

OPEN

# A comprehensive analysis of coding and non-coding transcriptomic changes in cutaneous squamous cell carcinoma

Kunal Das Mahapatra<sup>1,2</sup>, Lorenzo Pasquali<sup>1,2</sup>, Jonas Nørskov Søndergaard<sup>3</sup>, Jan Lapins<sup>4</sup>, István Balazs Nemeth<sup>5</sup>, Eszter Baltás<sup>5</sup>, Lajos Kemény<sup>5</sup>, Bernhard Homey<sup>6</sup>, Liviu-Ionut Moldovan<sup>7,8</sup>, Jørgen Kjems<sup>7,8</sup>, Claudia Kutter<sup>3</sup>, Enikő Sonkoly<sup>1,2,4</sup>, Lasse Sommer Kristensen<sup>7,8</sup> & Andor Pivarcsi<sup>1,2,9\*</sup>

Cutaneous Squamous Cell Carcinoma (cSCC) is the most common and fastest-increasing cancer with metastatic potential. Long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) are novel regulators of gene expression. To identify mRNAs, lncRNAs and circRNAs, which can be involved in cSCC, RNA-seq was performed on nine cSCCs and seven healthy skin samples. Representative transcripts were validated by NanoString nCounter assays using an extended cohort, which also included samples from pre-cancerous skin lesions (actinic keratosis). 5,352 protein-coding genes, 908 lncRNAs and 55 circular RNAs were identified to be differentially expressed in cSCC. Targets of 519 transcription factors were enriched among differentially expressed genes, 105 of which displayed altered level in cSCCs, including fundamental regulators of skin development (MYC, RELA, ETS1, TP63). Pathways related to cell cycle, apoptosis, inflammation and epidermal differentiation were enriched. In addition to known oncogenic lncRNAs (PVT1, LUCAT1, CASC9), a set of skin-specific lncRNAs were identified to be dysregulated. A global downregulation of circRNAs was observed in cSCC, and novel skin-enriched circRNAs, circ\_IFFO2 and circ\_POF1B, were identified and validated. In conclusion, a reference set of coding and non-coding transcripts were identified in cSCC, which may become potential therapeutic targets or biomarkers.

Cutaneous squamous cell carcinoma (cSCC) is one of the most common human malignancies worldwide, with an yearly 700,000 diagnosed cases in the US alone<sup>1</sup>. This keratinocyte-derived cancer develops mostly on sun-exposed skin, proceeds as a progressively invasive malignancy, starting from precancerous lesions, actinic keratosis (AK), which can progress into invasive cSCC<sup>2</sup>. The incidence of cSCC is increasing at an alarming rate worldwide majorly due to life-style changes, an ageing population and an increase in organ transplantations, which is a major risk factor for aggressive and multiple cSCCs<sup>3</sup>. Although early stage cSCC is curable by surgical excision, metastatic cSCC has a poor long-term survival rate of 10–20% due to inefficacy of systemic chemotherapy<sup>4</sup>. Thus, there is an urgent need to identify new druggable targets and pathways in cSCC.

The most important risk factor for cSCC is a cumulative lifetime exposure to UV-B radiation. Consequently, cSCC has an extremely high mutational burden with approximately 50 mutations per mega-base coding sequence and driver mutations in key tumor suppressor genes (*TP53* and *NOTCH1/2*) and oncogenes (*HRAS* and *KRAS*)<sup>5</sup>. These early mutational events prime for oncogenic transformation through altered cell cycle, decreased apoptosis,

<sup>1</sup>Dermatology and Venereology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.

<sup>2</sup>Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>3</sup>Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Science for Life Laboratory, SE-171 77, Stockholm, Sweden. <sup>4</sup>Unit of Dermatology, Karolinska University Hospital, SE-17176, Stockholm, Sweden. <sup>5</sup>Department of Dermatology and Allergology, Faculty of Medicine, University of Szeged, Szeged, Hungary. <sup>6</sup>Department of Dermatology, University Hospital Düsseldorf, Düsseldorf, Germany. <sup>7</sup>Interdisciplinary Nanoscience Center (iNANO), Aarhus University, DK-8000, Aarhus, Denmark. <sup>8</sup>Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark. <sup>9</sup>Research Institute of Translational Biomedicine, University of Szeged, Szeged, Hungary. \*email: [andor.pivarcsi@ki.se](mailto:andor.pivarcsi@ki.se)

all cumulatively leading towards uncontrolled proliferation of keratinocytes. Subsequent genomic alterations in oncogenes such as *EGFR* and *MYC* and tumor suppressors such as *TP63* and *NOTCH3* further contribute to the disease progression<sup>5,6</sup>. Previous transcriptome analyses have revealed thousands of protein-coding transcripts with altered expression in cSCC, but much less is known about the alterations in other types of RNAs<sup>7,8</sup>.

Long non-coding RNAs (lncRNAs) are a functionally diverse group of regulatory RNAs with transcript length of 200 nucleotides or longer<sup>9</sup>. The expression of lncRNAs is often stringently regulated in spatio-temporal manner during development<sup>10</sup>. Recent studies have convincingly shown vital roles for several lncRNAs not only in tissue homeostasis but also in tumor initiation, growth and metastasis<sup>11</sup>.

Circular RNAs (circRNAs) have recently been implicated in the regulation of gene networks with tissue-specific expression patterns<sup>12</sup>. CircRNAs are formed by a head-to-tail splicing event joining a 5' splice site to an upstream 3' splice site<sup>13</sup>. These molecules are exceptionally stable due to the lack of free ends and their functions are likely to be related to this structural feature. Any circRNAs have been shown to regulate gene expression in cancer via various modes of action such as decoys to sponge miRNAs and as regulators of transcription and alternative splicing<sup>14</sup>.

The goal of our study was to identify a reliable set of differentially expressed transcripts, including mRNAs, lncRNAs and circRNAs, in cSCC. To this end, we performed a RNA-seq analysis of cSCC and healthy skin at an unprecedented depth. Our analysis identified a large number of differentially expressed transcripts that included mRNAs, lncRNAs and circRNAs with previously uncharacterized roles in cSCC.

## Results

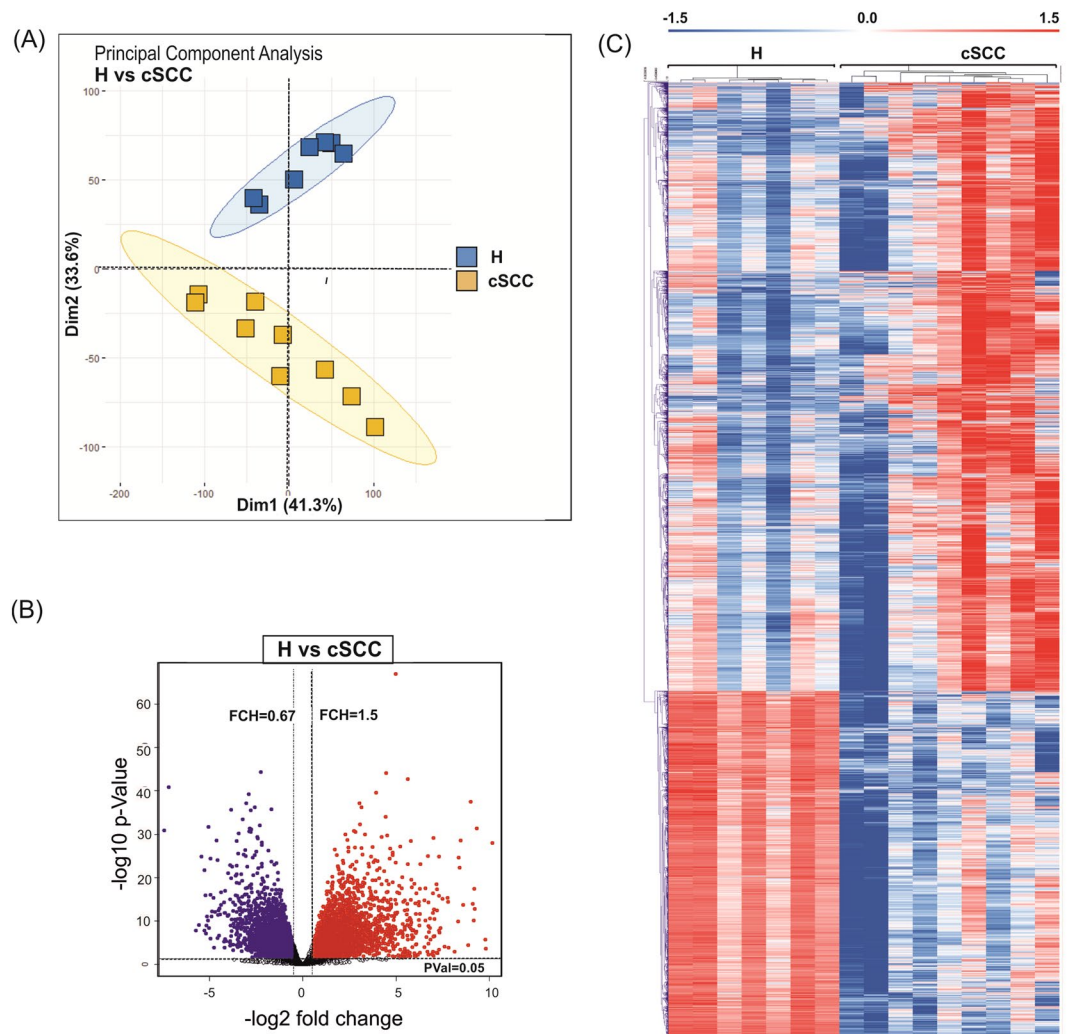
**Whole transcriptome profiling by RNA sequencing in cSCC and healthy skin.** In order to identify alterations in the expression of protein-coding as well as non-coding genes in cSCC, RNA sequencing of cSCCs (n = 9) and unmatched healthy skin samples (n = 7) was performed using the NextSeq500-platform, generating 800 million total reads (Supplementary Table S1), which to our knowledge represents the deepest transcriptomic analysis of cSCC to date. On average 49.8 million 100 base pair (bp) paired-end reads were obtained from each sample and genome mapping was on average 55% for all samples. We performed the subsequent analysis of coding sequences (mRNAs), non-coding transcripts (lncRNAs) and circular RNAs (circRNAs) separately.

