

# Медицинская геномика

Василий Евгеньевич Раменский  
Анастасия Александровна Жарикова

`ramensky@gmail.com, azharikova89@gmail.com`

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

2025

# Менделевские (моногенные) заболевания: поиск генов и диагностика

## План лекции

- 1 Болезни, синдромы и другие определения
- 2 Поиск генетических основ заболеваний
- 3 Менделевские заболевания и виды наследования
- 4 Эффект вариантов: приобретение или потеря функции
- 5 Пенетрантность, относительный риск, соотношение шансов
- 6 Поиск генов менделевских заболеваний
- 7 Клиническая значимость вариантов генома
- 8 Генетическая диагностика менделевских заболеваний

# Определения



## Генетические основы заболеваний

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

–Claussnitzer (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 (пример: 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 **накопления генетических изменений, происходящих на протяжении жизни** индивидуума, при этом на этот процесс 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 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.

NB: 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 [сложных] 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 [сложных] взаимодействий нескольких (или многих) генетических факторов и факторов окружающей среды (диабет 2 типа, ишемическая болезнь сердца, шизофрения).

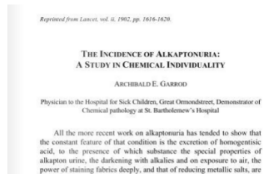
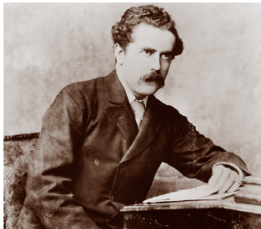
**Наследуемость:** относительный вклад генетических факторов в заболевание/фенотип.

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 гиперхолестеринемия, наследственные кардиомиопатии, нарушения ритма сердца	<i>Примеры:</i> ИБС, артериальная фибрилляция, гипертензия, шизофрения, сердечная недостаточность

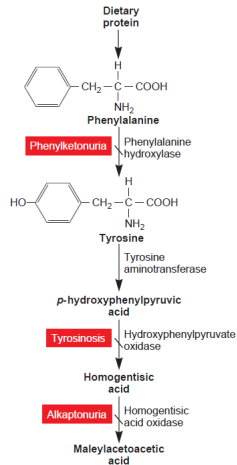
## Пример: алкаптонурия



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 фермента → врожденное нарушение метаболизма



Brooker – *Genetics, Analysis and Principles*

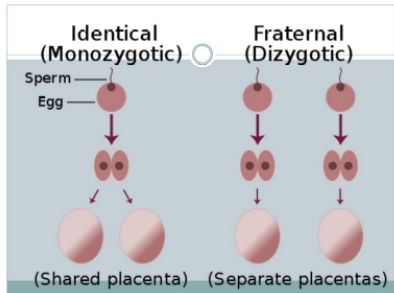
# Генетические ослывы?

## 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

**TABLE 15.3 AN ADOPTION STUDY IN SCHIZOPHRENIA**

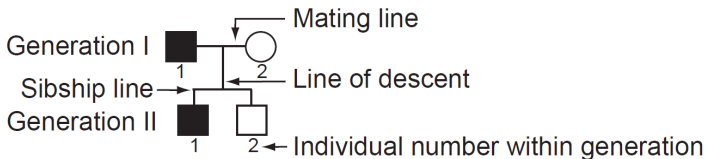
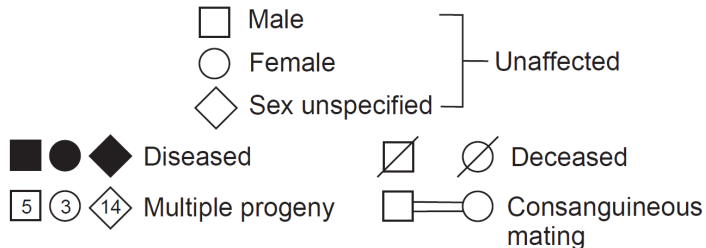
Case types	Schizophrenia cases among biological relatives	Schizophrenia cases among adoptive relatives
Index cases (47 chronic schizophrenic adoptees)	44/279 (15.8%)	2/111 (1.8%)
Control adoptees (matched for age, sex, social status of adoptive family, and number of years in institutional care before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

# Виды наследования

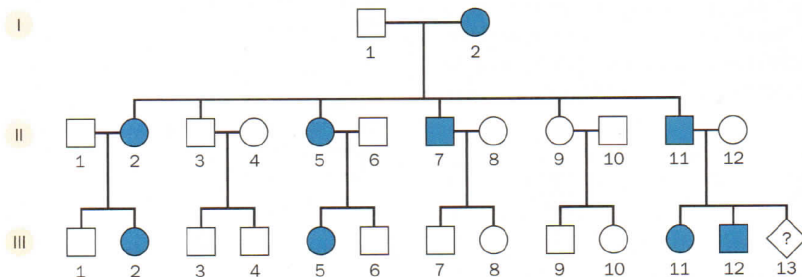
## Символы в схемах родословной



Hartwell – *Genetics. From genes to genomes*

## Аутосомно-доминантное наследование (AD): обзор

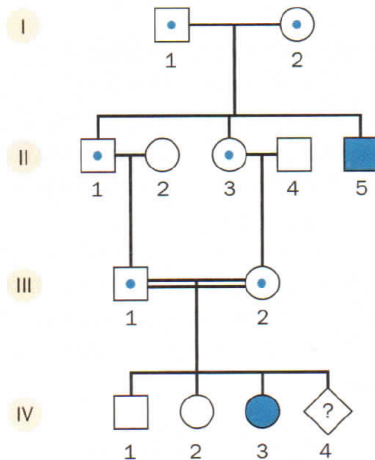
- Больной (пробанд) обычно имеет хотя бы одного больного родителя
- Поражает оба пола
- Ребенок с одним больным и одним здоровым родителем имеет 50% вероятность родиться больным; часто, *de novo*
- Каузальный вариант является вариантом приобретения или потери функции, если ген гаплонедостаточный



Strachan, Read – *Human Molecular Genetics*

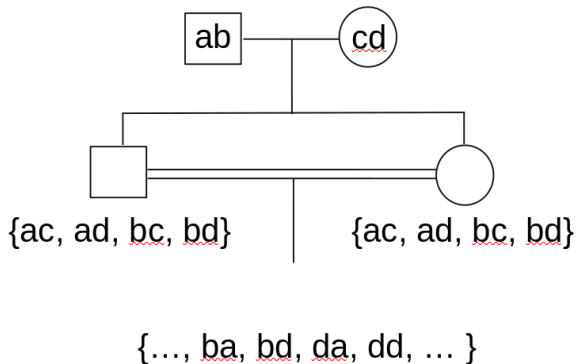
## Аутосомно-рецессивное наследование (AR): обзор

- 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

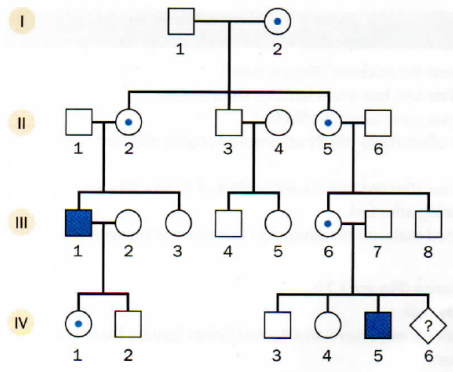
- CFTR:p.F508del, муковисцидоз
- GJB2:c.35delG, потеря слуха



## X-сцепленное [рецессивное] наследование

Как правило:

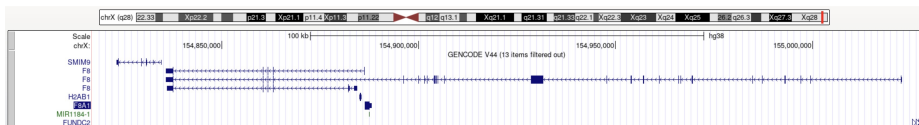
- Поражает мужчин
- Женщины могут быть больными только если (а) отец был больным, а мать — носителем; (б) произошла неслучайная инактивация X-хромосомы
- Больные потомки мужского пола рождаются у здоровых матерей-носителей
- Нет передачи от мужчины к мужчине в родословной



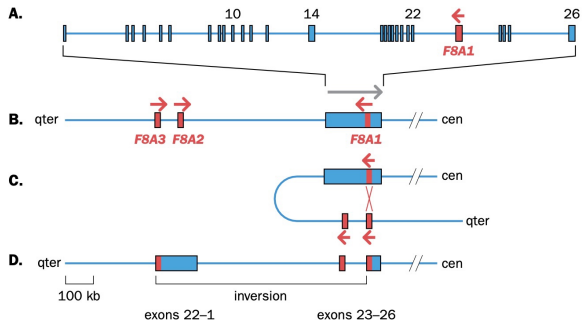
Strachan, Read – *Human Molecular Genetics*

## Пример X-сцепленного наследования: гемофилия А (OMIM:306700)

- Нарушение свертывания крови (коагулопатия): длительные кровотечения после любых повреждений кожи, самопроизвольное возникновение подкожных гематом
- Известна более 1,500 лет: рекомендации в Талмуде по отмене обрезания при наличии заболевания у братьев // Rosner (1969) *Ann Int Med*
- 50% вероятность передачи от матери сыну
- Частота встречаемости у новорожденных мальчиков:  $\sim 1/4,000$ – $1/5,000$
- Гемофилия А: фактор свертывания крови VIII (*F8A*) на X-хромосоме // Современная терапия: инъекции фактора
- Фактор свертывания крови IX на X-хромосоме: гемофилия В
- Гемофилия А, примерно в половине случаев: инверсия мобильного элемента LINE1, или L1



## Пример X-сцепленного наследования: гемофилия А (OMIM:306700)



**Figure 16.1 Hemophilia A can be caused by an inversion that disrupts the F8A gene.** (A) There is a repetitive sequence in intron 22 of the F8A gene (F8A1, red bar); (B) two additional copies are located 360 kb and 435 kb upstream of the F8A gene. Arrows indicate the relative orientations of the three copies. (C) During male meiosis, this part of the X chromosome has no homologous pairing partner. The F8A repeats may pair, forming a loop. (D) A crossover between paired F8A repeats causes inversion of a 500 kb segment. Although the F8A gene is disrupted and nonfunctional, each individual exon and its flanking intronic sequence is still intact.

