The mouse papillomavirus epigenetic signature is characterised by DNA hypermethylation after lesion regression

Allison M. Tschirley¹, Peter A. Stockwell¹, Euan J. Rodger¹, Oliver Eltherington¹, Ian M. Morison¹, Neil Christensen², Aniruddha Chatterjee^{1*}, Marilyn Hibma^{1*Ψ}.

Affiliations:

¹Department of Pathology, University of Otago, Dunedin, Otago, New Zealand.

²Department of Pathology, Pennsylvania State University, State College, PA, USA.

*Joint senior authors

ΨCorresponding author: Marilyn Hibma, Department of Pathology, Otago Medical School, University of Otago, Dunedin 9054, New Zealand

Department of Pathology

Otago Medical School, Dunedin

Dunedin 9054

New Zealand

merilyn.hibma@otago.ac.nz

Ph. +64 3 479 7726

Short title: The epigenetic signature of mouse papillomavirus

Key words: DNA methylation, papillomavirus, epigenetics, hypermethylation

ABSTRACT

The β genus of human papillomaviruses (HPVs) infect cutaneous epidermis. They contribute to the development of cutaneous squamous cell carcinoma (cSCC) in individuals with epidermodysplasia verruciformis, and increase susceptibility to UV-induced cSCC. This has been demonstrated in UV-exposed mice previously infected with mouse papillomavirus (MmuPV1). However, the mechanism by which β -HPVs contribute to cSCC is unclear. We propose that viral infection leaves a DNA methylation signature following resolution of the active lesion that may contribute to increased susceptibility to UV-induced cSCC.

To test this, we carried out Reduced Representation Bisulphite Sequencing on DNA from tail skin of mice with actively infected lesions, MmuPV1-infected then healed lesions (regressed infection), and mock-infected control mice. Genome-scale DNA methylation libraries were generated and analysed for differentially methylated regions throughout the genome, and for HPV sequences.

We found that DNA of active lesions was not differentially methylated compared to matched control mice. In contrast, 834 differentially methylated fragments were identified in regressed lesions compared to mock-infected control skin. An analysis of MmuPV1 viral DNA demonstrated retention of viral DNA in some of the lesions that had regressed. Overall, the viral sequences identified showed over-representation of sequences from the E4 region. The DNA hypermethylation that we found in regressed MmuPV1 lesions may be a factor in the increased susceptibility of mice to UV-induced cSCC.

AUTHOR SUMMARY

Papillomavirus infections can be asymptomatic, can cause warts, and in some cases can lead to cancer. There is direct evidence for mouse papillomavirus infection resulting in increased susceptibility to UV-induced cutaneous squamous cell carcinoma in a mouse model. We propose that DNA methylation following viral infection may contribute to the increased susceptibility. We describe the DNA methylation landscape during an active infection with mouse papillomavirus and following regression of the lesion. We found that there were very few differentially methylated DNA fragments during active infection. In contrast, over 800 differentially methylated DNA fragments were identified following regression of the lesion. This is the first description of the genome-wide DNA methylation landscape for mouse papillomavirus, to our knowledge. The dramatic DNA hypermethylation that we observe following resolution of infection may contribute to a ‘hit and run’ mechanism for the increased susceptibility to UV-induced cancer by papillomaviruses.

INTRODUCTION

Cutaneous squamous cell carcinoma (cSCC) is one of the most common cancers world-wide (1). Mortality is at least ten-fold higher in immunosuppressed individuals compared with those who are immune competent (2). Although ultra-violet (UV) light is the primary initiator of oncogenesis for cSCC, the β -human papillomavirus (β -HPV) genus has been implicated in cSCC causality. However, this proposition is confounded by the observation that, unlike cervical cancer, β -HPV DNA is not always detected in the resultant tumour. This has led to the formulation of a viral ‘hit and run’ hypothesis, whereby the β -HPV contributes to the process of carcinogenesis but is then lost from the cSCC (3).

β -HPVs are almost ubiquitously present on the skin of individuals. Unlike the mucosal family of α -HPVs, some members of which have been shown to directly cause a variety of anogenital cancers, β -HPVs are directly linked to the causality of skin malignancies only in special cases, particular systemic epidermodysplasia verruciformis (EV) (4). Several lines of *in vitro* and *in vivo* evidence have implicated β -HPVs in the development of UV-induced cSCC (5-8). For example, EV patients are more likely to develop SCC (4) and keratinocytes exposed to β -HPV E6 proteins are protected from the apoptotic pathway triggered by UVB irradiation (9).

Until recently, *in vivo* experimentation to assess the role of papillomavirus in UV-induced cSCC has been hindered by a lack of naturally occurring, accessible murine papillomavirus models. Mouse papillomavirus type 1 (MmuPV1) has more recently been described by Ingle and colleagues in Mumbai, India (10, 11). This virus is a murine homologue of the HPVs, bearing particular resemblance to the β -HPV genus in several ways including causing the formation of cutaneous papillomas following infection of immunosuppressed mice.

Immunosuppression during MmuPV1 infection leads to an increased risk of cSCC from chemical carcinogens or UV light (12, 13). It is proposed that the virus induces epigenetic changes within the host that increase the risk of UV carcinogenesis (8). It is also proposed that immunosuppression suppresses antigen-specific immune defence involving the viral E7 protein that is a CD8 T-cell target in MmuPV1 infected cells (14, 15).

The circular DNA of MmuPV1 encodes seven open reading frames (ORFs) found in other HPV types (E1, E2, E4, E6, E7, L1 and L2). Each of these ORFs is differentially expressed in the viral life cycle, similarly to other papillomaviruses (11). It has not yet been established whether all of the functions of the MmuPV1 viral proteins function the same way as HPVs, but importantly it appears that E6 and E7 are still required for oncogenesis. Similar to the cutaneous HPV types in the β family, MmuPV1 does not contain an E5 ORF (16). The E4 protein is known to accumulate at high levels during productive infection (17). The loss of E4 expression in β -HPV lesions correlates with an increase in cervical intraepithelial neoplasia (CIN) grade in cervical SCC patients (18).

