ACIN
ASOCIACIÓN COLOMBIANA DE INFECTOLOGÍA

infectio

ARTÍCULO ORIGINAL

In silico analysis of genomic variables associated to HPV16 integration sites

Nicole Díaz-Moreno[1], Julio César Osorio[2], Herney Andrés García-Perdomo[3], Andrés Castillo[4,*]

Abstract

Background: Despite current prophylactic interventions, a significant proportion of patients suffers a cancer-specific mortality, leading to a global awareness of the importance of identifying factors associated to the etiology of HPV-associated cancer. According to this, HPV-DNA integration into human genome is an important event in the pathogenesis.

Purpose: To identify *in silico*, molecular regions of the genome where the HPV integration events occur

Methods: We performed a bioinformatic study based on a systematic search in Medline through PubMed, Embase and Lilacs from inception to April 2019. We used the UCSC Genome Browser Home (https://genome.ucsc.edu) to evaluate the genetic environment.

Results: HPV integration sites by anatomical location related to cervical cancer were 374 (61%). In addition, 325 (87%) of these integration sites had HPV-16, 21 (5%) had HPV-18 and 28 (7%) had another type of genotype. Oro-pharyngeal cavity was the second anatomic site with 162 (26%) integration sites. It is noteworthy that the HPV-16 was found integrated into 160 (99%) analyzed sites.

Conclusion: Our results suggest that many of the integration sites reported in the scientific literature are HPV 16 from squamous cell carcinomas and 50% of HPV16 were integrated into transcriptional units that might affect the expression of gene target.

Keywords: Papillomaviridae; Human papillomavirus 16; virus integration; Genomic Structural Variation; computational biology

Análisis *in silico* de variables genómicas asociadas a sitios de integración de HPV16

Resumen

Antecedentes: A pesar de las intervenciones profilácticas actuales, una proporción significativa de pacientes muere debido al cáncer, lo que aumenta la conciencia global de la importancia de identificar los factores asociados a la etiología del cáncer asociado al VPH. Según esto, la integración del ADN-VPH en el genoma humano es un evento importante en la patogénesis.

Propósito: Identificar *in silico*, las regiones moleculares del genoma donde ocurren los eventos de integración del VPH

Métodos: Realizamos un estudio bioinformático basado en una búsqueda sistemática en Medline a través de PubMed, Embase y Lilacs desde el inicio hasta abril de 2019. Utilizamos el UCSC Genome Browser Home (https://genome.ucsc.edu) para evaluar el entorno genético.

Resultados: Los sitios de integración del VPH relacionados con el cáncer de cuello uterino fueron 374 (61%). Además, 325 (87%) de estos sitios de integración tenían VPH-16, 21 (5%) tenían VPH-18 y 28 (7%) tenían otro tipo de genotipo. La cavidad orofaríngea fue el segundo sitio anatómico con 162 (26%) sitios de integración. Es de destacar que el VPH-16 se encontró integrado en 160 (99%) sitios analizados.

Conclusión: Nuestros resultados sugieren que muchos de los sitios de integración reportados en la literatura científica que presentan al VPH-16 son carcinomas de células escamosas y que el 50% de estos VPH-16 se integraron en unidades transcripcionales que podrían afectar la expresión de algún gen objetivo.

Palabras clave: Papillomaviridae; Virus del papiloma humano 16; integración de virus; Variación estructural genómica; Biología Computacional

Introduction

Worldwide, the diagnosis of HPV-associated cancer is increasing[2]. Despite current prophylactic interventions, a significant proportion of patients suffers a cancer-specific mortality[3], which leads to a global awareness of the importance of identifying factors associated in the etiology of cancer. The latter could provide a better understanding of preventive, predictive and public health purposes.

According to this, HPV-DNA integration into human genome is frequently an important event in the pathogenesis of HPV-associated cancer[4]. The virus-mediated carcinogenesis has been mediated by two important pathways: Deregulation of viral gene expression and genomic instability of the host[5]. Accordingly, there is experiments describing that integration could result in augmented levels of oncogene (E6/E7) transcripts[6]. The selective growth and the genomic instability of the cell come from the integration process[7]. Furthermo-

1. Universidad Icesi
2. Universidad del Valle
3. Associate Professor Universidad del Valle.
4. Department of Biology, Faculty of Natural and Exact Sciences. Universidad del Valle Universidad del Valle.

