scientific reports

Check for updates

OPEN

Clustering of RNA co-expression network identifies novel long non-coding RNA biomarkers in squamous cell carcinoma

Liisa Nissinen[1,2], Josefiina Haalisto[1,2], Pilvi Riihilä[1,2], Minna Piipponen[1,2] & Veli-Matti Kähäri[1,2]✉

Long non-coding RNAs (lncRNAs) have emerged as important players in cancer progression. Cutaneous squamous cell carcinoma (cSCC) is the most common metastatic skin cancer with increasing incidence worldwide. The prognosis of the metastatic cSCC is poor, and currently there are no established biomarkers to predict metastasis risk or specific therapeutic targets for advanced or metastatic cSCC. To elucidate the role of lncRNAs in cSCC, RNA sequencing of patient derived cSCC cell lines and normal human epidermal keratinocytes was performed. The correlation analysis of differentially expressed lncRNAs and protein-coding genes revealed six distinct gene clusters with one of the upregulated clusters featuring genes associated with cell motility. Upregulation of the expression of lncRNAs linked to cSCC cell motility in cSCC and head and neck SCC (HNSCC) cells was confirmed using qRT-PCR. Elevated expression of *HOTTIP* and *LINC00543* was also noted in SCC tumors in vivo and was associated with poorer prognosis in HNSCC and lung SCC cohorts within TCGA data, respectively. Altogether, these findings uncover a novel set of lncRNAs implicated in cSCC cell locomotion. These lncRNAs may serve as potential novel biomarkers and as putative therapeutic targets for locally advanced and metastatic cSCC.

Keywords lncRNA, Matrix metalloproteinase, Squamous cell carcinoma, Cell motility, Invasion

Abbreviations

cSCC	Cutaneous squamous cell carcinoma
ESCA	Esophageal carcinoma
HNSCC	Head and neck squamous cell carcinoma
lncRNA	Long non-coding RNA
linc	Long intergenic non-coding
LUSCC	Lung squamous cell carcinoma
MMP	Matrix metalloproteinase
qRT-PCR	Quantitative real-time reverse transcriptase-PCR

Long non-protein coding RNAs (lncRNAs) have been recognized as crucial cellular regulators influencing tissue homeostasis, along with playing significant roles in pathological conditions such as cancer initiation, growth, and metastasis[1]. These lncRNAs, typically with a transcript length of 200 nucleotides or longer, form a diverse group of regulatory RNAs capable of interacting with DNA, proteins, or other RNAs[2]. Their precise spatial and temporal regulation lends them versatility as regulators of cellular processes in various cellular compartments, presenting them as interesting targets for therapeutic interventions in pathological conditions[3].

Keratinocyte-derived cutaneous squamous cell carcinoma (cSCC) is the most common metastatic skin cancer globally, with its incidence on the rise due to lifestyle shifts and aging populations[4]. The metastasis rate of primary cSCC is approximately 3–5%, contributing to at least 20% of skin cancer-related deaths[5,6]. Cumulative exposure to UV radiation stands as a primary risk factor for cSCC, attributed to its high mutation rate[7]. Mutations in Tumor protein 53 (*TP53*) gene occur early in keratinocyte carcinogenesis[8–10], followed by subsequent genomic

[1]Department of Dermatology, University of Turku and Turku University Hospital, Hämeentie 11 TE6, FI-20520 Turku, Finland. [2]FICAN West Cancer Centre Research Laboratory, University of Turku and Turku University Hospital, FI-20520 Turku, Finland. ✉email: veli-matti.kahari@utu.fi

alterations in oncogenes like *NOTCH1*, *HRAS*, *CDKN2A*, and *EGFR* that further drive cSCC progression[11,12]. Since epidermal keratinocytes in chronically sun-exposed normal skin already carry several mutations in genes linked to cSCC progression, changes in the tumor microenvironment play a critical role in the initiation and progression of cSCC[13,14].

While the role of lncRNAs in cSCC has been subject to extensive investigation, their specific impact on cSCC progression remains largely unexplored[15]. We have previously characterized and named three lncRNAs that are upregulated in cSCC and regulate its progression through different mechanisms: *PICSAR* (p38 inhibited cutaneous squamous cell carcinoma-associated lincRNA)[16,17], *PRECSIT* (p53 regulated carcinoma-associated STAT3-activating long intergenic non-protein coding transcript)[18], and *SERLOC* (super enhancer and ERK1/2-regulated long intergenic non-protein coding transcript overexpressed in carcinomas)[19]. Additionally, *PVT1* was identified as an lncRNA that promotes cSCC progression by suppressing cellular senescence, inhibiting *CDKN1A* expression, and preventing cell cycle arrest[20,21].

In this study, to elucidate the role of lncRNAs in cSCC, RNA sequencing (RNA-seq) was performed on patient-derived cSCC cell lines and normal human epidermal keratinocytes (NHEK). Correlation analysis of differentially expressed lncRNAs and protein-coding genes identified six clusters. Cluster 1 was upregulated in cSCC cells and contained genes related to cell motility. Further investigation of the lncRNAs in this cSCC cell motility cluster revealed that they were also upregulated in patient-derived head and neck SCC (HNSCC) cells. Notably, higher expression of *HOTTIP* and *LINC00543* was associated with worse prognosis in head and neck SCC (HNSCC) and lung SCC (LUSCC) cohorts in the TCGA data, respectively. These findings uncover a novel set of lncRNAs involved in cSCC cell locomotion, which may serve as biomarkers and potential therapeutic targets in advanced cSCC.

Materials and methods

Ethical issues

The use of tumor-derived SCC cell lines and NHEKs, as well as the collection of cSCC tissues, was approved by the Ethics Committee of the Hospital District of Southwest Finland (187/2006, 138/2007). All participants provided written informed consent, and the study was conducted with the permission of Turku University Hospital, in accordance with the Declaration of Helsinki.

Cell culture

NHEKs (n = 4) were isolated from the skin of healthy individuals undergoing mammoplasty[16]. NHEK-PC, NHEK (adult pooled), NHEK (adult single donor), HEK, NHEK were from PromoCell (Heidelberg, Germany), while HEK and human oral keratinocytes (HOK) were acquired from ScienCell Research Laboratories (Carlsbad, CA, USA). Primary non-metastatic (UT-SCC12A, UT-SCC91, UT-SCC105, UT-SCC111, and UT-SCC118) and metastatic (UT-SCC7, UT-SCC59A, and UT-SCC115) cSCC cell lines were established from surgically removed cSCCs at Turku University Hospital[22,23]. The authenticity of cSCC cell lines was verified by short tandem repeat profiling as described previously[22]. HNSCC cell lines (n = 29) were established from surgically removed HNSCCs at Turku University Hospital[24,25]. The immortalized non-tumorigenic human keratinocyte–derived cell line (HaCaT) and three Ha-*ras*-transformed tumorigenic HaCaT cell lines (A5, II-4, and RT3) were kindly provided by Dr Norbert Fusenig (Deutsche Krebsforschungszentrum, Heidelberg, Germany). A5 cells form benign tumors, II-4 cells invasive tumors and RT3 cells metastatic tumors in nude mice in vivo[26]. Cells were cultured as described previously[18]. For growth factor treatment, cSCC cell lines (n = 2–8) were maintained in serum-free medium for 24 h and then treated with transforming growth factor-β1 (5 ng/mL; Sigma Aldrich, St Louis, MO) for 24 h.

Tissue samples

Normal human skin samples (n = 6) were obtained from the upper arm of healthy volunteers or during mammoplasty operation at Turku University Hospital. Primary cSCC samples (n = 6) were collected from surgically removed tumors in Turku University Hospital[16].

RNA sequencing

RNA was isolated from cSCC cell lines (n = 8) and NHEKs (n = 4) using miRNeasy Mini Kit (Qiagen), and the RNA-seq analysis was performed using Illumina HiSeq2500 system using paired-end sequencing chemistry with 100 bp read length (Illumina, San Diego, CA) at the Finnish Functional Genomics Centre, University of Turku, Åbo Akademi University and Biocenter Finland. The number of reads was ~ 60 M or above for all samples (Supplementary Table S1). Fastq read files from the same sample were first merged in order to obtain one fastq file pair per sample. Before pooling the reads, the quality of all read files were inspected using FastQC software v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)[27] to ensure consistent quality of the reads (Supplementary Fig. S1). The reads were aligned against the human GRCh38 (hg38) genome assembly with Homo_sapiens.GRCh38.88.gtf transcript annotation file using STAR alignerr, version 2.6.1 (http://code.google.com/p/rna-star/)[28] (Supplementary Fig. S2, Supplementary Table S2; GeneVia Technologies, Tampere, Finland). Gene-level read counts were obtained simultaneously with the alignment process. Read count data was filtered to exclude genes with very low read counts (less than ten reads in total in all samples). The obtained read counts were normalized using regularized log transformation function of DESeq2 R package, v. 1.20.0 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html)[29], which transforms the count data to the log2 scale in a way that minimizes differences between samples for rows with small counts and also normalizes the data with respect to library size. DESeq2 R package, v. 1.20.0[29] uses raw read counts as the input and performs internal normalization using the median of ratios method. Differential expression (DE) analysis was performed using DESeq2, v. 1.20.0. *P*-values were adjusted using the Benjamini–Hochberg multiple testing adjustment method[30].