Strachan, Read – *Human Molecular Genetics*

## Пример X-сцепленного наследования: гемофилия А (OMIM:306700)



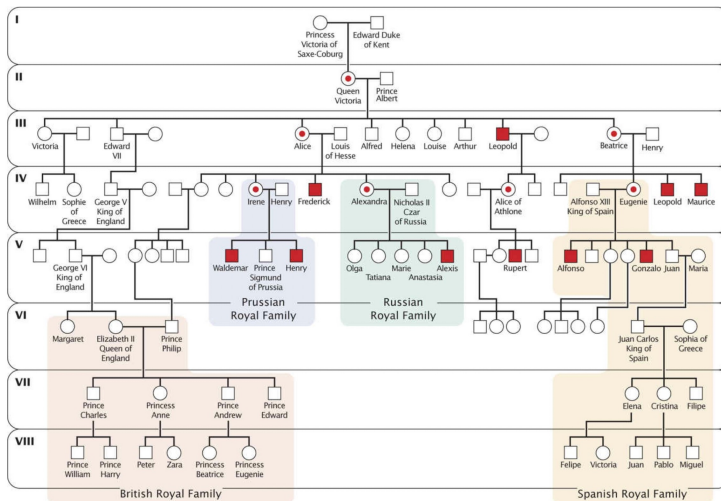
**Family of Tsar Nicholas Romanov II of Russia. The boy in front is the tsar's son Alexei, who suffered from hemophilia.**

[Mondadori Portfolio/Getty Images.]

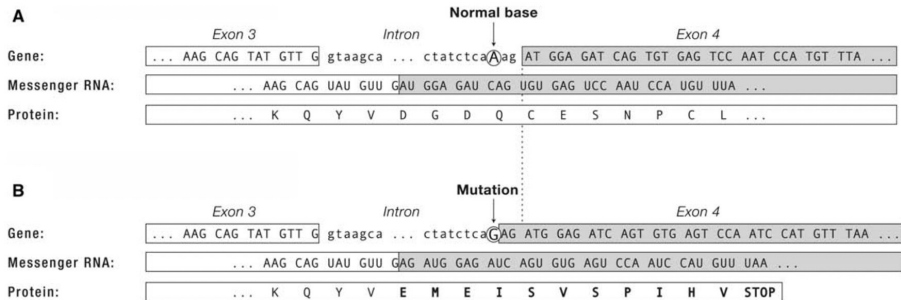
### A Royal Disease

**O**n August 12, 1904, Tsar Nicholas Romanov II of Russia wrote in his diary: "A great never-to-be forgotten day when the mercy of God has visited us so clearly." That day, Alexei, Nicholas's first son and heir to the Russian throne, had been born.

# Пример X-сцепленного наследования: гемофилия А (OMIM:306700)

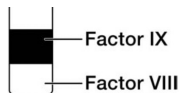


Pierce – *Genetics Essentials. Concepts and Connections*

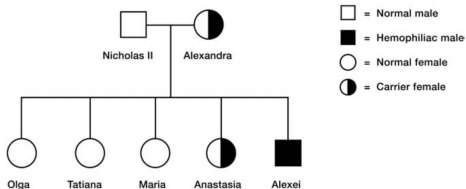
Царская семья: гемофилия А или В? // Rogaev et al. (2009) *Science*

**Figure 3.** The mutation that caused hemophilia in Queen Victoria and her descendants. Figure 3A shows the sequence of DNA in the normal Factor IX gene at the end of exon 3 and the beginning of exon 4, with the ends of the intron between them. The next lines show the messenger RNA coded for by the normal gene, followed by the sequence of amino acids coded for by this part of the normal gene. Figure 3B shows the sequence of DNA in the mutated Factor IX gene, with the point mutation (A to G) highlighted. This mutation alters the splicing, so that the last two bases of the intron (AG) are now included in the mRNA, as shown. The next lines show the messenger RNA and the order of amino acids in the mutated protein. Notice that every amino acid after the faulty splice is different, and that after 11 altered amino acids there is a stop codon, so the last part of the protein is not produced. The changed amino acids are shown in bold. The point mutation, A to

Offner (2013) doi:10.1525/abt.2013.75.9.5

Царская семья: гемофилия А или В? // Rogaev et al. (2009) *Science*

**Figure 1.** Partial map of the human X chromosome (idiogram copyright 1994 by David Adler, <http://www.pathology.washington.edu/research/cytopages/idiograms/human>; used with permission).



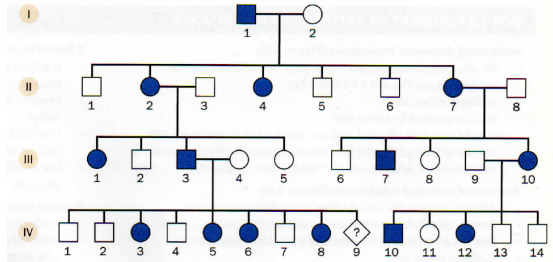
**Figure 2.** The Russian royal family. Alexandra was a granddaughter of Queen Victoria. Alexei had hemophilia, so it was assumed that his mother, Alexandra, was a carrier. The work of Rogaev et al. (2009) confirmed that Alexandra was a carrier and showed that Olga, Tatiana, and Maria were normal, while Anastasia was a carrier.

sequences that are spliced out before the final mRNA is that  $\leq 40\%$  of the DNA in the whereas only  $\sim 1.5\%$  of the D exons and actually codes for i not highly conserved, because the order of amino acids in a i folding or properties of protei usually not selected against by otide sequences at the ends of that splice out the introns. I because n coded for a mutatio faulty spl and a very In stu philia, yo in an exo would exq amino aci ever, occu the splice two-base to the tr mutation nature of why the I are critica

Offner (2013) doi:10.1525/abt.2013.75.9.5

## X-сцепленное доминантное наследование

- Поражает оба пола
- Женщины поражены слабее и более вариабельно, чем мужчины
- Поражен хотя бы один родитель
- 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*

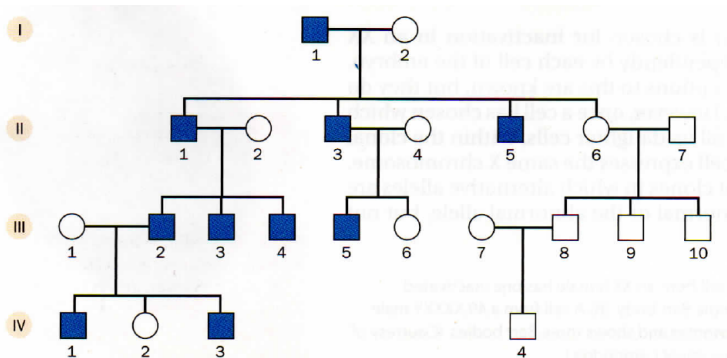
Вопрос

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</i> / <i>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
	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
X-linked recessive	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.

## 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

## 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

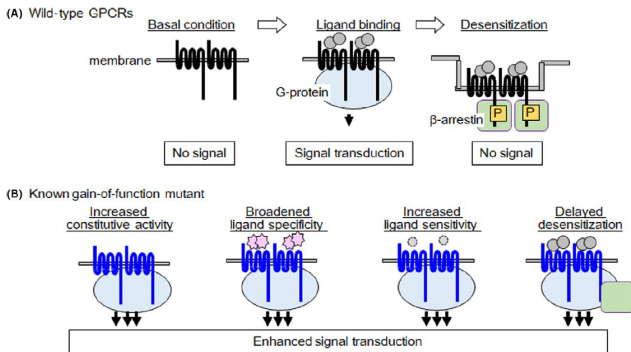
- В основном миссенсы, но также делеции без сдвига рамки считывания.
- Наличие нормального аллеля не может остановить аномальное поведение мутантного аллеля  $\Rightarrow$  доминантный?

### Вопрос

Приобретение функции из-за сдвига рамки считывания?

## Пример приобретения функции: 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.

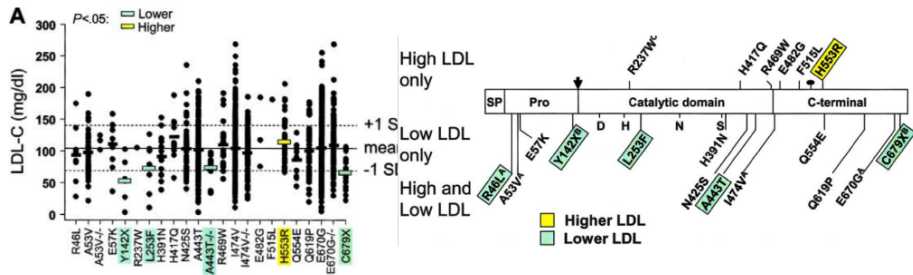


Fukami (2018) *Clin Endocrinol*

## Пример потери или приобретения функции: PCSK9

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

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

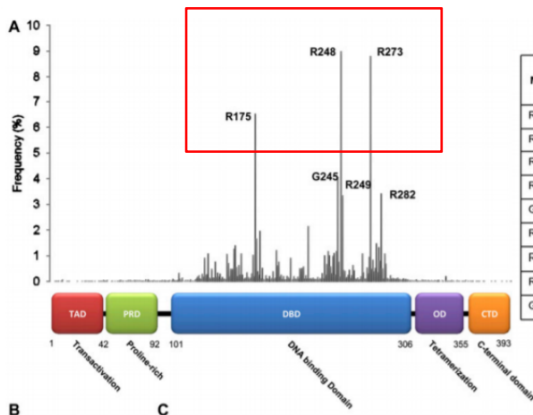


Kotowski (2006) *Am J Hum Genet*

## Пример приобретения функции: 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

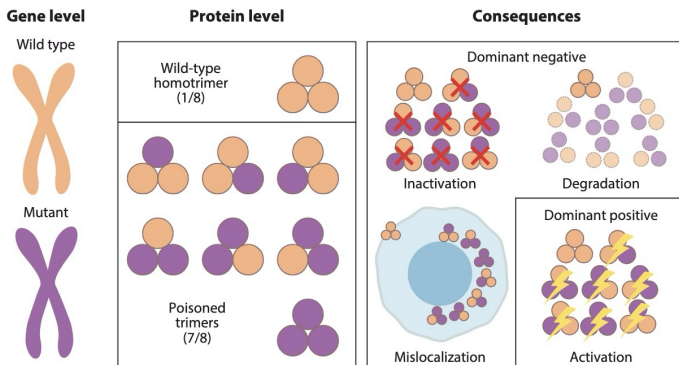
William A. Freed – Pastor (2012) *Genes & Development*

## Виды наследования и механизмы болезнетворности мутаций

Механизм	Число генов	Комментарии	Примеры
Autosomal recessive			
AR	648	Как правило, LoF. Enzymes; NA binding; receptors	ACO2, PNPT1, PNKP, LAMA2, GSS
Autosomal dominant			
HI	67	HaploInsufficiency $\approx$ LoF. NA binding; TF	FBN1, JAG1, EP300, PKD1, GATA4
DN	54	Dominant Negative, «отравление» белок-белковых комплексов. NA binding; Transporters; TFs	TNNT2, KCNQ4, GABBR2, PAX8, OTX2
GoF	74	Gain-of-function, приобретение функции. Transporters; Enzymes; NA binding	SNCA, FGFR2, PRSS1, JAK1, PSEN1
Unk	126	Неизвестно. NA binding; Transporters; TFs	ACTC1, CLTC, HNF4A, KCNJ2, GABRB2

