

RESEARCH

Open Access



# Exploring the dynamics and interplay of human papillomavirus and cervical tumorigenesis by integrating biological data into a mathematical model

Wenting Wu<sup>1</sup>, Lei Song<sup>2</sup>, Yongtao Yang<sup>3</sup>, Jianxin Wang<sup>4</sup>, Hongtu Liu<sup>5\*</sup> and Le Zhang<sup>1,6\*</sup>

From 6th International Work-Conference on Bioinformatics and Biomedical Engineering  
Granada, Spain. 25-27 April 2018

## Abstract

**Background:** Cervical cancer is the fourth most common tumor in women worldwide, mostly resulting from high-risk human papillomavirus (HR-HPV) with persistent infection.

**Results:** The present discoveries are comprised of the following: (i) A total of 16.64% of the individuals were positive for HR-HPV infection, with 13.04% having a single HR-HPV type and 3.60% having multiple HR-HPV types. (ii) Cluster analysis showed that the infection rate trends of HPV31 and HPV33 in all infections as well as HPV33 and HPV35 in single infections in precancerous stages were very similar. (iii) The single/multiple infection proportions of HR-HPV demonstrated a trend that the multiple infections rates of HR-HPV increased as the disease developed.

**Conclusions:** The HR-HPV prevalence in outpatients was 16.64%, and the predominant HR-HPV types in the study were HPV52, HPV58 and HPV16. HR-HPV subtypes with common biological properties had similar infection rate trends in precancerous stages. Especially, as the disease development of precancer evolved, defense against HPV infection broke, meanwhile, the potential of more HPV infection increased, which resulted in increase of multiple infections of HPV.

**Keywords:** Cervical tumorigenesis, Human papillomavirus, Cluster analysis, Poisson regression

## Background

Cervical cancer is the fourth most common tumor in women worldwide [1], mostly resulting from human papillomavirus (HPV). The current report [2] estimates that cancer cases related to HPV infection account for 4.5% of the total number of new cancers worldwide, of which cervical cancer accounts for 83% of these HPV

infection-related cancers, posing a great threat to women's health, especially in developing countries. Thirteen HPV genotypes denoted as high-risk HPV (HR-HPV) are essential factors for cervical tumorigenesis [2], so the dynamics of the HPV genotypes described here reflect the relationship between an individual HPV genotype and the development of cervical cancer, i.e., precancerous stages. Since it takes approximately 20 years for the carcinogenesis of HR-HPV with persistent infection [3], cervical cancer is the only malignant tumor that can be prevented and treated early through HPV-type screening, which plays a significant role in improving

\* Correspondence: [liuht@chinacdc.cn](mailto:liuht@chinacdc.cn); [zhangle06@scu.edu.cn](mailto:zhangle06@scu.edu.cn)

<sup>5</sup>National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

<sup>1</sup>College of Computer Science, Sichuan University, Chengdu 610065, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the prognosis of patients [3–5]. As China has become one of the countries with a high incidence of cervical cancer and HPV infection is widespread in females [6], it is very important to investigate HPV-type infections for Chinese population.

Generally, both the liquid-based cytology test (LCT) and the ThinPrep cytology test (TCT) are used to screen cervical cancer, but they do not effectively detect specific HR-HPV genotypes in infections. It is noted that we usually use the commercial names of LCT and TCT to represent the cytology tests, since they are from different manufacturers. However, HPV genotyping can easily detect the genotypes better than the LCT and TCT methods, and distinguish the difference between single infection (Denoted by Table 1) and multiple infections (Denoted by Table 1), once HPV infection occurs. Since the handicap of using cytology tests or HPV genotyping alone, we always employ the combination of cytology tests and HPV genotyping based on the significantly high sensitivity and lower false-negative rates achieved [7–11]. For example, Catteau et al. [12] calculated the prevalence rates of 13 HR-HPV-type infections in different precancerous stages among Belgian women, and Ying et al. [13] mainly employed the prevalence rates of different HR-HPV types in all precancerous stages to describe the distribution of the major infectious types in Beijing China after collecting the related data by both LCT/TCT and HR-HPV genotyping methods. Nevertheless, reported by the previous studies [12, 13], the prevalence of each HR-HPV type are different in the same precancerous stage. And for the same kind of HR-HPV type, the prevalence is not consistent in different precancerous stages [13]. Therefore, it is inaccurate to employ the total precancerous stage data to describe the relationship between HR-HPV types and precancerous stages.

Furthermore, it is still unclear whether multiple infections are more risky than single infection of HR-HPV [13–15]. For example, Chaturvedi et al. [14] investigated the coinfection patterns of 25 HPV genotypes and computed the odds ratios for each genotype with 24 other genotypes. The results showed that the disease risk of multiple infections is close to the total estimated risk of

individual infections. However, both Ying et al. [13] and Dickson et al. [15] indicated that women with multiple infections have a significantly higher risk of cervical disease than women with single infections. Since previous research collect data with different HPV types, patient ages and other related factors [13–15], they result in inconsistent conclusions for the risk of cervical lesions caused by multiple infections [16, 17]. Moreover, most previous studies [13–15, 18] employed cohort analysis for cervical cancer without considering the proportion of single and multiple infections in different precancerous stages for different HPV types as well as the impact of HR-HPV genotypes and precancerous stages on the infections.

