



# ГЕНОМИКА: ПРОШЛОЕ, НАСТОЯЩЕЕ И БУДУЩЕЕ

---

## Лекция 1

# Преподаватели



- Лагоненко Александр Леонидович  
Доцент кафедры. к 126, 318 (ф. радиофизики)



- Валентович Леонид Николаевич
- Доцент кафедры. Институт микробиологии НАН Б.



- Горовик Юрий Николаевич
- Ассистент кафедры. к 318 (ф. радиофизики)

# Что вас ждет...

<http://bio.bsu.by/2017a/>

- Написать эссе на тему «Придумайте геном бактерии (организма), которая сможет спасти экономику Беларуси». Объём эссе – до 25 000 знаков. Сдавать в напечатанном или написанном от руки виде. Обязательно должен присутствовать титульный лист и список литературы. Эссе сдаётся до окончания всех практических занятий.
- Написать отчёт о проделанной индивидуальной работе. Полученные в результате компьютерного анализа данные настоятельно рекомендуется приводить в виде иллюстраций (выравниваний, карт последовательностей и т.д.) или таблиц, которые необходимо уметь объяснять и на их основании уметь делать выводы. Вся информация в отчёте должна быть только на РУССКОМ или БЕЛОРУССКОМ языке. **ОБЯЗАТЕЛЬНО** в конце отчёта следует привести краткую характеристику проанализированных вами биологических последовательностей, написать вывод. Необходимо показать в ЭЛЕКТРОННОМ виде отчёт и аннотацию в формате Genbank преподавателю практики, уметь ответить на вопросы, доказав своими знаниями, что всё сделано и написано вами. После этого с распечатанным отчётом приходить на экзамен (зачёт).



# Геном

---

- Термин предложен Хансом Винклером в 1920 году для описания совокупности генов, заключенных в гаплоидном наборе хромосом.
- **Геном** — совокупность наследственного материала, заключенного в клетке организма.



# Геномика

---

- Genomics is “the study of functions and interactions of all the genes in the genome, including their interactions with environmental factors.”
  - Collins, Francis, and Alan Guttmacher. “Genomic Medicine—A Primer,” NEJM, Vol. 347:1512-1520.)
- **Геномика** - наука, изучающая организацию, функции и эволюцию геномов.



# Хронология развития геномики

---

- 1975 – Ф. Сенгер и независимо А. Максам и У. Гилберт разработали методы секвенирования ДНК.
- 1977 – Секвенирован первый полный геном - бактериофаг фХ-174 (5368 н).
- 1980 – Верховный суд США принял решение о возможности патентования генетически модифицированных бактерий, что послужило основой для патентования генов.
- 1981 – Секвенирована митохондриальная ДНК человека: 16569 п.о.
- 1982 – Опубликован релиз GenBank; секвенирован геном фага  $\lambda$  (Фред Сенгер и соавторы).
- 1984 – Секвенирован геном вируса Эпштейна-Барра: 172281 п.о.
- 1990 – Запущен международный проект по геному человека с намеченным сроком 15 лет.



# Хронология развития геномики

---

- 1995 – TIGR получили первую последовательность генома бактерии, *Haemophilus influenzae*.
- 1996 – Полностью определена последовательность генома дрожжей, первого генома эукариот.
- 1999 – Корпорация Celera объявила о завершении секвенирования генома *Drosophila melanogaster*.
- 1999 – Опубликована первая полная последовательность одной из хромосом человека.
- 26 июня 2000 – Совместное заявление о полной расшифровке генома человека (на самом деле определено 98% генома).

# Гномика

## Три основных направления:

---

- **Структурная геномика:** генетическое, физическое картирование, секвенирование геномов
- **Функциональная геномика:** функции генов и некодирующих последовательностей в геномах
- **Сравнительная геномика:** сравнение геномов различных организмов. Эволюционная задача



# Гномика

## Интегративные направления:

---

- Медицинская геномика и фармакогеномика
- Нутригеномика
- Спортивная геномика
- Этногеномика
- Палеогеномика
- ...





## «-ОМ» означает «имеющий природу»

---

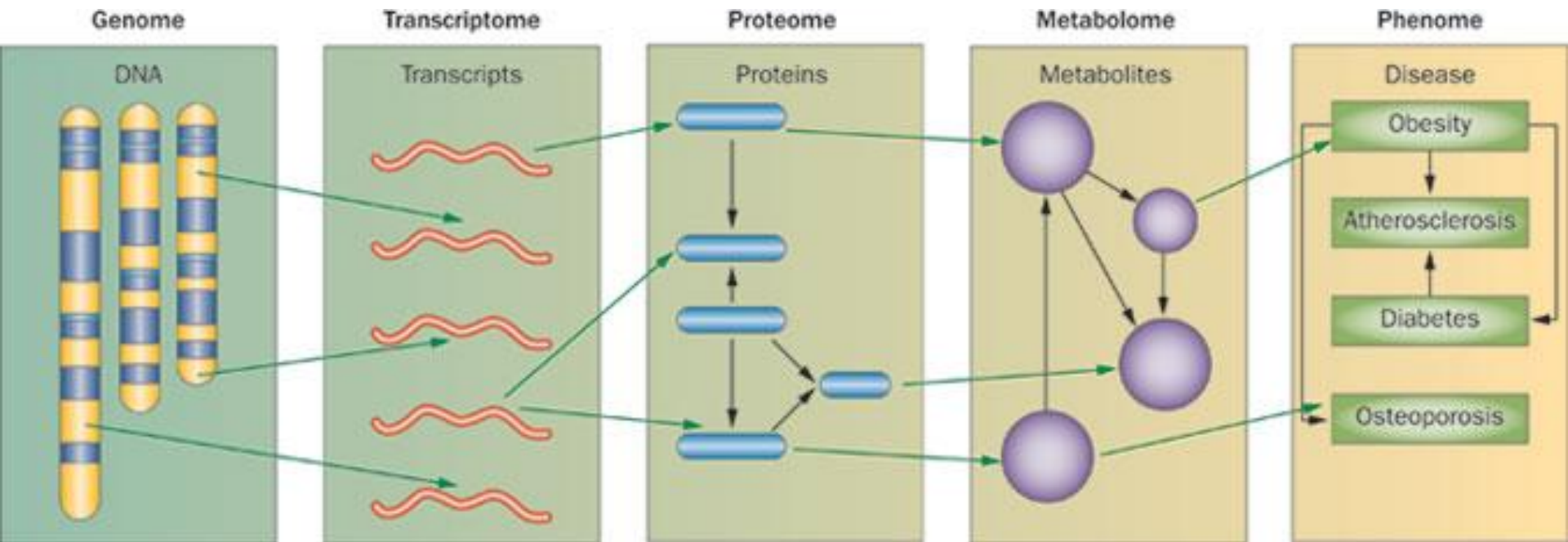
- *«Используя суффикс „-ом“, вы показываете, что принадлежите к абсолютно новой увлекательной области науки»*

