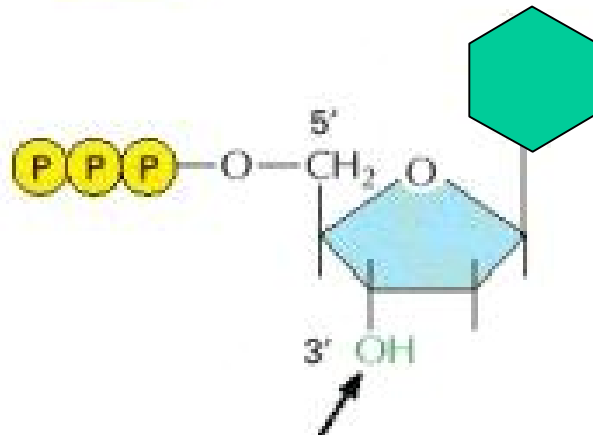
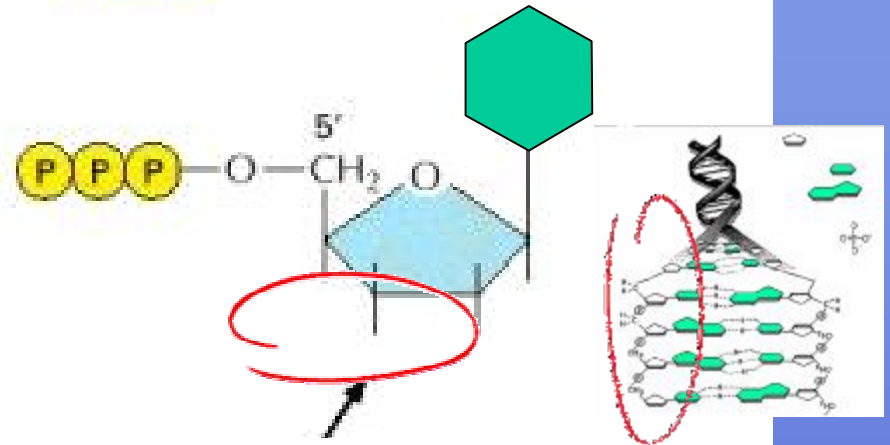


«1D_{ДНК} > 1D_{ДНК}»

деоксирибонуклеозид трифосфат

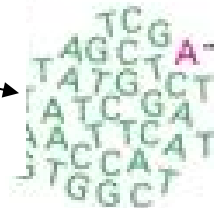


дидеоксирибонуклеозид трифосфат



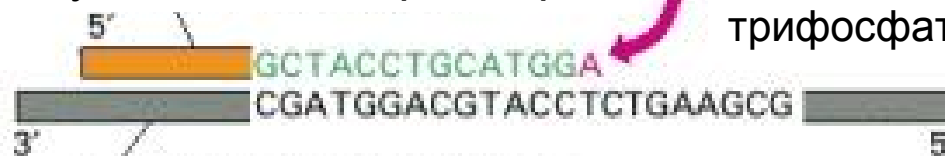
необходим для роста цепи (элонгации) **ЭЛОНГАЦИЯ ХИМИЧЕСКИ НЕВОЗМОЖНА**

