Novel association analysis between *HLA-DQB1* polymorphisms and rectal cancer based on a cross-validation design

F.Z. Tong[1], W.J. Yu[2] and H. Liu[3]

[1]Dalian Fifth Hospital, Dalian, China
[2]Dalian Blood Center, Dalian, China
[3]Department of Clinical Immunology, Dalian Medical University, Dalian, China

Corresponding author: H. Liu
E-mail: liuhui60@sina.com / immunology@dlmedu.edu.cn

ABSTRACT. A new study design based on cross-validation of the age at the onset of rectal cancer and the differences between the frequency distributions of relevant genes in 2 groups was developed for identification of disease-related HLA. Patients with rectal cancer were recruited and their age at the time of the first surgery was recorded. The genetic variants of HLA-DQB1 were genotyped using an HLA-DQB1 PCR-SSP typing kit. Allele frequencies were compared with control population. The mean age of patients with and without the alleles was compared. The frequency values of *HLA-DQB1*02* were 12.3% higher in the cancer group than in the control population (P < 0.05). The median ages of the subjects with and without *HLA-DQB1*02* were 54.0 and 61.0 years, respectively, with significant difference observed between the ages for these groups (P < 0.05). The median ages of the subjects with and without *HLA-DQB1*03* were 62.0 and 58.0 years, respectively, and a significant difference was observed. The cross-validation of the 2 above mentioned analytical results showed that a

statistically significant difference was noted for *HLA-DQB1*02* (P < 0.05), whereas no such statistically significant difference was observed for *HLA-DQB1*03*. HLA-DQB1*02 allele was related to cancer susceptibility. The new analysis method may be an efficient and reliable approach for the identification of disease-related HLA.

Key words: Rectal cancer; HLA; Gene identification; Complex disease

INTRODUCTION

Human leukocyte antigen complex (HLA) genes are located on the short arm of chromosome 6 and are the most polymorphic loci within the human genome. The primary function of HLA is to allow the immune system to identify infectious pathogens and eliminate them. HLA genes are the most significant genetic predisposition factor or genetic markers on chromosome 6 for diseases and cancers affecting the immune system (Jeffery and Bangham, 2000; Brewerton, 2003; Madeleine et al., 2008). In particular, the status of HLA alleles has a significant role in immune responses and immunologic tolerance (Fischer and Mayr, 2001). Generally, high-throughput genotyping assays and association analysis are used for identifying such HLA alleles. However, when many alleles are observed in same time is not possible to confirm whether a statistical difference is due to a 5% sampling error; therefore, an allele that shows a statistically significant relation with a complex disease may be not a true complex disease-related allele. Although the Bonferroni correction is used for multiple comparison, some recent studies provide arguments that Bonferroni correction is too rigid with a lower power (Nyholt, 2004; Teriokhin et al., 2007; Vialatte and Cichocki, 2008). This impact is especially problematic for high-multidimensional data, such as association analysis of disease-related HLA allele. Therefore, we developed a new study design based on the link between the age at the onset of disease and the differences between the frequency distributions of relevant alleles in patients and controls.

The pathogenic effects of a gene on individuals should include two factors: 1) the differences between the frequency distributions of the gene in disease and control groups (differences in frequency distribution increase with an increase in the effect of a gene) and 2) the age at disease onset in gene carriers (increase in pathogenic effects corresponds to earlier onset). These 2 factors provide information in 2 dimensions (frequency distribution and age at disease onset); therefore, cross-validation of this information should improve the efficacy of association analysis on the identification of disease-related HLA alleles comparing with corrections for multiple comparisons. Here, we have used association analysis for *HLA-DQB1* and rectal cancer to illustrate our novel cross-validation design for identification of disease-related genes.

MATERIAL AND METHODS

Subjects

A total of 79 patients - 44 males and 35 females; mean age, 61.1 ± 11.6 years - with rectal cancer (as evidenced by surgical biopsy and pathological analysis) were recruited from the Dalian University Hospital, Dalian, China. All the subjects were unrelated ethnic Han

Chinese individuals and had provided informed consent for participation in this study. The age of first onset was recorded.

HLA-DQB1 alleles typing

Genomic DNA was extracted from white blood cells using standard techniques for HLA typing. The samples were typed using an *HLA-DQB1* "low resolution" PCR-SSP typing kit (Pel-Freez Clinical Systems) including allele-specific primers for *DQB1*02*, *DQB1*03*, *DQB1*04*, *DQB1*05*, *DQB1*06*. All commercial tests were run according to manufacturer instructions. Products were separated by electrophoresis on 2% agarose and visualized with ethidium bromide staining and UV transillumination. Automated gel reading was performed using the Pel-Freez software.

