

The genomic distribution map of human papillomavirus in China

Ling Chen^{1‡}, Yan Dong^{2, 3‡}, Jiao Li^{4‡}, Dan Wang^{5‡}, Li Xu^{6‡}, Jinqiu Zhao⁷, Huandong Liu⁸, Jungao Lu⁹, Zuoyi Yao¹⁰, Xiaosong Li^{2†*}

¹Key Laboratory of Molecular Biology of Infectious Diseases, Ministry of Education, Chongqing Medical University, Chongqing 400016, China

²Clinical Molecular Medicine Testing Center, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

³Department of Microbiology and Immunology, Kunming Medical University, Kunming 650031, China

⁴Department of Laboratory Medicine, the First Affiliated Hospital of Xi'an Jiaotong University

⁵Department of Clinical Laboratory, People's Hospital of Rongchang District, Rongchang 402460, China

⁶Department of Clinical Laboratory, the First Affiliated Hospital of AMU, Chongqing 400038, China

⁷Department of Infectious Diseases, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

⁸Department of Neurosurgery, People's Hospital of Tibet Autonomous Region, Lhasa, 850000, China

⁹Department of Clinical Laboratory, The Third Affiliated Hospital of Guizhou Medical University, Duyun 558000, China

¹⁰Department of General Surgery, The Chengdu Fifth People's Hospital, Chengdu 611130, China

† Current address

‡ These authors contributed equally to this study.

* Corresponding author:

Xiaosong Li MD., PhD., Clinical Molecular Medicine Testing Center, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

Tel: +86 023 6848 6780 Fax: +86 023 6848 6780

E-mail: aleex24@163.com

Abstract

Background: Human papillomavirus (HPV) infection plays a key role in the development of cervical cancer that ranks as the third most common gynecological cancer in China. We attempt to investigate the age and genotype-specific prevalence of HPV DNA in western Chinese females. **Methods:** A total of 301,880 woman were recruited from 4 different regions of western China. Routine cervical samples were collected and HPV screening was performed using 21 HPV Genotyping test (HybriBio). The overall prevalence, age-specific prevalence, and genotype distribution were analyzed. **Results:** The overall prevalence of HPV was 18.01%. The high-risk HPV infection rate was 79.14%, the low-risk HPV infection rate was 12.56%, and the mixed HPV infection rate was 8.30%. The most common 4 HR HPV subtypes were HPV-52, 16, 58 and 53, which accounted for 20.49 %, 19.93 %, 14.54 % and 10.01 %. In LR HPV genotype, HPV-6 ranked the highest (28.17 %), followed by HPV-81 (9.09 %), HPV-11 (3.78 %). HPV genotype subgroup analysis also showed that single-type infections had the highest prevalence rate (77.26%) among HPV positive individuals. Among multi-infection genotype, double infection was most common with frequencies of 76.04%. **Conclusion:** This large report shows that the overall prevalence of HPV was high in China, whose distribution exhibits different patterns across different particular age and regions. Viral genotypes HPV53, 6 were frequently detected in this population, which is worth of significant clinical attention.

Keywords: Human papillomavirus (HPV); cervical cancer; prevalence; genotypes; China

1. Introduction

Cervical cancer, a leading genital cancer, is considered as the third most common female gynecologic malignancy and the fourth most common cause of female death from cancer, with the estimated 570,000 new cases and 311,000 new deaths in 2018 (GLOBOCAN, 2018)[1,2]. Compared with developed countries, the age-standardized

incidence rate for cervical cancer is more frequent in developing countries (8.0 per 100,000 vs 5.0 per 100,000 women-years, respectively)[3]. China accounts for 14 % of the world's annual incidence of cervical cancer[4]. Thus, in China, cervical cancer remains a relatively heavy burden of public hygiene management with increasing morbidity and mortality of cervical cancer in young women[5,6].

Human papillomavirus (HPV), a sexually transmitted DNA virus belonging to the Papovaviridae family, has been confirmed as the essentially causative agent for cervical cancer[7]. It is estimated that 80 % of sexually active adults have been infected with at least one HPV Genotype[8]. If the infection with the high-risk HPV strains persists which then could be a well-established cause of cervical cancer[9]. More than 200 distinct HPV genotypes have been characterized to date, of which, approximately 40 infect the mucosal epithelium of anus and genital tract[10]. Generally, they are classified as high-risk HPV (carcinogenic HPV types, HR-HPV), low risk HPV (non-carcinogenic HPV types, LR-HPV) and intermediate-risk HPV (IR-HPV), according to their carcinogenic risk or potentially pathogenicity have been reported to exist[11].

It is well known that this type of malignancy is one of the most preventable cancers. Currently, part of comprehensive strategies aimed at control of cervical cancer is based on vaccination against HPV and HPV-based screening programs, which have been demonstrated to effectively eliminate the burden of cervical cancer worldwide[12,13]. Nonetheless, the prevalence and genotype distribution of HPV infections are heterogeneous widespread (differences varies between nations and regions, as well as within a country), which makes progress towards prevention often frustrating[14]. Hence, an accurate understanding of the regional distribution characteristics of HPV genotypes is extremely important for both prophylactic vaccine-based HPV development and for HPV-based cervical cancer screening strategy-making.

Reports about large-scale data on the genotypic spectrum of HPV infection in China are limited. Therefore, we conducted a retrospective summary of enormous amount of HPV genotypes distribution data in China, the overall prevalence,

age-specific prevalence, and genotype distribution of HPV in different regions were also calculated and analyzed. This study would provide guidance for the development of future screening and prevention programs.

2. Results

2.1. Overall HPV infection prevalence and genotype distribution

A total of 301880 samples were collected and detected by HPV genotype. The genotype test showed 55071 samples were HPV positive, the overall HPV prevalence was 18.01 %, in which 28.14 % in Tibet Autonomous Region, 18.59 % in Chongqing Municipality, 10.33 % in Guizhou Province and 29.09 % in Shaanxi Province. Among the four different provinces, autonomous regions and municipalities, there were significant differences in the HPV prevalence ($X^2=6120.54$, $P<0.001$). The details of genotypes distribution of was shown respectively in the following contents.

