

# Медицинская геномика

Василий Евгеньевич Раменский  
Анастасия Александровна Жарикова и Мария Ильинична Зайченко

ramensky@gmail.com, azharikova89@gmail.com

НМИЦ Терапии и профилактической медицины  
Факультет биоинженерии и биоинформатики МГУ  
Институт искусственного интеллекта МГУ

2024

# Менделевские (моногенные) заболевания: поиск генов и диагностика. Часть 1

## План лекции

- 1 Болезни, синдромы и другие определения
- 2 Поиск генетических основ заболеваний
- 3 Менделевские заболевания: обзор и виды наследования
- 4 Эффект вариантов: приобретение или потеря функции
- 5 Пенетрантность, относительный риск, соотношение шансов
- 6 Поиск генов, вызывающих заболевания

## Генетические основы заболеваний

Почти для всех заболеваний человека индивидуальная предрасположенность, в некоторой степени, определяется генетической изменчивостью.

–Clausnitzer (2020) *Nature*

(1) Индивидуальные различия в ДНК, индивидуально или в комбинациях, могут сделать индивидуума **более предрасположенным к одному заболеванию**

(например, типу рака), но и могут сделать того же самого индивидуума **менее предрасположенным к развитию другого заболевания** (например, диабета).

(2) Окружение (в том числе и образ жизни) играет значимую роль во многих нарушениях (например, взаимосвязь диеты и физической активности и диабета), но наши клеточные ответы и реакции организма **на окружение могут различаться в зависимости от нашего генома**.

(3) **Генетика иммунной системы**, с огромной изменчивостью в популяции, определяет наш ответ на инфицирование патогенами.

(4) Большинство случаев рака развивается из-за **аккумуляции генетических изменений**, происходящих на протяжении жизни индивидуума, при этом на них могут влиять факторы окружающей среды.

– Jackson (2018) *Essays in Biochemistry*

## Некоторые определения

**Заболевание:** медицинское состояние тела, которое нарушает нормальные функциональные и физиологические процессы. **Генетическое заболевание** вызвано одной или несколькими аномалиями в геноме.

**Наследственные:** передается от родителей к потомкам.

**Генетическое:** унаследованное или *de novo*.

**Спорадические:** нарушение, произошедшее случайно (генетическое или же нет).

**Врожденное (vs. приобретенное):** нарушение, присутствующее при рождении.

**Фенокопия:** фенотипическая изменчивость, которая напоминает проявление генотипа, но вызвана факторами окружающей среды.

**Синдромом** называется набор симптомов, которые зачастую ассоциированы с определенным заболеванием. В генетических случаях синдром  $\approx$  заболевание.

*Пример:* синдром CHARGE (ген *CHD7*), синдром Дауна (трисомия по 21 хромосоме), синдром Туретта (причина неизвестна). Стокгольмский синдром.

## Заболевания, синдромы и другие определения

- 1. Менделевские (моногенные) заболевания** зависят от генотипа в одном локусе, наследование подчиняется законам Менделя (муковисцидоз, гемофилия А).
- 2. Комплексные (мультифакторные) нарушения:** результат комплексных взаимодействий нескольких (или многих) генетических факторов и факторов окружающей среды (диабет 2 типа, ишемическая болезнь сердца, шизофрения).

**Наследуемость:** относительный вклад генетических факторов в заболевание/фенотип.

## Заболевания, синдромы и другие определения

1. **Менделевские (моногенные) заболевания** зависят от генотипа в одном локусе, наследование подчиняется законам Менделя (муковисцидоз, гемофилия А).
2. **Комплексные (мультифакторные) нарушения:** результат комплексных взаимодействий нескольких (или многих) генетических факторов и факторов окружающей среды (диабет 2 типа, ишемическая болезнь сердца, шизофрения).

**Наследуемость:** относительный вклад генетических факторов в заболевание/фенотип.

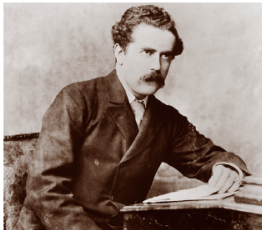
3. **Митохондриальные заболевания** происходят из-за мутаций в мтДНК.
4. **Хромосомные заболевания** происходят, когда целые хромосомы или их части отсутствуют или заменяются.
5. **Эпигенетические заболевания** – заболевания, связанные с нарушениями в активности генов, нежели чем с мутацией в структуре ДНК.

## Менделевские vs комплексные заболевания

Менделевские	Комплексные
Индивидуально редки в популяции	Часты в популяции
Паттерны наследования в семье: AD, AR, и т.д.	Нет четких паттернов в семье; существуют в популяции
Один или несколько генов с сильным эффектом	Несколько локусов, нет одного необходимого и достаточного локуса
Вызывается аллелем с высокой или полной пенетрантностью	Комбинация генетических факторов и факторов окружающей среды
Аллельная гетерогенность	Сложная аллельная архитектура
<i>Примеры:</i> муковисцидоз, семейная гиперхолестеремия, наследственные кардиомиопатии, нарушения ритма сердца	<i>Примеры:</i> ИБС, артериальная фибрилляция, гипертензия, шизофрения, сердечная недостаточность



## Пример: алкаптонурия



*Reprinted from Lancet, vol. ii, 1902, pp. 1619-1620.*

### THE INCIDENCE OF ALKAPTONURIA: A STUDY IN CHEMICAL INDIVIDUALITY

ARCHIBALD E. GARROD

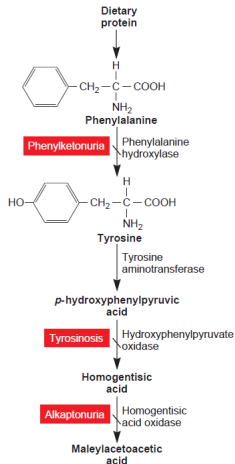
Physician to the Hospital for Sick Children, Great Ormondstreet, Demonstrator of  
Chemical pathology at St. Bartholomew's Hospital

All the more recent work on alkaptonuria has tended to show that the constant feature of that condition is the excretion of homogentisic acid, to the presence of which substance the special properties of alkapton urine, the darkening with alkalis and on exposure to air, the power of staining fabrics deeply, and that of reducing metallic salts, are

Аномальные уровни гомогентизиновой кислоты (алкаптона), которые выделяются в мочу, вызывая ее почернение при контакте с воздухом.

