New approach reveals CD28 and IFNG gene interaction in the susceptibility to cervical cancer

Valeska B. Guzman1,†, Anatoly Yambartsev2,‡, Amador Goncalves-Primo1,†, Ismael D.C.G. Silva3, Carmen R.N. Carvalho3, Julisa C.L. Ribalta3, Luiz Ricardo Goulart4, Natalia Shulzhenko1,5, Maria Gerbase-DeLima1,‡ and Andrey Morgun1,5,†

1Immunogenetics Division, Pediatrics Department, Federal University of São Paulo, SP, Brazil, 2Mathematics and Statistics Institute, University of São Paulo, Brazil, 3Department of Gynecology, Federal University of São Paulo, SP, Brazil, 4Genetics Department, Federal University of Uberlândia, MG, Brazil and 5TCMTS “Ghost lab”, LCMI, NIAID, NIH, Bethesda, MD, USA

Received September 19, 2007; Revised and Accepted March 5, 2008

Cervical cancer is a complex disease with multiple environmental and genetic determinants. In this study, we sought an association between polymorphisms in immune response genes and cervical cancer using both single-locus and multi-locus analysis approaches. A total of 14 single nucleotide polymorphisms (SNPs) distributed in CD28, CTLA4, ICOS, PDCD1, FAS, TNFA, IL6, IFNG, TGFB1 and IL10 genes were determined in patients and healthy individuals from three independent case/control sets. The first two sets comprised White individuals (one group with 82 cases and 85 controls, the other with 83 cases and 85 controls) and the third was constituted by non-white individuals (64 cases and 75 controls). The multi-locus analysis revealed higher frequencies in cancer patients of three three-genotype combinations [CD2817(TT)/IFNG1874(AA)/TNFA-308(GG), CD2817(TT)/IFNG1847(AA)/PDCD1-7785(CT), and CD2817(TT)/IFNG1874(AA)/ICOS1564(TT)] (P < 0.01, Monte Carlo simulation). We hypothesized that this two-genotype [CD28(TT) and IFNG(AA)] combination could have a major contribution to the observed association. To address this question, we analyzed the frequency of the CD28(TT), IFNG(AA) genotype combination in the three groups combined, and observed its increase in patients (P = 0.0011 by Fisher’s exact test). The contribution of a third polymorphism did not reach statistical significance (P = 0.1). Further analysis suggested that gene–gene interaction between CD28 and IFNG might contribute to susceptibility to cervical cancer. Our results showed an epistatic effect between CD28 and IFNG genes in susceptibility to cervical cancer, a finding that might be relevant for a better understanding of the disease pathogenesis. In addition, the novel analytical approach herein proposed might be useful for increasing the statistical power of future genome-wide multi-locus studies.

INTRODUCTION

There is overwhelming evidence that prolonged infection with oncogenic human papillomavirus is the major factor associated with development of cervical cancer (1). It is conceivable that other environmental and/or genetic factors play a role in susceptibility, as only a relatively small proportion of infected women develop cervical cancer (2). Since the immune response has an important role in the defense against viruses and tumors polymorphisms in genes that potentially affect the immune response are candidates for influencing the susceptibility to cervical cancer.

A series of publications on polymorphisms of HLA class II genes and cervical cancer and/or its precursor lesions, cervical intraepithelial neoplasia (CIN), show that some HLA alleles are associated with protection, while others are associated

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.
‡M.G.D. and A.M. are co-senior authors.

Published by Oxford University Press 2008.
with susceptibility (3–10). These associations are probably explained by the role of HLA II molecules in presenting viral- or tumor-derived epitopes to T CD4+ cells. More recently, interesting findings were reported concerning resistance/susceptibility to cervical cancer mediated by combinations of killer immunoglobulin-like receptor (KIR) and their HLA class I ligands (11).

As polymorphisms in genes coding for cytokines, cytokine receptors and co-stimulatory molecules may affect the immune response (12–14), they are natural candidates to influence the susceptibility to various diseases, including cancer. The majority of studies of association between immune response genes and cervical cancer have analyzed SNPs in genes coding for tumor necrosis factor alpha (TNF-α), interferon gamma (IFN-γ), interleukin 6 (IL-6), interleukin 10 (IL-10) and transforming growth factor beta (TGF-β) (15–18). Though some significant associations have been reported in some studies, they were not consistently confirmed in other studies (19–23).

