RNA sequencing analysis identifies novel spliced transcripts but does not indicate quantitative or qualitative changes of viral transcripts during progression of cottontail rabbit papillomavirus-induced tumours

Sonja Probst-Hunczek, Günter Jäger, Markus Schneider, Ekaterina Notz, Frank Stubenrauch and Thomas Iftner

1Division of Experimental Virology, Institute of Medical Virology and Epidemiology of Viral Diseases, University Hospital Tübingen, Elfriede-Aulhorn-Strasse 6, Tübingen 72076, Germany
2Centre for Bioinformatics Tübingen (ZBIT), University of Tübingen, Tübingen 72070, Germany
3Department of Medical Genetics, University Hospital Tübingen, Calwerstrasse 7, Tübingen 72076, Germany

Persistent infections with high-risk human papillomaviruses (HPVs) can result in the development of cancer of the cervix uteri and other malignancies. The underlying molecular mechanisms leading to the progression of HPV-induced lesions are, however, not well understood. Cottontail rabbit papillomavirus (CRPV) induces papillomas in domestic rabbits which progress at a very high rate to cancer. Using this model, we compared the transcriptional patterns of CRPV in papillomas and carcinomas by RNA sequencing (RNA-seq). The most abundant transcripts can encode E7, short E6 and E1^E4, followed by full-length E6, E2, E1 and E9^E2C. In addition, we identified two rare, novel splice junctions 7810/3714 and 1751/3065 in both papillomas and carcinomas which have been described for other papillomaviruses. Neither RNA-seq nor quantitative real-time PCR-based assays identified qualitative or quantitative changes of viral transcription between papillomas and carcinomas. In summary, our analyses confirmed that papillomaviruses have highly similar transcriptional patterns, but they do not suggest that changes in these patterns contribute to the progression of CRPV-induced tumours.

INTRODUCTION

Infections with high-risk human papillomaviruses (hrHPVs), mainly HPV16, can lead to cervical, anal, and head and neck cancer (Haedicke & Iftner, 2013). The majority of hrHPV infections are cleared and only a minority become persistent. A fraction of persistently infected individuals develop precancer and cancer after several years. The events that cause persistence and progression of hrHPV-induced lesions are not well understood. hrHPVs encode E6 and E7 oncoproteins which have immortalizing activities (Moody & Laimins, 2010). In addition, E6 and E7 induce genomic instability which induces aberrations in the infected cell (Moody & Laimins, 2010). Furthermore, E6 and E7 have been implicated in the induction of epigenetic changes of the host cell (Steenbergen et al., 2014). Consistent with the idea that E6 and E7 are the major drivers of carcinoma development, the overexpression of only HPV16 E6 and E7 gives rise to carcinomas in transgenic mice (Lambert et al., 1993). However, during the normal infection cycle hrHPVs express a large number of differentially spliced transcripts that encode, in addition to E6 and E7, the viral E1, E2, E5, L1, L2, E1^E4 and E8^E2C proteins. Some of these may contribute to progression as they are regulators of viral transcription (E1, E2 and E8^E2C), regulators of cell growth (E1, E2 and E1^E4) or have oncogenic activity (E5) (Bergvall et al., 2013; DiMaio & Petti, 2013; Doorbar, 2013; McBride, 2013). Furthermore, the analysis of viral transcripts by sensitive methods such as reverse transcription (RT)-PCR and RACE have suggested that HPV16, for example, expresses additional viral transcripts whose relevance for progression has not been addressed (http://pave.niaid.nih.gov/#explore/transcript_maps/hpv16).

The analysis of HPV-related progression mechanisms has been hampered by the failure of hrHPVs to replicate in animal model systems and then cause cancer. In contrast, infection of domestic rabbits (Oryctolagus cuniculus) with...
the cottontail rabbit papillomavirus (CRPV) leads to the formation of papillomas that progress within months at high rates to carcinomas (Jeckel et al., 2002, 2003). Papilloma formation is dependent upon the E6 and E7 genes, but also on the presence of the viral replication proteins E1 and E2, suggesting that viral replication is at least initially required for tumour formation (Jeckel et al., 2002; Meyers et al., 1992; Muench et al., 2010; Wu et al., 1994). CRPV E7 immortalizes primary rabbit keratinocytes and this is enhanced by CRPV E6. CRPV E6 and E7 interfere with the functions of p53 and pRb (Ganzenmueller et al., 2008; Muench et al., 2010; Schmitt et al., 1994). In summary, this suggests that the CRPV/domestic rabbit model mimics several aspects of hrHPV-induced cancer development in humans.

