RESEARCH ARTICLE

Mapping of microRNAs related to cervical cancer in Latin American human genomic variants [version 1; peer review: 2 approved with reservations]

Milena Guerrero Flórez ^{1,2}, Olivia Alexandra Guerrero Gómez^{1,2},
Jaqueline Mena Huertas^{1,2}, María Clara Yépez Chamorro ¹

¹Department of Biology, Center for Health Studies at the University of Nariño (CESUN), University of Nariño, Pasto, Nariño, Colombia

²Department of Biology, University of Nariño, Pasto, Nariño, Colombia

V1 First published: 20 Jun 2017, 6:946
<https://doi.org/10.12688/f1000research.10138.1>
 Latest published: 05 Dec 2018, 6:946
<https://doi.org/10.12688/f1000research.10138.2>

Abstract

Background: MicroRNAs are related to human cancers, including cervical cancer (CC), which is mainly caused by human papillomavirus (HPV) infection. In 2012, approximately 70000 cases and 28000 deaths from this cancer were registered in Latin America according to GLOBOCAN reports. The most frequent genotype worldwide is HPV-16. The main molecular mechanism of HPV in CC is related to integration of viral DNA into the hosts' genome. However, the different variants in the human genome can result in different integration mechanisms, specifically involving microRNAs (miRNAs).

Methods: miRNA sequences associated with CC and four human genome variants from Latin American populations were obtained from miRBase and the 1000 Genomes Browser, respectively. HPV integration sites near cell cycle regulatory genes were identified. miRNAs were mapped on human genomic variants. miRSNPs (single nucleotide polymorphisms in miRNAs) were identified in the miRNA sequences located at HPV integration sites on the human genomic Latin American variants.

Results: Two hundred seventy-two miRNAs associated with CC were identified in 139 reports from different geographic locations. By mapping with the Blast-Like Alignment Tool (BLAT), 2028 binding sites were identified from these miRNAs on the human genome (version GRCh38/hg38); 42 miRNAs were located on unique integration sites; and miR-5095, miR-548c-5p and miR-548d-5p were involved with multiple genes related to the cell cycle. Thirty-seven miRNAs were mapped on the human Latin American genomic variants (PUR, MXL, CLM and PEL), but only miR-11-3p, miR-31-3p, miR-107, miR-133a-3p, miR-133a-5p, miR-133b, miR-215-5p, miR-491-3p, miR-548d-5p and miR-944 were conserved.

Conclusions: 10 miRNAs were conserved in the four human genome variants, and in the remaining 27 miRNAs, substitutions, deletions or

Open Peer Review

Approval Status

	1	2
version 2		
(revision)		
05 Dec 2018	view	
version 1		
20 Jun 2017	view	view

1. **Juan Manuel Anzola** , Corporación
CorpoGen, Bogotá, Colombia

2. **Subhash Mohan Agarwal**, ICMR-National
Institute of Cancer Prevention and Research,
Noida, India

Any reports and responses or comments on the article can be found at the end of the article.

insertions were observed in the nucleotide sequences. This variability can imply differentiated mechanisms towards each genomic variant in human populations, relative to specific genomic patterns and geographic features. These findings may be decisive in determining susceptibility to the development of CC. Further identification of cellular genes and signalling pathways involved in CC progression could lead to the development of new therapeutic strategies based on miRNAs.

Keywords

cervical cancer, HPV, HPV integration sites, microRNAs, miRNAs, secondary structure, human genome variants, bioinformatics tools

Corresponding author: Milena Guerrero Flórez (mlenague@udenar.edu.co)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2017 Guerrero Flórez M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Guerrero Flórez M, Guerrero Gómez OA, Mena Huertas J and Yépez Chamorro MC. **Mapping of microRNAs related to cervical cancer in Latin American human genomic variants [version 1; peer review: 2 approved with reservations]** F1000Research 2017, 6:946 <https://doi.org/10.12688/f1000research.10138.1>

First published: 20 Jun 2017, 6:946 <https://doi.org/10.12688/f1000research.10138.1>

Introduction

Cervical cancer (CC) is the second most common malignancy in women worldwide. According to GLOBOCAN reports, approximately 530,000 women are diagnosed with CC and 265,672 die from it each year¹. Infection by human papillomavirus (HPV) has been recognized as the major risk factor in this pathology^{2,3}, but the virus presence is not the main cause for the development of this cancer^{4,5}. Viral DNA integration into the host cell genome is considered a conducive factor for cervical intraepithelial neoplasia (CIN) to develop into CC⁵⁻⁷.

Numerous microRNAs (miRNAs) have been identified in proximity to HPV integration sites^{8,9}. miRNAs are a class of small (18 to 26 nucleotides length), noncoding, evolutionarily conserved RNAs that are processed from longer transcripts known as pre-miRNAs (60 to 100 nucleotides in length)¹⁰. They are located on regions known as fragile sites and distributed in intergenic, intronic and exonic segments of the human genome involved in cancer^{11,12}. Functionally, they regulate post-transcriptional expression levels of up to 60% of total protein-encoding genes by binding their seed sequences (2–8 nucleotides length). The 5'-UTR end of the miRNA seed sequence is complementary to the 3'-UTR end of the target mRNAs¹³. This recognition event can affect the expression of important regulatory genes. Deregulation of genes such as tumour suppressor genes and oncogenes can lead to cancer development, including CC¹⁴⁻¹⁶.

Human genome variants generate different patterns of miRNA deregulation¹⁷, which can contribute to cancer development susceptibility, treatment efficacy and patient prognosis¹⁸⁻²⁰. 99% of the human genome is genetically identical, and the remaining 1% is responsible for all human diversity. miRNAs represent a major part of this genetic variation²¹. miRSNPs (single nucleotide polymorphisms in miRNAs) are human polymorphisms at or near predicted miRNA target sites²². The occurrence of miRSNPs can influence miRNA functionality on all levels, including transcription, maturation, and mRNA target binding.

Knowledge on miRNAs related to CC development in human genome variants from Latin American populations is scarce. Thus, in this study, we mapped miRNAs associated with CC in human genome variants obtained from Colombia, Mexico, Peru and Puerto Rico. Complete genomes were included in this study. Additionally, the relationships between HPV integration sites, genes close to these sites, mapping profiles and mutation patterns for each of the miRNAs were estimated for each of the genome sequences. The objective of this research was to analyse how genetic variation of CC-associated miRNAs identified in previously reported HPV integration sites affects cell cycle regulatory genes in human genomic variants from Latin America.

Methods

miRNA sequences associated with cervical cancer

Two hundred and seventy-two miRNAs associated with CC were selected as described in the systematic review published by Guerrero & Guerrero²³. With the information contained in miRBase²⁴⁻²⁶, miRNAmap²⁷ and miRNAstart, features, such as length, chromosomal and genomic location of pre-miRNAs and mature miRNAs, were analysed. The mature miRNA reference

sequences were obtained in FASTA format from the miRBase database (Dataset 1²⁸).

Latin American human genomic variants

Four human genome sequences were obtained from randomly selected female participants in the 1000 Genomes Project from Latin American populations^{22,29}. Their codes were CLM (from Medellin in Colombia), MXL (from Los Angeles and of Mexican ancestry in the USA), PEL (from Lima in Peru) and PUR (from Puerto Rico). The control sequence was a variant that is phylogenetically distant to Latin American variants and identified with the code BEB (from Bangladesh and of Bengali ancestry). Access codes were obtained from the 1000 Genomes Project resources^{21,30}. This information is summarized in Table 1.

Selection, identification and analysis of HPV integration sites near cell cycle regulatory genes

Viral insertion sites and nearby genes on the human genome were identified with the UCSC Genome Bioinformatics search engine^{31,32}. To establish possible functional relationships with the development of CC, functional information on the associated functions of these human genes was obtained from UniProt^{33,34}.

Mapping miRNAs and chromosomal locations on the human genome

According to Xia *et al.*³⁵, the mature miRNA sequences are located in regions with pre-miRNA secondary structure complementarity (3' and 5'). In total, 445 miRNA sequences were analysed. The Blast-Like Alignment Tool (BLAT) available on the UCSC Genome Bioinformatics website was used for mapping the miRNAs associated with the full human genome with the following parameters: (a) genome, human; (b) assembly, Dec. 2013 (GRCh38/hg38); (c) query type, DNA; (d) sort output, query; and (e) score and output, hyperlinks. A matrix of chromosomal location data was built with Microsoft Excel 2013 ('Matrix of data' in Dataset 2³⁶). From this matrix, the miRNAs over HPV integration sites were manually identified.

Identification of miRNAs in Latin American human genomic variants

To identify miRNA mutations in the four Latin American human genome variants, the available tools, including ideogram view, subjects and exon navigator, in the NCBI 1000 Genomes Browser (Phase 3, version 3.7) were used. The code for each female genetic

Table 1. Access codes of the four Latin American human genome variants obtained from the NCBI 1000 genomes project.

