



Klymenko, T., Hernandez-Lopez, H., MacDonald, A.I., Bodily, J. M., and Graham, S. V. (2016) Human papillomavirus E2 regulates SRSF3 (SRp20) to promote capsid protein expression in infected differentiated keratinocytes. *Journal of Virology*, 90(10), pp. 5047-5058.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/117176/>

Deposited on: 01 April 2016

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

1

2 **Human papillomavirus E2 regulates SRSF3 (SRp20) to promote**
3 **capsid protein expression in infected differentiated keratinocytes.**

4 Klymenko, T.³, Hernandez-Lopez, H.⁴, MacDonald, A.I.¹, Bodily J.M.² and Graham,
5 S.V.^{1*}

6 ¹ MRC-University of Glasgow Centre for Virus Research; Institute of Infection, Immunity and
7 Inflammation; College of Medical Veterinary and Life Sciences, University of Glasgow,
8 Garscube Estate, Glasgow, G61 1QH, Scotland, UK

9

10 ² Louisiana State University Health Sciences Center, Department of Microbiology and
11 Immunology, B 2-213, 1501 Kings Highway, Shreveport, LA 71130, USA

12

13 **Current addresses**

14 ³ Barts Cancer Institute, Queen Mary, University of London, John Vane Science Centre,
15 Charterhouse Square, London EC1M 6BQ

16 ⁴ Janssen de México, 247 Av. Miguel Angel de Quevedo, Col. Romero de Terreros, México,
17 D.F.C.P. 04310

18

19 ***Corresponding author.**

20 Rm 253, Jarrett Building, Garscube Estate, University of Glasgow, Glasgow, G61 1QH,
21 Scotland, UK.

22 Tel: 44 141 330 6256; Fax: 44 141 330 5602;

23 e-mail: Sheila.Graham@gla.ac.uk

24

25 **Running title: HPV16 late gene expression is controlled by SRSF3.**

26 Key words: HPV, E2, SRSF3 (SRp20), capsid proteins, RNA processing

27 Abstract word count: 212

28 Text Word count: 6473

29 **ABSTRACT**

30 The human papillomavirus (HPV) life cycle is tightly linked to differentiation of the infected
31 epithelial cell suggesting a sophisticated interplay between host cell metabolism and virus
32 replication. Previously we demonstrated in differentiated keratinocytes *in vitro* and *in vivo*
33 that HPV16 infection caused increased levels of the cellular SR splicing factors (SRSFs)
34 SRSF1 (ASF/SF2), SRSF2 (SC35) and SRSF3 (SRp20). Moreover, the viral E2 transcription
35 and replication factor that is expressed at high levels in differentiating keratinocytes could
36 bind and control activity of the SRSF1 gene promoter. Here we reveal that E2 proteins of
37 HPV16 and HPV31 control expression of SRSFs 1, 2 and 3 in a differentiation-dependent
38 manner. E2 has the greatest trans-activation effect on expression of SRSF3. siRNA
39 depletion experiments in two different models of the HPV16 life cycle (W12E and NIKS16)
40 and one model of the HPV31 life cycle (CIN612-9E) revealed that only SRSF3 contributed
41 significantly to regulation of late events in the virus life cycle. Increased levels of SRSF3 are
42 required for L1 mRNA and capsid protein expression. Capsid protein expression was
43 regulated specifically by SRSF3 and appeared independent of other SRSFs. Taken together
44 these data suggest a significant role of the HPV E2 protein in regulating late events in the
45 HPV life cycle through transcriptional regulation of SRSF3 expression.

46

47 **IMPORTANCE**

48 Human papillomavirus replication is accomplished in concert with differentiation of the
49 infected epithelium. Virus capsid protein expression is confined to the upper epithelial layers
50 so as to avoid immune detection. In this study we demonstrate that the viral E2 transcription
51 factor activates the promoter of the cellular SRSF3 RNA processing factor. SRSF3 is
52 required for expression of the E4^{L1} mRNA and so controls expression of the HPV L1
53 capsid protein. Thus we reveal a new dimension of virus-host interaction crucial for
54 production of infectious virus. SRSF proteins are known drug targets. Therefore, this study
55 provides an excellent basis for developing strategies to regulate capsid protein production in
56 the infected epithelium and production of new virions.

57

58

59 **INTRODUCTION**

60

61 Human papillomaviruses (HPVs) infect epithelia causing benign lesions or warts. For the so-
62 called “high risk” (HR) HPVs such as HPV type 16 (HPV16), the most prevalent HPV,
63 persistent infection causes cervical and other anogenital lesions and head and neck lesions
64 that may progress to cancer [1]. Although prophylactic vaccines against the most prevalent
65 HR-HPVs are available, therapies are still required to treat infected individuals who are not
66 vaccinated. A more comprehensive understanding of the HR-HPV replication cycle could
67 help in the development of novel therapeutic approaches.

68

69 The HPV16 life cycle is tightly linked to differentiation of the epithelium the virus infects [2].
70 Initial infection is within basal epithelial cells where the episomal viral genome is maintained
71 at around 50-100 nuclear copies [3]. Differentiation of infected epithelial cells leads to
72 activation of early, then late gene expression [4]. The viral replication and transcription
73 factor, E2 is expressed at greatest levels in the mid to upper layers of the epithelium [5,6]
74 where together with the viral DNA helicase, E1 it facilitates vegetative viral DNA replication
75 leading to production of thousands of viral genome copies [7]. The viral late proteins,
76 including the capsid proteins L1 and L2 are synthesized in the uppermost granular layer of
77 the epithelium to encapsidate the newly replicated genomes [4]. This spatial restriction of
78 production of the highly immunogenic capsid proteins is important as it avoids triggering an
79 immune response as a result of low immune surveillance in the upper epithelial layers.

80

81 Expression of virus capsid proteins is known to be controlled not only at the level of
82 transcription initiation but also at various post-transcriptional levels including
83 polyadenylation, alternative splicing, nuclear export, mRNA stability and translation [2,8–12].
84 Notably, at least thirteen mRNAs are produced late in the virus life cycle that contain open
85 reading frames encoding the capsid proteins [10]. Seven mRNAs contain the L1 but not the

86 L2 open reading frame and six mRNAs contain both L1 and L2 open reading frames.
87 Alternative splicing regulates the proportions of the various late mRNAs encoding the capsid
88 proteins [12]. In particular, viral capsid protein L1 is thought to be encoded by an E4^{L1}
89 mRNA produced by splicing from a splice donor site at the end of the E4 open reading frame
90 to the splice acceptor site at the start of the L1 coding region while the L2 capsid protein is
91 encoded by a read through L2L1 RNA [10].

92

93 Constitutive splicing is the process whereby introns are removed from pre-mRNAs and
94 exons are spliced together to form a protein-coding mRNA. Alternative splicing is a
95 mechanism used by mammalian and viral genomes to maximise coding potential [13,14]. A
96 gene is transcribed to give a single primary transcript but from this precursor RNA (pre-
97 mRNA) different mature mRNA isoforms are produced by differential inclusion or exclusion
98 of exons and introns. Each alternatively spliced mRNA isoform can encode a different
99 protein. Alternative splicing is directly regulated by the following two classes of proteins, SR
100 proteins (serine-arginine-rich splicing factors, SRSFs) and hnRNP proteins (heterogenous
101 ribonucleoproteins) [15]. SR proteins can bind to exonic sequence enhancers (ESEs) to
102 stimulate recognition of adjacent splice sites by the splicing machinery while hnRNP proteins
103 recognise exonic splicing silencers (ESSs) to repress splice site utilisation [16]. Therefore
104 the combinations of *cis*-acting ESEs and ESSs in open reading frames together with the
105 relative concentrations of *trans*-acting SR and hnRNP proteins that can access these sites
106 determine the ultimate mRNA isoforms produced from a single gene.

107

108 Aside from splicing, SR proteins can potentially regulate other processes that control protein
109 production from an mRNA including transcription elongation, polyadenylation, nonsense-
110 mediated decay, nuclear export mRNA stability, and translation [17]. There are nine
111 classical SR proteins described, named SRSF1-9. Previous data have demonstrated the
112 importance of a number of these proteins in production of HPV mRNAs [18]. SRSF1
113 (SF2/ASF), SRSF3 (SRp20) and SRSF9 (SRp30c) have been shown to bind viral mRNAs in

114 the E4 open reading frame [19–22]. SRSF2 (SC35) and SRSF3 have been shown to control
115 production of the viral E6 and E7 oncoprotein mRNAs [19,23]. Previous studies that have
116 examined SR protein regulation of viral RNA splicing in tumour cells transiently expressing
117 subgenomic reporter constructs have demonstrated that SRSF1 controls expression of early
118 and late viral mRNAs [20,22] while SRSF3 can regulate BPV1 and HPV16 early and late
119 gene expression [19]. SRSF9 can also control HPV16 late gene expression [21].

120

121 Previously we demonstrated *in vitro* and *in vivo* that HPV16 infection up-regulated
122 expression of a specific subset of SR proteins in differentiating epithelial cells, these being
123 the three smallest; SRSF1, SRSF2, and SRSF3, [24]. Subsequently we showed that the
124 viral transcription factor E2 activated the SRSF1 promoter via its *trans*-activation domain
125 [25]. Here we demonstrate that HPV16 E2 also specifically *trans*-activates the promoters of
126 SRSF2 and SRSF3. We have extended the work to show that HPV31 E2 can also
127 transcriptionally control SR protein expression in an epithelial differentiation-stage specific
128 manner. siRNA depletion experiments were used to determine which HPV-regulated SR
129 protein(s) was important for HPV16 late mRNA and capsid protein production. Data revealed
130 that SRSF3 depletion resulted in significantly reduced levels of L1 protein suggesting that
131 E2-regulated SRSF3 is required for capsid protein expression. SRSF3 was required for
132 expression of the viral E4^{L1} spliced mRNA but appeared to repress expression of the L2L1
133 unspliced read through RNA. Taken together, these data indicate that the E2 protein links
134 the viral replication cycle to epithelial differentiation via SRSF3, a key cellular regulator of
135 HR-HPV gene expression.

136

137 **EXPERIMENTAL**

138

139 ***Cell lines***

140

141 W12E cells are non-tumour cervical epithelial cells (clone 20863: [26]) that if maintained at
142 low passage (<17) contain 50-100 nuclear episomal copies of the HPV16 genome. NIKS16
143 cells are normal immortalised keratinocytes (NIKS) stably transfected with the HPV16
144 genome. Clone 2L maintains episomal HPV16 genomes [27]. CIN612-9E cells are cervical
145 keratinocytes containing episomal HPV31 genomes [28]. All three lines form tissues in raft
146 culture that mimic a low grade cervical lesion. W12E, NIKS, NIKS16 and CIN612-9E cells
147 were cultured in E-medium and differentiated as described (2×10^5 cells per 10 cm plate) on
148 mitomycin C-treated J2 3T3 mouse fibroblasts [10,26]. Cells were differentiated by growing
149 to high density in the presence of 1.88 mM Ca^{2+} as previously described [26]. The 3T3 cells
150 were grown in DMEM with 10% donor calf serum. Prior to harvesting for RNA or protein
151 preparation, 3T3 cells were removed by trypsinisation and cells layers washed twice with
152 PBS. U2OS osteosarcoma cells and U2OS clones stably expressing HPV16 E2 (U2OSA4,
153 U2OSB1) were cultured in DMEM with 10% foetal calf serum. All cells were maintained
154 under humidified 5% CO_2 95% air at 37°C.

155

156 ***Cell transfection and siRNA depletion***

157 Cells were seeded at 2×10^5 per 10 cm dish and grown for 4 days on mitomycin C-treated J2
158 3T3 mouse fibroblasts as describe above. At day 4 fibroblasts were removed by brief
159 trypsinisation. Keratinocytes were washed twice with PBS and transfected with plasmids
160 using Lipofectamine 3000 according to the manufacturer's protocol. 8 h after transfection
161 freshly prepared mitomycin C-treated J2 3T3 mouse fibroblasts were added back to the
162 keratinocyte cultures in fresh E-medium. After 72 hours cells were harvested into 400 μl
163 protein loading buffer (125 mM Tris (pH 6.8), 4% SDS, 20% glycerol, 10% mercaptoethanol
164 and 0.006% bromophenol blue, fresh protein inhibitor cocktail (Roche, UK). Lysates were

165 passed through a 21 gauge needle 15 times then sonicated in a sonibath for 3 x 30 sec.
166 Cells were seeded at 2×10^5 per well in a six well plate 24 h prior to transfection in antibiotic-
167 free medium. siRNA (10 nM) and Lipofectamine RNAi MAX (Invitrogen) were diluted in Opti-
168 MEM serum-free medium (Invitrogen). Cells were transfected for 48 h according to the
169 manufacturer's protocol. siGENOME SMART pools (consisting of four siRNAs designed to
170 minimise off-target effects) specific for each of the SRSFs tested were purchased from
171 Dharmacon. Transfection efficiencies as calculated by co-transfection with siGLO
172 (Dharmacon) or a GFP expression plasmid were between 70-80%.