Алкаптонурия (АКУ) является наследственной и подчиняется рецессивному паттерну наследования.

Сэр Арчибальд Гаррод (1902): мутация → потеря энзима → врожденное нарушение метаболизма



Brooker – *Genetics, Analysis and Principles*

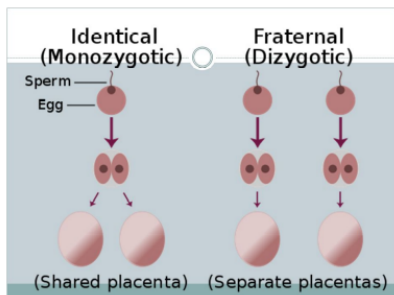


## Поиск генетических основ заболеваний: близнецы

**Монозиготные близнецы (MZ)** развиваются из одной зиготы, которая разделяется и формирует два эмбриона.

**Дизиготные близнецы (DZ)** развиваются из разных яйцеклеток, каждая из которых оплодотворяется своим сперматозоидом.

<i>Disease type</i>	<i>Concordance</i>	
	<i>MZ</i>	<i>DZ</i>
Monogenic	100%	50%
Complex	70%	25%
Non-genetic	X%	X%



## Поиск генетических основ заболеваний: эпидемиология

**Семейная агрегация:** появляется ли заболевание в семье чаще, чем ожидалось бы при случайном появлении? У родственников общие не только генетические варианты, но и факторы окружающей среды: диета, воспитание, экология.

- Паттерны сегрегации (типы наследования)
- Исследования близнецов (напр., разделенных монозиготных близнецов)
- Исследования после усыновления: больные/здоровые родители или потомки

**Описательная [генетическая] эпидемиология:** изменчивость в риске развития заболеваний между странами; исследования мигрантов; admixture

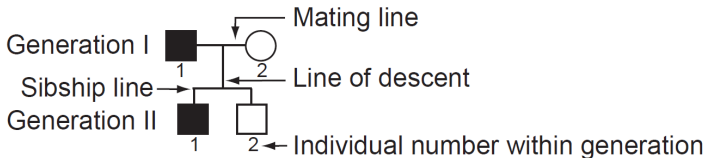
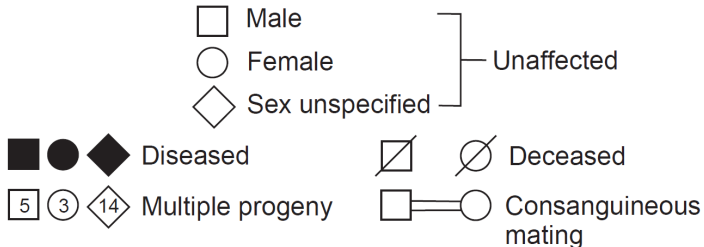
Case types	Schizophrenia cases among biological relatives	Schizophrenia cases among adoptive relatives
Index cases (47 chronic schizophrenic adoptees)	44/279 (15.8%)	2/111 (1.8%)
Control adoptees (matched for age, sex, social status of adoptive family, and number of years in institutional care before adoption)	5/234 (2.1%)	2/117 (1.7%)

The study involved 14,427 adopted persons aged 20–40 years in Denmark; 47 of them were diagnosed as chronic schizophrenic. The 47 were matched with 47 non-schizophrenic control subjects from the same set of adoptees. [Data from Kety SS, Wender PH, Jacobsen B et al. (1994) *Arch. Gen. Psychiatry* 51, 442–455.]

Strachan, Read – *Human Molecular Genetics*



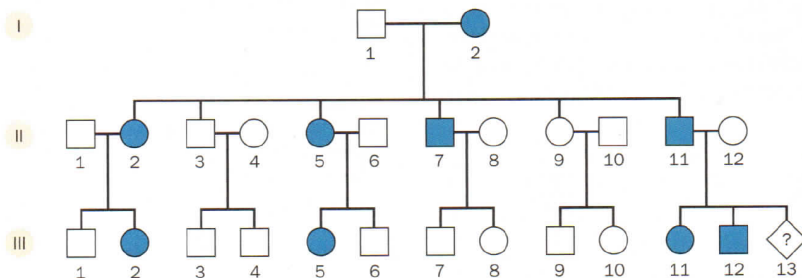
## Символы в схемах родословной



Hartwell – *Genetics. From genes to genomes*

## Аутосомно-доминантное наследование (AD): обзор

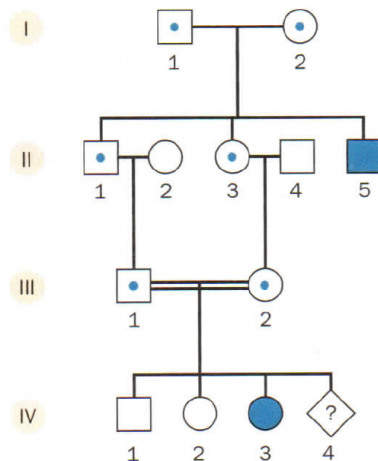
- Больной (пробанд) обычно имеет хотя бы одного больного родителя
- Поражает оба пола
- Ребенок с одним больным и одним здоровым родителем имеет 50% вероятность родиться больным; часто, *de novo*
- Каузальный вариант является вариантом приобретения или потери функции, если ген гаплонедостаточный



Strachan, Read – *Human Molecular Genetics*

## Аутосомно-рецессивное наследование (AR): обзор

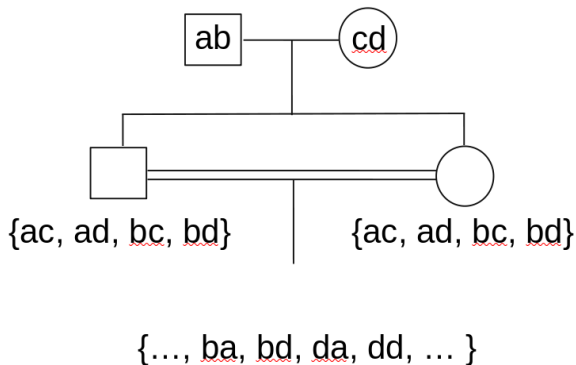
- Больные чаще всего рождаются у здоровых родителей, которые часто являются носителями без симптомов
- Поражает оба пола
- У ребенка 25% вероятность родиться больным
- Каузальным вариантом является вариант потери функции
- Часты случаи родственной связи родителей



Strachan, Read – *Human Molecular Genetics*



## AR: родственность и гомозиготность



## Упражнение

Укажите все возможные генотипы близкородственного потомка и рассчитайте вероятность гомозиготности, т.е. коэффициент инбридинга  $F$

## AR: родственность и гомозиготность

Области гомозиготности (ROH): сегменты генома, в которых наблюдается протяженная гомозиготность (без перемежающейся гетерозиготности)



**Fig. 1.** ROH detected by SNP microarray analysis (Affymetrix Cytoscan HD) in a male child who was the offspring of a brother-sister mating. Each block on the right of the chromosome represents a genomic region at least 3 Mb in size. The laboratory-reported autosomal Froh was >21%.