SEQUENCE TYPE	SEQUENCE NAME	DATABASE	ACCESS CODE
Genomic sequence	CLM	NCBI 1000 Genomes Project	HG01432
	MXL		NA19749
	PEL		HG01566
	PUR		HG00554
	BEB (Control)		HG03589

variant selection (Colombia, Mexico, Peru, Puerto Rico and Bangladesh) was inserted and the sequence of each miRNA identified in viral integration sites was introduced and the mapped nucleotide positions were selected. Using WebLogo 3³⁷, logos were created to view the nucleotide differences. The bioinformatics workflow is summarized in Figure 1.

Dataset 1. The mature miRNA reference sequences were obtained in FASTA format from the miRBase database

<http://dx.doi.org/10.5256/f1000research.10138.d164732>

Dataset 2. Matrix of data containing all the necessary components for the validation of data on CC-associated miRNAs in HPV integration sites in Latin American human genomic variants

<http://dx.doi.org/10.5256/f1000research.10138.d164736>

Results

HPV integration sites and chromosomal distribution

A total of 44 publications were identified between 1987 and 2015 related to HPV integration sites in the human genome. The most frequent types of HPV associated with CC were HPV-16 and HPV-18. Details of these articles are outlined in Supplementary File 1. Five hundred and seventy-eight integration sites for 8 types

of HPV associated with different histological cervical conditions were identified, of which 63.84% were HPV-16 (Figure 2 and 'HPV integration sites' in Dataset 2³⁶).

HPV-16 and HPV-18 have integration sites on all human chromosomes. HPV-16 has more integration sites on chromosomes 2, 1, 3, 6, 9, 5, 8 and 4, while HPV-18 has more on chromosomes 2, 1, 8, 12, 5, 10, 4, 6 and 9. Some less frequently oncogenic HPV types have integration sites on specific chromosomes, such as HPV-45 on 2, 1, 3, 9, 4, 7 and 13; HPV-33 on 9, 13, 5, 6, 8, 11, 16, 18 and X; HPV-58 on 4, 12 and 18; HPV-31 on 2 and 17; HPV-67 on 4 and 13; and HPV-68 on chromosome 18. Chromosomes 1 and 2 displayed a higher number of viral insertion sites (41 and 45, respectively), while chromosomes 13 and 18 displayed insertion sites for 5 different HPV genotypes. The chromosomal loci with the highest numbers of HPV integration sites are presented in Table 2.

Analysis of HPV integration sites near cell cycle regulatory genes

Information on the associated functions of genes located near HPV integration sites obtained from UniProt showed that 86.1% of the genes located in close proximity were involved in apoptosis, cell adhesion, cell differentiation, ion transport and metabolic processes. Fifty-four genes were involved in direct regulation of the cell cycle. Twenty-six of these were tumour suppressor genes, 8 were oncogenes, 8 were proto-oncogenes and 12 did not have a determined functionality in the development of this neoplasia (Figure 3).

Figure 1. Bioinformatic workflow for mapping of miRNAs related to CC on Latin American human genomic variants.

Figure 2. Chromosomal distribution of integration sites of HPV types (HPV 16, 18, 31, 33, 45, 58, 67 and 68) most frequently reported in the literature.

Table 2. Chromosomal loci with the highest numbers of HPV integration sites.

CHROMOSOMAL LOCUS	HPV INTEGRATION SITES	HPV TYPES
8q24.21	23	16, 18, 45
3q28 y 13q22.1	9	16, 18, 45
4q13.3	7	16, 45
2q34	6	16, 18
2q22.3 y 20p12.1	5	16, 18
13q21 y 17q12	5	16

Figure 3. Functional classification of cellular genes in HPV integration sites.

Mapping miRNAs associated with cervical cancer

The 2028 miRNA binding sites associated with CC in the human genome were identified from BLAT mapping using previously identified miRNAs²³, including 432 sites previously reported in miRBase ('Results of mapping with BLAT' in [Dataset 2³⁶](#)). These sites were located on both DNA strands (52.97% on the positive strand and 47.03% on the negative strand). 1881 binding sites were fully complementary (100% sequence identity) to miRNA sequences, while 1, 24, and 122 binding sites had 96.2%, 95.7% and 95.5% sequence identity, respectively.

miR-5095 was mapped onto 853 binding sites on 23 chromosomes. Four hundred and twenty-four mature miRNAs sequences (98.15%) mapped to one, two, three and even ten different binding sites. miR-522-5p and miR-523-5p binding sites mapped only a single chromosome (Chr. 19). [Table 3](#) shows the chromosomal location and number of binding sites for each specific miRNA associated with CC.

The distribution of the 2028 binding sites was not homogeneous along the human genome. 41% of the total binding sites were identified on chromosomes 1, 19, 5, 2, 3, 14, 7 and X. Although the number of miRNA binding sites correlated with the size of each

chromosome, some short chromosomes, such as 19 and X, had more miRNA binding sites when compared to other larger chromosomes ([Table 4](#)).

14.89% (302) of binding sites grouped into the following 19 specific chromosomal locations: (1) 19q13.42 (51 sites/14 miRNAs), (2) 14q32.31 (34 sites/16 miRNAs), (3) 13q31.3 (16 sites/11 miRNAs), (4) 14q32.2 (16 sites/9 miRNAs), (5) 4q25 (16 sites/7 miRNAs), (6) 20q13.33 (15 sites/7 miRNAs), (7) 16p13.3 (15 sites/4 miRNAs), (8) Xq26.2 (14 sites/8 miRNAs), (9) 7q22.1 (14 sites/6 miRNAs) and (10) 1p31.3 (14 sites/6 miRNAs). The remaining 9 chromosomal locations contained between 10 and 13 binding sites ([Supplementary File 2](#)). 92% (1865/2028) of the binding sites were distributed into 250 groups along the human genome; the remaining 8% (163/2028) of binding sites for various miRNAs including miR-5095 were distributed along the human genome without being distributed into any groups.

Each group contains between 2 and 7 miRNA binding sites, although some groups contain between 8 and 16 ([Figure 4](#)). The majority of the groups are located on chromosomes 1, 2, 3, 5, 10 and 11. The biggest groups are located on chromosome 19, with 51 binding sites for 25 miRNAs involved in CC development.

Table 3. Chromosomal location and frequency of miRNA binding sites associated with CC.

miRNA ASSOCIATED WITH CC	miRNAs BINDING SITES	CHROMOSOMAL LOCATION
hsa-miR-5095	853	All chromosome
hsa-miR-548c-5p	194	All, except 9
hsa-miR-548d-5p	188	All, except X, Y
hsa-miR-548b-5p	87	All, except 3, 4, 5, 6, X, Y
hsa-miR-574-5p	62	All, except 16, 21, Y
hsa-miR-576-3p	15	4, 5, 8, 9, 12, 13, 15, 18, 22, X
hsa-miR-548c-3p	13	2, 4, 5, 7, 8, 13, 14, X, Y
hsa-miR-1273g-5p	11	1, 3, 7, 9, 10, 11, 13, 14, 15
hsa-miR-95-5p	10	1, 2, 4, 6, 7, 13, X
hsa-miR-1244	9	2, 3, 5, 7, 12, 13, 14, 20
hsa-miR-545-3p	8	3, 5, 7, 10, 12, X
hsa-miR-378a-3p	7	3, 5, 10, 11, 14, 17, 18
hsa-miR-522-5p, -523-5p	7	19
hsa-miR-518f-5p	6	5, 19
hsa-miR-545-5p	6	2, 3, 5, 14, 17, X
hsa-miR-151a-5p	5	1, 4, 8, 19, X
hsa-miR-339-5p	5	5, 7, 20, 22
hsa-miR-603	4	10, 13, 14, 16
hsa-miR-7-5p	4	9, 10, 15, 19
hsa-miR-584-5p	4	4, 5, 9, 19

Table 4. Distribution of binding sites in chromosomes identified in miRNAs associated with CC. CHR= Chromosome.

CHR.	NUMBER OF miRNA BINDING SITES	(%)
1	175	8.63
2	108	5.33
3	106	5.23
4	89	4.39
5	111	5.47
6	87	4.29
7	103	5.08
8	81	3.99
9	79	3.90
10	92	4.54
11	93	4.59
12	93	4.59
13	71	3.50
14	106	5.23
15	66	3.25
16	81	3.99
17	94	4.64
18	57	2.81
19	131	6.46
20	42	2.07
21	27	1.33
22	29	1.43
X	100	4.93

58.8% of miRNA binding sites associated with CC (1194 binding sites) are located in intergenic regions, 39.65% (804 binding sites) in intronic regions, 1.28% (26 binding sites) in exonic regions and 0.19% (4 binding sites) between intronic and exonic regions (mixed miRNAs). [Figure 5](#) shows the variation in the number of intergenic, exonic and intronic miRNAs associated with CC.

miRNA identification in selected HPV integration sites

Thirty-eight integration sites were found for six types of oncogenic HPV (HPV-16, -18, -33, -45, -58 and -68) in miRNA binding sites and cell cycle regulatory genes associated with CC ([Table 5](#)). The largest number of HPV integration sites was found for miR-5095 (33 sites), followed by miR-548c-5p (11 sites) and miR-548d-5p (11 sites) ([Table 5](#)). In 14 integration sites, no miRNA binding sites were detected. The highest number of miRNA binding sites was found in chromosome regions 18q11.2 and 19p13.12 ([Supplementary File 2](#)).