деокси-
рибонуклеотид
фосфаты



небольшое количество
дидеоксирибонуклеозид
трифосфата

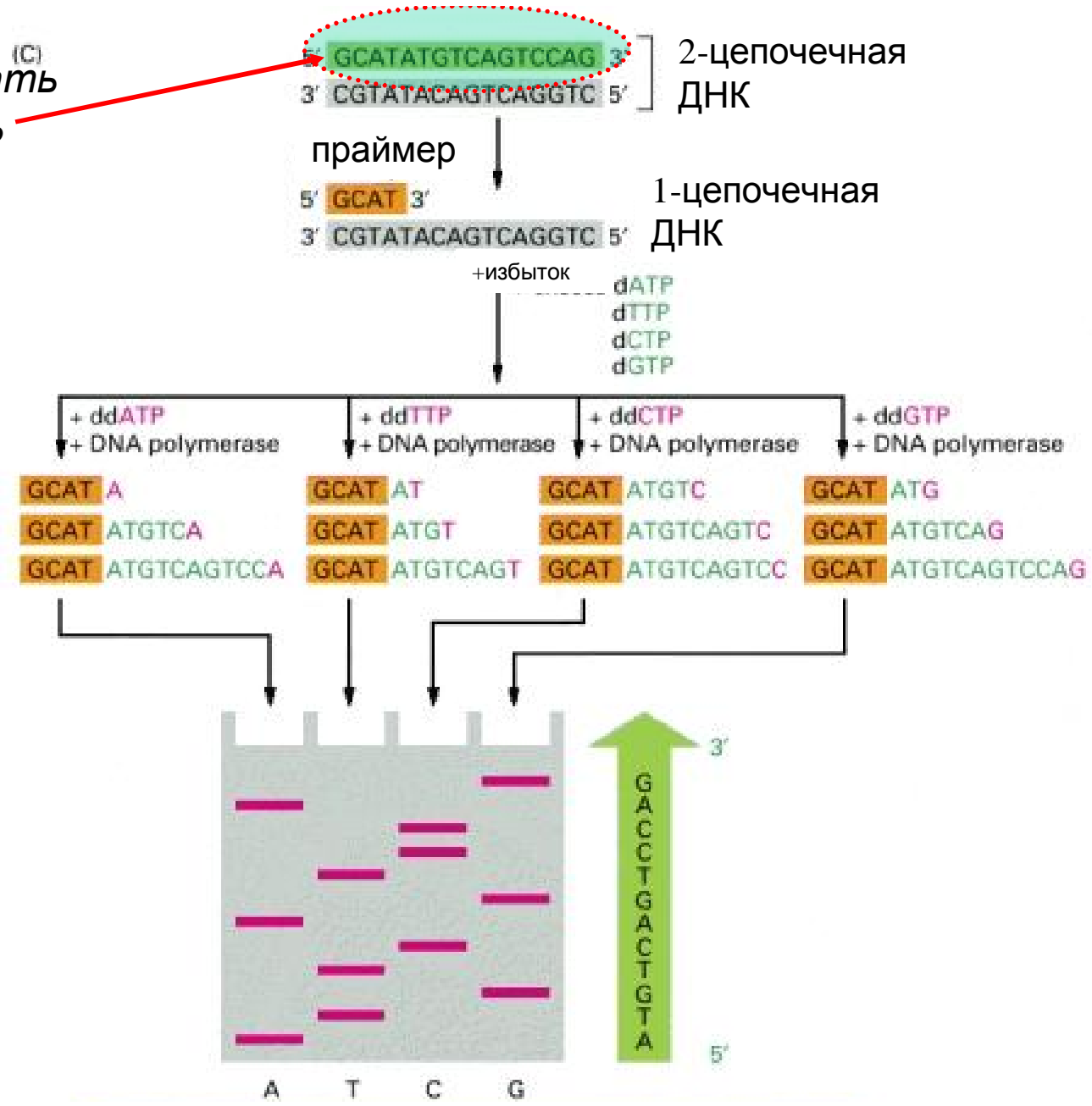
олигонуклеотидный праймер



включение дидеокси-
рибонуклеозид
трифосфата

секвенируемая ДНК

(C)
будем читать
эту цепь

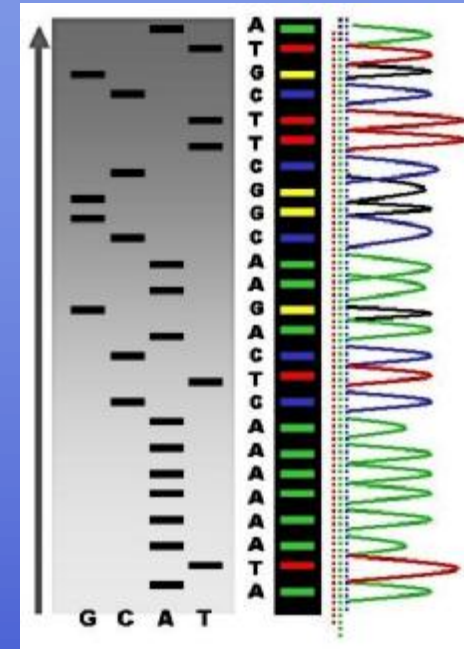


чтение последовательности ДНК посредством гель-электрофореза

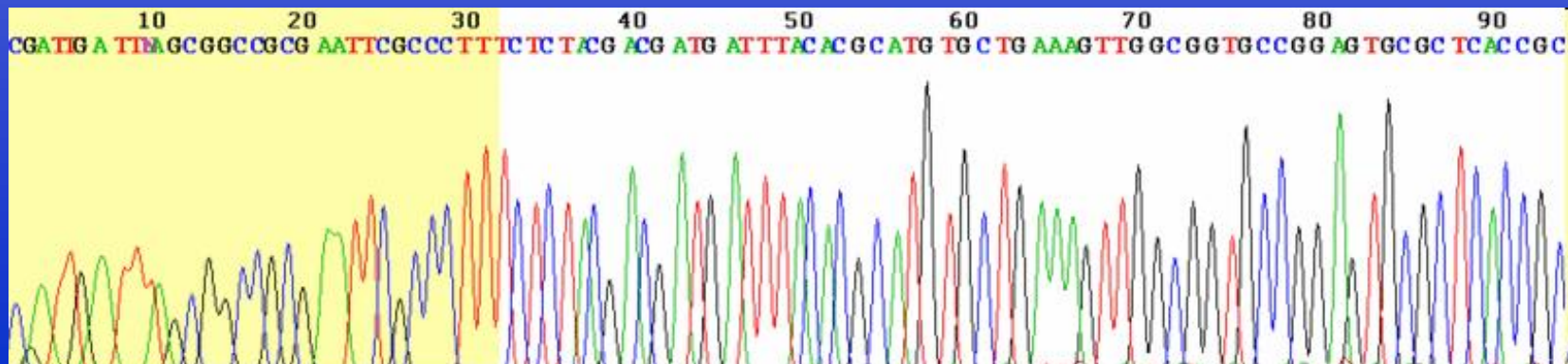


Ошибки секвенирования

- ДНК полимераза в контексте клетки – 10^{-9}
- Современное секвенирование – $10^{-3}..10^{-5}$



радиоактивные метки флуоресцентные метки



Общая теория алгоритмов
выравнивания



Алгоритмы выравнивания

Точные

**Установление и
устранение
противоречивости**

**Установление
соответствий ДНК-
РНК-белок**

Быстрые

**Поиск по базам данных
последовательностей**

**Задачи классификации
последовательностей**

«Точные»

Динамическое программирование

Принцип оптимальности

$$\left. \begin{array}{l} F_{0j} = d * j \\ F_{i0} = d * i \end{array} \right\} d - \text{штраф за вставки/делеции}$$

