miR-34b associates with epithelial-mesenchymal transition and regulates *BMP7*, *CAV1*, *ID2* and *FN1* in human cervical cancer

Nalini Venkatesan[1], Ashley Xavier[1,2], Sindhu K.J.[1], Sandhya Sundhram[3], Himanshu Sinha[1,2,4,*], Devarajan Karunagaran[1,*]

[1] Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India

[2] Centre for Integrative Biology and Systems Medicine, Indian Institute of Technology Madras, Chennai, India

[3] Department of Pathology, Sri Ramachandra University, Chennai, India

[4] Robert Bosch Centre for Data Science and Artificial Intelligence, Indian Institute of Technology Madras, Chennai, India

* Corresponding authors: Devarajan Karunagaran (karuna@iitm.ac.in), Himanshu Sinha (sinha@iitm.ac.in)

Abstract

Purpose: We aimed to identify the epithelial-mesenchymal transition and associated putative targets of the *miR-34* family in transcriptomic data of cervical epithelial squamous cell carcinoma (CESC) to find new therapeutic targets for better disease management.

Methods: A combined computational analysis of the *miR-34* family; gene expression in heterogeneous primary CESCs derived from TCGA; and the integration of *miR-34b* and EMT-regulated genes was performed. Four EMT-associated *miR-34b* gene targets were analysed in primary human CESC and non-cancerous cervical tissues by qRT-PCR. Effects of endogenous *miR-34b* expression and its associated gene modulations in cervical cancer cells (C33A and HeLa) were analysed using qRT-PCR, western blotting and immunofluorescence, transwell migration and invasion assays.

Results: The results showed that the *miR-34* family might regulate the mTOR pathway, cell cycle (*CCND2*) and cell adhesion functions (*FZD4*). Further, we showed that a low negative correlation (r^2 = -0.07) between *miR-34b*/EMT score and four EMT signature genes (*BMP7*, *CAV1*, *ID2*, *FN1*) were significantly regulated by *miR-34b*. Also, *FN1* was indicated as its putative target with the highest negative EMT score and high binding energy between MRE/3'UTR. Further, these transcriptomic signatures in CESC revealed a significant inverse correlation across the stages of primary human CESC. These genes were repressed at transcriptional and translational levels in *miR-34b-3p* expressing C33A and HeLa cells, contributing to their reduced cell migration and invasive properties.

Conclusions: Our studies revealed the potential targets of the *miR-34* family, especially *miR-34b*, that can emerge as potential biomarkers and promising therapeutic targets in CESC disease management.

Keywords: miRNA, Cervical Epithelial Squamous cell Carcinoma (CESC), *miR-34*, EMT, EMT score

1. Introduction

Carcinoma of the uterine cervix, the second largest cancer in women, is on a steady rise, leading to cancer-related deaths worldwide[1,2]. Despite the awareness of cervical cancer prevention and improvement in screening systems, there is a higher incidence of cervical cancer in developing countries, especially in India, accounting for 16% of total global cases[3]. Indian women develop cervical cancer with a cumulative risk of 1.6% and face a cumulative death risk of 1.0%, contributing to about one-third of the global cervical cancer deaths[4]. Advanced and recurrent cervical cancer results in a poor prognosis with a one-year survival rate of 10-20%[5]. The existence of invasion, metastasis and drug resistance in cervical cancer results in treatment failure and induces most of its related deaths. A major contributor to the poor prognosis of cervical cancer is Epithelial-Mesenchymal Transition (EMT) initiation, triggered by the cellular processes, which include the dysregulation of certain tumour suppressors, oncogenes, transcription factors, growth factor signalling, and miRNAs. EMT, a pivotal and intricate process, plays an essential role in tumour cell invasion and migration by involving numerous cytokines, transcription factors and cellular pathways such as *NF-κB*, *PI3K*, Wnt, and *TGF*β, etc.[6] During the EMT process, the tumour cells undergo a "cadherin switch", lose their epithelial molecule (E-cadherin) and gain the mesenchymal molecule (N-cadherin), attributing to tumour invasion and metastasis[7]. So, the dysregulation of these EMT markers indicates the poor prognosis of cancer patients, especially epithelial cancer cell types like cervical cancer. A positive correlation exists between EMT phenotype and increased tumour progression, invasion, and metastasis in primary cervical cancers[8]. Despite many studies showing stimulation of *IL6*[9] and increased *REX1*[10] and *ATG5*[11] expression in the EMT induction, there is a lack of concrete molecular insights involved in the EMT regulatory factors to halt cervical cancer progression. In this context, microRNAs (miRNAs), small 18-22 nucleotides long non-coding RNAs, as potential modulators of tumour suppressors and oncogenes acquire attention as they act as RNA silencers or post-transcriptional regulators of messenger RNAs (mRNAs)[12,13]. Changes in miRNA and gene expression have been documented in tumour initiation and progression[14], such as decreased

expression of miRNAs, explaining its intrinsic role in tumour suppression[15]. Multiple miRNAs target the EMT regulatory factors inhibiting or promoting EMT in cervical cancers. These multiple miRNAs include *miR-200* family, *miR-141*, *miR-10a*, *miR-19a*, *miR-29a*, *miR-361-5p*, *miR-429* and *miR-155*[16-19]. Earlier, several studies have indicated the role of *miR-34b* in many cancers, including cervical cancer[20-22]. In addition, differential expression of *miR-34b* was reported between high-grade cervical intraepithelial neoplasia and normal cervical epithelial samples[23]. *miR-34b,* along with *miR-34a* and *miR-34c,* belong to the *miR-34* family. *miR-34a* is encoded by exon 2 of chromosome 1 (ch1p36.22), while *miR-34b/c* exists as the polycistronic transcript located within intron 1 and exon 2 on chromosome 11 (ch11q23.1)[24]. Interestingly, the genes of the *miR-34* family harbour a significant conserved sequence in the putative promoter regions of *p53* binding sites contributing to its pro-apoptotic/anti-proliferative functions[25-27]. The switching of the inverse regulation between the *p53* inactivation and *miR-34a/b/c* contributes to the mesenchymal induction and, thereby, the metastatic state of the cancer cells[28]. Further, a significant association exists between *miR-34b* and phospho-Met, *p53* (phospho S392), and *MDM2* in non-small cell lung cancer (NSCLC)[29]. In a recent report, *miR-34b* was shown to mediate *TGF-β1* regulation of cervical cancer cell proliferation and apoptosis[23]. In yet another study, Lu et al.[30] have indicated the significance of *miR-34b* as a diagnostic biomarker and a potential molecular target for ovarian cancer treatment by using exosomal *miR-34b*-mediated inhibition of cell proliferation and EMT. Also, Rivas et al.[20] have suggested the involvement of other regulatory genes in the cellular process of cervical cancer mediated by *miR-34*. While all these reports show that *miR-34b* plays a vital role in the EMT regulation of solid tumours, including cervical cancer, exploring the list of *miR-34b*/gene modulations and their underlying mechanisms would contribute to the identification of newer therapeutic targets for better disease management. To address this research gap, we applied a computational approach that includes three processes to identify targets of *miR-34b* contributing to EMT progression in cervical cancer from The Cancer Genome Atlas (TCGA), TCGA-CESC dataset, as shown in Figure 1a. Further, we validated the identified candidate *miR-34b* targets in primary CESC

derived from the South Asian Indian population and cervical cancer cell lines to show the correlation between the expression of *miR-34b* and target genes. The study also revealed the repression of candidate genes/proteins in the ectopically *miR-34b* expressing stable cervical cancer cells indicating the role of *miR-34b* in cervical cancer progression. As many gene aberrations contribute to cancer progression and metastasis, advanced treatment regimens targeting multiple gene modulations acquire importance in treating cervical cancer. As EMT offers insight into novel strategies for cancer treatment, the outcome of this study could aid in identifying the multigene aberrations contributing to the advancement of biotherapy and adjuvant therapy in cervical cancer.

2. Results and Discussion

2.1. *miR-34* family regulates the candidate genes in the mTOR pathway, cell cycle (*CCND2*) and cell adhesion functions (*FZD4*)

miRNA target prediction strategies alone do not indicate whether a particular pair of miRNA and mRNA interactions are biologically relevant. Therefore, a negative correlation between the team in independent TCGA-CESC cancer samples was used as a proxy for a regulatory effect (Figure 1b). This study used the mirDIP4.1 target prediction database, which includes 30 different databases for microRNA-mRNA target prediction to identify targets of the *miR-34*(*a/b/c*) family. A set of 23 databases had targets for all the three *miR-34* family miRNAs, and to obtain a confident set of target prediction genes, we iterated a number of databases in which a gene was predicted as a target. About 19,675 predicted genes were obtained with the cut-off of exactly five databases, of which 1,648 genes revealed a significant negative correlation after the Benjamini-Hochberg correction. However, about 4,029 predicted targets exhibited a significant negative correlation out of 27,591 genes with a cut-off of ten databases and these 4,029 genes were considered for further analysis.

Gene ontology analyses for the biological process for the above gene list identified 18 biological processes, including epithelial cell migration and adhesion (q-value ≤ 0.05, Supplementary File 1a).

Using Reactome, a total of 10 different pathways were obtained significantly, of which the four biologically significant pathways included mTOR-associated pathways, G1 cell cycle, epithelial cell migration (ECM) and cell adhesion (q-value ≤ 0.05, Figure 1c, Supplementary File 1b). *miR-34a* was predicted to interact with the mTOR pathway genes *PRKAB2*, *CAB39*, and *PPM1A*; *miR-34b* with genes such as *PRKAA2*, *CAB39L* and *miR-34c* with genes like *PRKAA2*, *CAB39L*, and *STRADA* (Supplementary Figure S1).

We observed that *miR34a* and *miR-34c* regulated the key cell cycle genes *CCND1*, *CCND2*, *CDKN1B*, *E2F3*, *E2F5*, and *ABL1* (Supplementary Figure S2). *CCND2*, a regulator of cell cycle G1/S transition, is also expressed in cancers[31], and miRNA-mediated repression of *CCND2* lowered the proliferation, migration and invasion properties of cervical cancer cells[32]. Our findings support the scope of studying the *miR-34a/c* and *CCND2* interactions in cervical cancer. The genes involved in ECM and *miR-34* family showed the interaction of *miR-34a* with many vital genes such as *TGFB1*, *TGFB2*, *GATA3*, *MAP3K3*, *MAP4K4* and *CDH13*. Also, *miR-34b* and *miR-34c* regulated the genes like *PTPRTG*, *ARSB*, and *PIK3R3* (Supplementary Figure S3). Further, the regulatory networks with the genes involved in cell adhesion function revealed a unique gene cluster for *miR-34a*. At the same time, *miR-34b* and *miR-34c* regulate *FZD4* and *LIMCH1* genes (Supplementary Figure S4). This unique cluster of *miR-34a* includes the key genes, namely *SMAD3*, *JAG1*, *ITGA6*, *CD44*, and *CDH13*. All these genes were widely studied for their role in the progression of various cancers and are also inversely correlated with the *miR-34* family in cervical cancer. Thus, these mRNA/*miR-34* family interactions can be further explored by the experimental validation contributing to gene therapy.

2.2. Epithelial genes show a positive correlation, while mesenchymal genes exhibit a negative correlation with *miR-34b* in cervical cancer

Further, the regulatory targets of *miR-34b* were specifically explored as its expression was observed to be lower than *miR-34a* and *miR-34c* in TCGA-CESC datasets (Figure 1d). Interestingly, similar

results were observed in two other epithelial cancer types: ovarian cancer (TCGA-OV) and uterine corpus endometrial carcinoma (TCGA-UCEC, Figure 1d). Next, a similar strategy was applied to the *miR-34* family using the mirDIP4.1 target prediction database. About 17,786 genes out of 27,591 in 26 databases and 6,064 in more than five databases were predicted for any given interaction. Out of 6,064 genes, we observed a significant correlation of *miR-34b* with 1,385 genes, of which 544 genes showed a significant negative correlation. In comparison, 841 genes showed a significant positive correlation after the Benjamini-Hochberg correction for multiple comparisons. Out of 1,385 genes, a total of 344 genes were known to be involved in regulating the EMT pathway. Figure 1e shows the schematic representation of the steps followed in this analysis. As decreased *miR-34b* levels in the three epithelial cancers (TCGA-CESC, TCGA-OV and TCGA-UCEC) were observed, the interaction between *miR-34b* versus epithelial and mesenchymal markers was the focus of the study.