Ninety-six possible interactions were identified between 37 mature miRNAs associated with CC and 42 cell cycle regulatory genes located in proximity to the viral insertion sites. The network of interactions is presented in [Figure 6](#). 35.42% of the interactions involved miR-5095, 12.5% involved miR-548c-5p and 12.5% miR-548d-5p.

38.1% of genes identified in HPV integration sites have binding sites for a single miRNA, and 61.9% have binding sites for more than two miRNAs. [Table 6](#) displays genes with more than five miRNA binding sites.

A gene may have binding sites for both regions of complementarity (3' and 5') of a miRNA³⁸. In this study, we found that the TTC39C gene has binding sites for miR-133a-3p and miR-133a-5p and MAP3K1 has binding sites for miR-449b-3p and miR-449b-5p, though some mature sequences from one miRNA also showed binding sites to different genes ([Figure 6](#)). As an example, the

Figure 4. Chromosomal distribution of groups of identified miRNA binding sites.

Figure 5. Variation in the number of intergenic, exonic and intronic miRNAs associated with cervical cancer.

Table 5. miRNAs in HPV integration sites and their correlation with cell cycle regulatory genes.

HPV TYPES	HPV INTEGRATION SITES	miRNAs PRESENT AT SITES OF INTEGRATION OF HPV ¹	CELLULAR GENES ²	CL. ³
18	1p22.2	miR-548c-5p (-)	<i>CDC7</i> (+)	--
18	1p31.2	-	<i>GADD45A</i> (+)	ST
16	1p34.1	-	<i>PLK3</i> (+)	--
16	1p34.3	miR-5095 (3; -, -, +), -548b-5p (-), -548c-5p (2, -, -), -548d-5p (-)	<i>CDCA8</i> (+)	OG
16	1q25	-	<i>TPR</i> (-)	--
16	1q36.32	-	<i>TP73</i> (+)	ST
16, 18	1q41	miR-5095 (2, +, +), -194-5p (-), -215-3p (-), -215-5p (-), -548b-5p (-)	<i>PROX1</i> (+)	ST
18	2p15	miR-5095 (-)	<i>XPO1</i> (-)	ST
16	2q33.1	miR-152-5p(-), -548d-5p(-)	<i>ORC2</i> (-)	--
			<i>BZW1</i> (+)	--
16	2q33.3	miR-5095 (+)	<i>PARD3B</i> (+)	ST
16	2q34	miR-5095 (-)	<i>BARD1</i> (-)	ST
16	3p21.31	miR-5095 (3; -, -, +), -191-3p (-), -191-5p (-), -425-3p (-), -425-5p (-)	<i>MAP4</i> (-)	--
16	3q26.33	miR-5095 (2; -, +)	<i>SOX2</i> (+)	OG
16	3q28	miR-5095 (-), -944 (+), -28-3p (+), -28-5p (+)	<i>P3H2</i> (-)	ST
			<i>TP63</i> (+)	ST
16, 45	4q13.3	-	<i>CXCL8</i> (+)	PO
16	4q23	-	<i>EIF4E</i> (-)	OG
16	4q31.21	miR-548c-5p (+)	<i>FBXW7</i> (-)	ST
16	5q11.2	miR-5095 (3; -, -, +), -449a (-), -449b-3p (-), -449b-5p (-), -548c-3p (+), -548d-5p (+), -581 (-)	<i>MAP3K1</i> (+)	ST
16	5q31.1	miR-5095 (-)	<i>PPP2CA</i> (-)	ST
16	6p21.31	miR-5095 (+)	<i>BAK1</i> (-)	ST

HPV TYPES	HPV INTEGRATION SITES	miRNAs PRESENT AT SITES OF INTEGRATION OF HPV ¹	CELLULAR GENES ²	CL. ³
16	6p22.3	miR-5095 (4; -, -, +, +), -548c-5p (+), -548d-5p (2; +, +)	<i>ID4</i> (+)	ST
16	6q22.32	-	<i>CENPW</i> (+)	--
16	6q23.3	miR-5095 (3; -, +, +)	<i>CITED2</i> (-)	ST
16	7p21.1	-	<i>AHR</i> (+)	ST
18	7q36.2	miR-5095 (-)	<i>RHEB</i> (-)	PO
18	8q21.2	-	<i>E2F5</i> (+)	--
16, 18	8q21.3	-	<i>NBN</i> (-)	ST
16, 18, 45	8q24.21	miR-5095 (-), -548d-5p (-)	<i>MYC</i> (+)	PO
16	8q24.21	miR-5095 (-), -548d-5p (-)	<i>PVT1</i> (+)	OG
18	9p21.3	miR-5095 (+), -31-3p (-), -31-5p (-), -491-3p (+), -491-5p (+)	<i>CDKN2A</i> (-)	ST
16	9q22.2	miR-5095 (+), -576-3p (2; +, +)	<i>CKS2</i> (+)	OG
16, 18	10q23.31	miR-5095 (-), -107 (-), -103a-3p (-), -548b-5p (2; -, -), -548d-5p (2; -, -)	<i>PTEN</i> (+)	ST
16	10q24.2	miR-5095 (-), -1287-5p (-)	<i>MARVELD1</i> (+)	ST
16	12q14.3	miR-574-5p (-)	<i>CDK4</i> (-)	OG
			<i>MDM2</i> (+)	OG
18	12q15	-	<i>HMGA2</i> (+)	PO
58	12q24.33	-	<i>ZNF268</i> (+)	ST
18	14q11.2	miR-5095 (+), -548c-3p (+), -574-5p (+)	<i>HAUS4</i> (-)	--
18, 45	14q24.1	miR-5095 (2, -, +), -548c-5p (+)	<i>RAD51B</i> (+)	ST
18	15q21.3	miR-5095 (2; -, +), -574-5p (-)	<i>CCNB2</i> (+)	PO
16	16p13.3	miR-5095 (12; (7 -, 5+,)), -548c-5p (+), -572 (-), -940 (+)	<i>TSC2</i> (+)	ST
16	17q21.31	miR-5095 (3; -, +, +)	<i>BRCA1</i> (-)	ST
33	18q11.2	miR-5095 (-), -1-3p (-), -133a-3p, -133a-5p (-), -133b, -378a-3p (+), -548b-5p (-), -548d-5p (-)	<i>TTC39C</i> (+)	--
68	18q21.1	miR-5095 (3; -, +, +), -548c-5p (+), -548d-5p (+), -574-5p (+)	<i>ZBTB7C</i> (-)	ST
18	18q21.33	miR-5095 (-), -548b-5p (+), -548c-5p (-), -548d-5p (+)	<i>BCL2</i> (-)	PO
16	19p13.12	miR-5095 (-), -23a-3p (-), -23a-5p (-), -27a-3p (-), -27a-5p (-), -181c-3p (+), -181c-5p (+), -584-5p (+)	<i>NANOS3</i> (+)	--
16	20q11.21	-	<i>TPX2</i> (+)	ST
16	20q13.2	miR-5095 (-)	<i>SRC</i> (+)	PO
16	21q22.13	miR-5095(+), -548d-5p (-)	<i>DYRK1A</i> (+)	--
16	22q12.1	miR-548c-5p (+)	<i>CHEK2</i> (-)	ST
16, 18, 45	22q13.1	miR-5095 (2, -, -)	<i>MCM5</i> (+)	PO
16	Xq25	miR-5095 (-), -574-5p (-)	<i>DCAF12L2</i> (-)	OG

¹The information in brackets shows the number of miRNA binding sites, and whether the miRNAs are located on the positive sense or negative sense DNA strand.

²The information in brackets shows and whether the cell cycle regulatory genes are located on the positive sense or negative sense DNA

³Cl: Classification of cellular genes; ST: tumor suppressors; OG: Oncogenes; PO: Proto-oncogenes.

Figure 6. The network of interactions between cervical cancer-associated miRNAs and cell cycle regulatory genes present at HPV integration sites. Rectangles of various colors represent the cell cycle regulatory genes, and color depend on their classification (ST - ■, OG - ■, POG - ■ e IND - ■). The arrows represent the interactions between miRNAs and genes involved in cell cycle regulation; each arrow's color depends on the DNA chain where miRNAs and cell cycle regulatory genes are located.

Table 6. Genes associated to 5 or more binding sites of miRNAs.

NUMBER OF miRNA BINDING SITES	miRNAs	GENE
5 sites	miR-103a-3p, -107, -548b-5p, -548d-5p and -5095	<i>PTEN</i>
	miR-194-5p, -215-3p, -215-5p, -548b-5p and -5095	<i>PROX1</i>
7 sites	miR-449a, -449b-3p, -449b-5p, -548c-3p, -548d-5p, -581 and -5095	<i>MAP3K1</i>
8 sites	miR-1-3p, -133a-3p, -133a-5p, -133b, -378a-3p, -548b-5p, -548d-5p and -5095	<i>TTC39C</i>
	miR-23a-3p, -23a-5p, -27a-3p, -27a-5p, -181c-3p, -181c-5p, -584-5p and -5095	<i>NANOS3</i>

miR-548c-3p mature chain has binding sites in the HAUS4 gene as well as in the MAP3K1, CDCA8, BCL2, ID4, cMYC, RAD51B, TSC2, ZBTB7C, FBXW7, CHEK2 and CDC7 genes (Figure 6).