Gerasimavicius (2022) *Nat Comm*

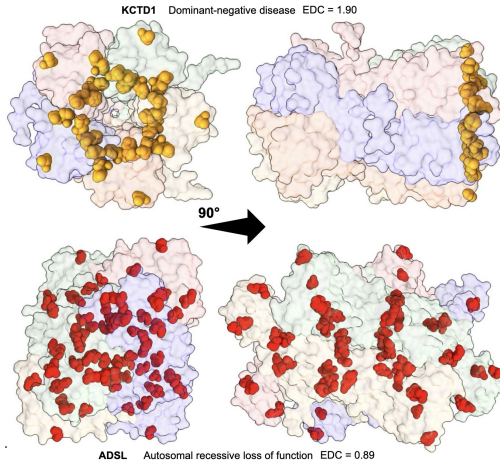
# Виды наследования и механизмы болезнетворности мутаций



**Figure 5**

Assembly-mediated dominant-negative and dominant-positive effects. For a hypothetical homotrimeric protein complex, where wild-type and mutant subunits are expressed at equal levels, seven of the eight possible trimers that form will contain at least one mutant subunit. The presence of a mutant subunit in a trimer can poison its activity and cause a dominant-negative effect in various ways, including inactivation, mislocalization, and degradation. A mutant subunit can also cause increased activation, leading to a dominant-positive effect.

## Пример: кластеризация доминантных мутаций



Pathogenic mutations in genes associated with **dominant-negative and gain-of-function** mechanisms are significantly more clustered in space than those associated with **loss-of-function** mechanisms // Gerasimavicius (2022) *Nat Comm*



# Попытки систематизации приобретения функции

**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
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 и пр.

# Пенетрантность, относительный риск, отношение шансов

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

**Риск заболевания, или пенетрантность:** вероятность заболевания при наличии мутации:  $\frac{D_m}{D_m + H_m}$  // NB: не учитывает риск заболевания в отсутствие мутации!

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

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

## Упражнение

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

Пример: OR или  $\beta$  в UK Biobank

Gene	UKB ID	Position (GRCh37)	HGV5	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*

## Пример: протромбин F2, тромбоз церебральных вен

	Cases With CVT	Controls Without CVT	Totals
20210 G>A allele present	23	4	27
20210 G>A allele absent	97	116	213
Total	120	120	240

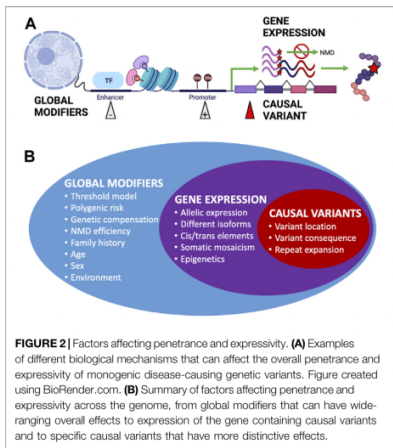
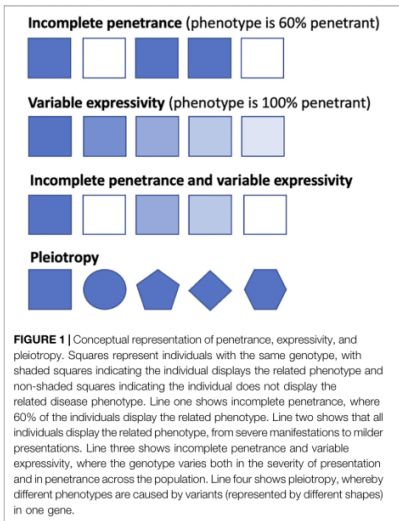
CVT, Cerebral vein thrombosis.

Because this is a case-control study, we will calculate an odds ratio:  $OR = (23/4)/(97/116) = \approx 6.9$  with 95% confidence limits of 2.3 to 20.6. The effect size of 6.9 is substantial, and 95% confidence limits exclude 1.0, thereby demonstrating a strong and statistically significant association between the 20210 G>A allele and CVT. Stated simply, individuals carrying the prothrombin 20210 G>A allele have nearly seven times greater odds of having the disease than do those who do not carry this allele.

Thompson & Thompson (2024) *Genetics and Genomics in Medicine*, 9th Ed.

NB: Oral contraceptive use in a heterozygote for prothrombin 20210 G>A raises the relative risk for cerebral vein thrombosis 30- to 150-fold!

# Пенетрантность: определения



# Пенетрантность: определения

**Incomplete penetrance** (phenotype is 60% penetrant)



**Variable expressivity** (phenotype is 100% penetrant)



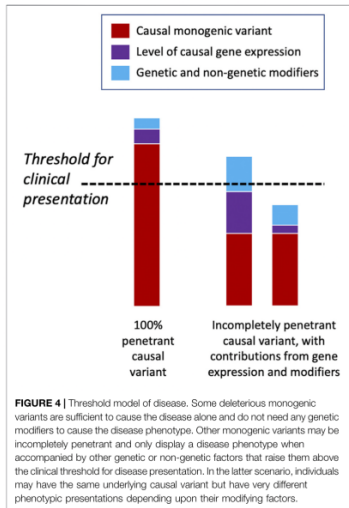
**Incomplete penetrance and variable expressivity**



**Pleiotropy**



**FIGURE 1** | Conceptual representation of penetrance, expressivity, and pleiotropy. Squares represent individuals with the same genotype, with shaded squares indicating the individual displays the related phenotype and non-shaded squares indicating the individual does not display the related disease phenotype. Line one shows incomplete penetrance, where 60% of the individuals display the related phenotype. Line two shows that all individuals display the related phenotype, from severe manifestations to milder presentations. Line three shows incomplete penetrance and variable expressivity, where the genotype varies both in the severity of presentation and in penetrance across the population. Line four shows pleiotropy, whereby different phenotypes are caused by variants (represented by different shapes) in one gene.

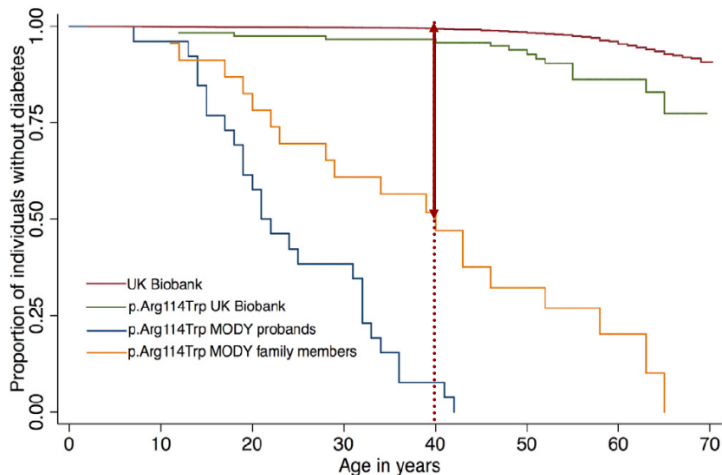


# Пенетрантность: примеры из ClinVar

Gene, variant, ClinVar ID	Disease	Penetrance
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</b> and <b>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).

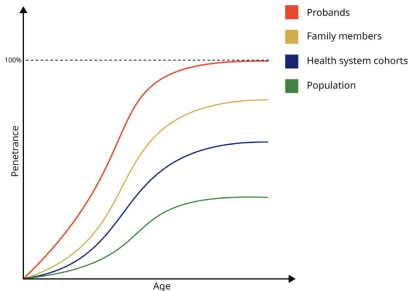


# Пенетрантность: еще пример // Wright (2019) *AJHG*



Comparison of Penetrance Estimate for HNF4A p.Arg114Trp in UK Biobank versus Previously Published Estimates from MODY Cohort Studies

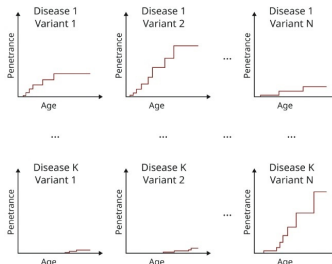
# Пенетрантность: к чему стремиться



Zaichenoka (2025) *Circulation: Genomic and Precision Medicine*



Sequencing at birth  
with subsequent  
pathogenic variant  
identification



# Пенетрантность: что появляется // McGurk (2023) *Am J Hum Genet*

## Box 1. Case study: The *MYBPC3* c.1504C>T (p.Arg502Trp) Northwestern European variant

The variant *MYBPC3* c.1504C>T (p.Arg502Trp) (GenBank: NM\_000256.3) was found in our cohort 159 times in individuals referred for HCM genetic panel sequencing (3.7% of total observations; 1.5% total case frequency). To date, the variant has been classified on ClinVar 15 times as pathogenic (ClinVar ID 42540). Penetrance has been previously estimated as ~50% (increased relative risk of 340) by 45 years old in a clinical setting, and major adverse clinical events in heterozygotes are significantly more likely when another sarcomeric variant is present.<sup>52</sup>

In our case cohort, heterozygotes of this variant were reported as broadly European ancestry (Oxford, n = 59; London, n = 11; Belfast, n = 30; LMM, n = 45; GDX, n = 14). In gnomAD, the variant was identified ten times, of which seven heterozygotes were non-Finnish Northwestern Europeans (NWE; plus one African; one South Asian, and one other), and in the UK Biobank, the variant was found 77 times, of which 68 heterozygotes were NWE (plus eight other Europeans and one other). The population frequency of the variant in Ensembl population genetics showed that the variant (rs375882485) is only found multiple times in NWE ancestry sub-cohorts. Thus, the variant is most common in NWE populations: the UK, Ireland, Belgium, the Netherlands, Luxembourg, Northern France, Germany, Denmark, Norway, Sweden, and Iceland.

We use this relatively common variant to highlight the effect of ancestry on estimated variant penetrance (see related figure in this text box):

we estimated the penetrance as 6.4% (4.6%–9.0%) with the UK Biobank cohort (93% European) and this is inflated to 35.1% (18.2%–67.5%) when we estimated the penetrance with the gnomAD dataset (45% European) as a result of the difference in the proportion of individuals with NWE ancestry. In individuals of NWE ancestry only, the penetrance of this variant is 6.4% (4.6%–9.0%). Penetrance estimated from the NWE subset of gnomAD or UKBB do not differ significantly.

