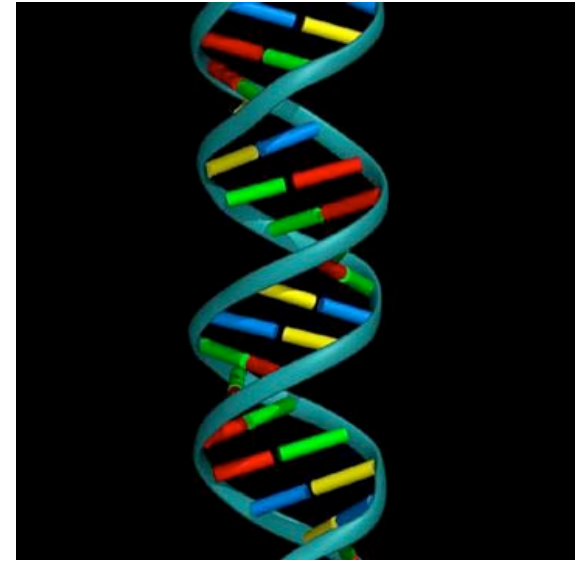
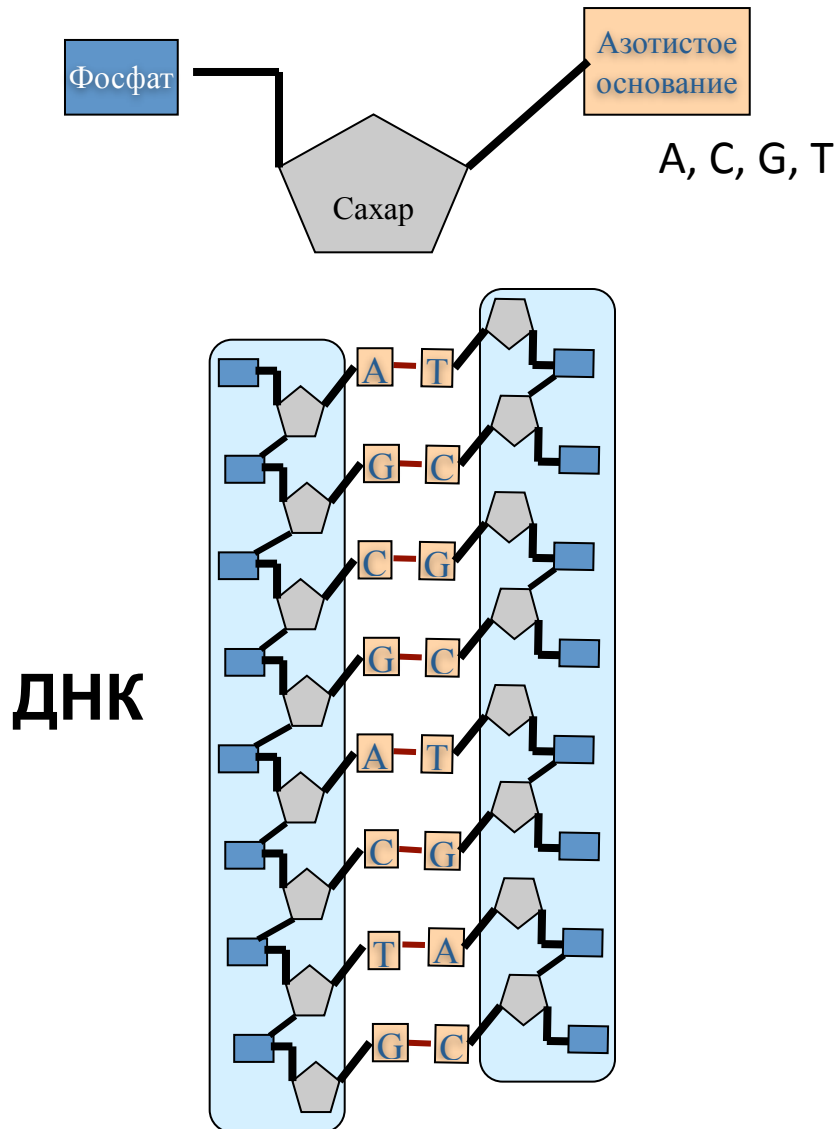


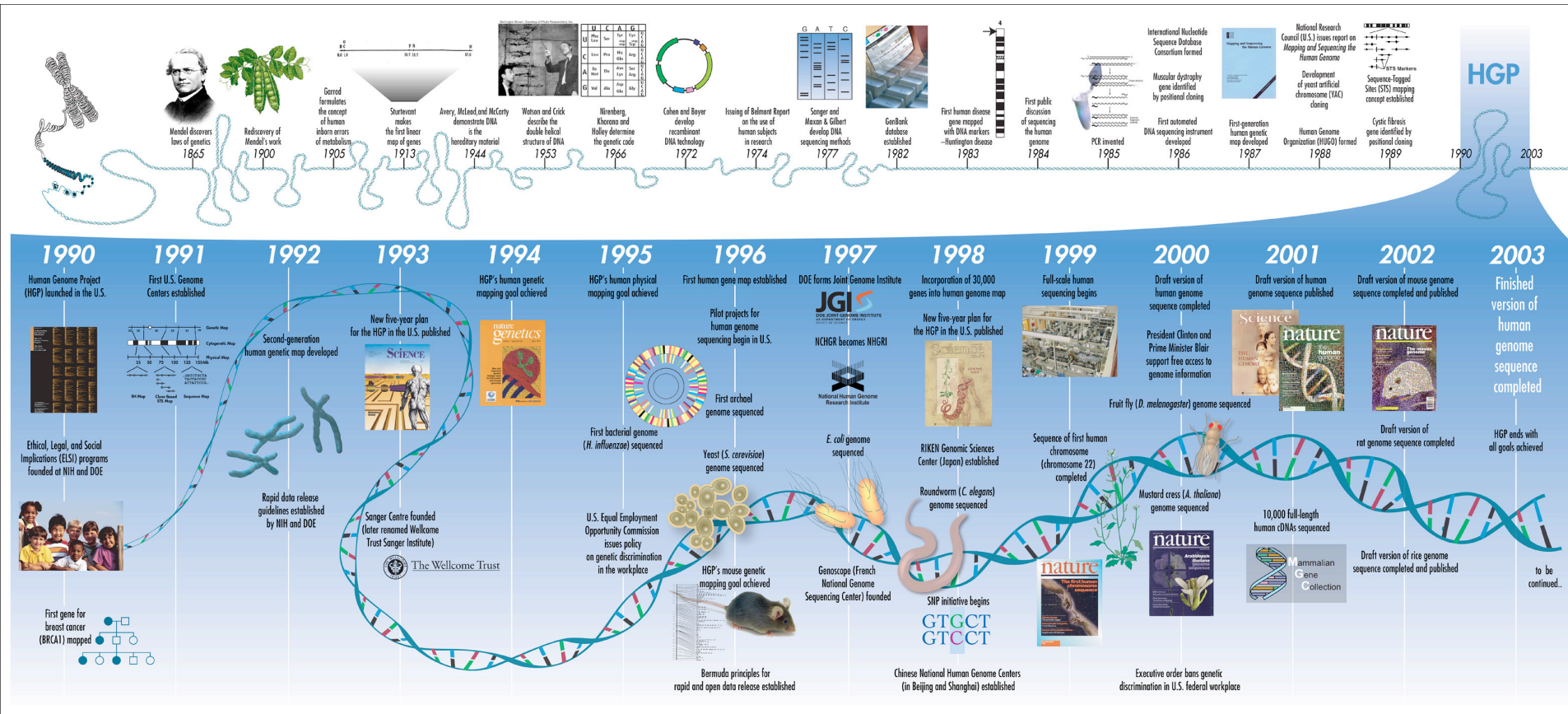
Введение: основы молекулярной биологии

Структура двойной спирали ДНК



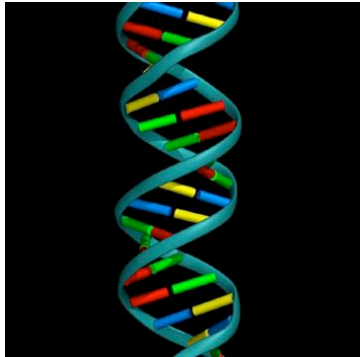
Courtesy of Cold Spring Harbor Laboratory Archives. Noncommercial, educational use only.

