

Полногеномные исследования в генетике

Елена Набиева, Ph.D.

Сколковский институт науки и технологий

e.nabieva@skoltech.ru

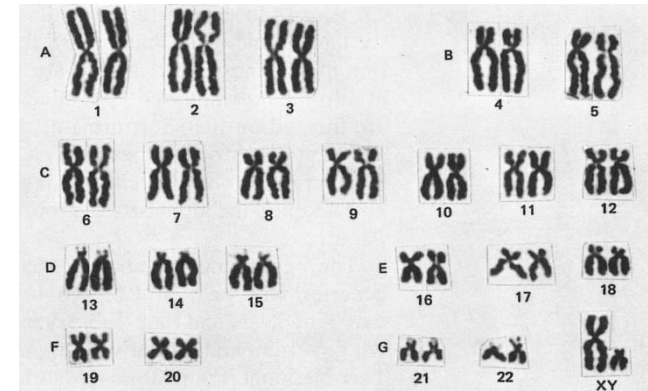
enabieva@gmail.com

План рассказа

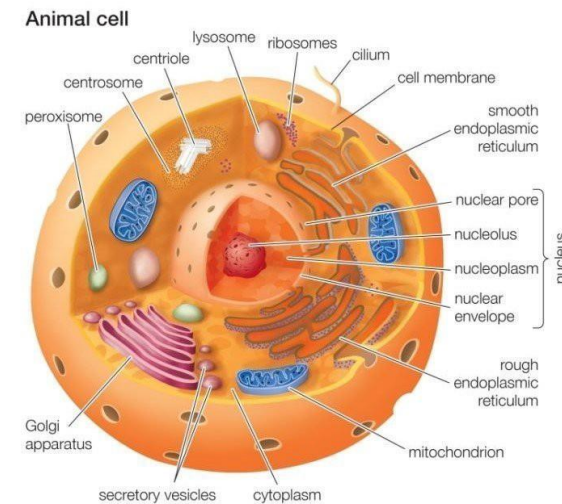
- Немного биологии
- Секвенирование генома человека
 - История
- Полногеномные методы
 - ДНК-чипы
 - Полногеномные исследования
 - Пример: предсказание роста
 - Персонализированная геномика:
 - Полигенная оценка риска заболеваний
- Что дальше?

Геном человека

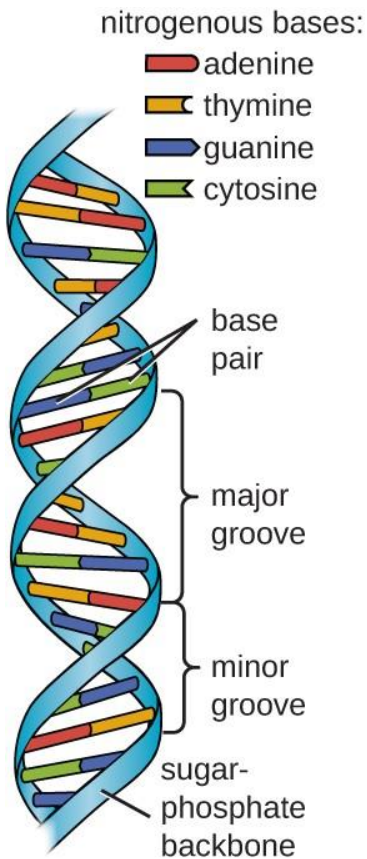
- 22 пары хромосом (аутосом)
- 2 половые хромосомы: XX или XY
- + митохондрии с собственной хромосомой
- Ребенок получает:
 - Одну аутосому от одного из родителей
 - Одну X-хромосому от мамы
 - X или Y хромосому от папы
- Митохондрии с их хромосомами – от мамы



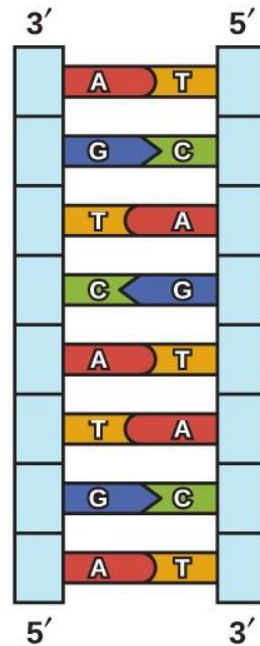
https://www.mun.ca/biology/scarr/Human_Karyotype.html



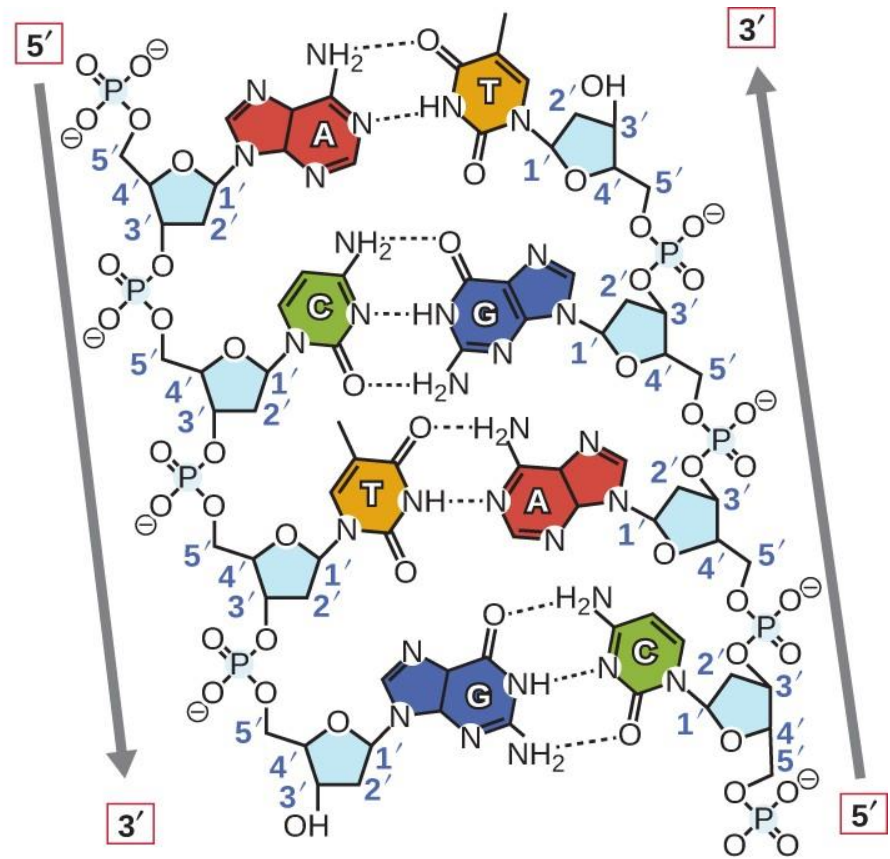
Вспомним: структура ДНК



(a)

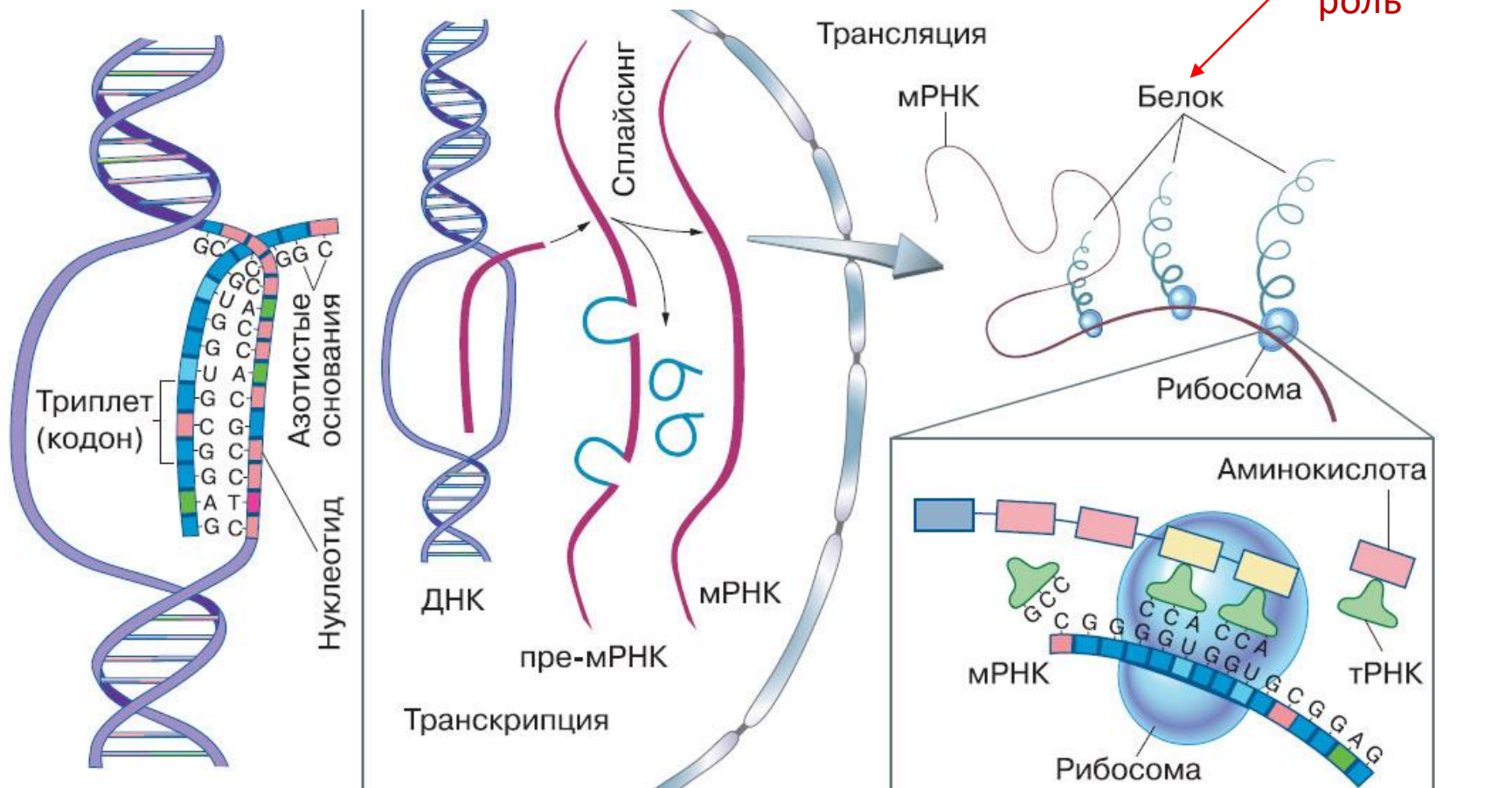


(b)



