Cite this article: *Atramentova L., Ehyakonandeh H. Molecular genetic data in terms of associative and population genetics. The Journal of V. N. Karazin Kharkiv National University, Series "Biology", 2021, 36, 35–40.*

••• ГЕНЕТИКА ••• GENETICS •••

UDC: 575.17+61

Molecular genetic data in terms of associative and population genetics

L. Atramentova, H. Ehyakonandeh

In studies on associative genetics of various multifactorial diseases, it is most often found that the minor allele's frequency in the group of patients is higher than in the group of healthy people. Due to reduced adaptation, the minor allele manifests itself as a disease. In the group of patients, the number of homozygotes by major allele is reduced, the number of heterozygous carriers of the provocative allele is increased, and the frequency of homozygotes by the provocative allele is significantly increased. The aim of this article was to analyze the unusual result for SNP 1298*A/C* of the *MTHFR* gene in multiple sclerosis, previously obtained by one of the authors. The allele frequencies in the control group and in the group of multiple sclerosis do not differ, but the distribution of genotypes in the patients does not correspond to the Hardy–Weinberg ratio in compare to healthy people. Among patients, the number of heterozygotes is increased and the number of each homozygote is decreased. The increase in the proportion of heterozygotes can be explained by the presence of one provocative allele, but the large shortage of homozygotes for the minor allele is unclear. To explain this fact, the composition of the group of patients was analysed. The patients are of different ages, this group is heterogeneous in sex and form of multiple sclerosis, but none of these indicators has not be taken into account in the analysis of the distribution of genotypes. The age of the disease is a diagnostic sign and may depend on the genotype. The manifestation of multifactorial diseases depends on gender as well. Clinical forms of the disease usually have a different genetic basis. Due to the neglect of these conditions, a genetically heterogeneous group is formed, and any result, difficult for explanation, can be obtained. The lack of *CC* genotypes may be due to increased mortality, which reduces the probability of patients to be investigated in the sample. These facts once again indicate the need to form homogeneous groups for research on associative genetics.

Key words: *multiple sclerosis, SNP 1298A/C MTHFR, associative genetics, science methodology, research design, sample formation, natural selection, Hardy–Weinberg equilibrium.*

About the authors:
L. Atramentova – V.N. Karazin Kharkiv National University, Svobody Sq., 4, Kharkiv, Ukraine, 61022, lubov.atramentova@gmail.com, https://orcid.org/0000-0002-7143-9411
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

Introduction

Associative genetics typically uses a case/control pattern. The most important condition for a reliable conclusion from such is an investigation of representative samples. When forming a group of patients and a control group, their comparability by sex, age, ethnic origin, and lifestyle must be ensured. When it comes to multifactorial pathological signs with varying age of manifestation, the formation of comparable groups becomes especially difficult. In such research works, the most frequent errors are generated as a result of incorrect methodology which distort the result. A frequent drawback of many works on associative genetics is the use of biased samples. Usually they are heterogeneous by gender, age, origin of group members, and the variance of these indicators in the compared groups does not coincide. When characterizing the groups being compared, authors often only state their comparability, without giving any numerical characteristics, or giving only fragmentary information from which it is impossible to make an idea of the true state of affairs. It is known from practice that a group of patients is usually more heterogeneous in age and the statement about the equality of middle age and/or other biologically important indicators is not a serious argument for the methodological correctness of the study. The distribution of subjects by gender is also a big problem. Usually in the main group, the proportion of men and women reflects this ratio in the population. In the control group, this ratio is determined by the unequal willingness of healthy men and women to participate in the study.

DOI: 10.26565/2075-5457-2021-36-4

It is not difficult to perform a formal genetic analysis according to molecular genotyping data, but it can be difficult to explain the result obtained in such samples. This also raises the question of the reliability of the source data. The analysis of the questions that arise when reading articles devoted to the description of the results of molecular genetic analysis was the aim of this article.

Materials and methods

For the analysis, we used the results of genotyping by SNP 1298*A/C* of the *MTHFR* gene obtained by one of the authors of this article (Naghibalhossaini et al., 2015). 180 patients with multiple sclerosis and 186 healthy residents of southern Iran were examined. The algorithm of genetic-statistical analysis was described in our previous work (Ehyakonandeh, Atramentova, 2020). The prevalence index of multiple sclerosis obtained from (Izadi et al., 2015) was used in calculations of some statistical indicators (Atramentova, 2015).