Основные вехи развития молекулярной биологии и генетики. Геном человека



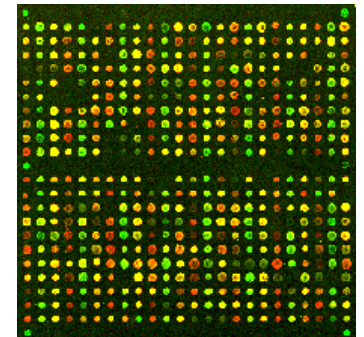
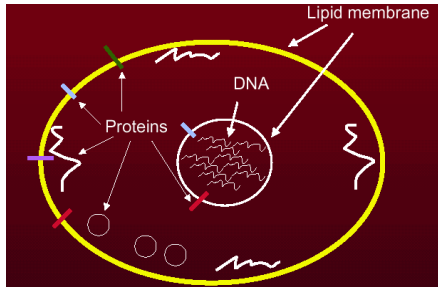
21 век: развитие технологий приводит к информационной революции в биологии и медицине

- Секвенирование (расшифровка) ДНК



AGTAGCACAGACTACGACGAGA
CGATCGTGCGAGCGACGGCGTA
GTGTGCTGTACTGTCGTGTGTG
TGTA CTCTCCCTCTCTAGTCT
ACGTGCTGTATGCGTTAGTGTC
GTCGTCTAGTAGTCGCGATGCT
CTGATGTTAGAGGATGCACGAT
GCTGCTGCTACTAGCGTGCTGC
TGCGATGTAGCTGTCGTACGTG
TAGTGTGCTGTAAGTCGAGTGT
AGCTGGCGATGTATCGTGGT

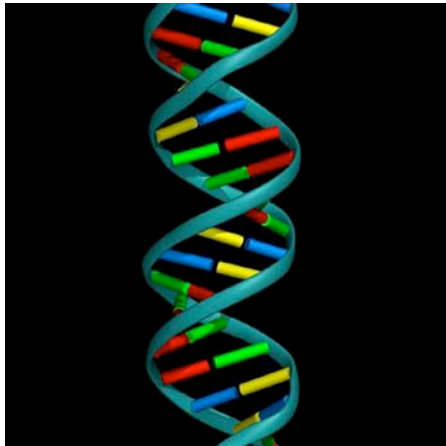
- ДНК-микрочипы: измерение экспрессии генов



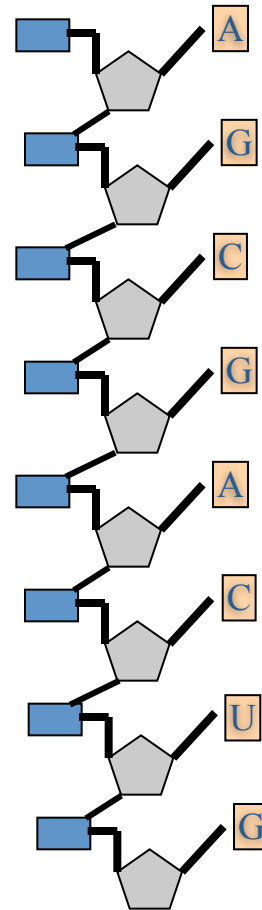
ДНК → РНК → белок ("Central dogma")

РНК: промежуточный переносчик информации

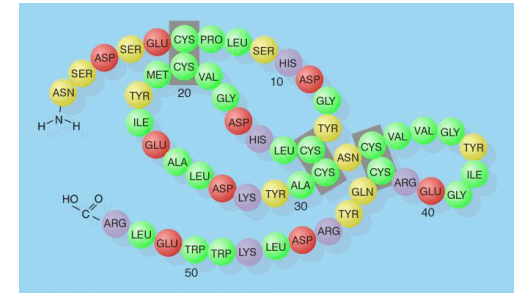
ДНК, длина у человека $\sim 3 \times 10^9$
Содержит $\sim 22,000$ генов



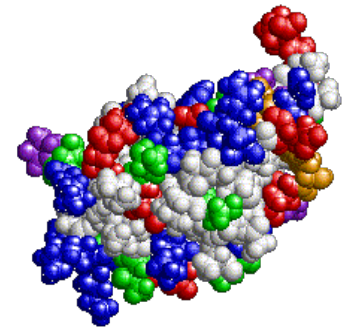
транскрипция



→ трансляция

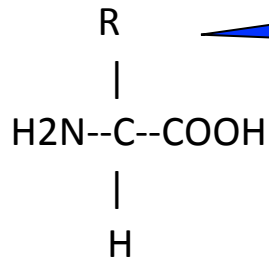


 folding



Белки

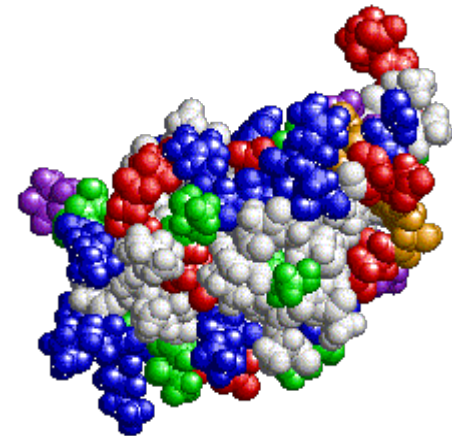
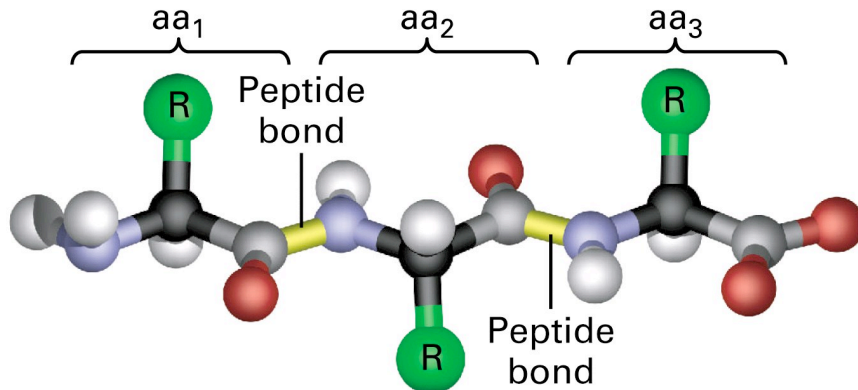
Полимер, состоящий из аминокислот