**Altered expression of protein-coding genes in cSCC.** Principal component analysis (PCA) of all detected genes clearly separated cSCC from healthy skin samples (Fig. 1A). More variation was observed among cSCC samples as compared to samples obtained from healthy skin (H), potentially arising from an inherent heterogeneity of the disease caused by its exceptionally high mutational burden. Differential expression analysis identified 5,352 differentially expressed genes (DEGs) of which 3,419 were upregulated and 1,933 were down-regulated in cSCC (linear fold-change (FCH) > 1.5, false discovery rate (FDR) < 0.05) (Fig. 1B, Supplementary Table S2). Unsupervised hierarchical clustering of protein-coding genes separated the healthy skin and cSCC samples (Fig. 1C). The DEGs included several well-known genes related to skin carcinogenesis with roles in cell motility (e.g. *SNAI2*, *TGFBR1*), extracellular matrix remodeling (e.g. *BMP*, *MMP10*), cell proliferation (e.g. *MKI67*, *PCNA*), apoptosis (e.g. *BCL2*, *DDR1*), epidermal differentiation (e.g. *LCE2D*, *KRT10*, *MAF*), stemness (e.g. *ITGA6* and *ITGB1*) and inflammation (e.g. *IFNGR1*, *IL-8/CXCL8*) (Supplemental Fig. S1).

**Functional classification of deregulated protein-coding genes in cSCC.** In order to get an insight into the altered biological processes in cSCC, we performed Gene Ontology (GO) enrichment analysis on the identified DEGs. Because genes with increased expression can have different regulatory and functional function than downregulated genes, the gene enrichment analysis was performed on up- and downregulated genes separately (Supplementary Table S3A,B). The top enriched GO terms for the upregulated DEGs were “cellular response to interferon gamma” (GO:0071346,  $P < 0.0001$ ), “extracellular matrix organization” (GO:0030196,  $P < 0.0001$ ), “negative regulation of programmed cell death” (GO:0043169,  $P < 0.0001$ ), “positive regulation of NF-kappaB TF activity” (GO:0051092,  $P < 0.0001$ ) and “DNA damage response by p53 class mediator” (GO:0030330,  $P < 0.0001$ ) (Fig. 2A). Interestingly, a distinct signature for interferon-response was observed among DEGs with increased expression of interferon gamma receptors (*IFN-YR1/IFNGR1* and *IFN-YR2/IFNGR2*), downstream effector *STAT1* and a number of CC (*CCL2*, *CCL4*, *CCL5*, *CCL8*) and CXC chemokines (*CXCL1*, *CXCL10* and *CXCL11*) (Supplementary Fig. S1). For the down-regulated DEGs, “establishment of skin barrier” (GO:0061436,  $P = 0.003$ ) and “regulation of water loss via skin” (GO:0033561,  $P = 0.004$ ) were among the top enriched GO Biological processes along with processes such as “acetyl-CoA metabolic process” (GO:0006084,  $P = 0.001$ ) and “positive regulation of canonical *Wnt* signaling pathway” (GO:0035413,  $P < 0.05$ ).

Analysis of differentially expressed genes for enriched oncogenic signatures from the Molecular Signature Data Base revealed a strong enrichment of *EGFR*, *KRAS*, *mTOR*, *MEK* and *TP53*-gene signatures (Fig. 2B, Supplementary Table S3C,D). This observation supports the prevailing understanding of the molecular pathogenesis of cSCC according to which inactivation of *TP53* and activation of *EGFR*/*MAPK*-signaling pathways are frequent events in cSCC that lead to activation of *RAS*-*RAF*-*MEK* signaling pathway<sup>5</sup>. Interestingly, we also observed an enrichment of *SNF5*-regulated genes among the DEGs. *SNF5* has been previously shown to act as a part of the chromatin modifier- *SWI*/*SNF* complex to regulate cell survival in a context of *p53*-deficit<sup>15</sup>.

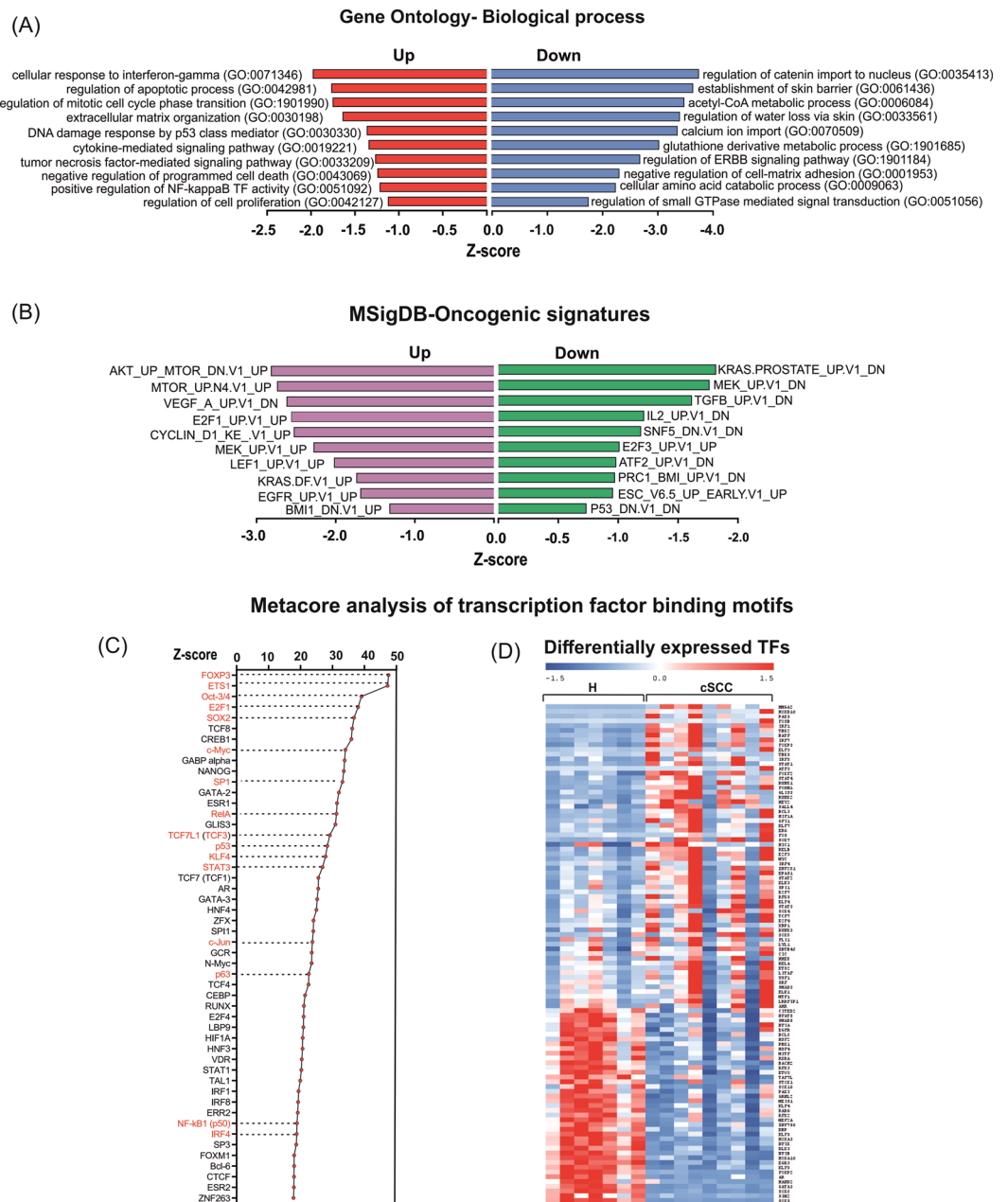
**Identification of differentially expressed transcriptional regulators in cSCC and validation of protein-coding transcripts.** We next aimed to identify the putative upstream regulators, which could be responsible for the observed transcriptomic changes in cSCC. To this end, we performed an enrichment analysis of transcription factor (TF) binding sites among the differentially expressed genes using MetaCore. This motif-based analysis identified 519 TFs whose target genes were significantly enriched among the DEGs ( $P < 0.05$ , Supplementary Table S4A) out of which 105 TFs were deregulated (64 up- and 41 downregulated) at the transcript level in our RNA-seq data (Fig. 2C,D) potentially explaining a part of the observed transcriptomic



