



СЕКВЕНИРОВАНИЕ ПО СЭНГЕРУ

Екатерина Рюмина, Анастасия Жарикова
ФББ МГУ 2023

Секвенирование

- Определение последовательности некоторого нерегулярного биологического гетерополимера – белка или нуклеиновой кислоты

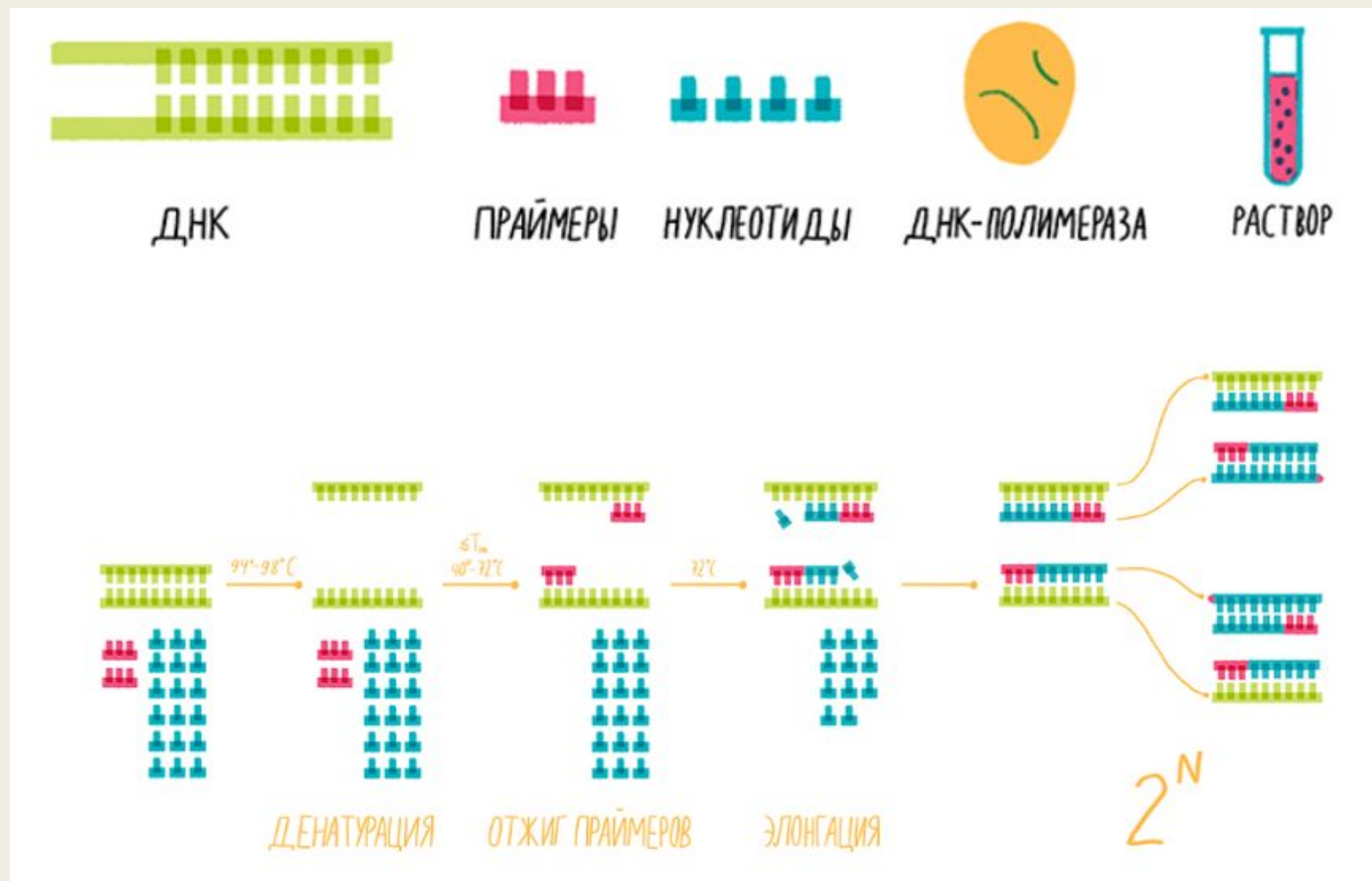
Экскурс в историю

- Секвенирование по Сэнгеру (1975 – 1977)
 - *Фредерик Сэнгер, Нобелевская премия 1980 (вторая!) (с Уолтером Гильбертом и Полем Бергом)*
- Полимеразная цепная реакция (1985 – 1986)
 - *Кэрри Муллис, Нобелевская премия 1993 (совместно с Майклом Смитом)*
- NGS — совокупность методов (454, Illumina, Oxford Nanopore, PacBio и проч.) (2005 – наст. вр.)

Полимеразная цепная реакция - ПЦР

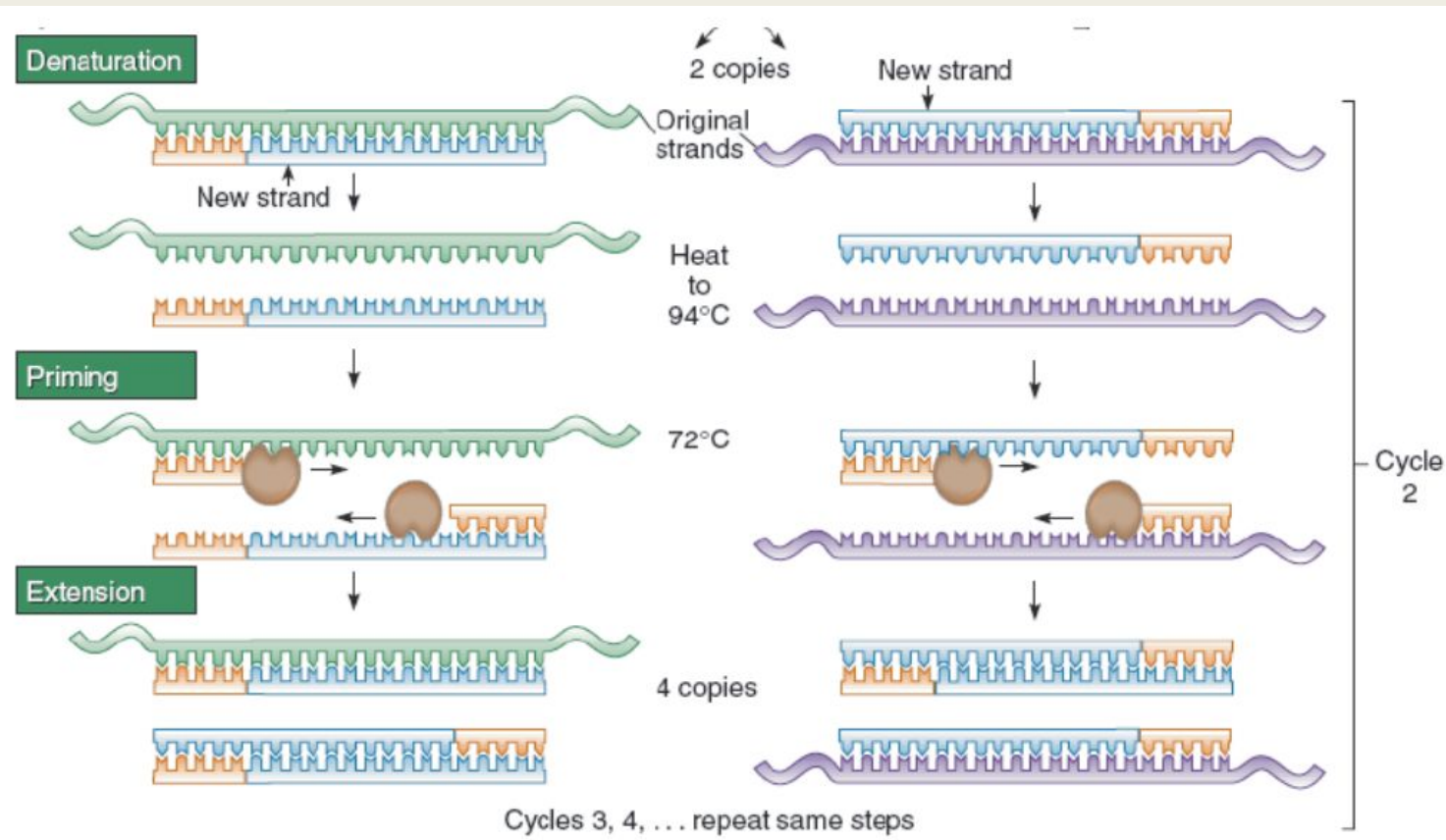
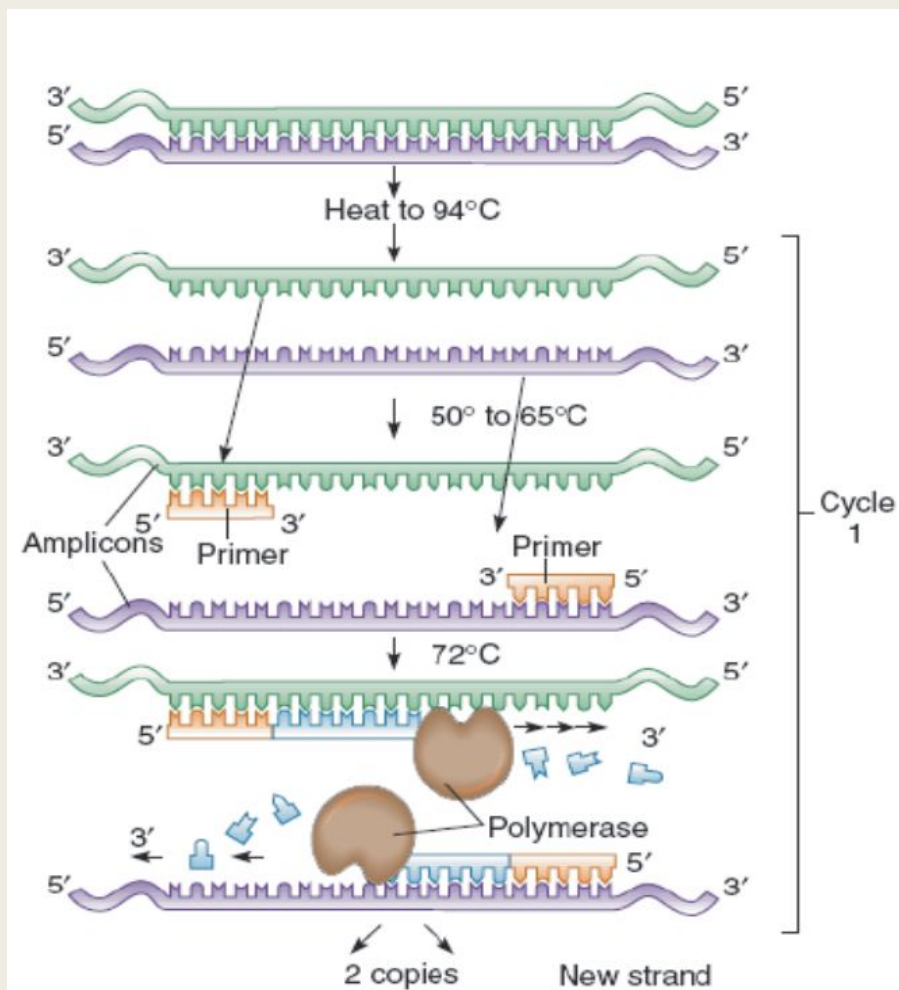
- <https://www.youtube.com/watch?v=DkT6XHWne6E>
- <https://www.dnalc.org/view/15475-The-cycles-of-the-polymerase-chain-reaction-PCR-3D-animation.html>
- <https://www.youtube.com/watch?v=iQsu3Kz9NYo>
- Для чего нужна?
- Что на вход?
- Времена этапов процесса

