



# Zenome.io

IT на службе у современной науки

*к.б.н. Николай Александрович Кулемин*  
Смоленск, 23.06.2018



Об авторе

# Кулемин Николай Александрович

В 2013 году окончил Факультет биологической и медицинской физики МФТИ ГУ с отличием, кандидат биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.



С 2008 по 2011 годы:

Младший Научный Сотрудник Центра Фотохимии РАН.  
**Флуориметрия, фотоспектрометрия и прочее.**

С 2011 по настоящее время:

Научный Сотрудник ФНКЦ Физико-Химической медицины ФМБА России.  
Руководитель проекта «Zenome Platform» (zenome.io).  
Автор более 20 высокоцитируемых международных научных публикаций.  
Генотипирование, секвенирование, базы данных, математические модели: **генетика человека**.  
Разработка Web-Сервисов, сетевые технологии, обслуживание вычислительных систем: **бионформатика** в самом широком смысле.

# А как Вы оказались в IT?

Увлекались компьютерной техникой / программированием в школе.

Поступили в технический ВУЗ.

Осознали, что в науке мало платят и занялись прикладным коммерческим программированием или заинтересовались реализацией интересного проекта.

**Наверное, каждому IT-специалисту хотелось разработать что-то новое и общественно-полезное для человечества.**

**Возможно, стоит присоединиться к наукоемкому стартапу?**





В 2016 году фонд «Сколково» опубликовал 100 профессий по отраслям, которые будут очень востребованы в 2030 году:

## **Биомедицинская отрасль:**

Клинический биоинформатик

IT-генетик

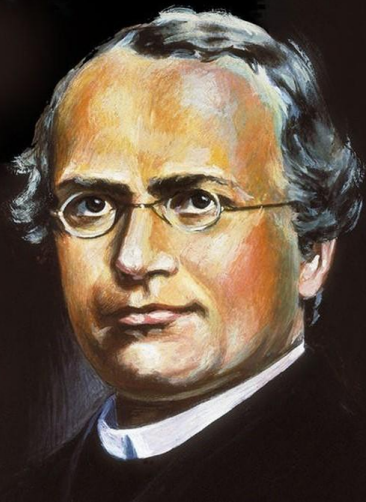
Молекулярный диетолог

Эксперт персонифицированной медицины

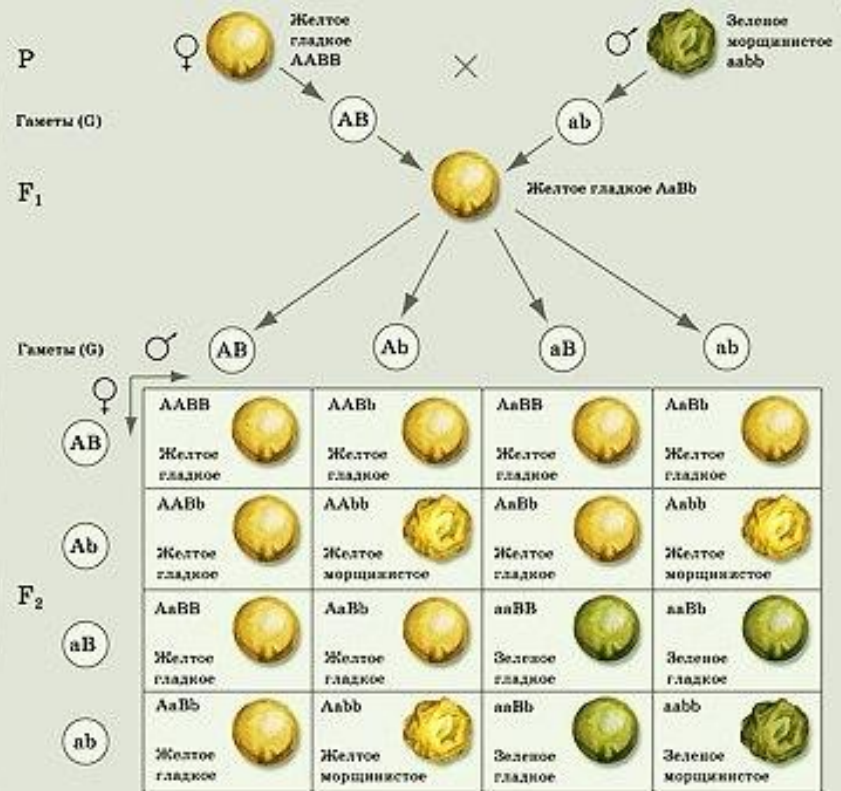
Генетический консультант

Оператор омиксного оборудования

**Кто все эти люди?  
Зачем нужны эти профессии  
человечеству?**



## ДИГИБРИДНОЕ СКРЕЩИВАНИЕ. НЕЗАВИСИМОЕ НАСЛЕДОВАНИЕ ПРИЗНАКОВ



### ФЕНОТИПИЧЕСКИЙ РАДИКАЛ ДИГИБРИДНОГО СКРЕЩИВАНИЯ

$9A\_B\_ : 3A\_bb : 3aaB\_ : 1aabb$

Желтые гладкие      Желтые морщинистые      Зеленые гладкие      Зеленые морщинистые

$A\_ : aa = 3 : 1$

$B\_ : bb = 3 : 1$

Желтые Зеленые

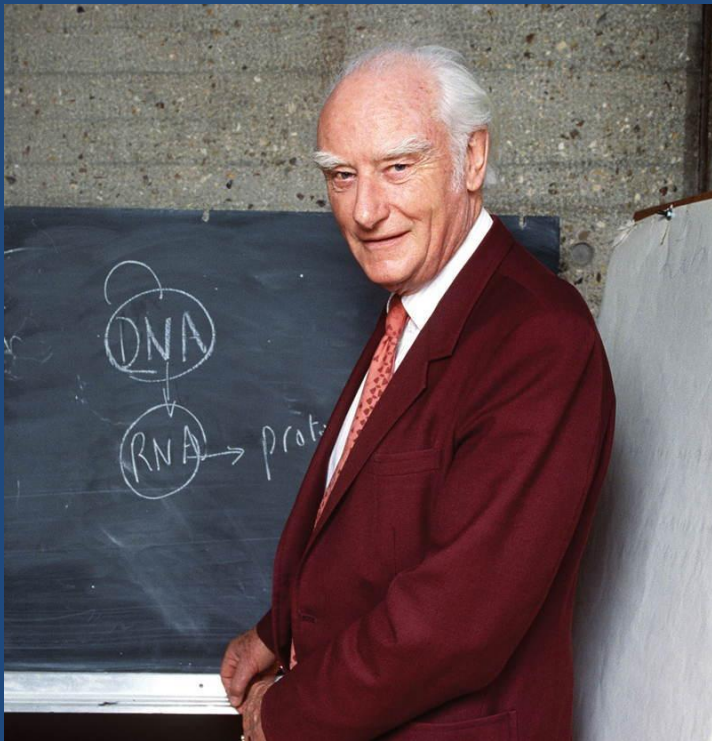
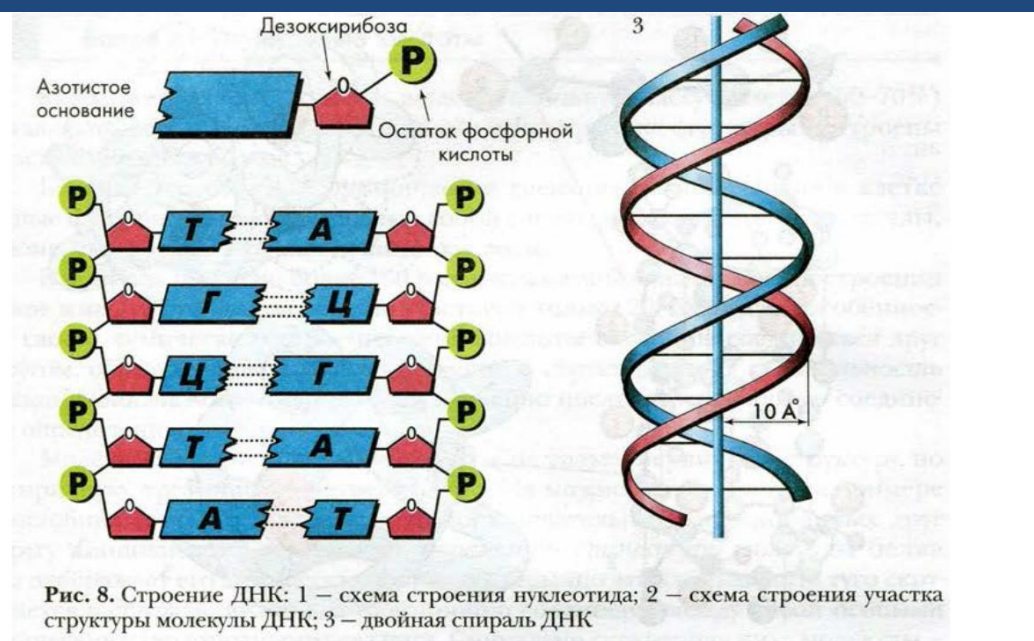
Гладкие Морщинистые