2.2. Age and genotype distribution of HPV infection in different region

2.2.1. The HPV prevalence in Tibet Autonomous Region

There were 36073 samples involving in Tibet Autonomous Region and the HPV prevalence was identified as 28.14 % ($n = 10150$). The highest HPV detection rate was found in age group 36-50 years old (detection rate = 53.40 %) (Fig 1). Among HPV positive women, the HR- HPV infection accounted for 73.51 % ($n=7461$). In HR- HPV, type 16 (19.44%) and 52 (21.50%) were the most prevalent genotypes, and followed by 53 (11.23%) and 58 (11.83%). Among the 1428 case of LR- HPV infection, the two most common detected LR- HPV were type 6 (14.13%) and 81 (11.15%) in descending order. The rates of single and multiple (infection with ≥ 2 different HPV genotypes) infections were 70.35% ($n=7141$) and 29.65% ($n=3009$), respectively (among the multiple infections, the rate of double infections was 71.32%). Among the mixed (co-infection with HR- HPV and LR- HPV) infection cases, double HPV genotype infection accounted for more than a half ($n=761$). And in multiple infection of HR- HPV, the double HPV genotype infection was also the highest with 1298 cases. The detailed genotype distribution of HPV infection was described in Fig 2 and Fig 3. Overall, the prevalence of HPV infection in Tibet Autonomous Region was dominated by single HR- HPV infection, especially type 16 or 52.

2.2.2. HPV infection in Chongqing Municipality

A total of 37389 women were identified as HPV positive among the 201,089 samples in Chongqing Municipality, with a HPV infection rate of 18.59 %. Similar to Tibet Autonomous Region, the highest HPV detection rate was found in age group 36–50 years old (detection rate = 52.96 %) (Fig 4). Among the infected women, HR-HPV infection accounted for 81.32% (n= 30406). Among HR- HPV infection, the three most prevalent types were type 16, 52 and 58, with frequencies of 20.82 %, 20.80 % and 15.64 %, respectively. The most common LR- HPV types were HPV 6 (35.65 %), even exceeded the most prevalent HR- HPV genotype (Fig 5 and Fig 6). The rates of single and multiple infections were 79.00 % (n=29537) and 21.00% (n=7852), respectively (among the multiple infections, the rate of double infections was 77.06 %).

2.2.3. HPV infection in Guizhou Province

Among 60205 female of Guizhou Province, the prevalence of HPV infection was 10.33 % with 6219 cases identified as HPV positive. In all tested age groups, the highest HPV detection rate was found in age group 20-30 years old (detection rate = 43.18 %) in Guizhou Province (Fig 7). Among HPV positive female, the HR- HPV infection accounted for 62.58 % (n=3892). Among HR- HPV infection, the four most prevalent types were type 52, 16, 58, and 53 in descending order. The most common LR- HPV types were HPV 6 and 11 (Fig 8 and Fig 9). The rates of single and multiple infections were 79.71% (n=4957) and 20.29% (n=1262), respectively. Among the multiple genotypes infections, the rate of double type infection was 82.96% (n=1047). The HR- HPV infection was also the dominant in Guizhou Province with 3892 cases (62.58%).

2.2.4. HPV infection in Shaanxi Province

There were 4513 samples recorded in Shaanxi Province, with 1313 sample were identified as HPV positive (29.09 %). In all tested age groups, the highest HPV detection rate was found in age group 26-55 years old (detection rate = 79.06 %) in Shaanxi Province (Fig 10). Among HPV positive female, the HR- HPV infection accounted for 80.05% (n=1051). The four most prevalent HR- HPV types were type

16, 52, 58, 53 in descending order. The most common LR- HPV types were HPV 6 and 81 (Fig 11 and Fig 12). The rates of single and multiple infections were 69.92% (n=918) and 30.08% (n=395), respectively. Among the multiple infection, double genotypes infection was the dominant types, which counted for 69.62 % (n=275).

2.3. HPV genotype distribution in all positive samples

There were 55071 samples in this study. We also analyzed the genotype distribution so that we could make better decision on the HPV vaccine tragedy. Notably, HPV 52 (11282, 20.49 %), 16 (10974, 19.93%), 58 (8007, 14.54%), 53 (5513, 10.01%) were the top four prevalent genotypes among HR- HPV and HPV 6 (15515, 28.17 %), 81 (5008, 9.09 %), 11 (2082, 3.78 %) were the top three prevalent genotypes among LR- HPV. Table 1 showed the detailed genotype distribution.

2.4. Age distribution in all positive samples

To better understand the age distribution of HPV infection, after dividing the subjects into 14 age groups (≤ 15 , 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, ≥ 76 years), the HPV infection prevalence of different region in each age group was analyzed (Table 2). It could be directly observed that HPV infection happened in all ages, and women aged 26-50 years accounted for a main part (Fig 13). Elder women (> 50) and young girls (≤ 25) had a relatively lower HPV infection proportion. Besides, we conducted a chi-square test to determine whether the difference of age distribution was significant in different region. The result showed that there was a significant difference among these groups ($X^2=32.83$, $P<0.001$). We summarized the age distribution in three HPV infection patterns, including HR- HPV only, LR- HPV only, mixed HR- HPV and LR- HPV HPV infections. It was showed that the pattern of HR- HPV only was the dominate infection pattern (Table 3).

2.5. Age distribution of dominant types

According to the above results, the most common 5 types of HPV were as follows: the HR- HPV genotype HPV 52, HPV 16, HPV 58, HPV 53, and and LR- HPV 6. As the most five prevalent genotype, we decided to further explore their distribution characteristics in different age. Among type 52, 16, 58 and 53, the number

of infection cases increased gradually and reached its highest point in the group of 41-45 years old, while after age 41-45, the number of infection cases was reduced. Interesting, in type 6, the most common infection age was found in women 21-30 years old (Fig 14). Women aged 36-40 years old and 46-50 years old also had an abundant number in these 5 genotypes HPVs. According to these results, we could come to the conclusion that women aged 36-50 years old account for the vast majority of these five dominant HPV infection.