173

174 ***Cloning the SRSF3 expression vector***

175 pEGFPSRp20 was a gift of Dr R. Sandri-Goldin. The insert fragment from pEGFPSRp20
176 was cleaved out using BamHI/EcoRI restriction digest. The fragment was then inserted using
177 the same sites into pcDNA3.1 to give plasmid pcDNA3.1SRSF3.

178

179 ***Cloning the SRSF promoters***

180 SRSF promoter regions were amplified from HeLa cell DNA. For SRSF1pr the forward
181 primer was 5'-GATC***CTCGAG***GTTACGGTTCTCACATCCATTTTGC-3' and the reverse
182 primer was 5'-GTGC***AAGCTT***TCTCCCGCGGCCCTCCAAAATG-3' (amplifying nucleotides
183 -886 to +143 relative to the transcription initiation site at +1). For SRSF2pr the forward
184 primer was 5'-GGGT***GGTACCG***TACAGCTCTCCTCGGGGCGAAG-3' and the reverse primer
185 was 5'-GTAC***AAGCTT***TCTCAGGCAGTTGCCTTCGCG-3' (amplifying nucleotides -985 to
186 +96 relative to the transcription initiation site at +1). For SRSF3pr the forward primer was 5'-
187 GATC***GGTACCG***GCGGCTCTGTCTTCGTAAGGG-3' and the reverse primer was 5'-
188 GTGC***AAGCTT***TCTCTCACTACCCGGCGTCC-3' (amplifying nucleotides -720 to +81
189 relative to the transcription initiation site at +1). For SRSF7pr the forward primer was 5'-
190 CATC***CTCGAG***ACCAACTAGGCCTGCTTTCC-3' and the reverse primer was 5'-
191 GTGC***AAGCTT***AAACAGCCAAGAAACGACGC-3' (amplifying nucleotides -1036 to +69)
192 relative to the transcription initiation site at +1). Nucleotides in bold italics indicate the Xho I

193 of Hind III restriction enzyme sites that were used for cloning. A four base pair overhang was
194 added to the end of each primer sequence to facilitate restriction enzyme cleavage. PCR
195 was performed with High Fidelity Platinum® Taq polymerase (Invitrogen). Single amplicons
196 were obtained with the correct theoretical length. PCR products were ligated into pGL3 basic
197 vector (Promega) and transformed into super competent *E. Coli* DH5α (Invitrogen). Insert
198 sequences of clones obtained were confirmed by DNA sequencing.

199 ***Transcription assays***

200 U2OS and U2OS clones A4 and B1 cells that stably express HPV16-E2 [29]) were seeded in
201 10mm well plates at 1×10^5 cells per well without antibiotics 24 h prior to transfection.
202 Transfections were carried out using Lipofectamine 2000 (Invitrogen). An EGFP expression
203 vector (pMAXEGFP) was used as a transfection efficiency control using. Protein was
204 extracted 48-72 h post-transfection as follows. Cells were washed twice in ice cold PBS after
205 the removal of the culture medium. NP40 lysis buffer (0.5% NP40, 150mM NaCl, 50mM Tris
206 HCl pH 8) with protease (Roche Diagnostics) and phosphatase (Roche Diagnostics)
207 inhibitors was added to the cells and the cells scraped into the buffer on ice. The lysed cells
208 were transferred to a 1.5ml eppendorf tube and incubated on ice for 30 min with periodic
209 vortexing. The extracts were centrifuged at 10000g for 10 min and the supernatant was
210 subsequently stored at -80°C. A luciferase assay system (Promega) and a GloMax–Multi
211 Detection System luminometer were used to detect luciferase activity according to
212 manufacturer's instructions. Samples were normalized by protein concentration. At least 3
213 independent experiments were performed.

214

215 ***Cloning the HPV31 genome E2:I73L mutant***

216 The I73L mutation in the HPV31 genome was created using the QuickChange XL site
217 directed mutagenesis kit (Agilent) using the plasmid pLit-HPV31 as template and the primers
218 5'GCCAAAGCCT TACAAGCTcT TGAACTACAA ATGATGTTGG3' and 5'CCAACATCAT

219 TTGTAGTTCA AgAGCTTGTA AGGCTTTGGC 3' according to the manufacturer's
220 directions. Primary human foreskin keratinocytes were transfected with recircularized HPV31
221 genomes and expanded following drug selection as previously described [30].

222

223 ***Western blotting***

224 Cells were scraped into protein-loading buffer (125 mM Tris (pH 6.8), 4% SDS, 20%
225 glycerol, 10% mercaptoethanol and 0.006% bromophenol blue with fresh protein inhibitor
226 cocktail (Roche, UK)). Lysates were passed through a 21 gauge needle 15 times then
227 sonicated in a sonibath for 3 x 30 sec. Subsequently 10 µg protein (or 5, 10 20 µg where
228 indicated) was resolved by polyacrylamide gel electrophoresis using NOVEX precast 4-12%
229 gradient gels (Invitrogen) then transferred to nitrocellulose membrane using an i-blotter
230 (Invitrogen). The membrane was preincubated for 1 h at room temperature with 5% dried
231 milk powder in PBS-0.1% Tween, before overnight incubation at 4°C with diluted primary
232 antibody in PBS-Tween, 1% dried milk powder. The following antibodies were used: SRSF1
233 (clone 96) (Invitrogen, 32-4500) 1:1000. SRSF2 (BD Pharmingen, 556363) 1:250. SRSF3
234 (clone 7B4) (Invitrogen, 33-4200) 1:250. SRSF7 (clone 98) gift of Dr James Stevenin
235 (IGBMC, Strasbourg, France) 1:100. GAPDH (clone 6CS) (Biosdesign International,
236 H86504M) 1:5000. Involucrin (Sigma 19018) 1:1000. HPV16 E2 (clone TVG261) (Abcam
237 ab17185) 1:250 and HPV L1 (clone K1H8) (Dako M3528). Mab104 (used neat) detects a
238 phosphoepitope on SRSFs 1-6) (ATTC hybridoma supernatant). Because SRSFs1 and 2
239 have a similar apparent molecular mass on SDSPAGE, as well as probing with Mab104, we
240 also probed blots with specific monoclonal antibodies against these SR proteins as above
241 Horseradish peroxidase (HRP)-conjugated secondary antibodies (Pierce-ECL) were diluted
242 1:2000 in PBS-Tween and incubated for 1 h. Blots were developed using Pierce enhanced
243 chemi-luminescence (ECL) kit and exposed to Kodak X-OMAT film. Blots were quantified by
244 scanning at 300 dpi and image analysis using Image J.

245

246 ***Immunohistochemistry***

247 Archival paraffin-embedded cervical biopsy samples were obtained with ethical permission
248 (Glasgow Royal Infirmary: RN04PC003). Diagnosis was made by a gynaecological
249 pathologist. HPV presence was confirmed by PCR. Immunohistochemistry was carried out
250 by the University of Glasgow Veterinary Diagnostic Services. SRSF3 antibody (clone 7B4)
251 (Invitrogen, 33-4200) was used at a dilution of 1:100

252

253 ***RNA extraction***

254 Cells were scraped into TRIzol reagent (Invitrogen) and total RNA was extracted according
255 to the manufacturer's protocol. Polyadenylated RNA was isolated using an oligo-dT-based
256 mRNA extraction kit (Oligotex, Qiagen) according to the manufacturer's instructions. DNA
257 was removed from all RNAs using the Promega RQ1 DNase kit according to the
258 manufacturer's instructions. DNase-treated RNA was reverse transcribed using the
259 Superscript III kit (Invitrogen) according to the protocol provided.

260

261 ***PCR***

262 cDNA was amplified using 200 nM primers, 200 μM dNTPs, 1.5 mM MgCl₂ and 2 units Taq
263 polymerase (Invitrogen). Primers for E4^ΔL1 amplification were E4 forward (nts 3518-3540)
264 5'-GTTGTTGCACAGAGACTCAGTGG-3' matched with L1 reverse (nts 6918-6997) 5'-
265 GACAAGCAATTGCCTGGGTTAC-3'. Primers for L2L1 amplification were L2 forward (nts
266 5465-5475) 5-GTATCAGGTCCTGATATACCC-3', matched with L1 reverse B (nts 5690-
267 5669) 5'-TACTGGGACAGGAGGCAAGTAG-3', GAPDH forward 5'-
268 TCCACCACCCTGTTGCTGTA-3'; GAPDH reverse 5'- ACCACAGTCCATGCCATCAC-3'.
269 PCR products were separated on a 6% acrylamide gel and post-stained with ethidium
270 bromide.

271

272

273

274 **qPCR**

275 cDNA was amplified using an Applied Biosystems 7500 qPCR machine. The Stratagene
276 Brilliant qPCR mastermix was used for qPCR reactions. cDNA (100ng) was amplified in
277 each reaction in triplicate and three different experiments were performed. Reaction mixes
278 (25 μ l) contained 1xMastermix (Stratagene), 900 nM primers, 100 nM probe, 300 nM
279 reference dye (Stratagene). E4 forward primer (nts 3594-3615) 5'-
280 CTGTAATAGTAACACTACACCC-3' E4^{L1} probe (nts 3620-3631:5637-5656) 5'-
281 TACATTTAAAAGTGTCTCTTTGGCTGCCTAG -3'. L2 forward primer (nts 5547-5569) 5'-
282 CAATTATTGCTGATGCAGGTGAC-3', L2L1 probe (nts 5600-5622) 5'-
283 CGAAAACGACGTAAACGTTTACC -3'. L1 reverse (nts 5690-5669) 5'-
284 TACTGGGACAGGAGGCAAGTAG-3'. The amplification protocol was 95 °C for 15 sec, 50°C
285 for 60 sec for 40 cycles. Expression was quantified using $\Delta\Delta C_t$ relative to GAPDH values.

286

287 **RESULTS**

288

289

290 **HPV E2 regulates the promoters of the genes encoding SRSF1-3**

291 Using cell line models of epithelial differentiation and HPV infection and backed up by data
292 from virus-infected patient tissue, we and others have demonstrated that SRSF1 (SF2/ASF)
293 SRSF2 (SC35) and SRSF3 (SRp20) are specifically upregulated in HPV-infected,
294 differentiated epithelial cells [24,25,31,32]. For SRSF1, we showed previously that HPV16
295 E2 controlled the promoter of the SRSF1 gene by interacting within a region from -565 to -
296 363 nts with respect to the transcription initiation site (this was originally named -689 to -482
297 in our previous paper [25] according to the numbering of a previously sequenced version of
298 the SRSF1 promoter). Transcription activation was through the E2 *trans*-activation domain
299 [25]. E2 was only able to trans-activate the SRSF1 promoter in the A4 clone of the
300 osteosarcoma cell line U2OS that stably expresses low E2 levels (Figure 1A) [25, 29]. We
301 proposed previously that E2 may begin to trans-repress the SRSF1 promoter at
302 higher concentrations as it has been shown to do for the BPV4 long control region
303 promoter [25]. To examine whether E2 also controlled expression of the genes encoding
304 the other two SR proteins, SRSF2 and SRSF3, both of which are upregulated in HPV-
305 infected cells, we cloned their gene promoters into luciferase expression (pGL3-luciferase)
306 vectors for transcription analysis. The SRSF1 promoter (previously assayed in
307 chloramphenicol acetyl transferase (CAT) construct) was cloned in the same vector in order
308 to use as a known E2-regulated control. The promoter for SRSF7 (9G8) was also cloned and
309 analysed in order to compare an SRSF family member that we have shown is not
310 upregulated during differentiation of HPV-infected cervical epithelial cells [24]. A non-SR
311 protein promoter control, the hTERT promoter whose expression is down regulated by HPV
312 E2 [33], and a pGL3 empty vector with no promoter sequence was also used. U2OS
313 subclones A4 and B1 , stably express different amounts of HPV16 E2 protein: A4 expresses
314 low and B1 expresses high levels (Figure 1A) [25,29]. Transient transfections of the different
315 luciferase expression vectors were performed in U2OS, U2OSA4 and U2OSB1 cells. An

316 EGFP expression vector was co-transfected in each transfection experiment in order to
317 normalise transfection efficiency. Figure 1B shows quantification of luciferase expression
318 from the different promoters indicated on the X-axis. Control transfections of the
319 promoterless pGL3 vector showed very low luciferase activity (pGL-3 empty). hTERT
320 promoter activity was low in all the experiments. This is due to U2OS cells displaying low
321 telomerase activity [34]. Despite the low levels of reporter gene expression some decrease
322 was detected in the U2OSB1 cell line that expresses higher levels of E2 as expected [33].
323 Similar to what we reported previously, the SRSF1 promoter was *trans*-activated by E2
324 protein in the U2OSA4 cell clone which expresses low levels of HPV16 E2. However, the
325 higher levels of E2 expressed by the U2OSB1 cell clone did not upregulate the SRSF1
326 promoter confirming our previous data [25]. The SRSF2 promoter also showed a small but
327 statistically significant upregulation by HPV16 E2 protein. Under control of the SRSF2
328 promoter, luciferase activity was increased 1.35-fold in U2OSA4 cells and 1.24-fold in
329 U2OSB1 cells ($p < 0.05$). However, a greater increase in activity of the SRSF3 promoter than
330 that observed for SRSF1 and SRSF2 in the presence of E2 was observed. The basal activity
331 of the SRSF3 promoter in U2OS cells was lower than either the SRSF1 or 2 promoters and
332 HPV E2 protein gave a greater *trans*-activation response, showing a larger absolute
333 difference in fold change in luciferase activity. Luciferase activity was increased 2.93-fold in
334 U2OSA4 cells and 2.69-fold in U2OSB1 cells ($p < 0.001$) (Figure 1B). In contrast to the
335 repressive effect seen on SRSF1 promoter activity, high levels of E2 expression did not
336 abrogate the *trans*-activation of the SRSF3 promoter. The SRSF7 promoter is not controlled
337 by HPV16 infection [24] and showed no statistically significant change in activity in the
338 presence of E2 demonstrating that the E2-mediated up-regulation of SRSF1-3 is specific
339 (Figure 1B).