**Грегор Иоганн Мендель** (нем. Gregor Johann Mendel; 20 июля 1822, Хейнцендорф, Силезия, Австрийская империя — 6 января 1884, Брюнн, Австро-Венгрия) — австрийский биолог и ботаник, монах-августинец, аббат.

Основоположник учения о наследственности, позже названного по его имени менделизмом. Открытие им закономерностей наследования моногенных признаков (эти закономерности известны теперь как Законы Менделя) стало первым шагом на пути к современной генетике.







**Джеймс Дьюи Уотсон** (англ. James Dewey Watson, род. 6 апреля 1928 года в Чикаго, Иллинойс) — американский биолог. Лауреат Нобелевской премии по физиологии и медицине 1962 года — совместно с Фрэнсисом Криком и Морисом Х. Ф. Уилкинсом за открытие структуры молекулы ДНК.

**Фрэнсис Крик** (англ. Francis Harry Compton Crick; 8 июня 1916, Нортгемптон, Англия — 28 июля 2004, Сан-Диего, Калифорния, США) — британский молекулярный биолог, биофизик и нейробиолог. Лауреат Нобелевской премии по физиологии и медицине 1962 года — совместно с Джеймсом Д. Уотсоном и Морисом Х. Ф. Уилкинсом с формулировкой «за открытия, касающиеся молекулярной структуры нуклеиновых кислот и их значения для передачи информации в живых».

# Математика генома человека

## Statistics

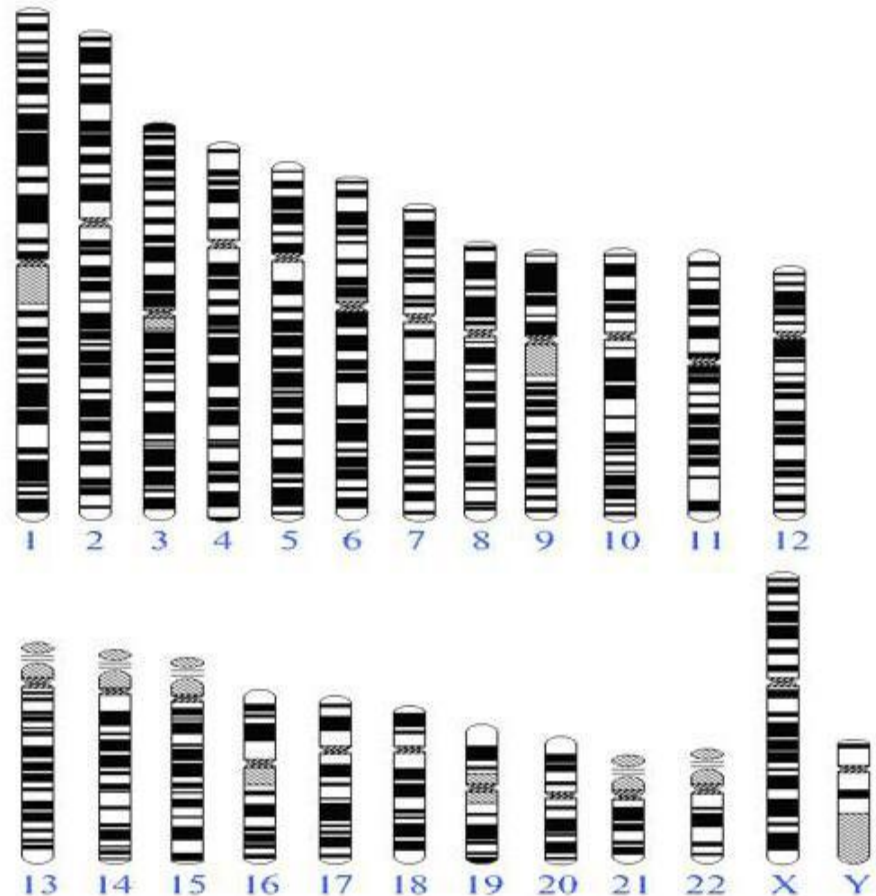
### Summary

<b>Assembly:</b>	GRCh37.p10, Feb 2009
<b>Database version:</b>	71.37
<b>Base Pairs:</b>	3,320,602,131
<b>Golden Path Length:</b>	3,101,804,739
<b>Genebuild by:</b>	Ensembl
<b>Genebuild method:</b>	Full genebuild
<b>Genebuild started:</b>	Jul 2010
<b>Genebuild released:</b>	Apr 2011
<b>Genebuild last updated/patched:</b>	Feb 2013

### Gene counts (Primary assembly)

<b>Coding genes:</b>	20,806
<b>Non coding genes:</b>	22,476
<b>Pseudogenes:</b>	13,413
<b>Gene transcripts:</b>	194,010

## ВСЕ ХРОМОСОМЫ ЧЕЛОВЕКА





# Проявления особенностей

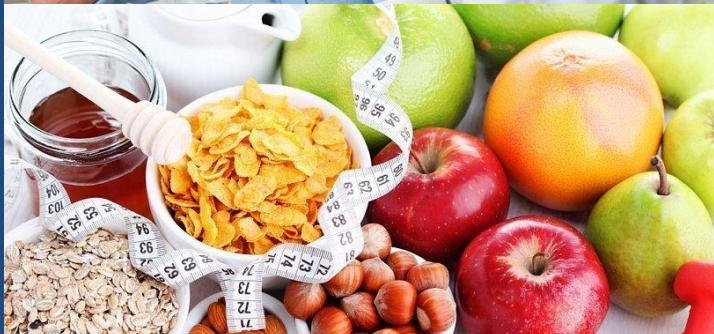
## генома

В развитии любого проявления присутствует вклад из трех компонент:

- 1) Генетика – наследственность, не изменяемая по жизни.
- 2) Образ жизни – регулярные действия человека, возможно изменение.
- 3) Случайность – неконтролируемые процессы, такие как падение кирпича или попадание вируса.