(c)

«Центральная догма» молекулярной биологии



Последовательность генома человека: 2001 год



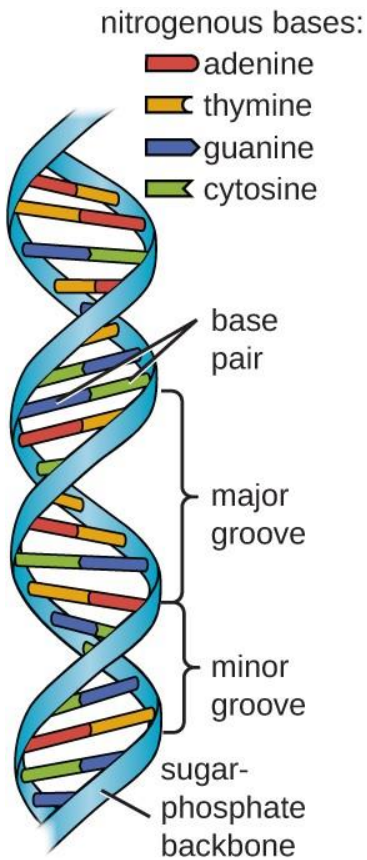
Некоммерческий Human Genome Project



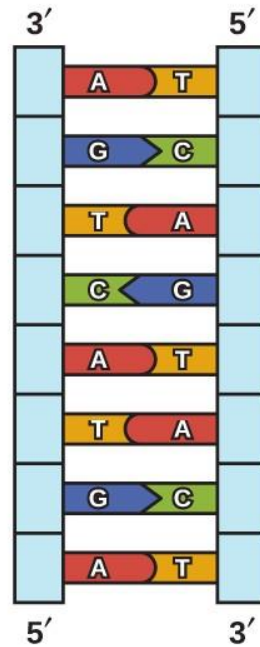
Коммерческий проект (Celera)

Как это делалось?

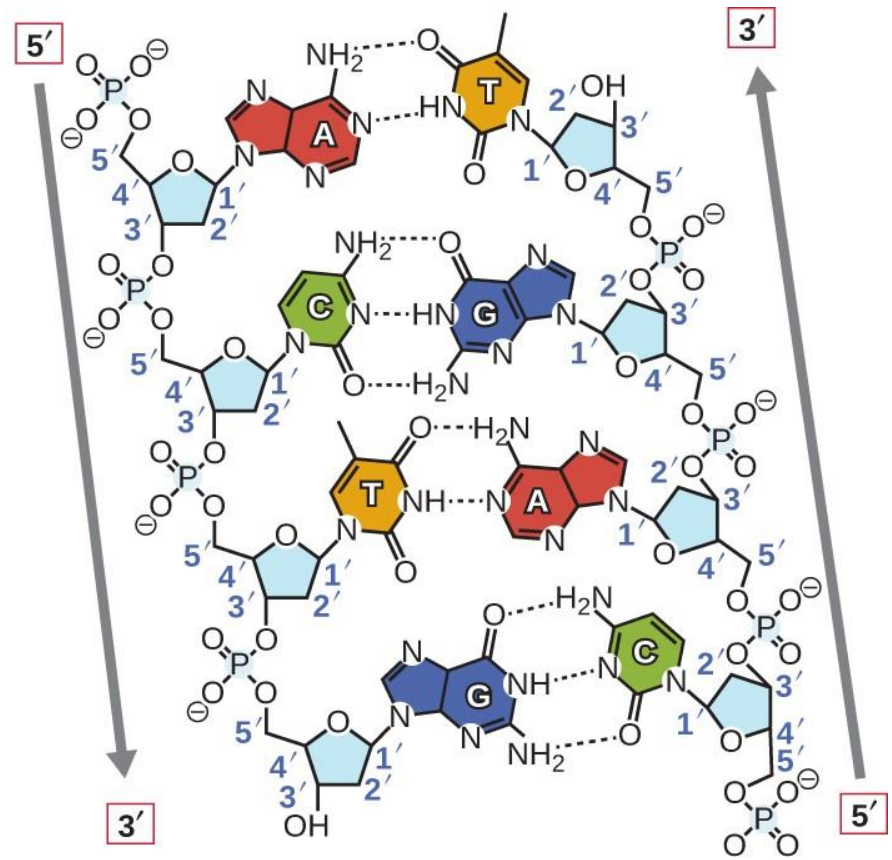
Вспомним: структура ДНК



(a)



(b)

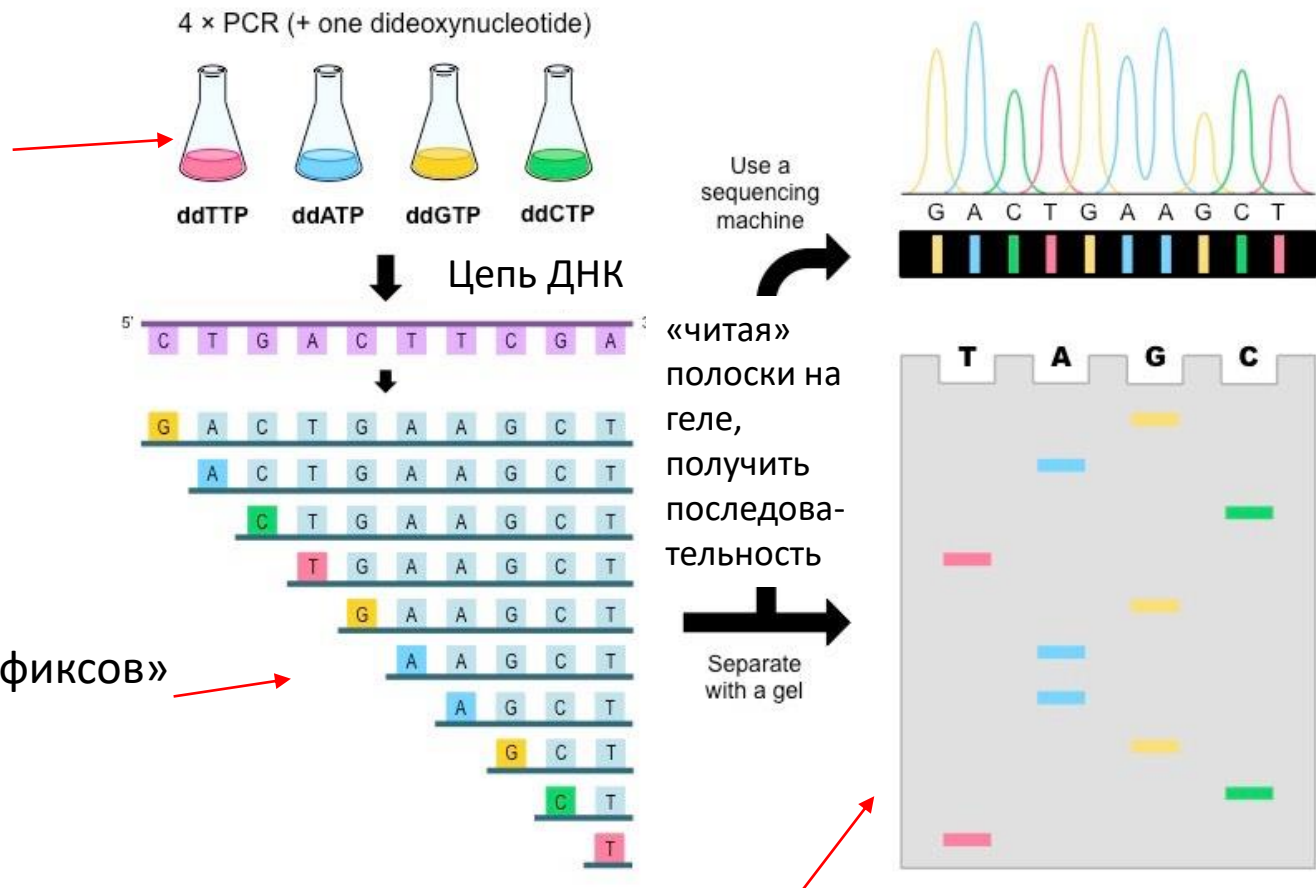


(c)

Секвенирование: Фред Сэнгер (1977)

Измененные «кирпичики» ДНК: могут встраиваться в растущую цепь, но дальше синтез не идет.

