Published in final edited form as:

Int J Cancer. 2016 November 01; 139(9): 2001–2011. doi:10.1002/ijc.30243.

Genomic characterization of viral integration sites in HPV-related cancers

Clara Bodelon^{1,*}, Michael E. Untereiner¹, Mitchell J. Machiela¹, Svetlana Vinokurova², Nicolas Wentzensen^{1,*}

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.

²Laboratory of Molecular Biology of viruses, NN Blokhin Russian Cancer Research Center, Moscow, Russia.

Abstract

Persistent infection with carcinogenic human papillomaviruses (HPV) is known to cause the majority of anogenital cancers and a subset of head and neck cancers. The HPV genome is frequently found integrated into the host genome of invasive cancers. The mechanisms of how it may promote disease progression are not well understood. Thoroughly characterizing integration events can provide insights into HPV carcinogenesis. Individual studies have reported limited number of integration sites in cell lines and human samples. We performed a systematic review of published integration sites in HPV-related cancers and conducted a pooled analysis to formally test for integration hotspots and genomic features enriched in integration events using data from the Encyclopedia of DNA Elements (ENCODE). Over 1,500 integration sites were reported in the literature, of which 90.8% (N=1,407) were in human tissues. We found 10 cytobands enriched for integration events, three previously reported ones (3q28, 8q24.21 and 13q22.1) and seven additional ones (2q22.3, 3p14.2, 8q24.22, 14q24.1, 17p11.1, 17q23.1 and 17q23.2). Cervical infections with HPV18 were more likely to have breakpoints in 8q24.21 ($P=7.68 \times 10^{-4}$) than those with HPV16. Overall, integration sites were more likely to be in gene regions than expected by chance ($P=1.36 \times 10^{-13}$). They were also significantly closer to CpG regions, fragile sites, transcriptionally active regions and enhancers. Few integration events occurred within 50Kb of known cervical cancer driver genes. This suggests that HPV integrates in accessible regions of the genome, preferentially genes and enhancers, which may affect the expression of target genes.

Keywords

HPV; integration

*Correspondence: Clara Bodelon, PhD, Division of Cancer Epidemiology and Genetics, 9609 Medical Center Dr., Rm 7E-236, Bethesda, MD 20892, Phone: 240-276-7327, Fax: 240-276-7838, clara.bodelon@nih.gov, Nicolas Wentzensen MD, PhD, MS, Division of Cancer Epidemiology and Genetics, 9609 Medical Center Dr., Rm 7E-E114, Bethesda, MD 20892, Phone: 240-276-7303, Fax: 240-276-7838, wentzenn@mail.nih.gov.

Disclaimers

The authors have no conflicts of interest.

Introduction

Persistent infection with carcinogenic human papillomaviruses (HPV) is known to cause the majority of anogenital cancers^{1–3} as well as some head and neck cancers⁴. During its natural life cycle, the HPV genome persists and replicates as a circular DNA molecule or episome. However, with the progression from pre-cancer to invasive cancers, the HPV genome is frequently found integrated into the host genome^{5–7}. The biological mechanisms leading to HPV integration remain elusive. It has been proposed that HPV integration is an important step in the progression to cancer. However, the mechanism by which integration may promote this progression is not well understood.

The HPV genome encodes oncoproteins E6 and E7 which target important cellular pathways, including inhibition of the p53 and pRB function^{8,9}. Viral E6 and E7 oncoproteins from episomal and integrated genomes are known to cause chromosomal instability by interfering with the host replication machinery¹⁰; it has been suggested that oncogene transcripts from integrated genomes may have longer half-lives compared to transcripts from episomes⁹. Some studies have reported integration sites near known tumor suppressor genes or oncogenes and suggested that HPV-integration may promote progression and invasion by disrupting the function of these genes¹¹. In addition, studies have reported that host genomes with HPV DNA integrated also have different expression and methylation profiles as compared to non-integrated genomes¹².

While HPV integration events have been observed to occur over the entire genome, some regions of the human genome have been repeatedly targeted by HPV integration^{13–15}. In particular, the regions surrounding the *MYC* oncogene, 3q28 and 13q22.2 have been suggested to be integration hotspots^{13, 14, 16}. Individual studies typically only include a small number of integration events. To have a better understanding of the locations and the role HPV integration in the host genome, it is important to conduct large pooled analyses of integration sites.

New technologies, such as next-generation sequencing approaches, have led to a recent increase in reporting of HPV integration sites^{12, 15, 17, 18}. As of 2004, 192 HPV integration sites had been described, many of them in cell lines¹³. In the last decade more than 1,500 integration sites have been reported, preferentially in tumor tissues. This provides an opportunity to conduct a well-powered analysis of HPV integration sites using recent efforts to describe the functional elements of the human genome, such as the Encyclopedia of DNA Elements (ENCODE)¹⁹. We analyzed genomic characteristics of the human host genome surrounding HPV integration sites in an effort to better understand common molecular features that may lead to HPV integration into the human genome and that may contribute to progression and invasion.

Results

Our analysis included integration sites from 33 articles out of 516 articles^{12–18, 20–45}, including a previous systematic review, as well as integration sites obtained from TCGA⁴⁶ or our own unpublished data (Table 1). In total, we collected information on 1,551

integration events across several cancer organ sites and cell lines (Supplemental Table 1). New technologies, such as next-generation sequencing (NGS), whole-genome sequencing (WGS) or RNA sequencing, were more likely to determine a greater number of integration sites per sample (about 2 per sample) compared to fluorescence *in situ* hybridization (FISH) or polymerase chain reaction (PCR)-based technologies (about 1.4 per sample), such as amplification of papillomavirus oncogene transcript (APOT) or detection of integrated papillomavirus sequencing (DIPS) (Supplemental Table 2).

The number of integration sites reported in human tissues was 1,407 (90.7%) and 143 (9.2%) in cell lines (Figure 1). Among the reported integration events in human specimens, 1,262 of 1,407 reported breakpoint events were from cervical tissues (81.4%) while 122 were from head and neck cancers (7.9%). The majority of HPV integration events were reported in invasive cancers (89.1%), a subset was described in pre-cancerous lesions (10.9%). Most of the reported HPV integration events involved HPV16 (74.1%), followed by HPV18 (15.0%) and HPV45 (3.5%) (Figure 1B). The human chromosomes with the most HPV integration events were chromosome 3 (8.4%), followed by chromosome 2 (8.0%), chromosome 8 (7.8%) and chromosome 1 (7.7%) (Figure 1C).