* Autor para correspondencia.
Correo electrónico: andres.castillo.g@correounivalle.edu.co

Recibido: 12/02/2019; Actualizado: 07/10/2019; Aceptado: 09/11/2019

Cómo citar este artículo: N. Díaz-Moreno, *et al*. *In silico* analysis of genomic variables associated to HPV16 integration sites. Infectio 2020; 24(2):76-80
http://dx.doi.org/10.22354/in.v24i2.836

re, host genomes with integration of HPV have also different expression and methylation profiles compared with nonintegrated genomes[8].

HPV integration is considered a process which occurs randomly in almost all chromosomes[9]. However, some important regions of the genome have been found to have repeatedly integration of HPV[10]. Other genetics elements surrounding the integration site are involved in the integration event and enhance the genomic instability. These elements are part of genomic environment, for example, it has been indicated that HPV integration occurred within - or neighboring - sequences of Alu-repeats[11]. Additionally, due to persistent infection of HPV, it becomes susceptible to epigenetic modification (CpG Islands and rich CG regions)[12]. Another element is DNase I hypersensitive sites where studies provide evidence that DNase I sites are surrounding the host-viral junction[13,14]. Finally, identifying genes with HPV integration and evaluating subsequent alterations in the expression, turn the transcriptional regions into important elements to analyze. However, the intergenic sites must be taken into account because viral integration into these frequently occurs in cervical cancer and it has also been identified in oral squamous cell carcinoma[15].

The aim of this study was to identify *in silico*, molecular regions of the genome where the HPV16 integration events occur.

Methods

We performed a bioinformatic study based on a systematic search in Medline through PubMed, Embase and Lilacs from inception to April 2019. Our search included the following terms: "((HPV) OR human papillomavirus) AND ((cancer) OR carcinoma) AND ((integration) OR breakpoint)".

Inclusion criteria: We included studies that reported HPV integration places and focused on published articles. We did not impose any language restriction. Two reviewers selected the studies by title, abstract and full-text. The genome browser positions or nucleotide sequences adjacent to the viral integration site of the selected articles were extracted.

Exclusion criteria: Those articles showing no information regarding integration places.

In-silico methods

It consists of two components: 1) The search and retrieval of the integration sites reported for different HPV genotypes, and 2) The evaluation of the integration sites in a descriptive way (chromosomal distribution and clinical variables) and through bioinformatics tool. We used the UCSC Genome Browser Home (https://genome.ucsc.edu) to evaluate the transcriptional units, intergenic sites, DNase I hypersensitive sites (DHSs), Alu sequences and CpG islands present on viral breakpoint. The reference genome GRCh38-2013 was used for the alignment of nucleotide sequences.

A descriptive analysis of the data obtained from each of the variables was performed related to the HPV-16 genotype. Additionally, through *IBM SPSS Statistics 20* a statistical analysis was carried out to identify significant differences between the content of CG dinucleotides in exons and introns with integration sites, to test for major integration events in hypomethylated regions. Finally, the genes where an HPV integration site was found were analyzed through gene set enrichment analysis. This analysis can find significant associations with functional gene sets.

Results

We found 1808 studies according to the search strategy; then we included integration sites from 20 studies. The integration sites were distributed on all chromosomes and 667 integration site were obtained (Figure 1). Chromosome 19 had the highest number of integration sites (23 integration sites).

When analyzing the HPV integration sites by anatomical location, we found that 374 (61%) were related to cervical cancer. In addition, 325 (87%) of these integration sites had HPV-16, 21 (5%) had HPV-18 and 28 (7%) had another type of genotype. Oro-pharyngeal cavity was the second anatomic site with 162 (26%) integration sites. It is noteworthy that the HPV-16 was found integrated into 160 (99%) analyzed sites. The oral cavity presented 59 (10%) integration sites where HPV-18 showed a higher frequency (49%) compared to HPV-16 (31%). A remaining 20% presented other genotypes of HPV. Finally, the anatomical sites anus and larynx showed 11 (2%) and 5 (1%) integration sites respectively (Table 1).