Methylation of DNA is an epigenetic change that can alter expression of genes that has been linked to many cancers including cSCC (19-21). Several DNA tumour viruses such as EBV, HBV and HPV have been found to dysregulate host gene DNA methylation (22). α -HPVs regulate methylation (23), however the effects of this methylation on host susceptibility to cancer once the virus infection has resolved is unknown. Methylation occurs predominantly at CpG dinucleotides, where a cytosine is followed by a guanine. CpG dinucleotides are generally underrepresented in the genomes of both DNA viruses and their hosts (24).

115 Cellular gene methylation in the context of high-risk α -HPVs or cSCC is well-documented (19,
116 25-27). Cancers associated with HPV are similar to most other cancers in their propensity towards
117 global hypomethylation (28). Some of the most frequently hypermethylated genes in cSCC include
118 *CADMI*, *CDHI* and *DAPKI* (19). Methylation of *DAPKI* is associated with cSCC development
119 and progression.

120

121 There is a paucity of data on DNA methylation in the context of MmuPV1 infections, and this
122 study is the first to describe DNA methylation in resolved lesions following MmuPV1 infection.

123

RESULTS

Immune-suppressed BALB/c mice develop and maintain active MmuPV1 infection

BALB/c mice demonstrate intermediate susceptibility to MmuPV1 infection when immunosuppressed with CsA (15). Handisurya *et al.* (15) found that five out of eight (63 %) BALB/c mice produced MmuPV1 lesions following infection. In this study, the proportion of mice that were infected was much lower (20 %), and some mice required a second dose of virus for a visible lesion to appear. Lesions visible to the naked eye appeared on the tail of CsA treated mice between 49 and 63 days post infection, and any mice in the infected groups that did not develop visible lesions were excluded from the study (Fig. 1A). The mean lesion length on the infected mice was 2.4 ± 0.3 mm. CsA treatment was withdrawn for the regressor group to promote lesion regression. All mice in the regressed group had healed their lesions (assessed by visual and tactile tests) by 31 days post CsA withdrawal.

Reduced Representation Bisulphite Sequencing (RRBS) was carried out on tissue samples from treated and control mice, in order to determine the effects of viral infection and regression on the methylation profile, and CsA treatment and scarification. On average, we generated $9.3 \times 10^6 \pm 0.5 \times 10^6$ uniquely aligned reads for each of the samples tested (Table 1). The mean mapping efficiency was $47 \% \pm 1.3 \%$, which is consistent with the expected level of efficiency for a bisulphite converted genome.

When we compared tissues from the regressed mice and their matched controls we found that there was a marked difference in DNA methylation between the groups. The comparison of the regressed (R1 – 3) versus the regressed control group (RC1 – 3) yielded 834 significantly differentially methylated fragments (DMFs) with a False Discovery Rate (FDR) of 5 % or less using a stringent multiple test correction (Benjamini-Hochberg (Table 1).

Table 1. Methylation alignment and efficiency by individual sample and number of differentially methylated fragments (DMFs) using Benjamini-Hochberg correction for each group comparison.

Comparison	Individual	No. of unique alignments	Mapping efficiency	No. of DMFs at 5% FDR
Regressed	R1	6,942,498	44.3%	834
	R2	7,861,430	48.6%	
	R3	7,240,969	47.0%	
Reg. Control	RC1	11,964,235	50.6%	
	RC2	12,197,807	49.8%	
	RC3	9,824,427	39.5%	
Active	A1	7,449,228	49.6%	0
	A2	10,712,796	33.7%	
	A3	6,442,790	49.5%	
Act. Control	AC1	13,217,138	38.3%	
	AC2	9,198,760	49.8%	
	AC3	10,374,183	51.9%	
No CsA	NCsA1	6,434,780	47.8%	0
	NCsA2	6,853,519	49.1%	
CsA	CsA1	9,428,602	49.5%	
	CsA2	12,426,447	52.7%	
Non-scarified	CsA1, CsA2			0
Scarified	AC1, AC2, AC3			
Active	A1, A2, A3			0
Regressed	R1, R2, R3			

In contrast, there was little difference in DNA methylation between the tissues from actively infected mice and their controls. Only five DMFs were identified in the actively infected group (A1 – 3) when compared to the matched control samples (AC1 – 3). Treatment of mice with CsA alone (CsA1 – 2 v. NCsA1 – 2) or scarification alone (A1 – 3 v. CsA1 – 2) had little effect on differential methylation between the condition and the control.

To assess the relationship of global RRBS methylation between the samples that were tested, we performed hierarchical clustering of all of the analysed high coverage autosomal fragments from the 16 RRBS methylomes in this study (Fig. 1C). We found that the infected samples grouped into two clusters, irrespective of whether they were actively infected or if the lesion had regressed.

Samples from uninfected mice generally grouped into two other clusters, separated from the infected mice.

Genome-scale DNA methylation identifies extensive hypermethylation in regressed skin

We carried out further analysis to determine if the DMFs were hyper- or hypomethylated. The striking finding from this analysis was that 98% of the DMFs identified in the regressed group were hypermethylated (Mann-Whitney U, $P < 0.0001$) (Fig. 2A).

To further probe the DMFs in the regressed MmuPV1 lesions relative to their location in the genome, we generated DNA methylation maps for the specific genomic regions. We found that methylation was increased overall in the DMFs from all genomic regions that we assessed in regressed skin when compared with controls (Fig. 2B). Significantly increased methylation was detected in DMFs genome-wide (Mann-Whitney U, $P < 0.0001$), proximal to gene promoters (defined as +/- 1 kb of transcription start site) (Mann-Whitney U, $P < 0.0001$), in the core promoter region (defined as +/- 500 bp of transcription start site) (Mann-Whitney U, $P < 0.0001$), in gene introns (Mann-Whitney U, $P < 0.0001$) and gene exons (Mann-Whitney U, $P < 0.0001$), and in intergenic regions (Mann-Whitney U, $P < 0.0001$).