Three early (P1–P3) and one late promoter (PL) have been mapped for CRPV, and the corresponding polycistronic transcripts are processed at splice donor (SD) and splice acceptor (SA) sites conserved among papillomaviruses. In contrast to hrHPV, the PL resides in the regulatory region (and not in the E7 gene) and two additional promoters are located within the E6 gene giving rise to a truncated version of E6 [short E6 (SE6)] and full-length E7 (Nasseri & Wettstein, 1984; Wettstein et al., 1987). Northern blot, S1 nuclease, primer extension and in situ hybridization analyses have not provided strong evidence that the levels of known viral transcripts differ between papillomas and carcinomas (Nasseri & Wettstein, 1984; Phelps et al., 1985; Wettstein et al., 1987; Zeltner et al., 1994). However, an RT-PCR approach has revealed that CRPV also expresses a spliced transcript equivalent to the HPV E8^E2C transcript that was not detected by other methods (Jeckel et al., 2003). This led us to speculate that additional viral transcripts may exist that are modulated during progression. To identify such transcripts the transcriptomes of several papillomas and carcinomas were analysed in an unbiased manner by RNA sequencing (RNA-seq).

Two novel splice junctions could be identified. In the first one, the late upstream regulatory region (URR) exon is linked to the E2/E4 exon; in the second one, the E9 exon is linked to the E2 exon. Both transcripts are expressed at low levels and their expression does not change with progression. Furthermore, RNA-seq and quantitative PCR analyses confirmed that all previously described transcripts also do not change during progression, suggesting that transcriptional changes are not contributing to progression of CRPV-induced skin cancers.

**RESULTS**

First, the histology of several CRPV-induced tumours was evaluated by haematoxylin/eosin staining (Fig. 1). Two of the carcinomas were undifferentiated squamous epithelial tumours that also contained highly differentiated areas with keratin aggregates, whereas the other carcinomas were of undifferentiated origin. All papillomas had an intact basal cell layer with a clear border line and separated from the stroma. Total RNA was isolated from three papillomas and three carcinomas, and subjected to RNA-seq analysis. Reads were then mapped to the CRPV genome reference sequence (GenBank accession number NC_001541.1) which was linearized at nt 7345 (the stop codon of L1) and extended to nt 4377 (the ATG of L2) to include the viral PL in the URR and to cover the complete early region of CRPV (Fig. 2). The late region with L2 and L1 was excluded as the CRPV genome was cloned via SalI at the beginning of L2 and thus the viral late region was separated by vector sequences from the URR/early region. The numbers of mapped reads ranged from 45 560–83 630 for individual biopsies. In general, no obvious quantitative differences in transcription levels could be observed between papillomas and carcinomas (Fig. 2). Low levels of reads were found in all samples mapping to the region between nt 7527 and 7810, which was consistent with transcription of the URR exon where transcription starts at the PL (nt 7519–7523) and the RNA is processed at the SD site at nt 7810 (Wettstein et al., 1987). The drop in reads to almost background levels after nt 7809 suggested that the majority of transcripts initiated at PL were processed at SD7810. Reads were also present in all samples starting at nt 31 and their numbers increased continuously up to nt 225 where they plateaued and then slightly decreased until nt 900–920.
After that, read numbers increased again and sharply declined after nt 1371. This was consistent with the presence of the P1 (major start site: nt 95), P2 (major start site: nt 165) and P3 (major start site: nt 908) promoters (Nasseri & Wettstein, 1984; Wettstein, et al., 1987). The mean normalized counts for the long E6 (LE6) ATG were ~10.8-fold lower than for the SE6 ATG and 18.8-fold lower than for the E7 ATG in papillomas, suggesting that more E7 than full-length E6 protein was present. Low but constant levels of reads were found from nt 1372 to 3064, indicating the presence of transcripts encoding full-length E1. An increase in read levels occurred after nt 3065 which remained constant up to nt 3713 consistent with a SA site at nt 3065 (see below). At nt 3714, a sharp increase in reads to levels similar to the region between nt 900 and 1371 was observed until nt 4255–4280, after which a steady decline to background levels took place which was consistent with a polyadenylation sequence at nt 4348–4343. In addition to collinear reads, splice junctions were detected at different frequencies (Table 1). In the most abundant splice junction, SD1371 was linked to SA3714 which was described previously (Nasseri & Wettstein, 1984). The SD1371/SA3065 junction was detected on average 20-fold less often than SD1371/SA3714 and this RNA serves to express a full-length E2 protein (Wettstein et al., 1987). The third-most abundant splice junction was SD1751/SA3714 which was detected in five of six samples and gives rise to the E9^E2C fusion protein, which is the homologue of HPV E8^E2C proteins (Jeckel et al., 2003). In addition, two minor splice junctions were detected with similar frequencies which have not been previously described for CRPV. SD7810/SA3714 was observed in four of six samples and SD1751/SA3065 in five of six samples (Table 1). SD7810/SA3714 links the URR exon derived from the PL to the E2/E4 exon. These data indicated that neither papilloma- nor carcinoma-specific spliced transcripts exist at detectable levels. Furthermore, no strong indication could be found that viral transcription differed quantitatively between papillomas and carcinomas.

