COMPUTER MODELING OF GENE-GENE AND GENE-ENVIRONMENT INTERACTION IN ESSENTIAL HYPERTENSION

Olga Pavlova[1], Vladimir Malugin [2], Svetlana Ogurtsova[3], Alexander Novopolcev[2], Tatjana Gorbat[1], Maria Liventseva[1], Alexander Mrochek[1]

[1] Republican Scientific and Practical Centre «Cardiology»,
[2] Belarusian State University, Department of Mathematical Modeling and Data Analysis
[3] Institute of Bioorganic Chemistry of National Academy of Sciences Republic of Belarus
Minsk, Belarus
ol_skorochod@yahoo.com, malugin@bsu.by

Abstract. In this study the computer model of the gene-gene and gene-environment interaction among the polymorphisms nine candidate genes relating with essential hypertension (EH) and environmental cardiovascular risk factors such as overweight, abdominal obesity, smoking, insufficient physical activity was performed. There were determined five significant genetic patterns for hypertensive patients using APSampler software based on Markov Chain Monte-Carlo method with a specially adapted algorithm Metropolis – Hastings. The binary logit model with the high sensitivity demonstrated the cumulative effects the multilocus combinations and environmental factors associated with EH. This model allows classifying of subjects into two classes: the healthy and the hypertensive patients. Performance evaluation of the model by means of statistical tests indicates an acceptable accuracy of classification and prediction.

Keywords: Essential hypertension, Markov Chain Monte-Carlo method, Metropolis – Hastings algorithm, Binary dependent variable logit model**.**

1. Introduction

Essential hypertension (EH) is a complex disorder influenced by multiple genetic and environmental factors [1]. The renin-angiotensin-aldosterone system (RAAS), vascular endothelial and kallikrein-kinin systems have a vital role in the blood pressure (BP) regulation and the pathogenesis of EH. The polymorphisms (SNP) genes encoding components these systems are associated with EH and it was showed from previous studies [2]. But these associations are often not reproducible and depend from ethnicity and environmental factors. Environmental factors such as smoking, harmful use of alcohol, insufficient physical activity (PA), overweight, abdominal obesity (AO), psychological stress and depression also make influence on the development EH. The interaction of mutations' candidate genes and environmental factors may substantially increase susceptibility to EH. Objectives our study were to examine the gene-gene interaction among the polymorphisms nine

candidate genes of RAAS, vascular endothelial and kallikrein-kinin systems - angiotensin-converting enzyme (ACE), angiotensinogen (AGT), angiotensin II type 1 and 2 receptor (AGTR1, AGTR2), aldosterone synthase gene (CYP11B2), renin (REN), β2-bradykinin receptor gene (BKR2), methylenetetrahydrofolate reductase (MTHFR), endothelial nitric oxide synthase (eNOS) and gene-environment interaction in patients (pts) with EH.

2. Methods

A total of 532 subjects are included (356 hypertensive pts and 176 normotensives). Genotyping for ACE-I/D, AGT-M235T, AGTR1-A1166C, AGTR2-C3123A, CYP11B2-C344T, REN-19-83G/A, BKR2-T58C, eNOS-E298D, MTHFR-C677T polymorphisms were performed by polymerase chain reaction and restriction digestion. The following environmental (biological, behavioral and psychosocial) factors were assessed: office BP, overweight, AO, smoking, alcohol consumption, physical activity, level of psychological stress and depression using the international scales Psychological Stress Measure (PSM-25) and Center for Epidemiologic Studies Depression (CES-D), accordingly. The study included a search of polygenic associations to determine the genetic pattern, i.e. combination of alleles or genotypes of different locus associated with a phenotypic trait (the gene-gene interaction analysis) using APSampler software based on Markov chains (Markov Chain Monte-Carlo – MCMC method) with a specially adapted algorithm Metropois – Hastings (Metropois – Hastings algorithm) [3,4]. Then a binary logit model was built for estimate the cumulative effects the genetic and environmental factors associated with EH (the gene-environment interaction analysis). This model allows classifying of subjects into two classes: the healthy and the hypertensive patients. Performance evaluation of the model by means of statistical tests indicates an acceptable accuracy of classification and prediction.