Regarding to the previous shortcomings, we develop three innovations to overcome them: (1) we collected the clinical data for 13 HR-HPV types in 4 precancerous stages by integrating TCT into HR-HPV genotype detection; (2) we performed cluster analysis for 13 genotypes in 4 precancerous stages; (3) we investigated the proportion of single/multiple infections at 4 precancerous stages for each HR-HPV genotype, and explore the impact of HR-HPV genotypes and precancerous stages on the infections by Poisson regression [19].

A total of 16,693 patients were studied from July 2016 to July 2017 in the outpatient department of the General Hospital of the People's Liberation Army. We first statistically analyzed infection data for 13 HR-HPV types in 4 precancerous stages. The results showed that the overall prevalence rate of the 13 HR-HPV types (16.64%) is less than the previous, but HPV52, HPV58 and HPV16 still have the greatest impact on the health of women in China. Next, we found that biological homology results in similar infection rate trends in precancerous stages by our k-means [20] cluster analysis. Finally, we not only found that the multiple infection proportion of HR-HPV increased as the disease developed, but also demonstrated that only the precancerous stages were statistically significant [19] by considering the impact of both HR-HPV genotypes and precancerous stages on infection results. Finally, we discuss the limitations and future study.

**Table 1** Nomenclature

Terms or equations	Description
All infections	It includes all subjects with HPV infections containing a specified HR-HPV genotype
Single infection	It refers to such infection with one HR-HPV genotype alone
Multiple infections	It refers to the difference between all infections and single infection for a HR-HPV type, i.e., HPV infected with two or more HR-HPV genotypes
All_Infection_Set	The dataset for all infections numbers of the 13 HR-HPV types in the four precancerous stages (see Additional file 1)
Single_Infection_Set	The dataset for single infection numbers of the 13 HR-HPV types in the four precancerous stages (see Additional file 1)
Multiple_Infection_Set	The dataset for multiple infections numbers of the 13 HR-HPV types in the four precancerous stages (see Additional file 1)

**Methods**

**Pathological examination**

Cervical cells were detected by TCT and the results of cytological pathology were diagnosed by senior physicians according to the Bethesda System of cervical cytology [21]. The precancerous stages are classified as follows [21]: (1) Normal; (2) Atypical squamous cells of undetermined significance (ASC-US); (3) Low-grade squamous intraepithelial lesions (LSIL); and (4) High-grade squamous intraepithelial lesions (HSIL).