To further investigate the methylation landscape, we performed an analysis of the wider CpG island topography (10 kb upstream to 1 kb downstream of transcription start sites). CpG islands are defined as stretches of DNA 500–1500 bp long with a CG:GC ratio of more than 0.6 (29). We found that most of the hypermethylation was in the CpG island cores. Of 328 hypermethylated fragments that were detected within the region of interest, 66 % were found in the core, 4 % were in the CpG island shore (up to 2 kb upstream and 2 kb downstream of the CpG island core), 1 % were in the interface between core and shore (termed core-shore) and <1 % were in the shelf (up

to 2 kb upstream and 2 kb downstream of the shore). 27 % of the regions were outside of these CpG island features (Fig. 2C). Of the nine hypomethylated fragments that were detected, five were in the CpG island core (Fig. 2D).

Genes in the promoter region are skewed extensively towards hypermethylation in regressed skin

As promoter DNA methylation strongly influences expression from the corresponding gene, we specifically investigated differential methylation at the core promoter regions. Of the 221 DMFs in these regions, 214 were hypermethylated and seven were hypomethylated. When comparing individual samples within the groups, all genes with > 20 % methylation difference were hypermethylated in the regressed group (Fig. 3A). This was also the case at > 40 % methylation difference (Fig. 3B), where all genes were hypermethylated in regressed tissue compared to control tissue.

Hypermethylated genes in regressed skin are enriched for cellular senescence and cancer related pathways, and for CTCF and RNA Polymerase II binding

To document the function of the genes that were hypermethylated in the core promoter regions of regressed skin, we performed functional enrichment analysis using two different sources (KEGG and Wiki Pathways for Mouse). We found genes involved in cellular senescence/cell cycle, mRNA processing, and p53 signalling were significantly enriched ($P < 0.05$) in both the analyses (Fig. 4A). Disease associated pathways, particularly cancer pathways, were also enriched in our analysis. To place our findings in a broader epigenomic context, we utilized histone modification ChIP-Seq data from the ENCODE project as well as from the Epigenomics roadmap. The genes harbouring hypermethylation in the core promoters of regressed skin were predominantly enriched for the repressive chromatin mark H3K9me3 (in several cell types, see Fig. 4B). We also found

significant enrichment for several active histone marks, especially H3K27ac and H3K29ac (Fig. 4B). Overlap analysis of ENCODE transcription factor ChIP-seq data identified enrichment of several transcription factors for the hypermethylated genes (Fig. 4C). The most enriched (significant in many cell types) transcription factors were CTCF and POLR2A (RNA Polymerase II Subunit) (Fig. 4C).

MmuPV1 viral reads are highly concentrated in the E4 portion of the genome

We probed the data for MmuPV1 viral reads to determine whether viral DNA remained after lesion regression (Fig. 5A). Traces of the viral genome were present in all actively infected tissue (5.7 %, 0.2 % and 5.4 % of viral reads were viral in origin for samples A1 to A3, respectively). Interestingly, viral reads were also detected in two of three regressed tissues (7.2 %, 0.0 % and 5.3 % for samples R1 to R3, respectively). The small number of background reads mapping to the MmuPV1 genome in uninfected groups most likely occurred through the misalignment of bisulphite reads, as previously reported (30) (S1 Table). The most intriguing finding of viral read analysis was the high proportion of viral reads within the E4 region of the MmuPV1 genome (Fig. 5B, C). In two of the actively infected samples (A1 and A3), more than half of all reads aligned to this region. In one of the regressed lesions, sample R1, 74 % of the viral reads mapped to E4. The L2 region also had high clustering, contributing up to 25 % of all viral reads in the actively infected samples.

237 S1 Table. Viral reads were present within regressed lesions. Total and percent viral reads within
238 E4 region.

Group	Sample	Total reads	Viral reads	% viral reads	% in E4	Infection status
Regressed	R1	15,674,849	1,129,427	7.2	74	Regressed lesion
	R2	16,166,560	5,871	0	0	Regressed lesion
	R3	15,392,780	819,114	5.3	59.9	Regressed lesion
Reg. Control	RC1	23,666,799	3,713	0	0	Mock-infected
	RC2	24,496,535	2,868	0	0	Mock-infected
	RC3	24,877,419	4,143	0	0	Mock-infected
Active	A1	15,020,374	849,677	5.7	52.1	Active lesion
	A2	31,835,719	48,521	0.2	11.4	Active lesion
	A3	13,025,548	708,976	5.4	56	Active lesion
Act. Control	AC1	34,512,945	6,246	0	0	Mock-infected
	AC2	18,479,427	2,534	0	0	Mock-infected
	AC3	19,996,210	2,475	0	0	Mock-infected
CsA	CsA1	19,044,226	1,271	0	0	CsA/No infection
	CsA2	23,584,967	1,829	0	0	CsA/No infection
No CsA	NCsA1	13,470,444	1,106	0	0	No CsA/Not inf.
	NCsA2	13,946,497	1,282	0	0	No CsA/Not inf.

239

240

241

242

243

244

245

246

247

248

249

250

251

252

DISCUSSION

Here we show the first reported analysis of DNA methylation following regression of MmuPV1 infection. Interestingly, differences in methylation were only detected following lesion regression, and not during active infection. These differences were highly skewed towards hypermethylation.

In our analysis, *DAPK1* was hypermethylated in core promoters in regressed skin compared to control tissue. Hypermethylation is associated with cSCC as reported by Li and colleagues, who found that methylation of *DAPK1* was associated with cSCC but in their case it did not correlate with the presence of cutaneous HPV (19). *DAPK1* is considered likely to be a tumour suppressor. Our finding is supported by RNAseq analysis performed by Hu and colleagues on MmuPV1 infected tissue where *DAPK1* was downregulated by 0.76 log₂ fold change from infected (with no lesion) to infected (with lesion) (31). The hypermethylation of the tumour suppressor gene *DAPK1* indicates that this gene is ‘switched off’ in MmuPV1 regressed lesions. *DAPK1* presence in the core promoter region suggests that its hypermethylation may influence the *DAPK1* expression profile, and thus could be a possible route to increased susceptibility to cSCC after regression of HPV infection.