Because molecules involved in the immune response do not act in isolation, but rather form a complex network of interacting proteins, the net immunological response is most likely the product of variation in many polymorphic genes. It is therefore of great importance to test the combination of polymorphisms in different genes as a risk factor for diseases (24–28). Any multi-locus approach, however, faces the problem that has been referred to as the ‘curse of dimensionality’ because, though the number of polymorphisms under study may not be very large, the number of possible combinations between them turns out to be extremely high. Some statistical approaches for this problem were recently reviewed (29).

The purpose of this study was to investigate the association between invasive cervical cancer and polymorphisms in genes coding for the following immune response molecules: CD28, CTLA-4 (cytotoxic T-lymphocyte-associated protein 4), ICOS (inducible T cell co-stimulator), PDCD1 (programmed cell death receptor-1, also called PD-1) FAS, TNFα, IL-6, IFNγ, TGFβ1 and IL-10. In addition to analysis of the association with each polymorphism, we searched for association with two- and three-polymorphism combinations, utilizing a new statistical approach.

RESULTS

No deviation from Hardy–Weinberg equilibrium was detected regarding any SNP in patients or in controls. Linkage disequilibrium (LD) was detected between three alleles of *IL10* (−1082 A, −819 C and −592 T) and between two alleles of *TGFB1* (+869 T and +915 G) in all patient and control groups, confirming the previous findings (30–32). In addition, LD was found between *CD28* +17 C and *CTLA4* −319 T, both in cases and in controls, confirming our previous finding in healthy individuals (30). Considering the LD, SNPs at positions *CTLA4* −319, *IL10* −921, *IL10* −819 and *TGFB1* +915 were excluded from the analysis. Thus, only 10 SNPs were included in the association analysis.

Among the 14 SNPs studied, we found no association between any isolated SNP and cervical cancer risk after correction for multiple testing using false discovery rate (FDR).

The delete-d-jackknife method revealed three three-genotype combinations with frequencies in patients higher than in controls, in all three case–controls sets: [CD28(TT)/IFNG(AA)/TNFA(GG), CD28(TT)/IFNG(AA)/PDCD1(CT) and CD28(TT)/IFNG(AA)/ICOS(TT)] (Fig. 1). The estimated probability of finding at least three of any three-genotype combinations by chance in the three study sets, according to the Monte Carlo simulation, is <0.01.
The aim of the present study was to search for association of single and/or combined polymorphisms in 10 immune response genes with cervical cancer.

In agreement with the majority of previous association studies between cervical cancer and FAS, TNFA, IFNG and IL10 polymorphisms (20, 31–33), we found no association between any of these polymorphisms and cervical cancer risk. Also, our study is the first investigation of CD28, ICOS, PDCD1 or CTLA4 polymorphisms in cervical cancer patients. Importantly, we have revealed that the CD28(TT)/IFNG(AA) combination is associated with the disease.

In order to identify associations of two- or three-genotype combinations with cervical cancer, we applied a new statistical analysis comprising three steps: (i) use of a procedure (delete-djackknife) for sorting out combinations of any two- and/or three-marker combination that could be associated with cervical cancer in each one of the three case–control sets; (ii) selection of the combinations in common to all three case–control sets; (iii) calculation of the probability of finding the common combination pattern by chance (Monte Carlo simulation). The major advantage of applying this new approach is its ability to increase the searching capacity (i.e. power) of the test (Step i), while keeping first type error low by accepting only associations that appeared in all three independent case–control sets (Step ii).

Although the multi-locus analysis herein reported uses a combinatorial method based on multi-dimensional reduction (MDR) method (34), the direct comparison of the efficiency of these two procedures could not be performed in our dataset. The first problem is that MDR does not accept missing values, and this restriction decreases the number of individuals available for analysis in our dataset. The main difference between MDR and our analysis is that, while the first was devised for finding the best combination in a single dataset, the second is based on the consistency of findings in three separate datasets. Additionally, we think the MDR method does not provide an adequate solution for the multiple hypothesis problem.

