Characterization of Naturally Occurring HPV16 Integration Sites Isolated from Cervical Keratinocytes under Noncompetitive Conditions

Keltie L. Dall,[1] Cinzia G. Scarpini,[1] Ian Roberts,[1] David M. Winder,[1] Margaret A. Stanley,[2] Balaji Muralidhar,[1] M. Trent Herdman,[1] Mark R. Pett,[1] and Nicholas Coleman[1,2]

[1]Medical Research Council-Cancer Cell Unit, Hutchison/Medical Research Council Research Center; [2]Department of Pathology, University of Cambridge, Cambridge, United Kingdom

Abstract

As the high-risk human papillomavirus (HPV) integrants seen in anogenital carcinomas represent the end-point of a clonal selection process, we used the W12 model to study the naturally occurring integration events that exist in HPV16-infected cervical keratinocytes before integrant selection. We performed limiting dilution cloning to identify integrants present in cells that also maintain episomes. Such integrants arise in a natural context and exist in a noncompetitive environment, as they are transcriptionally repressed by episome-derived E2. We found that integration can occur at any time during episome maintenance, providing biological support for epidemiologic observations that persistent HPV infection is a major risk factor in cervical carcinogenesis. Of 24 different integration sites isolated from a single nonclonal population of W12, 12 (50%) occurred within chromosome bands containing a common fragile site (CFS), similar to observations for selected integrants *in vivo*. This suggests that such regions represent relatively accessible sites for insertion of foreign DNA, rather than conferring a selective advantage when disrupted. Interestingly, however, integrants and CFSs did not accurately colocalize. We further observed that local DNA rearrangements occur frequently and rapidly after the integration event. The majority of integrants were in chromosome bands containing a cancer-associated coding gene or microRNA, indicating that integration occurs commonly in these regions, regardless of selective pressure. The cancer-associated genes were generally a considerable distance from the integration site, and there was no evidence for altered expression of nine strong candidate genes. These latter observations do not support an important role for HPV16 integration in causing insertional mutagenesis. [Cancer Res 2008;68(20):8249–59]

Introduction

Cervical carcinoma is caused by persistent infection with high-risk human papillomavirus (HR-HPV), most commonly HPV16 and HPV18(1). Squamous cell carcinoma (SCC) accounts for 80% to 85% of cases(2) and arises from noninvasive squamous intraepithelial lesions (SIL). Integration of HR-HPV into the host genome is seen in most cervical SCC (3, 4). Cervical neoplasms are clonal, usually with a single integration site (5), suggesting that particular integration events are selected during carcinogenesis. Mechanisms of integration are not well understood, but the process most likely involves double-strand breaks (DSB) in viral and host genomes, followed by DNA ligation by host proteins (6).

Previous studies of the sites of HPV integration into host chromosomes have used clinical samples of SCC and SIL, as well as cell lines derived from SCC (4, 5, 7). The integrants in such specimens reflect the end point of a clonal selection process. It is not known how typical they are of the range of integration events that occur in HR-HPV infected cells nor what factors confer a selective advantage to particular integrants. Over 200 selected HPV16 and HPV18 integration sites have been reported. These are widely distributed across the genome, although many are mapped at low resolution. There seems to be preferential integration near common fragile sites (CFS), specific chromosomal loci that are particularly prone to forming DSBs (8, 9), with around 50% of selected HPV16 and HPV18 integration sites being in the same chromosomal band as a CFS (5, 7, 10, 11).

Important unanswered questions in HR-HPV biology concern the mechanisms by which the site of integration contributes to selection of a particular integrant. One area of controversy is whether CFSs are frequently associated with selected integrants because integration at these sites confers a competitive advantage to the cell or simply because CFSs are relatively accessible sites for integration (4, 5, 11). In addition, it is not well understood whether integrated HPV affects transcription of adjacent host genes by a process of insertional mutagenesis and/or whether the site of integration affects viral transcription. There is some evidence to suggest that both scenarios may occur (10, 12, 13), although there is very little supporting functional data. Many previous studies have suffered from a variety of limitations, including a lack of suitably matched controls, use of cells in which chromosome breaks are induced or occur at a high spontaneous rate, and minimal investigation of host protein levels. In addition, and of particular importance, analysis of HR-HPV integrants in clinical samples does not allow distinction between the process of integration itself and the factors associated with selection.

In the present study, we have used the W12 cervical keratinocyte model to examine the range of HPV16 integration sites that occur before integrant selection during cervical neoplastic progression. W12 is a nonclonal cell culture, propagated from a cervical low-grade SIL (LSIL) that arose after natural cervical infection with HPV16, which represents a unique system for studying early events in HPV-associated carcinogenesis (14). At early passages, Southern blotting reveals ~100 HPV16 episomes per cell with no detectable

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Nicholas Coleman, Medical Research Council-Cancer Cell Unit, Medical Research Council/Hutchison Research Center, Hills Road, Cambridge, CB2 0XZ, United Kingdom. Phone: 44-1223-763285; Fax: 44-1223-763284; E-mail: nc109@cam.ac.uk.

doi:10.1158/0008-5472.CAN-08-1741

integrants, and the cells recapitulate an LSIL in organotypic tissue culture. Long-term *in vitro* culture series are characterized by spontaneous clearance of episomes and the emergence of cells containing integrated HPV16, with different integrants being selected in different culture series. These changes are associated with the development of high-level genomic instability and phenotypic progression through high-grade SIL to SCC (15).

We recently showed that early passage W12 cells also contain latent integrants that are repressed by episomally derived E2 protein (16). After episome clearance by mechanisms such as induction of IFN-stimulated genes, cells containing only episomes undergo apoptosis whereas those containing latent integration events emerge by virtue of retention of E6 and E7 expression (17). This creates a competitive environment in which the integration event(s) conferring the strongest growth advantage comes to dominate the population. We reasoned that within a noncompetitive environment, wherein E2 is retained, a range of integrants would be present, regardless of whether or not they confer a selective growth advantage after episome loss. Such integrants would represent those that occur naturally after HPV16 infection, as they arise in a background of episomally expressed E6 and E7, rather than in a setting where chromosome breaks are chemically induced or occur at a high spontaneous rate (7).

To characterize the range of naturally occurring HPV16 integration sites that exist in a noncompetitive environment, we isolated single-cell clones from an early point of a nonclonal W12 culture series (W12 Series 2; ref. 16) and observed 24 different integration sites across the genome. We undertook high-resolution mapping of the integration sites, which we performed directly after clone isolation to minimize any confounding effects of integrant-associated genomic instability. Our data provide detailed information concerning the relationships between naturally occurring HPV16 integrants and the host genome.

Materials and Methods

Cell culture and single-cell cloning. The starting population from which clones were isolated was nonclonal W12 Series 2 at passage 12 or 13 (W12Ser2p12 or W12Ser2p13), which is diploid and harbors ~100 HPV16 episomes per cell, with no integrated HPV16 DNA detectable by Southern blot (16, 18). During previous long-term passage of this nonclonal population, cells with HPV16 integrated at 8q24.21 emerged and were selected by passage 24 (this integrant is, henceforth, called the "selected 8q24.21 integrant").

W12 cells were propagated in monolayer culture, as described in detail elsewhere(14, 19). The cloning strategy is summarized in Fig. 1*A*. For single-cell cloning, cells were seeded on 96-well plates at a density of 0.2 cells per well on a layer of lethally irradiated mouse 3T3J2 fibroblast feeder cells (density 1×10^5 cells/cm^2). Colonies that arose were expanded for ~30 population doublings in total from the initial single cell of origin and DNA then extracted for analysis of HPV16 physical state. In all but one experiment, feeder cells were maintained for all ~30 population doublings. The exception was an initial training round of cloning (see Results), in which feeder cells were allowed to diminish over the first ~15 population doublings (to test the hypothesis that this would encourage episome loss), after which they were reintroduced for ~15 population doublings to encourage growth (Fig. 1*A*). The clones isolated from this round were named A(f)-Q(f) after isolation, with the suffix "f" to indicate that their expansion had included feeder cell support and the parentheses to denote that feeders had not been present throughout. In the four subsequent cloning rounds, we used the suffix f without parentheses, as feeders were present for the whole expansion phase. Clones from round 2 were named Xf-Zf and A2f-U2F, whereas clones from rounds 3, 4, and 5 were named A3f-J3f, A4f-F4f, and A5f-C5f, respectively.