2.6. Figures and Tables

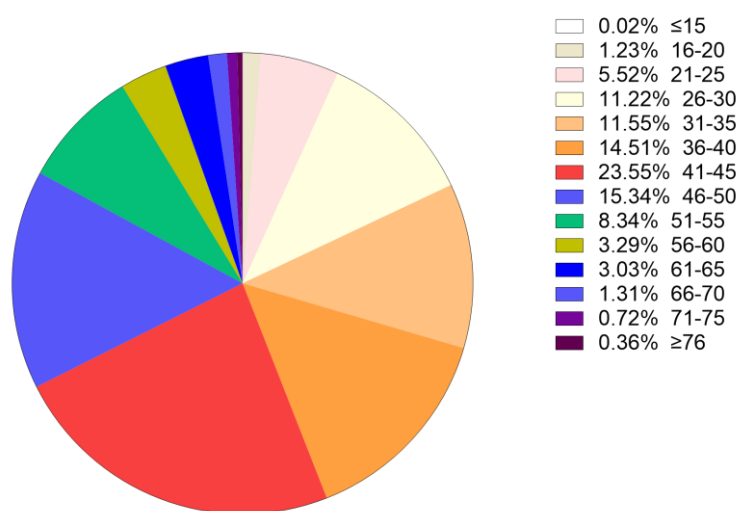


Figure 1. Age distribution of HPV infection in Tibet Autonomous Region

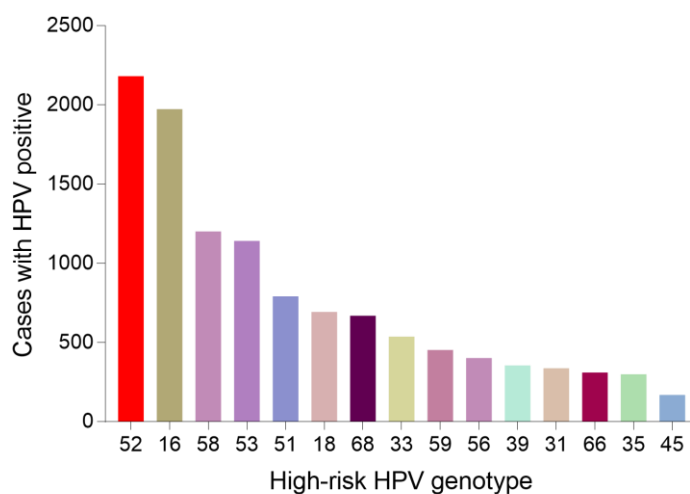


Figure 2. Genotype distribution of high-risk HPV infection in Tibet Autonomous Region

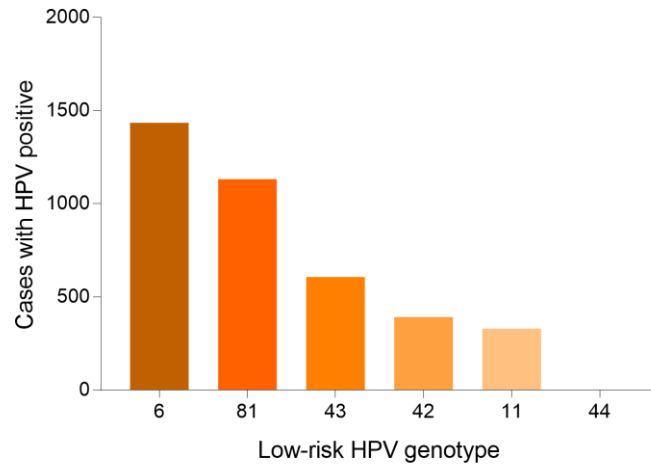


Figure 3. Genotype distribution of low-risk HPV infection in Tibet Autonomous Region

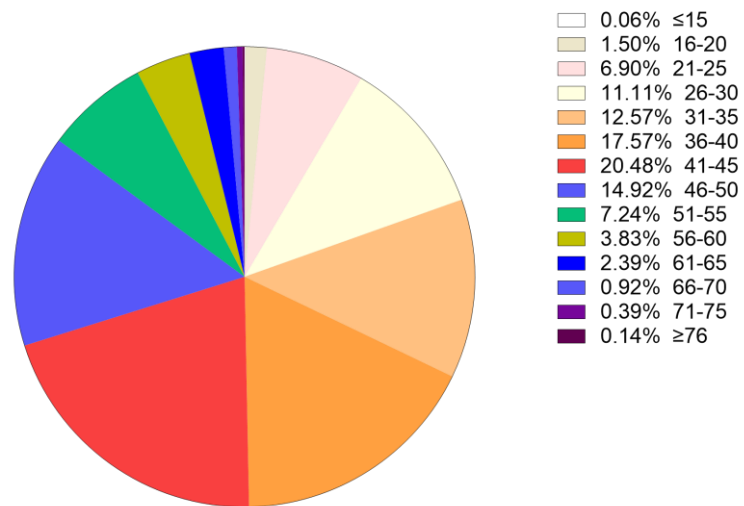


Figure 4. Age distribution of HPV infection in Chongqing Municipality

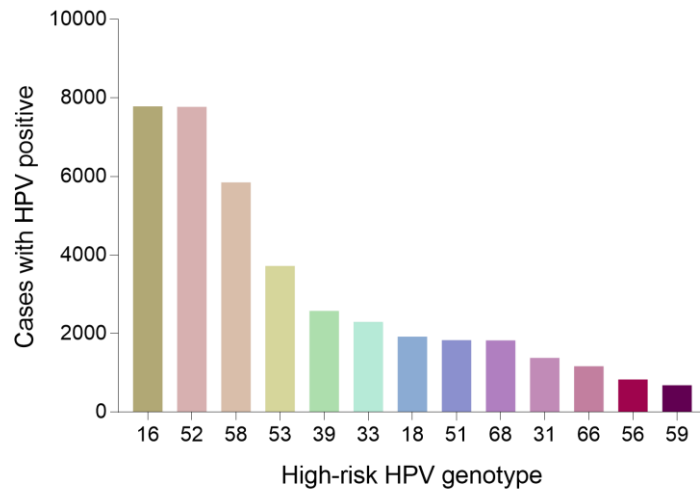


Figure 5. Genotype distribution of high-risk HPV infection in Chongqing Municipality

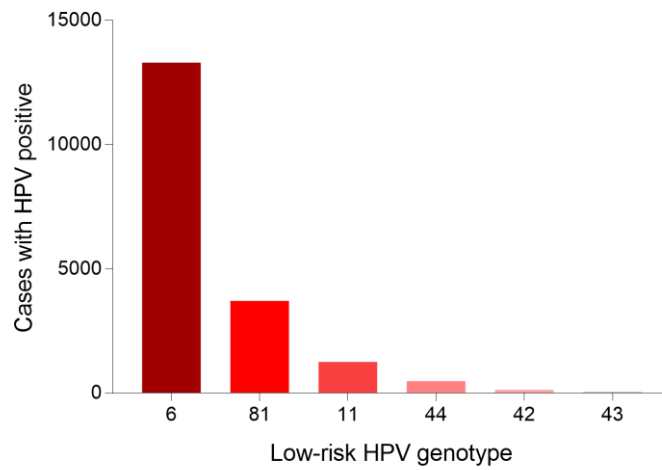


Figure 6. Genotype distribution of low-risk HPV infection in Chongqing Municipality

