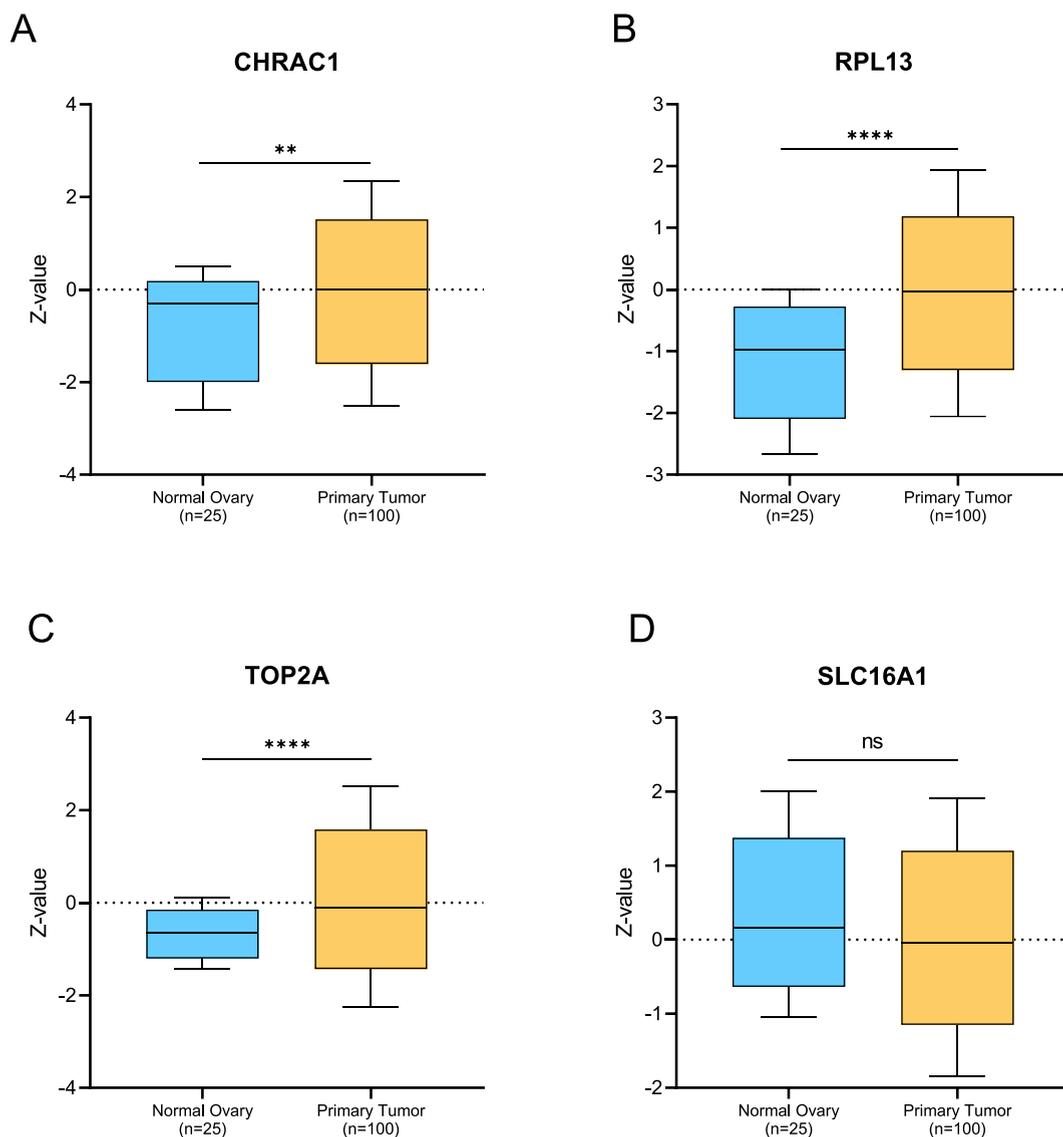
#### 3.1. Cervical cancers

Our results showed significant changes in the expression levels of 3'UTR isoforms in cervical cancer datasets, but shared events were rare due to the heterogeneity of patients. When we analyzed a cohort with an extended follow-up of only HPV16-infected patients, we identified 3'UTR shortening and lengthening events unique to the progressed patients compared to HPV16-positive patients who did



