

RESEARCH

Open Access



Biological and clinical relevance of correlated expression levels of coding and long noncoding RNAs in HPV16 positive cervical cancers

Abhisikta Ghosh^{1†}, Abarna Sinha^{1†}, Arnab Ghosh¹, Somrita Roy¹, Sumana Mallick¹, Vinoth Kumar¹, Sonia Mathai², Jaydip Bhaumik², Asima Mukhopadhyay³, Saugata Sen², Aditi Chandra², Arindam Maitra¹, Nidhan K. Biswas¹, Partha P. Majumder^{1,4} and Sharmila Sengupta^{1,5*}

Abstract

Human papillomavirus (HPV) drives cervical cancer (CaCx) pathogenesis and viral oncoproteins jeopardize global gene expression in such cancers. In this study, our aim was to identify differentially expressed coding (DEcGs) and long noncoding RNA genes (DElncGs) specifically sense intronic and Natural Antisense Transcripts as they are located in the genic regions and may have a direct influence on the expression pattern of their neighbouring coding genes. We compared HPV16-positive CaCx patients (N=44) with HPV-negative normal individuals (N=34) by employing strand-specific RNA-seq and determined the relationships between DEcGs and DElncGs and their clinical implications. By performing Gene set enrichment and protein–protein interaction (PPI) analyses of DEcGs, we identified enrichment of processes crucial for abortive virus life cycle and cancer progression. The DEcGs formed 16 gene clusters which we identified through Molecular Complex Detection (MCODE) plugin of Cytoscape. All the gene clusters portrayed cancer-related functions. We recorded significantly correlated expression levels of 79 DElncGs with DEcGs at proximal genomic loci based on Pearson's Correlation coefficients. Of these gene pairs, 24 pairs portrayed significantly altered correlation coefficients among patients, compared to normal individuals. Of these, 6 DEcGs of 6 such gene pairs, belonged to 5 of the identified gene clusters, one of which was survival-associated. Out of the 24 correlated DEcG: DElncG pairs, we identified 3 pairs, where expression of both members was significantly associated with patient overall survival. The findings justify the cooperative roles of these gene pairs, in patient prognostication, thereby bearing immense potential for translation. Thus, elucidation of correlative strengths between paired DElncGs and DEcGs in patient and normal samples, could serve as a foundation for identification of therapeutic and prognostic targets of HPV16-positive CaCx.

Keywords Coding gene, HPV16-positive cervical cancer, Long noncoding RNA gene, Noncoding Natural Antisense Transcript, Patient overall survival, Sense intronic, Strand-specific RNA-sequencing

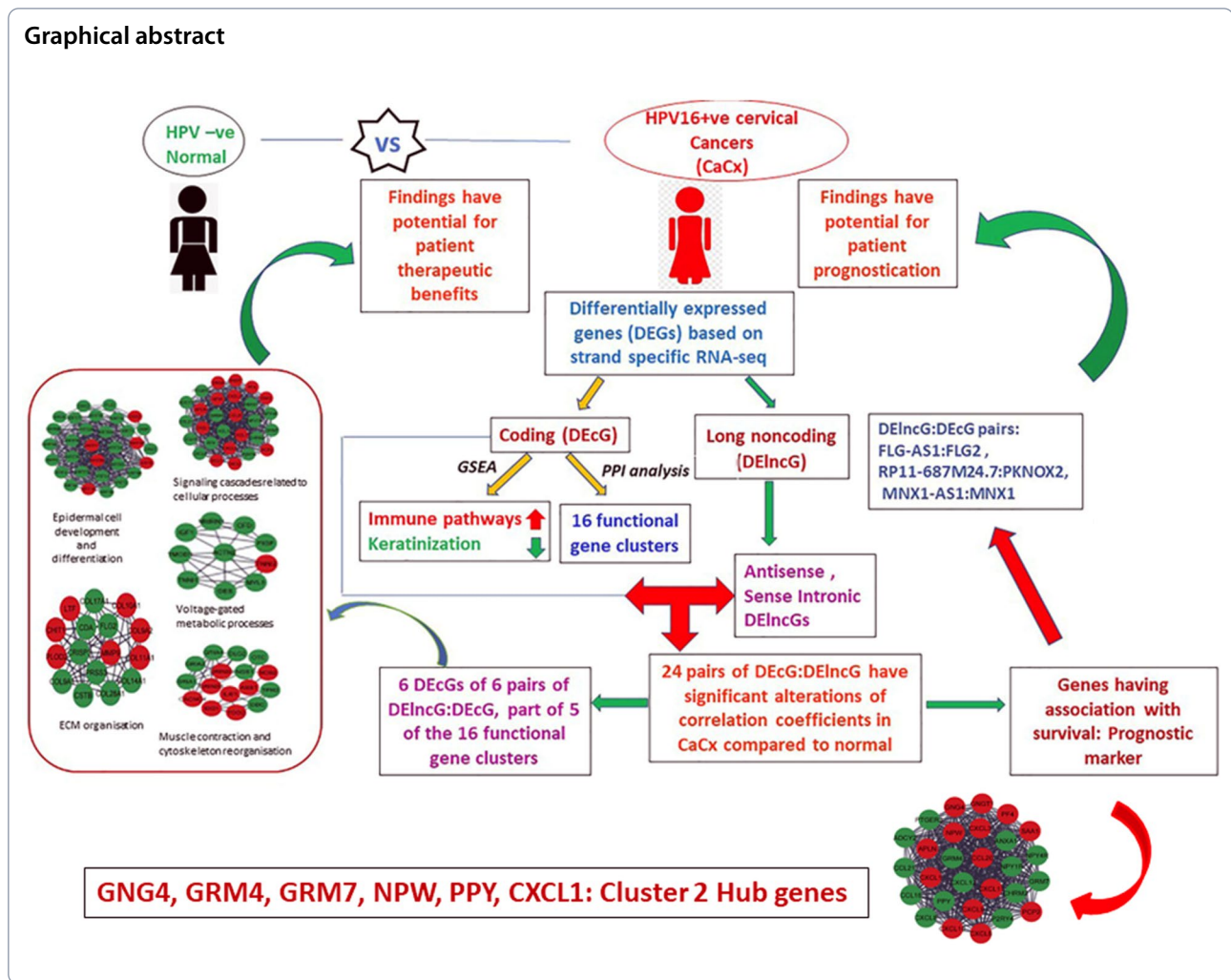
[†]Abhisikta Ghosh and Abarna Sinha have contributed equally to this manuscript and are considered as joint first authors.

*Correspondence:
Sharmila Sengupta
sharmilasg@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



Introduction

Cervical cancer (CaCx) continues to be the second most common cancer among Indian women with high mortality rates (<http://cancerindia.org.in/globocan-2020-india-factsheet/>). Majority of such cancers (>70%) are caused by persistent infection with high-risk human papillomavirus (HPV), especially HPV16 [1]. Enhanced expression of HPV16 encoded oncoproteins E6 and E7 in the cervical epithelium, perturbs the p53 and pRb pathways in the host, leading to genomic instability and causing transformation of the cervical epithelium, that ultimately leads to invasive cancer [2, 3]. Thus, crosstalk between viral and host factors that impact upon host gene expressions are likely to influence HPV related CaCx pathogenesis.

In an earlier study [4], employing microarray-based assay, we showed that HPV16 E7 orchestrates the gene expression profiles of HPV16-positive CaCx cases by interacting with a long noncoding RNA (lncRNA), HOTAIR. A report on HPV16 positive CaCx [5]

employing 3 such tumor tissues and matched adjacent normal samples identified the differential expression of lncRNAs in HPV-driven CaCx, with the potential of further being developed as diagnostic, prognostic and therapeutic markers of such cancers. The involvement of lncRNAs in HPV-driven cancers is also clearly depicted in a review by Casarotto et al. [6]. Current evidence also indicates that lncRNAs play a regulatory role in cancer progression [7]. Specifically, lncRNAs, epigenetically regulate gene expression, support macromolecular complex assembly through scaffold formation, bind and inactivate miRNAs through a sponging effect and regulate mRNA stability [8]. A proportion of lncRNAs also function as Natural Antisense Transcripts (ncNATs) and regulate the expression of sense coding gene at the same locus by various mechanisms [9, 10] These are *cis*-acting in nature and are situated within the genic regions, possessing the capability to regulate the expression of their adjacent coding genes. Another major proportion of the lncRNAs

are long intergenic noncoding RNAs or lincRNAs, which are mostly *trans*-acting and located in the intergenic regions. They may regulate the expression pattern of one or more coding genes, located in the vicinity or at distant loci. We have already performed a detailed study on the lincRNAs [11], employing the same sets of samples used in our current study, to demonstrate that they play a major role in the regulation of the coding genes in HPV16 related CaCx pathogenesis. Thus, prompted by such observations we undertook the current study to dissect the gene expression profiles of both coding genes and *cis*-acting lincRNAs (ncNATs and sense intronics) associated with HPV16-positive CaCx patients employing high-throughput strand-specific RNA sequencing (ssRNA-seq) technology, through which we can estimate the precise expression of various lincRNA types.

