

Практикум 6 - Секвенирование по Сенгеру

Демидов Иван

Мне достался файл [13_F.ab1](#). Анализ проводился с помощью UGENE.

1. Основная информация

Длина хроматограммы - 717 нуклеотидов. Координаты **начального трудно читаемого** фрагмента я “на глаз” оценил как 1-94 (то есть **его длина - 94 нуклеотидов**). В целом, фрагмент 95-135 тоже явно хуже, чем в среднем по хроматограмме, а на участке 123-136 сильные проблемы, но он также явно лучше, чем участок 1-94, поэтому я решил не относить его к начальному трудно читаемому фрагменту (хотя возможно стоило).

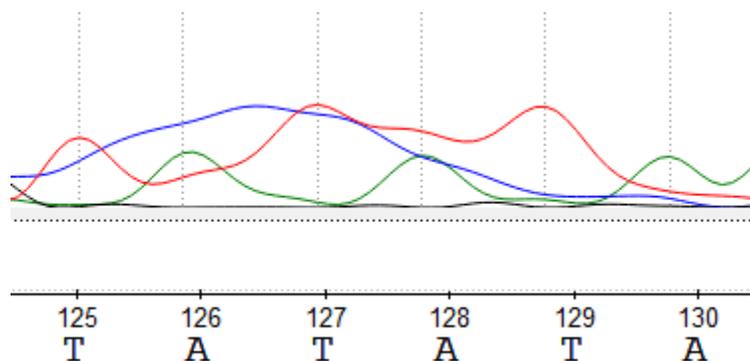


Рис.1. Участок 125-130

Мне кажется, **конечный трудно читаемый** фрагмент начинается примерно с 675 нуклеотида (и до конца хроматограммы, то есть **его длина - 43 нуклеотида**). Вообще может показаться, что он не так уж и плох, но в целом хуже, чем в среднем по хроматограмме.

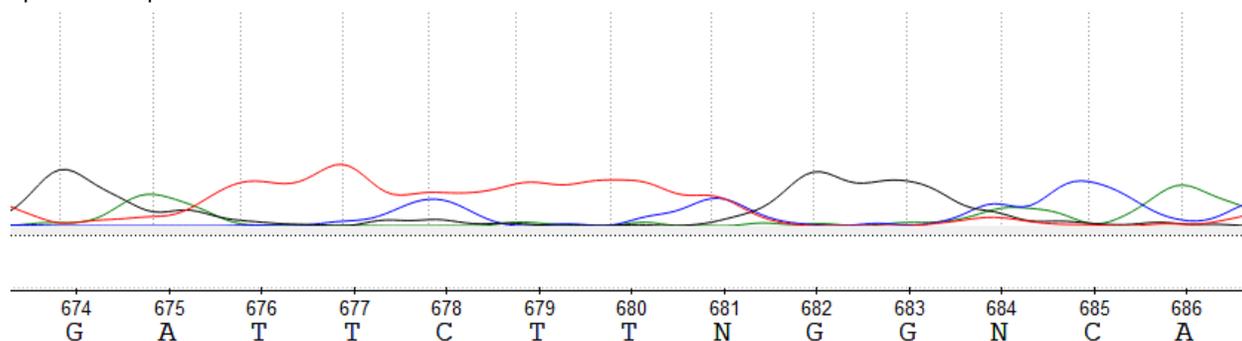


Рис.2. Участок 674-686

2. “ШУМ”

Не считая начального и конечного трудно читаемых фрагментов, хроматограмма характеризуется довольно низким уровнем шума, особенно в “центральной” части хроматограммы.