EMT is a complex process involving morphological changes and a cascade of molecular events contributing to metastasis in cervical cancer, but its regulatory networks are poorly understood. To better understand the *miR-34b* dependent regulatory network modulating the EMT process, we analysed TCGA-CESC data sets and CESC primary tumour tissues (n = 21) obtained from the Asian Indian population along with non-cancerous cervical tissues (n = 5) as control. First, qPCR analyses were performed for *miR-34b*, *CDH1* (epithelial signature gene) and *VIM* (mesenchymal signature gene) expressions in these select CESC and non-cancerous cervical tissues. A decreased expression in the median of *miR-34b* (Figure 2a), *CDH1* gene (Figure 2b), and increased expression in the median of *VIM* gene (Figure 2c) in Stage III of the Asian Indian population-derived CESC was observed. These results corroborate with the ones obtained from the TCGA-CESC datasets as described in the section below (*miR-34b* has a positive correlation with epithelial gene and negatively correlates with mesenchymal gene, Figures 2d, 2e).

Upon comparing correlation values of EMT markers across the stages of TCGA-CESC, we observed about nine epithelial markers such as *ANAX1*, *BMP7*, *CAV1*, *CDH13*, *CLDN1*, *LYPD3*, *PKP3*, *PTPN14* and *SDC1*, had a positive correlation across the stages of CESC (Figure 2d, Supplementary File 2). *BMP7* was known to inactivate the EMT-associated genes, thereby contributing to the reduced *TGFβ1*-mediated cell growth and metastasis in breast cancer[33] and a significant association of decreased *BMP7* in primary breast cancer cells with clinically overt bone metastasis was observed earlier[34]. Similarly, *CAV1* positively correlated with E-cadherin (*CDH1*) gene expression, thereby regulating EMT in gastric cancers[6,35]. Therefore, based on the high positive correlation across stages, gene expression and significant expression in the survival analysis, *BMP7* and *CAV1* were selected out of the nine genes identified for further study (Supplementary File 2).

First, we analysed the expression of *BMP7* across the TCGA-CESC data sets and found no significant change across the stages (Figure 3a). In contrast, a significant difference was observed across G2 versus G3 grades of TCGA-CESC (Figure 3d). We decided to analyse *BMP7* expression in CESC primary tumour tissues (n = 21). The results showed a significant decrease in *BMP7* expression in Stage III CESC, while no significant change was observed in Stage II CESC compared to the control cervical tissues Figure 3g). Also, a significant decrease in the survival period was observed in patients having lower *BMP7* expression (p-value = 0.017, Figure 3j). Similarly, a previous study has shown that treating cervical cancer cells with human recombinant *BMP7* resulted in tumour growth arrest by triggering hTERT gene repression[36]. Therefore, there is significant evidence for *miR-34b*-mediated *BMP7* modulation, which has the potential to contribute to CESC growth arrest.

Next, we analysed the expression of the *CAV1* gene across the TCGA-CESC datasets; the results showed an insignificant change across stages and grades of TCGA-CESC datasets (Figures 3b, 3e). Further, qPCR analysis of *CAV1*gene expression in CESC revealed a significant decrease in Stage III than Stage II cervical cancer tissues (Figure 3h). A recent study on cervical cancer cell lines has

reported a decreased level of *CAV1* gene expression, and the *miR-96*-mediated restoration of the *CAV1* has reduced cell proliferation, migration, and invasion[37]. Also, other studies have indicated the tumour suppressor role of *CAV1* in various cancers, including cervical cancer[38-40]. However, in the present study, a significant decrease in survival period was observed in patients having higher *CAV1* gene expression in TCGA-CESC datasets (Figure 3k). So, as the current study indicates a positive correlation between *miR-34b* and *CAV1*, further experimental validation of *miR-34b* mediated *CAV1* modulation would aid in understanding their role in CESC tumour management.

Following epithelial markers, we observed a significant negative correlation of eight mesenchymal markers with *miR-34b*, namely *RHOA*, *GAB2*, SON, *PROM1*, *TCF7*, *ID2*, *FLT1*, and *LATS1* (Figure 2e, Supplementary File 2). All genes, except *PROM1*, had the MRE in their 3'UTR for *miR-34b*, indicating that they may be directly regulated by *miR-34b*.

Among these eight mesenchymal markers, the *ID2* gene, which had a relatively high negative correlation (r^2 = -0.27; p-value < 0.001) than other genes, was further analysed to understand its significance in CESC tumour progression. *ID2* has been reported as an EMT attenuator in triple-negative breast cancers[41] and multiple myeloma[42]. Also, *ID2* has been stated as a potential therapeutic target in salivary gland carcinoma where its repression has reduced Vimentin, N-cadherin, and Snail with induction of E-cadherin expression leading to a more differentiated phenotype[43]. In the present study, an insignificant change in *ID2* gene expression was observed between the stages (Figure 3c), while a significant difference was observed between the grades of TCGA-CESC datasets (Figure 3f). Further, the qPCR analysis in primary cervical cancer tissues of the South Asian Indian population showed an insignificant change in *ID2* gene expression across stages of CESC (Figure 3i). However, a minor difference in the patient's survival between the low and high *ID2* gene expressions was also observed (Figure 3l). Therefore, these results indicate sufficient

evidence for exploring further the *miR-34b*-mediated *ID2* modulation and their potential to contribute to tumour regression in CESC.

In addition to mesenchymal markers showing a negative correlation, a total of 25 mesenchymal markers had a significant positive correlation (r^2 = 0.2-0.25, p-value ≤ 0.01), including genes such as *HSPB1*, *IGF1R*, *SNAI2*, and *SMAD3* (Figure 2e) and this observation can be further studied with experimental validation on its association with *miR-34b*.

To further understand the relationship between putative EMT markers regulated by *miR-34b*, we analysed EMT scores, indicating the distribution of hybrid Epithelial/Mesenchymal (E/M) cells in the primary tumours and aiding in identifying the putative therapeutic targets and therapy resistance in cervical cancer. Previous studies have discussed the various methods to derive EMT score[44]. A differential correlation analysis of *miR-34b* expression with EMT scores across the stages and grades of TCGA-CESC datasets was performed in the present study (Supplementary File 3). As this analysis precisely involves the quantitative measurement of EMT signature genes, its outcome would aid us in identifying relevant genes regulated by *miR-34b* involved in EMT. The overall correlation revealed a significant low negative correlation (r^2 = -0.07) between *miR-34b* and EMT score in TCGA-CESC. Also, a significant low negative correlation was observed across stages (Figure 4a) and a varied correlation across grades (Figure 4b). These results indicate that if there is an increase in *miR-34b* expression, we may obtain a more negative EMT score, showing an increase in epithelial signature genes than the mesenchymal genes. So *miR-34b* can be further explored for its potential to regulate the EMT signature genes, thereby improving the management of CESC metastasis.

Based on the correlation analysis between *miR-34b* and these EMT signature genes, we found Fibronectin 1 (*FN1*), a mesenchymal marker, had a significant high negative correlation (r^2 = 0.0028). *FN1* expression in TCGA-CESC datasets showed an insignificant difference in expression across stages

(Figure 4c) while a significant change across the grades (Figure 4d). Also, the qPCR analysis in primary cervical cancer tissues of the South Asian Indian population showed a significant change in *FN1* gene expression across stages of CESC (Figure 4e). In addition, a significant decrease in the patient's survival with high *FN1* expression was also observed (Figure 4f). Also, *FN1* has MRE in its 3'UTR for *miR-34b* (Figure 4g), indicating that it is inversely regulated by *miR-34b*. Further, the deregulation of *FN1* protein by *miR-34b* was analysed by immunofluorescence stain. Here, a significant decrease of Corrected Total Cell Fluorescence (CTCF) in the presence of *miR-34b* expressing cervical cancer cells compared to the non-target sequence vector (NTSV) transfected control cells (Figure 4h, 4i) indicated the repression of *FN1* protein expression. These results prove that *FN1* could be a potential target of *miR-34b* in CESC.

To ascertain the select candidate gene expression in the presence of *miR-34b*, the cervical cancer cells (C33A and HeLa) were stably expressed with the *miR-34b-3p* expression construct. The stable cells expressed *miR-34b-3p* up to 4-5 folds in cervical cancer cells (C33A and HeLa) after 48h of induction (Figure 5a). At this time point, a significant increase in the *CDH1* (Figure 5b) and a decrease in *VIM* (Figure 5c) compared to the NTSV transfected control cells was observed. This change in *CDH1* and *VIM* gene expression was also supported by their protein expressions (Figure 5h). Thus, the increase of epithelial marker (*CDH1*) and a decrease of mesenchymal marker (*VIM*) in the presence of *miR-34b* indicate its association with EMT regulation of cervical cancer cells.

Further, the *miR-34b-3p* expressing cervical cancer cells revealed a significant increase of *BMP7* gene expression in C33A cells (1.5-fold) while a decreased expression in HeLa cells (0.9-fold) compared to the transfected control cells (Figure 5d). A similar result was also significantly observed for the Bmp7 protein expression (Figure 5h). The difference in the cell lines phenotype may contribute to the variation of *BMP7* expression. The observed decrease in *BMP7* gene expression in the HeLa cells corroborated an earlier report where the *BMP7* knockdown inhibited EMT by inducing E-cadherin

and decreasing Vimentin[45]. However, the increased *BMP7* gene levels observed in C33A trigger further evaluation of the molecular insights as *BMP7* was reported to inactivate the EMT-associated genes resulting in reduced *TGFβ1*-mediated breast cancer cell growth and metastasis[33]. Next, a significant increase of *CAV1* at the mRNA level (Figure 5e) and protein levels (Figure 5h) was observed. As *CAV1* is reported for tumour suppressor activity in other cancers, its modulation by *miR-34b* in the present study may contribute to the increased tumour suppressor activity in cervical cancer cells. Next, a significant decrease of *ID2* (Figure 5f) was observed in the cervical cancer cells in the presence of *miR-34b*. *ID2,* being reported as an EMT attenuator in other cancers, its suppression by *miR-34b* in cervical cancer cells may contribute to better tumour management. Also, the decrease of *FN1* mRNA levels in the cervical cancer cells expressing *miR-34b-3p* strongly indicates the regulation of *FN1* by *miR-34b-3p* (Figure 5g).

Thus, the inverse changes of select candidate gene expressions in the ectopically *miR-34b-3p* expressing cervical cancer cells (Figure 5) to that of primary cervical cancers (Figures 2, 3) indicate their association. In addition, the regulation of these genes or proteins resulting in a significant decrease of migration and invasion properties of *miR-34b-3p* expressing cervical cancer cells suggests that the select candidate genes are potentially modulated by *miR-34b* to contribute to tumour regression in CESC (Figures 5i -5n). However, further experimental validation needs to be carried out to understand the regulation of these candidate genes by *miR-34b* in a cascade.

3. Conclusions

A wide range of genetic aberrations regulated by the *miR-34* family involving EMT in cervical cancer has been indicated in this study. This study reveals the evidence for regulating candidate genes in the mTOR pathway, cell cycle (*CCND2*) and cell adhesion functions (*FZD4*) by the *miR-34* family. In addition, the correlation of *miR-34b* with the signature genes involved in EMT modulation has identified the new putative targets, namely *CDH1*, *BMP7*, *CAV1,* and *ID2*. Further, using EMT scores,

we show that *FN1* could be a direct target regulated by *miR-34b*. Following this identification, the experimental validation using gene expression studies in ectopically expressed *miR-34b-3p* cervical cancer cells proves their association with each other. In the present study, an integrated approach was used to reveal the novel targets for the *miR-34* family, especially *miR-34b* and experimentally showed the relationship between *miR-34b* and its target genes in CESC. Also, the derived observations depend entirely on the absolute gene expressions of primary TCGA-CESC datasets, as they are robust to be carried out for further experimental validation. In addition, these results were supported by the evidence obtained from a small subset of the Asian Indian population. Finally, the results obtained from this integrated analysis provide convincing evidence for the role of *miR-34b* in EMT modulation by regulating the candidate signature genes (*BMP7*, *CAV1, ID2,* and *FN1*) and pave the way for a more detailed study to understand their interactions and help in epithelial cancer cell metastasis management of cervical cancer.