Identification of miRNAs on Latin American human genomic variants

26.31% (10/42) of the miRNAs analysed (miR-11-3p, miR-31-3p, miR-107, miR-133a-3p, miR-133a-5p, miR-133b, miR-215-5p, miR-491-3p, miR-548d-5p and miR-944) were identical across the Latin American human genome variants, and 73.69% showed a genetic mutation (substitution or deletion of nucleotides) (Figure 7, Panels A and B).

When mapping the sequences of these miRNAs to the selected Latin American human genome variants (Supplementary File 3), 88 miRSNPs related to miRNAs or miRNA binding sites were identified on the Latin American variants compared with 33 on the reference variant. Twenty-one miRSNPs were located in the miRNA seed sequences of Latin American variants compared with 3 located in the reference variant. The most representative mapping results are shown in Table 6.

Types of nucleotide substitutions in the miRNA sequences associated with CC in the selected human genome variants showed that there were more frequent transversions than transitions and that the most frequent nucleotide substitutions were G→U (16.9%), followed by A→C (15.7%), C→A (15.7%) and G→A (10.8%) (Figure 7).

Between one and 18 nucleotide deletions were detected in miR-27a-3p, miR-31-5p, miR-103a-3p, miR-191-3p, miR-215-3p and miR-574. The sequences of miR-28, miR-152, miR-548c-5p, miR-572 and miR-5095 only mapped to reference sequences (version GRCh38/hg38), but not to any of the Latin American genomic variants. miR-152 did not map to the PUR variant (Table 6).

Table 7 displays the nucleotide variations from human genome variants obtained from Colombia, Mexico, Peru and Puerto Rico and Bangladesh, which was the control variant.

Discussion

HPV integration sites

According to the literature, approximately 570 integration sites have been identified for eight oncogenic HPV types associated with CC (Figure 2). HPV integration into cellular DNA and consequent deregulation of genes is considered a crucial step in cancer progression. Genotype HPV-16 is the most studied for its relationship with CC, as it is responsible for 70% of cases worldwide³⁹. This could be a consequence of the greater proportion of integration sites reported for this genotype. In contrast, low risk genotypes, such as HPV-45, -66 and -93 reported in Colombia, are frequent in CC^{40–44}.

HPV integration into the host genome occurs in regions well-known as fragile sites, breakpoints or transcriptionally active regions⁴⁵. This integration induces functional alterations of cellular genes in close proximity^{12,46–48}. According to our results, the 8q24.21

chromosome region is the most affected by HPV integration. If we take into account that proto-oncogenes such as the MYC gene are located here⁴⁹ (as displayed in Figure 3) and that MYC represents a family of genes overexpressed in several tumours including CC^{49–51}, inhibition of MYC expression can induce cancer cell destruction⁵⁰. In this context, the MYC gene could be both a tumour biomarker and potential treatment target for several tumours⁵¹ (Table 2).

Chromosomes 1, 14, 19 and X contain significantly more mature miRNAs than others, and chromosome 18 contains fewer miRNAs. The 19q13.4 chromosome region contains the largest group of human miRNAs (known as the group of miRNAs on chromosome 19 “C19 MC”), with alterations in several that have been previously reported in cancer⁵². Studies have reported associations between chromosome 1 and malignant transformation in cancers, including CC⁵³.

The 578 integration sites identified in eight HPV types associated with CC were located in cell cycle regulatory genes, including the tumour suppressor genes TP73, P3H2, TP63, NBN, PTEN, BRCA1, and TPX2; the oncogenes EIF4E, CDCA8, MDM2, and PVT1; and the proto-oncogenes SRC, MYC, MCM5, CXCL8, and BCL2. Their deregulation could explain the progression of CC (Figure 3).

miRNA binding sites associated with cervical cancer

In 2011, Reshmi *et al.* used BLAT to determine the exact location of four miRNA binding sites associated with CC using bioinformatics programmes and computational tools⁵⁴. To the best of our knowledge, this study is the first to use BLAT to identify miRNA binding sites in proximity to HPV integration sites involved in CC progression. In this study, 2028 binding sites from 272 CC-associated miRNAs were identified.

Identification of the target mRNAs of these miRNAs is considered a key step in their structural and functional analysis to establish possible interactions and consequently, cellular processes that may be altered in CC progression^{55–57}. miRNAs located in the two strands of cellular DNA (5' and 3' strands) demonstrate their ability to interact in both orientations with the two strands of DNA and form triple helix structures to enhance RNA stability^{58,59}.

Each CC-associated miRNA showed a different number of binding sites in the human genome (Table 3, Supplementary File 2), and in the human genomic variants^{17,21,60,61}; miRNAs were distributed throughout the genomes in both intronic or exonic regions¹³. In this study, CC-associated miRNAs were distributed in the karyosome, with chromosomes 1, 19, 5, 2, 3, 14, 7 and X having the largest number of miRNA binding sites (Table 4). These results are consistent with those reported by Calin *et al.*¹². Because some chromosomes have a greater number of miRNA binding sites, it provides evidence of a non-random distribution of miRNAs within the chromosomes.

Our results showed a low number of exonic miRNAs. These exonic miRNAs are considered rare miRNAs⁶², which are important candidates for gaining a better comprehension of interaction networks between miRNAs and their CC-associated targets.

Figure 7. **A)** Number of miRNAs and nucleotide substitutions found in each human genomic variant; **B)** Number of miRNAs with between 1 and 7 nucleotide substitutions; **C)** Number of miRNAs with nucleotide substitutions in one, two or three genomic variants in the Latin American human genome, and **D)** Types of nucleotide substitutions in the miRNA sequences associated with CC in the selected human genome variants.

Table 7. miRNAs identified in HPV integration sites, displaying the nucleotide variations in the selected Latin American human genome variants and the control variant. More data is available in [Supplementary File 3](#).

HG ¹	miRNAs IDENTIFIED IN HPV INTEGRATION SITES (Cromosomal location (Chain)) ²	
	hsa-mir-1-3p (18q11.2 (-))	hsa-mir-23a-3p (19p13.12 (-))
CLM	UGGAAUGUAAAGAAGUAUGUAU	AUCACAUUGCCAGGGGAUUUCC
MXL	UGGAAUGUAAAGAAGUAUGUAU	AUCACAUUGCCAGGGGAUUUCC
PEL	UGGAAUGUAAAGAAGUAUGUAU	AUAACAUUGCAAGGGGAUUUCC
PUR	UGGAAUGUAAAGAAGUAUGUAU	AUCACAUUGCCAGGGGAUUUCC
BEB	UGGAAUGUAAAGAAGUAUGUAU	AUCACAUCCAGGGGAUUUCC

	Conserved	Nucleotide substitution
	hsa-mir-31-5p (9p21.3 (-))	hsa-mir-152 (17q21.32 (-))
CLM	AGGCAAGAUGCUGGCAUAGCU	CGGGUCUGUGCUACACUCCGACU
MXL	AGGCAAGAUGCUGGCAUAGCU	CGACU
PEL	AGGCAAGAUGCUGGCAU	AGGUUCUGUAUACACUACGACU
PUR	AGGCAAGAUGCUGGCAUAGCU	AGGUUCUGUUGUGCACUCUGACU
BEB	AGGCAAGAUGCUGGCAUAGCU

	Nucleotide deletion	Absence of the miRNA sequence

¹HG: Human genome; CLM: variant of Medellin, Colombia; MXL: Los Angeles with Mexican ancestry; PEL: Lima, Peru; PUR: of Puerto Rico; BEB: Bengali, Bangladesh.

²The size of each letter indicates the enrichment of each nucleotide in Latin American variants of the human genome, displayed through WebLogo.

The miRNA binding sites are within a short distance of each other in the chromosome, indicating that they tend to cluster^{63–66}. Altuvia *et al.* reported miRNAs in groups of two or three⁶⁴. This coincides with our results on CC-associated miRNA binding sites, as we found that miRNAs are capable of forming groups of more than 6 miRNAs on both strands of human DNA (Figure 4). We identified an important group of 16 miRNAs that can form these clusters and are located on chromosome 14 region 14q32.31. They include hsa-miR-134, miR-299, miR-323a, miR-329, miR-376a, miR-376c, miR-379, miR-411, miR-485, miR-487a, miR-487b, miR-494, miR-495, miR-539, miR-654 and miR-5095 (Supplementary File 2). Understanding their individual and collective roles is important when studying the development of this neoplasia.

miR-5095 had the highest number of binding sites distributed throughout the human genome (Table 3), which is in accordance with previously reported data^{66–68} where approximately 900 binding sites were identified; they are probably related to the expression of many target mRNAs and biological processes. Based on its extensive genomic distribution and low specificity in CC, miR-5095 is a good candidate to be used as an indicator of genetic variability within the human population.

miRNAs located in HPV integration sites

To identify the role of miRNAs, HPV integration sites located in cell cycle-controlling genes were analysed. Thirty-seven miRNAs were identified in HPV integration sites close to cell cycle-controlling genes (Table 5). Nambaru *et al.* and Schmitz *et al.* identified numerous miRNAs in the proximity of HPV integration sites and reported that approximately 65% of these were involved in cervical carcinogenesis^{8,9}. Inactivation of tumour suppressor genes by viral integration increases genomic instability and leads to cervical malignant neoplasm progression⁶⁹.