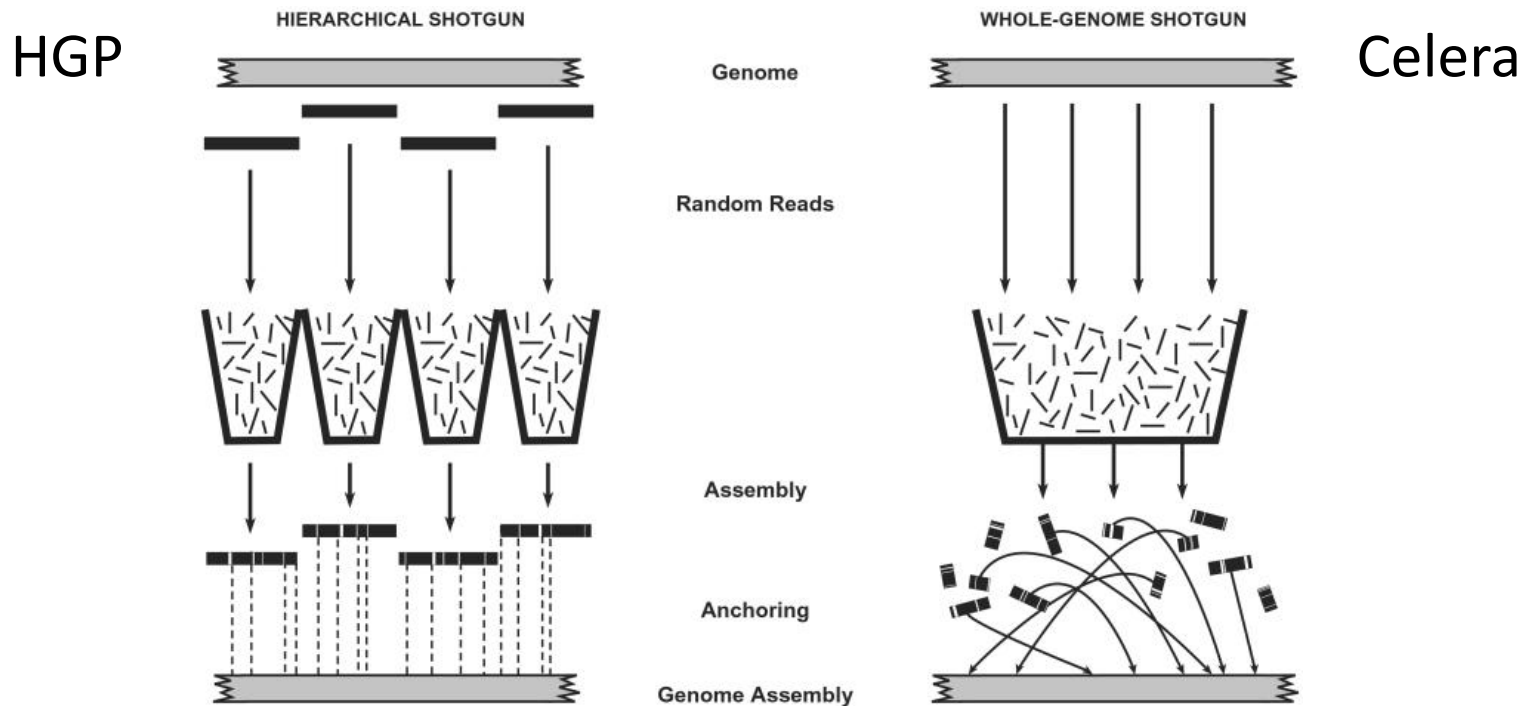
Много коротких «префиксов» цепи разной длины



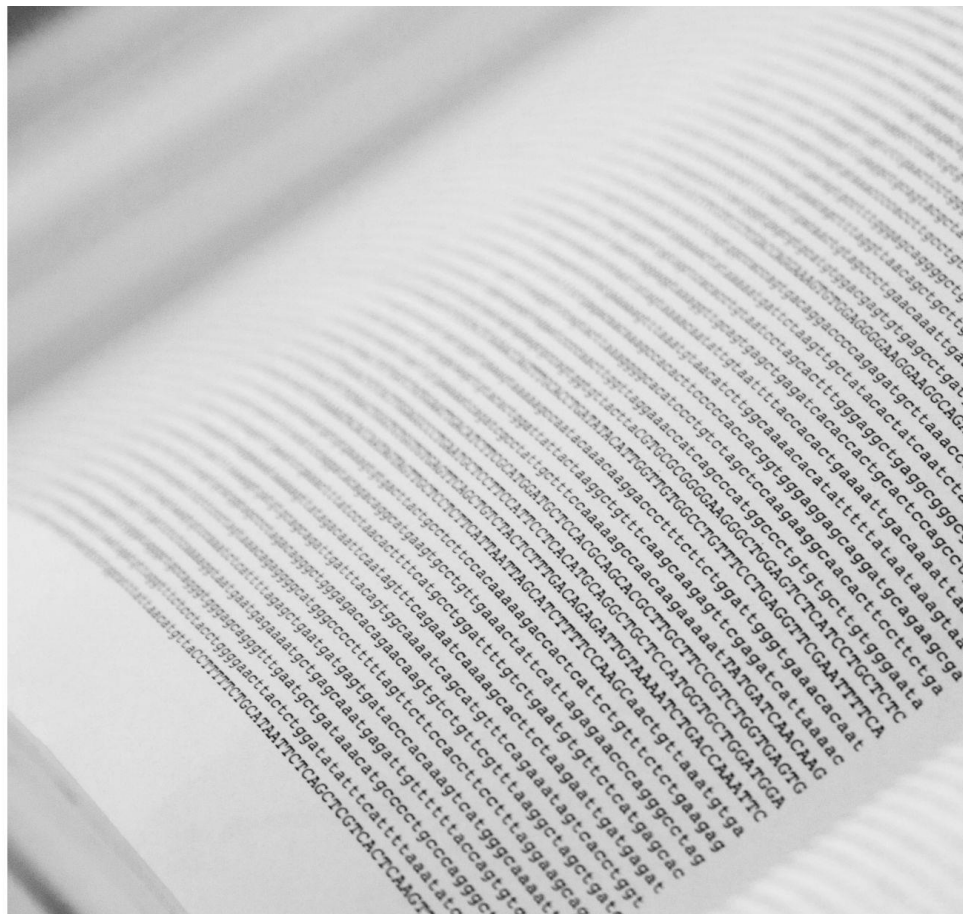
Они мигрируют с разной скорости в геле электрофореза и оказываются на разном расстоянии от начала

Секвенирование генома человека

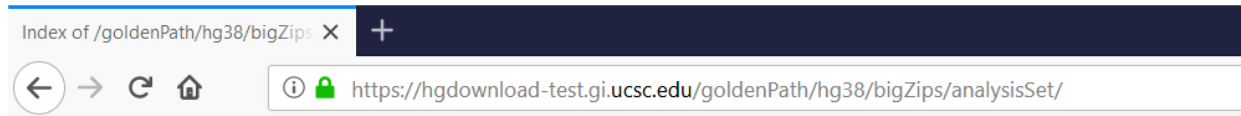
- Методом Сэнгера можно отсеквенировать порядка 500-700 нуклеотидов за раз
- Как отсеквенировать весь геном?



Последовательность генома человека



Последовательность генома человека



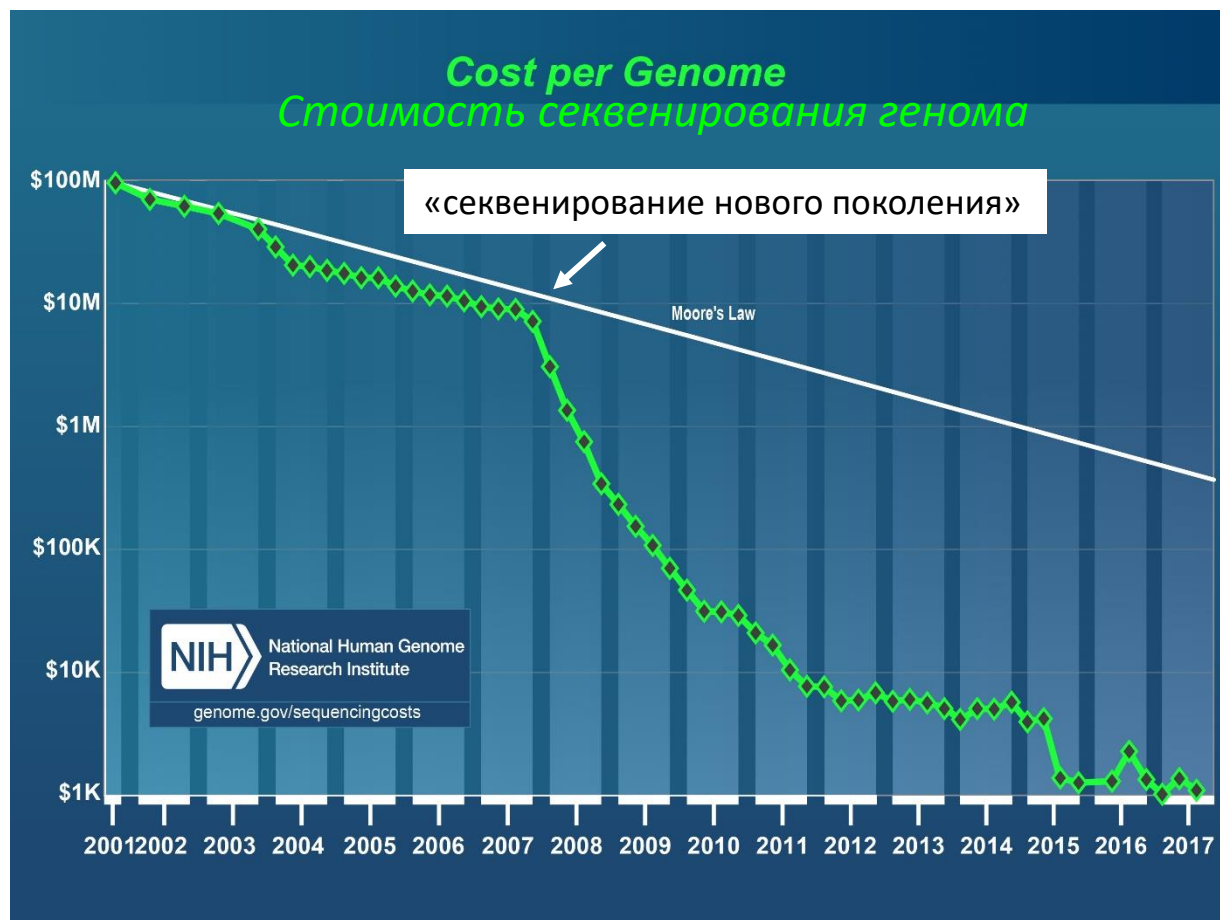
Index of /goldenPath/hg38/bigZips/analysisSet

