Society of Medical Education & Research

2025, Volume 5, Issue 1, Page No: 161-178
ISSN: 3108-4834

Archive of International Journal of Cancer and Allied Science

Alternative 5′UTR Isoforms Drive Enhanced Translational Efficiency via TOP/PRTE Motif Switches in Squamous Cell Carcinoma

Maria Teresa Fonseca[1]*, Joana Cristina Matos[1], Ricardo Manuel Alves[1]

[1]Oncology Unit, Centro Hospitalar de Lisboa Norte, Lisbon, Portugal.

***E-mail** ✉ m.fonseca.chln@outlook.com

Abstract

The regulation of gene expression relies heavily on both transcriptional and translational control, yet how these two layers are coordinated remains largely unresolved. In this work, we applied Nanopore long-read sequencing together with cap analysis of gene expression (CAGE-seq) to chart the repertoire of 5′ and 3′ untranslated region (UTR) isoforms and transcription start sites (TSSs) in epidermal stem cells, normal keratinocytes, and squamous cell carcinoma cells. In squamous cell carcinomas, we observed that a select group of genes possessing alternative 5′UTR isoforms displayed markedly higher translational efficiency and were particularly enriched in ribosomal proteins and splicing-related factors. By integrating polysome fractionation with CAGE-seq, we investigated two of these genes sharing identical coding sequences and found that variability in TSS usage frequently resulted in switches between 5′ terminal oligopyrimidine (TOP) and pyrimidine-rich translational element (PRTE) motifs, promoting mTORC1-dependent translation. At a genome-wide scale, transcripts with higher translation in squamous cell carcinoma preferentially used 5′TOP and PRTE motifs, had shorter 5′UTRs, and showed reduced RNA secondary structure. Importantly, only the two RPL21 isoforms containing 5′TOP motifs, not the TOP-lacking versions, significantly correlated with patient overall survival in head and neck squamous cell carcinoma. These findings highlight the need for isoform-specific analysis in cancer studies and indicate that switching 5′UTR isoforms offers a straightforward mechanism to adjust protein synthesis rates, tune mTORC1 responsiveness, and regulate the translational potential of mRNAs through their 5′UTR sequence.

Keywords: Keratinocytes, RNA, Isoform-specific analysis, Cancer

Introduction

Precise spatial and temporal control of gene expression is fundamental for proper cellular function. Although transcription and translation are recognized as major regulatory points, their coordinated influence on gene expression remains incompletely defined. Recent evidence in yeast demonstrates a non-canonical regulatory strategy whereby numerous genes alternate between short and long 5′UTRs while maintaining identical coding sequences [1-4]. Because 5′UTRs are critical for ribosome recruitment and translation initiation [5], these alternative isoforms can lead to substantial differences in translational efficiency. For instance, during yeast meiosis, roughly 8% of genes were regulated via extended 5′UTRs, which were inefficiently translated due to upstream open reading frames (uORFs) [2]. These examples illustrate how precise 5′UTR composition can directly control protein output, emphasizing the importance of accurately mapping transcription start sites and 5′UTR variants.

The 5′UTR sequence also interfaces with the mammalian target of rapamycin complex 1 (mTORC1) nutrient-sensing pathway, which selectively enhances translation of mRNAs containing 5′TOP and PRTE motifs [6, 7]. Dysregulated mTORC1 activity is frequently observed in cancers and drives tumor progression [8]. Translational reprogramming under mTORC1 control is considered a

Access this article online https://smerpub.com/

Received: 03 December 2024; **Accepted:** 16 March 2025

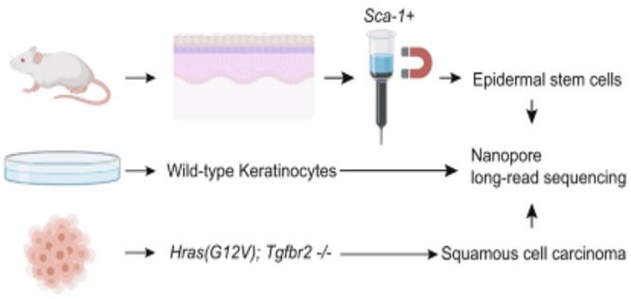
How to cite this article: Fonseca MT, Matos JC, Alves RM. Alternative 5′UTR Isoforms Drive Enhanced Translational Efficiency via TOP/PRTE Motif Switches in Squamous Cell Carcinoma. Arch Int J Cancer Allied Sci. 2025;5(1):161-78. https://doi.org/10.51847/mnH8oA90dc

key process in cancer development [6, 9], making it essential to identify which mRNAs mediate this pathway. Here, we applied Nanopore long-read sequencing together with CAGE-seq to profile 5′ and 3′UTR isoforms in vivo in epidermal stem cells, wild-type keratinocytes, and cultured squamous cell carcinoma cells (SCCc). The combined dataset and de novo transcriptome annotations are provided via an accessible genome browser (Materials & Methods). By combining polysome profiling with CAGE-seq, we quantified the translational efficiency of individual 5′ isoforms, showing that highly translated SCC transcripts favor 5′UTR motifs that drive mTORC1-dependent translation. Remarkably, only the two RPL21 isoforms containing TOP motifs, not the TOP-lacking isoforms, strongly stratified overall survival in head and neck squamous cell carcinoma patients, with survival differences up to 3.5 years.

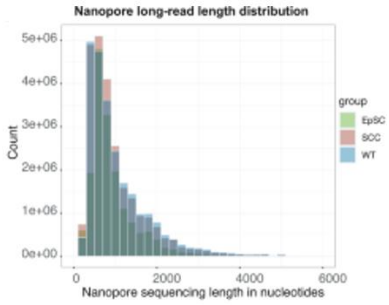
These findings reveal the intricacy of TSS selection in mammals, demonstrating that minor differences in TSS location and 5′UTR isoform can profoundly influence translational output. Given the widespread heterogeneity of promoters, reliance solely on reference transcripts may miss crucial regulatory variation. Switching between TOP/PRTE and non-TOP/PRTE 5′UTRs provides a robust strategy to regulate translation of cancer-associated genes. More broadly, alternative TSS usage [10] and splicing events [11] suggest that post-transcriptional elements such as RNA-binding protein sites, uORFs [2, 3], and 5′TOP motifs may be extensively exploited in cancer to coordinate transcription with translation, thereby determining mRNA translational potential based on precise 5′ transcript sequence.

Results and Discussion

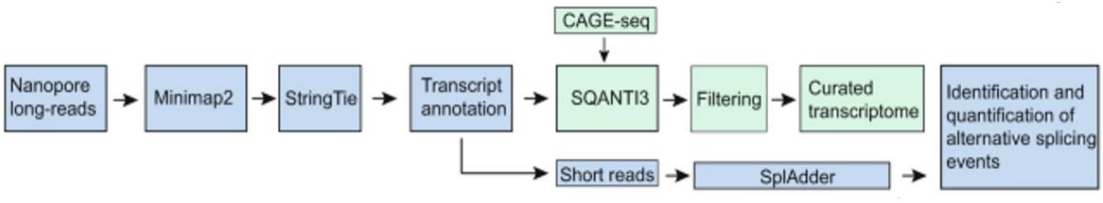
To comprehensively capture full-length transcriptomes, we performed Nanopore long-read RNA sequencing on mouse epidermal stem cells in vivo and cultured squamous cell carcinoma cells (SCCc), representing distinct biological contexts. Epidermal stem cells (EpSCs) were isolated from P60 (postnatal day 60) mice via rapid magnetic-activated cell sorting using anti-stem cell antigen-1 microbeads (Sca-1, Ly6a gene) [12, 13]. SCCc were derived from an established tumor allograft model combining oncogenic HrasG12V expression with TGFβ receptor II loss, which forms invasive squamous cell carcinomas when transplanted into immunocompromised mice [14]. Wild-type keratinocytes served as a reference **(Figure 1a).**



a)



b)



c)