Имея информацию о своем геноме возможно:

- 1) Найти своих кровных родственников.
- 2) Заранее иметь представление о возможных наследственных заболеваниях и рисках их развития.
- 3) Контролировать диету для наилучшего самочувствия и поддержания формы.
- 4) Выбирать те виды спорта и сферы деятельности, к которым есть наиболее выраженные предрасположенности.





# Пример отчета (HTML – форма)



## #DODECAD: Этнический анализ с помощью DIYDodecad v2.1

Источник: <http://dodecad.blogspot.ru>

Проанализировано: 5607/165615=3% модели.

Анализируемые гены: *Мультигенная модель*

Описание нозологии: Этнический тест по современным популяциям. Показывает долю каждой коренной популяции в Вашем происхождении.

Популяция	Доля в %
Восточно-Центральная Европа (Польша, Словакия, Чехия)	38%
Восточная Европа (Белорусь, Россия)	24%
Финно-скандинавы	22%
Восточные Балканы (Болгария, Македония)	16%

[Оглавление](#)

## Онкологические заболевания (C00-D48)

### #CAN01: Рак молочной железы, наследуемый риск

Источник: Кулемин, SNPedia ([http://www.snpedia.com/index.php/Breast\\_cancer](http://www.snpedia.com/index.php/Breast_cancer))

Проанализировано: 27/35=77% модели.

Анализируемые гены: *BRCA1, BRCA2, ATM, TP53* и др.

Описание нозологии: Рак молочной железы, наследуемый риск

Персональные рекомендации: Ниже среднепопуляционного уровня риска

Список определившихся полиморфизмов:

Ген	Полиморфизм	Хромосома	Позиция	Генотип	Частота	Качество
TP53	rs1042522	chr17	7579472	CC	0%	<u>BEST</u>
CASP8	rs1045485	chr2	202149589	GG	0%	<u>MEDIUM</u>
BRCA2	rs11571746	chr13	32945108	TT	0%	<u>MEDIUM</u>
BRCA2	rs11571747	chr13	32945172	AA	0%	<u>MEDIUM</u>

# Пример отчета (HTML – форма)



AIP rs267606549 chr11 67256808 -- -- I II MEDIUM

Ген AIP проанализирован с высокой точностью (57.58 шт/кБ).

[Оглавление](#)

## #MON009: Болезнь Фабри

**Источник:** Кулемин, Сколтех, 2016.

**Проанализировано:** 545/548=99% модели.

**Анализируемые гены:** GLA

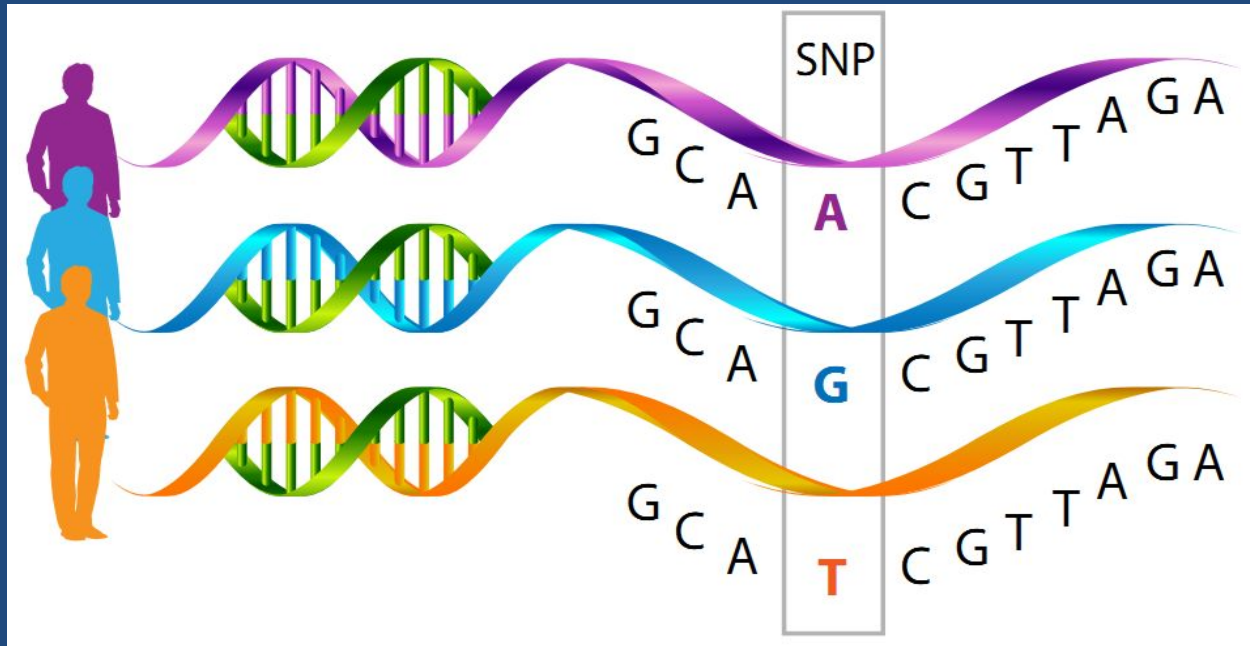
**Описание нозологии:** Болезнь Фабри (болезнь Андерсона-Фабри) - редкое заболевание из группы лизосомных болезней накопления. Данное заболевание вызвано нарушением метаболизма сфинголипидов и обладает широким спектром клинических симптомов. Частота встречаемости заболевания составляет 1 на 70000 пациентов.

**Персональные рекомендации:** Риск заболевания отсутствует.

[Список определившихся полиморфизмов и мутаций, известных как патогенные](#)

Ген	Полиморфизм	Хромосома	Позиция	Нуклеотидная	Аминокислотная	Референс	Генотип	Качество
GLA	mut58921	chrX	100652800	--	--	--	--	NA
GLA	mut58920	chrX	100652809	c.1277_1278delAA	--	I	II	MEDIUM
GLA	rs112341092	chrX	100652843	c.1244T>C	Leu415Pro	A	AA	MEDIUM
GLA	mut58783	chrX	100652846	c.1241T>C	Leu414Ser	A	AA	MEDIUM
GLA	mut58919	chrX	100652851	c.1235_1236delCT	--	I	II	MEDIUM
GLA	mut58782	chrX	100652855	c.1232G>A	Gly411Asp	C	CC	MEDIUM
GLA	mut58779	chrX	100652858	c.1229C>T	Thr410Ile	G	GG	MEDIUM
GLA	rs104894852	chrX	100652859	c.1228A>G	Thr410Ala	T	TT	MEDIUM
GLA	mut58775	chrX	100652862	c.1225C>G	Pro409Ala	G	GG	MEDIUM
GLA	mut58918	chrX	100652864	c.1223delA	--	I	II	MEDIUM
GLA	mut58774	chrX	100652867	c.1220T>A	Ile407Lys	A	AA	MEDIUM
GLA	mut58917	chrX	100652873	c.1212_1214delAAG	--	I	II	MEDIUM
GLA	mut58773	chrX	100652879	c.1208T>C	Leu403Ser	A	AA	MEDIUM
GLA	mut58916	chrX	100652879	c.1208delT	--	I	II	MEDIUM
GLA	mut58953	chrX	100652884	c.1197_1202dupGACTTC	--	D	DD	MEDIUM
GLA	mut58772	chrX	100652885	c.1202C>G	Ser401Term	G	GG	MEDIUM
GLA	mut58954	chrX	100652885	c.1201dupT	--	D	DD	MEDIUM
GLA	mut58771	chrX	100652891	c.1196G>A	Trp399Term	C	CC	MEDIUM
GLA	mut58915	chrX	100652893	c.1194delA	--	I	II	MEDIUM
GLA	rs104894844	chrX	100652895	c.1192G>A	Glu398Lys	C	CC	MEDIUM
GLA	mut58914	chrX	100652899	c.1188delC	--	I	II	MEDIUM
GLA	mut58952	chrX	100652899	c.1187dupT	--	D	DD	MEDIUM
GLA	mut58913	chrX	100652900	c.1187delT	--	I	II	MEDIUM
GLA	mut58768	chrX	100652903	c.1184G>C	Gly395Ala	C	CC	MEDIUM
GLA	mut58911	chrX	100652908	c.1176_1179delGAAG	--	I	II	MEDIUM