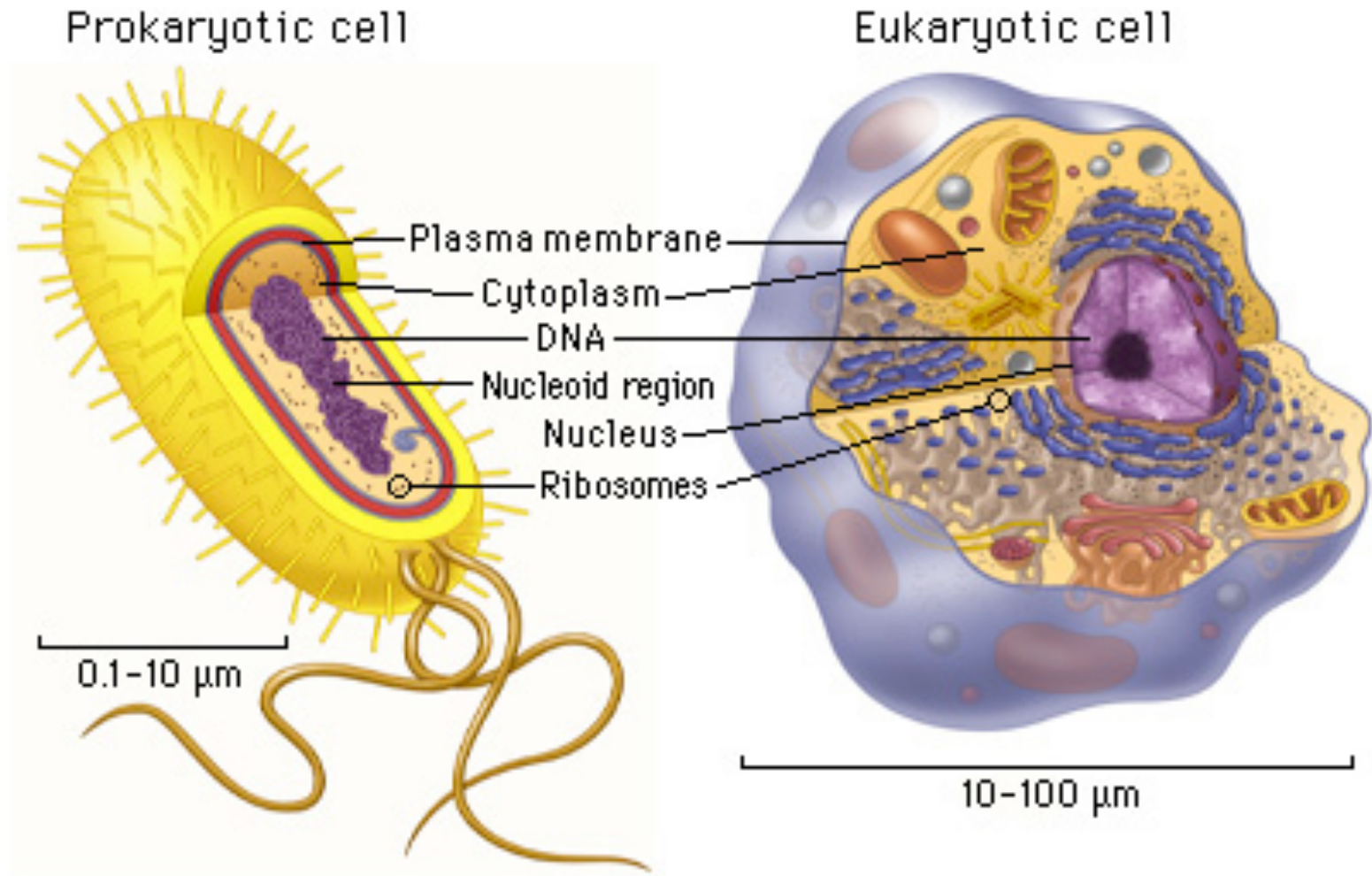
20 возможных оснований

Аминокислотная последовательность
сворачиваясь образует сложную
трехмерную структуру.

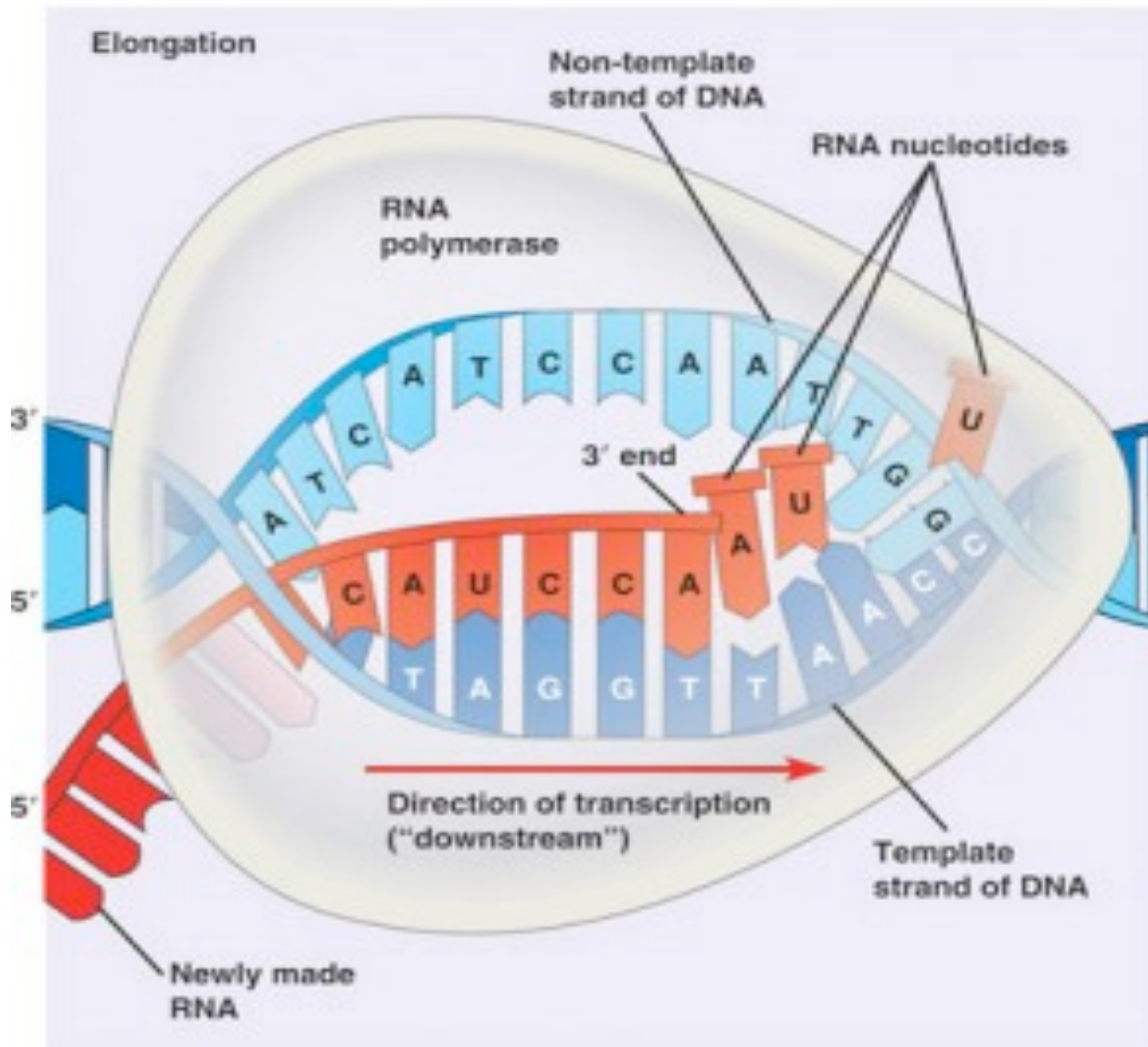
Пространственная структура белка
определяет его функцию.