Some recent genome-wide studies [9, 12] have provided evidence that within the same tissues or cells, there is a positive correlation between expression of sense genes and the corresponding antisense genes rather than a negative correlation. Therefore, perturbation of the expression of one or both partners of the ncNATs: protein coding gene pairs, associated with disease pathogenesis, could be of immense therapeutic relevance for various diseases including cancers [13].

Thus, we hypothesize that harnessing the expression profiles of lincRNAs using ssRNA-seq, together with the protein coding transcripts in CaCx tissues, is likely to provide novel insights on CaCx pathogenesis, and to pinpoint clinically relevant targets. To test this, we determined the biologically relevant genes and pathways portraying dysregulated expression in HPV16-positive CaCx patients, in comparison to HPV-negative normal individuals. We also investigated the nature of correlation in the expression levels between co-expressed ncNAT/sense intronic and protein coding gene pairs in patients and normal individuals. Subsequently, we considered only those gene pairs that revealed significantly altered correlation co-efficient in patients, as opposed to the normal individuals, for further interpretation of their biological relevance. We, thereby, identified putative actionable target genes of the correlated ncNAT/sense intronic: protein coding gene pairs in patients, which belonged to biologically relevant gene clusters, identified through protein–protein interaction (PPI) analysis of our dataset. Finally, we determined the impact of the correlated genes on patient survival using the survival data available in our cohort.

Materials and methods

Subjects and samples

Tissue samples from married non-pregnant women, aged 28–50 years (median age: 43 years), who underwent

hysterectomy and did not have any prior history of cervical dysplasia or malignancy were considered as healthy individuals or controls. Tissues from CaCx patients were collected from women aged 35–78 years (median age: 54 years) and considered as patients or cases. The cervical biopsies were histopathologically confirmed as malignant or non-malignant. All samples, from both healthy individuals and cancer patients, were collected with written informed consent. The study was approved by the Institutional Ethics Committees of our clinical collaborators such as Tata Medical Center, Kolkata and College of Medicine and Jawaharlal Nehru Memorial Hospital, Kalyani and the National Institute of Biomedical Genomics (NIBMG), Kalyani.

Sample processing and sequencing

Cervical biopsy tissues were collected in RNAlater. Genomic DNA and total RNA were isolated from these tissues using QIAamp DNA mini kit and RNeasy mini kit (Qiagen) respectively, following the manufacturer's protocol. Quality and integrity of RNA was determined using Agilent Bioanalyzer 2100. Details regarding DNA isolation, HPV screening and type identification, are described in our earlier studies [4, 14]. All the samples were tested for the presence of HPV and classified as HPV-negative or positive. Of the HPV positive samples, we selected only those with HPV16 infection as HPV16 is the most prevalent type in CaCx cases in India. Samples showing presence of both HPV18 and HPV16 were excluded from the study. For this study, we compared the HPV-negative normal samples (n=34) with HPV16-positive CaCx samples (n=44). We excluded the HPV positive normal samples because of uncertainty in ascertaining the transient or persistent nature of the infection.

About 500 ng of total RNA isolated from these samples with RNA Integrity Number (RIN) ≥ 5.5 , was considered for library preparation using TruSeq Stranded Total RNA Library Prep kit (Illumina). The library quality was assessed using Agilent Bioanalyzer 2100 system and then sequenced in Nova-Seq 6000 (Illumina) to generate paired-end reads of 100 bases. The details of the sequencing quality and coverage are provided in Supplementary Table S1.

Alignment of RNA sequencing data and identification of differentially expressed genes (DEGs)

A community standard approach was adopted for analysis that included using FastQC-0.11.7 (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>) for QC analysis, STAR aligner (STAR-2.6.0c) [15] for alignment of RNA-sequence reads against human reference sequence (GRCh37 primary assembly) and HTSeq (HTSeq-0.11.0) [16] for generation of raw expression

count data for all coding and noncoding genes, with stranded mode using Homo sapiens GRCh37.87 GTF (Ensembl) as the base sequence. To filter out genes with low count, only genes that were expressed (at least one transcript) in at least 50% samples in any of the two groups (a) HPV negative normal individuals, and (b) HPV16 positive CaCx patients or both, were included in further analysis. Differential gene expression analysis and generation of normalised expression counts were performed with DESeq2 [17] package using HTSeq-count data for both coding and lncRNA genes.

As discussed earlier, there are different types of lncRNAs. But for the present analysis, we considered only the ncNATs and sense intronic transcripts that were of length >200 nucleotides, as identified from LNCipedia and Ensembl database. The transcripts that did not fulfil these criteria were removed from the study. Single exon transcripts, if identified, were removed because most of such noncoding genes, by virtue of being novel, remain to be experimentally validated and may represent background noise from DNA contamination because these transcripts lack a splice junction [18].

For our analysis, we considered only the genes (coding and lncRNA) for which there was significant ($p < 0.05$ after correction using the Benjamini-Hochberg [BH] procedure) differential expression and a high expression fold-change (FC), $|\log_2(\text{FC})| \geq 2$. TPM (transcripts per million) values were calculated using the TPMCalculator tool [19].

We also downloaded the STAR counts for HPV16 positive CaCx patients ($n = 165$) and adjacent normal tissues ($n = 3$) from the GDC portal (<https://portal.gdc.cancer.gov/>). DEGs between patients and normal tissues were identified using DESeq2 based on the same criteria ($|\log_2(\text{FC})| \geq 2$, FDR corrected $p < 0.05$) as employed in our study.

Expression analysis of DEGs in HPV16-positive patients compared to HPV-negative normal individuals using quantitative Real time PCR (qRT-PCR)

cDNA was generated from RNA (about 400 ng) as described in our earlier study [4]. The gene expression of the DEGs was determined in the HPV-negative healthy individuals ($N = 26$) as well as in HPV16-positive CaCx patient tissue ($N = 26$) samples using SYBR Green based qRT-PCR on ABI-Quant Studio 5. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as internal control. The fold change was calculated by $2^{(-\Delta\Delta C_t)}$ method. The list of the primers used for the assay is provided in Supplementary Table S2.

Visualisation and characterisation of DEGs

For the visualisation of DEGs, volcano plot and heatmap was generated using ggplot and pheatmap package (<https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf>), respectively in R.. Pathway analysis was performed using Gene Set Enrichment Analysis (GSEA) [20] and pathways were considered as significant if their FDR corrected p was < 0.05 . PPI network of the DEGs were predicted using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database Version 11.0 [21], considering an interaction score ≥ 0.7 (high confidence). The PPI network was visualised using Cytoscape (version 3.8.2) and the plug-in MCODE (Molecular Complex Detection) was used to extract functional gene clusters in the PPI network using default parameters. Hub genes of the functional clusters were identified through Cytoscape, based on highest degree of connectivity. Functional enrichment of the gene members of the clusters was carried out using DAVID (Database for Annotation, Visualisation and Integrated Discovery) [22, 23]. lncGs and their coding gene partners were identified from the Ensembl database. Survival analysis of the gene pairs was done employing KM Plotter [24], using default settings.

Statistical analysis

Statistical analyses were done using R and online available tools. Pairwise Pearson's correlation coefficients (r) between genes (antisense or sense intronic lncRNA genes and their sense coding gene partners) were calculated using their normalised expression, i.e., TPM values for both CaCx patients ($n = 44$) and normal individuals ($n = 34$). Differential correlation analysis of the gene pairs between cancer patients and healthy individuals was performed using Fisher's z transformation of r (<http://vassarstats.net/rdiff.html>). The crude p values thus generated were further used to determine the adjusted p -values (after FDR correction) using an in-house developed R code. All the statistical tests were two-sided and a result was considered as significant when the adjusted p value was < 0.05 (after FDR correction).