*Alexa McCray*

- *«Люди пытаются убедить окружающих, что область их исследований — самостоятельная отрасль науки, и что она заслуживает особого финансирования»*

*Jonathan Eisen*

# Центральная догма и -омики



**Основная цель:**

**Понять, как функционирует живой организм**

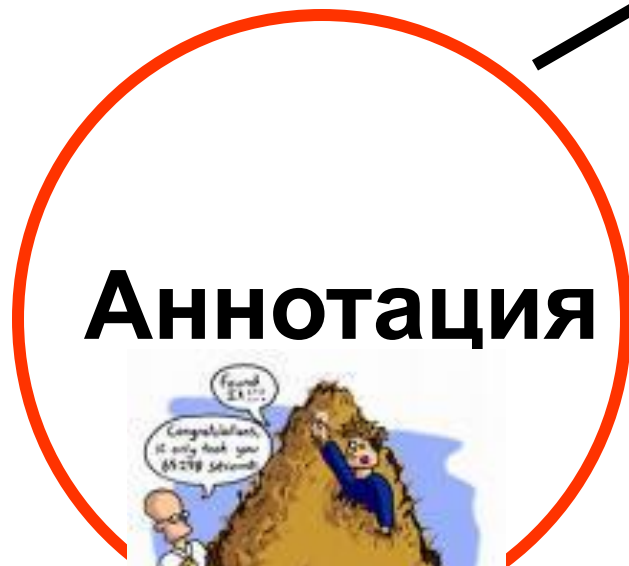


# Аннотация последовательности

---

CCTGACAAATTCGACGTGCGGCATTGCATGCAGACGTGCATG  
CGTGCAAATAATCAATGTGGACTTTTCTGCGATTATGGAAGAA  
CTTTGTTACGCGTTTTTGTTCATGGCTTTGGTCCCGCTTTGTTC  
AGAATGCTTTTAATAAGCGGGGTACCGGTTTGGTTAGCGAGA  
AGAGCCAGTAAAAGACGCAGTGACGGAGATGTCTGATG CAA  
TAT GGA CAA TTG GTT TCT TCT CTG AAT . . . . .  
. . . . . TGAAAAACGTA

# Аннотация последовательности



Идентифицировать гены в данной последовательности ДНК

Идентифицировать Регуляторные участки

Предсказать функцию

promoter

TF binding site

CCTGACAAATTCGACGTGCGGCAATTGCATGCAGACGTGCATG  
CGTGCAAATAATCAATGTGGACTTTTCTGCGATTATGGAAGAA  
CTTTGTTACGCGTTTTTGTTCATGGCTTTGGTCCCGCTTTGTTC  
AGAATGCTTTTAAATAAGCGGGGTTACCGGTTTGGTTAGCGAGA  
AGAGCCAGTAAAAGACGCAGTGACGGAGATGTCTGATG CAA  
TAT GGA CAA TTG GTT TCT TCT CTG AAT

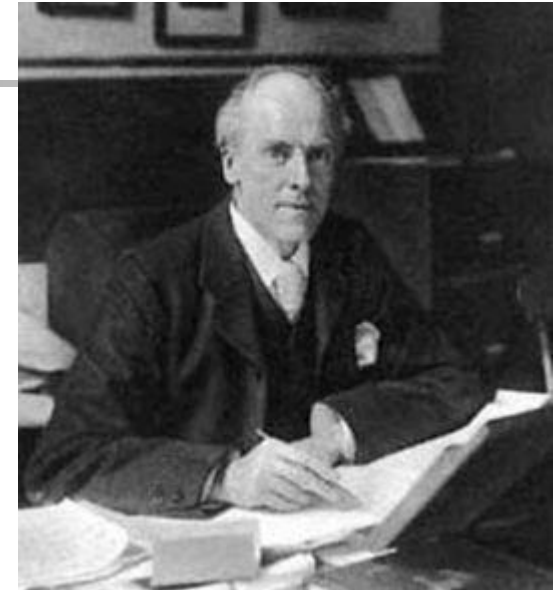
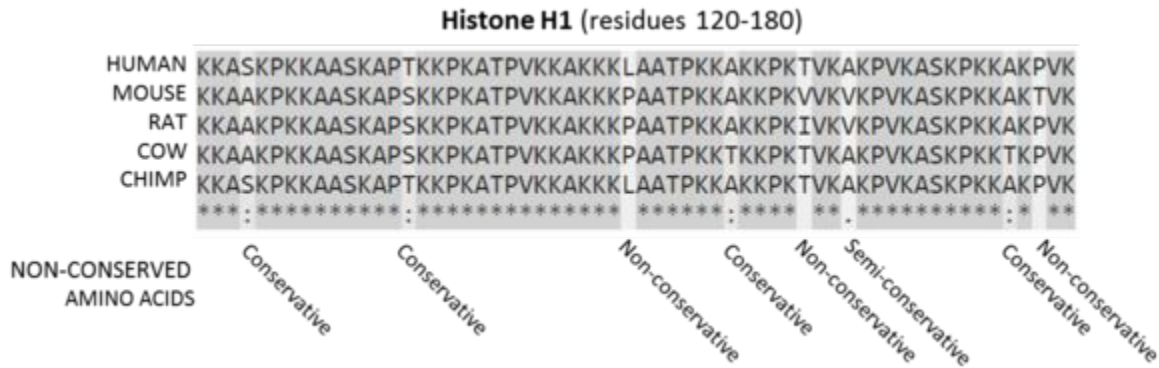
Transcription  
Start Site