ПЦР

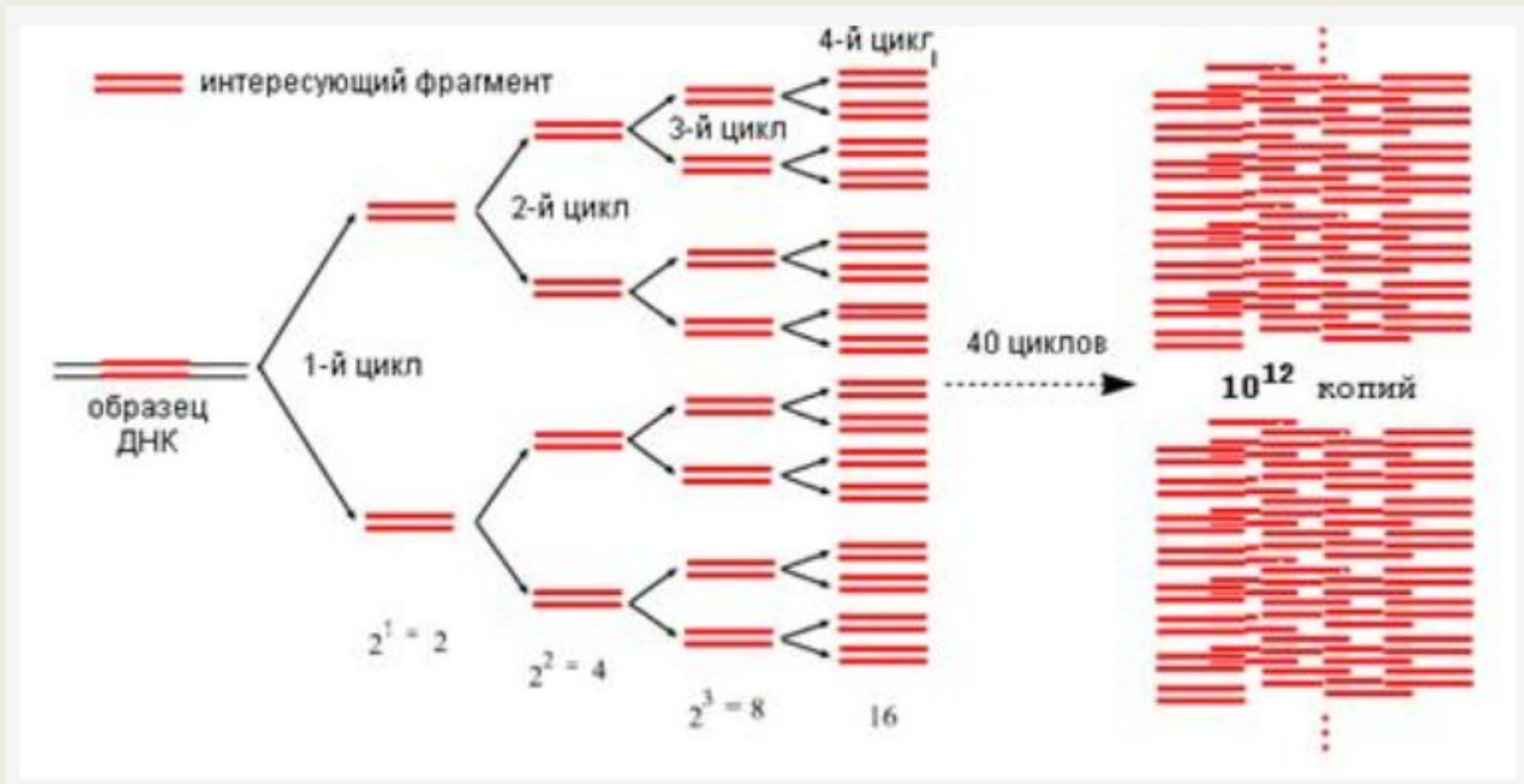


<https://biomolecula.ru/articles/metody-v-kartinkakh-polimeraznaia-tsepnaia-reaktsii>

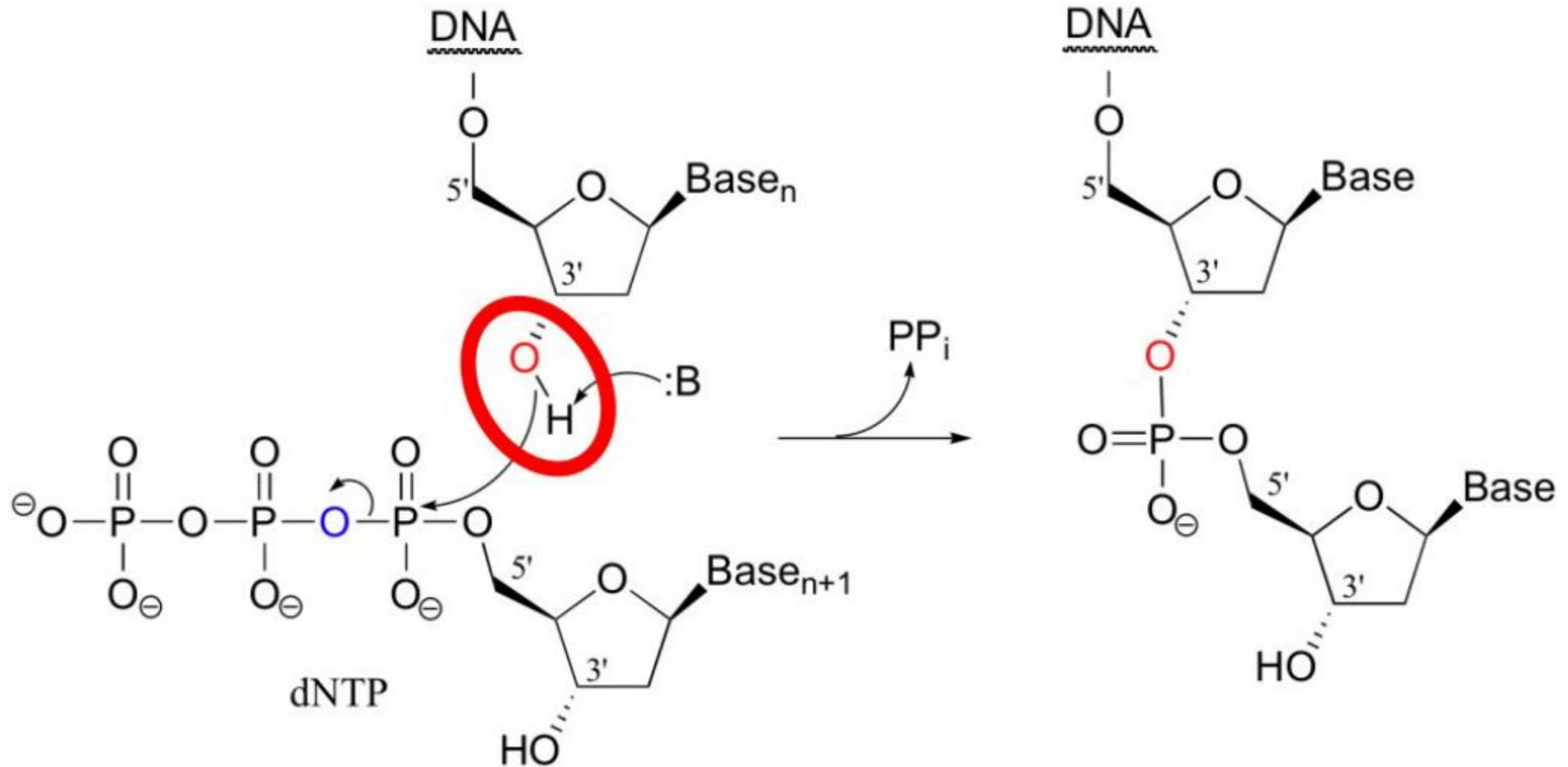
ПЦР



ПЦР

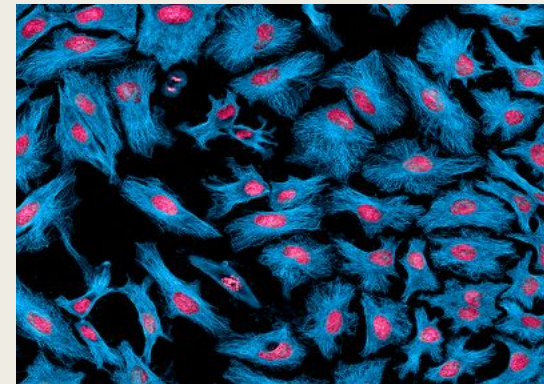


Реакция полимеризации



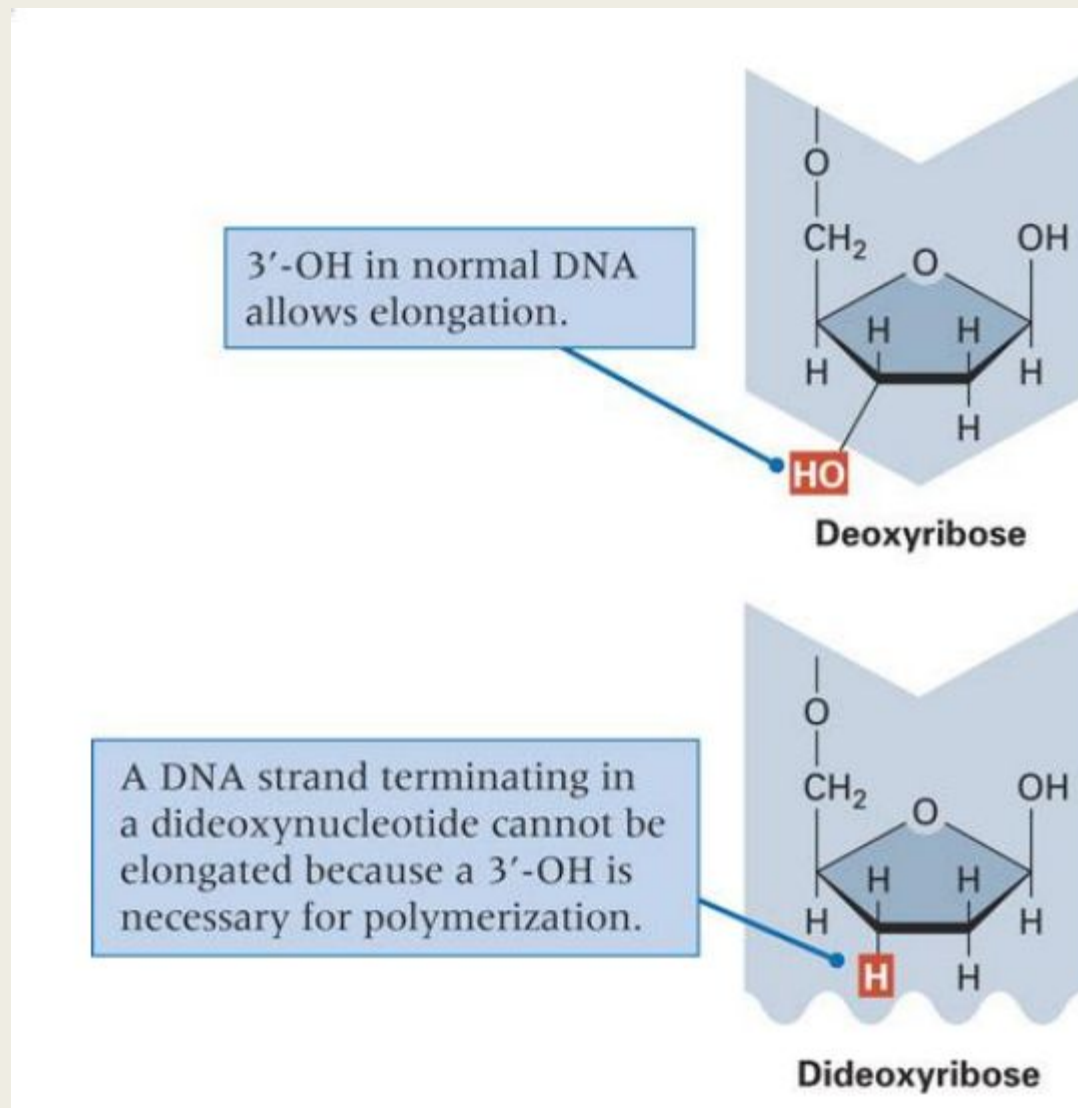
Секвенирование ДНК

- Выделение ДНК



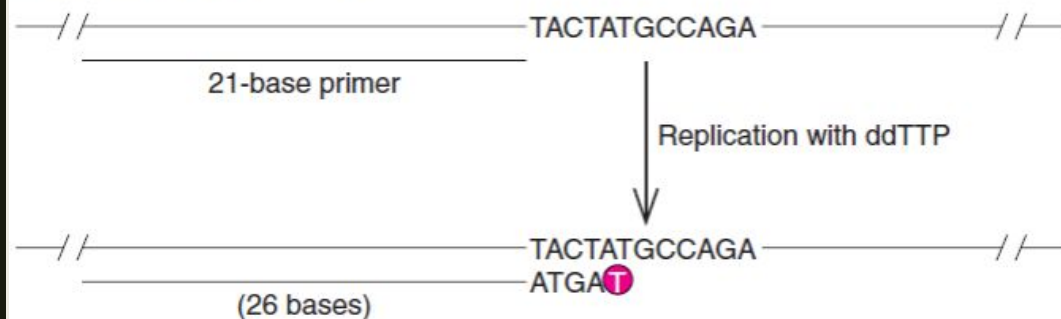
Секвенирование ДНК

- Метод «терминаторов»



Секвенсовая реакция

(a) Primer extension reaction:



Ферменты

- ДНК полимеразы I
- Секвенза – ДНК полимеразы со сниженной 3'→5'-эндонуклеазной активностью и повышенной процессивностью

(b) Products of the four reactions:

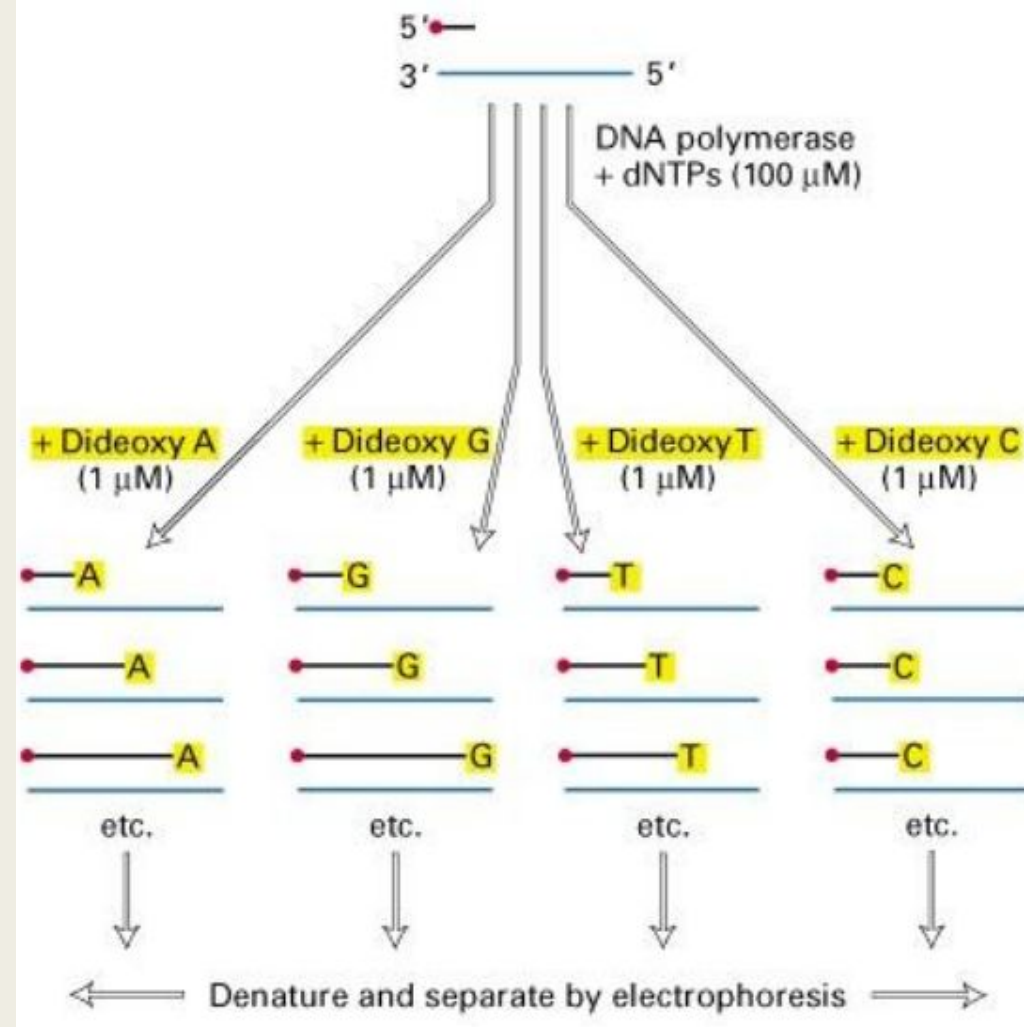
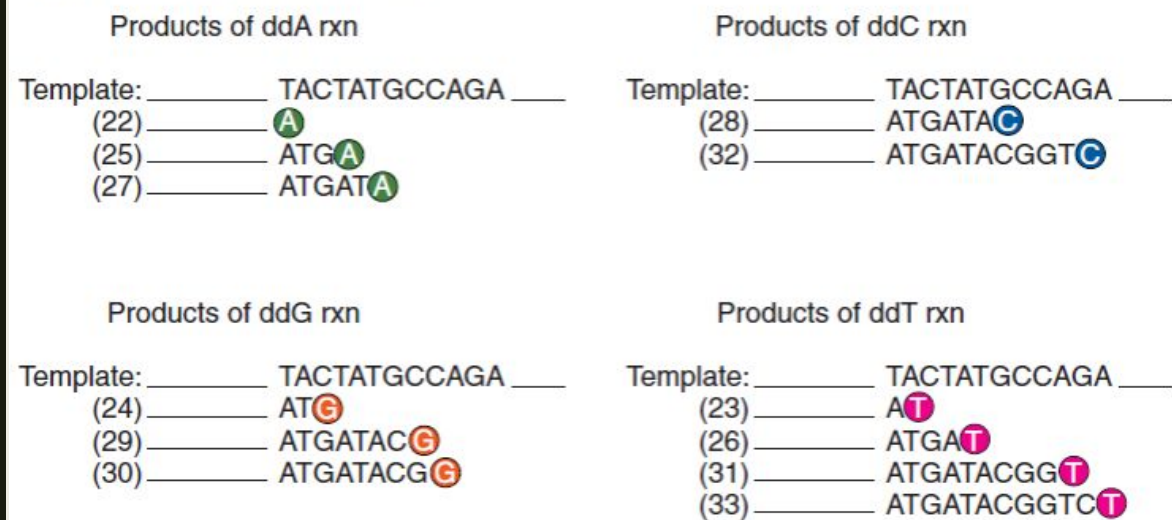
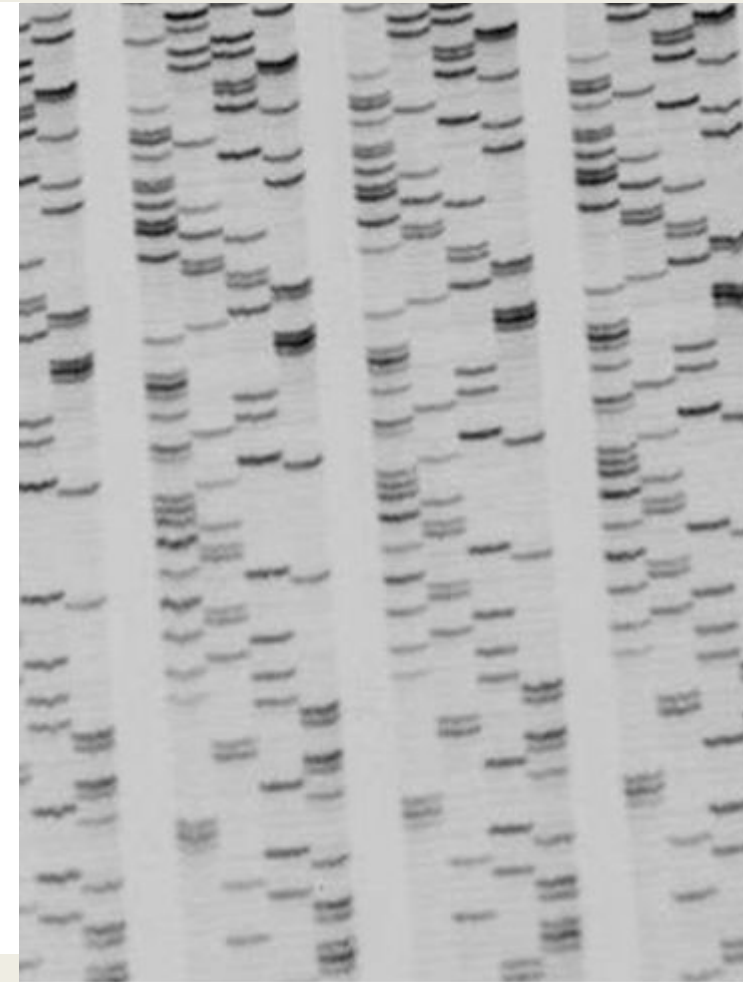
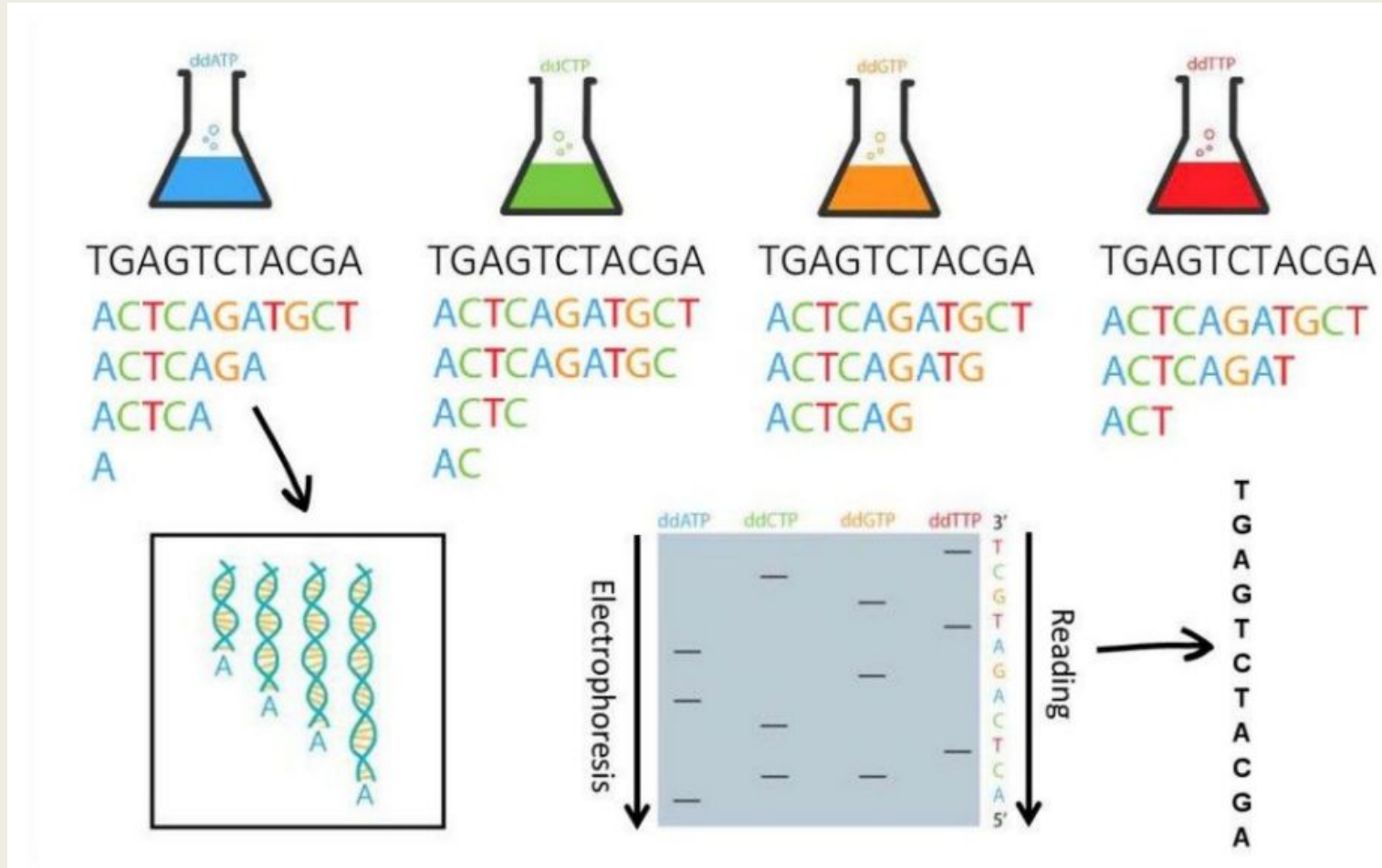
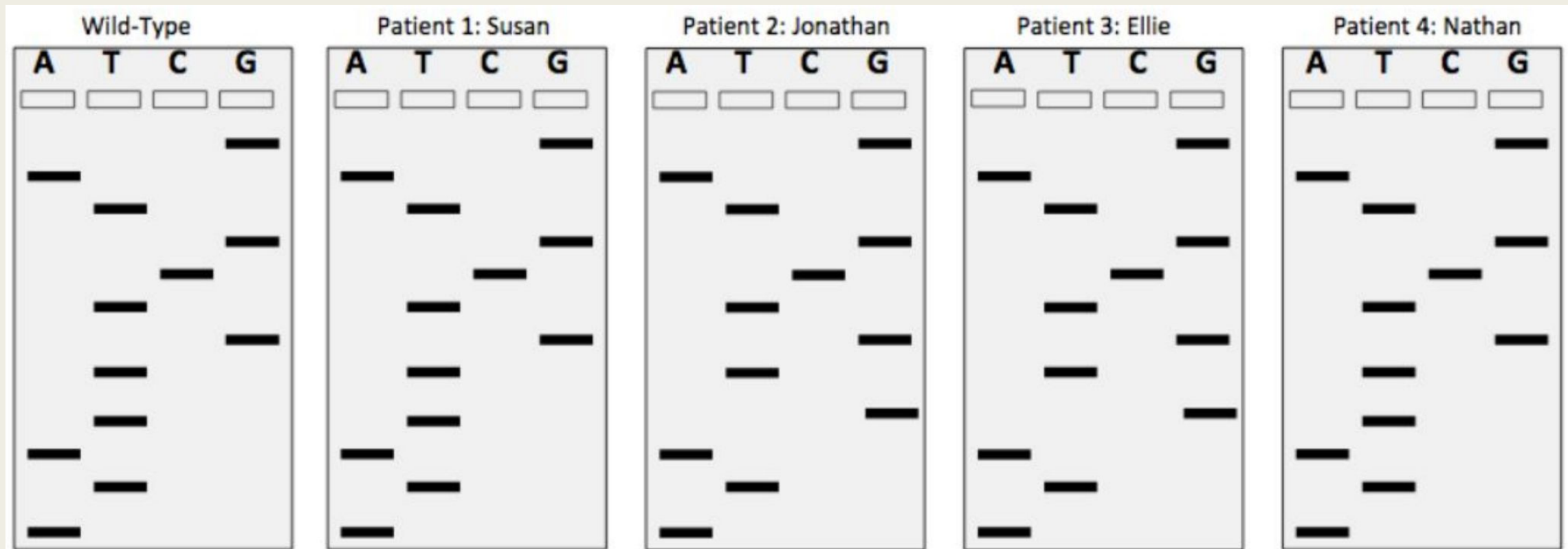


Схема процесса секвенирования по Сэнгеру



<https://www.youtube.com/watch?v=FvHRio1yyhQ>

Ниже даны результаты секвенирования по Сэнгеру четырёх пациентов. У кого из них мутация?