Results

Expression patterns of coding and long noncoding RNA genes in HPV16-positive CaCx patients compared to HPV-negative healthy individuals

About 100 million RNA sequencing reads per sample were mapped to a total of 55,638 genes (coding and noncoding). Of these, 22,208 genes expressed at least 1 transcript in at least 50% of HPV16-positive patients or HPV-negative normal individuals. These genes comprised 16,708 protein coding and 5,070 lncRNA genes

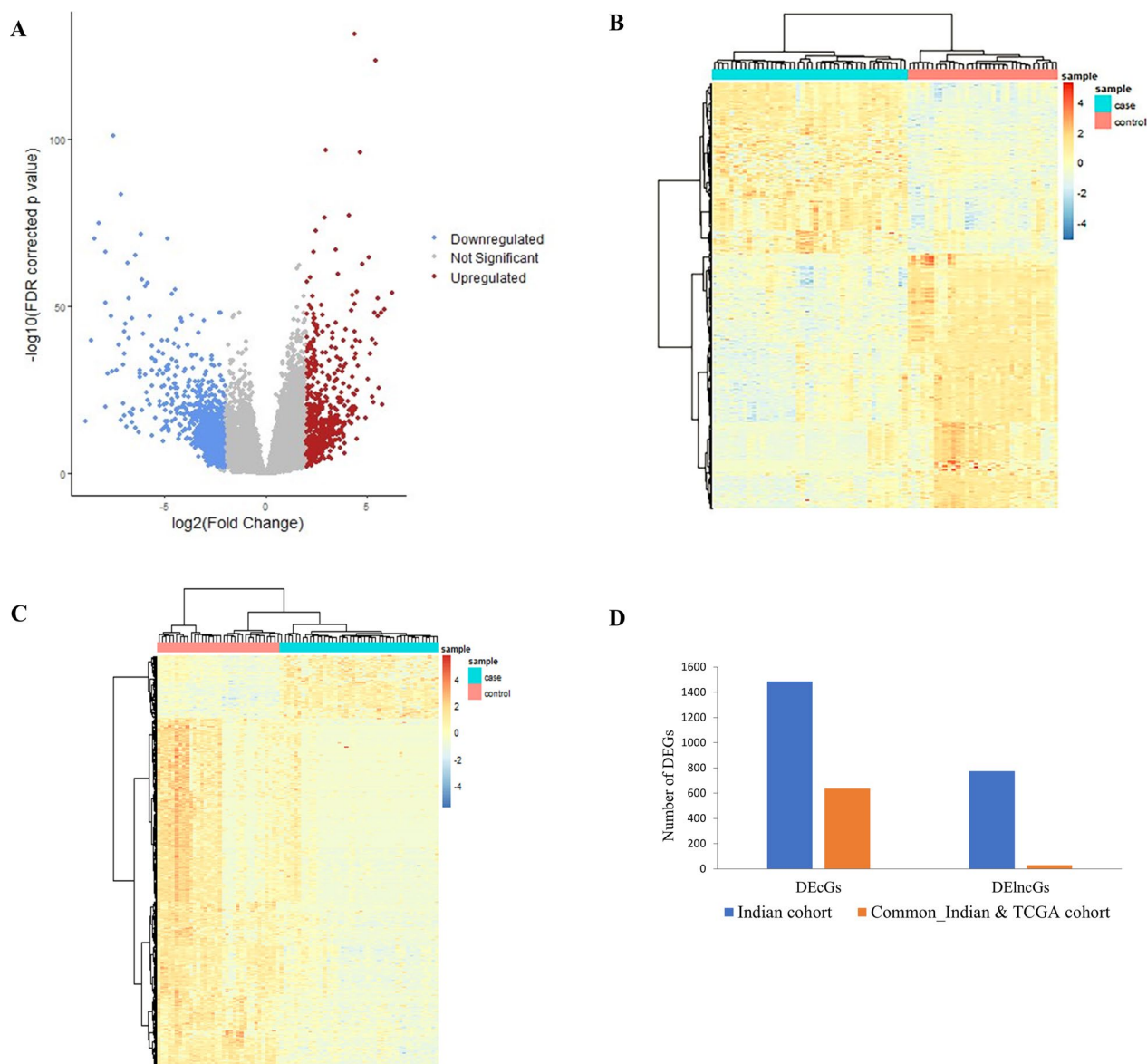


Fig. 1 Visualisation of DEGs. **A** Volcano plot depicting the Differentially Expressed Genes (DEGs) in HPV16-positive CaCx patients ($n = 44$) as compared to HPV-negative normal individuals ($n = 34$). The X-axis represents $\log_2(\text{Fold Change})$ and Y-axis the $-\log_{10}(\text{FDR corrected } p \text{ value})$. Red dots represent the differentially expressed upregulated genes [FDR corrected $p < 0.05$ and $\log_2(\text{Fold Change}) \geq 2$]. Blue dots represent the differentially expressed downregulated genes [FDR [BH] corrected $p < 0.05$ and $\log_2(\text{Fold Change}) \leq -2$]. Grey dots represent genes that were not differentially expressed. **B–C** Heatmap demarcating the gene expression profiles of CaCx patients from the normal individuals. The Heatmap represent normalized counts, which were \log_2 transformed after adding constant 1 to all values of **B** the differentially expressed coding genes (DEcGs) and **C** the differentially expressed lncRNA genes (DElncGs). The rows represent the gene names, and the columns represent the samples (blue bar represents the CaCx patient samples and pink bar represents the samples from normal individuals). **D** Comparison of the DEcGs and DElncGs between our cohort and TCGA-CESC cohort

(Fig. 1A). For subsequent analysis, genes that were differentially expressed in CaCx patients as compared to HPV-negative healthy individuals, were identified based on $|\log_2(\text{FC})| \geq 2$, with FDR corrected $p < 0.05$. We identified 1486 DEcGs and 775 DElncGs, which are marked in red (upregulated genes) and blue (downregulated genes)

on Fig. 1A. The details of the DEGs are provided in Supplementary Table S3 and the expression pattern is visually represented in Fig. 1B (for DEcGs) and Fig. 1C (for DElncGs).

We also identified the DEGs (both coding and lncRNA genes) between HPV16 positive CaCx patients ($n = 165$)

of the TCGA- CESC cohort compared to normal individuals ($n=3$) based on the same criteria i.e. $|\log_2(\text{FC})| \geq 2$, with FDR corrected $p < 0.05$. Thereby, we identified 3166 DEcGs (1828 upregulated and 1338 downregulated) and 1385 DElncGs (838 upregulated and 547 downregulated). About 637 DEcGs (42.8%) were found to be common between our cohort and TCGA-CESC cohort whereas for lncRNA genes the number of differentially expressed common genes was very low (3.74%), as represented in Fig. 1D. Such differences may be attributable to difference in experimental designs and analytic pipeline (unstranded mode in TCGA and stranded approach in our cohort) as well as ethnic differences between the 2 cohorts.

Validation of the expression of some DEGs in HPV16-positive CaCx patients compared to HPV-negative normal individuals using qRT-PCR

We validated the expression levels of some of the DEGs (both coding and noncoding) using qRT-PCR. For validation, we selected the upregulated DEcG and DElncG (FOXD3 and RP4-792G4.2) out of the 3 pairs that were associated with patient-survival. Upon analysis, significant upregulation of both FOXD3 and RP4-792G4.2 were observed among HPV16 positive CaCx patients compared to normal individuals, validating our finding from RNA-seq data. The fold change of the DEGs in patients and normal individuals are depicted in Supplementary Figure S1.

Pathways enriched and depleted in HPV16-positive CaCx patients in comparison with HPV-negative normal individuals

We then performed GSEA from MSigdb (msigdb.v7.4.symbols.gmt) considering only the GO (Gene Ontology) Biological Processes. In comparison to healthy individuals, genes involved in both humoral and cell mediated immune processes (Immunoglobulin family members, VTCN1, CD70, FOXJ1, LAMP3, ICAM4 etc.) and cell cycle processes like cell migration, chromosome segregation (CDKN2A, SMC1B, MCM2, STAG3, E2F1) appeared to be significantly (FDR corrected $p < 0.05$) enriched in CaCx patients (Supplementary Figure S2). Similarly, the processes that were found to be significantly depleted in the CaCx patients included keratinization (KRT family genes), cornification (LCE family genes) and epithelial cell development and differentiation (Supplementary Figure S3). The details of all the processes are shown in Supplementary Table S4. This clearly reflected the functional distinctiveness of the depleted processes, as compared to the enriched processes.

