

РАЗРАБОТКА «ЧАСОВ СТАРЕНИЯ» НА ОСНОВЕ НОВЫХ АЛЛЕЛЬНЫХ ПОЛИМОРФИЗМОВ

Мироненко Я.В.^{1*}, Примак А.Л.², Карагяур М.Н.³, Арбатский М.С.¹

¹ ФГАОУ ВО РНИМУ им. Н.И.Пирогова Минздрава России, Российский геронтологический научно-клинический центр, Москва, Россия

² Факультет фундаментальной медицины, МГУ им. М.В. Ломоносова, Москва, Россия

³ Институт регенеративной медицины, Медицинский научно-образовательный центр, МГУ им. М.В. Ломоносова, Москва, Россия

Цель исследования — создание «часов старения», учитывающих новые аллельные полиморфизмы в результатах полногеномного секвенирования, для определения биологического возраста.

Для изучения структуры данных были использованы .vcf-файлы 21 образца из исследования, посвященного установлению роли отдельных геномных вариантов факторов морфогенеза головного мозга в развитии когнитивных нарушений. В качестве исходных данных собственного исследования будут использованы результаты полногеномного секвенирования (GWAS) в формате .vcf. Для анализа данных будут применены методы корреляционного анализа и машинного обучения. Предлагается следующий алгоритм обработки данных: препроцессинг сырых данных, сравнение найденных полиморфизмов с известными, установление зависимости между обнаруженными полиморфизмами и известными возраст-ассоциированными заболеваниями, оценка статистической значимости найденных новых полиморфизмов с возрастными показателями (использование обученных моделей), выбор модели, проверка работы «часов».

В процессе разработки алгоритма была изучена структура .vcf-файлов с целью автоматизации поиска новых аллельных полиморфизмов в генах с известным участием в развитии возраст-ассоциированных заболеваний. Анализ результатов полногеномного секвенирования 21 образца позволил идентифицировать несколько геномных вариантов (rs6265, rs4760, rs4758443 и др.), ассоциированных, по данным литературы, с нарушением развития головного мозга и возникновением когнитивных нарушений, а также возраст-ассоциированных заболеваний. Предварительные результаты анализа демонстрируют, что предложенный метод обработки биоинформатических данных потенциально может быть использован для оценки биологического возраста исследуемого образца, однако малый объем выборки не позволяет установить точность предложенного метода анализа. По этой причине планируется провести анализ 5000 дополнительных образцов, в ходе которого будет проведена оптимизация алгоритмов предлагаемых «часов старения» и оценена их точность.

Анализ большей выборки биологических образцов от пациентов с возраст-ассоциированными заболеваниями позволит выявить новые аллельные полиморфизмы, что в дальнейшем повысит точность «часов старения».

Ключевые слова: «часы старения»; биологический возраст; биоинформатика старения; генетика старения; GWAS.

Для цитирования: Мироненко Я.В., Примак А.Л., Карагяур М.Н., Арбатский М.С. Разработка «часов старения» на основе новых аллельных полиморфизмов. *Проблемы геронауки*. 2023; 4: 237–239.

DEVELOPING "AGING CLOCKS" BASED ON NOVEL ALLELIC POLYMORPHISMS

Mironenko I.V.^{1*}, Primak A.L.², Karagyaour V.N.³, Arbatsky M.S.¹

¹ Russian Gerontology Research and Clinical Centre, Pirogov Russian National Research Medical University, Moscow, Russia

² Faculty of Fundamental Medicine, Lomonosov Moscow State University, Moscow, Russia

³ Institute for Regenerative Medicine, Medical Research and Education Center, Lomonosov Moscow State University, Moscow, Russia

The aim of this study is to create an "aging clock" that accounts for novel allelic polymorphisms in full-genome sequencing results to determine biological age.

The .vcf files of 21 samples from a study investigating the role of individual genomic variants of brain morphogenesis factors in the development of cognitive impairment were used to examine the data structure. Full genome sequencing (GWAS) results in .vcf format will be used as raw data from our own study. Correlation analysis and machine learning methods will be applied to analyze the data. The following algorithm for data processing is proposed: preprocessing of raw data, comparison of found polymorphisms with known polymorphisms, establishment of correlation between found polymorphisms and known age-associated diseases, assessment of statistical significance of found new polymorphisms with age-associated parameters (using trained models), model selection, and "clock" validation.

In the process of algorithm development, the structure of .vcf files was studied in order to automate the search for new allelic polymorphisms in genes with known involvement in the development of age-associated diseases. Analysis of full-exome sequencing results from 21 samples identified several genomic variants (rs6265, rs4760, rs4758443, etc.) associated with impaired brain development and the occurrence of cognitive impairment and age-associated diseases according to the literature. Preliminary results of the analysis demonstrate that the proposed bioinformatics data processing method can potentially be used to estimate the biological age of the study sample; however, the small sample size does not allow us to establish the accuracy of the proposed analysis method. For this reason, an analysis of 5000 additional samples is planned to optimize the proposed "aging clock" algorithms and evaluate their accuracy.

Analysis of a larger sample of biological samples from patients with age-associated diseases will make it possible to identify new allelic polymorphisms, which will further improve the accuracy of the "aging clock".

Keywords: "aging clocks"; biological age; bioinformatics of aging; genetics of aging; GWAS.

For citation: Mironenko I.V., Primak A.L., Karagyaur V.N., Arbatsky M.S. Developing "Aging Clocks" Based on Novel Allelic Polymorphisms. *Problems of Geroscience*. 2023; 4: 237–239.

СПИСОК ЛИТЕРАТУРЫ

1. Johnson AA, English BW, Shokhirev MN, Sinclair DA, Cuellar TL. Human age reversal: Fact or fiction? *Aging Cell*. 2022 Aug;21(8):e13664. doi: 10.1111/ace1.13664. Epub 2022 Jul 2. PMID: 35778957; PMCID: PMC9381899.
2. Buckley, M.T., Sun, E.D., George, B.M. et al. Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nat Aging* 3, 121–137 (2023). <https://doi.org/10.1038/s43587-022-00335-4>
3. Higgins-Chen AT, Thrush KL, Wang Y, Minter CJ, Kuo PL, Wang M, Niimi P, Sturm G, Lin J, Moore AZ, Bandinelli S, Vinkers CH, Vermetten E, Rutten BPF, Geuze E, Okhuijsen-Pfeifer C, van der Horst MZ, Schreiter S, Gutwinski S, Luykx JJ, Picard M, Ferrucci L, Crimmins EM, Boks MP, Hägg S, Hu-Seliger TT, Levine ME. A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nat Aging*. 2022 Jul;2(7):644–661. doi: 10.1038/s43587-022-00248-2. Epub 2022 Jul 15. PMID: 36277076; PMCID: PMC9586209.
4. Reed E, Nunez S, Kulp D, Qian J, Reilly MP, Foulkes AS. A guide to genome-wide association analysis and post-analytic interrogation. *Stat Med*. 2015 Dec 10;34(28):3769–92. doi: 10.1002/sim.6605. Epub 2015 Sep 6. PMID: 26343929; PMCID: PMC5019244.
5. Chang M, He L, Cai L. An Overview of Genome-Wide Association Studies. *Methods Mol Biol*. 2018;1754:97–108. doi: 10.1007/978-1-4939-7717-8_6. PMID: 29536439

***Автор, ответственный за переписку:** Мироненко Ярослав Владимирович. E-mail: mironenko_iv@rgnkc.ru