To investigate this by a different approach, quantitative (q)PCR experiments were carried out. Primer pairs were designed to detect the LE6 (nt 106–306), SE6 (nt 330–467), E7 (nt 1197–1325), E1 (nt 2032–2179), E2 (nt 3312–3426), and spliced RNAs that use the junctions SD7810/SA3714, SD1371/SA3065, SD1751/SA3065 and SD1751/SA3714. In addition to the biopsies used for RNA-seq, one papilloma and one carcinoma were included. Consistent with the RNA-seq experiments, the most abundant transcript encoded E7 followed by SE6, E2, E6 and E1 transcripts (Fig. 3). Furthermore, E7, SE6, E2 and E6

---

**Table 1. CPM normalized counts for different splice junctions in papillomas and carcinomas**

<table>
<thead>
<tr>
<th>Splice junction</th>
<th>Papilloma 1</th>
<th>Carcinoma 1</th>
<th>Papilloma 2</th>
<th>Carcinoma 2</th>
<th>Papilloma 3</th>
<th>Carcinoma 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD7810/SA3714</td>
<td>109.75</td>
<td>75.39</td>
<td>0</td>
<td>0</td>
<td>136.18</td>
<td>35.87</td>
</tr>
<tr>
<td>SD1371/SA3065</td>
<td>2019.32</td>
<td>1623.34</td>
<td>2629.68</td>
<td>2430.09</td>
<td>2133.42</td>
<td>1877.32</td>
</tr>
<tr>
<td>SD1371/SA3714</td>
<td>36 918.35</td>
<td>38 401.81</td>
<td>34 041.79</td>
<td>32 464.45</td>
<td>36 086.61</td>
<td>41 623.82</td>
</tr>
<tr>
<td>SD1751/SA3065</td>
<td>65.85</td>
<td>60.31</td>
<td>0</td>
<td>94.93</td>
<td>60.52</td>
<td>47.83</td>
</tr>
<tr>
<td>SD1751/SA3714</td>
<td>109.75</td>
<td>271.39</td>
<td>0</td>
<td>75.94</td>
<td>181.57</td>
<td>191.32</td>
</tr>
</tbody>
</table>

---
transcripts were expressed at slightly lower levels in carcinomas (80, 91, 68 and 63 %, respectively, of the levels in papillomas), but this was not statistically significant. The most abundant splice junction of the ones tested was SD1371/SA3065, followed by SD1751/SA3714, SD7810/SA3065 and SD1751/SA3065 (Fig. 3). The amounts of SD7810/SA3065 transcripts were slightly increased to 129 %, whereas SD1371/SA3065, SD1751/SA3065 and SD1751/SA3714 transcripts were slightly decreased (72, 74 and 72 % of the levels in papillomas). These minor differences were also not statistically significant. Generally, the RNA-seq data were in accordance with the qPCR analyses and confirmed that no major transcriptional changes occurred during the progression of CRPV-induced tumours in vivo.

