International Journal of Infectious Diseases 28 (2014) 8–12

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

A risk evaluation model of cervical cancer based on etiology and human leukocyte antigen allele susceptibility

Bicheng Hu [a,b,1], Ning Tao [a,c,1], Fanyu Zeng [a,b], Min Zhao [a], Lixin Qiu [a], Wen Chen [a], Yun Tan [a], Yun Wei [a], Xufeng Wu [d,*], Xinxing Wu [a,*]

[a] Institute of Virology, School of Medicine, State Key Laboratory of Virology, Wuhan University, Wuhan 430071, Hubei, China
[b] The Clinical Laboratory, Wuhan No. 1 Hospital, Wuhan, Hubei, China
[c] Institute of Biophysics, Chinese Academy of Sciences, Beijing, China
[d] Hospital for Women and Children of Hubei, Wuhan 430070, Hubei, China

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form 9 May 2014
Accepted 19 May 2014

Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:
HPV
HLA II class allele
Evaluation model
Cervical cancer

SUMMARY

Background: There are no reliable risk factors to accurately predict progression to cervical cancer in patients with chronic cervicitis infected with human papillomavirus (HPV). The aim of this study was to create a validated predictive model based on the risk factors for cervical cancer. A model to estimate the risk of cervical cancer may help select patients for intervention therapy in order to reduce the occurrence of cervical cancer after HPV infection.

Methods: This retrospective analysis included 68 patients with cervical cancer and 202 healthy female controls. HPV infection and human leukocyte antigen (HLA) class II alleles in HLA-DRB1, 3–7, and 9 were detected. Other information was collected, including level of education and age at first parturition. Multiple regression analysis and an artificial neural network (ANN) were performed to identify the independent risk factors for cervical cancer, and based on these, an evaluation model for the prediction of the incidence of cervical cancer was formed.

Results: This model showed HPV to be a pivotal player in cervical cancer that increased the risk by 7.6-fold. The presence of the HLA-DRB1*13-2 and HLA-DRB1*3(17) alleles was associated with an increased risk of developing cervical cancer. Conversely, the HLA-DRB1*09012 and HLA-DRB1*1201 alleles were found to be associated with a reduced cervical cancer risk. In addition, other factors, such as age at first parturition and education level, had significant effects on cervical cancer risk. The model was applied to conduct a risk assessment of women in the mountain area of Wufeng County, Hubei Province in China. The sensitivity and specificity of our model both exceeded 95%.

Conclusions: This model, based on etiology and HLA allele susceptibility, can estimate the risk of cervical cancer in chronic cervicitis patients after HPV infection. It combines genetic and environmental factors and significantly enhances the accuracy of risk evaluation for cervical cancer. This model could be used to select patients for intervention therapy and to guide patient classification management.

1. Introduction

A risk evaluation model is crucial for efficient cancer screening among high-risk populations.[1] Environmental factors, such as education level, age at sexual debut, parity, and body mass index (BMI), are related to the risk of developing cervical cancer. Some genetic factors (e.g., interferon regulatory factor 1 (IRF-1)) in which specific polymorphisms correlate with cervical cancer have been reported.[2] Several other genes, such as codon 72 of p53, codon 31 of p21, and fragile histidine triad (FHIT), have been examined for their association with cervical cancer. Cervical carcinogenesis is a multifactorial disease that may result from environmental and genetic factors. To improve the predictive accuracy for determining the cervical cancer risk, we developed a risk evaluation model comprising both genetic and environmental factors.

* Corresponding author. Tel.: +86 27 68758898; fax: +86 27 68758766.
E-mail address: wuxinxing9755@163.com (X. Wu).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ijid.2014.05.015

In previous studies, we proposed a risk evaluation model for cervical cancer based on several well-known environmental contributors, such as infection with high-risk genotype(s) of human papillomavirus (HPV),[3–5] young age at first parturition,[6] and low education level of the subject and their spouse.[7,8] Emerging evidence suggests that human leukocyte antigen (HLA) class II alleles are associated with cervical cancer.[9] In this study, we combined HLA class II alleles with several risk factors to establish a risk evaluation model using multiple logistic regression analysis and an artificial neural network. This new evaluation model could significantly improve the accuracy of risk evaluation for cervical cancer.