**Fig. 4.** SLR changes in highly malignant serous ovarian tumors compared with low malignant tumors. **A.** Heatmap of SLR fold changes (BFL) in highly malignant tumors (n = 225) compared to low malignant tumors (LMP) (n = 18). **B.** SLR values of BFL isoforms that best distinguish highly malignant tumors from low-risk tumors. **C.** Average probe intensity graphs showing isoform level expression in malignant and LMP tumors. Proximal probe sets recognize total transcripts, and distal probes recognize longer 3'UTR isoforms of BFL transcripts. (Unpaired *t*-test, Kolmogorov-Smirnov test, ns: not significant) \*\*\*\* (p < 0.0001), \*\*\* (0.0001 < p < 0.001), \*\* (0.001 < p < 0.01).

not develop cancer. The proximal poly(A) site usage was enhanced for *FOXP1* and *OGT* in HPV16-positive patients who had progressed. For *VPS4B*, the longer 3'UTR isoform was increased more in the progression group. Interestingly, these three proteins have potential implications for cervical cancer. For example, FOXP proteins regulate the transcription of differentiation and immune



**Fig. 5.** Protein levels of **A. CHRAC1**, **B. RPL13**, **C. TOP2A**, and **D. SLC16A1** in ovarian tumors. Protein expression in ovarian tumors (n = 100) in the CPTAC data of UALCAN is shown compared to normal ovary tissue (n = 25). Log<sub>2</sub> spectral count ratio values from CPTAC were first normalized within each sample profile and then normalized across samples. Z-values represent standard deviations from the median (p values were calculated by UALCAN; Welch's *t*-test, ns: not significant, \*\*\*\* (p < 0.0001), \*\* (0.001 < p < 0.01).

system-related genes and have been implicated in angiogenesis and tumorigenesis [36]. VPS4B belongs to the AAA (ATPase associated with diverse cellular activities) protein family and functions during the endosomal sorting and lysosomal degradation of membrane proteins. VPS4B also participates in cytokinesis and virus budding [37,38]. On the other hand, O-linked *N*-acetylglucosamine (GlcNAc) transferase (OGT) catalyzes the reversible addition of the O-GlcNAc to target proteins. Recently, overexpression of OGT was reported in colon and rectum adenocarcinomas [39]. High OGT levels were linked to colon cancer metastasis and poor prognosis [40]. Increased OGT was also reported in HPV-associated cervical neoplasms. Interestingly, HPV *E6* upregulates OGT, causing increased O-GlcNAc of MYC, enhancing the oncogenic activity of HPV [41]. High OGT activity may also alter the stability and function of many other target proteins. In support of an oncogenic role, inhibiting OGT slows tumor growth in animal models [41]. Hence, alternative polyadenylation and 3'UTR shortening could contribute to the upregulation of OGT protein levels in cervical cancers.

In short, the roles of FOXP1, VPS4B, and OGT in neoplastic transformation within the specific context of HPV-infected cells merit further investigation. The functional consequence and predictive value of these isoform level changes in high-grade cervical lesions must be explored in larger patient cohorts.

Of note, we would like to emphasize that, despite the significantly different isoform ratios in patients, overall mRNA levels are not different for the high-risk and low-risk patient groups (Fig. 2D). This contradiction highlights the need to study gene expression at the isoform level. An isoform-aware gene expression quantification may allow the identification of new oncogene activation cases without

genomic mutations. Hence, isoform switches in cancer transcriptomes are a promising strategy for discovering new cancer genes with biological impact [8,23].