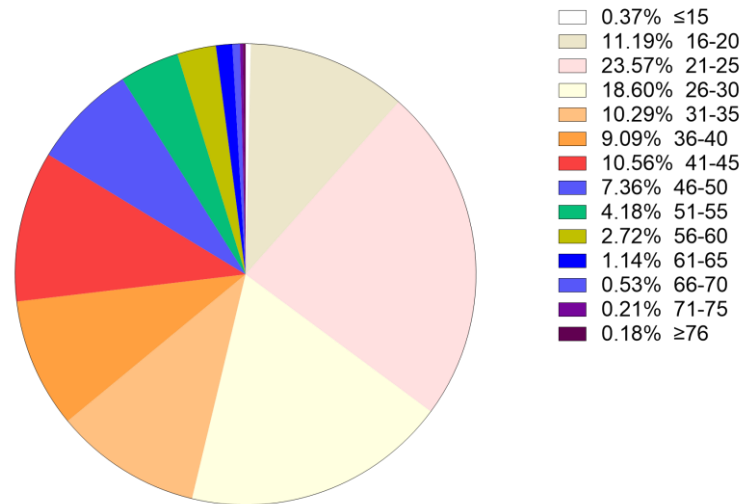


Figure 7. Age distribution of HPV infection in Guizhou Province

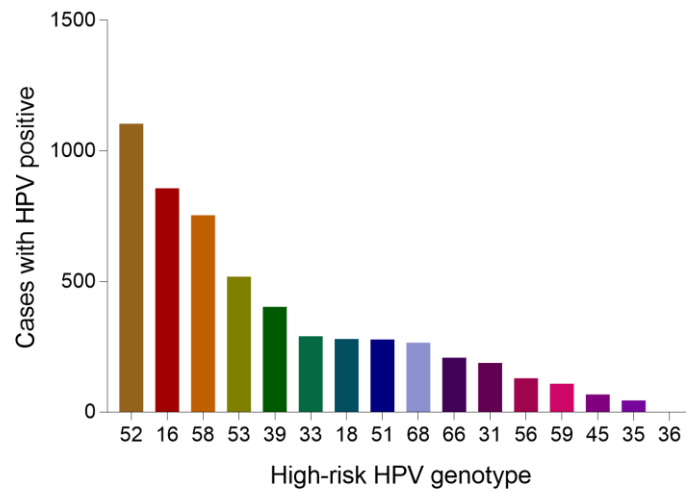


Figure 8. Genotype distribution of high-risk HPV infection in Guizhou Province

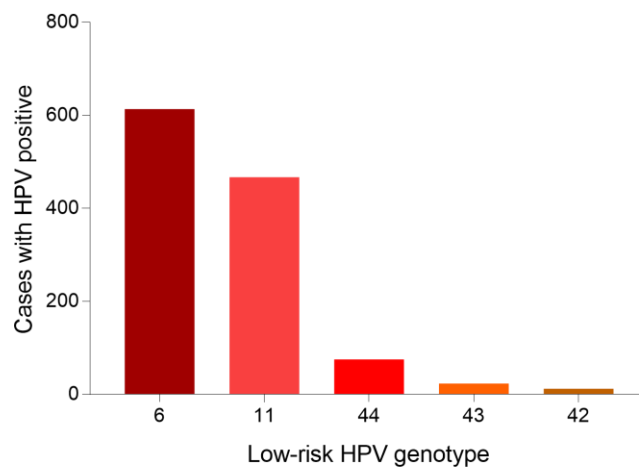


Figure 9. Genotype distribution of low-risk HPV infection in Guizhou Province

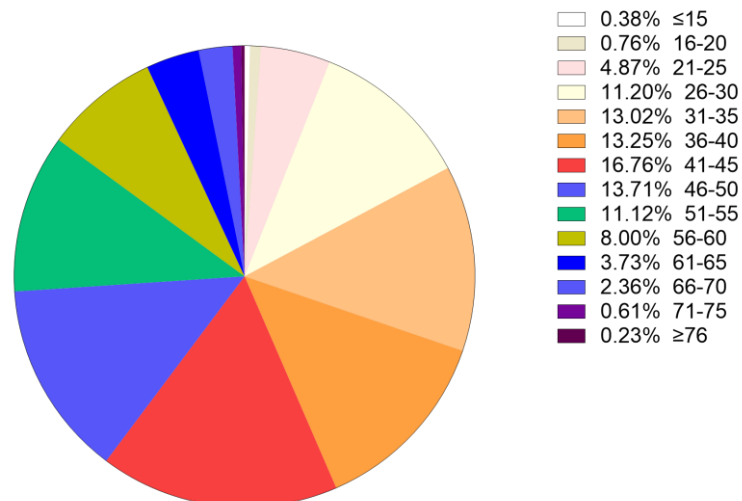


Figure 10. Age distribution of HPV infection in Shaanxi Province

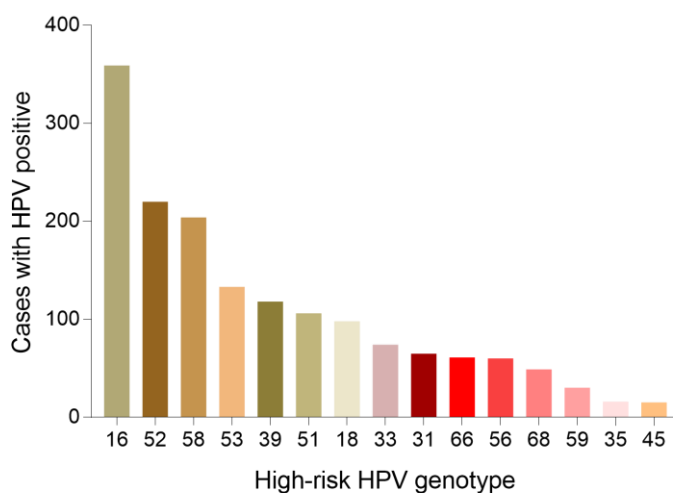


Figure 11. Genotype distribution of high-risk HPV infection in Shaanxi Province

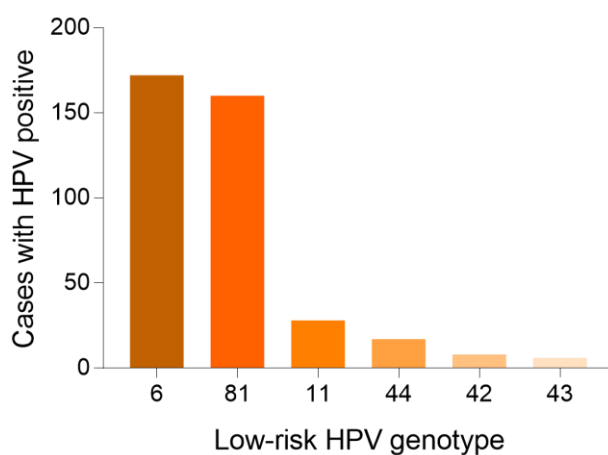


Figure12. Genotype distribution of low-risk HPV infection in Shaanxi Province

Table 1. Genotype distribution of the total population

	Genotype	Cases	Prevalence (%)
High-risk HPV	16	10974	19.93
	18	2998	5.44
	31	1971	3.58
	33	3200	5.81
	35	667	1.21
	39	3454	6.27
	45	630	1.14
	51	3013	5.47
	52	11282	20.49
	53	5513	10.01
	56	1419	2.58
	58	8007	14.54
	59	1276	2.32
	66	1749	3.18
68	2813	5.11	
Low-risk HPV	6	15515	28.17
	11	2082	3.78
	42	536	0.97
	43	691	1.25
	44	571	1.04
	81	5008	9.09