Пенетрантность: что появляется // Gudmundsson (2025) *Nat Comm*

## Article

<https://doi.org/10.1038/s41467-025-61698-x>

# Exploring penetrance of clinically relevant variants in over 800,000 humans from the Genome Aggregation Database

Received: 12 June 2024

Accepted: 23 June 2025

Published online: 31 October 2025

 Check for updates

Sanna Gudmundsson<sup>1,2,3,4</sup>✉, Moriel Singer-Berk<sup>1</sup>, Sarah L. Stenton<sup>1,2,3</sup>, Julia K. Goodrich<sup>1</sup>, Michael W. Wilson<sup>1</sup>, Jonah Einson<sup>5</sup>, Nicholas A. Watts<sup>1</sup>, Genome Aggregation Database Consortium\*, Tuuli Lappalainen<sup>4,5</sup>, Heidi L. Rehm<sup>1,2</sup>, Daniel G. MacArthur<sup>1,6,7</sup> & Anne O'Donnell-Luria<sup>1,2,3</sup>✉

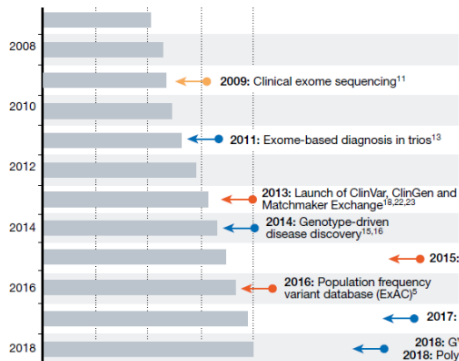
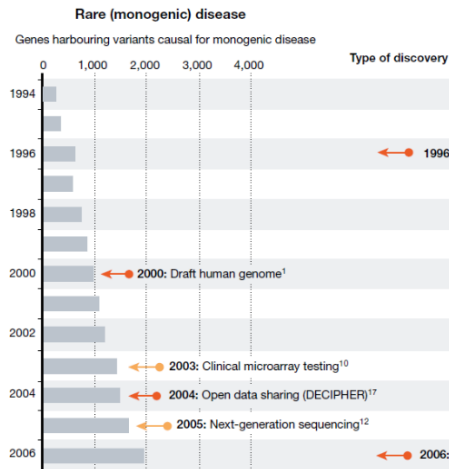
Incomplete penetrance, or absence of disease phenotype in an individual with a disease-associated variant, is a major challenge in variant interpretation. Studying individuals with apparent incomplete penetrance can shed light on underlying drivers of altered phenotype penetrance. Here, we investigate clinically relevant variants from ClinVar in 807,162 individuals from the Genome Aggregation Database (gnomAD), demonstrating improved representation in gnomAD version 4. We then conduct a comprehensive case-by-case assessment of 734 predicted loss of function variants in 77 genes associated with severe, early-onset, highly penetrant haploinsufficient disease. Here, we identify explanations for the presumed lack of disease manifestation in 701 of 734 variants (95%). Individuals with unexplained lack of disease manifestation in this set of disorders are rare, underscoring the need and power of deep case-



# Поиск генов

## Трудности при поиске генов моногенных заболеваний

- Редкие клинические наблюдения; сложно наблюдать менделевские паттерны наследования
- *De novo* мутации: нет сегрегации в семьях (AD)
- Доминантные или составные гетерозиготы (AR)
- Сдвиг в сторону очевидных вариантов; нужны последующие функциональные исследования
- Все возможные отклонения от менделевского наследования: локусная, аллельная и клиническая гетерогенность; неполная пенетрантность и вариабельная экспрессия; импринтинг; фенкопии и мозаицизм

Поиск генов, хронология 🏠// Claussnitzer (2020) *Nature*

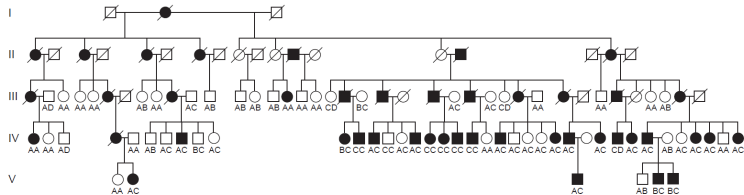
Поиск генов, хронология 🏠// Rabbani (2012) *J Hum Genet*

Table 1 Landmark events from DNA structure identification to new NGS reports

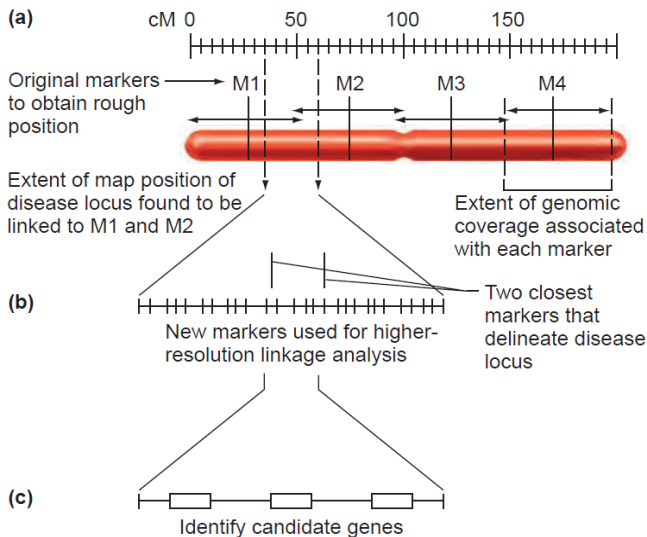
Year	Event	Reference
1953	Watson and Crick infer DNA's structure	Watson and Crick <sup>65</sup>
1964	The first nucleotide sequence of the gene encoding yeast alanine tRNA was reported	Holley <i>et al.</i> <sup>9</sup>
1977	Initial DNA sequencing methods were introduced by Sanger, Maxam and Gilbert	Sanger <i>et al.</i> <sup>10</sup> Maxam and Gilbert <sup>11</sup>
1980	First human linkage map based on restriction fragment length polymorphism	Botstein <i>et al.</i> <sup>66</sup>
1983	First dominant disease locus on the basis of linkage	Gusella <i>et al.</i> <sup>15</sup>
1985	Mullis discovered PCR technique	Mullis <i>et al.</i> <sup>67</sup>
1986	The idea of human genome sequencing was proposed	Smith <i>et al.</i> <sup>68</sup>
	The first human disease gene was cloned	Royer-Pokora <i>et al.</i> <sup>69</sup>
1987	The first homozygosity mapping was done	Lander and Botstein <sup>16</sup>
1989	First positional cloning of a recessive disease gene on the basis of linkage	Riordan <i>et al.</i> <sup>14</sup>
1993	A first-generation physical map of the human genome	Cohen <i>et al.</i> <sup>70</sup>
1995	First-genome sequence of an organism ( <i>Hemophilus influenza</i> ) was reported	Fleischmann <i>et al.</i> <sup>71</sup>
1999	First human chromosome was sequenced	Dunham <i>et al.</i> <sup>72</sup>
2000	Fruit fly genome was sequenced	Adams <i>et al.</i> <sup>73</sup>
	First assembly of the human genome was completed	Myers <i>et al.</i> <sup>74</sup>
2001	The first draft of human genome sequence was published	Venter <i>et al.</i> <sup>75</sup> Lander <i>et al.</i> <sup>76</sup>
2003	The human genome sequence was completed	Jasny and Roberts 2003
2004	Massively parallel sequencing platforms giving rise to the 'next-generation sequencing' were introduced	<a href="http://www.genome.gov/12513210">http://www.genome.gov/12513210</a>
2005	The first NGS instrument was on market	Margulies <i>et al.</i> <sup>77</sup>
2008	First individual genome based on NGS was published	Wheeler <i>et al.</i> <sup>78</sup>
2009	Proof of principle: disease-gene identification by WES	Ng <i>et al.</i> <sup>79</sup>
2010	The first successful application of WES to identify the gene for a rare Mendelian disorder	Ng <i>et al.</i> <sup>12</sup>



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

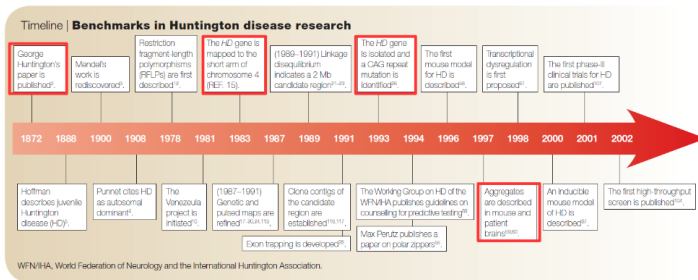
Hartwell – *Genetics. From genes to genomes*

## Поиск генов с помощью сцепления 🏠



## Начало: ген болезни Хантингтона 🏠

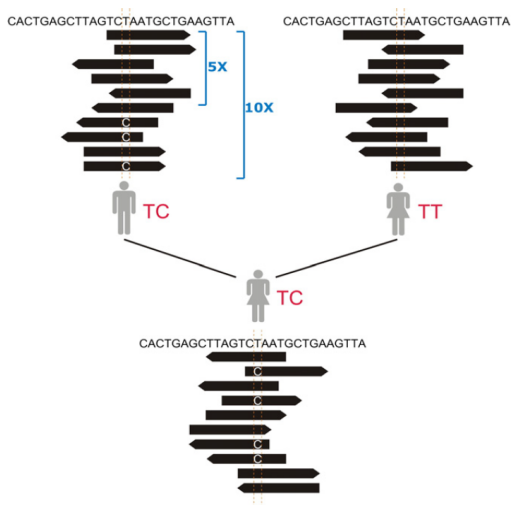
- Нейродегенеративное заболевание, начинается во взрослом возрасте (30-45 лет) и прогрессирует за 15-20 лет
- Психиатрические нарушения, нарушения двигательной функции и ухудшение когнитивных функций
- Доминантное наследование, нет спорадических форм
- Первый локус, ассоциированный с болезнью, который картировали на хромосому (1983)
- До сих пор (2018) нет лечения, кроме симптоматического // ? 2024 ?