In all cases, cells were next transferred to feeder-free conditions to encourage episomal loss or further loss in the case of the initial training experiment (Fig. 1*A*). After four passages (~25 population doublings) without feeder support, genomic DNA (gDNA) was extracted for analysis of HPV16 physical state and integration site. From this point, the cells were called clones A-Q, etc., without a suffix. Finally, total RNA and protein were extracted for use in viral physical state assessment and quantification of host gene expression. To establish culture conditions more representative of the *in vivo* environment before RNA/protein isolation, cells were returned to growth with feeder support for a single passage (about six population doublings; Fig. 1*A*).

Normal ectocervical keratinocytes were obtained after hysterectomy for disease unrelated to the cervix and cultured as described in detail elsewhere (19). Other cells used were the cervical carcinoma lines SiHa (HPV16-positive), C33A (HPV-negative), and HeLa (HPV18-positive), the Burkitt's lymphoma line BJAB, and the alveolar rhabdomyosarcoma line SJRH30. All were obtained from American Type Culture Collection and grown as recommended.

Southern analysis of HPV16. Southern blotting was performed as described in detail elsewhere (18). A total of 5 to 10 μg of gDNA was restriction enzyme digested, electrophoresed through a 1% agarose gel, and transferred to a Hybond-N$^+$ nylon membrane for subsequent hybridization. The restriction enzymes used were *Hin*dIII, *Bam*HI, *Eco*RI, and *Pst*I. Probe was prepared by excision of full-length HPV16 DNA from the pspHPV16 plasmid (14), followed by labeling with [α-32P]dCTP by random priming.

Restriction site PCR. Restriction site PCR (RS-PCR) enables amplification of unknown nucleotide sequences adjacent to known nucleotide sequences (20). The use of RS-PCR to generate PCR products spanning HR-HPV host junction fragments in cervical carcinomas has been described previously (11). A total of 100 ng gDNA from each clone was amplified with various combinations of eight HPV16-specific primers and six restriction site oligonucleotide primers, using low-stringency cycling conditions followed by high-stringency nested PCR. PCR products were then sequenced using the appropriate HPV16-specific sequencing primers. All primers and PCR conditions were as described previously (11).

Amplification of papillomavirus oncogene transcripts. Cells were lysed *in situ* by application of Bio-RNA Xcell2 solution (BioGene), and total RNA extracts were prepared according to the manufacturer's instructions. Contaminating DNA was removed using Turbo DNA-free (Ambion). Amplification of papillomavirus oncogene transcripts (APOT) was performed as described previously using the same primers and PCR cycling conditions (21). Briefly, 1 μg total RNA was reverse transcribed and then used as a template for low-stringency nested PCR. Products were analyzed by gel separation, gel excision, and purification, followed by direct sequencing (Medical Research Council Geneservice Ltd.).

PCR screening for integration sites and quantification of host RNA and protein levels. PCR screening for integration sites detected by RS-PCR was applied to individual clones, as well as to the starting W12Ser2p12 nonclonal population. Levels of host mRNA and protein were determined by quantitative PCR and quantitative Western blotting, respectively. For full details of these techniques, refer to Supplementary Methods.

Results

Isolation of HPV Integrants in Noncompetitive Conditions

To assess the range of naturally occurring HPV16 integration events in low-passage W12Ser2, we performed limiting dilution cloning to sample single cells within the population. We reasoned that the integrants present would coexist with episomes expressing inhibitory E2.

Initial cloning round. In the initial training experiment (Fig. 1*A*) after cloning on feeder cells, we deliberately allowed the feeder layer to diminish to test the hypothesis that this would encourage loss of episomes through a stress response (17). We predicted that episome clearance would induce apoptosis of clones

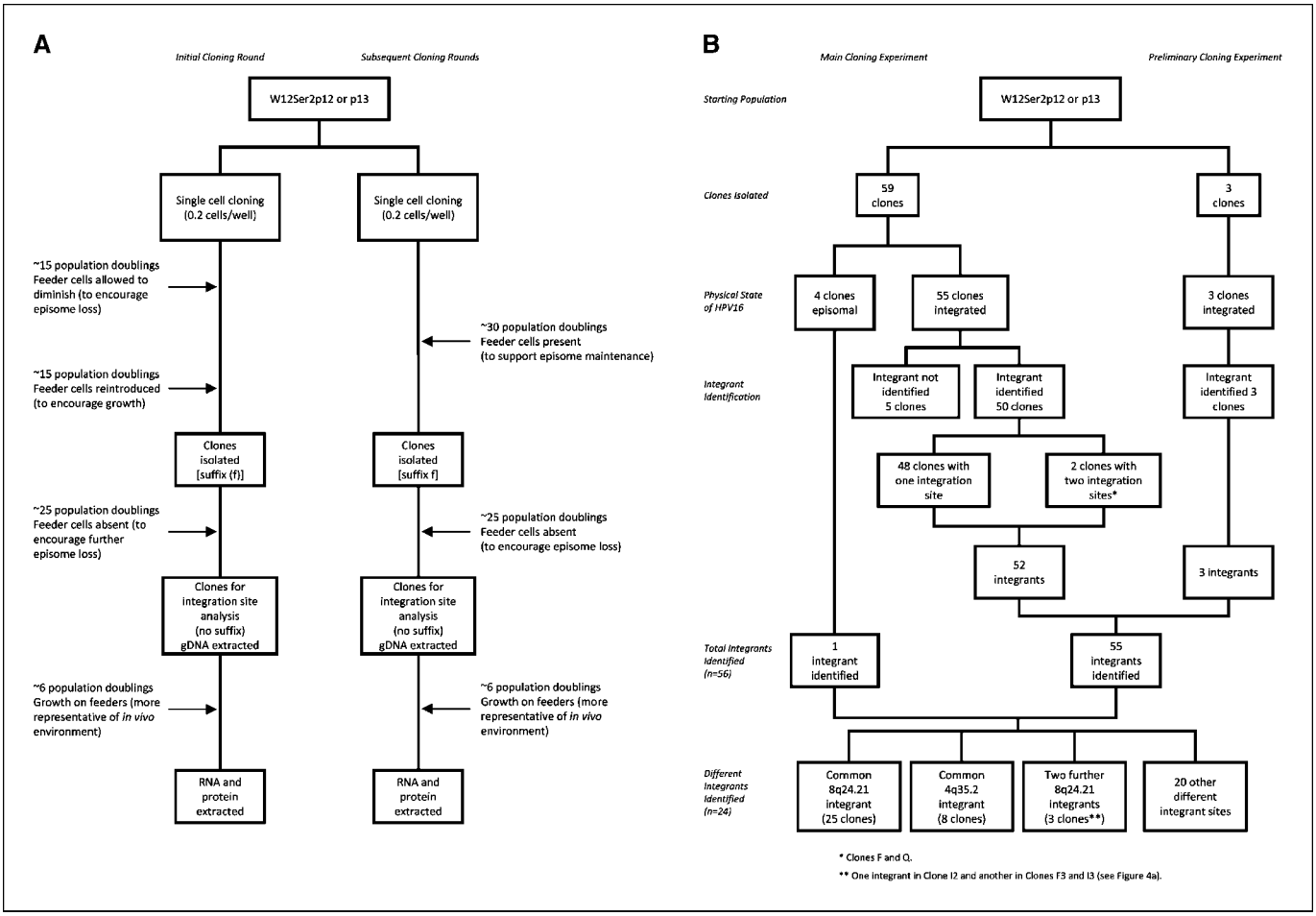

Figure 1. Summary of W12 cell cloning strategy and HPV16 integration sites identified. *A,* summarizes the cloning strategy used in the initial training experiment (*left stem*) and subsequent cloning rounds (*right stem*). *B,* summarizes the routes by which the 24 different HPV16 integration sites were identified. The preliminary cloning experiments referred to on the right of this panel were separate from those shown in *A* but used the method of single-cell cloning from W12Ser2p12 shown in the right stem of *A*.

that contained episomes only and allow emergence of clones containing integrants (17) whether or not the integrants would confer a selective advantage in a mixed population. Colonies that emerged using this strategy (after ~15 population doublings from the originating cell) were expanded for a further ~15 population doublings (using feeder cells to encourage growth; Fig. 1*A*) and then screened by Southern blot to assess HPV16 physical state. Consistent with our predictions, episomes were maintained in only 5 of the 17 clones (29%) isolated in this round [named A(f)-Q(f)]. Moreover, 11 of the 12 clones with episome loss showed the same banding pattern as the selected 8q24.21 integrant (i.e., the integrant that emerged during long-term passage of nonclonal W12Ser2).