340

341 **Differentiation of HPV31-positive keratinocytes increases SRSF1-3 levels.**

342 W12 cells are cervical epithelial cells derived from a patient with an HPV16-positive low
343 grade cervical lesion [35]. Cells of the non-tumour subclone W12E (clone 20863), when

344 maintained at low passage ($p > 17$), contain around 50-100 episomal copies of the HPV16
345 genome [26,35]. W12E cells can be induced to differentiate to keratinocytes that express
346 markers of terminal differentiation [36], viral late mRNAs and late proteins [10] by culturing to
347 high density in the presence of 1.88 mM Ca^{2+} . Previously we showed that a subset of SRSF
348 proteins, including SRSF1, 2 and 3, were upregulated upon differentiation of W12E cells
349 [24]. To determine if control of SRSF expression during epithelial differentiation was
350 associated with other HR-HPV infections we examined SR protein levels in HPV31-infected
351 CIN612-9E cervical epithelial cells [28]. CIN612-9E cells were established from an HPV31-
352 infected low grade cervical lesion and contain episomal HPV31 viral genome copies.
353 CIN612-9E cells can also be induced to differentiate in culture using the same protocol as for
354 W12 cells (by culturing to high density in the presence of 1.88 mM Ca^{2+}) to express viral late
355 mRNAs and synthesise virions [37, 38]. Monoclonal antibodies against SRSF1, SRSF3 and
356 SRSF7 perform well in western blotting. However, to detect the remainder of the SR proteins
357 we used Mab104 that detects a phosphoepitope on most of the classical SR proteins
358 (SRSF1, 2, 4, 5 and 6). Therefore, in Figure 2A detection of SRSF1 is shown both with
359 Mab104 and the SRSF1-specific antibody Mab96. Semi-quantitative western blot analysis of
360 protein titrations (5, 10, 20 μ g) showed that levels of SRSF1, 2 and 3, were higher in
361 differentiated compared to undifferentiated HPV31-positive cells (Figure 2A), similar to what
362 was found in W12E cells [24]. Unlike W12E cells, CIN612-9E cells expressed SRSF4
363 (SRp75) whose levels also increased upon differentiation. Levels of SRSF5 (SRp40) and
364 SRSF6 (SRp55) increased slightly upon differentiation. Levels of SRSF7 did not change
365 significantly. Cell differentiation was revealed by detection of increased levels of the
366 epithelial differentiation marker involucrin in the differentiated keratinocytes (Figure 2A).

367

368 **The E2 transactivation domain is required for control of SRSF1-3 levels.**

369 The data support the hypothesis that HR-HPV E2, which is expressed at highest levels in
370 differentiating HPV-infected epithelial cells [5,6], controls expression of SR proteins. To test
371 this directly, we compared levels of SR proteins in normal foreskin keratinocytes (NFKs)

372 stably transfected with wild type HPV genomes or with genomes containing an inactivating
373 point mutation in the trans-activation domain of E2. HPV16 genomes containing such a
374 mutant cannot be maintained episomally in keratinocytes [39]. However, keratinocytes
375 containing HPV31 genomes with this mutation are available (mutant E2:I73L) [30]. If E2
376 *trans*-activated the SRSF1-3 promoters there would be reduced levels of the proteins in
377 these keratinocytes. Western blot quantification indeed revealed decreased levels of
378 SRSF1-3 in the keratinocytes expressing the mutant E2 protein compared to keratinocytes
379 expressing wild type E2 but there was no change in levels of SRSF7 as expected as it is not
380 regulated by E2 (Figure 2B). There were similar viral genome copy numbers in E2 wt and
381 E2:I73L NFKs (data not shown). Figure 2C shows levels of E2 protein in each NFK line.
382 Comparing with levels of GAPDH, there were similar levels of E2 protein expressed in E2wt
383 compared to E2:I73L cells (Figure 2C, lanes 3 & 4). The NFK E2wt and E2:I73L protein
384 extracts used in Figure 2B were prepared from differentiated NFK clones because both
385 expressed involucrin, a marker of epithelial differentiation (Figure 2C, lanes 3 & 4)). Both
386 E2wt and E2:I73L NFKs clearly expressed more involucrin than undifferentiated HPV31-
387 positive CIN6129E cells (Figure 2C, lanes 2) and levels were almost as high as fully
388 differentiated CIN6129E cells (Figure 2C, lanes 1). Levels of SRSF1 in the E2:I73L
389 expressing cells were 52% (± 3.2)% of wild type while SRSF2 and 3 were 31% (± 6.5)% and
390 29% (± 5.0)% of wild type respectively (Figure 2D). These data indicate that E2 *trans*-
391 activates the SRSF gene promoters to control expression of SRSF1-3 proteins in infected
392 epithelial cells. E2:I73L mutant HPV31 genomes express around 80% less late mRNA than
393 wild type genomes [30] but whether this leads to reduced L1 protein expression has not
394 been tested. In accordance with the reported change in late mRNA production, western
395 blotting revealed that there was a reduction of around 75% in L1 protein expression from
396 E2:I73L genomes compared to E2 wild type genomes ($p < 0.01$) (Figure 2E). Our data
397 suggest that HR-HPV E2 upregulates expression of SR proteins 1, 2 and 3 in infected,
398 differentiated keratinocytes
399

400 **SRSF3 controls expression of the HPV16 L1 capsid protein in differentiated**
401 **keratinocytes.**

402 E2-mediated upregulation of cellular SR proteins 1-3 in differentiated infected epithelial cells
403 could indicate that increased levels of these RNA processing proteins are required for
404 completion of the viral replication cycle. Therefore we next determined which of the
405 differentiation stage-specific controlled SR proteins controlled HPV capsid protein
406 expression during the HPV life cycle in differentiating keratinocytes. For these experiments
407 we used both the HPV16 infected W12E life cycle model we had used previously [24] and a
408 second model of the HPV16 infectious life cycle (NIKS16) to corroborate our data with W12
409 cells. NIKS16 cells are normal immortalised foreskin keratinocytes (NIKS) stably transfected
410 with episomal HPV16 genomes [27]. If used at low passage, NIKS16 clone 2L stably
411 maintains episomal genomes. Like W12E and CIN612 9E cells they can differentiate in
412 monolayer culture and display a CIN1-like phenotype in organotypic raft culture [27].
413 Differentiation of W12E, NIKS16 and the parental NIKS cells can be induced by culturing in
414 1.88 mM Ca²⁺ to high density [10,26]. At the end of the differentiation protocol between 75-
415 85% of NIKS or NIKS16 cells expressed the differentiation marker involucrin similar to that
416 observed with W12E cells [10]. Using Mab104 that simultaneously detects SRSFs1, 2, 4, 5
417 and 6 and antibody 7B4 that detects specifically SRSF3, semi-quantitative western blotting
418 of 5, 10, 20 µg HPV-negative NIKS protein extracts revealed very little change in levels of
419 SR proteins upon differentiation, similar to what we observed previously using HPV-negative
420 HaCaT cells [25] (Figure 3A). However, semi-quantitative western blotting of 5, 10, 20 µg
421 HPV16-positive NIKS16 protein extract revealed a pattern of change in SRSF expression
422 between undifferentiated and differentiated NIKS16 cells similar to that observed for W12E
423 [24] and CIN612-9E cells (Figure 2A). Figure 3B shows that levels of SRSFs1, 2 and 3 were
424 higher in differentiated NIK16 cells. Levels of SRSF4 and 6 also appeared to increase
425 slightly while SRSF5 and 7 levels (SRSF7 was detected with specific monoclonal antibody
426 clone 98) did not change. Quantification of changes in SRSF3 levels between

427 undifferentiated and differentiated HPV-negative NIKS cells revealed a reduction of 30%
428 upon differentiation (Figure 3C). In contrast, SRSF3 levels increased four-fold upon
429 differentiation of HPV16-positive NIKS16 cells similar to the increase in SRSF3 levels we
430 quantified in W12 cells previously [24] (Figure 3D). Immunohistochemistry staining of
431 normal cervical epithelium revealed staining in the basal layer cells and some in the
432 suprabasal layers (Figure 3E). Immunohistochemistry staining of a representative low grade
433 cervical lesion revealed that while more cells in the basal epithelial layers stained positive for
434 SRSF3 compared to cells in the upper layers, the intensity of staining of cells in the mid to
435 upper layers was strong indicating high levels of SRSF3 expression (Figure 3F).

436

437 Having demonstrated that NIKS16 cells control expression of SRSFs1, 2, and 3 in a similar
438 manner to W12E and CIN612-9E cells [24], these SR proteins were depleted using siRNA in
439 NIKS16 keratinocytes. SRSFs 5 and 7 were also depleted as controls for SR proteins that
440 are not regulated by E2 or HPV infection. Figure 4A shows that levels of each of the SR
441 proteins were significantly reduced by their respective siRNAs. SRSF1, 2 and 5 were
442 detected with Mab104 (top panel) that detects most of the classical SR proteins. The second
443 panel from the top shows detection of SRSF1 using specific antibody Mab96. A second
444 panel showing SRSF1 and 2 detected together is included as the third panel down from the
445 top because SRSF2 depletion is not clearly seen in the top panel. The bottom two panels
446 show SRSF3 detected by specific antibody 7B4 and SRSF7 detected by specific antibody
447 clone 98. Levels of L1 protein were assessed as a marker of late events in the virus life cycle
448 in the SRSF-depleted cells by western blotting with and HPV L1 antibody. L1 was expressed
449 in differentiated (Figure 4B lane 2) but not in undifferentiated (Figure 4B lane 1) NIKS16 cells
450 as expected from data on raft cultures of this line [27]. Differentiation was assessed by
451 increased expression of the epithelial differentiation marker involucrin (Figure 4B compare
452 lanes 1 and 2). There was no significant change in levels of L1 protein in differentiated
453 NIKS16 cells when either SRSF1, 2, 5 or 7 were depleted. However, depletion of SRSF3
454 (Figure 4B, lane 5) caused a significant decrease in L1 protein expression. Quantification of

455 three separate experiments showed that loss of SRSF3 caused a decrease of 53% \pm 11%
456 ($p < 0.05$) in L1 protein expression in differentiated NIKS16 cells (Figure 4C).

457

458 Next we confirmed the SRSF3 regulation of L1 expression in the W12E model of the HPV16
459 life cycle. SRSF1 was used as a control SR protein as this did not significantly affect the
460 levels of HPV16 L1 protein in NIKS16 cells (Figures 4B and 4C). SRSF1 and SRSF3 were
461 successfully depleted using specific siRNAs in undifferentiated (W12U) and differentiated
462 (W12D) W12E cells (Figure 5A). Involucrin expression was higher in the differentiated cell
463 population as expected (Figure 5A, lanes 4-6). No L1 protein expression was detected in
464 undifferentiated W12 cells but high levels of L1 protein were detected in differentiated W12E
465 cells (Figure 5A lane 4). Depletion of SRSF1 had a small reduction effect on L1 protein
466 expression in the differentiated cells (Figure 5A, lane 5) but, consistent with the data using
467 NIKS16 cells, SRSF3 depletion reduced L1 protein levels significantly (Figure 5A, lane 6).
468 Quantification of data from three separate experiments showed that L1 protein levels were
469 reduced by 55% \pm 11% ($p < 0.05$) in differentiated W12E cells (Figure 5B).