Next, we explored the PPI network among the DEcGs by using the STRING database. We then determined the

GO Biological processes corresponding to the upregulated and downregulated DEcGs from STRING, the details of which are provided in Supplementary Table S5 and S6, respectively. The findings replicated the processes identified through GSEA. In addition, the upregulated genes were mostly involved in chromatin assembly, cellular differentiation, cytokine mediated signalling pathways, DNA replication, G1-S transition, Extracellular Matrix Remodelling and Organization. The downregulated genes, on the other hand, were associated with cell development and differentiation, cellular signalling, cell-cell adhesion, cell death, lipoygenase activity, etc.

Significantly correlated expression levels of DElncGs (ncNATs and sense intronic) with their corresponding sense DEcGs at the same location

We have previously identified global transcriptional reprogramming mediated by HPV16 encoded oncoprotein E7 through expression modulation and functional inactivation of lncRNA, HOTAIR [4]. Therefore, in this study we focussed on DElncGs to determine their crosstalk with DEcGs in CaCx pathogenesis. Of the 775 DElncGs among the CaCx patients, 274 encoded ncNATs, 45 were sense intronic, while the remaining were other types of lncRNA genes. For further analysis, we considered only the ncNATs and sense intronic genes, because they are known to regulate their protein coding gene counterparts at the same locus [13], thus bearing biological relevance. We identified coding genes in the neighbourhood of these 319 DElncGs (274 antisense and 45 sense intronic), through Ensembl database, followed by identification of 83 DElncG:DEcG pairs and a single sense intronic: ncNAT gene pair, for which both members of the pair were significantly and differentially expressed ($|\log_2(\text{FC})| \geq 2$ and FDR corrected $p < 0.05$).

Using the TPM values, we further determined the significantly correlated DElncG:DEcG pairs in CaCx patients ($n=44$) as well as in normal individuals ($n=34$). We observed that of the 83 such DElncG:DEcG pairs, 79 pairs were significantly and positively correlated (FDR corrected $p < 0.05$), i.e., co-expressed, in both the CaCx patients and normal individuals, one pair (RP11-256L6.3: HAL) showed significant positive correlation among patients only, while two pairs (NOVA1-AS1:NOVA1 and RP1-40E16.11: TUBB2A) showed significant positive correlation only in the normal individuals. The remaining pair RP11-96C23.10: FAM25A failed to show significant correlation among both the patients and normal individuals. Both members of all these gene pairs showed expression dysregulation in the same direction, excepting for eight pairs. The details of the findings are provided in Supplementary Table S7. As functions of most of the noncoding partners remain unknown, correlation of

these ncNATs/sense intronic genes with their neighbouring DEcGs, reveal their association in the regulation of their respective coding gene expression.

Of the 79 significantly correlated gene pairs in both patients and normal individuals, we identified 24 pairs that showed significant difference in the correlation coefficient between the patients and normal individuals. This gene set included 22 pairs of ncNATs: coding genes and 2 pairs of sense intronic: coding genes (RP11-687M24.7: PKNOX2 and RP11-483C6.1: NOVA1). The correlations were significantly higher in the patients than in the normal individuals for 17 pairs, while for the remaining the pattern was the opposite (Supplementary Table S7). The significant alteration in the strength of correlation between such gene pairs in tumorigenesis, calls for functional validation of the finding. The remaining 55 correlated pairs, which did not show any difference in correlative strength between the normal and patient samples, may jointly play a role in cervical tissue specific functions.

Identification of putative actionable target gene pairs of DElncGs (ncNATs and sense intronic) and DEcGs associated with tumorigenesis

To decipher the functional relevance of the ncNAT/sense intronic: coding gene pairs in CaCx, we considered the 24 gene pairs that showed significant alteration in strength of correlation between the patients and the normal individuals (Supplementary Table S7), with 70% showing increased correlated coefficients as opposed to 30% with reduced correlated coefficients. We further considered these for determining their potential as actionable targets of HPV16 related CaCx.

To accomplish this, we attempted to identify if any of the coding gene counterparts of these 24 gene pairs were part of functionally relevant gene clusters. Thus, we performed PPI network analysis of 1486 DEcGs recorded in our study, using STRING database (interaction score > 0.7). This revealed 1324 nodes (genes) and 3445 edges (interactions), as shown in Supplementary Figure S4. The genes clustered into 16 modules identified through MCODE plug-in (Degree cut-off ≥ 2 , node score cut-off 0.2, K-core 2 and MCODE score ≥ 5) of

Cytoscape. Thereby we determined the GO processes corresponding to these clusters employing DAVID (FDR corrected $p < 0.05$), revealing their association with cancer related functions. Of the 24 tumor specific ncNAT/sense intronic: coding gene pairs, 6 coding genes appeared to be a part of some of these gene clusters as depicted in Fig. 2A–E and the detailed analysis of these 6 clusters is summarised in Table 1. The details of the remaining clusters, which did not incorporate any of the coding genes of the 24-tumor specific ncNAT/sense intronic: coding gene pairs, are provided in Supplementary Table S8. Two of these 6 gene pairs had concordant upregulated expression (CTD-3214H19.6: PCP2 and RP11-845M18.6: KRT86), 3 pairs had concordant downregulated expression (AC019349.5: KRT13; FLG-AS1: FLG2 and AC053503.6: DES), and one pair (RP11-74H8.1: CACNG4) portrayed discordant expression with upregulation of the coding gene and downregulation of the noncoding partner, with respect to the healthy individuals. These gene clusters were associated with biological processes such as, signalling cascades related to cellular processes, Epidermal cell development and differentiation, ECM organisation, Voltage-gated metabolic processes and Muscle contraction and cytoskeleton reorganisation as indicated in Table 1. These findings indicate that perturbation of the correlated expression levels of these six noncoding: coding gene pairs is likely to impact the expression of other genes of the clusters, thereby, affecting key biological processes related to cancer development and progression.

Deciphering the prognostic implication of tumor-specific significantly correlated DElncGs (ncNATs and sense intronic):DEcGs pairs in CaCx patients

To estimate the prognostic value of the 24 co-expressed and significantly correlated DElncG: DEcG pairs in CaCx patients, we determined the impact of the expression status of these DEcGs and DElncGs on patient overall survival (~4 years) in our dataset, employing KM Plotter with default settings. Thereby, we identified 2 gene pairs with concordant downregulated expression (FLG-AS1:FLG2 and RP11-687M24.7:PKNOX2) and 1 pair (MNX1-AS1:MNX1) with concordant upregulated

(See figure on next page.)

Fig. 2 Visualisation of the gene clusters formed by considering the coding gene counterparts of the gene pairs DElncG (ncNAT/sense intronic): DEcG that revealed significantly altered correlation among HPV16-positive CaCx patients ($n = 44$) in comparison with HPV-negative normal individuals ($n = 34$). The gene clusters correspond to **A** PCP2 (CTD-3214H19.6: PCP2), **B** KRT13 (KRT13:AC019349.5) and KRT86 (KRT86:RP11-845M18.6), **C** FLG2 (FLG2: FLG-AS1), **D** DES (DES: AC053503.6), **E** CACNG4 (CACNG4:RP11-74H8.1). The green nodes in the clusters represent genes with downregulated expression, whereas red nodes represent the genes with upregulated expression in CaCx patients as compared to normal individuals. This was generated using the MCODE-plugin of Cytoscape using the parameters-Degree cut-off ≥ 2 , node score cut-off 0.2, K-core 2 and MCODE score ≥ 5