All clones were further expanded in feeder-free conditions for ~25 population doublings, after which they were renamed clones A-Q (Fig. 1*A*). This step induced episomal loss in the five clones that had retained episomes. All 17 clones then underwent Southern blotting and/or sequence analysis using RS-PCR. The 8q24.21 integrant was seen in 14 clones overall, two of which contained a second integrant; at 4q13.3 in clone F and 12q14.3 in clone Q. The remaining three clones each contained an integrant at a different site, namely 3q24, 4q21.23, and 4q35.2. Subcloning of clones F and Q showed that the two detected integrants coexisted in the same cells rather than being present in different subpopulations (data not shown).

Based on RS-PCR data, we designed PCR primers to amplify the host-virus junctions identified. Whereas the selected 8q24.21 integrant and the 4q35.2 integrant were detectable in the starting W12 Series 2 population (and were therefore latent integrants), the other integrants were not (Supplementary Fig. S1). Although we could not exclude that these other integrants were present in an extremely small number of cells in the starting population, we considered it more likely that the relevant integration events occurred *de novo* after cell cloning in a background of E2-expressing episomes, as has been observed previously in W12 (6).

Subsequent cloning rounds. Based on these observations, further revised cloning rounds were designed to encourage the clonal expansion of cells containing episomes. We predicted that this would allow further *de novo* integration events to occur in a noncompetitive environment, as well as enable isolation of any other latent integration sites present in the starting population. In our revised protocol, the feeder cell layer was maintained throughout the cloning and expansion phase (Fig. 1*A*). In four cloning rounds performed (using 2,208 wells, with ~442 cells seeded), W12 grown in feeders showed an overall colony forming efficiency of 10.5%, a value similar to that of nonclonal early

passage W12 when seeded at standard density (15), supporting the notion that the feeder cell layer did not exert significant cell stress. Isolation of integration sites was then achieved by feeder-free culture to encourage episome loss, as before.

In these cloning rounds, 33 of the 42 isolated clones (79%) retained episomes before growth in feeder-free conditions (Fig. 2*A*). After removal of feeders, only 11 of 42 clones (26%) contained the 8q24.21 integrant, whereas the majority of the other clones contained a single, different integration site. Given that cells containing these different sites were able to grow in feeder-free conditions, we concluded that they had the potential to be isolated during the initial cloning round had they been present in the starting population. The increase in the number of different isolated integration sites (relative to the 8q24.21 integrant), therefore, suggested that the subsequent cloning rounds allowed the development and isolation of naturally occurring integration events.

Summary of Clones Derived

In total, 59 clones were isolated from the five limiting dilution cloning rounds combined (Fig. 1*B*). Of these, 55 cleared episomes and showed virus-host junction fragments on Southern analysis. Interestingly, four clones retained episomes, suggesting that they were resistant to stress induced by feeder-free conditions. Using a variety of techniques (see Materials and Methods), we identified HPV16 integration sites in 50 of the 55 clones that cleared episomes, with some sites being confirmed by more than one method (Fig. 1*B*). The integration site of five clones could not be found by any of the techniques applied. Two clones (F and Q) contained two integration sites on different chromosomes, giving a total of 52 integration sites found. The integration sites were initially identified by RS-PCR (n = 23), APOT (n = 6), PCR screening for integration sites identified by the other approaches (n = 6), and Southern blotting using a panel of four restriction enzymes (n = 17). In the latter technique, the integration site was identified based on identical banding patterns to other clones where host-virus junction sequence information was available.

An additional three clones that had been isolated in preliminary experiments (separate to those described here but using the same method of single-cell cloning from W12Ser2p12) were added to the analysis at this stage (Fig. 1*B*). These clones were labeled clone 1, clone 3, and clone 5. The integration site was identified in each clone using RS-PCR. Finally, an integration site was identified by RS-PCR screening in one of the four clones that failed to clear episomes, clone B5, and this was also added to the analysis, giving a total of 56 naturally occurring HPV16 integration sites, isolated in a noncompetitive environment (Fig. 1*B*).

From this final panel, we identified 24 different integration sites (Table 1 and Fig. 2*B* and *C*, for example). Details of the junctional sequence used to identify each site by BLAST searching are given as supplementary data. Two of the 24 sites were seen in more than two clones (i.e., were "common" integration sites). Both of these were shown by RS-PCR to have been latent integrants in the starting W12Ser2p12 population (see above), suggesting that they were isolated at high frequency by virtue of being available for clonal expansion in feeder-free conditions rather than representing discrete *de novo* integration events. The first of these sites was at 8q24.21, the same as the selected 8q24.21 integrant that emerged in

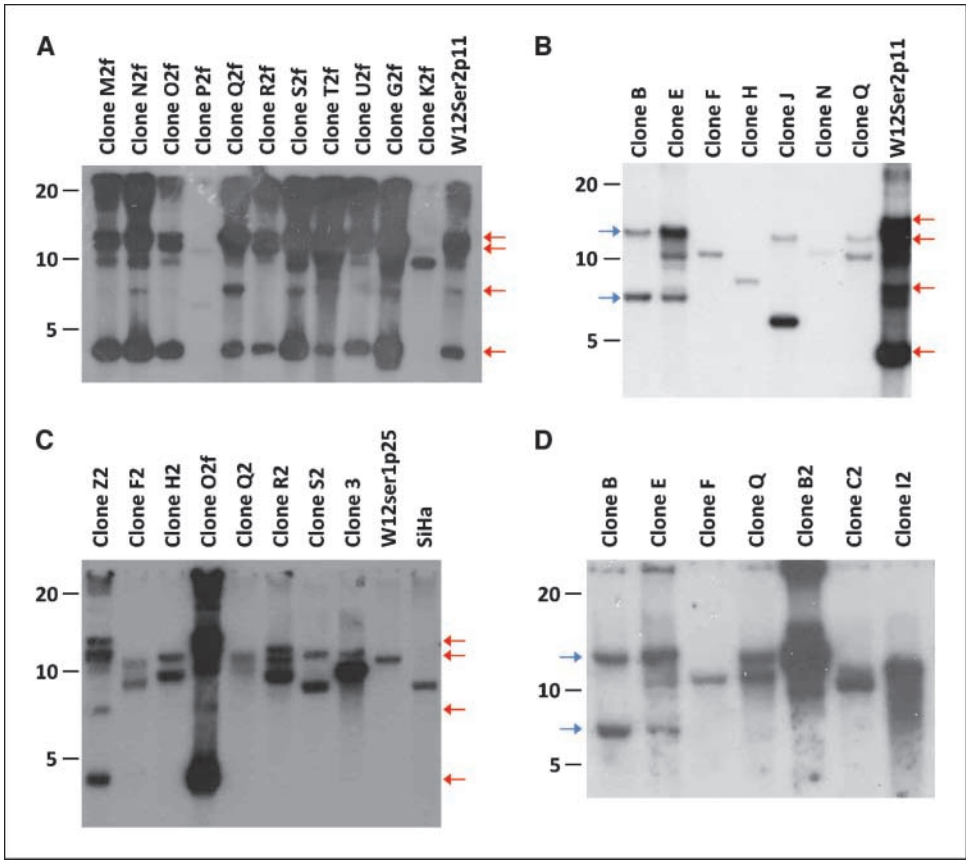

Figure 2. Southern analysis of HPV16 physical state in W12 clones. The panels show representative blots after *Hin*dIII digests. *Red arrows,* sizes of the fragments expected after digestion of HPV16 episomes; *blue arrows,* fragments seen after digestion of clones with HPV16 integrated at the 8q24.21 common integration site. Molecular weight markers (kb) are shown on the left of each blot. The control samples were W12Ser2p11, in which only HPV16 episomes were detectable; W12Ser1p25, which contained HPV16 integrated at 5p15 and had cleared all episomes (15); and the cervical SCC cell line SiHa, which contained HPV16 integrated at 13q21 and no episomes. *A,* clones after expansion on a well-maintained feeder layer before transfer to feeder-free conditions. Most retain episomes, exceptions being clones P2f and K2f. The 8-kb band (*third red arrow from top*) is due to the presence of nicked (linearized) episomes, which vary in their representation in different clones. *B* and *C,* clones with unique integrant bands identified after transfer to feeder-free conditions. Note that clone O2f is included in *C* to show the bands produced by episomal HPV16. *D,* clones in which the same host-viral junction (at 128,476,710 bp of 8q24.21 adjoined to 1,757 bp in HPV16) was isolated by RS-PCR. The different banding patterns indicate local rearrangement at the site of integration.