3. Results

Mean ages of hypertensive pts was 49.0±11.4 years (192 men, 164 women) and healthy subjects - 43,0±10,5 years (78 men, 98 women). Differences BP between the patient and control groups were significant: average systolic/diastolic BP were 141.3±18.9/90.6±12.3 mmHg in hypertensive group and 120.8±8.7/77.0±7.4 mmHg in normotensives ($p<0.05$). The average BMI and waist circumference were significantly higher in hypertensives than in the healthy subjects (30.0±4.7 kg/m^2 vs 25.5±4.2 kg/m^2, $p<0.05$ and 98.1±12.9 cm vs 86.3±12.4 cm, $p<0.05$). The percentage of smokers in the past and subjects with insufficient PA (<1 time a week) were higher in pts than in normotensives (27.2% vs 17.6%, $p<0.01$ and 53.1% vs 40.3%, $p<0.01$; accordingly). The genotypes and alleles distribution of genetic polymorphisms did not differ between the groups. Firstly, the search of combinations of alleles and genotypes the studied polymorphisms nine genes was held in the total sample of study participants. Secondly, as originally it was observed differences in the two groups by

age and sex, the search for genetic patterns below have been carried out separately in age groups under 45 years old and after, and then, depending on the gender in each of the groups. Then it was obtained the significant genetic patterns for the examined groups in patients with EH and healthy subjects (Table 1).

Table 1. Significant genetic patterns for the hypertensive group

№	**Locus combinations**
1.	T allele AGT-M235T/AA genotype AGTR2-C3123A
2.	T allele AGT-M235T/T allele CYP11B2-C344T
3.	TT genotype AGT -M235T/D allele ACE-ID
4.	D allele eNOS-E298D/C allele BKR2-T58C
5.	D allele eNOS-E298D/ D allele ACE-ID

The binary dependent variable logit model [4] was constructed to explore the cumulative effects the significant multilocus combinations and environmental (biological, behavioral and psychosocial) factors. The results of estimation and testing the model (Table 2) indicate the statistical significance of all factors at the level near 0.05 and below as well as the adequacy of the model as a whole.
The model allows classifying of subjects into two classes: the healthy (depended variables Dep=0) and the hypertensive patients (Dep=1). Expectation-prediction evaluation of the model based on the classification tables give the following estimates of accuracy of the classification: overall, the estimated model correctly predicts 74.08% of the observations; the percentage of correct decisions in the classification of hypertensive patients is 86.42%.

Table 2. Estimation results for the binary logit model with genetic and non-genetic factors

Variables	**β-coefficients**	**Std. Error**	***z*-Statistic**	***p*-value**
Age	0.928626	0.224505	4.136318	0.0000
Smoking	0.438137	0.220897	1.983440	0.0473
AO	0.723835	0.273650	2.645118	0.0082
BMI	1.380139	0.310520	4.444605	0.0000
Insufficient PA	0.631120	0.218890	2.883269	0.0039
T allele AGT-M235T/AA genotype AGTR2-C3123A	0.491607	0.253880	1.936374	0.0528
D allele eNOS-E298D/ D allele ACE-ID	0.713285	0.223251	3.194985	0.0014
C	-1.295533	0.253726	-5.106024	0.0000
McFadden *R*-squared	0.185426	*LR* statistic (*p*-value)		121.6938 (0.0000)

We also used two tests to evaluate the goodness-of fit properties of the model: Hosmer – Lemeshow test [5] and Andrews test [6]. These tests allow comparing the fitted expected values to the actual values by grouping based upon predicted risk. The hypothesis of the proximity of the observed and expected values of the number of patients for all groups of cases is not rejected: *p*-values greater than 0.05 for both criteria (Table 3).

Table 3. Goodness-of-Fit testing statistics

Test	**Statistic value**	**Chi-Square statistic: d.f., *p*-value**
Hosmer – Lemeshow	4.8973	d.f. = 8, 0.7685
Andrews	10.2924	d.f. = 10, 0.4152

4. Conclusions

The interaction of genetic and environmental factors such as the multilocus combinations D allele of eNOS3-E298D and D allele of ACE-ID, T allele AGT-M235T and genotype AA of AGTR2-C3123A with overweight, abdominal obesity, insufficient physical activity, smoking may increase the susceptibility to the development to EH.

5. References

1. Lifton RP, Gharavia G, Geller DS. Molecular mechanisms of human hypertension. Cell.2001; 104(4):545-556.
2. Abbate R, Sticchi E, Fatini C. Genetics of cardiovascular disease. Clin Cases Miner Bone Metab. 2008(Jan); 5(1):63–66.
3. Favorov A.V., Andreewski T.V., Sudomoina M.A., Favorova O.O., Parmigiani G., Ochs M.F. A Markov Chain Monte Carlo technique for identification of combinations of allelic variants underlying complex diseases in Humans. Genetics. 2005 Dec;171(4):2113-21.
4. Markov Chain Monte Carlo in Practice / eds: W. R. Gilks, S. Richardson and. D. J. Spiegelhalter. Chapman & Hall, 1996. – 486 p.
5. Greene, William H. (2008). Econometric Analysis, 6th Edition, Upper Saddle River, NJ: Prentice-Hall.
6. Hosmer, David W. Jr. and Stanley Lemeshow (1989). Applied Logistic Regression, New York: John Wiley &Sons.
7. Andrews, Donald W. K. (1988). "Chi-Square Diagnostic Tests for Econometric Models: Introduction and Applications", Journal of Econometrics, 37, 135–156.