Results

The frequency distribution of genotypes in the group of patients with multiple sclerosis differs significantly from the distribution of genotypes in the control group (table 1). The number of *AC* heterozygotes in this group is almost one and a half times higher than the control value (68 % vs 46 %), and the proportion of homozygotes is reduced: *AA* is almost one and a half times (29 % vs 42 %), *CC* is more than four times (3 % vs 12 %). The frequency distribution of genotypes in the control group ideally corresponds to the Hardy–Weinberg proportion ($p^2_{AA} : 2pq_{AC} : q^2_{CC}$).

Table 1. **The distribution of genotypes in healthy people and patients**

Group	Index	Genotypes *1298A/C*			
		AA	*AC*	*CC*	Total
Control	*n**	79	85	22	186
	%	42.5	45.7	11.8	100
	95% CI	35.6 – 49.7	38.7 – 52.9	7.9 – 17.3	
Patients	*n**	53	122	5	180
	%	29.4	67.8	2.8	100
	95% *CI*	23.3 – 36.5	60.6 – 74.2	1.0 – 6.5	
	Δ	↓1.44	↑1.48	↓4.21	
	RR	0.69	1.48	0.24	

** – Naghibalhossaini et al., 2015, CI – confidence interval, RR – relative risk, Δ – change from control, ↓ – decrease, ↑ – increase.*

Allele *A* is the major one in the population of southern Iran. Allele frequency ratio (p_A = 0.653, q_C = 0.347) significantly differs from the 1 : 1 ratio ($p < 0.05$), that may indicate the adaptive disambiguity of these alleles with a selective advantage of one of them, although the drift of genes cannot be ruled out. Though the allele frequencies in patients were almost the same as in the control group, the distribution of genotypes did not correspond to the Hardy–Weinberg proportion.

Table 2. The allele frequencies

Group	Allele	Frequency	95 % *CI*
Control	*A*	0.653	0.604 – 0.700
	C	0.347	0.300 – 0.397
Patients	*A*	0.633	0.582 – 0.682
	C	0.347	0.319 – 0.418

CI – confidence interval.

Confidence intervals indicate the significance of all calculated odds ratio (*OR*) (table 3).

Table 3. Odds ratio for different genotypes

Testing genotype	*a*	*b*	*c*	*d*	*OR*	ln*OR*	s_{lnOR}	95% *CI* ln*OR*	95% *CI OR*
AA	53	79	127	107	0.565	–0.571	0.22	–1.011....– 0.131	0.36 – 0.88
AC	122	85	58	101	2.500	0.916	0.22	0.476 – 1.356	1.61 – 3.88
CC	5	22	175	164	0,213	-1.532	0.51	–2.552… –0.512	0.08 – 0.60
AA+CC	58	101	122	85	0,400	-0.916	0.22	–1.356… –0.476	0.26 – 0.48

a – the number of patients with the testing genotype, b – the number of healthy people with the testing genotype, c – the number of patients without the testing genotype, d – the number of healthy people without the testing genotype, OR – the odds ratio, s_{lnOR} *– the standard error of lnOR, CI – the confidence interval.*

The *RR* value shows how much the risk of the disease is changing for the given genotype in comparison to the empirical risk (π) calculated as πRR. So, if we assume that the empirical probability for a resident of southern Iran to get multiple sclerosis is approximately 0.072 %, then the risk for *AA* and *CC* genotypes is reduced, and for the genotype *AC* is increased. The feasibility of using this method for screening programs can be estimated based on the statistical characteristics of the test (table 4).

Table 4. Statistic characteristics of the tests

Characteristics	Genotype		
	AA	*CC*	*AC*
Prevalence, π = 0.000721*			
Index of sensitivity, %	40.2	18.5	58.9
Index of specificity, %	59.8	81.5	41.1
Prognostic value of positive result, %	0.05	0.03	1.2
Prognostic value of negative result, %	99.91	99.92	99.95

** Izadi et al., 2015*

As can be seen, the prognostic significance of the results of testing the genotype for polymorphism 1298*A/C* of the *MTHFR* gene is very low, and this method is not advisable to include in screening programs. Nevertheless, it can be useful for consulting patients when choosing a treatment method taking into account the characteristics of the genotype, if any are found in special studies.

Discussion

In studies on associative genetics of various multifactorial diseases, it is most often found that the minor allele's frequency in the group of patients is higher than in the group of healthy people. This is not surprising, since it logically follows from the fact of the minority of the allele, due to its reduced adaptability, which manifests itself as a disease. In such cases, the minor allele is designated as provocative, and the major one as preventive in relation to the studied disease. The change in allele frequencies in the group of patients is caused by a shift in the ratio of genotypes compared to the control group. In the group of patients, the frequency of heterozygotes with one dose of a provocative allele is usually increased, and the frequency of homozygotes in the minor allele with two doses of a provocative allele is even more increased. The frequency of homozygotes for the major preventive allele is lower than in the control. This ratio of genotypes in the main and control groups is observed with the same viability of all genotypes. The similar result was found for the SNP 677*T/C* genotype of the same gene (Naghibalhossaini et al., 2015).