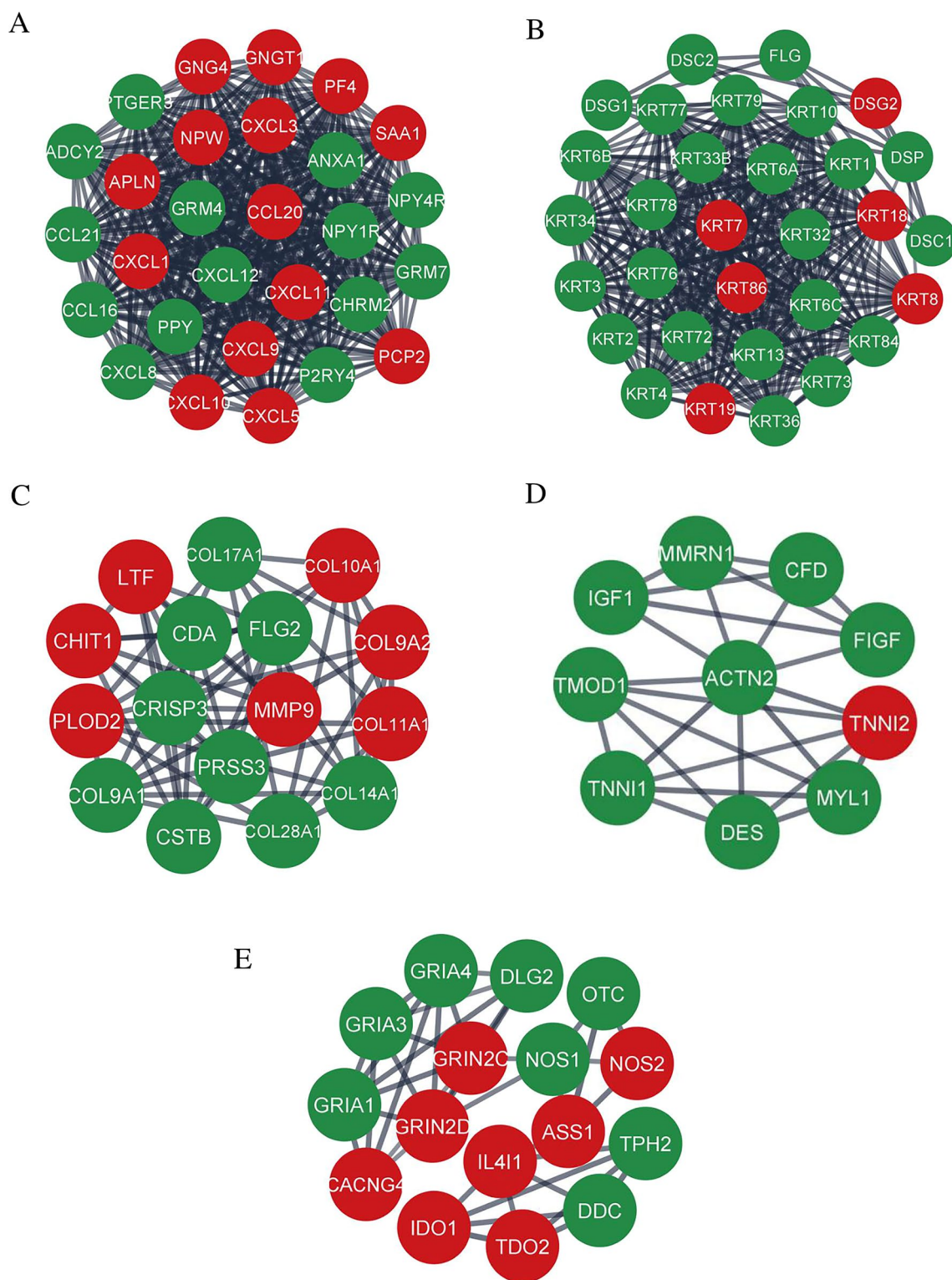


Fig. 2 (See legend on previous page.)

expression among patients, as survival associated genes. Both members of the 3 gene pairs at high expression levels, were significantly (logrank $p < 0.05$) associated

with increased patient overall survival compared to those patients with low expression levels, as indicated in Fig. 3A–F.

Putative actionable target gene pairs ncNATs: coding genes	# of nodes and cluster number	Functions of the cluster	Name of the genes in the cluster
CTD-3214H19.6: PCP2 (Concordant expression)	Nodes: 28 Cluster No: 2	Involved in signaling processes like GPCR, chemokine mediated signaling, neutrophil chemotaxis, inflammatory response, cell-cell signaling, positive regulation of T cell migration	ANXA1, APLN, CCL16, CCL20, CCL21, CHRM2, CXCL1, CXCL10, CXCL11, CXCL12, CXCL3, CXCL5, CXCL8, CXCL9, GNG4, GNGT1, GRM4, GRM7, NPW, NPY1R, NPY4R, P2RY4, PF4, PPY, PTGER3, SAA1, ADCY2, PCP2
AC019349.5: KRT13 (Concordant expression)	Nodes: 31 Cluster No: 3	Involved in cornification, keratinisation, epithelial cell differentiation, epidermis development, cytoskeleton organisation, cell-cell adhesion, peptide crosslinking	KRT1, KRT10, KRT18, KRT8, KRT13 , KRT19, KRT2, KRT3, KRT32, KRT33B, KRT34, KRT36, KRT4, KRT6A, KRT6B, KRT6C, KRT7, KRT72, KRT73, KRT76, KRT77, KRT78, KRT79, KRT84, KRT86 , DSP, FLG, DSC1, DSC2, DSG1, DSG2
RP11-845M18.6: KRT86 (Concordant expression)			
FLG-AS1: FLG2 (Concordant expression)	Nodes: 16 Cluster No: 8	Involved in collagen fibril organisation, extracellular matrix organisation, neutrophil degranulation	COL14A1, MMP9, CDA, CHIT1, COL10A1, COL11A1, COL17A1, COL28A1, COL9A1, COL9A2, CRISP3, CSTB, FLG2 , LTF, PLOD2, PRSS3
AC053503.6: DES (Concordant expression)	Nodes: 10 Cluster No: 9	Involved in muscle contraction and platelet degranulation, cardiac muscle contraction	ACTN2, CFD, DES , FIGF, IGF1, MMRN1, MYL1, TMOD1, TNNI1, TNNI2

Table 1 Clusters of coding genes (red- upregulated expression; green- downregulated expression) associated with specific cancer related functions and incorporating the coding genes of the ncNATs: coding gene pairs portraying significant altered correlated expression levels in CaCx patients as compared to normal individuals

Furthermore, to understand the prognostic relevance of the 5 gene clusters, we identified the hub genes of each of these 5 clusters based on the higher degree of connectivity. The hub genes of each cluster and the corresponding degree of connectivity are represented in Supplementary Table S9. The prognostic value of each of these hub genes was explored by KM-Plotter. Among these clusters, all genes belonging to Cluster 2 were densely interconnected with each other, based on which, these were considered as the hub genes. Most of the genes (PPY, CXCL1, GRM4, GRM7, GNG4 and NPW) belonging to this cluster showed significant association with patient survival. Association of higher expression with better patient survival was recorded for PPY, GRM4, GRM7, GNG4 and NPW, while CXCL1 was the only hub gene that was associated with better patient survival at lower expression. Therefore, this cluster of genes appears to be significantly associated with patient survival. The survival curves are depicted in Fig. 4A–F. Moreover, this cluster incorporated the hub gene PCP2, which was significantly correlated with its ncNAT partner, CTD-3214H19.6 and both genes failed to reveal any association with patient survival. On the other hand, the survival associated gene pair, FLG-AS1:FLG2, belonged to cluster 8, where none of the coding hub genes revealed any association with patient survival.

Thus, a substantial proportion of the coding genes and/or their noncoding (antisense/sense intronic) counterparts of such gene pairs with altered relative strength in patients, appear to be of prognostic relevance in our patient cohort. The observation thereby points towards the biological relevance of such correlations in HPV16-positive CaCx patients.

Discussion

To our knowledge, this is the first study to report the global transcriptome profile (including both coding and noncoding) of HPV16 related CaCx in Indian population employing patient and normal individuals with the help of ssRNA-seq. The functions of a large majority of identified DEIncGs remain to be characterized and reported in CaCx. Therefore, we aimed to identify crosstalk, if any, between the DEIncGs (focussing on only the ncNATs and sense intronic genes) and their corresponding DEcG counterparts, to draw insights

related to their functions based on those of the coding gene partners. Further, we determined the biological relevance of such interactions pertaining to identification of putative actionable targets and prognostic markers of HPV16-positive CaCx.

The ncNATs and sense intronic lncRNAs are known to affect expression of their proximal coding genes through multiple ways [25] and their dysregulated expression have been recorded among various cancer types [26, 27] including CaCx [28, 29]. These genes, therefore, play an important role in cancer development, progression [25, 30, 31], therapeutics [13] and could also serve as biomarkers of cancer pathogenesis [32]. In a recent study from our laboratory we delved into the mechanistic aspect underlying the regulatory role of an ncNAT on the expression of its adjacent coding gene through various functional analyses [10]. The analyzed gene pair (AC103563.8: MAL) was one out of the identified 24 co-expressed and significantly correlated DEIncG: DEcG pairs in CaCx patients. We established that AC103563.8-E7-EZH2 axis regulates the expression of MAL, through chromatin inactivation in HPV16 positive CaCx pathogenesis [10].