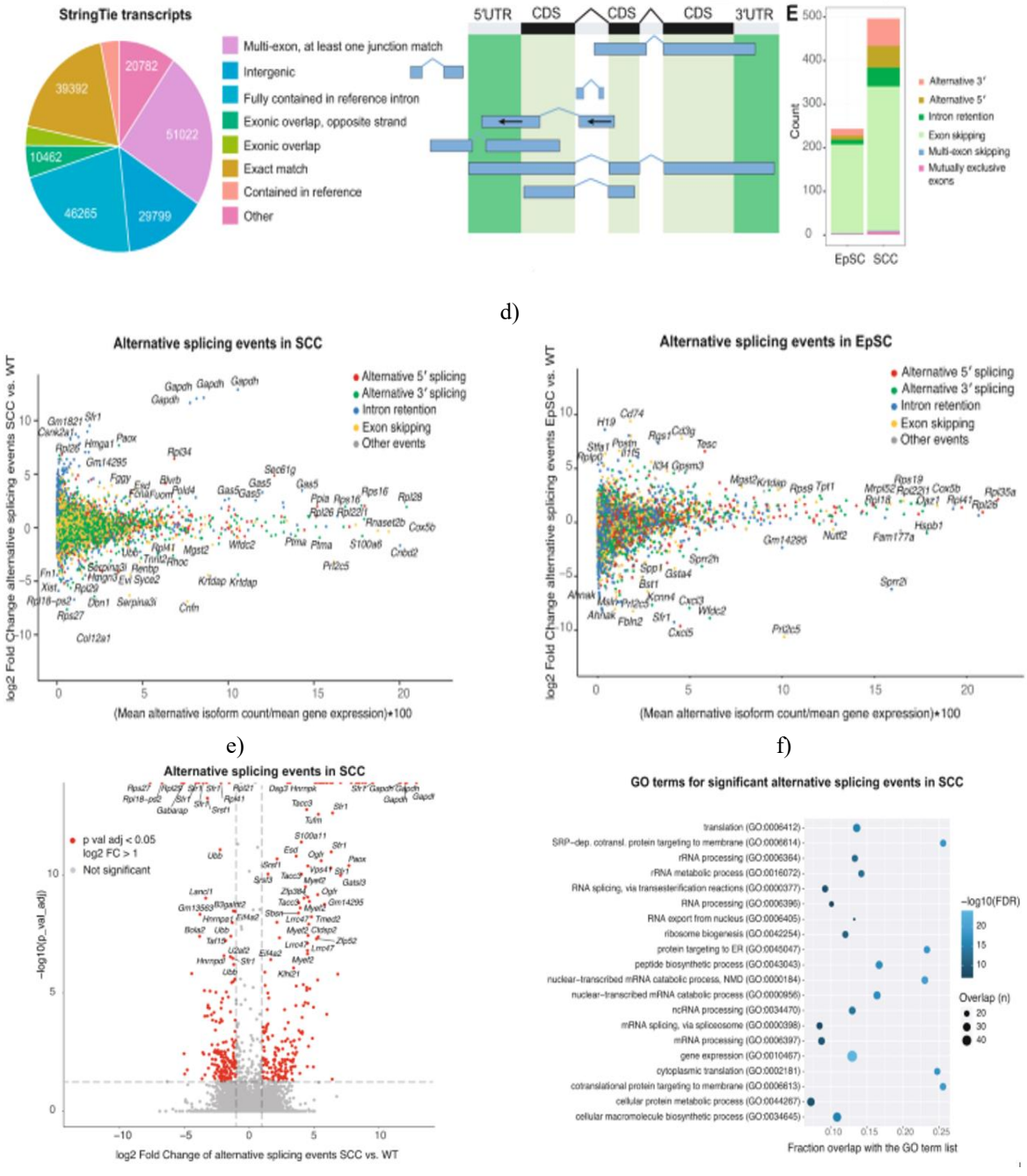


Figure 1. Mapping alternative mRNA isoforms in mouse skin via Nanopore long-read sequencing

a Schematic of the experimental workflow used to isolate SCA-1+ epidermal stem cells (EpSCs), wild-type keratinocytes (WT), and cultured squamous cell carcinoma cells (SCCc) for long-read sequencing.

b Length distribution of Nanopore sequencing reads for EpSCs, WT keratinocytes, and SCCc. Average read lengths ranged from 943 to 1035 bp (EpSC: 943 bp; WT: 1035 bp; SCCc: 994 bp).

c Analysis pipeline for identifying and quantifying alternative transcript isoforms. Nanopore reads were aligned to the mouse genome, and transcripts were reconstructed with StringTie. SQANTI3 was applied for transcript classification and quality filtering, followed by

validation with CAGE-seq TSSs to generate a high-confidence transcriptome. SplAdder was then used alongside short-read RNA-seq data to detect and quantify alternative splicing events.

d Overview of the total number of transcripts identified by StringTie in EpSCs, WT keratinocytes, and SCCc prior to SQANTI3 and CAGE-seq filtering. The left panel shows transcript counts per category, and the right panel illustrates representative examples of the classification.

e Count and type of significantly altered splicing events in EpSCs and SCCc relative to WT keratinocytes, identified by SplAdder.

f, g Distribution of alternative splicing events in SCCc (F) or EpSCs (G) versus WT keratinocytes. Color codes represent alternative 5′ or 3′ splice sites, exon skipping, and intron retention. The x-axis indicates isoform proportion relative to total gene expression, and the y-axis shows fold change of the splicing event.

h Volcano plot displaying fold changes for significantly altered splicing events in SCCc compared with WT keratinocytes. Red dots highlight events with >2× fold change.

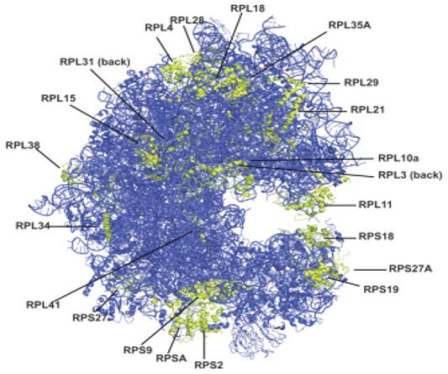
i Gene ontology (GO:0010467) enrichment for genes with significant splicing changes in SCCc, highlighting ribosomal proteins and splicing factors. The top 20 GO terms are shown alphabetically; circle size represents the number of overlapping genes, x-axis shows the fraction of overlap, and blue tones indicate FDR.

Across all samples, 158.1 million reads were generated, with mean lengths of 943–1035 bp (**Figure 1b**). De novo transcriptome assembly with StringTie [15] reconstructed 210,630 transcripts, including 39,392 perfectly matching the reference genome intron chain (**Figures 1c and 1d**). Around 62% of genes produced a single transcript, 17% produced two or three, and the remaining 21% had four or more isoforms (**Figure 1a**).

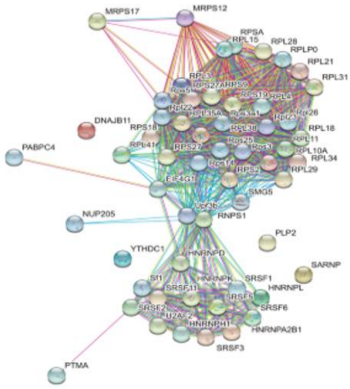
Applying SQANTI3 for classification and filtering yielded 149,434 high-confidence transcripts (1 QC in Materials & Methods). Further filtering using CAGE-seq data resulted in a curated set of 77,017 isoforms **(Figures 1c and 1b).**

Using the SplAdder toolbox [16], we quantified alternative isoforms and their differential expression in EpSCs and SCCc relative to WT **(Figures 1e–h).** At stringent confidence thresholds, 243 and 496 significantly altered splicing events were detected in EpSCs and SCCc, respectively. The majority involved 5′ and 3′ alternative splice sites, exon skipping, or intron retention **(Figure 1e).** In EpSCs, exon skipping predominated, whereas SCCc additionally showed 50 alternative 5′ and 63 alternative 3′ splice events. Alternative splicing events occurred across a wide range of relative isoform abundances compared to total gene expression (**Figures 1f and 1g**).

Genes with significant splicing changes in SCCc were enriched for ribosomal proteins and splicing factors **(Figures 1h and 1i),** including 23 genes encoding 40S and 60S ribosomal subunits (**Figures 2a, b1c and d**). Many of these proteins are located on the ribosome surface, including Rpl38, which has been linked to selective translation of specific mRNA subsets [17]. The 5′UTRs of these 23 ribosomal proteins exhibited higher GC content and more structured secondary elements than other ribosomal proteins (**Figure 1i**). The total number of ribosomal protein isoforms increased from 264 in WT to 290 in SCCc. Collectively, these analyses provide a comprehensive view of alternative transcript isoforms and suggest that a subset of translation-related genes undergoes differential splicing in SCCc.



a)



b)