Genes with absolute log2 fold change > 1 and adjusted $p < 0.05$ were considered as significantly differentially expressed. Differentially expressed genes were annotated with HGNC gene symbols, gene description and gene biotype using biomaRt, v. 2.36.1 (http://www.bioconductor.org/biocLite.R)[31]. The results were split into differentially expressed protein-coding genes and lncRNAs based on the gene biotype annotations, using Ensembl biotype definitions as guidance. RNA sequencing data (accession number GSE252347) have been deposited in the public database GEO (Gene Expression Omnibus, NCBI; http://www.ncbi.nlm.nih.gov/geo/).

Correlation and clustering analysis

A combined dataset of significantly differentially expressed lncRNAs and protein coding genes was generated and pairwise Pearson's correlations of the expression profiles of each gene were computed (GeneVia Technologies). Correlation distances (1—correlation coefficient) were used as input for k-means clustering with 6 centres, generating clusters of co-expressed genes, i.e. genes with similar expression profiles across all samples. Correlation distance matrix was subjected to classical multidimensional scaling (MDS), or principal coordinates analysis, to be able to visualize the clustering result in a two-dimensional plot generated using ggplot2 1.0 (https://ggplot2.tidyverse.org/articles/ggplot2.html).[32]. Correlation densities within each of the clusters were visualized as histograms using basic R functions and custom R scripts (http://www.R-project.org/). Distributions of protein coding gene and lncRNA expression values in each cluster were examined using box plots generated using ggplot2 1.0 (https://ggplot2.tidyverse.org/articles/ggplot2.html)[32]. Gene expression values of all six clusters were also visualized as a heatmap using R package pheatmap[33].

Functional analysis

Enrichment of Gene ontology (GO)[34,35] terms in the genes in k-means clusters was studied using R package clusterProfiler, v.3.8.1 (http://bioconductor.org/ packages /release /bioc/ html/clusterProfiler.html[36]. The *p*-values of enrichment analysis were corrected for multiple testing using Benjamini–Hochberg multiple testing adjustment procedure[30]. For the cluster genes, the enrichment analysis was conducted using both lncRNAs and mRNAs.

Ingenuity Pathway Analysis (IPA) (QIAGEN, Germantown, MD; QIAGEN Inc., https://www.QIAGENbioinformatics.com/products)[37] was performed separately for the k-means gene cluster including both lncRNAs and mRNAs (GeneVia Technologies). As before, genes with absolute log2 fold change > 1 and adjusted $p < 0.05$ were considered significantly differentially expressed and were included in the analyses. *P*-values were corrected for multiple testing using Benjamini–Hochberg method[30]. IPA also calculated a z-score for each pathway, indicating predicted pathway activation if the z-score > 2 or inhibition if z-score < -2.

DAVID Bioinformatics Resources (https://david.ncifcrf.gov/tools.jsp) was used to study the genes in k-means clusters[38,39]. It calculates the probability that particular GO annotations are overrepresented in a given gene list using a Fisher exact probability test. Molecular function and biological process GO terms with a *p* value < 0.05 containing at least three genes were considered significant.

Transcription factor annotation

Experimentally validated transcription factors (TFs) from Transcriptional Regulatory Element Database (TRED)[40] via RegNetwork[41], Encyclopedia of DNA Elements (ENCODE) and ChIP-X Enrichment Analysis (CHEA) database[42] were collected (GeneVia Technologies). The data from the three sources were combined and redundant entries removed to generate a list of unique TF-TG pairs consisting of 376 TFs and 23,430 TF targets. For each TG, a list of potential TFs was then extracted, and the list was used for annotating DE genes with their potential upstream TFs. To visualize the enriched TFs in cluster 1 https://www.wordclouds.com was used.

Real-time quantitative PCR

Total RNA was extracted from cultured NHEKs and cSCC cells using RNeasy mini kit (Qiagen, Germantown, MD, USA), and 1 µg of total RNA was reverse transcribed into cDNA with random hexamer and M-MLV Reverse Transcriptase H Minus (both from Promega, Madison, WI, USA) for real-time quantitative reverse transcriptase-PCR (qRT-PCR) analysis. Primers and probe for *LINC01361* (Hs03084701_cn, Cat Nro 4400291), *LINC01558* (Hs00205026_m1, Cat Nro 4351372), *LINC01460* (Hs03405179_cn, Cat Nro 4400291), *LINC00543* (Hs00905601_s1, Cat Nro 4426961), *LINC00702* (Hs05184596_cn, Cat Nro 4400291) and *HOTTIP* (Hs04965481_cn, Cat Nro 4400291) were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Primers and probes for *MMP13* and β-actin (*ACTB*) were designed as previously described[25]. qRT-PCR reactions were performed using the QuantStudio 12 K Flex (Thermo Fisher Scientific) at the Finnish Functional Genomics Centre in Turku, Finland. qRT-PCR amplification was executed with the following protocol: hold stage for 2 min at 50 °C, 10 min at 95 °C, and PCR stage for 40 cycles of 0.15 min at 95 °C and 1 min at 60 °C. Samples were analyzed using the standard curve method in 2–3 parallel reactions with threshold cycle values < 5% of the mean threshold cycle.

GEO dataset analyses

Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) was utilized to investigate the expression of LINC RNAs in cSCCs in vivo. RDEBSCC gene expression profile dataset GSE111582[8] was downloaded from the publicly available GEO database and used to analyze the expression of *LINC00702*, *LINC01361* and as a control gene *MMP13* in RDEBSCC (n = 8) and non-SCC (n = 10) samples investigated in our study. Another dataset we utilized was GSE45216[43]. This dataset included samples from immunosuppressed patients (cSCC, n = 21, AK n = 7) and sporadic cSCC (n = 9) and AK (n = 3). The GEO2R online tool was employed to analyze the gene expression profile.

Gene expression profiling interactive analysis

The online Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/) analysis tool was utilized to analyze the expression of *HOTTIP* and *LINC00543* in SCCs in the database and the relationship between *HOTTIP* and *LINC00543* expression and prognosis of HNSCC, LUSCC, and esophageal carcinoma (ESCA) in The Cancer Genome Atlas (TCGA) data[44,45].

Statistical analysis

The R package Limma was used for statistical analysis of RNA-seq data. Statistical analyses of qRT-PCR results were performed using SPSS Statistics software for Windows, version 27 (IBM, Armonk, NY). To determine the significance of differences between two sample groups, the *U*-test was used. TCGA data was analyzed using the GEPIA tool[46].

Declaration of AI use

ChatGPT-4 (https://chat.chatbotapp.ai/) was used to revise the language of some paragraphs during the final round of editing the revised manuscript.

Results

The expression profiles of lncRNAs and protein-coding genes in cSCC cells and NHEKs

To determine differentially expressed lncRNAs and protein-coding genes, the RNA-seq was performed on primary non-metastatic (n = 5) and metastatic (n = 3) cSCC cell lines, as well as NHEKs. Genes were considered significantly differentially expressed if the adjusted *p*-value (FDR) was less than 0.05 and the absolute log2 fold change in expression level was greater than 1. Genes in two categories—lncRNAs or protein-coding genes—were further characterized. The number of differentially expressed lncRNA genes was 723, while the number of differentially expressed protein-coding genes was 1667. To illustrate the distribution of differentially expressed lncRNAs (Supplementary Fig. S3A) and protein coding genes (Supplementary Fig. S3B) volcano blots were constructed. The top 50 most significantly differentially expressed lncRNAs and protein coding genes are presented in a heatmap (Fig. 1A). As previously shown in cSCC, *LINC00520*[47] was among the top downregulated lncRNAs (Fig. 1A) and *PRECSIT* (*LINC00346*)[18] was among top upregulated lncRNAs (Fig. 1A). Additionally, inflammasome component *AIM2*[48] was among top upregulated mRNAs (Fig. 1A).

Regarding lncRNA expression with more than an eightfold change (log2), there were 11 upregulated genes, 5 of which were novel transcripts (Fig. 1B, Supplementary Table S3). One antisense lncRNA, *ZNF667-AS1* (ZNF667 antisense RNA 1) was > 10 FC (log2) downregulated (Fig. 1B, Supplementary Table S3). *LINC01361*, *HOTTIP,* and *LINC01558* were among the upregulated lncRNAs (FC (log2) > 5 and > 4, respectively) (Fig. 1B, Supplementary Table S3). The majority of the differentially expressed genes belonged to the 2–3 FC (log2) group, including *LINC00543*, *LINC00702* and *LINC01460* (Fig. 1B, Supplementary Table S3). Similarly, for differentially expressed protein coding genes, 5 upregulated genes (e.g. *MMP13*) and 4 downregulated genes were identified with > 10 FC (log2), and the group with most differentially expressed protein coding genes exhibited a 2–3 FC (log2) (Fig. 1B, Supplementary Table S4).

Classification of differentially expressed lncRNA biotypes

Altogether 568 lncRNAs were significantly upregulated (adjusted *p*-value < 0.05, FC (log2) 1) in cSCC cells. The majority of these were either antisense RNAs (48%) or lincRNAs (41%) (Fig. 1C). In contrast, 155 lncRNAs were significantly downregulated (adjusted *p*-value < 0.05, FC (log2) 1) in cSCC cells, with the vast majority being antisense RNAs (55%) (Fig. 1C). Additionally, 35% of the significantly downregulated lncRNAs were lincRNAs (Fig. 1C).