Table 2. Region stratified HPV prevalence in 55071 positive womens

Age Group, y	Total No (%)	2		3		4		6	
		Positive No	Positive, %	Positive No.	Positive, %	Positive No.	Positive, %	Positive No.	Positive, %
≤15	51(0.09%)	2	0.02%	21	0.06%	23	0.37%	5	0.38%
16-20	1391(2.53%)	125	1.23%	560	1.50%	696	11.19%	10	0.76%
21-25	4669(8.48%)	560	5.52%	2579	6.90%	1466	23.57%	64	4.87%
26-30	6597(11.98%)	1139	11.22%	4154	11.11%	1157	18.60%	147	11.20%
31-35	6681(12.13%)	1172	11.55%	4698	12.57%	640	10.29%	171	13.02%
36-40	8780(15.94%)	1473	14.51%	6568	17.57%	565	9.09%	174	13.25%
41-45	10924(19.84%)	2390	23.55%	7657	20.48%	657	10.56%	220	16.76%
46-50	7773(14.11%)	1557	15.34%	5578	14.92%	458	7.36%	180	13.71%
51-55	3959(7.19%)	847	8.34%	2706	7.24%	260	4.18%	146	11.12%
56-60	2041(3.71%)	334	3.29%	1433	3.83%	169	2.72%	105	8.00%
61-65	1320(2.40%)	308	3.03%	892	2.39%	71	1.14%	49	3.73%
66-70	542(0.98%)	133	1.31%	345	0.92%	33	0.53%	31	2.36%
71-75	239(0.43%)	73	0.72%	145	0.39%	13	0.21%	8	0.61%
≥76	104(0.19%)	37	0.36%	53	0.14%	11	0.18%	3	0.23%
Total	55071(100%)	2	0.02%	21	0.06%	23	0.37%	5	0.38%

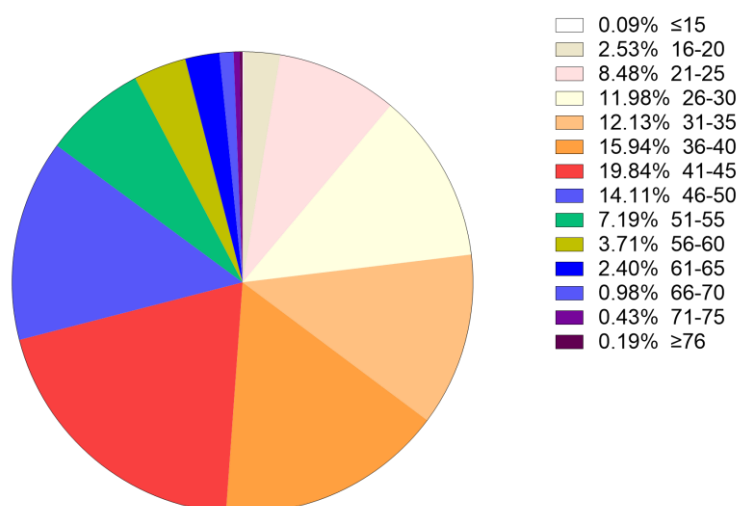
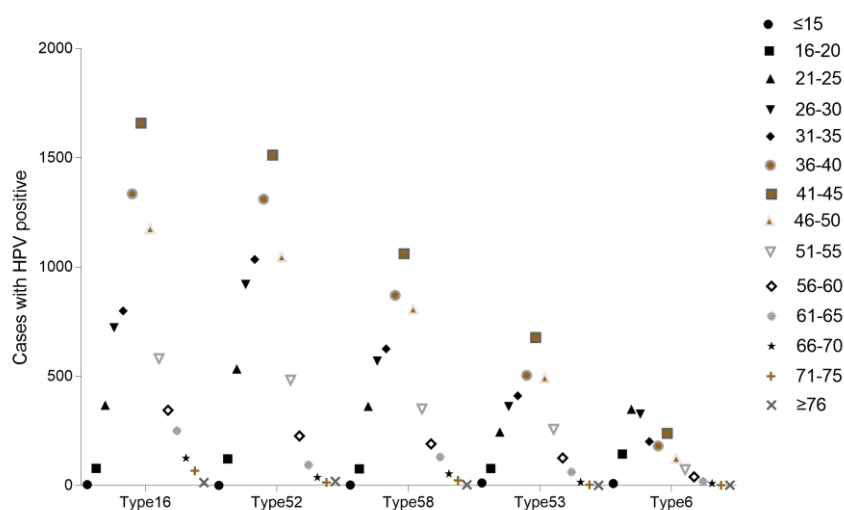
**Figure 13.** Pie chart of the age distribution in total population

Table 3. Age stratified HPV prevalence in 55071 positive women

Age Group, y	Total No.	High-risk only		Low-risk only		mixed	
		Positive No.	Positive, %	Positive No.	Positive, %	Positive No.	Positive, %
≤15	51	32	0.07%	13	0.19%	6	0.13%
16-20	1391	852	1.95%	313	4.53%	226	4.95%
21-25	4669	3266	7.49%	823	11.90%	580	12.69%
26-30	6597	5047	11.58%	993	14.36%	557	12.19%
31-35	6681	5418	12.43%	839	12.13%	424	9.28%
36-40	8780	7187	16.49%	968	13.99%	625	13.68%
41-45	10924	8955	20.55%	1275	18.43%	694	15.19%
46-50	7773	6313	14.48%	893	12.91%	567	12.41%
51-55	3959	3137	7.20%	450	6.51%	372	8.14%
56-60	2041	1633	3.75%	177	2.56%	231	5.06%
61-65	1320	1043	2.39%	107	1.55%	170	3.72%
66-70	542	428	0.98%	46	0.67%	68	1.49%
71-75	239	196	0.45%	12	0.17%	31	0.68%
≥76	104	78	0.18%	8	0.12%	18	0.39%
Total	55071	43585	86.30%	6917	13.70%	4569	9.05%

**Figure 14.** Age distribution of dominant HPV genotypes

3. Discussion

A pooled estimate among women worldwide showed that HPV prevalence in eastern Asia (which includes China) was significant higher than that of both southeastern Asia and south-central Asia where HPV prevalence were 13.6%, 6.2%, 7.5%, respectively[14]. In addition, the prevalence of HPV in less developed

countries (15.5%) was higher than that in more developed countries (10.0%), and the most the heavily burdened HPV regions of Asia are eastern Asia[15]. The number of HPV infection cases varies widely among eastern Asian countries, as the most populous developing country, China faces serious burden[16]. Moreover, it has been shown that the western China ranks as the first most mortality rate and the second most cervical cancer incidence rate in China, thus the primary prevention of cervical cancer of western China is particularly important[17].