The multiple miRNA binding sites on a target may decrease the levels of mRNA translation and improve the specificity of gene regulation. For example, one miRNA can have multiple target genes and each individual mRNA can be regulated by numerous miRNAs^{13,70,71}. Ninety-seven interactions were identified between miRNAs and cell cycle regulatory genes (Table 4–Table 5, Figure 4–Figure 6); miR-5095, -548c-5p and -548d-5p showed the highest number of interactions with these kinds of genes.

Ivashchenko *et al.* identified miR-5095 binding sites in the BRCA1 gene⁶⁷. In this study, miR-5095 was also found to have binding sites

in the BAK1, BARD1, CITED2, MDM5, SRC, PARD3B, PPP2CA, RHEB, SOX2 and XPO1 genes (Table 5 and Figure 6). Our findings provide a basis for searching for other interactions, gene targets, and CC-associated miRNAs.

During miRNA biogenesis, each pre-miRNA produces two mature miRNAs, such as miRNA-5p and miRNA-3p⁷². Mature miRNA deregulation can have an important role in tumour development, suggesting the need to analyse each mature sequence (miRNA-5p and -3p). In this study, binding sites were analysed for both mature miRNA sequences (-5p and -3p) in several interactions (Figure 6). A mature miRNA sequence, such as miR-548c, demonstrated binding sites in different cellular genes. Thus, this miRNA could serve as candidate biomarker for CC prognosis and diagnosis.

Han *et al.* characterized the two mature chains of miR-21 and their oncogenic roles in cervical cancer⁷³. The regulation of the mature 5p and 3p chains from several miRNAs has been investigated in other cancers, including colorectal, gastric, breast, lung, kidney, and bladder^{36,72,74–77}, suggesting the need to focus further studies on the two mature chains from the 272 miRNAs reported in this study.

Figure 6 shows the complexity of the interactions between miRNAs and tumour suppressor genes, proto-oncogenes and oncogenes. The study of interaction networks between cell cycle genes and miRNAs involved in cancer is one of the most recent challenges in systems biology and is important for elucidating the control mechanisms for cancer biological process^{78–81}.

miRNAs in HPV integration sites and Latin American human genome variants

The differences in miRNA expression profiles between normal and cancerous tissues have led to the identification of clinical biomarkers for the early detection of many diseases, including various cancers and their precursor stages^{79,82,83}. Research on miRNAs associated with cancer has not taken into account the genetic variability in human populations, which influences the structure, expression and function of miRNAs in populations from different ethnic backgrounds. Studies on genetic variability are relevant to designing strategies for the diagnosis and prognosis of various diseases.

miR-11-3p, miR-31-3p, miR-107, miR-133a-3p, miR-133a-5p, miR-133b, miR-215-5p, miR-491-3p, miR-548d-5p and miR-944 were conserved in the four human genome variants. In the remaining 27 miRNAs, substitutions, deletions or insertions were observed in the nucleotide sequences, indicating that this variability can be decisive when determining susceptibility to the development of CC (Table 7 and Supplementary File 3).

There are numerous studies that analyse miRSNPs in different malignancies^{84–86}, but there is no available data on the correlation of SNPs in CC-associated miRNAs located in HPV integration sites in Latin American human genomic variants.

According to our results, the genomes from Latin America showed a lower miR SNP frequency compared to the control genome (BEB), although the Colombian (CLM) genome frequency was more

similar to the BEB genome. Latin American populations have experienced migrations from European, Asian and African individuals⁸⁷. Thus, our results could be a result of the specific interracial mixing of Colombian populations but also due to migration patterns during human settlement in Latin America.

miRSNPs can affect the structure and function of miRNAs by impacting interactions between miRNAs and their mRNA targets or interfering with the expression levels of individual miRNAs^{20–22,88,89}. miRSNPs could cause the loss or gain of binding sites for the co-evolution of miRNAs and their target mRNA and even influence cell processes related to tumour progression, disease phenotypes or susceptibility to developing a specific disease.

More studies are needed to clarify the role, targets and transcriptional regulatory mechanisms of cellular events in which miRNA are involved, including differentiation, apoptosis, metabolism and carcinogenesis. The expression and deregulation of miRNAs in cancer as well as their role as biological markers in diagnosis and treatment of CC should be explored. Further identification of cellular genes and signalling pathways involved in CC progression could lead to the development of new therapeutic strategies based on miRNAs^{90,91}. Additional biomarkers associated with apoptosis, necrosis and possible interactions with CRISPR complex sequences can be explored in order to develop therapeutic strategies in the future.

Data availability

Dataset 1. The mature miRNA reference sequences were obtained in FASTA format from the miRBase database. DOI, [10.5256/f1000research.10138.d164732](https://doi.org/10.5256/f1000research.10138.d164732)²⁸

Dataset 2. Matrix of data containing all the necessary components for the validation of data on CC-associated miRNAs in HPV integration sites in Latin American human genomic variants. DOI, [10.5256/f1000research.10138.d164736](https://doi.org/10.5256/f1000research.10138.d164736)³⁶

Author contributions

MGF directed all the research and bioinformatics analysis, wrote the article and made the final edits. OAGG developed the methodology and bioinformatics analysis and edited the article. JMH co-advised the research, wrote the article and made the final edits, and MCYC wrote the article.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgments

The authors thank the recommendations and suggestions of cPhD. Guillermo Torres from Kiel University (Germany) to improve the bioinformatics approach in this research.

Supplementary material

Supplementary File 1. Articles that mention HPV integration sites, detailing the most frequent types of HPV associated with CC.

[Click here to access the data.](#)

Supplementary File 2. Diagram indicating the regions on all chromosomes with miRNA binding sites that are associated with cervical cancer.

[Click here to access the data.](#)

Supplementary File 3. miRNAs identified in HPV integration sites, displaying the nucleotide variations in the selected Latin American human genome variants and in the control variant.

[Click here to access the data.](#)

References

1. Ferlay J, Soerjomataram I, Dikshit R, *et al.*: **Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012.** *Int J Cancer*. 2015; **136**(5): E359–86.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Bernard HU, Calleja-Macias IE, Dunn ST: **Genome variation of Human Papillomavirus types: Phylogenetic and medical implications.** *Int J Cancer*. 2006; **118**(5): 1071–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Burd EM: **Human Papillomavirus and Cervical Cancer.** *Clin Microbiol Rev*. 2003; **16**(1): 1–17.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Richardson H, Kelsall G, Tellier P: **The Natural History of Type-specific Human Papillomavirus Infections in Female University Students.** *Cancer Epidemiol Biomarkers Prev*. 2003; **12**(6): 485–90.
[PubMed Abstract](#)
5. Woodman CB, Collins SI, Young LS: **The natural history of cervical HPV infection: unresolved issues.** *Nat Rev Cancer*. 2007; **7**(1): 11–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. 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**(11): 3878–84.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Pett M, Coleman N: **Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?** *J Pathol*. 2007; **212**(2): 356–67.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Namburu L, Meenakumari B, Swaminathan R, *et al.*: **Prognostic significance of HPV physical status and integration sites in cervical cancer.** *Asian Pac J Cancer Prev*. 2009; **10**(3): 355–60.
[PubMed Abstract](#)
9. Schmitz M, Driesch C, Jansen L, *et al.*: **Non-random integration of the HPV genome in cervical cancer.** *PLoS One*. 2012; **7**(6): e39632.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Bartel DP: **MicroRNAs: Genomics, Biogenesis, Mechanism, and Function.** *Cell*. 2004; **116**(2): 281–97.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Rodriguez A, Griffiths-Jones S, Ashurst JL, *et al.*: **Identification of mammalian microRNA host genes and transcription units.** *Genome Res*. 2004; **14**(10A): 1902–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. 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**(9): 2999–3004.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Bartel DP: **MicroRNA: Target Recognition and Regulatory Functions.** *Cell*. 2009; **136**(2): 215–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Sharma G, Dua P, Agarwal SM: **A Comprehensive Review of Dysregulated miRNAs Involved in Cervical Cancer.** *Curr Genomics*. 2014; **15**(4): 310–23.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Adams BD, Kasinski AL, Slack FJ: **Aberrant regulation and function of MicroRNAs in Cancer.** *Curr Biol*. 2014; **24**(16): R762–76.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Melo SA, Esteller M: **Dysregulation of microRNAs in cancer: Playing with fire.** *FEBS Lett*. 2011; **585**(13): 2087–99.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Cammaerts S, Strazisar M, De Rijk P, *et al.*: **Genetic variants in microRNA genes: Impact on microRNA expression, function, and disease.** *Front Genet*. Frontiers Media SA; 2015; **6**: 186.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Sudmant PH, Rausch T, Gardner EJ, *et al.*: **An integrated map of structural variation in 2,504 human genomes.** *Nature*. 2015; **526**(7571): 75–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, *et al.*: **A map of human genome variation from population-scale sequencing.** *Nature*. 2010; **467**(7319): 1061–73.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Liu C, Rennie WA, Carmack CS, *et al.*: **Effects of genetic variations on microRNA: Target interactions.** *Nucleic Acids Res*. 2014; **42**(15): 9543–52.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Torruella-Loran I, Laayouni H, Dobon B, *et al.*: **MicroRNA Genetic Variation: From Population Analysis to Functional Implications of Three Allele Variants Associated with Cancer.** *Hum Mutat*. 2016; **37**(10): 1060–73.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Wu M, Jolicœur N, Li Z, *et al.*: **Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs.** *Carcinogenesis*. 2008; **29**(9): 1710–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Guerrero A, Guerrero M: **MicroRNAs asociados al Cáncer de Cuello Uterino y sus lesiones precursoras: Una revisión sistemática MicroRNAs associated with Cervical Cancer and its precursor lesions: A systematic Review.** *Rev Univ y Salud*. 2016; **28**(2): 1–26.
[Reference Source](#)
24. Kozomara A, Griffiths-Jones S: **miRBase: annotating high confidence microRNAs using deep sequencing data.** *Nucleic Acids Res*. 2014; **42**(Database issue): D68–73.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Van Peer G, Lefever S, Anckaert J, *et al.*: **miRBase Tracker: keeping track of microRNA annotation changes.** *Database (Oxford)*. 2014; **2014**: pii: bau080.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acids Res*. 2011; **39**(Database issue): D152–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Hsu PW, Huang HD, Hsu SD, *et al.*: **miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes.** *Nucleic Acids Res*. 2006; **34**(Database issue): D135–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Guerrero Flórez M, Guerrero Gómez OA, Mena Huertas J, *et al.*: **Dataset 1 in: Mapping of microRNAs related to cervical cancer in Latin American human genomic variants.** *F1000Research*. 2017.
[Data Source](#)
29. 1000 Genomes Project Consortium, Abecasis GR, Auton A, *et al.*: **An integrated map of genetic variation from 1,092 human genomes.** *Nature*. 2012; **491**(7422): 56–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. International HapMap Consortium: **The International HapMap Project.** *Nature*. 2003; **426**(6968): 789–96.
[PubMed Abstract](#) | [Publisher Full Text](#)