470

471 If SRSF3 depletion in differentiated NIKS16 cells results in loss of L1 expression then
472 overexpression of SFSF3 in undifferentiated NIKS16 cells might activated L1 expression. To
473 test this we transfected undifferentiated NIK16 cells with an expression construct for SRSF3
474 (pcDNA3.1SRSF3). Cells were maintained at low cell density in medium with low calcium
475 concentrations over the course of the experiment to minimise differentiation [26]. 72 hours
476 following transfection cells were harvested and protein lysates were prepared. Western blot
477 analysis revealed that cells transfected with vector alone did not express L1 protein (Figure
478 5C, lane 2). In contrast, cells transfected with the SRSF3 expression construct showed some
479 L1 expression (Figure 5C, lane 3) but not to the level detected in fully differentiated NIKS16
480 cells (Figure 5C, lane 1). The transfected NIKS16 cells had not been induced to differentiate
481 because levels of involucrin differentiation marker were low in the transfected cells (Figure

482 5C, lanes 2 & 3) compared to the fully differentiated NIKS16 cell control (Figure 5C, lane 1).

483 Taken together the data reveal that SRSF3 is a key regulator of HPV16 L1 protein levels.

484

485 **SRSF3 depletion results in loss of the E4^{L1} spliced RNA**

486 SRSF3 can regulate mRNA expression at a number of different levels [40] and it is known to

487 be a master regulator of alternative splicing [41]. Alternative splicing is a key control

488 mechanism of HPV gene expression. HPV L1 capsid protein is thought to be expressed from

489 an E4^{L1} spliced transcript (RNA B) but may also be expressed from an E1^{L1} (RNA A) or

490 the bicistronic L2L1 transcript (RNA C) (Figure 5A) [10,42]. Therefore, we examined if levels

491 of any of these mRNAs were affected by SRSF3 depletion during the HPV16 life cycle. We

492 mock-depleted or siRNA-depleted SRSFs 1 and 3 in W12E cells and allowed the cells to

493 differentiate. siRNA treatment successfully depleted levels of SRSF1 and 3 (Figure 5B). We

494 used primer pairs to PCR amplify across splice junctions E1^{L1} and E4^{L1} or across the

495 junction of the open reading frames in the read-through L2L1 mRNA. E1^{L1} proved very

496 difficult to detect indicating that it is an mRNA present at very low levels in HPV16-positive

497 keratinocytes. In viral late mRNAs, the end of the E4 open reading frame contains a splice

498 donor site at nt 3632 that allows splicing to the L1 splice acceptor site at nt 5639. E4^{L1}

499 mRNAs were detected readily. Compared to the effect of SRSF1 depletion that has only a

500 small effect on L1 protein production, SRSF3 depletion caused a remarkable decrease in

501 levels of late mRNA E4^{L1} (RNA B) in differentiated W12E cells (Figure 5C, lane 6). SRSF1

502 depletion did not alter production of this RNA (Figure 5C, lane 4). In contrast, depletion of

503 SRSF1 or SRSF3 appeared to increase the levels of the L2/L1 bicistronic mRNA (RNA C)

504 (Figure 5C, lower panel, lanes 4 and 6). Levels of GAPDH did not change upon SRSF1 or

505 SRSF3 depletion. qRT-PCR analysis using previously published probe/primer sets [43]

506 confirmed that E4^{L1} mRNA levels were reduced significantly upon SRSF3 depletion in

507 differentiated cells (Figure 5D) while the bicistronic L2L1 mRNA levels increased when either

508 SRSF1 (2.0-fold) or SRSF3 (1.7-fold) were depleted (Figure 5E). Taken together these data

509 suggest that SRSF3 can control production of L1 capsid protein via enhancing expression of
510 the E4^{L1} mRNA.

511

512

513

514 **DISCUSSION**

515
516 Focusing on the paradigm SR protein SRSF1 we demonstrated previously that E2 bound the
517 SRSF1 promoter and its trans-activation domain was essential for upregulation of SRSF1
518 transcription [25]. Here we have shown that E2 transcriptionally activates the genes
519 encoding not only SRSF1 but also SRSF2 and 3. The SRSF3 promoter displayed the
520 highest level of transactivation in our transcription assays. E2 regulated the SRSF3 promoter
521 in a dose-dependent manner. This is different to the SRSF1 promoter that we found was
522 activated by low levels but repressed by high levels of E2 (this manuscript, [25]). We do not
523 yet understand how E2 transactivates the SRSF promoters because similar to what we have
524 found for the SRSF1 promoter [24] the SRSF2 and 3 promoters do not contain a cognate E2
525 binding site. *In silico* analysis of the proximal promoter sequences of SRSF1-3 revealed that
526 they are each predicted to bind the E2-interacting factors Sp1 and cEBP (data not shown)
527 but it remains to be tested if E2 binds and transactivates via these proteins and if other
528 factors are also involved. It is even possible that the array of cellular transcription factors that
529 control each promoter may be different and mediate different transcriptional responses to
530 E2. SRSF7 was used in the transcription assays as a control promoter as it is not regulated
531 by E2 but we did not clone and test E2 control of other SRSFs. Therefore we cannot
532 conclusively rule out the possibility that E2 controls the expression of additional SRSF
533 proteins. However, our data clearly reveal that the cellular SRSF1, 2 and 3 genes are
534 upregulated by HPV16 E2. This is the first reported cloning of these SRSF promoters.

535
536 Semi-quantitative western blotting experiments revealed a similar profile of changes in SRSF
537 protein levels upon differentiation of HPV16-positive W12E cells [24], NIKS16 cells (Figure
538 3A) and the HPV31-positive CIN612-9E cells. This allowed us to test the mechanism of E2
539 control of the SRSF promoters by comparing levels of the SRSF proteins in normal foreskin
540 keratinocytes stably transfected with a wild type HPV31 genome or a genome containing a
541 single inactivating point mutation in the E2 transactivation domain [30]. There was a
542 significant decrease in expression of SRSFs1-3 in keratinocytes expressing the point mutant

543 E2. These data support our hypothesis that HR-HPV E2 controls the promoters of these SR
544 proteins via its trans-activation domain. SRSF3 levels showed the greatest change with a
545 71% decrease detected in cells expressing the mutant E2 compared to wild type E2. The
546 reported increased levels of E2 in the mid to upper layers of HPV-infected epithelia [5,6] fits
547 well with the increase in SRSF1-3 expression in differentiating compared to undifferentiated
548 W12E cells [24], NIKS16 and CIN612-9E cells (this study). It will be of interest to determine if
549 other high risk and low risk HPV types control SRSFs during the infectious life cycle. Our
550 data imply that E2 control of SRSF1-3 is likely at the level of transcription initiation. However,
551 because E2 has been shown to possess some properties of SR proteins [44-46] we cannot
552 discount a direct E2-mediated regulation of HPV mRNA processing or indeed E2 control of
553 alternatively spliced isoforms of SR protein mRNAs that are more stimulatory of L1 splicing.
554 Also, although E2 is a regulator of SRSFs we cannot rule out the possibility that other HPV
555 proteins could be involved in controlling SRSF expression. For example, E6 and E7 may
556 activate cellular transcription factors and control SRSF levels indirectly in differentiated
557 keratinocytes [47,48].

558

559 Our data suggest that a specific subset of SR proteins (SRSFs1-3) is regulated by E2,
560 indicating that these proteins may have important functions during the HPV replicative life
561 cycle. Reasoning that increased SR protein expression in differentiated keratinocytes might
562 impact late events in the viral life cycle we compared the effect of siRNA depletion of each of
563 the E2-regulated SRSFs1-3 with SRSF5 and 7 (the latter two of which are not controlled by
564 E2 in the NIKS16 model) on capsid protein expression. Depletion of SR proteins 2, 5 and 7
565 had no significant effect on L1 protein expression. Depletion of SRSF1 caused a modest
566 reduction in L1 protein expression in NIKS16 and W12 cells as predicted by a previous study
567 [20]. However, SRSF3 depletion caused a major reduction in L1 protein expression in the
568 differentiated NIKS16 cells. SRSF3 control of L1 expression was confirmed in a second
569 HPV16 life cycle model using W12E cells. Conversely, overexpression of SRSF3 in
570 undifferentiated NIKS16 cells resulted in some induction of L1 protein expression. We were

571 unable to repeat the experiment in the W12 cell model due to problems inhibiting
572 differentiation in these cells during the course of the overexpression experiment. Although
573 further work is required to determine whether the effect of SRSF3 on L1 induction in
574 undifferentiated epithelial cells is at the level of mRNA processing, our data indicate that
575 SRSF3 is a key regulator of viral capsid protein expression.

576

577 Although we did not carry out SR protein depletion experiments in CIN612-9E cells, we
578 observed a significant decrease in capsid protein production in human keratinocytes
579 maintaining E2:I73L mutant genomes, with a concomitant 50% reduction in SRSF3
580 expression, compared to cells containing wild type HPV31 genomes. This does not prove
581 that SRSF proteins control L1 protein production in HPV31-positive keratinocytes but it is
582 correlative with the observations in the HPV16-positive lines. The approximately 4-fold level
583 of L1 protein depletion we observed (Figure 2D) agrees with the 80% reduction observed
584 previously in L1 mRNA production in foreskin keratinocytes expressing the HPV31 E2:I73L
585 mutant genome [30]. The greater effect of the E2 inactivating point mutation on L1 protein
586 expression compared to SRSF3 depletion could be due to incomplete knock down of SRSF3
587 in our experiments. However, it may be due to other effects of E2 on HPV or cellular gene
588 expression. E2 appears to be required for induction of HPV late mRNA expression by
589 inhibiting the viral early polyadenylation signal through interaction with CPSF30 [11]. While in
590 our study an HPV31 E2 transactivation domain point mutant was all that was required to
591 reduce virus capsid protein expression, in the study by Johannson et al. both the E2
592 transactivation domain and the hinge region were required for control of L1 and L2 mRNA
593 levels [11]. This suggests that there at least two E2-regulated mechanisms for controlling the
594 viral late gene expression and these are not mutually exclusive. Apart from these important
595 mechanisms, E2 can bind other RNA processing factors, including SR proteins SRSFs 1, 2,
596 and 7 [49] and could modulate their splicing activities as well as or instead of controlling
597 transcription of *SRSF* genes. Moreover, E2 possesses SR protein-like properties itself as it
598 can bind RNA and affect splicing *in vitro* [44, 45]. Importantly, a recent study revealed that in

599 U2OS cells E2 can control human gene expression by inducing alternative splice isoforms of
600 mRNAs that encode proteins involved in cancer formation and cell motility [46]. Therefore, it
601 is possible that E2 could control late mRNA expression directly. A complex combination of
602 different E2-regulated mechanisms may be essential in differentiation stage-specific control
603 of HPV gene expression.

604

605 Several previous studies have reported the effects of SR proteins on HPV gene expression.
606 These reports have delineated ESEs and ESSs and their interacting proteins on viral RNAs.
607 SRSF1 binds a downstream ESE in the E4 open reading frame to control the viral splice
608 acceptor site 3258 [20]. SRSF2 controls HPV16 oncoprotein expression by regulating mRNA
609 stability [23]. SRSF3 also controls HPV16 oncoprotein expression [19] but to a lesser extent
610 than SRSF2 [23]. SRSF9 can activate the E4 splice donor 3632 while inhibiting splice
611 acceptor 3358 at the 5' end of the E4 open reading frame [21]. Of interest, another splicing
612 control element in the E4 coding region [50] binds SRSF3 [19]. Knocking down SRSF3 in
613 U2OS osteosarcoma cells increased levels of both L1 and L2 mRNAs while mutating the
614 SRSF3 binding site increased L1 mRNA production almost two-fold [19]. In contrast to this
615 study, we found that SRSF3 depletion in undifferentiated epithelial cells did not induce L1
616 protein expression (Figure 3E). Moreover, our data indicate that SRSF3 is required for the
617 L1-encoding E4^{L1} mRNA and L1 protein expression in differentiating keratinocytes.
618 Although some of the effects we have observed are rather modest, the differences in our
619 data compared to the previous studies can be attributed to our use of differentiating, HPV-
620 infected non-tumour epithelial cells (three different model systems) that better mimic the
621 natural HPV life cycle. Further, wild type HPV genomes are present in NIKS16 and W12E
622 and CIN612-9E cells. In contrast, in the previous studies, subgenomic constructs with
623 deletions in important *cis*-acting elements (e.g. the late negative regulatory region [51])
624 controlling gene expression were used [20,21,50]. In our HPV-infected lines, viral
625 transcription is driven from the natural viral promoters [10,38] which are relatively weakly
626 active. Expression of HPV late mRNAs from the subgenomic constructs is under control of

627 the strong CMV promoter. Of importance, RNA polymerase II elongation rate, determined by
628 promoter strength, is a major determinant of the use of weak over strong splice sites
629 regardless of ESE/ESS activity [52,53]. Finally, the exact levels of SR proteins in cells are
630 known to be vital to their function in splicing [54] meaning the effects on HPV late gene
631 expression of the SRSF levels found in undifferentiated tumour cells may be different to their
632 effects in differentiated keratinocytes.