We observed that several DEIncGs in the CaCx patients showed correlated expression levels with the corresponding proximal DEcGs, concomitant with altered relative relationships (both stronger and weaker), as compared to the normal individuals. Our study, therefore, is the first of its kind to reveal the significant joint role of correlated ncNATs/sense intronic genes and their coding gene counterparts at the expression level in HPV16 related CaCx pathogenesis. This underscores the existence of regulatory circuits between such noncoding and coding gene pairs. Such findings are similar to recent studies on other cancers [9, 33], which recorded that a large majority of the ncNATs revealed positive and significant correlated expression levels with their corresponding sense genes, with stronger correlations among the tumors than in the normal samples. We also identified a large majority of such correlations, which failed to differ in relative strength between cancer and normal individuals, justifying their roles in cervical tissue house-keeping functions. Taken together, our analysis strongly supports the biological relevance of correlated expression levels of ncNATs and sense intronic genes with their

(See figure on next page.)

Fig. 3 Patient overall survival in relation to the expression status of coding and noncoding counterparts, of the DEIncG (ncNAT/sense intronic): DEcG gene pairs that revealed significantly altered correlation (FDR corrected $p < 0.05$) among HPV16-positive CaCx patients ($n = 44$) in comparison with HPV-negative normal individuals ($n = 34$). **A–F** Kaplan–Meier overall survival curve depicting CaCx patients with high expression and low expression of (A) FLG-AS1 (DEIncG), **B** FLG2 (DEcG), **C** RP11-687M24.7 (DEIncG), **D** PKNOX2 (DEcG), **E** MNX1-AS1 (DEIncG) and **F** MNX1 (DEcG), employing KM-Plotter. KM plots with logrank p value < 0.05 were considered significant

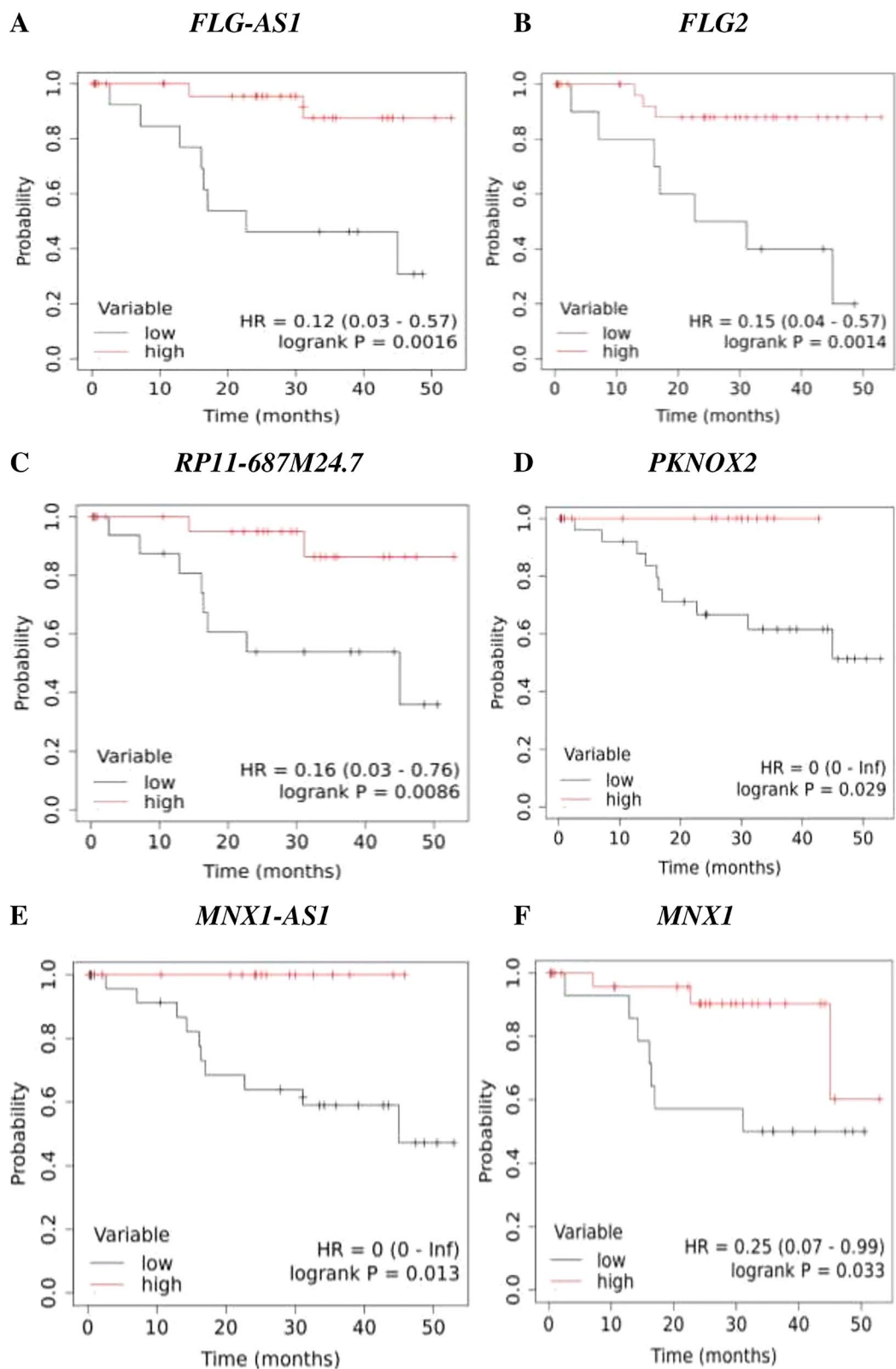


Fig. 3 (See legend on previous page.)