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
Parent Directory	-	-	-
hg38.analysisSet.2bit	2014-01-27 10:40	770M	
hg38.analysisSet.chroms.tar.gz	2014-01-27 11:02	905M	
hg38.fullAnalysisSet.2bit	2014-03-18 13:23	797M	
hg38.fullAnalysisSet.chroms.tar.gz	2014-03-18 13:41	936M	
md5sum.txt	2014-03-18 13:41	250	

```
enabieva@mgmt:~/grch38
CCACCTCTTATTTAGTGGACAGACAGTAAGTGTGAGAAGACAGGGGATTTTGTGGCGGAAAAAAAAA
TTTATCAAAAGTCGTCTTCTATCAGGGAGTTTATGAGAAACCCTAGCTCCTCAGTTCACAGTGGGTAA
CTGTAATTCATTCTAGGTCTGCGATATTTCTGCCTATCCATTTTGTTAACTCTTCAATGCATTCCACAA
ATACCTAAGTATTCTTTAATAATGGTGGTTTTTTTTTTTTTTTGCATCTATGAAGTTTTTCAAATTCCT
TTTAAGTGACAAAACCTTGACATGTGTATCGCTCAATATTTCTAGTCGACAGCACTGCTTTCGAGAATGT
AAACCGTGCACCTCCAGGAAATGCAGACACAGCACGCCTCTTTGGGACCGCGGTTTATACCTTCGAAGT
GCTCGGAGCCCTTCTCCAGACCGTTCTCCACACCCCGCTCCAGGTCTCTCCCGAGTTACAAGCCTC
GCTGTAGGCCCCGGGAACCAACGCGGTGTCAGAGAAGTGGGGTCCCTACGAGGGACAGGAGCTCCGG
GCGGGCAGCAGCTGCGGAAGAGCCGCGCAGGCTTCCAGAACCCGGCAGGGGCGGGAAGACGCAGGAGT
GGGGAGGCGGAACCGGGACCCCGCAGAGCCCGGTCCCTGCGCCCCACAAGCCTTGGCTTCCCTGCTAGG
GCCGGGCAAGGCCGGGTGCAGGGCGCGGCTCCAGGGAGGAAGTCCGCGGGCAGCCCCAAGACGCCTCCCG
GGCGGTGCGGGCCAGCGGCGGCGTTTCGAGTGAGCCGGGCACCGGGCAGCGGCCGCGGAACACCAGCT
TGGCGCAGGCTTCTCGGTGAGAACGGTCCCGGGCCTCCCGCCCGCTCCCTCCAGCCCCCTCCGGGTCCC
CTACTTCGCCCCGCGAGGCCCCACGACCTACTTCCCGCGGCCCGGACGCCTCCTCACCTGCGAGCCG
CCCTCCCGGAAGCTCCCGCCGCGCTTCCGCTCTGCCGGAGCCGCTGGGTCTAGCCCCGCGCCCCCAG
TCCGCCCGCGCTCCGGGTCTTAACGCCGCGCTCGCCCTCCACTGCGCCCTCCCGAGCGCGGCTCCAG
GACCCCGTGCAGCCGAGCGCTGTCTGTGCGGCCGAGTCGCGGGCCTGGGCACGGAACCTACGCTCACT
CCGAGCTCCCGACGTGCACACGGCTCCCATGCGTTGTCTTCGAGCGTCAGGCCGCCCCCTACCCGTGCTT
TCTGCTCTGCAGACCTCTTCTAGACCTCCGTCTTTGTCCCATCGCTGCCTTCCCTCAAGCTCAGGG
CCAAGCTGTCCGCCAACCTCGGCTCCTCCGGGCAGCCCTCGCCCGGGTGCGCCCCGGGGCAGGACCCCC
AGCCCACGCCCAGGGCCCCGCCCTGCCCTCCAGCCCTACGCCTTGACCCGCTTTCTGCGTCTCTCAGCC
TACCTGACCTTGTCTTTACCTCTGTGGGCAGCTCCCTTGTGATCTGCTTAGTTCCACCCCCCTTTAAGA
ATTCAATAGAGAAGCCAGACGCAAACTACAGATATCGTATGAGTCCAGTTTTGTGAAGTGCCTAGAATA
--More-- (0%)
```

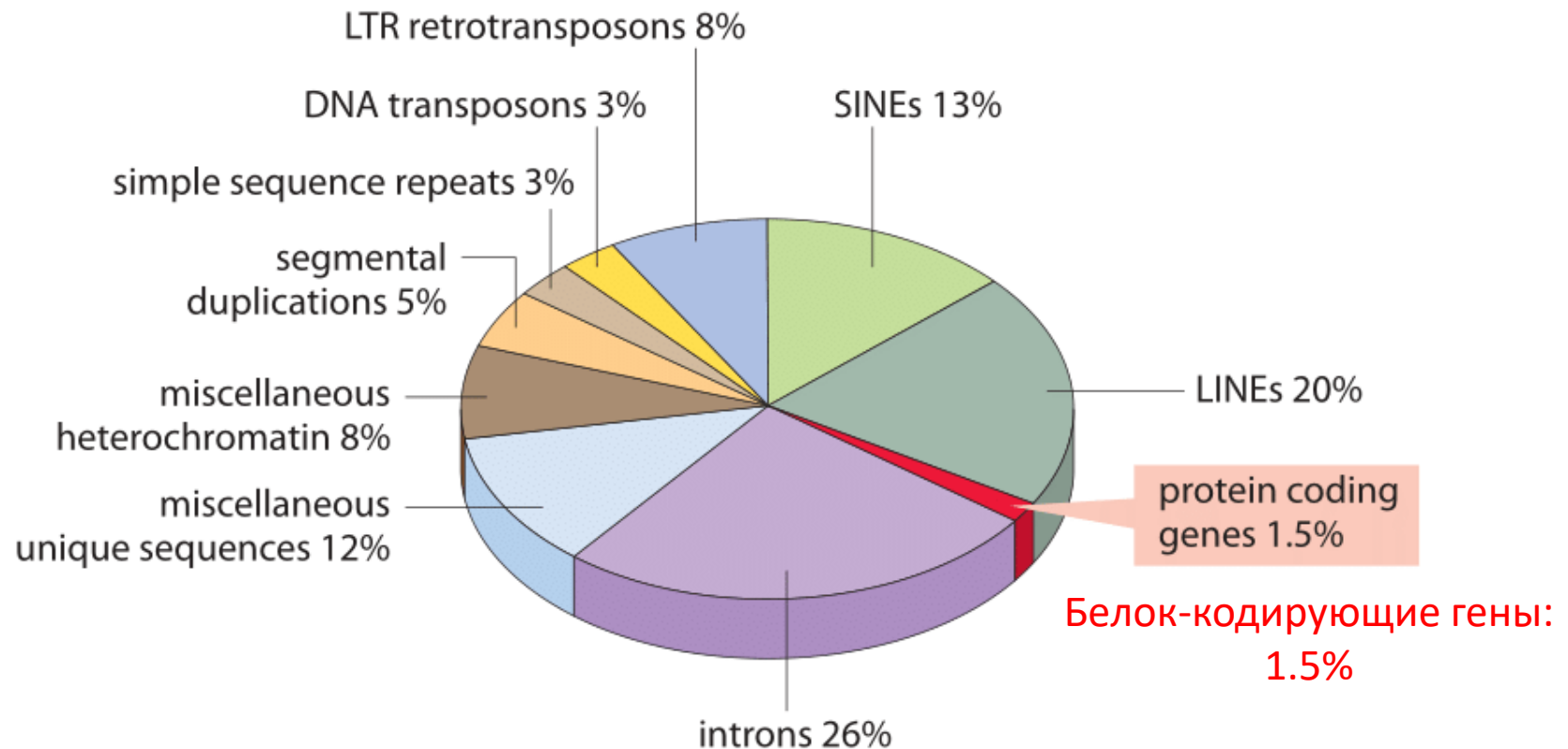
Секвенирование нового поколения

- «высокопроизводительное секвенирование»



Из чего состоит геном человека?

