

Вам предоставлен доступ к файлу с дизайном таргетной панели. Формат файла bed detail. В файле указаны координаты регионов для таргетного обогащения методом высоко-мультиплексной ПЦР (указаны координаты фрагментов без праймеров).

Задание №1

Установите, какие регионы генома (гены, экзоны) включены в данную таргетную панель. Добавьте в файл идентификатор (название гена и номер экзона) для каждого региона.

Предположите для диагностики каких заболеваний/групп заболеваний может быть использована данная панель.

Задание №2

Выравнивание коротких прочтений, получаемых в результате высокопроизводительного секвенирования, осложняется в случае наличия в геноме регионов с высокой степенью гомологии, например, псевдогенов. 100% гомология между исследуемым и нецелевым регионом, приводит к невозможности однозначно картировать прочтение, и, как следствие, установить генотип. Определите, есть ли данной таргетной панели нецелевые регионы генома со 100% гомологией? В случае наличия таких регионов, укажите координаты целевых и гомологичных нецелевых регионов.

При выполнении заданий предложите универсальный (масштабируемый) алгоритм выполнения. Укажите использованные инструменты и команды запуска.