

***De novo* идентификация и таксономический анализ пептидов крови человека, не являющихся продуктами известных генов человека**

**М.С. Осетрова<sup>1</sup>, Г.П. Арапиди<sup>1,2</sup>, Т.М. Савельева<sup>1</sup>, О.М. Иванова<sup>2</sup>, И.О. Бутенко<sup>3</sup>,  
В.М. Говорун<sup>1,2,3</sup>**

<sup>1</sup>Московский физико-технический институт (государственный университет)

<sup>2</sup>Институт биоорганической химии им. академиков М. М. Шемякина и Ю. А. Овчинникова РАН

<sup>3</sup>Федеральный научно-клинический центр физико-химической медицины ФМБА

По современным оценкам, в человеческом организме лишь около 10% клеток являются собственно человеческими – оставшуюся часть составляют клетки симбиотических микроорганизмов, населяющих кишечник, слизистые оболочки и кожу человека. Это позволяет говорить о так называемом «*суперорганизме*», представляющем совокупность множества различных видов, лишь один из которых – человек [1, 2]. Поскольку кровь представляет соединительную ткань организма, которая контактирует с большинством органов и тканей, она содержит в своем составе белки, отражающие состояние организма в целом [3], а также, потенциально, – белки и пептиды симбиотических организмов [4, 5, 6].

Комплексный подход к анализу протеома и пептидома сыворотки крови позволяет выявить функциональные зависимости представленности тех или иных групп белков и пептидов в различных состояниях организма. На сегодняшний день разработан ряд методик для анализа метапротеома человека – совокупности всех белков человека и его микробиоты, многие из которых, однако, обладают принципиальным недостатком: идентификация пептидов микроорганизмов сильно зависит от выбора базы данных белковых последовательностей.

В нашем исследовании мы предложили алгоритм, позволяющий избавиться от этого недостатка. Для этого мы использовали *de novo* подход к идентификации пептидов по масс-спектрометрическим данным, не ограниченный подаваемой на вход базой данных.

В работе использовались 17 образцов плазмы крови: 10 из них принадлежали здоровым мужчинам (4 индивидуальных образца и 6 пулов от 10 доноров) и 6 здоровым женщинам (4 индивидуальных образца и 3 пула от 10 доноров). В ходе экспериментов пептиды из образцов плазмы крови получали по отработанной методике, включающей в себя следующие этапы: фракционирование на катионообменных частицах, десорбцию пептидов с поверхности мажорных белков плазмы при термической обработке, ультрафильтрацию с использованием картриджей с порогом пропускания в 10 кДа и твердофазную экстракцию на обращенно-фазовом сорбенте. Полученные элюаты анализировали на масс-спектрометрах Sciex TripleTOF 5600+. *De novo* идентификацию по полученным масс-спектрам проводили при помощи программного обеспечения PEAKS 7.5 [7].

Для отбора спектров и относящихся к ним пептидов для дальнейшего анализа использовались два фильтра, определяющие значение характеристики ALC (average local confidence) – средней достоверности определения каждой из аминокислот в пептидной последовательности. Отбирались только те спектры, у которых значение ALC для лучшего пептида было не меньше 80%, кроме того, был выставлен общий порог по ALC для всех пептидов в 50%. Затем проводилась фильтрация *de novo* идентификаций по времени выхода. В качестве «правильных» идентификаций для построения калибровочной кривой использовались пептиды человека, идентифицированные по классическому алгоритму с использованием базы данных в этих же образцах.

Для всего списка пептидов, прошедших фильтрацию, затем проводился поиск точных совпадений в базе данных всех известных белков NCBI nr при помощи разработанного нами алгоритма. Для организмов, к белкам которых относились идентифицированные пептиды, затем извлекалась таксономическая информация.

