

# БИОИНФОРМАТИЧЕСКИЙ ПАЙПЛАЙН ДЛЯ ОЦЕНКИ ИММУНОГЕННОСТИ БЕЛКОВ ПАТОГЕНОВ И РАКОВЫХ ОПУХОЛЕЙ

И.С. Трусов, Е.В. Гузова, В.В. Гринев

Белорусский государственный университет, Минск, Беларусь  
E-mail: ivan.trusau@yandex.by

Данная исследовательская работа посвящена разработке биоинформационного пайплайна для оценки иммуногенности белков. В работе описывается пошаговая сборка пайплайна, а также проводится его сравнение со схожими программными решениями. Разработанный пайплайн превосходит аналоги по точности и F1-score и позволяет в автоматическом режиме генерировать список пептидов целевого белка, ранжированных по иммуногенности.

**Ключевые слова:** иммуногенность белков; прогнозирование; биоинформатика.

Ежегодно всё больше достижений отмечается в области персонализированной медицины, в частности, благодаря значительному прогрессу в создании персонализированных вакцин, в основе которых лежат генные и дендритно-клеточные технологии [1]. Однако темпы разработки новых вакцин остаются низкими, что является главным препятствием к их массовому и/или персонализированному применению. Одна из причин – проблемы с выявлением иммуногенных пептидов (эпитопов, специфических последовательностей в белках, способных вызывать иммунный ответ). В таком контексте появление биоинформатического инструмента, способного с высокой точностью и чувствительностью предсказывать иммуногенные пептиды, да к тому же ещё и учитывающего различные аспекты процессинга белка, могло бы стать краеугольным камнем при создании вакцин нового поколения.

## МАТЕРИАЛЫ И МЕТОДЫ

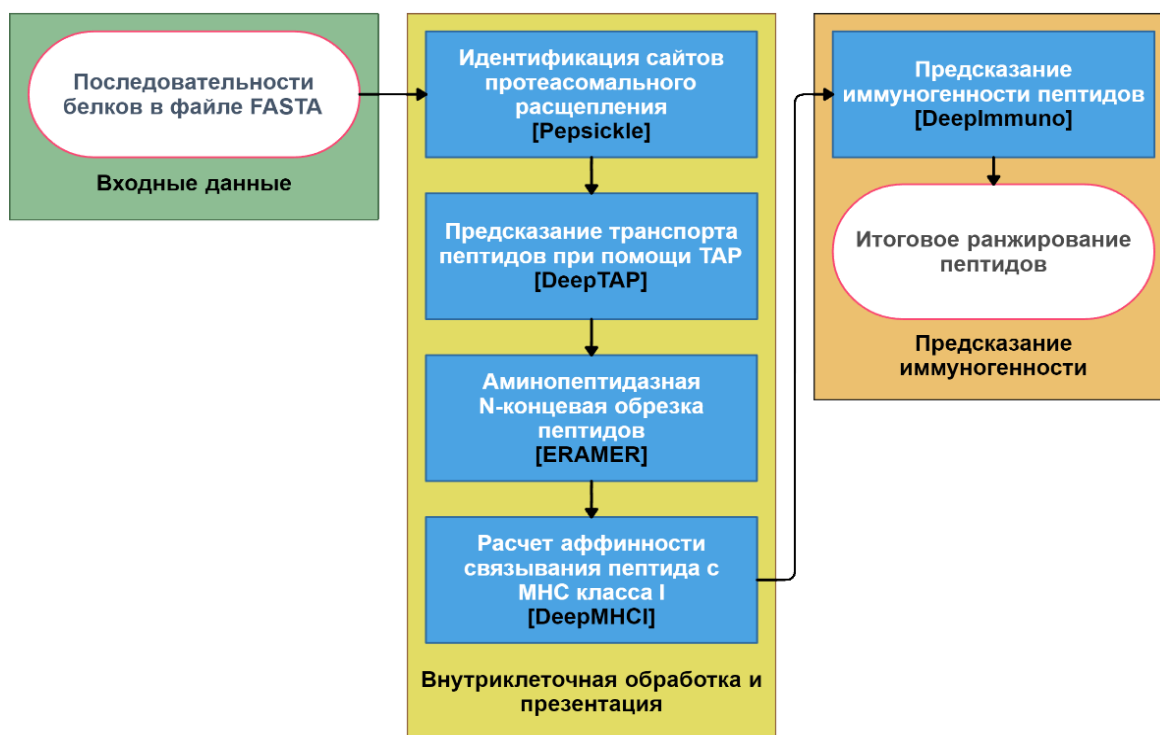
Для проектирования пайплайна были отобраны опубликованные независимые программные пакеты с открытым исходным кодом и наилучшими значениями *AUC* при сравнении с аналогами. Все пакеты были написаны на языке *Python* с применением библиотек *tensorflow*, *pytorch*, *numpy*, *pandas* и *scipy*. Для тестирования пакетов были взяты данные по иммуногенности вирусных пептидов, доступные в базе *IEDB* (англ. immune epitope database) [3]. В тестах учитывались только пептиды, которые имели длину в 9-10 нативных аминокислотных остатков. Иммуногенные свойства таких пептидов были подтверждены экспериментально.