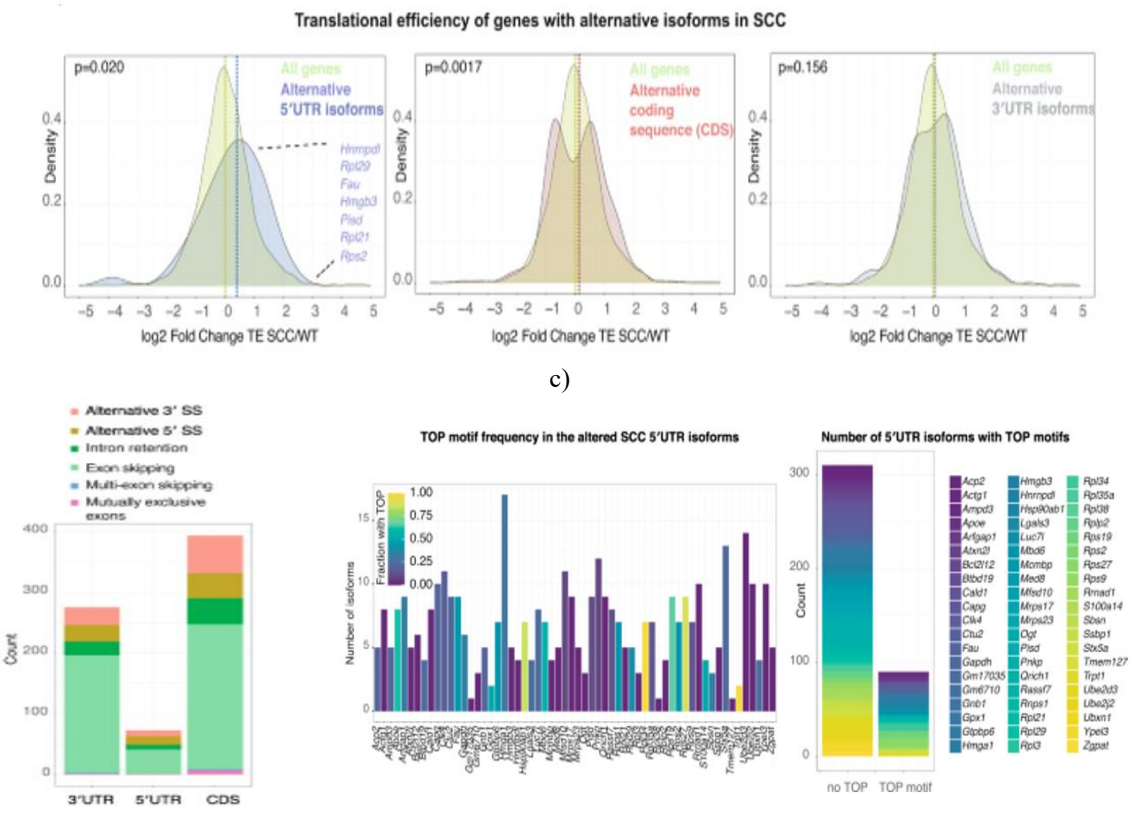


Figure 2. Alternative 5′UTR isoforms correlate with enhanced translational efficiency in squamous cell carcinomas

a Mapping of ribosomal proteins on the human 80S ribosome, highlighting those whose mRNAs exhibit significant alternative splicing in SCCc. Note that RPLP0 and RPLP2 were not represented in the structural model.
b STRING network analysis depicting protein–protein interactions among alternatively spliced genes associated with the GO term "gene expression" (GO:0010467) in SCCc relative to WT keratinocytes.
c Translational efficiency (TE) of genes with significant alternative splicing events in the 5′UTR is elevated in SCCc. TE was calculated as the ratio of ribosome profiling reads to RNA-seq reads and analyzed using the LRT test in DESeq2. Fold-change comparisons were made for genes with alternative splicing in the 5′UTR, CDS, or 3′UTR versus all genes. P-values derive from a two-sample Kolmogorov–Smirnov test comparing TE distributions.
d Counts and categories of alternative splicing events in SCCc affecting the 5′UTR, CDS, or 3′UTR. SS indicates splice sites.
e Most SCCc genes with alternative 5′UTR isoforms express transcripts that either include or exclude 5′TOP motifs. TOP motifs were defined by a +1 cytidine followed by 4–16 consecutive pyrimidines. Left panel: fraction of StringTie-derived 5′UTR isoforms containing a TOP motif for each gene (0–1, corresponding to 0–100%). Right panel: number of isoforms with or without TOP motifs in the cohort of SCCc genes with alternative 5′UTRs.

Given the critical role of the 5′UTR in regulating translation [5], we examined how alternative 5′UTR isoforms in SCCc impact translational efficiency. Among significant alternative splicing events, 73 altered 5′UTRs and 275 affected 3′UTRs. Of the 23 ribosomal proteins with differential splicing, 12 showed alternative 5′UTR isoforms. Most splicing events involved alternative 5′ or 3′ splice sites, intron retention, or exon skipping (**Figure 2d**).

We used previously published ribosome profiling and RNA-seq data from WT keratinocytes and HrasG12V; Tgfbr2-null SCCc to calculate TE as a proxy for protein

output per mRNA [18]. TE was assessed at the gene level, noting that low-abundance isoforms might not be captured in overall TE shifts. Splicing changes in 3′UTRs did not significantly alter gene-level TE, which remained similar to the global gene set (**Figures 2c and 2d**). In contrast, genes with alternative 5′UTRs showed a marked increase in TE (median log2 fold change 0.23 vs. 0.011 for all genes). Genes with alternative CDS isoforms exhibited a bimodal TE distribution with only modest median changes (median log2 fold change 0.13 vs. 0.011; **Figures 2c and d**). These findings indicate that SCCc selectively utilizes alternative 5′UTR isoforms to enhance translation for a subset of genes.

Translation factor mRNAs often carry 5′ terminal oligopyrimidine (TOP) motifs, which coordinate translation of the TOP gene family [19]. TOP-mediated translation is controlled by the mTORC1 nutrient-sensing pathway [6, 7]. TOP motifs are defined as a +1 cytidine adjacent to the 5′ cap, followed by 4–16 consecutive pyrimidines. To assess TOP motif content, we analyzed de novo 5′ sequences from StringTie transcripts of SCCc genes with alternative 5′UTRs. Most genes expressed both TOP-containing and non-TOP isoforms (**Figure 2e**), suggesting the potential for toggling between mTORC1-dependent and -independent translation.

Previous work has shown that 5′TOP motifs exist along a continuum of transcription start sites (TSSs), with the combination of TOP and non-TOP isoforms influencing mTORC1/LARP1-mediated translational regulation [19]. A “TOPscore” metric incorporates TSS heterogeneity and motif length to predict mTORC1 sensitivity. Genome-wide mapping has also revealed over a thousand additional mRNAs harboring 5′TOP motifs beyond canonical TOP genes [20].

To validate TSS positions from our long-read sequencing, we captured capped 5′ transcript ends from EpSCs, WT keratinocytes, and SCCc using CAGE-seq (**Figure 3a**) [21]. Large-scale 5′ end mapping enables precise TSS identification, promoter prediction, and calculation of TOPscores in different cell types. Mammalian promoters are classified as sharp, TATA-box-enriched promoters or broad, CpG-rich promoters [22]. Using the CAGEr pipeline [23], we observed that WT keratinocytes and SCCc exhibited increased promoter width compared to EpSCs, consistent with greater use of flexible TSSs in differentiated cells relative to stem cells (**Figure 3b**).