**Figure 1.** Analysis of the protein-coding transcriptome in cSCC. **(A)** Principal component analysis of samples obtained from healthy skin samples (H; blue) and cSCC (cSCC; yellow) based on RNA-seq data. **(B)** Volcano plot shows the result of EdgeR-analysis of all detected mRNAs ( $\log_2$  fold change versus  $\log_{10}$  nominal P-value for all detected genes). Vertical lines denote the fold change cutoff, while the horizontal line denotes the P-value cutoff. Red color represents upregulated and blue color represents downregulated coding transcripts. **(C)** Heatmap and hierarchical clustering of all differentially expressed protein-coding genes in cSCC (FDR < 0.05 and FCH > 1.5).

changes (Supplementary Table S4B). One of the most overrepresented TF was ETS1, which has been previously shown to block terminal differentiation in keratinocytes along with induction of matrix metalloproteases<sup>16</sup>. Target genes of TP63, a master regulator of epidermal development and an important player in stemness and skin tumorigenesis, were also found to be significantly enriched<sup>17</sup>. In addition to TP63, overrepresentation of KLF4-target genes highlighted a impaired epidermal-differentiation program, which is often a primary event in cSCC<sup>18</sup>. TP53, the most frequently mutated gene in cSCC was also identified<sup>19</sup>. Several TFs with known (e.g. E2F1, c-MYC and SP1) and currently unreported functions (e.g. FOXP3, IRF1, USF1 and FOXM1) in cSCC were identified<sup>20–22</sup>. Enrichment of STATs (STAT3 and STAT5A) and IRF1-regulated genes can possibly contribute to the strong type I interferon response signature in our dataset. RELA was another over-represented TF, which highlighted an activation of NF-kappaB signaling in cSCC (Fig. 2A, Supplementary Fig. S1)<sup>23</sup>.

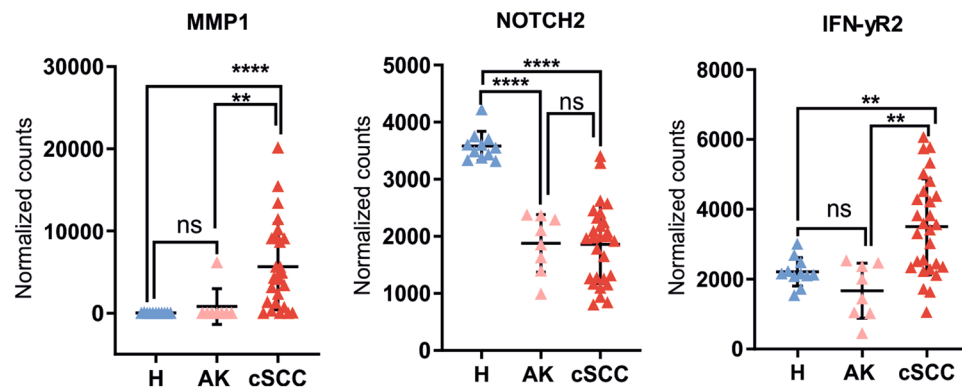
Next, to validate the results of the RNA-seq analysis, we performed NanoString nCounter assays for three DEGs on a larger cohort consisting of healthy skin (H, n = 11), cSCCs, (n = 28) and AKs (n = 8) (Fig. 3, Supplementary Table S5A). These three genes were representatives for enriched biological processes such as invasiveness (MMP1), elevated immune response (IFN- $\gamma$ 2), and altered epidermal differentiation (NOTCH2). NanoString analysis confirmed the significant upregulation of MMP1 (undetected in healthy skin) and IFN- $\gamma$ 2 (1.6-fold,  $P < 0.01$ ) in cSCC with no detectable alteration in precancerous skin lesions (AK), whereas NOTCH2 was found to be downregulated not only in cSCC (1.8-fold,  $P < 0.0001$ ) but even in pre-cancerous skin lesions (AK; 1.79-fold,  $P < 0.0001$ ).



**Figure 2.** Functional classification of deregulated mRNAs and identification of differentially expressed transcription factors in cSCC. **(A)** Top 10 ( $P < 0.05$ ) biological processes (Gene Ontology, sorted on Z-score) for up- and down-regulated coding genes in cSCC. **(B)** Top 10 most relevant ( $P < 0.05$ ) Molecular Signature Database (MSigDB) gene sets for up- and down-regulated coding genes in cSCC. **(C)** Top 50 transcription factors with overrepresented binding sites among differentially expressed coding genes in cSCC. Transcription factors in red color have previously been associated with cSCC pathogenesis. **(D)** Heatmap of transcription factors which are differentially expressed in cSCC (FDR  $< 0.05$  and FCH  $> 1.5$ ).

### Analysis of the long non-coding RNA landscape reveals differential expression of oncogenic and skin-specific lncRNAs.

Most of the previous transcriptomic studies in cSCC have only assessed protein-coding transcripts and little is known about the involvement of lncRNAs in cSCC. RNA-seq revealed a generally lower transcript abundance for lncRNAs compared to protein-coding transcripts both in healthy skin and in cSCC: while the median FPKM<sub>mRNAs</sub> was 757 and 847 in healthy skin and cSCC respectively, median FPKM<sub>lncRNAs</sub> was 153 and 75 ( $P = 0.01$  and  $0.003$  for H and cSCC, respectively) (Fig. 4A), in line with previous observations in other tissues<sup>24</sup>. Differential expression analysis identified 908 annotated lncRNAs with significantly altered expression in cSCC (FCH  $> 1.5$ , FDR  $< 0.05$ ), of which 319 were upregulated and 589 were downregulated in cSCC (Supplementary Table S6). Similar to protein-coding genes, unsupervised hierarchical clustering of differentially expressed lncRNAs clearly separated the samples into healthy and cSCC groups (Fig. 4B). The most upregulated lncRNA in cSCC was RP11-493L12.5 (46.77-fold), while the most



**Figure 3.** Validation of selected differentially expressed protein-coding genes. MMP1, NOTCH2 and IFN- $\gamma$ R2 in healthy skin (H,  $n = 11$ ), AK ( $n = 8$ ) and cSCC ( $n = 28$ ) by NanoString nCounter assay. Target gene expression is presented as background-corrected and normalized count (threshold count of 35). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , Mann-Whitney  $U$  test.

downregulated one was KB-1410C5.3/lnc-GRHL2 (0.005-fold). Of note, a number of lncRNAs with broad oncogenic (SNHG12, CASC9, LUCAT1 and PVT1) or tumor suppressor (such as TINCR) function were also identified to be differentially expressed in cSCC<sup>25–28</sup>.

Next, we classified differentially expressed lncRNAs into different subgroups based on their relative position to the nearby protein-coding genes: the most prominent classes of detected lncRNAs were intergenic and antisense (60.7% and 28.3%) followed by processed transcripts (7.7%), sense overlapping transcripts (2.5%) and sense intronic (0.4%) (Fig. 4C).