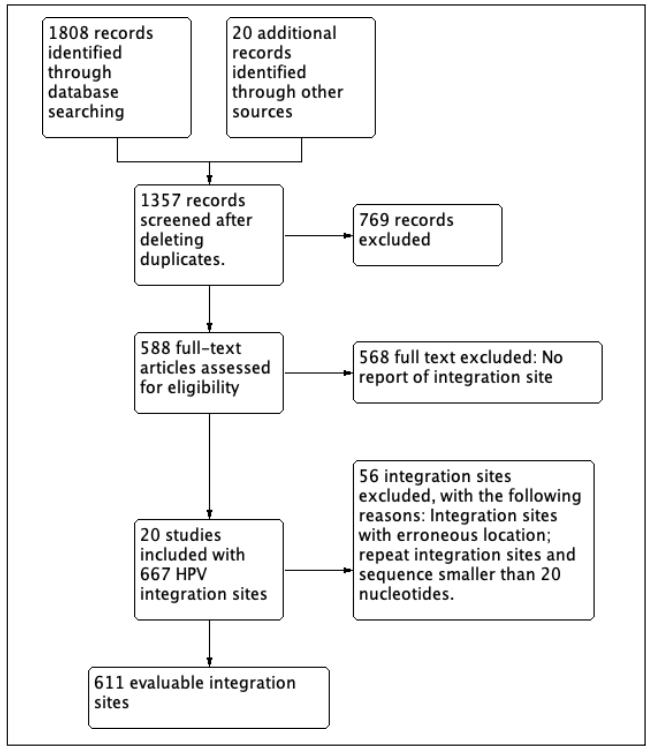

Figure 1. Flow diagram of In *silico* methodology. In total, 611 integration sites were analyzed for HPV-16

Table 1. HPV integration site by anatomical site

Anatomic Site	Integration site	HPV16	HPV18	Other
Cervical cancer	374	325	21	28
Oropharynx	162	160	0	2
Oral	59	31	20	8
Anus	11	11	0	0
Larynx	5	5	0	0
TOTAL	611	532	41	38

When analyzing the different genomic variables, 50% of HPV16 integrates into transcriptional units (Table 2). 306 sites, coming from this group were divided into 24 integration events in exonic regions (8%, corresponding to 19 genes) and 282 integration events in intronic regions (92%, corresponding to 163 genes).

Table 2. Genomic variables present at HPV16 integration sites.

Category number	Genomic Variables	Integration Sites n (%)
1	Transcriptional units	268 (50)
2	DNase I hypersensitive sites	93 (18)
3	Alu sequences	53 (10)
4	CpG islands	5 (1)
5	Intergenic sites	113 (21)

From 182 genes present in the sites of viral integration, four presented integration events in both exonic and intronic regions, therefore, a total of 178 genes involved in viral integration events were shown. Additionally, we found that the exons tend to be more methylated which could explain the tendency of HPV to integrate into intronic regions. The percentage of CG in the exonic region was 55.91% (95%CI: 45.34-66.47) while in the intronic region was 44.03% (95%CI:41.32-46.73). The percentages of CG in the two regions were significantly different (p = 0.015).

Furthermore, three out of 178 genes, were tumor suppressor genes (TP63, PRBM1, and FHIT) and four were oncogenes (ERBB4, MECOM, MLLT1 and USP4) (Figure 2).

Intergenic sites and sites of DHSs were present in approximately 20% of the integration sites; Alu sequences were present in 10% of integration sites and CG rich regions were shown in only five (1%) of the integration sites included in the study.

Discussion

In the *in-silico* review, 611 of 667 integration sites for the HPV 16 genotype were collected, with a high proportion of squamous cell carcinomas (91%) and 60% in cervical-uterine anatomical sites, followed by 30% in oropharyngeal anatomical sites.