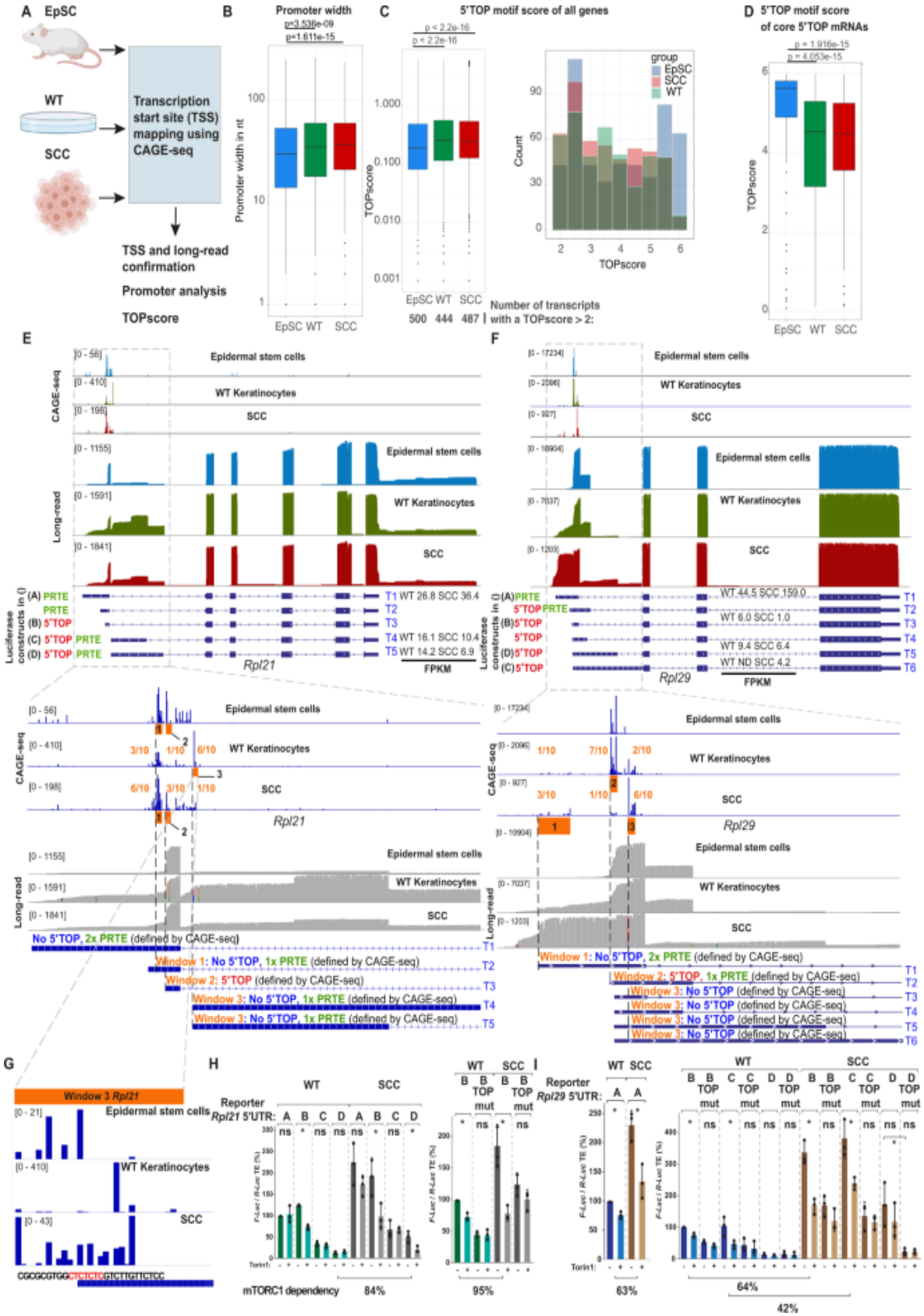


Figure 3. 5′UTR isoform switches of Rpl21 and Rpl29 enhance translation in squamous cell carcinomas

A Overview of the cap analysis of gene expression (CAGE-seq) workflow for mapping transcription start sites (TSSs) in SCA-1+ epidermal stem cells (EpSCs), wild-type keratinocytes (WT), and cultured SCCc.
B Genome-wide promoter width is broader in WT keratinocytes and SCCc. Promoter width was computed using the CAGEr pipeline, and P-values correspond to Wilcoxon tests comparing promoter width distributions.
C Distribution of TOP scores shows that WT and SCCc have higher median values, although fewer transcripts exceed a TOP score >2. TOP scores were calculated using CAGE-seq data and the previously reported script [19]. Left panel: overall distribution and count of transcripts with TOP score >2 (numbers below). Right panel: distribution of transcripts exceeding TOP score >2, including only genes with >500 average CAGE-seq reads. Wilcoxon tests compare distributions.
D TOP scores for the 97 core 5′TOP mRNAs in EpSCs, WT keratinocytes, and SCCc. P-values indicate Wilcoxon test comparisons between cell types.
E, F Visualization of Nanopore long-read and CAGE-seq data for Rpl21 and Rpl29, with transcript annotations. Red indicates 5′TOP motifs (defined as a +1 C followed by 4–16 consecutive pyrimidines). PRTE motifs are defined by nine consecutive pyrimidines with a conserved U at position 6. Letters next to transcripts correspond to luciferase constructs tested in H and I. Lower panels: orange windows highlight major TSS regions in EpSCs, WT, and SCCc; fractions indicate the proportion of CAGE-seq reads in these windows. Labeling of 5′TOP, non-TOP, or PRTE refers to major CAGE peaks rather than annotated transcripts. FPKM values were obtained from long-read sequencing.
G CAGE-seq read distribution in Rpl21 window 3. Although the annotated transcript starts with a 5′TOP motif, WT keratinocytes display a major TSS 8 nucleotides downstream, lacking a 5′TOP motif.
H, I Translational efficiencies of Rpl21 and Rpl29 5′UTR isoforms. WT or SCCc cells were transfected with Firefly luciferase reporters containing Rpl21 or Rpl29 5′UTRs, along with a Renilla control. Cells were treated for 3 h with 500 nM Torin 1 (+) or DMSO (−) before harvesting. Isoform labels correspond to panels E and F. The first WT construct was set to 100%. “TOP mut” indicates constructs in which the 5′TOP motif was mutated. Data represent mean ± s.d. of three independent experiments. Asterisks indicate $P < 0.05$ (ANOVA). mTORC1 dependency was calculated as: (SCC no Torin 1 – SCC Torin 1)/(SCC no Torin 1 – WT Torin 1).

To investigate the use of TOP motifs, we applied the TOPscore metric, which incorporates both consecutive C/U content and the peak height at each 5′CAGE-seq end [19]. WT and SCCc displayed higher median TOPscores compared to EpSCs (**Figure 3c**), suggesting that differentiated cells and SCCc utilize more 5′ motifs subject to mTORC1/LARP1-mediated translational regulation. Core 5′TOP mRNAs typically have TOP scores between 2 and 6 [19]. Interestingly, although EpSCs have a lower median TOPscore, the number of transcripts exceeding a score of 2 was higher than in WT or SCCc (**Figure 3c**, numbers below left panel). Broad promoter regions in WT and SCCc (**Figure 3b**) corresponded to less strict 5′TOP motif usage, also observed among “core” 5′TOP mRNAs (**Figure 3d**).
To experimentally assess the impact of 5′UTR isoforms on translation, we focused on Rpl21 and Rpl29, which exhibited distinct 5′UTR isoforms in the Nanopore dataset and were highlighted in the SplAdder analysis (**Figures 3e–g, 1f and 1g**). Both genes encode multiple 5′UTR isoforms with identical coding sequences and 3′UTRs. Several Rpl21 and Rpl29 5′UTR isoforms were significantly altered in SCCc based on RNA-seq data and Ensembl annotations (**Figure 1e**). Rpl21 and Rpl29 also exhibited increased gene-level translational efficiencies in SCCc compared to WT keratinocytes (log2 fold change: 1.38 and 1.25, respectively).
Using our de novo transcriptome and CAGE-seq data, we updated the 5′UTR annotations. Unlike other ribosomal proteins with sharp promoters, such as Rpl10 (**Figure 1h**), Rpl21 and Rpl29 have broad promoter regions. We cloned major 5′UTR isoforms into Firefly luciferase reporters and transfected them into WT and SCCc to measure translational efficiency via dual-luciferase assays, normalized to Renilla control. To test mTORC1 dependence, cells were treated with Torin 1 or subjected to 5′TOP motif mutation.
Aligned with the translational efficiency measurements obtained via ribosome profiling and RNA-seq, SCCc exhibited generally elevated TEs for the 5′UTR reporter constructs (**Figures 3h and 3i**). Given the higher quantitative resolution of CAGE-seq compared with long-read sequencing, we examined the distribution of CAGE-seq peaks for **Rpl21**, revealing a relative shift in TSS usage from window 3 in WT keratinocytes toward windows 1 and 2 in SCCc (**Figures 3e and 3f**). This TSS reallocation coincided with increased TE in SCCc, as evidenced by the luciferase construct corresponding to window B showing greater translation efficiency than

constructs C and D, which map to the WT window 3 (**Figure 3h**, left panel, B vs. C/D).

Nevertheless, this analysis also highlighted a limitation: for broad promoters, 5′UTR luciferase reporters may not fully capture regulatory nuances, since minor variations in the precise TSS can determine the presence of 5′TOP motifs and thereby influence translational control. For example, in **Rpl21** window 3, the predominant TSS in WT keratinocytes is located eight nucleotides downstream of the annotated 5′TOP, generating a non-TOP isoform. In SCCc, this TSS region broadens to include isoforms containing 5′TOP motifs (**Figure 3g**). Despite these complexities, mTORC1 dependency of the 5′TOP-containing constructs (**Rpl21-B/D**, **Rpl29-B/C/D**) was validated through treatment with Torin 1 or mutation of the TOP motif (**Figures 3h and 3i**), whereas the TOP-less construct Rpl21-A displayed no mTORC1 sensitivity. Interestingly, Rpl21-C, which contains a 5′TOP motif, did not show reduced TE upon Torin 1 treatment (**Figures 3h and 3i**). Prior studies have reported that over 90% of mTORC1-regulated mRNAs possess either a 5′TOP or a pyrimidine-rich translational element (PRTE), implying that either motif predicts mTORC1 responsiveness [6]. Consistently, although Rpl21-A lacks a 5′TOP and its PRTE motif does not confer sensitivity, the TOP-less Rpl29-A harbors a PRTE, explaining its mTORC1 regulation despite lacking a canonical TOP motif.

To overcome potential limitations of reporter assays and the overexpression bias of transient transfection, we analyzed **endogenous mRNAs**. WT and SCCc were subjected to sucrose gradient fractionation, separating transcripts into light polysome fractions (≤3 ribosomes) and heavy polysome fractions (≥4 ribosomes), representing low and high translational efficiencies, respectively. CAGE-seq was then performed on mRNAs from each fraction to map precise TSSs (**Figure 4a**). In agreement with earlier O-propargyl-puromycin (OPP) incorporation studies, SCCc showed reduced polysome occupancy relative to WT keratinocytes, indicating diminished global translation, likely due to increased eIF2α phosphorylation [18]. Nonetheless, SCCc demonstrated elevated mTORC1 activity, as reflected by higher phosphorylation of canonical targets 4E-BP1 and S6K (**Figure 2a**). This indicates that mTORC1 can selectively enhance translation of specific mRNA subsets even when overall protein synthesis is reduced.