**DISCUSSION**

Here, we investigated whether quantitative or qualitative changes in viral transcription can be detected by deep sequencing of RNA derived from CRPV-induced papillomas or carcinomas. Our data largely confirm and extend the transcriptional patterns of CRPV in vivo (Jeckel et al., 2003; Nasser & Wettstein, 1984; Phelps et al., 1985; Wettstein et al., 1987; Zeltner et al., 1994). The P2 and P3 promoters, which are downstream of the LE6 ATG and upstream of the E7 ATG, respectively, have the highest activities, whereas the P1 promoter immediately upstream of the LE6 ATG is ∼10-fold less active. The PL promoter is 10- and 400-fold less active than the P1 and P3 promoters, respectively. We also were able to detect similar levels of reads throughout the E1 gene, indicating that low levels of transcripts exist that encode the full-length E1 protein. Furthermore, we detected the known splice junctions SD1371/SA3065, SD1371/SA3714 and SD1751/SA3714 by RNA-seq. We also identified for the first time, to the best of our knowledge, the splice junctions SD7810/SA3714 and SD1751/SA3065, which are present at even lower levels than the SD1751/SA3714 junction. We validated the RNA-seq data by qPCR using nine different amplicons to detect collinear and spliced transcripts, and also increased the number of biopsies. Nevertheless, neither papilloma- nor carcinoma-specific CRPV transcripts could be detected by RNA-seq nor were quantitative changes of the different viral transcripts obvious by RNA-seq or qPCR. As even novel, rare splicing events were consistently detected in the majority of samples, we believe it is very unlikely that additional viral transcripts exist that make a major contribution to progression.

We have identified two novel splice junctions in which previously described SDs and SAs are connected. The splice junction SD7810/SA3714 links the URR exon transcribed by the PL to the E2/E4 exon. Comparable transcripts have also been described for bovine papillomavirus (BPV) 1-, HPV1a-, HPV5- and HPV8-positive skin lesions, and are most likely involved in the expression of the L1 and L2 capsid proteins (Baker & Howley, 1987; Chow et al., 1987; Haller et al., 1995; Stubenrauch et al., 1992). The splice junction SD1751/SA3714 links the E9 exon to the E2 exon, and has been described previously for BPV1, HPV5 and HPV16 (Choe et al., 1989; Sankovski et al., 2014; Straub et al., 2015). This transcript may encode an E9^E1C fusion protein and/or E2. However, the inactivation of SD3714 in the context of the CRPV genome did not give rise to a phenotype, suggesting that this spliced transcript does not have a major impact on tumour formation by CRPV in vivo (Jeckel et al., 2003). In summary, our data confirm and extend the transcriptional pattern of CRPV in vivo, and indicate that the transcriptional patterns of papillomaviruses are highly conserved (Fig. 2).

A recent study addressed the transcriptional patterns of HPV16 in CIN2, CIN3 and squamous cell carcinoma (SCC) human cervical lesions by RNA-seq comparable to our study (Chen et al., 2014). The authors came to the conclusion that a major transcriptional switch occurs from CIN2 to CIN3 lesions. Transcripts encoding E6 and E7 are upregulated, whereas E2 and L1 transcripts are strongly downregulated in CIN3 and SCC lesions. However, a study by Häfner et al. (2008) using qPCR to quantify E6 or E7 transcript levels did not observe differences between precursor lesions and cervical cancer samples. These differences are most likely not due to the different methodologies used to quantify viral transcripts as our data indicate that RNA-seq and qPCR analyses of the same RNA samples give rise to very similar quantitative results.
The expression of hrHPV E6 and E7 in tissue culture models results in genomic instability (Moody & Laimins, 2010). Consistent with this, a large-scale study identified recurrent genomic alterations in cervical carcinomas and HPV-induced head and neck SCCs (Cancer Genome Atlas, 2013; Ojesina et al., 2014). Thus, the continuous expression of hrHPV E6 and E7 (irrespective of their levels) most likely induces mutations which result in the progression to invasive cancer. Future studies are required to address if similar genomic alterations occur in CRPV-induced cancers as described for HPV-induced SCCs which would allow evaluating their functional contribution to the development of the malignant phenotype in vivo.