### 3.2. Ovarian cancers

Ovarian cancer is the leading cause of mortality in gynecologic malignancies, and most ovarian cancers are diagnosed at a late stage. As a result, ovarian cancer has one of the lowest survival rates among all cancers [42]. Hence, effective screening methods are needed for early detection and diagnosis. We screened ovarian cancer subtypes for differentially expressed 3'UTR isoforms to contribute to these efforts. As expected, histologically distinct subtypes had unique and shared SLR changes. For example, the common pattern of 3'UTR lengthening of *WBP5* in different subtypes is potentially interesting because *WBP5* has already been implicated in various malignancies [43,44].

We also identified SLR changes in a highly malignant group of ovarian tumors compared to a low malignancy group. The isoform ratio changes detected in malignant tumors are potentially interesting for identifying high-risk patients. While some of these transcripts have not been investigated within the context of cancer, *TOP2A* is a well-known cancer gene. *TOP2A* (Topoisomerase II alpha) regulates DNA topology during transcription, replication, and repair [45]. High *TOP2A* expression correlates with poor prognosis, and overexpression of *TOP2A* is reported in various cancers [46,47]. DNA topoisomerase II (*TOP2*) inhibitors are clinically used drugs that cause cancer cell death by inducing DNA damage. *TOP2A* expression levels are known to affect the effectiveness of these topoisomerase inhibitors [48]. The 3'UTR shortening may contribute to the overexpression of *TOP2A* in a group of ovarian tumors that may not harbor DNA level alterations. These findings may be relevant for the therapeutic potential of *TOP2* inhibitors and combination therapies with DNA-damaging agents.

Overall, these findings show malignancy-related isoform level changes that conventional gene expression analyses may not detect. Unfortunately, these 3'UTR isoforms are generally under-detected. However, they may be informative in evaluating gene expression changes in cancer cells. Shorter or longer 3'UTRs are targeted differently by *trans*-factors, including miRNAs and RNA-binding proteins. The increased expression of a specific isoform more than other isoforms may significantly alter protein levels and/or function. Unfortunately, the overall quantification of an mRNA by disregarding isoforms may hinder the detection of isoform-level changes and cause loss of valuable information with protein-level implications. Our group has identified such an isoform switch for *HNRNPA1*. The total mRNA levels were no different in breast cancers compared to normal breast tissue, but APADetect revealed upregulation of a stable mRNA isoform, whereas the normally expressed and unstable isoform was downregulated in breast cancers [23].

In short, our results propose a mechanism where alternative polyadenylation of mRNA isoform provides additional ways to deregulate protein levels in cancers. In addition, transcript-level diagnostic and prognostic tests are widely used for different cancers. We think an isoform-focused view of cancer transcriptomes and our data can open up new research areas that will discover novel targets for diagnostic and prognostic applications, which could help develop early detection tools and help improve patient outcomes.

## 4. Methods

### 4.1. Datasets

Cancer patient datasets (GSE9750, GSE6791, GSE63514, GSE75132, GSE6008, GSE9891) from the Expression Project for Oncology from National Center for Biotechnology Information Gene Expression Omnibus (GEO) were re-analyzed to detect differentially expressed 3'-end isoforms.

### 4.2. Detection and quantification of 3'UTR isoforms

We used the APADetect algorithm to detect and quantify mRNA isoforms that differ at their 3'-ends [21]. We used CEL files of Human Genome U133A (HG133A, GPL96) and U133 Plus 2.0 arrays (HG133Plus2, GPL570) to determine the mean signal intensities of proximal and distal probe sets for each transcript. Means of proximal and distal probe sets were calculated as the "short" to "long" ratio (SLR). Patient SLRs were compared to normal tissues. Significant SLR changes were determined using SAM [49] and TM4 Multiple Array Viewer [50] (SLR >1.5 for shortening events or SLR <0.6 for lengthening events). All SLR events presented in supplementary data are APADetect and SAM outputs. Probes with no reads in more than 15% of patients were filtered out for reliability.