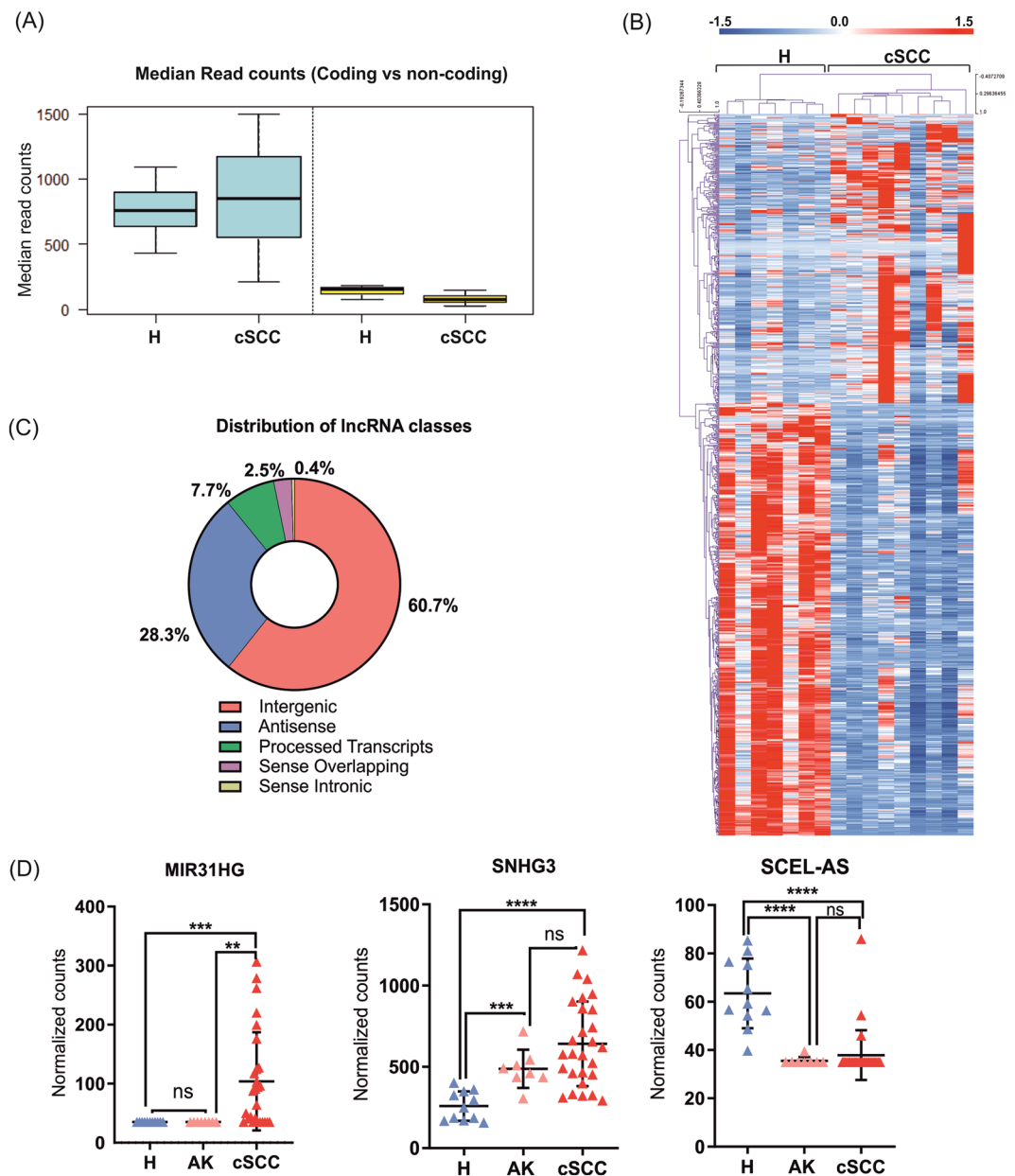
To get further insights into tissue-specific transcriptomic alterations, we calculated the tissue enrichment score (T) for all the differentially expressed protein-coding genes and lncRNAs using publicly available RNA-seq data for 28 tissues, including skin, from the GTEx database (Supplementary Table S7A,C)<sup>29</sup>. Although we obtained skin-enriched transcripts for both protein-coding genes and lncRNAs, lncRNAs displayed a higher degree of skin-specificity ( $T_s > 0.9$ ) as indicated by the relative frequency of their skin enrichment scores (Supplementary Fig. S3, Table S7B,D). Interestingly, analysis of the top 50 skin-enriched mRNAs and lncRNAs revealed that most of them were downregulated in cSCC (Supplementary Figs. S4 and S5). These skin-specific transcripts (e.g. RP13-455A7.1, CHODL-AS1, RP11-73G16.2) are largely uncharacterized in cSCC and therefore require further investigation.

Next, we aimed to identify the possible regulators of the differentially expressed lncRNAs in cSCC. Motif enrichment analysis (Supplementary Fig. S5) using their upstream regulatory regions resulted in a significantly high enrichment for CREB, ATF1 and NRF1 associated motifs for upregulated lncRNAs. A previous study has shown that CREB - a transcriptional co-activator of c-Jun regulates keratinocyte proliferation and differentiation<sup>30</sup>. For the downregulated lncRNAs, we observed a significant motif enrichment for SP1/SP4 transcription factors which have been demonstrated to regulate cell-specific gene expression during keratinocyte differentiation<sup>31</sup>.

Three differentially expressed lncRNAs were validated in an expanded cohort of patient samples that also included pre-cancerous skin lesions (AK), using NanoString nCounter assays (Fig. 4D). SCEL-AS1 was chosen as a representative of skin-enriched lncRNAs. MIR31HG and SNHG3, both of which have been reported to play oncogenic role in other cancers, were 2.5- and 2.3-fold upregulated ( $P < 0.0002$  and  $P < 0.0001$ ), respectively. Although MIR31HG was not altered, SNHG3 was found to be upregulated (1.9-fold,  $P = 0.0003$ ) already in the AK samples. SCEL-AS1, found to be significantly downregulated in cSCC and AK samples (1.5-fold,  $P < 0.0001$ ) is located antisense to Scellin (*SCEL*) with enriched expression in human skin and esophagus (<https://www.ncbi.nlm.nih.gov/gene/104355296/?report=expression>).

**Identification of novel circRNAs in healthy skin and cSCC.** Detection and quantification of circRNAs in our RNA seq dataset was done using a stringent version of the find\_circ pipeline as previously described<sup>32</sup>. In total, we identified 227 and 150 circRNAs supported by at least an average of five backsplicing junction-spanning reads in healthy skin and cSCCs, respectively (Supplementary Table S8A,B). Notably, most of these, 197/227 (86.8%) and 139/150 (92.7%), were also detected by the circExplorer pipeline indicating that these are high-confidence circRNAs (data not shown). Among the top 50 most abundant circRNAs in either sample type, we observed several circRNAs (e.g. CDR1as, circ\_HIPK3, circ\_CDYL, circ\_FAT1, circ\_LPAR1 and circ\_SLC8A1) that are highly expressed in several tissues in RNA-seq data from the ENCODE consortium (known as the top 10 alpha circRNAs (Supplementary Fig. S6A,B))<sup>33</sup>. In addition to annotated circRNAs, our analysis also identified six novel circRNA candidates, which are not annotated in circBase, derived from *IFFO2*, *PLIN4*, *DMKN*, *METRNL*, *KRT1* and *POF1B* genes (Supplementary Table S8). Of note, two of the novel circRNA (circ\_*IFFO2* and circ\_*PLIN4*) were relatively abundant ( $>1$  RPM) in the normal skin.

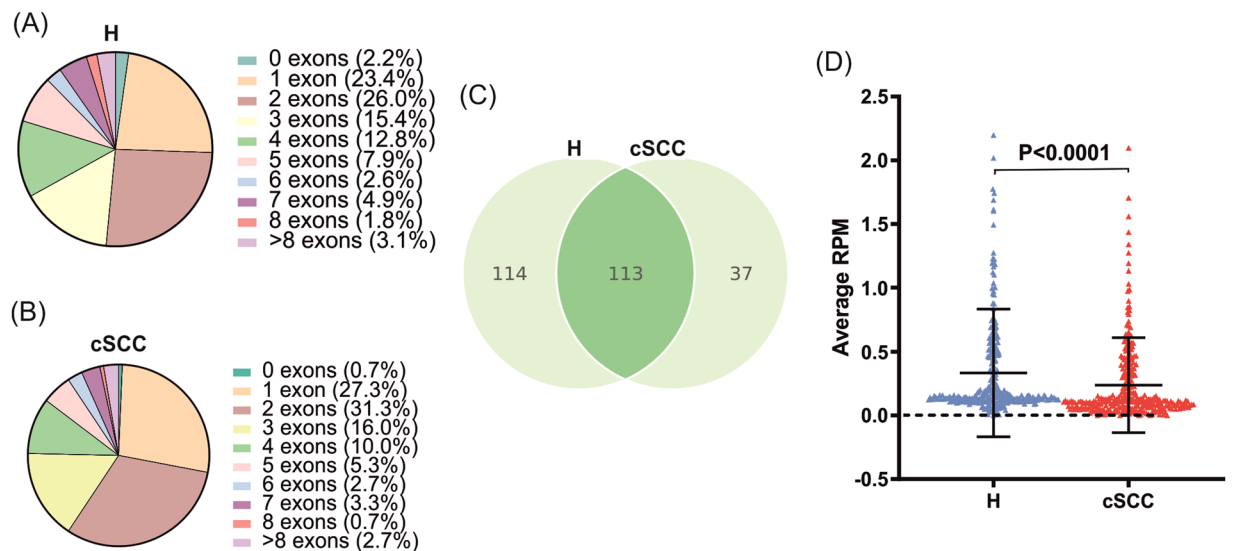
Evaluation of exon structures of the detected circRNAs revealed that circRNAs consisting of two exons was the most frequently observed population in both healthy and cSCC samples (Fig. 5A,B), similar to previous observations in other tissues and tumor types<sup>32</sup>. In terms of circRNA producing-loci, find\_circ identified that the 227 circRNAs in healthy skin from 210 host genes and that the 150 circRNAs in cSCC originated from 142 host



**Figure 4.** Analysis of differentially expressed long non-coding RNAs in cSCC. **(A)** Distribution of median read counts between coding and non-coding genes in healthy skin and cSCC. **(B)** Heatmap and hierarchical clustering of all differentially expressed lncRNAs in cSCC (FDR < 0.05 and FCH > 1.5). Red color represents upregulated and blue color represents downregulated lncRNAs. **(C)** Pie-chart depicting the proportion of different classes of lncRNAs with altered expression in cSCC. **(D)** Validation of selected differentially expressed lncRNAs (MIR31HG, SNGH3 and SCCL-AS) in healthy skin (H, n = 11), AK (n = 8) and cSCC (n = 28) by NanoString nCounter assay. Target gene expression is presented as background-corrected and normalized count (threshold count of 35). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , Mann-Whitney  $U$  test.

genes. Of note, 37 (16.3%) and 22 (14.7%) circRNAs were on average expressed at higher levels than their linear host genes (Circular to Linear ratio (CTL) > 1) in the normal skin and cSCC, suggesting their slow turnover or increased transcriptional output.