Statistical analysis of allele frequency distributions

Allele frequencies were calculated and compared with control population coming from same geographic region of subjects participating in the study (a total of 718 subjects were included in that study) (Hei et al., 2009). Constituent ratio of allele frequencies in the cancer group were compared with that in control population using the chi-square test. If significant locus was determined, the significant difference of allele frequencies between cancer group and control group were examined using the chi-square test. A P value <0.05 indicated significant difference. The analyses were performed by the SPSS 13.0 statistical software package.

Statistical analysis of the age at the onset of cancer

The basic concept of the analysis is to consider the age of onset as a pathogenic effect of a gene. The association between the gene and the disease can be ascertained by comparing the average age of patients carrying a gene and that of patients not carrying the gene. In the present study, the age of the patients was considered as the dependent variable and that of the allele carriers (subjects carrying the allele or those not carrying the allele) was considered as the grouping variable; this analysis was performed using the SPSS 13.0 statistical software package. The Mann-Whitney *U* test, a nonparametric test, was used to assess the difference between the mean age of the subjects carrying the alleles and that of subjects that did not carry the alleles. A P value of <0.05 indicated significant difference. If the average age of the patients who carry a certain allele is significantly lower or higher than that of the patients who did not carry the allele, then the allele can be regarded as a disease-related allele (susceptibility or resistance allele). If the total number of genotype-positive chromosomes in the subjects is less than 5, the sample size is considered too small for statistical analysis.

Cross-validation of the 2 sets of analytical results

According to the probability principle of multiplication, the P value (type I error) is as follows:

$$P = 1 - (1 - P1) \times (1 - P2),$$

where *P1* and *P2* are the P values obtained with the results of the 2 analyses, i.e., the allele frequency distributions and the age at the onset of cancer, respectively, for 1 allele. A P value of <0.05 indicated significant difference.

RESULTS

The alleles identified in our experiments are summarized in Table 1. The frequency values of *HLA-DQB1*02* was higher in the cancer group than in the control population (P < 0.05). No statistical difference was observed between the 2 groups with respect to the other alleles.

Table 1. Allele distributions for *HLA-DQB1* in the rectal cancer and control groups.

HLA-DQB1*	Frequency of allele (%)		P
	Cancer	Control	
02	21.4	12.3	0.002
03	38.5	44.2	0.050
04	6.6	5.6	0.544
05	9.9	14.4	0.148
06	23.6	23.5	0.648
P	0.008		-

The age differences between the gene groups for *HLA-DQB1* are shown in Table 2. The median ages of the subjects with and without *HLA-DQB1*02* were 54.0 and 61.0 years, respectively, and a significant difference was observed between the ages for these groups (P < 0.05). The median ages of the subjects with and without *HLA-DQB1*03* were 62.0 and 58.0 years, respectively, and a significant difference was observed. No significant age differences were observed between the gene groups for the other alleles.

Table 2. Age differences between the 2 gene groups for *HLA-DQB1*.

HLA-DQB1*	Age (percentiles)			P
	25th	50th	75th	
2+	49.0	54.0	66.3	0.010
2-	55.0	61.0	72.0	
3+	57.0	62.0	73.0	0.037
3-	52.5	58.0	67.5	
4+	56.0	60.0	67.0	0.935
4-	54.0	60.0	71.0	
5+	53.0	66.0	76.0	0.129
5-	54.0	59.0	69.3	
6+	54.0	57.5	67.0	0.323
6-	53.0	61.0	72.3	

The cross-validation of the 2 above-mentioned analytical results is shown in Table 3. A statistically significant difference was noted between the results of the 2 analyses for *HLA-DQB1*02* (P < 0.05), whereas no such statistically significant difference was observed for *HLA-DQB1*03*.

Table 3. Cross-validation of the 2 analytical results.

HLA-DQB1*	P1	P2	P
2	0.002	0.010	0.012
3	0.05	0.037	0.085

P1 = P value obtained on analyzing allele frequency distribution; P2 = P value obtained on analyzing the age at the onset of cancer; $P = 1 - (1 - P1) \times (1 - P2)$.