Co-expression networks of the lncRNAs and protein-coding genes

The differentially expressed protein-coding genes and lncRNAs (adjusted *p*-value < 0.05, FC (log2) 1) were combined and pairwise Pearson's correlation coefficients were computed for their expression values across the samples (Fig. 2). k-means clustering, using six centers, was performed based on the correlation distances. The clustering of the genes is visualized as an MDS plot with different colors representing the different clusters (Fig. 2A). Pairwise Pearson's correlations of the expression values across the samples within each cluster were calculated to determine the expression profile similarity among the cluster genes, and the correlation distributions were visualized using histograms (Fig. 2B). Average correlations for the clusters were also calculated (Supplementary Table S5). The within-cluster correlations showed that Cluster 6 had the highest average correlation and the narrowest distribution of correlation values (Supplementary Table S5). In contrast, Cluster 5 exhibited a bimodal distribution of correlation values. Additionally, in the MDS plot, the genes in Cluster 5 did not form a clearly defined group, suggesting that these genes had variable expression profiles that lacked clear correlation with the expression profiles of the other genes (Fig. 2A,B Supplementary Table S5).

Genes in Cluster 1 and 2 are upregulated in cSCC cells

The gene expression profiles of the clusters were visualized by generating a heatmap of the expression values of all clustered genes across the samples (Fig. 2C). Genes in Clusters 1 and 2 were upregulated in cSCC cells, whereas genes in Clusters 3, 4, 5 and 6 were downregulated compared to NHEKs (Fig. 2C). To further inspect the expression values of the cluster genes, box plot visualizations were generated. These box plots display the expression value distributions in cSCC cells and NHEKs separately for the protein-coding genes and lncRNAs within the same cluster (Fig. 2D).

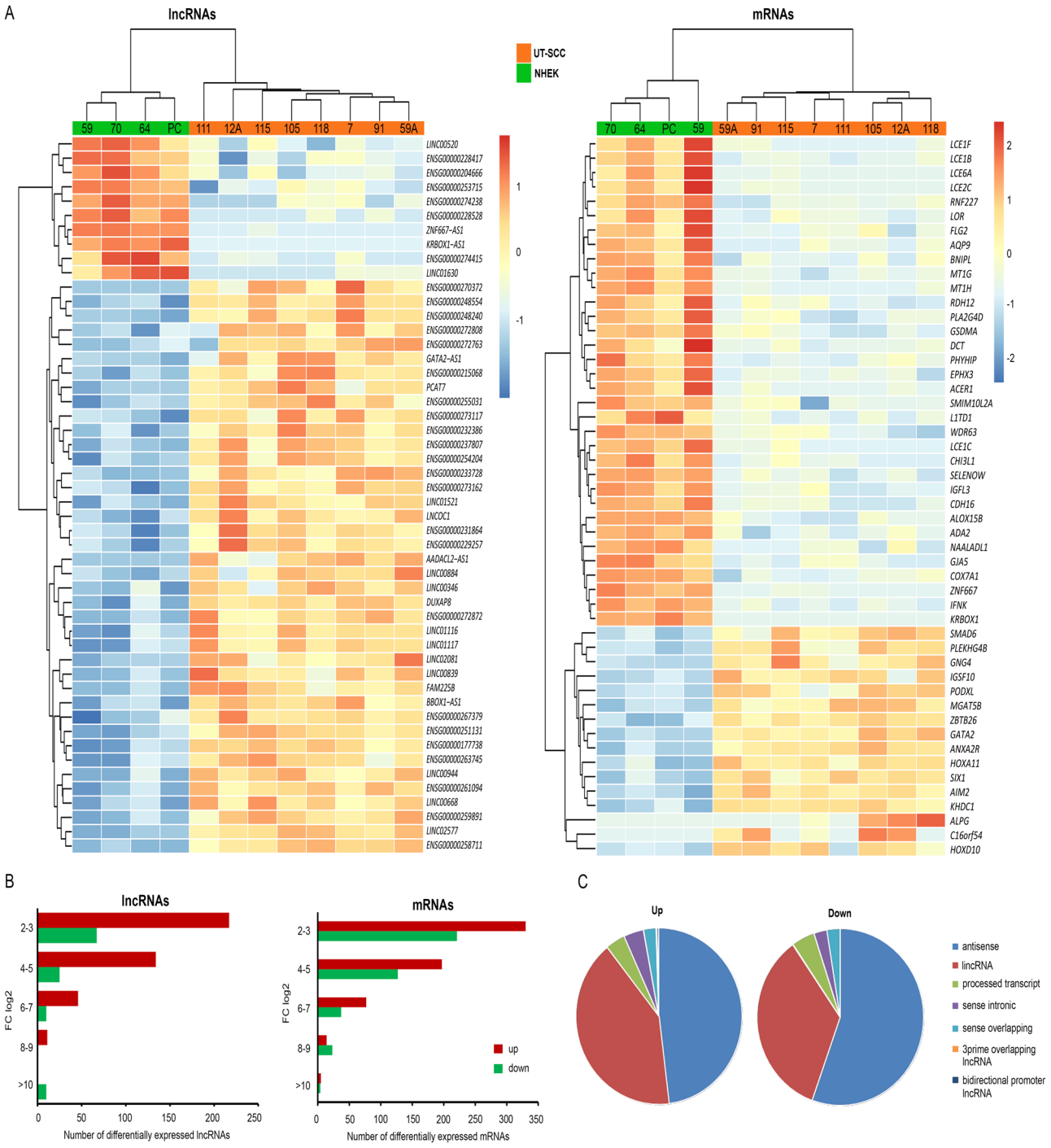

Figure 1. Profile of the differentially expressed lncRNAs and mRNAs in cSCC cells compared to NHEKs. RNA-seq was performed on patient-derived cSCC cell lines (n = 8) and normal human epidermal keratinocytes (NHEK, n = 4). (**A**) Heatmaps and the hierarchical clustering of the top differentially expressed lncRNAs and mRNAs in cSCC cells compared to NHEKs. (**B**) The number of differentially expressed lncRNAs and mRNAs based on fold changes (FC log2). (**C**) Biotype classification of upregulated and downregulated lncRNAs. All biotypes identified in the data are shown.

Genes in Cluster 1 are involved in cSCC cell motility

The upregulated Cluster 1 (Fig. 3) was studied in more detail. Pathway analysis of the genes associated with Cluster 1 revealed significant upregulation of the IPA biofunction categories *Cell movement*, *Migration of cells* and *Invasion of cells* (Fig. 3A). This association was also supported by GO analysis. GO terms related to Biological process associated with Cluster 1 genes included *Cell migration* and *Locomotion* (Fig. 3B). GO terms related to Cellular component associated with Cluster 1 genes included *Cell periphery* and *Plasma membrane* (Fig. 3C). David analysis of Cluster 1 genes showed the association of the protein coding genes with terms such as