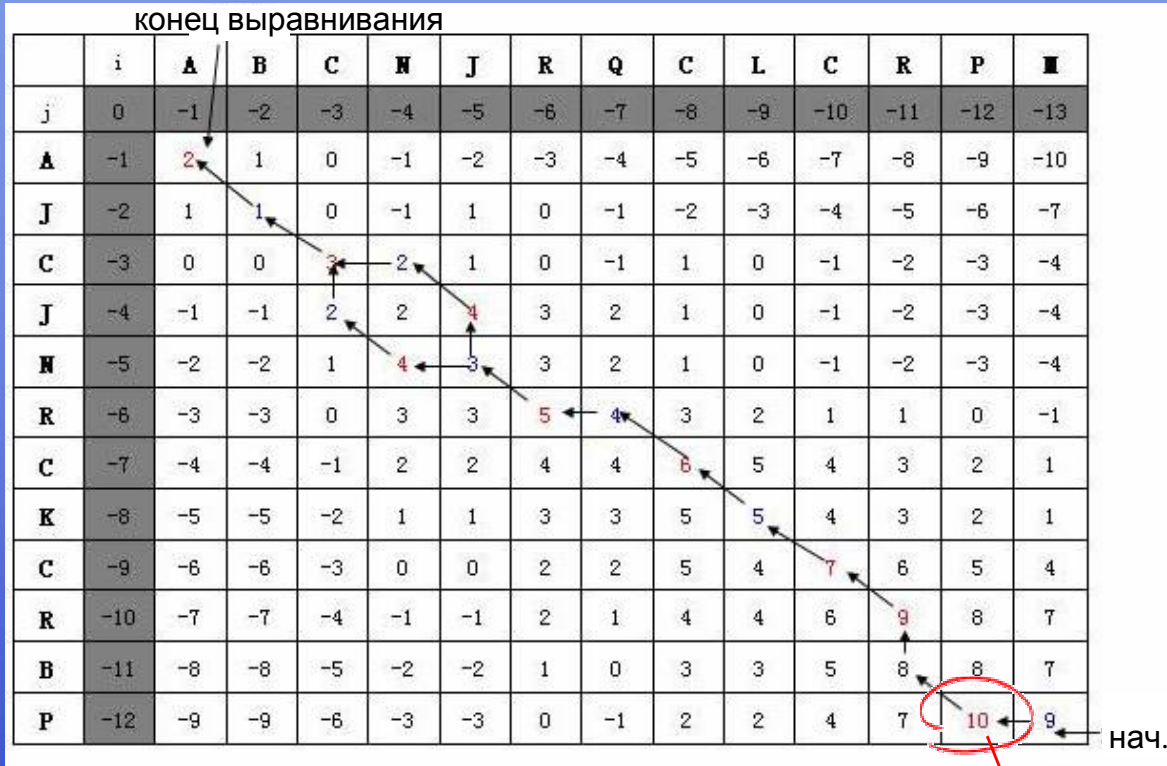
$$F_{ij} = \max(F_{i-1,j-1} + S(A_i, B_j), F_{i,j-1} + d, F_{i-1,j} + d)$$

из матрицы
схожести

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$F_{ij} = \max(F_{i-1,j-1} + S(A_i, B_j), F_{i,j-1} + d, F_{i-1,j} + d)$$

