

Федеральное государственное автономное образовательное учреждение
высшего образования

«Национальный исследовательский университет

«Высшая школа экономики»

Факультет компьютерных наук

**Отчёт о прохождении производственной (научно-
исследовательской) практики**

Выполнил студент:	Горохов Никита Сергеевич
Образовательной программы:	Анализ данных в биологии и медицине

Организация:
НИУ ВШЭ

**Отчет проверил
руководитель
практики от
организации:** **Пятницкий Михаил Алексеевич**

**Отчет проверил
руководитель
практики от НИУ
ВШЭ:** **Спирин Сергей Александрович**

Подпись студента:



Москва, 2019

Введение

Персонализированная терапия представляет собой использование данных генетического анализа конкретного больного для выбора наиболее подходящей для заболевания терапии. Другими словами назначение правильного препарата, правильному больному, в нужное время и в нужном месте. Для определения того, какой препарат необходим (или комбинация препаратов) онкологическому больному, необходимо оценить активность сигнальных путей в клетке опухоли, а также выявить клинически значимые генетические варианты. Для выявления всего этого используются технологии NGS. При развитии опухоли часто происходят мутации - нарушения в структуре определенных генов, кодирующих молекулы, которые влияют на восприимчивость опухолевых клеток к действию препаратов, скорость роста опухоли и метастазирование.

Целью моей научно-исследовательской практики является погружение в область персонализированной медицины, детальное знакомство с данными результатов секвенирования и базой данных клинически значимых генетических вариантов. И в качестве закрепления новых знаний создание вспомогательного инструмента для контроля качества результатов секвенирования опухолевой ткани.

Основная часть

Формат BAM представляет собой бинарный формат для хранения биологических последовательностей, выровненных по эталонной последовательности (референсной). Этот формат широко используется для хранения таких данных, как фрагменты нуклеотидных последовательностей (риды), полученные с помощью технологии секвенирования нового поколения.

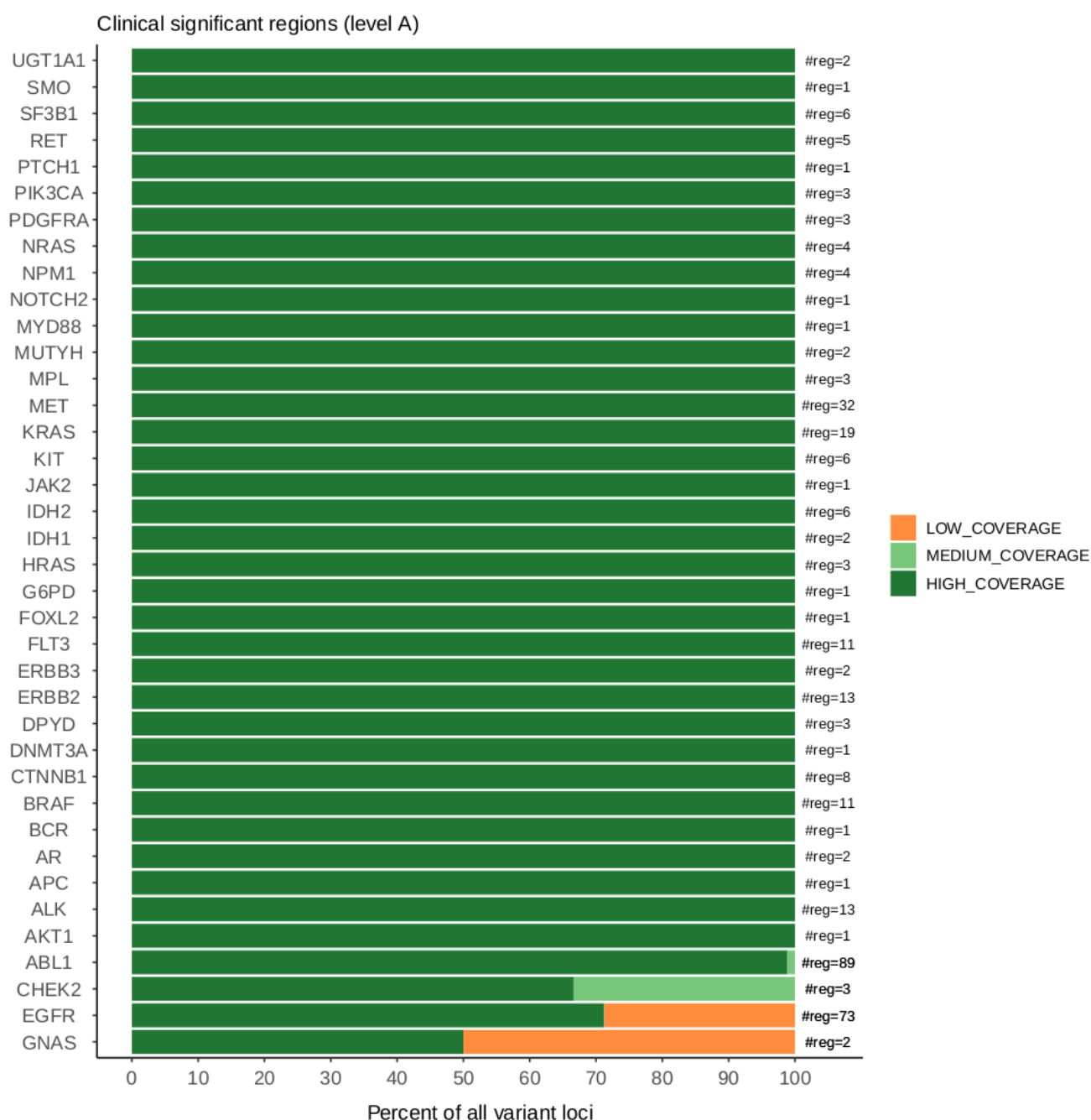
Опухолевые клетки из одного органа могут очень сильно отличаться структурой ДНК. Из-за этого для онкологических данных важную роль играет «покрытие» участка. Чем больше покрытие, тем надежнее определена конкретная мутация.