Все живые организмы состоят из клеток



Транскрипция ДНК



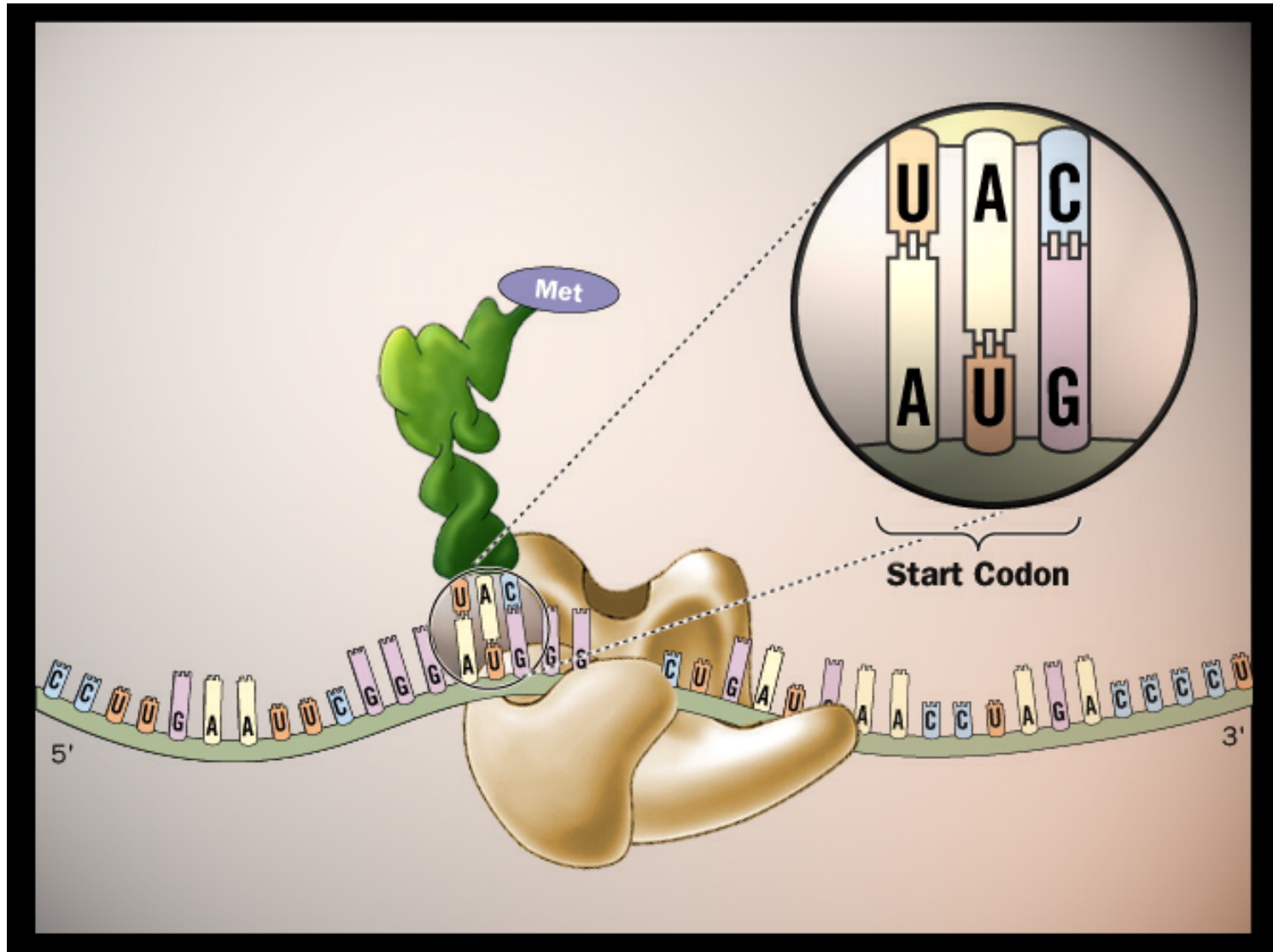
Транскрипция ДНК

<http://www.sector3.iitp.ru/video/transcription.mov>

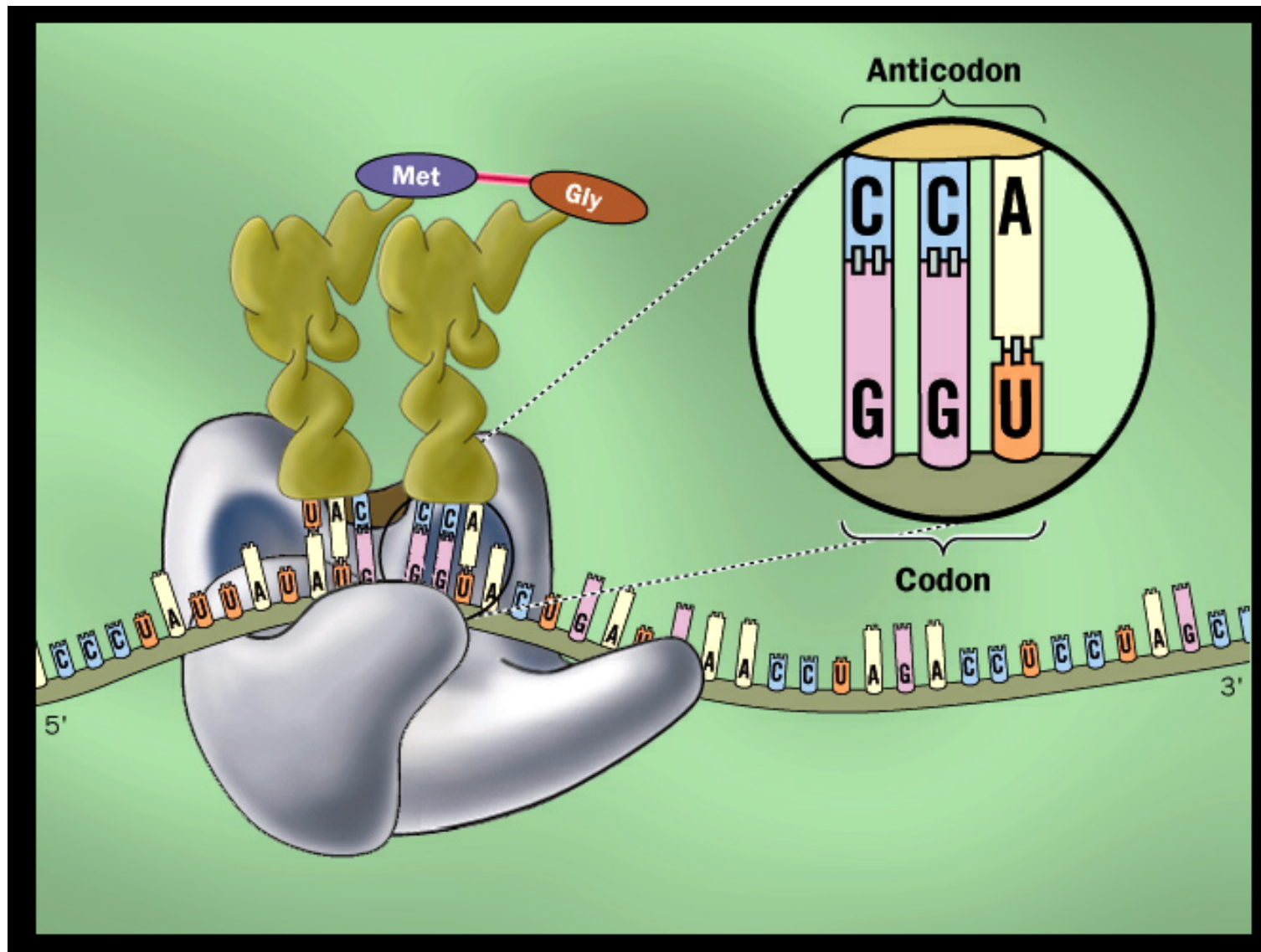
Генетический код

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	Third letter	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }		U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }		U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }		U C A G