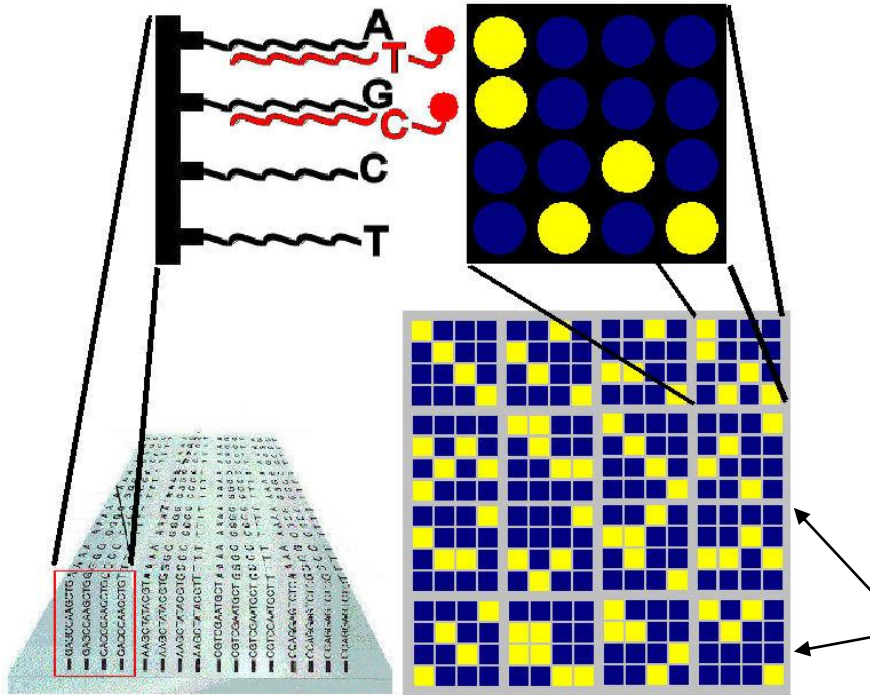
main components of the human genome



ДНК-микрочипы



ДНК-микрочипы

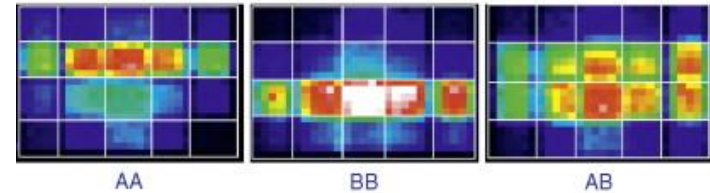


https://www.mun.ca/biology/scarr/DNA_Chips.html

Target sequence (250–2000 bp)
...CAGACAGAAGTCTTG[A/C]AATCTATTTCTCATA...

Probe sequence (25 bp)

PM _A	: TGTCTTCAGAAC	T	TTAGATAAAAGAG
MM _A	: TGTCTTCAGAAC	A	TTAGATAAAAGAG
PM _B	: TGTCTTCAGAAC	G	TTAGATAAAAGAG
MM _B	: TGTCTTCAGAAC	C	TTAGATAAAAGAG

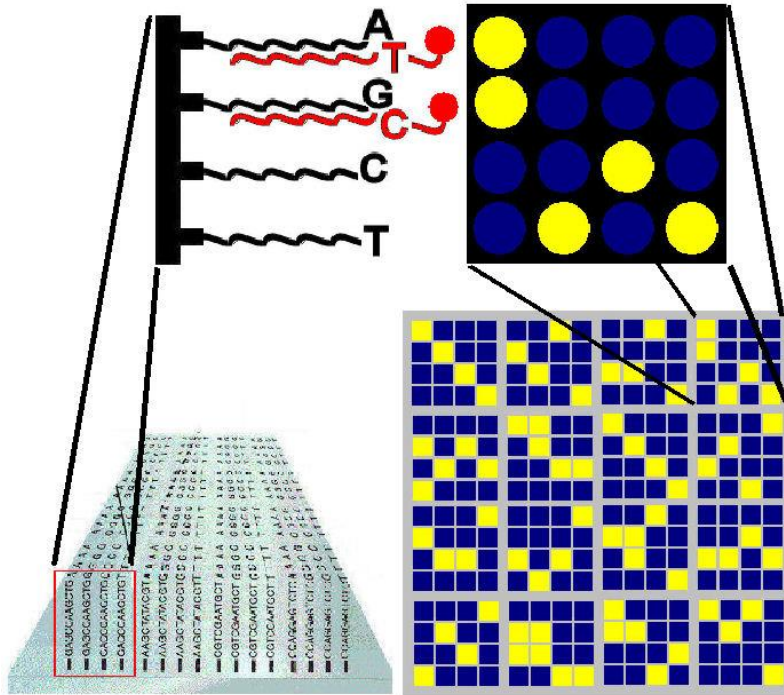


<https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/snp-array>

Выявляют аллели в
определенных участках генома

Современные чипы: до
миллиона маркеров

ДНК-микрочипы

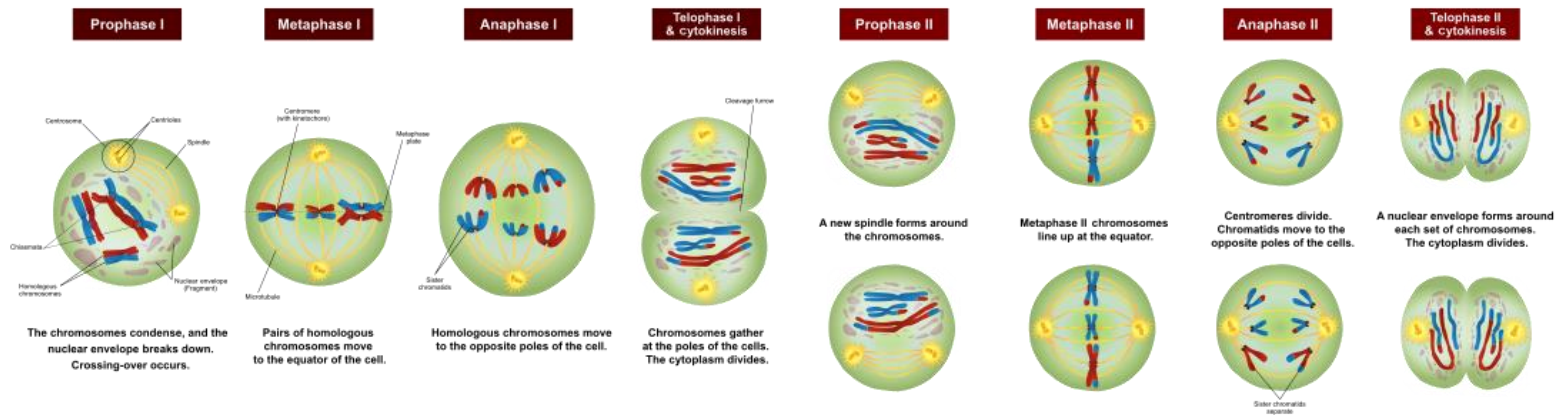


https://www.mun.ca/biology/scarr/DNA_Chips.html

CTAGCTCCTCACCTGCTTCCATCCTCCCTGGTGCGGGGTGGGCCCAGTGATATCAGCTGCCTGCTGTTCC
 CCAGATGTGCCAAGTGCATTCTTGTGTGCTTGCATCTCATGGAACGCCATTTCCCCAGACATCCCTGTGG
 CTGGCTCCTGATGCCCCAGGCCAAGTGTCTGATGCTTTAAGGCACATCACCCCACTCATGCTTTTCCAT
 GTTCTTTGGCCGAGCAAGGCCGCTCTCACTGCAAAGTTAACTCTGATGCGTGTGTAACACAACATCCTC
 TCCCCAGTCGCCCCGTAGCTCCCCCTACCTCCAAGAGCCAGCCCTTGCCACAGGGCCACACTCCACGT
 GCAGAGCAGCCTCAGCACTCACCGGGCAGAGCGAGCCTGTGTGGTGCAGGGATGAGAAGGCAGAGGC
 GCGACTGGGGTTCATGAGGAAGGGCAGGAGGAGGTGTGGGATGGTGGAGGGGTTTGAGAAGGCAGAGGC
 GCGACTGGGGTTCATGAGGAAGGGAGGGGAGGATGTGGGATGGTGGAGGGGCTGCAGACTCTGGGCTA
 GGGAAAGCTGGGATGTCTCTAAAGGTTGGAATGAATGGCCTAGAATCCGACCCAATAAGCCAAGCCACT
 TCCACCAACGTTAGAAGGCCTTGGCCCCCAGAGAGCCAATTTACAATCCAGAAGTCCCCGTGCCCTAAA
 GGGTCTGCCCTGATTACTCCTGGCTCCTTGTGTGCAGGGGGCTCAGGCATGGCAGGGCTGGGAGTACCAG
 CAGGCACTCAAGCGGCTTAAGTGTTCATGACAGACTGGTATGAAGGTGGCCACAATTCAGAAAGAAAAA
 AGAAGAGCACCATCTCCTTCCAGTGAGGAAGCGGGACCACCCAGCGTGCTGCCATCTTTCTGCGCT
 GGGGAGAGGCCTTCATCTGTGTAAAGGGTCTCCAGCACAAGCTGTCTTAATTGACCTAGTTCACAGG
 GCAGCCTCGTTCTGCCTTGGGTGCTGACACGACCTTCGGTAGGTGCATAAGCTCTGCATTCGAGGTCCAC
 AGGGGCACTGGGAGGGAACTGAGACTGGGGAGGGACAAAGGCTGCTCTGCTCGGTGCTCCCAAAAGGA
 GAAGGGCTGATCACTCAAAGTTGCGAACACCAAGCTCAACAATGAGCCCTGGAATAATTCTGGAATGGAT
 TATTAACAGAGAGTCTGTAAGCACTTAGAAAAGCGCGGTGAGTCCCAGGGGCCAGCACTGCTCGAAA
 TGTACAGCATTTCTCTTTGTAACAGGATTATTAGCCTGCTGTGCCGGGGAAAACATGCAGCACAGTGCA
 TCTCGAGTCAGCAGGATTTTGACGGCTCTTAACAAAATCTTGTAGAACAGATGGAGCTATGGGGGTGGA
 GGAGAGAACATATAGAAAAATCAGAGCCAATGAACCACAGCCCCAAAGGGCACAGTTGAACAATGGAC
 TGATTCCAGCCTTGACGAGGAGGATCTGGCAGAGTCCATCCAGTTTCAATCAACACCTGGTTAGAAAAGT
 GGGCCAGCACACAGGGGAAGGGTAAGCTGGTTTTCATGATCGAATCAAGGCTCAGACAATTTTTAAAGGCC
 AGAGGGTAGACTGCAATCACCAAGATGAAATTTACAAGGAACAAATGTGAAGCCCAACATTTAGGTTTTA
 AAAATCAAGCGTATAAATACAGAAGGTGGAGGGAAGTGTCTTTAGACACAGTTTCAAGTGAAGAAAGACCT
 GGAAACTTCTGTTAACTATAAGCTCAGTAGGGGCTAAAAGCATGTTAATCGGCATAAAAAGGCAATGAGA
 TCTTAGGGCACACAGTCCCCGCCCTCTTCTGCCCTTCATCCTTCTTTCAATCAGCAGGACCGTGCAC