Analyzing the proportion of integration sites by chromosome size, a greater proportion of integration sites were found on the small chromosomes 19, 20 and 21, followed by chromosome 9. Also, chromosomes 9, X, 20 and 21 were found to have a high frequency of affected gene regions, according to the number of total genes per chromosome. The previous results show that chromosomes 9, 20 and 21 have a possible relevance in the integration event because they present the highest proportions of integration sites according to their size and number of total genes, also, it could be suggested that that integration is not a random event such as indicate some authors[9,16].

With respect to the integration sites by anatomical site, it is clear that cervical cancer presents a higher frequency of integration sites reported in the literature. Cervical carcinoma is one of the most commonly occurring cancers in women worldwide, accordingly, greater than 99% contains HPV sequences[17]. On the other hand, oropharyngeal cancer has also reported a large number of integration sites in the literature. HPV infection of the mouth and oropharynx can be acquired by a variety of sexual and social forms of transmission. HPV-16 accounted for a larger majority of HPV-positive oropharyngeal SCCs (86.7%) than HPV-positive oral (68.2%) and laryngeal SCCs (69.2%). HPV-18 was the second most frequent type detected: 2.8% in oropharyngeal, 34.1% in oral, and 17% in laryngeal SCCs. Other oncogenic HPVs were rarely detected in HNSCC[18]. Our data on HPV-18 could suggest that the frequencies reported in some scientific articles may be underestimated.

A significant result in the analysis of genomic variables was that 50% of HPV16 integration events occur in coding regions or transcriptional units. This result agrees with the found by Zhang and colleagues from the analysis of 14 cervical cancer publications where these integration sites showed preference for transcriptionally active regions and intragenic areas[19]. Furthermore, a higher proportion of integration events occur in intronic regions (92%), results that are consistent with the reported by Christiansen and colleagues in 2015[20]. A total of 178 coding genes were affected by the HPV integration event.

The 178 coding genes were subjected to a Gene Set Enhancement Analysis (GSEA) focused on their biological function. 59 of these genes were part of the following biological processes: tissue development, expression of genes of active regulation, the formation of anatomical structures, biosynthetic processes of positive regulation, regulation of transcription, cell cycle and development of epithelium. The other 119 genes each are part of biological processes from many other metabolic pathways.

In the 178 genes, three were tumor suppressor genes (FHIT, PRBM1 and TP63) and four were oncogenes (ERBB4, MECOM, MLLT1 and USP4). The importance and relevance of integration sites on the tumor behavior is still unclear nowa-