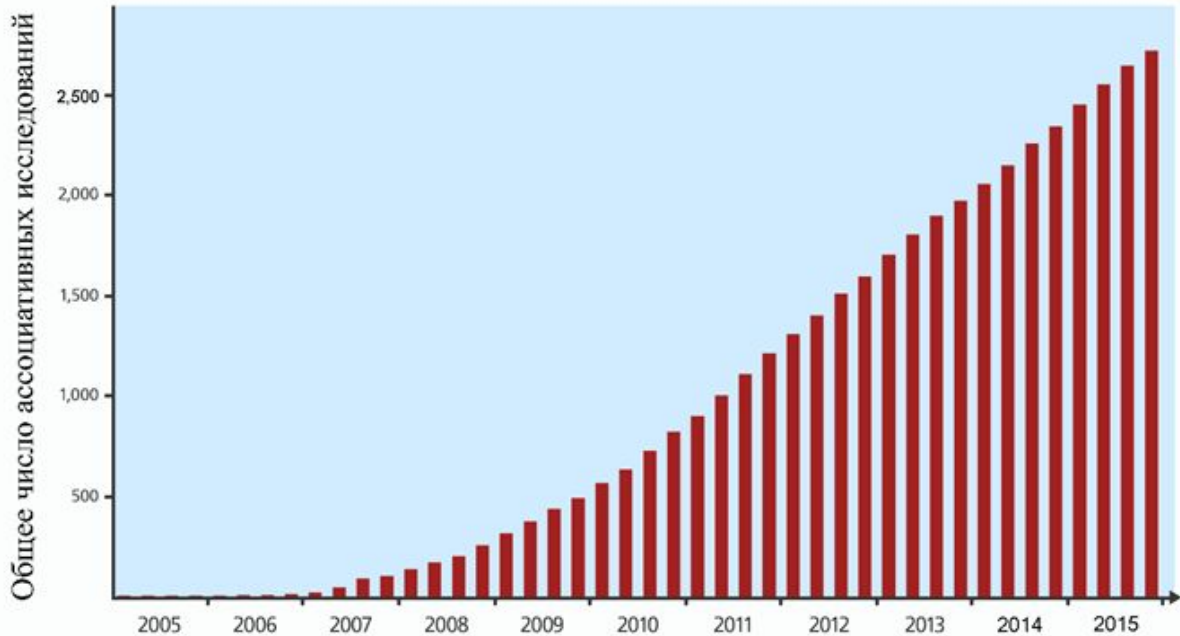
# Геномные данные – IT взгляд



- 4 нуклеотида, четверичная система счисления.
- «Программа», с помощью которой записано поведение организма в тех или иных ситуациях.



# Ассоциативные исследования в генетике



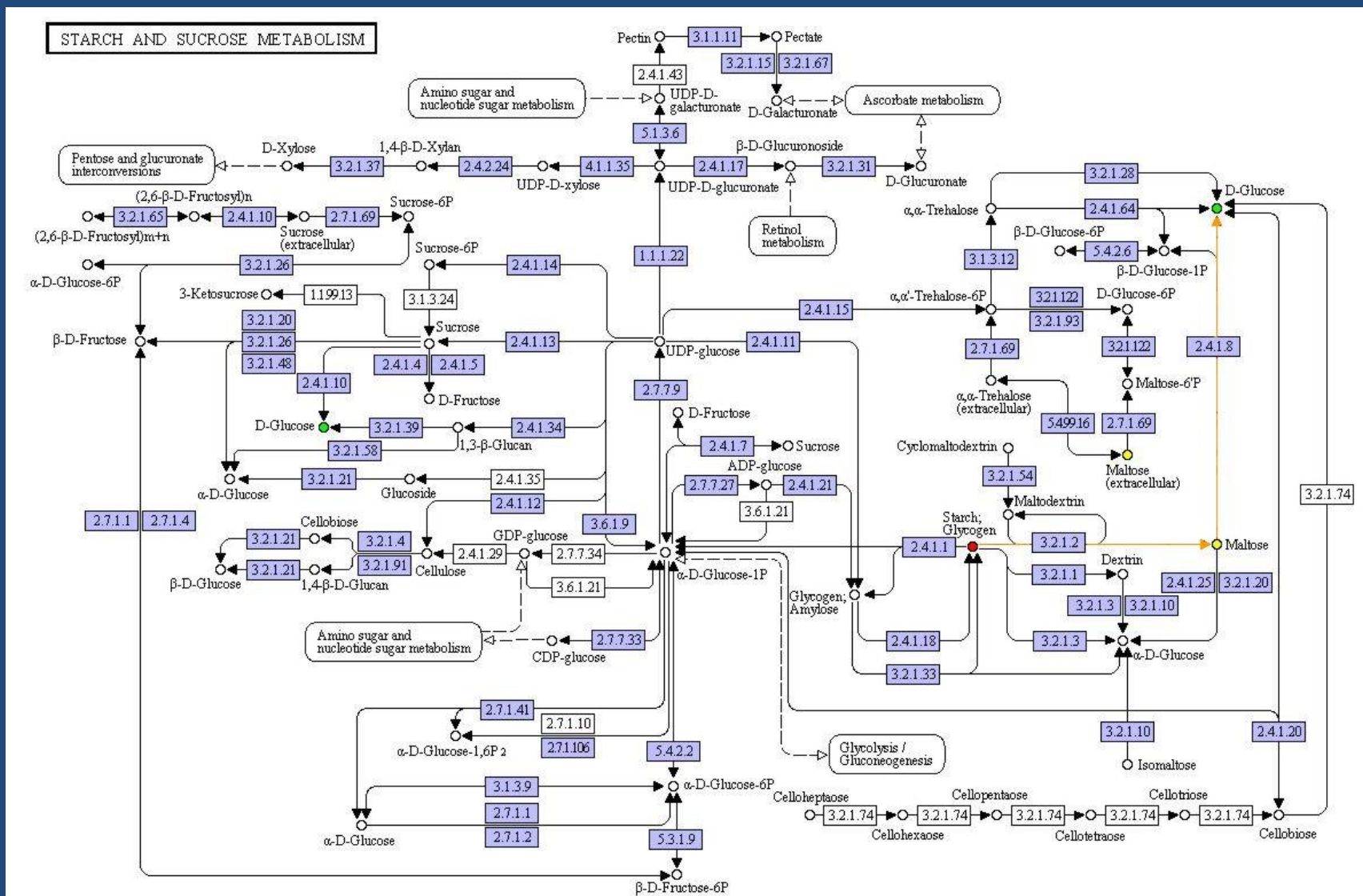
Рост числа ассоциативных исследований по годам