Correlation-based feature selection subset evaluation (CfsSubsetEval) was used to avoid overfitting and dimensionality problems as implemented in WEKA [24]. CfsSubsetEval assessed the performance of SLR values based on the predictive ability of unique SLR events to distinguish tumors from normal tissue. We used the BestFirst algorithm with default parameters in WEKA [24]. Cancer-specific SLR values were listed as the best first list (BFL) using at least 5 of the ten cross-validations to identify SLR values that best discriminate between normal and cancer samples. For statistical analysis, unpaired *t*-test and Kolmogorov-Smirnov tests were used, \*\*\*\* ( $p < 0.0001$ ), \*\*\* ( $0.0001 < p < 0.001$ ), \*\* ( $0.001 < p < 0.01$ ).

### 4.3. Protein levels

We retrieved protein expression data, when available, for ovarian tumors from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) and UALCAN (<http://ualcan.path.uab.edu/analysis.html>) [33]. Log<sub>2</sub> Spectral count ratio values from CPTAC were first normalized within each sample profile and then normalized across samples. Z-values represent standard deviations from the median.

## 5. Limitations of the study

Early diagnosis is a challenge for gynecological cancers. Because tumors are diagnosed at late stages, available gene expression data is limited to reflect all disease stages or rare subtypes. Hence it is difficult to define early vs. late molecular changes. In addition to this difficulty, disregarding isoforms in gene expression analysis hinders the true complexity of the cancer transcriptome. Our approach utilizes existing microarray datasets to determine isoform level expression changes in tumors. Hence, we were limited by the availability of datasets and by pre-designed probes, as well as low sample sizes for normal tissues and tumor samples. However, tailored library preparation or long-read RNA sequencing are more likely to reveal the true transcriptome complexity, including at the 3'-ends. Isoform-level deciphering of cancer transcriptomes is expected to contribute to a more comprehensive understanding of gynecological cancers.

## Data and code availability

We used expression data from tumors or corresponding normal tissue as publicly accessible data from GEO. The accession numbers for the datasets are listed in [Table S1](#). APADetect algorithm is available at <https://github.com/tolgacan/APADetect>.

Additional information required to use APADetect or re-analyze the data reported in this paper is available from the lead author upon request.

## Author contributions

Didem Naz Dioken: Performed the experiments, Analyzed and interpreted the data.

Ibrahim Ozgul: Performed the experiments, Analyzed and interpreted the data.

Gozde Koksall Bicakci: Performed the experiments, Analyzed and interpreted the data.

Kemal Gol: Analyzed and interpreted the data.

Tolga Can: Performed the experiments, Analyzed and interpreted the data.

Ayse Elif Erson-Bensan: Conceived and designed the experiments, Analyzed and interpreted the data, Wrote the paper.

## Funding statement

This project was funded by TUBITAK 119Z075.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e20035>.

## References

- [1] J. Lonsdale, et al., The genotype-tissue expression (GTEx) project, *Nat. Genet.* 45 (6) (Jun. 2013), <https://doi.org/10.1038/ng.2653>. Art. no. 6.
- [2] E.A. Obeng, C. Stewart, O. Abdel-Wahab, Altered RNA processing in cancer pathogenesis and therapy, *Cancer Discov.* 9 (11) (Nov. 2019) 1493–1510, <https://doi.org/10.1158/2159-8290.CD-19-0399>.
- [3] P. Chen, et al., Comprehensive analysis of prognostic alternative splicing signatures in endometrial cancer, *Front. Genet.* 11 (2020). Accessed: Jul. 05, 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00456>.
- [4] J. Liu, D. Lv, X. Wang, R. Wang, X. Li, Systematic profiling of alternative splicing events in ovarian cancer, *Front. Oncol.* 11 (Mar. 2021), 622805, <https://doi.org/10.3389/fonc.2021.622805>.
- [5] C. Yue, T. Zhao, S. Zhang, Y. Liu, G. Zheng, Y. Zhang, Comprehensive characterization of 11 prognostic alternative splicing events in ovarian cancer interacted with the immune microenvironment, *Sci. Rep.* 12 (1) (Jan. 2022) 980, <https://doi.org/10.1038/s41598-021-03836-1>.
- [6] D. Sun, et al., Comprehensive characterization of the alternative splicing landscape in ovarian cancer reveals novel events associated with tumor-immune microenvironment, *Biosci. Rep.* 42 (2) (Feb. 2022), BSR20212090, <https://doi.org/10.1042/BSR20212090>.
- [7] E. Szostak, F. Gebauer, Translational control by 3'-UTR-binding proteins, *Brief Funct Genomics* 12 (1) (Jan. 2013) 58–65, <https://doi.org/10.1093/bfgp/els056>.
- [8] E.T. Wang, et al., Alternative isoform regulation in human tissue transcriptomes, *Nature* 456 (7221) (Nov. 2008) 470–476, <https://doi.org/10.1038/nature07509>.
- [9] J.J. Chan, et al., Pan-cancer pervasive upregulation of 3' UTR splicing drives tumorigenesis, *Nat. Cell Biol.* 24 (6) (Jun. 2022) 928–939, <https://doi.org/10.1038/s41556-022-00913-z>.
- [10] C. Mayr, D.P. Bartel, Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells, *Cell* 138 (4) (Aug. 2009) 673–684, <https://doi.org/10.1016/j.cell.2009.06.016>.
- [11] H.J. Park, et al., 3' UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk, *Nat. Genet.* 50 (6) (Jun. 2018) 783–789, <https://doi.org/10.1038/s41588-018-0118-8>.