METHODS

Animal biopsies. All experimental animal procedures were approved by the Regierungspräsidium Tübingen, Baden-Württemberg, Germany (permit number H1/08) and experiments were performed according to institutional guidelines. The backs of New Zealand white rabbits (Charles River) were infected with purified plasmid DNA by particle-mediated gene delivery as described previously (Jeckel et al., 2003; Probst et al., 2013). Papilloma and carcinoma biopsies were obtained from our biopsy collection of WT CRPV (CRPV-pLAII, 2

Table 2. Primers used for qPCR

<table>
<thead>
<tr>
<th>Primer designation</th>
<th>RNA target</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRPV 7794–7810/3714-3715 F</td>
<td>SD7810/SA3714</td>
<td>GCCAGGTTGTCATGACTCT</td>
</tr>
<tr>
<td>CRPV 3831 R</td>
<td>SD7810/SA3714</td>
<td>CTGGTGACCTTGACCTGTC</td>
</tr>
<tr>
<td>CRPV 1723 F</td>
<td>SD1751/SA3065</td>
<td>TGGTGTGTAGTCTGACAGG</td>
</tr>
<tr>
<td>CRPV 3281 R</td>
<td>SD1751/SA3065</td>
<td>GGTAGCCGCATGCTGTAGT</td>
</tr>
<tr>
<td>rGAPDH F 448</td>
<td>GAPDH</td>
<td>TGGCACCAACACTGTAGTC</td>
</tr>
<tr>
<td>rGAPDH R 535</td>
<td>GAPDH</td>
<td>GGCTGCGACCTGTGCTAGG</td>
</tr>
<tr>
<td>CRPV LE6 For 106</td>
<td>LE6</td>
<td>AAGCAGTTGGCAGGATT</td>
</tr>
<tr>
<td>CRPV LE6 Rev 306</td>
<td>LE6</td>
<td>ATACGGCCACCCCTTATGC</td>
</tr>
<tr>
<td>CRPV SE6/LE6 For 330</td>
<td>SE6</td>
<td>CTCGAGAGACTGCTGACTTGT</td>
</tr>
<tr>
<td>CRPV SE6/LE6 Rev 467</td>
<td>SE6</td>
<td>CCTCCGAGAACTGTCACAAG</td>
</tr>
<tr>
<td>CRPV SE6/LE6/E7 For 1197</td>
<td>E7</td>
<td>TAGACAGTTGGCAGGAGAGG</td>
</tr>
<tr>
<td>CRPV SE6/LE6/E7 Rev 1325</td>
<td>E7</td>
<td>ATGCCGATAGCCAGCTATC</td>
</tr>
<tr>
<td>CRPV E1 For 2032</td>
<td>E1</td>
<td>ATGGTTCATCTGCTGCTG</td>
</tr>
<tr>
<td>CRPV E1 Rev 2179</td>
<td>E1</td>
<td>ACAATGCAACAGGACACCT</td>
</tr>
<tr>
<td>CRPV E2 For 1306</td>
<td>SD1371/SA3065</td>
<td>AATCGACTGCTATCCGCATC</td>
</tr>
<tr>
<td>CRPV E2 Rev 3085</td>
<td>SD1371/SA3065</td>
<td>AGTTCTAAGGTGGACCAGA</td>
</tr>
<tr>
<td>CRPV E2N-term 3312 For</td>
<td>E2</td>
<td>GTGTCGAAAGCGACCATAG</td>
</tr>
<tr>
<td>CRPV E2N-term 3426 Rev</td>
<td>E2</td>
<td>AGGCGCTTCAACCTTCTTC</td>
</tr>
<tr>
<td>CRPV E9/E2C_1730 For</td>
<td>SD1751/SA3714</td>
<td>TCCTGGTGTGACCTGTTACA</td>
</tr>
<tr>
<td>CRPV E9/E2C_3965 Rev</td>
<td>SD1751/SA3714</td>
<td>TCGTAGTCTTCGTGCAACG</td>
</tr>
</tbody>
</table>
National Center for Biotechnology Information database and passed to the STAR aligner. After mapping, potential PCR duplicates were removed using SAMtools rmdup (Li et al., 2009). To decrease the discovery of false-positive splices, a minimum overlap of 10 bases at each side of the spliced region was required for spliced reads and all splice sites covered by less than three reads were removed. Read counts were normalized using the CPM (counts per million mapped reads) method to allow for a comparative analysis of mapped reads for the different samples. Read coverage at each genomic position in the CRPV genome was accessed using the IGV (Integrative genomics viewer) genome browser (Robinson et al., 2011).