Современная методика

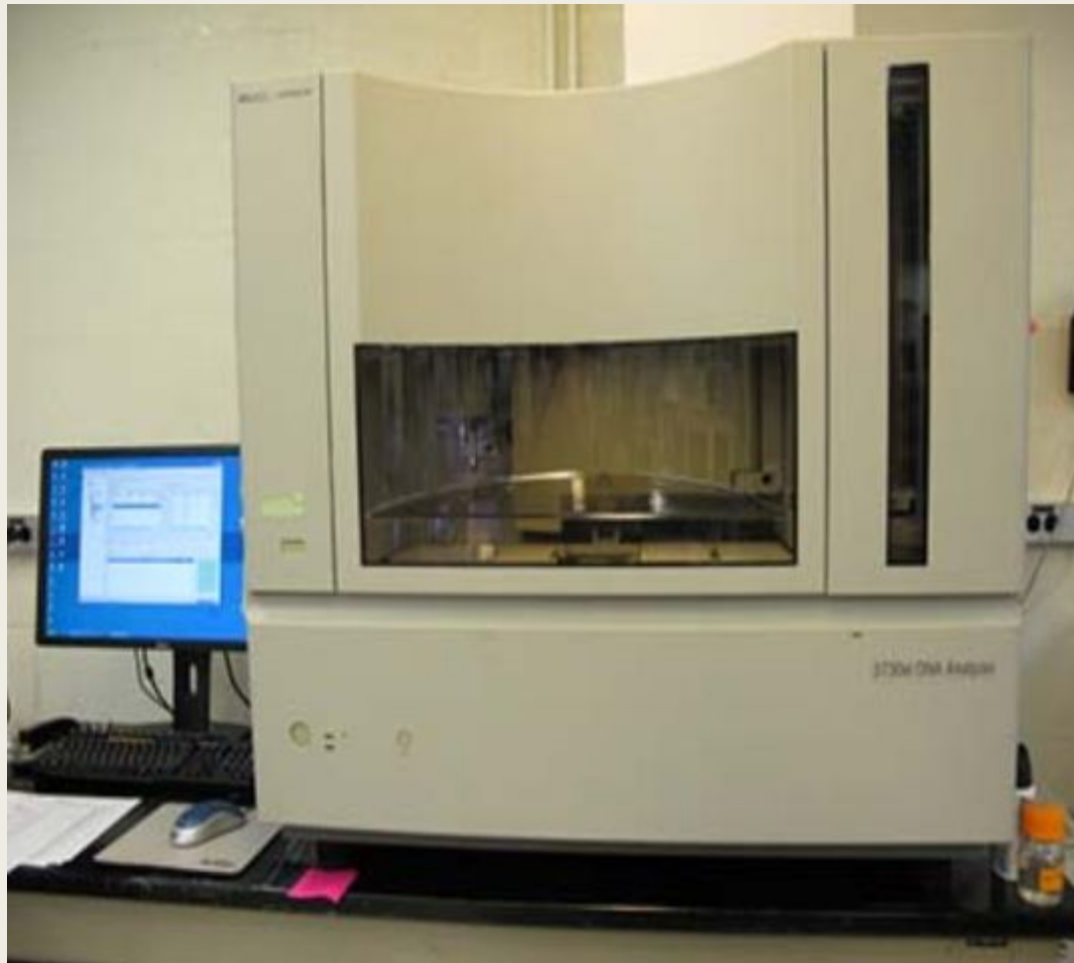
Сейчас вместо того, чтобы проводить фореуз в четырёх пробирках с разными ddNTP, проводят один фореуз на смеси, которая содержала все четыре типа ddNTP, каждый из которых окрашен по-своему

Фореуз проводится в капиллярах (а не в геле) — так у него получается больше разрешение

Максимальная длина прочтения — не больше 1000-1200 нуклеотидов

Почему до сих пор применяется, если есть более производительные методы секвенирования?

Современный капиллярный секвенатор



Много капилляров – секвенируем параллельно несколько фрагментов ДНК

ddNTP с красителями

Для секвенирования к терминирующему нуклеозиду в зависимости от их типа – А, Т, G, С – присоединяют флюорофоры, излучающие в

разных длинах волн
Primer

ACGTACGTACTCAGATGCT
ACGTACGTACTCAGATGC
ACGTACGTACTCAGATG
ACGTACGTACTCAGAT
ACGTACGTACTCAGA
ACGTACGTACTCAG
ACGTACGTACTCA
ACGTACGTACTC
ACGTACGTACT
ACGTACGTAC
ACGTACGTA

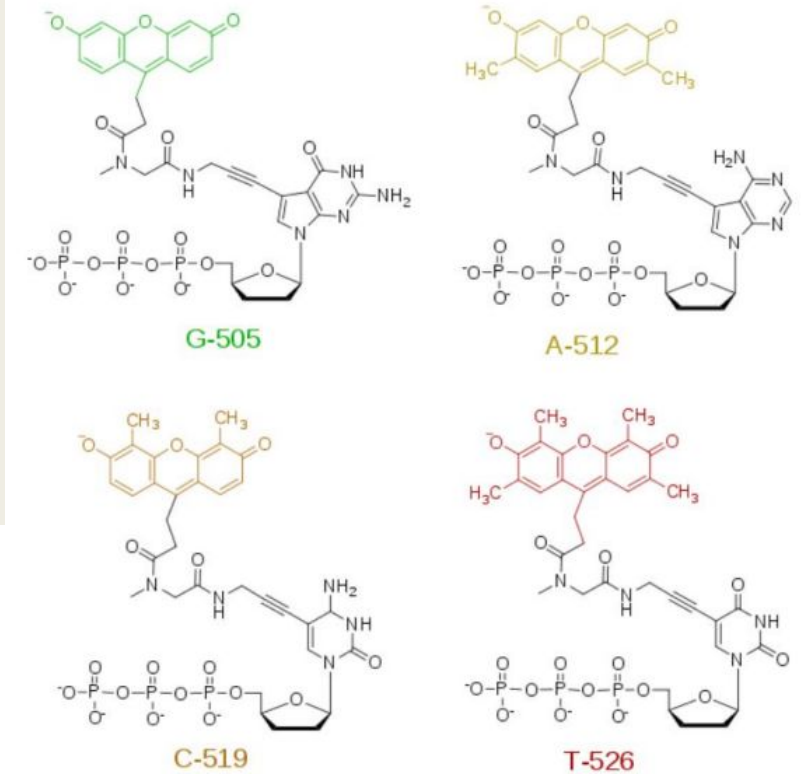
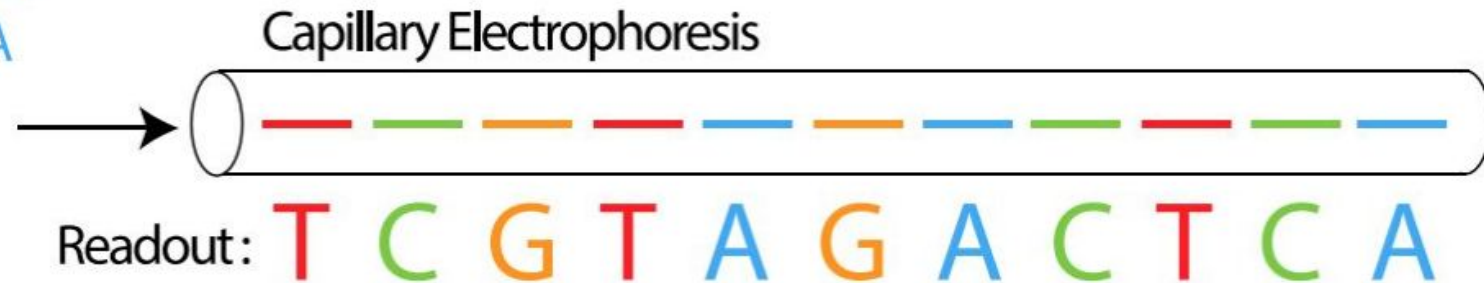
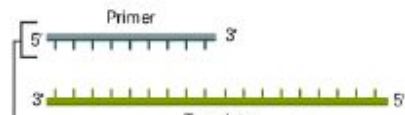


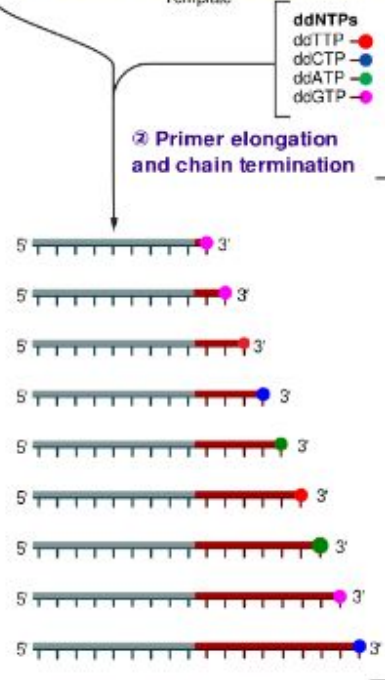
Схема эксперимента

① Reaction mixture

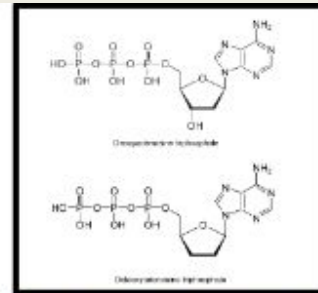
- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouochromes ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



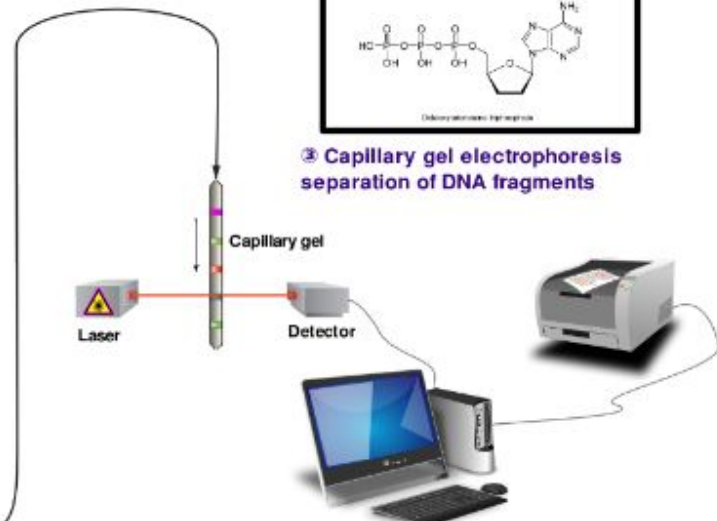
② Primer elongation and chain termination



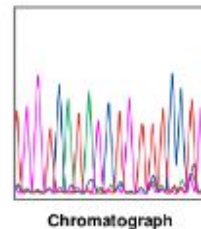
Результат основной реакции
(в пробирке)



③ Capillary gel electrophoresis separation of DNA fragments



④ Laser detection of flouochromes and computational sequence analysis

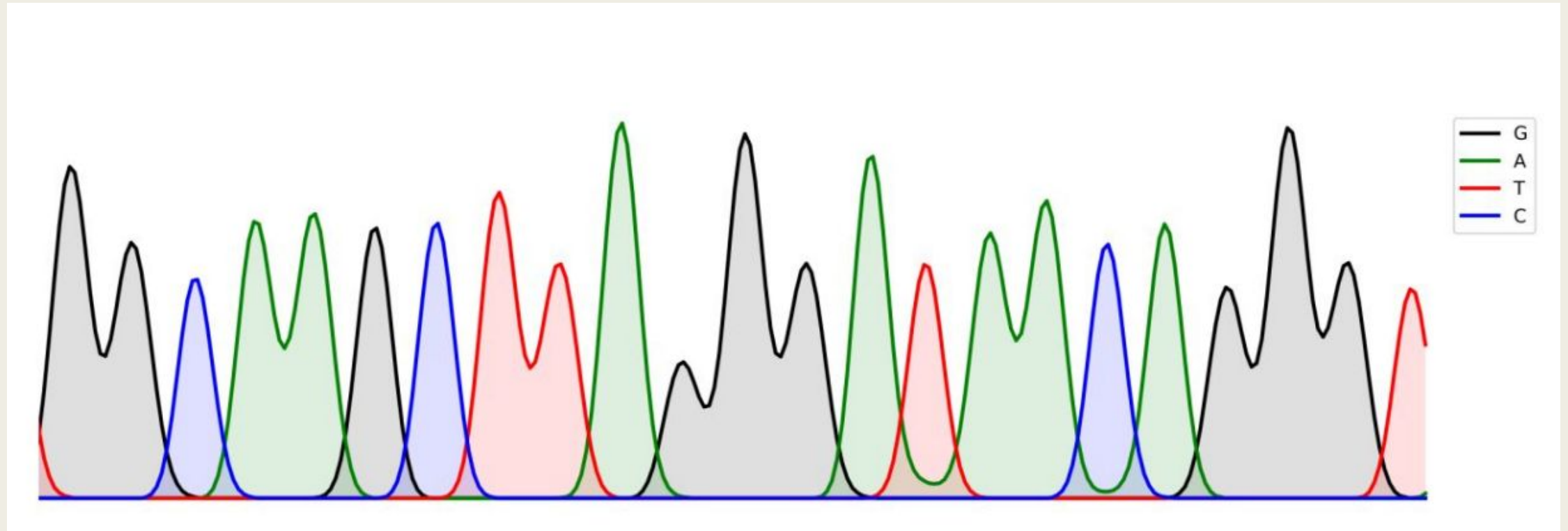


Результат капиллярного электрофореза
(в компьютере)

Полоски могут быть размазанными и наезжать друг на друга.