The identification of hypermethylation of POLR2A is consistent with the observation that one of the enriched pathways that we identified was mRNA processing, suggesting a role of hypermethylation in RNA processing in regressed skin. We also identified hypermethylation of the CTCF promoter. Variable CTCF binding has been strongly linked with differential DNA methylation in multiple human cell types and it has been shown that CTCF binding patterns are markedly different in normal versus immortal human cells. In immortal cells, disruption of CTCF binding was strongly associated with hypermethylation (32). Our analysis provides evidence for

277 hypermethylation and disruption of CTCF binding in regressed skin compared to normal skin
278 controls.

279

280 Interestingly, *MSH6* and *PAPD7* are DNA mismatch repair genes that were also hypermethylated
281 within the core promoter region of regressed skin (5 % and 6.5 % methylation difference,
282 respectively). *MSH6* is important for UVB-induced apoptosis (33, 34). Hypermethylation of *MSH6*
283 in regressed tissue prevents skin cells from effectively undergoing UVB-induced apoptosis.
284 *PAPD7* has been found to be upregulated by HSV1 infection (35). Its hypermethylation indicates
285 that the damage response no longer occurs. Indeed, Uberoi *et al.* (13) found that immune-
286 competent mice infected with MmuPV1 and exposed to UV radiation develop lesions that progress
287 to cSCC. Key genes such as *MSH6*, whose promoters are hypermethylated by MmuPV1, may
288 render the skin more sensitive to the damaging effects of UV, providing the virally mediated
289 susceptibility ‘hit’ that promotes UV-induced SCC.

290

291 Our analysis of viral reads found evidence of viral DNA in regressed tissues with no remaining
292 visible lesion. To our knowledge, this is the first evidence of the presence of MmuPV1 DNA in
293 infected tissue where lesions were no longer observable. Viral gene expression has been detected
294 by Xue *et al.* (17) well before lesions are observed. These investigators also observed
295 asynchronous papillomavirus growth, which was similarly found by our analysis of viral reads,
296 where one visible lesion in the actively infected group (sample A2) yielded only 0.2% viral reads
297 while skin with no visible lesion in the regressed group (sample R1) showed the highest level of
298 viral DNA at 7.2 %.

299

300 The low viral reads of sample A2 could be partly explained by the size of the lesion, which was
301 smallest of the three active lesions (height: 0 mm, length: 2 mm, width: 1 mm) and could contribute

to its methylation profile clustering with two of the samples from the regressed lesion group (Fig. 1C). Interestingly, R1 which had the highest level of viral DNA clustered with A2 (0.2 %), R2 (0 %) and NCsA1 (did not receive virus). Therefore, it is not solely the amount of viral DNA that dictates the methylation patterns of the samples, nor is it the current or past immune suppression, as NCsA1 did not receive cyclosporin at any stage. Rabbit oral papillomavirus DNA and RNA has been found to persist for at least a year post lesion regression (36). The high viral reads in two out of three regressed lesions suggests viral latency in MmuPV1 is also possible. The diversity in the quantity of viral DNA found could be explained by the stage of epithelial differentiation that the cells were in when collected. A progression of this research could involve re-establishment of immunosuppression in regressed animals to see if the latent infection re-emerges at the site of the original lesion.

The most striking finding within the analysis of viral reads was the high clustering of reads in the E4 region. E4 protein is found to be abundant in upper layers of the epithelium of productive lesions (37). E4 expression is followed by the late structural proteins, explaining the concurrent, though considerably lower, expression of L2. This pattern was also found by Xue *et al.* (17), who looked solely at active MmuPV1 infections in the tail, muzzle and ear of mice.

There is considerable variability in the pattern of expression of different papillomavirus types, depending on the animal model and tissue tropism (38). High levels of E4 transcripts in the upper layers of the epithelium, similarly to the DNA viral reads found here, follow the pattern found in low-grade squamous intraepithelial lesions of high-risk cervical HPV infections (39). In an analysis of 92 different HPV genomes, the E4 region was found to have the highest proportion of CpG sites (40). This is despite the fact that CpG sites tend to be under-represented in viruses,

326 which may be a way to avoid methylation by host methyltransferases or CpG mediated immune
327 responses.

328

329 In summary, our study outlined the methylation landscape and viral content and distribution of
330 MmuPV1 in regressed lesion tissue. We highlight several genes hypermethylated in the core
331 promoter regions of regressed skin that could influence susceptibility to cSCC without the need
332 for active lesions to be present. In future it would be useful to leave regressed skin for longer and
333 continue to monitor viral content. Experiments using UV as a co-factor would also aid in validating
334 the function or loss of function of the identified genes through methylation.

335

336

337

338

339

340

341

342

MATERIALS AND METHODS

Animals

Female specific pathogen-free (SPF) BALB/c mice, aged 8 - 12 weeks, were obtained from the Hercus Taieri Research Unit at the University of Otago and housed in a SPF environment. A total of 63 mice were initially in the study. For systemic immune suppression, Cyclosporin A (CsA, Novartis Neoral®, 100 mg/mL) diluted in PBS was administered to the 30 control and 30 infected mice by gavage five times per week (75 mg/kg, 11.25 mg/mL) starting one week prior to MmuPV1 infection. The remaining three mice were the non-CsA treated mice. Thirty of the mice were infected with MmuPV1 and were monitored for the development of visible lesions and the other thirty mice were mock infected with PBS. Mice that had not developed visible lesions were re-infected at day 49. Six animals developed visible lesions, and the remainder of the infected mice that did not develop visible lesions were excluded from the study. The final groups (Fig. 1) consisted firstly of mice (n = 3) that were harvested at day 122 with “active” lesions (A1, A2, A3) and three matched mock-infected controls (AC1, AC2, AC3), and secondly, mice that had active lesions but were harvested at day 133 with “regressed” lesions (R1, R2, R3) following the withdrawal of CsA at day 102, and three matched CsA treated, mock-infected controls (RC1, RC2, RC3). Groups of two CsA treated mice (CsA1, CsA2) and two controls (NCsA1, NCsA2) were included in the study to test whether any changes in DNA methylation occurred as a result of CsA treatment alone. Commencing at day -3, Baytril was added to drinking water for all immune suppressed mice to reduce the risk of bacterial infection.