Ribosome binding Site

ORF=Open Reading Frame  
CDS=Coding Sequence

TGAAAAACGTA

# Биоинформатика и принцип Пирсона



**Карл Пирсон**  
(1857 — 1936).



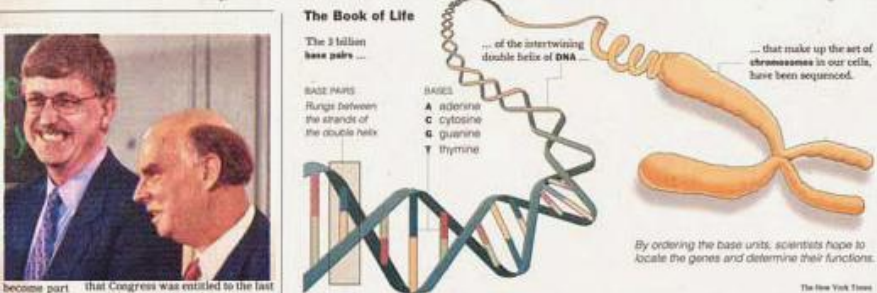
# Лидеры геномной революции 90-х



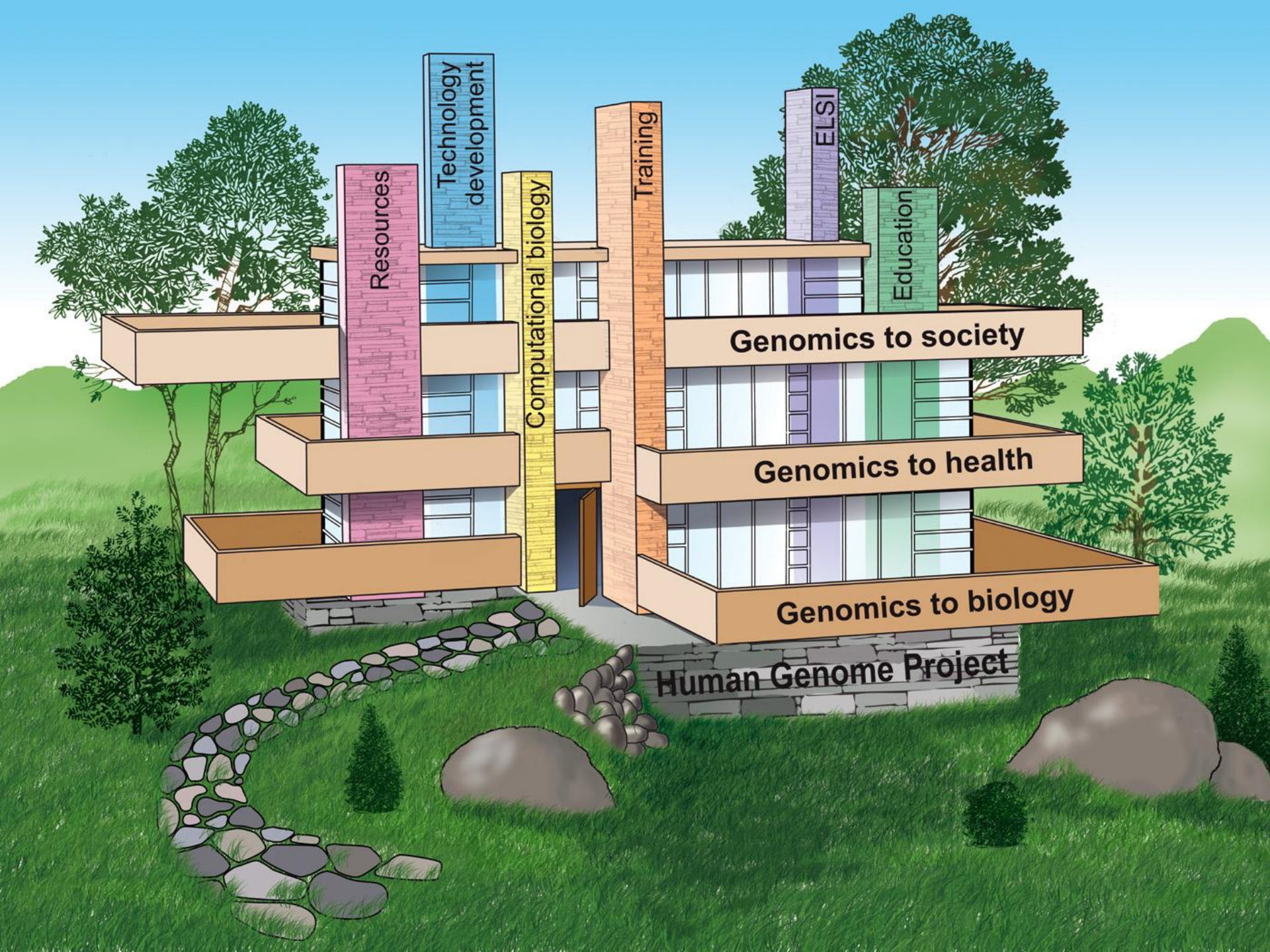
Francis S. Collins



J. Craig Venter







Resources

Technology development

Computational biology

Training

ELSI

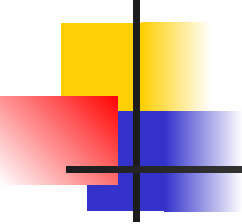
Education

Genomics to society

Genomics to health

Genomics to biology

Human Genome Project



# Геномика - биологии

---

- Выявление структуры variability генома - карта гаплотипов
- Секвенирование геномов многих видов
- Развитие технологий секвенирования, генотипирования, анализа экспрессии генов, протеомики
- Выявление всех функциональных элементов в геноме человека (генов, регуляторных участков)
- Выявление всех белков в клетке и их взаимодействий (**Протеом**)
- Разработка моделей клетки и взаимодействия клеточных компонентов (**Целлом**)
- Выявление всех метаболитов и моделирование метаболических процессов (**Метаболом**)
- Развитие математических и компьютерных методов анализа генома и реализации генетической информации (**Биоинформатика**)



# Геномика - здравоохранению

---

- Выявление генетических и внешних факторов всех распространенных болезней
- Разработка подходов к ранней диагностике и профилактике болезней
- Разработка подходов к генотерапии наследственных болезней
- Разработка лекарственных препаратов «направленного действия» и генотип-специфических препаратов (Фармакогеномика)
- Разработка «молекулярной» классификации болезней человека



# Геномика – обществу

## Этические, юридические и социальные вопросы (ELSI)

- Доступ к генетической информации и защита от генетической дискриминации (*генетические межэтнические различия, накапливаются данные о том, что гены детерминируют IQ, самостоятельность и зависимость, чувствительность и толерантность к стрессам, агрессивность и сексуальность, спортивные возможности*)
- Патентование и лицензирование генетической информации
- Выявление взаимоотношений генетического разнообразия, расовой и этнической принадлежности
- Выявление генетических факторов, обуславливающих поведение
- Границы применения геномики в немедицинской области (клонирование)