In total, 264 high abundance circRNAs were detected in cSCC and normal skin samples combined. Although the overlap between the circRNAs detected in cSCC and in normal skin was substantial (42%), many were unique, mainly expressed in the normal skin (Fig. 5C). To get an insight into their expression pattern, we plotted the expression value of these 264 circRNAs and observed a significant reduction in cSCC samples compared to healthy skin (Fig. 5D). This apparent downregulation in circRNA abundance prompted us to investigate whether circRNA-biogenesis could be altered in cSCC. Indeed, we observed that among the DEGs key negative regulator of circRNA-biogenesis ADAR is significantly upregulated (2.03-fold, FDR =  $4 \times 10^{-8}$ ) in cSCC while positive regulators such as MBNL (0.66-fold, FDR =  $4.2 \times 10^{-5}$ ) and ESRP1 (0.52-fold, FDR =  $1.2 \times 10^{-4}$ ) were downregulated



**Figure 5.** Detection of abundant circRNAs in healthy skin and cSCC. Pie charts showing the distribution of the numbers of exons annotated within the back-splicing junction of the high abundance circRNAs in healthy skin (A) and in cSCC (B). (C) Venn-diagram showing the overlap between the high abundance circRNAs detected in the normal skin biopsies (left) and the cSCC biopsies (right). (D) Column scatter plot showing the average RPM for the 264 unique high abundance circRNAs detected in the normal skin- and cSCC biopsies combined. Mann-Whitney  $U$  test.

(Supplementary Fig. S7) thereby providing a potential explanation for the large-scale decrease in circRNA abundance.

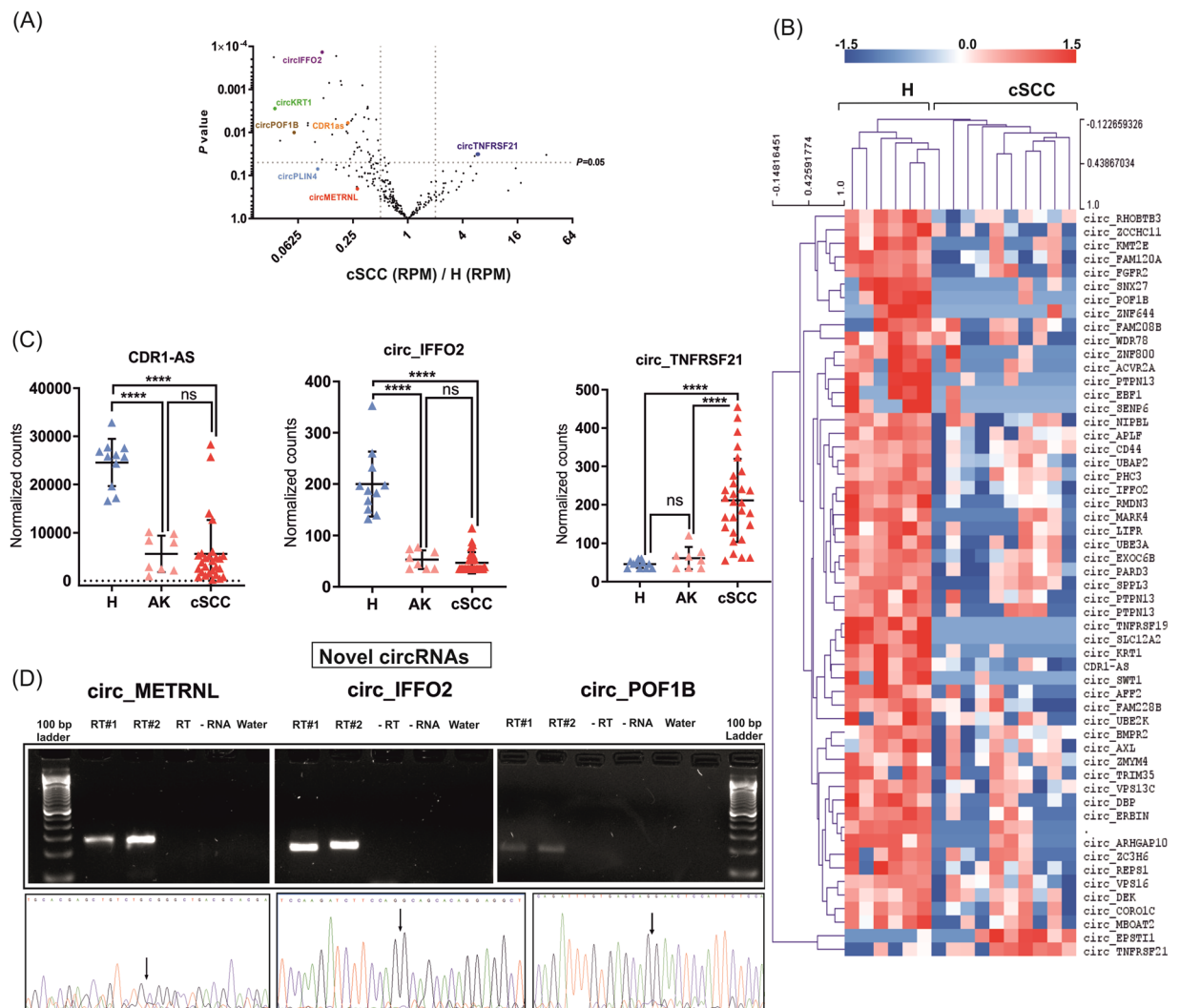
**Differentially expressed circRNAs are mostly downregulated in cSCC.** Differential expression analysis identified 55 circRNAs with significantly ( $P < 0.05$ ) altered expression in cSCC. In line with the generally lower abundance of circRNAs in cSCC (Fig. 5D) almost all differentially expressed circRNAs, 53/55 were downregulated and only two were upregulated in cSCC relative to healthy skin (Supplementary Table S8C and Fig. 6A). Hierarchical clustering of differentially expressed circRNAs clearly separated the samples into two groups, corresponding to healthy and cSCC samples (Fig. 6B). The most downregulated circRNAs included both well-characterized cancer-associated circRNAs, such as CDR1as, as well as novel circRNAs, which were derived from host genes *IFFO2*, *KRT1* and *POF1B*<sup>34</sup>. In terms of fold change, the most significantly downregulated circRNA was *IFFO2* (0.11-fold,  $P < 0.0001$ ) while the most upregulated was circ\_EPSTI with a 33-fold increase in expression ( $P < 0.03$ ; Fig. 6A).

To investigate whether the altered expression of circRNAs reflects the altered expression level of their host genes, fold changes in RPM were plotted against circular to linear (CTL) ratios (Supplementary Fig. S8). Comparison of linear and circular transcript produced from the same gene showed that the majority of the differentially expressed circRNAs were changed independently of their host genes.

**Validation of annotated and novel circRNAs in cSCC.** To validate the findings obtained from circRNA identification and differential expression analysis, next we analyzed the expression of three differentially expressed circRNAs. The expression of CDR1as, the most well-characterized member of the circRNA family, as well as two novel circRNAs identified in this study, circ\_ *IFFO2* and circ\_ *TNFRSF21*, in healthy skin (H,  $n = 11$ ), precancerous lesions (AK,  $n = 8$ ) and cSCC ( $n = 28$ ) was analyzed using NanoString assays (Fig. 6C) as this assay is particularly well-suited for circRNA quantification since no enzymes are used, which can introduce amplification bias<sup>35</sup>. In accordance with the results of RNA-seq, CDR1as and circ\_ *IFFO2* were found to be 6.7-fold and 5.3-fold downregulated ( $P < 0.0001$ ), while circ\_ *TNFRSF21* was 4.6-fold ( $P < 0.0001$ ) upregulated in cSCC compared to healthy skin (Fig. 6C). Interestingly, our validation data also suggested that downregulation of CDR1as and circ\_ *IFFO2* might be an early event in cSCC progression as they were downregulated in precancerous skin lesions 4.7 and 3.7-fold downregulation in AK, respectively,  $P < 0.0001$ ). Using divergent primer-based PCR amplification and Sanger sequencing, the back-spliced junction of three of the novel circRNAs detected in our analysis were confirmed (circ\_ *IFFO2*, circ\_ *METRNL* and circ\_ *POF1B*) (Fig. 6D, Supplementary Table S9, Fig. 9).

