bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title

Identification and validation of putative target genes regulated by *miR-34* in cervical cancer

Nalini Venkatesan[1], Ashley Xavier[1,2], Sindhu K.J.[1], Himanshu Sinha[1,2,3,*], Karunagaran Devarajan[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]Robert Bosch Centre for Data Science and Artificial Intelligence, Indian Institute of Technology Madras, Chennai, India

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

List of abbreviations		
S.No	**Abbreviation**	**Description**
1	API	Application Programming Interface
2	CESC	Cervical Epithelial Squamous cell Carcinoma
3	CPM	Counts Per Million
4	Ct	Cycle threshold
5	dbEMT	Database for EMT genes
6	DMEM	Dulbecco's Modified Eagle's Medium
7	ECM	Epithelial Cell Migration
8	EMT	Epithelial-Mesenchymal Transition
9	FDR	False Discovery Rate
10	FBS	Foetal Bovine Serum
11	FPKM	Fragments Per Kilo-base of exon per Million mapped
12	FIGO	International Federation of Gynaecology and Obstetrics
13	GO	Gene ontology
14	GDC	Genomic Data Commons
15	GFP	Green Fluorescence Protein
16	ID	Identity
17	mRNA	Messenger Ribonucleic Acid
18	miRNA	Micro Ribonucleic Acid
19	MRE	miRNA Response Element
20	OV	Ovarian cancer
21	qPCR	quantitative Polymerase Chain Reaction
22	TCGA	The Cancer Genome Atlas
23	UCEC	Uterine Corpus Endometrial Carcinoma
24	3'UTR	3' Untranslated Regions

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The emergence of large-scale transcriptomic data provides the opportunity for identifying novel putative targets of microRNAs (miRNAs). In this study, we followed a computational pipeline to predict the candidate gene targets of the *miR-34* family. This approach integrates the expressions of *miR-34* with genes of heterogeneous primary cervical epithelial squamous cell carcinomas (CESC). Integration of *miR-34b* and epithelial-mesenchymal transition (EMT) regulated genes has also been focussed, EMT being a reversible process that fuels cancer metastasis. An *in-silico* approach involving three processes was carried out with CESC datasets of the cancer atlas genome (TCGA), which includes correlation analysis, target prediction database lookup, functional enrichment, network analysis, survival analysis, and EMT score derivation. The results indicate that the *miR-34* family may regulate the candidate genes of the mTOR pathway, cell cycle (*CCND2*) and cell adhesion functions (*FZD4*). Further, the study reveals the possible regulation of EMT signature genes, namely *BMP7*, *CAV1* and *ID2*by *miR-34b*. Further, these transcriptomic signatures were validated in a subset of CESC from the South Asian Indian population (n = 10) and in non-cancerous cervical tissues (n = 5). Upon stably expressing *miR-34b* in cervical cancer cells (C33A and HeLa), we found repression of these candidate genes and a low negative correlation (r^2 = -0.07) between *miR-34b* and EMT score indicating *FN1* as its putative target. Together, these studies revealed the potential targets of the *miR-34* family, especially *miR-34b*, with the hope that they would emerge as potential biomarkers and/or promising therapeutic targets in CESC.

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Brief Description

A combined analysis of *miR-34* and gene expression in heterogeneous primary CESC, along with the integration of *miR-34b* and EMT regulated genes, was used to predict the candidate gene targets of the *miR-34* family. The results show that the *miR-34* family may regulate the mTOR pathway, cell cycle, and cell adhesion functions. Further, we showed that EMT signature genes (*BMP7*, *CAV1*, *ID2*, *FN1*) were regulated by *miR-34b* in CESC and cervical cancer cells.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Carcinoma of the uterine cervix, the most common health problem in women, is on a steady rise, leading to cancer-related deaths worldwide[1]. 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[2]. 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[3]. Although the current strategies in diagnosing and treating cervical cancers have improved, it remains the second largest cancer in women with high mortality. Also, advanced and recurrent cervical cancer results in a poor prognosis with 10-20% of the one-year survival rate[4]. 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 poor prognosis is EMT initiation, triggered by the cellular processes, which include the dysregulation of certain tumour suppressors, oncogenes, transcription factors, growth factor signalling, and miRNAs.

There exists a positive correlation between EMT phenotype and increased tumour progression, invasion, and metastasis in primary cervical cancers[5]. EMT is known to play an essential role in tumour cell invasion and migration by involving numerous cytokines, transcription factors and cellular pathways such as nuclear factor-kappa B (*NF-κB*), phosphatidylinositol 3-kinase (*PI3K*), Wnt, and transforming growth factor-β (*TGF-β*)[6], etc. During the EMT process, the tumour cells lose their epithelial molecules (E-Cadherin, keratin) and gain the mesenchymal molecules (Vimentin, Fibronectin, and 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cervical cancer. Despite many studies that have been conducted to understand the molecular insights involved in the EMT regulatory factors, there is a lack of concrete findings to halt cervical cancer progression.

microRNAs (miRNAs), small 18-22 nucleotides long non-coding RNAs, act as RNA silencers or post-transcriptional regulators of messenger RNAs (mRNAs). They interact with specific microRNA Response Elements (MREs) on the 3' Untranslated Regions (3'UTR) of mRNAs through complete or partial complementarity[8-13]. Changes in microRNA and gene expressions have been documented in tumour initiation and progression[14], such as decreased expression of miRNAs explaining its intrinsic role in tumour suppression[15]. However, the underlying mechanisms involved in the miRNA-mRNA modulation contributing to cervical cancer progression remain to be explored.

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* encoded by exon 2 of chromosome 1 (ch1p36.22) is well characterised. Meanwhile, *miR-34b/c* exists as the polycistronic transcript located within intron1 and exon 2 on chromosome 11 (ch11q23.1)[24]. Interestingly, the *miR-34* family is known to regulate p53 directly. 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]. Further, a significant association exists between

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

miR-34b and phospho-Met, *p53* (phospho S392), Mdm2 in non-small cell lung cancer (NSCLC)[28]. In a recent report, *miR-34b* was shown to mediate TGF- β1 regulation of cervical cancer cell proliferation and apoptosis[23]. While these reports show that *miR-34b* plays a vital role in the EMT regulation of solid tumours, including cervical cancer, targets to *miR-34b* affecting cervical cancer and its progression are not known.

In this study, we describe a computational approach that includes three processes to identify targets of *miR-34b* contributing to EMT progression in cervical cancer from the TCGA-CESC dataset, as shown in Figure1A. We validate candidate *miR-34b* targets of cervical cancer progression and EMT pathway in primary CESC and cervical cancer cell lines derived from the South Asian Indian population to show the correlation between the expression of *miR-34b* and target genes. We also show that candidate genes are repressed in the ectopically *miR-34b* expressing stable cervical cancer cells. We, therefore, establish the role of *miR-34b* in cervical cancer progression and identify potential therapeutic targets for further validation.

Methods

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 openly accessible 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sample were downloaded through the GDC Application Programming Interface (API). Supplementary Table 1 describes the summary of 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.

Identify potential gene targets of *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[29]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mirDIP 4.1 database was downloaded and used to compare the target prediction across 30 individual resources[30]. The mirDIP 4.1 database has been used to compare within 26 different databases of microRNA-mRNA target prediction tools. 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[31]. 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[32,33]. Only predictions that were categorised as either good or better according to the miDIP database were used. As microRNA target prediction databases are known to have a large number of 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 stronger indicator.

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 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 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 of the correlation was significant (p-value < 0.01).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The EMT genes from a first literature-based database for EMT human genes (dbEMT) by Zhao et al. (2015) was used in the study[34]. 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 ≤ of 0.01 were selected and analysed across stages/grades of CESC. The analysed results are presented in Supplementary File 2 and derived a co-expression network using Cytoscape v3.7.1. The International Federation of Gynaecology and Obstetrics (FIGO) staging system, mainly based on clinical examination and certain procedures (imaging), was used in the study[35]. The tumour stage of each sample was classified into Stage I, Stage II, Stage III, and Stage 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)[35]. The expression of the candidate genes was checked across stage 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.

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)[36]. The epithelial marker genes were E-Cadherin (*CDH1*), *DSP*, and *TJP1*; while the mesenchymal marker genes were Fibronectin (*FN1*), Vimentin (*VIM*), N-Cadherin (*CDH2*), Integrin beta 6

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(*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 there was no data 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 went up, 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.

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.

***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 control (Cat# VSC11651; Dharmacon™) was purchased as glycerol stocks. The packaging lentiviral particle, transduction and puromycin

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

selection was carried out as per the manufacture's protocol. The stable cell lines were established for HeLa and C33A cervical cancer cell lines. Human cervical cancer cell lines, C33A and HeLa, 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.

The total RNA is extracted from the stable cell lines (C33A/HeLa) using Trizol (RNA Iso Plus reagent, Takara) extraction and converted to cDNA using Primescript RT reagent kit (Cat# RR037A, DSS Takara Bio India Private Ltd) as per the manufacturer's instructions. A total RNA of fresh human cervical cancer tissues (n = 10) and normal cervical tissues (n = 5) were included in this study. 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 expression of the genes using a real-time PCR system (Applied BiosystemQuant Studio 7 Flex) with the following protocol (Denaturation:95 °C/5mins; Amplification for 40 cycles: 95 °C/30s, 60 °C/45s, 72°C/30s). The primers used in this study are given in Supplementary File 1. For expression analysis, the relative Ct (cycle threshold) quantification of genes in tumour and normal cell lines were analysed. All samples were assayed in triplicates**.** The unpaired t-test was performed using GraphPad Prism 6.