Sund & Rehder (2014) *Hum Hered*

## Эффект вариантов: рецессивность и доминантность

### Доминантные

- Эффект наблюдается как у гомозигот, так и у гетерозигот
- Частота варианта  $\approx$  встречаемость заболевания
- Передается от одного родителя или же *de novo*

### Примеры

- CHD7:p.Trp2332Ter, синдром CHARGE
- KMT2D:p.Arg5179His, синдром Кабуки

### Рецессивные

- Эффект наблюдается только у гомозигот
- Частота варианта  $\gg$  встречаемость заболевания
- Передается от обоих родителей

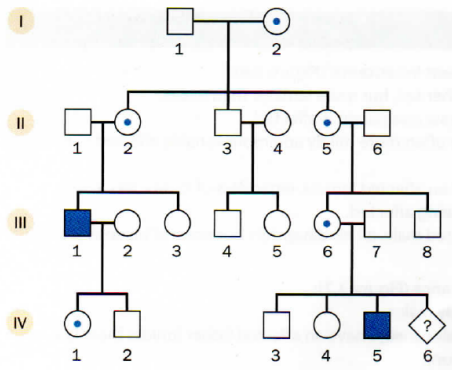
### Примеры

- CFTR:p.F508del, муковисцидоз
- GJB2:c.35delG, потеря слуха

## X-сцепленное наследование

Как правило:

- Поражает мужчин
- Больные потомки мужского пола рождаются у здоровых родителей-носителей
- Мать является бессимптомным носителем
- Нет передачи от мужчины к мужчине в родословной
- Женщины могут быть больными только если (а) отец был больным и мать носителем; (б) иногда результат неслучайной инактивации X-хромосомы

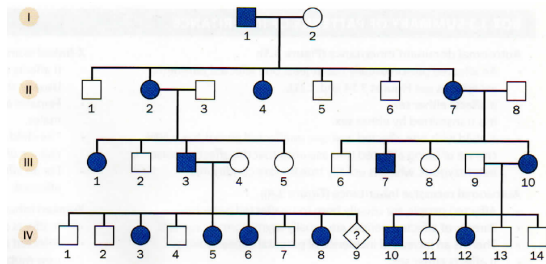


Strachan, Read – *Human Molecular Genetics*

## X-сцепленное доминантное наследование

Чаще всего:

- Женщин, чем мужчин
- Поражен хотя бы один родитель
- Женщины поражены слабее и более вариабельно, чем мужчины
- Ребенок больной женщины, вне зависимости от пола, имеет вероятность 50% родиться больным
- У больного мужчины дочь будет больной, а все сыновья – здоровыми



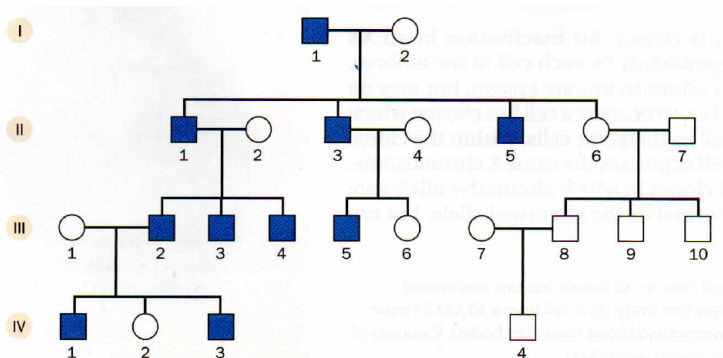
Strachan, Read – *Human Molecular Genetics*

Вопрос

Почему?

## Y-сцепленное наследование

- Поражает только мужчин
- У больных мужчин отец всегда болен, только если это не *de novo* мутация
- Все сыновья больного мужчины будут больны



Strachan, Read – *Human Molecular Genetics*

## Моногенных заболевания: общие сведения

**Менделевские (моногенные) заболевания** зависят от генотипа в одном локусе (гене); наследование подчиняется законам Менделя.

Распространенность моногенного наследования заболеваний на 1,000 новорожденных: // Blencowe (2018) *J Community Genet*:

- Аутосомно-доминантное: 1.40
- Аутосомно-рецессивное:  $1.84 + F * 650$  (связано с кровным родством)
- X-сцепленное рецессивное: 0.05
- X-сцепленное доминантное: ?
- Y-сцепленное: ?
- Неизвестно: 1.16

Общая встречаемость:  $\sim 0,4\%$  живых новорожденных

## Примеры моногенных заболеваний

Inheritance pattern	Disease	Gene/region	Nature of variants	Estimated frequency
Autosomal dominant	Glut1 deficiency (De Vivo disease)	<i>SLC2A1</i>	Mutations reduce or eliminate function	Rare, approximately 1/90000
	Osteogenesis imperfecta (brittle bone disease)	<i>COL1A1</i> or <i>COL1A2</i> (90%) (also <i>CRTAP</i> or <i>P3H1</i> )	<i>COL1A1/COL1A2</i> – usually missense mutations that lead to protein (collagen) of altered structure	6–7/100000
Autosomal recessive	Achondroplasia	<i>FGFR3</i>	Activating point mutations	1/15000 to 1/40000
	Phenylketonuria	<i>PAH</i>	Many different mutations, including missense, non-sense, splicing mutations	1/10000 to 1/15000
	Cystic fibrosis	<i>CFTR</i>	Over 2000 different variants known	1/2500 to 1/3500 in Caucasians, less common in other ethnic groups
X-linked recessive	Sickle-cell anaemia	<i>HBB</i>	Various missense variants, gene deletions	1/70000 to 1/80000 in the U.S.A., more common in other countries
	Haemophilia A	<i>F8</i>	Missense and nonsense mutations	1/4000 to 1/5000 males
	Duchenne muscular dystrophy	<i>DMD</i>	Usually deletions or duplications	1/3500 to 1/5000 (Duchenne and Becker muscular dystrophy together)
X-linked dominant	Fragile X syndrome	<i>FMR1</i>	CGG trinucleotide repeat expansion	1/4000 (males), 1/8000 (females)
	Rett syndrome	<i>MECP2</i>	Missense mutations, abnormal epigenetic regulation	1/8500 females
	X-linked hypophosphatemic rickets	<i>PHEX</i>	Deletions, insertions, missense, nonsense, splicing mutations	1/20000
Y-linked	Nonobstructive spermatogenic failure	<i>USP9Y</i>	Most commonly deletions	1/2000 to 1/3000