Integration events occurred across the genome, but three loci had a substantial accumulation of integration events: 3q28 with 31 integration events, 8q24.21 with 41 integration events and 13q22.1 with 38 integration events (Figure 2A). All of these regions are gene-rich regions hosting important oncogenic or tumor suppressor genes. Several genes were directly affected by integration events, including *TP63* a member of the p53 family, *TPRG1* a regulator of *TP63* and *IL1RAP* a pro-inflammatory activation gene in 3q28 (Figure 2B); the *MYC* oncogene in 8q24.21 (Figure 2C); and important regulators of gene expression during development and carcinogenesis (*KLF5* and *KLF12*) in 13q22.1 (Figure 2D). These clusters were observed in cervical tissues but not in head and neck cancers (Supplemental Figure 1). When comparing invasive cancers with precancers, we also observed that they clustered differently, with most integration events reported in precancers in chromosome 9 (Supplemental Figure 2). The pattern also differed by HPV type (Supplemental Figure 3). Cervical infections with HPV18 were more likely to have breakpoints in chromosome 8 (χ^2 test $P=2.72 \times 10^{-5}$), especially in 8q24.21 (χ^2 test $P=7.68 \times 10^{-4}$). This pattern was also observed when phylogenetic similar HPV types were combined.

We formally investigated which cytobands had significantly higher number of observed integration events compared to what would be expected based on the 1,407 integration events found in human tissues (Table 2). In addition to the cytobands 3q28, 8q24.21 and 13q22.1 previously reported, we also found seven new cytobands, 2q22.3, 3p14.2, 8q24.22, 14q24.1, 17p11.1, 17q23.1 and 17q23.2, that were significantly enriched for integration events.

Next, we investigated the relation between integration sites with exact know position and genes (Figure 3). For this analysis, integration sites determined with transcript-based assays that did not map to an exon or untranslated region (UTR) were excluded, leaving 639 integration events for the analysis. We found that integration sites tended to be located in genes ($P=6.79 \times 10^{-9}$; Figure 3A), and more specifically in 5'UTR and exons ($P=4.51 \times 10^{-5}$

and $P=6.94 \times 10^{-35}$, respectively). Because some breakpoints were detected with transcript-based assays, which may have biased to find integration events in genes, we stratified by assay type and still observed a similar pattern (Supplemental Figure 4). Intergenic breakpoints were significantly closer to genes than expected by chance ($P=0.0097$; Figure 3B). This pattern was also observed when we stratified integration events by HPV16 and HPV18 (Supplemental Figure 5). Nine of the 639 integration events occurred near reported significantly mutated genes (1%) compared to the 0.1% expected ($P=6.41 \times 10^{-7}$). The majority of these integration events were in chromosome 17 near *TP53*.

We explored the location of integration events with a number of genomic features (Table 3). Integration sites tended to be closer to CpG regions ($P=9.38 \times 10^{-10}$) and fragile sites than expected by chance ($P=2.05 \times 10^{-6}$). The relationship between integration events and repetitive elements was complex, being further from LINE elements, but closer to SINE-Alu and short tandem repeats. They were closer to open chromatin sites ($P=3.57 \times 10^{-19}$ and 6.65×10^{-30} for FAIRE and DnaseI, respectively) and to sites of initiation of DNA replication ($P=1.74 \times 10^{-10}$). Integration events were also closer to sites with histone makers than expected by chance, except for H3K27me3, which is a maker of repressed genes. Using the chromatin genome segmentation information from ENCODE, which combines open chromatin and histone data, we found that integration sites were closer to enhancers ($P=1.68 \times 10^{-33}$) and transcription start sites ($P=3.77 \times 10^{-19}$).

We investigated whether specific host DNA sequences were more likely to be observed at loci of integration, but a specific motif was not found (Supplemental Figure 6).

Finally, we explored which biological processes that were enriched by human genes affected integration events. Gene Ontology analysis identified epithelial cell related processes to be especially affected (Supplemental Table 3). Several genes were over-represented in these pathways, including *ELF3*, *LEF1*, *NOTCH1*, *TNF*, *TP63* and *WT1*.

Discussion

We conducted a comprehensive analysis of over 1,500 HPV integration events, focusing on events found in human tissues. Our analysis revealed seven new integration hotspots. In addition, integration breakpoints tended to occur in genomic regions with enrichment for genes, CpG sites, open chromatin and transcriptionally active regions. Based on histone markers, integration sites were more likely to be close to enhancers. We also saw that integration events tended to cluster differently by HPV type and found no evidence that a host cellular motif is present at HPV integration sites.

An important goal when analyzing HPV integration is to understand how it affects carcinogenesis. Studying the location in the host genome where HPV integration is more likely to occur can provide insights into affected pathways. Several studies have reported putative HPV integration hotspots based on relatively small numbers^{13–15}. However, whether a certain locus is a hotspot has previously not been tested formally. Our analysis identified several integration hotspots, including previously suggested ones (3q28, 8q24.21 and 13q22.1)^{14, 15} as well as new ones: 2q22.3, 3p14.2, 8q24.22, 14q24.1, 17p11.1, 17q23.1

and 17q23.2. Integration events in these loci were reported in several independent papers, discovered in different laboratories and different technologies, which indicates robustness of these findings. Many of the new loci were identified using new technologies, predominantly using RNA-based technologies, which suggest a relationship between recurrent integration events and affected gene expression.

A recent study reported putative hotspots in *HMGA2*, *DLG2* and *SEMA3D* genes that were not confirmed in our analysis (14). There may be several reasons for this. First, the previous study did not account for interdependence of integration events. This study suggested that the viral genome can be frequently fragmented and several of these fragments integrated in close proximity of each other within a sample, which can be misinterpreted as multiple independent events¹⁵. This is frequently observed when integration events are determined using NGS techniques. It is important to note that an integration hotspot must be at a site where re-occurring events are found in different samples, and most importantly across studies and laboratories. These should be taken into account to reduce biases of methods and sample contamination within a study. After taking the clustering into account, we only observed 2 events in 12q14.1 (where *HMGA2* is located) and one event in 11q14.1 (where *DLG2* is located). There were none in 7q21.11 (where *SEMA3D* is located). The reason for no finding integration events in 7q21.11 is because it had very low number of supporting reads and we only included integration events from Hu *et al.* with 10 or more supporting reads (see methods). Therefore, *HMGA2*, *DLG2* and *SEMA3D* are unlikely to be recurrent HPV integration loci.