Analysis of promoter width revealed that transcripts with higher translation efficiency generally originate from narrower promoters compared with transcripts in light polysomes (**Figure 4b**). Genome-wide assessment of 5′TOP and PRTE motifs within light and heavy polysome fractions, guided by precise CAGE-seq peak mapping, showed that SCCc consistently had higher TOP and PRTE scores than WT across both fractions (**Figures 4c and 2e**). This trend was also reflected in the number of transcripts with TOPscore >2 (**Figure 4c**) [19]. Furthermore, heavy polysome fractions exhibited greater TOP and PRTE scores than light fractions in both WT and SCCc. Overall, these results indicate that SCCc undergoes a genome-wide shift toward increased usage of 5′TOP and PRTE motifs, facilitating mTORC1-mediated translational control and generally correlating with higher translational efficiency.

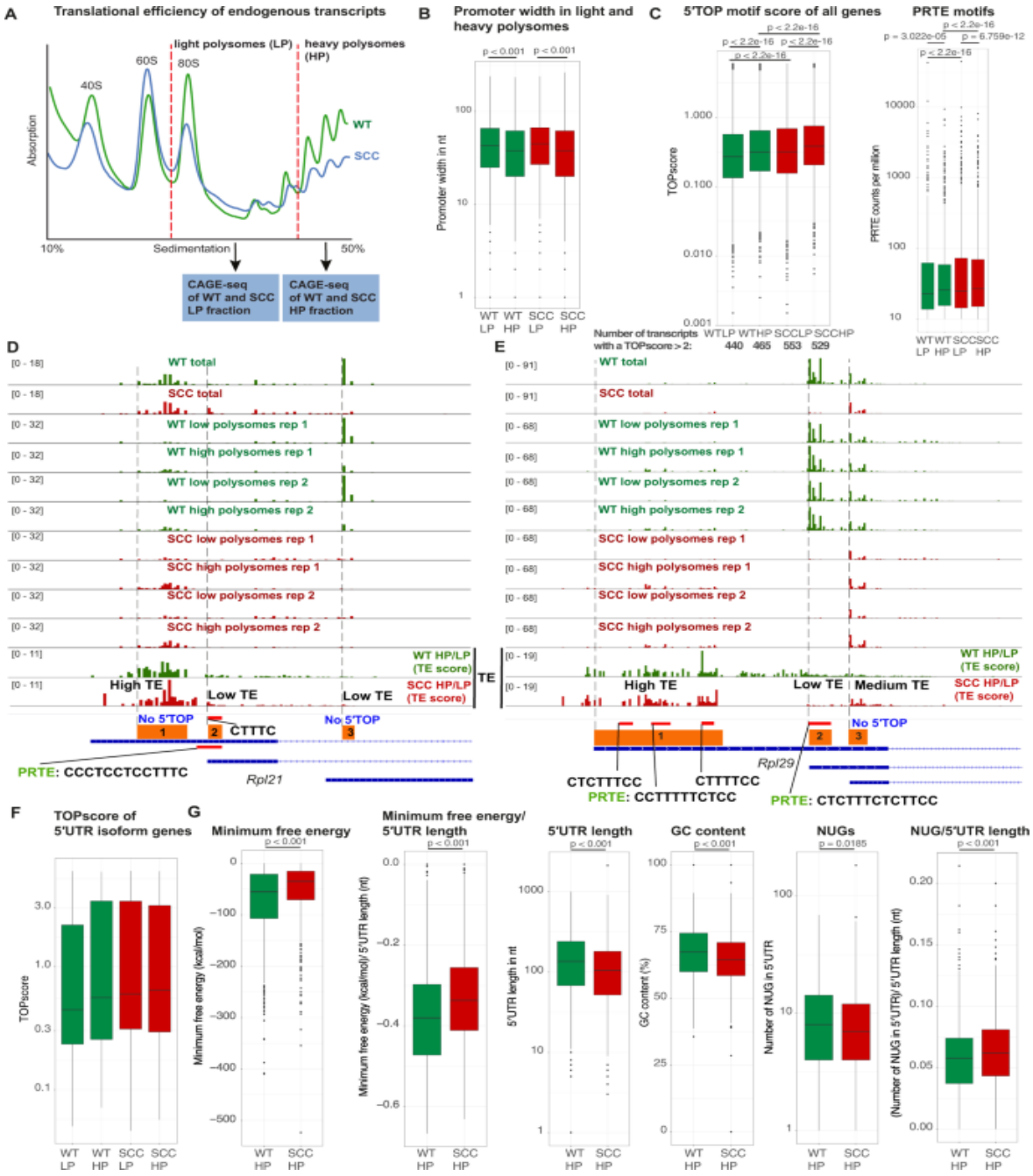


Figure 4. High translational potential of 5′UTR isoforms in squamous cell carcinoma

A Strategy combining polysome fractionation with cap analysis of gene expression (CAGE-seq) for mapping transcription start sites (TSS) in wild-type (WT) keratinocytes and SCCc. Lysates were separated on sucrose density gradients, collecting light polysome (LP; ≤3 ribosomes) and heavy polysome (HP; ≥4 ribosomes) fractions. RNA extracted from these fractions was used to generate CAGE libraries. Results represent averages from two independent experiments.

B Promoter width is narrower in heavy polysome fractions for both WT and SCCc.

C Median TOP and PRTE scores were elevated in SCCc transcripts across LP and HP fractions, along with an increased number of transcripts with TOPscore >2 (total transcript counts shown below the graph). TOPscores were computed using WT and SCCc LP/HP CAGE-seq data with the published script [19]. PRTE scores were calculated by identifying PRTE motifs and normalizing

by total reads: median PRTE scores were 22.9 (WT LP), 25.8 (WT HP), 24.9 (SCCc LP), and 26.9 (SCCc HP). PRTE-containing 5′UTRs were defined as PRTE score >10, with motifs consisting of nine consecutive pyrimidines and a conserved uridine at position six.

D, E Distribution of CAGE-seq peaks in total, LP, and HP fractions for the three main TSS windows in WT and SCCc, normalized and autoscaled for direct comparison. Ratios of HP/LP were used as proxies for translational efficiency (TE), categorized as low (<1), medium (1–3), and high (>3). Red bars indicate potential 5′TOP TSSs or PRTE-containing sequences.

F TOPscores across LP and HP fractions for the set of 5′UTR isoforms.

G Heavy polysome-associated 5′UTRs in SCCc show reduced RNA secondary structure and fewer potential upstream open reading frames (uORFs). CAGE ctss were clustered into promoter regions using the CAGEr pipeline, and DESeq2 identified significantly altered clusters. Corresponding 5′UTRs were examined for minimum free energy, length, GC content, and potential uORF start sites (NUG: AUG, CUG, GUG, UUG). P-values were determined using the Wilcoxon test.

Analysis of **Rpl21** and **Rpl29** revealed that the main WT TSS windows (Rpl21-3, Rpl29-2) correspond to low TE (HP/LP ratios), whereas SCCc preferentially use TSS windows Rpl21-1, Rpl29-1, and Rpl29-3, showing medium to high TE (**Figures 4d and 4e**). Notably, Rpl21-1 has high TE without 5′TOP motifs. Similarly, Rpl29-1 lacks 5′TOP at the annotated transcript start, but its broad TSS window includes three stretches of 5′TOP motifs with high TE (**Figures 4d and 4e**). These observations highlight that broad promoter regions may harbor additional functional 5′TOP motifs not reflected in annotated TSS. Since either 5′TOP or PRTE motifs predict mTORC1 responsiveness [6], we also assessed PRTE presence; Rpl21-1 and Rpl29-1 contained PRTE motifs, consistent with potential mTORC1-dependent translation. Across all 73 5′UTR isoforms, TOPscores trended higher in SCCc and in HP fractions, though heterogeneity prevented statistical significance (**Figure 4f**). These results suggest that SCCc preferentially select TSSs associated with high TE and genome-wide enrichment of 5′TOP and PRTE motifs (**Figure 4c**).

To further characterize features of efficiently translated 5′UTR isoforms, HP fractions from WT and SCCc were compared. CAGE ctss were grouped into tag cluster promoter regions, and DESeq2 identified significantly altered clusters. Corresponding 5′UTRs in SCCc HP fractions exhibited less negative folding free energy, lower GC content, and, when normalized for length, reduced secondary structure (**Figure 4g**). Although NUG density was higher in SCCc, shorter 5′UTRs resulted in fewer total potential uORFs, supporting improved scanning and translation initiation (**Figure 4g**). These data indicate that SCCc not only increase 5′TOP/PRTE motif usage but also express 5′UTRs optimized for translation efficiency.