Jackson (2018) *Essays in Biochemistry*



## Примеры моногенных заболеваний

**Huntington disease (HD)** is one of the trinucleotide repeat expansion disorders where the CAG repeat encodes a polyglutamine tract within the coding region of the huntingtin gene HTT on chromosome 4p16. It is a progressive neurodegenerative disorder with patients suffering from progressive neural cell loss and atrophy. Symptoms start with personality and mood changes, followed by a steady deterioration of physical and mental abilities. The function of the huntingtin protein is unclear, but it is essential for development.

Inheritance follows an autosomal dominant pattern, caused by a gain-of-function associated with the repeat expansion. Unaffected individuals carry between 9 and 35 CAG repeats, incomplete penetrance occurs in carriers of 36–39 repeats, while the disease is fully penetrant when 40 or more repeats are present. Alleles containing 250 and more repeats have been reported. While repeat alleles of 9–30 are almost always transmitted without change to the next generation, larger alleles show instability, both in somatic tissues and in the germline, with a tendency towards expansion from one generation to the next. There is a correlation between the number of repeats and the severity of disease and also an inverse correlation between the number of repeats and the age of disease onset. The degree of repeat instability is also largely proportional to the number of repeats, and is also affected by the sex of the transmitting parent, with larger expansions occurring in male transmission. This leads to 'anticipation' where an apparently healthy individual might have a child with late onset HD and a grandchild with more severe symptoms and an earlier onset, and so on.

Jackson (2018) *Essays in Biochemistry*

## Примеры моногенных заболеваний 🏠

**Achondroplasia (ACH)** is the most common form of dwarfism in humans and is inherited in an autosomal dominant fashion with 100% penetrance. Individuals with ACH have shortened limbs, a large head, and a trunk of relatively normal size. ACH is caused by specific variants in *FGFR3*, the gene for fibroblast growth factor (FGF) receptor 3 (*FGFR3*), on chromosome 4p16.

Almost all individuals with ACH are heterozygous for a variant p.Gly380Arg in the mature protein. 80% of ACH cases are due to spontaneous, de novo mutations, often occurring during spermatogenesis. *FGFR3* is a transmembrane receptor protein which binds to FGF ligands and triggers intracellular signalling processes. One of these processes is the inhibition of chondrocyte proliferation in the growth plate of long bones. The p.Gly380Arg variant in *FGFR3* generates a constitutively active version of the receptor which can be further activated by binding of FGF. Therefore, this variant acts as a gain-of-function mutation. Consequently, chondrocyte proliferation in growth plates is constitutively inhibited. While one such variant allele (in the heterozygous state) leads to ACH, homozygosity is lethal before birth or perinatally.

Interestingly, loss-of-function variants in *FGFR3* have also been described which cause a different condition, **camptodactyly**, tall stature and hearing loss (CATSHL) syndrome. This is an example where different variants of the same gene result in different phenotypes, so-called 'allelic disorders'.

Jackson (2018) *Essays in Biochemistry*

## Примеры моногенных заболеваний 🏠

**Cystic fibrosis (CF)** mostly affects the lungs (resulting in breathing difficulty and frequent lung infections) and the pancreas, but the liver, kidney, intestines and male reproductive system are also frequently affected. It is the most common lethal genetic disease among Caucasians, and is inherited in an autosomal recessive pattern.

CF is caused by pathogenic variants in the CFTR gene, which encodes the CF transmembrane conductance regulator, a transmembrane protein which functions as a selective chloride channel. If the CFTR protein does not function properly, the chloride balance between the inside and outside of cells becomes disrupted, leading to the build-up of mucus in narrow passages in affected organs such as the lungs. The CFTR gene is located on chromosome 7q31 and encodes a protein of 1480 amino acids with >2000 pathogenic variants have been identified in its sequence. These variants fall into different classes (e.g. those where protein synthesis is defective, those where reduced amounts of normal protein is made, and others). As long as an individual carries one functional allele of CFTR, they may show no or only very mild symptoms, but an individual carrying two pathogenic variants will display symptoms that depend on the amount of functional protein generated.

The most common pathogenic variant, representing approximately 70% of Caucasian CF alleles, is a deletion p.Phe508del. This particular variant leads to the synthesis of a protein which does not fold properly into its 3D shape, and is degraded by the cell before it can reach the membrane, therefore representing a loss of function.

Jackson (2018) *Essays in Biochemistry*

## Отклонения от моногенного наследования

- **Гетерогенность локуса:** некоторые клинические фенотипы могут проявиться от мутации в любом из нескольких локусов
- **Аллельная гетерогенность:** много разных мутаций в одном гене вызывают одно и то же заболевание
- **Клиническая гетерогенность:** мутации в одном гене вызывают два или более заболеваний у разных людей. // Не то же самое, что **плейотропия**. *Пример:* мутации в гене *HPRT* могут вызывать или форму подагры [OMIM:300323], или синдром Леша-Нихена, проявляющегося в задержке развития с поведенческими проблемами [OMIM:300322]
- **Неполная пенетрантность:** у человека, который является носителем вызывающего болезнь генотипа, заболевание не всегда проявляется. В частности, пенетрантность, зависящая от возраста у заболеваний с поздним развитием.