2. Materials and methods

2.1. Subjects

Demographic information and blood samples were collected from 68 patients with cervical cancer and 202 healthy female controls admitted to the Healthcare Hospital for Women and Children of Wufeng County in Hubei Province, China between March 2002 and December 2009. All patients were diagnosed with cervical squamous cell carcinoma based on pathological examination. The average age of the patients was 51.8 ± 10.0 years (mean ± standard deviation) and ranged from 35 to 75 years. The average age of the controls was 42.5 ± 8.1 years and ranged from 22 to 73 years. All data from patients and controls were assigned randomly to one of two subsets. One subset comprised data from 63 patients and 192 controls and was used to establish the risk evaluation model. The other subset included data from five patients and 10 controls and was used to test the model. All patients and healthy controls gave written informed consent for the use of the specimens obtained for medical research. The study was approved by the ethics committee of the local institution and the basic medical school of Wuhan University.

2.2. Detection of HPV infection

HPV infection was identified by the amplification of HPV DNA from cervical cell scrapings, as described previously.[10]

2.3. Questionnaire and health examination

A questionnaire and health examination were conducted, as described previously.[7]

2.4. Genotyping of HLA-II alleles

HLA class II genotypes of the study subjects were determined by DNA sequencing.[10] HLA gene typing was performed for 68 cervical cancer cases and 202 controls. DNA was extracted by the phenol/chloroform method using peripheral blood mononuclear cells. DNA samples were typed at the HLA-DRB1, DRB3, DRB4, DRB5, DRB6, DRB7, and DRB9 loci using an HLA-DRB gene typing chip (United Gene, Shanghai, China) with a sequence-specific oligonucleotide probe (SSOP). Each HLA II allele was amplified and hybridized individually using locus-specific probe arrays. SSOP reactions for each sample were submitted electronically to an HLA analysis program to deduce the HLA type. HLA-DR genes were identified and named in concordance with the 12th International Histocompatibility Workshop and Conference.

2.5. Backpropagation

Backpropagation (BP) is a common algorithmic approach by which an artificial neural network (ANN) is instructed to perform a given task.[11] The BP algorithm consists of the following two parts: propagation and weight update. We used the Levenberg–Marquardt method, an optimized BP algorithm with a three-layered ANN, and nine hidden neurons. The input variables included environmental risk factors (high-risk HPV infection, education level of the subject and their spouse, and age at first parturition) and genetic contributors (HLA-DRB1*09012, HLA-DRB1*1201, HLA-DRB1*13-2, and DRB1*3(17)). The method for normalization of the input sample is described by the following equation: $x' = x/\max(x)$.

The initial weighted value of the neural network was gained from an initiff process. The transfer function between the input and output layer was a logarithm S-shaped function (logic) and it between the hidden and output layer was a linear function (purelin). For other indices, one hidden layer was used. The independent variables are described in Table 1. The deviation index (eg) was 0.09 and the maximum training steps (me) was 1000.

Table 1
Coding table of variables

Variable	Status		Coding
HPV high risk (16, 18/45, 52, 58)	Negative		0
	Positive		1
Genetic factor: HLA-DRB1*09012 + DRB1*1201 + DRB1*13-2 + DRB1*3(17)	DRB1*09012	Yes	0
		No	1
	DRB1*1201	Yes	0
		No	1
	DRB1*13-2	Yes	1
		No	0
	DRB1*3(17)	Yes	1
		No	0
Education level coding in reverse direction	Below primary school		5
	Primary school		4
	Junior high school		3
	High school		2
	College and above		1

HPV, human papillomavirus.