Помните: различие между полосками определяется Единственным нуклеотидом!

Хроматограмма



Итого

Сначала мы готовим препарат:

нам необходима чистая фракция ДНК, которую надо отсеквенировать;
она амплифицирована с помощью ПЦР,
а также на неё имеются специфические праймеры

Раствор для секвенирования:

- матрица ДНК в большом количестве;
- нуклеотиды (dATP, dGTP, dCTP, dTTP) — много;
- терминирующие нуклеотиды (ddATP, ddCTP, ddGTP, ddTTP) — мало;
- один праймер;
- ДНК-полимераза

Расшифровка хроматограммы

Полученную хроматограмму надо верно проанализировать, чтобы восстановить, какая последовательность ДНК у нас была в образце

Чаще всего программное оборудование секвенатора делает это само.

Этот процесс называется base calling

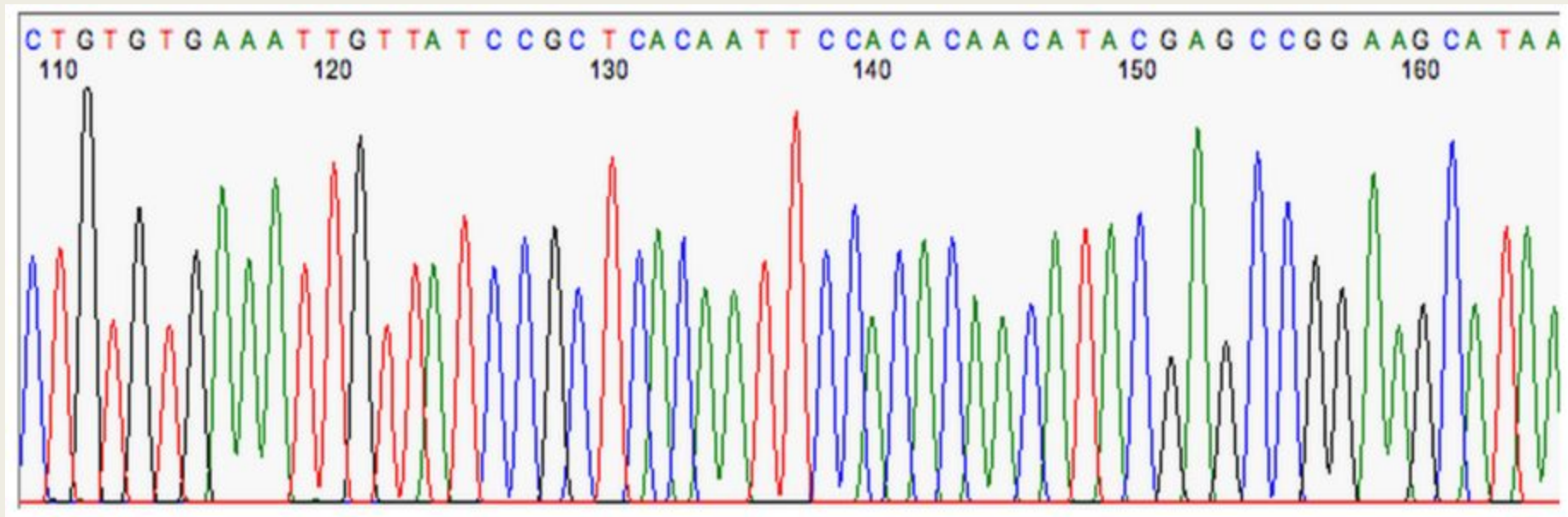
Во время base calling'а могут происходить ошибки

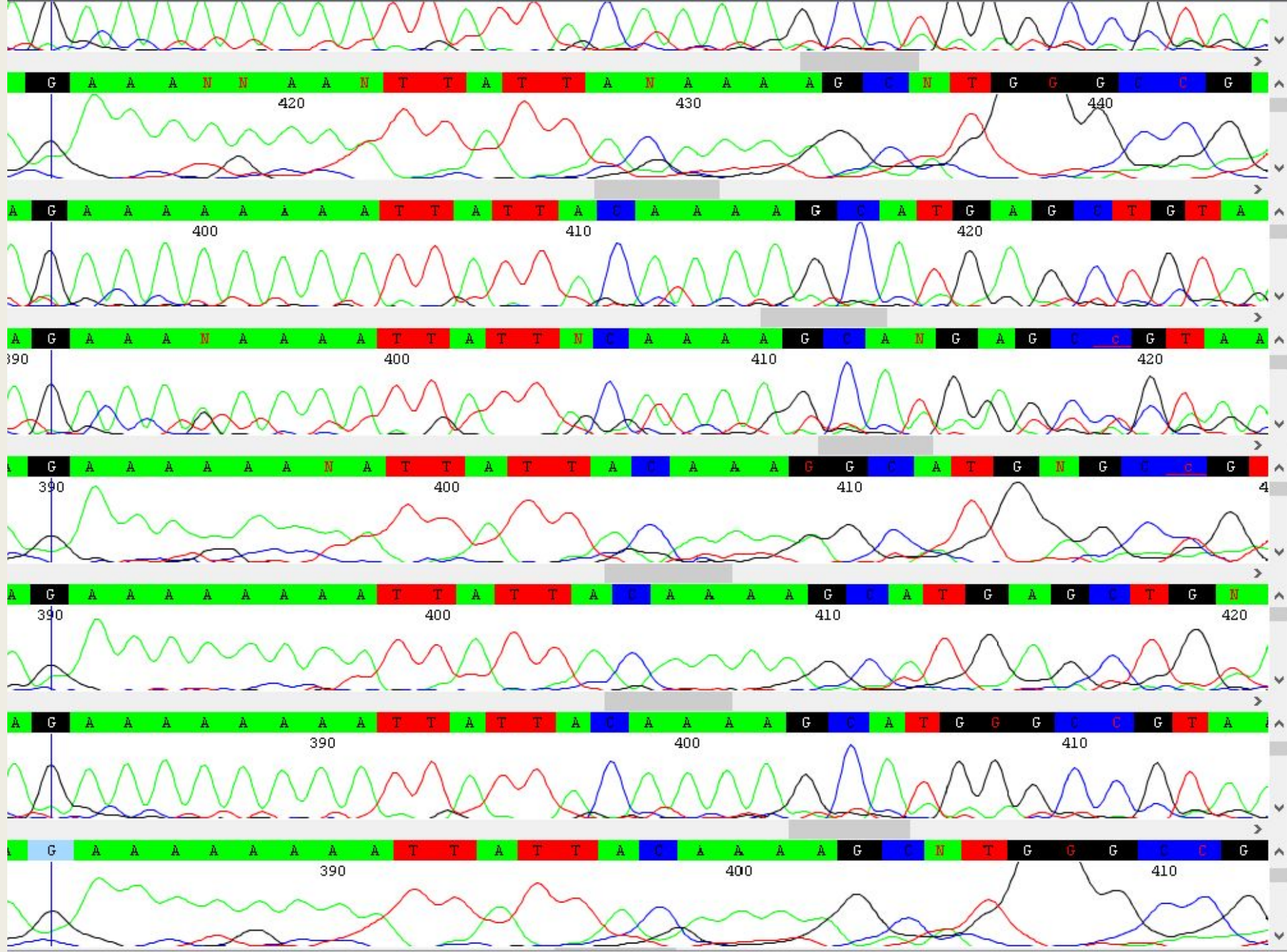
Результат автоматического секвенирования по Сэнгеру



Идеальная хроматограмма

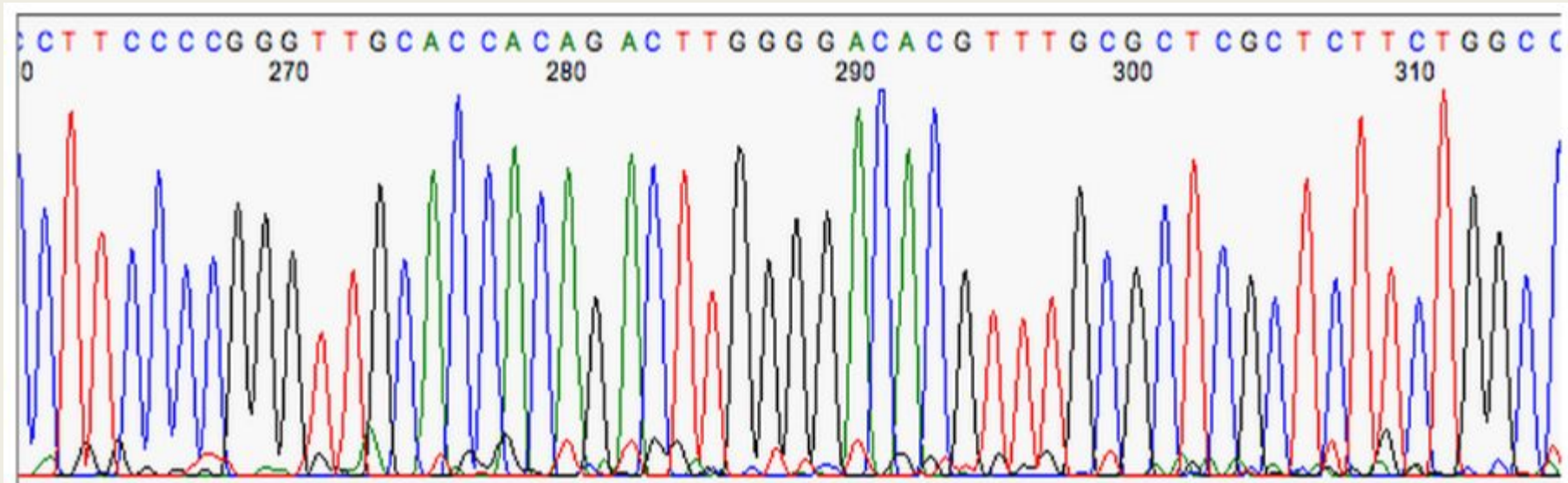
- Сила сигнала – это интенсивность детектируемого света в данный момент времени
- На графике отражена высотой





Шум

- В разы ниже сигнала (в идеале)
- Это световой сигнал от посторонних молекул ДНК, загрязнения.
 - Например, праймеры при ПЦР отжигались и на посторонний фрагмент ДНК.
 - Или малая доля праймеров в основной реакции отождася на другое место в целевой ДНК.