Different assays have been used to detect integration events, all of them with advantages and limitations. DNA-based assays can, in theory, detect any viral-cellular junction. Transcripts based-assays can only detect integration events that have been transcribed. The type of assay is important as to where in the genome integration events can be found. In particular, transcript based assays would bias into finding integration events in exons of genes (although enhancers in intergenic and intronic regions can also be transcribed). Without a formal analysis taking the assay type into account, this could lead to the conclusion that integration events occur closer to genes than expected. We looked closely into this issue and observed that integration events were significantly enriched in genes, independently of the assay type. However, the distributions were slightly different with more integration events detected in exons from transcript-based assays, as expected. Although a few previous studies were not able to find such a significant association, this likely due to their small sample size^{12, 17}.

Some studies have recently investigated how the HPV genome may be integrated into the host genome^{12, 15, 17}. Amplification and rearrangements have been found to be located at or near integration sites^{12, 18, 45, 46}, near genomic events related to genomic instability¹⁷ and translocations¹². It has also been reported that some breakpoints occurred in regions where the viral and host genomes had similar short DNA sequences (micro-homologies) spanning 1–10 base pairs^{15, 17}, although further analysis may be required to clarify this mechanism⁴⁷. This may suggest that local genomic instability at the breakpoint may have led to the action of DNA repair pathways and, due to the micro-homology observation of the HPV and host sequences, HPV genomes might be integrated into the host genome. Given that

approximately 60% of breakpoints occur in nucleosome-depleted regions, our findings suggest that integration of breakpoints occur in regions where the host DNA is accessible. This may be during cell repair, cell division or transcription.

The mechanisms relating HPV integration to carcinogenesis are still not well understood. One potential mechanism for the HPV carcinogenesis is the stabilization of HPV oncogene expression^{18, 46}. Integration could also be related to the disruption of critical host genomic sequences^{13, 48}, given that integration events tend to occur near genes. It has also been observed that integration is related to greater expression of nearby genes compared to non-integrated tumor samples¹⁸. Another possibility is that integration events, or at least some, do not affect carcinogenesis. Our study had the power to thoroughly investigate the flanking genomic regions and we found significant associations with genes, fragile sites, CpG enrich regions, and open chromatin, improving our understanding of the human genomic features surrounding HPV integration sites. In particular, chromatin segmentation of the genome implicated enhancers to be significantly targeted by integration events. This suggests that integration may have an important regulatory function. This is in agreement with other studies that looked at histone makers in a subset of integration events^{49, 50}. Finally, we found little overlap between prevalent mutations reported in cervical cancer^{18, 46} and integration events, suggesting that these may be unrelated mechanism that alters the host genome in HPV-related carcinogenesis.

We observed that integration events differ by HPV type. It is known that cervical cancer can occur in the absence of HPV integration after infection with HPV16. For reasons that are not well understood, cancers infected with HPV18 tend to have a higher prevalence of integration⁴⁶. Our analysis stratifying by HPV type revealed that integration events in HPV18 samples were more likely to be in 8q24.21 where the *MYC* oncogene is located than HPV16 infected samples⁵¹. *MYC* codes for a protein with roles in cell cycle progression, apoptosis and cellular transformation. It is possible that integration at this locus disrupts the normal function of *MYC* and may provide growth advantage to those clones. However, the reasons underlying the observed difference by HPV type are not known. As mentioned before, HPV integration occurs near amplifications^{18, 45}. In particular, gains in 3q have been reported for HPV16^{45, 52} and to a lesser degree for HPV18⁵². This difference may explain the cluster of integration events in 3q for HPV16 but mostly absent for HPV18. Whether a difference in gains occurs in 8q is observed according to HPV type has not been investigated. If there were such difference for chromosome 8, with higher prevalence of gains for HPV18, it could potentially explain the integration locus observed for HPV18 in 8q.

We did not find specific clusters of integration events in head and neck clusters, which were mostly related to HPV16, but this may be due to the low number of events reported so far. There was a suggestive cluster in precancers in chromosome 9p, although the number of integration events in precancers was small. We did not observe a cluster in chromosome 9 among invasive cancers. The reasons for this are unclear. Future studies should further explore whether integration in 9p is an important event in carcinogenesis.

Our study has several strengths including the largest number of pooled, independent HPV integration events to date. In addition, when possible, we obtained DNA sequences to map the exact integration locations. These pooled data allowed us to have the power to formally test for new integration hotspots and to look at genomic features for which individual studies did not have the power to do. Also, our analyses focus on integration events in human samples. Many previous studies have based their analysis in cell lines, but this can introduce artifacts and not represent the natural history of cervical cancer. While individual studies used different criteria in order to report what genomic features are affected by integration events (e.g., genes closer than 50Kb to integration sites) potentially leading to different results, we comprehensively summarized genomic features across studies. Our study also has some limitations. We did not have subject level data to relate integration loci with other cancer-specific genomic features and relied on databases, and for several features on the HeLA S3 clone, to characterize the host genomic regions. Future studies should measure integration, gene expression and copy number variation directly in the same samples, which will be important to confirm our findings.

In summary, our data found that the HPV genome integrates into several hotspots and may do so differently by HPV type. Several genomic regions enriched by integration events were also enriched for genes, open chromatin, actively transcribed regions and enhancers. This suggests that accessible regions of the genome can be preferentially targeted by HPV integration and may lead to functional alteration of gene expression in the development of HPV-related cancers.

Materials and methods

A systematic review of the literature was performed to identify studies in the NCBI PubMed database that reported HPV integration loci as of June 30, 2015 in the English language. Our searches included the following keywords: (HPV OR human papillomavirus) AND (integration) AND (locus OR loci OR chromosome OR genome OR host genome) AND (cancer OR cancers OR carcinoma OR carcinomas OR precancer). Review of titles, abstract and full-text was done by CB and MU. Since integration sites published prior to 2004 were reported in the systematic review by Wentzensen *et al.*¹³, we limited our search to papers published after 1/1/2004 and combined the new integration sites with those from the previously review for this analysis. Papers included in our systematic review were required to report at least a cytoband where the integration occurred. When multiple integration events were reported in the same sample and chromosome, only one event per 500Kb was included in the analysis so that only independent integration events were more likely to be included⁴⁶. The same criterion was applied to cell lines across multiple papers, so that only one integration event per 500Kb per cell line was included in our analysis. A recent paper reported integration breakpoints from a new next-generation sequencing assay using a questionable low threshold regarding number of supporting reads¹⁵. To assure the inclusion of robust breakpoint events, we restricted to events with the number of supporting reads being 10 or greater.