31. Karolchik D, Hinrichs AS, Kent WJ: **The UCSC Genome Browser**. *Curr Protoc Bioinformatics*. 2009; Chapter 1: Unit1.4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Karolchik D, Baertsch R, Diekhans M, *et al.*: **The UCSC Genome Browser Database**. *Nucleic Acids Res*. 2003; **31**(1): 51–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Apweiler R, Bairoch A, Wu CH, *et al.*: **UniProt: the Universal Protein knowledgebase**. *Nucleic Acids Res*. 2004; **32**(Database issue): D115–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Magrane M, UniProt Consortium: **UniProt Knowledgebase: a hub of integrated protein data**. *Database (Oxford)*. 2011; **2011**: bar009.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Xia H, Li F, He T, *et al.*: **Distribution of Mature MicroRNA on Its Precursor: A New Character for MicroRNA Prediction**. *Int J Inf Technol*. 2005; **11**(8).
[Reference Source](#)
36. Guerrero Flórez M, Guerrero Gómez OA, Mena Huertas J, *et al.*: **Dataset 2 in: Mapping of microRNAs related to cervical cancer in Latin American human genomic variants**. *F1000Research*. 2017.
[Data Source](#)
37. Crooks GE, Hon G, Chandonia JM, *et al.*: **WebLogo: A sequence logo generator**. *Genome Res*. 2004; **14**(6): 1188–90.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Kuo WT, Su MW, Lee YL, *et al.*: **Bioinformatic Interrogation of 5p-arm and 3p-arm Specific miRNA Expression Using TCGA Datasets**. *J Clin Med*. 2015; **4**(9): 1798–814.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Muñoz N, Bravo LE: **Epidemiology of cervical cancer in Colombia**. *Salud Publica Mex*. 2014; **56**(5): 431–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Angulo A: **Análisis bioinformático de secuencias L1, E6, E7 de VPH de alto y bajo riesgo más frecuentes Latinoamérica**. Universidad de Nariño; 2014.
41. Sanchez C, Suarez K, Yopez M, *et al.*: **Infección por VPH en mujeres del municipio de Pasto, Colombia con resultados de citología normal**. *Rev Univ y Salud*. 2013; **15**(1): 7–21.
[Reference Source](#)
42. Nicola SN: **Tipificación del Virus del Papiloma Humano-VPH y su relación con características poblacionales y lesiones en Cáncer de Cuello Uterino en mujeres del Municipio de Pasto**. Universidad de Nariño; 2014.
43. Bodelon C, Untereiner ME, Machiela MJ, *et al.*: **Genomic characterization of viral integration sites in HPV-related cancers**. *Int J Cancer*. 2016; **119**(1): 1–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
44. Soto-De Leon SC, Camargo M, Sanchez R, *et al.*: **Prevalence of infection with high-risk human papillomavirus in women in Colombia**. *Clin Microbiol Infect*. 2009; **15**(1): 100–2.
[PubMed Abstract](#) | [Publisher Full Text](#)
45. 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**(7): 2514–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
46. Thorland EC, Myers SL, Gostout BS, *et al.*: **Common fragile sites are preferential targets for HPV16 integrations in cervical tumors**. *Oncogene*. 2003; **22**(8): 1225–37.
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Dall KL, Scarpini CG, Roberts I, *et al.*: **Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions**. *Cancer Res*. 2008; **68**(20): 8249–59.
[PubMed Abstract](#) | [Publisher Full Text](#)
48. 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**(46): 7233–42.
[PubMed Abstract](#) | [Publisher Full Text](#)
49. Haws BT, Cui W, Persons DL, *et al.*: **Clinical and Pathologic Correlation of Increased MYC Gene Copy Number in Diffuse Large B-Cell Lymphoma**. *Clin Lymphoma Myeloma Leuk*. 2016; **16**(12): 679–683.
[PubMed Abstract](#) | [Publisher Full Text](#)
50. Lee KS, Kwak Y, Nam KH, *et al.*: **Favorable prognosis in colorectal cancer patients with co-expression of c-MYC and β -catenin**. *BMC Cancer*. 2016; **16**(1): 730.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Wolfer A, Wittner BS, Irimia D, *et al.*: **MYC regulation of a “poor-prognosis” metastatic cancer cell state**. *Proc Natl Acad Sci U S A*. 2010; **107**(8): 3698–703.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
52. Rao PH, Arias-Pulido H, Lu XY, *et al.*: **Chromosomal amplifications, 3q gain and deletions of 2q33-q37 are the frequent genetic changes in cervical carcinoma**. *BMC Cancer*. 2004; **4**(1): 5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Wilting SM, Snijders PJ, Verlaet W, *et al.*: **Altered microRNA expression associated with chromosomal changes contributes to cervical carcinogenesis**. *Oncogene*. 2013; **32**(1): 106–16.
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Reshmi G, Chandra SS, Babu VJ, *et al.*: **Identification and analysis of novel microRNAs from fragile sites of human cervical cancer: Computational and experimental approach**. *Genomics*. 2011; **97**(6): 333–40.
[PubMed Abstract](#) | [Publisher Full Text](#)
55. Peter ME: **Targeting of mRNAs by multiple miRNAs: the next step**. *Oncogene*. 2010; **29**(15): 2161–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
56. Carleton M, Cleary MA, Linsley PS: **MicroRNAs and cell cycle regulation**. *Cell Cycle*. 2007; **6**(17): 2127–32.
[PubMed Abstract](#) | [Publisher Full Text](#)
57. Devi KJ, Chakraborty S, Deb B, *et al.*: **Computational identification and functional annotation of microRNAs and their targets from expressed sequence tags (ESTs) and genome survey sequences (GSSs) of coffee (*Coffea arabica* L.)**. *Plant Gene*. 2016; **6**: 30–42.
[Publisher Full Text](#)
58. Trafton A: **Shrinking tumors with an RNA triple-helix hydrogel glue**. 2015; 1–3.
[Reference Source](#)
59. Conde J, Oliva N, Atilano M, *et al.*: **Self-assembled RNA-triple-helix hydrogel scaffold for microRNA modulation in the tumour microenvironment**. *Nat Mater*. 2016; **15**(3): 353–63.
[PubMed Abstract](#) | [Publisher Full Text](#)
60. Kertesz M, Iovino N, Unnerstall U, *et al.*: **The role of site accessibility in microRNA target recognition**. *Nat Genet*. 2007; **39**(10): 1278–84.
[PubMed Abstract](#) | [Publisher Full Text](#)
61. Bulik-Sullivan B, Selitsky S, Sethupathy P: **Prioritization of genetic variants in the microRNA regulome as functional candidates in genome-wide association studies**. *Hum Mutat*. 2013; **34**(8): 1049–56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
62. Slezak-Prochazka I, Kluiver J, de Jong D, *et al.*: **Cellular localization and processing of primary transcripts of exonic microRNAs**. Wilusz CJ, editor. *PLoS One*. 2013; **8**(9): e76647.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
63. Concepcion CP, Bonetti C, Ventura A: **The microRNA-17-92 family of microRNA clusters in development and disease**. *Cancer J*. 2012; **18**(3): 262–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
64. Altuvia Y, Landgraf P, Lithwick G, *et al.*: **Clustering and conservation patterns of human microRNAs**. *Nucleic Acids Res*. 2005; **33**(8): 2697–706.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
65. Cai N, Wang YD, Zheng PS: **The microRNA-302-367 cluster suppresses the proliferation of cervical carcinoma cells through the novel target AKT1**. *RNA*. 2013; **19**(1): 85–95.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
66. Ivashchenko A, Berillo O, Pyrkova A, *et al.*: **The properties of binding sites of miR-619-5p, miR-5095, miR-5096, and miR-5585-3p in the mRNAs of human genes**. *Biomed Res Int*. 2014; **2014**: 720715.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
67. Ivashchenko A, Berillo O, Pyrkova A, *et al.*: **The arrangements of the locations of miR-619, miR-5095, miR-5096 and miR-5585 binding sites in the human mRNAs**. *Recent Adv Biomed Chem Eng Mater Sci*. 2014; 144–9.
[Reference Source](#)
68. Ivashchenko A, Berillo O, Pyrkova A, *et al.*: **The properties of binding sites of miR-619-5p, miR-5095, miR-5096, and miR-5585-3p in the mRNAs of human genes**. *Biomed Res Int*. 2014; **2014**: 720715.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
69. Schmitz M, Driesch C, Beer-Grondke K, *et al.*: **Loss of gene function as a consequence of human papillomavirus DNA integration**. *Int J Cancer*. 2012; **131**(5): E593–602.
[PubMed Abstract](#) | [Publisher Full Text](#)
70. Dweep H, Sticht C, Gretz N: **In-Silico Algorithms for the Screening of Possible microRNA Binding Sites and Their Interactions**. *Curr Genomics*. 2013; **14**(2): 127–36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
71. Palmero EI, de Campos SG, Campos M, *et al.*: **Mechanisms and role of microRNA deregulation in cancer onset and progression**. *Genet Mol Biol*. 2011; **34**(3): 363–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
72. Choo KB, Soon YL, Nguyen PN, *et al.*: **MicroRNA-5p and -3p co-expression and cross-targeting in colon cancer cells**. *J Biomed Sci*. 2014; **21**(1): 95.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
73. Han Y, Xu GX, Lu H, *et al.*: **Dysregulation of miRNA-21 and their potential as biomarkers for the diagnosis of cervical cancer**. *Int J Clin Exp Pathol*. 2015; **8**(6): 7131–9.
[PubMed Abstract](#) | [Free Full Text](#)
74. Uchino K, Takeshita F, Takahashi RU, *et al.*: **Therapeutic Effects of MicroRNA-582-5p and -3p on the Inhibition of Bladder Cancer Progression**. *Mol Ther*. 2013; **21**(3): 610–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
75. Mlcochova J, Faltejskova-Vychytilova P, Ferracin M, *et al.*: **MicroRNA expression profiling identifies miR-31-5p/3p as associated with time to progression in wild-type RAS metastatic colorectal cancer treated with cetuximab**. *Oncotarget*. 2015; **6**(36): 38695–704.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
76. Muti P, Sacconi A, Hossain A, *et al.*: **Downregulation of microRNAs 145-3p and 145-5p is a long-term predictor of postmenopausal breast cancer risk: The**