Почему это информативно?

- «Неравновесное сцепление»
 - Варианты наследуются вместе с хромосомами
- Но:

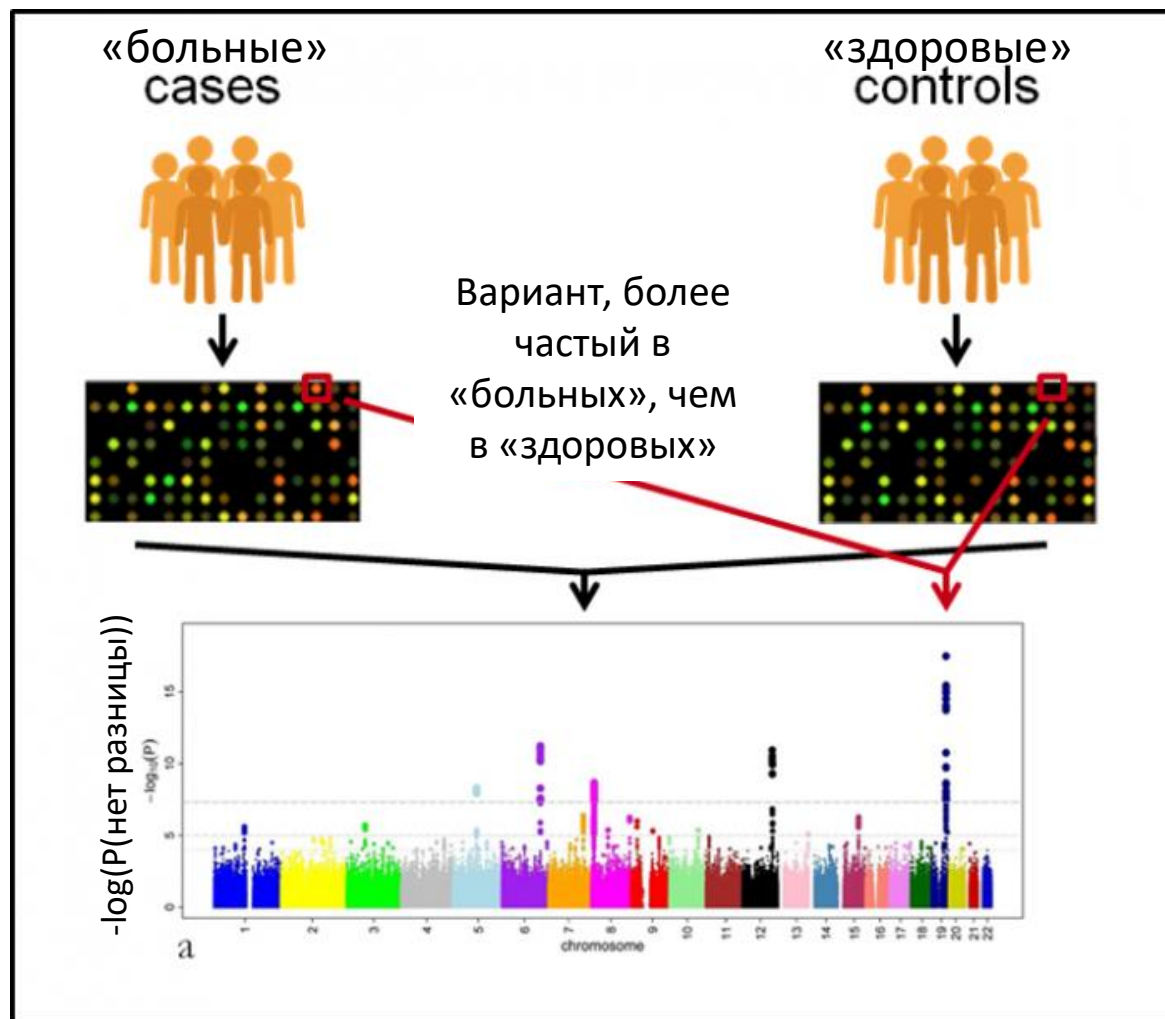


<https://en.wikipedia.org/wiki/Meiosis>

Рекомбинация: гомологичные (парные) хромосомы обмениваются кусками

- Случайный процесс: $\sim 1.6/\text{хромосому}$
- Чем ближе два участка хромосомы тем меньше вероятность, что их разделит рекомбинация (в первом приближении)

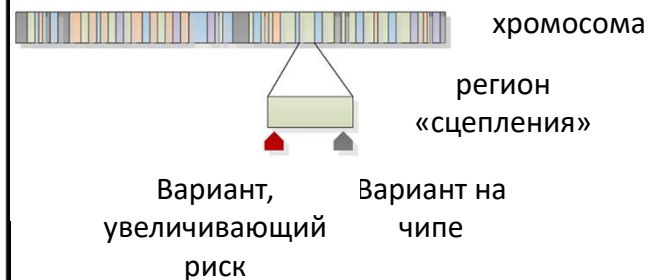
Полногеномный поиск ассоциаций



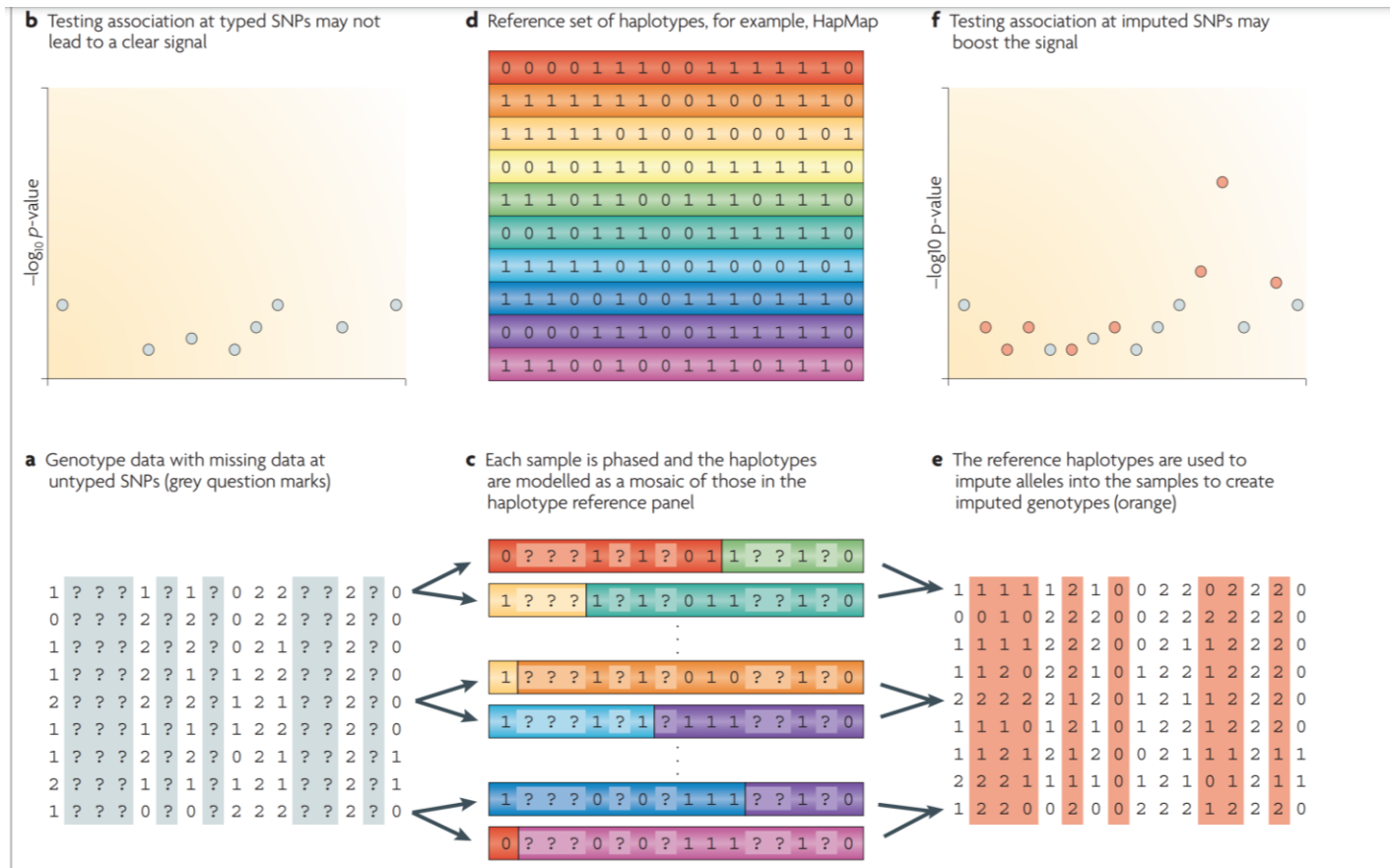
“GWAS”