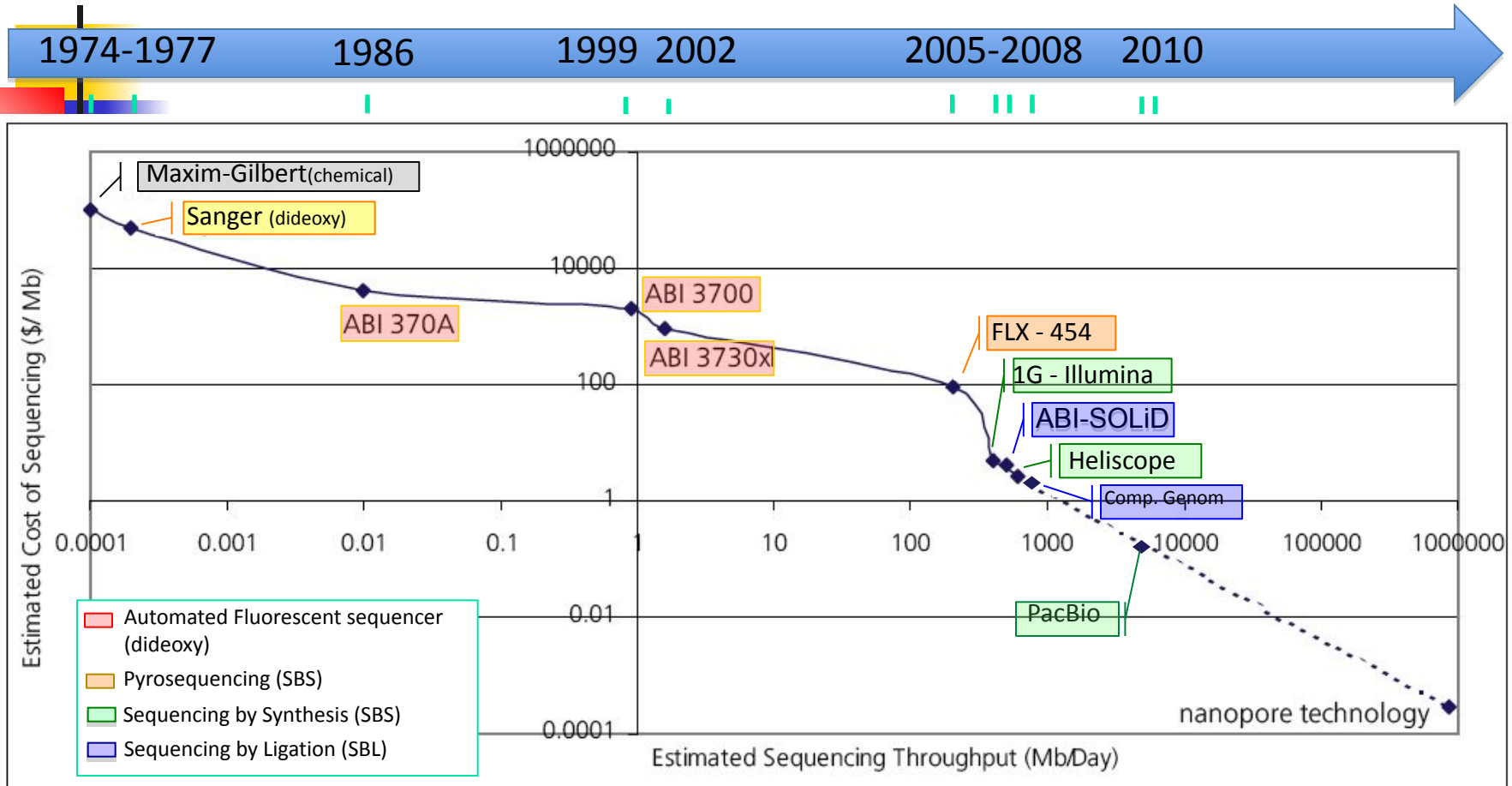


# Технологии секвенирования

---

- Технологии секвенирования первого поколения
  - Sanger Sequencing
- Технологии секвенирования второго поколения
  - Roche - 454
  - Illumina – GA II
  - SOLiD
- Технологии секвенирования третьего поколения
  - Helicose
  - PacBio
  - Illumina HiSeq, MiSeq
  - Ion Torrent
  - Oxford Nanopore

# DNA Sequencing Cost vs Throughput Timeline



## •\$10M genome x-prize

- 100 human genomes in  $\leq 10$  days
- 99.99999% accurate with 98% coverage
- \$10K/genome

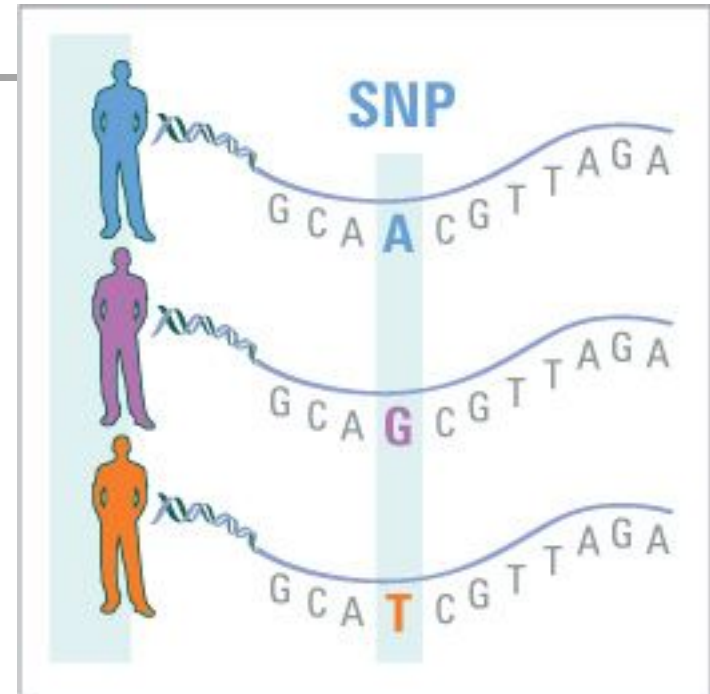
# Oxford Nanopore Technology



- ❖ Latest sequencing technology announced last month
- ❖ Size of USB drive
- ❖ May drive the next revolution in genomics
- ❖ Whole genome sequencing in 15 minutes for less than \$1,000
- ❖ Commercially available by the end of this year