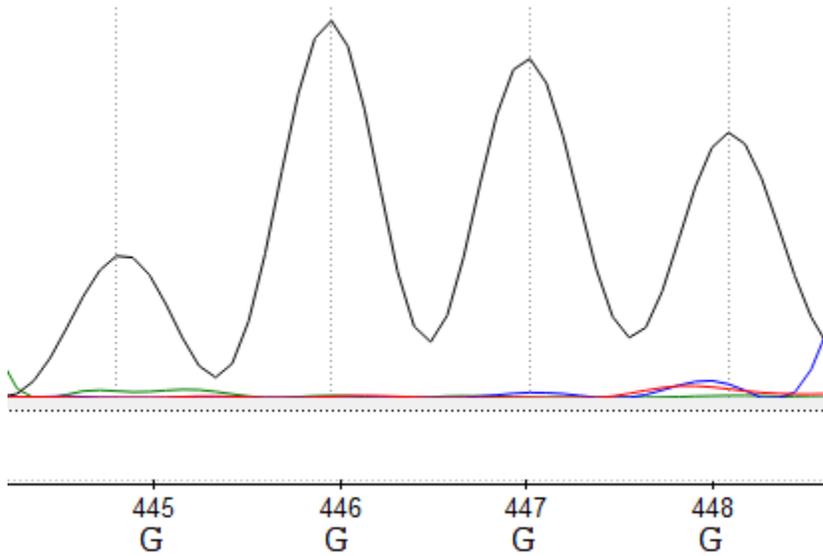


Рис. 3. Участок 445-448 - “шума” почти нет

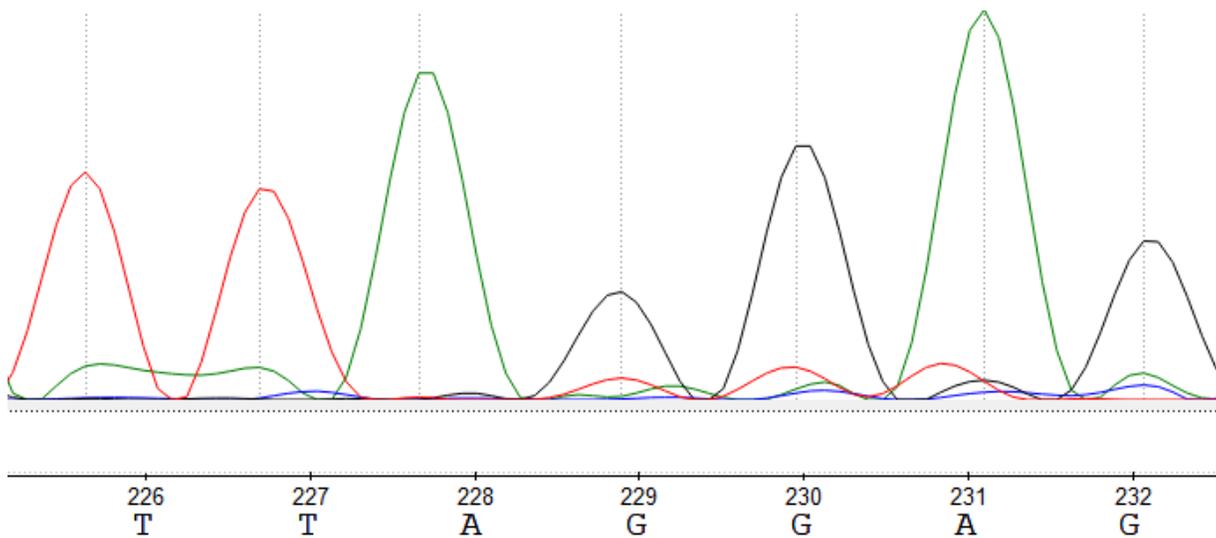


Рис. 4. Участок 445-448 - “шум” есть, но не мешает интерпретации сигнала

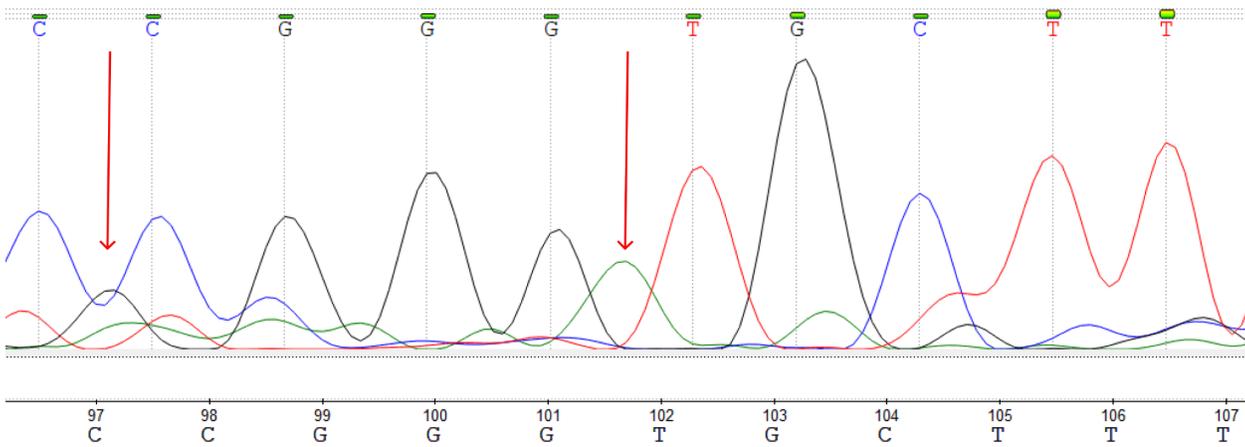


Рис. 4. Участок 97-107 - “шум” мешает интерпретации сигнала - например, мне кажется, нельзя сделать однозначный вывод об отсутствии или наличии нуклеотида