Это необязательно биологически функциональный вариант:

Непрямая ассоциация



Секвенирование помогает «заполнить пробелы»



Особенности полногеномного поиска ассоциаций:

- Способность установить связь между генетическим вариантом и изучаемым признаком **пропорциональна**:
 - силе эффекта варианта, влияющего на признак
 - количеству участников исследования
 - сцепленности между вариантом на чипе и вариантом, влияющим на признак

=> Чем больше участников исследования, тем больше вариантов можно найти

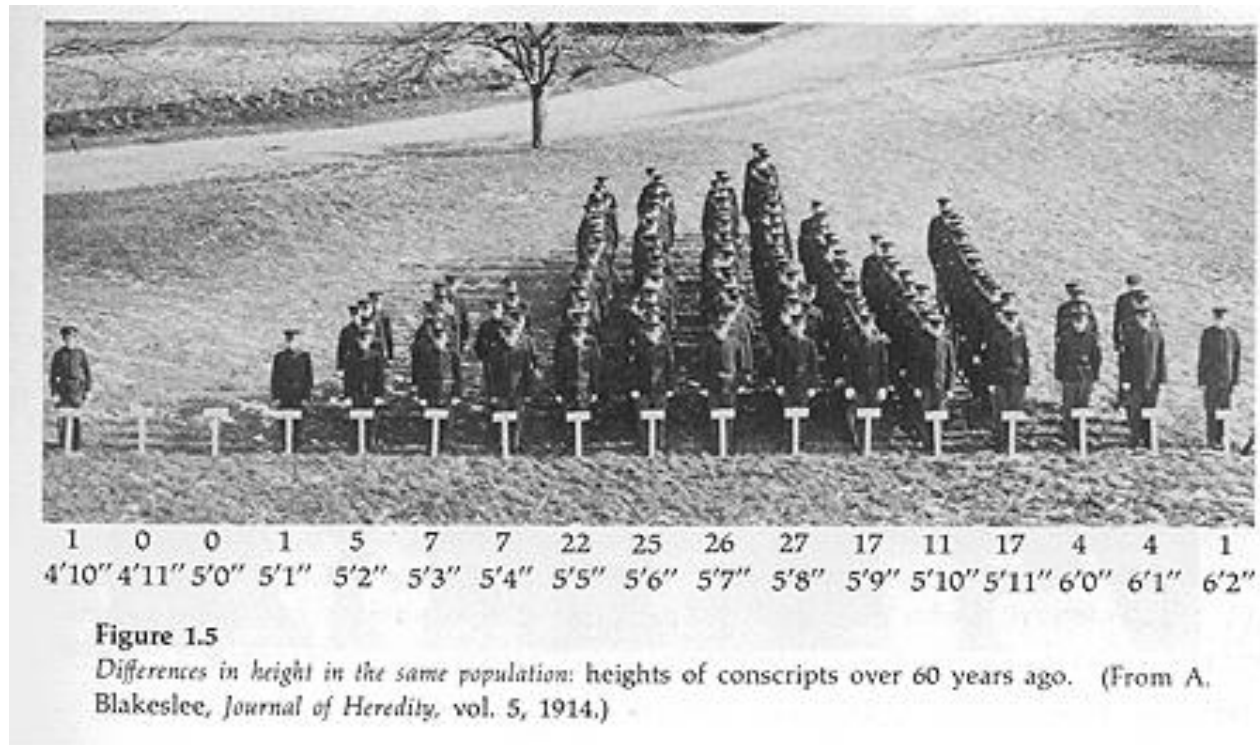
Уроки

На «сложные признаки» влияет **множество** генетических вариантов

- Предрасположенность к шизофрении:
 - 2009: 3 000 «больных» => 1 вариант
 - 2014: 35 000 «больных» => 108 вариантов

Многие варианты влияют на множество признаков

История: умеем ли мы предсказывать рост?



Умеем ли мы предсказывать рост?

- Традиционная генетика: рост наследуется на ~80%
- Полногеномные исследования:
 - 2008: 40 вариантов объясняют 5% наследуемости
 - 2014: ~700 вариантов объясняют 20%+ наследуемости

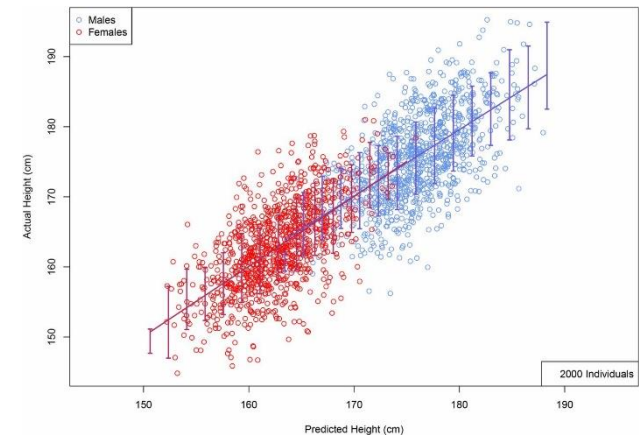
[Visscher et al. Am J Hum Genet.](#) 2017 Jul 6; 101(1): 5–22.

Потерянная наследуемость?



А что если использовать все варианты?

- 2011 (Yang et al), 2018 (Lello et al): статистические модели, использующие все варианты с частотой $> 1\%$, объясняют $\sim 50\%$ наследуемости
- 2019 (Wainschtein et al, bioRxiv): использование редких вариантов, полученных методами *полногеномного секвенирования* объясняет 79% наследуемости



Lello et al 2018

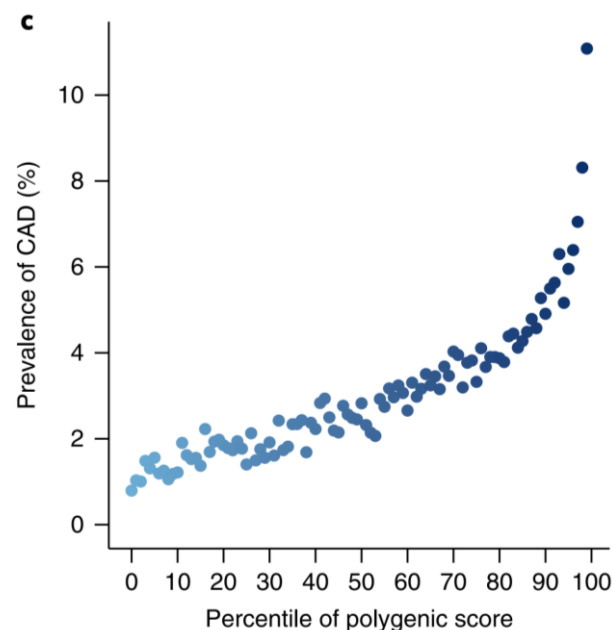
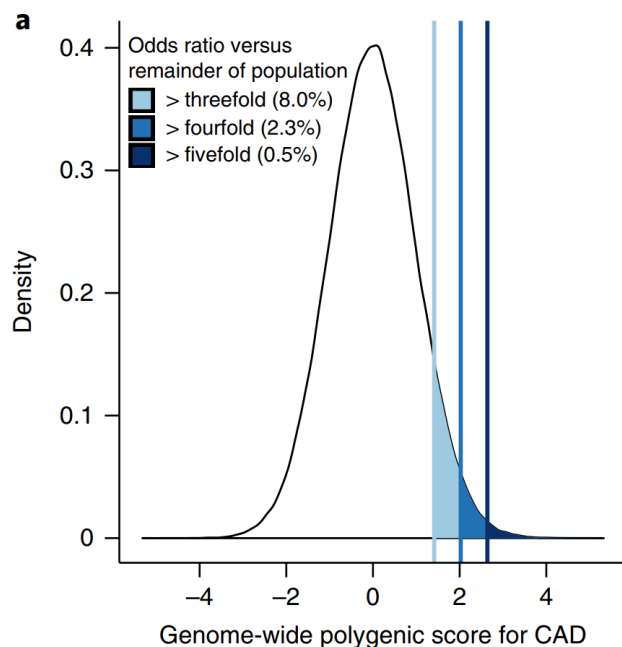




Персонализированная медицина

Полигенная оценка риска

Риск развития ишемической болезни сердца:



Сравнимо с риском от мутации, вызывающей наследственную гиперхолестеринемию, но выявляет в 20 раз больше людей!

