UDC: 575.17+61

Prognostic value of markers in associative genetics

H. Ehyakonandeh, L.A. Atramentova

The article presents an algorithm for genetic-statistical analysis of a rare chronic disease with a hereditary predisposition. The results of genotyping obtained by one of the authors by the method of associative genetics were used. Multiple sclerosis is a disease with a hereditary predisposition. At present, the association of multiple sclerosis and a large number of genes has already been discovered, but the genetic control is not completely clear. An important step in combating the disease can be preventive measures among groups of people with hereditary predisposition. In this article, using the example of the *C677T* genetic polymorphism of the *MTHFR* gene, we show how to determine the effectiveness of identifying individuals at increased risk. Multiple sclerotic patients (n=180) and healthy (n=231) residents of southern Iran were examined by one of the authors previously. Based on these data, population-genetic indicators and statistical characteristics of the test were calculated. The distribution of genotypes in healthy people: *CC* – 65 %, *CT* – 29 %, *TT* – 6 %, in patients with multiple sclerosis *CC* – 35 %, *CT* – 46 %, *TT* – 19 %. The major allele in the population of southern Iran is *C* (p_C = 0.797; q_T = 0.203). The frequency of the minor *T* allele is doubled in the group of patients compared with healthy ones (q_T = 0.419). The *T* allele is considered to be provocative; the allele *C* is protective. The *CC* genotype reduces the likelihood of multiple sclerosis by almost half compared with the empirical risk. In heterozygotes of *CT*, the risk is increased by more than one and half times, in homozygotes of *TT* more than three times. 95 % *CI* confidence intervals for the *OR* odds ratio indicator are: *CC* (0.19–0.44), *CT* (1.36–3.10), *TT* (1.99–7.61), *CT* + *TT* (2.29–5.21). The statistical characteristics of the test indicate its low power when used in screening programs. The sensitivity when testing carriers of the *T* allele (*CT* + *TT* genotypes) is 65 %. The very low prognostic value of a positive test makes it inappropriate to use for screening, but this test may be useful in individual genetic counselling for patients with multiple sclerosis, as well as their relatives. The analysis scheme can be used in other studies of traits with a genetic component.

Key words: *genetic polymorphism, associative genetics, statistical characteristics of the test.*

About the authors:
H. Ehyakonandeh – V.N.Karazin Kharkiv National University, Svobody Sq., 4, Kharkiv, Ukraine, 61022, hesam.ehyakonandeh@yahoo.com, https://orcid.org/0000-0002-2964-901X
L.A. Atramentova – V.N.Karazin Kharkiv National University, Svobody Sq., 4, Kharkiv, Ukraine, 61022, lubov.atramentova@gmail.com, https://orcid.org/0000-0002-7143-9411

Introduction

Multiple sclerosis is a complex neurological disease that affects the central nervous system resulting in debilitating neuropathology. Pathogenesis is primarily defined by inflammation and demyelination of nerve axons. Although multiple sclerosis was first described in patients over 150 years ago, the exact etiology and pathogenesis of the disease remain unclear. There are certainly immunological factors, which are involved in the disease pathogenesis. However, epidemiological studies suggest that still unknown genetic factors also can contribute to the etiology of this disease. The precise etiology of multiple sclerosis remains elusive with a complex interplay between environmental factors, genetic susceptibility, and age-dependant exposure to viral infection (Levin et al., 2005).

Genetic predisposition has long been suspected in the etiology of this disease. At present, its associations with a large number of genes have already been discovered, but the genetic control is not completely clear. An important step in combating the disease can be preventive measures among groups of people with hereditary predisposition. The association between *MTHFR* polymorphisms and multiple sclerosis has been investigated in different ethnic groups (Alatab et al., 2011; De Marco et al., 2002; Jonasdottir et al., 2003; Fekih Mrissa et al., 2013; Klotz et al., 2010). The association between *MTHFR C677T* variants and multiple sclerosis has been revealed in the southern Iranian population (Naghibalhossaini et al., 2015). We used these results to continue the analysis in terms of population genetics and examined whether this polymorphism can be used in a screening programs and in the genetic counseling.