3. Results

Based on multiple regression analysis, the risk factors for cervical cancer included high-risk HPV infection, low education level of the individual and their spouse, young age at first parturition, HLA class II susceptibility alleles, and non-protective HLA alleles. Compared with uninfected women, the risk for cervical cancer in women with high-risk HPV infection increased approximately 7.6-fold (Table 2). High-risk HLA alleles also increased the risk of cervical cancer 2.3-fold in women lacking protective HLA alleles (Table 2). When women and their spouses had a lower level of education, the risk of cervical cancer increased 2.0- and 3.8-fold, respectively (Table 2). Backfitting indicated that the accuracy of predicting cervical cancer was 88.9% in the patient group and 98.4% in the control group (Table 3). The testing database showed that the accuracy of the risk model was 100% in the cancer group and reached 90% in the control group (Table 4), suggesting that this model has good predictive specificity and sensitivity.

Table 2
Results of multiple regression analysis for cervical cancer risk factors

	β_i	SE (β_i)	Wald	p-Value	OR
HPV infection	2.032	0.652	9.703	0.002	7.626
Genetic factor	0.854	0.454	3.544	0.060	2.349
Education level of woman	0.710	0.351	4.096	0.043	2.034
Education level of spouse	1.345	0.367	13.443	0.001	3.837
Age at first birth	−0.380	0.201	3.575	0.059	0.684
Constant	−1.174	4.838	0.059	0.808	-

SE, standard error; OR, odds ratio; HPV, human papillomavirus.

Table 3
Backfitting result of the multiple logistic regression model for cervical cancer classification[a]

			Evaluation of model	
			Cancer	
			0	1
Clinical data	Cancer	0	189 (98.4%)	3
		1	7	56 (88.9%)

[a] The number in brackets is the accuracy of classification.

An alternative risk evaluation model for cervical cancer was developed using the BP of an ANN. To test the capacity of this model to predict and classify unknown data, we applied the model to the test data using the coding rule listed in Table 2. The ANN model also demonstrated good classification ability. The predictive results on the test data showed that the accuracy in predicting cervical cancer was 80% in the patient group and 90% in the control group (Table 5), and the back substitution fitting had a high accuracy of classification. A sensitivity of 95.2% and a specificity of 99% were achieved with the ANN model (Table 6). This optimized model is compatible with input parameters.

4. Discussion

Epidemiological investigations have confirmed that more than 70% of women are infected with HPV, but most infection is transient. Only a small proportion progress to persistent infection and then develop cervical cancer.[12] A predictive model is urgently needed to evaluate the cervical cancer-associated risk factors.

Recent studies have demonstrated that the host genetic background may play a crucial role in the development of cervical cancer.[13] In humans, the major histocompatibility complex is referred to as the HLA system. This group of genes resides on chromosome 6 and encodes several proteins, including the HLA class II histocompatibility antigens, which are essential for the functions of the human immune system. HLA class II molecules are expressed on the surface of antigen-presenting cells and consist of an alpha chain (DRA) and a beta chain (DRB). The most prevalent beta subunit of HLA-DR is DRB1-9, encoded by the HLA-DRB1 gene, which plays a central role in the immune system by presenting peptides derived from extracellular proteins.

Previous reports[14–18] have suggested that the risk of cervical cancer may be associated with specific HLA alleles or HLA-linked genes. Several studies have indicated that the cellular immune response is essential for clearance of HPV infection.[19,20] Different combinations of HLA class II molecules and antigenic peptides may influence cytokine production during the early stage of the immune response to HPV infection. The peptide epitopes that are encoded by HLA susceptibility alleles are unable to present cervical cancer-derived antigens to T cells, and the cervical cancer risk may be increased or decreased depending on the HLA class II alleles.[19,21,22] In a previous study we indicated that HLA-A*0206 is a protective allele against the development of cervical cancer by presenting HPV16E7$_{11\text{-}20}$ and the peptide epitope of E7$_{86\text{-}93}$ to antigen-presenting cells and provoking cytotoxic T lymphocytes to destroy target cells.[3] In this study, we focused on HLA class II alleles that are associated with the presentation of cancer antigens.[23]