# Вариации в геноме человека – Single nucleotide polymorphisms (SNPs)

- ~ 97 % генома двух человеческих особей идентичны
  - ~ 1% вариаций - SNPs
  - ~2% другие изменения – изменение количества копий, делеции.
- Сейчас идентифицировано 11-12 миллиона SNPs





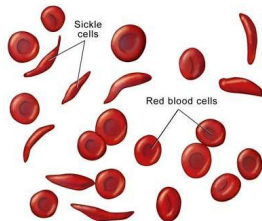
# SNPs могут приводить к изменению структуры и функции белка

## ■ SNPs приводят к фенотипическим вариациям

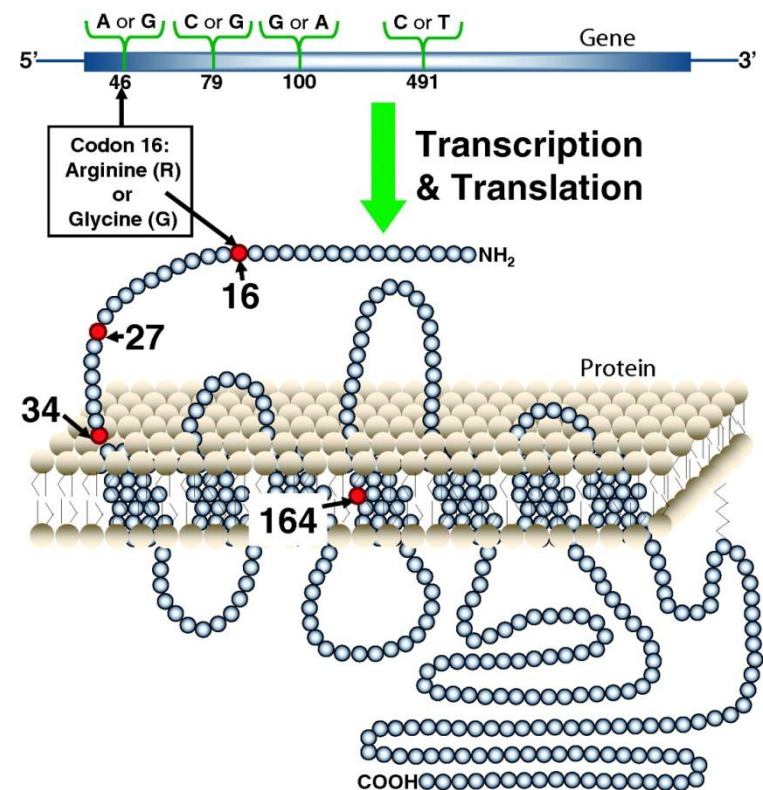
- Врождённая непереносимость алкоголя (полиморфизм в гене, кодирующем ацетальдегиддегидрогеназу)
- Серповидноклеточная анемия

Normal HbA      ATGGTGCACCTGACTCCTGTGGAGAAGTC  
 Disease HbS    ATGGTGCACCTGACTCCTG**A**GGAGAAGTC

Normal HbA    MVHLTPVEKSAVTA  
 Disease HbS   MVHLTP**E**EKSAVTA



## Polymorphisms of the Human $\beta_2$ AR

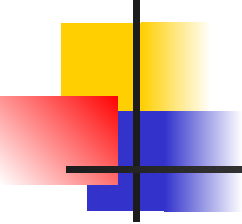




- According to the analysis, Osbourne has about 300,000 novel variants, a figure similar to that of other newly sequenced genomes.

*"One of the unusual things we found in your genome was a spelling in the regulatory segment of your ADH4 gene, which metabolises alcohol," said Dr Nathan. "It could make you more able to break down alcohol than the average person. Or less able."*

*I used to drink four bottles of cognac a day. I'm not sure I need a Harvard scientist to get to the bottom of that mystery.*



# Здравоохранение и персонализированная медицина

---

- **Диагностика, предсказание и предотвращение**
- **Разработка и применение лекарственных препаратов (фармакогеномика)**