Materials and methods

The distributions of genotypes in the groups of patients and healthy people were compared using the criterion χ^2:

DOI: 10.26565/2075-5457-2020-34-7

$$\chi^2 = \frac{N^2}{n_1 n_2}\left(\sum_{i=1}^{k}\frac{f_1^2}{f_1+f_2} - \frac{n_1^2}{N}\right);\ df = k - 1$$

n_1 and n_2 – the number of observations in the compared groups ($N = n_1 + n_2$), f_1 and f_2 – the number of genotypes in the compared groups, *df* – the degree of freedom, *k* – the number of classes.

The coefficient of association *r* between allele and disease was calculated as:

$$r = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}};\quad \chi^2 = r^2 n\ ;\ df = 1$$

a – the number of *T* alleles in the group of patients, *b* – the number of *C* alleles in the group of patients, *c* – the number of *T* alleles in the group of healthy people, *d* – the number of *C* alleles in the group of healthy people, *n* – the total number of alleles in two groups, *df* – the degree of freedom.

Allele frequencies p_C and q_T were calculated as:

$$p_C = \frac{2n_{CC}+n_{CT}}{2(n_{CC}+n_{CT}+n_{TT})};\ q_T = \frac{2n_{TT}+n_{CT}}{2(n_{CC}+n_{CT}+n_{TT})}$$

n – the number of the genotype.

The odds ratio (*OR*) was calculated:

$$OR = \frac{ad}{bc}$$

a – the number of patients with the predisposition genotype, *b* – the number of patients without the predisposition genotype, *c* – the number of healthy people with the predisposition genotype, *d* – the number of healthy people without the predisposition genotype.

The confidence interval (*CI*) of *OR* was carried out in the logarithmic scale.

$$CI:\ \ln OR \pm t s_{\ln OR}$$

t – the Student coefficient (for 95 % *CI* $t = 1.96$), $s_{\ln OR}$ – the statistical error for ln*OR*.

The statistical error ln*OR* was calculated as:

$$s_{\ln OR} = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

The s_{lnOR} was used to obtain the 95 % confidence interval of ln*OR*. The limit values of the confidence interval ln*OR* were converted to the boundary values of the confidence interval *OR* by a back procedure: *exp* ln*OR* (Armitage, Berry, 1994).

The index of sensitivity (*Sen*) and specificity (*Spe*) of the tests was calculated using the formulas:

$$Sen = \frac{a}{a+c},\ Spe = \frac{d}{b+d}$$

The prevalence of multiple sclerosis obtained from (Izadi et al., 2015) is used for the calculations of the follow statistical characteristics. The calculation of the prognostic value of the positive result (*PVP*) and

the prognostic value of the negative result (*PVN*) was carried out according to the formulas (Atramentova, 2015):

$$PVP = \frac{a}{a + b^*}; \quad PVN = \frac{d^*}{c + d^*};$$

$$b^* = \frac{b}{b + d}(N^* - a - c), \; N^* = \frac{a + c}{\pi}; \; d^* = N^*(1 - \pi) - b^*$$

a – the number of patients with the test genotype, *b* – the number of healthy people with the test genotype, *c* – the number of patients without the test genotype, *d* – the number of healthy people without the test genotype.

Results and discussion

The group of healthy people due to the negligible prevalence of multiple sclerosis (Izadi et al., 2015) can be considered as a sample from the whole population. In the group of patients, the proportion of *CC* genotype is almost two times less than in the group of healthy people. The number of patients with *CT* genotype is one and a half times more than in the control group, the number of people with *TT* genotype is three times more than in control group, so the presence of the *T* allele in the genotype is considered as an indicator of increased risk for multiple sclerosis in the population of southern Iran.

The major allele in the Iranian population is *C* (*pC* = 0.797). The frequency of the minor allele *T* in patients (*qT* = 0.419) is two times higher than in the group of healthy people (*qT* = 0.203; table 1). The distribution of genotypes in this group corresponds to the Hardy – Weinberg equilibrium.

Table 1.