Using multiple logistic regression analysis and an ANN we established a risk evaluation model for cervical cancer. Our data showed that HPV plays a pivotal role in cervical cancer, as reported previously, increasing the risk 7.6-fold. The presence of the HLA-DRB1*13-2 and HLA-DRB1*3(17) alleles was found to be associated with an increased risk of cervical cancer carcinogenesis. Conversely, the HLA-DRB1*09012 and HLA-DRB1*1201 alleles were found to be associated with a reduced risk of cervical cancer.[10] In addition, other factors, such as age at birth of first child and education level, had significant effects on cervical cancer risk.

The model was applied to conduct a risk assessment of women in the mountain area of Wufeng County, Hubei Province in China. The sensitivity and specificity of this model exceeded 95%. There were two cases among the cohort of 192 normal cases that the ANN model judged as high risk and three cases with a clinical diagnosis of cervical cancer that the model judged as low risk. We considered that the primary reason for this is related to individual differences in genetic background. Another reason could be the heavy weight of high-risk HPV in the model, and we only incorporated common high-risk HPV types (HPV16, 18, 45, 52, 58) in the model, which may therefore miss the other HPV type infections. Along with the improvements in people's living conditions and healthcare consciousness, the patient with chronic cervicitis and an HPV infection could undergo active treatment as an intervention, which could significantly delay or even completely prevent cervical carcinogenesis.

Our results support the hypothesis that HLA class II DRB alleles influence the risk of developing invasive cervical cancer. This model takes HLA class II alleles as genetic risk factors and the other risk factors correlated with cervical cancer into consideration,

Table 4
Predictive results for cervical cancer with the multiple logistic regression evaluation model on the test data

Clinical group	Predictive group	Predictive value[a]	X1 (HPV infection)	X2 (Genetic factor)	X3 (Reverse coding of education level of woman)	X4 (Reverse coding of education level of spouse)	X5 (Age at first birth)
Control	Control	−1.8	0	1	3	3	20
Control	Control	−0.7	0	1	5	3	21
Control	Control	−5.0	0	1	3	2	25
Control	Control	−6.7	0	1	2	1	24
Control	Control	−6.2	0	2	2	1	25
Control	Case	0.4	1	2	3	3	22
Control	Control	−8.1	0	1	1	1	26
Control	Control	−6.4	0	3	1	1	26
Control	Control	−1.8	0	1	4	3	22
Control	Control	−6.5	0	2	1	1	24
Case	Case	1.7	0	4	4	4	23
Case	Case	0.8	1	2	3	3	21
Case	Case	0.6	0	1	5	4	21
Case	Case	6.4	1	2	5	5	17
Case	Case	4.1	1	2	5	5	23

HPV, human papillomavirus.
[a] If the predictive value is ≤0, then the predictive group is the control group; if the predictive value is >0, then the predictive group is the case group.

Table 5
Predictive results for cervical cancer with the ANN model on the test data[a]

Practical group	Predictive group	Predictive value[b]	X1 (HPV infection)	X2 (Reverse coding of education level of woman)	X3 (Reverse coding of education level of spouse)	X4 (Age at first birth)	X5 (Genetic factor)
Control	Control	−0.0000	0	0.6	0.6	0.67	0.25
Control	Control	0.0377	0	1	0.6	0.70	0.25
Control	Control	−0.0000	0	0.6	0.4	0.83	0.25
Control	Control	0.0000	0	0.4	0.2	0.80	0.25
Control	Control	−0.0001	0	0.4	0.2	0.83	0.5
Control	Case	1.0000	1	0.6	0.6	0.73	0.5
Control	Control	−0.0000	0	0.2	0.2	0.87	0.25
Control	Control	0.0000	0	0.2	0.2	0.87	0.75
Control	Control	−0.0000	0	0.8	0.6	0.73	0.25
Control	Control	0.0000	0	0.2	0.2	0.80	0.5
Case	Case	1.0000	0	0.8	0.8	0.77	1
Case	Case	1.2049	1	0.6	0.6	0.70	0.5
Case	Control	0.0055	0	1	0.8	0.70	0.25
Case	Case	1.0000	1	1	1	0.57	0.5
Case	Case	1.0000	1	1	1	0.77	0.5

ANN, artificial neural network; HPV, human papillomavirus.
[a] The data in Table 5 are from the testing data, which are the same as the testing data for the logistic regression model, and these were normalized.
[b] If the predictive value is <0.5, then the predictive group is the control group; if the predictive value is ≥0.5, then the predictive group is the case group.