633

634 However, it is important to consider that because we have analysed HPV gene expression
635 together with keratinocyte differentiation upon SRSF knockdown, some of the effects we
636 observed could be an indirect result of SRSF-mediate control of other cellular factors or of
637 epithelial differentiation itself. For example, in Figure 6A we observed an increase in
638 expression of the epithelial differentiation marker involucrin upon SRSF3 depletion implying
639 that SRSF3 may repress involucrin expression upon epithelial differentiation. Further work
640 will be required to determine whether SRSF3 is controlling HPV late gene expression directly
641 or whether it does so through other cellular pathways.

642

643 Immunohistochemical staining of HPV16-infected low grade cervical lesions has revealed a
644 distinct pattern of SRSF3 expression [24](this study). SRSF3 is expressed in the majority of
645 basal epithelial cells. Although fewer cells in the upper epithelial layers express SRSF3
646 those that do express the protein contain very densely stained nuclei (Figure 3F). We
647 propose that these cells are HPV-infected and supporting late events in the viral replication
648 cycle including E2-mediated control of SRSF3 expression. In preparation for capsid protein
649 production, viral late mRNAs may be produced in epithelial cells below the granular layer [8]
650 colocalizing with high levels of E2 and SRSF3. SRSF3 appears to be required for expression
651 of the E4^{L1} late mRNA that encodes the L1 capsid protein. Conversely, SRSF1 and
652 SRSF3 negatively regulate production of the L2L1 read through RNA that likely encodes L2.
653 This could be one regulatory mechanism that ensures high levels of L1 and low levels of L2
654 protein production to provide the correct ratio of L1 to L2 protein for capsid formation.

655 SRSF3 can cross-regulate SRSF2, SRSF5 and SRSF7 and has been proposed as a master
656 regulator of RNA splicing [41]. The cooperative activity of SR proteins imposes positive and
657 negative combinatorial control on splicing [54]. Therefore, in future it will be important to
658 examine the effect of siRNA depletion of combinations of E2-controlled SR proteins on the
659 HPV gene expression programme.

660

661 In conclusion, if HPV-regulated SR proteins are key to completion of the viral replication
662 cycle they may represent a useful target for antiviral therapy. Small molecule inhibitors of SR
663 protein kinases that modulate the functions of SR proteins are available. As they have been
664 shown to be effective in inhibiting replication of HIV, hepatitis C virus and Sindbis virus [14]
665 they have potential to be effective against HPV replication. However, it will also be important
666 also to understand any potential effects of these therapies on the switch from the viral
667 replicative cycle to persistent or latent infection that underlies tumour progression.
668 Nevertheless, such targets may prove effective for future therapeutic study.

669

670 **FUNDING INFORMATION**

671 This work was funded by the Wellcome Trust, grant number Wtd004098. The funders had
672 no role in study design, data collection and interpretation, or the decision to submit the work
673 for publication.

674

675 **ACKNOWLEDGEMENTS**

676 We would like to thank Margaret Stanley (University of Cambridge), Paul Lambert and
677 Denis Lee (University of Wisconsin) for provision of the W12 cells. We are grateful to Craig
678 Meyers (Penn State University) for the gift of CIN612-9E cells and to John Doorbar
679 (University of Cambridge) for the NIKS16 cells. Iain Morgan (Virginia Commonwealth
680 University) kindly provided the U2OS cells lines transfected with HPV16 E2. We are
681 grateful to Lou Laimins (Northwestern University, Chicago) for agreeing to provide the

682 protein extracts of the wild type and E2:I73L mutant HPV31-transfected keratinocytes. The

683 authors declare no conflicts of interest.

684

685 **FIGURE LEGENDS**

686

687 **Figure 1 HPV16 E2 trans-activates the promoters of SRSFs 1-3.** (A) Western blot
688 analysis of levels of E2 protein in two clones of U2OS cells (U2OSA4, U2OSB1) stably
689 transfected with an HPV16 E2 expression vector [29]. U2OSv, cells transfected with empty
690 vector. GAPDH is shown as a control for protein loading levels. (B) Luciferase transcription
691 assays reveal E2-transactivation of the SRSF1-3 promoters. Graph of luciferase activity in
692 U2OSv (dark grey bars), U2OSA4 (light grey bars) and U2OSB1 (mid grey bars) cells
693 transiently transfected with luciferase expression vectors under control of the promoters
694 shown on the X-axis or the promoterless vector, pGL3, as a negative control. The mean and
695 standard deviation from the mean of three separate experiments is shown. Asterisks show
696 statistically significant change (p value) using a student's T-test (two-tailed). * p<0.05. ***
697 p<0.001.

698 **Figure 2. The transactivation domain of HPV31 E2 is required for control of SRSF1-3**
699 **and production of L1 protein.** (A) SRSF levels are increased in differentiated HPV31-
700 infected CIN612-9E cells. Semi-quantitative western blot analysis of levels of SR proteins in
701 undifferentiated (U) and differentiated (D) CIN612-9E cells. 5, 10 or 20µg protein extracts
702 were loaded as indicated above the blots. SRSF1, 3, and 7 were detected with specific
703 monoclonal antibodies as indicated on the right hand side of the blots. SRSF1, 2, 4, 5 and 6
704 were detected with Mab104 that detects phosphoepitopes of all the classical SR proteins
705 except SRSF9. Involucrin was detected as a control for differentiated epithelial cells.
706 Involucrin is detected in the undifferentiated cells due to around 20% of these having
707 undergone differentiation. GAPDH was detected as a protein loading control. A single gel
708 was blotted and probed with antibodies against SRSF1 and involucrin. The same samples
709 were electrophoresed on identical gels for probing with GAPDH and SRSF7 or with Mab104
710 and SRSF3. This experiment was repeated three times with very similar results from each

711 western blot. (B) Keratinocytes containing a transactivation-negative E2 point mutant HPV31
712 genome have reduced levels of SRSF1-3. Western blot analysis of SRSF1-3 levels in
713 normal human foreskin keratinocytes (NFKs) stably transfected with wild type (E2wt) or point
714 mutant E2:I73L HPV31 genomes (E2:IL73). SRSF7 was detected as a control for an SR
715 protein whose levels are not changed upon epithelial differentiation or significantly trans-
716 activated by HPV16E2. Levels of L1 protein in the two lines are also shown. The experiment
717 was carried out three times using two different clones of each E2-expressing keratinocyte
718 line. Very similar results were obtained in each experiment. (C) Western blot analysis of
719 levels of E2 protein and differentiation status of NFKs stably transfected with wild type
720 (E2wt) or point mutant E2:I73L HPV31 genomes (E2:IL73). As a marker of differentiation,
721 levels of involucrin are shown in undifferentiated (U) and differentiated (D) CIN6129E
722 keratinocytes for comparison. The experiment was carried out twice using two different
723 clones of each E2-expressing keratinocyte line. Very similar results were obtained in each
724 experiment. (D) Quantification of levels of the SR proteins shown in (B). The graph shows
725 the mean and standard deviation from the mean of three separate experiments. Values were
726 calculated relative to GAPDH levels. Very similar data were obtained with two different
727 clones of each keratinocyte line. (E) Western blot quantification of levels of L1 protein in
728 differentiated NFKs stably maintaining wild type HPV31 and in NFKs stably maintaining point
729 mutant E2:I73L HPV31 genomes. The graph shows the mean and standard deviation from
730 the mean from three separate experiments. Values were calculated relative to GAPDH
731 levels. P-values were calculated using a student's T-test.

732 **Figure 3 SRSF3 levels increase upon differentiation of HPV16-infected cells** (A)
733 SRSF1-3 levels do not alter upon differentiation of HPV-negative NIKS cells. Semi-
734 quantitative western blot analysis of levels of SR proteins in undifferentiated (U) and
735 differentiated (D) NIKS cells. 5, 10 or 20µg protein extracts were loaded as indicated above
736 the blots. SRSF3 was detected with a specific monoclonal antibody 7B4. SRSF1, 2, 4, 5 and
737 6 were detected with Mab104. SRSF7 was not tested in this experiment because its levels

738 did not change upon HPV infection [24]. Involucrin was detected as a control for
739 differentiation of epithelial cells. GAPDH was detected as a protein loading control.
740 (B)SRSF1-3 levels are increased in differentiated NIKS16 cells. Semi-quantitative western
741 blot analysis of levels of SR proteins in undifferentiated (U) and differentiated (D) NIKS16
742 cells. 5, 10 or 20µg protein extracts were loaded as indicated above the blots. SRSF3 and 7
743 were detected with specific monoclonal antibodies. SRSF1, 2, 4, 5 and 6 were detected with
744 Mab104. Involucrin was detected as a control for differentiated epithelial cells. GAPDH was
745 detected as a protein loading control. The experiments in (A) and (B) were carried out at
746 least three times with very similar results obtained each time. (C) Quantification of SRSF3
747 levels in undifferentiated (U) and differentiated (D) HPV-negative NIKS cells. The graph
748 shows the mean and standard deviation from the mean from three separate experiments.
749 Values were calculated relative to GAPDH levels. (D) Quantification of SRSF3 levels in
750 undifferentiated (U) and differentiated (D) NIKS16 cells. The graph shows the mean and
751 standard deviation from the mean from five separate experiments. Values were calculated
752 relative to GAPDH levels. (E) Immunohistochemical staining of a representative normal
753 cervical epithelium (10 normal lesions were stained) with an antibody against SRSF3. Note
754 the strongly stained nuclei are present mainly in the lower epithelial layers. (F)
755 Immunohistochemical staining of a representative low grade cervical lesion (10 low grade
756 lesions were stained) with an antibody against SRSF3. Note the strongly stained nuclei in
757 the mid to upper epithelial layers. The picture in (E) is taken at lower magnification than that
758 in (F) to show that the staining pattern is consistent over a wide area of tissue.

759 **Figure 4. SRSF3 is required for L1 protein expression.** (A) Western blot analysis of SR
760 protein depletion for the experiment shown in (E). SRSF1, 3 and 7 were detected with
761 specific antibodies. SRSF1, 2 and 5 were detected with Mab104. The panel showing SRSF2
762 depletion was probed with Mab104. Only the area of the blot showing SRSF2 and SRSF1
763 bands is shown. SRSF2 is the upper band. U, undifferentiated, D, differentiated NIKS16
764 cells. Cntrl, cells transfected with a control siRNA. (B) SRSF3 depletion results in reduced L1

765 protein levels. Western blot analysis of L1 protein in, differentiated keratinocytes (D) with the
766 various SR proteins depleted. (U), undifferentiated NIKS16 cell extracts were analysed as a
767 negative control for L1 protein expression. Cntrl, cells transfected with a control siRNA.
768 Involucrin was detected as a control for differentiated epithelial cells. GAPDH was detected
769 as a protein loading control. (C) Quantification of levels of L1 protein in NIK16 keratinocytes
770 transfected with a control siRNA (cntrl) or siRNAs against each of the SR proteins shown on
771 the X-axis. U, undifferentiated keratinocytes D, differentiated keratinocytes. The graph
772 shows the mean and standard deviation from the mean from three separate experiments.
773 Values were calculated relative to GAPDH levels. P-value was calculated using a student's
774 T-test.

775

776 **Figure 5 SRSF3 controls L1 protein levels in W12E cells.** (A) Western blot analysis of
777 SRSF1 and 3 protein depletion by siRNA in U, undifferentiated and D, differentiated W12E
778 cells. Much higher levels of SRSF1 and SRSF3 were detected in differentiated W12E cells
779 as expected [24]. L1 protein was not detected in U, undifferentiated W12E cells even with
780 SRSF1 or SRSF3 depletion. L1 protein was detected in D, differentiating W12E cells but
781 levels were reduced in the cells treated with siRNA against SRSF3. Cntrl, cells transfected
782 with a control siRNA. Involucrin was detected as a control for differentiated epithelial cells.
783 GAPDH was detected as a protein loading control This experiment was carried out five times
784 and very similar data were obtained each time. (B) Western blot quantification of levels of L1
785 protein in differentiated W12E keratinocytes transfected with a control siRNA (cntrl) or
786 siRNAs against SRSF1 or SRSF3. The graph shows the mean and standard deviation from
787 the mean from three separate experiments. Values were calculated relative to GAPDH
788 levels. P-value was calculated using a student's T-test. (C) Undifferentiated NIKS16 cells
789 were transfected with vector control (Cntrl: lane 2) or a plasmid encoding the SRSF3 protein
790 (SRSF3 transfect: lane 3). Differentiated NIK16 cell extract (DNIKS16, lane 1) was probed
791 as a positive control for L1 expression. GAPDH was used as a loading control. Involucrin

792 was detected as a marker of epithelial differentiation. The level of involucrin was so high in
793 the differentiated compared to the undifferentiated NIK16 cell extracts that the blot was
794 bleached out in lane 1. The experiment was repeated twice and very similar data were
795 obtained each time.