## Discussion

The molecular characterization of cSCC is an important step towards the understanding of dysregulated pathways and identifying key drivers of the disease. Although previous profiling experiments in cSCC have identified a large number of differentially expressed genes, the field has been hampered by the little consensus among these studies<sup>7,8,36</sup>. Here, we performed deep sequencing-based extensive analysis of transcriptomic changes in cSCC, which allowed us to identify differentially expressed protein-coding genes, including many, which were not detected in previous transcriptomic analyses, as well as a large number of lncRNAs and circular RNAs. In



**Figure 6.** Analysis of differentially expressed circRNAs in cSCC. **(A)** Volcano plot showing the fold changes in RPM vs  $P$ -values for the 264 unique high abundance circRNAs with the exception of circRNAs that were not expressed in either of the sample groups. **(B)** Heatmap and hierarchical clustering of all 55 differentially expressed circRNAs in cSCC. Red color represents upregulated and blue color represents downregulated circRNAs. **(C)** Validation of selected differentially expressed circRNAs (CDR-AS1, circ\_IFFO2 and circ\_TNFRSF21) by NanoString nCounter assay in healthy skin (H,  $n=11$ ), AK ( $n=8$ ) and cSCC ( $n=28$ ). Target gene expression is presented as background-corrected and normalized counts (threshold count of 35). \*\*\*\* $P < 0.0001$ , Mann-Whitney  $U$  test. **(D)** Validation of the novel circRNAs (circ\_METRNL, circ\_IFFO2 and circ\_POF1B) detected in our analysis by PCR using divergent primers and Sanger sequencing. The agarose gel image showed the expected size of PCR product present only in the reverse-transcribed samples. Below the agarose gel images, corresponding Sanger sequencing chromatograms across the backsplicing junction are shown. Arrows indicate the back-splicing junctions. Gel images for respective circRNA transcripts were cropped from the different parts of a single gel.

accordance with previous observation, we also identified a small overlap of commonly regulated differentially expressed genes in our dataset as compared to previous profiling studies (data not shown) which is possibly due to the heterogeneous nature of cSCC, along with variation in sampling and profiling methods. Interestingly, we identified a set of differentially expressed mRNAs and lncRNAs, which display skin-specific expression, suggesting function in skin development and epidermal differentiation and representing interesting targets for functional investigation.

The functional annotation of differentially expressed protein-coding genes identified enrichment of biological pathways related to altered barrier function, extracellular matrix assembly, decreased apoptosis and increased cell proliferation, indicating the breakdown of the differentiation program and associated cancer hallmarks, in line with the previous studies<sup>5</sup>. Interestingly, a large number of genes related to the inflammatory response and cytokine mediated signaling pathway were enriched among genes overexpressed in cSCC, including pro-angiogenic CXC chemokines such as CXCL1, CXCL10 and CXCL11 which could be involved in neo-angiogenesis<sup>37</sup>. Similarly, a number of CC chemokines (e.g. CCL2 and CCL5) with known oncogenic capacity



were significantly upregulated<sup>38</sup>. Not all chemokines were overexpressed in cSCC: RNA seq results confirmed the decreased expression of CCL27 in cSCC, a chemokine earlier identified to play a role in tumor immune escape<sup>39</sup>.

Analysis of oncogenic signatures revealed *EGFR* and *KRAS* activation, consistently with the key role of the EGFR/MAPK pathway in epidermal carcinogenesis<sup>5</sup>. We observed an upregulation of the key markers of progenitor-cell related genes, e.g. *ITGA6*, *ITGB1*, *CDKN1A* and *CDKN2A* in cSCC, suggesting an expansion of progenitor-cell like population<sup>40</sup>. Moreover, activation of NF-kappaB signaling was also identified, mirrored by the overexpression of p65 and upregulation of several direct targets (*ICAM1*, *NFKB2*, *NFKBIA*, *RELB*, *IRF1* and *BIRC3*) consistent with its proinflammatory and prosurvival role in cSCC<sup>23</sup>.

We identified a large number of transcription factors with altered expression in cSCC and thereby offer a potential mechanistic explanation for the expression changes we observed for thousands of mRNAs and lncRNAs. These differentially expressed transcription factors also represent a potential link between the genomic alterations in a relatively small number of driver genes in cSCC and the large-scale transcriptomic alterations in the disease. Many of the dysregulated transcription factors are related to skin development or skin carcinogenesis (e.g. *ETS1*, *SP1*, *TP63*, *TP53*, *AP1* and *TCF3*)<sup>19,22,41,42</sup>. Similarly, several known regulators (e.g. *KLF4*, *ZNF750*, *GRHL2*) of the proliferation-differentiation switch in the epidermis were also altered at the transcript level. Importantly, we observed altered expression of key regulators of pluripotency and stemness (e.g. *MYC*, *SOX2* and *OCT-3/4*), potentially revealing an activation of the stem-cell like program in cSCC, a feature associated with cellular transformation<sup>40</sup>. Increased expression of the key components of the JAK-STAT pathway (*STAT1* and *STAT3*) further corroborates the prominent cytokine activation signature observed in the GO analysis.

In addition to the transcription factors with established roles in the disease, several less characterized transcription factors were predicted to regulate a large number of the genes in cSCC. An example is *FOXM1* that has been shown to regulate cell proliferation and senescence in healthy keratinocytes and implicated in cell invasion and metastasis in head and neck squamous cell carcinoma, a disease resembling cSCC in several aspects<sup>43</sup>. Another less well-characterized TF whose expression was increased in cSCC was *USF1*, a leucine-zipper-family TF, which has been described to regulate p53 stability<sup>44</sup>. However, its role in cSCC has not been investigated yet.

In terms of tissue specificity, although protein-coding genes were largely non-specific (85%, Supplementary Table S7), a number of skin-specific transcripts were also identified (0.23%). As expected, majority of them were downregulated in cSCC, with many of them functioning as key players in epidermal differentiation (*LCE1D*, *FLG*, *KRT77*, *KRT10*, *ALOEX3* etc.). In addition to the expected downregulation of terminal differentiation-associated transcripts, this meta-analysis also revealed that *IL-37*, a negative regulator of innate immunity with hitherto unexplored function in cSCC, has a strong skin-specific expression<sup>45</sup>. Future studies should address their functional role or prognostic value in cSCC.

Although many lncRNAs display highly lineage and cell type-specific expression pattern, we observed an increased expression of several lncRNAs dysregulated in a broad range of solid tumors, such as *MIR31HG*, which regulates senescence, as well as *CYTOR* (*LINC00152*), *PVT1*, *HOXA-AS2* and *SNHG12*, which can regulate various cancer hallmarks<sup>25,28,46,47</sup>. Moreover several lncRNAs, reported to be upregulated in the primary and metastatic cSCC cell lines compared to progenitor keratinocytes<sup>48</sup> (e.g. *LINC00346*, *PVT1*, *SNHG12*, *ZFAS1* and *TUG1*) were significantly upregulated in cSCC, suggesting that some of the disease-related molecular alterations are comparable between cell lines and primary cells. Comparison of the differentially expressed lncRNAs with a list of altered lncRNAs in cSCC from a previous microarray-based study with a smaller sample size ( $n = 3$ ), revealed a set of 51 upregulated (e.g. *PVT1*, *HIF1A-AS2*, *HOXD-AS2*, *CYTOR*) and 73 downregulated lncRNAs (e.g. *SNRK-AS1*, *BDNF-AS*, *GUSBP11*) consistently detected across the two studies<sup>49</sup>, representing a common set of cSCC-associated lncRNAs.

We identified a set of skin-enriched lncRNAs (e.g. *RP13-455A7.1*, *RP11-73G16.2*, *CHODL-AS*) with dysregulated expression in cSCCs. Since tissue-restricted lncRNAs often play crucial homeostatic function, the identified skin-enriched lncRNAs with altered expression in cSCCs, may also have important function in skin development and homeostasis. In this list, we also observed the presence of *TINCR* – a lncRNA which is required for proper epidermal differentiation<sup>50</sup>.