Ethics statement

All animal experiments were approved by the Animal Ethics Committee at the University of Otago (ethical approval AEC56/14) and were performed in accordance with the New Zealand Animal

Welfare Act (1999) and adhered to National Animal Ethics Committee (Ministry of Primary Industries, NZ) guidelines.

MmuPVI infection

Mouse papillomavirus was isolated from lesions on the tails of Foxn1^{nu}/Foxn1^{nu} immunocompromised mice and diluted 1:10 in PBS (41). On day -3, tail skin was scarified using a hand-held rotary tool with a felt wheel attached (42). Animals were anaesthetised with Isoflurane, and a 1.5 cm long mark was placed at the base of the tail using a felt tip pen, and the rotary tool was moved 10 times along the mark at a speed of approximately 15,000 rpm to remove the upper epidermal layer. Marcaïn was applied dropwise over the wound area for pain relief. At day 0, mice were placed in restrainers and the scarified region was scored lightly with a needle tip. Inoculation was performed by placing 20 µL of prepared virus onto the scored region of the scarified area (Fig. 1A). Mock-infected mice received PBS. At day 49, animals who had yet to show signs of a lesion were re-infected with 20 µL of prepared virus. At day 102, CsA was discontinued for the regressor group, to allow the lesions to heal naturally.

Sample preparation

Beginning at day 102 for the regressor group, the lesions were outlined with marker pen for identification of tissue to be harvested once lesions regressed. The actively infected group mice were culled at day 122 post-infection. Tail skin was removed and epidermal dissociation of the epithelium was performed following the protocol of Lichti *et al.* (43). DNA was extracted from epidermal tissue (approximately 3 - 5 mm² per lesion) using a DNAeasy blood and tissue kit (Qiagen). Regressed and CsA group mice were culled at day 133 post-infection, once lesions were no longer palpable, and the regressed group was culled at 31 days post withdrawal of CsA.

393 *RRBS library preparation*

394 RRBS was performed on 16 samples of extracted genomic DNA as previously described (44-46).
 395 Briefly, DNA was digested into fragments using the MspI enzyme, followed by end-repair and
 396 ligation of sequencing adaptors. The fragments were size selected (150 - 325 bp) and bisulphite
 397 converted prior to a PCR amplification step. The quality of the 16 libraries was confirmed using
 398 bioanalyser traces before being sequenced on an Illumina HiSeq2500 machine (Table 1).

400 *DNA methylation analysis*

401 Differential methylation analysis was performed using the previously reported differential
 402 methylation analysis package (DMP) (47). Briefly, only fragments with at least two CpG sites
 403 and at least ten hits per CpG were included in the data (hereon called high coverage fragments).
 404 We have previously described this fragment-based approach (48-51). High coverage fragments
 405 were mapped to the mouse genome and the closest protein coding genes were annotated. An
 406 ANOVA test was applied to the fragments of the experimental and control groups and regions
 407 showing methylation differences with a fold change of ≥ 1.5 and significant *P*-values were
 408 identified. Five different group comparisons were performed (Regressed vs. Control, Active vs.
 409 Control, Active vs. Regressed, CsA vs. no CsA, Scarified vs. Non-scarified) (Fig. 1B). Active and
 410 regressed tissues were the experimental groups. Non-scarified vs. scarified was a methylation
 411 comparison only and compared animal samples from the CsA and actively infected control groups.
 412 These control tissues were tested to determine whether scarification impacted on methylation. CsA
 413 vs. no CsA tissues was assessed to determine whether CsA administration impacted methylation.

415 *Analysis of viral reads and pathway analysis*

416 Methylation data were aligned to the MmuPV1 genome using bowtie in the Bismark Bisulphite
 417 Read Mapper (52). Sequences that mapped to the MmuPV1 genome were also aligned to the *Mus*

musculus genome to ascertain whether mapping was unique to the viral genome. The mapper BSMAP(z) was then used to verify the alignments shown by Bismark. This mapping software confirmed that the reads were not indicating artefactual behaviour of Bismark alone. The Integrative Genomics Viewer (IGV) was used to visualise reads within the MmuPV1 genome (53). Lanes were auto-scaled to show proportionate representation between the samples.

Unsupervised hierarchical clustering was performed in the R environment using the Euclidean distance metric of all analysed high coverage autosomal fragments ($n = 126,681$) from the 16 RRBS methylomes. Pathway analysis was performed on a gene list containing the 221 genes differentially expressed in the core promoter region using the online platform Enrichr (54). The enrichment P values calculated by Enrichr are from a modified Fisher's exact test, which is a proportion test that assumes a binomial distribution and independence for probability of any gene belonging to any set.

Statistical analyses

All data are presented as the mean $\pm$ standard error of the mean and statistical analyses were done in GraphPad Prism7®. Mann-Whitney U was used to test for differences between groups. The original FDR method of Benjamini and Hochberg under the multiple T tests function was used to increase the power of the data and identify significant P values with a false discovery rate (FDR) of 5%.

Data availability

The RRBS data generated as part of this study were submitted to Gene Expression Omnibus (GEO) with the accession number GSE160868.

443 **AUTHOR CONTRIBUTIONS**

444 Conceptualisation of research: MH, AT, AC, NC, IM, OE; Data Curation: AT, PS, AC; Formal
445 Analysis: PS, AT, MH, AC, ER; Methodology: AT, NC, MH, AJ; Software: AC, PS, ER; Funding
446 acquisition: MH, AT, AC, IM; Writing: AT, OE, MH, AC, IM, NC.

447

448 **ACKNOWLEDGMENTS**

449 We thank Jiafen Hu (Pennsylvania State University) for supplying additional RNAseq data from
450 MmuPV1 tissues. The authors declare no conflicts of interest. AC would like to thank the
451 Rutherford Discovery Fellowship (Royal Society of NZ) for their support.