4. Methods

4.1. Identify miRNAs whose expression is correlated with cervical cancer

Previously generated miRNA and mRNA expression levels by RNASeq within TCGA Research Network (http://cancergenome.nih.gov/) were used in the study. The processed count and raw read data were downloaded for cervical cancer from TCGA Genomics Data Commons (GDC) portal (https://gdc.cancer.gov/). TCGA-CESC expression data were derived from 307 cervical carcinomas (epithelial squamous cell and endocervical adenocarcinoma) and two non-cancerous cervical epithelial cells. The samples were identified using the TCGA barcode associated with each sample. The clinical metadata files such as tumour stage, patient's age, gender, and survival statistics of each TCGA-CESC sample were downloaded through the GDC Application Programming Interface (API). Supplementary Table 1 summarises the clinical metadata of all the CESC samples (n = 307) from the TCGA project. Gene expression (Fragments per Kilo-base of exon per Million mapped, FPKM) and

miRNA expression (Counts per Million, CPM) quantification data files available from cervical epithelial squamous cell carcinoma (TCGA-CESC) were utilised from TCGA.

4.2. Identify potential gene targets of the *miR-34* family and *miR-34b* through TCGA cervical cancer gene expression dataset

RNA-sequence expression data generated by the TCGA project available from the Genomics Data Common repository (https://portal.gdc.cancer.gov) was used. A TCGA file list, 'TCGA-CESC sample sheet', containing information such as sample ID and file ID for each CESC sample, was also downloaded by querying the GDC API. This file list sample sheet was used to download the files in Ras input for the pipeline in R. The library sizes for gene expressions and miRNA expressions were 60,483 and 1,881, respectively. Expression criterion of genes with FPKM ≥ 1.0 and miRNA with CPM ≥ 1.0 across all samples were applied as described earlier[46]. The cut-off values represent a mean frequency ≥1 in a million reads mapped for microRNA and a mean frequency ≥1 in a million reads after transcript length normalisation for mRNA. There were 14,509 genes and 407 miRNAs for which the expression was above the cut-off in TCGA-CESC datasets.

The pipeline was built using R version 3.6.1, which includes data downloading, correlation analysis, target prediction database lookup, and functional enrichment. The instructions for running the pipeline are presented as R script and can be accessed from the GitHub repository (https://github.com/HimanshuLab/*miR-34b*-in-pan-Gynaecological-Cancers). The entire mirDIP4.1 database was downloaded and used to compare the target prediction across 30 individual resources[47]. The mirDIP4.1 database has been used to compare 26 microRNA-mRNA target prediction tools databases. These databases cover a wide variety of target prediction strategies, including seed region matching, evolutionary conservation of regulatory region on mRNA, binding energy, site accessibility, CLIP experimental evidence, and machine learning methods. DIANA-TarBase (v8.0; http://www.microrna.gr/tarbase) was used to lookup experimentally validated

microRNA-mRNA interactions[48]. Additionally, TargetScan (v7.0; www.targetscan.org), MiRanda-miRSVR (v3.3a; http://www.microrna.org) and DIANA-microT-CDS (v5.0; http://diana.imis.athenainnovation.gr/DianaTools/index.php) databases were utilised[49,50]. Only predictions categorised as either good or better according to the miDIP database were used. As microRNA target prediction databases are known to have many false positives, we have considered microRNA-mRNA interactions predicted to occur in at least five databases as a weak indicator and being predicted in over ten bases as a strong indicator.

4.3. miRNA:mRNA correlation analysis with epithelial and mesenchymal genes

As one of the key functions of miRNAs is to potentially destabilise the mRNAs, we focussed on the significant miRNA-mRNA negative correlation as an indication of a regulatory effect in subsequent analysis. Spearman rank correlations were calculated for *miR-34(a/b/c)* family versus gene expression using the rcorr function from the Hmisc R package. Hmisc::rcorrfunction also provides asymptotic p-values for the correlations, corrected for multiple comparisons using the Benjamini-Hochberg method. *miR-34(a/b/c)* family and gene pairs were only considered if they were anti-correlated (Spearman correlation; $r^2 < 0$) and if the corrected p-value was significant (p-value < 0.01).

The EMT genes from the first literature-based database[51] for EMT human genes (dbEMT) was used in this study. The correlations of *miR-34b* with EMT genes were generated with adjusted p-values using the Benjamini-Hochberg method. The correlations with a p-value ≤ 0.01 were selected and analysed across stages/grades of CESC. The analysed results are presented in Supplementary File 1 and derived a co-expression network using Cytoscape v3.7.1. The International Federation of Gynaecology and Obstetrics (FIGO) staging system based on clinical examination and specific procedures (imaging) was used in the study[52]. The tumour stage of each sample was classified into Stages I, II, III, and IV by binning the various subtypes in each stage. Independent of the staging system, the histopathologic parameter based on the tumour cell differentiation was also used to

study the candidate gene expression levels. The tumour grade of each sample was classified as GX (grade cannot be assessed), G1 (well-differentiated), G2 (moderately differentiated) and G3 (poorly or undifferentiated)[52]. The candidate gene expressions were checked across Stages I/II/III and grade G1/G2/G3 as the increase in stage and grade directly proportionate to the metastasis and poor prognosis of CESC.

4.4. Calculation of EMT scores from TCGA cervical cancer gene expression dataset

EMT score is a quantitative measure of the sample characteristic on the Epithelial-Mesenchymal spectrum. The EMT score for each sample was derived by subtracting the mean gene expression of epithelial markers from the mean gene expression of mesenchymal markers. TCGA-CESC datasets were used to derive the gene expression RNA sequencing z-scores for the epithelial markers (n = 3) and mesenchymal markers (n = 13)[53]. The epithelial markers were E-Cadherin (*CDH1*), *DSP*, and *TJP1*; while the mesenchymal markers were Fibronectin (*FN1*), Vimentin (*VIM*), N-Cadherin (*CDH2*), Integrin beta 6 (*ITBG6)*, Forkhead box protein C2 (*FOXC2)*, Matrix metalloproteinase 2 (*MMP2)*, Snail family transcriptional repressor 1 (*SNAI1)*, Matrix metalloproteinase 3 (*MMP3)*, Snail family transcriptional repressor *2 (SNAI2)*, Matrix metalloproteinase 9 (*MMP9)*, Twist-related protein 1(*TWIST1)*, SRY-Box transcription factor10(*SOX10)*, and Goosecoid homeobox (*GSC)*. The z-scores for three mesenchymal genes (*ITBG6*, *SOX10*, and *GSC*) were not obtained as no data was available in TCGA-CESC datasets. A positive EMT score was thus associated with the mesenchymal cell type, while a negative score was associated with the epithelial cell type. A subgroup analysis using grades and stages of CESC for correlation between *miR-34b* expressions and EMT score was also derived. A negative correlation indicates that as the gene's expression increased, the sample was more likely to take on epithelial characteristics (a more negative EMT score). The derived EMT scores for each sample of TCGA–CESC are given in Supplementary File 3.

4.5. Gene ontology and pathway analysis

Functional enrichment was done using Reactome Pathways (https://reactome.org/) using the R/Reactome PA package. Ontology enrichment was done using gene ontology (GO) using the R/cluster Profiler package. The Reactome and gene ontology results are filtered using q-values calculated using adjusted p-value and false discovery rate (FDR). A cut-off q-value of 0.05 was used. The obtained results are presented in Supplementary File 1. The resulting network was plotted using Cytoscape v3.7.1 network analysis of gene interactions between the genes of interest was done using the STRING-db v11.0 webpage (https://string-db.org/cgi/).

4.6. Survival analysis

Survival Analysis was done using Kaplan-Meier (KM) plots using the R packages survminer, survival, and RTCGA.clinical. The survival probability was compared between the samples in the upper and lower quartiles of expression for a given gene, and the adjusted p-values were derived using the Benjamini-Hochberg method.

4.7. Primary cervical cancer tumour tissues of the South Asian population

Twenty-one primary cervical cancer tumour tissues and five normal cervical tissues collected from the South Asian population reported at the regional institutes were included in the study. These primary tumour tissues (n = 21) include fresh human cervical cancer tissues (n = 10) and formalin-fixed paraffin-embedded tissues (n = 11) with a median age of 55. Ten of these tumour samples were Stage II, while eleven were Stage III (Supplementary Table 2, Supplementary Figure S5). Respective non-cancerous cervical biopsy tissues were used as normalised controls (n = 3).

4.8. *miR-34b* over-expression in cervical cancer cell lines

The shMIMIC inducible Lentiviral microRNA, tetracycline-inducible expression system (Tet-On 3G) with TurboGFP as reporter and mCMV as promoter for both hsa-*miR-34b-3p* (MIMAT0004676;

5'CAAUCACUAACUCCACUGCCAU3'; Cat# GSH11929-224638820; Dharmacon™) and for non-targeting sequence control (Cat# VSC11651; Dharmacon™) was purchased. The packaging lentiviral particle, transduction and puromycin selection were carried out per the manufacturer's protocol. The stable cell lines were established for HeLa (RRID: CVCL_0030) and C33A (RRID: CVCL_1094) cervical cancer cell lines (*Mycoplasma sps* free). Both the cell lines were maintained in a growth medium containing Dulbecco's Modified Eagle's Medium (DMEM, Life Technologies, Carlsbad, CA, USA); 10% FBS (Life Technologies) and antibiotics (100 U ml^{-1} penicillin and 100 µg ml^{-1} streptomycin) at 37 °C with a humidified atmosphere of 5% CO_2. Both cell lines have been authenticated using STR profiling within the last three years.

4.9. Immunofluorescence

About 1 x 10^5 stable cells of C33A/HeLa lines were seeded in a 12-well cell culture plate and incubated overnight at 37 °C with a humidified atmosphere of 5% CO_2. Following incubation, doxycycline (C33A:750ng/ml / HeLa: 800ng/ml was added and incubated for 48h at 37 °C with 5% CO_2. Next, the cells were fixed with chilled methanol (10 min), washed with ice-cold 1XPBS (3 times) and incubated with 1% BSA in PBST (PBS+ 0.1% Tween 20). The cells were incubated with the diluted FN1 antibody (1:50, Cat# SC-9068, Santacruz) in a humidified chamber overnight at 4 °C. Then, the cells were probed with diluted Alexa flour conjugated secondary antibody (Alexa Flour™ 568, Cat# A-11011, Invitrogen) for 1h in the dark. The cells were intermittently washed three times with chilled 1XPBST. Finally, the cells were stained with DAPI (DNA stain) for 1 min, and the cells were imaged using an Olympus IX-73 microscope and analysed using ImageJ software. The corrected total cell fluorescence (CTCF) was calculated using the equation, $CTCF = Integrated\ Density - (Area\ of\ selected\ cell \times Mean\ fluorescence\ of\ background\ readings)$, described by Jakic et al.[54]

4.10. RNA isolation, cDNA synthesis, and real-time quantitative PCR (qRT PCR)

The total RNA is extracted from the stable cell lines (C33A/HeLa) using a trizol reagent (RNA Iso Plus reagent, Takara) and from paraffin-embedded cervical cancer tumour tissue sections using the RNeasy FFPE kit (catalogue number 73504, Qiagen) according to the manufacturer's protocol. The total RNA was quantified using a NanoDrop spectrophotometer (ND2000, Thermo Scientific, MA, USA).and converted to cDNA using a Primescript RT reagent kit (Cat# RR037A, DSS Takara Bio India Private Ltd) as per the manufacturer's instructions. The stem-loop primers for miRNA cDNA conversion and oligodT primers for mRNA cDNA conversion were used. The converted cDNA was amplified using gene-specific primers and TB Green®Premix Ex Taq™ II (Cat# RR820A; DSS Takara Bio India Private Ltd) to quantify the expression of the genes using a real-time PCR system (Applied Biosystem Quant Studio 7 Flex) with the following protocol (Denaturation: 95°C/5 min; Amplification for 40 cycles: 95°C/30 sec, 60°C/45 sec, 72 °C/30 sec). The primers used in this study are given in Supplementary File 4. The relative C_t (cycle threshold) quantification of genes in tumour and normal cell lines was analysed for expression analysis. All samples were assayed in triplicates. The unpaired t-test was performed using GraphPad Prism 6.