**Особую актуальность** имеют генетические исследования стратифицированных и ограниченных групп лиц с уникальными физиологическими характеристиками, например, **групп профессиональных спортсменов.**

# Ассоциативные исследования в генетике

## Анализ метаболических путей



Мутация => Проблемы с метаболизмом => Заболевание

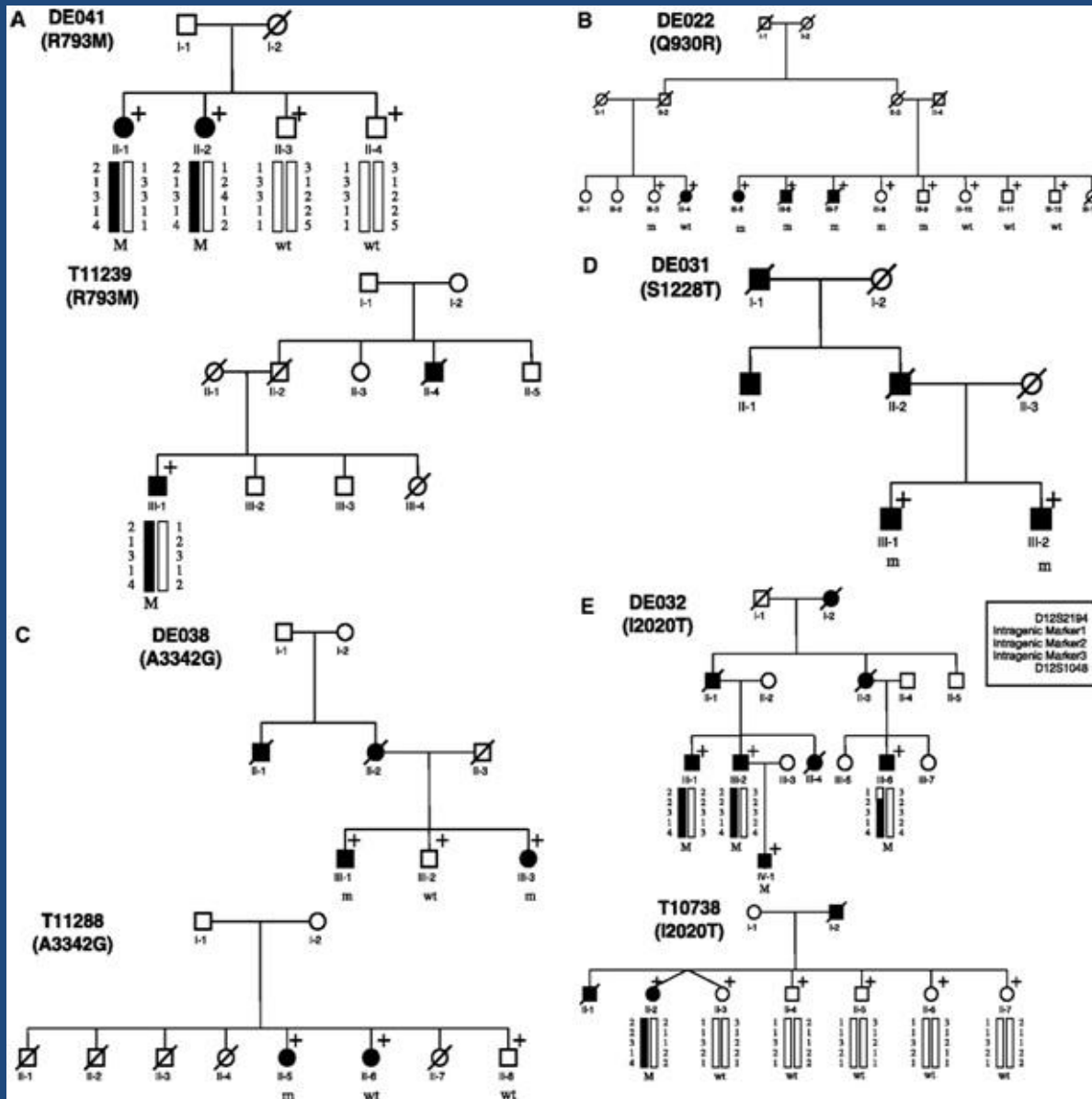
# Ассоциативные исследования в генетике

## Семейный анализ

Изучение родственников и поиск одинаковых фенотипов.

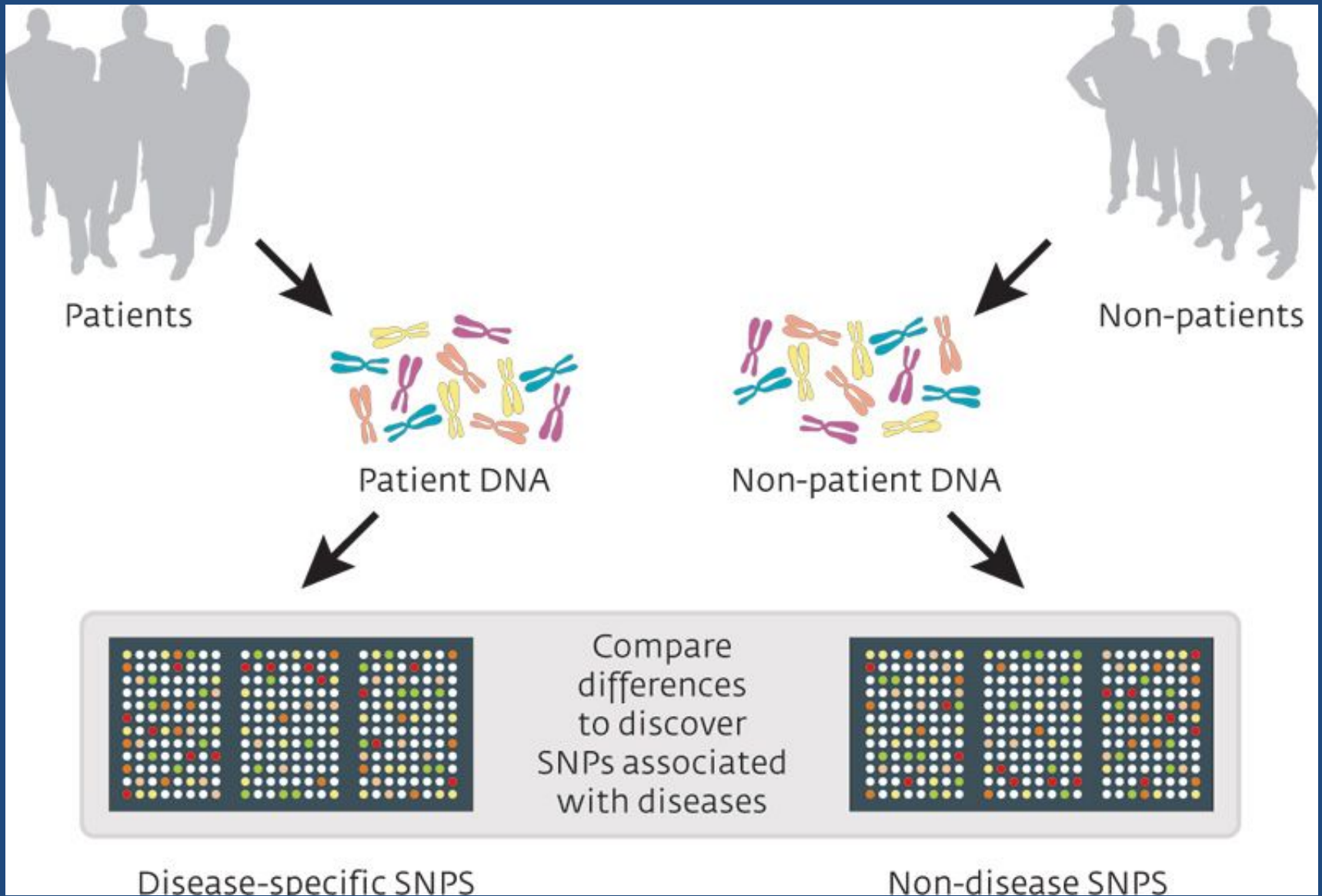
Генотипирование и поиск одинаковых мутаций.