452

453 **CONFLICT OF INTEREST**

454 The authors declare no conflicts of interest.

REFERENCES

1. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. *Br J Dermatol*. 2012;166(5):1069-80.
2. Mittal A, Colegio OR. Skin Cancers in Organ Transplant Recipients. *Am J Transplant*. 2017;17(10):2509-30.
3. Hufbauer M, Akgül B. Molecular Mechanisms of Human Papillomavirus Induced Skin Carcinogenesis. *Viruses*. 2017;9(7).
4. de Jong SJ, Imahorn E, Itin P, Uitto J, Orth G, Jouanguy E, et al. Epidermodysplasia verruciformis: inborn errors of immunity to human beta-papillomaviruses. *Front Microbiol*. 2018;9:1222.
5. Tommasino M. HPV and skin carcinogenesis. *Papillomavirus Research*. 2019;7:129-31.
6. Bavinck JNB, Plasmeijer EI, Feltkamp MCW. β -Papillomavirus Infection and Skin Cancer. *J Invest Dermatol*. 2008;128(6):1355-8.
7. Karagas MR, Nelson HH, Sehr P, Waterboer T, Stukel TA, Andrew A, et al. Human papillomavirus infection and incidence of squamous cell and basal cell carcinomas of the skin. *J Natl Cancer Inst*. 2006;98(6):389-95.
8. Underbrink MP, Howie HL, Bedard KM, Koop JI, Galloway DA. E6 proteins from multiple human betapapillomavirus types degrade Bak and protect keratinocytes from apoptosis after UVB irradiation. *J Virol*. 2008;82(21):10408-17.
9. Underbrink MP, Howie HL, Bedard KM, Koop JI, Galloway DA. E6 proteins from multiple human betapapillomavirus types degrade Bak and protect keratinocytes from apoptosis after UVB irradiation. *J Virol*. 2008;82(21):10408-17.
10. Ingle A, Ghim S, Joh J, Chepkoech I, Bennett Jenson A, Sundberg JP. Novel laboratory mouse papillomavirus (MusPV) infection. *Vet Pathol*. 2011;48(2):500-5.

- 479 11. Joh J, Jenson AB, King W, Proctor M, Ingle A, Sundberg JP, et al. Genomic analysis of
480 the first laboratory-mouse papillomavirus. *J Gen Virol*. 2011;92(3):692-8.
- 481 12. Spurgeon ME, Uberoi A, McGregor SM, Wei T, Ward-Shaw E, Lambert PF. A novel in
482 vivo infection model to study papillomavirus-mediated disease of the female reproductive tract.
483 *MBio*. 2019;10(2).
- 484 13. Uberoi A, Yoshida S, Frazer IH, Pitot HC, Lambert PF. Role of Ultraviolet Radiation in
485 Papillomavirus-Induced Disease. *PLoS Pathog*. 2016;12(5):e1005664.
- 486 14. Joh J, Ghim S-j, Chilton PM, Sundberg JP, Park J, Wilcher SA, et al. MmuPV1 infection
487 and tumor development of T cell-deficient mice is prevented by passively transferred
488 hyperimmune sera from normal congenic mice immunized with MmuPV1 virus-like particles
489 (VLPs). *Exp Mol Pathol*. 2016;100(1):212-9.
- 490 15. Handisurya A, Day PM, Thompson CD, Bonelli M, Lowy DR, Schiller JT. Strain-Specific
491 Properties and T Cells Regulate the Susceptibility to Papilloma Induction by *Mus musculus*
492 Papillomavirus 1. *PLoS Pathog*. 2014;10(8):e1004314.
- 493 16. Doorbar J, Egawa N, Griffin H, Kranjec C, Murakami I. Human papillomavirus molecular
494 biology and disease association. *Rev Med Virol*. 2015;25(S1):2-23.
- 495 17. Xue X-Y, Majerciak V, Uberoi A, Kim B-H, Gotte D, Chen X, et al. The full transcription
496 map of mouse papillomavirus type 1 (MmuPV1) in mouse wart tissues. *PLoS Pathog*.
497 2017;13(11):e1006715.
- 498 18. Cerasuolo A, Annunziata C, Tortora M, Starita N, Stellato G, Greggi S, et al. Comparative
499 analysis of HPV16 gene expression profiles in cervical and in oropharyngeal squamous cell
500 carcinoma. *Oncotarget*. 2017;8(21):34070-81.
- 501 19. Li L, Jiang M, Feng Q, Kiviat NB, Stern JE, Hawes S, et al. Aberrant methylation changes
502 detected in cutaneous squamous cell carcinoma of immunocompetent individuals. *Cell Biochem*
503 *Biophys*. 2015;72(2):599-604.

504 20. Hervás-Marín D, Higgins F, Sanmartín O, López-Guerrero JA, Bañó MC, Igual JC, et al.
505 Genome wide DNA methylation profiling identifies specific epigenetic features in high-risk
506 cutaneous squamous cell carcinoma. PLoS One. 2019;14(12):e0223341.

507 21. Rodríguez-Paredes M, Bormann F, Raddatz G, Gutekunst J, Lucena-Porcel C, Köhler F, et
508 al. Methylation profiling identifies two subclasses of squamous cell carcinoma related to distinct
509 cells of origin. Nature Communications. 2018;9(1):577.

510 22. Kuss-Duerkop SK, Westrich JA, Pyeon D. DNA Tumor Virus Regulation of Host DNA
511 Methylation and Its Implications for Immune Evasion and Oncogenesis. Viruses. 2018;10(2).

512 23. Durzynska J, Lesniewicz K, Poreba E. Human papillomaviruses in epigenetic regulations.
513 Mutation Research/Reviews in Mutation Research. 2017;772:36-50.

514 24. Balakrishnan L, Milavetz B. Epigenetic Regulation of Viral Biological Processes. Viruses.
515 2017;9(11):346.

516 25. Leonard SM, Wei W, Collins SI, Pereira M, Diyaf A, Constandinou-Williams C, et al.
517 Oncogenic human papillomavirus imposes an instructive pattern of DNA methylation changes
518 which parallel the natural history of cervical HPV infection in young women. Carcinogenesis.
519 2012;33(7):1286-93.