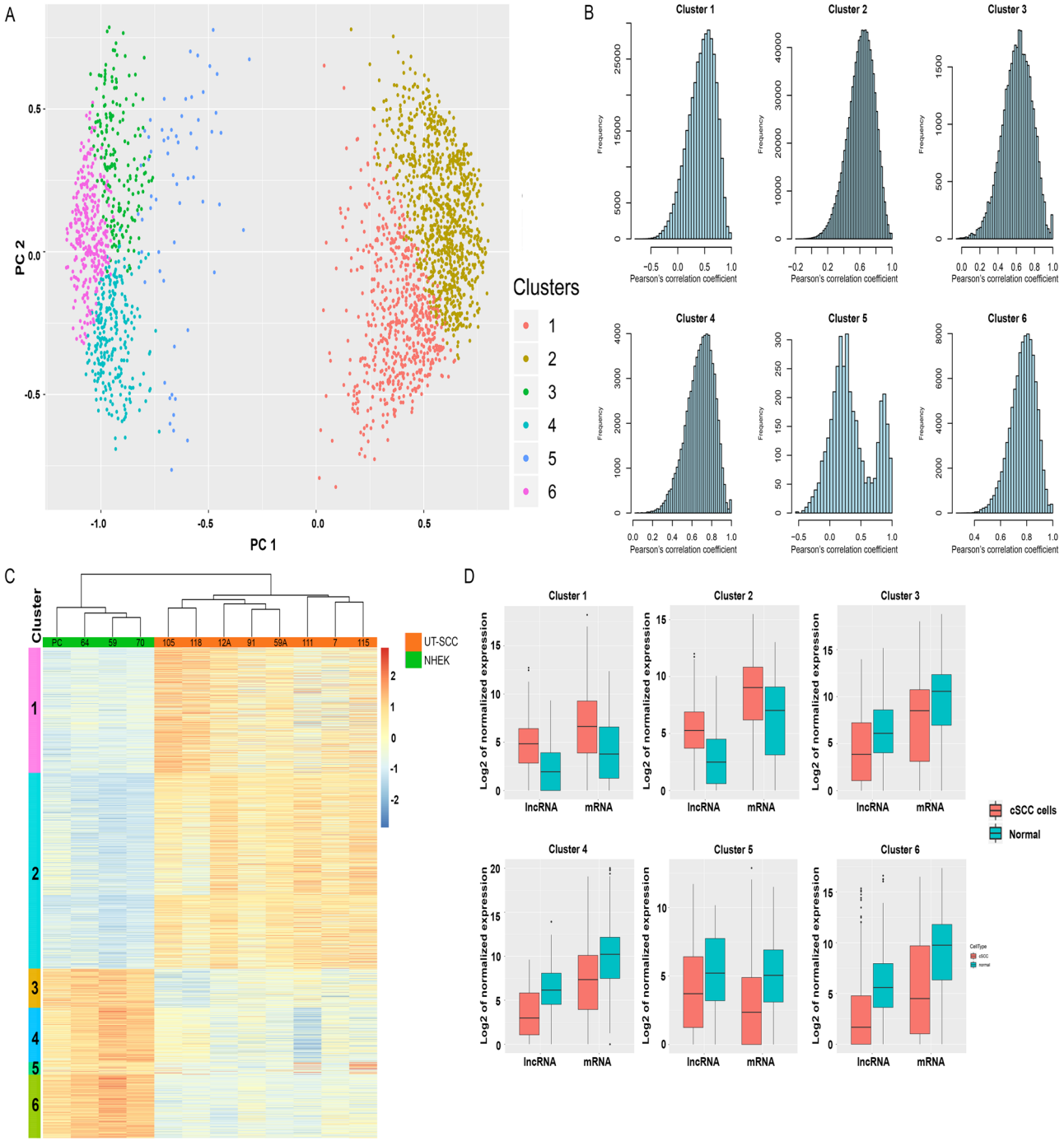

Figure 2. Construction of the co-expression network of lncRNAs and mRNAs. Differentially expressed lncRNAs and mRNA (adjusted p-value < 0.05, FC (log2) 1) in cSCC cells compared to NHEKs were joined and pairwise Pearson's correlation coefficients were computed using their expression values. k-means clustering with six centers was performed based on the correlation distances. (**A**) Visualization of gene clustering in a multidimensional scaling (MDS) plot. Genes are represented as dots with their principal coordinates based on pairwise correlation distances colored by clusters. (**B**) Calculation of pairwise Pearson's correlations within each cluster to assess expression profile similarities among the clustered genes. Correlation distributions were visualized using histograms. (**C**) Visualization of gene expression profiles in a heatmap displaying expression values of all clustered genes across the samples. (**D**) Box plot presentation illustrating the distribution of expression values in cSCC cell lines (n = 8) and NHEKs (normal, n = 4) for lncRNAs and mRNAs within one cluster.

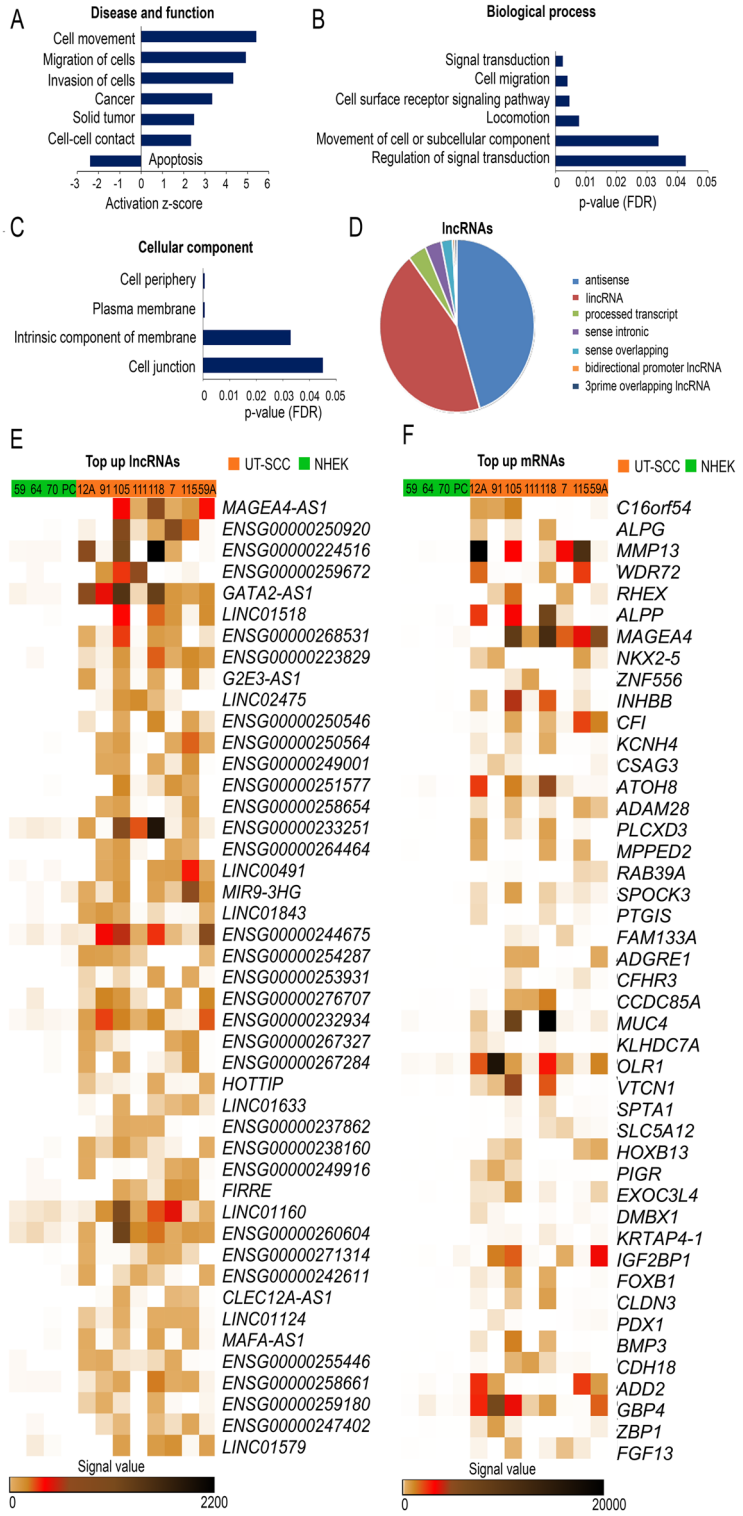

Figure 3. Genes within Cluster 1 genes play a role in cSCC cell motility. Summary of (**A**) significantly regulated IPA biofunctions, (**B**) the Gene ontology (GO) Biological process and (**C**) GO Cellular component associated with genes in Cluster 1. (**D**) Biotype classification of lncRNAs within Cluster 1 displaying all biotypes discovered in the data. Heatmap visualization of the top upregulated (**E**) lncRNAs and (**F**) mRNAs in cSCC cell lines (n = 8) compared to NHEKs (n = 4).

Endopeptidase activity, *Cell migration* and *Extracellular matrix disassembly* (Supplementary Table S6). A network of the Cluster 1 genes related to the term *Extracellular space* (Supplementary Fig. S4) was created by integrating protein–protein functional interaction data from the STRING database (https://string-db.org/, August 16, 2023)[49]. Interestingly, the previously known cSCC cell invasion related matrix metalloproteinase, MMP-13, was associated with this term (Supplementary Fig. S4)[50].

Classification of differentially expressed lncRNA biotypes in Cluster 1

LncRNA biotypes in Cluster 1 were studied in more detail. Altogether, there were 208 lncRNAs in Cluster 1, the majority of which were antisense or lincRNAs (45% and 44%, respectively) (Fig. 3D). Among the top regulated lncRNAs in Cluster 1, there were several novel transcripts and previously uncharacterized lincRNAs, as well as some previously characterized ones, such as *HOTTIP*[51] and *LINC00491*[52], which have been shown to be involved in cancer cell invasion (Fig. 3E) (Supplementary Table S7). Interestingly, Cluster 1 included novel transcripts such as *LINC01361*, *LINC01460,* and *LINC01558,* as well as RNA transcripts *LINC00543* and *LINC00702,* which have not previously been identified in SCCs (Supplementary Table S7). The expression of selected lncRNAs, both novel and those not previously characterized in cSCC within cluster 1 in cSCC cells and NHEKs was demonstrated as a heatmap by RNA-seq (Supplementary Fig. S5). Notably, the cSCC invasion-related *MMP13* and complement factor I (*CFI*) were among the top upregulated protein-coding genes in Cluster 1 (Fig. 3F) (Supplementary Table S7).

Cluster 1 genes were also annotated with potential upstream regulator transcription factors (TFs) (Supplementary Fig. S6, Supplementary Table S8). Only experimentally validated TFs were included in the annotations. Myc was shown to be upstream regulator for 377 genes and EZH2 for 368 genes (Supplementary Fig. S6). These TFs were shown to activate the transcription of genes coding for molecules potentially related to cancer progression, such as Myc regulating *CFI* and EZH2 regulating *MMP13* and *CFI* (Supplementary Table S8).

Validation of differentially expressed lncRNAs in cSCC

The expression of selected lncRNAs in cSCC cells was verified using qRT-PCR (Fig. 4A). *HOTTIP*, *LINC00543*, *LINC00702*, *LINC01361*, *LINC01460*, and *LINC01558* were significantly upregulated in cSCC cells compared to NHEKs (Fig. 4A). As a control, the mRNA levels of *MMP13* were also observed to be upregulated in cSCC cells (Fig. 4A).

The expression of lncRNAs in HNSCC cell lines

The expression of these selected lncRNAs was also investigated in patient-derived HNSCC cells using qRT-PCR (Fig. 4A). Significant overexpression of *LINC00543*, *LINC01361*, *LINC01460*, *LINC01558* and *HOTTIP* was detected in HNSCC cells compared to normal keratinocytes (Fig. 4B). Additionally, *MMP13* was significantly upregulated in HNSCC cells, as previously shown[24,25] (Fig. 4B).