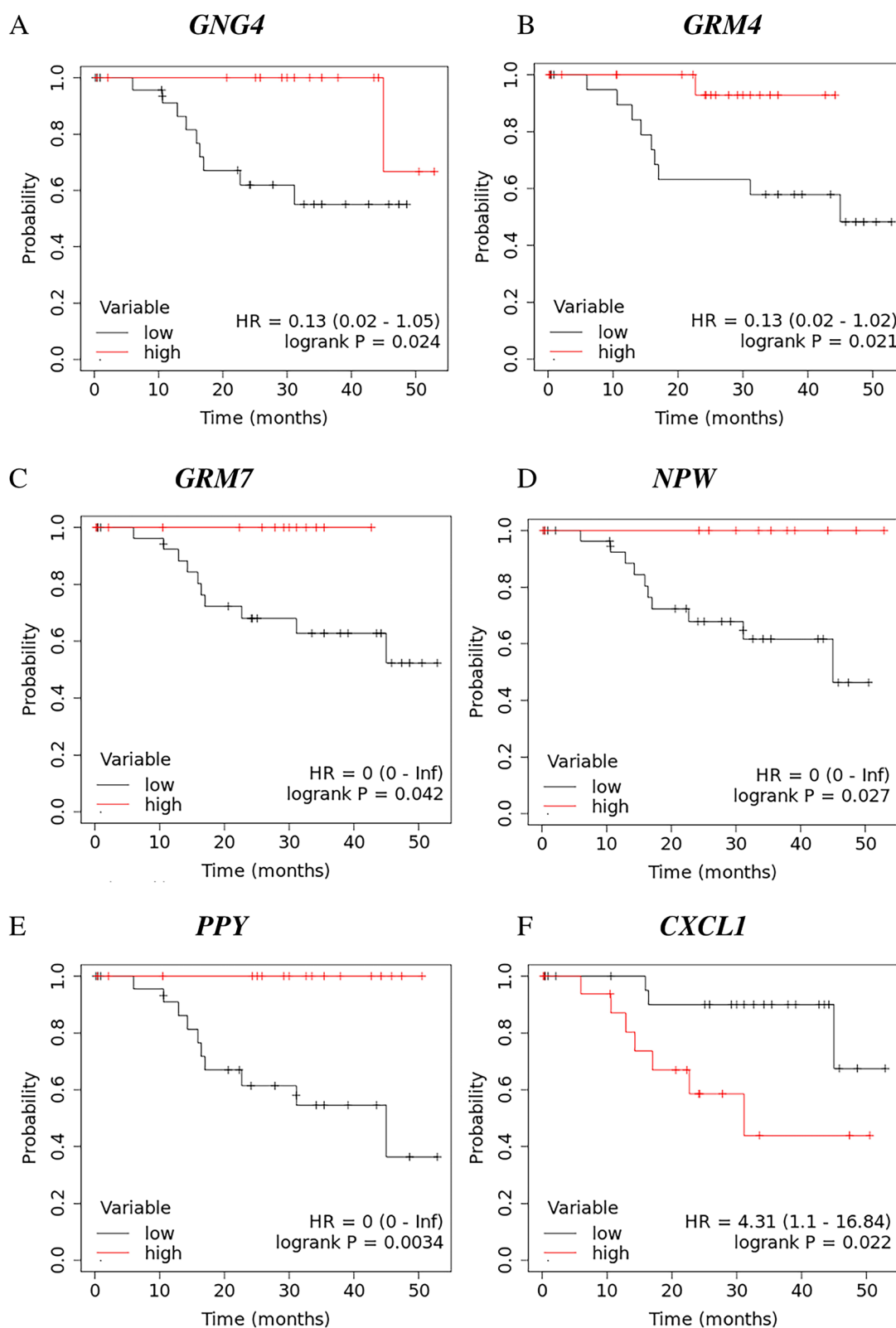


Fig. 4 Patient overall survival in relation to the expression status of coding hub genes of the cluster 2 that were partners of the ncNATs: coding gene pairs portraying significant altered correlated expression in CaCx patients (n=44), as compared to normal individuals (n=34). **A–F** Kaplan–Meier overall survival curve depicting CaCx patients with high expression and low expression of **A** GNG4, **B** GRM4, **C** GRM7, **D** NPW, **E** PPY and **F** CXCL1, employing KM-Plotter. KM plots with logrank p value < 0.05 were considered significant

sense coding genes, both in cervical carcinogenesis as well as in cervical tissue specific functions.

In addition, we identified some functionally relevant gene clusters through PPI analysis of the DEcGs. These clusters were associated with the cancer related processes like keratinocyte and epidermal cell differentiation, cellular signalling cascades for proliferation, migration, cell cycle regulation, post translational modifications, interferon and neutrophils mediated immune pathways, extracellular matrix organisation etc. Thus, these gene clusters could potentially be highlighted as therapeutic targets, subjected to identification of ways and means to target them. We addressed this issue by examining the cross-talk between DElncG (ncNATs and sense intronic):DEcG pairs where the DEcG counterparts belong to such clusters. Our analysis reflected that the coding genes of 6 such gene pairs, were part of 5 large coding gene clusters, comprising of functionally related genes of cancer-associated pathways. Based on this observation, we propose that perturbations of the differential correlative relationships of the noncoding:coding gene pairs, where the coding gene counterpart appears to be a part of a functionally relevant gene cluster associated with CaCx pathogenesis, could serve as an option of disrupting the entire cluster (Table 1). This could thereby open up an interesting avenue of tackling such cancers.

In recent times, targeting ncNATs, to perturb the correlated ncNAT: coding gene pairs by use of antisense oligos (ASOs), is under exploration and some ASOs have been approved by FDA for the treatment of various diseases including cancers [34]. Apart from this, some other possibilities based on advanced genome modification techniques [35] could also be employed for this. Our study therefore lays the foundation for exploring such possibilities. Also, the findings strongly confirm the participation of the DElncGs of these 6 gene pairs in cancer-associated mechanisms, pertaining to HPV16 positive CaCx cases.

Besides revealing the merit of altered significant correlated expression levels of the ncNAT/sense intronic and coding gene pairs in CaCx patients as therapeutically relevant, our study also depicted the prognostic relevance of such pairs, involving both members (FLG-AS1:FLG2, MNX1-AS1:MNX1 and RP11-687M24.7:PKNOX2). As these genes show association with cancer development and are correlated at the expression level, therefore they may also seem to be of importance as prognostic targets, identified for the first time in case of HPV16 related CaCx.

Recent Pan-cancer studies [36] have provided convincing data that underscores the identification of pathway based biomarkers, which happen to be more proficient than single gene biomarkers. With respect to this, we

moved a step further to show that of the 5 functionally relevant gene clusters in CaCx that incorporated 6 significantly correlated noncoding:coding gene pairs, cluster number 2 associated with various signaling processes and incorporating the CTD-3214H19.6: PCP2, also could be characterized as a patient survival-associated gene cluster. In fact, all genes of this cluster including PCP2, were hub genes and 6 such hub genes revealed association with patient survival. While both members of the gene pair CTD-3214H19.6: PCP2 were not associated with survival of patients, however, disruption of their correlated expression levels, could affect genes in cluster 2, and hence patient survival. Therefore, this gene pair could potentially be considered as a prognostic biomarker of HPV16 related CaCx.

Conclusion

Through GSEA and PPI analysis of DEcGs, we identified enrichment of the processes crucial for abortive virus life cycle, cancer progression, immunity, and depletion of epithelial cell differentiation. The dysregulated DEcGs also formed several clusters, with characteristic cancer-related functions. In addition, we unfurled significant correlated expression levels of DElncGs (sense intronic and ncNATs) with DEcGs at proximal genomic loci within our sample cohort. A subset of these gene pairs portrayed significant alterations of correlation coefficients among patients than normal samples. The DEcGs of some such gene pairs showing altered correlative strengths, were part of a few gene clusters implicated in CaCx, including a survival associated gene cluster. Thus, such gene pairs portray the potential for serving as both therapeutic and prognostic targets. Three such pairs were also identified, where expression of both the members of the pairs were significantly associated with overall patient survival. Our study strongly highlights the cooperative role of DElncGs and DEcGs in cervical carcinogenesis, with the potential of translation.

Abbreviations

DEG	Differentially expressed gene
DEcG	Differentially expressed coding gene
DElncG	Differentially expressed long noncoding RNA gene
FDR	False discovery rate
FIGO	International federation of gynecology and obstetrics
GSEA	Gene Set Enrichment Analysis
HPV	Human papillomavirus
LncRNA	Long noncoding RNA
ncNAT	Noncoding Natural Antisense Transcript
NGS	Next generation sequencing
PPI	Protein–protein interaction
RIN	RNA Integrity Number
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TPM	Transcripts per million

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-024-00660-2>.

Supplementary file 1.

Supplementary file 2.

Acknowledgements

We are grateful to the Government of India, Department of Biotechnology, Ministry of Science and Technology, for financial support to this study, through the Systems Medicine Cluster (SyMeC) project (BT/Med-II/NIBMG/SyMeC/2014/Vol. II). We are also grateful to all participating members of SyMeC for their inputs and contributions to this study. We acknowledge Tata Medical Center, Kolkata, and College of Medicine and J.N.M. Hospital, Kalyani, for providing us with the clinical samples for the study; all the members of the Core Technology Research Initiative (CoTeRI) of the National Institute of Biomedical Genomics, Kalyani, India for their technical support during the work; and special thanks also to Council of Scientific and Industrial Research (CSIR), India, for providing research fellowships to Abhisikta Ghosh (JRF and SRF), Department of Science and Technology (INSPIRE), Govt. of India for providing research fellowship to Abarna Sinha (JRF and SRF), Department of Biotechnology, Government of India, for funding Sumana Mallick (Project RA), and Department of Science and Technology, Govt. of India for funding Vinoth Kumar (National Post-doctoral Fellowship), to work on this project. We also acknowledge the support of Prof. Indranil Mukhopadhyay (Indian Statistical Institute, Kolkata), Dr. Samsiddhi Bhattacharjee and Ms. Sahana Ghosh (National Institute of Biomedical Genomics, Kalyani) and Dr. Shreoshi Sengupta (Indian Institute of Science, Bangalore) in data analyses.

Author contributions

A.G.*, A.S.* and S.S.G. conceived the study; A.G.*, A.S.*, S.R., S.M. and V.K. performed the experiments, A.G.*, A.S.*, N.K.B. and S.S.G. wrote the manuscript; N.K.B. and A.G.* developed the RNA-seq data analysis pipeline and was involved in raw data processing, A.G.*, A.S.* and S.S.G. performed downstream data analysis, S.M., J.B., A.M3, S.S., A.C. coordinated patient recruitment, clinical assessments and collection of data on patients and provided patient materials; A.M1. coordinated massively parallel sequencing experiments; P.P.M. provided guidance of the study design and analysis and S.S.G. provided overall guidance of the study; all authors contributed towards reviewing the manuscript.

Funding

The study was supported by Department of Biotechnology, Ministry of Science and Technology, Government of India (BT/Med-II/NIBMG/SyMeC/2014/Vol. II).