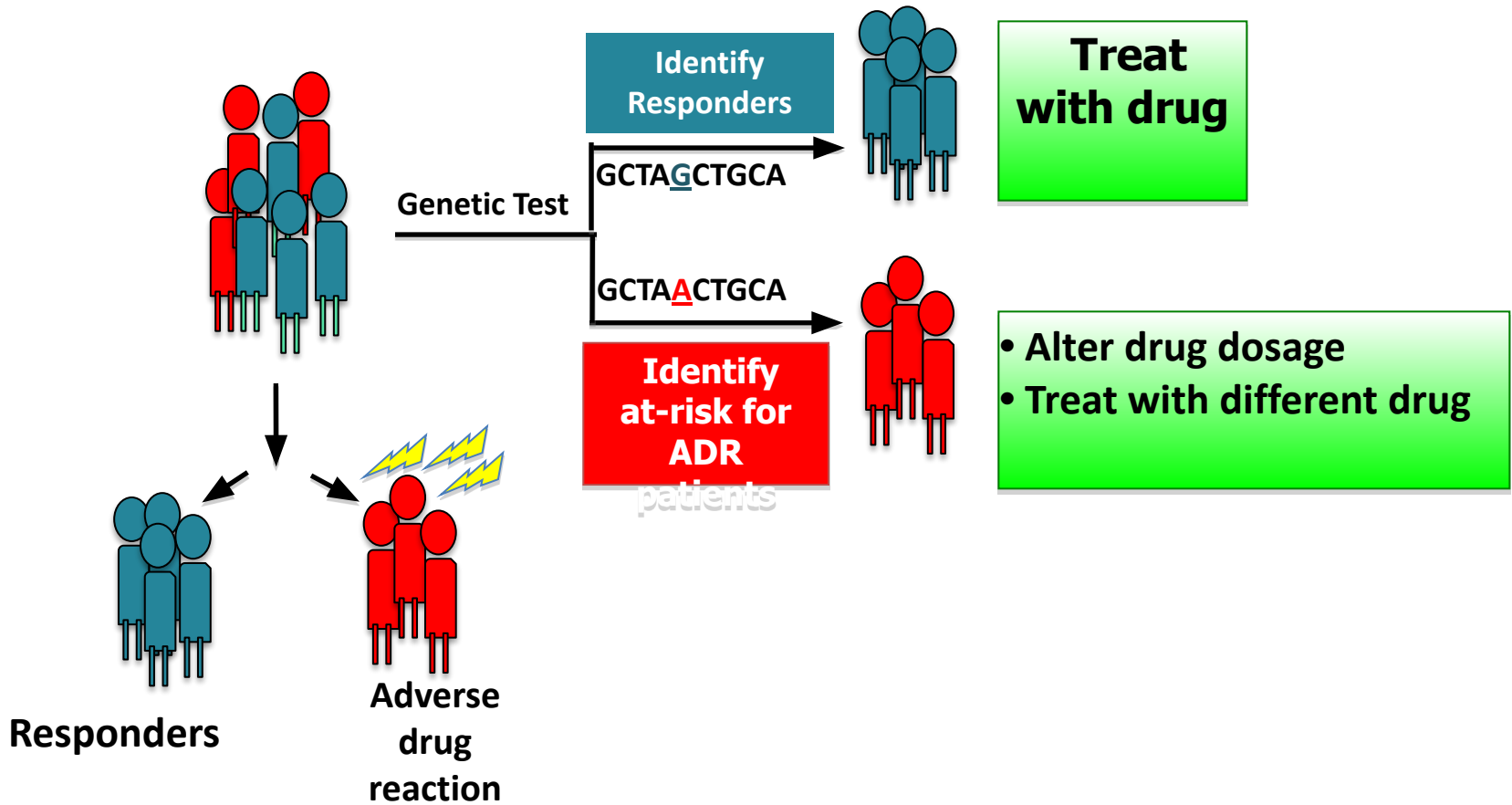


# Примеры индивидуальных ДНК-тестов

---

- **Тесты на носительство**
- **Тесты, оценивающие риски заболевания**
- **Тесты, оценивающие ответ организма на те или иные лекарственные препараты**

# Тесты, оценивающие ответ организма на те или иные лекарственные препараты



# Варфарин



- Широко используемый антикоагулянт
- Основной побочный эффект, связанный с дозировкой – кровоизлияния, причем иногда летальные.
- Определенные вариации (SNP) в генах *CYP2C9* и *VKORC1* напрямую влияют на побочный эффект этого препарата.

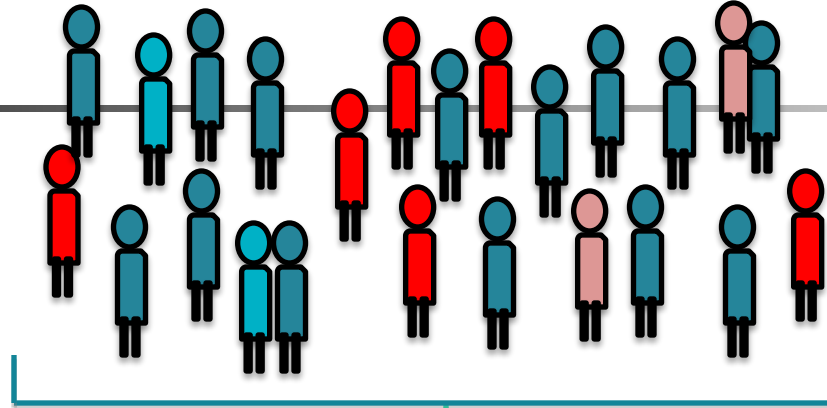
# Статины

- Используются для понижения содержания холестерина.
- Основной побочный эффект – миопатии.
- Определенные вариации (SNP) в гене *SLCOB1* приводят к побочному эффекту



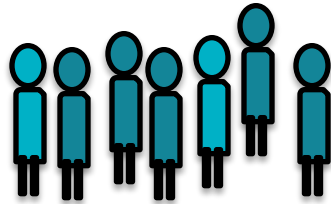
# Тесты, оценивающие риски заболевания

Individuals in population with varying risk levels for a disease



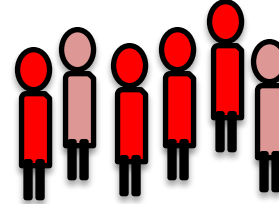
Genetic Testing

CGTCGATCGGC



Individuals with disease risk no greater than the population risk for a disease

CGTCGCTCGGC



Individuals with increased risk for a disease relative to the general population

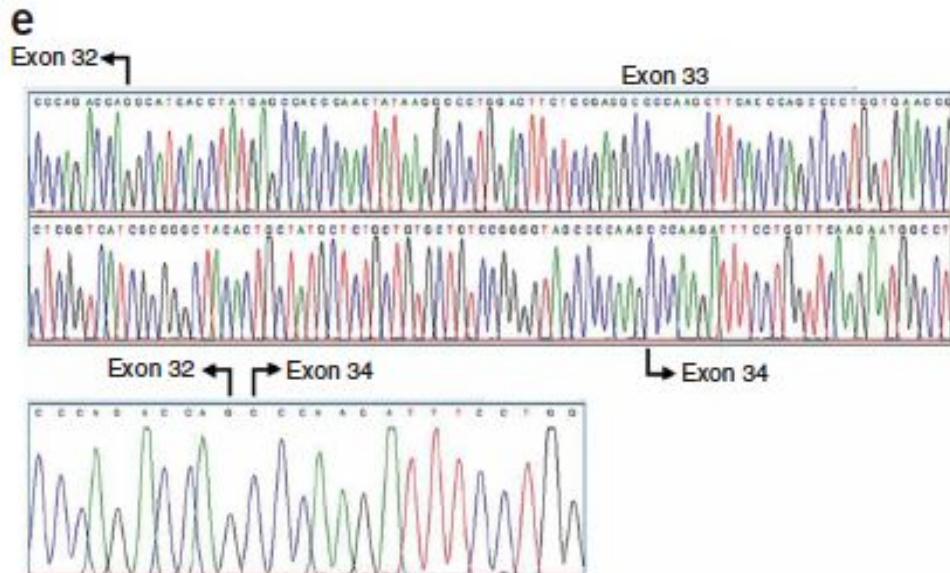
- Proactive monitoring for disease
- Preventive measures to mitigate disease occurrence