4.11. Western Analysis

Protein isolation was carried out by resuspending the PBS-washed cells in radioimmunoprecipitation assay (RIPA) buffer and vortexing intermittently. The protein lysate was subjected to high-speed centrifugation, resolved by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto a polyvinylidene difluoride membrane (Cat# 1620177, BioRad). Following the protein transfer, blots were blocked using bovine serum albumin (5%) and probed with the primary antibody at 4 °C overnight and on the next day with an HRP-linked secondary antibody for 1h at room temperature. The primary antibodies against *BMP7* (Cat# A0697, ABclonal), *CAV1* (Cat# A1555, ABclonal), E-cadherin (Cat# SC-8426, G-10, Santacruz), Vimentin (Cat# SC-33322, RV-202, Santacruz), Vinculin (Cat# 4650S, Cell signalling technologies), and *β*-actin (Cat# A2228, Ac-74, Sigma Aldrich) proteins were used in the study. Intermittent washing of the blots was carried out using Tris-

buffered saline containing 1% Tween-20 (1x TTBS). Following the incubation of antibodies, the immunoblot was subjected to imaging using an Enhanced Chemiluminescence Kit (Cat# 1705061, Bio-Rad) and ChemiDoc (BioRad).

4.12. Migration and invasion assay

About 2×10^6 pre-induced cells were plated in each well of Culture-Insert 3 Well in µ-Dish 35 mm (Cat# 80366, ibidi). After overnight incubation, a fresh medium containing 5 µg/ml mitomycin-c and doxycycline was added, and the three-well silicone insert was removed to form two defined cell-free gaps. The migration of cells was imaged using an Olympus IX-73 microscope at 0h, 24h, and 72h of incubation. The percentage of area covered by the cells was analysed using ImageJ software.

Trans-well cell culture inserts coated with GelTrex (24-well inserts: Cat# 140629, GelTrex: Cat# A15696-01, Thermo Fischer Scientific) were used for the study. About 50×10^3 pre-induced cells were seeded on the upper chamber of the cell culture insert with growth media containing 1% FBS, while the lower one had a complete medium (10% FBS). Cells were incubated to migrate for 48 h. After incubation, the cells were fixed with 100% chilled methanol, stained with 0.1% crystal violet solution, and imaged using an Olympus IX-73 microscope. After imaging, the insert membranes carrying cells were destained in 10% acetic acid solution, and absorbance was measured at 595 nm to quantify migrated cells[55].

Ethical approval

The study followed the protocol as per the Declaration of Helsinki. The Institutional Ethics Committee of Sri Ramachandra Medical College and Research Institute (SRMC&RI), Chennai and the Institute of Obstetrics and Gynaecology, Chennai (No:36433/E1/2008-1) approved this study.

Informed consent

All individual participants of the study have provided informed consent.

Competing interests

The authors declare no conflict of interest.

Authors' contributions

The authors confirm their contribution to the paper as follows: VN, AX, HS, DK: Study conception and design; VN, AX, KJS, SS: Data collection; SS, VN: Interpretation of primary tumour samples and sample selection; VN, AX, HS, DK: Analysis and interpretation of results; VN, AX, HS, DK: Draft manuscript preparation: All authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

The authors would like to thank the Indian Institute of Technology Madras and the Indian Council of Medical Research for the postdoctoral fellowship of Nalini Venkatesan (Project no. 2019-4566/CMBBMS); Dr A.K. Munirajan, Professor and Head, Department of Genetics, University of Madras, Chennai, for the cDNA of CESC from the South Asian Indian population.

Availability of data and materials

Supplementary data is available for the manuscript. The TCGA data is publicly available. The instructions for running the pipeline are presented as R script and can be accessed from the GitHub repository (https://github.com/HimanshuLab/miR-34b-in-pan-Gynaecological-Cancers).

This manuscript is available in the bioRxiv preprint server https://doi.org/10.1101/2021.09.02.458804.

References

1 Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87-108, doi:10.3322/caac.21262 (2015).

2 Beatriz Serrano, M. B., Francesc Xavier Bosch, Laia Bruni Epidemiology and burden of HPV-related disease. *Best Practice & research. Clinical Obstetrics & Gynaecology* **47.**, 14-26. (2018).

3 Srivastava, A. N., Misra, J. S., Srivastava, S., Das, B. C. & Gupta, S. Cervical cancer screening in rural India: Status & current concepts. *Indian J Med Res* **148**, 687-696, doi:10.4103/ijmr.IJMR_5_17IndianJMedRes_2018_148_6_687_252161 [pii] (2018).

4 Monica & Mishra, R. An epidemiological study of cervical and breast screening in India: district-level analysis. *BMC Women's Health* **20**, 225, doi:10.1186/s12905-020-01083-610.1186/s12905-020-01083-6 [pii] (2020).

5 Diaz-Padilla, I., Monk, B. J., Mackay, H. J. & Oaknin, A. Treatment of metastatic cervical cancer: future directions involving targeted agents. *Crit Rev Oncol Hematol* **85**, 303-314, doi:10.1016/j.critrevonc.2012.07.006S1040-8428(12)00155-2 [pii] (2013).

6 Zhang, K. *et al.* Decreased Expression of Caveolin-1 and E-Cadherin Correlates with the Clinicopathologic Features of Gastric Cancer and the EMT Process. *Recent Pat Anticancer Drug Discov* **11**, 236-244, DOI:PRA-EPUB-73309 [pii]10.2174/1574892811666160128151437 (2016).

7 Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-196, doi:10.1038/nrm3758nrm3758 [pii] (2014).

8 Lee, M. Y., Chou, C. Y., Tang, M. J. & Shen, M. R. Epithelial-mesenchymal transition in cervical cancer: correlation with tumor progression, epidermal growth factor receptor overexpression, and snail up-regulation. *Clin Cancer Res* **14**, 4743-4750, doi:10.1158/1078-0432.CCR-08-0234 14/15/4743 [pii] (2008).

9 Jin-Wei Miao , L.-J. L., Jie Huang. Interleukin-6-induced epithelial-mesenchymal transition through signal transducer and activator of transcription 3 in human cervical carcinoma. *Int J Oncol* **45(1):165-76**, 165-176 (2014).

10 Xiao-Fang Liu , W.-T. Y., Peng-Sheng Zheng. REX1 promotes EMT-induced cell metastasis by activating the JAK2/STAT3-signaling pathway by targeting SOCS1 in cervical cancer. *Oncogene* **38.**, 6940-6957. (2019).

11 Suna Zhou, X. W., Jiapei Ding, Haihua Yang, and Youyou Xie Increased ATG5 Expression Predicts Poor Prognosis and Promotes EMT in Cervical Carcinoma. *Front Cell Dev Biol.* **Nov 25;9:757184**, 757184 (2021).

12 Setlai, B. P. H., R.; Reis, R.M.; Agbor, C.; Ambele, M.A.; Mulaudzi, T.V.; Dlamini, Z. MicroRNA Interrelated Epithelial Mesenchymal Transition (EMT) in Glioblastoma. *Genes* **13**, 244 (2022).

13 Guangtao Pan, Y. L., Luorui Shang, Fangyuan Zhou, Shenglan Yang. EMT-associated microRNAs and their roles in cancer stemness and drug resistance. *Cancer Commun (Lond).* **Mar;41**, 199-217. (2021).

14 Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* **103**, 2257-2261, doi:0510565103 [pii] 10.1073/pnas.0510565103 (2006).

15 Calin, G. A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* **101**, 2999-3004, doi:10.1073/pnas.0307323101 0307323101 [pii] (2004).

16 Burk, U. *et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* **9**, 582-589, doi:10.1038/embor.2008.74 embor200874 [pii] (2008).

17 Gregory, P. A. *et al.* The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**, 593-601, doi:10.1038/ncb1722ncb1722 [pii] (2008).
18 Hu, Y. & Tang, H. MicroRNAs regulate the epithelial to mesenchymal transition (EMT) in cancer progression. *Microrna* **3**, 108-117, doi:MIRNA-EPUB-62715 [pii] (2014).
19 Korpal, M., Lee, E. S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* **283**, 14910-14914, doi:10.1074/jbc.C800074200 C800074200 [pii] (2008).
20 Cordova-Rivas, S. *et al.* 5p and 3p Strands of miR-34 Family Members Have Differential Effects in Cell Proliferation, Migration, and Invasion in Cervical Cancer Cells. *Int J Mol Sci* **20**, doi:E545 [pii] 10.3390/ijms20030545 ijms20030545 [pii] (2019).
21 Kim, J. S. *et al.* MiR-34a and miR-34b/c have distinct effects on the suppression of lung adenocarcinomas. *Exp Mol Med* **51**, 1-10, doi:10.1038/s12276-018-0203-1 10.1038/s12276-018-0203-1 [pii] (2019).
22 Ye, K., Xu, C. & Hui, T. MiR-34b inhibits the proliferation and promotes apoptosis in colon cancer cells by targeting Wnt/beta-catenin signaling pathway. *Biosci Rep* **39**, DOI:BSR20191799 [pii] 10.1042/BSR20191799 (2019).
23 Cao, Z., Zhang, G., Xie, C. & Zhou, Y. MiR-34b regulates cervical cancer cell proliferation and apoptosis. *Artif Cells Nanomed Biotechnol* **47**, 2042-2047, doi:10.1080/21691401.2019.1614013 (2019).
24 Maroof, H., Salajegheh, A., Smith, R. A. & Lam, A. K. Role of microRNA-34 family in cancer with particular reference to cancer angiogenesis. *Exp Mol Pathol* **97**, 298-304, doi:10.1016/j.yexmp.2014.08.002 S0014-4800(14)00128-2 [pii] (2014).
25 Corney, D. C., Flesken-Nikitin, A., Godwin, A. K., Wang, W. & Nikitin, A. Y. MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res* **67**, 8433-8438, doi:0008-5472.CAN-07-1585 [pii] 10.1158/0008-5472.CAN-07-1585 (2007).
26 He, L. *et al.* A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130-1134, DOI:nature05939 [pii] 10.1038/nature05939 (2007).
27 Kumamoto, K. *et al.* Nutlin-3a activates p53 to both down-regulate inhibitor of growth 2 and up-regulate mir-34a, mir-34b, and mir-34c expression, and induce senescence. *Cancer Res* **68**, 3193-3203, doi:10.1158/0008-5472.CAN-07-2780 68/9/3193 [pii] (2008).
28 Helge Siemens, R. J., Sabine Hünten, Markus Kaller, Antje Menssen, Ursula Götz, Heiko Hermeking. miR-34 and SNAIL form a double-negative feedback loop to regulate epithelial-mesenchymal transitions. *Cell Cycle* **Dec 15**, 4256-4271. (2011).
29 Wang, L. G. *et al.* MicroRNA-34b functions as a tumor suppressor and acts as a nodal point in the feedback loop with Met. *Int J Oncol* **42**, 957-962, doi:10.3892/ijo.2013.1767 (2013).
30 Lu S, L. W., Shi H, Zhou H. . Exosomal miR-34b inhibits proliferation and the epithelial-mesenchymal transition by targeting Notch2 in ovarian cancer. *Oncol Lett.* **Sep;20**, 2721-2728, doi:10.3892/ol.2020.11837. (2020).
31 Uzuka, Y. & Saito, Y. [Combination therapy of chronic myelogenous leukemia with busulfan and bestatin]. *Gan To Kagaku Ryoho* **18**, 571-578 (1991).
32 Meyerson, M. & Harlow, E. Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol Cell Biol* **14**, 2077-2086, doi:10.1128/mcb.14.3.2077 (1994).
33 Ying, X., Sun, Y. & He, P. Bone Morphogenetic Protein-7 Inhibits EMT-Associated Genes in Breast Cancer. *Cell Physiol Biochem* **37**, 1271-1278, doi:10.1159/000430249 000430249 [pii] (2015).
34 Buijs, J. T. *et al.* Bone morphogenetic protein 7 in the development and treatment of bone metastases from breast cancer. *Cancer Res* **67**, 8742-8751, doi:67/18/8742 [pii] 10.1158/0008-5472.CAN-06-2490 (2007).