796

797 **Figure 6 SRSF3 controls levels of the E4^{L1} mRNA.** (A) Schematic diagram of the
798 HPV16 genome. Open reading frames are represented with light grey boxes with the gene
799 names beneath. Promoters are indicated with arrowheads. Polyadenylation sites are
800 indicated with downward arrows (Poly(A)_{E/L}). The major late mRNA splicing events with the
801 open reading frames (dark grey boxes) and introns (dark grey lines) are indicated below the
802 genome map. Each mRNA is shown truncated at the 5' end to indicated the presence of
803 several possible initiation sites for transcription of these mRNAs [10]. The open reading
804 frames in each mRNA are indicated to the right hand side. (B) Western blot analysis of levels
805 of SRSF1 and SRSF3 in NIKS16 cells following siRNA knock down. siSRSF1, siRNA
806 against SRSF1. siSRSF3, siRNA against SRSF3. siCntrl, control siRNA. GAPDH was
807 detected as a protein loading control. (C) RT-PCR detection of HPV late mRNAs type B
808 (upper gel) or type C (lower gel) in cDNA synthesized from RNA isolated from the same cell
809 populations analysed in (B). Cntrl, control siRNA. SRSF1, siRNA against SRSF1. SRSF3,
810 siRNA against SRSF3. RT, reverse transcriptase. M, 1Kb DNA marker ladder. GAPDH was
811 amplified as an internal control for RNA concentrations. (D) Relative levels of E4^{L1} spliced
812 mRNA B compared to GAPDH mRNA determined by qPCR. The mean and standard
813 deviation from the mean are shown from three separate experiments. (E) Relative levels of
814 unspliced L2L1 mRNAs C compared to GAPDH mRNA determined by qPCR. The mean and
815 standard deviation from the mean are shown from three separate experiments. P-values
816 were calculated using a student's T-test.

817

818 **REFERENCES**

819

820 1. **zur Hausen H.** 2009. Papillomaviruses in the causation of human cancers - a brief
821 historical account. *Virology* **384**: 260-265.

822 2. **Graham SV.** 2010. Human papillomavirus: gene expression, regulation and
823 prospects for novel diagnostic methods and antiviral therapies. *Future Microbiol* **5**:
824 1493-1505.

825 3. **Groves IJ, Coleman N.** 2015. Pathogenesis of human papillomavirus-associated
826 mucosal disease. *J Pathol* **235**: 527-538.

827 4. **Doorbar J.** 2005. The papillomavirus life cycle. *J Clin Virol* **32S**: S7-S15.

828 5. **Xue Y, Bellanger S, Zhang W, Lim D, Low J, Lunny D, Thierry F.** 2010. HPV16 E2
829 is an immediate early marker of viral infection, preceding E7 expression in precursor
830 structures of cervical carcinoma. *Cancer Res* **70**: 5316-5325.

831 6. **Coupe VM, Gonzalez-Barreiro L, Gutierrez-Berzal J, Melian-Boveda AL, Lopez-**
832 **Rodriguez O, Alba-Dominguez J, Alba-Losada J.** 2012. Transcriptional analysis of
833 human papillomavirus type 16 in histological sections of cervical dysplasia by in situ
834 hybridisation. *J Clin Pathol* **65**: 164-170.

835 7. **McBride AA.** 2013. The papillomavirus E2 proteins. *Virology* **445**: 57-79.

836 8. **Stoler MH, Rhodes CR, Whitbeck A, Wolinsky SM, Chow LT, Broker TR.** 1992.
837 Human papillomavirus type 16 and 18 gene expression in cervical neoplasia. *Human*
838 *Pathol* **23**: 117-128.

839 9. **Zhou J, Lui.W.J., Peng SW, Sun XY, Frazer IH.** 1999. Papillomavirus capsid protein
840 expression levels depends on the match between codon usage and tRNA availability.
841 *J Virol* **73**: 4972-4982.

- 842 10. **Milligan SG, Veerapraditsin T, Ahamat B, Mole S, Graham SV.** 2007. Analysis of
843 novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical
844 epithelial cells. *Virology* **360**: 172-181.
- 845 11. **Johannson C, Somberg M, Li X, Winquist EB, Fay J, Ryan F, Pim D, Banks L.,**
846 **Schwartz S.** 2012. HPV-16 E2 contributes to induction of HPV-16 late gene
847 expression by inhibiting early polyadenylation. *EMBO J* **31**: 3212-3227.
- 848 12. **Johannson C, Schwartz S.** 2013. Regulation of human papillomavirus gene
849 expression by splicing and polyadenylation. *Nat Rev Microbiol* **11**: 239-251.
- 850 13. **Singh RK, Cooper TA.** 2012. Pre-mRNA splicing in disease and therapeutics.
851 *Trends Mol Med* **18**: 472-482.
- 852 14. **Hernandez-Lopez HR, Graham SV.** 2012. Alternative splicing in tumour viruses: a
853 therapeutic target? *Biochem J* **445**: 145-156.
- 854 15. **Busch A, Hertel KJ.** 2012. Evolution of SR proteins and hnRNP splicing regulatory
855 factors. *Wiley Interdiscip Rev RNA* **3**: 1-12.
- 856 16. **Roca X, Krainer AR, Eperon IC.** 2013. Pick one, but be quick:5' splice sites and the
857 problem of too many choices. *Genes Dev* **27**: 129-144.
- 858 17. **Long JC, Caceres JF.** 2009 The SR protein family of splicing factors: master
859 regulators of gene expression. *Biochem J* **417**: 15-27.
- 860 18. **Kajitani N, Schwartz S.** 2015. RNA binding proteins that control human
861 papillomavirus gene expression. *Biomolecules* **5**: 758-774.
- 862 19. **Jia R, Liu X, Tao M, Kruhlak M, Guo M, Meyers C, Baker CC, Zheng ZM.** 2009.
863 Control of the papillomavirus early-to-late switch by differentially expressed SRp20. *J*
864 *Virology* **83**: 167-180.

- 865 20. **Somberg M, Schwartz S.** 2010. Multiple ASF/SF2 sites in the human papillomavirus
866 type 16 (HPV-16) E4-coding region promote splicing to the most commonly used 3'-
867 splice Site on the HPV-16 genome. *J. Virol.* **84**: 8219-8230.
- 868 21. **Somberg M, Li X, Johansson C, Orru B, Chang R, Rush M, Fay J, Ryan F,**
869 **Schwartz S.** 2011. Serine/arginine-rich protein 30c activates human papillomavirus
870 type 16 L1 mRNA expression via a bimodal mechanism. *J Gen Virol* **92**: 2411-2421.
- 871 22. **Li X, Johansson C, Cardoso Palacios C, Mossberg A, Dhanjal S, Bergvall M,**
872 **Schwartz S.** 2013. Eight nucleotide substitutions inhibit splicing to HPV-16 3'-splice
873 site SA3358 and reduce the efficiency by which HPV-16 increases the life span of
874 primary human keratinocytes. *PLoS One* **8**: e72776.
- 875 23. **McFarlane M, MacDonald AI, Stevenson A, Graham SV.** 2015. Human
876 papillomavirus 16 oncoprotein expression is controlled by the cellular splicing factor
877 SRSF2 (SC35). *J Virol* **89**: 5276-5287.
- 878 24. **Mole S, McFarlane M, Chuen-Im T, Milligan SG, Millan D, Graham SV.** 2009. RNA
879 splicing factors regulated by HPV16 during cervical tumour progression. *J Pathol* **219**:
880 383-391.
- 881 25. **Mole S, Milligan SG, Graham SV.** 2009. Human papillomavirus type 16 E2 protein
882 transcriptionally activates the promoter of a key cellular splicing factor, SF2/ASF. *J*
883 *Virol* **83**: 357-367.
- 884 26. **Jeon S, Allen-Hoffman BL, Lambert PF.** 1995. Integration of human papillomavirus
885 type 16 into the human genome correlates with a selective growth advantage of cells.
886 *J Virol* **69**: 2989-2997.
- 887 27. **Weschler EI, Wang Q, Roberts I, Pagliarulo E, Jackson D, Untersperger C,**
888 **Coleman N, Griffin H, Doorbar J.** 2012. Reconstruction of human papillomavirus

- 889 type 16-mediated early-stage neoplasia implicated E6/E7 deeregulation and the loss
890 of contact inhibition in neoplastic progression. J Virol **86**: 6358-6364.
- 891 28. **Hummel M, Hudson JB, Laimins LA.** 1992. Differentiation-induced and constitutive
892 transcription of human papillomavirus type 31b in cell lines containing viral episomes.
893 J Virol **66**: 6070-6080.
- 894 29. **Taylor ER, Boner W, Dornan ES, Corr EM, Morgan IM.** 2003. UVB irradiation
895 reduces the half life and transactivation poteintial of the human papillomavirus 16 E2
896 proteins. Oncogene **22**: 4469-4477.
- 897 30. **Stubenrauch F, Colbert AME, Laimins LA.** 1998. Transactivation by the E2 protein
898 of oncogenic human papillomavirus type 31 is not essential for early and late viral
899 functions. J Virol **72**: 8115-8123.
- 900 31. **McPhillips MG, Veerapraditsin T, Cumming SA, Karali D, Milligan SG, Boner W,**
901 **Morgan IM, Graham SV.** 2004. SF2/ASF binds the human papillomavirus type 16
902 late RNA control element and is regulated during epithelial differentiation. J Virol **78**:
903 10598-10605.
- 904 32. **Fay J, Kelehan P, Lambkin H, Schwartz S.** 2009. Increased expression of cellular
905 RNA-binding proteins in HPV-induced neoplasia and cervical cancer. J Med Virol **81**:
906 897-907.
- 907 33. **Lee D, Kim H-Z, Jeong KW, Shim YS, Horikawa I, Barrett JC, Choe J.** 2002
908 Human papillomavirus E2 down-regulates the human telomerase reverse
909 transcriptase promoter. J Biol Chem **277**: 27748-27756.
- 910 34. **Frolkis M, Fischer MB, Wang Z, Lebkowski JS, Chiu C-P, Majumdar AS.** 2003.
911 Dendritic cells reconstituted with human telomerase gene induce potent cytotoxic T-
912 cell response against different types of tumors. Cancer Gene Ther **10**: 239-249.

- 913 35. **Stanley MA, Browne HM, Appelby M, Minson AC.** 1989. Properties of a non-
914 tumourigenic human cervical keratinocyte cell line. *Int J Cancer* **43**: 672-676.
- 915 36. **Chuen-lm T, Zhang J, Milligan SG, McPhillips MG, Graham SV.** 2008. The
916 alternative splicing factor hnRNP A1 is up-regulated during virus-infected epithelial
917 cell differentiation and binds the human papillomavirus type 16 late regulatory
918 element. *Virus Res* **131**: 189-198.
- 919 37. **Hummel M, Lim HB, Laimins LA.** 1995 Human papillomavirus type 31b late gene
920 expression is regulated through protein kinase C-mediated changes in RNA
921 processing. *J Virol* **69**: 3381-3388.
- 922 38. **Ozbun MA, Meyers C.** 1997. Characterisation of late gene transcripts expressed
923 during vegetative replication of human papillomavirus type 31b. *J Virol* **71**: 5161-
924 5172.
- 925 39. **Gauson EJ, Donaldson MM, Dornan ES, Wang X, Bristol M, Bodily JM, Morgan**
926 **IM.** 2015. Evidence supporting a role for TopBP1 and Brd4 in the initiation but not
927 continuation of human papillomavirus 16 E1/E2-mediated DNA replication. *J Virol* **89**:
928 4980-4990.
- 929 40. **Corbo C, Orru S, Salvatore F.** 2013. SRp20: An overview of its role in human
930 disease. *Biochem Biophys Res Comm* **436**: 1-5.
- 931 41. **Anko M-L, Muller-McNicoll M, Brandi H, Curk T, Gorup C, Henry I, Ule J,**
932 **Neugebauer KM.** 2012. The RNA-binding landscapes of two SR proteins reveal
933 unique functions and binding to diverse RNA classes. *Genome Biol* **13**: R17.
- 934 42. **Wang X, Meyers C, Wang H, Chow LT, Zheng ZM.** 2011. Construction of a full
935 transcription map of human papillomavirus type 18 during productive viral infection. *J*
936 *Virol* **85**: 8080-8092.