The expression of *LINC00702*, *LINC01558* and *HOTTIP* in Ha-ras-transformed tumorigenic HaCaT cells

To further elucidate the role of the invasion cluster related lncRNAs during the progression of cSCC, the expression of *LINC00702*, *LINC01558, HOTTIP* and *MMP13* was determined in an immortalized non-tumorigenic keratinocyte–derived cell line (HaCaT) lacking functional p53, as well as in three Ha-*ras*-transformed tumorigenic HaCaT-derived cell lines with varying levels of tumorigenicity in vivo (A5, II-4, and RT3)[26]. Other selected lncRNAs were not expressed in this in vitro model of cSCC progression. *LINC00702* was expressed only in RT3 cells, the most aggressive ras-transformed tumorigenic HaCaT-derived cell line (Fig. 5A). The expression of *LINC01558* was low in HaCaT and A5 cells, whereas markedly higher levels were observed in II-4 and RT3 cells (Fig. 5A). The expression of *HOTTIP* and *MMP13* was increased in A5 compared to HaCaT cells, but was low in II-4 and RT3 cells (Fig. 5A).

Regulation of lncRNAs *LINC00543* and *LINC00702* by TGF-β in cSCC cells

Transforming growth factor-β (TGF-β) signaling has been recognized as a regulator of cSCC cell invasion[53–56]. Therefore, the effect of TGF-β on the regulation of cell motility-related lncRNAs was examined. Cutaneous SCC cell lines were treated with TGF-β, and the regulation on the lncRNAs was assessed using qRT-PCR. Treatment with TGF-β resulted in the upregulation of *LINC00702* and a significant downregulation of *LINC00543* in most SCC cell lines (Fig. 5B). In contrast, *LINC01361* and *LINC01558* were downregulated, while *LINC01460* was upregulated in cSCC cells following TGF-β treatment (Supplementary Fig. S7).

The expression of lncRNAs in cSCC tumors in vivo

Overexpression of *LINC00543* was detected in cSCC tumor samples in vivo compared to normal skin using qRT-PCR (Fig. 6A). Additionally, qRT-PCR analysis indicated an upregulation of the expression of the protein-coding gene *MMP13* in cSCC tumors in vivo (Fig. 6B)[57]. Furthermore, by analyzing the GSE111582 database[8] we observed that *LINC00702*, *LINC01361* and *MMP13* were upregulated in recessive dystrophic epidermolysis bullosa-associated cSCC (RDEBSCC), an aggressive form of cSCC compared to non-tumor samples (Supplementary Fig. S8). Using another dataset (GSE45216)[43] we detected a slight increase of expression of *LINC01361* in immunocompetent patients with moderately and poorly differentiated cSCC compared to actinic keratoses (AKs), premalignant epidermal lesions (Supplementary Table S9). In cSCCs from immunosuppressed patients, we observed a slight increase in *LINC00702* expression when all cSCCs were compared to AKs. The expression of *LINC00702* was further upregulated when moderately and poorly differentiated cSCCs were specifically compared to AK (Supplementary Table S9).

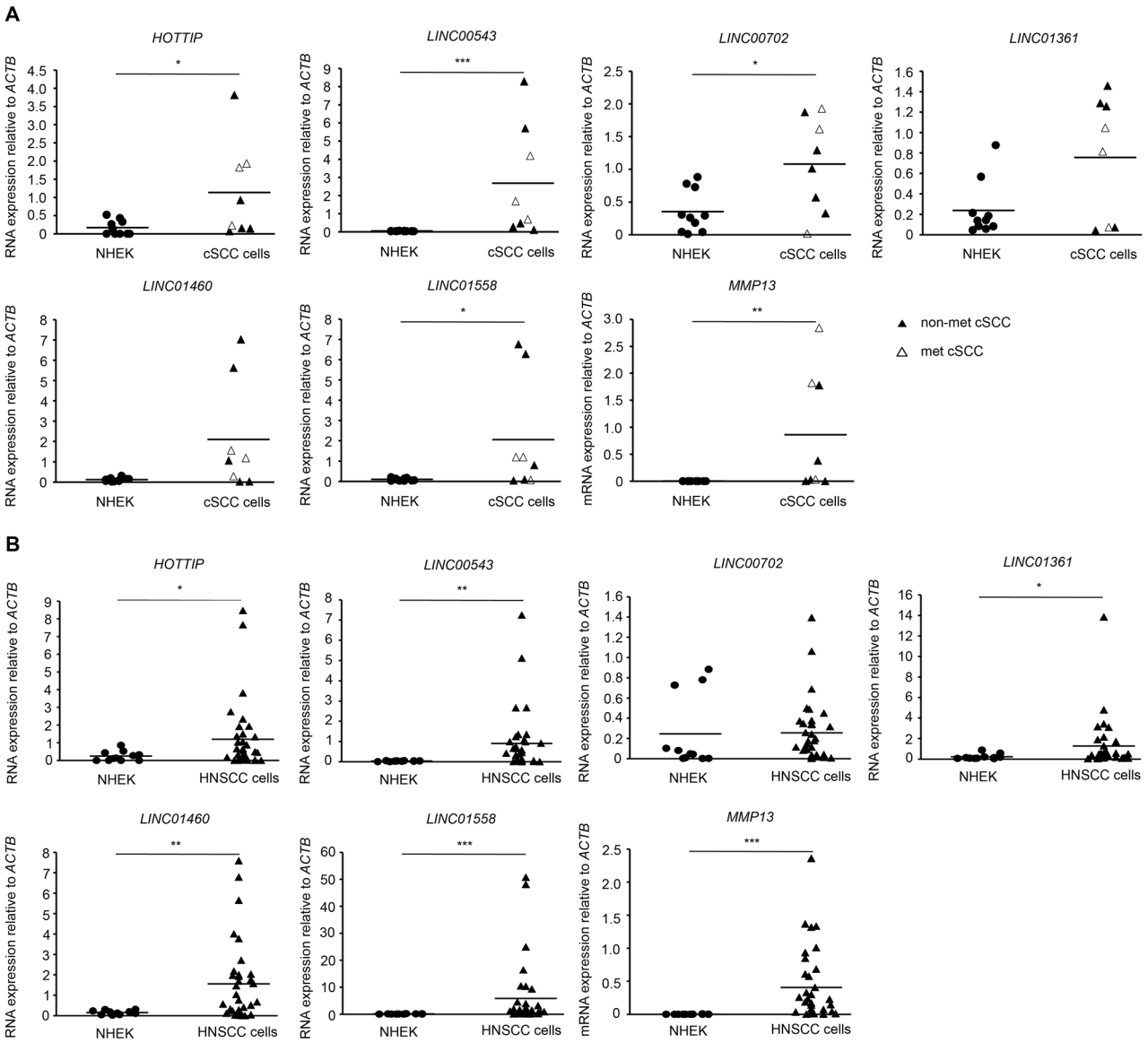

Figure 4. Validation of differentially expressed lncRNAs in cSCC and HNSCC cell lines. The RNA expression levels of *LINC00543*, *LINC00702*, *LINC01361*, *LINC01460*, *LINC01558*, *HOTTIP* and *MMP13* were determined using qRT-PCR in cSCC (**A**) and HNSCC (**B**) cells and NHEKs (**A**, n = 10; **B**, n = 11). The levels of β-actin (*ACTB*) mRNA were used as a reference gene. Statistical significance was determined by the Mann–Whitney U-test; $*p < 0.05$; $**p < 0.01$, $***p < 0.001$.

The expression of *LINC00543* and *HOTTIP* in other SCCs is associated with a poor prognosis

Analysis of data from the Cancer Genome Atlas (TCGA)[44,45] was conducted to study the expression of lncRNAs and *MMP13* in SCCs within the TCGA database, including HNSCC, ESCA and LUSCC[44,45]. Cluster 1-related *LINC00543*, *HOTTIP* and *MMP13* were upregulated in the investigated SCCs (Fig. 6C). Furthermore, the impact of lncRNA genes expression on the overall survival of SCC patients was assessed using available TCGA data[44,45]. Elevated expression of genes encoding *LINC00543* and *HOTTIP* was found to correlate with a poor prognosis, resulting in shorter overall survival in LUSCC and HNSCC, respectively (Fig. 6D).

Discussion

The prognosis for metastatic cSCC is poor, with no established biomarkers to predict metastatic risk or specific therapeutic targets for advanced or metastatic cSCCs. Therefore, it is important to characterize cSCC progression at the molecular level to uncover dysregulated pathways and to identify key drivers of the disease. While lncRNAs are being extensively studied in cSCC, their precise role in the progression of cSCC remains largely unknown[15]. Several recent studies have identified lncRNAs, such as *MALAT1*, *PVT1*, *LINC00319*, that play roles in cSCC progression[20,21,58,59]. Additionally, our previous studies have highlighted the role of lncRNAs *PICSAR*, *PRECSIT* and *SERLOC* in cSCC[16–19].