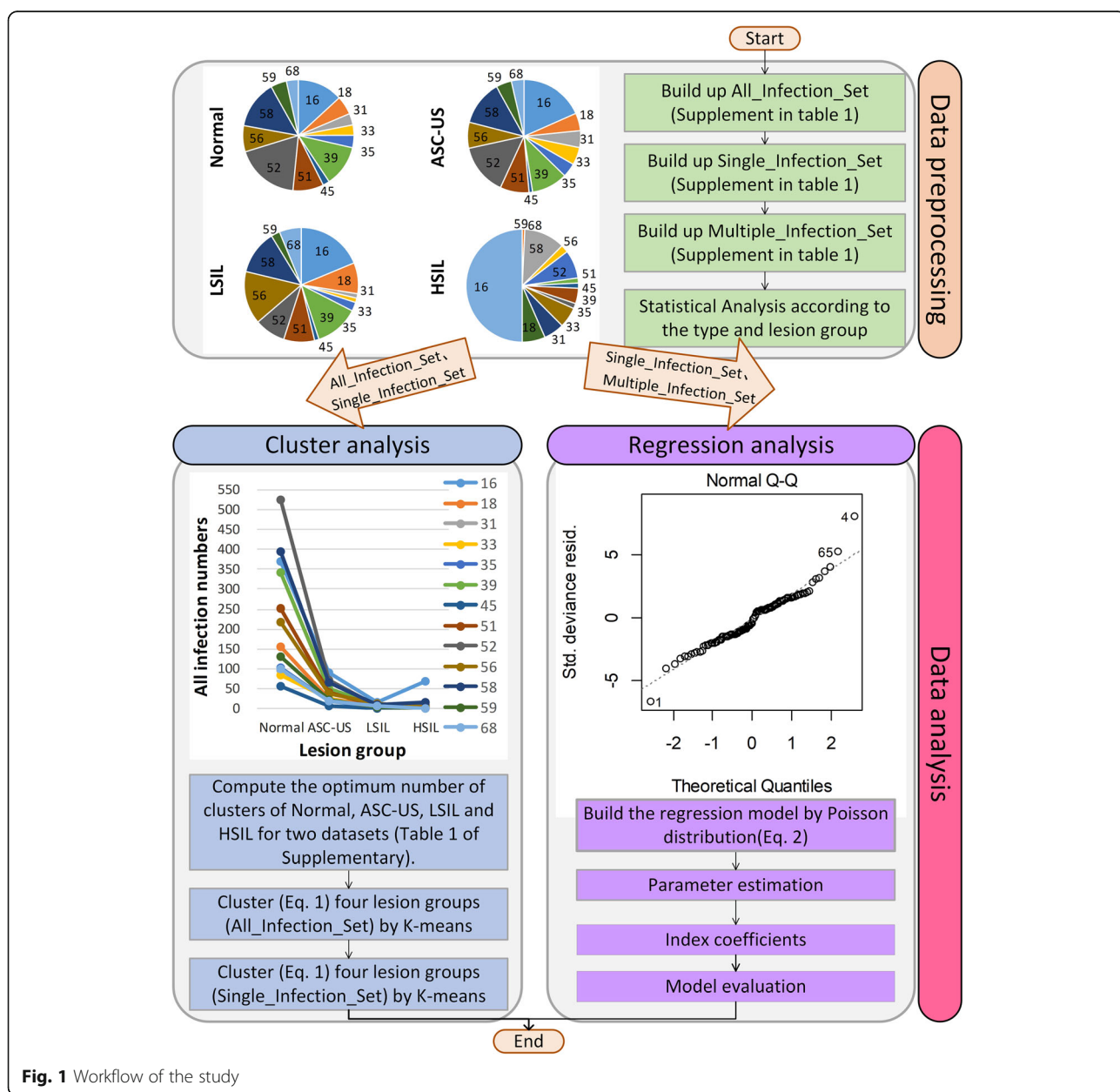
**Detection of HPV genotypes**

Thirteen HR-HPV genotypes (HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, HPV59 and HPV68)

HPV56, HPV58, HPV59 and HPV68) were detected with the real-time polymerase chain reaction kit for high-risk HPV genotypes from Shanghai ZJ biotechnology Company ([http://www.liferiver.com.cn/productinfo/p15\\_62.html](http://www.liferiver.com.cn/productinfo/p15_62.html)). The specific steps are strictly in accordance with the instructions of the kit. If the viral load of HPV-DNA was greater than or equal to  $10^4$  copies/ml, it was positive, otherwise it was negative.

**Research subjects**

The data in this study are from 16,693 patients who all underwent biopsies in the outpatient department of the General Hospital of the People's Liberation Army from July 2016 to July 2017. The cervical



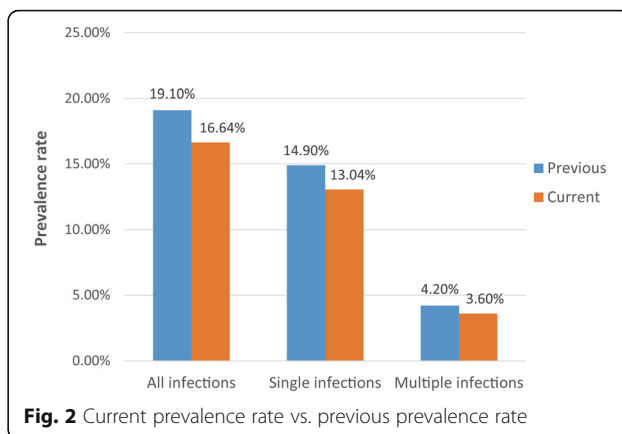
**Fig. 1** Workflow of the study

samples were collected and detected with Riverlife Bio kits ([http://www.liferiver.com.cn/productinfor/p15\\_62.html](http://www.liferiver.com.cn/productinfor/p15_62.html)). Here, we have 15,706 Normal, 785 ASC-US, 69 LSIL and 133 HSIL cases. Also, we conducted quantitative detection for 13 HR-HPV subtypes in 16,693 cases to diagnose specific infection of HR-HPV genotypes.

**Workflow of the study**

Figure 1 and Table 1 describe the workflow of the study and the nomenclature, respectively. The workflow consists of Data preprocessing and Data analysis steps. Data preprocessing step process the raw datasets for all infections (denoted by Table 1), single infection and multiple infections of the 13 HR-HPV types in the four precancerous stages by using a pie chart (left panel of data preprocessing component in Fig. 1) to describe the classical statistical analysis results.

Data analysis (Fig. 1) comprises cluster and regression analysis. Although the 13 HR-HPV genotypes are biologically independent, some of them may have common biological properties, resulting in similarity in the phenotypes (i.e., similarity in the number of infected people). Therefore, we use cluster analysis (left panel of data analysis in Fig. 1) to investigate the similarity of infection for 13 different HR-HPV types in precancerous stages. It is well known that the cluster analysis [22] consists of hierarchical clustering and nonhierarchical clustering. Since the aim of the study is to investigate which HR-HPV types have similar infections in the precancerous stages, we consider that the classical K-means [20] is suitable for this study. Here, K-means uses Euclidean distance (Eq. 1) [23] to measure the distance between two observed values:



**Fig. 2** Current prevalence rate vs. previous prevalence rate

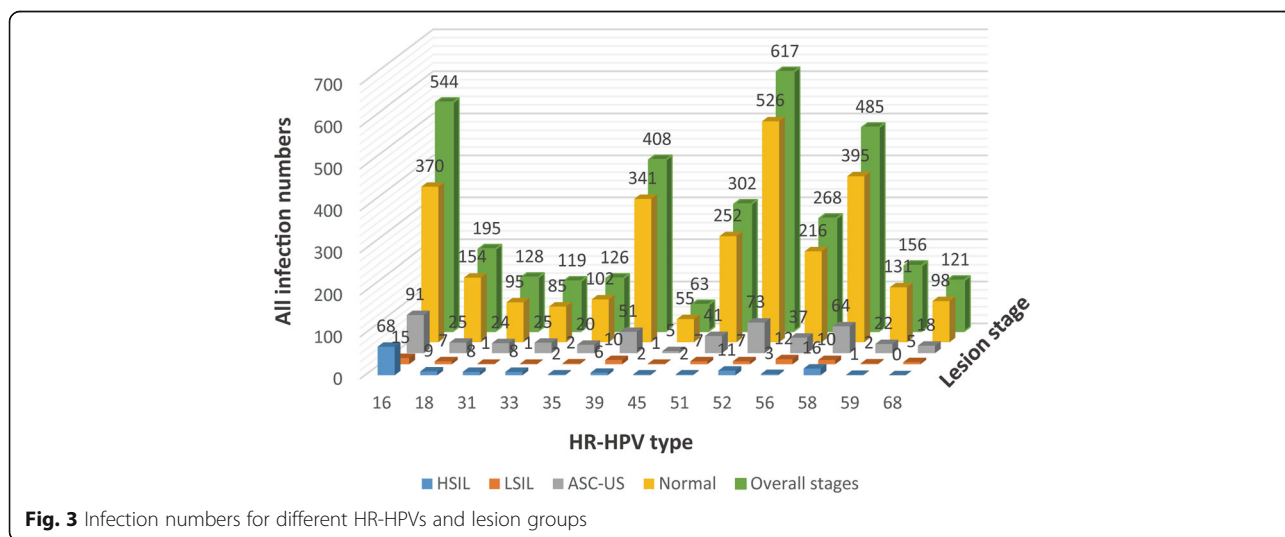
$$d_{ij} = \sqrt{(x_i - x_j)^2} \tag{1}$$

$d_{ij}$  represents the distance between observations of  $i$ th and  $j$ th HR-HPV genotypes.  $x_i$  and  $x_j$  represent the number of infected  $i$ th and  $j$ th HR-HPV genotypes, respectively.

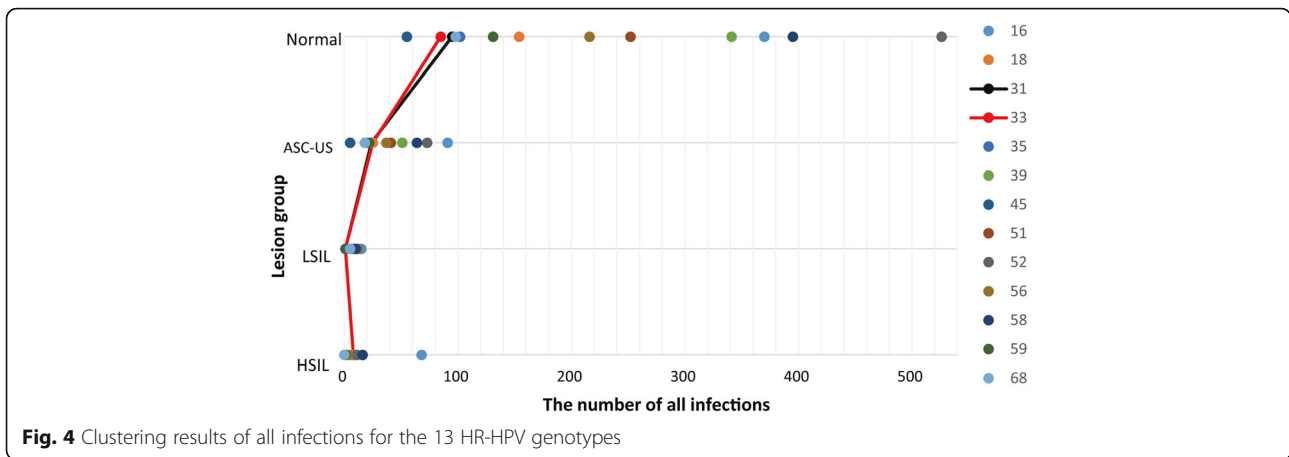
Currently, Poisson regression is widely used for clinical data analysis [24]. For instance, Rochon et al. [25] used Poisson regression analysis to study the number of rejection reactions in patients after transplantation within a certain time, and Vonesh et al. [26] analyzed the potential risk factors related to the number of peritoneal bacterial infections. Here, we used Poisson regression (Eq. 2) to investigate the impact of HR-HPV genotypes and precancerous stages on infection [23].

$$\log_e(\lambda) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{2}$$

Here, we set the infection number ( $\lambda$ ) as the outcome variable and  $\log_e(\lambda)$  as the connection function in R



**Fig. 3** Infection numbers for different HR-HPVs and lesion groups



software [27].  $X_1$  and  $X_2$  respectively represent the HR-HPV genotypes and precancerous stages as the prediction variables.  $\beta_0$  is the intercept.  $\beta_1$  and  $\beta_2$  are coefficients for prediction variables.

**Results**

**Comparison of the prevalence rates**

Figure 2 shows the comparison of the prevalence rates of the 13 HR-HPV types in all infections, single infection and multiple infections for the previous [13] and current study.