между 97-С и 98-С и между 101-Г и 102-Т, так как эти пики не сильно отличаются от “шума” по высоте (я не знаю почему шкала снизу съехала, извините)

3. Проблемные позиции

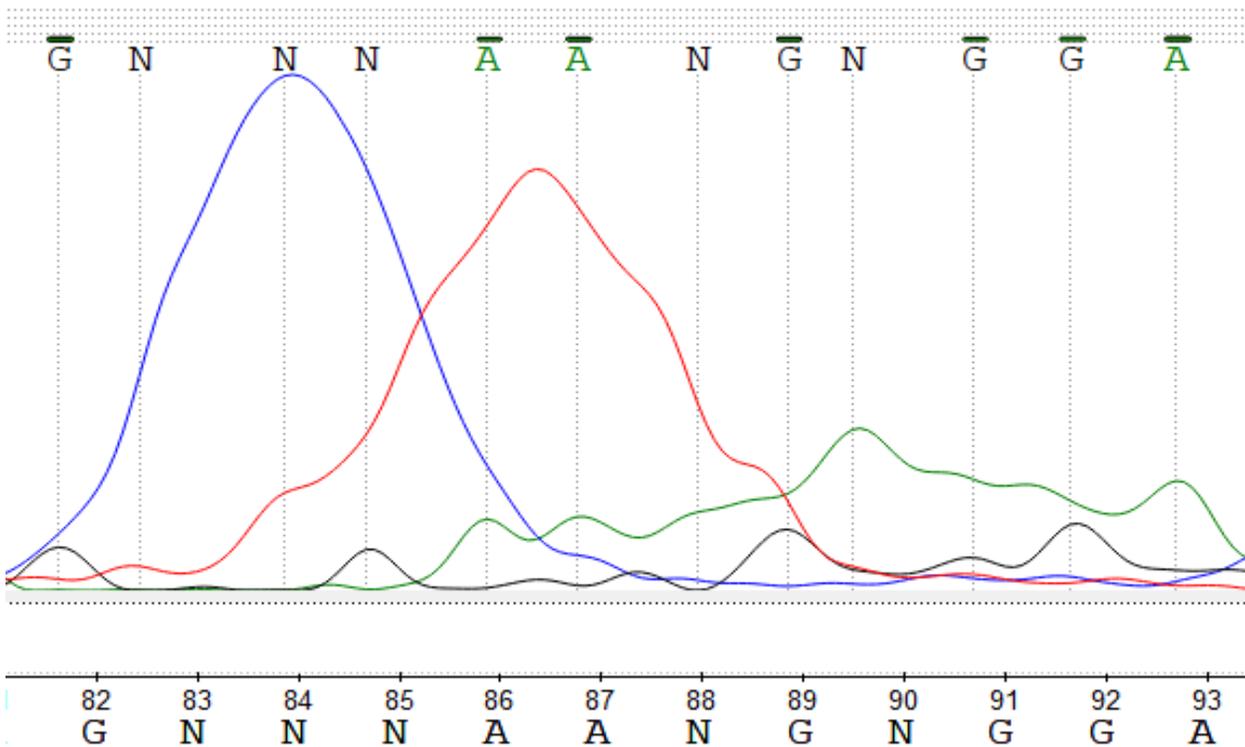


Рис. 5. Участок 82-89 (82-93 - не уверен насчет зеленого пятна) - Пятно краски - нигде в хроматограмме больше нет сигналов такой интенсивности. Вердикт: фрагмент не подлежит интерпретации.