Finally, the relevance of 5′UTR isoform switching in human cancer was assessed in 519 HNSCC patients. Elevated **RPL21**, but not **RPL29**, total mRNA levels correlated with reduced overall survival (**Figures 5a–c, 2b, 3a and b**). RPL21 encodes four isoforms differing only in 5′UTRs while producing identical proteins. The two TOP motif-containing isoforms were significantly associated with survival (**Figures 5b and 4**). Patients with low RPL21-201 expression had a median survival of 89 months versus 48 months for high expressers, an increase of ~3.5 years. Non-TOP isoforms showed no correlation, while TOP-like isoforms weakly correlated with survival (**Figure 5b and 4**). Analysis of alternative 5′UTR splicing (RPL21-201/204 vs. RPL21-203) also stratified overall survival in HNSCC patients (**Figures 5d and 4**). Collectively, these findings support that 5′UTR isoform switching may contribute to cancer progression and warrant isoform-specific investigation in human datasets.

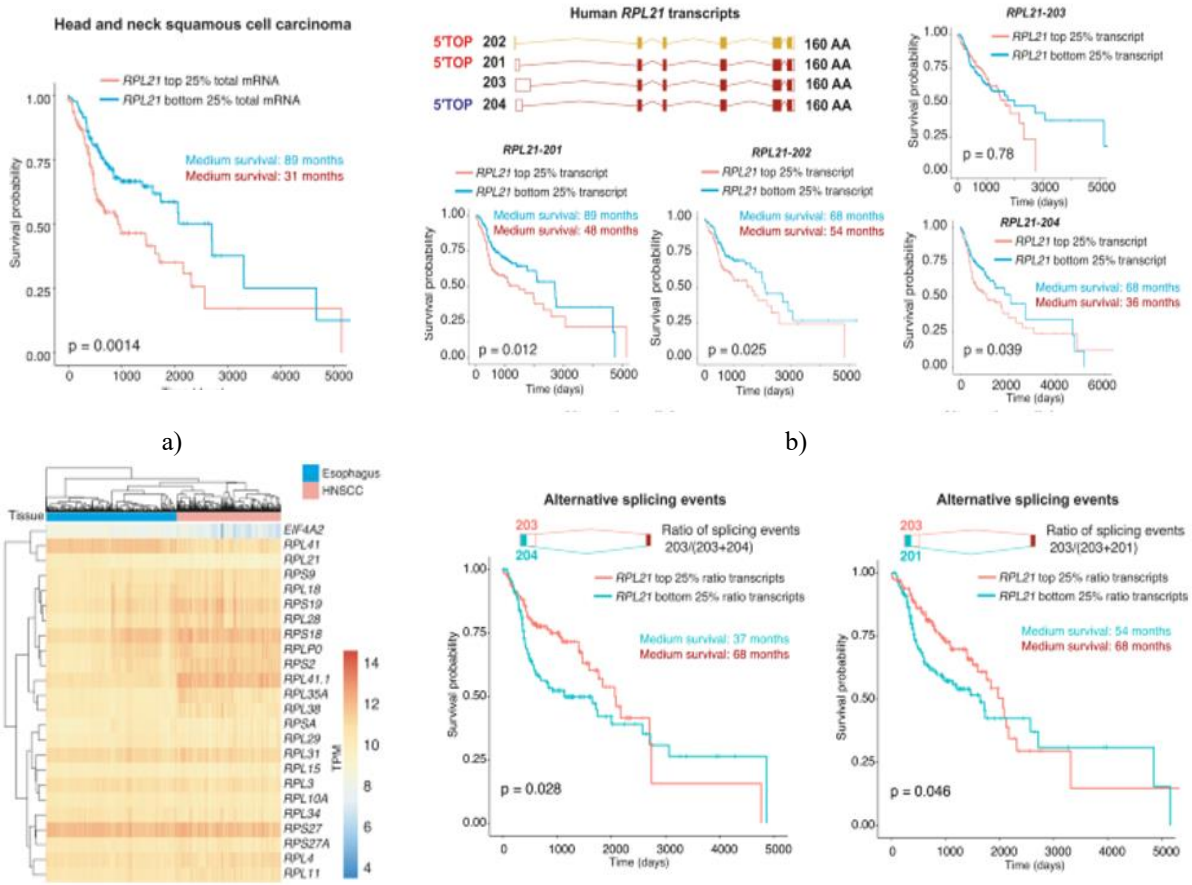


Figure 5. Association of RPL21 5′TOP-containing isoforms with survival in HNSCC patients

a Higher RPL21 mRNA expression correlates with decreased overall survival in head and neck squamous cell carcinoma (HNSCC). Shown are the top versus the bottom quartiles in TCGA HNSCC samples (n = 519). Cox regression hazard ratio = 1.4629.

b Among RPL21 isoforms, only the TOP motif-containing transcripts 201 and 202, not the TOP-less 203, are linked to shorter overall survival. Quartile expression in TCGA HNSCC (n = 519). The four primary isoforms all encode the same 160-amino-acid protein; two feature canonical 5′TOP motifs (red), one carries a TOP-like motif (blue), and one lacks a TOP motif.

c Expression of translation-related, alternatively spliced genes in TCGA HNSCC (n = 519) versus non-diseased esophageal tissue (no pharynx controls available).

d Alternative splicing distinguishes survival outcomes. The ratio of events for survival-associated transcripts RPL21-201/204 versus non-correlating RPL21-203 was used to stratify top and bottom quartiles of overall survival in HNSCC patients (n = 519).

Transcription and translation jointly dictate nearly 90% of cellular protein abundance [24], yet the extent of their direct coordination across biological contexts remains unclear. Here, we combined Nanopore long-read sequencing with polysome-associated CAGE-seq to define 5′UTR isoform usage and promoter architecture in epidermal stem cells and cultured SCCc. The resulting annotated transcriptomes are publicly accessible through a genome browser.

We found that a limited subset of 5′UTR isoforms in SCCc modulates translational efficiency (TE) of their coding sequences. Globally, SCCc transcripts with high TE preferentially utilize 5′UTR motifs known to enhance mTORC1-dependent translation. Crucially, elevated levels of TOP motif-containing—but not TOP-less—RPL21 isoforms significantly associate with reduced

survival in HNSCC, emphasizing the functional importance of specific 5′UTR configurations.

These data suggest that, for translation-related genes, isoform identity may be a stronger determinant of protein output than total mRNA quantity. Supporting this, genome-wide correlations between mRNA and protein in HNSCC averaged rs = 0.52, but ribosomal proteins, including RPL21 and RPL29, exhibited weak correlations (**Figures 2c and 2d**) [25]. This highlights the role of post-transcriptional control, particularly via 5′UTR-driven translational potential, in regulating ribosomal protein levels in tumors. Despite the small number of 5′UTR switches observed, genome-wide translation regulators likely remain dominant for overall protein synthesis.

Switching 5′UTR isoforms represents a non-canonical, efficient mechanism for controlling translation [1-3, 26]. Our study adds to this framework by showing that exposing or masking 5′TOP and PRTE motifs can dramatically adjust mRNA TE in cancer cells. Even minor shifts in TSS selection can reveal or hide regulatory 5′UTR elements, altering mTORC1 dependency [6, 7]. Given the prevalence of alternative splicing and heterogeneous TSS usage [10, 11], isoform switching could allow tumors to dynamically control translational potential by encoding uORFs, RBP motifs, and 5′TOP/PRTE elements in their 5′UTRs at different stages of progression.

Materials and Methods

Genome browser access

All Nanopore long-read RNA sequencing data, CAGE datasets, and de novo assembled transcriptomes can be explored via the UCSC genome browser at: https://genome-euro.ucsc.edu/s/umeshghosh/cage_ont. Nanopore expression levels are presented as TPM (transcripts per kilobase million), while CAGE signals are displayed as tags per million. Note that bigwig files are used instead of bam files, which may slightly change visual representation. Updated browser links will also be maintained on GitHub: https://github.com/ugdastider/long_read_paper.

Nanopore long-read RNA sequencing

RNA extraction was performed using TRIzol LS (Thermo Fisher, 10296010), followed by purification with the Direct-zol™ RNA Miniprep Kit (Zymo Research, R2050). Libraries were constructed according to the SQK-PCS109 protocol (Oxford Nanopore Technologies, UK). In brief, polyadenylated RNA was enriched with oligo(dT) primers, annealed, and reverse transcribed. After template-switching, PCR amplification was performed using rapid attachment primers, followed by ligation of rapid 1D adapters. Sequencing was carried out on a PromethION device (Oxford Nanopore Technologies, UK).