Strachan, Read – *Human Molecular Genetics*

## Отклонения от моногенного наследования (продолжение)

- **Вариабельная экспрессия:** у разных членов семьи проявляются разные признаки заболевания
- **Импринтинг:** у мутации есть эффект, только если она унаследована от родителя определенного пола. *Пример:* аутосомно-доминантное наследование параганглиом [OMIM:168000], наследуется только от отца; синдром Беквита-Видемана [OMIM:130650], наследуется только от матери
- **Фенокопия:** случаи, когда заболевание наблюдается без вызывающего его фенотипа. Пример: глухота
- **De novo мутации** усложняют менделевское наследование
- **Мозаицизм** в линии соматических клеток

Strachan, Read – *Human Molecular Genetics*

## Эффект вариантов: приобретение или потеря функции

**Потеря функции (loss of function):** у продукта гена ограничена или утеряна функция.

*Примеры:* факторы транскрипции, нарушение функции катализа у ферментов

- Укорачивающие белок, миссенсы?
- Рецессивное наследование, но в некоторых случаях (гаплонедостаточность) и доминантное

**Приобретение функции (gain of function):** продукт гена делает «что-то положительно аномальное»

*Примеры:* транскрипционные факторы, мутации приобретения функции в GPCR

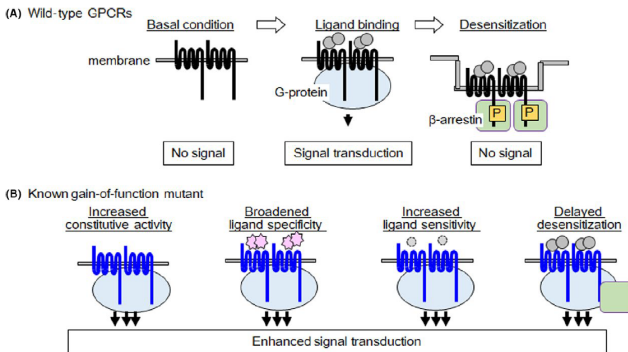
- В основном миссенсы, но также делеции без сдвига рамки считывания.
- Наличие нормального аллеля не может остановить аномальное поведение мутантного аллеля ⇒ доминантный?

### Вопрос

Приобретение функции из-за сдвига рамки считывания?

## Пример: GPCR

- **G-protein coupled receptors** являются сенсорами внутренних стимулов: гормонов, ионов и хемокинов; света, запаха и вкуса. G-белок-связанные рецепторы играют важную роль в эндокринной системе.
- Человеческий геном: >700 различных GPCR.
- Ассоциированы с различными заболеваниями, в т.ч. эндокринными.

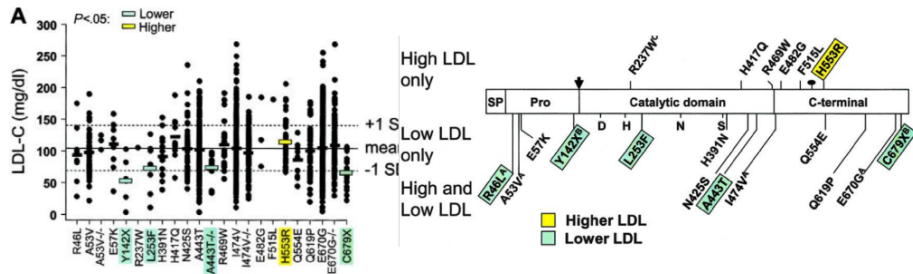


Fukami (2018) *Clin Endocrinol*

## Пример: PCSK9

Сериновая протеаза PCSK9 (пропротеиновая конвертаза субтилизин-кексинового типа 9) регулирует холестерин липопротеинов низкой плотности (ХС ЛПНП).

Высокие уровни ХС ЛПНП  $\Rightarrow$  атеросклероз  $\Rightarrow$  инфаркт или инсульт



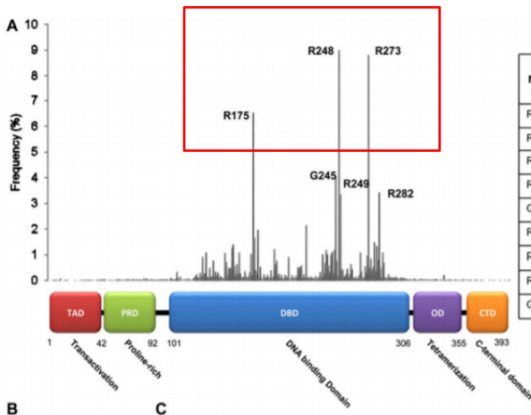
Kotowski (2006) *Am J Hum Genet*



## Пример: TP53 🏠

## TP53 mutational spectrum in human cancers

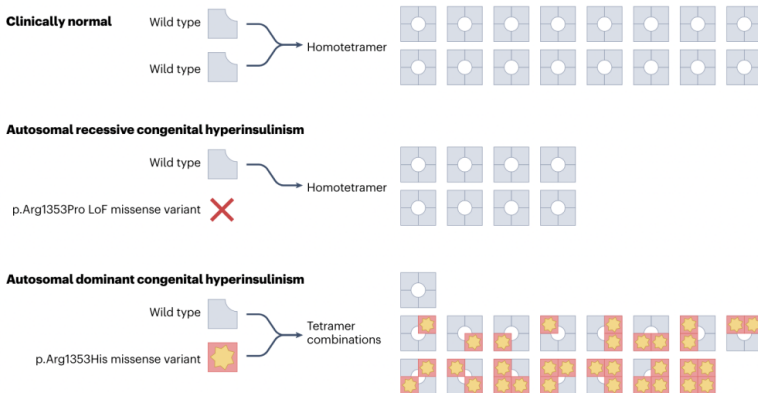
## Gain-of-function



Mutation	Overall Frequency	Wild-Type Codon	Mutant Codon	CpG	Class
R175H	4.6%	CGC	CAC	Yes	Conformation
R248Q	3.5%	CGG	CAG/CAA	Yes	DNA Contact
R273H	3.1%	CGT	CAT	Yes	DNA Contact
R248W	2.8%	CGG	TGG	Yes	DNA Contact
G245S	2.8%	GGC	AGC	Yes	Conformation
R273C	2.7%	CGT	TGT	Yes	DNA Contact
R282W	2.4%	CGG	TGG	Yes	DNA Contact
R249S	1.8%	AGG	AGT	No	Conformation
G245D	0.68%	GGC	GAC	No	Conformation