Data availability

Sequence data that support the findings of this study have been deposited in the European Nucleotide Archive with the primary accession code PRJEB40877 and secondary accession number: ERP124576. Raw data will be made available upon request.

Declarations

Ethics approval and consent to participate

All samples, from both healthy individuals and cancer patients, were collected with written informed consent. The study was approved by the Institutional Ethics Committees of the collaborating institutions- College of Medicine and Jawaharlal Nehru Memorial Hospital, Kalyani (Ref No.: F-24/PR/COMJNM/IEC/20/50/1 dated February 7, 2020) and Tata Medical Center, Kolkata (Ref No.: EC/GOVT/18/17 dated January 3, 2018). As National Institute of Biomedical Genomics (NIBMG), Kalyani did not independently collect the clinical samples, the Institutional Ethical Committee (IEC) of NIBMG was satisfied that NIBMG members can participate in the research as IECs of the two primary institutions College of Medicine and Jawaharlal Nehru Memorial Hospital, Kalyani and Tata Medical Center, Kolkata collecting data and clinical samples, had approved the undertaking of this research.

Competing interests

The authors declare no competing interests.

Author details

¹National Institute of Biomedical Genomics, PO.: N.S.S. Kalyani, West Bengal 741251, India. ²Tata Medical Center, Kolkata, West Bengal, India. ³Kolkata Gynecological Oncology Trials and Translational Research Group, Kolkata, West Bengal, India. ⁴John C. Martin Center for Liver Research and Innovations, Kolkata, West Bengal, India. ⁵Saroj Gupta Cancer Centre & Research Institute, Kolkata, West Bengal, India.

Received: 24 February 2024 Accepted: 16 August 2024

Published online: 29 August 2024

References

- Muñoz N, Castellsagué X, de González AB, Gissmann L. HPV in the etiology of human cancer. *Vaccine*. 2006;24:S1-10.
- Yim EK, Park JS. The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis. *Cancer Res Treatment Off J Korean Cancer Assoc*. 2005;37(6):319–24.
- Narisawa-Saito M, Kiyono T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins. *Cancer Sci*. 2007;98(10):1505–11.
- Sharma S, Mandal P, Sadhukhan T, Roy Chowdhury R, Ranjan Mondal N, Chakravarty B, et al. Bridging links between long noncoding RNA HOTAIR and HPV oncoprotein E7 in cervical cancer pathogenesis. *Sci Rep*. 2015;5(1):1–15.
- Wang H, Zhao Y, Chen M, Cui J. Identification of novel long non-coding and circular RNAs in human papillomavirus-mediated cervical cancer. *Front Microbiol*. 2017;8:1720.
- Casarotto M, Fanetti G, Guerrieri R, Palazzari E, Lupato V, Steffan A, et al. Beyond microRNAs: emerging role of other non-coding RNAs in HPV-driven cancers. *Cancers*. 2020;12(5):1246.
- Xie L, Yao Z, Zhang Y, Li D, Hu F, Liao Y, et al. Deep RNA sequencing reveals the dynamic regulation of miRNA, lncRNAs, and mRNAs in osteosarcoma tumorigenesis and pulmonary metastasis. *Cell Death Dis*. 2018;9(7):1–16.
- Sharma S, Munger K. The role of long noncoding RNAs in human papillomavirus-associated pathogenesis. *Pathogens*. 2020;9(4):289.
- Choi H, Lee S, Lee M, Park D, Choi SS. Investigation of the putative role of antisense transcripts as regulators of sense transcripts by correlation analysis of sense-antisense pairs in colorectal cancers. *FASEB J*. 2021;35(4):e21482.
- Sinha A, Ghosh A, Ghosh A, Mathai S, Bhaumik J, Mukhopadhyay A, et al. MAL expression downregulation through suppressive H3K27me3 marks at the promoter in HPV16-related cervical cancers is prognostically relevant and manifested by the interplay of novel MAL antisense long noncoding RNA AC103563. 8, E7 oncoprotein and EZH2. *Clin Epigenet*. 2024;16(1):40.
- Sinha A, Ghosh S, Ghosh A, Ghosh A, Mathai S, Bhaumik J, et al. Unfurling the functional association between long intergenic noncoding RNAs (lincRNAs) and HPV16-related cervical cancer pathogenesis through weighted gene co-expression network analysis of differentially expressed lincRNAs and coding genes. *Carcinogenesis*. 2024;45(7):451–462. <https://doi.org/10.1093/carcin/bgae019>.
- Balbin OA, Malik R, Dhanasekaran SM, Prensner JR, Cao X, Wu YM, et al. The landscape of antisense gene expression in human cancers. *Genome Res*. 2015;25(7):1068–79.
- Halley P, Khorkova O, Wahlestedt C. Natural antisense transcripts as therapeutic targets. *Drug Discov Today Ther Strategies*. 2013;10(3): e119.
- Laikangbam P, Sengupta S, Bhattacharya P, Duttgupta C, Singh TD, Verma Y, et al. A comparative profile of the prevalence and age distribution of human papillomavirus type 16/18 infections among three states of India with focus on northeast India. *Int J Gynecol Cancer*. 2007;17(1):107.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–9.

17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):1–21.
18. Cabanski CR, White NM, Dang HX, Silva-Fisher JM, Rauck CE, Cicka D, et al. Pan-cancer transcriptome analysis reveals long noncoding RNAs with conserved function. *RNA Biol.* 2015;12(6):628–42.
19. Vera Alvarez R, Pongor LS, Mariño-Ramírez L, Landsman D. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics.* 2019;35(11):1960–2.
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci.* 2005;102(43):15545–50.
21. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–13.
22. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57.
23. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37(1):1–13.
24. Lánczky A, Gyórfy B. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. *J Med Internet Res.* 2021;23(7): e27633.
25. Zhao S, Zhang X, Chen S, Zhang S. Natural antisense transcripts in the biological hallmarks of cancer: powerful regulators hidden in the dark. *J Exp Clin Cancer Res.* 2020;39(1):1–18.
26. Huang B, Song JH, Cheng Y, Abraham JM, Ibrahim S, Sun Z, et al. Long non-coding antisense RNA KRT7-AS is activated in gastric cancers and supports cancer cell progression by increasing KRT7 expression. *Oncogene.* 2016;35(37):4927–36.
27. Jadalaha M, Gholamalamdari O, Tang W, Zhang Y, Petracovici A, Hao Q, et al. A natural antisense lncRNA controls breast cancer progression by promoting tumor suppressor gene mRNA stability. *PLoS Genet.* 2018;14(11): e1007802.
28. He J, Huang B, Zhang K, Liu M, Xu T. Long non-coding RNA in cervical cancer: from biology to therapeutic opportunity. *Biomed Pharmacother.* 2020;127: 110209.
29. Chen S, Wang J. HAND2-AS1 inhibits invasion and metastasis of cervical cancer cells via microRNA-330-5p-mediated LDOC1. *Cancer Cell Int.* 2019;19(1):1–16.
30. Reis EM, Nakaya HI, Louro R, Canavez FC, Flatschart ÁV, Almeida GT, et al. Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene.* 2004;23(39):6684–92.
31. Perez DS, Hoage TR, Pritchett JR, Ducharme-Smith AL, Halling ML, Ganapathiraju SC, et al. Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Mol Genet.* 2008;17(5):642–55.
32. Mallardo M, Poltronieri P, D'Urso OF. Non-protein coding RNA biomarkers and differential expression in cancers: a review. *J Exp Clin Cancer Res.* 2008;27(1):1–12.
33. Wu W, Wagner EK, Hao Y, Rao X, Dai H, Han J, et al. Tissue-specific co-expression of long non-coding and coding RNAs associated with breast cancer. *Sci Rep.* 2016;6(1):1–13.
34. Stein CA, Castanotto D. FDA-approved oligonucleotide therapies in 2017. *Mol Ther.* 2017;25(5):1069–75.
35. Yang J, Meng X, Pan J, Jiang N, Zhou C, Wu Z, et al. CRISPR/Cas9-mediated noncoding RNA editing in human cancers. *RNA Biol.* 2018;15(1):35–43.
36. Su K, Yu Q, Shen R, Sun SY, Moreno CS, Li X, et al. Pan-cancer analysis of pathway-based gene expression pattern at the individual level reveals biomarkers of clinical prognosis. *Cell reports methods.* 2021;1(4): 100050.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.