Models that compare several genotype combinations generate a series of hypotheses and, therefore, increase the number of false discoveries. It is not clear how this problem can be addressed using traditional statistical methods (e.g. FDR), because these methods are applicable only in situations where the hypotheses are independent. This is not the case for multi-locus analysis, and there is no widely accepted analytical solution for this problem. Thus, we employed an experimental statistical approach to control for the falsely included genotype combinations in the model. The results of our screening were significant (Monte Carlo simulation, P < 0.01): CD28(TT)/IFNG(AA)/TNFA(GG), CD28(TT)/IFNG(AA)/PDCD1(CT) and CD28(TT)/IFNG(AA)/ICOS(TT) genotype combinations are candidates to be associated with cervical cancer.
CD28(TT) and IFNG(AA) polymorphisms are associated with cervical cancer only when both polymorphisms are present. (A), proportion of CD28(TT) in cases and controls in the presence or absence of IFNG(AA). (B), proportion of IFNG (AA) in cases and controls in the presence or absence of CD28(TT). The X-axis represents the odds ratio (OR) scale.

Noticing that the CD28(TT)/IFNG(AA) genotype combination was present in all the three three-genotype combinations, we wondered why this two-genotype combination was not sorted out as significant by the delete-d-jackknife procedure. In order to better understand this issue, we examined the frequencies of CD28(TT)/IFNG(AA) genotypes combination in all three case–control sets. Though the association was, as expected, significant (P = 0.0011). We therefore concluded that the CD28(TT)/IFNG(AA) combination is a better marker than CTLA4(CC)/IFNG(AA) for susceptibility to cervical cancer.

In order to examine the interaction effect between the CD28(TT) and IFNG(AA) genotypes, we compared the frequency of CD28(TT) between patients and controls stratified for the presence of the IFNG(AA) genotype and vice-versa (Fig. 3). This allows us to illustrate the interaction and to perform statistical evaluation of the effect. When using penalized logistic regression (35), we also observed the interaction of CD28 and IFNG, thus confirming our finding with another method for interaction analysis.

The biological relationship between the CD28/IFNG interaction and cervical cancer may rely on multiple mechanisms. The first potential mechanism is the effect of the product of one gene on the product of the other gene. This possibility is supported by several observations. Voigt et al. (36) reported that, in CD28-deficient mice, IFN-γ-producing CD8+ T cells were not only reduced in number, but were also less potent in lysing their respective target cells. It is also known that regulation of the NFκB transcription family in T cells involves signaling through CD28 (37). NFκB may not only regulate the induction of IFN-γ-induced genes, but also regulate the antiviral and immunomodulatory activities of IFN-γ (38).

A second potential mechanism is an independent functional contribution of each one of the genotypes, with the presence of both seemingly necessary to trigger the studied phenotype. Indeed, there is an abundance of literature showing genetic factors substantially influencing the production of cytokines, and showing that the anti-inflammatory cytokine profile may contribute to the disease process (39). The presence of allele A on position +874 intron 1 of the IFNG gene disables a putative NfκB binding site and can result in a lower level of IFN-γ production (40). Accordingly, several studies observed decreased levels of IFNG mRNA or protein in persistent HPV infection (41–43), or in invasive carcinoma (44,45). El-Sherif et al. (46) demonstrated a decreased level of IFNG mRNA in HPV-16-associated epithelium, and also decreased sub-epithelial IFNG mRNA with the progression of lesions. Although we did not find any study addressing the question of the influence of CD28 +17 T>C polymorphism in CD28 mRNA or protein expression, this SNP is located near the...
splice acceptor site. It could therefore influence splicing events which, in turn, could affect CD28 signaling and T cell activation. We did not find any study concerning CD28 expression or polymorphisms in cervical cancer.

Another potential mechanism for interactions of these genes involves direct physical contact between these two loci. Although IFNG and CD28 genes are located on different chromosomes (12q14 and 2q33, respectively), recent works discovered the possibility of interchromosomal interactions between two separate loci (47,48). The presence of the IFNG and CD28 polymorphisms could promote an interaction that changes expression of these genes.