- [12] O. Begik, M. Oyken, T. Cinkilli Alican, T. Can, A.E. Erson-Bensan, Alternative polyadenylation patterns for novel gene discovery and classification in cancer, *Neoplasia* 19 (7) (Jul. 2017) 574–582, <https://doi.org/10.1016/j.neo.2017.04.008>.
- [13] D. Burri, M. Zavolan, Shortening of 3' UTRs in most cell types composing tumor tissues implicates alternative polyadenylation in protein metabolism, *RNA* 27 (12) (Dec. 2021) 1459–1470, <https://doi.org/10.1261/rna.078886.121>.
- [14] A.E. Erson-Bensan, RNA-biology ruling cancer progression? Focus on 3'UTRs and splicing, *Cancer Metastasis Rev.* 39 (3) (Sep. 2020) 887–901, <https://doi.org/10.1007/s10555-020-09884-9>.
- [15] I. Pereira-Castro, A. Moreira, On the function and relevance of alternative 3'-UTRs in gene expression regulation, *WIREs RNA* 12 (5) (2021) e1653, <https://doi.org/10.1002/wrna.1653>.
- [16] Z. Huang, E.C. Teeling, ExUTR: a novel pipeline for large-scale prediction of 3'-UTR sequences from NGS data, *BMC Genom.* 18 (1) (Nov. 2017) 847, <https://doi.org/10.1186/s12864-017-4241-1>.
- [17] N. Kandhari, C.A. Kraupner-Taylor, P.F. Harrison, D.R. Powell, T.H. Beilharz, The detection and bioinformatic analysis of alternative 3' UTR isoforms as potential cancer biomarkers, *Int. J. Mol. Sci.* 22 (10) (May 2021) 5322, <https://doi.org/10.3390/ijms22105322>.
- [18] C. Mayr, Evolution and biological roles of alternative 3'UTRs, *Trends Cell Biol.* 26 (3) (Mar. 2016) 227–237, <https://doi.org/10.1016/j.tcb.2015.10.012>.
- [19] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nat. Rev. Genet.* 10 (1) (Jan. 2009) 57–63, <https://doi.org/10.1038/nrg2484>.
- [20] B. Tian, J.L. Manley, Alternative polyadenylation of mRNA precursors, *Nat. Rev. Mol. Cell Biol.* 18 (1) (Jan. 2017) 18–30, <https://doi.org/10.1038/nrm.2016.116>.
- [21] B.H. Akman, T. Can, A.E. Erson-Bensan, Estrogen-induced upregulation and 3'-UTR shortening of CDC6, *Nucleic Acids Res.* 40 (21) (Nov. 2012) 10679–10688, <https://doi.org/10.1093/nar/gks855>.
- [22] H.B. Akman, M. Oyken, T. Tuncer, T. Can, A.E. Erson-Bensan, 3'UTR shortening and EGF signaling: implications for breast cancer, *Hum. Mol. Genet.* 24 (24) (Dec. 2015) 6910–6920, <https://doi.org/10.1093/hmg/ddv391>.
- [23] M. Erdem, et al., Identification of an mRNA isoform switch for HNRNPA1 in breast cancers, *Sci. Rep.* 11 (1) (Dec. 2021), 24444, <https://doi.org/10.1038/s41598-021-04007-y>.
- [24] M.A. Hall, Correlation-based Feature Selection of Discrete and Numeric Class Machine Learning, Working Paper, University of Waikato, Department of Computer Science, May 2000. Accessed: Jul. 05, 2022. [Online]. Available: <https://researchcommons.waikato.ac.nz/handle/10289/1024>.
