

Систематический анализ генетических вариантов с конфликтующими интерпретациями клинической значимости

Лазарева Т.*, Барбитов Ю.

ФГБНУ «НИИ акушерства, гинекологии и репродуктологии им. Д.О. Отта»,
Санкт-Петербург, Россия

* tatiana.ev.lazareva@gmail.com

Ключевые слова: генетический вариант; интерпретация вариантов; ClinVar

Мотивация и цель: Секвенирование нового поколения (NGS) все чаще используется в биомедицинских исследованиях как мощный инструмент для выявления генетических причин наследственных заболеваний. Несмотря на разработанные рекомендации и предложенные стандартизированные классификации генетических вариантов Американского колледжа медицинской генетики (ACMG), расхождения в клинической интерпретации изменений последовательности ДНК между лабораториями сохраняются [1–3]. В настоящем исследовании на основе общедоступной базы данных ClinVar мы систематически проанализировали свойства генетических вариантов и свойства генов, для которых в базе данных о клинической значимости генетических вариантов зафиксированы расхождения в клинической интерпретации, что в конечном итоге позволит разработать стратегии для повышения согласованности интерпретации данных NGS [4].

Методы и алгоритмы: Для анализа были собраны данные ClinVar за период апрель 2018 – апрель 2024 г. о вариантах в генах, связанных с редкими заболеваниями, описанными в OMIM и представленными в Ensembl BioMart [5]. Под вариантами, для которых описан конфликт интерпретаций клинической значимости (COI), в настоящем исследовании подразумеваются варианты, которые были аннотированы как COI по крайней мере один раз за выбранный период. С помощью гипергеометрического теста были выявлены гены, обогащенные COI вариантами. Проанализированы частоты аллелей (на основе данных gnomAD genomes v2.1 [6]) и эффекты вариантов на продукт гена. Оценена эволюционная консервативность, уровень экспрессии изоформ, длина канонического транскрипта, количество экзонов, вовлеченность в биологические процессы, а также количество и тип наследования ассоциированных заболеваний, связанных с генами, для которых (не) описаны COI варианты.

Результаты: Всего для 4731 гена с описанным в OMIM наследственным заболеванием обнаружено 2 296 245 вариантов в ClinVar, в 78 % которых идентифицирован как минимум один COI вариант. Статистически значимое обогащение COI вариантов наблюдалось для 285 генов. Такие гены более сложно организованы, что следует из большего числа экзонов и большей длины транскриптов ($p\text{-value} < 2.2 \times 10^{-16}$). Интересно, что эти гены имеют и большее число связанных заболеваний ($p\text{-value} < 2.2 \times 10^{-16}$), преимущественно с аутосомно-доминантным

типом наследования ($p\text{-value} = 3.78 \times 10^{-9}$). Выявленные гены обогащены терминами, ассоциированными со строением и функционированием мышечной ткани (например, развитие и сокращение поперечно-полосатой скелетной и сердечной мускулатуры, связывание актина и кальмодулина, функционирование трансмембранных ионных каналов).

Выводы: Полученные наблюдения согласуются с предыдущим исследованием феномена «фенотипической гетерогенности», варибельности тяжести заболевания или даже развитием различных заболеваний у лиц, являющихся носителями патогенных вариантов в одном и том же гене [7]. Данный феномен может объяснять накопление конфликтующих интерпретаций вариантов в генах с несколькими ассоциированными заболеваниями (преимущественно из групп нервно-мышечной и сердечно-сосудистой патологии), с преобладанием аутосомно-доминантного типа наследования. Обнаруженные ассоциации между характеристиками генов и обнаружением вариантов с COI могут быть использованы для создания дополнительных правил аннотации и приоритизации вариантов в этих генах.

Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (проект «Многоцентровая научно-исследовательская коллекция биоресурсов» «Репродуктивное здоровье человека» (контракт № 075-15-2021-1058 от 28 сентября 2021 г.).

Systematic analysis of genetic variants with conflicting interpretations of pathogenicity

Lazareva T.*, Barbitoff Y.

Research Institute of Obstetrics, Gynecology and Reproductology named after D.O. Ott,
St. Petersburg, Russia

* tatiana.ev.lazareva@gmail.com

Key words: genetic variant; variant interpretation; ClinVar

Motivation and Aim: Next-generation sequencing (NGS) has revolutionized biomedical research by enabling the identification of genetic causes of inherited diseases. However, discrepancies persist in the clinical interpretation of DNA variants between laboratories, despite standardized classifications proposed by the American College of Medical Genetics (ACMG) [1–3]. The study aimed to systematically analyze properties of genetic variants and their associated genes for which discordance of variant interpretation have been reported in the ClinVar database [4]. It will allow us to develop strategies for improving the consistency of NGS data interpretation.

Methods and Algorithms: We collected ClinVar data (April 2018 – April 2024) on variants in genes linked to rare diseases from OMIM (Online Mendelian Inheritance in Man) and submitted to Ensembl BioMart [5]. Variants with conflicting interpretations (COI) reported at least once during this period were defined as discordant. We employed the hypergeometric test to identify genes enriched for COI variants. Allele frequencies (gnomAD v2.1 genomes [6]) and variant effects on gene products were analyzed. Additionally, we evaluated evolutionary conservation, isoform expression levels, transcript length, exon number, involvement in biological processes, and the number and type of inheritance of diseases associated with genes harboring (or lacking) COI variants.

Results: Our analysis of 2,296,245 variants from 4,731 OMIM-linked disease genes in ClinVar revealed that 78 % contained at least one COI variant. We identified 285 genes with statistically significant enrichment of COI variants. These genes exhibited increased structural complexity, evidenced by a higher number of exons and longer transcripts ($p\text{-value} < 2.2 \times 10^{-16}$). Notably, these genes were also associated with a higher number of diseases per gene ($p\text{-value} < 2.2 \times 10^{-16}$), predominantly with autosomal dominant inheritance ($p\text{-value} = 3.78 \times 10^{-9}$). Moreover, these genes were enriched for terms related to muscle tissue structure and function.

Conclusion: Our findings align with previous studies on “phenotypic heterogeneity”, where individuals with pathogenic variants in the same gene can exhibit variable disease severity or even develop different diseases [7]. This phenomenon might explain the accumulation of COI variants in genes with multiple associated diseases, particularly those related to neuromuscular and cardiovascular pathologies, often exhibiting autosomal dominant inheritance. The observed associations between gene characteristics and COI variant detection could be utilized for establishing additional rules for annotation and prioritization of variants in these genes.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (project “Multicenter research bioresource collection” “Human Reproductive Health” contract No. 075-15-2021-1058 from September 28, 2021).

Список литературы/References

1. Richards S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-423. doi 10.1038/gim.2015.30
2. Harrison S., Dolinsky J., Knight Johnson A. et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med.* 2017;19:1096-1104. doi 10.1038/gim.2017.14
3. Bland A., Harrington E., Dunn K. et al. Clinically impactful differences in variant interpretation between clinicians and testing laboratories: a single-center experience. *Genet Med.* 2018;20:369-373. doi 10.1038/gim.2017.212
4. Landrum M.J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42(D1):D980-D985. doi 10.1093/nar/gkt1113
5. Kinsella Rh. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011;2011:bar030. doi 10.1093/database/bar030
6. Karczewski K.J., Francioli L.C., Tiao G. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581:434-443. doi 10.1038/s41586-020-2308-7
7. Lazareva T.E. et al. Statistical dissection of the genetic determinants of phenotypic heterogeneity in genes with multiple associated rare diseases. *Genes.* 2023;14(11):2100. doi 10.3390/genes14112100