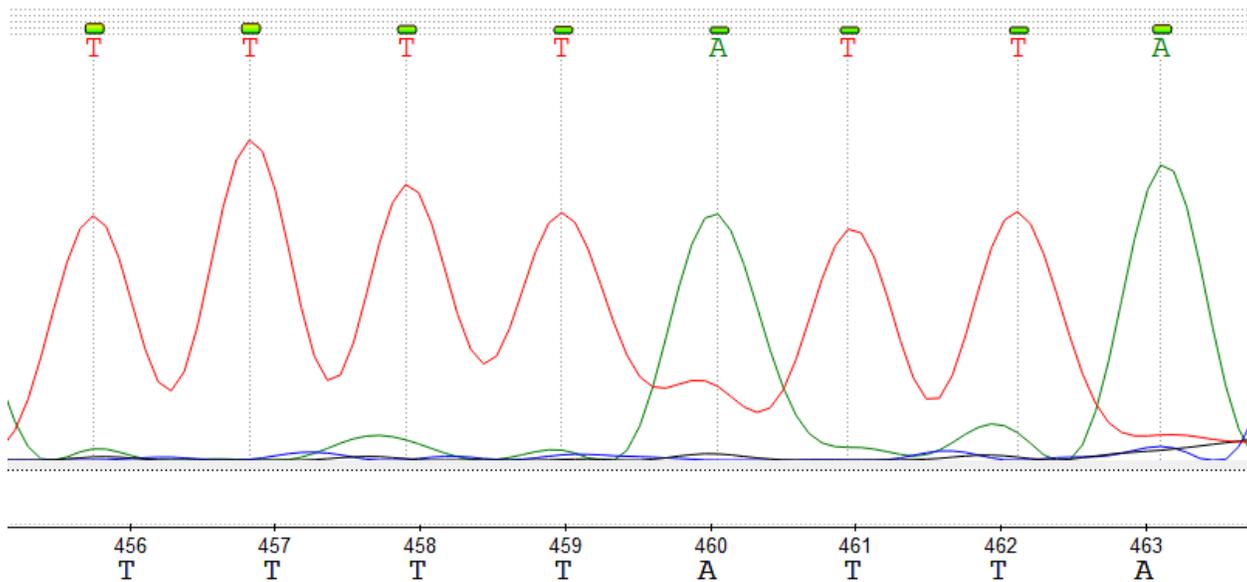


Рис. 6. Участок 456-463. Интерес представляет позиция 460. В целом для фрагмента характерно малое количество “шума”, поэтому дополнительный красный пик может настораживать и наводить на мысли о полиморфизме. Но мне кажется, причиной его возникновения может быть то, что позиция “окружена” тиминами, и поэтому тиминный сигнал не успевает затухать перед тем, как снова появится (я не знаю как это нормально описать, но мне не верится в полиморфизм данной позиции). Вердикт: не можем утверждать о полиморфизме.