# Cardiomyopathy risk

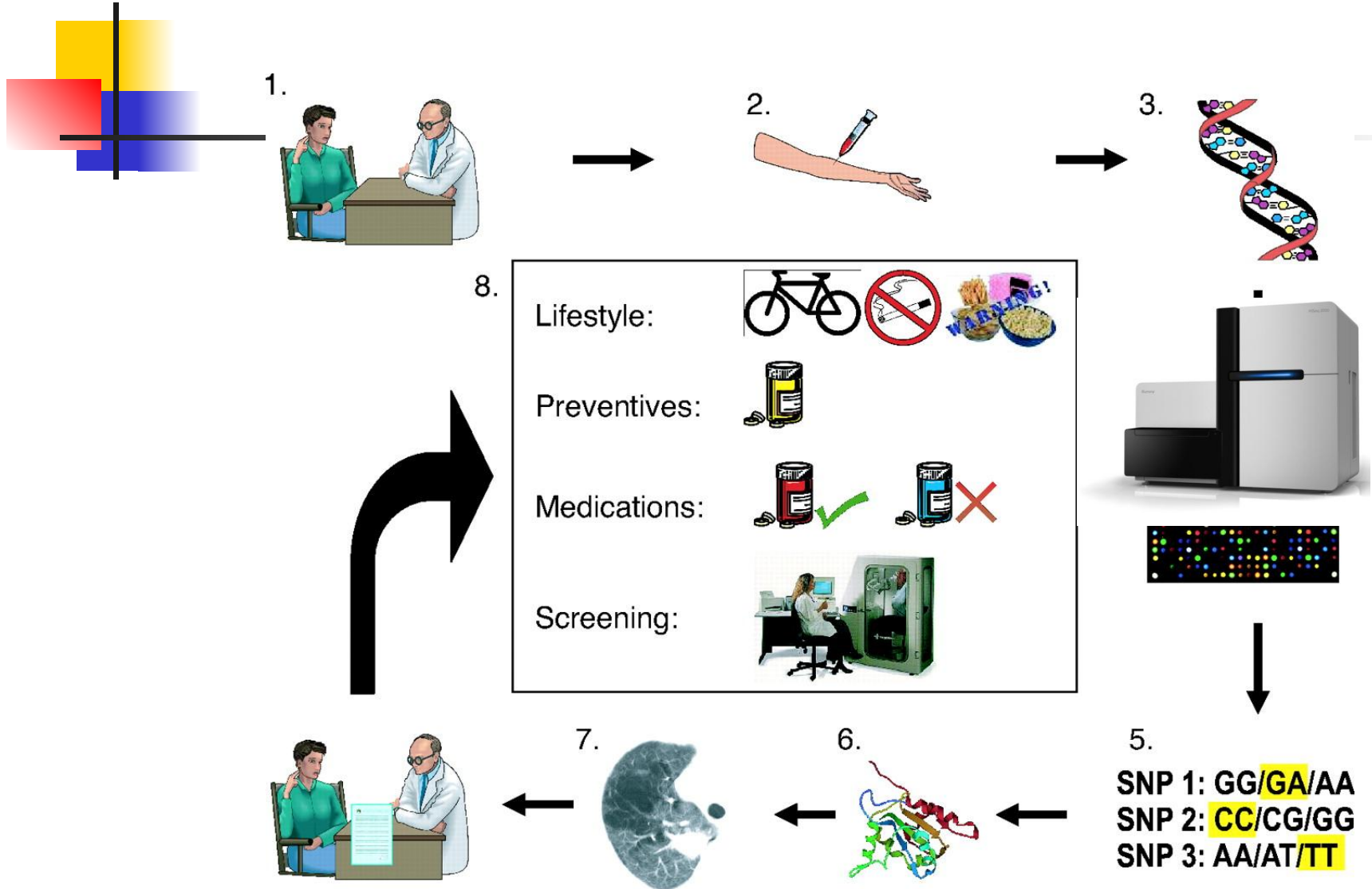
## A common *MYBPC3* (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia

Perundurair S Dhandapan<sup>1</sup>, Sakthivel Sadayappan<sup>2</sup>, Yali Xue<sup>3</sup>, Gareth T Powell<sup>3</sup>, Deepa Selvi Rani<sup>4</sup>, Prathiba Nallari<sup>5</sup>, Taranjit Singh Rai<sup>6</sup>, Madhu Khullar<sup>6</sup>, Pedro Soares<sup>7</sup>, Ajay Bahl<sup>6</sup>, Jagan Mohan Tharkan<sup>8</sup>, Pradeep Vaideeswar<sup>9</sup>, Andiappan Rathinavel<sup>10</sup>, Calambur Narasimhan<sup>11</sup>, Dharma Rakshak Ayapati<sup>12</sup>, Qasim Ayub<sup>13,18</sup>, S Qasim Mehdi<sup>13,18</sup>, Stephen Oppenheimer<sup>14</sup>, Martin B Richards<sup>7</sup>, Alkes L Price<sup>15</sup>, Nick Patterson<sup>16</sup>, David Reich<sup>16,17</sup>, Lalji Singh<sup>4</sup>, Chris Tyler-Smith<sup>3</sup> & Kumarasamy Thangaraj<sup>4</sup>





# Геномная медицина будущего



# Микробиом



- Человеческое тело - 10 триллионов клеток
- Но мы носим 90 триллионов клеток микроорганизмов
- Многие из них индивидуальны для каждого человека



# Метагеномика

---

- Метагеномика изучает набор генов всех микроорганизмов, находящихся в образце среды — *метагеном*. Метагеномный анализ позволяет определить видовое разнообразие исследуемого образца без необходимости выделения и культивирования микроорганизмов.







## Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,<sup>1\*</sup> Karin Remington,<sup>1</sup> John Heidelberg,<sup>3</sup> Aaron L. Halpern,<sup>2</sup> Doug Rusch,<sup>2</sup> Jonathan A. Eisen,<sup>3</sup> Dongying Wu,<sup>3</sup> Ian Paulsen,<sup>3</sup> Karen E. Nelson,<sup>3</sup> William Nelson,<sup>3</sup> Derrick E. Fouts,<sup>3</sup> Samuel Levy,<sup>2</sup> Anthony H. Knap,<sup>6</sup> Michael W. Lomas,<sup>6</sup> Ken Nealson,<sup>5</sup> Owen White,<sup>3</sup> Jeremy Peterson,<sup>3</sup> Jeff Hoffman,<sup>1</sup> Rachel Parsons,<sup>6</sup> Holly Baden-Tillson,<sup>1</sup> Cynthia Pfannkoch,<sup>1</sup> Yu-Hui Rogers,<sup>4</sup> Hamilton O. Smith<sup>1</sup>

<sup>1</sup>The Institute for Biological Energy Alternatives, <sup>2</sup>The Center for the Advancement of Genomics, 1901 Research Boulevard, Rockville, MD 20850, USA. <sup>3</sup>The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA.