Table 1. Summary of the 24 different HPV16 integrants isolated from W12

Clone*	Integration site	Technique	CFS †	Nearest gene in direction of transcription	Cancer genes ‡	miRNAs
Clone Q2	1q44	RS-PCR	FRA1I	HNPRU (1,074 kb)		
Clone 3	2p24.1	RS-PCR	FRA2C 2p24.2 (1,633 kb)	HS1BP3 (98 kb)	MYCN (4,814 kb)	
Clone N	3q24	RS-PCR		Within an intron of SLC9A9	GMPS (12,343 kb)	
Clone C5	3q28	RS-PCR		Within an intron of TP73L	LPP (1,741 kb)	mir-28 (1,205 kb)
Clone F	4q13.3 & 8q24.21	RS-PCR		Within intron 1-2 of RASSF6		
Clone H	4q21.23	RS-PCR		Within intron 11-12 of MAPK10	MLLT2 (873 kb)	mir-575 (4q21.22) 3,308 kb
Clone F2	4q31.21	RS-PCR	FRA4C 4q31.1 (1,758 kb)	XR_017925.1 (538 kb)	FBXW7 (10,013 kb)	
Clone J	4q35.2 (two sites)	RS-PCR		FAT (836 kb)	DUX4 (2,434 kb)	
Clone Z	5q11.2	APOT		FST (110 kb)		mir-581 (581.6 kb) 449 and 449b (1,800 kb)
Clone A5	8p11.21	RS-PCR		mir-486 (11.9 kb)	MYST3 (282 kb)	mir-486 (11.9 kb)
Clone B	8q24.21 (two sites)	RS-PCR	FRA8C/D 8q24.1/3 (1176 kb)	POU5F1P1 (pseudogene) 20.1 kb	C-MYC (330 kb)	mir-30b, 30d (8q24.22) mir-548d, 151, 661 (8q24.13/3)
Clone I2	8q24.21	RS-PCR				
Clone F3	8q24.21	RS-PCR				
Clone E3	9p24.3	APOT		DOCK8 (82 kb)	JAK2 (4,795 kb)	
Clone R2	10q22.1	RS-PCR	FRA10D	U6 spliceosomal RNA (111 kb)	MYST4 (2,544 kb)	mir-606 (10q22.2) 3,271 kb
Clone Q	12q14.3& 8q24.21	RS-PCR		WIF1 (484.7 kb)	MDM2 (3,204 kb) CDK4 (7,850 kb)	mir-26a,let-7i,548c (12q14.1/2)
Clone 5	12q14.3	RS-PCR		HMGA2 (309 kb)	MDM2 (3,204 kb) CDK4 (7,850 kb)	mir-26a,let-7i,548c (12q14.1/2)
Clone B5 (episome retaining)	16p13.3	RS-PCR	FRA16A 16p13.1 (5,724 kb)	MRGN1 (39.8 kb)	TSC2 (2,537 kb) CIITA (6,293 kb)	mir-193b, 365 (16p13.1)
Clone H2	17q12 (two sites)	RS-PCR		Within PLXDC1	Nf1 (7,800 kb) and RARA (1,192 kb)	
Clone D2	18q21.2	RS-PCR	FRA18B 18q21.3 (335 kb)	WDR7 (304.5 kb)	MALT1 (2,324 kb)	mir-122a (18q21.31)
Clone O2	19q13.31	APOT	FRA19A (19q13)	Isoform 3 of Q8N6Q3 (71 kb)	BCL3 (1,456 kb) ERCC2 (2,078 kb) AKT2 (3,058 kb)	mir-330, 642, 769, 150 (19q13.32)
Clone G2	21q22.1	APOT		BACH1 (15.7 kb)		mir-802 (21q22.12)
Clone S2	22q12.1	RS-PCR	FRA22B 22q12.2 (1,188 kb)	Within TTC28	CHEK2 (675.7 kb) EWSR1 (1,256 kb)	
Clone 1	Xp22.32	RS-PCR	FRAXB Xp22.31 (671 kb)	ENSG00000210216 (225 kb)		

*The 4q35.2 common integrant (listed for clone J) was seen in eight clones, whereas the 8q24.21 common integrant (listed for clone B) was seen in 25 clones, including clones F and Q. The integrant in clone F3 was also seen in a second clone, I3.
†A CFS is listed if it maps to the same chromosomal band as the integrant. Where the CFS does not map to the same sub-band as the integrant, the location of the CFS and its distance from the integration site is given in brackets. Note that some chromosomal bands are not divided into sub-bands.
‡The cancer-associated coding gene and miR gene nearest to each integration site are also listed. Where a coding gene is orientated in the direction of transcription from the viral promoter, it is shown in green. Where this is not the case, it is shown in red. The *C-MYC* gene at 8q24.21 is uncolored as the integrants at this locus are in both orientations.

long-term nonclonal culture of the starting population; 25 clones had this integration site (including clones B, F, and Q in Table 1). These represented 18 clones that gave an identical pattern of bands on Southern blot, as well as seven further clones that gave unique bands, indicating that rearrangement had occurred at the site of integration (Fig. 2*D*). The second common integration site was at 4q35.2. All eight clones with this integration site gave unique bands on Southern blot, indicating a different pattern of rearrangement

in each clone. The representative clone in Table 1 with this integration site is clone J.

In several clones, we identified two host-virus junctions within very close proximity. The available evidence in these cases was most consistent with a single integration event with local rearrangement. Firstly, clone J contained two sites of integration at 4q35.2, one of which showed a complex pattern of host and viral sequences in different orientations (Fig. 3). Importantly, identical breakpoints in host and viral DNA were seen at both integration sites. Secondly, clone B contained the 8q24.21 common integration site and a second junction 418 bp further upstream in 8q24.21. The banding pattern of clone B on Southern blotting was identical to that of other clones containing the 8q24.21 common integrant, suggesting that both host-virus junctions were related to a single integration event. Thirdly, clone H2 contained two host-viral junctions at 17q12 that were separated by 1,170 kb. We cannot exclude the possibility that two separate integration events occurred in this case, although the close proximity of the two junctions in a single clone is more likely to represent local rearrangement similar to that in clones J and B. In our further analysis, clone H2, as well as clones J and B, was regarded as containing a single integration event with local rearrangement.