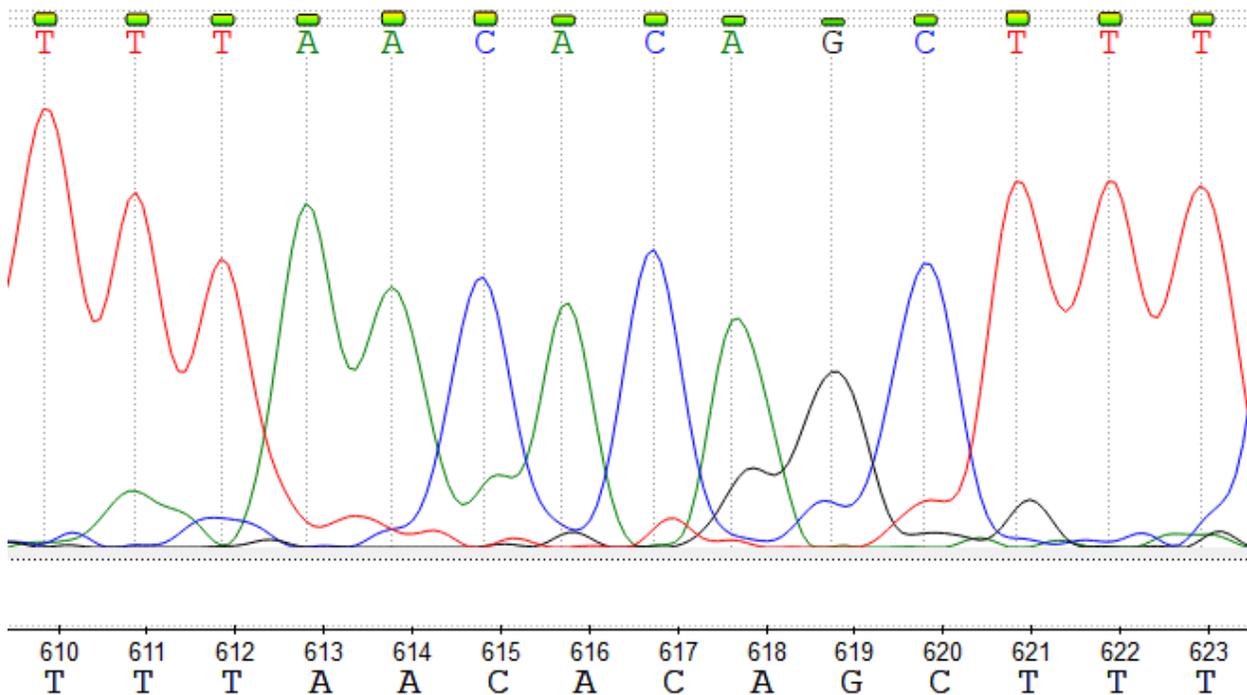


Рис. 7. Участок 610-623. Данный участок изобилует проблемными позициями. Возможно все они объясняются высоким уровнем “шума”, но пусть будет. Например, позиция 615 мне кажется аналогичной той, что я описывал на прошлом рисунке. Но вот позиция 618 мне кажется вполне вероятно может быть полиморфизмом, так как пик вроде сильно

выше, чем “шум”. Но, опять же, из-за сильного “шума” однозначный вывод сделать нельзя. Также небольшие пики в позициях 611, 619, 621 могут намекать на полиморфизмы, но они еще ниже и по ним точно нельзя ничего сказать (скорее всего, просто шум). Вердикт: нельзя сделать однозначного вывода ни по одной позиции.

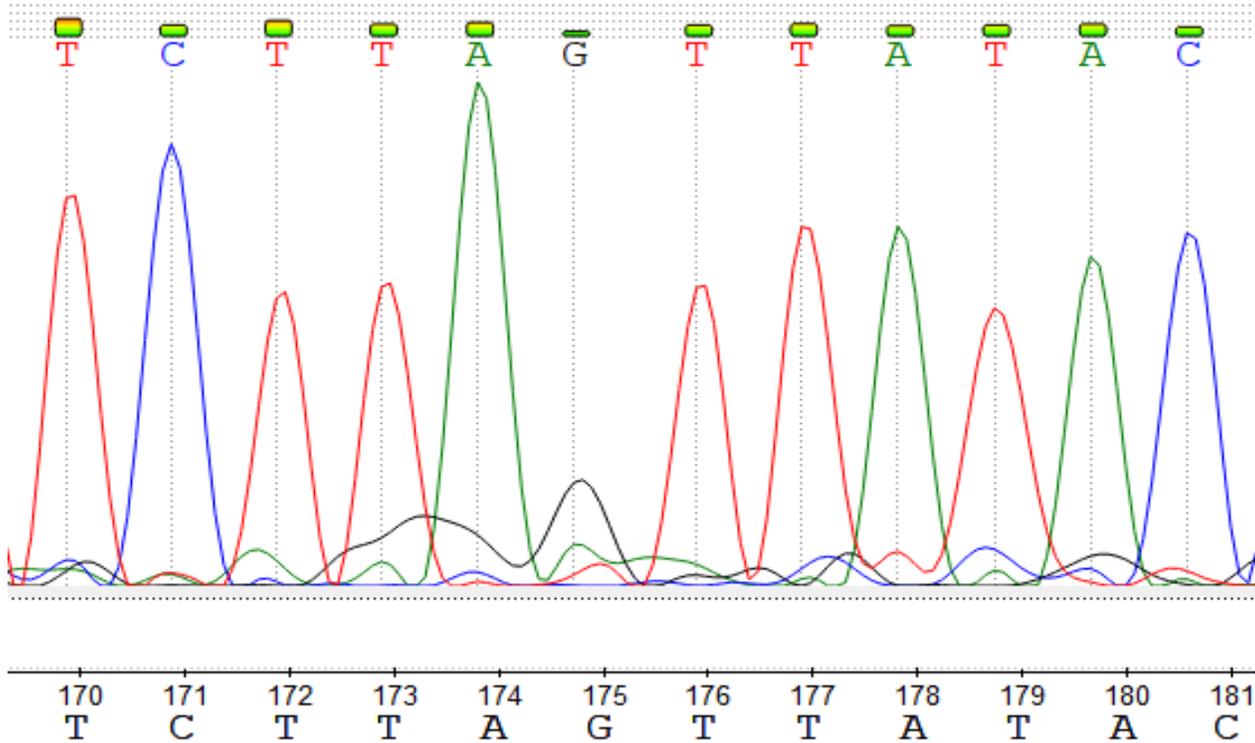


Рис. 8. Участок 170-181. Небольшие опасения может вызывать 175-G, так как пик не сильно выше, чем “черный шум” между 173-T и 174-A. Но, в целом, габитуально смотрится как нормальный (так себе объяснение). Вердикт: мне кажется, можно довериться base calling в этом случае, но повод для сомнений есть