Isolation of adult mid-telogen Epidermal Stem Cells (EpSCs)

Female C57BL6 mice aged P56–60 were purchased from Janvier. Epidermal stem cells were isolated from telogen back skin following a modified version of the previously reported method [27]. Subcutaneous fat and muscle were removed, and skin was incubated in 0.5% Trypsin-EDTA (10X; Gibco, 15400054) at 37 °C for 25 min on an orbital shaker. Skin was scraped to obtain a single-cell suspension, and trypsin was neutralized with PBS containing 2% chelexed FBS (Gibco, 10010-015). Cells were filtered through 70 µm and 40 µm strainers (Corning, 431750, 431751) and pelleted. SCA-1+ EpSCs were enriched using magnetic-activated cell sorting (MACS) with Anti-SCA-1 microbeads (Miltenyi Biotec, 130-106-641), MS columns, and a MultiStand system (Miltenyi Biotec, 130-042-201). Purified EpSCs were pelleted, resuspended in Trizol LS, and RNA was extracted using Direct-zol™ RNA MiniPrep Kit (Zymo Research, R2050). RNA concentration was measured using the Qubit™ RNA BR assay (Invitrogen, Q10210). All procedures were approved by the Veterinary Office of the Canton of Zurich, Switzerland (License ZH233/2019).

Cell culture conditions

Cells were maintained in 0.05 mM Ca^{2+} E-media formulated in-house using DMEM/F12 supplemented with 15% chelated FBS, 5 µg/mL insulin, 5 µg/mL transferrin, 2 nM triiodothyroxine, 40 µg/mL hydrocortisone, and 10 nM cholera toxin. Cultures were incubated at 37 °C in 5% CO_2. HrasG12V; Tgfbr2 knockout cells were previously established [14]. Primary newborn mouse keratinocytes from wild-type mice were cultured on 3T3-S2 feeders in 0.05 mM Ca^{2+} E-media with 15% serum [27]. Cells were periodically tested for mycoplasma contamination every three months using the Mycoplasma PCR Detection Kit (Sigma, D9307).

Luciferase reporter assays

0.25 × 10^6 cells were seeded in 6-well plates with 2 mL E-media 24 h before transfection. Transfection mixes contained 1.5 µg Rpl29/Rpl21-F-Luc and 0.5 µg R-Luc control plasmid, delivered with Lipofectamine 2000. Two hours post-transfection, cells were treated with 500 nM Torin 1 or DMSO and harvested 5 h after transfection. Cells were washed briefly with PBS and lysed in passive lysis buffer. Firefly and Renilla luciferase activities were measured at room temperature using the Dual-Luciferase Reporter Assay System (Promega, E1980) on a Tecan Infinite M1000Pro plate reader. Experiments were carried out in biological triplicate.

Quantitative real-time PCR

Cells were lysed in TRIzol LS (Thermo Fisher, 10296010), and RNA was extracted via chloroform separation, followed by isopropanol precipitation and resuspension in H_2O. For cDNA synthesis, 0.5 µg RNA was combined with 0.5 µg random hexamers (N6), denatured at 72 °C for 5 min, and reverse transcribed with 1× SSIII RT buffer, 1 mM dNTPs, 5 mM DTT, and 0.5 µL SSIII (Invitrogen, 18080093) to 20 µL total volume. The RT reaction was incubated at 55 °C for 1 h and terminated at 70 °C for 10 min. qPCR was performed in 10 µL reactions using 1× iTaq Universal SYBR Green Supermix (Biorad, 1725121), 0.4 µM of each primer, and 1 µL cDNA.

Bioinformatic analysis

All computational analyses were conducted on an Ubuntu 18.04.5 cluster equipped with 32 CPU cores and 128 GB RAM. Unless otherwise noted, software was run with default settings. Python version 3.8.6 was used unless specified.

Nanopore long-read data processing

Raw long-read sequencing outputs were processed using the Nextflow nanoseq v1.1 pipeline [28] together with a custom analysis workflow. Quality metrics for fastq files and sequencing performance were evaluated using FastQC and NanoPlot [29]. Reads were mapped to the mouse reference genome Gencode GRCm38 (version M25) via minimap2, which supports both spliced and unspliced alignments. Isoform assembly and abundance estimation were performed using StringTie2 [15], and merged transcriptomes were compared with GFFcompare v0.11.2 [30]. The datasets are available in GEO GSE179525: wild-type keratinocytes (GSM5419823), SCCc (GSM5419824), and epidermal stem cells (GSM5419822).

Short-Read RNA-seq data handling

Short-read RNA sequencing data were processed with the Nextflow RNA-seq pipeline [28] combined with custom scripts. Fastq files underwent quality checks via FastQC, and adapter trimming and quality filtering were performed using TrimGalore. STAR v2.6.1 [31] was used for alignment to GRCm38, followed by transcript quantification with Salmon [32]. Coverage visualization tracks in BigWig format were generated with BEDTools, and MultiQC compiled the quality assessments across all analyses.

Differential expression analysis

Raw counts at the gene and transcript levels were analyzed using DESeq2 [33] to detect differentially expressed (DE) genes across conditions. The top 500 most variable genes were selected for hierarchical clustering and PCA plots after variance-stabilizing transformation (vst) normalization to identify outliers. Counts were normalized using the median-of-ratios approach, dispersion was estimated, and generalized linear models were fitted for each gene. Wald test p-values were adjusted with the Benjamini–Hochberg method, and genes with FDR <0.05 were used in downstream analyses.

Ribosome profiling data

Ribosome profiling datasets were processed similarly to short-read RNA-seq, with an additional filtering step to remove reads mapping to a combined reference of ribosomal, mitochondrial, or tRNA sequences [34] using Bowtie2 v2.4.1 [35].

Sample details for ribosome profiling and RNA-Seq

Previously published datasets [18] of ribosome profiling and short-read RNA-seq from wild-type keratinocytes and Hras(G12V); Tgfbr2 null SCCc were employed, with data available at GEO GSE83332 and GSE179525:

- Wild-type keratinocytes: S11, S13, S14 (Ribosome profiling, GSM2199591, GSM2199593, GSM2199594); S27, S28 (RNA-seq, GSM2199607, GSM2199608)
- SCCc: S23, 197, S24 (Ribosome profiling, GSM2199603, GSM5419825, GSM2199604); 211, 212 (RNA-seq, GSM5419826, GSM5419827)

- Epidermal stem cells: S7, S8 (RNA-seq, GSM2199587, GSM2199588) representing P4 epidermis enriched for stem cells. Note: P60 SCA-1+ EpSCs ribosome profiling data were not available.

Curated transcriptome generation

For the curated transcriptome, isoform quantification was conducted using SQANTI3 [36], integrating short-read RNA-seq sequentially. The initial StringTie transcriptome was quantified, filtered to remove low-support junctions and antisense artifacts, and remapped to enhance counts. Functional annotations were transferred from tappAS annotations based on PacBio data using IsoAnnotLite v2.6 [37].

Translational efficiency estimation

RNA-seq and ribosome profiling reads were quantified over exons and CDS using Plastid v0.4.7 [38] (Python 2.7). Translational efficiency (TE) was calculated in R v4.0.2 using LRT-tests from DESeq2, following the full/reduced model approach [38]. Genes with RNA-seq RPKM >1 were considered, and for comparison with SplAdder gene lists (**Figure 2c**), TE calculations were limited to base mean >25. Raw counts, not RPKM, were used to normalize sample differences.

Alternative isoform analysis

Alternative splicing events were identified using SplAdder v2.4.3 [16] with level 3 confidence parameters and StringTie annotations. Only events with adjusted p-value <0.05 were retained.

Pathway enrichment analysis

Gene ontology enrichment was assessed using the pre-ranked GSEA method, enabling simultaneous detection of up- and down-regulated genes. Overrepresentation analyses were performed using EnrichR and custom scripts.

TCGA and GTEx data analysis

We retrieved gene expression profiles for HNSCC and esophagus tissues from UCSC Xena Toil (xena.ucsc.edu). Survival analysis for TCGA HNSCC patients was performed in R using the *survival* and *survminer* packages. Gene expression was modeled against patient survival using a Cox proportional hazards regression to calculate hazard ratios. Kaplan–Meier survival curves were plotted for patients in the top and bottom quartiles of gene expression, and significance was assessed using a log-rank test.

Sucrose gradient polysome fractionation

To isolate light and heavy polysome fractions, lysates from WT keratinocytes and SCCc were prepared in a buffer containing 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, 0.5% NP-40, 1 mM DTT, and 100 µg/mL cycloheximide. Polysomes were separated on 10–50% sucrose gradients (prepared in 20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 100 µg/mL cycloheximide) by ultracentrifugation at 41,000 rpm for 2 h. Fractions were collected using a Biocomp density gradient system. For RNA extraction, pooled fractions were mixed with three volumes of Trizol LS (Ambion) for 10 min at room temperature, followed by the addition of four volumes of 100% ethanol. RNA purification was performed with the Direct-zol™ RNA MiniPrep kit (Zymo Research, R2050).

CAGE-seq library preparation

RNA integrity was verified using an Agilent Bioanalyzer, retaining samples with RIN >7. cDNA synthesis employed random primers. 5′ RNA cap structures were oxidized and biotinylated, followed by streptavidin-based capture. After RNA digestion with RNaseONE/H and adaptor ligation to both cDNA ends, double-stranded cDNA libraries were constructed. Sequencing was performed on a NextSeq 500 (single-end, 75 nt). Reads were mapped to the mm10 mouse genome. CAGE-seq for polysome fractions was performed in biological duplicates, while total lysate libraries were single-replicate. Data are deposited on GEO (GSE201308).