Bates (2005) *Nat Rev Genet*

## Начало: ген болезни Хантингтона 🏠

- Высокая встречаемость в Венесуэле, один основатель
- 1983, лаборатория Джеймса Гузеллы: изучение 8 локусов, которые связали с болезнью Хантингтона, в ~4 сМ от теломерного участка 4 хромосомы
- Не было технологий для «прогулки» по хромосоме дальше 100-200 т.п.о.
- 1993: ген *HTT* (хантингтин) был клонирован совместными усилиями 9 лабораторий; транскрипт длиной 10,366 п.о. с **повторением CAG триплета (Gln)** в первом экзоне, который был полиморфным в нормальных хромосомах, и увеличивался при болезни Хантингтона
- 1993-1996: **Количество** (CAG)<sub>n</sub>: 6-35 безвредны; более 40 полностью пенетрантны и вызывают болезнь Хантингтона в течение жизни; более 70: раннее начало болезни
- Родительская антисипация: количество повторов увеличивается при передаче из поколения в поколение
- Поли-Gln повторы в болезнетворном количестве спонтанно агрегируют в амилоидные фибриллы ⇒ нейронная дегенерация
- Тестирование в детстве на начинающиеся во взрослом возрасте неизлечимые заболевания может принести больше вреда, чем пользы

# Секвенирование короткими прочтениями



Casals (2012) *J Neuroimmunology*

# Первый успех: синдром Миллера

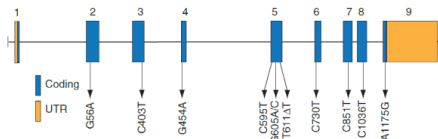
## Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng<sup>1,10</sup>, Kati J Buckingham<sup>2,10</sup>, Choli Lee<sup>1</sup>, Abigail W Bigham<sup>2</sup>, Holly K Tabor<sup>2,3</sup>, Karin M Dent<sup>4</sup>, Chad D Huff<sup>5</sup>, Paul T Shannon<sup>6</sup>, Ethylin Wang Jabs<sup>7,8</sup>, Deborah A Nickerson<sup>1</sup>, Jay Shendure<sup>1</sup> & Michael J Bamshad<sup>1,2,9</sup>

We demonstrate the first successful application of exome sequencing to discover the gene for a rare mendelian disorder of unknown cause, Miller syndrome (MIM#263750). For four affected individuals in three independent kindreds, we captured and sequenced coding regions to a mean coverage of 40× and sufficient depth to call variants at ~97% of each targeted exome. Filtering against public SNP databases and eight HapMap exomes for genes with two previously unknown variants in each of the four individuals identified a single candidate gene, *DHODH*, which encodes a key enzyme in the pyrimidine *de novo* biosynthesis pathway. Sanger sequencing confirmed the presence of *DHODH* mutations in three additional families with Miller syndrome. Exome sequencing of a small number of unrelated affected individuals is a powerful, efficient strategy for identifying the genes underlying rare mendelian disorders and will likely transform the genetic analysis of monogenic traits.

30

VOLUME 42 | NUMBER 1 | JANUARY 2010 | NATURE GENETICS



**Figure 1.** Clinical characteristics of an individual with Miller syndrome and an individual with methotrexate embryopathy.

**Figure 2.** Genomic structure of the exons encoding the open reading frame of *DHODH*. Arrows indicate the locations of 11 different mutations found in 6 families with Miller syndrome.

## По следам первого успеха

Table 2 Mendelian disease-gene identifications by exome sequencing

<i>Year</i>	<i>Disorder</i>	<i>MI</i>	<i>Location</i>	<i>Gene</i>	<i>Number exome sequenced</i>
2010					
1	Miller syndrome	AR	16q22	<i>DHODH</i>	4
2	Autoimmune lymphoproliferative syndrome	AR	11q13.3	<i>FADD</i>	1
3	Nonsyndromic hearing loss	AR	1p13.3	<i>GPSM2</i>	1
4	Combined hypolipidemia	AR	1p31.1–p22.3	<i>ANGPTL3</i>	2
5	Perrault syndrome	AR	5q21	<i>HSD17B4</i>	1
6	Complex I deficiency	AR	3q21.3	<i>ACAD9</i>	1
7	Hyperphosphatasia mental retardation syndrome	AR	1p36.11	<i>PIGV</i>	3
8	Sensenbrenner syndrome	AR	2p24.1	<i>WDR35</i>	2
9	Cerebral cortical malformations	AR	19q13.12	<i>WDR62</i>	3
10	3MC syndrome	AR	3q27–q28	<i>MASP1</i>	2
11	Kabuki syndrome	AD	12q13.12	<i>MLL2</i>	10
12	Schinz-Giedion syndrome	AD	18q21.1	<i>SETBP1</i>	4
13	Spinocerebellar ataxia	AD	20p13	<i>TGM6</i>	4
14	Terminal osseous dysplasia	XLD	Xq28	<i>FLNA</i>	2
2011					
15	Nonsyndromic mental retardation	AR	19p13.12	<i>TECR</i>	6
16	Retinitis pigmentosa	AR	1p36.11	<i>DHDDS</i>	4

## Поиск «иголки в стоге сена»

Необходимо определить, какой из выявленных вариантов наиболее вероятно [нарушает функцию гена и] является причиной болезни.

<i><b>SNVs</b></i>	<i><b>Average</b></i>	<i><b>Deviation</b></i>
PTV <i>HIGH</i>	97	6
Missense <i>MODERATE</i>	6291	139
Synonymous <i>LOW</i>	7192	88
Other <i>MODIFIER</i>	561	13
<i><b>Indels</b></i>		
Frameshift	69	3
Other	41	3

<i><b>SNVs</b></i>	<i><b>Average</b></i>	<i><b>Deviation</b></i>
Singleton	18	13
<0.01%	177	30
0.01-1%	273	23
1-10%	1308	72
>10%	12365	109
<i><b>Indels</b></i>		
<=5%	15	5
>5%	151	6

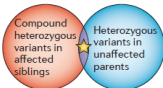
Варианты в индивидуальном экзOME: ExAC, Lek (2016), *Nature*



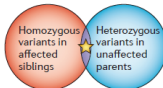
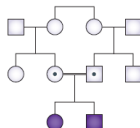
## Стратегии поиска

## a Inherited mutations

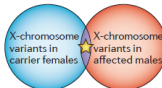
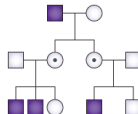
Autosomal recessive



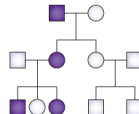
Consanguineous autosomal recessive



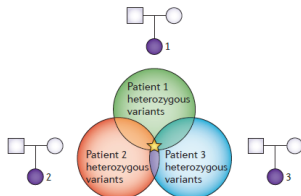
X-linked recessive



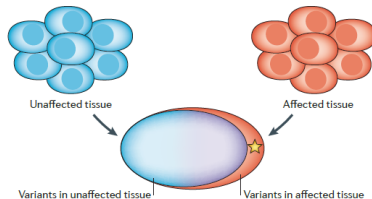
Autosomal dominant



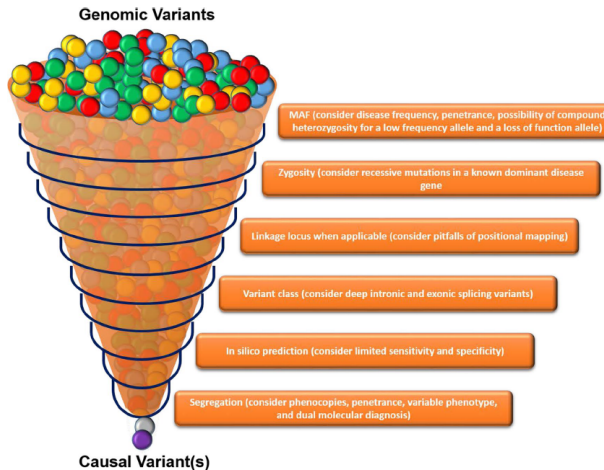
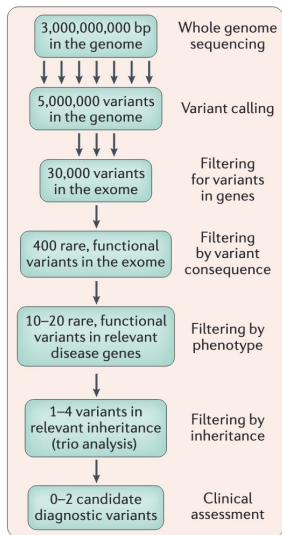
## b De novo dominant mutations



## c Mosaic mutations

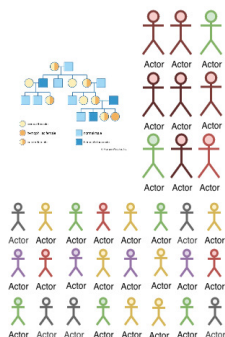
Boycott (2013) *Nat Rev Genet*

## Стратегии поиска



Wright (2018) *Nat Rev Genet*, Alkuraya (2016) *Hum Genet*

# Куда попадают результаты поиска?



**NGS  
+  
VEP**

**PubMed**

## ClinVar

Public archive of reports of human variations classified for diseases and drug responses, with supporting evidence.

## OMIM

An Online Catalog of Human Genes and Genetic Disorders

## Рекомендации: ClinGen, ACMG, ESHG

ClinGen: A central resource that defines **the clinical relevance of genes and variants** for use in precision medicine and research.

## gnomAD

Genome Aggregation Database

## UK Biobank

The world's most comprehensive dataset of biological, health and lifestyle

## HPO

The **Human Phenotype Ontology**: a vocabulary of phenotypic abnormalities encountered in human disease.

**Люди**

**Варианты**

**Гены**

## Сложности и ограничения

### Основные сложности:

- Редкие заболевания: малое количество случаев и/или семей
- Варибельная пенетрантность
- Неизвестный тип наследования
- Гетерогенность локусов и фенотипов
- *De novo* или унаследованные варианты (спорадические vs семейные случаи)
- Частота аллелей может быть обманчивой
- Ограниченность предсказательных алгоритмов *in silico*

### Ограничения полноэкзомного секвенирования (WES)

- Множество не кодирующих вариантов не детектируется
- Сложности в детекции структурных вариантов и CNV
- Ложноотрицательные (покрытие) и ложноположительные (паралоги?) варианты
- Большое количество вариантов-кандидатов, необходима фильтрация

### Ограничения полногеномного секвенирования (WGS)

- Слишком много данных, необходимо еще больше фильтрации
- Стоимость секвенирования и обработки данных

# Клиническая значимость

Стандарты и рекомендации // Richards (2015) *Genetics in Medicine*

## Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Клиническая значимость любого варианта последовательности попадает на точку спектра, который простирается от скорее всего безвредного до тех, которые скорее всего безвредны.