Попытка интерпретации обнаруженной мутации с использованием метаболических цепей.





# Ассоциативные исследования в генетике GWAS



# Ассоциативные исследования в генетике

## GWAS

	<b>GG</b>	<b>GT</b>	<b>TT</b>	<b>Total</b>
<b>Cases</b>	$r_0$	$r_1$	$r_2$	$R$
<b>Controls</b>	$s_0$	$s_1$	$s_2$	$S$
<b>Total</b>	$n_0$	$n_1$	$n_2$	$N$

### Odds Ratio (OR)

Contingency (or 2 x 2) Table

	Cases	Controls	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

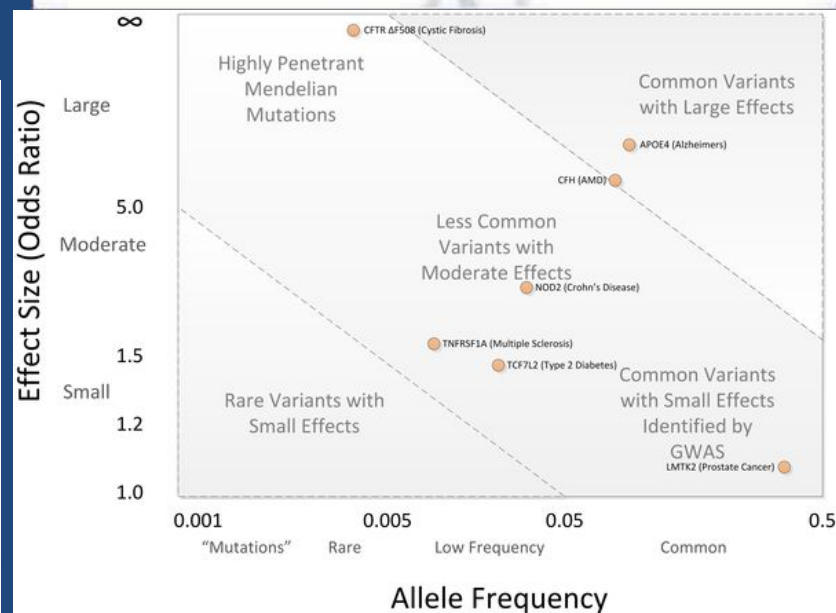
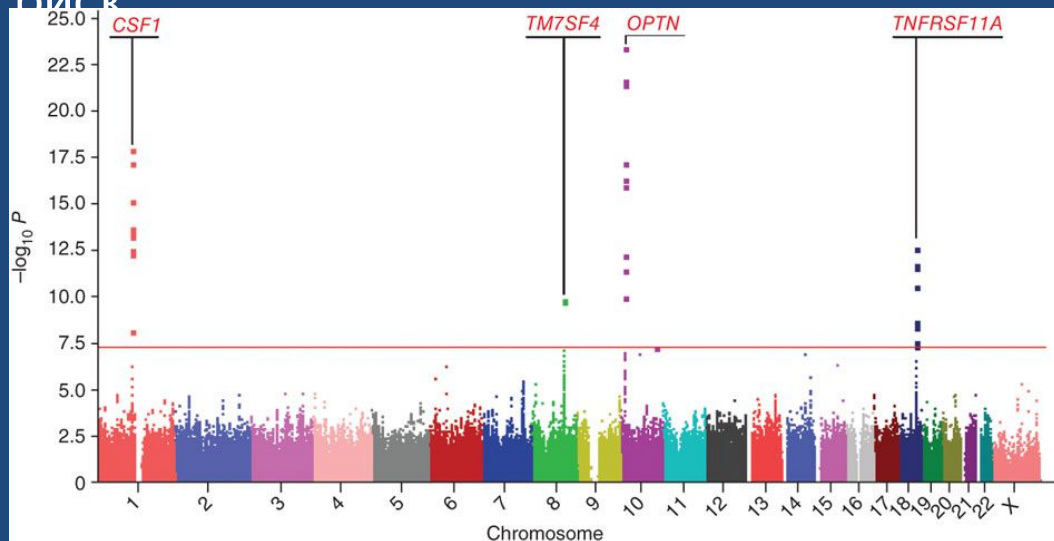
$$OR = (a/c) / (b/d)$$

$$= (a*d) / (b*c)$$

OR = 1: ассоциации нет.

OR > 1: обнаруженная аллель повышает риск.

OR < 1: обнаруженная аллель понижает риск.



# Использование геномных данных



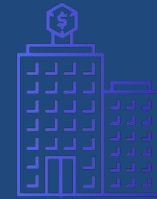
## Для пользователей:

- Улучшение качества жизни
  - Получение важной медицинской информации
  - Подбор партнера для зачатия здорового ребенка
- Монетизация собственной генетической информации



## Для науки:

- Проведение ассоциативных и статистических исследований



## Для корпораций:

- Разработка и продажа препаратов
- Электронная коммерция
- Сервисы знакомств
- Таргетная реклама
- Разработка персонализированных продуктов



# У кого сейчас есть геномные данные?

Геномный тест стоит от 1 до 350 тыс. рублей, в зависимости от числа прочитанных нуклеотидов.

В случае, если требуется подтвердить наследственный диагноз стоимость теста невысока, достаточно «прочитать» от нескольких нуклеотидов до нескольких генов.

Для «здоровых» людей информативную ценность несут тесты стоимостью от 50 тыс. рублей («читается» порядка 30 – 100 млн. нуклеотидов).