**Plasmids.** Plasmids pSG CRPV E2, pSG CRPV E9^E2C and CRPVshLuc-pLAII have been described previously (Jeckel et al., 2002, 2003; Leiprecht et al., 2014; Probst et al., 2013). The coding sequence of rabbit glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was amplified with EcoRl- and BamHI-linked primers (rabbitGAPDH For_EcoRI: 5'-CCGAGCCGAATTCATGTTGAAGGTCTCGAGTGAA-3'; rabbitGAPDH Rev_BamHI: 5'-GCAATTGGATCCATCCCTGGAGGCGCATGT-3', restriction sites in italics) from cDNA derived from CRPV E6/E7 immortalized rabbit keratinocytes and cloned into EcoRl/BamHI-digested pSG5, resulting in pSG-rGAPDH. To obtain plasmid pGL3b-CRPV E2 spliced (CRPV nt 810–1371/3065–3387), the viral fragment was amplified with primers 330 F/3426 and pGL3b-CRPV E2 spliced, pGL3b-CRPV E9^E2C spliced, pGL3b-CRPV E2 spliced, and ligated into pGL3-basic. To obtain plasmids pET-706F-7182/3714-4295 and pET-706F-7182/3714-4295, cDNA from a CRPV-induced immortalized keratinocyte clone (CRPV immortalization of normal rabbit keratinocytes and reduction of pRb levels, while E6 and E7 proteins of the CRPV genome were accessed using the IGV (Integrative genomics viewer) genome browser (Robinson et al., 2011).

**RT-qPCR.** cDNA (0.6 μg) was reverse transcribed using a Quantitect Reverse Transcription kit (Qiagen) and amplified with CRPV-specific primers. qPCR analyses were carried out with 30 ng cDNA in duplicate reactions using 0.3 μM gene-specific primers (Table 2), 1 × LightCycler 480 SYBR Green I Master (Roche Applied Science) in a total volume of 20 μl. Reactions were performed in a LightCycler 480 (Roche Applied Science) using an thermal profile of 10 min at 95 °C followed by 45 cycles for 10 s at 95 °C, 30 s at 60 °C, 90 °C and cooling 30 s at 40 °C, and analysed using LightCycler 480 software v1.5 (Roche Applied Science). Expression levels were determined using standard curves of plasmids CRPVshLuc-pLAII, pSG CRPV E2, pSG CRPV E9^E2C, pGL3b-CRPV E2 spliced, pGL3b-CRPV E2 spliced, and ligated into pGL3-basic. To obtain plasmids pET-706F-7182/3714-4295, cDNA from a CRPV-induced immortalized keratinocyte clone (CRPV immortalization of normal rabbit keratinocytes and reduction of pRb levels, while E6 and E7 proteins of the CRPV genome were accessed using the IGV (Integrative genomics viewer) genome browser (Robinson et al., 2011).

**ACKNOWLEDGEMENTS.**

This work was funded by a grant from the Deutsche Forschungsgemeinschaft (DFG SFB 773 B4) to T. I.

**REFERENCES.**