<sup>4</sup>The J. Craig Venter Science Foundation Joint Technology Center, 5 Research Place, Rockville, MD 20850, USA. <sup>5</sup>University of Southern California, 223 Science Hall, Los Angeles, CA 90089–0740, USA. <sup>6</sup>Bermuda Biological Station for Research, Inc., 17 Biological Lane, St George GE 01, Bermuda.

\*To whom correspondence should be addressed. E-mail: jcventer@tcag.org

**We have applied “whole genome shotgun sequencing” to microbial populations collected *en mass* on tangential flow and impact filters from sea water samples collected from the Sargasso Sea near Bermuda. A total of 1.045 billion basepairs of non-redundant sequence was generated, annotated and analyzed to elucidate the gene content, and diversity and relative abundance of the organisms within these environmental samples. These data are estimated to derive from at least 1800 genomic species based on sequence relatedness, including 148 novel bacterial phylotypes. We have identified over 1.2 million new genes represented in these samples, including more than 782 new rhodopsin-like photoreceptors. Variation in species present and stoichiometry suggests substantial oceanic microbial diversity.**

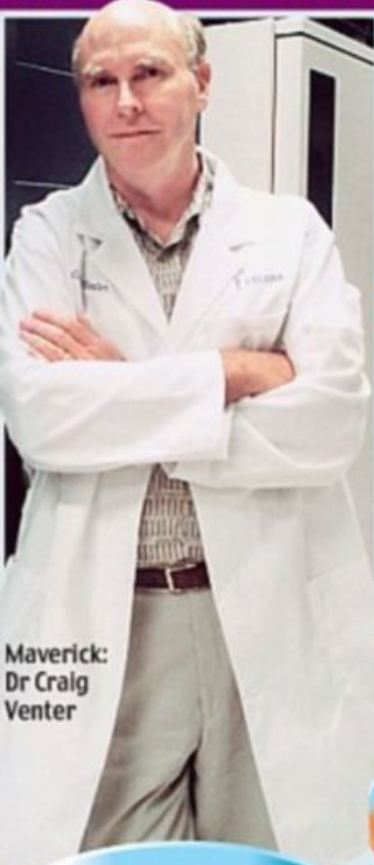
environment. Further, we concentrated on the genetic material captured on filters sized to isolate primarily microbial inhabitants of the environment, leaving detailed analysis of dissolved DNA and viral particles on one end of the size spectrum, and eukaryotic inhabitants on the other, for subsequent studies.

**The Sargasso Sea.** The northwest Sargasso Sea, at the Bermuda Atlantic Time-series Study site (BATS), is one of the best-studied and arguably most well-characterized regions of the global ocean. The Gulf Stream represents the western and northern boundaries of this region and provides a strong physical boundary separating the low nutrient, oligotrophic, openocean from the more nutrient-rich waters of the U.S. continental shelf. The Sargasso Sea has been intensively studied as part of the 50-year time-series of ocean physics and





# HOW TO MAKE ARTIFICIAL LIFE



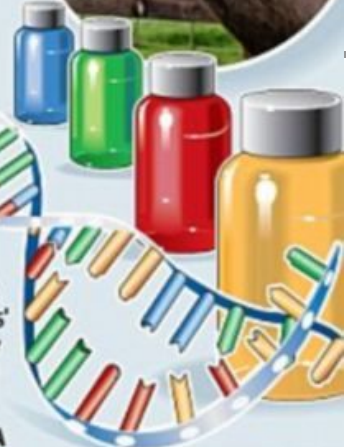
Maverick:  
Dr Craig  
Venter

**1** Entire DNA of *Mycoplasma mycoides*, a bug that usually infects goats, is decoded.

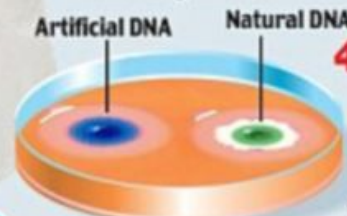


Synthetic  
DNA code

**2** Researchers buy fragments of DNA from a mail order catalogue. Each of the four bottles of chemicals contains a section of the code.



**3** The fragments are put into yeast, which 'stitches' them together, gradually building a synthetic copy of the original DNA.



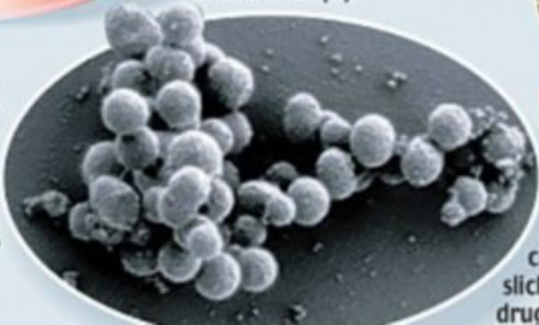
**4** The artificial DNA is put into a recipient bacterium, which then grows and divides, creating two daughter cells, one with the artificial DNA and one with the natural DNA.



**5** Antibiotics in the petri dish kill the bacterium with the natural DNA, leaving the one with the synthetic DNA to multiply.



**6** Within just a few hours, all traces of the recipient bug are wiped out and bugs with artificial DNA thrive. New life has been created.



**7** Possible uses are bugs capable of producing clean fuels and sucking carbon dioxide out of the atmosphere. Also microbes capable of mopping up oil slicks (above) or generating drugs, including the flu vaccine

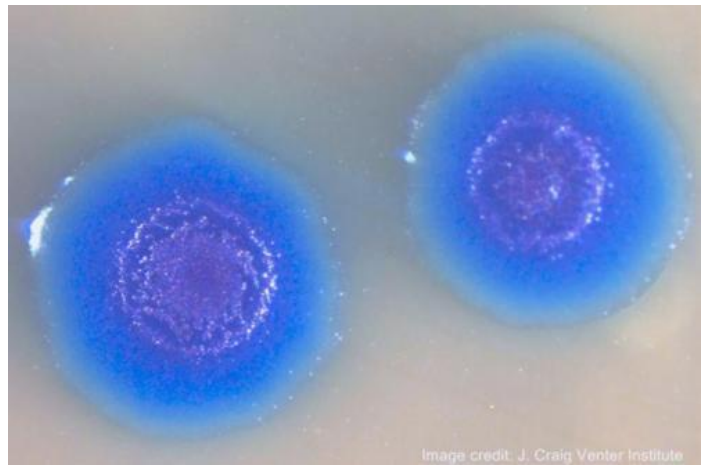


Image credit: J. Craig Venter Institute