When DNA sequences were available, the genomic integration sites were determined using the Basic Local Alignment Search Tool (BLAST, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) for

the GRCh37/hg19 build. If only the cytoband where the integration site occurred was reported, we approximated the integration site as the midpoint in the cytoband.

Statistical analysis

Integration sites were summarized according to cancer site, HPV type and HPV integration locations in the genome. χ^2 statistical tests were computed to assess statistical significance differences between groups.

To determine integration hotspots, we randomly generated 1,000,000 integration points across the genome (excluding from occurring in telomeres, centromeres and other large constitutive heterochromatin regions) and compared observed integration events to the expected number of events at each cytoband (cytoband information was downloaded from ENCODE: <ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/cytoband.txt.gz>). To test for significant difference p-values were computed using Fisher's exact test. The Bonferroni correction was used to account for multiple tests.

To characterize detailed genomic features surrounding integration sites, we restricted our analysis to integration sites for which we had the exact genomic coordinates. We downloaded data on genomic feature locations from the ENCODE Project (<ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/>) and information on fragile sites was obtained from Lukusa and Fryns⁵³. For each genomic feature, the distribution of minimal distances from the observed integration sites to the genomic feature was calculated and compared to the expected distribution of minimal distance from 10,000 randomly generated integration sites across the genome excluding telomeres, centromeres and other large constitutive heterochromatin regions. These distributions were compared using the Wilcoxon rank-sum test. Nineteen significant mutated genes in cervical cancer have been reported in the literature^{18, 46} and include the following genes: *ARID1A*, *ELF3*, *NFE2L2*, *CASP8*, *TGFBR2*, *PIK3CA*, *FBXW7*, *HLA.A*, *HLA.B*, *PTEN*, *KRAS*, *ERBB3*, *CBFB*, *TP53*, *MED1*, *ERBB2*, *STK11*, *MAPK1*, *EP300*. The number of integration sites within a 50Kb window around any of these genes was compared to the number of integration sites lying outside of the 50Kb window.

We also explored whether there was a host-DNA motif related to integration. For each integration event for which we had an exact location, we mapped the DNA sequence from the hg19 build at the integration site and 5 nucleotides before and after the integration site to identify DNA sequences around the integration events. The probability of each nucleotide at each location was plotted.

Finally, ontology enrichment was performed using the Gene Ontology (GO) project⁵⁴ to find pathways affected by integration events, using the *topGO* Bioconductor package. Enrichment of genes within a 50Kb window around an integration event was compared with all genes in the RefSeq database. Fisher exact test p-values were obtained to determine enriched GO terms.

All statistical tests were two-sided and computed at the 5% level. Statistical analyses were performed using R software (version 3.1.1).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This research was supported by the Intramural Research Program of the National Cancer Institute.

References

1. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet* 2007;370: 890–907. [PubMed: 17826171]
2. Madsen BS, Jensen HL, van den Brule AJ, Wohlfahrt J, Frisch M. Risk factors for invasive squamous cell carcinoma of the vulva and vagina--population-based case-control study in Denmark. *Int J Cancer* 2008;122: 2827–34. [PubMed: 18348142]
3. Rubin MA, Kleter B, Zhou M, Ayala G, Cubilla AL, Quint WG, Pirog EC. Detection and typing of human papillomavirus DNA in penile carcinoma: evidence for multiple independent pathways of penile carcinogenesis. *The American journal of pathology* 2001;159: 1211–8. [PubMed: 11583947]
4. Goon PK, Stanley MA, Ebmeyer J, Steinstrasser L, Upile T, Jerjes W, Bernal-Sprekelsen M, Gorner M, Sudhoff HH. HPV & head and neck cancer: a descriptive update. *Head & neck oncology* 2009;1: 36. [PubMed: 19828033]
5. Klaes R, Woerner SM, Ridder R, Wentzensen N, Duerst M, Schneider A, Lotz B, Melsheimer P, von Knebel Doeberitz M. 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. [PubMed: 10626803]
6. Wentzensen N, Ridder R, Klaes R, Vinokurova S, Schaefer U, Doeberitz M. Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions. *Oncogene* 2002;21: 419–26. [PubMed: 11821954]
7. Vinokurova S, Wentzensen N, Kraus I, Klaes R, Driesch C, Melsheimer P, Kisseljov F, Durst M, Schneider A, von Knebel Doeberitz M. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Res* 2008;68: 307–13. [PubMed: 18172324]
8. Mantovani F, Banks L. The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene* 2001;20: 7874–87. [PubMed: 11753670]
9. Munger K, Basile JR, Duensing S, Eichten A, Gonzalez SL, Grace M, Zacny VL. Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. *Oncogene* 2001;20: 7888–98. [PubMed: 11753671]
10. Duensing A, Spardy N, Chatterjee P, Zheng L, Parry J, Cuevas R, Korzeniewski N, Duensing S. Centrosome overduplication, chromosomal instability, and human papillomavirus oncoproteins. *Environ Mol Mutagen* 2009;50: 741–7. [PubMed: 19326465]
11. Ziegert C, Wentzensen N, Vinokurova S, Kisseljov F, Einkenkel J, Hoeckel M, von Knebel Doeberitz M. A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques. *Oncogene* 2003;22: 3977–84. [PubMed: 12813471]
12. Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, Lee S, Hadjipanayis AG, Ivanova EV, Wilkerson MD, Protopopov A, Yang L, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A* 2014;111: 15544–9. [PubMed: 25313082]
13. 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. [PubMed: 15172997]