In the present analysis, the overall HPV prevalence of 301,880 western Chinese women with normal cervical cytology was estimated to 18.01 %, which was higher than the average global level, lower than that of many other countries, such as Eastern Africa, Russia, and higher than that of Japan, India[14,15]. A large comparable population-based study of HPV genotype prevalence nationwide, showed a similar overall HPV infection of 21.07 % where 120,772 sample of 37 cities in China were tested[18], which was relatively higher than the overall HPV prevalence 18.01 % in this study. Regarding the HPV prevalence in different region groups, this present survey also showed the high HPV incidences, for Tibet Autonomous Region (28.14 %) and Chongqing Municipality (18.59 %), Guizhou Province (10.33 %) and Shaanxi Province (29.09 %). Compared with region-based data, the rates obtained in our study were different from than those previously reported neighboring regional data of Chongqing (26.20 %)[19], Guizhou (16.95 %)[20] and Yunnan (12.90 %)[21] of western China. The reported results of HPV prevalence varies from study to study as it possibly caused by several variables, including the large Chinese population composition as well as its territories. Together, the overall HPV-positive rate in the present study involving 301,880 cases was found to have increased slightly.

When stratified by HPV genotype, the most common HR-HPV types detected in our analysis were HPV52 (20.49 %), which was inconsistent with the previous data generated by some Chinese population-specific investigations and some related studies reported that HPV 16 was identified as the most common HR HPV genotype and a prevalence of 54.4 % in the world[15,18,22-24]. The other three most prevalent HR/IR-HPV types were HPV16, 58 and 53, with frequencies of 19.93%, 14.52%,

10.01%. The characteristics of HPV distribution in our study were similar to a recent HPV result identified in Guizhou, China[20]. However, a recent study enrolling 37,722 females showed that the most four prevalent genotypes were HPV 16 (3.79 %), HPV 52 (2.47 %), HPV 58 (1.76 %), HPV 53 (1.35 %)[17], which was different from our data. In addition, compared with a nationwide data of a Chinese population-based investigation from 37 cities, except for HPV53 (not reported), the infection rates of HPV16 (4.82 %), HPV52 (4.52 %), and HPV58 (2.74 %) were all lower than that in our population[18]. The knowledge of HPV prevalence and subtype distribution in different regions might facilitate in the development of vaccination program implementation. In this study, the HPV infection types and their proportions were different in different regions: the top three HPV genotypes were HPV 52, 16, 58 in Tibet Autonomous Region, HPV 16, 52, 58 in both Chongqing Municipality and Shaanxi Province, HPV 52, 16, 58 in Guizhou Province. Therefore, different regions showed diversity and have their respective proportions with respect to HPV genotypes. The differences in economic conditions, geographical cultural habits, migrations and other multiple factors might affect lifestyles among different populations makes the observed HPV prevalence varies.

Interesting, among HR-HPV genotype infections, HPV53 type infection accounted for the top four in our population (10.01 %). HPV53, a traditionally non-vaccine genotype, was recognized as a probable high-risk genotype and recently demonstrated to be associated with the putative potency of viral carcinogenicity (odds ratio, 3.92)[18]. Moreover, the prevalence of the HPV53 genotype gradually elevated between 2011 and 2015[23]. The HPV53, also reported as the fifth most common HPV types detected in eastern Africa and Central and northern America[15]. Thus, HPV prophylactic vaccines, including HPV53, may offer more sufficient protection for women in China.

In the present study, we found that the most three common LR-HPV types in this study were HPV6 (28.17 %), HPV81 (9.09 %) and HPV11 (3.78 %). A recent analysis enrolling 94,489 eastern China female conducted in 2019, has showed that the dominate LR-HPV genotypes were HPV81 (2.8%) and HPV6 (2.3%)[15].

Furthermore, a previous cross-sectional survey conducted in Arab women reported that HPV 81, 11 and 6 were the most commonly identified LR HPV genotypes in decreasing order[27], which were inconsistent with our research. The high HPV 6 prevalence in our study is unexpected. The reasons for this deviation of LR-HPV distribution are unclear and may be due to cultural range of the nationalities. Therefore, the HPV vaccine in China might also consider including the HPV6 genotype.

When regarding the HPV infection proportion of different age subgroups, in this current study, the age distribution showed that the middle age group (26 – 50 years) presented the highest HPV detection rate (75.53 %, 71.82 %, 62.75 %, respectively) among HR HPV only, LR HPV only and mixed HPV genotype subgroups, then all relatively lowest occurred in elderly individuals (≥ 76 years), similar to few recent studies in China[20,28]. The most frequent age group of our HPV screening population was 25–50 years, which could explain the reason why the HPV types were so distributed. Interestingly, in our present survey, we also concluded that the age distribution varied across different HPV genotype. With respect to the four top HR-HPV (type 52, 16, 58 and 53), the highest HPV detection rate was found in age group 41-45 years old, while the highest HPV detection rate in LR- HPV 6 infection was found in age group 21-30 years old group, which was different from a previous study about HPV genotyping results in China[22]. However, our result was consistent with the clinical phenomenon that more and more frequently diagnosis of cervical cancer was in middle/old-aged Chinese women (40–64 years old). Therefore, it has become particularly important to disseminate information about cervical diseases as well as perform HPV infection screening in China according to the HPV distribution characteristics of each age group.

Among the patients infected with multiple subtypes, co-infections with two HPV types was most common with percentage (76.04 %) in our population, which was comparable to the regional results of Guangzhou, Sichuan, and Macao in China from the previous reports[22]. Many epidemiologic researches found that infection with multiple HPV genotypes seemed to increase the risk of developing the tissue

abnormalities or high-grade lesions that precede invasive cervical cancer considerably, because HPV types might interact synergistically, which might contribute to increase the baseline risk observed with single-type infections[27, 32]. Similar results have been reached by other investigations that confirmed its association with development of the cervical carcinogenesis and increased the duration of the viral infection[33]. Therefore, it was suggested to consider diagnoses of co-infection with HPV into the prediction outcomes of HPV infections.

This is the first HPV distribution study with large sample size in China. The principal strength of this study was that the large data enrolling 0.3 million women come from western China, thus this population may be representative of the general women population in China. One limitation of this study was sampling bias, because the most frequent age group of our HPV screening strategy was 25–50 years. Secondly, the women we enrolled in this study attended to clinic for seeking medical advice based on routine gynecological examination and HPV prevalence results, they always accompanied by some clinical symptoms, which may lead to an over-reporting of HPV prevalence in this research. In addition, no clinical characters and cervical cytology results being collected as part of the study, resulting in the specific risk factors for the cervical cancer and the correlation between HPV genotype and cervical cancer or precancerous lesions were unable to be accurately examined. Furthermore, a follow-up study should be conducted to track changes in genotype, cervical pathology and cytology as there was a close relationship between cervical carcinoma and long-term persistent HR-HPV infections.