William A. Freed – Pastor (2012) *Genes & Development*

## Пример: врожденный гиперинсулинизм (ВГИ)



**K<sub>ATP</sub> channel-related congenital hyperinsulinism:** blood sugar regulation by pancreatic  $\beta$ -cells depends on the normal function of an octameric K<sub>ATP</sub> channel (4xKir6.2 proteins + 4xSUR1 proteins), respectively encoded by the KCNJ11 and ABCC8 genes (top; Kir6.2 not depicted for simplicity). ATP-mediated closure of the K<sub>ATP</sub> channel causes insulin release. p.Arg1353Pro: a non-functional recessive loss of function (LoF) (haplosufficiency; middle). p.Arg1353His results in a stable abnormal SUR1 protein that interferes with the wild type (WT) protein and has a dominant negative effect (bottom)

Zschocke (2023) *Nat Rev Genet*

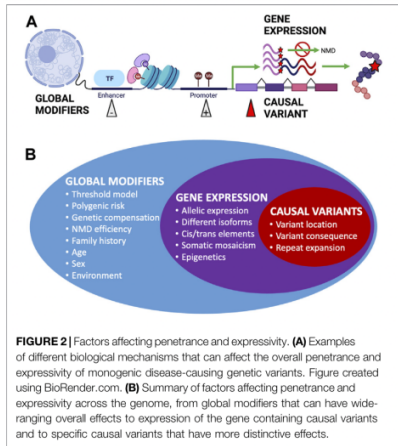
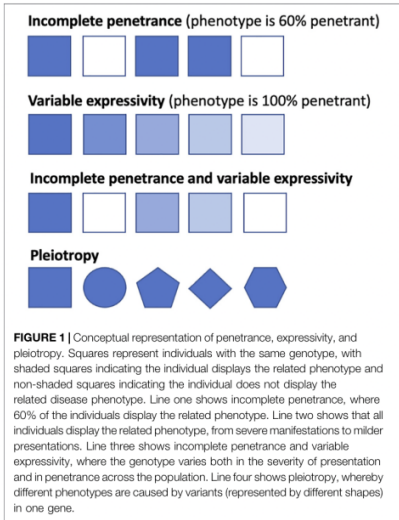
## Попытки систематизации

Table 1 | Examples of different gain of function effects

Principle	Basic mechanism	Detailed mechanism	Disease/trait	Gene(s)	Refs.
Protein activation and/or loss of protein control	Ligand-independent signalling increase	Constitutive activation by intermolecular cross-linking or loss of negative regulation	Osteoglophonic dysplasia, encephalocutaneous lipomatosis	<i>FGFR1</i>	171,172
	Ligand-dependent signalling increase	Increased binding affinity for physiological or non-physiological ligands	Pfeiffer syndrome	<i>FGFR1</i>	173
	Uncontrolled enzyme function	Intracellular autoactivation of the normally blocked serine protease domain	Periodontal Ehlers–Danlos syndrome	<i>C1R, C1S</i>	57
	Uncontrolled ion channel function	Loss of gating	Paramyotonia congenita, hypokalaemic and hyperkalaemic periodic paralysis	<i>SCN4A</i>	90
			Long QT syndrome type 3	<i>SCN5A</i>	91
	Transcription factor binding promiscuity	Mixed gain and loss of transcription factor binding specificity	Congenital dyserythropoietic anaemia type IV	<i>KLF1</i>	58
Activation of other protein functions	Decrease in the activation threshold of the pyrin inflammasome	Familial Mediterranean fever	<i>MEFV</i>	70	
Loss of expression control	Ectopic gene expression	Promoter activation	Exercise-induced hyperinsulinism	<i>SLC16A1</i>	59
		Enhancer activation	Pre-axial polydactyly	<i>SHH</i>	60
	Alteration of splicing	Disruption of alternative splicing	Apert and Pfeiffer syndromes	<i>FGFR2</i>	61
			Frasier syndrome	<i>WT1</i>	63
Alteration of topologically associating domains	Novel regulatory landscape, enhancer adoption	Acropectoral syndrome	<i>SHH</i>	153	
Non-specific effects of abnormal gene product	Abnormal mRNA effects	Detrimental interaction with repeat RNA-binding proteins, aberrant repeat-associated non-ATG translation	Myotonic dystrophy	<i>DMPK, CNBP</i>	64
	Toxic protein effect	Coding triplet repeat expansion (polyglutamine disorders)	Huntington disease	<i>HTT</i>	69
		Protein aggregation disorders (amyloidoses)	Hereditary transthyretin-related amyloidosis	<i>TTR</i>	66
Other functional effects	Novel protein function	Different substrate binding based on size of active centre	ABO blood groups	<i>ABO</i>	71

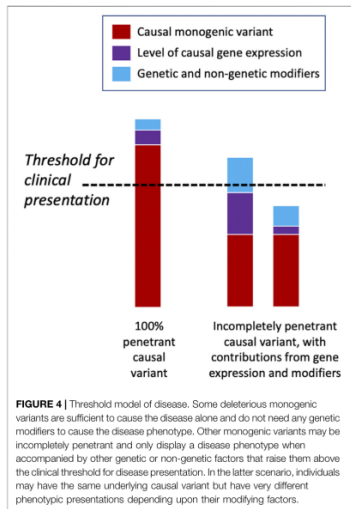
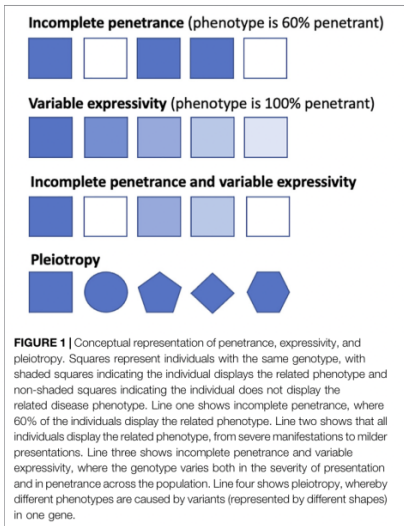
Note that this table is not exhaustive, and additional gain of function (GoF) mechanisms are well recognized, for example in tumour development.