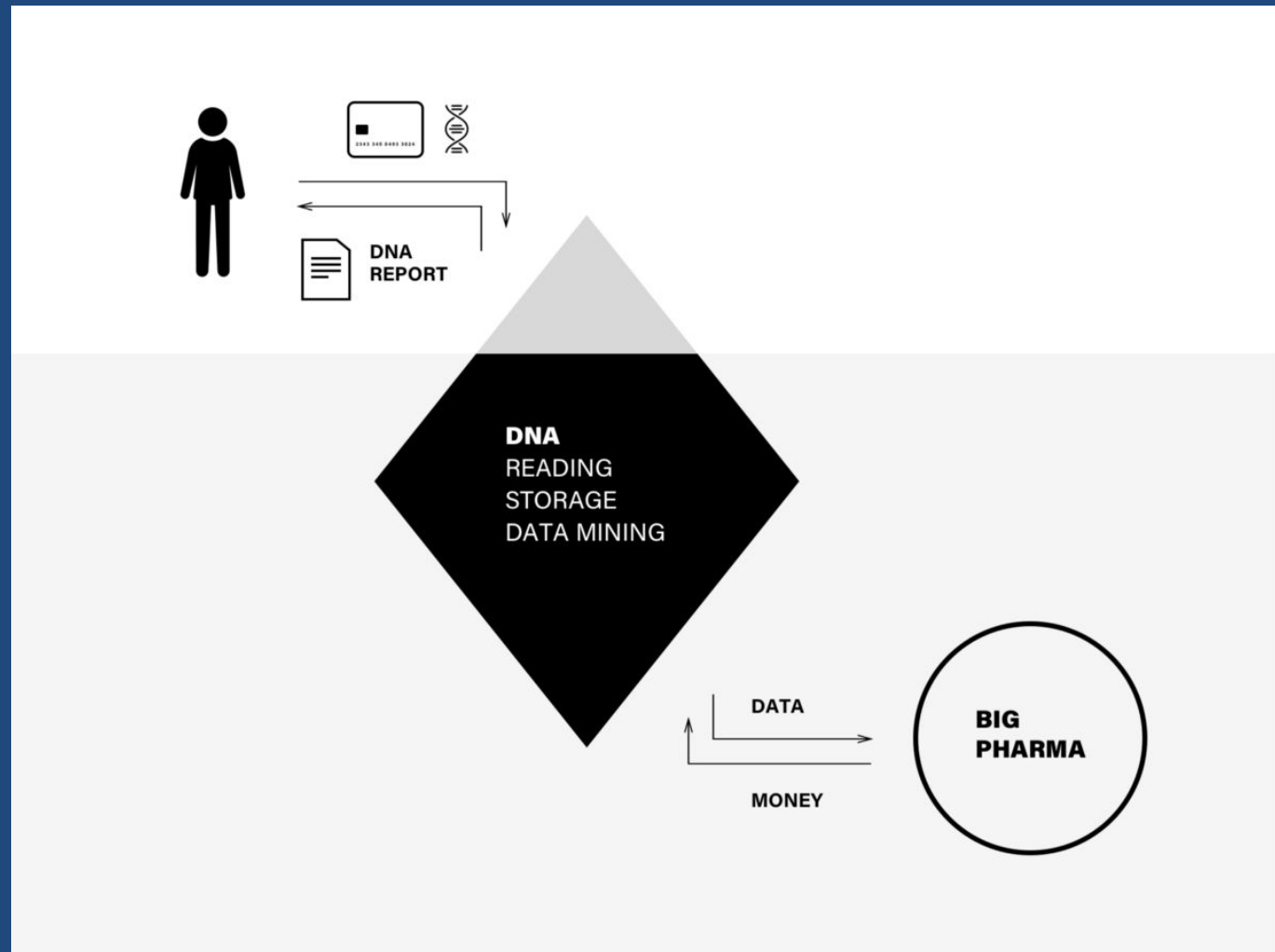
Поэтому, данные подходящие для научных исследований накапливаются:

- 1) У компаний, занимающихся коммерческим тестированием.
- 2) У научных организаций, проводящих медицинские исследования.
- 3) У людей, которые готовы потратить существенную сумму средств для дополнительного сохранения красоты и здоровья.

# 23andMe как оператор/агрегатор геномных данных



капитализация  
\$1 000 000 000



ETHAN HAWKE

UMA THURMAN

JUDE LAW

# GATTACA





# Геномные данные и приватность

В сентябре 2017 года Крейг Вентер, используя геномы и фенотипы 1061 человек, продемонстрировал, что возможно достаточно достоверно реконструировать фенотип человека используя геномные данные:

- Лицо и голос
- Рост, вес и индекс массы тела
- Цвет кожи и глаз



# Возможные решения

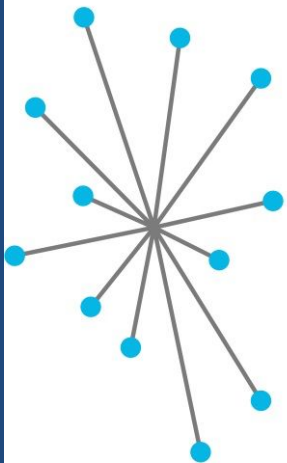
- 1) Запретить генотипирование.
- 2) Контролировать генотипирование правительством (жестче чем персональные данные).
- 3) Разработать систему обмена персональными и медицинскими данными.

Новая система с одной стороны не должна давать возможность дискредитации каждого человека, а с другой - должна допускать проведение медицинских исследований и персонализацию окружающей среды.

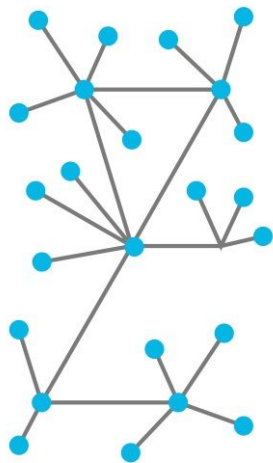
## Как это реализовать?

# Децентрализованные сети, блокчейн

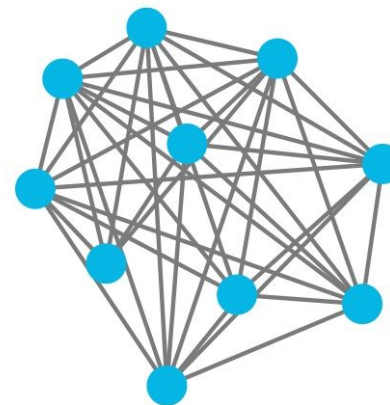
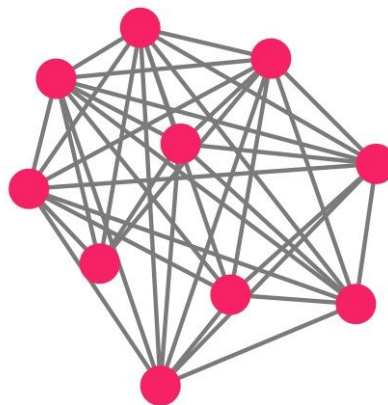
## Centralized



## Decentralized



## Distributed Ledgers



## The New Networks

Distributed ledgers can be public or private and vary in their structure and size.

Public blockchains

Require computer processing power to confirm transactions ("mining")

- Users (●) are anonymous

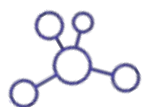
- Each user has a copy of the ledger and participates in confirming transactions independently

- Users (●) are not anonymous

- Permission is required for users to have a copy of the ledger and participate in confirming transactions

**Каждый участник сети независимо хранит информацию о состоянии сети, каждый участник сети независимо подтверждает изменение состояния сети («майнинг»).**

# Возможности сети Zenome



**Научное сообщество**

Популяционная статистика  
для исследований



Пользуется  
бесплатным API



**Пользователь**

Персональные рекомендации,  
Хранение данных



Зарабатывает ZNA  
предоставляя доступ к  
своим данным



**Покупатель данных  
(Unilever, Nestle)**

Персонализация продуктов,  
Лекарств



Приобретает ZNA для  
получения доступа к  
геномным и  
фенотипическим данным



**Сервис-провайдер  
(23andMe, SelfDecode)**

Предоставляет обработку,  
Формирует персональные  
рекомендации



Хранение геномные  
данные, получают  
новых клиентов



# Zenome: перспективы внедрения

Сейчас

В будущем

**Для пользователя:**

ограниченные возможности заказать генетический тест и получить интерпретацию.

**Для пользователя:**

возможность заказать секвенирование в любой лаборатории без интерпретации, а затем проанализировать в большинстве DTC компаний мира.

**Для научного сообщества:**

высокая сложность поиска пациентов с социально значимыми заболеваниями, обладающими близкими симптомами (фенотипами), проблемы стандартизации диагнозов.

**Для научного сообщества:**

возможность проведения популяционных и статистических исследований без секвенирования на основании уже имеющихся данных, доступ к большому массиву хорошо структурированных генетических, медицинских, социологических данных.