Филогенетический анализ и кластеризация полученных *de novo* идентификаций позволили выявить в образцах крови человека пептиды, относящиеся к симбиотическим микроорганизмам. Более 50%

спектров удалось идентифицировать с точностью до порядка, а более 80% - до определенного класса организмов. На рис.1 представлено распределение микроорганизмов на уровне типов. Наиболее представленными в крови оказались пептиды протеобактерий, актинобактерий и фирмикут. Все эти типы бактерий представлены в составе микробиоты человека.

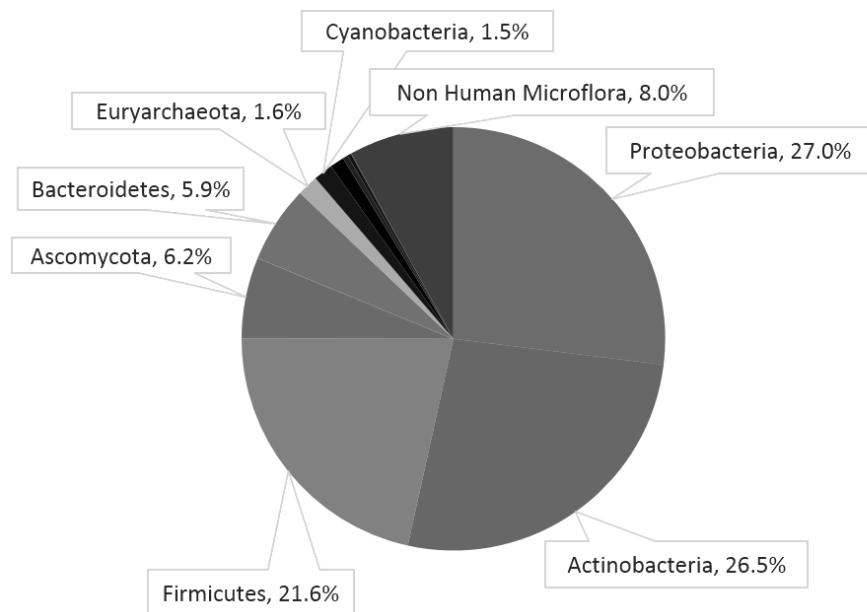


Рис. 1. Распределение микробиотических организмов, к которым относятся идентифицированные в плазме пептиды, на уровне типов.

Хотя тип фирмикуты является широко распространенным (его представителей находят в почве, горных породах, на дне Марианской впадины), среди идентифицированных в нашей работе пептидов ни один не относился к почвенным микроорганизмам, а большая часть идентификаций пришлась на два вида бактерий кишечной микрофлоры.

Таким образом, использование *de novo* подхода к идентификации масс-спектрометрических данных позволило выявить экзогенные пептиды в составе плазмы крови человека. Таксономический анализ показал значительную корреляцию распределения экзогенных пептидов по типам с известным распределением симбиотических микроорганизмов.

#### Литература

1. *Manimozhayan Arumugam et al.* Enterotypes of the human gut microbiome // Nature 2011. № 473 p. 174-180
2. *Tyakht, A.V. et al.* Human gut microbiota community structures in urban and rural populations in Russia // Nat. Commun. 2013. №4, article number: 2469
3. *Anderson NL, Anderson NG.* The human plasma proteome: history, character, and diagnostic prospects // Mol. Cell. Proteomics 2002. №1, p. 845-867.
4. *Rodland KD.* Proteomics and cancer diagnosis: the potential of mass spectrometry // Clin. Biochem 2004. №37, p. 579-583.
5. *Schulz-Knappe P. et al.* The peptidomics concept // Comb. Chem. & High Throughput Screen. 2005. №8, p. 697-704.
6. *Villanueva J. et al.* Differential exoprotease activities confer tumor-specific serum peptidome patterns // J. Clin. Invest. 2006. №116, p. 271-284.
7. *Ma B. et al.* PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry // Rapid Commun. Mass Spectrom. 2003. №17, p. 2337-2342.