A role for the interaction between CD28 and IFNG is biologically highly plausible, but our study does not rule out the possibility that the observed association is a result of LD between studied polymorphisms and other genes. Such important immune players as IL-26 and IL-22, for example, are close to IFNG gene and have a certain degree of LD with it (49). Considering the correlation between cervical cancer and persistent HPV infection (50,51), it is also important to note that the design of our study does not discriminate genetic association with cancer from association with persistent HPV infection. Thus, there is a possibility that we observed the association with persistent HPV infection that causes subsequent development of cervical cancer.

In conclusion, our results showed an epistatic effect between CD28 and IFNG genes in susceptibility to cervical cancer, a finding that might be relevant for a better understanding of the disease pathogenesis. In addition, the novel analytical approach herein proposed might be useful for increasing the statistical power of future genome-wide multi-locus studies.

MATERIALS AND METHODS

Patients and study design

The patients were Brazilian women with invasive squamous cell carcinoma of the uterine cervix diagnosed at the Department of Gynecology of São Paulo Hospital, São Paulo, Brazil and Cancer Hospital of Uberlândia, Brazil. The controls comprised ethnically matched Brazilian healthy unrelated individuals.

Three independent case–control sets were constituted. Case–control sets 1 (82 cases and 85 controls) and 2 (83 cases and 85 controls) comprised White individuals, while set 3 (64 cases and 51 to 75 controls, depending on the polymorphism) was constituted by non-Whites (Mulattos and Blacks). The additional control group of 83 women was genotyped for polymorphisms of IFNG and CD28 (Supplementary Material, Table S2). The ethnic classification was made by external phenotypic characteristics, as described in a publication of our group describing the frequencies of CD28, CTLA4 and ICOS polymorphisms in healthy individuals from three Brazilian ethnic groups (30).

The Medical Ethics Committee from the Federal University of São Paulo and the Federal University of Uberlândia approved the study (#1208/01, #035/01), and written informed consent was obtained from all participants.

Genotyping

Genomic DNA was extracted from EDTA-preserved peripheral blood using the dodecyltrimethylammonium bromide/hexadecyltrimethylammonium bromide (DTAB/CTAB) method (52). Subjects were genotyped for 14 polymorphisms. PCR-RFLP was applied to detect the CD28 intron 3 (+17 T>C), CTLA4 promoter (−319 C>T), CTLA4 exon 1 (+49 A>G), ICOS 3′-UTR (+1564 T>C), PDCD1 exon 5 (+7785 C>T) and FAS promoter (−670 G>A) SNPs. Each PCR reaction was carried out in 25 μl containing Master Mix (Eppendorf, Hamburg, Germany), 0.1 mM solution of each of the specific primers (Supplementary Material, Table S1) and 100 ng of genomic DNA. The cycling profile was: 95°C for 45 s, annealing for 30 s at temperatures shown in Supplementary Material, Table S1, 72°C for 30 s, 40 cycles. The PCR products were then incubated with the appropriate restriction enzymes at 37°C, overnight. DNA fragments were visualized on 2.5% agarose gels stained with ethidium bromide. The assignments of SNP genotypes were confirmed in 10% of randomly selected samples from each genotype by direct sequencing in an Applied Biosystems sequencer (ABI PRISM™, Model 3100 Avant). TNFα promoter (−308 G>A); IL6 promoter (−174 G>C); IFNG intron 1 (+874 A>T); TGFB1 codon 10 (+869 T>C), codon 25 (+915 G>C); and IIL10 promoter (−1082 A>G), (−819 C>T), (−592 C>A) SNPs were determined by polymerase chain reaction with sequence-specific primers (PCR-SSP) using the ‘Cytokine Genotyping Tray’ (One Lambda, Inc., Canoga Park, CA, USA). PCR conditions were as indicated by the manufacturer; the PCR products were then visualized by electrophoresis in 2.5% agarose gel.

Hardy–Weinberg equilibrium and linkage disequilibrium testing

Goodness-of-fit test to Hardy–Weinberg equilibrium was performed by calculating expected frequencies of each genotype and comparing them with the observed values. The FDR method (53) was used to correct the significance level of the Hardy–Weinberg equilibrium testing. LD between two SNPs was calculated by Arlequin software (54).

