

Исследование разнообразия и противомикробного потенциала кишечного вирома здоровых людей и пациентов с язвенным колитом

Федорец В.*, Тикунов А., Матвеев А., Тикунова Н.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия

* v.fedorets@g.nsu.ru

Ключевые слова: кишечный виром; язвенный колит; метагеном; эндолизины

Мотивация и цель: Известно, что виром кишечника человека представлен преимущественно бактериофагами, которые играют важную роль в формировании микробных сообществ, регулируя численность своих бактериальных хозяев [1]. В связи с этим все более широкую распространенность приобретают исследования, направленные на сравнение вирома здоровых людей и пациентов с различными хроническими заболеваниями кишечника, такими как язвенный колит (ЯК). Кроме того, в последнее время все острее встает проблема растущей антибиотикорезистентности бактерий, поэтому актуальной также является задача поиска новых противомикробных средств. Такими антибактериальными агентами могут являться эндолизины – ферменты бактериофагов, используемые вирусами для разрушения клеточной стенки бактерии. Целью данной работы было исследование кишечного вирома здоровых людей и пациентов с ЯК, а также поиск в нем генов потенциальных эндолизинов и оценка их разнообразия.

Методы и алгоритмы: Из 17 образцов стула (8 от здоровых добровольцев (ЗД) и 9 от пациентов с ЯК) была отделена обогащенная вирусными частицами фракция, из которой в дальнейшем была выделена и секвенирована тотальная ДНК. Полученные прочтения были отфильтрованы по качеству с помощью Trimmomatic и собраны *de novo* в контиги с помощью SPAdes. Потенциальные геномы бактериофагов и их фрагменты были выявлены с помощью программного обеспечения VIBRANT v.1.2.1 [2]. Статистическую обработку результатов, полученных при сравнении выборок, производили с помощью критерия Манна–Уитни. Также с помощью VIBRANT были определены (с использованием Prodigal) и аннотированы (с использованием баз данных KEGG, Pfam и VOG) открытые рамки считывания, содержащиеся в выявленных потенциальных геномах. На основании полученных аннотаций была собрана база данных, содержащая аминокислотные последовательности потенциальных эндолизинов. Затем для фаговых последовательностей, содержащих гены эндолизинов, осуществляли прогнозирование бактериальных хозяев с помощью интегрированной среды машинного обучения iPHoP [3], которая объединяет различные инструменты, используемые для решения этой задачи: поиск сходства между геномом вируса и геномом хозяина, анализ CRISPR кассет, анализ частот встречаемости k-меров, а также сравнение с базой данных уже изученных бактериофагов с известными хозяевами.

Результаты: В виромах, полученных из образцов от ЗД, медианное количество выявленных фаговых последовательностей составило 71, при этом распределение

по образцам было не равномерным – от 10 до 261. В образцах от пациентов с ЯК медианное количество фаговых последовательностей составило 115 (от 13 до 195), что оказалось достоверно выше, чем для ЗД (p -value < 0.05). Кроме того, количество выявленных полных кольцевых геномов бактериофагов было также достоверно выше в выборке пациентов с ЯК, чем ЗД (медианное значение 0.5 для ЗД, 5 для пациентов с ЯК, p -value < 0.05), что может объясняться более высоким покрытием, полученным в результате сборки контигов. На основании полученных с помощью VIBRANT аннотаций всего в виромах было выявлено 134 последовательности потенциальных эндолизин, 85 из которых – у пациентов с ЯК, 49 – у ЗД. Что касается предположительных бактериальных хозяев потенциальных бактериофагов, полные или частичные последовательности которых содержали гены эндолизин, для выборки ЗД наиболее распространенными оказались представители филума Bacteroidota, класса Bacteroidia, которые составили 54.5 % от общего количества хозяев, определенных с высокой степенью достоверности. В основном это были бактерии родов *Bacteroides*, *Phocaeicola*, *Prevotella*. Вторым распространенным таксоном бактерий оказался класс Clostridia, относящийся к филуму Bacillota (39.5 %), представленный большим количеством родов. Стоит отметить, что представители двух данных классов составляют подавляющую часть здоровой микробиоты кишечника человека. Что касается пациентов с ЯК, наиболее распространенными хозяевами бактериофагов у них оказались также представители классов Bacteroidia и Clostridia (36.4 и 34.5 % соответственно), однако обращает на себя внимание снижение относительной доли первого. При этом у пациентов были также выявлены потенциальные фаги, заражающие бактерии классов Actinomycetia (а именно род *Bifidobacterium*) и Gammaproteobacteria. Что касается структуры выявленных эндолизин, наиболее часто встречающимися каталитическими доменами были *Glyco_hydro_25* и *Glucosaminidase*, которые более характерны для эндолизин фагов грамположительных бактерий, а также *Phage_lysozyme* и *Muramidase*, более характерные для эндолизин, воздействующих на грамотрицательные бактерии. Также часто встречающимися были домены *Amidase_2* и *Peptidase_M23*, выявляющиеся в эндолизинах как грамположительных, так и грамотрицательных бактерий. Среди связывающих доменов наиболее распространенным был *LysM*. Выравнивание с помощью BLASTP показало, что значительная часть выявленных белков имеет сходство менее 50 % с уже охарактеризованными эндолизинами.

Выводы: Сравнительный анализ кишечных виромов показал, что общее количество выявленных потенциальных фаговых последовательностей у пациентов с ЯК достоверно выше, чем у ЗД, что подтверждает предположение о изменении состава вирусных сообществ кишечника при хронических воспалительных заболеваниях. Также у пациентов оказался шире спектр потенциальных бактериальных хозяев, что может быть как причиной, так и следствием нарушений в составе микробиоты у людей с ЯК. Кроме того, в исследуемых виромных данных были обнаружены последовательности 134 потенциальных эндолизин с разнообразной доменной структурой, многие из которых не имеют охарактеризованных гомологов, и их дальнейшее исследование может являться ступенью к разработке новых противомикробных средств.