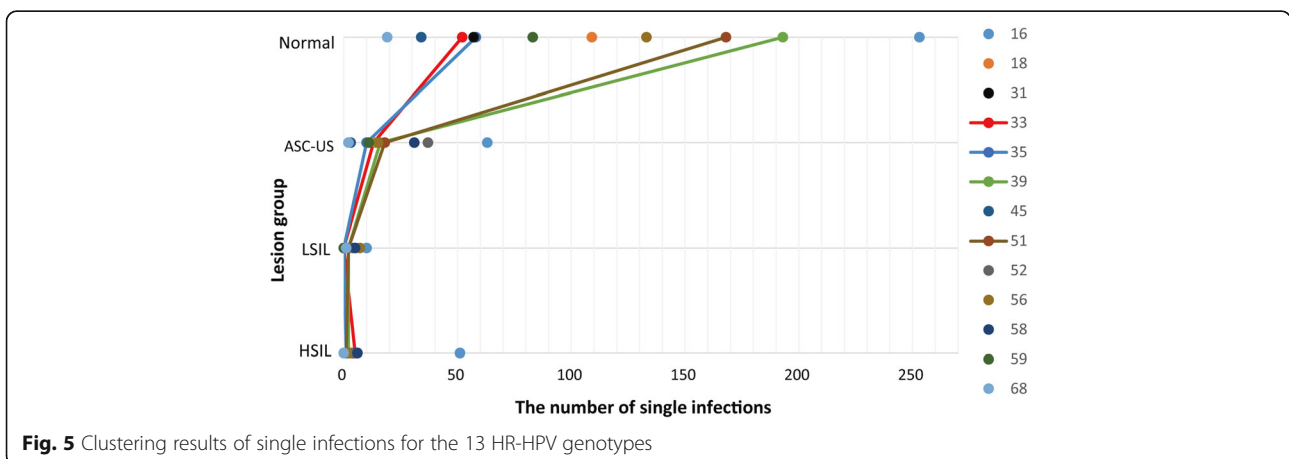
Figure 3 shows the dynamics of infections for each HR-HPV type at different precancerous stages.

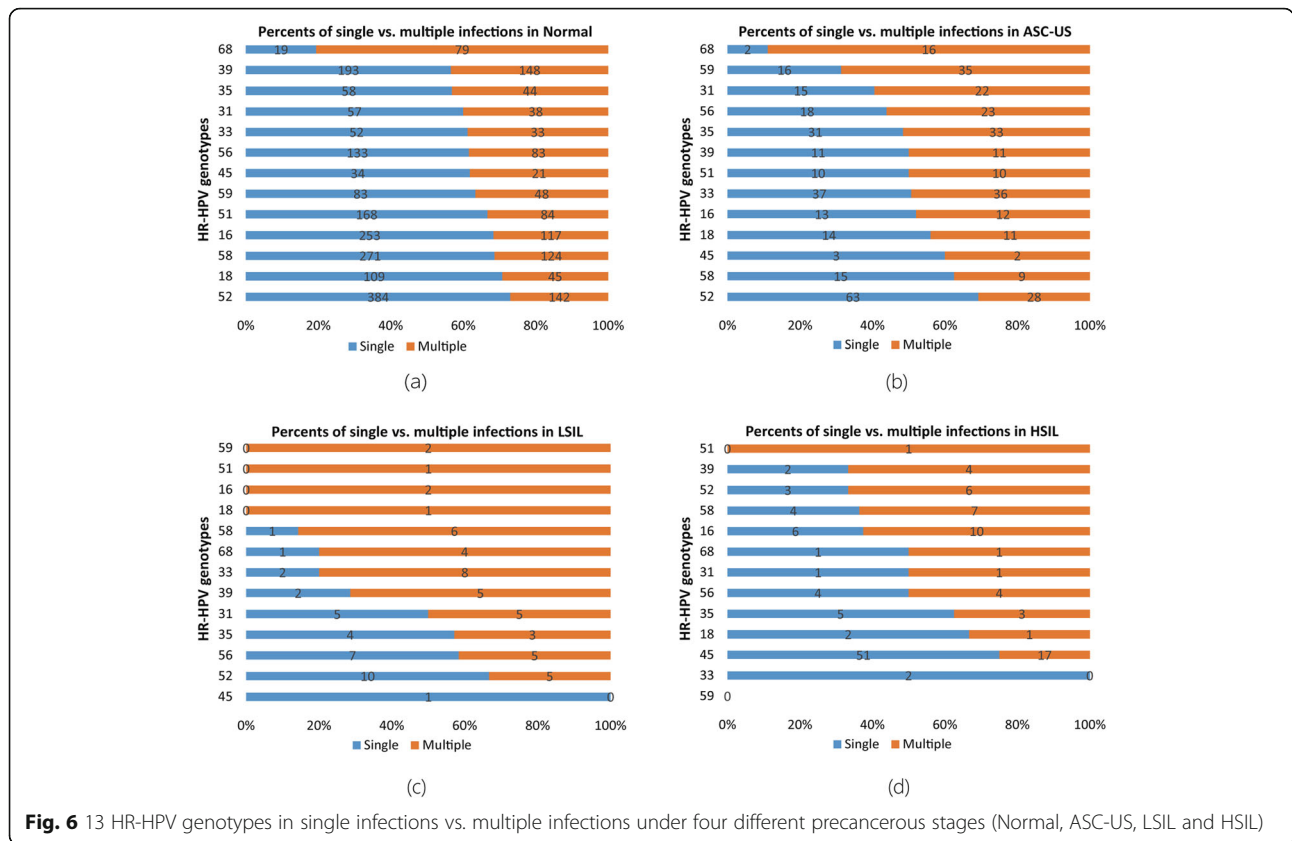
**Cluster analysis of all infections and single infections**