The 24 different integration sites were spread relatively evenly throughout the genome (Supplementary Fig. S2), although two clusters were apparent. The first represented a 600-bp region of 8q24.21, which encompassed each single integration site detected in clones I2, F3, and I3 (which carried the same integration site as F3), as well as the 8q24.21 common integration site (for which we observed evidence of local rearrangement in clone B; Fig. 4*A*). As clones I2 and F3/I3 contained discrete single integration sites,

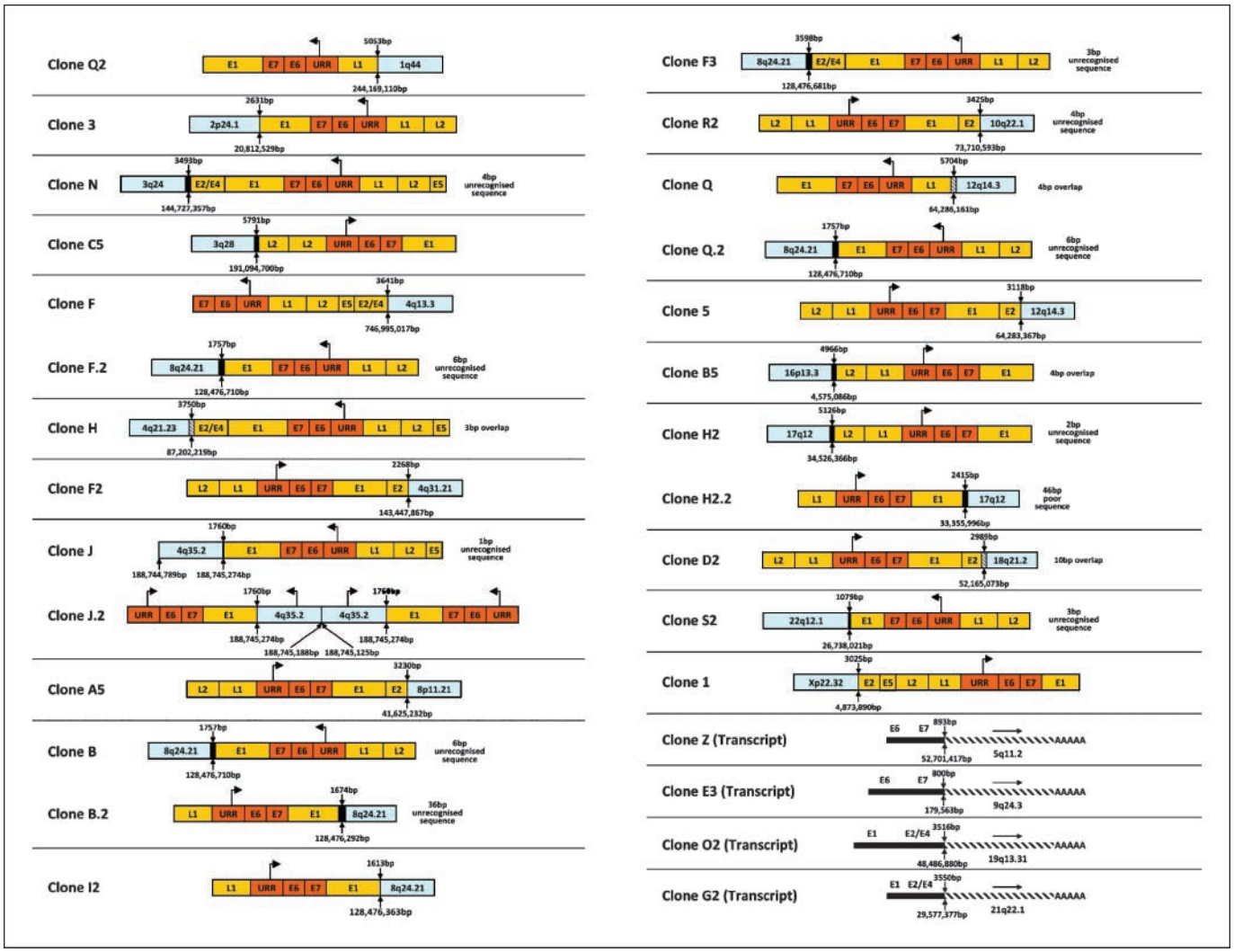

Figure 3. Schematics showing host-viral junctions at the different integration sites. In all schematics, host chromosomal DNA is shown in blue, orientated with the forward strand running from left to right. Integrated HPV16 DNA is shown in yellow, with the viral oncogenes and URR highlighted in orange, and the direction of transcription from the viral promoter shown by an arrow from the URR. The location of the viral breakpoint in base pairs is given above the junction, whereas the cellular DNA breakpoint in base pairs is given below the junction. Sequence overlap or unrecognized sequence at the host-viral junction is indicated where appropriate. Where two junctions were identified in a clone, the second is referred to using the suffix ".2". Note that for each integration site, only one host-virus junction was determined by RS-PCR, and although other viral genes not involved in the junction are shown in the schematics, they may have been partially or completely deleted. The last four schematics show viral-host fusion transcripts that were isolated by APOT and used to identify the integration sites in clones Z, E3, O2, and G2. Each virus-host fusion transcript is depicted in the direction of transcription from the viral promoter, i.e., 5′ to 3′ left to right, as depicted by the arrow. The viral and host portions of each transcript are shown by a solid black line and striped line, respectively. Note that the size of the fusion transcripts was not determined, and all are depicted at the same arbitrary length.

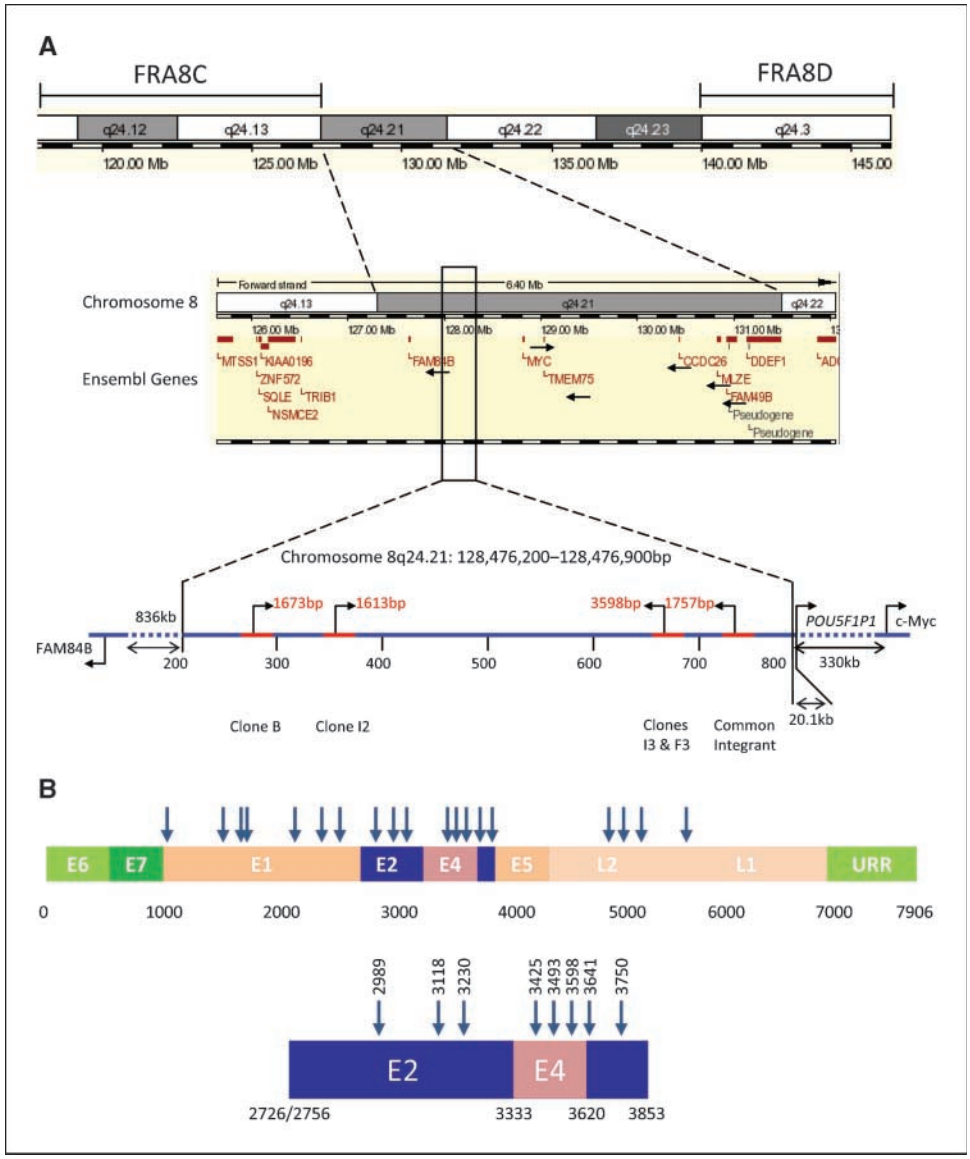

Figure 4. Detailed mapping of the 8q24.21 integrants and HPV16 breakpoints. *A,* the integration site at 8q24.21 is flanked by two CFSs (FRA8C and FRA8D; *top*). The Ensembl view of the region is shown in the middle panel, whereas the schematic in the bottom panel depicts the 600-bp region containing the cluster of integration sites. In this, the blue line represents cellular DNA and the red lines HPV16 DNA (not to scale), with the orientation of the viral promoter shown by the black arrows. The red numbers indicate viral breakpoints. Note that *POU5F1P1*, which maps to 8q24.21, is not shown by Ensembl. Clone B contained the common integrant, as well as the specific junction also shown. *B, top,* schematic representation of the linearized HPV16 genome, with the viral base pairs indicated below. Locations of the viral breakpoints in the 20 integrants mapped at the DNA level are indicated by the vertical blue arrows. *Bottom,* a detailed view of the E2/E4 region, with the base pair location for each viral breakpoint shown above the blue arrow.