$$S(A,A)=+2, S(A,B)=-1, d=-1$$



A	B	C	N	J		R	Q	C	L	C	R		P	M
A	J	C		J	N	R		C	K	C	R	B	P	

$$2-1+2-1+2-1+2-1+2-1+2+2-1+2-1=9$$

A	B	C		N	J	R	Q	C	L	C	R		P	M
A	J	C	J	N		R		C	K	C	R	B	P	

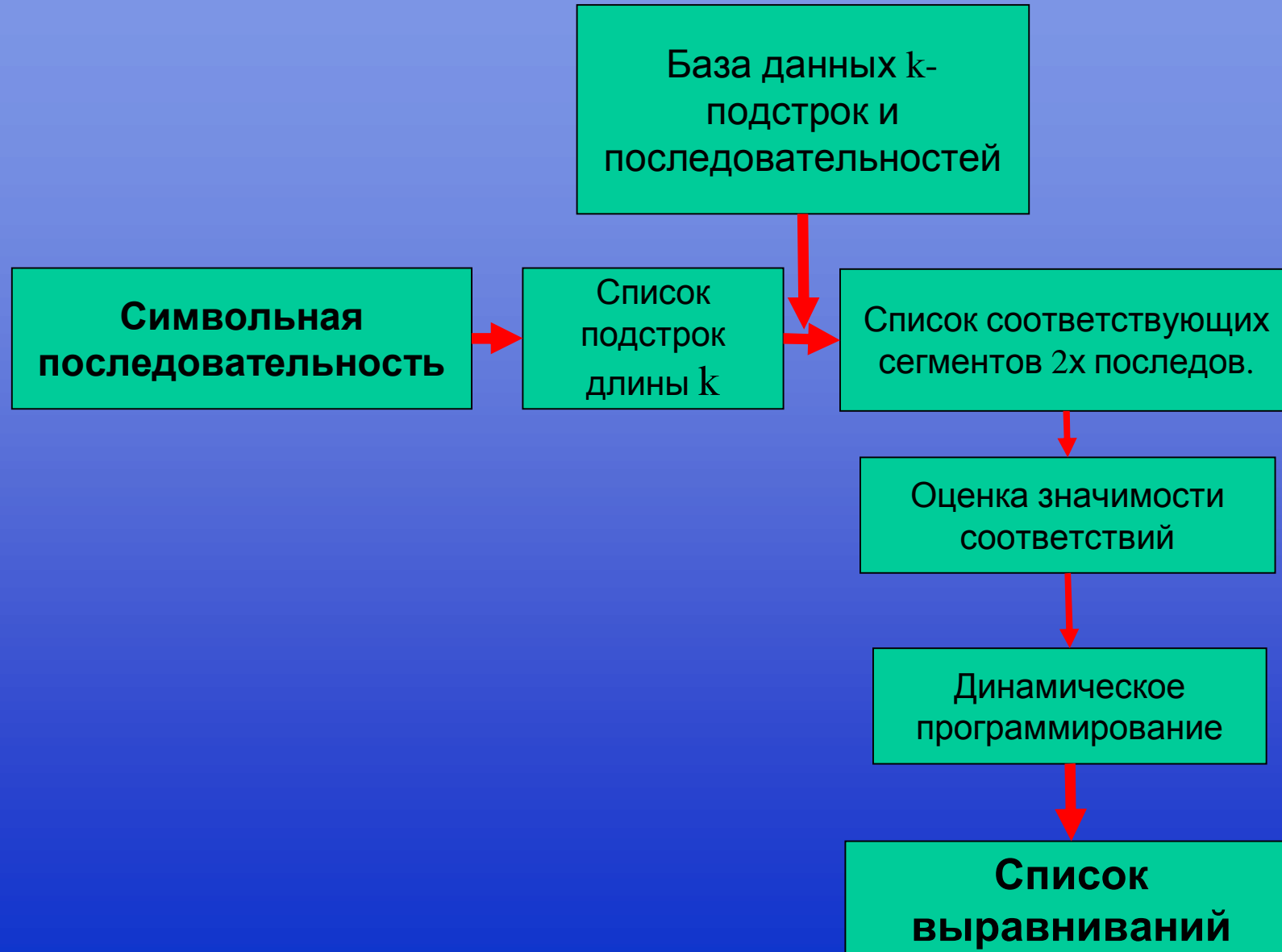
$$2-1+2-1+2-1+2-1+2-1+2+2-1+2-1=9$$

```

AlignmentA ← ""
AlignmentB ← ""
i ← length(A)
j ← length(B)
while (i > 0 and j > 0)
{
  Score ← F(i,j)
  ScoreDiag ← F(i - 1, j - 1)
  ScoreUp ← F(i, j - 1)
  ScoreLeft ← F(i - 1, j)
  if (Score == ScoreDiag + S(A(i), B(j)))
  {
    AlignmentA ← A(i-1) + AlignmentA
    AlignmentB ← B(j-1) + AlignmentB
    i ← i - 1
    j ← j - 1
  }
  else if (Score == ScoreLeft + d)
  {
    AlignmentA ← A(i-1) + AlignmentA
    AlignmentB ← "-" + AlignmentB
    i ← i - 1
  }
  otherwise (Score == ScoreUp + d)
  {
    AlignmentA ← "-" + AlignmentA
    AlignmentB ← B(j-1) + AlignmentB
    j ← j - 1
  }
}
while (i > 0)
{
  AlignmentA ← A(i-1) + AlignmentA
  AlignmentB ← "-" + AlignmentB
  i ← i - 1
}
while (j > 0)
{
  AlignmentA ← "-" + AlignmentA
  AlignmentB ← B(j-1) + AlignmentB
  j ← j - 1
}

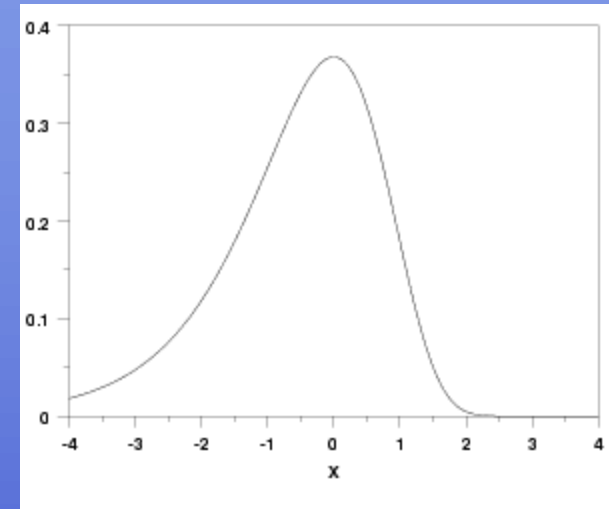
```