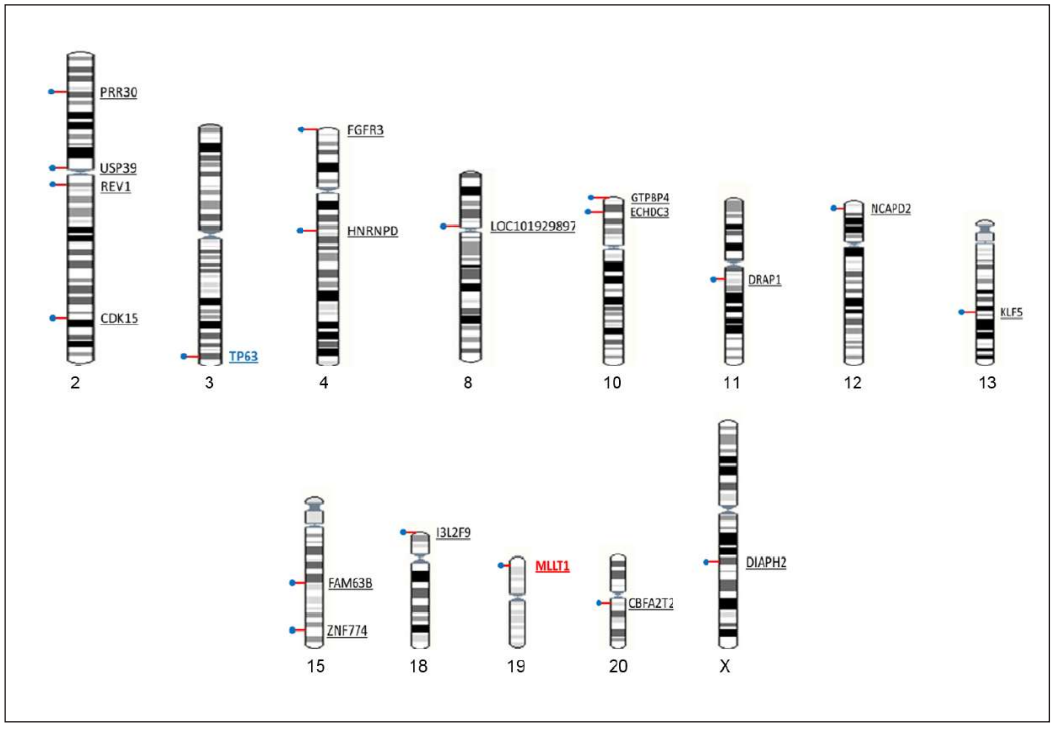

Figure 2. Viral integration sites in exonic regions. In total 21 viral integration sites were found in exonic regions of 19 transcriptional units. The human genes USP39 and TP63 showed two vira integration events. The TP63 is a suppressor gene and the MLLT1 is an oncogene.

days, however, understanding the complex of genes and viral integration with the disturbances of cellular expression might allow changing the oncologic therapies for patients[15].

Additionally, integration sites were found to coincide with DNase I hypersensitive sites in 18%, sites that have been assigned as regions available for the integration event[20] and associated with genomic instability, A characteristic of human cancers. Alu sequences were found in 9.5% of integration events, close to that reported by Zhu and colleagues (14%)[21].

As for regions rich in CG dinucleotides, only one integration site contained a CpG island, possibly because DNA methylation may act as a barrier to virus integration in these regions.[22] DNA methylation of the viral upstream regulatory region (URR) has been additionally associated with a latent and persistent infection[23].

There has been found a correlation between the extension of the viral quiescence, the DNA methylation, and the regulation of the viral genome expression in those viruses with the ability to maintain as a latent episome[24].

As a conclusion, our results suggest that many of the integration sites reported in the scientific literature are HPV 16 from squamous cell carcinomas (91%). 60% of these carcinomas belonged to the cervical anatomical sites and 30% to the oropharyngeal anatomical sites. It is also interesting to highlight that 50% of HPV16 integrates into transcriptional units and might affect the expression of target genes.

Ethical disclosure

Protection of human and animal subjects. This research do not use animal nor human material or data.

Confidentiality of data. Not applicable

Conflict of interest statement. This work was funded by authors, no was receivedexternal funding

Author contributions. Nicole Díaz Moreno designed the study, performed research, analyzed data, contributed new methods or models, and wrote the paper.

Julio César Osorio designed the study, performed research, analyzed data, contributed new methods or models, and wrote the paper.

Herney Andrés García-Perdomo designed the study, performed research, analyzed data, contributed new methods or models, and wrote the paper.

Andrés Castillo designed the study, performed research, analyzed data, contributed new methods or models, and wrote the paper.

Funding. Authors would like to thank Colciencias (Grant 110665741213)