- 937 43. **Schmitt M, Dalstein V, Waterboer T, Clavel.C., Gissman L, Pawlita M.** 2010.
938 Diagnosing cervical cancer and high grade precursors by HPV16 transcription
939 patterns. *Cancer Res* **70**: 249-256.
- 940 44. **Lai M-C, Teh BH, Tarn W-Y.** 1999 A human papillomavirus E2 transcriptional
941 activator. *J Biol Chem* **274**: 11832-11841.
- 942 45. **Bodaghi S, Jia R, Zheng Z.** 2009. Human papillomvirus type 16 E2 and E6 are
943 RNA-binding proteins and inhibit in vitro splicing of pre-mRNA with suboptimal splice
944 sites. *Virology* **386**: 32-43.
- 945 46. **Gauson EJ, Windle B, Donaldson MM, Caffarel MM, Dornan ES, Coleman N,**
946 **Henderson SC, Wang X, Morgan IM.** 2014. Regulation of human genome
947 expression and RNA splicing by human papillomavirus E2 protein. *Virology* 468-470:
948 10-18.
- 949 47. **Roman A, Munger K.** 2013. The papillomavirus E7 proteins. *Virology* **445**: 138-168.
- 950 48. **Vande Pol SB, Klingelhutz AJ.** 2013. Papillomavirus E6 proteins. *Virology* **445**: 115-
951 137.
- 952 49. **Muller M, Demert C.** 2012. The HPV E2-host protein-protein interactions: a complex
953 hijacking of the cellular network. *Open Virol J* **6**: 173-189.
- 954 50. **Rush M, Zhao X, Schwartz S.** 2005)=. A splicing enhancer in the E4 coding region
955 of human papillomavirus type 16 is required for early mRNA splicing and
956 polyadenylation as well as inhibition of premature late gene expression. *J Virol* **79**:
957 12002-12015.
- 958 51. **Graham SV.** 2008. Papillomavirus 3'UTR regulatory elements. *Frontiers Biosci.* **13**:
959 5646-5663.

960 52. **Proudfoot N.** 2000. Connecting transcription to messenger RNA processing. Trends
961 Biochem Sci **25**: 290-293.

962 53. **Moreno NN, Giono LE, Cambindo Botto AE, Munoz MJ, Kornblihtt AR.** 2015.
963 Chromatin, DNA structure and alternative splicing. FEBS Lett In press.

964 54. **Bradley T, Cook ME, Blanchette M.** 2015. SR proteins control a complex network of
965 RNA-processing events. RNA **21**: 75-92.

966

967

968

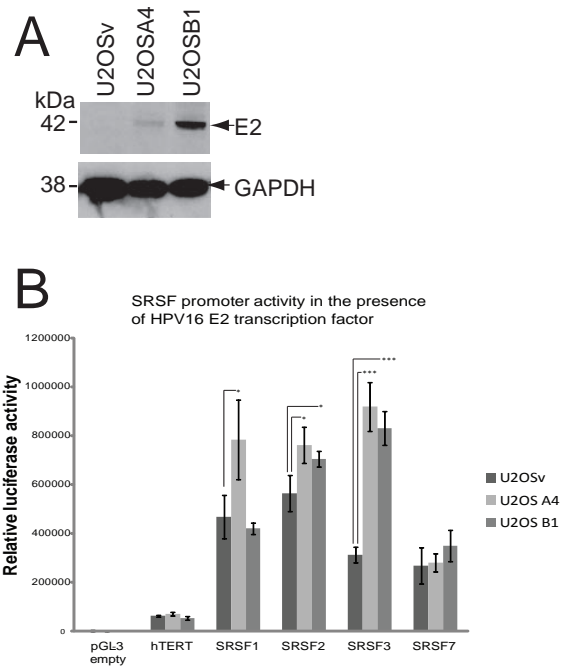


Figure 1 HPV16 E2 trans-activates the promoters of SRSFs 1-3. (A) Western blot analysis of levels of E2 protein in two clones of U2OS cells (U2OSA4, U2OSB1) stably transfected with an HPV16 E2 expression vector [29]. U2OSv, cells transfected with empty vector. GAPDH is shown as a control for protein loading levels. (B) Luciferase transcription assays reveal E2-transactivation of the SRSF1-3 promoters. Graph of luciferase activity in U2OSv (dark grey bars), U2OSA4 (light grey bars) and U2OSB1 (mid grey bars) cells transiently transfected with luciferase expression vectors under control of the promoters shown on the X-axis or the promoterless vector, pGL3, as a negative control. The mean and standard deviation from the mean of three separate experiments is shown. Asterisks show statistically significant change (p value) using a student's T-test (two-tailed). * $p < 0.05$. *** $p < 0.001$.

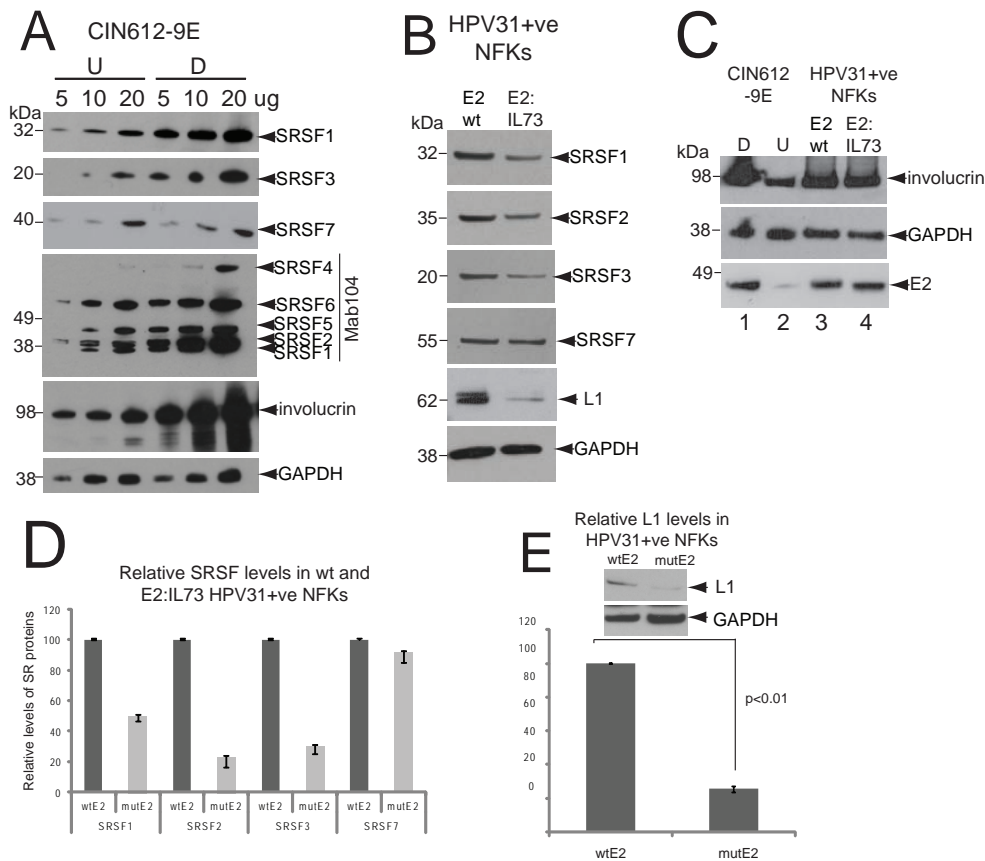


Figure 2. The transactivation domain of HPV31 E2 is required for control of SRSF1-3 and production of L1 protein. (A) SRSF levels are increased in differentiated HPV31-infected CIN612-9E cells. Semi-quantitative western blot analysis of levels of SR proteins in undifferentiated (U) and differentiated (D) CIN612-9E cells. 5, 10 or 20µg protein extracts were loaded as indicated above the blots. SRSF1, 3, and 7 were detected with specific monoclonal antibodies as indicated on the right hand side of the blots. SRSF1, 2, 4, 5 and 6 were detected with Mab104 that detects phosphoepitopes of all the classical SR proteins except SRSF9. Involucrin was detected as a control for differentiated epithelial cells. Involucrin is detected in the undifferentiated cells due to around 20% of these having undergone differentiation. GAPDH was detected as a protein loading control. A single gel was blotted and probed with antibodies against SRSF1 and involucrin. The same samples were electrophoresed on identical gels for probing with GAPDH and SRSF7 or with Mab104 and SRSF3. This experiment was repeated three times with very similar results from each western blot. (B) Keratinocytes containing a transactivation-negative E2 point mutant HPV31 genome have reduced levels of SRSF1-3. Western blot analysis of SRSF1-3 levels in normal human foreskin keratinocytes (NFKs) stably transfected with wild type (E2wt) or point mutant E2:I73L HPV31 genomes (E2IL73). SRSF7 was detected as a control for an SR protein whose levels are not changed upon epithelial differentiation or significantly trans-activated by HPV16E2. Levels of L1 protein in the two lines are also shown. The experiment was carried out three times using two different clones of each E2-expressing keratinocyte line. Very similar results were obtained in each experiment. (C) Western blot analysis of levels of E2 protein and differentiation status of NFKs stably transfected with wild type (E2wt) or point mutant E2:I73L HPV31 genomes (E2IL73). As a marker of differentiation, levels of involucrin are shown in undifferentiated (U) and differentiated (D) CIN6129E keratinocytes for comparison. The experiment was carried out twice using two different clones of each E2-expressing keratinocyte line. Very similar results were obtained in each experiment. (D) Quantification of levels of the SR proteins shown in (B). The graph shows the mean and standard deviation from the mean of three separate experiments. Values were calculated relative to GAPDH levels. Very similar data were obtained with two different clones of each keratinocyte line. (E) Western blot quantification of levels of L1 protein in differentiated NFKs stably maintaining wild type HPV31 and in NFKs stably maintaining point mutant E2:I73L HPV31 genomes. The graph shows the mean and standard deviation from the mean from three separate experiments. Values were calculated relative to GAPDH levels. P-values were calculated using a student's T-test.

Figure 3 Klymenko et al

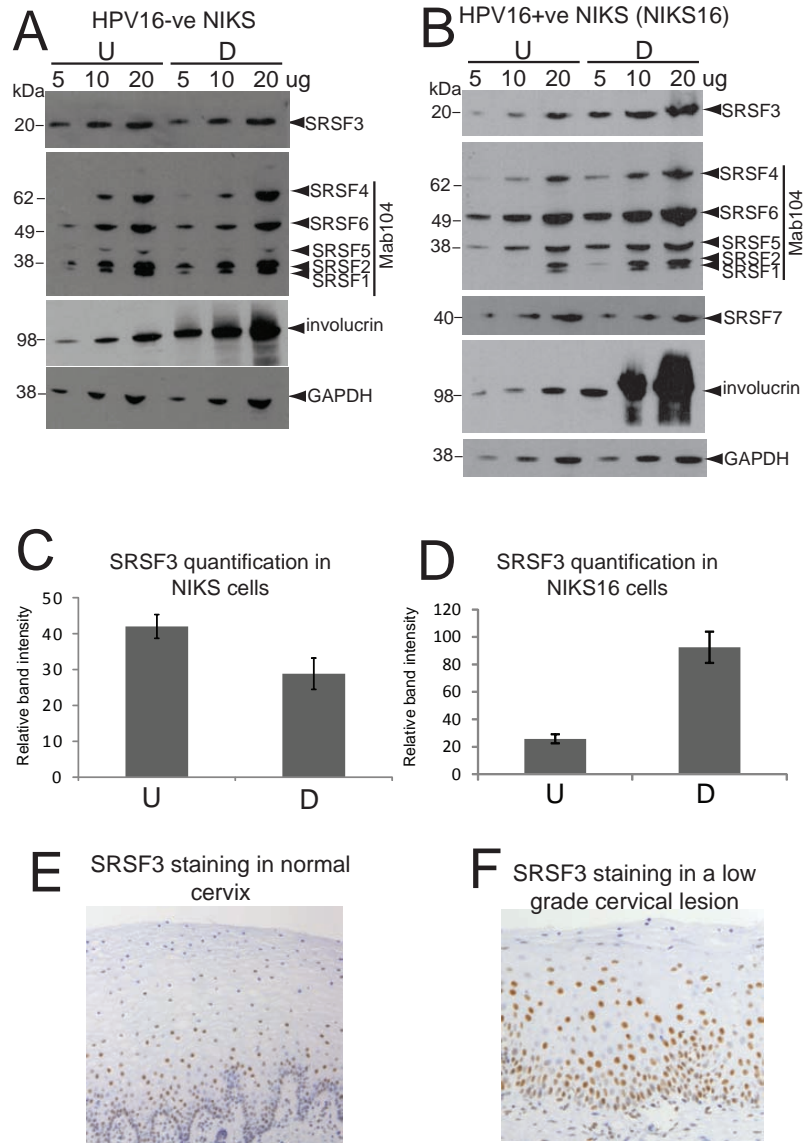


Figure 3 SRSF3 levels increase upon differentiation of HPV16-infected cells (A) SRSF1-3 levels do not alter upon differentiation of HPV-negative NIKS cells. Semi-quantitative western blot analysis of levels of SR proteins in undifferentiated (U) and differentiated (D) NIKS cells. 5, 10 or 20 μ g protein extracts were loaded as indicated above the blots. SRSF3 was detected with a specific monoclonal antibody B7H4. SRSF1, 2, 4, 5 and 6 were detected with Mab104. SRSF7 was not tested in this experiment because its levels did not change upon HPV infection [24]. Involucrin was detected as a control for differentiated epithelial cells. GAPDH was detected as a protein loading control. (B) SRSF1-3 levels are increased in differentiated NIKS16 cells. Semi-quantitative western blot analysis of levels of SR proteins in undifferentiated (U) and differentiated (D) NIKS16 cells. 5, 10 or 20 μ g protein extracts were loaded as indicated above the blots. SRSF3 and 7 were detected with specific monoclonal antibodies. SRSF1, 2, 4, 5 and 6 were detected with Mab104. Involucrin was detected as a control for differentiated epithelial cells. GAPDH was detected as a protein loading control. The experiments in (A) and (B) were carried out at least three times with very similar results obtained each time. (C) Quantification of SRSF3 levels in undifferentiated (U) and differentiated (D) HPV-negative NIKS cells. The graph shows the mean and standard deviation from the mean from three separate experiments. Values were calculated relative to GAPDH levels. (D) Quantification of SRSF3 levels in undifferentiated (U) and differentiated (D) NIKS16 cells. The graph shows the mean and standard deviation from the mean from five separate experiments. Values were calculated relative to GAPDH levels. (E) Immunohistochemical staining of a representative normal cervical epithelium (10 normal lesions were stained) with an antibody against SRSF3. Note the strongly stained nuclei are present mainly in the lower epithelial layers. (F) Immunohistochemical staining of a representative low grade cervical lesion (10 low grade lesions were stained) with an antibody against SRSF3. Note the strongly stained nuclei in the mid to upper epithelial layers. The picture in E is taken at lower magnification than that in F to show that the staining pattern is consistent over an area of tissue.