Table 6
Back substitution fitting results of the ANN evaluation classification model[a]

			Classification results by model	
			Case	
			0	1
Clinical data	Case	0	190 (99.0%)	2
		1	3	60 (95.2%)

[a] The number in brackets is the accuracy of classification.

which has not been reported previously; our results showed that HLA polymorphisms are closely related to genetic susceptibility to cervical cancer and HPV infection in the Chinese population. In particular, we found the HLA-DRB1*13-2 and HLA DRB1*3(17) alleles to be associated with an increased risk of developing cervical cancer. A better understanding of these associations would help to elucidate the role that HLA molecules play in the immune response to HPV infection and subsequent cervical cancer pathogenesis in the Chinese population. Moreover, this work may help to improve the current strategies for the prevention and treatment of cervical cancer through vaccination and immunotherapy. The model has theoretical and application value via screening high-risk HPV infections and monitoring HLA alleles to predict the risk of cervical cancer. This model could be used to select patients for intervention therapy (control of cervical erosion, microwave therapy, frozen treatment, loop electrosurgical excision procedure (LEEP) knife operation, etc.) and guide patient classification management.

Acknowledgements

We thank Dr Jinhu Xiong for revising this manuscript. All authors contributed to the study and approved the final manuscript. The authors wish to thank the clinical staff at the Healthcare Hospital for Women and Children of Wufeng County in Hubei Province, China and the outreach workers for providing their support in subject recruitment and survey interviews. The authors also thank all of the participants in the study for their time and sharing their information. This research was supported by the National Natural Science Foundation of China (No. 30772308 and No. 81072123).

Ethical approval: This research was approved by the Institutional Review Board of the Healthcare Hospital for Women and Children of Wufeng County Hubei Province and the basic medical school of Wuhan University and subjects gave informed consent to the work.

Conflict of interest: All the authors declare that they have no conflict of interest.