Our circRNA analysis detected 227 and 150 high confidence circRNAs in healthy skin and cSCCs, respectively. In addition to the near-ubiquitously expressed circRNAs such as *CDR1as* and *circSMARCA5*, we observed several circRNAs (*circ\_PARD3*, *circ\_APLF* and *circ\_MAP3K1*) with high expression in the skin<sup>51,52</sup>. Apart from annotated circRNAs, we identified novel candidates derived from *IFFO2*, *METRNL*, *KRT1* and *POF1B* genes, which with the exception of *circ\_KRT1*, we validated. Two of the host genes of these novel circRNAs- *POF1B* and *IFFO2*, are highly expressed in skin and esophagus tissues with stratified epithelia. Therefore, it would be interesting to study whether the circular counterparts also have such tissue-restricted expression and function.

We observed a general reduction of circRNA abundance in cSCC compared to normal human skin. One possibility for the overall decrease in circRNAs can be the dysregulation of their biogenesis. Indeed, we found that *ADAR*, a double-stranded RNA-binding protein that has been previously shown to suppress back-splicing events, was upregulated in the cSCC samples<sup>53</sup>. Involvement of pre-mRNA processing machinery led us to perform an intersection analysis with known splicing regulators where we found downregulation of *ESRP1* and *MBNL1* - two previously reported positive regulators of circRNA biogenesis<sup>54</sup>. Alternatively, it is also possible that the observed downregulation of circRNAs is caused by their high stability and inefficient biogenesis preventing them from reaching steady-state levels in the highly proliferative cancer cells<sup>55</sup>.

One of the most significantly downregulated circRNAs was *CDR1as* which has been previously shown to act by inhibiting tumor suppressor miR-7<sup>51</sup> and also can be negatively regulated by the same miRNA cellular context<sup>56</sup>. Since *CDR1as* has been previously shown to modulate *EGFR* and *KLF4* in breast cancer development, and these genes play important roles also in the skin, it will be interesting to assess its potential tumor suppressive effect in the context of cSCC. One of the most differentially expressed circRNAs- *circ\_EPSTI1* was highly overexpressed (32-fold) in cSCC. Its host gene (*EPSTI1*) is an IFN-response gene which has been shown to promote tumor

invasion and metastasis in various cancer types<sup>57</sup>. It is tempting to investigate whether the circRNA also have a function to synergize or antagonize with the host gene. Upregulation of circ\_TNFRSF21 seems to be co-regulated with the expression of its host gene in cSCC. In addition to CDR1as, two other downregulated circRNAs, circ\_UBAP2 and circ\_SNX27 have been previously demonstrated to have oncogenic effects in other cancer types<sup>58,59</sup>.

Our analysis showed that a number of circRNAs related to epidermal differentiation (circ\_MBOAT2, circ\_PTPN13 and circ\_ACVR2A) were decreased in cSCC (32). Since altered epidermal differentiation program is an early event during the development and progression of cSCC, these circRNAs are promising candidates for functional characterization<sup>32</sup>. Comparison of results from our circRNA-analysis with that of a hybridization array-based circRNA profiling study with smaller sample size ( $n = 3$ ) in cSCC revealed only one common deregulated circRNA (circ\_ACVR2)<sup>60</sup>. This might be due to our stringent filtering criteria (more than five backsplicing junctions), large difference in the sample number and the different detection methods, as array-based circRNA profiling does not have the potential to identify novel circRNAs.

In summary, our study provides a comprehensive dataset of deregulated protein-coding genes, along with linear and circular non-coding RNAs in cSCC. The findings entail the need for a future investigation to dissect the changes in lncRNAs and circRNA landscape during various stages of cSCC progression including precancerous lesions, cSCC *in situ* and different grades of cSCC. The present study identifies previously uncharacterized lncRNAs and circRNAs with altered expression, which could be utilized to identify novel therapeutic targets in cSCC and will contribute to a better understanding of the molecular pathogenesis of cSCC.

## Methods

**RNA isolation from patient samples and RNA integrity assessment.** Four mm punch biopsies were collected from 18 healthy donors and 28 patients with primary cSCC, at the department of Dermatology, Karolinska University Hospital, Stockholm, Sweden and the department of Dermatology and Allergology, Faculty of Medicine, University of Szeged, Szeged, Hungary. All the actinic keratosis samples included in the validation cohort were collected at the department of Dermatology, University Hospital Düsseldorf. All patients and healthy volunteers provided written informed consent for sample collection. Details of the samples used for whole transcriptome sequencing can be found in Supplementary Table S1 and Supplementary Table S10 (for extended cohort). Snap frozen punch biopsies were homogenized using TissueLyser LT (Qiagen) followed by RNA extraction using the miRNeasy mini kit (Qiagen). RNA concentrations were measured using a Nanodrop One (Thermo Fisher Scientific) instrument and RNA integrity was assessed on the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) using the Nano 6000 kit.

**Evaluation of tissue specificity.** Expression value of protein coding and non-coding transcripts were obtained from the GTEX portal. Transformed fibroblasts and lymphocytes were removed from the list, as they did not represent any single tissue. For adipose tissue, brain, heart and artery, average RPKM values were calculated as an average of the different components of the same tissue/organ. Similarly, average expression value was calculated from sun-exposed and non-exposed skin samples. The tissue specificity ( $T$ ) of a gene in tissue was calculated as the fraction of expression (RPKM) relative to the sum of its expression in all 28 tissues.

**NanoString nCounter assay.** A custom CodeSet of capture and reporter probes was designed to target a unique 100 nucleotide long exon spanning regions in mRNAs, lncRNAs and backsplicing junctions of differentially expressed circRNAs. Five reference genes (RPLP0, PUM1, SF3A1, GUSB and ALAS1) were included for normalization. Probe-set and RNA hybridization reaction were performed according to the manufacturer's instructions; complexes were purified, immobilized on nCounter cartridges and quantified using the digital analyzer. Mean of the negative controls were subtracted while analyzing the data using nSOLVER 3.0 software (NanoString Technologies, Seattle, WA, USA). Positive control normalization was performed using the geometric mean of all positive controls with the exception of the control named E, as recommended by the manufacturer. Finally, a second normalization was performed using the geometric mean of the three linear reference genes (ALAS1, PUM1 and SF3B1) having the lowest coefficient of variance percentage (%CV).

Further details are available in Supplementary Materials Section.

**Ethics approval and consent to participate.** The study was approved by the Regional Committee of Health Research Ethics (Stockholm, Sweden, Szeged, Hungary and Düsseldorf, Germany), and the study was performed in accordance with the Declaration of Helsinki. All patients and healthy volunteers provided written informed consent.

## Data availability

All data generated and analyzed in this study are included in this article and in the corresponding Supplementary Files. Normalized read counts, raw data and relevant metadata files are deposited to NCBI's Gene expression Omnibus (Accession number: GSE139505).

Received: 1 May 2019; Accepted: 31 January 2020;

Published online: 27 February 2020

## References

1. Karia, P. S., Han, J. & Schmults, C. D. Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012. *J. Am. Acad. Dermatol.* **68**, 957–966 (2013).
2. Alam, M. & Ratner, D. Cutaneous squamous-cell carcinoma. *N. Engl. J. Med.* **344**, 975–983 (2001).
3. Que, S. K. T., Zwald, F. O. & Schmults, C. D. Cutaneous squamous cell carcinoma: Incidence, risk factors, diagnosis, and staging. *J. Am. Acad. Dermatology* **78**, 237–247 (2018).