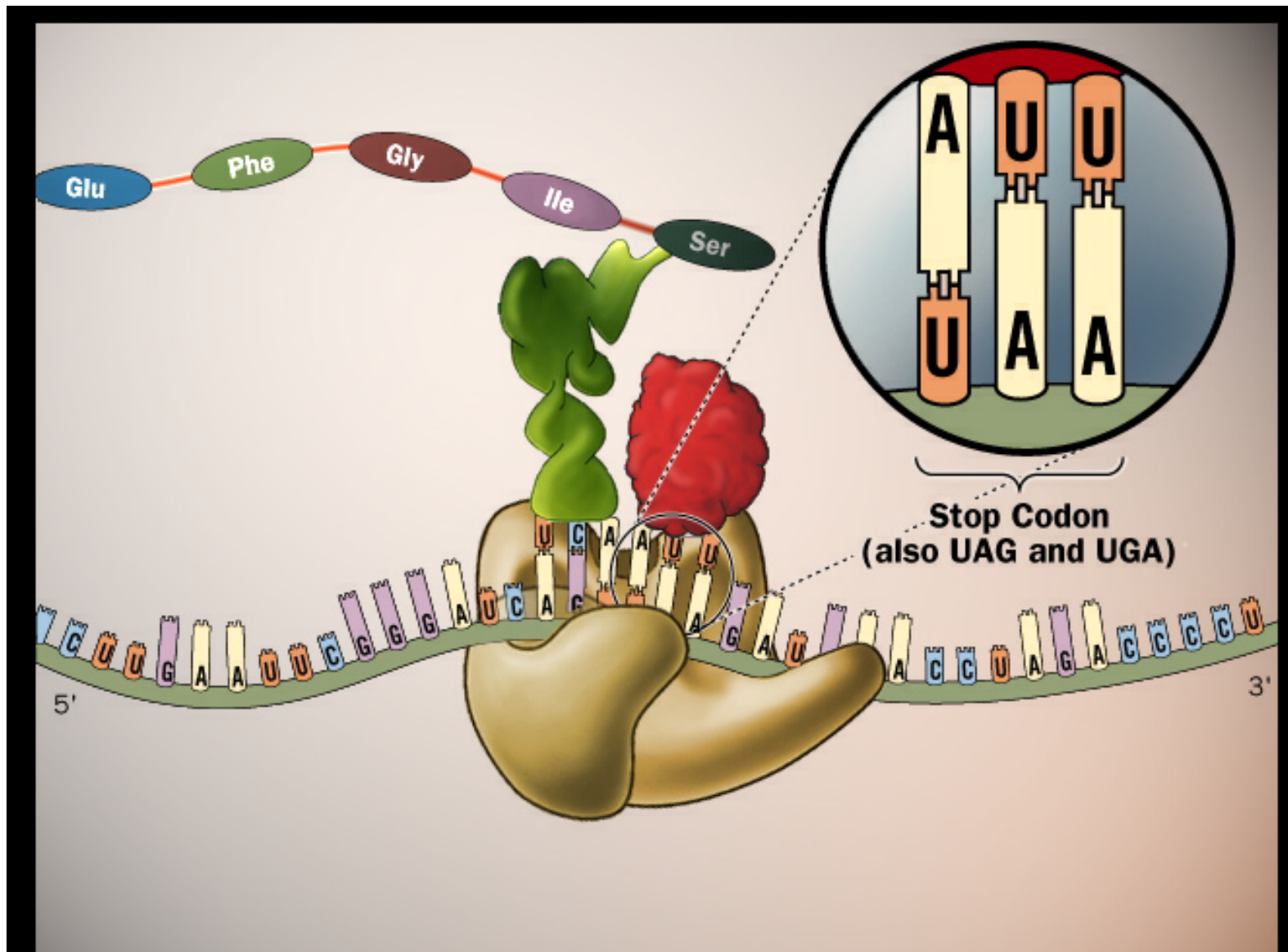
Трансляция: инициация



Трансляция: элонгация



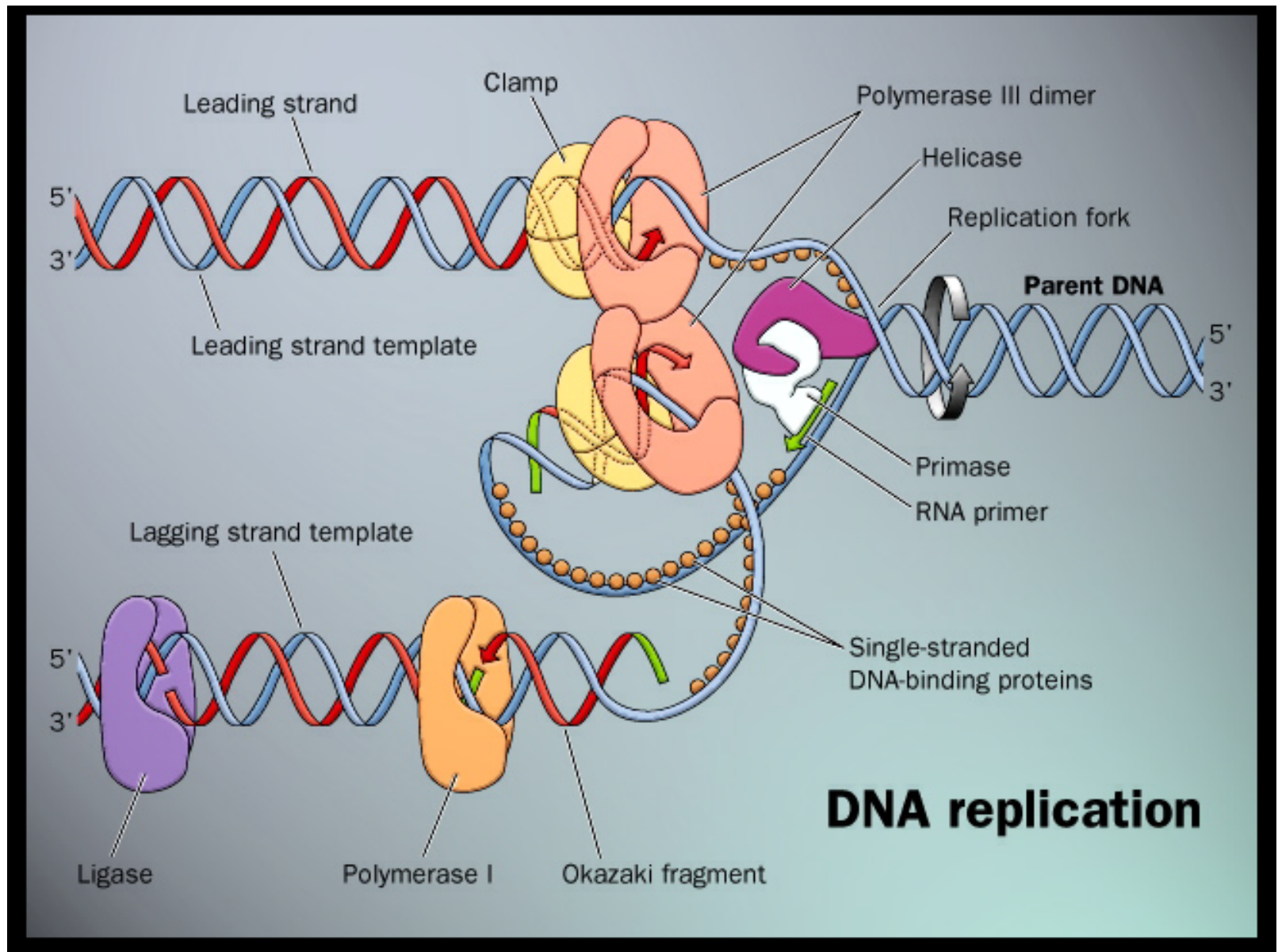
Трансляция: терминация



Трансляция

<http://www.sector3.iitp.ru/video/transcription.mov>

Репликация ДНК



Репликация ДНК

<http://www.sector3.iitp.ru/video/replication.mov>

Репликация ДНК

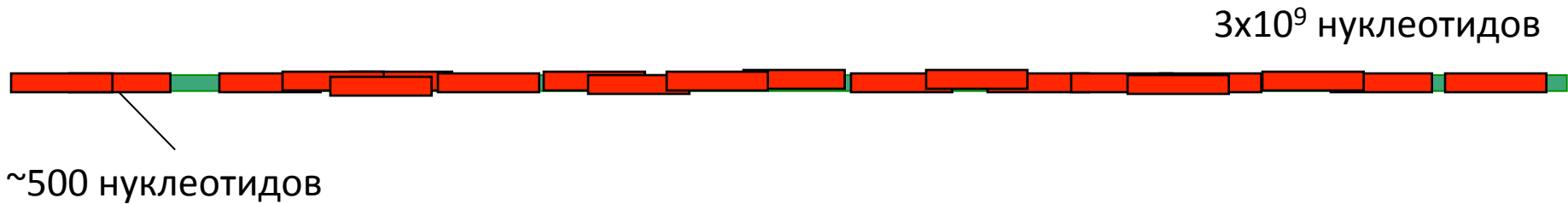
http://www.sector3.iitp.ru/video/DNA_Replication_Real_Time.mov

Основные процессы в клетке

http://www.sector3.iitp.ru/video/Journey_Inside_The_Cell.mov

Примеры задач биоинформатики

1. Секвенирование – сборка геномов



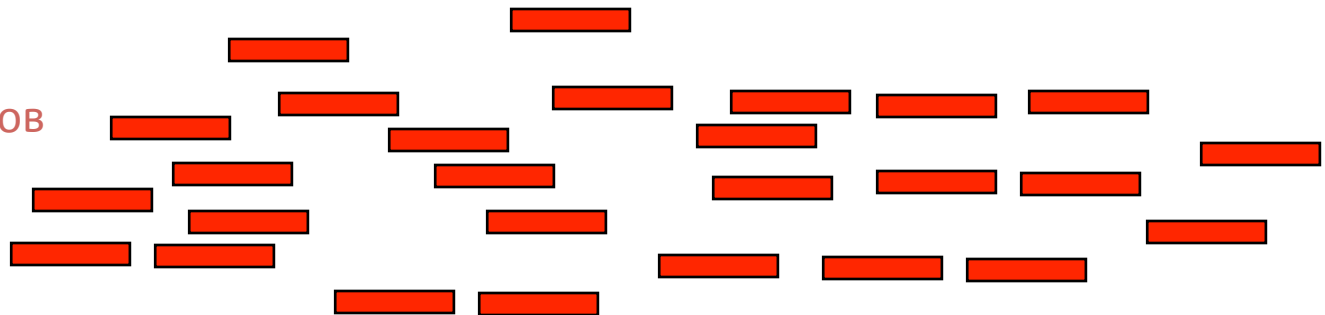
Примеры задач биоинформатики

1. Секвенирование – сборка геномов



3×10^9 нуклеотидов

Большой пазл
~60 миллионов кусков



Автоматизированная сборка геномов

Постановка проблемы ~1980

1995: сборка ДНК длиной до 1,000,000 нуклеотидов

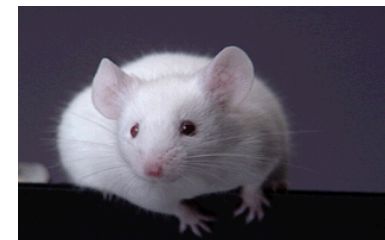
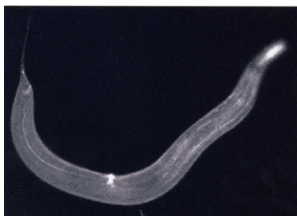
2000: сборка генома человека