with distinct virus and host breakpoints, we considered it unlikely that such sites were derivatives of the 8q24.21 common integrant. The second integrant cluster was at 12q14.3, where there were two integration events 3 kb apart. As for the 8q24.21 cluster, each 12q14.3 integrant was seen in isolation in an individual clone (clones Q and 5). Moreover, the viral genome was integrated in opposite orientations in the two clones with different viral breakpoints. These observations again suggested that the 12q14.3 junctions most likely represented different integration events within the same genomic region. We conclude that 8q24.21, and perhaps 12q14.3, may represent particularly susceptible sites for HR-HPV integration.

Clones F and Q had two integration sites in different chromosomes, as has been reported in a minority of cervical carcinoma samples (22–24). Several chromosome bands or sub-bands containing integration sites corresponded to those reported in clinical carcinoma samples (2p24.1, 3q28, 4q13.3, 4q31.21, 8q24.21, 10q22, and 17q12; refs. 10, 22, 25), which is of interest, as there are relatively few examples of recurrent sites of HPV16 integration in the literature.

Association of Integration Sites with CFSs and Other Regions of DSBs

Of the 24 different unselected integrants identified, 12 (50%) occurred within a chromosomal band containing a CFS, based on the 87 CFS listed by National Center for Biotechnology Information.[3] However, integration was not seen in bands containing FRA3B and FRA16D, the two most commonly expressed CFSs (26). Moreover, 9 of the 12 integrants were actually in different chromosomal sub-bands to the respective CFS, at a distance of 335 kb to 5.7 Mb from the sub-band reported to contain the CFS (Table 1). Only one integration event, in clone R2, occurred in the same sub-band as a CFS; at 10q22.1, although it should be noted that this is a large sub-band of 5.8 Mb. In another two clones, clones O2 and Q2, the relevant CFS was only mapped at the chromosomal band level, so there was insufficient information to determine the proximity of colocalization of CFS and integration

[3] http://www.ncbi.nlm.nih.gov/

site. In clone O2, the integrant was at 19q13.31, although the CFS is mapped only to 19q13, which covers 24.7 Mb. In clone Q2, HPV16 was integrated in 1q44, which is 5.5 Mb in length and not divided into sub-bands. FRA1I is also mapped to 1q44, but no further location information was available, and FRA1I may not cover the full 5.5 Mb.

We found no association between unselected HPV16 integration sites and other regions prone to forming DSBs, namely AT-rich regions and matrix attachment regions (see Supplementary Information and Supplementary Fig. S3).

Viral Breakpoints and Host-Viral Junctional Sequence

The HPV16 breakpoints determined from the RS-PCR and APOT sequences were all outside the E6 and E7 open reading frames (ORF) and the upstream regulatory region (URR; Fig. 4*B*). Of the 20 different integration sites where viral breakpoints were resolved at the DNA level (i.e., rather than by APOT alone), eight occurred within the E2/E4 ORFs, as is commonly seen in carcinomas *in vivo* (23), whereas a further eight occurred within E1, consistent with deletion of the downstream E2 gene. In the remaining four integrants, the breakpoints identified were within the L1/L2 region, most likely representing the upstream host-viral junctions, and it was not possible to determine whether or not E2 was deleted.

The 24 different integrants could be divided into three groups of approximately equal size, according to the sequence at the host-virus junction. In the first group (n = 8), the host sequence directly adjoined the viral sequence with no homology, suggesting a role for nonhomologous end joining (NHEJ) in the integration process (Table 2). In the second group (n = 7), regions of microhomology (1–10 bp) were seen between the host and viral sequences, suggesting microhomology-mediated NHEJ (6, 27, 28), whereas in the third group (n = 9), stretches of unrecognized "orphan" DNA (1–36 bp) were present between the host and viral sequences.

Significance of Host Genes at or near Integration Sites

Using the Ensembl genome browser, we determined the genomic distance between each integration site and the nearest known cancer-associated host gene[4] and whether or not the host gene was orientated in the direction of transcription from the viral promoter (Table 1). We also determined the nearest host gene [coding gene, pseudogene, or microRNAs (miR)] to each integration site in the direction of transcription using the Ensembl BIOMART function[5] and the miR-base website.[6] Twenty of the 24 different integration events (83%) occurred within the same chromosomal band as a cancer-associated gene, which included *C-MYC, TP73L(TP63), MAPK10, MYCN, RASSF6*, and *MDM2* (Table 1). However, the cancer-associated genes were generally a considerable distance from the integration site (median, 2,379 kb; range, 0–12,343 kb) and often in the opposite orientation to that of transcription from the viral promoter. For example, whereas integrants were seen in chromosome bands containing *C-MYC* and *MYCN* (genes previously implicated in HPV-related insertional mutagenesis; refs. 12, 29), the integration cluster at 8q24.21 was 330 kb from *C-MYC* and the integrant in clone 3 was 4,814 kb from *MYCN*. On the other hand, some integrants were present within intronic sequences—in clone H (*MAPK10*), clone N (*SLC9A9*), clone F (*RASSF6*), clone C5 (*TP73L*), clone S2 (*TTC28*), and clone H2 (*PLXDC1*).

[4] http://www.sanger.ac.uk/genetics/CGP/Census/
[5] http://www.ensembl.org/
[6] http://www.microrna.sanger.ac.uk/sequences/

Of the 24 different integration events, 14 were in the same chromosomal band as an miR gene. Five integrants were within 2.5 Mb of the miR gene (30), with a median separation for these five integrants of 1,205 kb (range, 11.9–2,347 kb; Table 1 and Supplementary Table S2).

Expression of Candidate Host Genes

To investigate further the possibility of insertional mutagenesis associated with HPV16, expression levels of eight candidate coding genes and one pseudogene were investigated in clones with HPV16 integrated in the same chromosomal band as the candidate gene, relative to clones with integration sites at other chromosomal loci and normal ectocervical keratinocytes. It was not possible to compare expression levels with those at the stage when the integrants coexisted with episomes (i.e., before growth in feeder-free conditions), as inadequate amounts of RNA were available from this point in the cloning process. No evidence for increased expression of c-Myc was found at the RNA or protein level in any of the clones with integration at 8q24.21 nor in the 8q24.21 selected integrant from nonclonal culture of W12 Series 2 (Supplementary Fig. S4). These clones also showed no evidence of overexpression of other candidate genes at 8q24: *FAM84B, Argonaute 2*, and the cancer-associated pseudogene *POU5F1P1/Oct4* (ref. 31; Supplementary Fig. S5). In addition, there was no evidence of altered expression of candidate genes at other integration sites, namely *MYCN, MDM2, MAPK10, SLC9A9,* and *RASSF6* (Table 1 and Supplementary Fig. S5), despite the last three showing HPV16 integration within introns.

Discussion

The HPV16 integration sites investigated in this study can be considered to represent natural events, because they occurred in naturally infected cervical keratinocytes containing viral episomes at copy numbers seen in basal squamous cells in productive infections *in vivo* (32). The viral oncogenes E6 and E7 would therefore be expected to be present at physiologic levels in episome-containing cells. In this regard, the W12 system compares favorably with previous studies that used cells in which DSBs were induced or were present at a high spontaneous level (7). Furthermore, HR-HPV integration in cells containing episomes expressing the inhibitory E2 protein represents a noncompetitive environment, from which integration events that would be out-competed in a mixed population can be isolated and studied.