References

1. Ho SH, Jee SH, Lee JE, Park JS. Analysis on risk factors for cervical cancer using induction technique. *Expert Syst Appl* 2004;**27**:97–105.
2. Lee JE, Lee SJ, Namkoong SE, Um SJ, Sull JW, Jee SH, et al. Gene–gene and gene–environmental interactions of p53, p21, and IRF-1 polymorphisms in Korean women with cervix cancer. *Int J Gynecol Cancer* 2004;**14**:118–25.
3. Qiu XP, Tao N, Tan Y, Wu XX. Constructing of the risk classification model of cervical cancer by artificial neural network. *Expert Syst Appl* 2007;**32**:1094–9.
4. Bosch FX, Manos MM, Muñoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J Natl Cancer Inst* 1995;**87**:796–802.
5. Andersson S, Rylander E, Larsson B, Strand A, Silfversvärd C, Wilander E. The role of human papillomavirus in cervical adenocarcinoma carcinogenesis. *Eur J Cancer* 2001;**37**:246–50.
6. Wilson HG, Curtis A, Marchbanks PA. Parity, age at first birth, and risk of invasive cervical cancer: meta-analyses. *Ann Epidemiol* 2002;**12**:490–1.
7. Tao N, Wu XF, Qiu XP, Zhao M, Tan Y, Wu XX. A study on the risk factors of cervical carcinoma in mountainous area of Wufeng County, Hubei Province, China. *Wuhan University Journal of Natural Sciences* 2005;**10**:759–66.
8. Rostad B, Schei B, Da Costa F. Risk factors for cervical cancer in Mozambican women. *Int J Gynecol Obstet* 2003;**80**:63–5.
9. Castro FA, Haimila K, Sareneva I, Schmitt M, Lorenzo J, Kunkel N, et al. Association of HLA-DRB1, interleukin-6 and cyclin D1 polymorphisms with cervical cancer in the Swedish population—a candidate gene approach. *Int J Cancer* 2009;**125**:1851–8.
10. Zhao M, Qiu L, Tao N, Zhang L, Wu X, She Q, et al. HLA DRB allele polymorphisms and risk of cervical cancer associated with human papillomavirus infection: a population study in China. *Eur J Gynaecol Oncol* 2013;**34**:54–9.
11. Demuth H, Beale M. MATLAB neural network toolbox user's guide: version 4. Natick, MA: The MathWorks; 2002.
12. Bekkers RL, Massuger LF, Bulten J, Melchers WJ. Epidemiological and clinical aspects of human papillomavirus detection in the prevention of cervical cancer. *Rev Med Virol* 2004;**14**:95–105.
13. de Freitas AC, Gurgel AP, Chagas BS, Coimbra EC, do Amaral CM. Susceptibility to cervical cancer: an overview. *Gynecol Oncol* 2012;**126**:304–11.
14. Wank R, Thomssen C. High risk of squamous cell carcinoma of the cervix for women with HLA-DQw3. *Nature* 1991;**352**:723–5.
15. Wu YP, Chen YL, Li LY, Cao Y, Liu Z, Liu B, et al. Polymorphic amino acids at codons 9 and 37 of HLA-DQB1 alleles may confer susceptibility to cervical cancer among Chinese women. *Int J Cancer* 2006;**118**:3006–11.
16. Carreon JD, Martin MP, Hildesheim A, Gao X, Schiffman M, Herrero R, et al. Human leukocyte antigen class I and II haplotypes and risk of cervical cancer. *Tissue Antigens* 2005;**66**:321–4.

17. Wang SS, Wheeler CM, Hildesheim A, Schiffman M, Herrero R, Bratti MC, et al. Human leukocyte antigen class I and II alleles and risk of cervical neoplasia: results from a population-based study in Costa Rica. *J Infect Dis* 2001;**184**:1310–4.
18. Hilders CG, Houbiers JG, Krul EJ, Fleuren GJ. The expression of histocompatibility related leukocyte antigens in the pathway to cervical carcinoma. *Am J Clin Pathol* 1994;**101**:5–12.
19. Cuzick J, Terry G, Ho L, Monaghan J, Lopes A, Clarkson P, et al. Association between high-risk HPV types, HLA DRB1* and DQB1* alleles and cervical cancer in British women. *Br J Cancer* 2000;**82**:1348–52.
20. Sastre-Garau X, Loste MN, Vincent-Salomon A, Favre M, Mouret E, de la Rochefordiere A, et al. Decreased frequency of HLA-DRB1 13 alleles in Frenchwomen with HPV-positive carcinoma of the cervix. *Int J Cancer* 1996;**69**:159–64.
21. Madeleine MM, Brumback B, Cushing-Haugen KL, Schwartz SM, Daling JR, Smith AG, et al. Human leukocyte antigen class II and cervical cancer risk: a population-based study. *J Infect Dis* 2002;**186**:1565–74.
22. Sastre-Garau X, Cartier I, Jourdan-Da Silva N, De Crémoux P, Lepage V, Charron D. Regression of low-grade cervical intraepithelial neoplasia in patients with HLA-DRB1*13 genotype. *Obstet Gynecol* 2004;**104**:751–5.
23. Maciag PC, Schlecht NF, Souza PS, Franco EL, Villa LL, Petzl-Erler ML. Major histocompatibility complex class II polymorphisms and risk of cervical cancer and human papillomavirus infection in Brazilian women. *Cancer Epidemiol Biomarkers Prev* 2000;**9**:1183–91.