14. Xu B, Chotewutmontri S, Wolf S, Klos U, Schmitz M, Durst M, Schwarz E. Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS One* 2013;8: e66693.
15. Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, Ding W, Yu L, Wang X, Wang L, Shen H, Zhang C, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet* 2015.
16. Schmitz M, Driesch C, Jansen L, Runnebaum IB, Dürst M. Non-Random Integration of the HPV Genome in Cervical Cancer. *PLoS One* 2012;7: e39632.
17. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, Rocco JW, Teknos TN, Kumar B, Wangsa D, He D, Ried T, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Research* 2014;24: 185–99. [PubMed: 24201445]
18. Ojesina AI, Lichtenstein L, Freeman SS, Peadarallu CS, Imaz-Rosshandler I, Pugh TJ, Cherniack AD, Ambrogio L, Cibulskis K, Bertelsen B, Romero-Cordoba S, Trevino V, et al. Landscape of genomic alterations in cervical carcinomas. *Nature* 2014;506: 371–5. [PubMed: 24390348]
19. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489: 57–74. [PubMed: 22955616]
20. Annunziata C, Buonaguro L, Buonaguro F, Tornesello M. Characterization of the Human Papillomavirus (HPV) Integration Sites into Genital Cancers. *Pathol Oncol Res* 2012;18: 803–8. [PubMed: 22451136]
21. Bryant D, Onions T, Raybould R, Flynn Á, Tristram A, Meyrick S, Giles P, Ashelford K, Hibbitts S, Fiander A, Powell N. mRNA sequencing of novel cell lines from human papillomavirus type-16 related vulval intraepithelial neoplasia: Consequences of expression of HPV16 E4 and E5. *Journal of Medical Virology* 2014;86: 1534–41. [PubMed: 24898764]
22. Chaiwongkot A, Pientong C, Ekalaksananan T, Vinokurova S, Kongyingyoes B, Chumworathayi B, Patarapadungkit N, Siriaunkgul S, von Knebel Doeberitz M. Detection of the human papillomavirus 58 physical state using the amplification of papillomavirus oncogene transcripts assay. *Journal of Virological Methods* 2013;189: 290–8. [PubMed: 23466631]
23. Dall KL, Scarpini CG, Roberts I, Winder DM, Stanley MA, Muralidhar B, Herdman MT, Pett MR, Coleman N. Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. *Cancer Res* 2008;68: 8249–59. [PubMed: 18922896]
24. Das P, Thomas A, Mahantshetty U, Shrivastava SK, Deodhar K, Mulherkar R. HPV genotyping and site of viral integration in cervical cancers in Indian women. *PLoS One* 2012;7: e41012.
25. Gao G, Johnson SH, Kasperbauer JL, Eckloff BW, Tombers NM, Vasmataz G, Smith DI. Mate pair sequencing of oropharyngeal squamous cell carcinomas reveals that HPV integration occurs much less frequently than in cervical cancer. *Journal of Clinical Virology* 2014;59: 195–200. [PubMed: 24440282]
26. Huebbers CU, Preuss SF, Kolligs J, Vent J, Stenner M, Wieland U, Silling S, Drebber U, Speel EJ, Klussmann JP. Integration of HPV6 and downregulation of AKR1C3 expression mark malignant transformation in a patient with juvenile-onset laryngeal papillomatosis. *PLoS One* 2013;8: e57207.
27. Kalantari M, Villa LL, Calleja-Macias IE, Bernard H-U. Human papillomavirus-16 and -18 in penile carcinomas: DNA methylation, chromosomal recombination and genomic variation. *International Journal of Cancer* 2008;123: 1832–40. [PubMed: 18688866]
28. Khoury JD, Tannir NM, Williams MD, Chen Y, Yao H, Zhang J, Thompson EJ, Network tT, Meric-Bernstam F, Medeiros LJ, Weinstein JN, Su X. Landscape of DNA Virus Associations across Human Malignant Cancers: Analysis of 3,775 Cases Using RNA-Seq. *Journal of Virology* 2013;87: 8916–26. [PubMed: 23740984]
29. Kraus I, Driesch C, Vinokurova S, Hovig E, Schneider A, von Knebel Doeberitz M, Durst M. The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. *Cancer Res* 2008;68: 2514–22. [PubMed: 18381461]
30. Lace MJ, Anson JR, Klussmann JP, Wang DH, Smith EM, Haugen TH, Turek LP. Human Papillomavirus Type 16 (HPV-16) Genomes Integrated in Head and Neck Cancers and in HPV-16-

Immortalized Human Keratinocyte Clones Express Chimeric Virus-Cell mRNAs Similar to Those Found in Cervical Cancers. *Journal of Virology* 2011;85: 1645–54. [PubMed: 21123375]