- [25] L. Scotto, et al., Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression, *Genes Chromosomes Cancer* 47 (9) (Sep. 2008) 755–765, <https://doi.org/10.1002/gcc.20577>.
- [26] D. Pyeon, et al., Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers, *Cancer Res.* 67 (10) (May 2007) 4605–4619, <https://doi.org/10.1158/0008-5472.CAN-06-3619>.
- [27] L. Luo, et al., Up-regulation of transcription factor 3 is correlated with poor prognosis in cervical carcinoma, *Int. J. Gynecol. Cancer* 27 (7) (Sep. 2017) 1422–1430, <https://doi.org/10.1097/IGC.0000000000001032>.
- [28] X. Yu, Z. Li, R. Bai, F. Tang, Transcriptional factor 3 binds to siRTUIN1 to activate the Wnt/ $\beta$ -catenin signaling in cervical cancer, *Bioengineered* 13 (5) (May 2022) 12516–12531, <https://doi.org/10.1080/21655979.2022.2076481>.
- [29] X. Castellsagué, Natural history and epidemiology of HPV infection and cervical cancer, *Gynecol. Oncol.* 110 (3 Suppl 2) (Sep. 2008) S4–S7, <https://doi.org/10.1016/j.ygyno.2008.07.045>.
- [30] Y. Li, C. Xu, Human papillomavirus-related cancers, *Adv. Exp. Med. Biol.* 1018 (2017) 23–34, [https://doi.org/10.1007/978-981-10-5765-6\\_3](https://doi.org/10.1007/978-981-10-5765-6_3).
- [31] A. Manawapat-Klopfner, et al., TMEM45A, SERPINB5 and p16INK4A transcript levels are predictive for development of high-grade cervical lesions, *Am. J. Cancer Res.* 6 (7) (2016) 1524–1536.
- [32] M.-A. Lisio, L. Fu, A. Goyeneche, Z.-H. Gao, C. Telleria, High-grade serous ovarian cancer: basic sciences, clinical and therapeutic standpoints, *Int. J. Mol. Sci.* 20 (4) (Feb. 2019) E952, <https://doi.org/10.3390/ijms20040952>.
- [33] D.S. Chandrashekar, et al., UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, *Neoplasia* 19 (8) (Aug. 2017) 649–658, <https://doi.org/10.1016/j.neo.2017.05.002>.
- [34] S. Lianoglou, V. Garg, J.L. Yang, C.S. Leslie, C. Mayr, Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression, *Genes Dev.* 27 (21) (Nov. 2013) 2380–2396, <https://doi.org/10.1101/gad.229328.113>.
- [35] A.E. Erson-Bensan, T. Can, Alternative polyadenylation: another foe in cancer, *Mol. Cancer Res.* 14 (6) (Jun. 2016) 507–517, <https://doi.org/10.1158/1541-7786.MCR-15-0489>.
- [36] J.-H. Kim, J. Hwang, J.H. Jung, H.-J. Lee, D.Y. Lee, S.-H. Kim, Molecular networks of FOXP family: dual biologic functions, interplay with other molecules and clinical implications in cancer progression, *Mol. Cancer* 18 (1) (Dec. 2019) 180, <https://doi.org/10.1186/s12943-019-1110-3>.
- [37] S. Urata, H. Yokosawa, J. Yasuda, Regulation of HTLV-1 gag budding by Vps4A, Vps4B, and AIP1/alix, *Virology* 366 (Jul. 2007) 66, <https://doi.org/10.1186/1743-422X-4-66>.