Выбор аллелей главного комплекса гистосовместимости (*MHC*, англ. major histocompatibility complex) соответствовал их использованию в боль-

шинстве исследований. Бенчмаркинг проводился в сравнении с широко известными программными решениями, находящимися в открытом доступе: *NetMHC 4.0*, *NetMHCpan 4.0*, *MHCflurry 2.0* и *MixMHCpred 2.2* [2]. Визуализация результатов осуществлялась с помощью библиотек *Python Matplotlib* и *Seaborn*. Расчёты проводились на компьютере, оснащённом процессором с 8 ядрами и тактовой частотой 3,2 ГГц каждое и 16 Гб оперативной памяти, а также *SSD*-диск.

## СТРУКТУРА И ОРАГНИЗАЦИЯ ПАЙПЛАЙНА

В нашем исследовании мы сосредоточились на внутриклеточном процессинге белков с участием только молекул МНС класса I. Такой процессинг включает в себя пять последовательных этапов: 1) расщепление белков протеасомами, 2) транспортировка полученных пептидов в эндоплазматическую сеть при помощи ТАР (англ. transporter associated with antigen processing), 3) дополнительное расщепление пептидов пептидазами, 4) связывание пептидов молекулами МНС, 5) распознавание комплекса пептида с МНС иммунными клетками. Для анализа этих этапов были выбраны пакеты *pepsickle*, *DeepTAP*, *ERAMER*, *DeepMHC1* и *DeepImmuno*, соответственно (см. рисунок).



Графическое изображение структуры и организации пайплайна

Работа пайплайна начинается с *pepsickle*, требующего последовательность белка в формате FASTA. *Pepsickle* возвращает координаты и вероятности разрезания белка протеасомами. Все возможные варианты получаемых

пептидов длиной 9-16 аминокислот передаются DeepTAP, который оценивает вероятность транспорта пептидов в эндоплазматическую сеть. Далее пептиды с длиной более 10 аминокислот дополнительно анализируются пакетом ERAMER, который определяет шансы пептида быть дополнительно расщеплённым ферментом ERAPI. Полученные пептиды длиной 9-10 аминокислот подвергаются анализу с помощью DeepMHC1, который предсказывает вероятность образования комплекса пептида с МНС. Отобранные комплексы затем анализируются с помощью DeepImmuno для ранжирования пептидов по иммуногенности.

В качестве порогового значения вероятностей при отборе пептидов использовались значения 0,5, 0,22, 0,1, 0,44 и 0,88 для pepstickle, DeepTAP, ERAMER, DeepMHC1 и DeepImmuno, соответственно. Конечным результатом работы пайплайна является файл формата CSV с отобранными пептидами и краткой информацией (аннотацией) по каждому из этапов обработки.

### ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПАЙПЛАЙНА

Для оценки эффективности вычислительных алгоритмов в выявлении иммуногенных пептидов было проведено сравнение разработанного нами пайплайна с известными программными решениями. Набор для тестирования включал 63 пептида вирусной природы из 12 исследований, при этом 40 пептидов были не иммуногенны. Каждый из пептидов тестировался в комбинации с 7 разными аллелями МНС. Сравнение проводилось по следующим метрикам: специфичность, чувствительность, метрика *F1* и *MCC* (англ. *Matthews correlation coefficient*).

Оценка пептидов программами *NetMHC*, *NetMHCpan* и *MHCflurry* проводилась путем расчета аффинности связывания пептидов с молекулами МНС по показателю *IC50* (концентрация полумаксимального ингибирования), выраженного в нмоль/л. При этом пептиды считались иммуногенными при стандартном пороговом значении *IC50* в 500 нмоль/л или менее. В программе *MixMHCpred* используется ранговая система оценки пептидов с пороговым значением 0,5. Данное значение было подобрано эмпирически – при нём демонстрировалась наилучшая точность предсказаний. Для нашего пайплайна пептиды считались иммуногенными, если они прошли все этапы и были представлены в конечном файле *CSV*.

Сравнительный анализ показал, что разработанный нами пайплайн демонстрирует лучшие результаты по сравнению с аналогами по показателям специфичности, метрике *F1* и *MCC* (см. таблицу). Очевидно, что наш пайплайн уступает аналогам только по такому показателю, как чувствительность: отношению количества обнаруженных программой событий к их истинному количеству. Однако этот показатель связан со специ-

фичностью, и снижение специфичности может повышать чувствительность. А поскольку для разработки эффективных вакцин более важным является показатель специфичности (для успешной вакцины достаточно, как правило, нескольких высокоиммуногенных пептидов), то мы не считаем более низкую чувствительность нашего пайплайна его существенным недостатком.

#### Статистические результаты тестирования программ

Метрика	<i>NetMHC</i>	<i>NetMHCpan</i>	<i>MHCflurry</i>	<i>MixMHCpred</i>	пайплайн
Специфичность	0,43	0,47	0,39	0,52	<b>0,84</b>
Чувствительность	0,83	0,91	<b>1,00</b>	0,70	0,70
<i>F1</i>	0,57	0,62	0,56	0,59	<b>0,76</b>
<i>MCC</i>	0,21	0,33	0,20	0,31	<b>0,65</b>

#### ЗАКЛЮЧЕНИЕ

Результатом нашего исследования стал биоинформатический пайплайн, позволяющий идентифицировать иммуногенные пептиды в изучаемом белке. Он основан на предсказательных моделях и программных пакетах, объединённых с помощью *Python*. Модульная архитектура пайплайна облегчает его настройку и интеграцию дополнительных функций, что делает пайплайн хорошей базой для дальнейших исследований. В тестах по оценке иммуногенности вирусных белков наш пайплайн продемонстрировал высокую специфичность, а также оказался лучше своих аналогов по метрикам *F1* и *MCC*. Полученные результаты указывают на перспективность дальнейшего развития данного подхода.

#### БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. IEDB and CEDAR: two sibling databases to serve the global scientific community. / Blazeska N., Kosaloglu Z.Y., Vita R. [et. al.] // *Computational Vaccine Design*. 2023. Vol. 2673. P. 133-149. DOI: 10.1007/978-1-0716-3239-0\_9
2. Трусов И.С. Биоинформатическое прогнозирование иммуногенности генно-инженерных белков для активной иммунотерапии // Матер. 80 науч. конф. студ. и аспирантов.: БГУ, 2023. С. 132-136.
3. Personalized dendritic cell vaccines – recent breakthroughs and encouraging clinical results. / Mastelic B.G., Balint K., Boudousquie C. [et. al.] // *Cancer Immunity and Immunotherapy*. 2019. Vol. 10. P. 1-6. DOI: 10.3389/fimmu.2019.00766