With regard to the *MTHFR* gene described by SNP 1298*A/C*, an unusual result was obtained. The allele frequencies in the control group and in patients with multiple sclerosis practically do not differ, but the distribution of genotypes in the group of patients is unusual. Among patients, the frequency of heterozygotes is increased and the frequency of both homozygotes is reduced. The formal phrase that

the observed distribution of genotypes corresponds to a certain genetic model gives little to understand this unusual result. It can be explained in different ways, for example by selection against heterozygotes. In this case, heterozygotes are characterized by reduced fitness compared to both homozygotes. The most famous example of the selection against heterozygotes is the hemolytic disease of newborns due to the incompatibility of Rh-positive fetus and Rh-negative mother. The physiological mechanism of hemolytic disease is well known, but how the reduced fitness of patients with the 1298*AC* genotype of the *MTHFR* gene is formed is not yet clear. Another reason for the unusual distribution of genotypes in the group of patients may be the uneven survival of patients with different genotypes (Fig. 1).

<table>
<tr><th>AA</th><th>AC</th><th>CC</th></tr>
<tr><td rowspan="2">Healthy</td><td>Healthy</td><td>Healthy</td></tr>
<tr><td rowspan="2">Multiple sclerosis</td><td>Deaths due to multiple sclerosis</td></tr>
<tr><td>Multiple sclerosis</td><td>Multiple sclerosis</td></tr>
</table>

Fig. 1. Scheme of population distribution and inclusion of patients in the sample

Following the logic described above, we assume that the major allele *A* of polymorphism *1298A/C MTHFR* is protective, and the minor *C* is provocative. Then in the group of patients we naturally expect the percentage of *AA* homozygotes to be lower, and the percentage of *AC* heterozygotes and *CC* homozygotes to be higher than in the control. Nevertheless, the result coincides with the expected only partially and the patient sample seems to be truncated due to the lack of *CC* genotypes. It can be assumed that the lack of *CC* genotypes is due to more severe manifestation of the disease with increased mortality rate, which reduces the likelihood of patients with the *CC* genotype falling into the researcher's field of vision. This explanation seems reasonable, since the increased mortality rate of patients with multiple sclerosis is the well-known fact (Lunde et al., 2017).

The methodological problem of this study is probably non-comparable sex ratio in studied groups. In the group of multiple sclerosis, women account for 74 %, and the sex ratio in the control group is not indicated at all. Exposure to multifactorial diseases is known to be modified by sex. This means that the genetic and statistical characteristics obtained in such samples are subject to a statistical artifact similar to the long-known Simpson's paradox (Armitage, Berry, 1994). Multiple sclerosis is an age-related disease. Although the groups are heterogeneous by age, statistical indicators of the age distribution of patients and healthy people are not indicated, namely age can be the cause of a shift in genotypes due to differential survival of patients. It is important in terms of genetics that the studied group of patients is phenotypically heterogeneous, it presents patients with three forms of multiple sclerosis: relapsing–remitting multiple sclerosis (MS) – 128 (71 %), secondary progressive MS – 43 (24 %), primary progressive MS – 9 (5 %). These forms of multiple sclerosis can be associated with various genetic variants, therefore it is always important to ensure their maximum possible uniformity when forming groups for genetic analysis.

When reviewing scientific publications, authors usually indicate that the results do not coincide. The reason for this, as we see, may be in a different way of sampling, in particular, differentiation of the clinical forms of the disease. As for the polymorphism 1298*A/C*, for a definitive answer to its role in the disease of multiple sclerosis, it is necessary to conduct research on a well-organized sample.

References

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)

Ehyakonandeh H., Atramentova L.A. (2020). Prognostic value of markers in associative genetics. *The Journal of V.N. Karazin Kharkiv National University. Series "Biology", 34*, 63–68. https://doi.org/10.26565/2075-5457-2020-34-7

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.
Lunde H., Assmus J., Myhr K.M. et al. (2017). Survival and cause of death in multiple sclerosis: a 60-year longitudinal population study. *Journal of Neurology, Neurosurgery, and Psychiatry, 88*(8), 621–625. https://doi.org/10.1136/jnnp-2016-315238
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.