Western blotting

Cells were rinsed with PBS and lysed in protein sample buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 0.2 M DTT), boiled at 95 °C for 5 min, and vortexed to shear DNA. Proteins were resolved on SDS-PAGE, transferred to nitrocellulose membranes, blocked with 5% BSA, and incubated with primary antibodies overnight at 4 °C. Secondary antibodies were applied for 3 h at 4 °C. Detection used freshly prepared ECL (GE Healthcare). Antibodies included: Phospho-p70 S6 Kinase (Thr389, Cell Signaling #9234), 4E-BP1 (Cell Signaling #9452), Phospho-4E-BP1 (Thr37/46, Cell Signaling #2855), and Tubulin (Sigma T6199). Band intensities were quantified using ImageJ.

Survival analysis of transcript usage

SUPPA2 [39] was used to quantify transcript-level alternative usage in TCGA HNSCC samples. Relative transcript abundances were computed using the *psiPerIsoform* function, then merged with survival data to identify top- and bottom-quartile differences. Significance was assessed with log-rank testing.

CAGE data processing and promoter analysis

The Nextflow *cageseq* pipeline [40] (https://nf-co.re/cageseq) was used to trim 5′ G nucleotides and align reads to the mm10 genome with STAR. BAM files were analyzed using the *CAGEr* package [23] to generate CTSS counts normalized to TPM. CTSS clustering was performed using the "distclu" approach. Promoter widths (0.1–0.9 quantiles) were computed with the *quantilePositions* function. DESeq2 [33] was applied to detect differential tag usage. 5′UTR sequence features, including GC content and NUG density, were analyzed using custom Python scripts, and minimum free energy (MFE) was calculated using the ViennaRNA Python library [41]. Sharp promoters were defined as IQR <10 bp, while broad promoters had IQR ≥10 bp.

SQANTI quality control reports

SQANTI3 QC reports for long-read transcriptomes are publicly available at: https://github.com/ugdastider/long_read_paper/tree/main/sqanti

TOPscore computation

TOPscores were derived from the CAGE-aligned BAM files using the *tss_analyzer* tool (available at https://github.com/carsonthoreen/tss_tools) [19]. Only transcripts with an average of more than 500 CAGE-seq reads across the samples were retained for further analysis.

PRTE motif quantification

CAGE-seq reads aligned to the mouse genome were analyzed to detect pyrimidine-rich translational element (PRTE) motifs. PRTEs were defined as stretches of 9 consecutive cytosine/thymine nucleotides with a thymine fixed at the sixth position. Using a custom Python script, we counted the total number of reads mapping to each gene that contained a PRTE motif. Counts were then normalized to library size and expressed as counts per million mapped reads. Only genes with PRTE abundance greater than 10 counts per million were included in downstream analyses.

Acknowledgments: This project was supported by an SNSF Professorship grant (grant number 176825), by the European Research Council (ERC) under the European Union′s Horizon 2020 research and innovation programme (grant agreement No 759006), the National Center of Competence in Research (NCCR) on RNA and Disease funded by the SNSF and received support from the Human Frontier Science Program Long-Term Fellowship (RW, LT001084/2021-L).

Conflict of Interest: None

Financial Support: Open access funding provided by University of Zurich.

Ethics Statement: None

References

1. Chen J, Tresenrider A, Chia M, McSwiggen DT, Spedale G, Jorgensen V, et al. Kinetochore inactivation by expression of a repressive mRNA. Elife. 2017;6. https://doi.org/10.7554/eLife.27417.
2. Cheng Z, Otto GM, Powers EN, Keskin A, Mertins P, Carr SA, et al. Pervasive, coordinated protein-level changes driven by transcript isoform switching during meiosis. Cell. 2018. https://doi.org/10.1016/j.cell.2018.01.035.
3. Tresenrider A, Morse K, Jorgensen V, Chia M, Liao H, van Werven FJ, et al. Integrated genomic analysis reveals key features of long undecoded transcript isoform-based gene repression. Mol Cell. 2021;81:2231–45.e11.
4. Hollerer I, Barker JC, Jorgensen V, Tresenrider A, Dugast-Darzacq C, Chan LY, et al. Evidence for an integrated gene repression mechanism based on mRNA isoform toggling in human cells. G3 Genes Genomes Genet. 2019;9:1045–53.
5. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5′-untranslated regions of eukaryotic mRNAs. Science. 2016. https://doi.org/10.1126/science.aad9868.
6. Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, et al. The translational landscape of mTOR signalling steers cancer initiation and metastasis. Nature. 2012. https://doi.org/10.1038/nature10912.

7. Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM. A unifying model for mTORC1-mediated regulation of mRNA translation. Nature. 2012;485:109–13.
8. Saxton RA, Sabatini DM. mTOR signaling in growth, metabolism, and disease. Cell. 2017;168:960–76.
9. Hsieh AC, Costa M, Zollo O, Davis C, Feldman ME, Testa JR, et al. Genetic dissection of the oncogenic mTOR pathway reveals druggable addiction to translational control via 4EBP-eIF4E. Cancer Cell. 2010;17:249–61.
10. Philippe L, van den Elzen AMG, Watson MJ, Thoreen CC. Global analysis of LARP1 translation targets reveals tunable and dynamic features of 5′ TOP motifs. Proc Natl Acad Sci USA. 2020;117:5319–28.
11. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. Nature. 2013. https://doi.org/10.1038/nature12121.
12. Kahles A, Lehmann KVan, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. Cancer Cell. 2018;34:211–24.e6.
13. Hu H, Gehart H, Artegiani B, LÖpez-Iglesias C, Dekkers F, Basak O, et al. Long-term expansion of functional mouse and human hepatocytes as 3D organoids. Cell. 2018;175:1591–1606.e19.
14. Joost S, Zeisel A, Jacob T, Sun X, La Manno G, Lönnerberg P, et al. Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity. Cell Syst. 2016;3:221–37.e9.
15. Yang H, Schramek D, Adam RC, Keyes BE, Wang P, Zheng D, et al. ETS family transcriptional regulators drive chromatin dynamics and malignancy in squamous cell carcinomas. Elife. 2015. https://doi.org/10.7554/eLife.10870.
16. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.
17. Kahles A, Ong CS, Zhong Y, Rätsch G. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. Bioinformatics. 2016;32:1840–7.
18. Kondrashov N, Pusic A, Stumpf CR, Shimizu K, Hsieh AC, Xue S, et al. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell. 2011;145:383–97.
19. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, et al. Translation from unconventional 5′ start sites drives tumour initiation. Nature. 2017. https://doi.org/10.1038/nature21036.
20. Yamashita R, Suzuki Y, Takeuchi N, Wakaguri H, Ueda T, Sugano S, et al. Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. Nucleic Acids Res. 2008;36:3707.
21. Murata M, Nishiyori-Sueki H, Kojima-Ishiyama M, Carninci P, Hayashizaki Y, Itoh M. Detecting expressed genes using CAGE. Methods Mol Biol. 2014;1164:67–85.
22. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet. 2006;38:626–35.
23. Haberle V, Forrest ARR, Hayashizaki Y, Carninci P, Lenhard B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. Nucleic Acids Res. 2015;43:e51–e51.
24. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011. https://doi.org/10.1038/nature10098.
25. Huang C, Chen L, Savage SR, Eguez RV, Dou Y, Li Y, et al. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. Cancer Cell. 2021;39:361–79.e16.
26. Chia M, Tresenrider A, Chen J, Spedale G, Jorgensen V, Ünal E et al. Transcription of a 5' extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter. Elife. 2017;6. https://doi.org/10.7554/eLife.27420.
27. Nowak JA, Fuchs E. Isolation and culture of epithelial stem cells. Methods Mol Biol. 2009;482:215–32.
28. Patel H, Wan YK, Wratten L, Sawyer C, Ying C, Alneberg J et al. nf-core/nanoseq: nf-core/nanoseq v1.1.0 - Green Panda. 2020. https://doi.org/10.5281/ZENODO.4249716.
29. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666–9.
30. Pertea M, Pertea G. GFF utilities: GffRead and GffCompare. F1000Research. 2020;9:304.

31. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
32. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417–9.
33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
34. Chan PP, Lowe TM. GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. 2016;44:D184–9.
35. Langmead B, Wilks C, Antonescu V, Charles R. Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics. 2019;35:421–32.
36. Tardaguila M, De La Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018;28:396–411.
37. De La Fuente L, Arzalluz-Luque Á, Tardáguila M, Del Risco H, Martí C, Tarazona S, et al. TappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. Genome Biol. 2020;21:1–32.
38. Dunn JG, Weissman JS. Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. BMC Genom. 2016. https://doi.org/10.1186/s12864-016-3278-x.
39. Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol. 2018;19. https://doi.org/10.1186/S13059-018-1417-1.
40. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38:276–8.
41. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:1–14.