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 2. 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/).

Results and Discussion

mRNA targets of the *miR-34* family by miRNA-mRNA correlation analysis

miRNA target prediction strategies alone do not indicate whether a particular pair of miRNA and mRNA interactions is biologically relevant. Therefore, a negative correlation between the pair 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 that 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 microRNAs, and to obtain a confident set of target prediction genes, we iterated the number of databases in which a gene was predicted as a target. If we set that at least ten databases predict a gene as a target, we got 4,029 predicted targets out of 27,591 genes. When we reduced this stringency to 5 or more databases, we obtained 10,886, while for five or fewer databases, the number was 14,889. Finally, when we set the cut off for the number of target databases to exactly 5, we got 19,675 predicted genes across the *miR-34* family. Of these 19,675 genes, we observed that the *miR-34* family was significantly negatively correlated with 1,648 genes at p-value ≤ 0.05 after Benjamini-Hochberg correction. We took the 4,029 genes obtained at a cut-off of 10 databases as the gene list for further analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gene ontology analyses for a biological process for the above gene list identified 18 biological processes, including epithelial cell migration and adhesion (q-value ≤ 0.05, Supplementary File 2). 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). *mi-34*a was predicted to interact with the mTOR pathway genes specifically; namely, Protein Kinase AMP-Activated Non-Catalytic Subunit Beta 2 (*PRKAB2*), Calcium Binding Protein 39 (*CAB39*), Protein Phosphatase 1A (*PPM1A*); *miR-34b* with genes such as 5'-AMP-activated protein kinase catalytic subunit alpha-2 (*PRKAA2*), Calcium Binding Protein 39 Like (*CAB39L*) and *miR-34c* with genes like 5'-AMP-activated protein Kinase catalytic subunit alpha-2 (*PRKAA2*), Calcium Binding Protein 39 Like (*CAB39L*), STE20 Related Adaptor Alpha (*STRADA*) has been observed (Supplementary Figure S1).

We observed that miR34a and *miR-34c* regulated the key cell cycle genes, Cyclin D1 (*CCND1*), Cyclin D2 (*CCND2*), Cyclin-dependent Kinase inhibitor 1B (*CDKN1B*), E2F Transcription Factor 3 (*E2F3*), E2F Transcription Factor 5 (*E2F5*) and ABL Proto-Oncogene 1 (*ABL1*; Supplementary Figure S2). *CCND2* is a regulator of cell cycle G1/S transition in cancers[37], and miRNA mediated repression of *CCND2* lowered the proliferation, migration and invasion properties of cervical cancer cells[38]. 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 key genes such as Transforming Growth Factor Beta 1 (*TGFB1*), Transforming Growth Factor Beta 2 (*TGFB2*), *GATA3*, Mitogen-Activated Protein Kinase Kinase Kinase 3 (*MAP3K3*), Mitogen-Activated Protein Kinase Kinase Kinase Kinase 4 (*MAP4K4*) and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cadherin-13 (*CDH13*). Also, *miR-34b* and *miR-34c* regulated the genes, like Protein Tyrosine Phosphatase Receptor Type D (*PTPRTG*), Arylsulfatase B (*ARSB*) and Phosphoinositide-3-Kinase Regulatory Subunit 3 (*PIK3R3*, Supplementary Figure S3).

Further, the regulatory networks with the genes involved in cell adhesion function revealed a unique gene cluster for *miR-34a* while *miR-34b* and *miR-34c* regulates Frizzled Class Receptor 4 (*FZD4*) and LIM and Calponin Homology Domains 1 (*LIMCH1*) genes (Supplementary Figure 4). This unique cluster of *miR-34a* includes the key genes, namely Mothers against DPP Homolog 3 (*SMAD3*), Jagged Canonical Notch Ligand 1 (*JAG1*), Integrin alpha-6 (*ITGA6*), *CD44* and *CDH13*. All these genes were widely studied for their role with respect to 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.

mRNA targets of *miR-34b*

Further, we were interested in exploring the regulatory targets of *miR-34b* specifically 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, namely ovarian cancer (TCGA-OV) and uterine corpus endometrial carcinoma (TCGA-UCEC, Figure 1D). Next, we applied a similar strategy to the *miR-34* family using the mirDIP4.1 target prediction database. We identified 17,786 genes out of 27,591 total genes in 26 databases and 6,064 in more than five databases, with a median of 4 databases predicting 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 significant negative correlations and 841 genes showed significant positive correlations after Benjamini-Hochberg correction for multiple

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

comparisons (p-value < 0.05). Out of 1,385 genes, a total of 344 genes were known to be involved in the regulation of the EMT pathway. Figure 1E shows the illustration of the steps followed in this analysis. As we observed a decreased level of *miR-34b* in the three epithelial cancers (TCGA-CESC, TCGA-OV and TCGA-UCEC), we focussed on the interaction between *miR-34b* versus epithelial and mesenchymal markers for our 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 EMT process, we analysed TCGA-CESC data sets and CESC primary tumour tissues (n = 10) obtained from the Asian Indian population along with non-cancerous cervical tissues (n = 5) as control. First, we carried out qPCR analyses for *miR-34b*, *CDH1* (epithelial signature gene) and *VIM* (mesenchymal signature gene) expressions in these select CESC and non-cancerous cervical tissues. Our results showed a decreased expression in the median of *miR-34b* (0.55-fold change; Figure 2A), *CDH1* gene (0.02-fold change; Figure 2B), and increased expression in the median of *VIM* gene (10.9-fold change; Figure 2C) in stage III of Asian Indian population derived CESC. These results corroborate with that obtained from the TCGA-CESC datasets as described in the section below (*miR-34b* positive correlation with epithelial gene, Figure 2D).

Epithelial genes show a positive correlation with miR34b in cervical cancer

Upon comparing correlation values of EMT markers across the stages of TCGA-CESC, we observed about nine epithelial markers such as Annexin A1 (*ANAX1*), Bone morphogenic protein 7 (*BMP7*), Caveolin-1 (*CAV1*), Cadherin 13 (*CDH13*), Claudin 1 (*CLDN1*), LY6/PLAUR Domain Containing 3 (*LYPD3*), Plakophilin-3 (*PKP3*), Protein Tyrosine

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phosphatase Non-receptor Type 14 (*PTPN14*) and Syndecan-1 (*SDC1*), had a positive correlation across the stages of CESC (Figure 2D, Supplementary File 3). *BMP7* was known to inactivate the EMT associated genes, thereby contributing to the reduced *TGFβ1*-mediated cell growth and metastasis in breast cancer[39] and a significant association of decreased *BMP7* in primary breast cancer cells with clinically overt bone metastasis was observed earlier[40]. Similarly, *CAV1* exhibited a positive correlation with E-cadherin (*CDH1*) gene expression, thereby regulating EMT in gastric cancers[6,41]. Therefore, based on the high positive correlation across stages, biological significance, 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 3).

First, we analysed the expression of *BMP7* across the TCGA-CESC data sets and found no significant change across the stages (Figure 3A), while a significant change was observed across G2 versus G3 grades of TCGA-CESC (Figure 3D). Based on these results, we decided to analyse *BMP7* expression in CESC primary tumour tissues (n = 10), and the results showed a significant decrease in *BMP7* expression in stage III CESC, while no change was observed in stage II CESC compared to the control cervical tissues (n = 5, Figure 3G). However, 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 xenograft cervical cancer mouse cells with human recombinant *BMP7* resulted in tumour growth arrest by triggering hTERT gene repression[42]. Therefore, there is significant evidence for *miR-34b*-mediated *BMP7* modulation, which has the potential to contribute to CESC growth arrest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 increase in stage II and an insignificant decrease in stage III than the non-cancerous cervical tissue (p-value < 0.05; Figure 3H). A recent study on cervical cancer cell lines has reported the decreased level of *CAV1* gene expression, and the miR-96 mediated restoration of the *CAV1* has resulted in the decrease of cell proliferation, migration, and invasion[43]. Also, other studies have indicated the tumour suppressor role of *CAV1* in various cancers, including cervical cancer[44-46]. However, in the present study, a significant decrease in survival period was observed in patients having higher *CAV1* gene expression in TCGA-CESC datasets (p-value = 0.03, 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 to understand their role in CESC tumour management.