Полигенный риск для других болезней

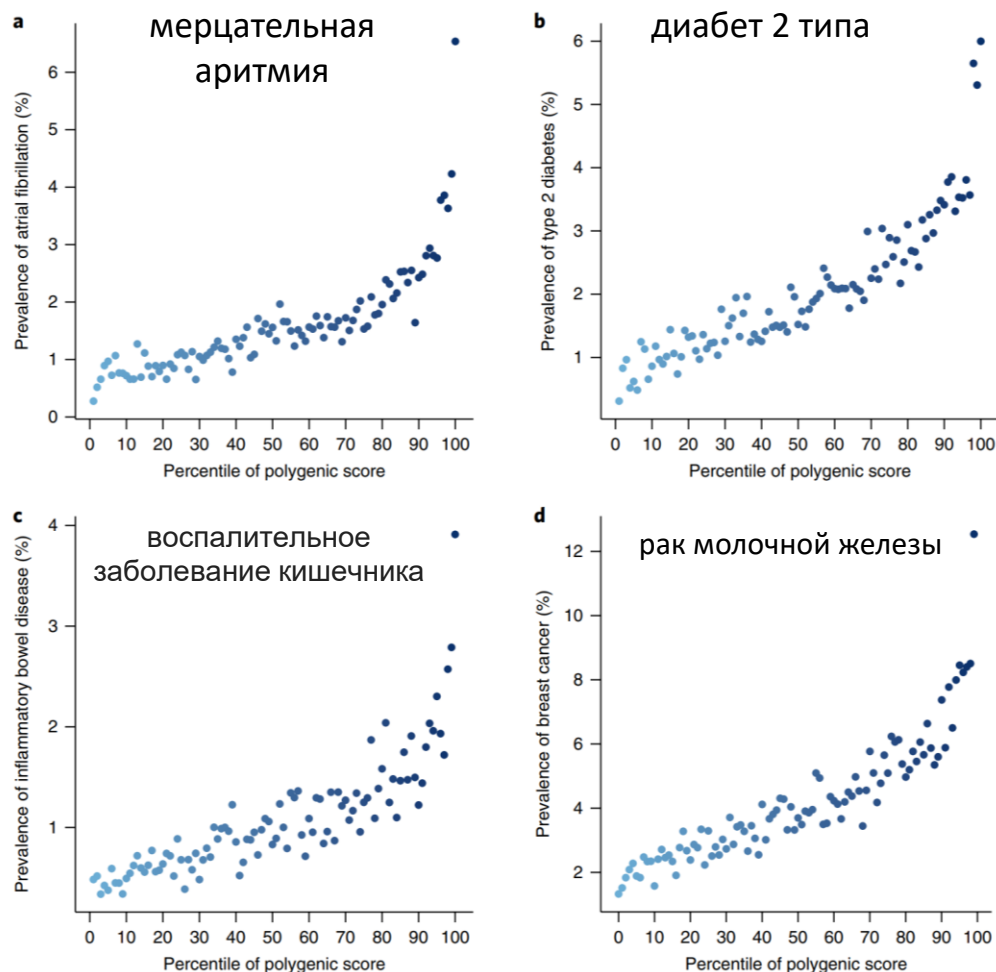
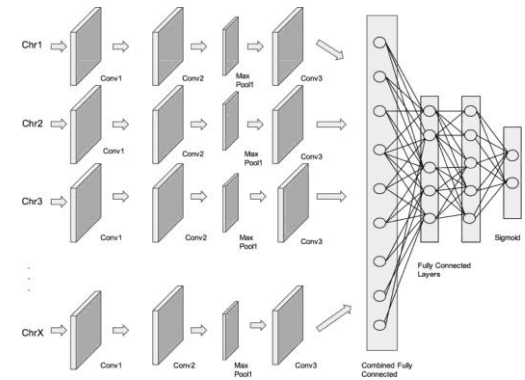


Fig. 3 | Risk gradient for disease according to the GPS percentile. 100 groups of the testing dataset were derived according to the percentile of the disease-specific GPS. **a-d**, Prevalence of disease displayed for the risk of atrial fibrillation (**a**), type 2 diabetes (**b**), inflammatory bowel disease (**c**), and breast cancer (**d**) according to the GPS percentile.

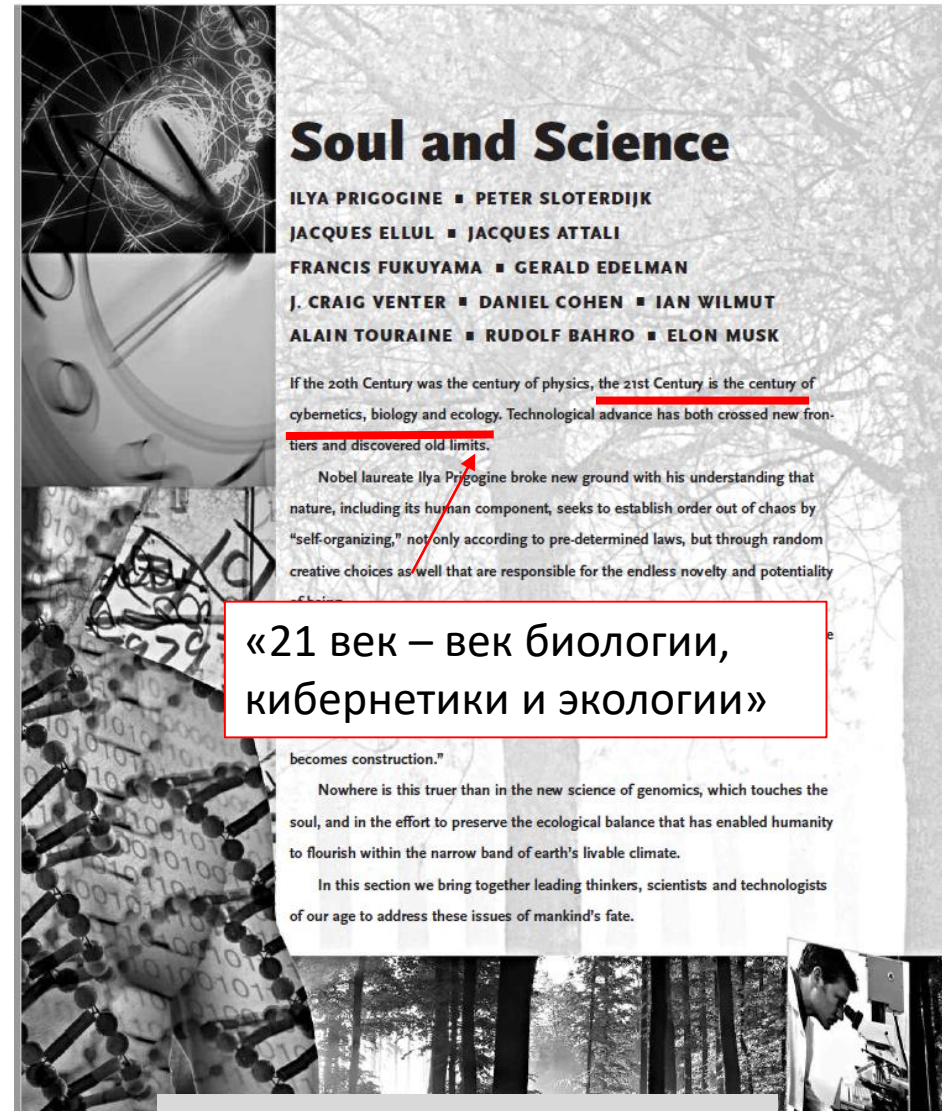
Что дальше?

- Возможно, не все признаки хорошо предсказываются линейными моделями
 - Линейные: складываем эффект вариантов с их «весами»
- Поможет ли машинное обучение?
- DeepBipolar, победитель конкурса CAGI-4
 - Sundaram Laksshman et al. Hum Mutat 2017
 - Выборка: 500 человек с биполярным расстройством, 500 здоровых
 - Данные: 60 ассоциированных генов + микрочипы
- Эту задачу нельзя назвать решенной!



Заключения: «век биологии»

- Технология как двигатель науки
- Биология как наука данных
- Информационные технологии для осмысления новых данных



Soul and Science

ILYA PRIGOGINE ■ PETER SLOTERDIJK
JACQUES ELLUL ■ JACQUES ATTALI
FRANCIS FUKUYAMA ■ GERALD EDELMAN
J. CRAIG VENTER ■ DANIEL COHEN ■ IAN WILMUT
ALAIN TOURAINE ■ RUDOLF BAHRO ■ ELON MUSK

If the 20th Century was the century of physics, the 21st Century is the century of cybernetics, biology and ecology. Technological advance has both crossed new frontiers and discovered old limits.

Nobel laureate Ilya Prigogine broke new ground with his understanding that nature, including its human component, seeks to establish order out of chaos by "self-organizing," not only according to pre-determined laws, but through random creative choices as well that are responsible for the endless novelty and potentiality of life.

«21 век – век биологии, кибернетики и экологии»

becomes construction."

Nowhere is this truer than in the new science of genomics, which touches the soul, and in the effort to preserve the ecological balance that has enabled humanity to flourish within the narrow band of earth's livable climate.

In this section we bring together leading thinkers, scientists and technologists of our age to address these issues of mankind's fate.