The distribution of genotypes in the groups of healthy people and patients

Group	Index	Total*	Genotypes			Allele frequencies	
			CC	*CT*	*TT*	*C*	*T*
Control	*n* (%)	231 (100)	150 (65.0)	68 (29.4)	13 (5.6)	0.797	0.203
Patients	*n (%)*	180 (100)	63 (35.0)	83 (46.1)	34 (18.9)	0.581	0.419
Relative risk			0,54	1,57	3,38	r_A=*0.23*	
Statistics	$\chi^2_{0,001\ (2)}$=13,8; χ^2=40,6; *p*<0,001					$\chi^2_{0,001\ (1)}$=10.8 χ^2=45.1; *p*<0,001	

**Naghibalhossaini et al., 2015.*

The prevalence of multiple sclerosis can be considered as an empirical risk for this disease if there is no other information. The risk changes if new information appears. *CC* genotype reduces the likelihood of the disease in comparison with the empirical risk almost twice (0.54, table 1). *TT* genotype increases the likelihood 3.38 times compared with empirical risk. In heterozygotes (*CT*) the probability of the disease is increased by 1.56 times. Thus, the *T* allele is considered to be provocative for multiple sclerosis, and the *C* allele is protective.

Table 2.

Odds ratio for different genotypes

Test genotype	*a*	*b*	*c*	*d*	*OR*	ln*OR*	$s_{\ln OR}$	95% *CI* ln*OR*	95% *CI OR*
CC	63	150	117	81	0.29	-1.24	0.21	-1.65 – -0.83	0.19 – 0.44
CT	83	68	97	163	2.05	0.72	0.21	0.31 – 1.13	1.36 – 3.10
TT	34	13	146	218	3.91	1.36	0.34	0.69 – 2.03	1.99 – 7.61
CT+TT	117	81	63	150	3.44	1.24	0.21	0.83 – 1.65	2.29 – 5.21

a – the number of patients with the test genotype, *b* – the number of healthy people with the test genotype, *c* – the number of patients without the test genotype, *d* – the number of healthy people without the test genotype, *OR* – the odds ratio, $s_{\ln OR}$ – the statistical error of ln*OR, CI* – the confidence interval.

Given that the probability of disease in heterozygotes *CT* is half that the probability of homozygotes *TT* it can be argued that the interaction of these alleles exhibits a half-dominant effect or incomplete dominance with regard to multiple sclerosis. Since the presence in the genotype *T* allele increases the risk of multiple sclerosis, we combined these two genotypes (*CT* + *TT*) and designated this common group as a group of increased risk for multiple sclerosis. The individuals with the *CC* genotype are designated as the anti-risk group.

Confidence intervals of OR do not include 1 (table 2), so all indices are statistically significant ($p<0.05$). Therefore, *MTHFR C677T* polymorphism can be used as a marker of genetic predisposition to multiple sclerosis. The statistical characteristics of the test were calculated in order to estimate the predictive value of this polymorphism. The prevalence of multiple sclerosis in this population ($\pi = 0.000721$, Izadi et al., 2015) was used for further calculation.

The sensitivity of this test is rather low. The test for homozygote genotype *TT* reveals only 19 % individuals, which have the hereditary predispositions to multiple sclerosis, and the test for heterozygosity reveals 46 %. The *T*-allele in whole without taking into account the genotype identifies 65 % of potential patients (table 3).

Table 3.

Statistic characteristics of tests

Characteristics	Genotype		
	TT	*CT*	*TT + CT*
π = 0.000721*			
Index of sensitivity, %	18.9	46.1	65.0
Index of specificity, %	93.2	70.6	64.9
Prognostic value of positive result, %	0.24	0.11	0.13
Prognostic value of negative result, %	99.9	99.9	99.9

*π – prevalence of multiple sclerosis in southern Iran, * – Izadi et al., 2015.*

The prognostic value of the positive result is negligible (less than 1 %) and is of no practical value. High prognostic value of the negative result (> 99.9 %) is of no practical significance as well (table 3). Thus, the using of this polymorphism for screening has no practical value. Nevertheless, this polymorphism can be useful in the individual genetic counseling of the multiple sclerotic patients and their relatives. The analysis algorithm described in the article can be used to study a variety of traits with a genetic component and diseases with a hereditary predisposition.

References

Alatab S., Hossein-nezhad A., Mirzaei K. et al. (2011). Inflammatory profile, age of onset, and the *MTHFR* polymorphism in patients with multiple sclerosis. *J. Mol. Neurosci., 44*(1), 6–11. https://doi.org/10.1007/s12031-010-9486-y

Armitage P., Berry G. (1994). *Statistical methods in medical research*. Blackwell Scientific Publications, Oxford, 620 p.

