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.