

## БИОИНФОРМАТИЧЕСКИЙ АНАЛИЗ РАСПРОСТРАНЕНИЯ МУТАЦИЙ SARS-COV-2 В РОССИИ

А.И. Чангалиди (Университет ИТМО),

А.А. Капитонова (Московский государственный университет им. М.В. Ломоносова),

Е.Е. Петренко (Санкт-Петербургский государственный университет),

Ф.М. Поляков (Высшая Школа Экономики)

Научный руководитель – в.н.с, к.б.н А.С. Комиссаров (Университет ИТМО)

В ходе работы мы проанализировали полногеномные сборки вируса SARS-CoV-2. Сравнили распространение клад во времени для России и ее крупных городов, нашли и проанализировали уникальные мутации для популяций отдельных стран, а также проинтерпретировали результаты.

**Введение.** Ни одно событие в 2020 году не может по своей значимости превзойти пандемию COVID-19, вызванную новым коронавирусом SARS-CoV-2. Наряду с кардинальными изменениями, которые она принесла в экономику, здравоохранение и повседневную жизнь общества, в мировой науке десятки научных групп включились в работу по изучению нового вируса в целях борьбы с заболеванием. Исследование генов и целых геномов SARS-CoV-2 необходимо для понимания динамики передачи вирусов и их эпидемиологии, а также отслеживания мутаций, которые могут потенциально влиять на трансмиссивность, патогенность и/или меры диагностики и лечения (ПЦР-тесты и вакцины). Полученные в ходе таких исследований результаты могут иметь ключевое значение в сфере здравоохранения для принятия решений по борьбе со вспышками заболеваний.

### Основная часть.

**Данные.** В ходе исследования было проанализировано 455149 полногеномныхборок вируса SARS-CoV-2 из образцов человеческого происхождения (источник: база данных EpiCov, которая ежедневно пополняется учеными со всего мира и является главным ресурсом для накопления и свободного распространения геномных данных нового коронавируса). Кроме самих геномов, каждая запись содержит метаинформацию (место и дата сбора образца, технология секвенирования, пол и возраст пациента и др.), которая была использована для дальнейшего анализа.

**Методы.** Каждая сборка была индивидуально выровнена на референсный геном SARS-CoV-2 (NC\_045512.2 NCBI) инструментом minimap2 v2.17, после чего полученные файлы с вариантами были объединены утилитой Picard v2.24.2. Полученный массив мутаций был обработан с помощью скриптов на Python собственной разработки: были отфильтрованы редкие варианты, которые могли быть результатом ошибок секвенирования. Оставшиеся 40816 мутации были проаннотированы с использованием SnpEff v4.3.

**Результаты.** В ходе исследования распространения клад во времени, было выявлено почти полное преобладание клады GR в России (хотя пандемия началась с других клад, которые позже были вытеснены GR). Это может являться следствием того, что в Россию вирус попал не сразу и был занесен из Европы (как раз там начала активно появляться кладка GR). Похожая картина распространения характерна и для двух крупнейших городов России - Москвы и Санкт-Петербурга. Скачки других клад в этих городах обусловлены, скорее, небольшим количеством данных, чем другими факторами. Преобладание вирусов кладки GR связано с наличием в их геномах новых мутаций в Spike-белке (в RBD), дающих эволюционное преимущество перед другими кладами.

Также в российских сборках были обнаружены две мутации бразильского штамма (B.1.1.248): C11824T (белок nsp6), G23012A (белок Spike, позиция E484K) и одна мутация британского штамма (B.1.1.7): C23604A (P681H). Эти штаммы примечательны большей

трансмиссивностью и меньшей эффективностью связывания по крайней мере с одной формой антител. Это в свою очередь может усложнить разработку вакцины. Образцы с мутацией E484K были обнаружены в европейской части России (Москва, Санкт-Петербург, Саратов).

Наконец, было изучено количество уникальных мутаций в разных странах (учтены мутации с встречаемостью не меньше десяти). Самое большое количество уникальных мутаций было найдено в Великобритании и США (9404 и 6989), но надо учитывать, что там было произведено больше всего секвенированных геномов (195088 и 90456 соответственно). В Китае, изолировавшемся с самого начала пандемии, было найдено всего 4 уникальных мутации (на 913 сборок).

### **Выводы.**

Результаты нашего исследования показывают важные закономерности распространения вируса SARS-CoV-2 в России и мире, и могут быть использованы для лучшего понимания его действия и необходимых мерах противодействия ему.