Our data suggest that many of the integrants identified represented *de novo* events, indicating that integration can occur at any time during the period of HR-HPV episome maintenance when there is no quantitative deregulation of viral oncogene expression. The identification of clones that contained both the 8q24.21 common integrant (which was present within the starting population), and another integrant showed that, providing episomes are maintained, integration events can occur more than once within a single cell. Persistent infections with HR-HPV episomes *in vivo* will therefore increase the probability of integration events, which would initially be latent and may or may not be selected if episome loss is subsequently induced. These data provide biological support to the epidemiologic observation that persistent infection with HR-HPV is a major risk factor for cervical neoplastic progression (33).

Our evidence that 50% of the naturally occurring HPV16 integration events isolated from W12 were in the same chromosomal

Table 2. Host-viral junctional sequences in the HPV16 integrants isolated

Clone	Sequence at junction*	Overlap/unrecognized sequence
Clone Q2	AATACTATTATGTGAAGAA	None
Clone 3	TTTACATTTCCTTGGGAGG	None
Clone N	AGACAATTATGTCATTATT	4 bp unrecognized
Clone C5	TGAAAACAATATACAATTA	None
Clone F (4q13 junction)	CATTTTAAAGTATTCAAAC	9 bp overlap
Clone H	AAAAGTGCAATCAGTAGTC	3 bp overlap
Clone F2	TACCTAAAAAAAATTGCAT	None
Clone J	ACTATTATGAGAGGTCTCT	1 bp unrecognized
Clone A5	GACAGAGAGAAATATAAA	None
Clone B (Common 8q24 Integrant)	ACTATTTACATTATATTTA	6 bp unrecognized
Clone I2	GTCTTGCTCCTGTCAGCTA	1 bp overlap
Clone F3	AGCCCAGTCACACAGTTAA	3 bp unrecognized
Clone R2	GAGACTCTGAATATGGGTC	4 bp unrecognized
Clone Q (12q14.3 junction)	CATCCGTGCTTCCAAGGTC	4 bp overlap
Clone 5	AGTGCAGTTCCCTTCCTGC	None
Clone B5	GGTGTTGC---CTATTGCTG	11 bp unrecognized
Clone H2 (junction 1)	TTAACCTCAGGGTTT	2 bp unrecognized
Clone D2	CATACGGTGATGTGTCGTT	10 bp overlap
Clone S2	GTTGTTTACCCTCCAGCCT	3 bp unrecognized
Clone 1	TACAAACTGGACCTTTCCT	None
Clone Z (transcript)	TGATCCTGCAGGGTTGTTC	4 bp overlap
Clone E3 (transcript)	TGGACAGTGCTCCAAGCCT	7 bp overlap
Clone O2 (transcript)	CCACACCACACCCATGCAAC	1 bp unrecognized
Clone G2 (transcript)	ACAGTGCTACATTATTAAG	None

*The sequences were determined at the DNA level by RS-PCR, except in the case of clones Z, E3, O2, and G2 where the host-viral junction of the fusion transcript is given, as determined by APOT. HPV16 sequence is in red, host gDNA sequence in blue, host sequence from fusion transcripts in purple, overlapping sequence in green underlined, and unrecognized sequence in black underlined.

band as an CFS is very similar to observations made for HR-HPV in integrants selected *in vivo* (5, 7). This suggests that HR-HPV integration occurs close to CFSs because these regions are especially permissive, rather than conferring a selective advantage on the host cell. However, further questions arise from our findings. Firstly, although the unselected integration sites in W12 are associated with CFSs, they do not accurately colocalize with them, based on available data. This may reflect the fact that CFSs are mapped at a low resolution. Of the few CFS that have been mapped closely, many are under 1 Mb in length (34–40), whereas chromosomal bands can cover 30 Mb and sub-bands over 5 Mb. To add further complexity, CFSs recently mapped at high resolution were found to be larger than previously thought and to extend beyond the chromosomal bands to which they were initially mapped (22, 26, 41, 42). Secondly, we saw no colocalization of HPV16 integrants with FRA3B and FRA16D, the most commonly expressed CFSs. Although individuals are thought to express only 7 to 20 CFSs after CFS induction *in vitro*, FRA3B and FRA16D seem to be expressed in all individuals (26) and would be expected to be available for integration in W12.

The variability of the host-viral junctional sequences that we observed suggests that more than one mechanism of HPV integration may occur and that regions of microhomology or even orphan DNA sequences may be required to ligate the host and viral sequences in some cases. Interestingly, while a similar analysis performed on 16 selected integrants in HPV16-associated and HPV18-associated anogenital neoplasia found the same three types of junction (43), the relative frequencies differed, with 2 incidences of clean joins, 2 with orphan DNA (4–9 bp), and 12 with regions of microhomology (1–6 bp). Together, these data raise the possibility that microhomology-mediated integration events may confer a selective advantage in mixed populations. The viral breakpoints that we observed were all outside E6, E7, and the URR, supporting the view that retention of viral oncogene expression is a prerequisite for growth of HPV16-containing cells. Assuming that HPV16 episomal DNA is opened at random during the process of integration, the E6/E7/URR region would be predicted to be disrupted in some integration events. However, it is unlikely that such integrants would have been detected using our cloning strategy, as they would not have shown viral oncogene expression after episome clearance and would therefore have been incapable of sustained growth.

Our findings indicate that local rearrangements occur frequently after natural HPV16 integration. They also seem to occur early, as we analyzed integration sites after only ~25 population doublings in feeder-free conditions. Local rearrangement has been described in previous cell line studies (29, 44, 45), but it is likely to have been underreported, particularly in clinical samples, as PCR-based methods of integration site analysis would not identify it. Interestingly, it has recently been shown that latent integrants may be able to induce genomic instability in the locality of the integration site, despite no quantitative deregulation of E6 and E7, as expression of E1 and E2 proteins in cells containing integrated HR-HPV origins can induce genomic rearrangements of flanking

cellular sequences (45). The presence of a CFS near an integrant may also contribute to local rearrangement, although this seems not to be essential, as 4q35.2 is not associated with a CFS.

We found no evidence to suggest that HPV16-associated insertional mutagenesis occurs frequently. Expression of strong candidate genes (including *C-MYC* and *MYCN*) was not affected by integration in the same chromosomal band, including integration within introns (*MAPK10, SLC9A9, RASSF6*) or a relatively modest 20.1 kb upstream in the same orientation (*POU5F1P1*). Our conclusions are supported by a paucity of published data showing altered levels of host gene expression related to HR-HPV integration. Despite evidence for preferential integration of HPV18 within 8q24 in selected cervical carcinoma cells (46), there is no convincing functional evidence that *C-MYC* transcription is increased in this setting (4) nor in cervical carcinoma samples compared with normal cervix (47). An alternative explanation may be required to explain the observation that HR-HPV integration occurs close to a range of cancer-associated genes. One possibility is that these are simply transcriptionally active and therefore open regions, which are more prone to DSBs. This would be consistent with the overall notion that HR-HPV integration occurs within relatively accessible regions of the genome.

It may nevertheless be the case that HPV16 integration affects transcription of alternative genes to the candidates chosen in the present study. The genes closest to the integrated HPV16 genomes in the direction of viral transcription included pseudogenes and genes of unknown function (Table 1). Furthermore, it is feasible that insertional mutagenesis could involve looping of DNA, affecting transcription of genes that are not close to the integration site in base-pair distance but are brought into close proximity by DNA folding. However, in the absence of a detailed understanding of nuclear DNA organization and adequate functional investigations of HR-HPV–associated insertional mutagenesis, it is not possible to predict the likelihood or extent of these effects.

We conclude that the profile of naturally occurring integrants that can be isolated under noncompetitive conditions from cervical keratinocytes infected with HPV16 bears close similarity to the integrants that are selected in cervical SCC *in vivo*. The integration sites observed seem to be particularly suitable for the process of integration, potentially through an increased density and/or likelihood of DSBs. We observed a frequent association between integrants and genomic regions containing CFSs, although detailed mapping did not indicate colocalization. The panel of clones that we have generated will provide a valuable isogenic system for future experiments examining virus and host factors that confer a selective advantage in a mixed population.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 5/8/2008; revised 7/11/2008; accepted 7/14/2008.