520 26. Sen P, Ganguly P, Ganguly N. Modulation of DNA methylation by human papillomavirus
521 E6 and E7 oncoproteins in cervical cancer. Oncology letters. 2018;15(1):11-22.

522 27. Soto D, Song C, McLaughlin-Drubin ME. Epigenetic Alterations in Human
523 Papillomavirus-Associated Cancers. Viruses. 2017;9(9):248.

524 28. Dueñas-González A, Lizano M, Candelaria M, Cetina L, Arce C, Cervera E. Epigenetics
525 of cervical cancer. An overview and therapeutic perspectives. Mol Cancer. 2005;4:38-.

526 29. Cross SH, Bird AP. CpG islands and genes. Curr Opin Genet Dev. 1995;5(3):309-14.

527 30. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC
528 Bioinformatics. 2009;10(1):232.

529 31. Hu J, Cladel NM, Budgeon LR, Balogh KK, Christensen ND. The Mouse Papillomavirus
530 Infection Model. *Viruses*. 2017;9(9):246.

531 32. Lang F, Li X, Vladimirova O, Hu B, Chen G, Xiao Y, et al. CTCF interacts with the lytic
532 HSV-1 genome to promote viral transcription. *Sci Rep*. 2017;7(1):39861.

533 33. Young L, Listgarten J, Trotter M, Andrew S, Tron V. Evidence that dysregulated DNA
534 mismatch repair characterizes human nonmelanoma skin cancer. *Br J Dermatol*. 2008;158(1):59-
535 69.

536 34. Young LC, Hays JB, Tron VA, Andrew SE. DNA mismatch repair proteins: potential
537 guardians against genomic instability and tumorigenesis induced by ultraviolet photoproducts. *J*
538 *Invest Dermatol*. 2003;121(3):435-40.

539 35. Hu B, Li X, Huo Y, Yu Y, Zhang Q, Chen G, et al. Cellular responses to HSV-1 infection
540 are linked to specific types of alterations in the host transcriptome. *Sci Rep*. 2016;6(1):28075.

541 36. Maglennon GA, McIntosh P, Doorbar J. Persistence of viral DNA in the epithelial basal
542 layer suggests a model for papillomavirus latency following immune regression. *Virology*.
543 2011;414(2):153-63.

544 37. Middleton K, Peh W, Southern S, Griffin H, Sotlar K, Nakahara T, et al. Organization of
545 human papillomavirus productive cycle during neoplastic progression provides a basis for
546 selection of diagnostic markers. *J Virol*. 2003;77(19):10186-201.

547 38. Peh WL, Middleton K, Christensen N, Nicholls P, Egawa K, Sotlar K, et al. Life Cycle
548 Heterogeneity in Animal Models of Human Papillomavirus-Associated Disease. *J Virol*.
549 2002;76(20):10401.

550 39. Kremer WW, Vink FJ, van Zummeren M, Dreyer G, Rozendaal L, Doorbar J, et al.
551 Characterization of cervical biopsies of women with HIV and HPV co-infection using p16ink4a,
552 ki-67 and HPV E4 immunohistochemistry and DNA methylation. *Modern pathology : an official*
553 *journal of the United States and Canadian Academy of Pathology, Inc*. 2020;33(10):1968-78.

554 40. Galván SC, Martínez-Salazar M, Galván VM, Méndez R, Díaz-Contreras GT, Alvarado-
555 Hermida M, et al. Analysis of CpG methylation sites and CGI among human papillomavirus DNA
556 genomes. BMC Genomics. 2011;12:580-.

557 41. Cladel NM, Budgeon LR, Cooper TK, Balogh KK, Hu J, Christensen ND. Secondary
558 infections, expanded tissue tropism, and evidence for malignant potential in immunocompromised
559 mice infected with Mus musculus papillomavirus 1 DNA and virus. J Virol. 2013;87(16):9391-5.

560 42. Handisurya A, Day PM, Thompson CD, Buck CB, Pang YY, Lowy DR, et al.
561 Characterization of Mus musculus papillomavirus 1 infection in situ reveals an unusual pattern of
562 late gene expression and capsid protein localization. J Virol. 2013;87(24):13214-25.

563 43. Lichti U, Anders J, Yuspa SH. Isolation and short-term culture of primary keratinocytes,
564 hair follicle populations and dermal cells from newborn mice and keratinocytes from adult mice
565 for in vitro analysis and for grafting to immunodeficient mice. Nat Protoc. 2008;3(5):799-810.

566 44. Chatterjee A, Rodger EJ, Stockwell PA, Weeks RJ, Morison IM. Technical considerations
567 for reduced representation bisulfite sequencing with multiplexed libraries. J Biomed Biotechnol.
568 2012;2012:741542-.

569 45. Chatterjee A, Rodger EJ, Stockwell PA, Le Mée G, Morison IM. Generating Multiple
570 Base-Resolution DNA Methylomes Using Reduced Representation Bisulfite Sequencing. In:
571 Seymour GJ, Cullinan MP, Heng NCK, editors. Oral Biology: Molecular Techniques and
572 Applications. New York, NY: Springer New York; 2017. p. 279-98.

573 46. Ludgate JL, Wright J, Stockwell PA, Morison IM, Eccles MR, Chatterjee A. A streamlined
574 method for analysing genome-wide DNA methylation patterns from low amounts of FFPE DNA.
575 BMC Med Genomics. 2017;10(1):54.

576 47. Stockwell PA, Chatterjee A, Rodger EJ, Morison IM. DMAP: differential methylation
577 analysis package for RRBS and WGBS data. Bioinformatics. 2014;30(13):1814-22.

578 48. Chatterjee A, Stockwell PA, Ahn A, Rodger EJ, Leichter AL, Eccles MR. Genome-wide
579 methylation sequencing of paired primary and metastatic cell lines identifies common DNA
580 methylation changes and a role for EBF3 as a candidate epigenetic driver of melanoma metastasis.
581 Oncotarget. 2017;8(4):6085-101.