Система классификации из 5 уровней для менделевских заболеваний:

- Pathogenic (P)
- Likely pathogenic (LP)
- Benign (B)
- Likely benign (LB)
- Variant of unknown significance (VUS)

Мы предлагаем термины *Likely pathogenic* и *Likely benign* для обозначения тех случаев, когда с более чем 90% уверенностью можно сказать, что вариант или безвредный, или безвредный, чтобы предоставить лабораториям **общее, пусть и субъективное, определение.**

## Стандарты и рекомендации: отечественная версия

МЕДИЦИНСКАЯ ГЕНЕТИКА. — 2017. — №7

# Руководство по интерпретации данных, полученных методами массового параллельного секвенирования (MPS)

Рыжкова О.П.<sup>1</sup>, Кардымон О.Л.<sup>2</sup>, Прохорчук Е.Б.<sup>3</sup>, Коновалов Ф.А.<sup>4</sup>, Масленников А.Б.<sup>1</sup>, Степанов В.А.<sup>6</sup>, Афанасьев А.А.<sup>7</sup>, Заклязьминская Е.В.<sup>8</sup>, Костарева А.А.<sup>9</sup>, Павлов А.Е.<sup>10</sup>, Голубенко М.В.<sup>6</sup>, Поляков А.В.<sup>1</sup>, Куцев С.И.<sup>1</sup>

<sup>1</sup> ФГБНУ «Медико-генетический научный центр», Москва; e-mail ryzhkova@dnalab.ru

## Терминология

Предлагается заменить широко используемые термины «мутация» и «полиморфизм» на термин «вариант нуклеотидной последовательности» со следующими характеристиками:

- патогенный (pathogenic);
- вероятно патогенный (likely pathogenic);
- неопределенного значения (uncertain significance);
- вероятно доброкачественный (likely benign);
- доброкачественный (benign).

## Стандарты и рекомендации: отечественная версия

**ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЕ АСПЕКТЫ*****Руководство по интерпретации данных  
последовательности ДНК человека, полученных  
методами массового параллельного секвенирования (МПС)  
(редакция 2018, версия 2)***

Рыжкова О.П.<sup>1</sup>, Кардымон О.Л.<sup>1</sup>, Прохорчук Е.Б.<sup>2</sup>, Коновалов Ф.А.<sup>3</sup>, Масленников А.Б.<sup>4</sup>,  
Степанов В.А.<sup>5</sup>, Афанасьев А.А.<sup>6</sup>, Заклязьминская Е.В.<sup>7</sup>, Ребриков Д.В.<sup>8</sup>, Савостьянов К.В.<sup>9</sup>,  
Готов А.С.<sup>10,11</sup>, Костарева А.А.<sup>12</sup>, Павлов А.Е.<sup>13</sup>, Голубенко М.В.<sup>5</sup>, Поляков А.В.<sup>1</sup>, Куцев С.И.<sup>1</sup>

1 — ФГБНУ «Медико-генетический научный центр»,

**Введение**

Основой для настоящего руководства послужили протоколы по интерпретации результатов массового параллельного секвенирования (MPS), разновидностью которого являются методы секвенирования нового поколения (NGS), разработанные

в США и Европе [1–8]. Эти протоколы были переработаны группой ведущих российских специалистов в области генетики и биоинформатики. Редакция руководства была опубликована в 2018 году в журнале «Медицинская генетика» [9]. Его обсу



# Стандарты и рекомендации: отечественная версия

## МЕДИЦИНСКАЯ ГЕНЕТИКА. 2019. №

- нужно быть осторожным при интерпретации вариантов, расположенных близко к 3' концу гена;
- требует осторожности интерпретация сплайс-вариантов, которые прогнозируемо приводят к пропуску экзонов, но оставляют оставшуюся часть белка интактной и не меняют рамку считывания.
- необходима осторожность при наличии нескольких транскриптов одного гена, так как один и тот же вариант может по-разному классифицироваться в зависимости от его положения в конкретном транскрипте.

### Критерий «сильный» (strong, PS 1-4)

**PS1.** Этому критерию соответствует вариант нуклеотидной последовательности, приводящий к замене на ту же аминокислоту в том же положении, что и вариант, ранее описанный как патогенный при этом заболевании. Например, если описана замена G>C, приводящая к замене Val>Leu, то замена G>T, приводящая к той же аминокислотной замене, соответствует критерию PS1. Следует принять во внимание, что варианты, патогенность которых обусловлена измене-

(например, в активном сайте фермента), в которых описаны доброкачественные изменения.

**PM2.** К этому критерию относят вариант, отсутствующий в контрольной выборке или встречающийся в ней с крайне низкой частотой: для аутосомно-доминантных заболеваний частота аллеля не должна превышать 0,01%, для аутосомно-рецессивных заболеваний — 0,5%, для доминантных X-сцепленных — 0,01%, для рецессивных X-сцепленных — 0,3%. При отнесении варианта к этому критерию следует учитывать, что в базах данных вариантов нуклеотидной последовательности, выявленных методами MPS, может содержаться неполная информация о частоте в популяциях делеций/инсерций. Кроме того, необходимо учитывать неполную пенетрантность, возраст начала заболевания, наличие фенкопий заболевания.

**PM3.** Этому критерию соответствует вариант, находящийся в транс-положении с описанным патогенным вариантом при рецессивных заболеваниях. Для подтверждения транс-положения вариантов требуется обследование родителей (или потомков) па-

Стандарты и рекомендации // Richards (2015) *Genetics in Medicine*

## Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Evidence of pathogenicity	Category
Very strong	<p>PVS1 null variant (nonsense, frameshift, canonical <math>\pm 1</math> or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease</p> <p>Caveats:</p> <ul style="list-style-type: none"> <li>• Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>)</li> <li>• Use caution interpreting LOF variants at the extreme 3' end of a gene</li> <li>• Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact</li> <li>• Use caution in the presence of multiple transcripts</li> </ul>
Strong	<p>PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change</p> <p>Example: Val→Leu caused by either G&gt;C or G&gt;T in the same codon</p> <p>Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level</p> <p>PS2 De novo (<u>both</u> maternity and paternity confirmed) in a patient with the disease and no family history</p> <p>Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity.</p> <p>PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product</p> <p>Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.</p> <p>PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls</p> <p>Note 1: Relative risk or OR, as obtained from case-control studies, is <math>&gt;5.0</math>, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance.</p> <p>Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.</p>
Moderate	<p>PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation</p>

Стандарты и рекомендации // Richards (2015) *Genetics in Medicine*

## Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Evidence of pathogenicity	Category
Very strong	<p>PVS1 null variant (nonsense, frameshift, canonical <math>\pm 1</math> or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease</p> <p>Caveats:</p> <ul style="list-style-type: none"> <li>Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>)</li> <li>Use caution interpreting LOF variants at the extreme 3' end of a gene</li> <li>Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact</li> <li>Use caution in the presence of multiple transcripts</li> </ul>
Strong	<p>PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change</p> <p>Example: Val→Leu caused by either G&gt;C or G&gt;T in the same codon</p> <p>Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level</p> <p>PS2 De novo (<b>both</b> maternity and paternity confirmed) in a patient with the disease and no family history</p> <p>Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity.</p> <p>PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product</p> <p>Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.</p> <p>PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls</p> <p>Note 1: Relative risk or OR, as obtained from case-control studies, is &gt;5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance.</p> <p>Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.</p>
Moderate	<p>PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation</p>

Стандарты и рекомендации // Richards (2015) *Genetics in Medicine*

## Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

## Evidence of benign impact

## Category

## Stand-alone

BA1 Allele frequency is &gt;5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium

## Strong

BS1 Allele frequency is greater than expected for disorder (see Table 6)

BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age

BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing

BS4 Lack of segregation in affected members of a family

Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.

## Supporting

BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease

BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in *cis* with a pathogenic variant in any inheritance pattern

BP3 In-frame deletions/insertions in a repetitive region without a known function

BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)

Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.

BP5 Variant found in a case with an alternate molecular basis for disease

BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation

BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus

Стандарты и рекомендации // Richards (2015) *Genetics in Medicine*

## Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

**Table 5** Rules for combining criteria to classify sequence variants

Pathogenic	<ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) AND               <ul style="list-style-type: none"> <li>(a) <math>\geq 1</math> Strong (PS1–PS4) OR</li> <li>(b) <math>\geq 2</math> Moderate (PM1–PM6) OR</li> <li>(c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) OR</li> <li>(d) <math>\geq 2</math> Supporting (PP1–PP5)</li> </ul> </li> <li>(ii) <math>\geq 2</math> Strong (PS1–PS4) OR</li> <li>(iii) 1 Strong (PS1–PS4) AND               <ul style="list-style-type: none"> <li>(a) <math>\geq 3</math> Moderate (PM1–PM6) OR</li> <li>(b) 2 Moderate (PM1–PM6) AND <math>\geq 2</math> Supporting (PP1–PP5) OR</li> <li>(c) 1 Moderate (PM1–PM6) AND <math>\geq 4</math> supporting (PP1–PP5)</li> </ul> </li> </ul>
Likely pathogenic	<ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) AND 1 moderate (PM1–PM6) OR</li> <li>(ii) 1 Strong (PS1–PS4) AND 1–2 moderate (PM1–PM6) OR</li> <li>(iii) 1 Strong (PS1–PS4) AND <math>\geq 2</math> supporting (PP1–PP5) OR</li> <li>(iv) <math>\geq 3</math> Moderate (PM1–PM6) OR</li> <li>(v) 2 Moderate (PM1–PM6) AND <math>\geq 2</math> supporting (PP1–PP5) OR</li> <li>(vi) 1 Moderate (PM1–PM6) AND <math>\geq 4</math> supporting (PP1–PP5)</li> </ul>
Benign	<ul style="list-style-type: none"> <li>(i) 1 Stand-alone (BA1) OR</li> <li>(ii) <math>\geq 2</math> Strong (BS1–BS4)</li> </ul>
Likely benign	<ul style="list-style-type: none"> <li>(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) OR</li> <li>(ii) <math>\geq 2</math> Supporting (BP1–BP7)</li> </ul>
Uncertain significance	<ul style="list-style-type: none"> <li>(i) Other criteria shown above are not met OR</li> <li>(ii) the criteria for benign and pathogenic are contradictory</li> </ul>

Стандарты и рекомендации // Richards (2015) *Genetics in Medicine*

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
<b>Population data</b>	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
<b>Computational and predictive data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product BP4  Missense in gene where only truncating cause disease BP1  Silent variant with non predicted splice impact BP7  In-frame indels in repeat w/out known function BP9	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5  Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
<b>Functional data</b>	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
<b>Segregation data</b>	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
<b>De novo data</b>				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
<b>Allelic data</b>		Observed in trans with a dominant variant BP2  Observed in cis with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
<b>Other database</b>		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
<b>Other data</b>		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

## Sherloc: уточнение критериев классификации вариантов

### Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria

The ACMG–AMP criteria were not capturing certain qualitative considerations. Therefore, we first posed a normative question: **“What kind of evidence, and how much, should be required for a pathogenic classification?”** We first recognized that there are two general types of evidence: clinical and functional.

1. **Clinical evidence** describes the correlation of the variant with disease (or absence of disease) in human populations, and includes observations in affected and unaffected individuals and families.
2. **Functional evidence** describes the molecular consequence of a variant on various gene products and includes the results of molecular and cellular experiments, and predictions about functional effects based on variant type or complex computational algorithms.