For all infections, Fig. 4 demonstrates that the infection trends of HPV31 and HPV33 are similar in the four precancerous stages. Regarding single infections, Fig. 5 shows that not only HPV39 and HPV51 but also HPV33 and HPV35 have similar infection trends in the 4 precancerous stages.

**The impact of HR-HPV genotypes and precancerous stages on infection**

Figure 6 describes the proportion of single and multiple infections for each HR-HPV genotype in different precancerous stages. Figure 6a demonstrates that the proportion of multiple infections for 12 HR-HPV genotypes is less than their single infection except for HPV68 under Normal stage. Figure 6b demonstrates that the proportion of multiple infections for 6 HR-HPV genotypes (HPV52, HPV58, HPV45, HPV18, HPV16 and HPV33) is less than their single infection under ASC-US stage. Figure 6c demonstrates that the proportion of multiple infections for 4 HR-HPV genotypes (HPV35, HPV56, HPV52 and HPV45) is less than their single infection under LSIL stage. Figure 6d demonstrates that the proportion of multiple infections for 4 genotypes (HPV35, HPV18, HPV45 and HPV33) is less than their single infection under HSIL stage. Next, the Poisson regression analysis demonstrates that only the precancerous stages are statistically significant, while the HR-HPV genotypes are not (Table 2).





**Fig. 6** 13 HR-HPV genotypes in single infections vs. multiple infections under four different precancerous stages (Normal, ASC-US, LSIL and HSIL)

**Discussion**

Since Fig. 2 shows that current total prevalence rate of the 13 HR-HPV types is less than the previous, we consider that the prevalence of both single and multiple infections was decreasing during these years in China. Additionally, since Fig. 3 shows that the prevalence rates of these HR-HPV types are inversely proportional to the severity of cervical lesions, we consider that most patients infected with HR-HPV types are in the early lesion stage (especially the squamous epithelial cells were still in the Normal stage). Thus, we have plenty of room to reduce the prevalence of HR-HPV types in China and should pay more attention to promoting cervical screening and HPV vaccine research.

Furthermore, both previous studies [3, 13] and Fig. 3 indicate that the top three greatest HR-HPV types threatening to China are HPV52, HPV58 and HPV16. Figure 3 also demonstrates that neither the proportion of the 13 HR-HPV types in the same precancerous stage nor the infection rate in different stages for the same HR-HPV type is similar, which implies that the infection of different HR-HPV types in different pathological stages is not consistent. Therefore, it is better to describe the phenomenon of HR-HPV infections in different precancerous stages and HR-HPV types, but not using the overall prevalence rate of HR-HPV types for each precancerous stage [12].

Moreover, the blue part in Fig. 3 indicates that the prevalence rates of HPV16, HPV58, HPV52 and HPV18 are greater than those of other types under the HSIL stage which are easily transformed into cervical cancer.

For cluster analysis, Fig. 4 demonstrates that the infection trends of HPV31 and HPV33 in the four precancerous stages are very similar in all infections. For single infection, Fig. 5 shows that the infection trends of HPV33 and HPV35 in the four precancerous stages are very similar. Since Villiers et al. [28] previously reported that high-risk subtypes such as HPV31, HPV33, HPV35, HPV52, HPV16 and HPV58 belong to alpha-papillomavirus ninth species, our results show that such HR-HPV subtypes with common biological properties could have similar infection rate trends in precancerous stages. As we described previously, the Normal, ASC-US, LSIL and HSIL are in a disease development in order respectively. Figure 6 implies that as the disease

**Table 2** Parameter estimation for Poisson regression

Parameters	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Genotypes	25	68,914	2757	1.58	0.0672
Precancerous	3	196,329	65,443	37.51	6.49e-15*
Residuals	75	130,868	1745		

**Note.** \*Pr < 0.05

development of precancer evolves, defense against HPV infection breaks, meanwhile, the potential of more HPV infection increases, which results in increase of multiple infection of HPV. In addition, only the precancerous stages are statistically significant when considering the impact of HR-HPV genotypes and precancerous stages on infection by regression analysis (Table 2).

## Conclusions

In conclusion, the overall prevalence rate of the 13 HR-HPV types (16.64%) is less than the previous, which results from the efforts to popularize knowledge of the high-risk HPV types and cervical cancer in recent years as well as the efforts to openly provide the HPV vaccine injection in China. However, HPV52, HPV58 and HPV16 still have the greatest impact on the health of women in China. Therefore, we should pay close attention to them through vaccine prevention and HPV genotype screening and treatment. And we consider that HPV52, HPV58 and HPV16 play a guiding role in reducing the prevalence rates of high-risk HPV types in China. In addition, we show that HR-HPV subtypes with common biological properties have similar infection rate trends in precancerous stages, and the impact of HR-HPV genotypes and precancerous stages on infection. Moreover, the single/multiple infection proportions of HPV demonstrate a trend that the multiple infections proportion of HPV increases as the disease develops.