Качество сигнала

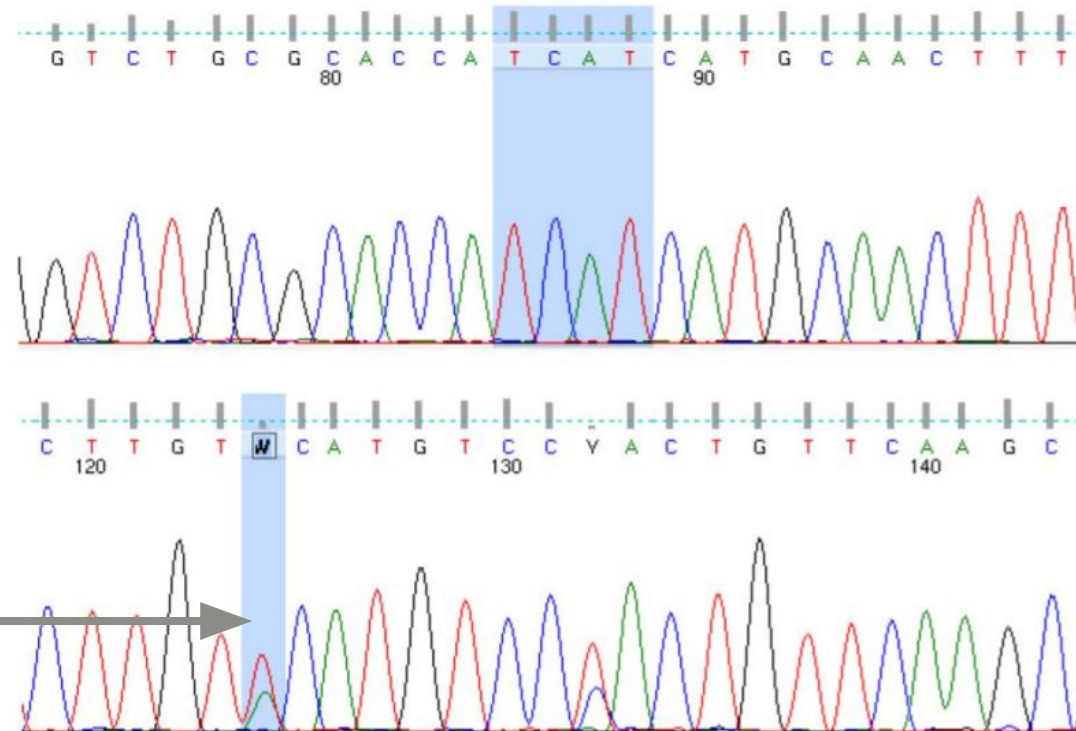
Phred quality score:

$$Q_{\text{sanger}} = -10 \log_{10} p$$

- Q = 10 - вероятность ошибки 1/10
- Q = 20 - вероятность ошибки 1/100
- Q = 30 - вероятность ошибки 1/1000

Подробности тут: [Phred quality score - Wikipedia](#)

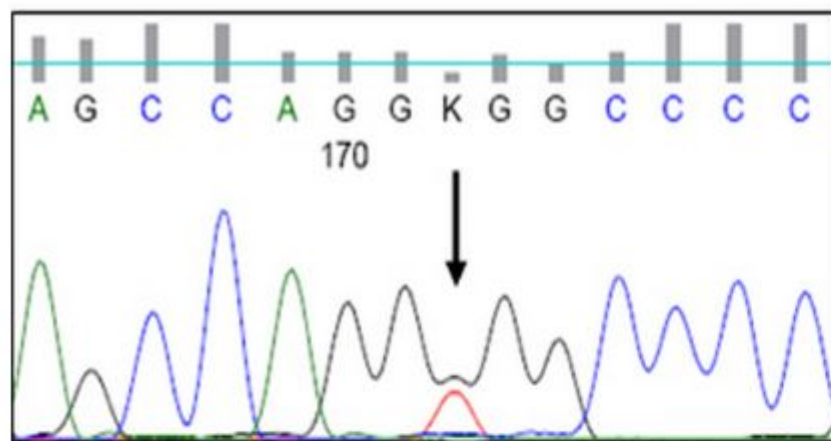
Качество аннотации каждой буквы отображается в программах для визуализации



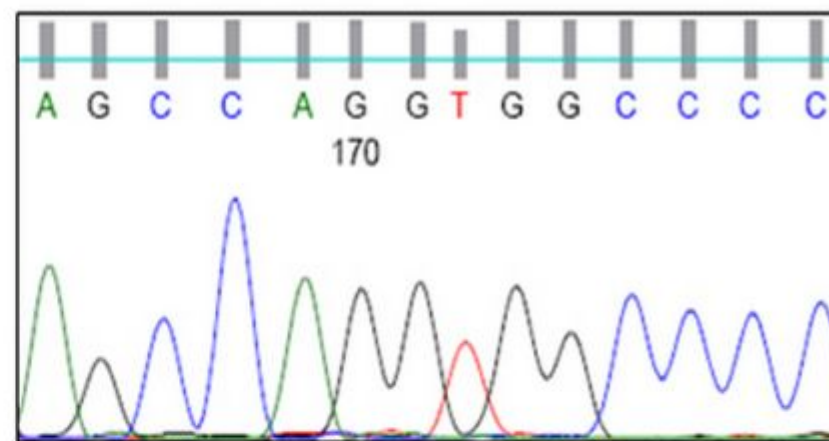
Источники полиморфизмов

- Гетерозиготы
- Соматические мутации
- Несколько разных образцов ДНК
- Несколько генов в разных локусах
- Ошибки во время ПЦР

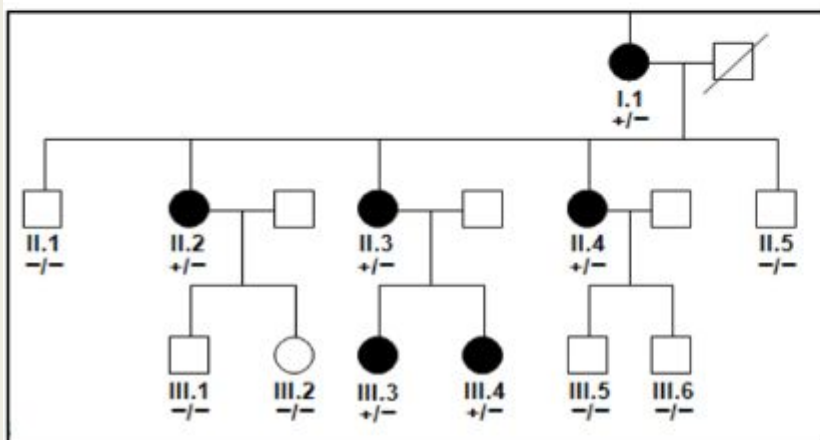
Полиморфизм в гене родопсина



c.1034T>G (p.Val345Gly) mutation



Wild-type sequence



Семья RPT100

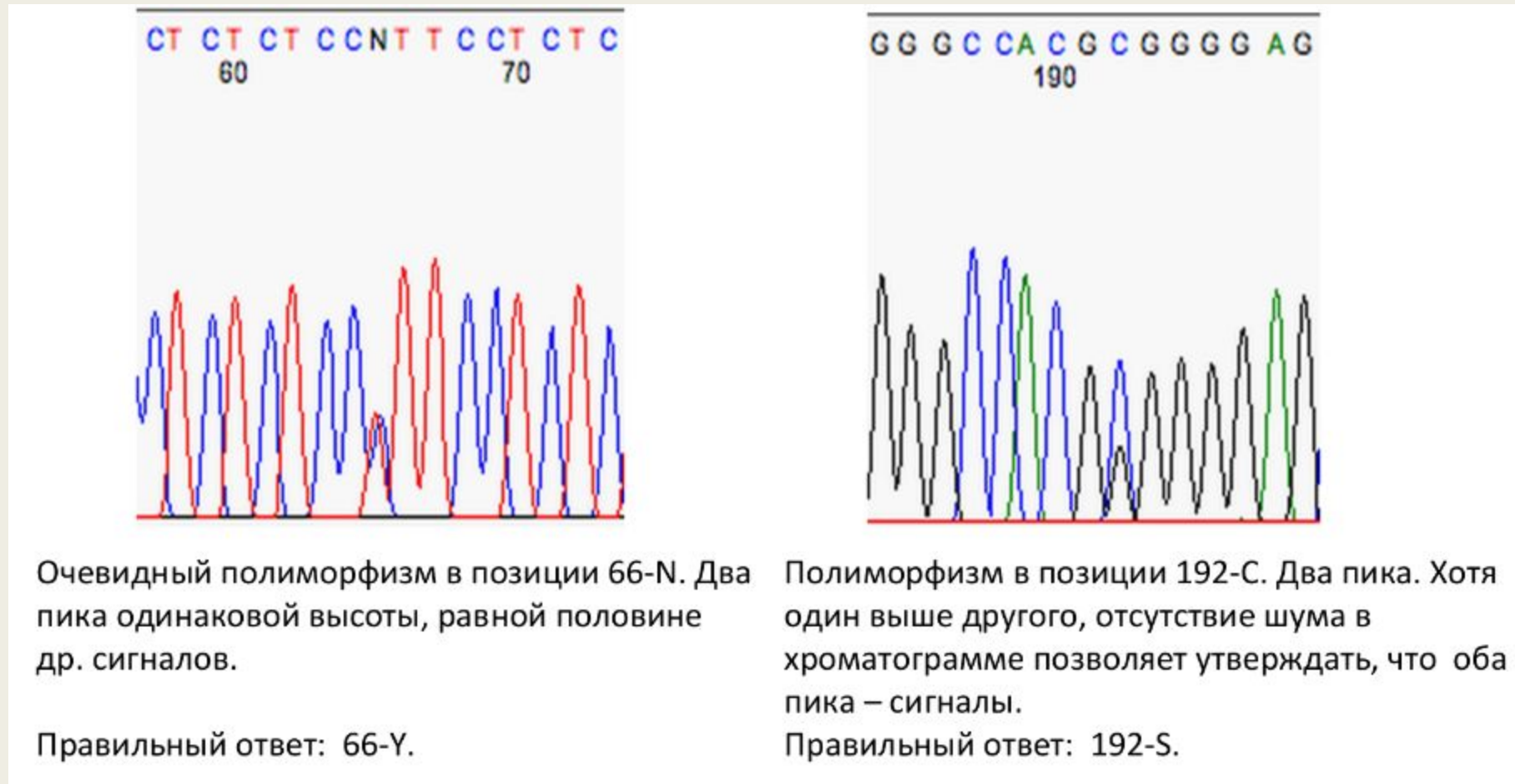
Квадратик – М

Кружок - Ж

Залиты черным – пациенты с полиморфизмом;
у всех признаки пигментного ретинита

de Sousa Diaz et al., Mol. Vis., 2015

Еще полиморфизмы



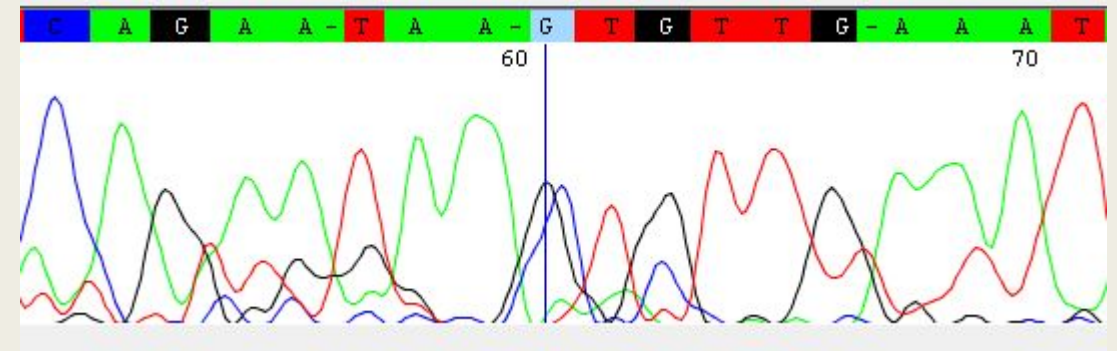
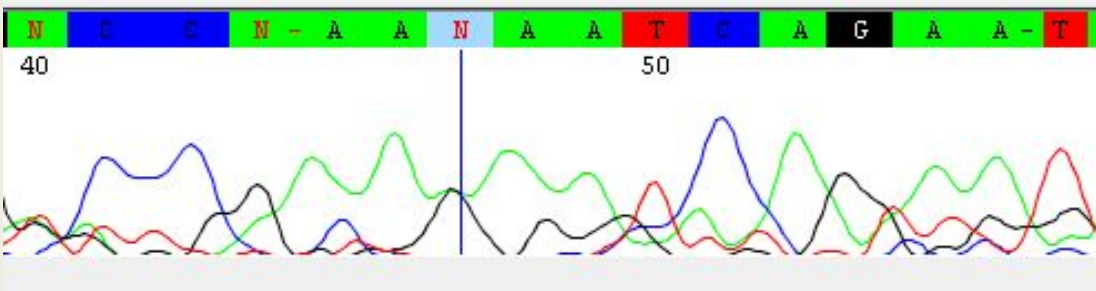
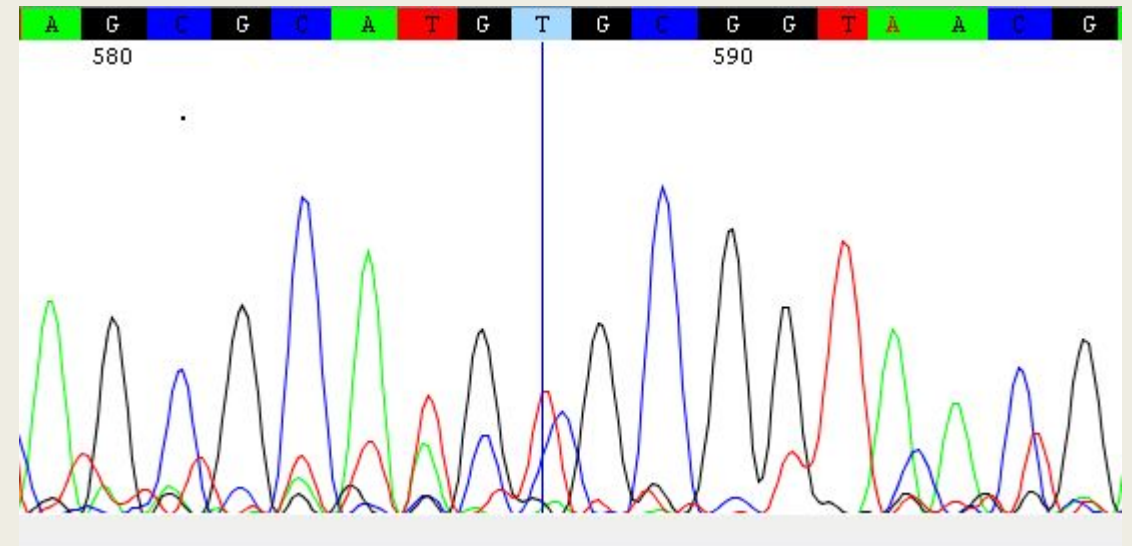
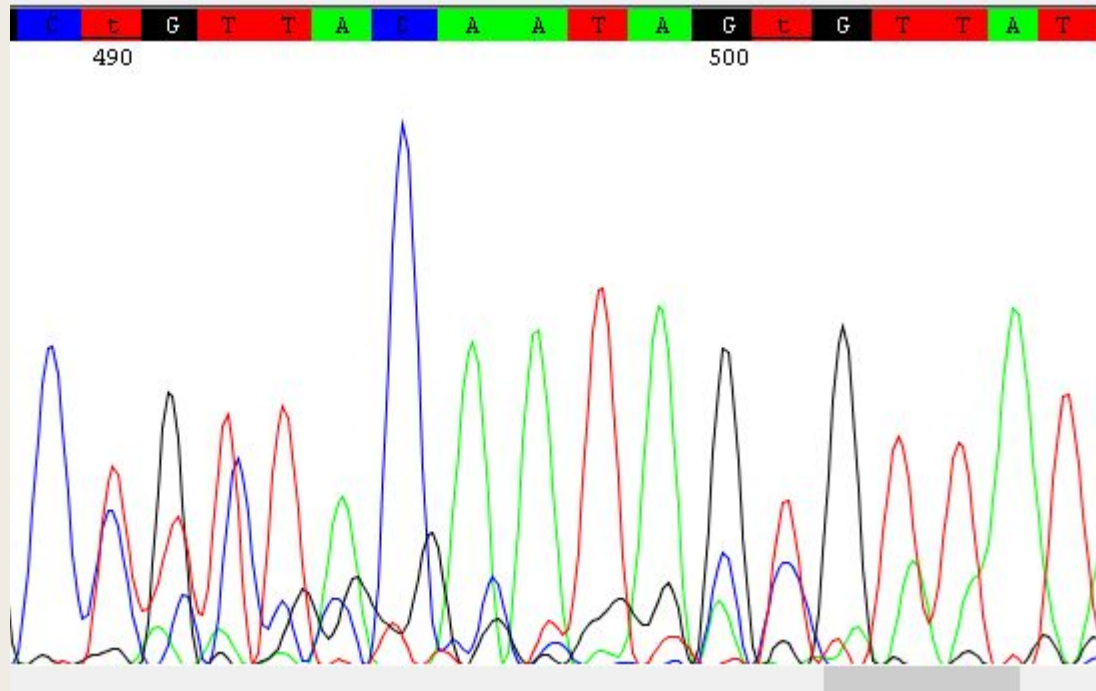
Очевидный полиморфизм в позиции 66-N. Два пика одинаковой высоты, равной половине др. сигналов.