4. Materials and methods

4.1. Ethical considerations, study population and sample collection

This investigation was approved (no. 2020-173) by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University Ethics Review Board and all participants provided informed consent for inclusion in the study. Briefly, this retrospective study involved 301880 samples from more than 8 clinical hospitals, women's health centers, clinics and physical examination centers located in 4 different provinces of China. For each woman, the HPV genotyping results and

relevant clinical information, including age and regional data were all collected.

4.2. DNA extraction and HPV genotyping

Exfoliated cervical cell samples were collected from women by gynecologists or obstetricians using a specialized cervical sampler brush and used for DNA isolation and purification. Based on rapid flow-through hybridization of nucleic acid molecules, a gene chip detection system for identification nucleic acids of 21 HPV types was provided by Kaipu Biochemical Company in Chaozhou, Guangdong, China. Among 21 HPV types, there were 14 HR-HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68), 6 LR-HPV types (6, 11, 42, 43, 44, and 81), and 1 IR-HPV type (53). Subsequently, all HPV tests were performed with a HPV genotyping panel (polymerase chain reaction (PCR)-reverse dot blot hybridization method). Briefly, according to manufacturer's instructions, PCR was performed in a 25- μ L reaction mixture containing 1- μ L extracted DNA, 0.75- μ L DNA Taq polymerase, and 23.25- μ L PCR-mix solution containing primer system. The PCR cycling parameters were as follows: an initial step at 95°C for 9 min, and followed by 40 amplification cycles (denaturation at 95°C for 20 s, annealing at 55°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 5 min). After amplification, HPV genotyping was performed by hybridization and RDB on the strips fixed with HPV type-specific probes. According to the manufacturer's instructions, the HPV type-specific probes immobilized on nylon membranes were used for reverse-blot hybridization and detection of all HPV genotypes in a single assay. Simultaneously, in order to validate the HPV test, sterile water was used as the negative control, and specimens with known HPV genotypes were used as the positive control.

4.3. Statistical analysis

Data were analyzed with IBM SPSS version 21.0. Sample characteristics including age (women with unknown ages were excluded), regionality, HPV infection result and genotypes distribution characters were summarized using frequency distributions to generate the numbers and percentages. The chi-square test was used to compare the HPV prevalence or proportions between different groups. A two-sided P-value of less than 0.05 was considered statistically significant.

5. Conclusion

This present study represented one of the most comprehensive study of the prevalence and genotype distribution of different HPV types in China to date. We examined the epidemiology of HPV infection in China and confirmed the high overall HPV prevalence of 18.01 % of all tested female. Moreover, in addition to common genotype HPV52, 16 and 58, particular attention should be paid to the high prevalence of non-vaccine genotypes (e.g. HPV53, HPV6) in China. Therefore, the future next-generation HPV prophylactic vaccines in China should also consider to include more HPV types (e.g. HPV53, HPV6). Regarding the age-specific distribution of HPV, the highest HPV detection rate was found in age group 26–50 years old (detection rate = 74.00 %). These results showed that the main proportion of HPV infections in China might be taking place in middle age, which remind us to pay more attention to middle age women in the prevention and control of HPV. Thus, our results indicate that the preventative strategies including HPV vaccine-based popularization and related educational campaigns should start in 25 years old female. This information might provide valuable information for estimation of the potential clinical benefit of HPV -based screening in China.

Authors' contributions: Yan Dong, Jiao Li, Dan Wang and Li Xu carried out the sample collection, laboratory detection and drafted the manuscript. Ling Chen participated in the design of the study and performed the statistical analysis. Huandong Liu, Jungao Lu, Zuoyi Yao and Xiaosong Li conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (grant no. 81871653) and Chongqing Medical Science Project (grant no. 2018MSXM065).

Acknowledgments: The authors would like to thank the patients who participated in this study, and their families, as well as the investigators and study teams.

Conflicts of Interest: The authors declare that they have no competing interests.

Abbreviations: HPV: Human papillomavirus; HR- HPV: High-risk HPVs; IR -HPV: intermediate-risk HPVs; LR -HPV: Low-risk HPVs.

Availability of data and materials: All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate: This study was approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University Ethics Review Board (no. 2020-173).

Consent for publication: Not applicable.

References

1. Freddie; Bray; Jacques; Ferlay; Isabelle; Soerjomataram; Rebecca; L; Siegel; Lindsey, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **2018**.
2. Forman, D.; De Martel, C.; Lacey, C.J.; Soerjomataram, I.; Lortet-Tieulent, J.; Bruni, L.; Vignat, J.; Ferlay, J.; Bray, F.; Plummer, M., Global Burden of Human Papillomavirus and Related Diseases. *Vaccine* **2012**, 30, F12-F23.
3. Schwalbe, N., Human Papillomavirus (HPV) and Related Cancers in the Global Alliance for Vaccines and Immunization (GAVI) Countries. A WHO/ICO HPV Information Centre Report Foreword. *Vaccine* **2012**, 30, (4), X-X.
4. Shi, J.F.; Qiao, Y.L.; Smith, J.S.; Dondog, B.; Bao, Y.P.; Dai, M.; Clifford, G.M.; Franceschi, S., Epidemiology and Prevention of Human Papillomavirus and Cervical Cancer in China and Mongolia. *Vaccine* **2008**, 26, (supp-S12).
5. Torre, L.A.; Bray, F.; Siegel, R.L.; Ferlay, J.; Lortet Tieulent, J.; Jemal, A., Global cancer statistics, 2012. *CA Cancer J Clin* **2015**, 65.
6. Arbyn, M.; Weiderpass, E.; Bruni, L.; de Sanjosé, S.; Bray, F., Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Global Health* **2019**, 8, (2).
7. Siddiqa, A.; Broniarczyk, J.; Banks, L., Papillomaviruses and Endocytic Trafficking. *Int J Mol Sci* **2018**, 19, (9).
8. Baseman, J.G.; Koutsky, L.A., The epidemiology of human papillomavirus infections. *J Clin Virol* **2005**, 32, (supp-S), 0-24.
9. Chan, C.K.; Aimagambetova, G.; Ukybassova, T.; Kongrtay, K.; Azizan, A., Human Papillomavirus Infection and Cervical Cancer: Epidemiology, Screening, and Vaccination-Review of Current Perspectives. *J Oncol* **2019**, 2019, 3257939.
10. Mu Oz, N.; Bosch, F.X.; Castellsagué, X.; Díaz, M.; Sanjose, S.D.; Hammouda, D.; Shah, K.V.; Meijer, C.J.L.M., Against which human papillomavirus types shall we vaccinate and screen? the international perspective. *Int J Cancer* **2004**, 111, (2), 278-285.
11. Bouvard, V.; Baan, R.; Straif, K.; Grosse, Y.; Cogliano, V., A Review of Human Carcinogens—Part B: Biological Agents. *Lancet Oncol* **2009**, 10, (4), 321-322.
12. De Martel, C.; Plummer, M.; Vignat, J.; Franceschi, S., Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer* **2017**.
13. Frati, E.R.; Fasoli, E.; Martinelli, M.; Colzani, D.; Bianchi, S.; Carnelli, L.; Amendola, A.; Olivani,