31. Li H, Zhang R, Cai Y, Li Y, Cheng X, Zhu B, Yang Y, Xiang Y. Determination of integrated HPV58 sequences in cervical lesions. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society* 2012;22: 1234–7. [PubMed: 22801029]
32. Li H, Yang Y, Zhang R, Cai Y, Yang X, Wang Z, Li Y, Cheng X, Ye X, Xiang Y, Zhu B. Preferential sites for the integration and disruption of human papillomavirus 16 in cervical lesions. *Journal of Clinical Virology* 2013;56: 342–7. [PubMed: 23290390]
33. Liang WS, Aldrich J, Nasser S, Kurdoglu A, Phillips L, Reiman R, McDonald J, Izatt T, Christoforides A, Baker A, Craig C, Egan JB, et al. Simultaneous Characterization of Somatic Events and HPV-18 Integration in a Metastatic Cervical Carcinoma Patient Using DNA and RNA Sequencing. *International Journal of Gynecological Cancer* 2014;24: 329–38 10.1097/IGC.000000000000049. [PubMed: 24418928]
34. Lu X, Lin Q, Lin M, Duan P, Ye L, Chen J, Chen X, Zhang L, Xue X. Multiple-integrations of HPV16 genome and altered transcription of viral oncogenes and cellular genes are associated with the development of cervical cancer. *PLoS One* 2014;9: e97588.
35. Matovina M, Sabol I, Grubisic G, Gasperov NM, Grce M. Identification of human papillomavirus type 16 integration sites in high-grade precancerous cervical lesions. *Gynecol Oncol* 2009;113: 120–7. [PubMed: 19157528]
36. Nakanishi G, Fujii K, Asagoe K, Tanaka T, Iwatsuki K. Human papillomavirus genome integration in multifocal vulvar Bowen's disease and squamous cell carcinoma. *Clinical and Experimental Dermatology* 2009;34: e965–e7. [PubMed: 20055874]
37. Nambaru L, Meenakumari B, Swaminathan R, Rajkumar T. Prognostic significance of HPV physical status and integration sites in cervical cancer. *Asian Pacific journal of cancer prevention : APJCP* 2009;10: 355–60. [PubMed: 19640172]
38. Olthof NC, Speel E-JM, Kolligs J, Haesevoets A, Henfling M, Ramaekers FCS, Preuss SF, Drebber U, Wieland U, Silling S, Lam WL, Vucic EA, et al. Comprehensive Analysis of HPV16 Integration in OSCC Reveals No Significant Impact of Physical Status on Viral Oncogene and Virally Disrupted Human Gene Expression. *PLoS ONE* 2014;9: e88718.
39. Olthof NC, Huebbers CU, Kolligs J, Henfling M, Ramaekers FC, Cornet I, van Lent-Albrechts JA, Stegmann AP, Silling S, Wieland U, Carey TE, Walline HM, et al. Viral load, gene expression and mapping of viral integration sites in HPV16-associated HNSCC cell lines. *Int J Cancer* 2015;136: E207–18. [PubMed: 25082736]
40. Peter M, Stransky N, Couturier J, Hupé P, Barillot E, de Cremoux P, Cottu P, Radvanyi F, Sastre-Garau X. Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. *The Journal of Pathology* 2010;221: 320–30. [PubMed: 20527025]
41. Ragin CC, Reshmi SC, Gollin SM. Mapping and analysis of HPV16 integration sites in a head and neck cancer cell line. *Int J Cancer* 2004;110: 701–9. [PubMed: 15146560]
42. Raybould R, Fiander A, Wilkinson GWG, Hibbitts S. HPV integration detection in CaSki and SiHa using detection of integrated papillomavirus sequences and restriction-site PCR. *Journal of Virological Methods* 2014;206: 51–4. [PubMed: 24880069]
43. Schmitz M, Driesch C, Beer-Grondke K, Jansen L, Runnebaum IB, Dürst M. Loss of gene function as a consequence of human papillomavirus DNA integration. *International Journal of Cancer* 2012;131: E593–E602. [PubMed: 22262398]
44. Yu T, Ferber MJ, Cheung TH, Chung TK, Wong YF, Smith DI. The role of viral integration in the development of cervical cancer. *Cancer genetics and cytogenetics* 2005;158: 27–34. [PubMed: 15771901]
45. Bodelon C, Vinokourova S, Sampson JN, den Boon JA, Walker JL, Horswill MA, Korthauer K, Schiffman M, Sherman ME, Zuna RE, Mitchell J, Zhang X, et al. Chromosomal copy number alterations and HPV integration in cervical precancer and invasive cancer. *Carcinogenesis* 2015.
46. The Cancer Genome Atlas Network. Integrated genomic and molecular characterization of cervical cancer. Submitted.
47. Dyer N, Young L, Ott S. Artifacts in the data of Hu et al. *Nat Genet* 2016;48: 2–3. [PubMed: 26711106]

48. Rusan M, Li YY, Hammerman PS. Genomic Landscape of Human Papillomavirus-Associated Cancers. *Clinical Cancer Research* 2015;21: 2009–19. [PubMed: 25779941]
49. Christiansen IK, Sandve GK, Schmitz M, Dürst M, Hovig E. Transcriptionally Active Regions Are the Preferred Targets for Chromosomal HPV Integration in Cervical Carcinogenesis. *PLoS ONE* 2015;10: e0119566.
50. Doolittle-Hall J, Cunningham Glasspoole D, Seaman W, Webster-Cyriaque J. Meta-Analysis of DNA Tumor-Viral Integration Site Selection Indicates a Role for Repeats, Gene Expression and Epigenetics. *Cancers* 2015;7: 0887.
51. Zhang R, Shen C, Zhao L, Wang J, McCrae M, Chen X, Lu F. Dysregulation of host cellular genes targeted by human papillomavirus (HPV) integration contributes to HPV-related cervical carcinogenesis. *Int J Cancer* 2016;138: 1163–74. [PubMed: 26417997]
52. Thomas LK, Bermejo JL, Vinokurova S, Jensen K, Bierkens M, Steenbergen R, Bergmann M, von Knebel Doeberitz M, Reuschenbach M. Chromosomal gains and losses in human papillomavirus-associated neoplasia of the lower genital tract - a systematic review and meta-analysis. *Eur J Cancer* 2014;50: 85–98. [PubMed: 24054023]
53. Lukusa T, Fryns JP. Human chromosome fragility. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2008;1779: 3–16. [PubMed: 18078840]
54. Gene Ontology C Creating the gene ontology resource: design and implementation. *Genome Res* 2001;11: 1425–33. [PubMed: 11483584]

Novelty and impact

Integration of human papilloma virus (HPV) DNA into the host genome is observed at the transition from precancer to cancer and is considered to be a driver of carcinogenesis. We identified several new integration hotspots that were not identified in individual studies. We evaluated differences of integration events by cancer type, HPV type and assay type. Finally, we analyzed genomic features of integration sites using ENCODE data and evaluated co-location with significantly mutated driver genes.

Legend Figure 1. Characteristics of reported HPV events.
A includes all reported HPV integration events form human specimens and cell lines. B and C include HPV integration events reported in human specimens only.

Figure 2 legend. Distribution of the location of HPV integration events in the human genome. Data for this plot was restricted to human specimens and with nucleotide information for the integration event. **A.** Distribution of the locations of the HPV integration events across the genome. The outer circle represents the locations of the 24 human chromosomes. The green bars in the inner circle depict the frequency of the HPV integration event at each location. Each green bar represents a bin size of 3 Kb. The scale for the frequency is represented by the light grey concentric circles, with each concentric circle representing 2 integration events. **B.** Position of HPV integration events and genes in 3q28 (only integration events with known position are plotted). Green bars are integration events. **C.** Position of HPV

integration events and genes in 8q24.21 (only integration events with known position are plotted). **D.** Position of HPV integration events and genes in 13q22.1 (only integration events with known position are plotted).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3 legend. Integration sites from human specimens with exact position with respected to RefSeq genes and commonly mutated genes.

Only breakpoints with exact nucleotide position were included (N=639). Breakpoints determined with transcript-based assays that did not map to an exon or UTR were excluded.

A. Integration breakpoints in relation to UCSC RefGenes, which include protein coding and non-protein coding genes. P-values are based on a test of proportions (Pearson's χ^2 test). **B.** Distribution of minimum distances from intergenic integration breakpoints to UCSC RefGenes. **C.** Integration breakpoints in relation to 19 significantly mutated genes (*ARID1A*, *ELF3*, *NFE2L2*, *CASP8*, *TGFBR2*, *PIK3CA*, *FBXW7*, *HLA.A*, *HLA.B*, *PTEN*,

KRAS, ERBB3, CFBF, TP53, MED1, ERBB2, STK11, MAPK1, EP300) from Ojesina *et al.* (17) and TCGA (19). P-value computed using Fisher’s exact test.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Integration sites included in this analysis.