Правильный ответ: 66-У.

Полиморфизм в позиции 192-С. Два пика. Хотя один выше другого, отсутствие шума в хроматограмме позволяет утверждать, что оба пика – сигналы.

Правильный ответ: 192-S.

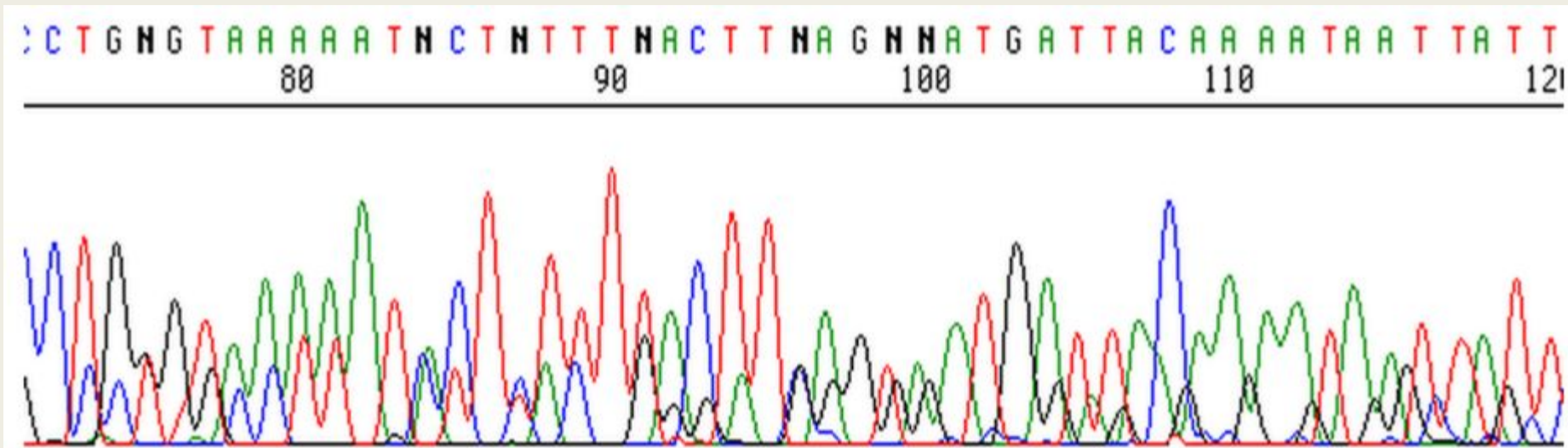
Полиморфизм, шум или одна буква?



A, T, G, C, ... ???!!!

Nucleotide symbol	Full Name
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil
R	Guanine / Adenine (purine)
Y	Cytosine / Thymine (pyrimidine)
K	Guanine / Thymine
M	Adenine / Cytosine
S	Guanine / Cytosine
W	Adenine / Thymine
B	Guanine / Thymine / Cytosine
D	Guanine / Adenine / Thymine
H	Adenine / Cytosine / Thymine
V	Guanine / Cytosine / Adenine
N	Adenine / Guanine / Cytosine / Thymine

Две разных ДНК в препарате



В препарате читаются одновременно два разных фрагмента.

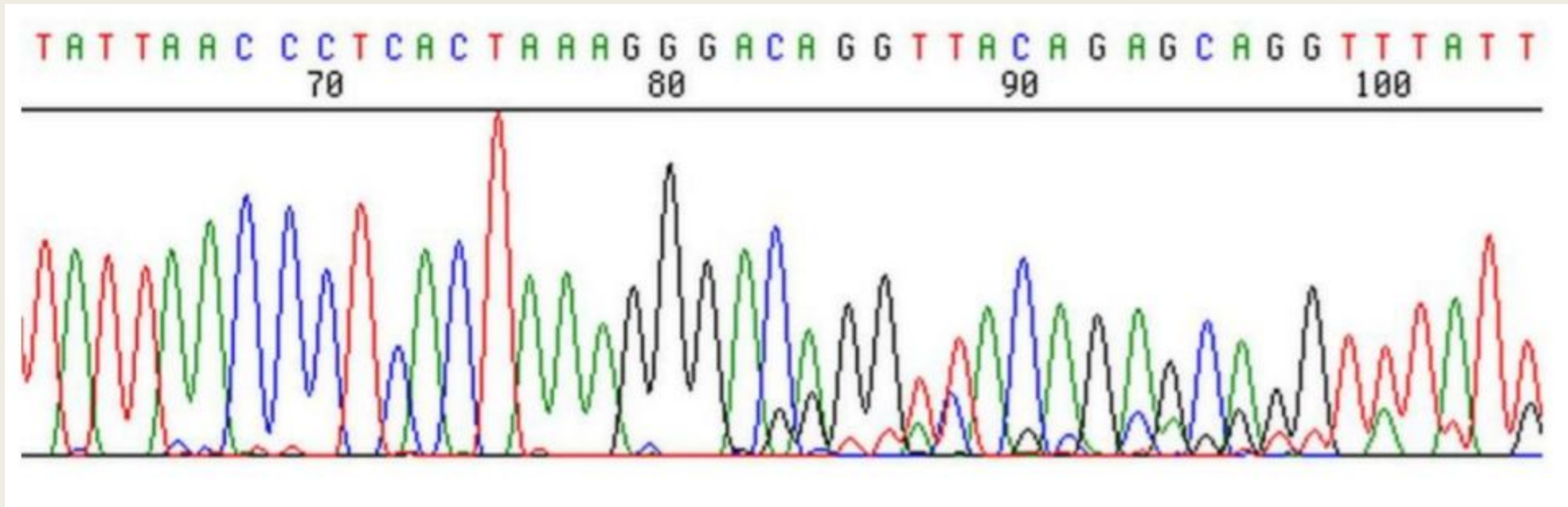
Либо праймер для секвенирования отжегся на два разных участка.

Либо при ПЦР поднялись два фрагмента из исходной ДНК.

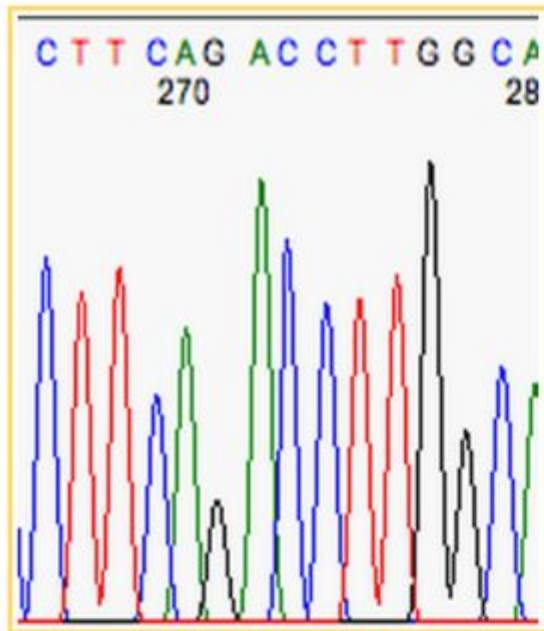
Последовательность прочесть невозможно. Автомат полностью ошибся.

Появление вторых пиков

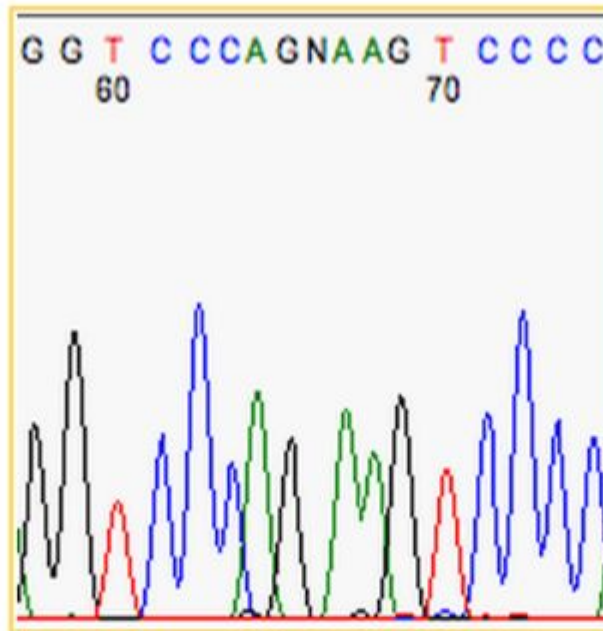
- В результате делеций на хроматограмме в неожиданный момент начнут систематически возникать вторые пики