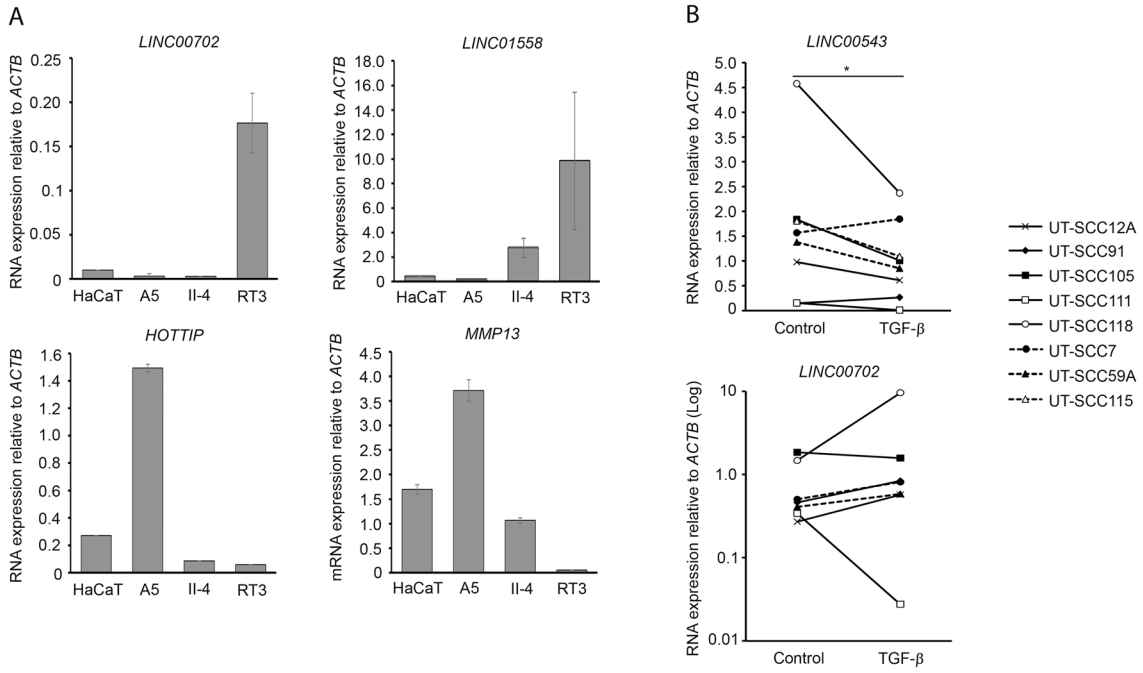

Figure 5. The regulation of lncRNA expression in Ha-*ras*-transformed HaCaT and cSCC cells. (**A**) Evaluation of *LINC00702* (n = 2–3), *LINC01558* (n = 3), *HOTTIP* (n = 2–3) and *MMP13* (n = 3) expression levels in HaCaT cells and tumorigenic Ha-ras-transformed HaCaT cell lines (A5, II-4, and RT3) with qRT-PCR normalized to β-actin (*ACTB*) mRNA levels. (**B**) Cutaneous SCC cell lines (n = 7–8) were treated with transforming growth factor-β1 (TGF-β1; 5 ng/mL) for 24 h. Subsequently, *LINC00543* and *LINC00702* levels were determined using qRT-PCR and normalized to β-actin (*ACTB*) mRNA levels. $^{*}p < 0.05$, Mann–Whitney U-test.

In this study, deep RNA-seq analysis of patient-derived cSCC cells and NHEKs was conducted to discover novel lincRNAs involved in cSCC progression. Among the identified lncRNAs and mRNAs that were most significantly downregulated, some were previously associated with cSCC progression. For instance, *LINC00520*, known for its inhibitory effects on cSCC progression, was among the top downregulated lncRNAs[47]. Conversely, the upregulated lncRNA *PRECSIT* (*LINC00346*) was previously shown to promote cSCC progression[18]. Additionally, the upregulation of *AIM2*, a protein-coding gene encoding an inflammasome component associated with promoting cSCC growth and invasion, further supported the findings of this RNA-seq analysis in accordance with previous research on cSCC cells[48].

To gain additional insight into cSCC progression, we utilized the deep RNA-seq data for integrative analysis of lncRNAs and protein-coding genes to predict their potential roles in cSCC progression. Integrative analysis is a commonly employed approach to predict the functions of lncRNAs in various biological processes[60]. Previously a potential reference set of lncRNAs, mRNAs and circular RNAs in cSCC was identified using whole transcriptome sequencing[61]. Furthermore, microarray analysis of cSCC precursor AK samples has revealed a lncRNA implicated in JAK-STAT3 signaling pathway in AK[62]. However, to our knowledge no integration analysis based on whole transcriptome sequencing of lncRNAs and protein coding genes has been conducted for genes regulated in cSCC samples. The integration analysis revealed six gene clusters, two of which were upregulated. We further investigated one of the upregulated clusters, specifically Cluster 1. Bioinformatics analysis of gene enrichment revealed that the genes in Cluster 1 were significantly enriched in terms and pathways related to cell motility. Therefore, we assigned the name "cSCC motility cluster" to Cluster 1.

We focused on cSCC motility cluster-related lincRNAs, which were the second largest biotype of lncRNAs after antisense RNAs. Previous studies in other cancers have also highlighted, that the two predominant categories of lncRNAs are antisense and lincRNAs[63]. Among the top regulated lncRNAs in Cluster 1 there were several novel transcripts and previously uncharacterized lincRNAs, along with some previously characterized cancer cell motility-related lncRNAs such as *HOTTIP*[51] and *LINC00491*[52]. *HOTTIP* has been shown to regulate the invasion of prostate cancer cells[64]. Additionally, *LINC00491* has been identified to regulate the migration and invasion of colon adenocarcinoma cells[65]. Furthermore, the previously known cSCC cell invasion-related matrix metalloproteinase MMP-13, and the complement inhibitor serine protease complement factor I (CFI) were found to be involved in this cluster of genes, emphasizing the functional significance of this cluster[50,66]. Myc and EZH2 were identified as the most frequent TFs in cell motility cluster. These findings are in accordance with previous studies showing that Myc can induce poor differentiation of cSCC, suggesting its role in the development of a more aggressive tumor phenotype[67]. EZH2 has been observed to enhance the progression of cSCC[68]. Interestingly, EZH2 has also been shown to interact with Myc and the coactivator (p300) via a cryptic transactivation domain (TAD), and this way induce gene activation and oncogenesis[69].

We investigated further the expression of selected lncRNAs and one mRNA, *MMP13*, in patient derived cSCC and HNSCC cells. The upregulation of previously uncharacterized lncRNAs, such as *LINC01361*, *LINC01460* and