Reference	Method	Cancer site	Sample type	Number of samples	Number of integration sites	Ratio # integra. sites to # samples
Akagi <i>et al.</i> (2014) Genome Res	WGS	Head and neck, cervix	Human samples, cell lines	8	29	3.6
Annunziata <i>et al.</i> (2012) Pathol Oncol Res.	Alu-HPV-based PCR	Penis	Human samples	2	2	1.0
Bodelon <i>et al.</i> (2015) Carcinogenesis	APOT, NGS	Cervix	Human samples	23	28	1.2
Bryant <i>et al.</i> (2014) J Med Virol	APOT, DIPS	Vulva	Cell lines	2	3	1.5
Chaiwongkot <i>et al.</i> (2013) J Virol Methods	APOT	Cervix	Human samples	3	3	1.0
Dall <i>et al.</i> (2008) Cancer Res	RS-PCR, APOT	?	Cell lines	24	27	1.1
Das <i>et al.</i> (2012) PLOS ONE	APOT	Cervix	Human samples	48	49	1.02
Gao <i>et al.</i> (2014) J Clin Virol	NGS	Oropharynx	Human samples	3	4	1.3
Hu <i>et al.</i> (2015) Nature Genetics	NGS	Cervix	Human samples, cell lines	83	310	3.7
Huebbers <i>et al.</i> (2013) PLOS ONE	APOT	Larynx	Human sample	1	1	1.0
Kalantari <i>et al.</i> (2008) Int J Cancer	rtPCR	Penis	Human samples	6	6	1.0
Khoury <i>et al.</i> (2013) J Virol	RNA-seq	Head and neck	Human samples	21	24	1.1
Kraus <i>et al.</i> (2008) Cancer Res	APOT	Cervix	Human samples	74	80	1.1
Lace <i>et al.</i> (2011) J Virol	Southern blot	Head and neck	Human samples	1	6	6
Li <i>et al.</i> (2012) Int J Gynecol Cancer	DIPS	Cervix	Human samples	5	8	1.6
Li <i>et al.</i> (2013) J Clin Virol	DIPS	Cervix	Human samples, cell lines	32	61	1.9
Liang <i>et al.</i> (2014) Int J Gynecol Cancer	WSG	Cervix	Human sample	1	1	1.0
Lu <i>et al.</i> (2014) PLOS ONE	APOT	Cervix	Human samples	23	57	2.5
Matovina <i>et al.</i> (2009) Gynecol Oncol	DIPS	Cervix	Human samples	11	11	1.0
Nakanishi <i>et al.</i> (2009) Clin Exp Dermatol	APOT	Vagina, vulva	Human samples	3	4	1.3
Nambaru <i>et al.</i> (2009) Asian Pac J Cancer Prev	APOT	Cervix	Human samples	37	37	1.0
Ojesina <i>et al.</i> (2014) Nature	RNA-seq, WGS	Cervix	Human samples	56	70	1.3
Olthof <i>et al.</i> (2014) PLOS ONE	APOT, DIPS	Oropharynx	Human samples	29	40	1.4
Olthof <i>et al.</i> (2015) Int J Cancer	APOT, DIPS	?	Cell lines	6	6	1.0
Parfenov <i>et al.</i> (2014) PNAS	WGS	Head and neck	Human samples	24	40	1.7
Peter <i>et al.</i> (2010) J Pathol	DIPS	Cervix	Human samples, cell lines	42	51	1.2
Ragin <i>et al.</i> (2004) Int J Cancer	FISH	Head and neck	Cell lines	2	23	11.5
Raybould <i>et al.</i> (2014) J Virol Methods	DIPS	Cervix	Cell lines	2	2	1.0
Schmitz <i>et al.</i> (2012) Int J Cancer	APOT, DIPS	Cervix	Human samples	10	12	
Schmitz <i>et al.</i> (2012) PLOS ONE	APOT	Cervix	Human samples	47	49	1.04

Reference	Method	Cancer site	Sample type	Number of samples	Number of integration sites	Ratio # integra. sites to # samples
TCGA –Cervical cancer (Submitted)	RNA-seq	Cervix	Human samples	142	221	1.6
Wentzensen <i>et al.</i> (2004), Cancer research	APOT, DIPS, FISH, PCR	Cervix, head and neck, penis, vagina, vulva	Human samples, cell lines	162	182	1.1
Wentzensen lab (unpublished data)	APOT, NGS	Cervix	Human samples	3	4	1.3
Xu <i>et al.</i> (2013) PLOS ONE	NGS	Cervix	Human samples	41	60	1.5
Yu <i>et al.</i> (2005) Cancer Genet Cytogenet	RS-PCR	Cervix	Human samples	35	40	1.1

APOT: Amplification of papillomavirus oncogene transcripts; DIPS: Detection of integrated papillomavirus sequences; FISH: Fluorescence *in situ* hybridization; NGS: Next generation sequencing; PCR: Polymerase chain reaction; rliPCR: Reverse ligation PCR; RS-PCR: Restriction site PCR; WGS: Whole genome sequence.

Table 2.

Integration hotspots.

Chromosome	Band	Start position	End position	Number of integration events observed	Percent of integration events observed (%)	Percent of integration events expected (%)	P-value *
13	q22.1	73,300,000	75,400,000	38	2.70	0.07	3.13×10^{-46}
8	q24.21	127,300,000	131,500,000	41	2.91	0.14	5.71×10^{-39}
3	q28	187,900,000	192,300,000	31	2.20	0.14	5.02×10^{-26}
17	q23.1	57,600,000	58,300,000	10	0.71	0.02	5.70×10^{-12}
17	q23.2	58,300,000	61,100,000	12	0.85	0.09	1.95×10^{-8}
14	q24.1	67,900,000	70,200,000	11	0.78	0.08	3.29×10^{-8}
2	q22.3	144,100,000	148,700,000	14	1.00	0.15	4.88×10^{-8}
17	p11.1	22,200,000	24,000,000	4	0.28	2.2×10^{-5}	5.66×10^{-8}
3	p14.2	58,600,000	63,700,000	14	1.00	0.18	3.95×10^{-7}
8	q24.22	131,500,000	136,400,000	11	0.78	0.17	5.10×10^{-5}

Based on 1,407 observed integration events in human samples and 1,000,000 randomly generated integration events. Ordered according to P-value

* P-value based on Fisher's exact test. Only shown results with P-values that passed Bonferroni correction (i.e., $P < 5.80 \times 10^{-5}$).