Although we obtained several interesting new findings, this study still has many limitations. For example, because the occurrence frequency of multiple infections is significantly affected by various factors [16, 17], the findings of multiple infections can only be used as a reference. Moreover, since we lack the related molecular data, the biological mechanism of multiple infections between HR-HPV types and the related time series analysis, survival, genome and signaling pathway analysis [29–32] remain to be studied in our future research.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12859-020-3454-5>.

**Additional file 1.** Collected Datasets. This file includes three datasheets about HR-HPV data from all infections, single infections and each HR-HPV type in each lesion group.

## Abbreviations

HR-HPV: High-risk human papillomavirus; HPV: Human papillomavirus; LCT: Liquid-based cytology test; TCT: ThinPrep cytology test; ASC-US: Atypical squamous cells of undetermined significance; LSIL: Low-grade squamous intraepithelial lesions; HSIL: High-grade squamous intraepithelial lesions

## Acknowledgements

This work was supported by the General Program from National Natural Science Foundation of China, Chongqing excellent youth award, and the National Science and Technology Major Project.

## About this supplement

This article has been published as part of *BMC Bioinformatics Volume 21 Supplement 7, 2020: Selected articles from the 6th International Work-Conference on Bioinformatics and Biomedical Engineering (part 2)*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-7>.

## Authors' contributions

Hongtu Liu and Le Zhang participated in the design of the study. Wenting Wu and Lei Song collected and analyzed the data. Wenting Wu, Yongtao Yang, Jianxin Wang and Le Zhang wrote and edited the manuscript. All authors read and approved the final manuscript.

## Funding

Publication costs are funded by the National Natural Science Foundation of China [61372138], the National Science and Technology Major Project [2018ZX10201002] and the Chinese Chongqing Distinguish Youth Funding [cstc2014jcyjqq40003].

## Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and the additional file.

## Ethics approval and consent to participate

Ethics documents were approved by the Ethics committee of the General Hospital of the People's Liberation Army, and each subject signed an informed consent form (ICF) when enrolled in this study.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>College of Computer Science, Sichuan University, Chengdu 610065, China. <sup>2</sup>Department of Obstetrics and Gynaecology PLA General Hospital, Beijing 100853, China. <sup>3</sup>BGI-Shenzhen, Shenzhen 518083, China. <sup>4</sup>School of Information Science and Engineering, Central South University, Changsha 410083, China. <sup>5</sup>National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China. <sup>6</sup>Medical Big Data Center of Sichuan University, Chengdu 610065, China.

Published: 5 May 2020

## References

- Vijayalakshmi R, Viveka TS, Malliga JS, Murugan K, Kanchana A, Arvind K. Use of fast transfer analysis cartridges for cervical sampling and real time PCR based high risk HPV testing in cervical Cancer prevention - a feasibility study from South India. *Asian Pacific Journal of Cancer Prevention Apjcp*. 2015;16(14):5993–9.
- Martel CD, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer*. 2017;141(4):664.
- Chang XL, Guo KJ, Zhang Y, Gynaecology DO, Hospital TF, University CM: Correlation analysis between HPV subtypes and cervical Cancerand precancerous lesions. *J Chin Med Univ*. 2014;43(8):720–3.
- Azuma Y, Kusumoto-Matsuo R, Takeuchi F, Uenoyama A, Kondo K, Tsunoda H, Nagasaka K, Kawana K, Morisada T, Iwata T. Human papillomavirus genotype distribution in cervical intraepithelial Neoplasia grade 2/3 and invasive cervical Cancer in Japanese women. *Jpn J Clin Oncol*. 2014;44(10):910–7.
- Genta MLND, Martins TR, Lopez RVM, Sadalla JC, Carvalho JPMD, Baracat EC, Levi JE, Carvalho JP. Multiple HPV genotype infection impact on invasive cervical cancer presentation and survival. *PLoS One*. 2017;12(8):e0182854.
- Li J, Wang YY, Nan X, et al. Prevalence of human papillomavirus genotypes among women with cervical lesions in the Shaanxi Province of China. *Genet Mol Res*. 2016;15(1):1–79. <https://doi.org/10.4238/gmr.15017181>.