- P.; Tanzi, E., Sexually Transmitted Infections: A Novel Screening Strategy for Improving Women's Health in Vulnerable Populations. *Int J Mol Sci* **2017**, *18*, (6).
14. Emma; J; Crosbie; Mark; H; Einstein; Silvia; Franceschi; Henry; C, Human papillomavirus and cervical cancer. *Lancet* **2013**.
15. de Sanjosé, S.; Diaz, M.; Castellsagué, X.; Clifford, G.; Bruni, L.; Muñoz, N.; Bosch, F.X., Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. *Lancet Infect Dis* **2007**, *7*, (7), 453-9.
16. Setiawan, D.; Luttjeboer, J.; Pouwels, K.B.; Wilschut, J.C.; Postma, M.J., Immunogenicity and safety of human papillomavirus (HPV) vaccination in Asian populations from six countries: a meta-analysis. *Jpn J Clin Oncol* **2017**, *47*, (3), 265-276.
17. Wang, J.; Tang, D.; Wang, K.; Wang, J.; Zhang, Z.; Chen, Y.; Zhang, X.; Ma, C., HPV genotype prevalence and distribution during 2009-2018 in Xinjiang, China: baseline surveys prior to mass HPV vaccination. *BMC Womens Health* **2019**, *19*, (1), 90.
18. Wang, R.; Guo, X.L.; Wisman, G.B.; Schuurin, E.; Wang, W.F.; Zeng, Z.Y.; Zhu, H.; Wu, S.W., Nationwide prevalence of human papillomavirus infection and viral genotype distribution in 37 cities in China. *BMC Infect Dis* **2015**, *15*, 257.
19. Tang, Y.; Zheng, L.; Yang, S.; Li, B.; Su, H., Epidemiology and genotype distribution of human papillomavirus (HPV) in Southwest China: a cross-sectional five years study in non-vaccinated women. *Virol J* **2017**, *14*, (1).
20. Chen, Z.; Li, Q.; Huang, Q.; Liu, H.; Luo, Q., Characteristics of human papillomaviruses distribution in Guizhou Province, China. *Virol J* **2019**, *16*, (1).
21. Li, Z.; Liu, F.; Cheng, S.; Shi, L.; Yan, Z.; Yang, J.; Shi, L.; Yao, Y.; Ma, Y., Prevalence of HPV infection among 28,457 Chinese women in Yunnan Province, southwest China. *Sci Rep* **2016**, *6*, 21039.
22. Zeng, Z.; Yang, H.; Li, Z.; He, X.; Griffith, C.C.; Chen, X.; Guo, X.; Zheng, B.; Wu, S.; Zhao, C., Prevalence and Genotype Distribution of HPV Infection in China: Analysis of 51,345 HPV Genotyping Results from China's Largest CAP Certified Laboratory. *J Cancer* **2016**, *7*, (9), 1037-43.
23. Chen, X.; Xu, H.; Xu, W.; Zeng, W.; Liu, J.; Wu, Q.; Zhao, X.; Jiang, T., Prevalence and genotype distribution of human papillomavirus in 961,029 screening tests in southeastern China (Zhejiang Province) between 2011 and 2015. *Sci Rep* **2017**, *7*, (1), 14813.
24. Wu, C.; Kajitani, N.; Schwartz, S., Splicing and Polyadenylation of Human Papillomavirus Type 16 mRNAs. *Int J Mol Sci* **2017**, *18*, (2).
25. Halc, G.; Alemany, L.; Lloveras, B.; Schmitt, M.; Alejo, M.; Bosch, F.X.; Tous, S.; Klaustermeier, J.E.; Guimerà, N.; Grabe, N., Pathogenic role of the eight probably/possibly carcinogenic HPV types 26, 53, 66, 67, 68, 70, 73 and 82 in cervical cancer. *J Pathol* **2014**, *234*, (4), 441-451.
26. Jiang, L.; Tian, X.; Peng, D.; Zhang, L.; Xie, F.; Bi, C.; Wang, R.; Wang, J.; Qi, D., HPV prevalence and genotype distribution among women in Shandong Province, China: Analysis of 94,489 HPV genotyping results from Shandong's largest independent pathology laboratory. *PLoS One* **2019**, *14*, (1), e0210311-e0210311.
27. Bansal, D.; Elmi, A.A.; Skariah, S.; Haddad, P.; Abu-Raddad, L.J.; Al, H.A.; Mohamed-Nady, N.; Affifi, N.M.; Ghedira, R.; Hassen, E.; Al-Thani, A.A.; Al-Ansari, A.A.; Sultan, A.A., Molecular epidemiology and genotype distribution of Human Papillomavirus (HPV) among Arab women in the State of Qatar. *J Transl Med* **2014**, *12*, 300.
28. Zhao, P.; Liu, S.; Zhong, Z.; Hou, J.; Lin, L.; Weng, R.; Su, L.; Lei, N.; Hou, T.; Yang, H., Prevalence and genotype distribution of human papillomavirus infection among women in northeastern

- Guangdong Province of China. *BMC Infect Dis* **2018**, 18, (1), 204.
29. Li, K.; Yin, R.; Li, Q.; Wang, D., Analysis of HPV distribution in patients with cervical precancerous lesions in Western China. *Medicine (Baltimore)* **2017**, 96, (29), e7304.
30. Zeng, X.X.; Yan, L.X.; Huang, X.X.; He, C.H.; Liu, Z.X., Prevalence and genotype distribution of human papillomavirus among Hakka women in China. *Ann Transl Med* **2016**, 4, (15), 276.
31. Prevalence and genotype distribution of cervical human papillomavirus infection in Macao. *J Med Virol* **2010**, 82, (10), 1724-1729.
32. Shen, Y.; Gong, J.M.; Li, Y.Q.; Gong, Y.M.; Lei, D.M.; Cheng, G.M.; Li, X.F., Epidemiology and genotype distribution of human papillomavirus (HPV) in women of Henan Province, China. *Clin Chim Acta* **2013**, 415, (none), 297-301.
33. Schmitt, M.; Depuydt, C.; Benoy, I.; Bogers, J.; Antoine, J.; Arbyn, M.; Pawlita, M., Multiple Human Papillomavirus Infections with High Viral Loads Are Associated with Cervical Lesions but Do Not Differentiate Grades of Cervical Abnormalities. *J Clin Microbiol* **2013**, 51, (5), 1458-1464.