BLAST (Basic Local Alignment Search Tool)



3. Оценка значимости соответствий

- Отсев сегментов с низким весом
 - $S > S_{\text{п}}$, пороговое значение веса сегмента
- Значимость веса каждого сегмента
 - E – вероятность ложно- позитивного совпадения в БД из D последовательностей
 - λ, K, H – параметры (0.318, 0.13, 0.4)
 - m', n' – эффективные длины последовательностей Q и DB



распределение
Гумбеля

$$p(S \geq x) = 1 - \exp\left(-e^{-\lambda(x-\mu)}\right)$$

$$\mu = \frac{[\log(Km'n')]}{\lambda}$$

$$m' \approx m - \frac{(\ln Kmn)}{H}$$

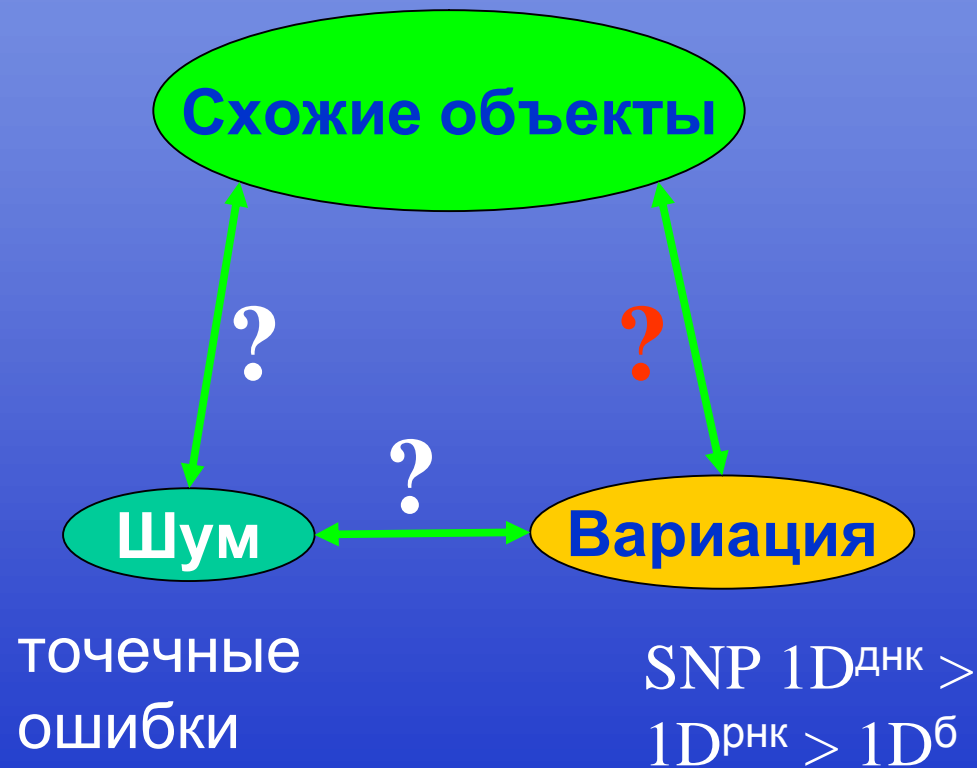
$$n' \approx n - \frac{(\ln Kmn)}{H}$$

$$E \approx 1 - e^{-p(s>x)D}$$

Как снизить ошибки при секвенировании?