7. Kir G, Seneldir H, Cosan Sarbay B: The clinical performance of computer-assisted liquid-based cytology, primary hrHPV screening, and cotesting at a Turkish Tertiary Care Hospital. *Diagn Cytopathol.* 2018;46(1):3–8.
8. Malila N, Leinonen M, Kotaniemi-Laitinen L, Laurila P, Tarkkanen J, Hakama M: The HPV test has similar sensitivity but more overdiagnosis than the pap test—a randomised health services study on cervical cancer screening in Finland. *Int J Cancer.* 2013;132(9):2141–7.
9. Blatt AJ, Kennedy R, Luff RD, Austin RM, Rabin DS: Comparison of cervical cancer screening results among 256,648 women in multiple clinical practices. *Cancer Cytopathol.* 2015;123(5):282–8.
10. Liu TY, Xie R, Luo L, Reilly KH, He C, Lin YZ, Chen G, Zheng XW, Zhang LL, Wang HB: Diagnostic validity of human papillomavirus E6/E7 mRNA test in cervical cytological samples. *J Virol Methods.* 2014;196(2):120–5.
11. Zhang L, Zheng CQ, Li T, Xing L, Zeng H, Li TT, Yang H, Cao J, Chen BD, Zhou ZY: Building up a robust risk mathematical platform to predict colorectal Cancer. *Complexity.* 2017;2017:14.
12. Catteau X, Simon P, Noël JC: Evaluation of the oncogenic human papillomavirus DNA test with liquid-based cytology in primary cervical Cancer screening and the importance of the ASC/SIL ratio: a Belgian study. *Isrn Obstet Gynecol.* 2015;2014:536495.
13. Ying LI, Ke H, Li JP, Lei S, Tu LH: Cervical infection of oncogenic human papillomavirus (HPV) types in Beijing, China. *Biomed Environ Sci.* 2016;29(10):734–41.
14. Chaturvedi AK, Katki HA, Hildesheim A, Rodríguez AC, Quint W, Schiffman M, Van Doorn LJ, Porras C, Wacholder S, Gonzalez P: Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease. *J Infect Dis.* 2011;203(7):910–20.
15. Dickson EL, Vogel RI, Geller MA, Downs JL: Cervical cytology and multiple type HPV infection: a study of 8182 women ages 31–65. *Gynecol Oncol.* 2014;133(3):405–8.
16. Dickson EL, Vogel RI, Bliss RL, Downs LS Jr: Multiple-type human papillomavirus (HPV) infections: a cross-sectional analysis of the prevalence of specific types in 309,000 women referred for HPV testing at the time of cervical cytology. *Int J Gynecol Cancer.* 2013;23(7):1295–302.
17. Vaccarella S, Franceschi S, Herrero R, Schiffman M, Rodríguez AC, Hildesheim A, Burk RD, Plummer M, Vaccarella S, et al: Clustering of multiple human papillomavirus infections in women from a population-based study in Guanacaste, Costa Rica. *J Infect Dis.* 2011;204:385–90.
18. Murdiyarsa LS, Kartawinata M, Jenie I, Widjajahakim G, Hidajat H, Sembiring R, Nasar IM, Cornain S, Sastranagara F, Utomo ARH: Single and multiple high-risk and low-risk human papillomavirus association with cervical lesions of 11,224 women in Jakarta. *Cancer Causes Control.* 2016;27(11):1371–9.
19. Coxe S, West SG, Aiken LS: The analysis of count data: a gentle introduction to poisson regression and its alternatives. *J Pers Assess.* 2009;91(2):121–36.
20. Hartigan JA: A K-means clustering algorithm. *Appl Stat.* 1979;28(1):100–8.
21. Diane S, Diane D, Robert K, Ann M, Dennis OC, Marianne P, Stephen R, Mark S, David W, Wright T Jr, Young N, Forum Group Members; Bethesda 2001 Workshop: The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA.* 2002;287(16):2114–9.
22. Nagai I, Takahashi K, Yanagihara H: Information criterion-based non-hierarchical clustering. *IJKESDP.* 2017;6(1):1–43.
23. Kabacoff R. *R in action*; 2011.
24. FaridKianifard G. *PP: Poisson regression analysis in clinical research.* *J Biopharm Stat.* 1995;5(1):115–29.
25. Rochon J: Analyzing the number of “rejection episodes” in human transplantation. *Control Clin Trials.* 1990;11(4):262.
26. Vonesh EF: Modelling peritonitis rates and associated risk factors for individuals on continuous ambulatory peritoneal dialysis. *Stat Med.* 2010;9(3):263–71.
27. Team R: *The R Project for Statistical Computing.* 2013:xiii–xx.
28. De EV, Fauquet C, Broker TR, Bernard HU, Zur HH: Classification of papillomaviruses. *Virology.* 2004;324(1):17–27.
29. Zhang L, Liu Y, Wang M, Wu Z, Li N, Zhang J, Yang C: EZH2-, CHD4-, and IDH-linked epigenetic perturbation and its association with survival in glioma patients. *J Mol Cell Biol.* 2017;9(6):477–88.
30. Zhang L, Qiao M, Gao H, Hu B, Tan H, Zhou X, Li CM: Investigation of mechanism of bone regeneration in a porous biodegradable calcium phosphate (CaP) scaffold by a combination of a multi-scale agent-based model and experimental optimization/validation. *Nanoscale.* 2016;8(31):14877–87.
31. Zhang L, Xiao M, Zhou J, Yu J: Lineage-associated underrepresented permutations (LAUPs) of mammalian genomic sequences based on a jellyfish-based LAUPs analysis application (JBLA). *Bioinformatics.* 2018;34(21):3624–30.
32. Zhang L, Zhang S: Using game theory to investigate the epigenetic control mechanisms of embryo development: comment on: “epigenetic game theory: how to compute the epigenetic control of maternal-to-zygotic transition” by Qian Wang et al. *Phys Life Rev.* 2017;20:140–2.

## Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