Секвенированные полные геномы



На данный момент
секвенировано более 1000
полных геномов

(большинство из них -
бактериальные геномы)



2. Распознавание генов

> HSKIIIBE, Human gene for casein kinase II subunit beta (EC 2.7.1.37).
ggggctgagatgtaaattagaggagctggagaggagtgcctcagagtttgggttgctttaagaaaggg
ggttcogaattctcccggtgttgaggggccgaatgtgggaggaggaggataccagaggcaggaagga
gaacttgagctttactgacactgttcttttctagctgacgtgaagatgagcagctcagaggaggtgtc
ctggatttcctggttctgtgggctccgtggcaatgaattcttctgtgaagtgagttctcttcaacctcc
ctacttgccagcttcacatatcttcccaccagacgttcccttcacatatccactctacactgttctct
aaagcttttatgggagagagtgtaggtgaactaggagagacacaagtactt
agaaacaagcacaacagatgcagttgtgttgatgataaggcatcacttagag
agatgaggattttgatatgggttccctcttggttccatgtcctgacaggtg
ggacaaatttaactcttactggactcaatgagcaggtccctcactatcgacaa
ggacctggagcctggtgaggcaccctcagggttgttttgtgtgtgtgcgtgc
atctctattcacttgccgtgaattttgccaaatttcccttggttctctgattt
tgctttattttgatcctccacctgactcttgtctagttttgtgacgtatct
tgcaagggtcagaagcccaggtttctgggtcccatgccagatgttgatgg
ggtgctaggcaaactgaatagcccgcagcccctggatatgggcagggcacct
agtagttgcatttgccggggtgtggttcagatgaagaactggaagacaacc
attgagcaggcagccgagatgctttatggattgatccacgcccgcctacatcc
gccagatggtgaggcctctctgctcctacctgcctcctctctgagcagtaagagacacaggttcttgc
gcaagaagtcatgtttaagccctgtttaaggaagctagctgagaagaggggaagaaccccagaacttgg
ccctgccctaatttggaagaaaggcaacacagaagtttgagagcccatctagtccagagaaggggacct
ctggacagagtttggaaggagtgcgcagacagagttgggtatgggttgggctgcgaaggaggttgccctctct
ttacatctacctgccaaaccttccattgtattcacctcagttggaaaagtagcagcaaggagactttg
gttactgtcctcgtgtgtactgtgagaaccagccaatgcttccattGgtgagtggtgaagaagggaaa
ggaaagcacctgtgtggcagttcttatgggaaggagtgggggtcaacacattggagcctgagtcctgagg
ggaggttaggttaggaatagggggataacctggcctgctgagtcctggctgtctcccaggccttccagacat
cccagggtgaagccatggtgaagctctactgcccgaagtgcattggatgtgtacacaccaagtcataag
acaccatcacacggatggcgccctacttcggcactggtttccctcacatgctcttcatggtgcattccga
gtacogggcccaagagacctgccaaccagtttgtgcccaggtagggagcagggagagtcattaaagggtca
aaggaaaggcccaagatccccagagaggggaggacagggcatggcccttcttgaggtctgcttctcc
cagaatcagggcctctccctgctgagtgcagctgggaaagtatttgattatctgtgcttgagttacct
tattgtagaatgttcttgagctgagaagtgggaaccacagaggtttagctctgagcaggtccatagag
gagctcaggtggggaggtgggaatgcaggtgactggcagggcctggatggggctcatgctgctgcctct
ctgacctctgccctggcctaggctctacggtttcaagatccatccgatggcctaccagctgcagctcca
agccgccagcaacttcaagagcccagtcagacgattcgctgatccctccccacctgtcctgcagtc
tttgtcttttcccttcttttttgccaccttccaggaacctgtatgggttttagtttaaataaagga
gtcgttatcgtggtgggaatatgaaataaagtagaagaaaaggccatgagctagctgctggtgcttgc
ggaaggggtggagcgtggccatggaaatcgggctccacggcccagggtatgg

Как определить границы генов?

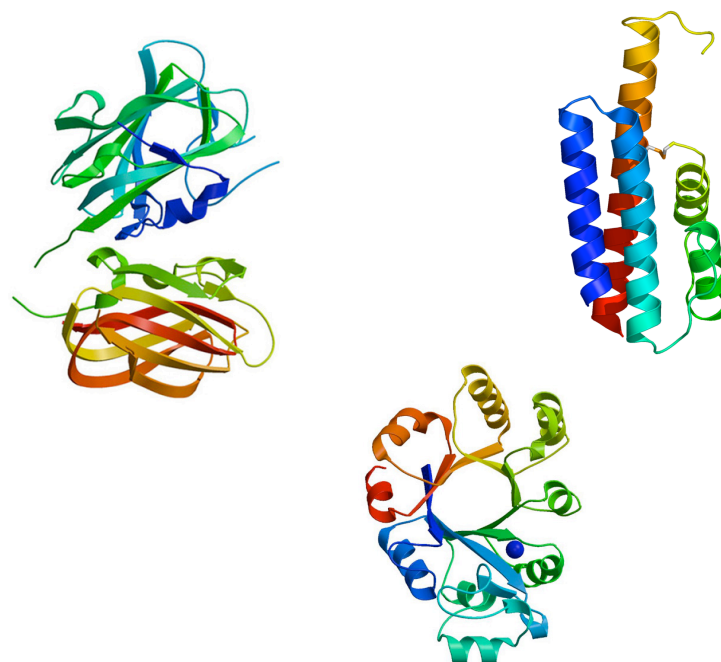
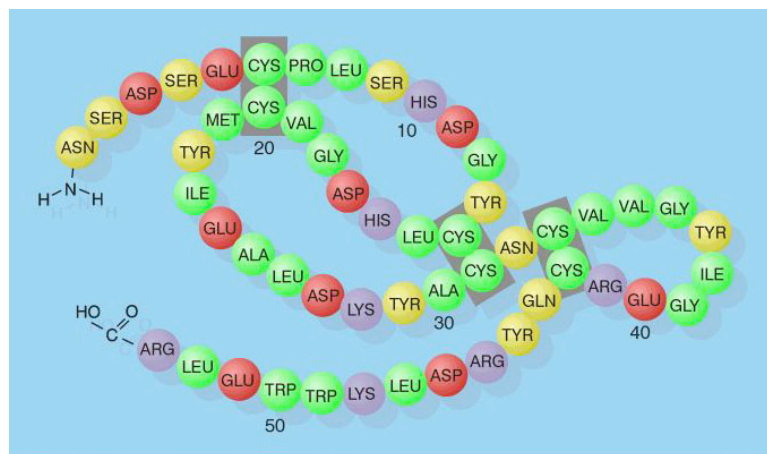
У человека:

~22,000 генов

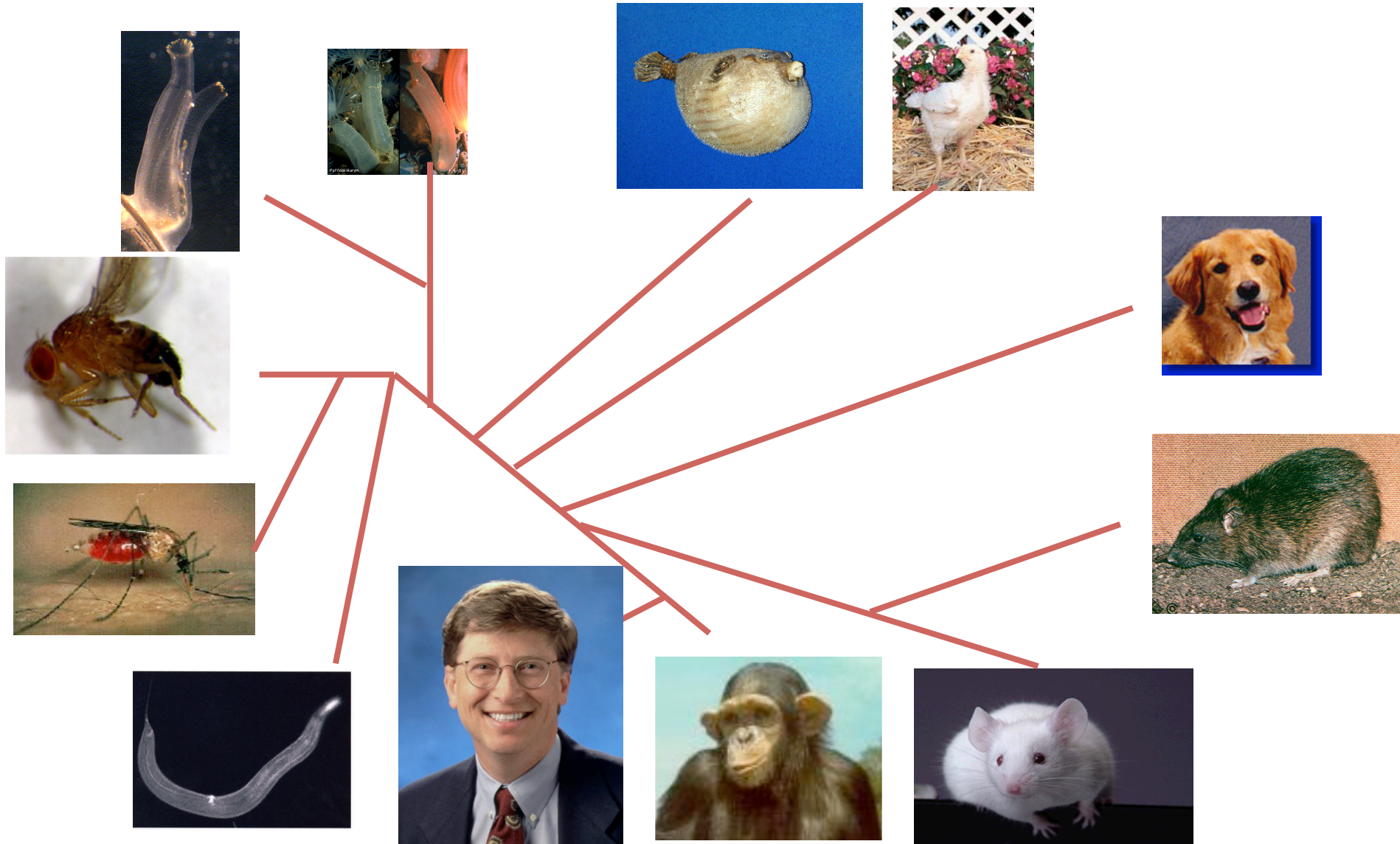
~1.5% ДНК

3. Сворачивание (folding) белков

- Аминокислотная последовательность белка определяет его тримерную структуру
- Пространственная структура белка определяет его функцию
- Можем ли мы предсказать пространственную структуру белка зная его аминокислотную последовательность?
 - Задача с 35-летней историей
 - Молекулярная динамика, методы машинного обучения, вычислительная геометрия

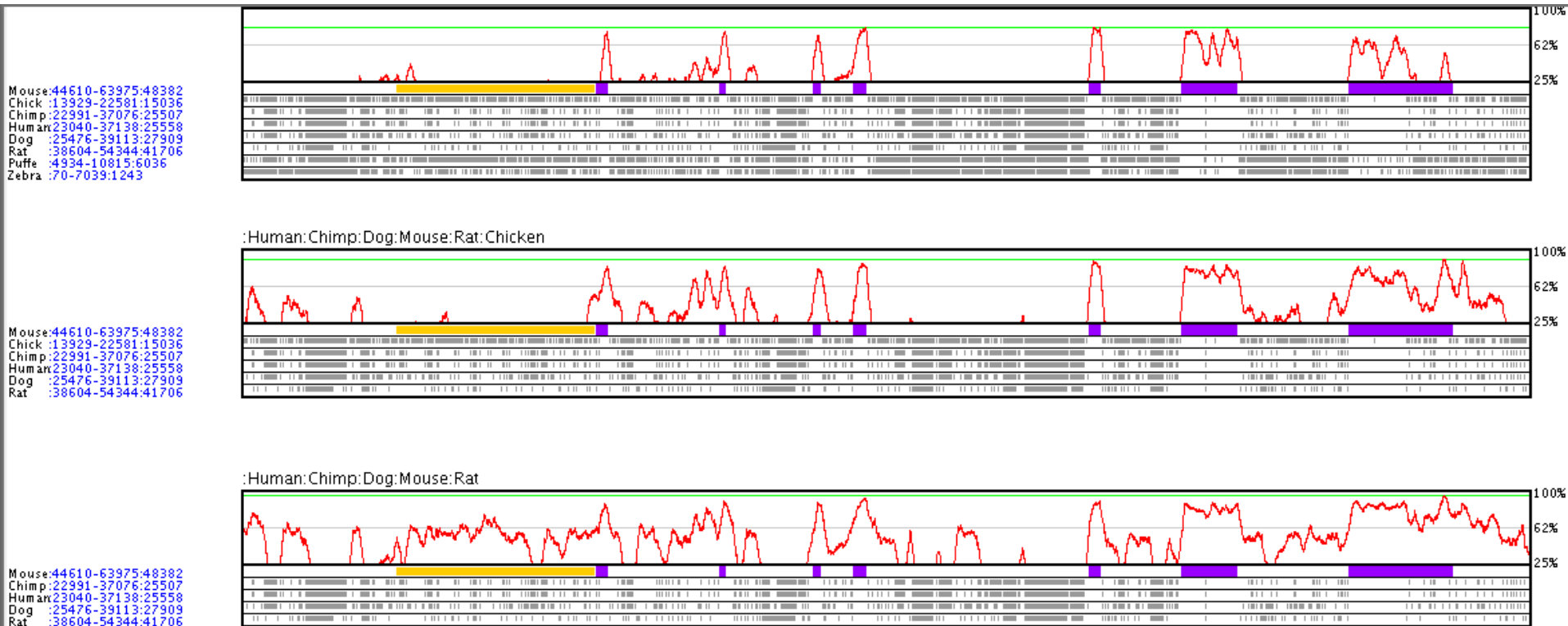


4. Эволюция



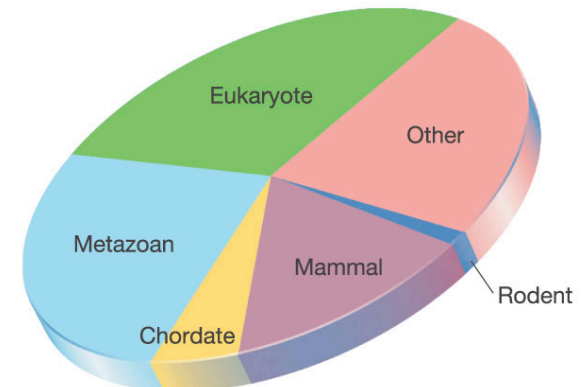
5. Сравнительная геномика

Консервативность последовательности – наличие функции



Сравнение биологических последовательностей применяется для

- поиска генов
- определения функции генов
- изучения эволюционных процессов



Сравнение последовательностей — выравнивание

AGGCTATCACCTGACCTCCAGGCCGATGCCC

TAGCTATCACGACCGCGGGTCGATTTGCCCGAC

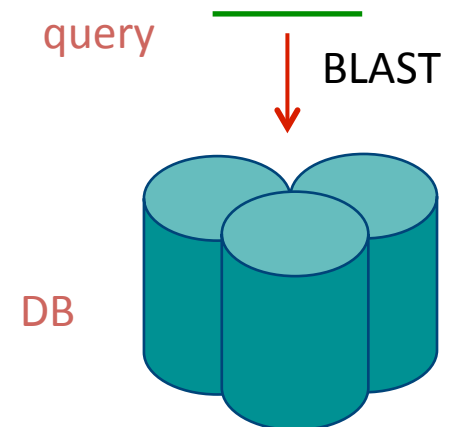
—AGGCTATCACCTGACCTCCAGGCCGA—TGCCC—
|
TAG—CTATCAC—GACCGC—GGTCGATTTGCCCGAC

Выравнивание последовательностей:

Постановка проблемы ~1970

BLAST: 1990, наиболее цитируемая статья

По-прежнему активная область исследований



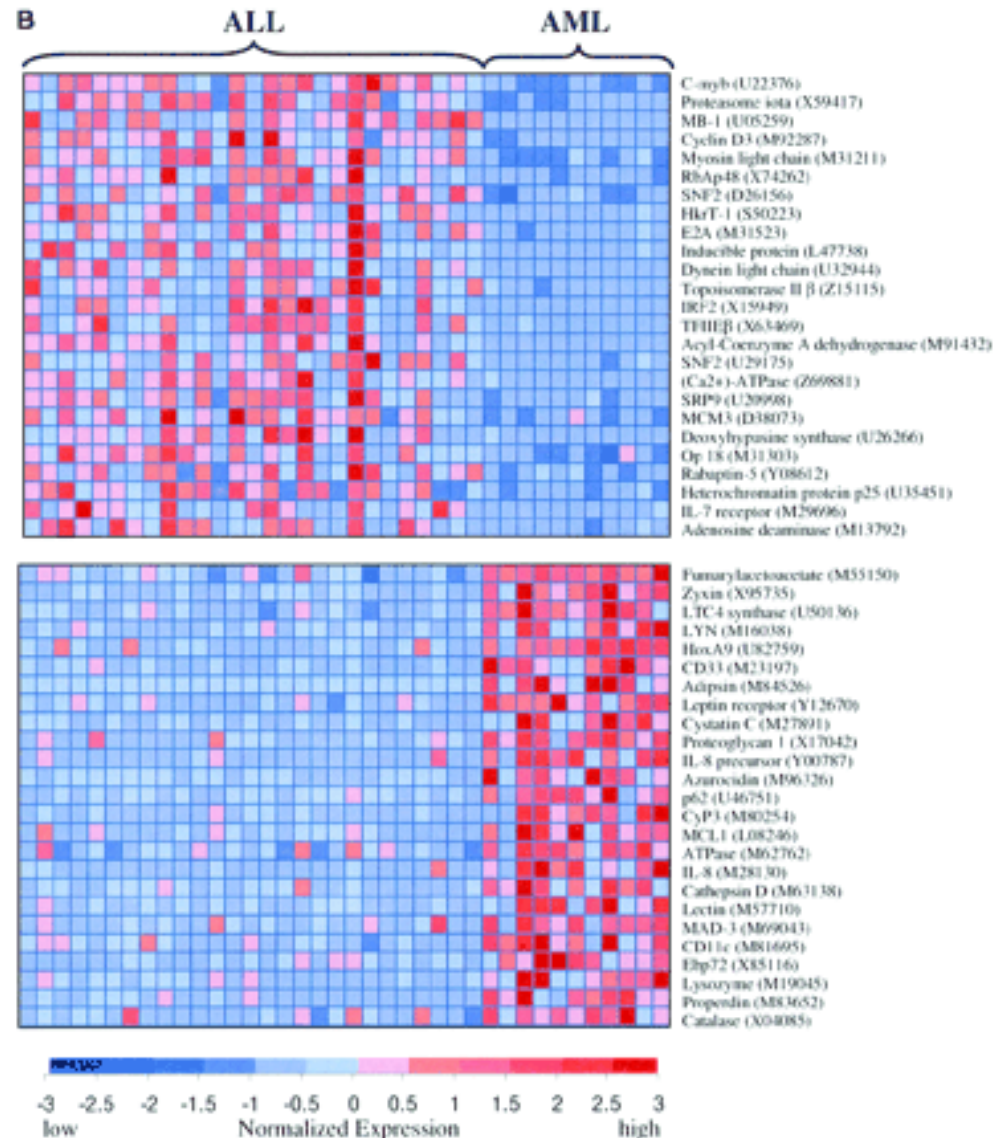
6. Кластеризация результатов ДНК-микрочипирования на примере предсказание типа лейкомии

- 2 типа лейкомии
 - Лимфоцитарная лейкомия (ALL)
 - Миелоидная лейкомия (AML)
- Тип определяет методы лечения
- Возможно ли предсказать тип до лечения?

Пробы костного мозга: ALL vs AML

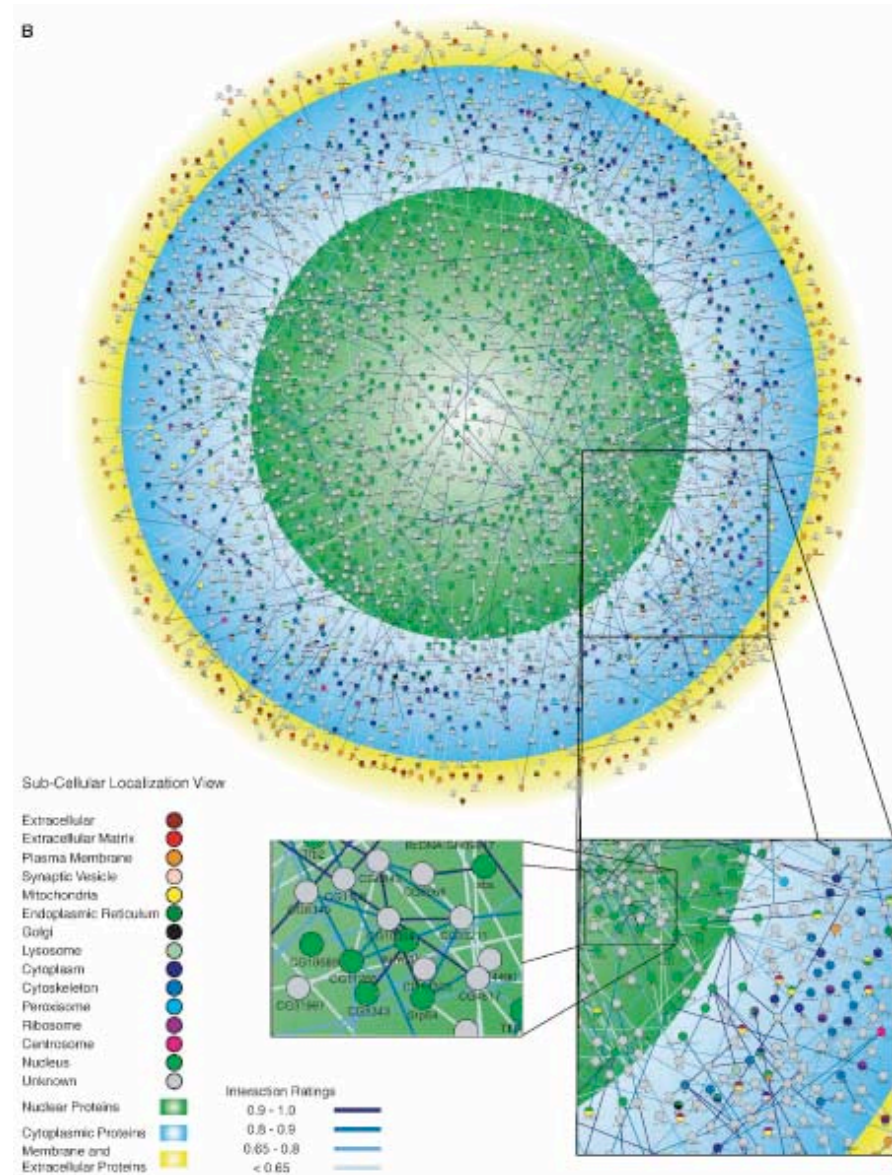


Измерение уровня экспрессии генов



7. Биологические сети

- Построение сетей на основе различных источников данных
- Сравнение сетей различных организмов
 - Статистика
 - Методы машинного обучения
 - Алгоритмы теории графов
 - Базы данных



Клеточные процессы

http://www.sector3.iitp.ru/video/Inner_life_of_a_Cell_FULL.mov

Благодарности

- При подготовке слайдов использовались материалы лекций:
 - Михаила Гельфанда (ИППИ)
 - Андрея Миронова (МГУ)
 - Serafim Batzoglou (Stanford)
 - Manolis Kellis (MIT)
 - Pavel Pevzner (UCSD)