582 49. Chatterjee A, Stockwell PA, Rodger EJ, Duncan EJ, Parry MF, Weeks RJ, et al. Genome-
583 wide DNA methylation map of human neutrophils reveals widespread inter-individual epigenetic
584 variation. Sci Rep. 2015;5:17328.

585 50. Chatterjee A, Stockwell PA, Rodger EJ, Morison IM. Genome-scale DNA methylome and
586 transcriptome profiling of human neutrophils. Scientific Data. 2016;3(1):160019.

587 51. Chatterjee A, Rodger EJ, Ahn A, Stockwell PA, Parry M, Motwani J, et al. Marked Global
588 DNA Hypomethylation Is Associated with Constitutive PD-L1 Expression in Melanoma. iScience.
589 2018;4:312-25.

590 52. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-
591 Seq applications. Bioinformatics. 2011;27(11):1571-2.

592 53. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-
593 performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.

594 54. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and
595 collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.

596

597

598

TABLES AND FIGURES

Fig. 1. Experimental setup for mouse model and preliminary methylation analysis. (A) Timeline of infection model. Colour-coded mouse images represent the three groups as shown in the boxes below. Boxes show sample names of each group and their associated control samples. (B) Representative images of the experimental groups show mouse tails at sample location; (i) Active, (ii) Regressed (original lesion location delineated in black), (iii) CsA. (C) Hierarchical clustering dendrogram of all high-coverage autosomal fragments for the 16 RRBS libraries generated.

Fig. 2. Global methylation description. (A) Difference plots of differentially methylated fragments (DMFs) in comparison of regressed to regressed control tissue (left) and active to active control tissue (right). Horizontal dotted line represents significance cut-off of 5 % FDR. (B) Proportion of methylated DMFs by genomic region for regressed (blue) and regressed control tissues (red). Data are represented in quartiles with whiskers indicating outliers outside of upper and lower quartiles. (C) Within CpG islands, differentially hypermethylated regions (n = 328) and (D) hypomethylated regions (n = 9) were located predominantly in the core.

Fig. 3. Extensive hypermethylation identified in regressed skin. (A) DMFs in the core promoter region with > 20% methylation difference between regressed and control groups by individual sample. (B) DMFs in the core promoter region with > 40% methylation difference between regressed and control groups. Error bars display standard error of the mean.

Fig. 4. Pathway and transcriptional function analysis of differentially methylated fragments (DMFs). Enrichr analysis with default parameters was used to identify pathways, histone modifications and transcription factor targets associated with DMFs in core gene promoter regions. Individual heatmaps show: (A) enriched KEGG and Wiki Pathways (P < 0.05); (B)

624 enriched histone modifications from Epigenomics Roadmap and ENCODE datasets ($P < 0.01$);
 625 and (C) Transcription factor targets from ENCODE ChIP-Seq and TRANSFAC/JASPAR
 626 positional weight matrices ($P < 0.01$).

627

628 Fig. 5. MmuPV1 viral reads within individual samples. (A) Images at time of harvest. Marker pen
 629 was used to delineate the area of the lesions at the height of infection (B - C). Coverage and
 630 distribution of viral reads visualised using IGV software. Coverage depth of reads set to group
 631 auto-scale for (B) regressed (0 - 266,605) and (C) active groups (0 - 541,120).

632

633

634

bioRxiv preprint doi: <https://doi.org/10.1101/2021.04.19.440429>; this version posted April 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1. Experimental setup for mouse model and preliminary methylation analysis. (A) Timeline of infection model. Colour-coded mouse images represent the three groups as shown in the boxes below. Boxes show sample names of each group and their associated control samples. (B) Representative images of the experimental groups show mouse tails at sample location; (i) Active, (ii) Regressed (original lesion location delineated in black), (iii) CsA. (C) Hierarchical clustering dendrogram of the 16 RRBS libraries generated.

A.

bioRxiv preprint doi: <https://doi.org/10.1101/2021.04.19.440429>; this version posted April 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

B.

C.

D.

Figure 2. Global methylation description. (A) Difference plots of differentially methylated fragments (DMFs) in comparison of regressed to regressed control tissue (left) and active to active control tissue (right). Horizontal dotted line represents significance cut-off of 5 % FDR. (B) Proportion of methylated DMFs by genomic region for regressed (blue) and regressed control tissues (red). Data are represented in quartiles with whiskers indicating outliers outside of upper and lower quartiles. (C) Within CpG islands, differentially hypermethylated regions (n=328) and (D) hypomethylated regions (n=9) were located predominantly in the core.

Figure 3. Extensive hypermethylation identified in regressed skin. (A) DMFs in the core promoter region with >20% methylation difference between regressed and control groups by individual sample. (B) DMFs in the core promoter region with >40% methylation difference between regressed and control groups. Error bars display standard error of the mean.

Figure 4. Pathway and transcriptional function analysis of differentially methylated fragments (DMFs). Enrichr analysis with default parameters was used to identify pathways, histone modifications and transcription factor targets associated with DMFs in core gene promoter regions. Individual heatmaps show: (A) enriched KEGG and Wiki Pathways ($P < 0.05$) (B) enriched histone modifications from Epigenomics Roadmap and ENCODE datasets ($P < 0.01$) (C) Transcription factor targets from ENCODE ChIP-Seq and TRANSFAC/JASPAR positional weight matrices ($P < 0.01$).

A.	Animal number	Lesion image	Viral reads
	R1		7.2%
	R2		0%
	R3		5.3%
	A1		5.7%
	A2		0.2%
	A3		5.4%

bioRxiv preprint doi: <https://doi.org/10.1101/2021.04.19.440429>; this version posted April 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 5. MmuPV1 viral reads within individual samples. (A) Images of mouse tails at time of harvest. Marker pen was used to delineate the area of the lesions at the height of infection. (B-C) Coverage and distribution of viral reads visualised using IGV software. Coverage depth of reads set to group autoscale for (B) regressed (0-266,605) and (C) active groups (0-541,120).