DISCUSSION

Cancer is a typical complex disorder to which a yet unknown number of genes contribute through interactions among each other and the environment (Velculescu, 2008; Christiani et al., 2008). Specific antigens of the HLA system are known to be associated with many solid tumors (Schirle, 2002; Seliger et al., 2003; Cabrera et al., 2003). Specifically, persistent pro-tumor immune responses, generally accepted to potentiate primary tumor development, are recognized as mediators of cancer metastasis (Sun et al., 2008; Kim et al., 2008). Obviously, the immune response may be an important mechanism against cancers, unfortunately, the identification of the relationship between *HLA* polymorphisms and cancer remains a great challenge. This lack of progress can be partly attributed to the association analysis (Gao et al., 2010). To this regard, despite being discouraging due to inconsistent findings and weak linkage signals, the results of our previous linkage studies indicated that in the presence of a large number of alleles observed, confirming whether a statistical difference is due to a sampling error or a true cancer-related gene is impossible. Widely used in testing statistical hypotheses, the Bonferroni multiple test has a rather low power that entails a high risk to accept falsely the overall null hypothesis, underestimating real effects. We believe that in case of hypothesis tests conducted on a large number of single nucleotide polymorphisms, the method of cross-validation with 2 indicators from different is the preferred one to solve the problem of association study. Therefore, we developed a method based on cross-validation of the age at the onset of cancer and the differences between the frequency distributions of relevant genes (*HLA-DQB1*) in rectal cancer patients and controls.

Our results indicated statistical differences in frequency values between the cancer group and the control population for *HLA-DQB1*02*, and showed a significant age difference between the patients with and without *HLA-A*02* and *HLA-A*03*. No significant age differences were observed between the gene groups for the other alleles. Thus, we confirmed that *HLA-DQB1*02* is a true rectal cancer-related HLA allele.

Our new analysis method may be a more efficient and reliable approach for the identification of complex disease-causing genes, compared to Bonferroni correction. We expect an increasing integration between genetic studies, epidemiological studies, and clinical trials through virtual data sharing among laboratories, leading to genetically informative designs that will not only identify susceptibility and resistance genes, but also clarify how they cross interact.

ACKNOWLEDGMENTS

REFERENCES

Brewerton DA (2003). Discovery: HLA and disease. *Curr. Opin. Rheumatol.* 15: 369-373.

Cabrera T, Lopez-Nevot MA, Gaforio JJ, Ruiz-Cabello F, et al. (2003). Analysis of HLA expression in human tumor tissues. *Cancer Immunol. Immunother.* 52: 1-9.

Christiani DC, Mehta AJ and Yu CL (2008). Genetic susceptibility to occupational exposures. *Occup. Environ. Med.* 65: 430-436.

Fischer GF and Mayr WR (2001). Molecular genetics of the HLA complex. *Wien. Klin. Wochenschr.* 113: 814-824.

Gao X, Becker LC, Becker DM, Starmer JD, et al. (2010). Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.* 34: 100-105.

Hei AL, Li W, Deng ZH, He J, et al. (2009). Analysis of high-resolution HLA-A, -B, -Cw, -DRB1, and -DQB1 alleles and haplotypes in 718 Chinese marrow donors based on donor-recipient confirmatory typings. *Int. J. Immunogenet.* 36: 275-282.

Jeffery KJ and Bangham CR (2000). Do infectious diseases drive MHC diversity? *Microbes Infect.* 2: 1335-1341.

Kim S, Sunwoo JB, Yang L, Choi T, et al. (2008). HLA alleles determine differences in human natural killer cell responsiveness and potency. *Proc. Natl. Acad. Sci. U. S. A.* 105: 3053-3058.

Madeleine MM, Johnson LG, Smith AG, Hansen JA, et al. (2008). Comprehensive analysis of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 loci and squamous cell cervical cancer risk. *Cancer Res.* 68: 3532-3539.

Nyholt DR (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* 74: 765-769.

Schirle M (2002). Identification of tumor-associated HLA-ligands in the post-genomic era. *J. Hematother. Stem Cell Res.* 11: 873-881.

Seliger B, Abken H and Ferrone S (2003). HLA-G and MIC expression in tumors and their role in anti-tumor immunity. *Trends Immunol.* 24: 82-87.

Sun Y, Song M, Jager E, Schwer C, et al. (2008). Human CD4+ T lymphocytes recognize a vascular endothelial growth factor receptor-2-derived epitope in association with HLA-DR. *Clin. Cancer Res.* 14: 4306-4315.

Teriokhin AT, de MT and Guegan JF (2007). On the power of some binomial modifications of the Bonferroni multiple test. *Zh. Obshch. Biol.* 68: 332-340.

Velculescu VE (2008). Defining the blueprint of the cancer genome. *Carcinogenesis* 29: 1087-1091.

Vialatte FB and Cichocki A (2008). Split-test Bonferroni correction for QEEG statistical maps. *Biol. Cybern.* 98: 295-303.