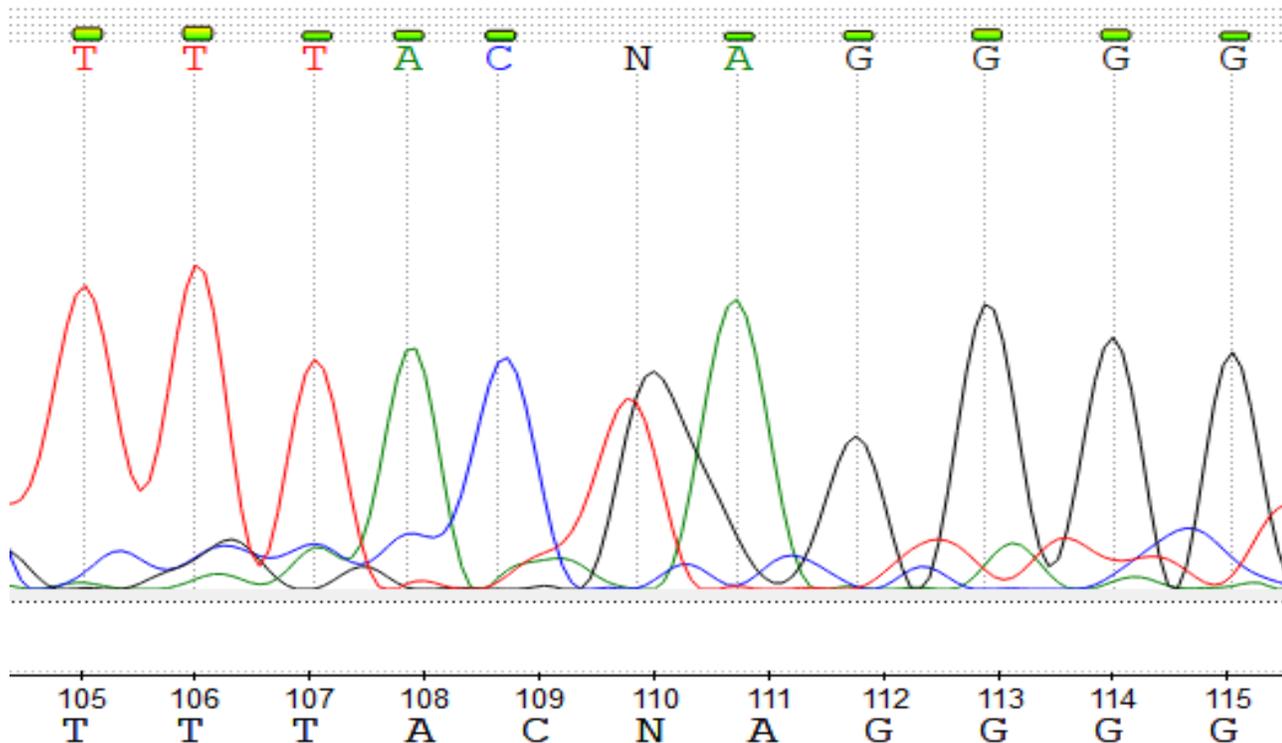


Рис. 9. Участок 105-115. На мой взгляд, самый интересный кейс. Для участка характерен высокий уровень “шума”, но, я думаю, в случае N-110 все же можно говорить о неразделенных пиках и , соответственно пропущенном нуклеотиде (так как пики все же существенно выше, чем окружающий “шум”). Вердикт: 110-T и пропущенный G между 110 и 111.

OVERALL: возможно, мне попалась не самая удачная хроматограмма, так как мне было довольно сложно найти проблемные позиции (скорее всего, я что-то проглядел). Приведенные примеры могут показаться не очень убедительными, но я все же не стал просить поменять файл. В целом, мне кажется, base calling справился довольно хорошо, а имеющийся уровень “шума” не позволил сделать достоверные предположения о его ошибках (в большей части случаев)