**Для фармкомпаний и производителей продукции повседневного спроса:**

персонализация товара возможна, но стоимость ее очень высока, так как требуется проводить подбор испытуемых по географическим и (в основном) требованиям, а затем секвенировать их геном.

**Для фармкомпаний и производителей продукции повседневного спроса:**

развитие рынка персонализированных продуктов за счет резкого снижения стоимости первичного исследования перспективности вывода продукта на рынок.

# Техническое устройство сети

Единица хранимой информации:

1. Последовательность ДНК длиной 10 000 нуклеотидов.
2. Ответ на вопрос анкеты.

Само значение доступно открыто, ему присвоен идентификатор.

Открытость «единиц» данных означает открытость обезличенной геномной и социологической статистики.

Данные хранятся в децентрализованной сети (DHT Kademlia) с количеством копий 3 и более на каждую единицу. Вычислительный узел – стационарный компьютер любого человека со свободным местом на жестком диске, требования стандартизированы с другими криптопроектами.

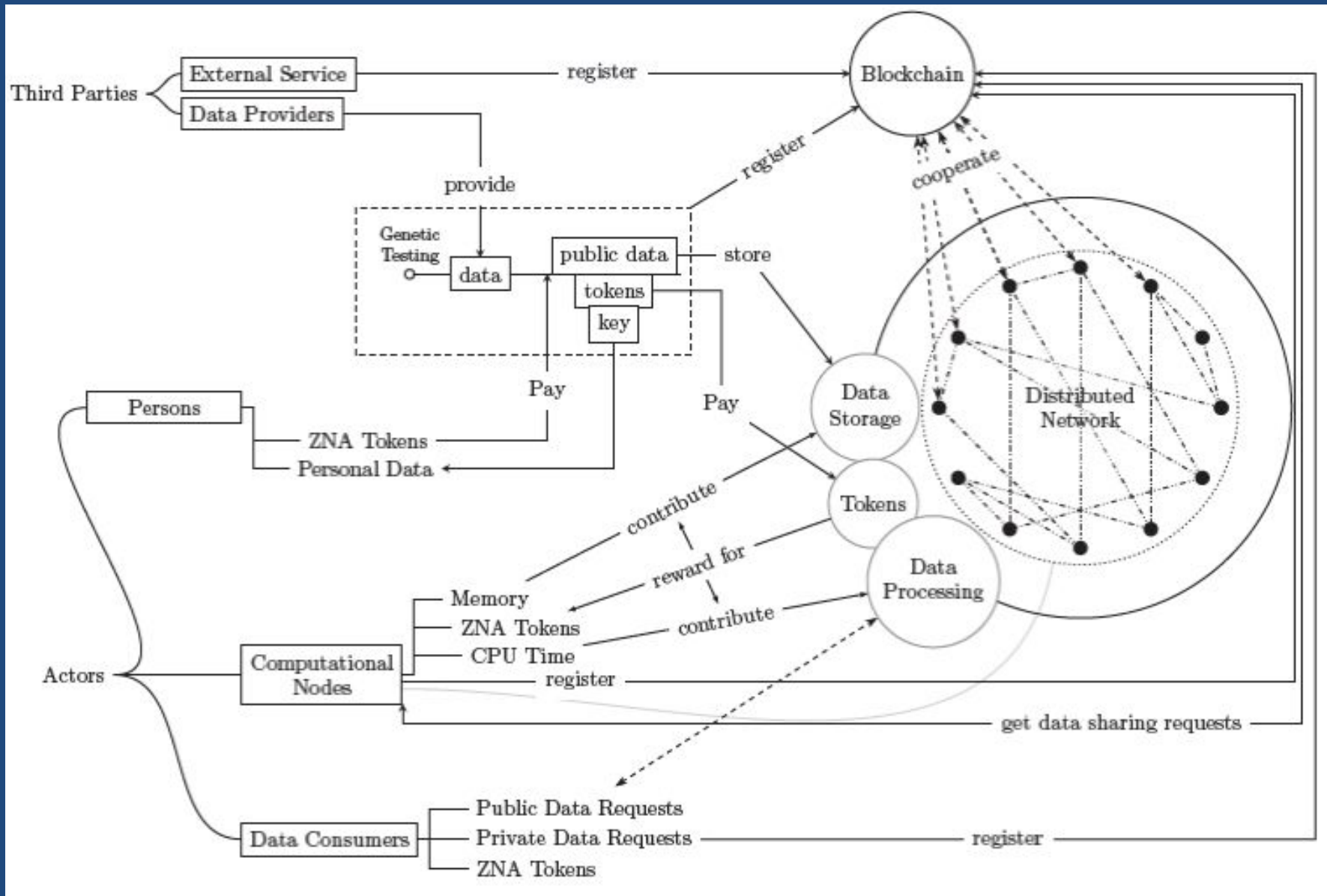
Сеть децентрализованная – это означает, что команда разрабатывает исключительно программное обеспечение, которое в будущем взаимодействует между собой без возможности вмешательства в процессы.

Только клиентская программа «знает» идентификаторы ответов «своего» пользователя. Объект продажи – ассоциация двух идентификаторов как принадлежащих одному человеку.

Для защиты от взлома сети хэш с указанием времени изменения хранится в блокчейне, а при продаже продавец передает покупателю ключ на единичную пару.

Блокчейн (Ethereum) также используется для фиксирования финансовых взаимодействий.

# Техническое устройство сети



# Текущее состояние проекта

**Август 2017** - Предложение идеи сообществу,  
публикация White Paper

**Октябрь 2017** - Предварительная продажа токенов ZNA

**Декабрь 2017** - Первичный релиз клиентского приложения,  
подключение первого сервис-провайдера (EnterOmics.com).

**Февраль 2018** - Первый раунд продажи токенов ZNA

**Май 2018** - Выход на публичные крипто-биржи

**Июнь 2018** - Первичный релиз децентрализованной сети, подключение  
второго сервис-провайдера (SelfDecode.com).

**Июль 2018** - Второй раунд продажи токенов ZNA

**Сентябрь 2018** - 1000 пользователей сети, соглашение с FMCG компаниями

**Март 2018** - Первые продажи доступа покупателей к данным сети.



# Наша команда и советники



**Алексей Горбачев**

Основатель, бизнес



**Николай Кулемин**

Основатель,  
биоинформатика



**Владимир Наумов**

Genomic data scientist



**Сергей Попов**

Блокчейн-разработчик,  
cypherpunk



**Анастасия Губина**

Community manager



**Алексей Кузменков**

Нейромаркетинг



**Дмитрий Квон**

Советник Zenomer



**Юрий Никольский**

Советник по  
развитию бизнеса



**Ричард Резник**

Старший советник



**Вадим Пушкарев**

Советник

# Будущее уже здесь.

Пора присоединиться.

## Zenome.io



<https://www.facebook.com/zenomeplatform>



<https://medium.com/zenome>



<https://twitter.com/ZenomePlatform>



<https://bitcointalk.org/index.php?topic=2191098.0>



<https://github.com/zenomeplatform>



<https://t.me/zenomeplatform>



<mailto:info@zenome.io>

THE INFORMATION PROVIDED IN THIS DOCUMENT IS NOT INTENDED FOR DISTRIBUTION  
TO, OR USE BY, ANY PERSON OR ENTITY,  
OR IN ANY JURISDICTION OR COUNTRY WHERE SUCH DISTRIBUTION OR USE WOULD BE CONTRARY TO ANY LAW