Mesenchymal genes show a negative correlation with *miR-34b* in cervical cancer

Following epithelial markers, we observed a significant negative correlation of eight mesenchymal markers with *miR-34b*, namely Ras homolog family member A (*RHOA*), GRB2 Associated Binding Protein 2 (*GAB2*), SON, Prominin 1 (*PROM1*), Transcription Factor 7 (*TCF7*), Inhibitor Of DNA Binding 2 (*ID2*), Fms Related Receptor Tyrosine Kinase 1 (*FLT1*) and Large Tumor Suppressor Kinase 1 (*LATS1*, Figure 2E, Supplementary File 3). Seven 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

its significance in CESC tumour progression. *ID2* has been reported as an EMT attenuator in triple-negative breast cancers[47] and multiple myeloma[48]. 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[49]. In the present study, an insignificant change in *ID2* gene expression was observed between the stages (Figure 3C), while a significant change was observed between the grades of TCGA-CESC datasets (G1 versus G2; G1 versus G3; p-value <0.05, Figure3F). Further, the qPCR analysis in primary cervical cancer tissues of the South Asian Indian population showed an insignificant change of *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 substantial 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 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 Heat shock protein beta-1 (*HSPB1*), Insulin-Like Growth Factor 1 Receptor (*IGF1R*), Snail Family Transcriptional Repressor 2 (*SNAI2*), Mothers against DPP Homolog 3 (*SMAD3*, Figure 2E) and this observation can be further studied with experimental validation on its association with *miR-34b*.

miR-34b and gene EMT score

Previous studies have discussed the various methods to derive EMT score[50] and to understand further 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

primary tumours and also aiding to identify the putative therapeutic targets and therapy resistance in cancers. In the present study, a differential correlation analysis of *miR-34b* expression with EMT scores across the stages and grades of TCGA-CESC datasets was performed (Supplementary File 4). As this analysis precisely involves the quantitative measurement of EMT signature genes, its outcome would aid us to identify 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 (Stage I: -0.013; Stage II:-0.176; Stage III: -0.325, Figure 4A) and a varied correlation across grades (G1: -0.050; G2: -0.173; and G3: 0.009, Figure 4B). These results indicate that if there is an increase in *miR-34b* expression, we may obtain a more negative EMT score, indicating 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 (G2 versus G3; p-value = 0.002, Figure 4D). In addition, a significant decrease in the patient's survival with high *FN1* expression was also observed (p-value ≤ 0.0001; Figure 4E). Also, *FN1* has MRE in its 3'UTR for *miR-34b* (Figure 6F), indicating it is inversely regulated by *miR-34b*. Based on these results, and with further experimental validation, one could prove that *FN1* could be a potential target of *miR-34b* in CESC.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Validation of the *in-silico* analysis in *miR-34b* expressing cervical cancer cells

To ascertain the select candidate genes 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* to 4-5 folds in cervical cancer cells (C33A and HeLa) after 48h of induction, as shown in Figure 5A. At this time point, a significant increase in the CDH1 (about 1.5-3.0-fold, Figure 5B) and a decrease in VIM (approximately 1.0-0.3 fold, Figure 5C) compared to transfected control cells (NTSV) was observed. This 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 (about 1.5-fold) while a decrease expression in HeLa cells (approximately 0.9-fold) compared to the transfected control cells (Figure 5D). The difference in the phenotype of cell lines may contribute to the variation of *BMP7* gene expression. The observed decrease in *BMP7* gene expression in the HeLa cells corroborates with an earlier report where the *BMP7* knockdown inhibits EMT by inducing E-cadherin and decreasing Vimentin[51]. However, the increased *BMP7* gene levels observed in C33A trigger for 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[39]. Next, a significant increase of *CAV1* (approximately 1.25-4.0-fold) and a significant decrease of *ID2* (about 0.87-0.45-fold) was observed in the cervical cancer cells (Figure 5E, F). 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, *ID2* being reported as an EMT attenuator in other cancers, its suppression by *miR-34b* in cervical cancer cells may contribute to better

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tumour management. Thus, the inverse changes of select candidate gene expressions in the ectopically *miR-34b* expressing cervical cancer cells (Figure 5) to that of primary cervical cancers (Figure 2, 3) indicate their association with each other. Therefore, these results are evidence for *miR-34b*-mediated select candidate genes to modulate and their potential to contribute to tumour regression in CESC. However, further experimental validation can be carried out to understand the regulation of these candidate genes by *miR-34b*.

Conclusions

Our 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. Further using EMT scores, we show that *FN1* can be a direct target regulated by *miR-34b*. Also, 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*. Following this identification, the experimental validation using gene expression studies in ectopically expressed *miR-34b* cervical cancer cells indicates their association with each other. The advantage of the present study is that it has used the integrated approach to reveal the novel targets for the *miR-34* family, especially *miR-34b* involving the experimental proof 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 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 and paves the way for a more detailed study to understand their interactions and help in epithelial cancer cell metastasis management of cervical cancer.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

The authors would like to thank the Indian Institute of Technology Madras and Indian Council of Medical Research for the postdoctoral fellowship of Dr Nalini Venkatesan; Dr A.K. Munnirajan, Professor and Head, Department of Genetics, University of Madras, Chennai, for the cDNA of CESC from South Asian Indian population.

Supplementary Files

Supplementary File1: Primers' details used in the study.

Supplementary File 2: Results obtained from functional enrichment analysis (Reactome and GO).

Supplementary File 3: Results obtained from the correlation analysis between EMT genes and *miR-34b*.

Supplementary File 4: Derived EMT scores of TCGA-CESC.

Supplementary Figures

Network analysis derived by using STRING-db v11.0 for

S1: mTOR pathway

S2: G1/Cell cycle regulation

S3: Epithelial cell migration (ECM)

S4: Cell adhesion

Supplementary Table 1

Summary of the clinical metadata such as median age, five-year survival statistic, tumour stage, and grade associated with each of the samples.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. 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 data set. (E) Illustration of the steps followed in the current study.

Figure 2

(A) Cluster graph of gene expression across stages (I/II/III) in South Asian Indian population for the *miR-34b*. (B) Cluster graph of gene expression across stages (I/II/III)in South Asian Indian population for the *CDH1*gene. (C) Cluster graph of gene expression across stages (I/II/III) in 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 * denotes p-value ≤0.05; ** denotes p-value ≤0.01.

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 stages (I/II/III)

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in 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.

Figure 4

(A) Correlation analysis between *miR-34b* and EMT score in TCGA-CESC. (B) Box plot graph across stages (I/II/III) of *FN1* expression in TCGA-CESC. (C) Box plot graph across grades (G1/G2/G3) of *FN1* expression in TCGA-CESC. (D) Survival analysis (KM plot) of TCGA-CESC based on their median *FN1* expression. (E) Sequence alignment analysis between *miR-34b* and 3'UTR of *FN1*.The symbol "NS" denotes non-significance; * indicates p-value ≤ 0.05.

Figure 5

(A-F) 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*. 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 * denotes p-value ≤ 0.05; ** denotes p-value ≤ 0.01;*** denotes p-value ≤ 0.001, **** denotes p-value ≤ 0.0001.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;**65**: 87-108.
2. Srivastava AN, Misra JS, Srivastava S, Das BC, Gupta S. Cervical cancer screening in rural India: Status & current concepts. *Indian J Med Res* 2018;**148**: 687-96.
3. Monica, Mishra R. An epidemiological study of cervical and breast screening in India: district-level analysis. *BMC Womens Health* 2020;**20**: 225.
4. Diaz-Padilla I, Monk BJ, Mackay HJ, Oaknin A. Treatment of metastatic cervical cancer: future directions involving targeted agents. *Crit Rev Oncol Hematol* 2013;**85**: 303-14.
5. Lee MY, Chou CY, Tang MJ, Shen MR. Epithelial-mesenchymal transition in cervical cancer: correlation with tumour progression, epidermal growth factor receptor overexpression, and snail up-regulation. *Clin Cancer Res* 2008;**14**: 4743-50.
6. Zhang K, Yang G, Wu W, Zhang J, Xia X, Jiang T, Cao J, Huang K, Qiu Z, Huang C. 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* 2016;**11**: 236-44.
7. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* 2014;**15**: 178-96.
8. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;**116**: 281-97.
9. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. *PLoS Biol* 2005;**3**: e85.
10. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* 2001;**293**: 834-8.
11. Kong W, Zhao JJ, He L, Cheng JQ. Strategies for profiling microRNA expression. *J Cell Physiol* 2009;**218**: 22-5.
12. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 2003;**425**: 415-9.
13. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 2003;**115**: 199-208.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

14. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* 2006;**103**: 2257-61.

15. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* 2004;**101**: 2999-3004.

16. Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S, Brabletz T. A reciprocal repression between *ZEB1* and members of the *miR-200* family promotes EMT and invasion in cancer cells. *EMBO Rep* 2008;**9**: 582-9.

17. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. The *miR-200* family and *miR-205* regulate epithelial to mesenchymal transition by targeting *ZEB1* and *SIP1*. *Nat Cell Biol* 2008;**10**: 593-601.

18. Hu Y, Tang H. MicroRNAs regulate the epithelial to mesenchymal transition (EMT) in cancer progression. *MicroRNA* 2014;**3**: 108-17.

19. Korpal M, Lee ES, 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* 2008;**283**: 14910-4.

20. Cordova-Rivas S, Fraire-Soto I, Mercado-Casas Torres A, Servin-Gonzalez LS, Granados-Lopez AJ, Lopez-Hernandez Y, Reyes-Estrada CA, Gutierrez-Hernandez R, Castaneda-Delgado JE, Ramirez-Hernandez L, Varela-Silva JA, Lopez JA. 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* 2019;**20**.