35 Sun, G. Y., Wu, J. X., Wu, J. S., Pan, Y. T. & Jin, R. Caveolin-1, E-cadherin and beta-catenin in Gastric Carcinoma, Precancerous Tissues and Chronic Non-atrophic Gastritis. *Chin J Cancer Res* **24**, 23-28, doi:10.1007/s11670-012-0023-0 cjcr-24-01-023 [pii] (2012).

36 Lucy Cassar, H. L., Alexander Ruvantha Pinto, Craig Nicholls, Sharyn Bayne and Jun-Ping Liu. Bone Morphogenetic Protein-7 Inhibits Telomerase Activity, Telomere Maintenance, and Cervical Tumor Growth. *Cancer Research* **68**, 9157–9166, doi:10.1158/0008-5472.CAN-08-1323 (2008).

37 Chen, Y. *et al.* miR96 exerts an oncogenic role in the progression of cervical cancer by targeting CAV1. *Mol Med Rep* **22**, 543-550, doi:10.3892/mmr.2020.11101 (2020).

38 Razani, B. *et al.* Caveolin-1 expression is down-regulated in cells transformed by the human papilloma virus in a p53-dependent manner. Replacement of caveolin-1 expression suppresses HPV-mediated cell transformation. *Biochemistry* **39**, 13916-13924, DOI:bi001489b [pii] 10.1021/bi001489b (2000).

39 Wiechen, K. *et al.* Caveolin-1 is down-regulated in human ovarian carcinoma and acts as a candidate tumor suppressor gene. *Am J Pathol* **159**, 1635-1643, DOI:S0002-9440(10)63010-6 [pii] 10.1016/S0002-9440(10)63010-6 (2001).

40 Zhang, T., Hu, Y., Wang, T. & Cai, P. Dihydroartemisinin inhibits the viability of cervical cancer cells by upregulating caveolin 1 and mitochondrial carrier homolog 2: Involvement of p53 activation and NAD(P)H:quinone oxidoreductase 1 downregulation. *Int J Mol Med* **40**, 21-30, doi:10.3892/ijmm.2017.2980 (2017).

41 Wen, X. F. *et al.* Inhibitor of DNA Binding 2 Inhibits Epithelial-Mesenchymal Transition via Up-Regulation of Notch3 in Breast Cancer. *Transl Oncol* **11**, 1259-1270, doi:S1936-5233(18)30166-9 [pii] 10.1016/j.tranon.2018.07.015 (2018).

42 Tommaso Perini, R. S., Mehmet Kemal Samur, Mariateresa Fulciniti, Michael A Lopez, Matthew Lawlor, Christopher J. Ott, Na Li, Yan Xu, Kenneth Wen, Nicola Amodio, Eugenio Morelli, Kenneth Anderson, Fabio Ciceri, Nikhil Munshi. Inhibitor of DNA Binding 2 (ID2) Plays a Key Tumor Suppressor Role in Promoting Oncogenic Transformation in Multiple Myeloma. *Blood,* **Volume 132, Supplement 1,** 60 (2018).

43 Sumida, T. *et al.* Targeting ID2 expression triggers a more differentiated phenotype and reduces aggressiveness in human salivary gland cancer cells. *Genes Cells* **21**, 915-920, doi:10.1111/gtc.12389 (2016).

44 Chakraborty, P., George, J. T., Tripathi, S., Levine, H. & Jolly, M. K. Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. *Front Bioeng Biotechnol* **8**, 220, doi:10.3389/fbioe.2020.00220 (2020).

45 Sun, R. *et al.* Expression of BMP7 in cervical cancer and inhibition of epithelial mesenchymal transition by BMP7 knockdown in HeLa cells. *Int J Mol Med* **45**, 1417-1424, doi:10.3892/ijmm.2020.4519 (2020).

46 Wagner, G. P., Kin, K. & Lynch, V. J. A model based criterion for gene expression calls using RNA-seq data. *Theory Biosci* **132**, 159-164, doi:10.1007/s12064-013-0178-3 (2013).

47 Tomas Tokar, C. P., Andrea E M Rossos, Mark Abovsky, Anne-Christin Hauschild, Mike Tsay, Richard Lu, Igor Jurisica. mirDIP 4.1—integrative database of human microRNA target predictions. *Nucleic Acids Research* **46**, D360-D370, doi:https://doi.org/10.1093/nar/gkx1144 (2018).

48 Vlachos, I. S. *et al.* DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res* **43**, D153-159, doi:10.1093/nar/gku1215 gku1215 [pii] (2015).

49 Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, doi:10.7554/eLife.05005 (2015).

50 Paraskevopoulou, M. D. *et al.* DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res* **41**, W169-173, doi:10.1093/nar/gkt393gkt393 [pii] (2013).

51 Zhao, M., Kong, L., Liu, Y. & Qu, H. dbEMT: an epithelial-mesenchymal transition associated gene resource. *Sci Rep* **5**, 11459, doi:10.1038/srep11459 srep11459 [pii] (2015).

52 Bhatla, N., Aoki, D., Sharma, D. N. & Sankaranarayanan, R. Cancer of the cervix uteri. *Int J Gynaecol Obstet* **143 Suppl 2**, 22-36, doi:10.1002/ijgo.12611 (2018).

53 Chae, Y. K. *et al.* Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). *Sci Rep* **8**, 2918, doi:10.1038/s41598-018-21061-1 10.1038/s41598-018-21061-1 [pii] (2018).

54 Jakic B, B. M., Cappellano G, Wick G. Elevated sodium leads to the increased expression of HSP60 and induces apoptosis in HUVECs. *PLoS One* **Jun 12;**, doi:10.1371/journal.pone.0179383 (2017).

55 K. Saito, T. O., N. Ata, H. Miyashiro, M. Hattori, I. Saiki. A modified and convenient method for assessing tumor cell invasion and migration and its application to screening for inhibitors. *Biol. Pharm. Bull.* **20** 345–348 (1997).

Figure legends

Figure 1

(a) Schematic representation describing the steps followed in the current study. The three processes followed in the pipeline are briefly presented. (b) Distribution of negative correlation of *miR-34* family (*a/b/c*) with transcripts of TCGA-CESC from TCGA datasets. The black colour in the graph indicates the cut-off value of the negative correlation considered in the study. (c) The pathway enrichment analysis using Reactome is presented. A total of 10 different significantly enriched pathways were observed (q-value ≤ 0.05). (d) Expression of *miR-34* family (*a/b/c*) in cervical cancer (CESC), ovarian cancer (OV) and uterine corpus endometrial carcinoma (UCEC) using TCGA datasets. (e) Illustration of the steps followed in the correlation analysis of miR-34b with epithelial/mesenchymal genes in cervical cancer.

Figure 2

(a) Cluster graph of miRNA expression across control and Stages (II/III) in the South Asian Indian population for the *miR-34b*. (b) Cluster graph of gene expression across control and Stages (II/III) in the South Asian Indian population for the *CDH1* gene. (c) Cluster graph of gene expression across control and Stages (II/III) in the South Asian Indian population for the *VIM* gene. (d) Correlation of *miR-34b* and epithelial genes across Stages (I/II/III) of TCGA-CESC (e) Correlation of *miR-34b* and mesenchymal genes across Stages (I/II/III) of TCGA-CESC. The symbols "NS" denotes non-significance; * denotes p-value ≤ 0.05; ** denotes p-value ≤ 0.01 and *** denotes p-value ≤ 0.001.

Figure 3

(a-f) Box plot graph across Stages (I/II/III) and grades (G1/G2/G3) of genes' expression in TCGA-CESC; (a) *BMP7*-stages; (b) *CAV1*-stages; (c) *ID2*-stages; (d) *BMP7*-grades; (e) *CAV1*-grades; (f) *ID2*-grades; (g-i) Cluster graph of genes' expression across control and Stages (II/III) in the South Asian Indian population: (g) *BMP7*, (h) *CAV1*, (i) *ID2*; (j-l) Survival analysis (KM plot) of TCGA-CESC based on their

median genes expressions of (j) *BMP7*, (k) *CAV1*, (l) *ID2*. The symbol "NS" denotes non-significance; * indicates p-value ≤ 0.05, and ** denotes p-value ≤ 0.01.

Figure 4

(a) Correlation analysis between *miR-34b* and EMT score in TCGA-CESC across Stages (I/II/III), (b) Correlation analysis between *miR-34b* and EMT score in TCGA-CESC across grades (G1/G2/G3), (c) Box plot graph across Stages (I/II/III) of *FN1* expression in TCGA-CESC, (d) Box plot graph across grades (G1/G2/G3) of *FN1* expression in TCGA-CESC, (e) Cluster graph of gene expression across control and Stages (II/III) in the South Asian Indian population, (f) Survival analysis (KM plot) of TCGA-CESC based on *FN1* median expression, (g) Sequence alignment analysis between *miR-34b* and 3'UTR of *FN1*, (h) Microscopic observation of immunofluorescence in transfected control (NTSV) and *miR-34b-3p* expressing cervical cancer cells (C33A and HeLa), (i) Bar graph of invasion assay in transfected control (NTSV) and *miR-34b-3p* expressing cervical cancer cells (C33A and HeLa). The black shaded bar shows the *miR-34b-3p*/gene expressions in cervical cancer cells carrying non-target sequence vector (NTSV), and the grey shaded bar shows the *miR-34b-3p*/gene expressions in cervical cancer cells carrying *miR-34b-3p* expressing vector. The symbol "NS" denotes non-significance; * indicates p-value ≤0.05.

Figure 5

(a-g) Bar graph of *miR-34b-3p*/gene expressions in *miR-34b-3p* expressing cervical cancer cells (C33A and HeLa). (a) *miR-34b-3p*, (b) *CDH1*, (c) *VIM*, (d) *BMP7*, (e) *CAV1*, (f) *ID2,* (g) *FN1,* (h) Western analysis of *BMP7*, *CAV1*, E-cadherin, *VIM*, Vinculin and β-actin proteins in transfected control (NTSV) and *miR-34b-3p* expressing cervical cancer cells (C33A and HeLa), (i and j) Microscopic observation of migration assay in transfected control (NTSV) and *miR-34b-3p* expressing cervical cancer cells at 0h, 48h and 72h (i: C33A and j: HeLa), (k) and (l) Bar graph of migration assay in transfected control (NTSV) and *miR-34b-3p* expressing cervical cancer cells at 48h (k:C33A and l: HeLa), (m) Microscopic

observation of invasion assay in transfected control (NTSV) and *miR-34b-3p* expressing cervical cancer cells at 48h (C33A and HeLa), (n) Bar graph of invasion assay in transfected control (NTSV) and *miR-34b-3p* expressing cervical cancer cells at 48h (C33A and HeLa). The black shaded bar shows the *miR-34b-3p*/gene expressions in cervical cancer cells carrying non-target sequence vector (NTSV), and the grey shaded bar shows the *miR-34b-3p*/gene expressions in cervical cancer cells carrying *miR-34b-3p* expressing vector. The symbols "NS" denotes non-significance;* denotes p-value ≤ 0.05; ** denotes p-value ≤ 0.01;*** denotes p-value ≤ 0.001, and **** denotes p-value ≤ 0.0001.

Supplementary Tables

Supplementary Table 1

Summary of the clinical metadata, including median age, five-year survival statistics, tumour stage, and grade associated with each sample.