Clearly, clinical and functional evidence are both important: a variant is pathogenic if it disrupts a gene product in a way that leads to human disease, and is benign if it has an effect that does not lead to disease in humans.

Nykamp (2017) *Genetics in Medicine*

## Sherloc: уточнение критериев классификации вариантов

### Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria

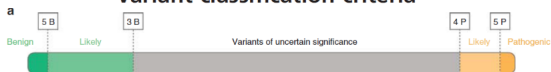
Although both clinical and functional evidence are relevant, they have a **hierarchical relationship**. Clinical data describe human disease directly, whereas functional data are relevant to disease only to the extent to which the measured property correlates with disease physiology. Therefore, when a discrepancy or conflict arises between clinical and functional observations, **the clinical observations should be considered more persuasive**. Broadly speaking, a variant should not be considered pathogenic if it is present in a large percentage of healthy individuals (clinical data), even if a measurable effect on protein function has been observed in an experimental assay (functional data). Conversely, a variant should be considered pathogenic if it is present in many affected individuals and has not been observed in healthy individuals (clinical data), even if it is predicted to be nondeleterious and has been demonstrated to have no effect on a measured protein property (functional data).

Nykamp (2017) *Genetics in Medicine*



# Sherloc: уточнение критериев классификации вариантов

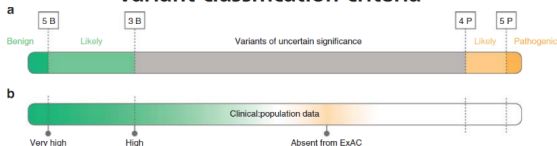
## Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria



Nykamp (2017) *Genetics in Medicine*

# Sherloc: уточнение критериев клин. классификации вариантов

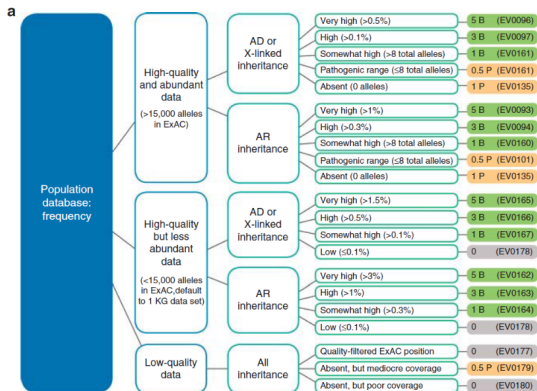
## Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria



Nykamp (2017) *Genetics in Medicine*

# Sherloc: уточнение критериев классификации вариантов

## Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria

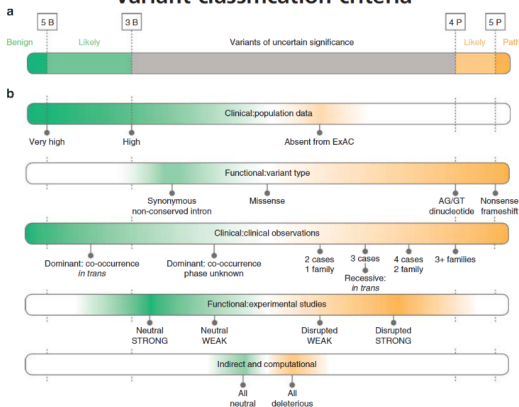


Population data: Sherloc criteria and decision tree

Nykamp (2017) *Genetics in Medicine*

# Sherloc: уточнение критериев клин. классификации вариантов

## Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria



Classification scoring thresholds and evidence categories

Nykamp (2017) *Genetics in Medicine*

## Sherloc: уточнение критериев классификации вариантов

### Пример 1: *TTC8* с.459G>A (p.Thr153=)

- Очень редкая синонимичная замена (0.02% в ExAC) в гене, который может вызвать синдром Барде-Бидля
- Предсказано, что нарушает нормальный сплайсинг
- Наблюдается в гомозиготном состоянии у трех больных братьев/сестер в одной семье
- Наблюдается в гомозиготном состоянии у неродственных пациентов и теперь классифицируется как **патогенная**

### Пример 2: *CDH1* с.1118C>T (p.Pro373Leu) вариант в гене, ассоциированным с наследственным диффузным раком желудка и дольковым РМЖ (лобулярная карцинома)

- Отсутствует в ExAC
- Поддерживается функциональными исследованиями: нарушает клеточно-клеточную адгезию и приводит к повышенной клеточной подвижности и активации *EGFR*, митоген-активируемой протеиновой киназы и Src-киназы
- Компьютерные предсказания подтверждают этот вывод
- Однако клинические наблюдения неоднозначны: этот вариант был найден и у больных, и у здоровых в одной семье.

Без поддерживающих клинических наблюдений, **классификация варианта как возможно патогенного преждевременна**

Nykamp (2017) *Genetics in Medicine*

## Sherloc: уточнение критериев клин. классификации вариантов

### Пример 3: *CDKN2A* c.9\_32dup24

- Дупликация без сдвига рамки считывания
- Предсказано, что не имеет влияния на функцию белка
- Показано, что не влияет на связывание CDK4 или CDK6
- Найдена у нескольких пациентов с меланомой
- Сегрегирует болезнь (неполная пенетрантность) в нескольких семьях с меланомой

**Изобилие положительных клинических доказательств превосходит отрицательные функциональные доказательства** (эффективность связывания CDK4/6 не является релевантным молекулярным последствием) **Пример 4: *SCN5A* c.3578G>A (p.Arg1193Gln)**

- Миссенс-замена в потенциалзависимом натриевом канале
- Показано, что дестабилизирует порог инактивации и приводит к постоянному току *in vitro*
- Глицин присутствует в аналогичной позиции в ортологе лошади
- Частота более 7% в Восточно-Азиатской популяции, с 17 гомозиготами, найденными в ExAC

**Изобилие отрицательных клинических доказательств превосходит положительные функциональные доказательства**

# Генетическая диагностика

## Генетическая диагностика: постановка задачи

**Применимость:** атипичное проявление; симптомы, схожие у нескольких нарушений; сложность подтверждения клиническими или лабораторными критериями

**Вход:** Клинические симптомы (НПО), медицинские записи

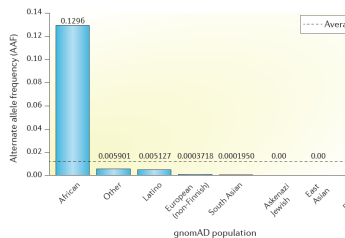
**Аннотация:** VEP; gnomAD; ClinVar, OMIM

**Фильтрация вариантов и приоритизация:**

20,000-100,000 → 50 - 1,000

известные патогенные варианты; редкие ( $MAF < 0.5\%$ ) или новые PTV; другие варианты, гены с ассоциированными фенотипами (ClinVar, OMIM, НПО)

**Выход:** клинический отчет с диагностикой, кандидатными генами/вариантами; направления; запрос секвенирования по Сенгеру



Частота варианта в различных популяциях // Bamshad (2011) *Nat Rev Gene*



Генетическая диагностика: примеры // Bamshad (2011) *Nat Rev Genet*

1. Новый гомозиготный вариант (Asp652Asn) в транспорте растворенных веществ 26, член 3 *SLC26A3* – ген, вызывающий **врожденную хлоридную диарею** – был найден у ребенка, у которого изначально подозревался другой диагноз – **синдром Барттера**
2. Новый вариант Cys203Tyr в X-сцепленном ингибиторе апоптоза (*XIAP*) у маленького мальчика с острым воспалительным заболеванием кишечника, точный диагноз было сложно поставить. Мутации в гене *XIAP* вызывают **X-сцепленный лимфопролиферативный синдром 2 типа (XLP2)**, однако острый колит является нетипичным симптомом XLP2. Точный диагноз XLP2 предложил путь лечения.

# Пример заключения по результатам генетического тестирования



**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ  
«НАЦИОНАЛЬНЫЙ МЕДИЦИНСКИЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР ТЕРАПИИ И  
ПРОФИЛАКТИЧЕСКОЙ МЕДИЦИНЫ»**

Министерства здравоохранения Российской Федерации  
(ФГБУ «НМИЦ ТПМ» Минздрава России)

101000, Москва, Петроверигский пер., д.10, стр.3, тел: (495) 623-86-36, факс: (495) 621-01-22

**Результаты генетического тестирования  
в рамках научного исследования  
«Изучение генетики сердечно-сосудистых заболеваний»**

<b>ФИО</b>	М	
<b>Пол</b>	мужской	
<b>Дата рождения</b>	XX.XX.1989	
<b>Дата взятия биоматериала</b>	15.03.2024	
<b>Вид биоматериала</b>	Кровь венозная	
<b>Фенотип/Диагноз</b>	Семейная гиперхолестеринемия, гетерозиготная форма	

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

**Таблица 1. Варианты нуклеотидной последовательности, имеющие вероятное отношение к фенотипу**

Ген (№ OMIM)	Положение (GRCh38/hg38)	Транскрипт	Генотип	Экзон	Изменение кДНК	Изменение белка	Частота аллеля*	dbSNP ID
<i>LDLR</i> (606945)	chr19:11 107433 G>A	NM_000527.5	G/A	6	c.859G>A	p.Gly287Ser	0,001675%	rs375495026

## Пример заключения по результатам генетического тестирования

**Таблица 2. Побочно выявленные варианты в генах, рекомендованных к проверке международными руководствами<sup>1,2</sup>**

Ген (№ OMIM)	Положение (GRCh38/hg38)	Транскрипт	Генотип	Экзон	Изменение кДНК	Изменение белка	Частота аллеля*	dbSNP ID
Не выявлено.								

\*согласно Genome Aggregation Database (gnomAD) v4.1.0

У пациента методом полногеномного секвенирования был проведен поиск вариантов нуклеотидной последовательности (ВНП), связанных с развитием наследственных дислипидемий.

В результате исследования выявлен ВНП в гене *LDLR* в гетерозиготном состоянии (Табл. 1). Патогенные и вероятно патогенные варианты в гене *LDLR* связаны с развитием семейной гиперхолестеринемии (№ OMIM 143890).

В соответствии с российскими рекомендациями для интерпретации данных, полученных методами массового параллельного секвенирования<sup>1</sup>, для оценки патогенности варианта были применены следующие критерии патогенности:

- PM2 (зарегистрирован в контрольной выборке gnomAD с частотой, не превышающей ожидаемую для моногенной патологии с аутосомно-доминантным типом наследования (0,01%))
- PM5 (миссенс-вариант, затрагивающий позицию, в которой ранее был описан другой патогенный ВНП c.859G>T, p.Gly287Cys (ClinVar ID 251489))<sup>3</sup>
- PP3 (компьютерные программы предсказания подтверждают его патогенность: REVEL score = 0,902<sup>4,5</sup>, MetaSVM rankscore = 0,99233<sup>6</sup>, MetaLR rankscore = 0,9806<sup>6</sup> (ВНП с REVEL score  $\geq 0,75$  или MetaSVM rankscore  $\geq 0,82257$  или MetaLR rankscore  $\geq 0,81101$  считали потенциально патогенным))
- PP4 (выявлен у анализируемого пациента с клинически подтвержденным диагнозом СГХС после исключения альтернативных причин гиперлипидемии («вероятный» диагноз по критериям Саймона Брума).