# Пенетрантность: определения

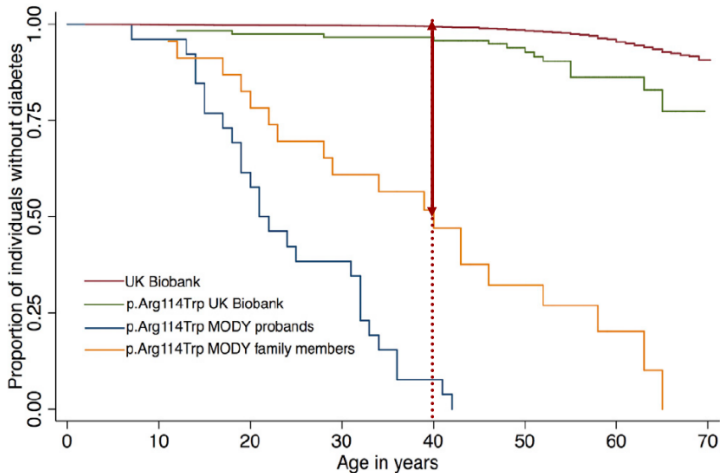


Kingdom and Wright (2022) *Frontiers Genet*

# Пенетрантность: определения



## Пенетрантность: примеры



Comparison of Penetrance Estimate for HNF4A p.Arg114Trp in UK Biobank versus Previously Published Estimates from MODY Cohort Studies // Wright (2019) *AJHG*

## Пенетрантность: примеры из ClinVar

<i>Gene, variant, ClinVar ID</i>	<i>Disease</i>	<i>Penetrance</i>
BRCA1 DNA Repair Associated <i>BRCA1</i> p.Arg1699Gln SCV000210198.11	Breast cancer, ovarian cancer	A study of 4,024 individuals from 129 families (Moghadasi 2017): a <b>20% risk of breast cancer</b> and a <b>6% risk of ovarian cancer</b> by age 70. Lifetime risks associated with typical BRCA1 variants are estimated to be <b>57 to 87% for female breast cancer and 24 to 54% for ovarian cancer</b> (Claus 1996, Antoniou 2003, King 2003, Risch 2006, Chen 2007)
Homeostatic Iron Regulator <i>HFE</i> p.Cys282Tyr SCV000221190.3	Hemochromatosis	Biochemically, 82% of p.Cys282Tyr homozygotes were shown to have elevated transferrin saturation (Pederson 2009); however, <b>&lt;5% of individuals</b> with biallelic pathogenic HFE variants exhibit clinical symptoms of HH (Beutler 2002, Gurrin 2009)
Leucine Rich Repeat Kinase 2 <i>LRRK2</i> p.Gly2019Ser SCV000640135.3	Parkinson's disease	This variant is clearly defined as a Parkinson's disease (PD) causative allele and is the most common known genetic cause of PD, having been observed in ~5% of familial and ~1-2% of sporadic PD cases (PMID: 18986508, 15726496, 22575234, 15680455). This variant exhibits age-dependent penetrance, with the probability of becoming affected increasing <b>from 20% at age 50 years to 80% at age 70 years</b> (PMID: 18986508, 15726496).

# Population-Based Penetrance of Deleterious Clinical Variants

Iain S. Forrest, BS; Kumardeep Chaudhary, PhD; Ha My T. Vy, PhD; Ben O. Petrazzini, BS; Shantanu Bafna, MS; Daniel M. Jordan, PhD; Ghislain Rocheleau, PhD; Ruth J. F. Loos, PhD; Girish N. Nadkarni, MD; Judy H. Cho, MD; Ron Do, PhD

## Key Points

**Question** What is the population-based penetrance of pathogenic and loss-of-function clinical variants?

**Findings** This cohort study included 72 434 participants from 2 biobanks who had alleles for pathogenic or loss-of-function variants reported for 157 diseases. Among the 5360 pathogenic/loss-of-function variants, 4795 (89%) were associated with less than or equal to 5% risk difference for disease in individuals with the variant allele; pathogenic variants were associated with 6.9% mean penetrance and benign variants were associated with 0.85% mean penetrance.

**Meaning** In these biobanks, the estimated penetrance of pathogenic/loss-of-function variants varied, but was generally associated with a small increase in the risk of disease.



## Пенетрантность, относительный риск, отношение шансов

	Больные	Здоровы
Есть мутация	$D_m$	$H_m$
Нет мутации	$D_0$	$H_0$

**Риск заболевания:** вероятность заболевания при наличии мутации:  $\frac{D_m}{D_m+H_m}$

Похоже на пенетрантность, не учитывает риск без мутации.

**Отношение риска:**  $RR = \frac{D_m(D_0+H_0)}{D_0(D_m+H_m)}$

**Отношение шансов:**  $OR = \frac{D_m/H_m}{D_0/H_0} = \frac{D_m H_0}{D_0 H_m}$  // вероятность события  $\frac{p}{1-p}$

## Упражнение

1. Когда  $OR \approx RR$ ?
2. Расчитайте  $OR, RR$  для  $D_m = 60, H_m = 40, D_0 = 2, H_0 = 48$