Supplementary Table 2

Details of the non-cancerous cervical tissue and primary CESC, of which *miR-34b-3p*/gene expression data presented in the study, are listed.

Supplementary Files

Supplementary File 1: Results from functional enrichment analysis – GO (1a) and Reactome (1b).

Supplementary File 2: Results obtained from the correlation analysis between EMT genes and *miR-34b*.

Supplementary File 3: Derived EMT scores of TCGA-CESC.

Supplementary File 4: Primer details used in the study.

Supplementary Figures

Network analysis was derived using STRING-db v11.0 for pathways, networks and functions below. The gene names in each figure are coloured based on the intensity directly proportional to their expression, as shown in the colour bar.

Supplementary Figure S1: mTOR pathway

Supplementary Figure S2: G1/Cell cycle regulation

Supplementary Figure S3: Epithelial cell migration (ECM)

Supplementary Figure S4: Cell adhesion

Supplementary Figure S5: Representative microscopic images of well-differentiated, moderately differentiated and poorly differentiated CESC (Haematoxylin and Eosin stain) included in the study are presented.

Figure 1

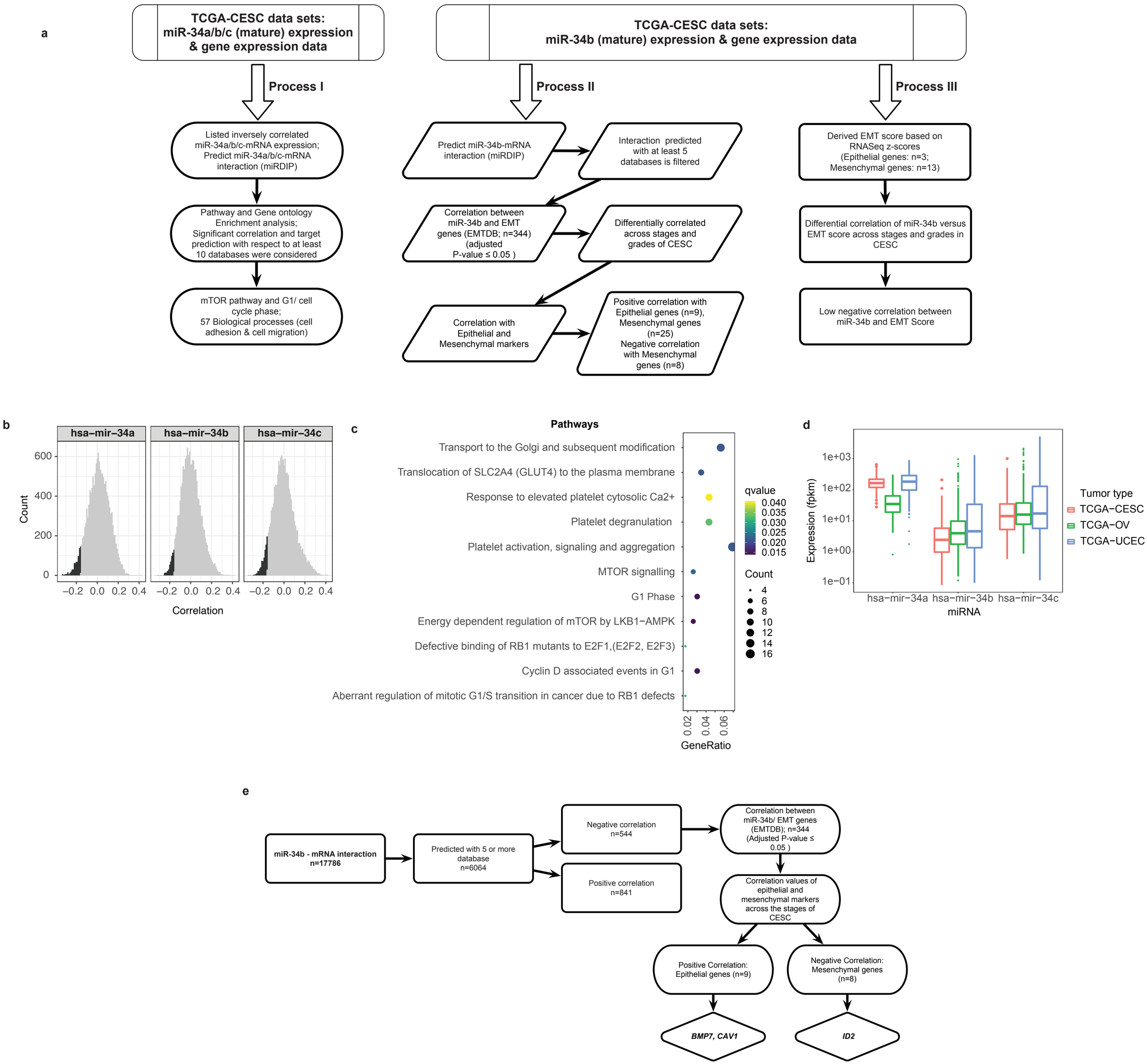
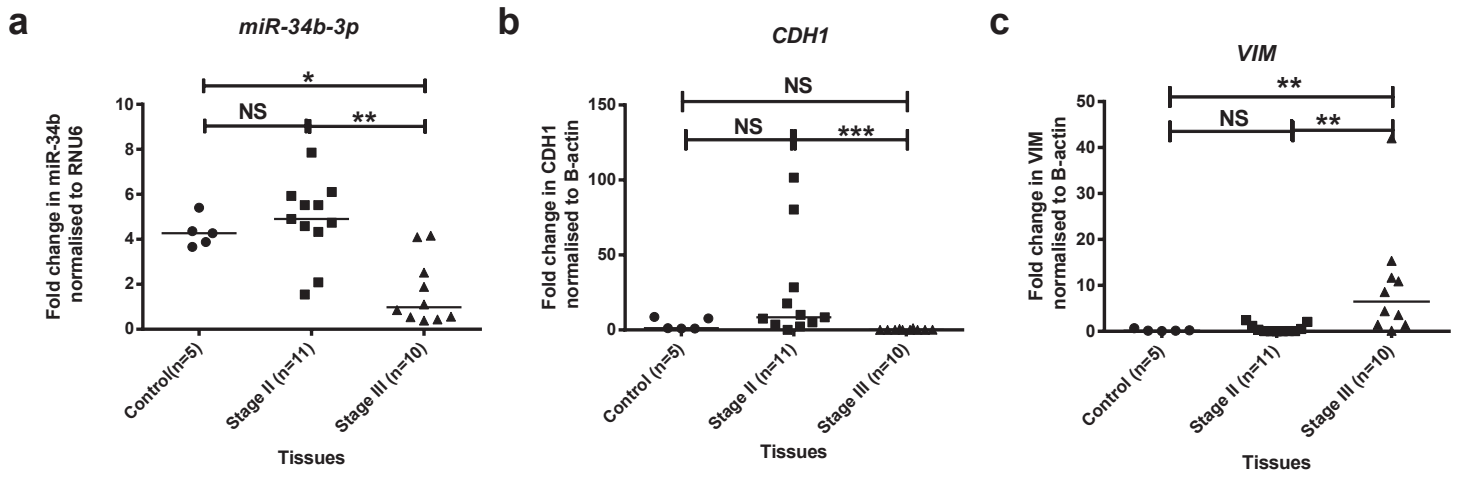
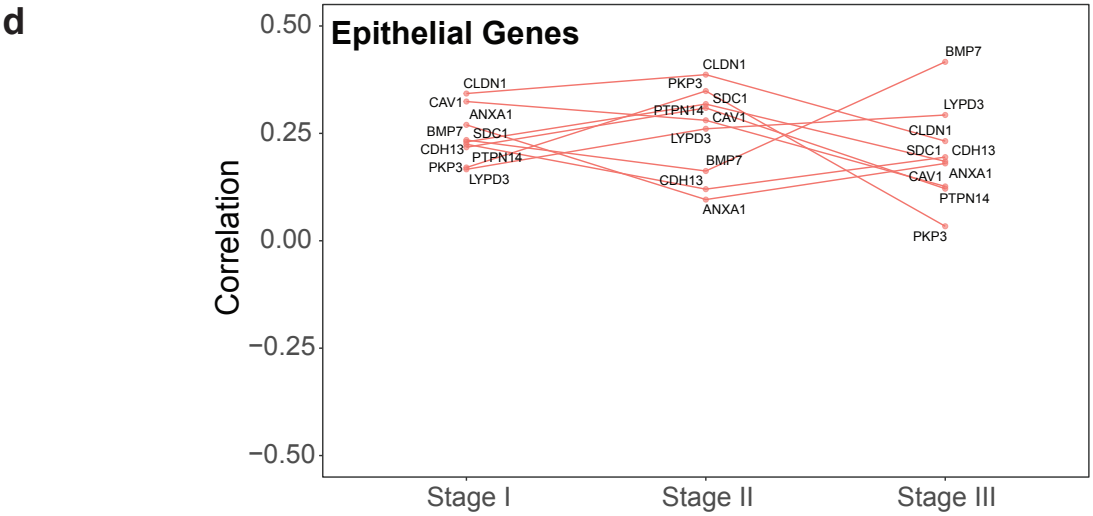

a
TCGA-CESC data sets:
miR-34a/b/c (mature) expression
& gene expression data
TCGA-CESC data sets:
miR-34b (mature) expression & gene expression data
Process I
Process II
Process III
Listed inversely correlated miR-34a/b/c-mRNA expression; Predict miR-34a/b/c-mRNA interaction (miRDIP)
Pathway and Gene ontology Enrichment analysis; Significant correlation and target prediction with respect to at least 10 databases were considered
mTOR pathway and G1/ cell cycle phase; 57 Biological processes (cell adhesion & cell migration)
Predict miR-34b-mRNA interaction (miRDIP)
Interaction predicted with at least 5 databases is filtered
Correlation between miR-34b and EMT genes (EMTDB; n=344) (adjusted P-value ≤ 0.05)
Differentially correlated across stages and grades of CESC
Correlation with Epithelial and Mesenchymal markers
Positive correlation with Epithelial genes (n=9), Mesenchymal genes (n=25) Negative correlation with Mesenchymal genes (n=8)
Derived EMT score based on RNASeq z-scores (Epithelial genes: n=3; Mesenchymal genes: n=13)
Differential correlation of miR-34b versus EMT score across stages and grades in CESC
Low negative correlation between miR-34b and EMT Score
b
hsa−mir−34a
hsa−mir−34b
hsa−mir−34c
Count
Correlation
c
Pathways
Transport to the Golgi and subsequent modification
Translocation of SLC2A4 (GLUT4) to the plasma membrane
Response to elevated platelet cytosolic Ca2+
Platelet degranulation
Platelet activation, signaling and aggregation
MTOR signalling
G1 Phase
Energy dependent regulation of mTOR by LKB1−AMPK
Defective binding of RB1 mutants to E2F1,(E2F2, E2F3)
Cyclin D associated events in G1
Aberrant regulation of mitotic G1/S transition in cancer due to RB1 defects
GeneRatio
qvalue
Count
d
Expression (fpkm)
miRNA
Tumor type
TCGA−CESC
TCGA−OV
TCGA−UCEC
e
miR-34b - mRNA interaction n=17786
Predicted with 5 or more database n=6064
Negative correlation n=544
Positive correlation n=841
Correlation between miR-34b/ EMT genes (EMTDB); n=344 (Adjusted P-value ≤ 0.05)
Correlation values of epithelial and mesenchymal markers across the stages of CESC
Positive Correlation: Epithelial genes (n=9)
Negative Correlation: Mesenchymal genes (n=8)
BMP7, CAV1
ID2