Молекулярно-генетичні дані з точки зору асоціативної та популяційної генетики

Л.О. Атраментова, Х. Ехьяконандех

У дослідженнях з асоціативної генетики мультифакторних захворювань зазвичай виявляється, що частота рідкісного в популяції алеля в групі пацієнтів вища, ніж у групі здорових людей. Через знижену пристосованість мінорний алель проявляється як хвороба. У таких випадках рідкісний алель функціонує як провокативний, а частий як протективний. В групі пацієнтів кількість гомозигот за частим алелем знижена, кількість гетерозиготних носіїв провокативного алеля збільшена, а частота гомозигот за провокативним алелем суттєво підвищена. Метою статті було проаналізувати незвичний результат щодо SNP 129*8A/C* гену *MTHFR* при розсіяному склерозі, отриманий раніше одним з авторів. Частота алелів у контрольній групі та у хворих на розсіяний склероз не відрізняється, але розподіл генотипів у пацієнтів порівняно із здоровими людьми не відповідає співвідношенню Харді–Вайнберга. Серед пацієнтів кількість гетерозигот збільшена, а кількість кожної з гомозигот зменшена. Збільшення частки гетерозигот можна пояснити наявністю одного провокативного алеля, але велика нестача гомозигот за мінорним алелем є незрозумілою. Для пояснення цього факту проаналізовано склад групи пацієнтів. Хоча пацієнти мають різний вік, група неоднорідна за статтю і за формою розсіяного склерозу, жоден з показників не врахований при аналізі розподілу генотипів. Вік маніфестації захворювання є діагностичною ознакою і може залежати від генотипу. Прояв мультифакторних захворювань залежить від статі. Клінічні форми захворювання зазвичай мають різну генетичну основу. Внаслідок нехтування переліченими умовами формується генетично неоднорідна група, на якій може бути одержано будь-який результат, що важко пояснити. Нестача генотипів *CC* може бути обумовлена підвищеною смертністю, що зменшує ймовірність потрапляння пацієнтів у вибірку. Проведений аналіз ще раз свідчить про необхідність для генетичного дослідження формувати однорідні групи пацієнтів.

Ключові слова: *розсіяний склероз, SNP 1298A/C MTHFR, асоціативна генетика, методологія науки, дизайн дослідження, формування вибірки, природний добір, рівновага Харді–Вайнберга.*

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

Молекулярно-генетические данные с точки зрения ассоциативной и популяционной генетики

Л.А. Атраментова, Х. Ехьяконандех

В исследованиях по ассоциативной генетике мультифакторных заболеваний обычно оказывается, что частота редкого в популяции аллеля в группе пациентов выше, чем в группе здоровых людей. Из-за пониженной приспособленности минорный аллель проявляется как болезнь. В таких случаях редкий аллель функционирует как провокативный, а частый как протективный. В группе пациентов количество гомозигот по частому аллелю снижено, количество гетерозиготных носителей провокативного аллеля увеличено, а частота гомозигот по провокативному аллелю существенно повышена. Целью статьи было проанализировать необычный результат по SNP 1298*A/C* гена *MTHFR* при рассеянном склерозе, полученный ранее одним из авторов. Частота аллелей в контрольной группе и у больных рассеянным склерозом не отличается, но распределение генотипов у пациентов по сравнению со здоровыми людьми не соответствует соотношению Харди-Вайнберга. Среди пациентов количество гетерозигот увеличено, а количество каждой из гомозигот

уменьшено. Увеличение доли гетерозигот можно объяснить наличием одного провокативного аллеля, но большая нехватка гомозигот по минорному аллелю непонятна. Для объяснения этого факта проанализирован состав группы пациентов. Хотя пациенты имеют разный возраст, группа неоднородна по полу и по форме рассеянного склероза, ни один из показателей не учтен при анализе распределения генотипов. Возраст манифестации заболевания является диагностическим признаком и может зависеть от генотипа. Проявление мультифакторных заболеваний зависит от пола. Клинические формы заболевания обычно имеют разную генетическую основу. Вследствие пренебрежения перечисленными условиями формируется генетически неоднородная группа, на которой может быть получен любой результат, который трудно объяснить. Нехватка генотипов *CC* может быть обусловлена повышенной смертностью, что уменьшает вероятность попадания пациентов в выборку. Проведенный анализ еще раз свидетельствует о необходимости для генетического исследования формировать однородные группы пациентов.

Ключевые слова: *рассеянный склероз, SNP 1298A/C MTHFR, ассоциативная генетика, методология науки, дизайн исследования, формирование выборки, естественный отбор, равновесие Харди-Вайнберга.*

Об авторах:
Л.А. 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
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

Подано до редакції / *Received: 16.10.2020*
Прорецензовано / *Revised: 12.11.2020*
Прийнято до друку / *Accepted: 14.05.2021*