Программа mosdepth позволяет для выбранного BAM файла и конкретной панели секвенирования подсчитать покрытие. Данная утилита работает быстрее, чем samtools.

База данных клинически значимых вариантов представляет собой список мутаций, которые каким-либо образом связаны с канцерогенезом. То есть наличие мутаций в этих вариантах является ключевым пунктом в определении терапии для больного. Существует несколько известных общедоступных таких баз: CIVIC, JAX, CGI и т. д. Также в этих базах есть связь клинически значимых вариантов с препаратами с определенным уровнем значимости. Это значит, что проводилось исследование, в котром тестировался препарат (или комбинация препаратов) для группы пациентов с наличием общих мутации в клинически значимых вариантах. Такая связь между вариантом, препаратом и исследованием (статьей) называется эвиденсом.

Передо мной была поставлена следующая задача: разработать программное обеспечение, которое бы наглядно (визуально) представляло информацию об уровне покрытия клинически значимых вариантов.

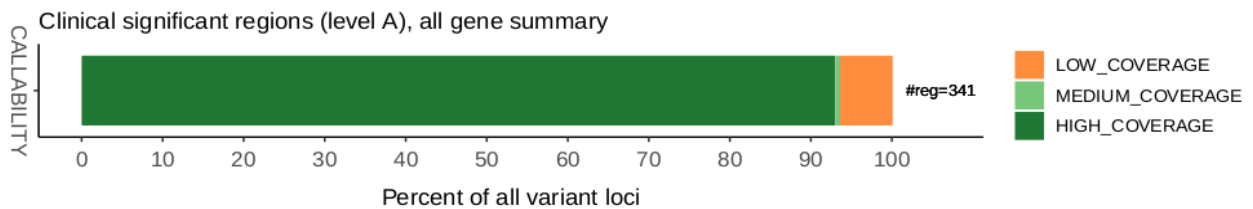
На рисунках ниже можно увидеть результат работы программы для эвидесев уровня А (А: Одобрение FDA, рекомендации NCCN, ASCO, ESMO, МинЗдрава, RUSSCO).



По оси ординат отложены гены, справа написано количество клинически значимых вариантов, найденных в BAM файле и найденных в нашей базе данных эвиденсов (#reg=2). Цвет означает уровень покрытий данного участка (LOW: <50, MEDIUM: >50 и <200, HIGH: >200).

Например, в гене EGFR найдено 73 клинически значимых генетических варианта, которые имеются в нашей базе данных эвиденсов. Около 70% этих вариантов имеют хорошее покрытие (>200) и около 30% имеют среднее покрытие (>50, <200).

На картинке ниже приведена суммарная информация по всем гена для уровня значимости A.



Заключение

Во время прохождения научно-исследовательской практики я узнал много нового из области персонализированной медицины, преобрел опыт работы с результатами секвенирования и баз данных клинически значимых генетических вариантов.

Я считаю, что данная практика прошла очень продуктивно. Скорее всего свою будущую деятельность я свяжу именно с медициной. И полученный опыт поможет мне в этом.

Список литературы

Mosdepth: <https://github.com/brentp/mosdepth>

CIVIC: <https://civicdb.org/>

JAX: <https://www.jax.org>