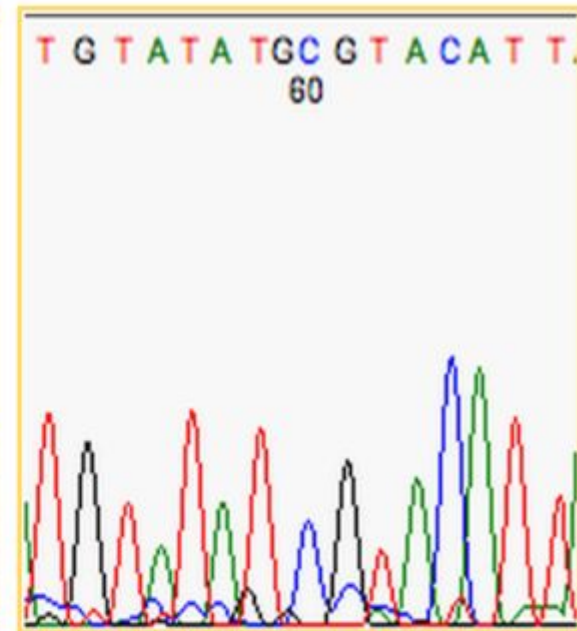
Разное расстояние между сигналами



Правильное прочтение автоматом



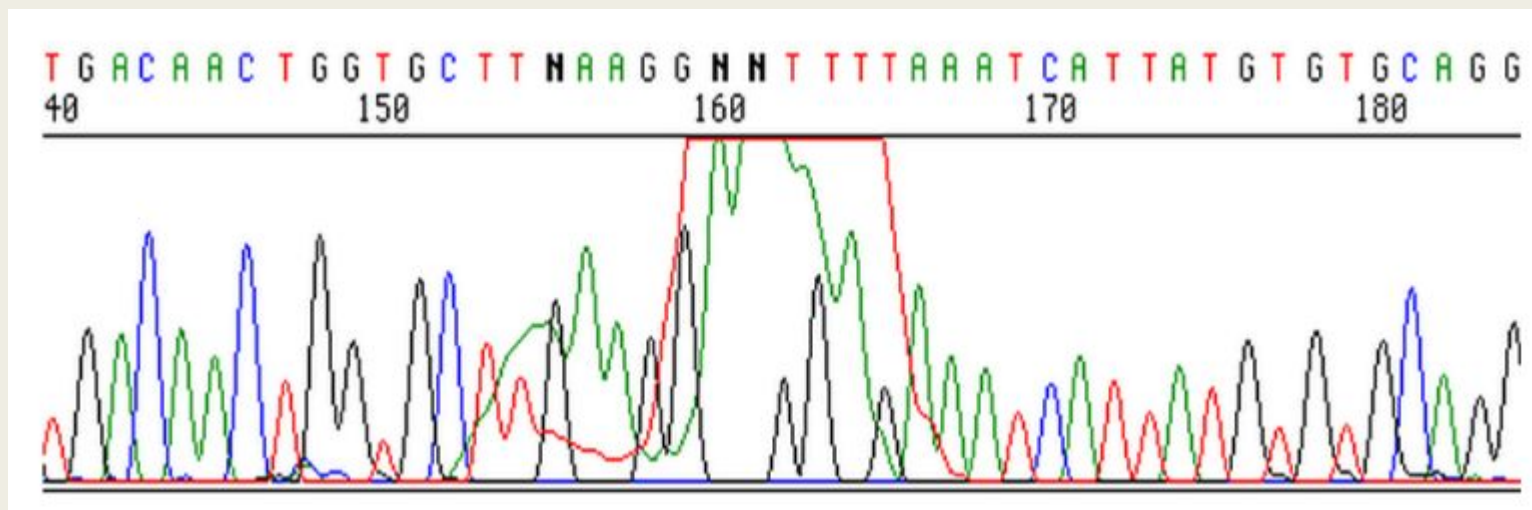
Ошибка автомата:
лишняя буква N
за 65-G!



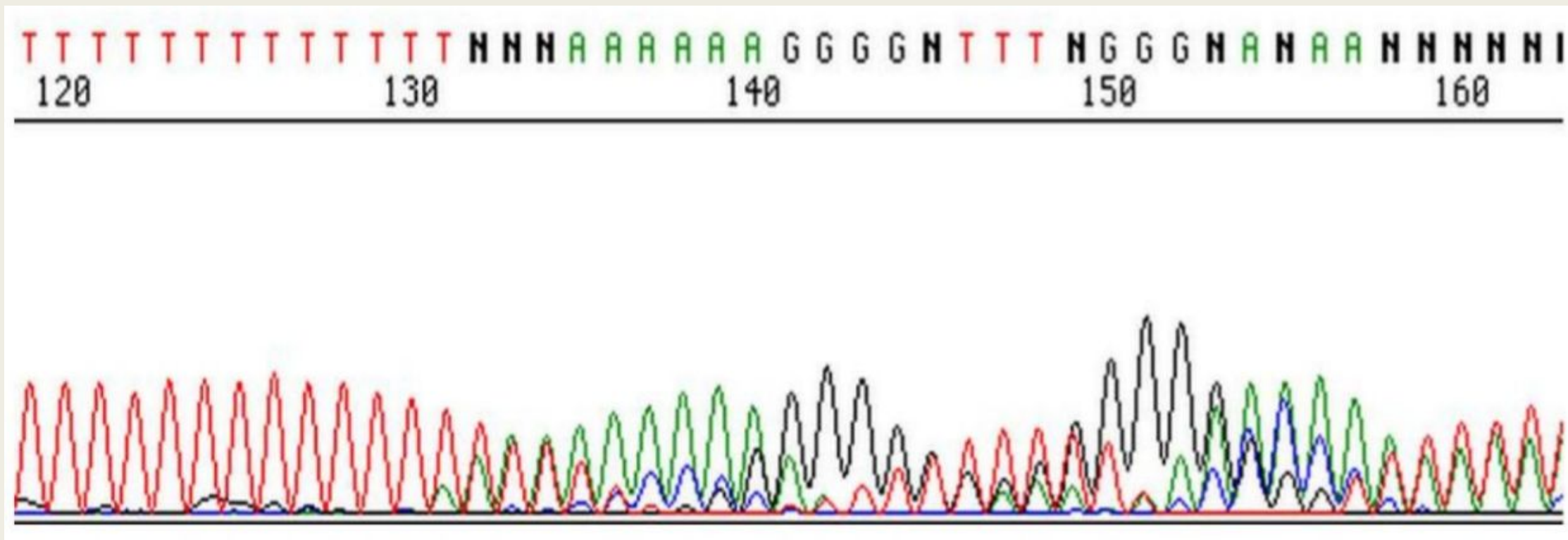
Ошибка автомата:
лишняя буква G
за 58-T!

Пятно краски

- Может возникнуть из-за ошибок во время фореаза

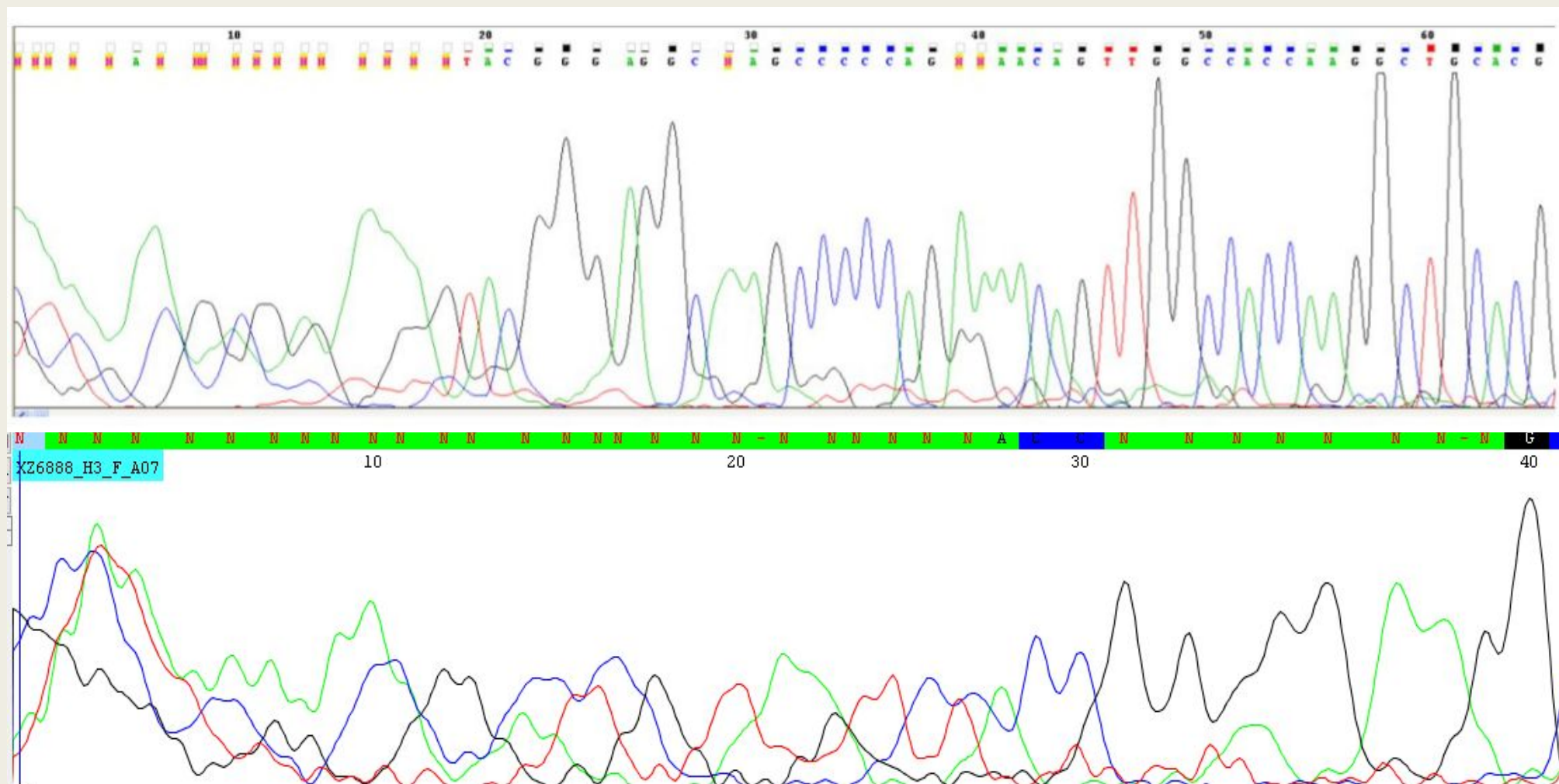


Проскальзывание полимеразы



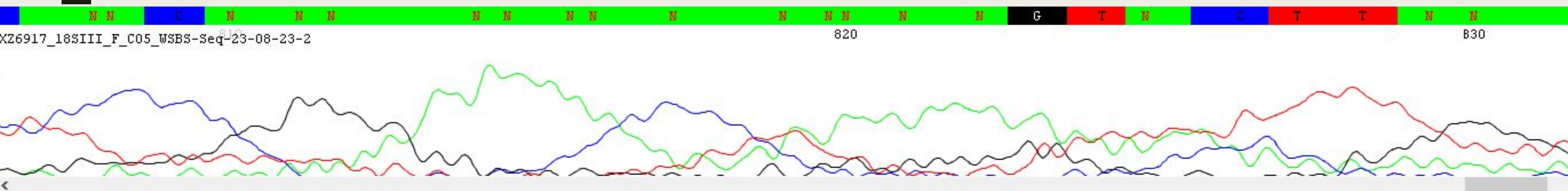
Проблемы в начале

- Короткие неспецифические участки, на которые могут отжигаться праймеры + короткие фрагменты могли не переосадиться после секвенсовой реакции



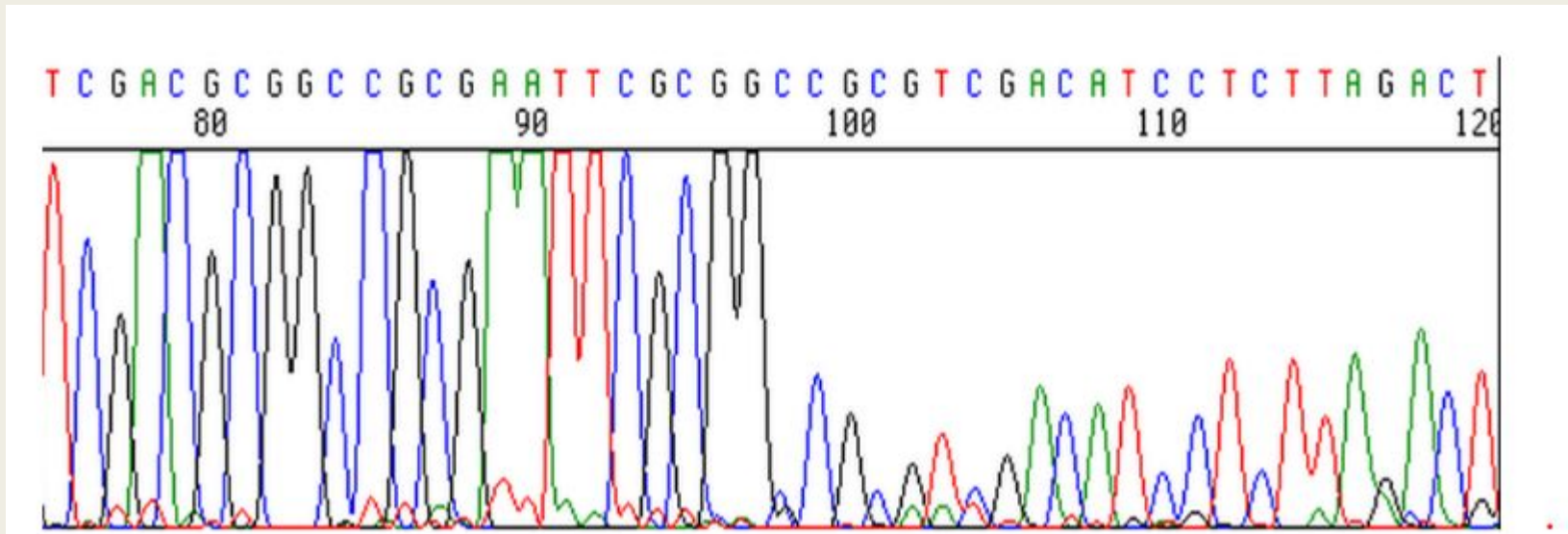
Проблемы в конце

- Фрагменты длиной 1000 и 1001 нуклеотидов на фореже различаются хуже, чем фрагменты длиной 51 и 51 нуклеотид



Резкое падение силы сигнала

- Может быть сложная структура в ДНК



Преждевременное падение силы сигнала

- Загрязнение препарата, примеси солей, ...