Atramentova L.A. (2015). *Design and statistics of biological research*. Kharkiv: NTMT, 269 p. (in Russian)

De Marco P., Calevo M.G., Moroni A. et al. (2002). Study of *MTHFR* and MS polymorphisms as risk factors for NTD in the Italian population. *J. Hum. Genet., 47*(6), 319–324. https://doi.org/10.1007/s100380200043

Fekih Mrissa N., Mrad M., Klai S. et al. (2013). Association of methylenetetrahydrofolate reductase A1298C polymorphism but not of C677T with multiple sclerosis in Tunisian patients. *Clin. Neurol. Neurosurg., 115*(9), 1657–1660. https://doi.org/10.1016/j.clineuro.2013.02.025

Izadi S., Nikseresht A.R., Poursadeghfard M. et al. (2015). Prevalence and incidence of multiple sclerosis in Fars Province, Southern Iran. *Iran J. Med. Sci., 40*(5), 390–395.

Jonasdottir A., Thorlacius T., Fossdal R. et al. (2003). A whole genome association study in Icelandic multiple sclerosis patients with 4804 markers. *J. Neuroimmunol., 143*(1–2), 88–92. https://doi.org/10.1016/j.jneuroim.2003.08.018

Klotz L., Farkas M., Bian N. et al. (2010). The variant methylenetetrahydrofolate reductase c.1298A>A(p.E429A) is associated with multiple sclerosis in German case-control study. *Neuroscience Lett., 468,* 183–185. https://doi.org/10.1016/j.neulet.2009.10.057
Levin L.I., Munger K.L., Ruberstone M.V. et al. (2005). Temporal relationship between elevation of Epstein–Barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis. *JAMA, 293*(20). – P. 2496–2500. https://doi.org/10.1001/jama.293.20.2496
Naghibalhossaini F., Ehyakonandeh H., Nikseresht A., Kamali E. (2015). Association between *MTHFR* genetic variants and multiple sclerosis in a Southern Iranian Population. *Int. J. Mol. Cell Med., 4*(2), 87–93.

Прогностичне значення маркерів в асоціативній генетиці

Х. Ехьяконандех, Л.О. Атраментова

В статті представлено алгоритм генетико-статистичного аналізу рідкісного хронічного захворювання із спадковою схильністю. Використані результати генотипування, одержані одним з авторів методом асоціативної генетики. Досліджені хворі на розсіяний склероз та здорові мешканці південного Ірану. Розсіяний склероз – захворювання із спадковою схильністю. До теперішнього часу вже визначені його асоціації з великою кількістю генів, однак генетичний контроль остаточно ще не з'ясовано. Важливим етапом протидії захворюванню з генетичною схильністю можуть бути профілактичні заходи серед осіб підвищеного ризику. У даній статті на прикладі генетичного поліморфізму *C677T* гена *MTHFR* показано, як з'ясувати ефективність виявлення осіб з підвищеним ризиком до захворювання, використовуючи знайдені асоціації захворювання з генетичним маркером. Були використані результати генотипування 180 хворих на розсіяний склероз і 231 здорового індивіда. За цими даними розраховані популяційно-генетичні показники і статистичні характеристики тесту. Розподіл генотипів у здорових людей: *CC* – 65 %, *CT* – 29 %, *TT* – 6 %, у хворих на розсіяний склероз *CC* – 35 %, *CT* – 46 %, *TT* – 19 %. Мажорним алелем в популяції південного Ірану є *C* (p_C = 0,797; q_T = 0.203). Частота мінорного алеля *T* підвищена в групі хворих у порівнянні зі здоровими в два рази (q_T = 0.419). Алель *T* розглядається як провокативний, алель *C* є протективним. Генотип *CC* знижує ймовірність розсіяного склерозу майже в два рази в порівнянні з емпіричним ризиком. У гетерозигот *CT* ризик збільшений у півтора рази, у гомозигот *TT* – в три рази. Довірчі інтервали 95 % *CI* для показника відношення шансів *OR* складають: *CC* (0,19–0,44), *CT* (1,36–3,10), *TT* (1,99–7,61), *CT* + *TT* (2,29–5,21). Чутливість при тестуванні на наявність алеля *T* в генотипі (*CT* + *TT*) становить 65 %. Дуже низьке прогностичне значення позитивного тесту (менше 1 %) робить недоцільним його використання для масового скринінгу, але цей тест може бути корисним при індивідуальному генетичному консультуванні пацієнтів з розсіяним склерозом, а також їхніх родичів. Схема аналізу може бути використана в інших дослідженнях ознак з генетичною компонентою.