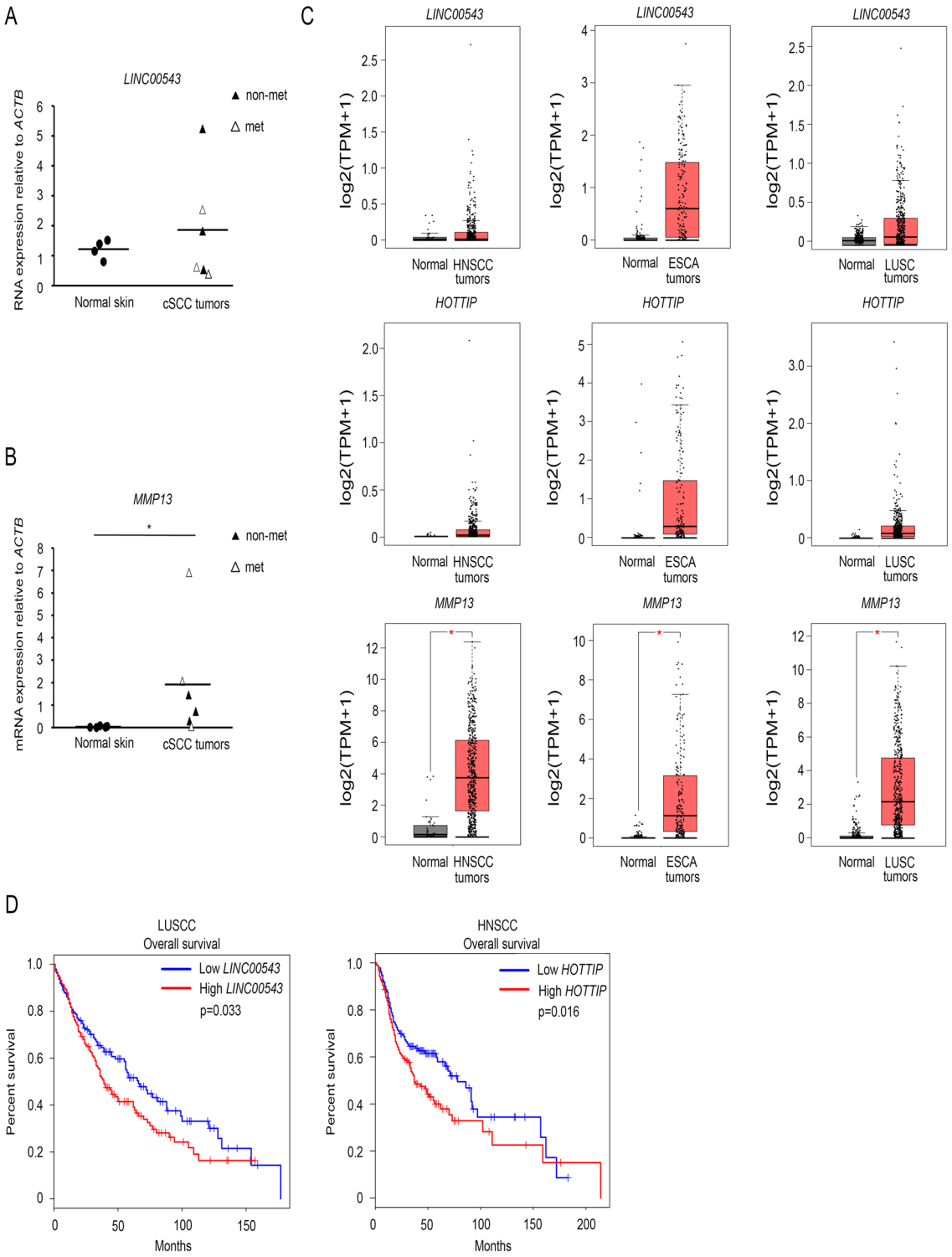

Figure 6. *LINC00543* and *HOTTIP* are linked to a poor prognosis in SCC. (**A, B**) The expression of (**A**) *LINC00543* and (**B**) *MMP13* was determined in primary non-metastatic (non-met) and metastatic (met) cSCC tumor samples (n = 6) in vivo compared to normal skin (n = 4–5) using qRT-PCR. Statistical significance (*$p < 0.05$) was determined using the Mann–Whitney U-test. (**C**) Analysis of The Cancer Genome Atlas (TCGA) data was conducted using the GEPIA tool to investigate the expression of *LINC00543*, *HOTTIP* and *MMP13* in squamous cell carcinomas within TCGA database (head and neck SCC (HNSCC), esophageal carcinoma (ESCA) and lung SCC (LUSCC)). (**D**) Evaluation of the impact of *LINC00543* and *HOTTIP* on the overall survival of SCC patients was performed using the GEPIA tool with available TCGA data.

LINC01558 as well as already characterized *HOTTIP*, *LINC00543* and *LINC00702* was confirmed in cSCC cells compared to NHEKs using qRT-PCR. These lncRNAs, except *LINC00702* were also upregulated in HNSCC cells. Additionally, the expression of *LINC00702*, *LINC01558* and *HOTTIP* in an in vitro model of cSCC progression confirmed their upregulation during epidermal keratinocyte carcinogenesis. Notably, *HOTTIP* has been demonstrated to regulate cancer progression by promoting the invasion of prostate cancer cells, while *LINC00543* promotes the metastasis of colorectal cancer[70]. On the other hand, *LINC00702* has been previously reported to be downregulated in gastric and bladder cancer[71,72] but to promote the pathogenesis of meningeoma[73].

TGF-β signaling pathway is involved in the invasion of SCC cells[53–55]. A decrease in TGF-β signaling has been observed to be linked with an increased invasion depth of cSCC[56]. In this study we demonstrate, that *LINC00543* and *LINC00702* are regulated by TGF-β. While TGF-β appears to play an important role in cell motility, our results support the notion, that these lncRNAs are part of the cell motility cluster.

The expression of selected lncRNAs was also assessed in cSCC tumor samples in vivo using qRT-PCR. *LINC00543* was found to be upregulated in cSCC tumor samples in vivo compared to normal skin. Moreover, qRT-PCR analysis demonstrated a significant upregulation of *MMP13* in cSCC compared to normal skin. Analysis of data from the GEO database revealed, that the expression of *LINC00702* and *LINC01361* was elevated in a more aggressive form of cSCC, RDEBSCC, compared to non-SCC samples. In immunosuppressed patients with cSCC, *LINC00702* was upregulated in moderately and poorly differentiated cSCCs compared to AK samples in vivo. These findings confirm the role of these lncRNAs as markers for aggressive and advanced disease. Additionally, TCGA data was used to investigate the expression of lncRNAs and *MMP13* in other SCCs. *LINC00543*, *HOTTIP* and *MMP13* were found to be upregulated in HNSCC, ESCA, and LUSCC. Furthermore, higher expression levels of *LINC00543* and *HOTTIP* were associated with a poorer prognosis, indicating shorter overall survival in LUSCC and HNSCC, respectively. These results suggest that upregulation of *LINC00543* and *HOTTIP* may serve as indicators of poor prognosis in SCC patients. Further studies with larger cSCC sample set are required to validate these results, as the cSCC expression data are not included in TCGA database.

In summary, we employed deep RNA-seq data for the integrated analysis of lncRNAs and mRNAs to predict their potential functions in cSCC progression. Based on these results, a set of lncRNAs was identified to be involved in cSCC cell motility cluster. The investigated lncRNAs were observed to be upregulated in patient-derived cSCC and HNSCC cell lines. Additionally, *LINC00543* and *HOTTIP* were found to be overexpressed in tumor samples in vivo and higher expression of *LINC00543* and *HOTTIP* was associated with a poor prognosis for patients with LUSCC or HNSCC. The findings of this study indicate that the cell motility cluster-related lncRNAs *HOTTIP*, *LINC00543*, *LINC00702*, *LINC01361*, *LINC01460* and *LINC01558* may serve as potential biomarkers and therapeutic targets in advanced and metastatic cSCC.

Data availability

The RNA-seq data of cSCC cell lines and NHEKs have been deposited in the GEO database with the accession code GSE252347.

Received: 5 January 2024; Accepted: 16 July 2024
Published online: 23 July 2024

References

1. Adnane, S., Marino, A. & Leucci, E. LncRNAs in human cancers: Signal from noise. *Trends Cell. Biol.* **32**, 565–573 (2022).
2. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
3. Marchese, F. P., Raimondi, I. & Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* **18**, 206 (2017).
4. Nehal, K. S. & Bichakjian, C. K. Update on keratinocyte carcinomas. *N. Engl. J. Med.* **379**, 363–374 (2018).
5. Burton, K. A., Ashack, K. A. & Khachemoune, A. Cutaneous squamous cell carcinoma: A review of high-risk and metastatic disease. *Am. J. Clin. Dermatol.* **17**, 491–508 (2016).
6. Knuutila, J. S., Riihilä, P., Kurki, S., Nissinen, L. & Kähäri, V. M. Risk factors and prognosis for metastatic cutaneous squamous cell carcinoma: A cohort study. *Acta Derm. Venereol.* **100**, adv00266 (2020).
7. Winge, M. C. G. *et al.* Advances in cutaneous squamous cell carcinoma. *Nat. Rev. Cancer* **23**, 430–449 (2023).
8. Cho, R. J. *et al.* APOBEC mutation drives early-onset squamous cell carcinomas in recessive dystrophic epidermolysis bullosa. *Sci. Transl. Med.* **10**, eaas9668 (2018).
9. Piipponen, M., Riihilä, P., Nissinen, L. & Kähäri, V. M. The Role of p53 in progression of cutaneous squamous cell carcinoma. *Cancers* **13**, 4507 (2021).
10. Hedberg, M. L. *et al.* Molecular mechanisms of cutaneous squamous cell carcinoma. *Int. J. Mol. Sci.* **23**, 3478 (2022).
11. Pickering, C. R. *et al.* Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin. Cancer Res.* **20**, 6582–6592 (2014).
12. South, A. P. *et al.* NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. *J. Invest. Dermatol.* **134**, 2630–2638 (2014).
13. Nissinen, L., Farshchian, M., Riihilä, P. & Kähäri, V. M. New perspectives on role of tumor microenvironment in progression of cutaneous squamous cell carcinoma. *Cell Tissue Res.* **365**, 691–702 (2016).
14. Riihilä, P., Nissinen, L. & Kähäri, V. M. Matrix metalloproteinases in keratinocyte carcinomas. *Exp. Dermatol.* **30**, 50–61 (2021).
15. Piipponen, M., Nissinen, L. & Kähäri, V. M. Long non-coding RNAs in cutaneous biology and keratinocyte carcinomas. *Cell. Mol. Life Sci.* **77**, 4601–4614 (2020).
16. Piipponen, M. *et al.* Long noncoding RNA PICSAR promotes growth of cutaneous squamous cell carcinoma by regulating ERK1/2 activity. *J. Invest. Dermatol.* **136**, 1701–1710 (2016).
17. Piipponen, M., Heino, J., Kähäri, V. M. & Nissinen, L. Long non-coding RNA PICSAR decreases adhesion and promotes migration of squamous carcinoma cells by downregulating α2β1 and α5β1 integrin expression. *Biol. Open* **7**, bio037044 (2018).
18. Piipponen, M. *et al.* p53-regulated long noncoding RNA PRECSIT Promotes progression of cutaneous squamous cell carcinoma via STAT3 Signaling. *Am. J. Pathol.* **190**, 503–517 (2020).