References

1. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM HD. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996-1006. Genome Research. 2002. p. 996–1006.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-tieulent J, Jemal A. Global Cancer Statistics, 2012. CA a cancer J Clin. 2015;65(2):87–108.
3. Wheeler CM. The Natural History of Cervical Human Papillomavirus Infections and Cervical Cancer Gaps in Knowledge and Future Horizons. Obstet Gynecol Clin NA. Elsevier Inc; 2017;40(2):165–76.
4. Williams VM, Filippova M, Soto U, Duerksen-Hughes PJ. HPV-DNA integration and carcinogenesis: putative roles for inflammation and oxidative stress. Future Virol. 2011;6(1):45–57.
5. Senapati R, Senapati NN, Dwibedi B. Molecular mechanisms of HPV mediated neoplastic progression. Infect Agent Cancer. Infectious Agents and Cancer; 2016;11(1):59.
6. Hafner N, Driesch C, Gajda M, Jansen L, Kirchmayr R, Runnebaum IB, et al. Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. Oncogene. 2008;27(11):1610–7.
7. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome Res. 2014;24(2):185–99.
8. Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. Proc Natl Acad Sci. 2014 Oct;111(43):15544–9.
9. Wentzensen N, Vinokurova S, Doeberitz MVK, von KDM. Systematic Review of Genomic Integration Sites of Human Papillomavirus Genomes in Epithelial Dysplasia and Invasive Cancer of the Female Lower Genital Tract Systematic Review of Genomic Integration Sites of Human Papillomavirus Genomes in Epithelial Dyspl. Cancer Res. 2004;64:3878–84.
10. Bodelon C, Untereiner ME, Machiela MJ, Vinokurova S, Wentzensen N. Genomic characterization of viral integration sites in HPV-related cancers. 1.
11. el Awady MK, Kaplan JB, O'Brien SJ, Burk RD. Molecular analysis of integrated human papillomavirus 16 sequences in the cervical cancer cell line SiHa. Virology. 1987;159(2):389–98.
12. Wilson GA, Lechner M, Köferle A, Caren H, Butcher LM, Feber A, et al. Integrated virus-host methylome analysis in head and neck squamous cell carcinoma. Epigenetics. 2013;8(9):953–61.
13. Bauer-Hofmann R, Borghouts C, Auvinen E, Bourda E, Rösl F, Alonso a. Genomic cloning and characterization of the nonoccupied allele corresponding to the integration site of human papillomavirus type 16 DNA in the cervical cancer cell line SiHa. Virology. 1996;217(1):33–41.
14. Gallego MI, Zimonjic DB, Popescu NC, DiPaolo JA, Lazo PA. Integration site of human papillomavirus type-18 DNA in chromosome band 8q22.1 of C4-I cervical carcinoma. Genes Chrom Cancer. 1994;9:28–32.
15. Walline HM, Komarck CM, Mchugh JB, Bellile EL, Brenner JC, Prince ME, et al. Genomic Integration of High-Risk HPV Alters Gene Expression in Oropharyngeal Squamous Cell Carcinoma. Mol Cancer Res. 2016;14(10):941–52.
16. Vojtechova Z, Sabol I, Salakova M, Turek L, Grega M, Smahelova J, et al. Analysis of the integration of human papillomaviruses in head and neck tumours in relation to patients' prognosis. Int J Cancer. 2016 Jan;138(2):386–95.
17. Liu Y, Lu Z, Xu R, Ke Y. Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology. Oncotarget. 2016 Feb 2;7(5):5852.
18. Munoz N, Castellsagué X, de González AB, Gissmann L. HPV in the etiology of human cancer. Vaccine. 2006 Aug 21;24:S1-0.
19. Zhang R, Shen C, Zhao L, Wang J, McCrae M, Chen X, et al. Dysregulation of host cellular genes targeted by human papillomavirus (HPV) integration contributes to HPV-related cervical carcinogenesis. Int J Cancer. 2016;138(5):1163–74.
20. 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(3):1–11.
21. Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, 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. Nature Publishing Group; 2015;47(2):158–63.
22. Durzynska J, Lesniewicz K, Poreba E. Human papillomaviruses in epigenetic regulations. Mutat Res Mutat Res [Internet]. 2017 [cited 2017 Jul 23];772:36–50. Available from: http://www.sciencedirect.com/science/article/pii/S1383574216301107
23. Steenbergen RDM, Snijders PJF, Heideman D a M, Meijer CJLM. Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. Nat Rev Cancer [Internet]. Nature Publishing Group; 2014;14(6):395–405. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24854082
24. Piyathilake CJ, Macaluso M, Alvarez RD, Chen M, Badiga S, Edberg JC, et al. A higher degree of methylation of the HPV 16 E6 gene is associated with a lower likelihood of being diagnosed with cervical intraepithelial neoplasia. Cancer [Internet]. NIH Public Access; 2011 Mar 1 [cited 2017 Jul 23];117(5):957–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20945322