21. Kim JS, Kim EJ, Lee S, Tan X, Liu X, Park S, Kang K, Yoon JS, Ko YH, Kurie JM, Ahn YH. *miR-34a* and *miR-34b/c* have distinct effects on the suppression of lung adenocarcinomas. *Exp Mol Med* 2019;**51**: 1-10.

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* 2019;**39**.

23. Cao Z, Zhang G, Xie C, Zhou Y. *miR-34b* regulates cervical cancer cell proliferation and apoptosis. *Artif Cells Nanomed Biotechnol* 2019;**47**: 2042-7.

24. Maroof H, Salajegheh A, Smith RA, Lam AK. Role of microRNA-34 family in cancer with particular reference to cancer angiogenesis. *Exp Mol Pathol* 2014;**97**: 298-304.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. Corney DC, Flesken-Nikitin A, Godwin AK, Wang W, Nikitin AY. MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res* 2007;**67**: 8433-8.

26. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D, Jackson AL, Linsley PS, et al. A microRNA component of the p53 tumour suppressor network. *Nature* 2007;**447**: 1130-4.

27. Kumamoto K, Spillare EA, Fujita K, Horikawa I, Yamashita T, Appella E, Nagashima M, Takenoshita S, Yokota J, Harris CC. 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* 2008;**68**: 3193-203.

28. Wang LG, Ni Y, Su BH, Mu XR, Shen HC, Du JJ. MicroRNA-34b functions as a tumor suppressor and acts as a nodal point in the feedback loop with Met. *Int J Oncol* 2013;**42**: 957-62.

29. Wagner GP, Kin K, Lynch VJ. A model based criterion for gene expression calls using RNA-seq data. *Theory Biosci* 2013;**132**: 159-64.

30. Tomas Tokar CP, 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* 2018;**46**: D360-D70.

31. Vlachos IS, Paraskevopoulou MD, Karagkouni D, Georgakilas G, Vergoulis T, Kanellos I, Anastasopoulos IL, Maniou S, Karathanou K, Kalfakakou D, Fevgas A, Dalamagas T, et al. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res* 2015;**43**: D153-9.

32. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015;**4**.

33. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res* 2013;**41**: W169-73.

34. Zhao M, Kong L, Liu Y, Qu H. dbEMT: an epithelial-mesenchymal transition associated gene resource. *Sci Rep* 2015;**5**: 11459.

35. Bhatla N, Aoki D, Sharma DN, Sankaranarayanan R. Cancer of the cervix uteri. *Int J Gynaecol Obstet* 2018;**143 Suppl 2**: 22-36.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

36. Chae YK, Chang S, Ko T, Anker J, Agte S, Iams W, Choi WM, Lee K, Cruz M. Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). *Sci Rep* 2018;**8**: 2918.

37. Uzuka Y, Saito Y. [Combination therapy of chronic myelogenous leukemia with busulfan and bestatin]. *Gan To Kagaku Ryoho* 1991;**18**: 571-8.

38. Meyerson M, Harlow E. Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol Cell Biol* 1994;**14**: 2077-86.

39. Ying X, Sun Y, He P. Bone Morphogenetic Protein-7 Inhibits EMT-Associated Genes in Breast Cancer. *Cell Physiol Biochem* 2015;**37**: 1271-8.

40. Buijs JT, Henriquez NV, van Overveld PG, van der Horst G, Que I, Schwaninger R, Rentsch C, Ten Dijke P, Cleton-Jansen AM, Driouch K, Lidereau R, Bachelier R, et al. Bone morphogenetic protein 7 in the development and treatment of bone metastases from breast cancer. *Cancer Res* 2007;**67**: 8742-51.

41. Sun GY, Wu JX, Wu JS, Pan YT, Jin R. Caveolin-1, E-cadherin and beta-catenin in Gastric Carcinoma, Precancerous Tissues and Chronic Non-atrophic Gastritis. *Chin J Cancer Res* 2012;**24**: 23-8.

42. Lucy Cassar HL, 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* 2008;**68**: 9157–66.

43. Chen Y, Liu C, Xie B, Chen S, Zhuang Y, Zhang S. miR96 exerts an oncogenic role in the progression of cervical cancer by targeting *CAV1*. *Mol Med Rep* 2020;**22**: 543-50.

44. Razani B, Altschuler Y, Zhu L, Pestell RG, Mostov KE, Lisanti MP. 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* 2000;**39**: 13916-24.

45. Wiechen K, Diatchenko L, Agoulnik A, Scharff KM, Schober H, Arlt K, Zhumabayeva B, Siebert PD, Dietel M, Schafer R, Sers C. Caveolin-1 is down-regulated in human ovarian carcinoma and acts as a candidate tumor suppressor gene. *Am J Pathol* 2001;**159**: 1635-43.

46. 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* 2017;**40**: 21-30.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. Wen XF, Chen M, Wu Y, Chen MN, Glogowska A, Klonisch T, Zhang GJ. Inhibitor of DNA Binding 2 Inhibits Epithelial-Mesenchymal Transition via Up-Regulation of Notch3 in Breast Cancer. *Transl Oncol* 2018;**11**: 1259-70.

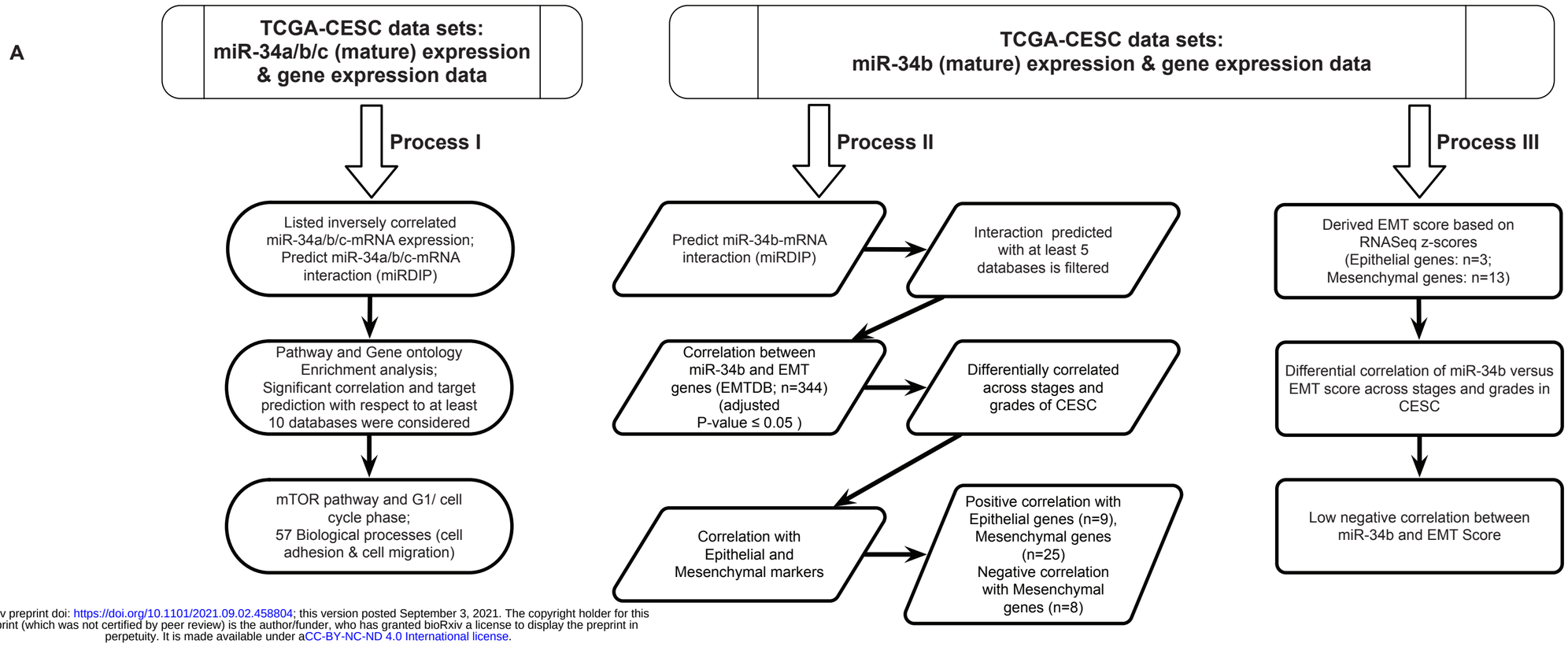
48. Tommaso Perini RS, Mehmet Kemal Samur, MariateresaFulciniti, Michael A Lopez, Matthew Lawlor, Christopher J. Ott, Na Li, Yan Xu, Kenneth Wen, Nicola Amodio, Eugenio Morelli, Kenneth Anderson, Fabio Ciceri, Nikhil Munshi. Inhibitor of DNA Binding 2 (*ID2*) Plays a Key Tumor Suppressor Role in Promoting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

49. Sumida T, Ishikawa A, Nakano H, Yamada T, Mori Y, Desprez PY. Targeting *ID2* expression triggers a more differentiated phenotype and reduces aggressiveness in human salivary gland cancer cells. *Genes Cells* 2016;**21**: 915-20.