## MAF или OR: пример UK Biobank

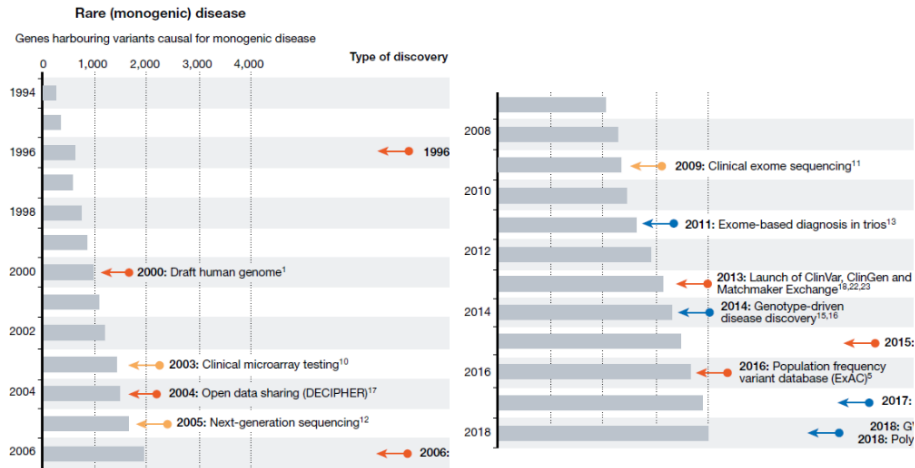
Gene	UKB ID	Position (GRCh37)	HGVs	MAF White British (%)	Significantly Associated Trait(s) in UKB (Units)	Odds Ratio or Beta [95% CI]	p value	Linked Disease (Mode of Inheritance)
<i>ACSF3</i>	dbSNP: rs141090143	chr16: 89220556 C>T	GenBank: NM_174917: c.C1672T:p.R558W	0.632	ease of sunburn (number of episodes)	0.31 [0.20, 0.42]	$4 \times 10^{-10}$	combined malonic and methylmalonic aciduria (AR)
<i>AR</i>	dbSNP: rs137852591	chrX: 66941751 C>G	GenBank: NM_000044: c.C2395G:p.Q799E	0.129	skeletal mass (SD)	-0.16 [-0.21, -0.11]	$1 \times 10^{-10}$	partial androgen insensitivity syndrome (XLR)
					height (cm)	-0.85 [-1.27, -0.43]	$1 \times 10^{-8}$	
	dbSNP: rs1800053	chrX: 66931295 C>A	GenBank: NM_000044: c.C1937A:p.A646D	0.269	balding pattern (males only)	-0.13 [-0.17, -0.08]	$1 \times 10^{-8}$	partial androgen insensitivity syndrome (XLR)
<i>ERCC4</i>	dbSNP: rs121913049	chr16: 14041848 C>T	GenBank: NM_005236: c.C2395T:p.R799W	0.060	ease of sunburn (number of episodes)	0.98 [0.64, 1.33]	$2 \times 10^{-8}$	xeroderma pigmentosum (AR)
<i>FLG</i>	dbSNP: rs150597413	chr1: 152277622 G>T	GenBank: NM_002016: c.C9740A:p.S3247X	0.369	eczema	1.66 [1.40, 1.98]	$9 \times 10^{-8}$	ichthyosis vulgaris (AD)
	dbSNP: rs138726443	chr1: 152280023 G>A	GenBank: NM_002016: c.C7339T:p.R2447X	0.446	eczema	1.96 [1.69, 2.27]	$5 \times 10^{-16}$	ichthyosis vulgaris (AD)
<i>GCK</i>	dbSNP: rs104894006	chr7: 44189591 G>A	GenBank: NM_000162: c.C556T:p.R186X	0.001	maturity-onset diabetes of the young	68 [14, 325]	$2 \times 10^{-8}$	diabetes mellitus (AD)

Wright (2019) *AJHG*

## Поиск генов, сложности

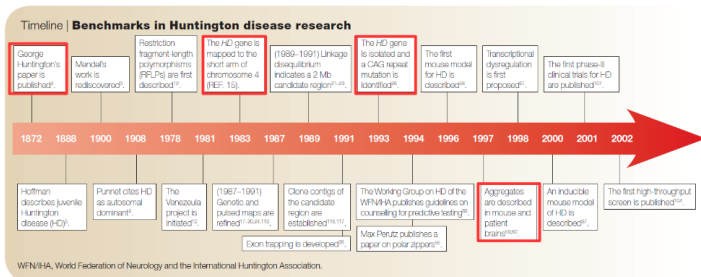
- Редкие клинические наблюдения: сложные в наблюдении повторения необходимы для менделевских паттернов
- *De novo* мутации: нет сегрегации в семьях // доминантные или составные гетерозиготы в случае рецессивных
- Нужны каузальные мутации, а не гены: необходимы функциональные исследования; сдвиг в сторону очевидных вариантов
- Все вышеупомянутые усложнения менделевского наследования: локусная, аллельная и клиническая гетерогенность; неполная пенетрантность и переменная экспрессия; импринтинг; фенкопии и мозаицизм

## Поиск генов, хронология

Claussnitzer (2020) *Nature*

## Начало: ген болезни Гентингтона

- Начинается во взрослом возрасте (30-45 лет), нейродегенеративное заболевание, прогрессирует за 15-20 лет
- Психиатрические нарушения, нарушения двигательной функции и ухудшение когнитивных функций
- Доминантное наследование, нет спорадических форм
- Первый локус, ассоциированный с болезнью, который сопоставили на хромосому (1983)
- До сих пор (2018) нет лечения, кроме симптоматического

Bates (2005) *Nat Rev Genet*

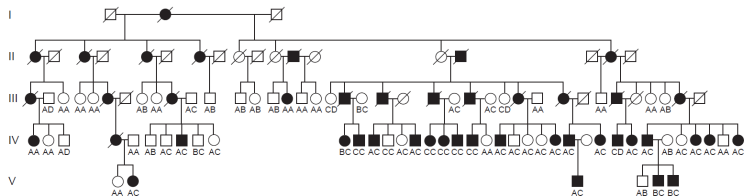
## Начало: ген болезни Гентингтона

- Высокая встречаемость в Венесуэле, один основатель
- 1983, лаборатория Джеймса Гузеллы: изучение 8 локусов, которые связали с болезнью Гентингтона, в  $\sim 4$  сМ от теломерного участка 4 хромосомы
- Не было технологий для изучения хромосомы на протяжении более 100-200 кб
- 1993: ген *HTT* (гентингтин) был клонирован совместными усилиями 9 лабораторий; 10 366 транскриптов с **повторением CAG триплета (Gln)** в первом экзоне который был полиморфным в нормальных хромосомах, и увеличивался при болезни Гентингтона
- 1993-1996: **Количество (CAG)<sub>n</sub>**: 6-35 безвредны; более 40 полностью пенетрантны и вызывают болезнь Гентингтона в течение жизни; более 70: раннее начало болезни
- Отцовская антисипация: количество повторов увеличивается при передаче
- Полиглутаминовые повторы в патогенном количестве спонтанно агрегируют в амилоидные фибриллы  $\Rightarrow$  нейронная дегенерация
- Тестирование в детстве на начинающиеся во взрослом возрасте неизлечимые заболевания может принести больше вреда, чем пользы

Bates (2005) *Nat Rev Genet*

## Поиск генов с помощью сцепления

- Несколько поколений с больными и здоровыми потомками
  - Карта полиморфных маркеров ДНК, для которых известны генетические расстояния
1. Найти маркеры ДНК, которые косегрегируют (в сцеплении) с признаком болезни в потомстве, идентифицировать возможный регион гена, связанного с заболеванием
  2. Секвенировать гены в сцепленном локусе для поиска болезнь-вызывающих аллелей, проверить аллели в здоровых индивидуумах
  3. Провести подтверждающие функциональные исследования в клеточных и животных моделях



Hartwell – *Genetics. From genes to genomes*

## Поиск генов с помощью сцепления