4. Stratigos, A. *et al.* Diagnosis and treatment of invasive squamous cell carcinoma of the skin: European consensus-based interdisciplinary guideline. *Eur. J. Cancer* **51**, 1989–2007 (2015).
5. Ratushny, V., Gober, M. D., Hick, R., Ridky, T. W. & Seykora, J. T. From keratinocyte to cancer: the pathogenesis and modeling of cutaneous squamous cell carcinoma. *J. Clin. Invest.* **122**, 464–472 (2012).
6. Li, Y. Y. *et al.* Genomic Analysis of Metastatic Cutaneous Squamous Cell Carcinoma. *Clin. Cancer Res.* **21**, 1447–1456 (2015).
7. Chitsazzadeh, V. *et al.* Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates. *Nature Communications* **7** (2016).
8. Hametman, L. *et al.* Molecular profiling of cutaneous squamous cell carcinomas and actinic keratoses from organ transplant recipients. *BMC Cancer* **13**, 58 (2013).
9. 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).
10. Mercer, T. R., Dinger, M. E., Sunken, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. USA* **105**, 716–721 (2008).
11. Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253–1261 (2015).
12. Xia, S. *et al.* Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief. Bioinforma.* **18**, 984–992 (2017).
13. Holdt, L. M., Kohlmaier, A. & Teupser, D. Molecular roles and function of circular RNAs in eukaryotic cells. *Cell. Mol. Life Sci.* **75**, 1071–1098 (2018).
14. Kristensen, L. S., Hansen, T. B., Venø, M. T. & Kjems, J. Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene* **37**, 555–565 (2018).
15. Xu, Y., Yan, W. & Chen, X. SNF5, a core component of the Swi/Snf complex, is necessary for p53 expression and cell survival in part via eIF4E. *Oncogene* **29**, 4090–4100 (2010).
16. Nagarajan, P. *et al.* Ets1 blocks terminal differentiation of keratinocytes and induces expression of matrix metalloproteases and innate immune mediators. *J. Cell Sci.* **123**, 3566–3575 (2010).
17. Lena, A. M. *et al.* miR-203 represses 'stemness' by repressing DeltaNp63. *Cell Death Differ.* **15**, 1187–1195 (2008).
18. Sen, G. L. *et al.* ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. *Dev. Cell* **22**, 669–677 (2012).
19. Berg, R. J. *et al.* Early p53 alterations in mouse skin carcinogenesis by UVB radiation: immunohistochemical detection of mutant p53 protein in clusters of preneoplastic epidermal cells. *Proc. Natl. Acad. Sci. USA* **93**, 274–278 (1996).
20. Pierce, A. M. *et al.* Increased E2F1 activity induces skin tumors in mice heterozygous and nullizygous for p53. *PNAS* **95**, 8858–8863 (1998).
21. Alvarez-Salas, L. M., Benitez-Hess, M. L. & Dipaolo, J. A. YY-1 and c-Jun transcription factors participate in the repression of the human involucrin promoter. *Int. J. Oncol.* **26**, 259–266 (2005).
22. Zhang, X., Wu, J., Luo, S., Lechler, T. & Zhang, J. Y. FRA1 promotes squamous cell carcinoma growth and metastasis through distinct AKT and c-Jun dependent mechanisms. *Oncotarget* **7**, 34371–34383 (2016).
23. Kim, C. & Pasparakis, M. Epidermal p65/NF- $\kappa$ B signalling is essential for skin carcinogenesis. *EMBO Mol. Med.* **6**, 970–983 (2014).
24. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes. Dev.* **25**, 1915–1927 (2011).
25. Jiang, B. *et al.* Identification of oncogenic long noncoding RNA SNHG12 and DUXAP8 in human bladder cancer through a comprehensive profiling analysis. *Biomedicine Pharmacotherapy* **108**, 500–507 (2018).
26. Xiao, H. *et al.* Long non-coding RNA Lucat1 is a poor prognostic factor and demonstrates malignant biological behavior in clear cell renal cell carcinoma. *Oncotarget* **8**, 113622–113634 (2017).
27. Pan, Z. *et al.* The long noncoding RNA CASC9 regulates migration and invasion in esophageal cancer. *Cancer Med.* **5**, 2442–2447 (2016).
28. Tseng, Y.-Y. *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nat.* **512**, 82–86 (2014).
29. Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Sci.* **348**, 648–660 (2015).
30. Rozenberg, J. M., Bhattacharya, P., Chatterjee, R., Glass, K. & Vinson, C. Combinatorial Recruitment of CREB, C/EBP $\beta$  and c-Jun Determines Activation of Promoters upon Keratinocyte Differentiation. *PLOS ONE* **8**, e78179 (2013).
31. Nakamura, Y. *et al.* The combination of ubiquitous transcription factors AP-1 and Sp1 directs keratinocyte-specific and differentiation-specific gene expression *in vitro*. *Exp. Dermatol.* **16**, 143–150 (2007).
32. Kristensen, L. S., Okholm, T. L. H., Venø, M. T. & Kjems, J. Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. *RNA Biol.* **15**, 280–291 (2018).
33. Stagsted, L. V., Nielsen, K. M., Daugaard, I. & Hansen, T. B. Noncoding AUG circRNAs constitute abundant and conserved subclass of circles. *Life Sci Alliance* **2**, (2019).
34. Hansen, T. B., Kjems, J. & Damgaard, C. K. Circular RNA and miR-7 in Cancer. *Cancer Res.* **73**, 5609–5612 (2013).
35. Dahl, M. *et al.* Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Lab. Invest.* **98**, 1657–1669 (2018).
36. Haren, R. V., Feldman, D. & Sinha, A. A. Systematic comparison of nonmelanoma skin cancer microarray datasets reveals lack of consensus genes. *Br. J. Dermatology* **161**, 1278–1287 (2009).
37. Keeley, E. C., Mehrad, B. & Strieter, R. M. CXC Chemokines in Cancer Angiogenesis and Metastases. *Adv. Cancer Res.* **106**, 91–111 (2010).
38. Chow, M. T. & Luster, A. D. Chemokines in Cancer. *Cancer Immunol. Res.* **2**, 1125–1131 (2014).
39. Pivarcsi, A. *et al.* Tumor immune escape by the loss of homeostatic chemokine expression. *PNAS* **104**, 19055–19060 (2007).
40. Jian, Z., Strait, A., Jimeno, A. & Wang, X.-J. Cancer Stem Cells in Squamous Cell Carcinoma. *J. Invest. Dermatol.* **137**, 31–37 (2017).
41. Nagarajan, P., Parikh, N., Garrett-Sinha, L. A. & Sinha, S. Ets1 Induces Dysplastic Changes When Expressed in Terminally-Differentiating Squamous Epidermal Cells. *PLoS One* **4** (2009).
42. Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* **127**, 171–183 (2006).
43. Tanaka, N. *et al.* Gain-of-function mutant p53 promotes the oncogenic potential of head and neck squamous cell carcinoma cells by targeting the transcription factors FOXO3a and FOXM1. *Oncogene* **37**, 1279 (2018).
44. Bouafia, A. *et al.* p53 Requires the Stress Sensor USF1 to Direct Appropriate Cell Fate Decision. *PLoS Genet* **10** (2014).
45. Nold, M. F. *et al.* IL-37 is a fundamental inhibitor of innate immunity. *Nat. Immunology* **11**, 1014–1022 (2010).
46. Montes, M. *et al.* The lncRNA MIR31HG regulates p16 INK4A expression to modulate senescence. *Nat. Commun.* **6**, 1–15 (2015).
47. Wang, X. *et al.* The long non-coding RNA CYTOR drives colorectal cancer progression by interacting with NCL and Sam68. *Mol. Cancer* **17**, 110 (2018).
48. 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).
49. Sand, M. *et al.* Expression profiles of long noncoding RNAs in cutaneous squamous cell carcinoma. *Epigenomics* **8**, 501–518 (2016).
50. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nat.* **493**, 231–235 (2013).

51. Weng, W. *et al.* Circular RNA ciRS-7-A Promising Prognostic Biomarker and a Potential Therapeutic Target in Colorectal Cancer. *Clin. Cancer Res.* **23**, 3918–3928 (2017).
52. Okholm, T. L. H. *et al.* Circular RNA expression is abundant and correlated to aggressiveness in early-stage bladder cancer. *NPJ Genom. Med.* **2**, 36 (2017).
53. Shi, L. *et al.* Circular RNA expression is suppressed by androgen receptor (AR)-regulated adenosine deaminase that acts on RNA (ADAR1) in human hepatocellular carcinoma. *Cell Death Dis.* **8**, e3171 (2017).
54. Li, M. *et al.* Biogenesis of circular RNAs and their roles in cardiovascular development and pathology. *FEBS J.* **285**, 220–232 (2018).
55. Bachmayr-Heyda, A. *et al.* Correlation of circular RNA abundance with proliferation – exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci. Rep.* **5**, 8057 (2015).
56. Kleaveland, B., Shi, C. Y., Stefano, J. & Bartel, D. P. A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell* **174**, 350–362.e17 (2018).
57. de Neergaard, M. *et al.* Epithelial-Stromal Interaction 1 (EPSTI1) Substitutes for Peritumoral Fibroblasts in the Tumor Microenvironment. *Am. J. Pathol.* **176**, 1229–1240 (2010).
58. Zhang, H. *et al.* Increased circular RNA UBAP2 acts as a sponge of miR-143 to promote osteosarcoma progression. *Oncotarget* **8**, 61687–61697 (2017).
59. Huang, X.-Y. *et al.* Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-100338/miR-141-3p pathway in hepatitis B-related hepatocellular carcinoma. *Sci Rep* **7** (2017).
60. Sand, M. *et al.* Circular RNA expression in cutaneous squamous cell carcinoma. *J. Dermatol. Sci.* **83**, 210–218 (2016).

## Acknowledgements

We express our gratitude to all the patients and control subjects who contributed to this study. The study was supported by the Swedish Research Council (VR Dnr. 2015-02844), the Swedish Cancer Society (CAN 2015/694 and CAN 2018/801), the Welander and Finsens Foundation (Hudfonden), the Swedish Society of Medicine (Svenska Läkaresällskapet). Open access funding provided by Karolinska Institute.

## Author contributions

J.L., I.N., E.B., L.K., E.S., B.H. and K.D.M. collected samples. K.D.M., L.M. and L.S.K. performed experiments. K.D.M., J.N.S., L.P. L.S.K. and A.P. analyzed the data. A.P., L.S.K., E.S. and K.D.M. planned the experiments. K.D.M., L.P., L.S.K., J.K., C.K., J.N.S., E.S. and A.P. contributed in writing the article.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-59660-6>.

**Correspondence** and requests for materials should be addressed to A.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020