The gut virome of healthy people and patients with ulcerative colitis: diversity and antimicrobial potential

Fedorets V.*, Tikunov A., Matveev A., Tikunova N.

Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

* v.fedorets@g.nsu.ru

Key words: gut virome; ulcerative colitis; metagenome; endolysins

Motivation and Aim: It is known that human intestinal virome is represented mainly by bacteriophages, which play an important role in the microbiota formation via regulating the number of their bacterial hosts [1]. In this regard, studies aimed at comparing the virome of healthy people and patients with various chronic intestinal diseases, such as ulcerative colitis (UC), are becoming increasingly common. Additionally, the problem of growing bacterial antibiotic resistance has become increasingly acute, so the issue of searching for new antimicrobial agents is also urgent. Such antibacterial agents can be endolysins, enzymes used by phages to destroy the bacterial cell wall. The purpose of this work was to study the intestinal virome of healthy people and patients with UC, as well as to search for potential endolysin genes and assess their diversity.

Methods and Algorithms: Stool samples from 8 healthy volunteers (HV) and 9 patients with UC were used to separate a fraction enriched in viral particles, from which total DNA was subsequently isolated and sequenced. The resulting reads were quality filtered using Trimmomatic and assembled *de novo* into contigs using SPAdes. Potential bacteriophage genomes and their fragments were identified using VIBRANT v.1.2.1 software [2]. Statistical processing of the results was carried out using the Mann–Whitney test. VIBRANT was also used to identify (via Prodigal) and annotate (via KEGG, Pfam and VOG databases) the open reading frames in the candidate genomes. Based on the annotations obtained, a database containing the amino acid sequences of putative endolysins was compiled. Then, for potential phage sequences containing endolysin genes, bacterial hosts were predicted using the integrated machine learning framework iPhoP [3], which combines various tools used to solve this problem: searching for similarities between the viral genome and the host genome, analysis of CRISPR cassettes, analysis of the k-mers occurrence frequency, as well as comparison with a database of already studied bacteriophages with known hosts.

Results: In viromes from HV, the median number of identified phage sequences was 71, while the distribution across samples was uneven – from 10 to 261. In samples from patients with UC, the median number of phage sequences was 115 (from 13 to 195), which turned out to be significantly higher than for HV (p -value < 0.05). In addition, the number of identified complete circular genomes of bacteriophages was also significantly higher in the sample of patients with UC than in HV (median value 0.5 for HV, 5 for patients with UC, p -value < 0.05), which may be explained by a higher coverage after contig assembly. Based on the annotations obtained using VIBRANT, a total of 134 potential endolysins sequences were identified in the viromes, 85 of them were from patients with UC, 49 from HV. As for the putative bacterial hosts of the identified potential bacteriophages, the full or partial sequences of which contained endolysin

genes, for the HV sample, the most common were representatives of the phylum Bacteroidota, class Bacteroidia, which accounted for 54.5 % of the total number of hosts identified with a high degree of confidence. These were mainly bacteria of the genera *Bacteroides*, *Phocaeicola*, and *Prevotella*. The second common taxon of bacteria was the class Clostridia, belonging to the phylum Bacillota (39.5 %), represented by a large number of genera. It is worth noting that representatives of these two classes make up the vast majority of the healthy human intestinal microbiota. As for patients with UC, the most common hosts of bacteriophages were also representatives of the classes Bacteroidia and Clostridia (36.4 and 34.5 %, respectively), however, a decrease in the relative proportion of the former was seen. At the same time, potential phages that infect bacteria of the classes ActinonOMICETIA (namely the genus *Bifidobacterium*) and Gammaproteobacteria were also identified in patients. Regarding the structure of the identified endolysins, the most frequently occurring catalytic domains were *Glyco_hydro_25* and *Glucosaminidase*, which are more common in phages of Gram-positive bacteria, as well as *Phage_lysozyme* and *Muraidase*, more common in endolysins affecting Gram-negative bacteria. Also frequently occurring were the *Amidase_2* and *Peptidase_M23* domains, found in endolysins of both Gram-positive and Gram-negative bacteria. Among the binding domains, *LysM* was the most abundant. Alignment using BLASTP showed that a significant proportion of the identified proteins have less than 50 % similarity to already characterized endolysins.

Conclusion: A comparative analysis of intestinal viromes showed that the total number of identified potential phage sequences in patients with UC was significantly higher than in HV, which confirms the assumption that the composition of intestinal viral communities changes in chronic inflammatory diseases. Patients also had a wider range of potential bacterial hosts, which may be both a cause and a consequence of the microbiota disturbances in people with UC. In addition, the sequences of 134 potential endolysins with diverse domain structures were discovered in the studied virome data, many of which do not have characterized homologues, and their further study may be a step towards the development of new antimicrobial agents.

Funding: The study is supported by State Assignment 0245-2022-0005 (No. 075-03-2022-371/4) of the ICBFM SB RAS.

Список литературы/References

1. Ogilvie L.A., Jones B.V. The human gut virome: A multifaceted majority. *Front Microbiol.* 2015;1(6):152433
2. Kieft K., Zhou Z., Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome.* 2020;8(1):90. doi 10.1186/s40168-020-00867-0
3. Roux S. et al. iPhoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* 2023;21(4):e3002083. doi 10.1371/journal.pbio.3002083