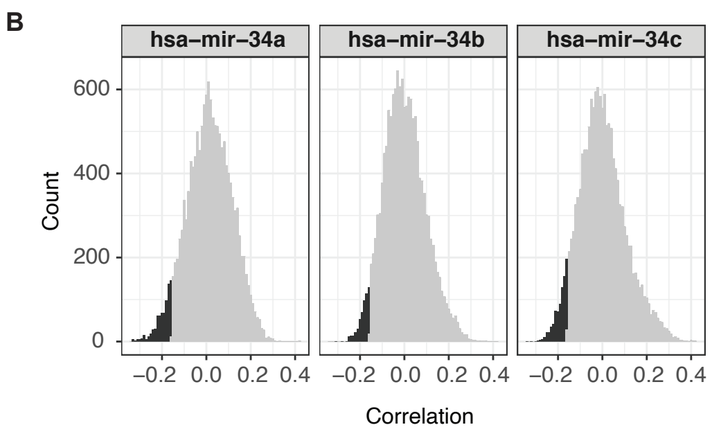
50. Chakraborty P, George JT, Tripathi S, Levine H, Jolly MK. Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. *Front Bioeng Biotechnol* 2020;**8**: 220.

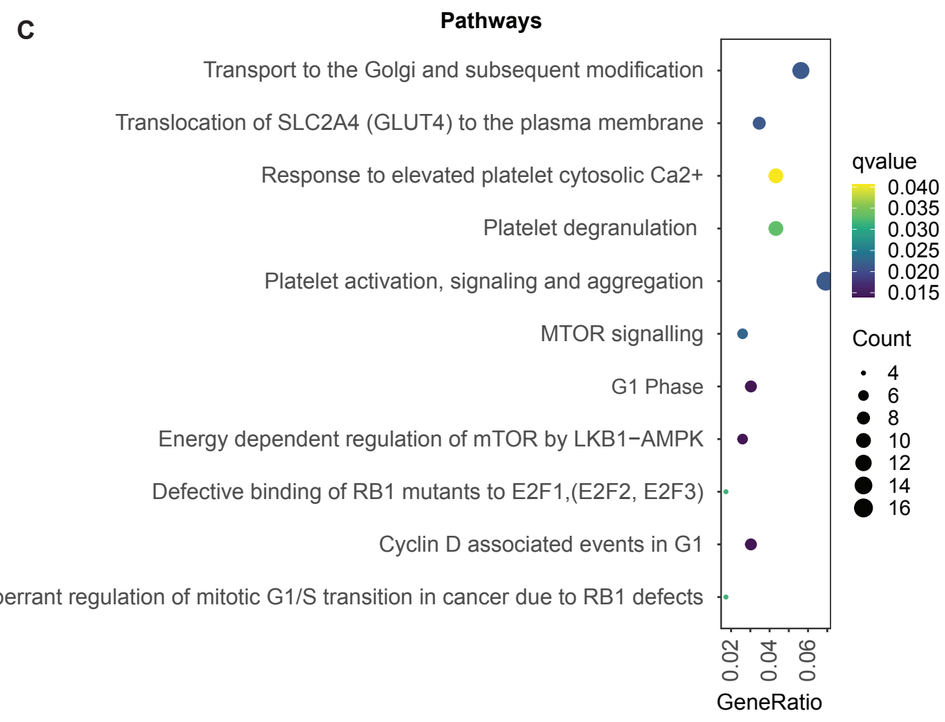
51. Sun R, Guan H, Liu W, Liang J, Wang F, Li C. Expression of *BMP7* in cervical cancer and inhibition of epithelial mesenchymal transition by *BMP7* knockdown in HeLa cells. *Int J Mol Med* 2020;**45**: 1417-24.