Table 3.

Characterization of genomic integration sites in human specimens with known nucleotide position (N=919).

Genomic feature	Observed distribution of minimum distances (Kb) Median (IQR) (N=919)	Expected distribution of minimum distances (Kb) Median (IQR) (N=10,000)	P-value *
CpG enriched regions	53.7 (13.6 – 195.7)	77.1 (23.0 – 256.1)	9.38×10^{-10}
Fragile sites ¹	2,361.2 (0 – 11,251.5)	4,380.9 (0 – 13,276.8)	2.05×10^{-6}
Recombination rates ²	209.5 (64.5 – 618.6)	240.3 (71.8 – 704.4)	0.040
Segmental duplications	192.7 (67.2 – 524.8)	215.8 (65.2 – 507.8)	0.78
Repetitive elements			
<i>Long Interspersed Nuclear Element (LINE)</i>	0.7 (0.1 – 1.9)	0.6 (0 – 1.7)	1.75×10^{-5}
<i>LINE-1</i>	1.9 (0.4 – 5.2)	1.4 (0.2 – 4.0)	1.63×10^{-6}
<i>Short Interspersed Nuclear Element (SINE)</i>	0.5 (0.1 – 1.3)	0.5 (0.1 – 1.5)	0.0065
<i>Alu</i>	0.9 (0.2 – 2.4)	1.0 (0.3 – 2.9)	0.0094
<i>Long Terminal Repeats (LTR)</i>	2.5 (0.8 – 7.4)	2.1 (0.5 – 5.6)	0.0013
<i>ERV1</i>	17.3 (5.4 – 39.6)	14.2 (4.7 – 35.0)	0.0066
<i>ERVK</i>	228.2 (86.7 – 528.4)	202.4 (72.7 – 451.3)	0.0035
<i>ERV1/ERV1-MaLR</i>	3.5 (1.3 – 9.5)	3.2 (0.9 – 7.7)	8.14×10^{-4}
<i>Short tandem repeats</i>	2.5 (0.9 – 5.6)	2.8 (1.1 – 5.9)	0.0079
Open chromatin ³			
<i>FAIRE</i>	8.1 (2.6 – 21.1)	12.9 (4.2 – 35.2)	3.57×10^{-19}
<i>DNaseI hypersensitivity</i>	7.8 (1.8 – 30.8)	18.3 (4.8 – 73.7)	6.65×10^{-30}
Peaks of DNA replication timing ^{3,4}	129.7 (60.7 – 241.7)	162.0 (76.7 – 300.4)	1.74×10^{-10}
Histone makers ^{3,5}			
<i>EZH2 (39875)</i>	16.5 (0 – 101.3)	26.7 (0.2 – 152.1)	5.81×10^{-5}
<i>H2A.Z</i>	5.3 (0 – 23.8)	11.2 (0.2 – 51.6)	3.47×10^{-13}
<i>H3K4me1</i>	6.8 (0.6 – 26.7)	21.3 (4.5 – 99.1)	1.97×10^{-43}
<i>H3K4me2</i>	8.9 (1.4 – 30.3)	21.3 (5.6 – 72.0)	8.79×10^{-34}
<i>H3K4me3</i>	18.1 (3.1 – 77.4)	48.9 (12.7 – 178.4)	9.35×10^{-38}
<i>H3K9ac</i>	20.6 (3.0 – 110.4)	69.7 (15.2 – 271.0)	4.14×10^{-44}
<i>H3K9me3</i>	6.3 (0 – 42.0)	10.7 (0 – 51.1)	1.72×10^{-4}
<i>H3K27ac</i>	18.9 (3.2 – 85.9)	54.5 (13.4 – 202.8)	3.35×10^{-40}
<i>H3K27me3</i>	85.6 (2.0 – 336.8)	60.5 (0 – 321.7)	0.064
<i>H3K36me3</i>	26.3 (0 – 215.2)	95.1 (3.6 – 398.1)	1.01×10^{-21}
<i>H3K79me2</i>	24.1 (2.1 – 93.5)	49.1 (10.3 – 160.9)	2.37×10^{-21}
<i>H4K20me1</i>	4.6 (0.3 – 15.7)	6.6 (1.3 – 20.6)	1.68×10^{-8}
<i>Pol2 (MMS-128P)</i>	2.5 (0 – 23.1)	9.7 (0 – 47.0)	1.04×10^{-20}
Chromatin genome segmentation ^{3,6}			
<i>CTCF enrich elements</i>	13.2 (4.7 – 32.4)	16.0 (5.7 – 38.0)	1.92×10^{-5}

Genomic feature	Observed distribution of minimum distances (Kb) Median (IQR) (N=919)	Expected distribution of minimum distances (Kb) Median (IQR) (N=10,000)	P-value *
<i>Enhancers</i>	11.9 (3.6 – 39.3)	29.6 (8.5 – 105.7)	1.68×10^{-33}
<i>Transcription Start Sites (TSS)</i>	73.9 (18.2 – 269.9)	142.3 (35.7 – 458.9)	3.77×10^{-19}
<i>Promoter flanking regions</i>	83.1 (26.2 – 223.5)	107.3 (33.9 – 310.4)	2.14×10^{-7}
<i>Transcribed regions</i>	0.4 (0.0 – 1.3)	0.5 (0.1 – 1.3)	0.11
<i>Repressed or low activity regions</i>	0 (0 – 0.3)	0 (0 – 0)	2.24×10^{-10}

IQR: Interquartile range. CTCF: CCCTC-binding factor (zinc finger protein).

* Based on the Wilcoxon rank-sum test comparing the observed to the expected distributions.

¹ Location of fragile sites from Lukusa T, Fryns JP. Human chromosome fragility. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2008;1779(1):3–16.

² Average of male and female specific recombination rates.

³ Data for the HeLa-S3 cell line.

⁴ Based on the ENCODE Repli-Seq data from the University of Washington.

⁵ Based on the ENCODE histone modification ChIP-seq peak data from the Broad Institute.

⁶ Based on the ENCODE chromatin state segmentation from the combined ChromHMM and Segway algorithms.