Figure 2


a
miR-34b-3p
Fold change in miR-34b normalised to RNU6
NS
*
**
Control(n=5)
Stage II (n=11)
Stage III (n=10)
Tissues
b
CDH1
Fold change in CDH1 normalised to B-actin
NS
NS
***
Control (n=5)
Stage II (n=11)
Stage III (n=10)
Tissues
c
VIM
Fold change in VIM normalised to B-actin
NS
**
**
Control (n=5)
Stage II (n=11)
Stage III (n=10)
Tissues


d
Epithelial Genes
Correlation
Stage I
Stage II
Stage III

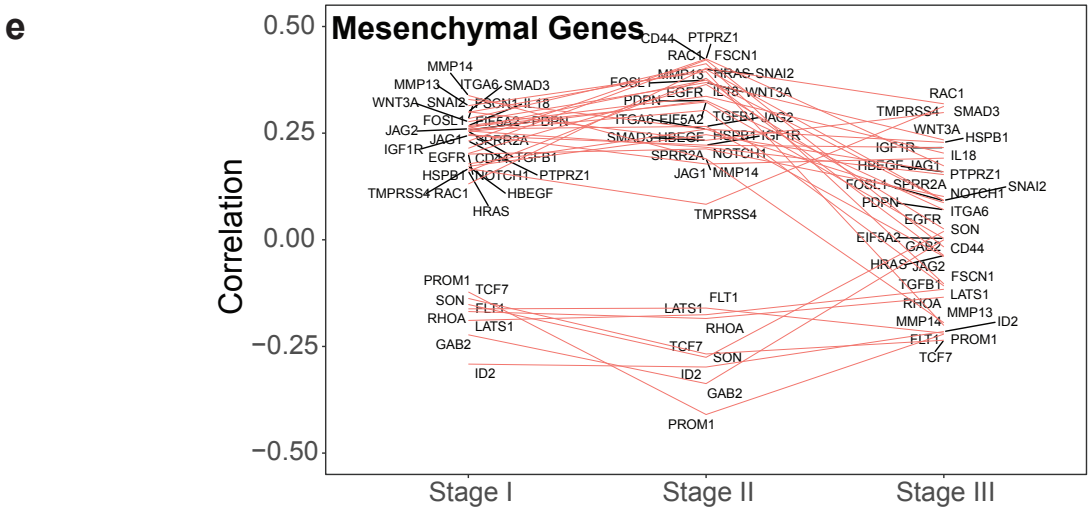

e
Mesenchymal Genes
Correlation
Stage I
Stage II
Stage III

Figure 3

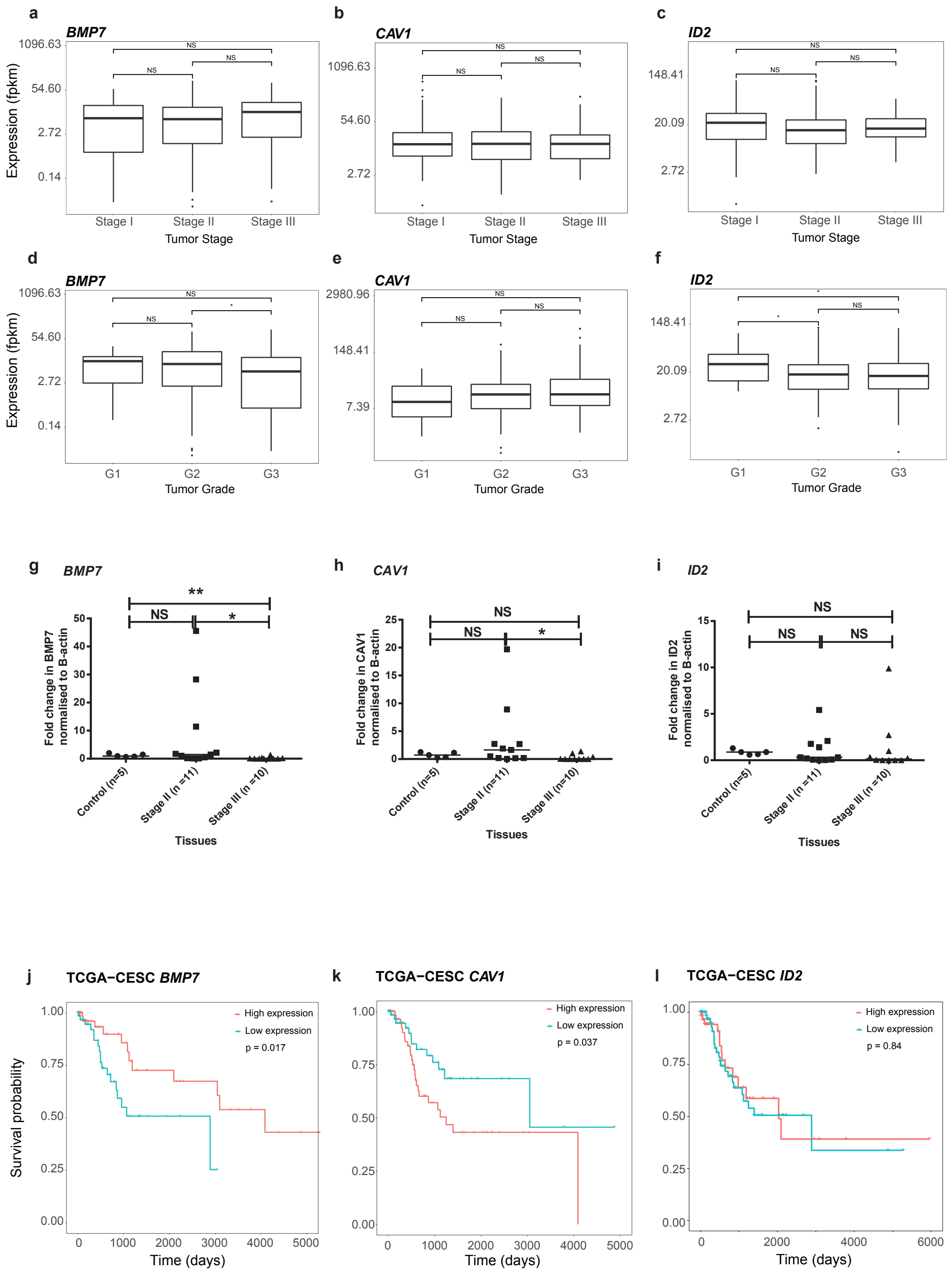
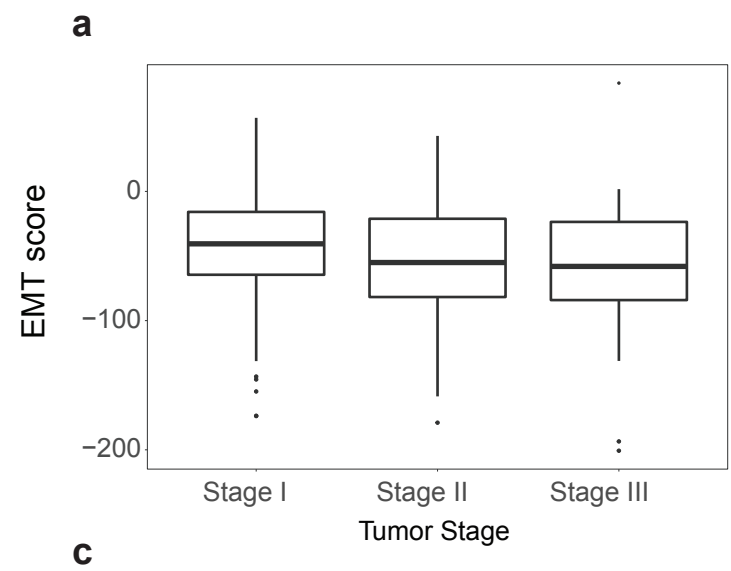
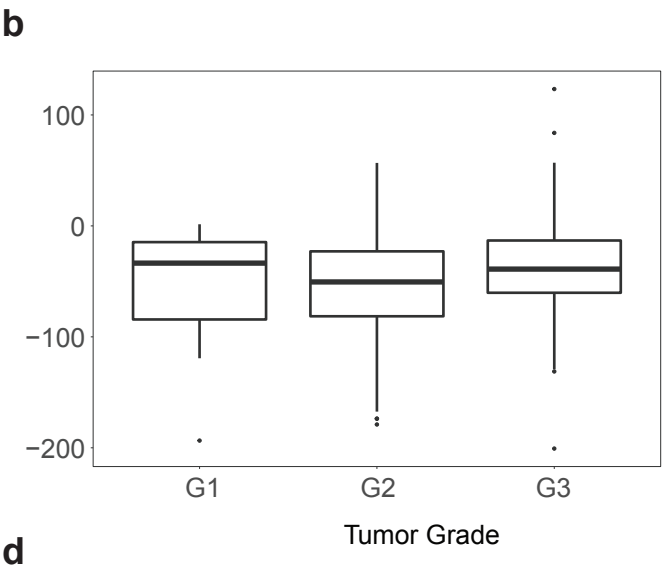
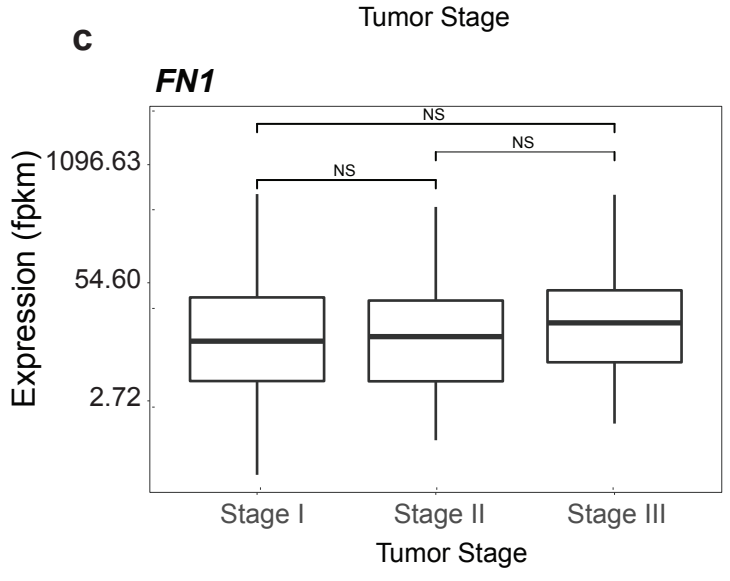
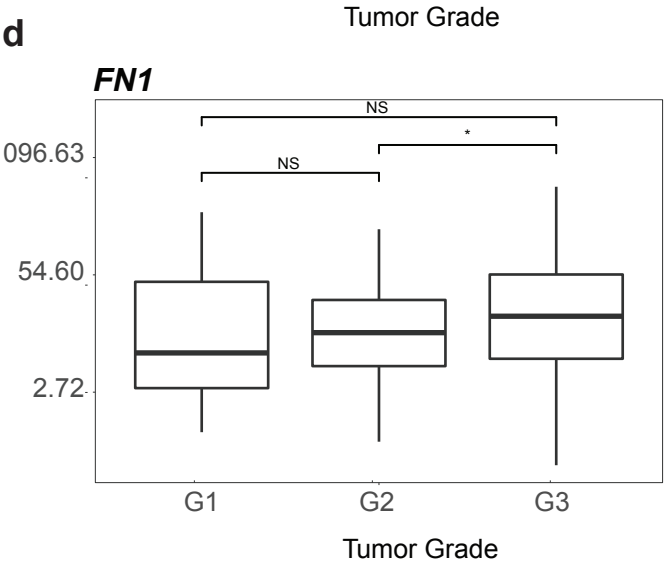

a
BMP7
Expression (fpkm)
1096.63
54.60
2.72
0.14
NS
NS
NS
Stage I
Stage II
Stage III
Tumor Stage
b
CAV1
1096.63
54.60
2.72
NS
NS
NS
Stage I
Stage II
Stage III
Tumor Stage
c
ID2
148.41
20.09
2.72
NS
NS
NS
Stage I
Stage II
Stage III
Tumor Stage
d
BMP7
Expression (fpkm)
1096.63
54.60
2.72
0.14
NS
NS
*
G1
G2
G3
Tumor Grade
e
CAV1
2980.96
148.41
7.39
NS
NS
NS
G1
G2
G3
Tumor Grade
f
ID2
148.41
20.09
2.72
*
*
NS
G1
G2
G3
Tumor Grade
g
BMP7
Fold change in BMP7 normalised to B-actin
50
40
30
20
10
0
**
NS
*
Control (n=5)
Stage II (n =11)
Stage III (n =10)
Tissues
h
CAV1
Fold change in CAV1 normalised to B-actin
25
20
15
10
5
0
NS
NS
*
Control (n=5)
Stage II (n=11)
Stage III (n=10)
Tissues
i
ID2
Fold change in ID2 normalised to B-actin
15
10
5
0
NS
NS
NS
Control (n=5)
Stage II (n =11)
Stage III (n =10)
Tissues
j
TCGA-CESC BMP7
Survival probability
1.00
0.75
0.50
0.25
0.00
High expression
Low expression
p = 0.017
0
1000
2000
3000
4000
5000
Time (days)
k
TCGA-CESC CAV1
1.00
0.75
0.50
0.25
0.00
High expression
Low expression
p = 0.037
0
1000
2000
3000
4000
5000
Time (days)
l
TCGA-CESC ID2
1.00
0.75
0.50
0.25
0.00
High expression
Low expression
p = 0.84
0
2000
4000
6000
Time (days)