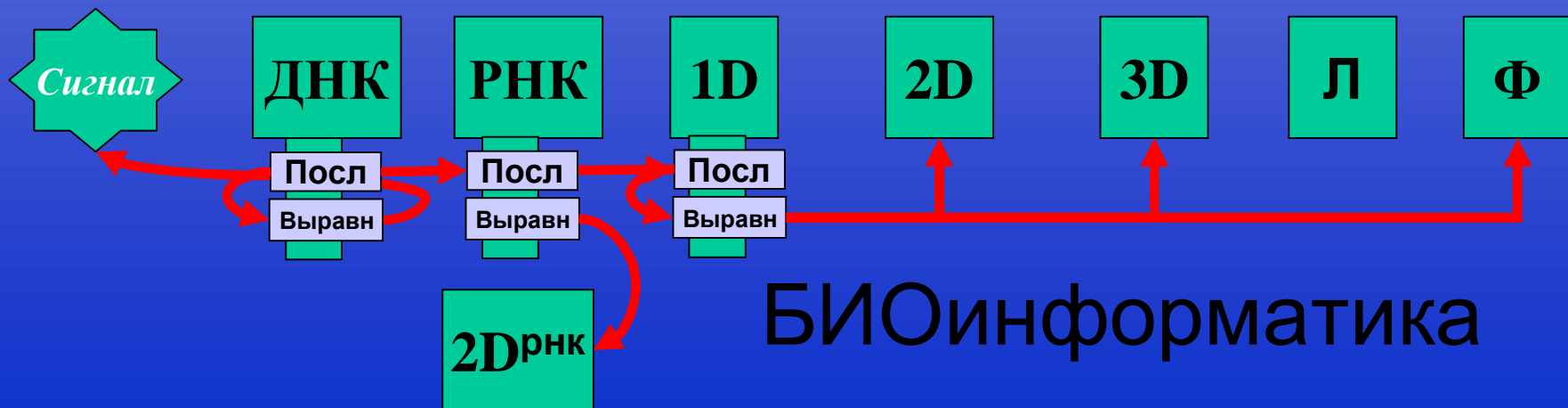
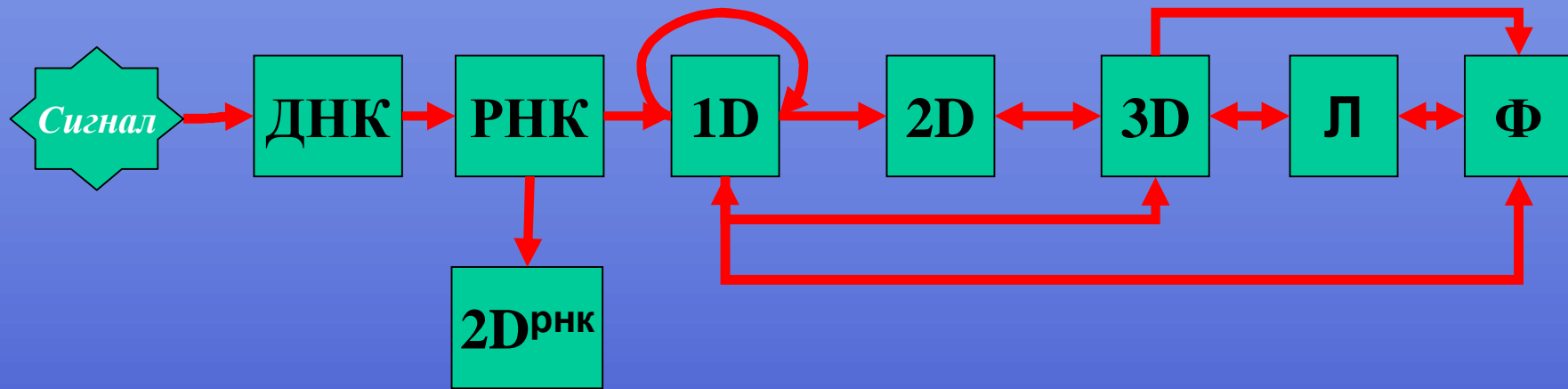
- Улучшение процедур распознавания профилей секвенирования
- Повторные эксперименты на одном объекте, одном приборе, выравнивание последовательностей
- Сопоставление 1D РНК и 1D белка, соответствующих данной 1D ДНК (**ген**)

Вариации объекта и «схожесть» объектов



Подход к решению задач 1D>...

биоИНФОРМАТИКА



БИОинформатика