Разработка методов оценки качества данных высокопроизводительного секвенирования

Максим Иванов | Атлас ОнкоДиагностика

Описание проекта

Существующие подходы к оценке качества данных высокопроизводительного секвенирования схожи с правилами Вестгарда для контроля качества аналитических тестов – для принятия решения о валидности или не-валидности данных для дальнейшего анализа последовательно исследуются несколько параметров, которые должны удовлетворять заданным критериям. Однако ни список параметров ни налагаемые на них критерии не являются унифицированными и стандартизированными. Таким образом, в различных лабораториях “удовлетворительные” для анализа данные означают разные понятия. Более того, текущий процесс контроля качества никак не определяет уровень качества данных и то, насколько отклонение от золотого стандарта может повлиять на конечный результат.

Учитывая комплексность как самого процесса, так и данных высокопроизводительного секвенирования, аналитические характеристики (чувствительность и специфичность детектирования мутаций) варьируются не только от самой платформы, инструмента и набора реагентов, но и от используемых методов обогащения по целевой ДНК, методов анализа и даже типов самих мутаций – к примеру, делеции детектируются с более низкой специфичностью, чем точечные нуклеотидные замены. Другими словами, аналитические характеристики могут меняться как в зависимости от лаборатории и непосредственного запуска, так и для каждого конкретного образца в рамках одного запуска.

В рамках проекта предлагается разработка инструментария для оценки качества данных высокопроизводительного секвенирования с ограниченной сферой приложения – целевое секвенирование небольших регионов ДНК с заранее известным спектром интересующих мутаций (к примеру, секвенирование генов EGFR/KRAS или BRCA1/2, CFTR), что позволяет симулировать оценку аналитических характеристик конкретного запуска.

Задачи в проекте

Разработка методов и инструментария для оценки качества данных высокопроизводительного секвенирования в клиническом приложении.

Требования/Пожелания к команде

- Linux/bash
- Программирование на среднем/высоком уровне (python/C++/java)
- Желательно знание принципов анализа данных NGS с целью поиска мутаций (препроцессинг/выравнивание/вариант коллинг/аннотация)

Желательное количество участников

3-4