Grant support: Medical Research Council and Cancer Research UK program grants.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank David I. Smith for helpful discussions and for advice on the RS-PCR technique and Elizabeth Gray and Nicolas Wentzensen for advice and assistance with the APOT technique.

References

1. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12–9.
2. Green J, Berrington de Gonzalez A, Sweetland S, et al. Risk factors for adenocarcinoma and squamous cell carcinoma of the cervix in women aged 20-44 years: the UK National Case-Control Study of Cervical Cancer. Br J Cancer 2003;89:2078–86.
3. Melsheimer P, Vinokurova S, Wentzensen N, Bastert G, von Knebel Doeberitz M. DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. Clin Cancer Res 2004;10:3059–63.
4. Pett M, Coleman N. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? J Pathol 2007;212:356–67.
5. Wentzensen N, Vinokurova S, von Knebel Doeberitz M. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res 2004;64:3878–84.
6. Winder DM, Pett MR, Foster N, et al. An increase in DNA double-strand breaks, induced by Ku70 depletion, is associated with human papillomavirus 16 episome loss and *de novo* viral integration events. J Pathol 2007;213:27–34.
7. Yu T, Ferber MJ, Cheung TH, Chung TK, Wong YF, Smith DI. The role of viral integration in the development of cervical cancer. Cancer Genet Cytogenet 2005;158:27–34.
8. Durkin SG, Glover TW. Chromosome fragile sites. Annu Rev Genet 2007;41:169–92.
9. Popescu NC. Genetic alterations in cancer as a result of breakage at fragile sites. Cancer Lett 2003;192:1–17.
10. Kraus I, Driesch C, Vinokurova S, et al. The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. Cancer Res 2008;68:2514–22.
11. Thorland EC, Myers SL, Persing DH, et al. Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. Cancer Res 2000;60:5916–21.
12. Peter M, Rosty C, Couturier J, Radvanyi F, Teshima H, Sastre-Garau X. MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. Oncogene 2006;25:5985–93.
13. von Knebel Doeberitz M, Bauknecht T, Bartsch D, zur Hausen H. Influence of chromosomal integration on glucocorticoid-regulated transcription of growth-stimulating papillomavirus genes E6 and E7 in cervical carcinoma cells. Proc Natl Acad Sci U S A 1991;88:1411–5.
14. Stanley MA, Browne HM, Appleby M, Minson AC. Properties of a non-tumorigenic human cervical keratinocyte cell line. Int J Cancer 1989;43:672–6.
15. Pett MR, Alazawi WO, Roberts I, et al. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. Cancer Res 2004;64:1359–68.
16. Pett MR, Herdman MT, Palmer RD, et al. Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. Proc Natl Acad Sci U S A 2006;103:3822–7.
17. Herdman MT, Pett MR, Roberts I, et al. Interferon-β treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants. Carcinogenesis 2006;27:2341–53.
18. Alazawi W, Pett M, Arch B, et al. Changes in cervical keratinocyte gene expression associated with integration of human papillomavirus 16. Cancer Res 2002;62:6959–65.
19. Coleman N, Greenfield IM, Hare J, Kruger-Gray H, Chain BM, Stanley MA. Characterization and functional analysis of the expression of intercellular adhesion molecule-1 in human papillomavirus-related disease of cervical keratinocytes. Am J Pathol 1993;143:355–67.
20. Sarkar G, Turner RT, Bolander ME. Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. PCR Methods Appl 1993;2:318–22.
21. Klaes R, Woerner SM, Ridder R, et al. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res 1999;59:6132–6.
22. Thorland EC, Myers SL, Gostout BS, Smith DI. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. Oncogene 2003;22:1225–37.
23. Choo KB, Pan CC, Han SH. Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. Virology 1987;161:259–61.
24. Luft F, Klaes R, Nees M, et al. Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. Int J Cancer 2001;92:9–17.
25. Wentzensen N, Ridder R, Klaes R, Vinokurova S, Schaefer U, Doeberitz MK. Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions. Oncogene 2002;21:419–26.
26. Denison SR, Simper RK, Greenbaum IF. How common are common fragile sites in humans: interindividual variation in the distribution of aphidicolin-induced fragile sites. Cytogenet Genome Res 2003;101:8–16.

27. Katsura Y, Sasaki S, Sato M, et al. Involvement of Ku80 in microhomology-mediated end joining for DNA double-strand breaks *in vivo*. DNA Repair Amst 2007;6: 639–48.

28. Kohno T, Yokota J. Molecular processes of chromosome 9p21 deletions causing inactivation of the p16 tumor suppressor gene in human cancer: deduction from structural analysis of breakpoints for deletions. DNA Repair Amst 2006;5:1273–81.

29. Couturier J, Sastre-Garau X, Schneider-Maunoury S, Labib A, Orth G. Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression. J Virol 1991; 65:4534–8.

30. Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A 2004;101:2999–3004.

31. Suo G, Han J, Wang X, et al. Oct4 pseudogenes are transcribed in cancers. Biochem Biophys Res Commun 2005;337:1047–51.

32. Stubenrauch F, Laimins LA. Human papillomavirus life cycle: active and latent phases. Semin Cancer Biol 1999;9:379–86.

33. Remmink AJ, Walboomers JM, Helmerhorst TJ, et al. The presence of persistent high-risk HPV genotypes in dysplastic cervical lesions is associated with progressive disease: natural history up to 36 months. Int J Cancer 1995;61:306–11.

34. Helmrich A, Stout-Weider K, Matthaei A, Hermann K, Heiden T, Schrock E. Identification of the human/mouse syntenic common fragile site FRA7K/Fra12C1-relation of FRA7K and other human common fragile sites on chromosome 7 to evolutionary breakpoints. Int J Cancer 2007;120:48–54.

35. Mishmar D, Rahat A, Scherer SW, et al. Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. Proc Natl Acad Sci U S A 1998;95: 8141–6.

36. Hormozian F, Schmitt JG, Sagulenko E, Schwab M, Savelyeva L. FRA1E common fragile site breaks map within a 370 kilobase pair region and disrupt the dihydropyrimidine dehydrogenase gene (DPYD). Cancer Lett 2007;246:82–91.

37. Huang H, Qian J, Proffit J, Wilber K, Jenkins R, Smith DI. FRA7G extends over a broad region: coincidence of human endogenous retroviral sequences (HERV-H) and small polydispersed circular DNAs (spcDNA) and fragile sites. Oncogene 1998;16:2311–9.

38. Savelyeva L, Sagulenko E, Schmitt JG, Schwab M. The neurobeachin gene spans the common fragile site FRA13A. Hum Genet 2006;118:551–8.

39. Arlt MF, Miller DE, Beer DG, Glover TW. Molecular characterization of FRAXB and comparative common fragile site instability in cancer cells. Genes Chromosomes Cancer 2002;33:82–92.

40. Sawinska M, Schmitt JG, Sagulenko E, Westermann F, Schwab M, Savelyeva L. Novel aphidicolin-inducible common fragile site FRA9G maps to 9p22.2, within the C9orf39 gene. Genes Chromosomes Cancer 2007;46: 991–9.

41. Curatolo A, Limongi ZM, Pelliccia F, Rocchi A. Molecular characterization of the human common fragile site FRA1H. Genes Chromosomes Cancer 2007; 46:487–93.

42. Becker NA, Thorland EC, Denison SR, Phillips LA, Smith DI. Evidence that instability within the FRA3B region extends four megabases. Oncogene 2002;21: 8713–22.

43. Ziegert C, Wentzensen N, Vinokurova S, et al. A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques. Oncogene 2003;22: 3977–84.

44. Herrick J, Conti C, Teissier S, et al. Genomic organization of amplified MYC genes suggests distinct mechanisms of amplification in tumorigenesis. Cancer Res 2005;65:1174–9.

45. Kadaja M, Sumerina A, Verst T, Ojarand M, Ustav E, Ustav M. Genomic instability of the host cell induced by the human papillomavirus replication machinery. Embo J 2007;26:2180–91.

46. Ferber MJ, Thorland EC, Brink AA, et al. Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. Oncogene 2003;22: 7233–42.

47. Soh LT, Heng D, Lee IW, Ho TH, Hui KM. The relevance of oncogenes as prognostic markers in cervical cancer. Int J Gynecol Cancer 2002;12:465–74.