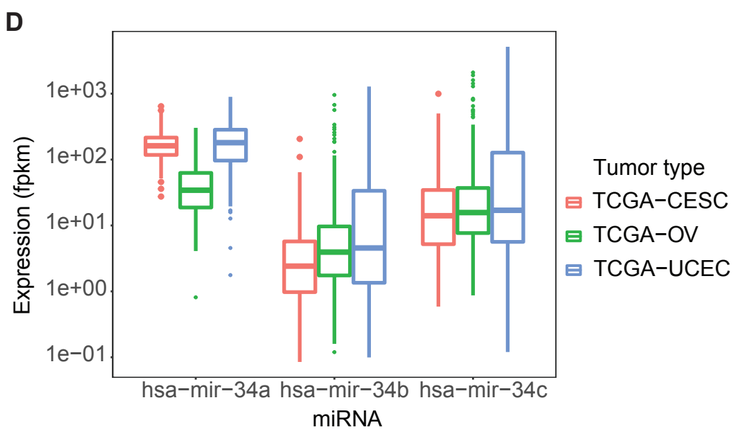
Figure 1

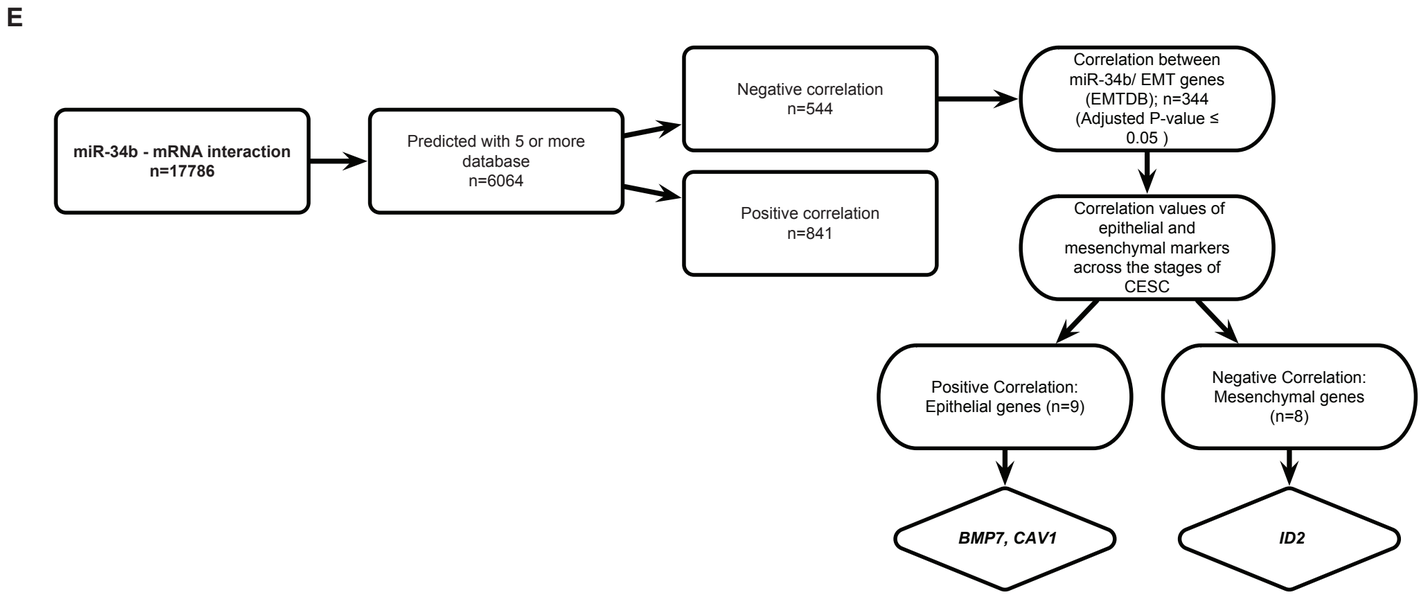


bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



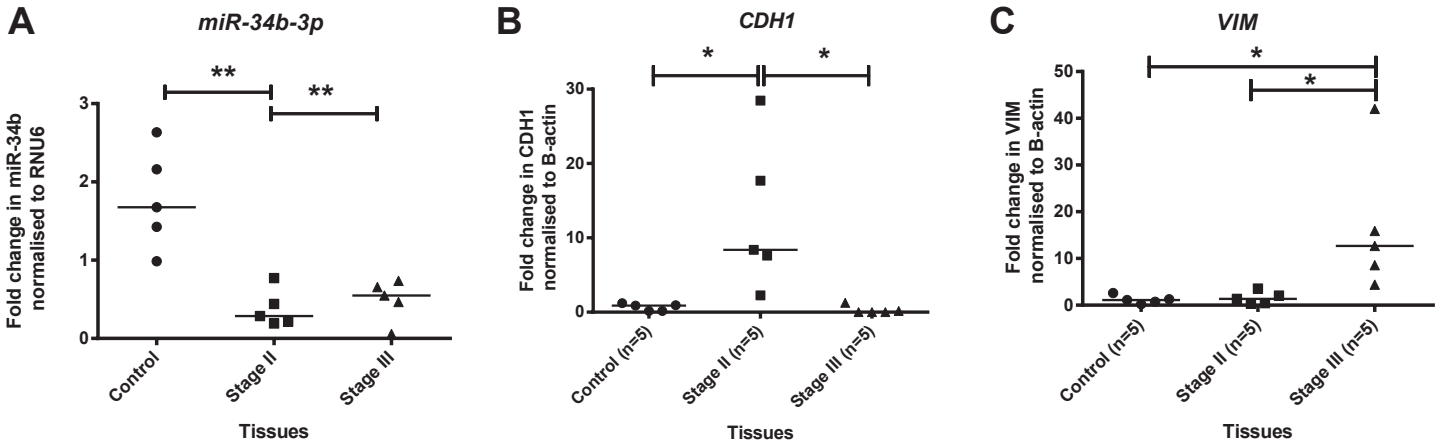


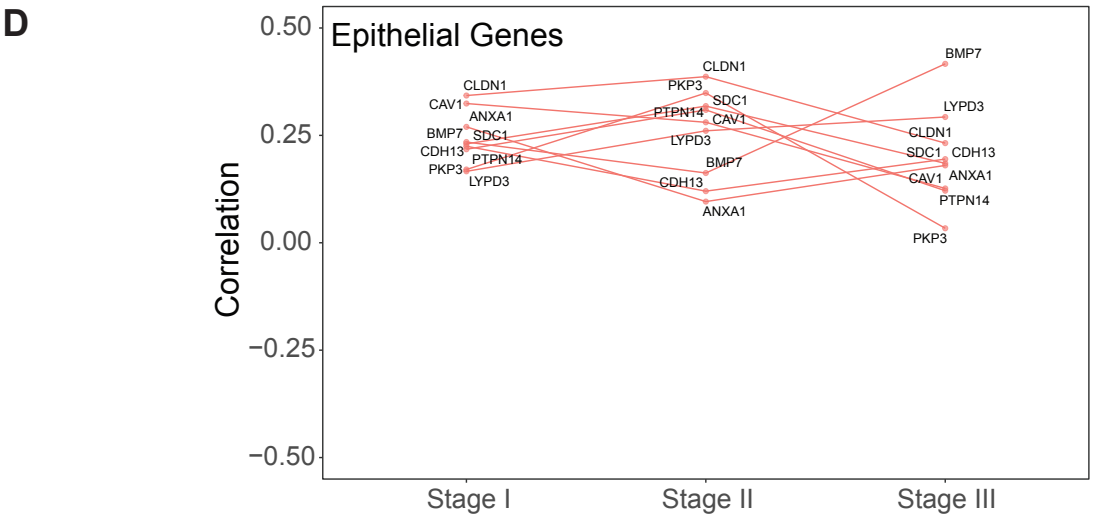




bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2





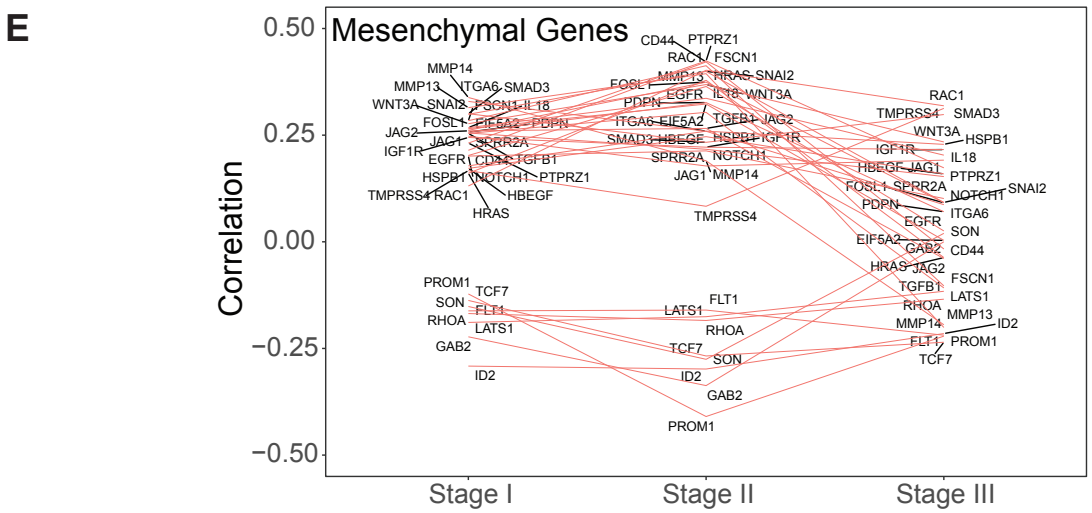
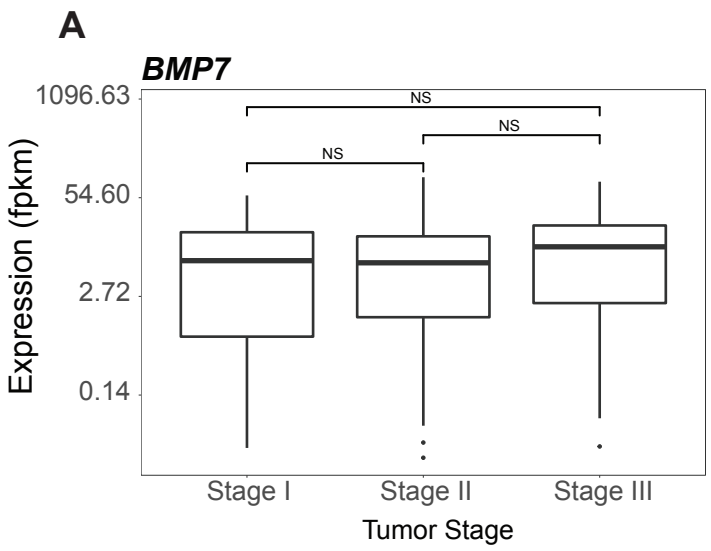
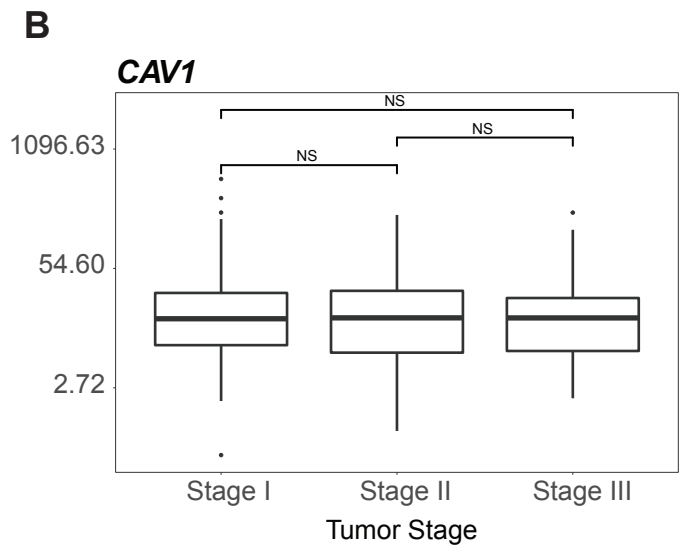
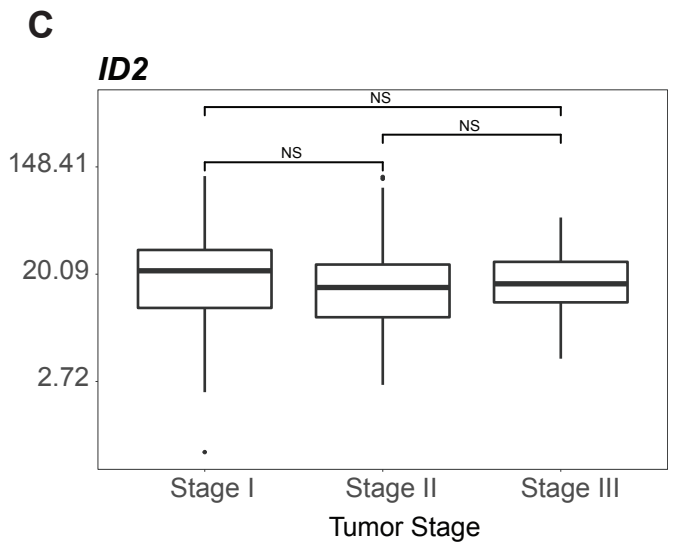
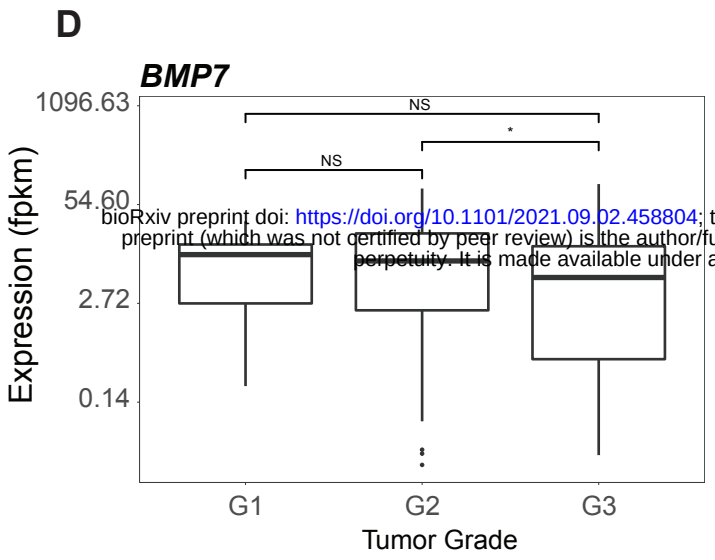