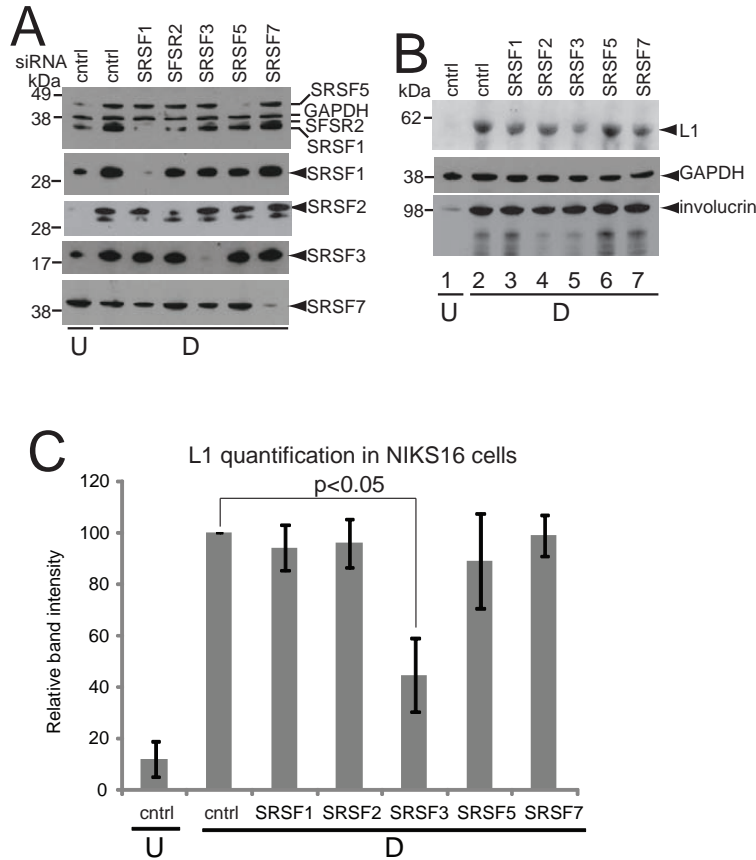


Figure 4. SRSF3 is required for L1 protein expression. (A) Western blot analysis of SR protein depletion for the experiment shown in (E). SRSF1, 2, 3 and 7 were detected with specific antibodies. SRSF1, 2 and 5 were detected with Mab104. The panel showing SRSF2 depletion was probed with Mab104. Only the area of the blot showing SRSF2 and SRSF1 bands is shown. SRSF2 is the upper band. U, undifferentiated, D, differentiated NIKS16 cells. Cntrl, cells transfected with a control siRNA. (B) SRSF3 depletion results in reduced L1 protein levels. Western blot analysis of L1 protein in, differentiated keratinocytes (D) with the various SR proteins depleted. (U), undifferentiated NIKS16 cell extracts were analysed as a negative control for L1 protein expression. Cntrl, cells transfected with a control siRNA. Involucrin was detected as a control for differentiated epithelial cells. GAPDH was detected as a protein loading control. (C) Quantification of levels of L1 protein in NIK16 keratinocytes transfected with a control siRNA (cntrl) or siRNAs against each of the SR proteins shown on the X-axis. U, undifferentiated keratinocytes D, differentiated keratinocytes. The graph shows the mean and standard deviation from the mean from three separate experiments. Values were calculated relative to GAPDH levels. P-value was calculated using a student's T-test.

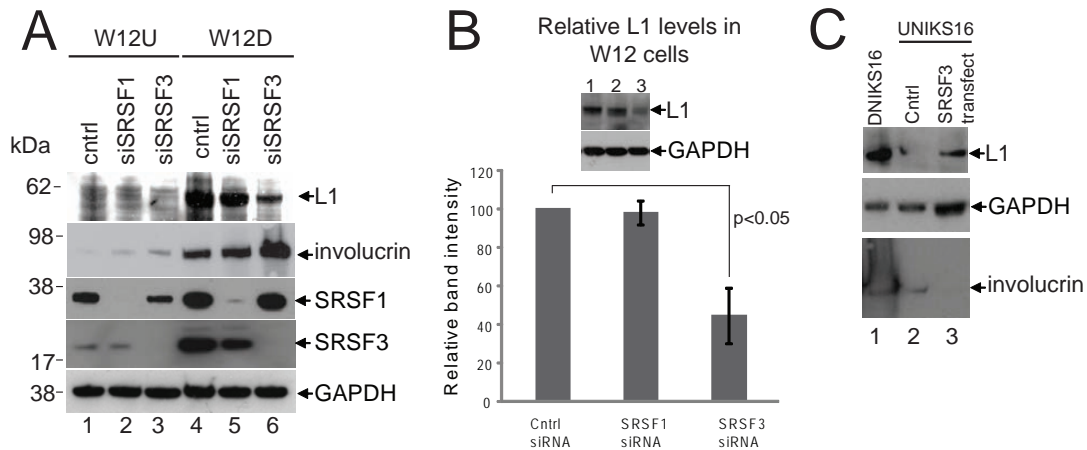


Figure 5 SRSF3 controls L1 protein levels in W12E cells. (A) Western blot analysis of SRSF1 and 3 protein depletion by siRNA in U, undifferentiated and D, differentiated W12E cells. Much higher levels of SRSF1 and SRSF3 were detected in differentiated W12E cells as expected [24]. L1 protein was not detected in U, undifferentiated W12E cells even with SRSF1 or SRSF3 depletion. L1 protein was detected in D, differentiating W12E cells but levels were reduced in the cells treated with siRNA against SRSF3. Cntrl, cells transfected with a control siRNA. Involucrin was detected as a control for differentiated epithelial cells. GAPDH was detected as a protein loading control. This experiment was carried out five times and very similar data were obtained each time. (B) Western blot quantification of levels of L1 protein in differentiated W12E keratinocytes transfected with a control siRNA (cntrl) or siRNAs against SRSF1 or SRSF3. The graph shows the mean and standard deviation from the mean from three separate experiments. Values were calculated relative to GAPDH levels. P-value was calculated using a student's T-test. (C) Undifferentiated NIKS16 cells were transfected with vector control (Cntrl) or a plasmid encoding the SRSF3 protein (SRSF3 transfect). Differentiated NIKS16 cell extracts (DNIKS16) were probed as a control for L1 expression. GAPDH was used as a loading control. Involucrin was detected as a marker of epithelial differentiation. The level of involucrin was so high in the differentiated compared to the undifferentiated NIKS16 cell extracts that the blot was bleached out in lane 1. The experiment was repeated twice and very similar data were obtained each time.

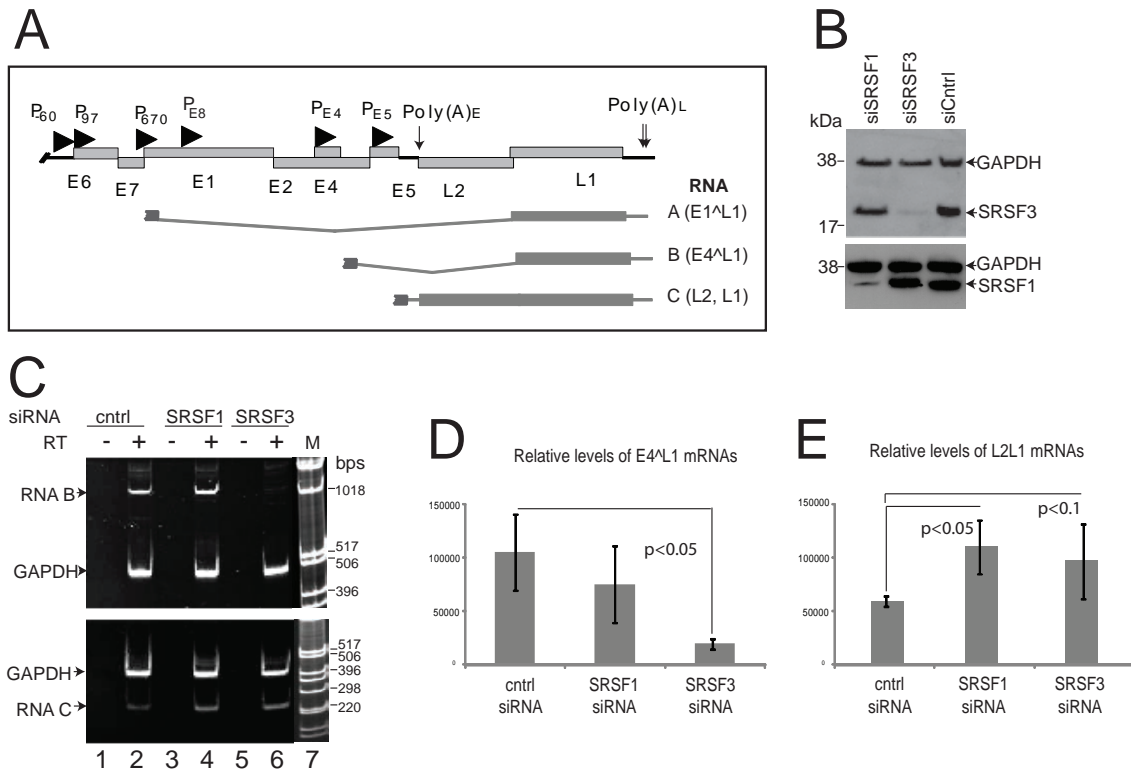


Figure 6 SRSF3 controls levels of the E4^{L1} mRNA. (A) Schematic diagram of the HPV16 genome. Open reading frames are represented with light grey boxes with the gene names beneath. Promoters are indicated with arrowheads. Polyadenylation sites are indicated with downward arrows (Poly(A)E/L). The major late mRNA splicing events with the open reading frames (dark grey boxes) and introns (dark grey lines) are indicated below the genome map. Each mRNA is shown truncated at the 5' end to indicate the presence of several possible initiation sites for transcription of these mRNAs [10]. The open reading frames in each mRNA are indicated to the right hand side. (B) Western blot analysis of levels of SRSF1 and SRSF3 in NIKS16 cells following siRNA knock down. siSRSF1, siRNA against SRSF1. siSRSF3, siRNA against SRSF3. siCtrl, control siRNA. GAPDH was detected as a protein loading control. (C) RT-PCR detection of HPV late mRNAs type B (upper gel) or type C (lower gel) in cDNA synthesized from RNA isolated from the same cell populations analysed in (B). Ctrl, control siRNA. SRSF1, siRNA against SRSF1. SRSF3, siRNA against SRSF3. RT, reverse transcriptase. M, 1Kb DNA marker ladder. GAPDH was amplified as an internal control for RNA concentrations. (D) Relative levels of E4^{L1} spliced mRNA B compared to GAPDH mRNA determined by qPCR. The mean and standard deviation from the mean are shown from three separate experiments. (E) Relative levels of unspliced L2L1 mRNAs C compared to GAPDH mRNA determined by qPCR. The mean and standard deviation from the mean are shown from three separate experiments. P-values were calculated using a student's T-test.