- ORDET prospective study.** *Cancer Epidemiol Biomarkers Prev.* 2014; **23**(11): 2471–81.
[PubMed Abstract](#) | [Publisher Full Text](#)
77. Lou C, Xiao M, Cheng S, *et al.*: **MIR-485-3p and miR-485-5p suppress breast cancer cell metastasis by inhibiting PGC-1 α expression.** *Cell Death Dis.* 2016; **7**(3): e2159.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 78. Watanabe Y, Tomita M, Kanai A: **Computational Methods for MicroRNA Target Prediction.** *Methods Enzymol.* 2007; **427**: 65–86.
[PubMed Abstract](#) | [Publisher Full Text](#)
 79. Pritchard CC, Cheng HH, Tewari M: **MicroRNA profiling: approaches and considerations.** *Nat Rev Genet.* 2012; **13**(5): 358–69.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 80. Wang N, Xu Z, Wang K, *et al.*: **Construction and analysis of regulatory genetic networks in cervical cancer based on involved microRNAs, target genes, transcription factors and host genes.** *Oncol Lett.* 2014; **7**(4): 1279–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 81. Yin Y, Song M, Gu B, *et al.*: **Systematic analysis of key miRNAs and related signaling pathways in colorectal tumorigenesis.** *Gene.* 2016; **578**(2): 177–84.
[PubMed Abstract](#) | [Publisher Full Text](#)
 82. Hayes J, Peruzzi PP, Lawler S: **MicroRNAs in cancer: Biomarkers, functions and therapy.** *Trends Mol Med.* 2014; **20**(8): 460–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
 83. Ma Q, Wan G, Wang S, *et al.*: **Serum microRNA-205 as a novel biomarker for cervical cancer patients.** *Cancer Cell Int.* 2014; **14**: 81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 84. Mu W, Zhang W: **Bioinformatic Resources of microRNA Sequences, Gene Targets, and Genetic Variation.** *Front Genet.* 2012; **3**: 31.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 85. Mi Y, Wang L, Zong L, *et al.*: **Genetic variants in microRNA target sites of 37 selected cancer-related genes and the risk of cervical cancer.** *PLoS One.* 2014; **9**(1): e86061.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 86. Hu Y, Yu CY, Wang JL, *et al.*: **MicroRNA sequence polymorphisms and the risk of different types of cancer.** *Sci Rep.* 2014; **4**: 3648.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 87. Homburger JR, Moreno-Estrada A, Gignoux CR, *et al.*: **Genomic Insights into the Ancestry and Demographic History of South America.** *PLoS Genet.* 2015; **11**(12): e1005602.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 88. Bhartiya D, Scaria V: **Genomic variations in non-coding RNAs: Structure, function and regulation.** *Genomics.* 2016; **107**(2–3): 59–68.
[PubMed Abstract](#) | [Publisher Full Text](#)
 89. Rawlings-Goss RA, Campbell MC, Tishkoff SA: **Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers.** *BMC Med Genomics.* 2014; **7**(1): 53.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 90. Ahmad J, Hasnain SE, Siddiqui MA, *et al.*: **MicroRNA in carcinogenesis & cancer diagnostics: a new paradigm.** *Indian J Med Res.* 2013; **137**(4): 680–94.
[PubMed Abstract](#) | [Free Full Text](#)
 91. Liu Z, Sall A, Yang D: **MicroRNA: An emerging therapeutic target and intervention tool.** *Int J Mol Sci.* 2008; **9**(6): 978–99.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 28 September 2017

<https://doi.org/10.5256/f1000research.10920.r24293>

© 2017 Agarwal S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Subhash Mohan Agarwal

Bioinformatics Division, ICMR-National Institute of Cancer Prevention and Research, Noida, Uttar Pradesh, India

In the present study the authors have mapped the miRNA involved in cervical cancer on to Latin American genome using in silico predictions. As cervical cancer has the highest mortality rates in low and middle income countries we do need to advance our understanding on mechanism of its progression. It is an interesting study however, there are few shortcomings in the current MS which need to be addressed.

1. It is not clear how human genes near to viral insertion sites have been identified. It was observed that near integration sites mostly only one or two genes are present. The method and parameters used for finding the genes should be detailed so that the results are reproducible. For example have the genes been identified within a particular distance of the insertion sites.
2. Why the authors have mapped the integration sites for 8 types of HPVs collectively and not HPV-16 and 18 alone which are the high risk HPV. Is there any basis for it?
3. The authors have stated that a total of 2028 miRNA binding sites of which 432 were detected in miRBase. In my opinion the analysis should have been restricted to only these sites as they are experimentally identified sites for miRNA binding.
4. As I understand the authors have mapped 42 miRNAs on Latin American genome. It is not clear how 42 miRNAs were selected for this subsequent step.

Minor comments:

1. In the supplementary data the headings of the tables should be in English.
2. Are there 578 or 568 integration sites. It appears from Dataset 2 that there are 568 integration sites. Sheet named "VPH integration sites"
3. Page 4 (last 2 lines) instead of 12 it should be 13. As per the data in figure 3 there are 13 genes in the intermediate category.

4. Methods in Abstract: miRNA sequences associated with CCwere obtained from miRBase. Shouldn't it be literature?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Aug 2018

Milena Guerrero, University of Nariño, Pasto, Nariño, Colombia

Dear Reviewer SMA.

We resubmitted the second version of the paper after addressing the various concerns raised.

We would like to thank for their time and for their constructive comments to help assist us in improving the manuscript.

We made the necessary changes in order to address all the specified concerns. The direct responses to the reviewer's comments are listed below

Reviewer. SMA

Q8. It is not clear how human genes near to viral insertion sites have been identified. It was observed that near integration sites mostly only one or two genes are present. The method and parameters used for finding the genes should be detailed so that the results are reproducible. For example have the genes been identified within a particular distance of the insertion sites.