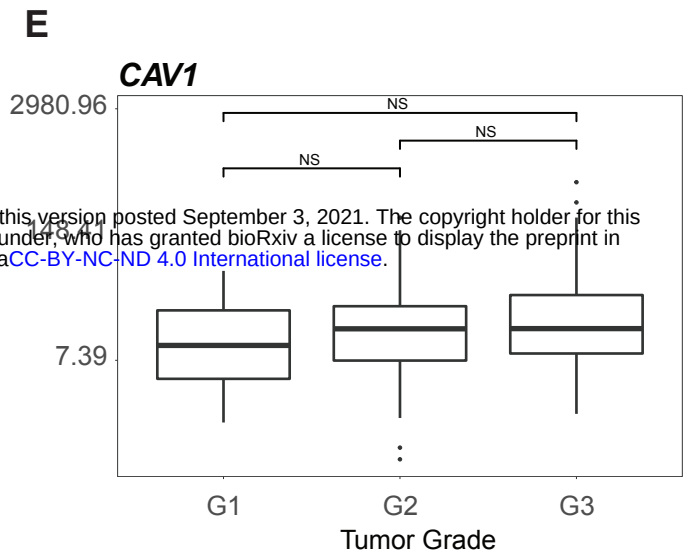
Figure 3

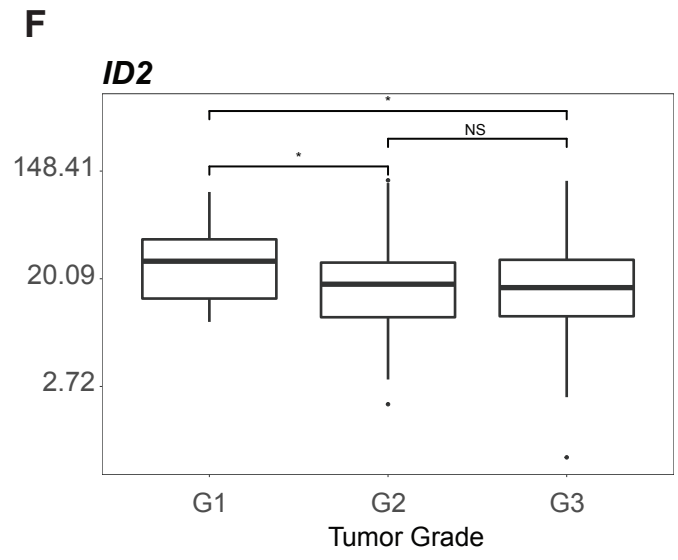




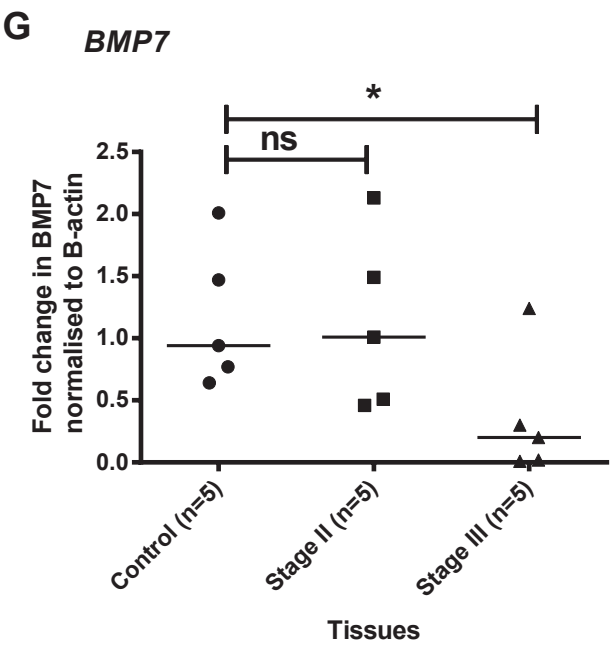


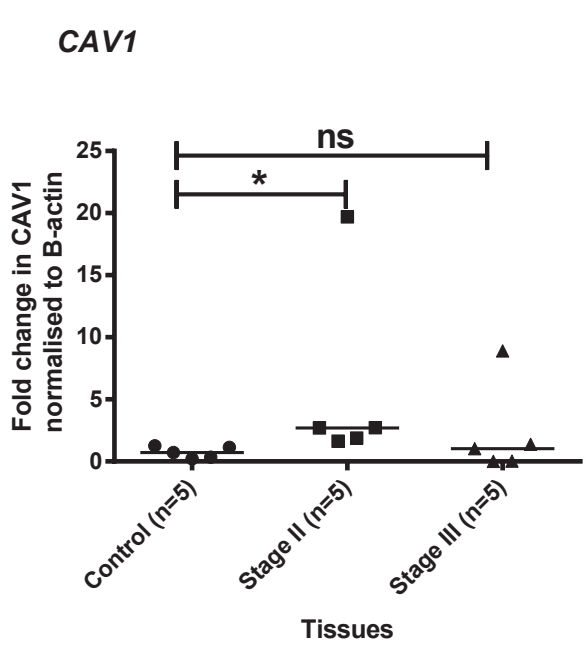


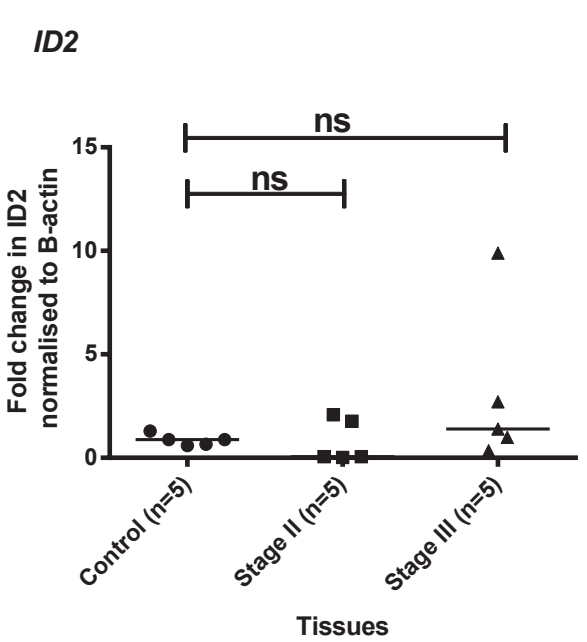


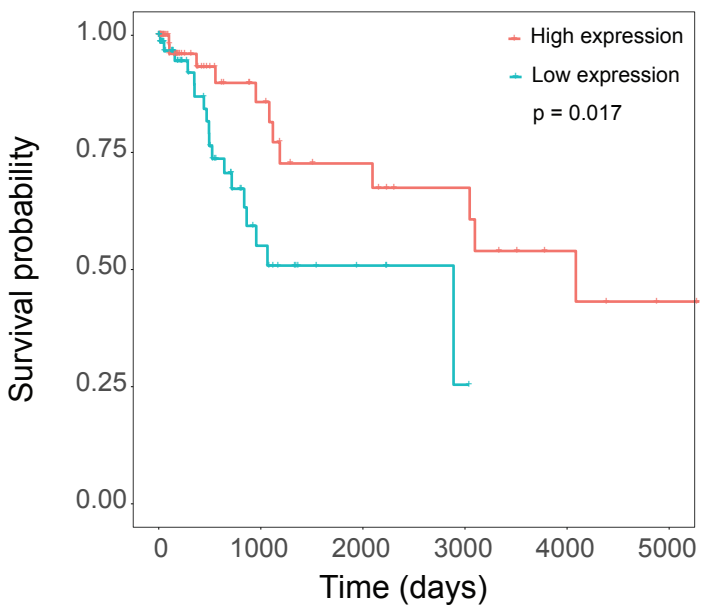


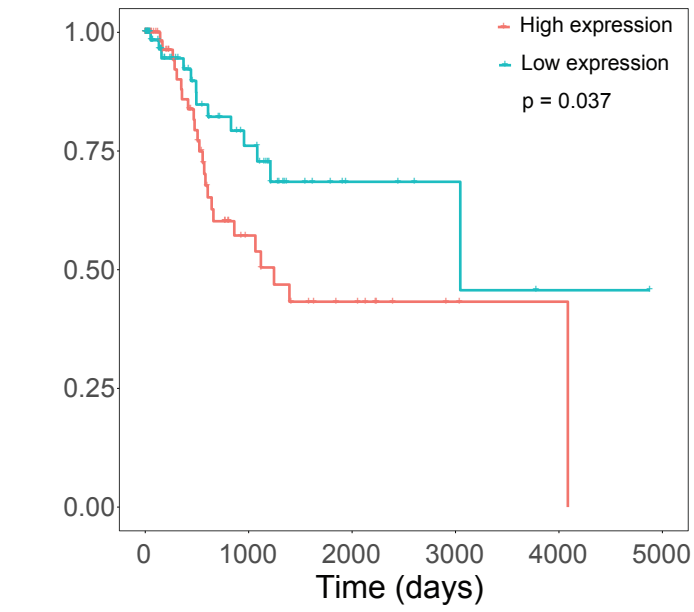
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.