Ключові слова: *генетичний поліморфізм, асоціативна генетика, статистичні характеристики тесту.*

Про авторів:
Х. Ехьяконандех – Харківський національний університет імені В.Н.Каразіна, пл. Свободи, 4, Харків, Україна, 61022, hesam.ehyakonandeh@yahoo.com, https://orcid.org/0000-0002-2964-901X
Л.О. Атраментова – Харківський національний університет імені В.Н.Каразіна, пл. Свободи, 4, Харків, Україна, 61022, lubov.atramentova@gmail.com, https://orcid.org/0000-0002-7143-9411

Прогностическое значение маркёров в ассоциативной генетике

Х. Эхьяконандех, Л.А. Атраментова

В статье представлен алгоритм генетико-статистического анализа редкого хронического заболевания с наследственной предрасположенностью. Использованы результаты генотипирования, полученные одним из авторов методом ассоциативной генетики. Исследованы больные рассеянным склерозом и здоровые жители южного Ирана. Рассеянный склероз – заболевание с наследственной предрасположенностью. В настоящее время уже обнаружены его ассоциации с большим количеством генов, однако генетический контроль окончательно ещё не выяснен. Важным этапом противодействия таким заболеваниям могут быть профилактические мероприятия среди лиц повышенного риска. В данной статье на примере генетического пролиморфизма *C677T* гена *MTHFR* показано, как выяснить эффективность выявления группы повышенного риска. Анализ выполнен на данных о генотипах 180 больных рассеянным склерозом и 231 здоровом индивиде из населения южного Ирана. Рассчитаны популяционно-генетические показатели и статистические характеристики теста. Распределение генотипов у здоровых людей: *CC* – 65 %, *CT* – 29 %, *TT* – 6 %, у больных рассеянным склерозом *CC* – 35 %, *CT* – 46 %, *TT* – 19 %. Мажорным аллелем в популяции южного Ирана является *C* (p_C = 0,797; q_T = 0,203). Частота минорного аллеля *T* повышена в группе больных по сравнению со здоровыми в два раза (q_T = 0,419). Аллель *T* рассматривается как провокативный, аллель *C* является протективным. Генотип *CC* снижает вероятность рассеянного склероза почти в два раза по сравнению с

эмпирическим риском. У гетерозигот *CT* риск увеличен в полтора раза, у гомозигот *TT* в три раза. Доверительные интервалы 95 % *CI* для показателя отношение шансов *OR* составляют: *CC* (0,19–0,44), *CT* (1,36–3,10), *TT* (1,99–7,61), *CT+TT* (2,29–5,21). Чувствительность при тестировании на присутствие в генотипе аллеля *T* (*CT+TT*) составляет 65 %. Очень низкое прогностическое значение положительного теста (менее 1 %) делает нецелесообразным его использование для скрининга, но этот тест может быть полезным при индивидуальном генетическом консультировании пациентов с рассеянным склерозом, а также их родственников. Схема анализа может быть использована в других исследованиях признаков с генетическим компонентом.

Ключевые слова: *генетический полиморфизм, ассоциативная генетика, генетико-статистические характеристики теста.*

Об авторах:
Х. Эхьяконандех – Харьковский национальный университет имени В.Н.Каразина, пл. Свободы, 4, Харьков, Украина, 61022, hesam.ehyakonandeh@yahoo.com, https://orcid.org/0000-0002-2964-901X
Л.О. Атраментова – Харьковский национальный университет имени В.Н.Каразина, пл. Свободы, 4, Харьков, Украина, 61022, lubov.atramentova@gmail.com, https://orcid.org/0000-0002-7143-9411

Представлено: О.В.Філіпцова / Presented by: O.V.Filiptsova
Рецензент: Д.А.Шабанов / Reviewer: D.A.Shabanov
Подано до редакції / Received: 13.04.2020