19. Piipponen, M. *et al.* Super enhancer-regulated LINC00094 (SERLOC) upregulates the expression of MMP-1 and MMP-13 and promotes invasion of cutaneous squamous cell carcinoma. *Cancers* **14**, 3980 (2022).
20. Li, C. *et al.* Long non-coding RNA PVT1 is overexpressed in cutaneous squamous cell carcinoma and exon 2 is critical for its oncogenicity. *Br. J. Dermatol.* **190**, 415–426 (2024).
21. Li, R. *et al.* The long non-coding RNA PVT1 promotes tumorigenesis of cutaneous squamous cell carcinoma via interaction with 4EBP1. *Cell Death Discov.* **9**, 101 (2023).
22. Farshchian, M., Nissinen, L., Grénman, R. & Kähäri, V. M. Dasatinib promotes apoptosis of cutaneous squamous carcinoma cells by regulating activation of ERK1/2. *Exp. Dermatol.* **26**, 89–92 (2017).
23. Nissinen, L. *et al.* C1s targeting antibodies inhibit the growth of cutaneous squamous carcinoma cells. *Sci. Rep.* **14**, 13465 (2024).
24. Johansson, N. *et al.* Expression of collagenase-3 (matrix metalloproteinase-13) in squamous cell carcinomas of the head and neck. *Am. J. Pathol.* **151**, 499–508 (1997).
25. Stokes, A. *et al.* Expression profiles and clinical correlations of degradome components in the tumor microenvironment of head and neck squamous cell carcinoma. *Clin. Cancer Res.* **16**, 2022–2035 (2010).
26. Mueller, M. M. *et al.* Tumor progression of skin carcinoma cells in vivo promoted by clonal selection, mutagenesis, and autocrine growth regulation by granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor. *Am. J. Pathol.* **159**, 1567–1579 (2001).
27. Andrews, S. *FastQC: a quality control tool for high throughput sequence data*. 1–1 (2010).
28. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
29. R Core Team. R. *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
30. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
31. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
32. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
33. Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.10. https://CRAN.R-project.org/package=pheatmap (2018).
34. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* **25**, 25–29 (2000).
35. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**(D1), D331–D338 (2017).
36. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
37. Krämer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530 (2014).
38. da Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
39. da Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
40. Jiang, C., Xuan, Z., Zhao, F. & Zhang, M. Q. TRED: A transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res.* **35**(Database issue), D137-140 (2007).
41. Liu, Z. P., Wu, C., Miao, H. & Wu, H. RegNetwork: An integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* https://doi.org/10.1093/database/bav095 (2015).
42. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
43. Lambert, S. R. *et al.* Key differences identified between actinic keratosis and cutaneous squamous cell carcinoma by transcriptome profiling. *Br. J. Cancer.* **110**, 520–529 (2014).
44. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).
45. Weinstein, J. N. *et al.* The Cancer Genome Atlas pan-cancer analysis project. *Nat. Genet.* **45**(10), 1113–1120 (2013).
46. Tang, Z. *et al.* GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**(W1), W98–W102 (2017).
47. Mei, X. L. & Zhong, S. Long noncoding RNA LINC00520 prevents the progression of cutaneous squamous cell carcinoma through the inactivation of the PI3K/Akt signaling pathway by downregulating EGFR. *Chin. Med. J.* **132**, 454–465 (2019).
48. Farshchian, M. *et al.* Tumor cell-specific AIM2 regulates growth and invasion of cutaneous squamous cell carcinoma. *Oncotarget* **8**, 45825–45836 (2017).
49. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**(Database issue), D561–D568 (2011).
50. Ala-aho, R. *et al.* Targeted inhibition of human collagenase-3 (MMP-13) expression inhibits squamous cell carcinoma growth in vivo. *Oncogene* **23**, 5111–5123 (2004).
51. Feng, H. *et al.* Long non-coding RNA HOTTIP exerts an oncogenic function by regulating HOXA13 in nasopharyngeal carcinoma. *Mol. Biol. Rep.* **50**, 6807–6818 (2023).
52. Wan, H., Lin, T., Shan, M., Lu, J. & Guo, Z. LINC00491 facilitates tumor progression of lung adenocarcinoma via Wnt/β-catenin-signaling pathway by regulating MTSS1 Ubiquitination. *Cells* **11**, 3737 (2022).
53. Leivonen, S. K. *et al.* Activation of Smad signaling enhances collagenase-3 (MMP-13) expression and invasion of head and neck squamous carcinoma cells. *Oncogene* **25**, 2588–2600 (2006).
54. Siljamäki, E. *et al.* H-Ras activation and fibroblast-induced TGF-β signaling promote laminin-332 accumulation and invasion in cutaneous squamous cell carcinoma. *Matrix Biol.* **87**, 26–47 (2020).
55. Siljamäki, E. *et al.* Inhibition of TGF-β signaling, invasion, and growth of cutaneous squamous cell carcinoma by PLX8394. *Oncogene* **42**, 3633–3647 (2023).
56. Rose, A. M. *et al.* Reduced SMAD2/3 activation independently predicts increased depth of human cutaneous squamous cell carcinoma. *Oncotarget* **9**, 14552–14566 (2018).
57. Viiklepp, K. *et al.* C1r upregulates production of matrix metalloproteinase-13 and promotes invasion of cutaneous squamous cell carcinoma. *J. Invest. Dermatol.* **142**, 1478–1488 (2022).
58. Zhang, C., Wang, J., Guo, L. & Peng, M. Long non-coding RNA MALAT1 regulates cell proliferation, invasion and apoptosis by modulating the Wnt signaling pathway in squamous cell carcinoma. *Am. J. Transl. Res.* **13**, 9233–9240 (2021).
59. Li, F., Liao, J., Duan, X., He, Y. & Liao, Y. Upregulation of LINC00319 indicates a poor prognosis and promotes cell proliferation and invasion in cutaneous squamous cell carcinoma. *J. Cell. Biochem.* **119**, 10393–10405 (2018).
60. Chen, H. *et al.* Comprehensive analysis of mRNA-lncRNA co-expression profile revealing crucial role of imprinted gene cluster DLK1-MEG3 in chordoma. *Oncotarget* **8**, 112623–112635 (2017).
61. Das Mahapatra, K. *et al.* A comprehensive analysis of coding and non-coding transcriptomic changes in cutaneous squamous cell carcinoma. *Sci. Rep.* **10**, 3637 (2020).
62. Luan, C. *et al.* Whole-genome identification and construction of the lncRNA-mRNA co-expression network in patients with actinic keratosis. *Transl. Cancer Res.* **11**, 4070–4078 (2022).

63. Siena, Á. D. D. *et al.* Whole transcriptome analysis reveals correlation of long noncoding RNA ZEB1-AS1 with invasive profile in melanoma. *Sci. Rep.* **9**, 11350 (2019).
64. Qian, Y. *et al.* Systematic fine-mapping and functional studies of prostate cancer risk variants. *iScience* **26**, 106497 (2023).
65. Wan, J. *et al.* LINC00491 as a new molecular marker can promote the proliferation, migration and invasion of colon adenocarcinoma cells. *Onco Targets Ther.* **12**, 6471–6480 (2019).
66. Rahmati Nezhad, P. *et al.* Complement factor I upregulates expression of matrix metalloproteinase-13 and -2 and promotes invasion of cutaneous squamous carcinoma cells. *Exp. Dermatol.* **30**, 1631–1641 (2021).
67. Toll, A. *et al.* MYC gene numerical aberrations in actinic keratosis and cutaneous squamous cell carcinoma. *Br. J. Dermatol.* **161**, 1112–1118 (2009).
68. Xie, Q. *et al.* Increased expression of enhancer of Zeste Homolog 2 (EZH2) differentiates squamous cell carcinoma from normal skin and actinic keratosis. *Eur. J. Dermatol.* **24**, 41–45 (2014).
69. Wang, J. *et al.* EZH2 noncanonically binds cMyc and p300 through a cryptic transactivation domain to mediate gene activation and promote oncogenesis. *Nat. Cell Biol.* **24**, 384–399 (2022).
70. Zheng, J. *et al.* LINC00543 promotes colorectal cancer metastasis by driving EMT and inducing the M2 polarization of tumor associated macrophages. *J. Transl. Med.* **21**, 153 (2023).
71. Ma, M., Li, J., Zeng, Z., Zheng, Z. & Kang, W. Integrated analysis from multicentre studies identities m7G-related lncRNA-derived molecular subtypes and risk stratification systems for gastric cancer. *Front. Immunol.* **14**, 1096488 (2023).
72. Pan, W. *et al.* LINC00702-mediated DUSP1 transcription in the prevention of bladder cancer progression: Implications in cancer cell proliferation and tumor inflammatory microenvironment. *Genomics* **114**, 110428 (2022).
73. Ghafouri-Fard, S., Abak, A., Hussen, B. M., Taheri, M. & Sharifi, G. The emerging role of non-coding RNAs in pituitary gland tumors and meningioma. *Cancers* **13**, 5987 (2021).

Acknowledgements

We thank Johanna Markola for expert technical assistance; Dr. Reidar Grénman for cSCC and HNSCC cell lines and tumor samples; Dr. Norbert E. Fusenig (The German Cancer Research Center, Heidelberg, Germany) for HaCaT cells; GeneVia Technologies (Tampere, Finland) for bioinformatics analysis and the Bioinformatics Unit of Turku Bioscience Centre for the gene expression profiling data analysis. The Unit is supported by the University of Turku, Åbo Akademi University, and Biocenter Finland.

Author contributions

Conception and design: LN, MP, VMK; Methodology: LN, JH, PR; Acquisition of data: LN, JH; Analysis and interpretation of data: LN, JH, VMK; Writing, review, and revision of the manuscript: LN, JH, PR, MP, VMK; Study supervision: LN, PR, VMK; Acquisition of funding: VMK.

Funding

This research was supported by Jane and Aatos Erkko Foundation, Sigrid Jusélius Foundation, Cancer Foundation Finland, Turku University Hospital (VTR Grant 13336), and Cancer Foundation of the Southwest Finland (PR) and The Finnish Medical Foundation (PR).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-67808-x.

Correspondence and requests for materials should be addressed to V.-M.K.