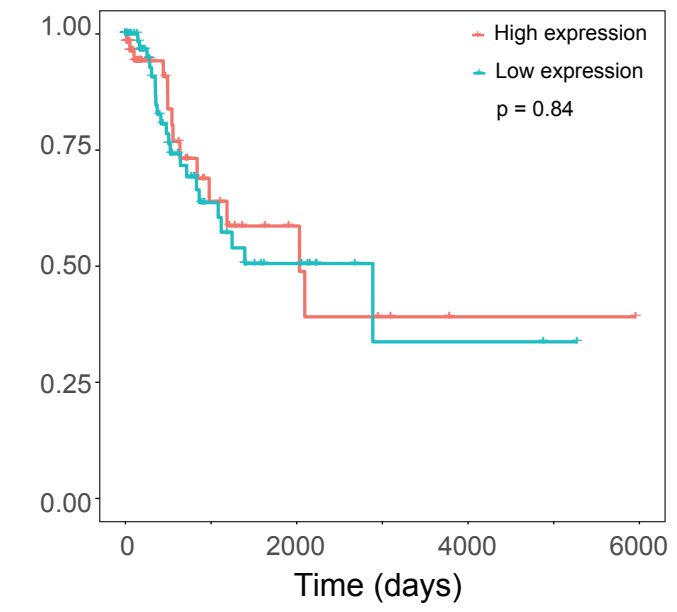
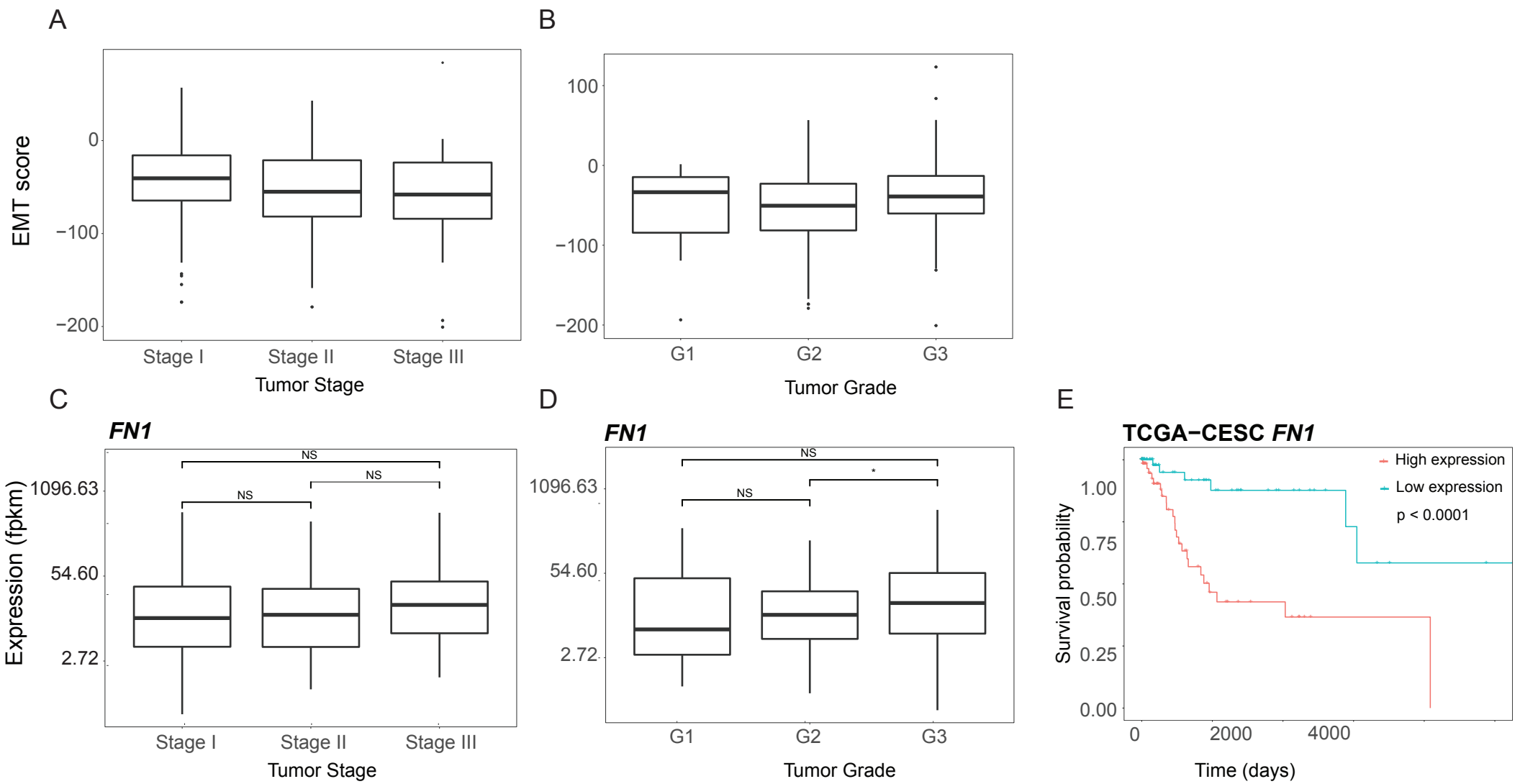


Figure 4



F

hsa-miR-34b/*FN1* alignment

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

mirSVR score: -0.9229
PhastCons score: 0.6314

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.02.458804; this version posted September 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

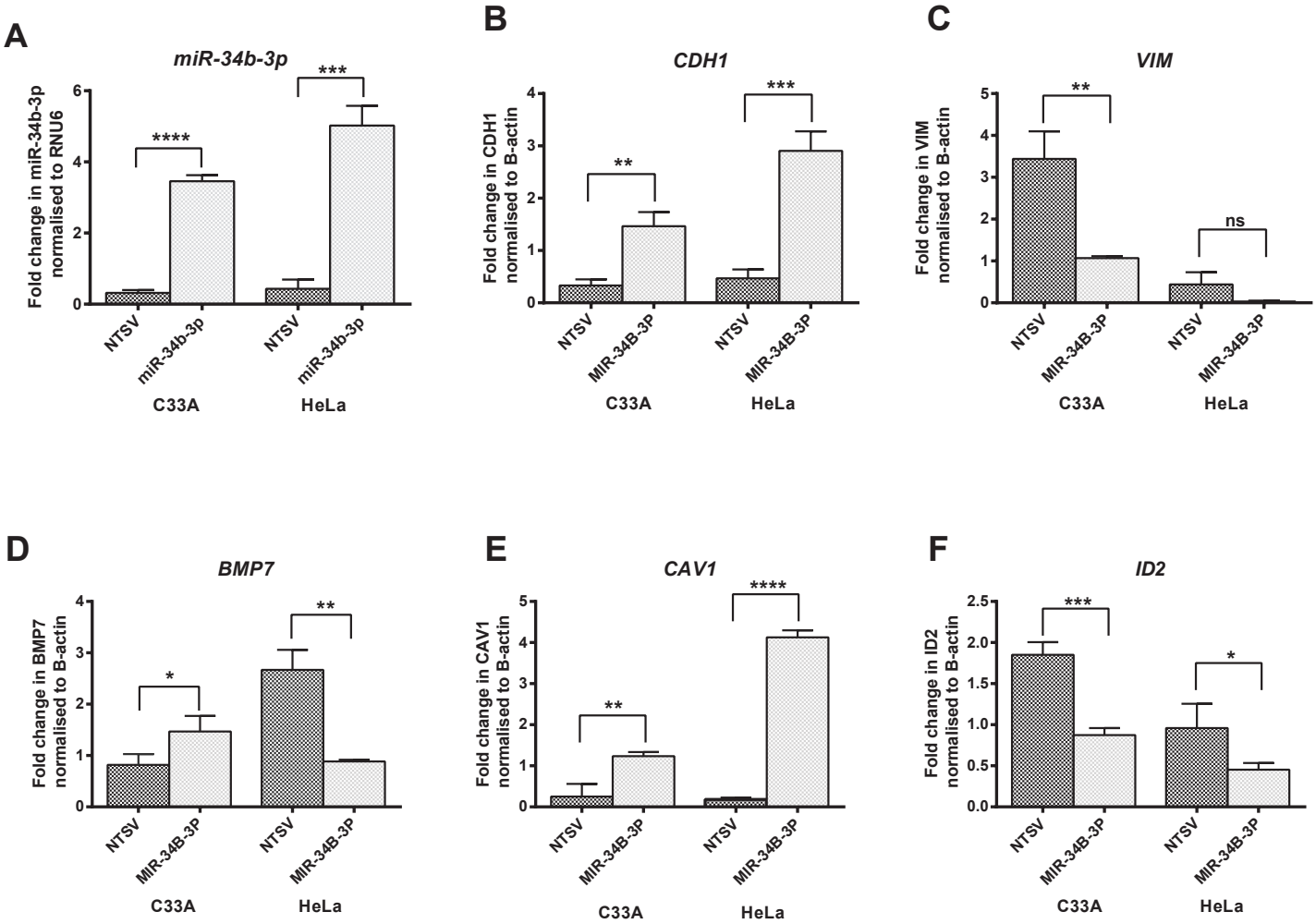



Figure 5