Statistics: single and multi-locus association analysis

Comparisons of single-allelic and single-genotype frequencies (obtained by direct count) between cases and controls in each one of the three study sets were performed by the Fisher’s exact test and the χ² test, respectively. We performed an adjustment for multiple testing using FDR.

In the multi-locus analysis, we used a new approach combining three independent case–control groups to decrease the probability of finding spurious association. We compared the frequencies (f) of combination for two and three genotypes between cases and controls. The first step consisted of the selection of gene-combination candidates by filtering two- and three-genotype combinations using the delete-d-jackknife method (55) in each of the three study sets. Genotypes with frequencies less than 10% among cases or controls were not considered. This threshold was determined taking into account the statistical power of future genome-wide multi-locus studies.
account the smallest sample size included in our study ($n = 48$, in control group 3). The strength ($S$) of the association for each genotype combination was calculated by $S = |P(\text{case} - f_{\text{control}}) < 0 - P(\text{case} - f_{\text{control}}) > 0|$, and all associations with $S > 0.990$ were filtered in each case–control set. Among the combinations filtered, the ones common to all three sets were selected.

We applied the Monte Carlo simulation for estimating the probability of sorting the same or higher number of combinations of any two and/or three genotypes in common to all three sets (56). The construction of the simulated case and control groups was performed as follows: (i) the experimental case and control groups were merged into one group; (ii) the frequencies of the genotypes were calculated in this group; (iii) the genotypes of the simulated case and control groups were randomly generated according to the frequencies of each genotype in the merged group. Three simulated case–control sets were constructed; (iv) the delete-d-jackknife method was applied in all three simulated case and control sets, as described before; (v) steps 3 and 4 were repeated a thousand times.

To investigate the nature of SNP–SNP interaction, we pooled all cases and all controls to generate a combined case–control set. The OR and its 95% confidence interval (CI) were calculated (Fisher’s exact test) as a measure of the association between genotypes, or genotype-combinations, and cervical cancer risk. The comparison between ORs was performed using Cochran’s Q-statistics (57). In addition, we used bootstrap to estimate the value of Q-statistics using its bootstrapping mean (55). The penalized logistic regression (35) also was applied in all three simulated case and control sets, as described.

We applied the Monte Carlo simulation for estimating the probability of sorting the same or higher number of combinations of any two and/or three genotypes in common to all three sets (56). The construction of the simulated case and control groups was performed as follows: (i) the experimental case and control groups were merged into one group; (ii) the frequencies of the genotypes were calculated in this group; (iii) the genotypes of the simulated case and control groups were randomly generated according to the frequencies of each genotype in the merged group. Three simulated case–control sets were constructed; (iv) the delete-d-jackknife method was applied in all three simulated case and control sets, as described before; (v) steps 3 and 4 were repeated a thousand times.

To investigate the nature of SNP–SNP interaction, we pooled all cases and all controls to generate a combined case–control set. The OR and its 95% confidence interval (CI) were calculated (Fisher’s exact test) as a measure of the association between genotypes, or genotype-combinations, and cervical cancer risk. The comparison between ORs was performed using Cochran’s Q-statistics (57). In addition, we used bootstrap to estimate the value of Q-statistics using its bootstrapping mean (55). The penalized logistic regression (35) also was applied in this analysis (see Supplementary Material).

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG Online.

**ACKNOWLEDGEMENTS**

We thank Dr Gerdine Sanson for introducing us to multi-locus analysis, as well as Carolina de Sa Primo and Kharen Yaerni Kawamura for their technical assistance. We thank Dr Chloé Camba Musatti, Dr Janeth Deolina Pena and Dr Rogério Agenor de Araújo for help in the logistics of sample collection and in preparation of DNA samples from case subjects. We also thank Max Behrens for help in grammatical and stylistic issues.

**Conflict of Interest statement.** None declared.

**FUNDING**

This study was supported by a grant from FAPESP # (01/10703-1), by intramural funding of Immunogenetics Division of UNIFESP and Intramural Research Program of the NIH, NIAID. GVB is recipient of a fellowship from FAPESP.

**REFERENCES**