- [38] S.R. White, B. Lauring, AAA+ ATPases: achieving diversity of function with conserved machinery, *Traffic* 8 (12) (Dec. 2007) 1657–1667, <https://doi.org/10.1111/j.1600-0854.2007.00642.x>.
- [39] I. Sheraj, N.T. Guray, S. Banerjee, A pan-cancer transcriptomic study showing tumor specific alterations in central metabolism, *Sci. Rep.* 11 (Jul. 2021), 13637, <https://doi.org/10.1038/s41598-021-93003-3>.
- [40] M. Jiang, et al., Correction: O-GlcNAcylation promotes colorectal cancer metastasis via the miR-101-O-GlcNAc/EZH2 regulatory feedback circuit, *Oncogene* 38 (28) (Jul. 2019) 5744–5745, <https://doi.org/10.1038/s41388-019-0834-2>.
- [41] Q. Zeng, et al., O-linked GlcNAcylation elevated by HPV E6 mediates viral oncogenesis, *Proc Natl Acad Sci U S A* 113 (33) (Aug. 2016) 9333–9338, <https://doi.org/10.1073/pnas.1606801113>.
- [42] L.A. Torre, et al., "Ovarian cancer statistics, 2018, *CA Cancer J Clin* 68 (4) (Jul. 2018) 284–296, <https://doi.org/10.3322/caac.21456>.
- [43] Y. Qian, et al., Clinicopathological and prognostic significance of WW domain binding protein 5 expression in papillary thyroid carcinoma, *BioMed Res. Int.* 2019 (2019), 1791065, <https://doi.org/10.1155/2019/1791065>.
- [44] C. Ward, P. Cauchy, P. Garcia, J. Frampton, M.A. Esteban, G. Volpe, High WBP5 expression correlates with elevation of HOX genes levels and is associated with inferior survival in patients with acute myeloid leukaemia, *Sci. Rep.* 10 (1) (Feb. 2020) 3505, <https://doi.org/10.1038/s41598-020-60480-x>.
- [45] Y. Pommier, A. Nussenzweig, S. Takeda, C. Austin, Human topoisomerases and their roles in genome stability and organization, *Nat. Rev. Mol. Cell Biol.* 23 (6) (Jun. 2022) 407–427, <https://doi.org/10.1038/s41580-022-00452-3>.
- [46] A. Faggad, et al., Topoisomerase II $\alpha$  mRNA and protein expression in ovarian carcinoma: correlation with clinicopathological factors and prognosis, *Mod. Pathol.* 22 (4) (Apr. 2009) 579–588, <https://doi.org/10.1038/modpathol.2009.14>.
- [47] Y. Gao, et al., TOP2A promotes tumorigenesis of high-grade serous ovarian cancer by regulating the TGF- $\beta$ /smad pathway, *J. Cancer* 11 (14) (2020) 4181–4192, <https://doi.org/10.7150/jca.42736>.
- [48] S. Srikantan, et al., Translational control of TOP2A influences doxorubicin efficacy, *Mol. Cell Biol.* 31 (18) (Sep. 2011) 3790–3801, <https://doi.org/10.1128/MCB.05639-11>.
- [49] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, *Proc. Natl. Acad. Sci. USA* 98 (9) (Apr. 2001) 5116–5121, <https://doi.org/10.1073/pnas.091062498>.
- [50] A.i. Saeed, et al., TM4: a free, open-source system for microarray data management and analysis, *Biotechniques* 34 (2) (Feb. 2003) 374–378, <https://doi.org/10.2144/03342mt01>.