ACMG-56 2.0: гены вторичных находок // Kalia (2017) *Genet Med*

- **ACMG-56:** список генов, случайные или вторичные находки в которых необходимо выносить в заключение, вне зависимости от причины прохождения генетического тестирования
- **Цель:** идентифицировать и управлять риском для отобранных высокопенетрантных генетических заболеваний [с доминантным типом наследования] за счет основанными вмешательств, направленных на предупреждение или значительное уменьшение заболеваемости и смертности
- **Обновления:** появился в 2013, в 2017 убрали один ген, добавили 4
- **Пример:** *ATP7B* ассоциирован с аутосомно-рецессивным заболеванием Вилсона (OMIM 277900). Заболеваемость среди гомозигот напрямую коррелирует с отложениями меди в печени, мозге и глазах. Болезнь прогрессирует, и, если остается без лечения, с большой вероятностью вызывает раннюю смерть. В некоторых случаях, отказ печени может быть знаком проявления. <...> Лечение для болезни Вилсона включает в себя назначение хелатирующих медь агентов и/или цинка для блокировки абсорбции меди в кишечнике; **лечение крайне эффективно, если назначено до проявления симптомов**

ACMG-56 2.0: гены вторичных находок // Kalia (2017) *Genet Med*

Table 1 ACMG SF v2.0 genes and associated phenotypes recommended for return of secondary findings in clinical sequencing

Phenotype	MIM disorder	PMID Gene Reviews entry	Typical age of onset	Gene	MIM gene	Inheritance <sup>a</sup>	Variants to report <sup>b</sup>
Hereditary breast and ovarian cancer	604370 612555	20301425	Adult	<i>BRCA1</i> <i>BRCA2</i>	113705 600185	AD	KP and EP
Li-Fraumeni syndrome	151623	20301488	Child/adult	<i>TP53</i>	191170	AD	KP and EP
Peutz-Jeghers syndrome	175200	20301443	Child/adult	<i>STK11</i>	602216	AD	KP and EP
Lynch syndrome	120435	20301390	Adult	<i>MLH1</i> <i>MSH2</i> <i>MSH6</i> <i>PMS2</i>	120436 609309 600678 600259	AD	KP and EP
Familial adenomatous polyposis	175100	20301519	Child/adult	<i>APC</i>	611731	AD	KP and EP
<i>MYH</i> -associated polyposis; adenomas, multiple colorectal, <i>FAP</i> type 2; colorectal adenomatous polyposis, autosomal recessive, with pilomatricomas	608456 132600	23035301	Adult	<i>MUTYH</i>	604933	AR <sup>c</sup>	KP and EP
Juvenile polyposis	174900	20301642	Child/adult	<i>BMPR1A</i> <i>SMAD4</i>	601299 600993	AD	KP and EP
Von Hippel–Lindau syndrome	193300	20301636	Child/adult	<i>VHL</i>	608537	AD	KP and EP
Multiple endocrine neoplasia type 1	131100	20301710	Child/adult	<i>MEN1</i>	613733	AD	KP and EP
Multiple endocrine neoplasia type 2	171400 162300	20301434	Child/adult	<i>RET</i>	164761	AD	KP
Familial medullary thyroid cancer <sup>d</sup>	1552401	20301434	Child/adult	<i>RET</i>	164761	AD	KP
<i>PTEN</i> hamartoma tumor syndrome	153480	20301661	Child/adult	<i>PTEN</i>	601728	AD	KP and EP
Retinoblastoma	180200	20301625	Child	<i>RB1</i>	614041	AD	KP and EP
Hereditary paraganglioma-pheochromocytoma syndrome	168000 (PGL1) 601650 (PGL2) 605373 (PGL3) 115310 (PGL4)	20301715	Child/adult	<i>SDHD</i> <i>SDHAF2</i> <i>SDHC</i> <i>SDHB</i>	602690 613019 602413 185470	AD	KP and EP KP KP and EP
Tuberous sclerosis complex	191100 613254	20301399	Child	<i>TSC1</i> <i>TSC2</i>	605284 191092	AD	KP and EP

# Переклассификация значимости // Shah (2018) *Am J Hum Genet*

## Identification of Misclassified ClinVar Variants via Disease Population Prevalence

Naisha Shah,<sup>1</sup> Ying-Chen Claire Hou,<sup>1</sup> Hung-Chun Yu,<sup>1</sup> Rachana Sainger,<sup>1</sup> C. Thomas Caskey,<sup>2</sup> J. Craig Venter,<sup>1,3,\*</sup> and Amalio Telenti<sup>3,\*</sup>

The American Journal of Human Genetics 102, 609–619, April 5, 2018 609

- Whole-genome sequence data from 10,495 unrelated individuals to contrast population frequency of pathogenic variants to the expected population prevalence of the disease
- 2.6% at risk for disease for 16 of the 26 ACMG-59 conditions,
- 4.9% were carriers for 17 of the 26 ACMG-59 conditions.
- 1.5%–6.5%, the estimated range of screened individuals that would have an incidental finding for the ACMG-56
- Allele frequency  $\times$  disease prevalence for 25,505 variants:  
many pathogenic variants have **low penetrance**  
or **incorrect pathogenicity**

# Некоторые перспективы WES/WGS-диагностики

## SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

### GENETIC DIAGNOSIS

# Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation

Michelle M. Clark<sup>1</sup>, Amber Hildreth<sup>1,2,3</sup>, Sergey Batalov<sup>1</sup>, Yan Ding<sup>1</sup>, Shimul Chowdhury<sup>1</sup>, Kelly Watkins<sup>1</sup>, Katarzyna Ellsworth<sup>1</sup>, Brandon Camp<sup>1</sup>, Cyrielle L. Kint<sup>4</sup>, Calum Vazouhian<sup>5</sup>

Use type	Retrospective patients								Prospective patients							
Subject ID	263	6124	3003	6194	290	352	362	374	7052	412						
Age	8 days	14 years	1 year	5 days	3 days	7 weeks	4 weeks	2 days	17 months	3 days						
Sex	♀	♂	♀	♀	♂	♀	♂	♂	♂	♂						
Abbreviated presentation	Neonatal seizures	Rhabdomyolysis	Dystonia, dev. delay	Hypoglycemia, seizures	Pulmonary hemorrhage, PPHN	Diabetic ketoacidosis	Neonatal seizures	HIE, anemia	Pseudomonas septic shock	Neonatal seizures						
Method	Auto.	Auto.	Auto.	Auto.	Auto.	Std.	Auto.	Std.	Auto.	Std.	Auto.	Std.	Auto.	Std.	Auto.	Std.
Number of phenotypic features	51	115	148	14	2	257	4	103	4	65	1	112	6	124	3	33
Molecular diagnosis	Early infantile epileptic encephalopathy 7	Glycogen storage disease V	Dopa-responsive dystonia	None	None	None	None	Permanent neonatal diabetes mellitus	None	None	None	None	X-linked agammaglobulinemia 1	Benign familial neonatal seizures 1		
Gene and causative variant(s)	KCNQ2 c.727C>G	PYGM c.2262delA c.1726C>T	TH c.785C>G c.541C>T	n.a.	n.a.	n.a.	n.a.	INS c.26C>G	n.a.	n.a.	n.a.	n.a.	BTX c.974+2 T>C	KCNQ2 c.1051C>G		
Total (hours)	20:25	19:56	19:20	19:14	20:42	56:03	19:29	48:46	19:11	42:04	19:10	57:21	31:02 <sup>†</sup>	34:38	22:04	38:37



## Выводы

- Менделевские (моногенные) заболевания зависят от генотипа в одном единственном локусе, наследование подчиняется законам Менделя
- При этом существует много отклонений от простого паттерна наследования
- Семейная агрегация, описательная эпидемиология и другие методы помогают установить генетические основы моногенных заболеваний. Поиск генов, ассоциированных с заболеваниями, значительно изменился и ускорился за счет высокопроизводительного секвенирования
- Важные паттерны наследования менделевских заболеваний: аутосомно-доминантный, аутосомно-рецессивный, Х-сцепленный рецессивный, Х-сцепленный доминантный, Y-сцепленное
- Пенетрантность, относительный риск и отношение шансов являются различными, но близкими мерами. Пенетрантность вариантов зачастую неизвестна или завышена
- Разработаны стандарты и руководства: от интерпретации вариантов до отчетов по вторичным находкам



## Список литературы

- Екатерина Померанцева — Генетическая диагностика и планирование семьи  
<https://youtu.be/TRLBiXgVqKg>
- Nykamp, K., Anderson, M., Powers, et al. (2017). Sherlock: a comprehensive refinement of the ACMG–AMP variant classification criteria. *Genet Med* 19, 1105–1117.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17, 405–424.
- Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., et al. (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet* 97, 199–215.
- Wright, C.F., West, B., Tuke, M., Jones, S.E., Patel, K., et al. (2019). Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet* 104, 275–286
- Blencowe, H., Moorthie, S., Petrou, M., et al. Rare single gene disorders: estimating baseline prevalence and outcomes worldwide. *J Community Genet.* 2018;9(4):397-406. doi:10.1007/s12687-018-0376-2
- Rabbani, B., Mahdih, N., Hosomichi, K., Nakaoka, H., and Inoue, I. (2012). Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* 57, 621–632

## Список литературы

- Bates, G.P. (2005). History of genetic disease: the molecular genetics of Huntington disease - a history. *Nat. Rev. Genet.* 6, 766–773.
- Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews Genetics* 14, 681–691.
- Alkuraya, F.S. (2016). Discovery of mutations for Mendelian disorders. *Hum Genet* 135, 615–623.
- Kalia, S.S., Adelman, K., Bale, S.J., Chung, W.K., et al. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* 19, 249–255.
- Shah, N., Hou, Y.-C.C., Yu, H.-C., Sainger, R., Caskey, C.T., Venter, J.C., and Telenti, A. (2018). Identification of Misclassified ClinVar Variants via Disease Population Prevalence. *Am J Hum Genet* 102, 609–619.
- Clark, M.M., Hildreth, A., Batalov, S., Ding, Y., Chowdhury, S., et al. (2019). Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Science Translational Medicine* 11, eaat6177
- Zschocke J, Byers PH, Wilkie AOM (2023) Mendelian inheritance revisited: dominance and recessiveness in medical genetics. *Nat Rev Genet* 1–22.