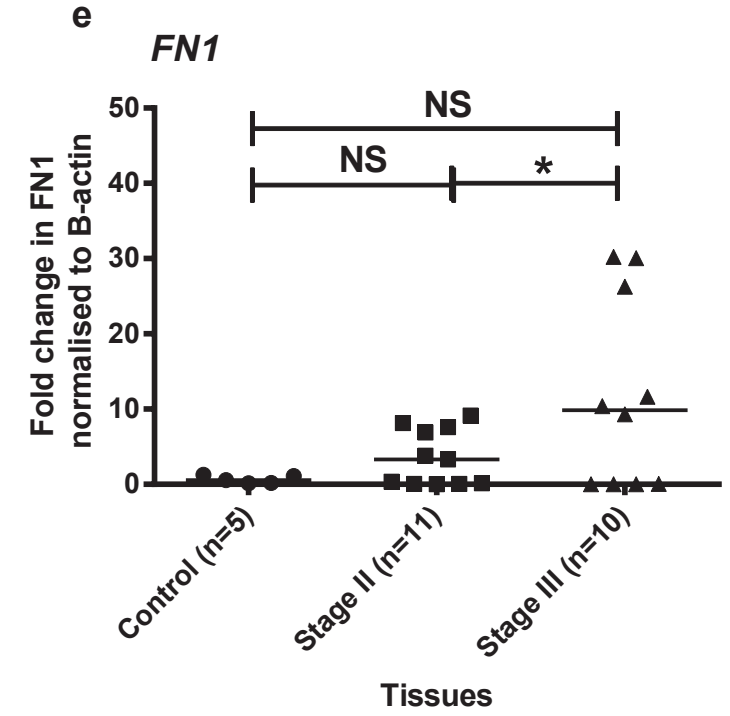
Figure 4

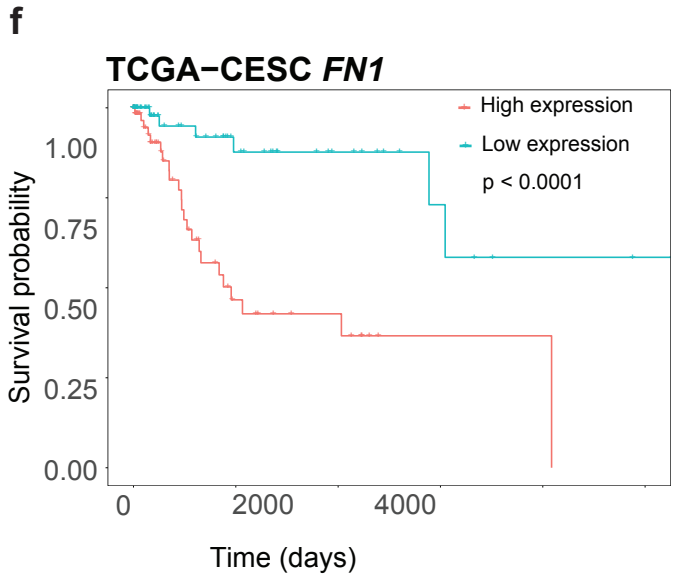










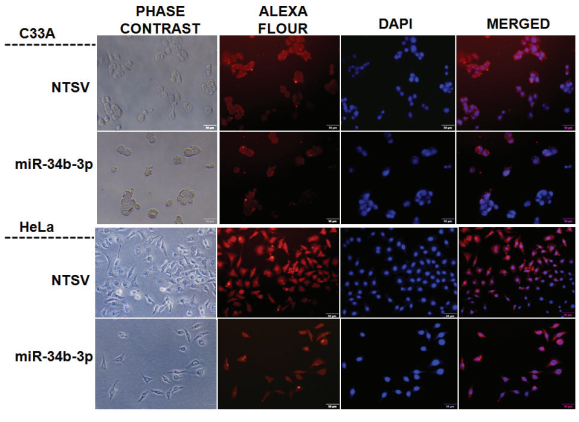


g

hsa-miR-34b/*FN1* alignment

```
     3′ uaccgucaccucaaUCACUAAc 5′ hsa-miR-34b
                        |||||||
246:5′ aguauuuuaaaugaAGUGAUUc 3′ FN1
```

mirSVR score: -0.9229
PhastCons score: 0.6314



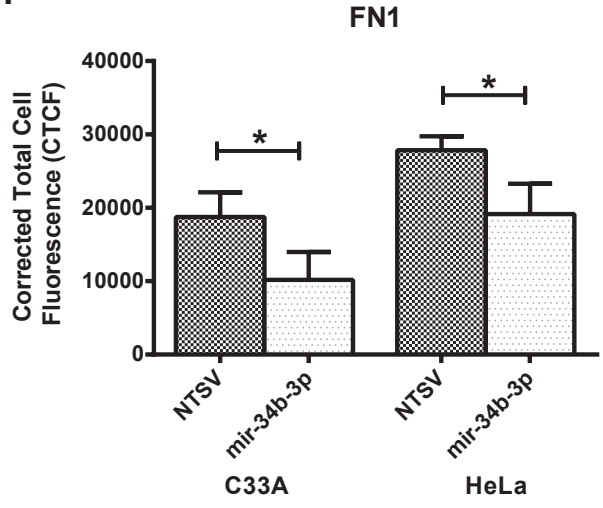
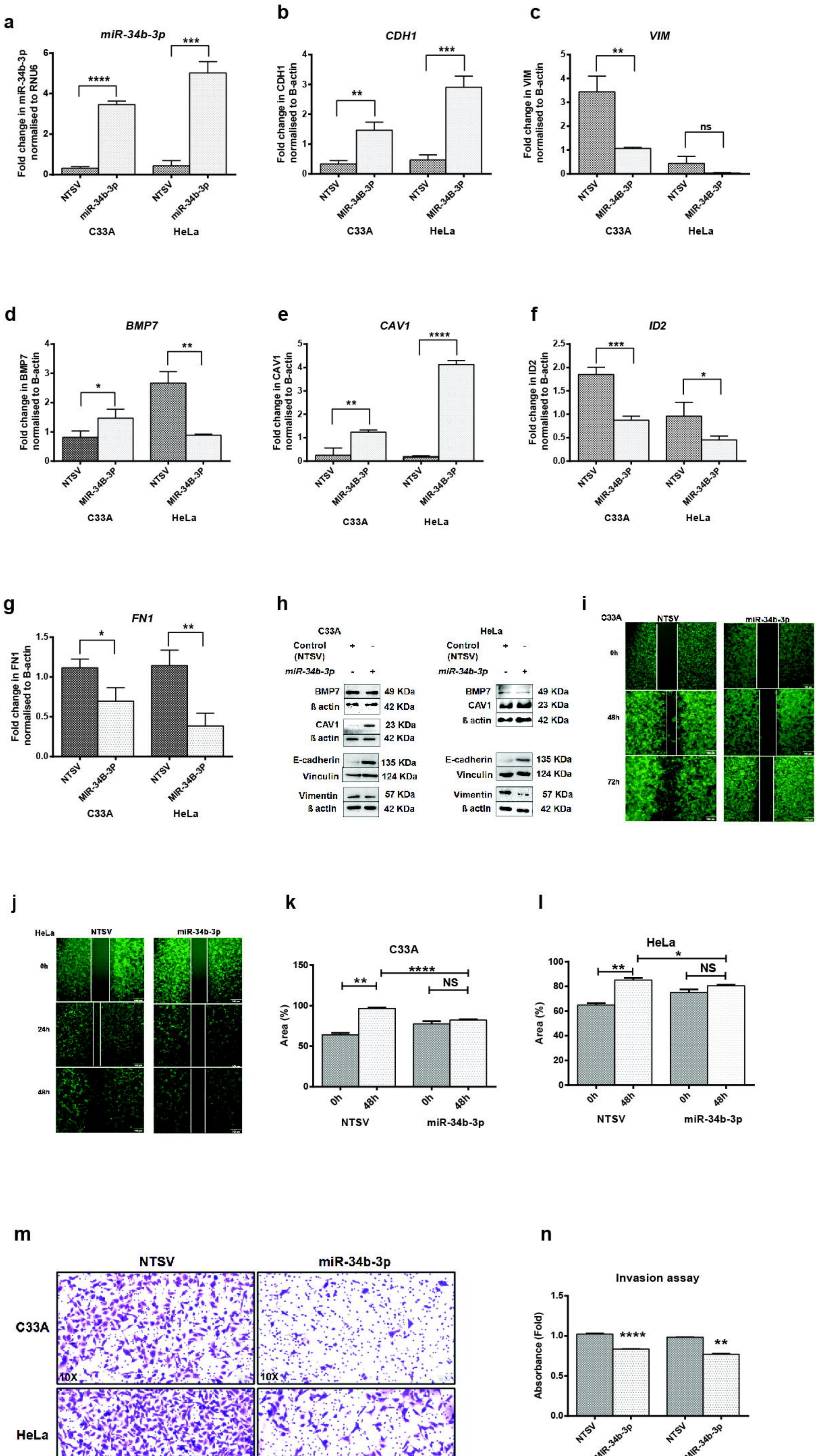


Figure 5


a
miR-34b-3p
Fold change in miR-34b-3p normalised to RNU6
****
***
NTSV
miR-34b-3p
C33A
HeLa
b
CDH1
Fold change in CDH1 normalised to B-actin
**
***
NTSV
MIR-34B-3P
C33A
HeLa
c
VIM
Fold change in VIM normalised to B-actin
**
ns
NTSV
MIR-34B-3P
C33A
HeLa
d
BMP7
Fold change in BMP7 normalised to B-actin
*
**
NTSV
MIR-34B-3P
C33A
HeLa
e
CAV1
Fold change in CAV1 normalised to B-actin
**
****
NTSV
MIR-34B-3P
C33A
HeLa
f
ID2
Fold change in ID2 normalised to B-actin
***
*
NTSV
MIR-34B-3P
C33A
HeLa
g
FN1
Fold change in FN1 normalised to B-actin
*
**
NTSV
MIR-34B-3P
C33A
HeLa
h
C33A
Control (NTSV)
miR-34b-3p
BMP7
ß actin
CAV1
ß actin
E-cadherin
Vinculin
Vimentin
ß actin
49 KDa
42 KDa
23 KDa
42 KDa
135 KDa
124 KDa
57 KDa
42 KDa
HeLa
Control (NTSV)
miR-34b-3p
BMP7
CAV1
ß actin
E-cadherin
Vinculin
Vimentin
ß actin
49 KDa
23 KDa
42 KDa
135 KDa
124 KDa
57 KDa
42 KDa
i
C33A
NTSV
miR-34b-3p
0h
48h
72h
j
HeLa
NTSV
miR-34b-3p
0h
24h
48h
k
C33A
Area (%)
**
****
NS
0h
48h
0h
48h
NTSV
miR-34b-3p
l
HeLa
Area (%)
**
*
NS
0h
48h
0h
48h
NTSV
miR-34b-3p
m
NTSV
miR-34b-3p
C33A
HeLa
10X
10X
10X
10X
n
Invasion assay
Absorbance (Fold)
****
**
NTSV
MIR-34b-3p
NTSV
MIR-34b-3p
C33A
HeLa