R8: Thanks for your valuable suggestion. In a previous work (published in Spanish) it was described Guerrero & Guerrero, 2016. We analyze 42 scientific reports including chromosomal bands, HPV genotype, molecular technique for experimental results, and expression profile of miRNA according to lesions in CC.

Here, we use the description and annotation of genes described in Data Bases, identified positions, regions, chromosomes, sequences and other characteristics to limit the regions of IS. All was manually curated for each chromosome along genome.

Q9: Why the authors have mapped the integration sites for 8 types of HPVs collectively and not HPV-16 and 18 alone which are the high risk HPV. Is there any basis for it?

R9: Thanks for your valuable comment. In fact, our reason were the results of two of our previous studies about genotyping of HPV in our region, which showed remarkable frequency, in addition to HPV16 and HPV 18, to other as HPV 45, 31, 33, 58, 67, 68. That was the main reason to include these genotypes in our analysis of mapping.

But also by scientific purposes, because, in the literature is highly frequent to find and to study HPV 16 and 18, but not other genotypes. This information is relevant for us, to understand more depth the natural history of HPV and its mechanisms.

Sánchez et al., 2013. Published in Spanish

Nicola N, 2014. Bachelor thesis published in Spanish

Q10: The authors have stated that a total of 2028 miRNA binding sites of which 432 were detected in miRBase. In my opinion the analysis should have been restricted to only these sites as they are experimentally identified sites for miRNA binding.

R10: Thanks for your valuable comment. To our knowledge, the 2028 binding sites to miRNA and its key role in cervical cancer were identified by the first time in this study, using BLAT mapping. Is an important finding to highlight, compared to 432 binding sites previously reported, and a valuable contribution of bioinformatic tools for this kind research.

Q11: As I understand the authors have mapped 42 miRNAs on Latin American genome. It is not clear how 42 miRNAs were selected for this subsequent step.

R11: Thanks for your valuable comment. Briefly the pipeline was:

- BLAT mapping of miRNAs on reference genome
- Identification of integration sites (IS) of HPV
- From IS HPV, looking for positions of genes of cell cycle, near to these IS.
- Functional analysis of IS HPV according to annotations described by Uniprot.
- With positions of near genes (regulators of cell cycle) and the positions of binding sites of miRNAs, manual mapping for each chromosome was done.
- Finally, miRNAs in proximity to cell cycle genes control, were identified.

Minor comments:

Q12. In the supplementary data the headings of the tables should be in English.

R12: Thanks for your detailed comment. It was corrected.

Q13: Are there 578 or 568 integration sites. It appears from Dataset 2 that there are 568 integration sites. Sheet named "VPH integration sites"

R13: Correct. There are 568 HPV integration sites. It was corrected.

Q14: Page 4 (last 2 lines) instead of 12 it should be 13. As per the data in figure 3 there are 13 genes in the intermediate category.

R14: Correct. Are 13 genes. It is corrected.

Q15: Methods in Abstract: miRNA sequences associated with CCwere obtained from miRBase. Shouldn't it be literature?

R15: Correct. The abstract is re written to include and express data in a clear way.

Competing Interests: No competing interests were disclosed.

Reviewer Report 11 July 2017

<https://doi.org/10.5256/f1000research.10920.r23646>

© 2017 Anzola J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Juan Manuel Anzola

Bioinformatics & Computational Biology, Corporación CorpoGen, Bogotá, Colombia

In this work, Guerrero et al. use mature microRNA in order to detect possible targets of these microRNAs in the human genome, and its population variants, including from Latin American, in order to determine possible associations with cervical cancer.

I found the paper sound and its results, analysis and conclusions within the reach of the methodology, however I find the methods lacking, in particular when it comes to the parameters used in the BLAT search. BLAT uses a default seed of 11 to do nucleotide searches (they call it tileSize). So it would be good if the authors state clearly what were the BLAT parameters used, in particular "tileSize" and "stepSize". If a 11-word was used for this analysis the authors are running the risk of not being sensitive enough in their searches. High Specificity, Low Sensitivity. It would be interesting to determine how many of the genes reported as being targets for microRNAs are not detected in your search.

microRNA have a particular set of rules when it comes to binding to their respective targets, with seeds between 6, 8 or 9 nucleotides. Nothing is stated in the paper to give an idea of how the rules

for target detection were used in this paper. See Mullany et al paper.

It is assumed throughout the paper that all the hits are true positives. There is no measure as to how good is BLAT to detect true vs false positives.

The paper:

In your introduction you mention that microRNAs are involved in cancer. The paragraph suggest this is the only role of microRNAs, however they are involved in processes such as development and morphogenesis, so please rephrase this paragraph because cancer is not the only role of microRNAs.

Figure 7D is better represented as percentage, as in the body of the paper.

Your phrase: "Because some chromosomes have a greater number of miRNA binding sites, it provides evidence of a non-random distribution of miRNAs within the chromosomes." could be the result of chromosome length. Please provide statistical support for your statement.

Page 14: Not all Pre microRNAs produce mature ones from both strands, in fact in the great majority of cases is only one strand that produces the mature one.

The paper will be ready for indexing once these observations are addressed.

References

1. Mullany LE, Herrick JS, Wolff RK, Slattery ML: MicroRNA Seed Region Length Impact on Target Messenger RNA Expression and Survival in Colorectal Cancer. *PLoS One*. 2016; **11** (4): e0154177
[PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Aug 2018

Milena Guerrero, University of Nariño, Pasto, Nariño, Colombia

We resubmitted the second version of paper after addressing the various concerns raised. We would like to thank you for their time and for their constructive comments to help assist us in improving the manuscript.

We made the necessary changes in order to address all the specified concerns. The direct responses to the reviewer's comments are listed below:

Reviewer. JMA

Q1. I found the paper sound and its results, analysis and conclusions within the reach of the methodology, however I find the methods lacking, in particular when it comes to the parameters used in the BLAT search. BLAT uses a default seed of 11 to do nucleotide searches (they call it tileSize). So it would be good if the authors state clearly what were the BLAT parameters used, in particular "tileSize" and "stepSize". If a 11-word was used for this analysis the authors are running the risk of not being sensitive enough in their searches. High Specificity, Low Sensitivity. It would be interesting to determine how many of the genes reported as being targets for microRNAs are not detected in your search.

R1: Thanks for your valuable suggestion. The methodology is re writer. In fact, BLAT only works with tile size 11. This mean that the average total length of mature miRNAs around 16 to 22, and consequently the seed sequence surely is represented at least 50% in mapping with this number of nucleotides.

About suggestion "to determine how many of genes reports as being for miRNAs are not detected in your search", we did the search, using programming R. And similar results of reported here we obtained. We not include this new focus on this paper, but If is need, we can send one of R mapping obtained for one chromosome.

Q2: microRNA have a particular set of rules when it comes to binding to their respective targets, with seeds between 6, 8 or 9 nucleotides. Nothing is stated in the paper to give an idea of how the rules for target detection were used in this paper. See Mullany et al paper.

R2: Thanks for the valuable comment. According to Mullany one of the most important "rules" for binding to mRNA and the role for cancer are length of seed sequence of miRNA. This condition is mentioned in second paragraph of introduction. Despite of this, the analysis no mentioned, but the authors analyzed seed sequences of miRNAs in terms of folding miRNAs (loop, folk, stem), length, 5'UTR extreme, the results was not included for this publication, because is part to another analysis.

Q3: It is assumed throughout the paper that all the hits are true positives. There is no measure as to how good is BLAT to detect true vs false positives.

R3: Thanks for your valuable annotation. Considering this probability, after that, we use R and bioconductor tools in order to be sure about the mapping results, we found a match between BLAT and R mapping. This data are under analysis ongoing.

The paper:

Q4. In your introduction you mention that microRNAs are involved in cancer. The paragraph suggest this is the only role of microRNAs, however they are involved in processes such as development and morphogenesis, so please rephrase this paragraph because cancer is not the only role of microRNAs.

R4. Correct. The text is re write.

Q5. Figure 7D is better represented as percentage, as in the body of the paper.

R5. Thanks for the valuable suggestion. The figure was modified, highlighting percentages instead numeric values. File of figures.

Q6. Your phrase: "Because some chromosomes have a greater number of miRNA binding sites, it provides evidence of a non-random distribution of miRNAs within the chromosomes." could be the result of chromosome length. Please provide statistical support for your statement.

R6: Thanks for the valuable suggestion. The authors included the statistical analysis and confirmed the results, for more clarity the paragraph is re-writing as follow:
"In order to confirm the distribution of miRNA binding sites, the analysis for each chromosomal following all chromosomes was done. The statistic W Shapiro-Wilk test, show a p-value 0.02; and the mean comparison analysis by ANOVA with a p-value 0.0046 allowed us to confirm the non-random distribution of miRNA binding sites along the genome".

Q7. Page 14: Not all Pre microRNAs produce mature ones from both strands, in fact in the great majority of cases is only one strand that produces the mature one.

R7: Thanks for the suggestion. It was adjusted.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research