Медицинская геномика

Василий Евгеньевич Раменский
Анастасия Александровна Жарикова и Мария Ильинична Зайченока

ramensky@gmail.com, azharikova89@gmail.com

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

2024

Менделевские (моногенные) заболевания: поиск генов и диагностика. Часть 2

План лекции

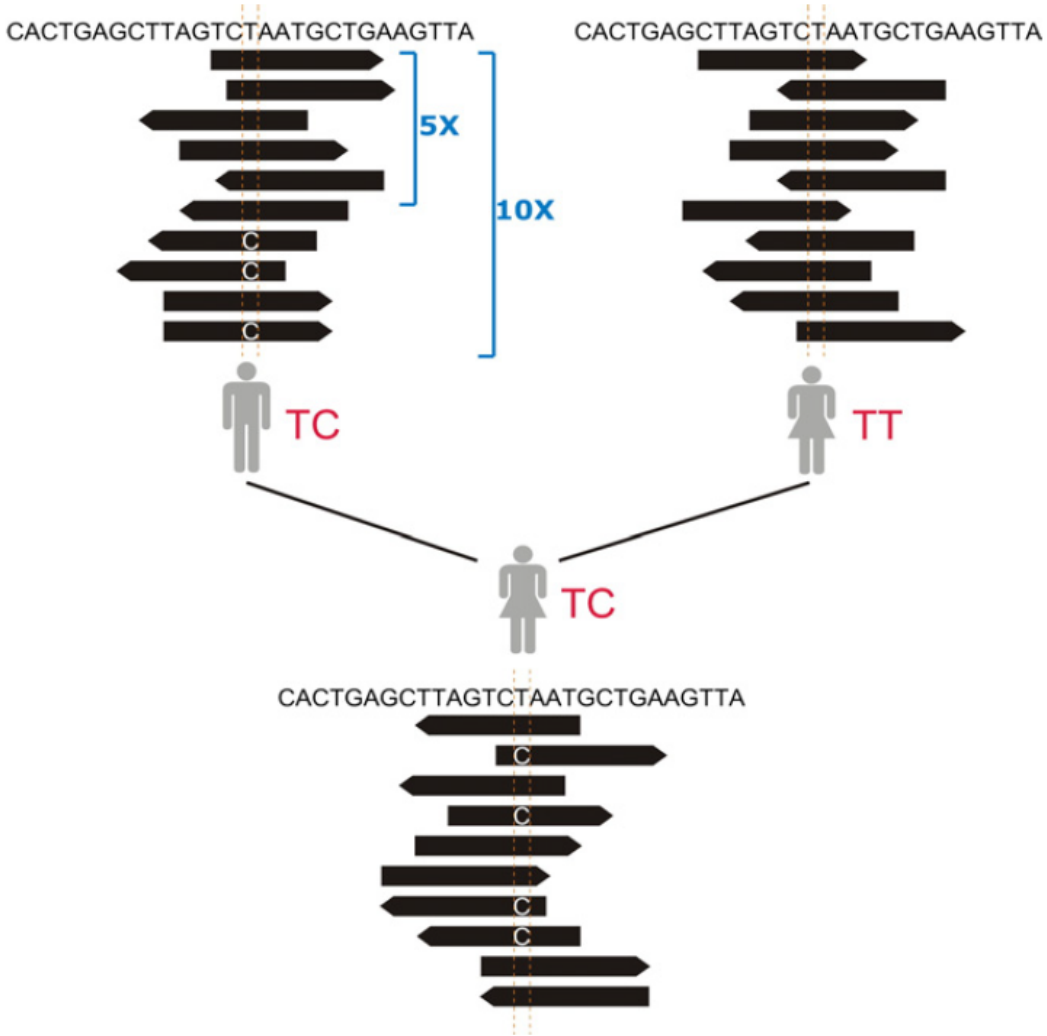
1. Поиск генов менделевских заболеваний с помощью NGS
2. Генетика моногенных заболеваний в пост-геномную эру
3. Клиническая значимость вариантов генома
4. WES-диагностика менделевских заболеваний

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

Напоминалка: NGS



Casals (2012) *J Neuroimmunology*

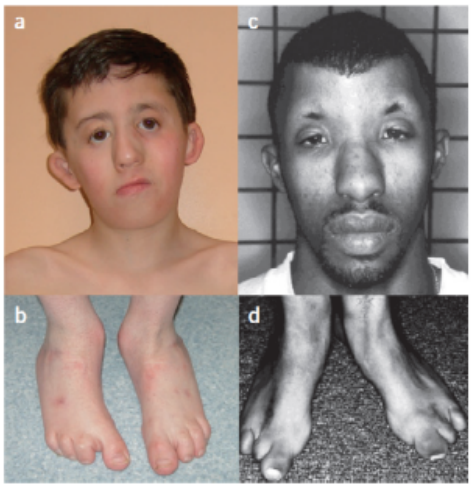
Первый успех: синдром Миллера

Exome sequencing identifies the cause of a mendelian disorder

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

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



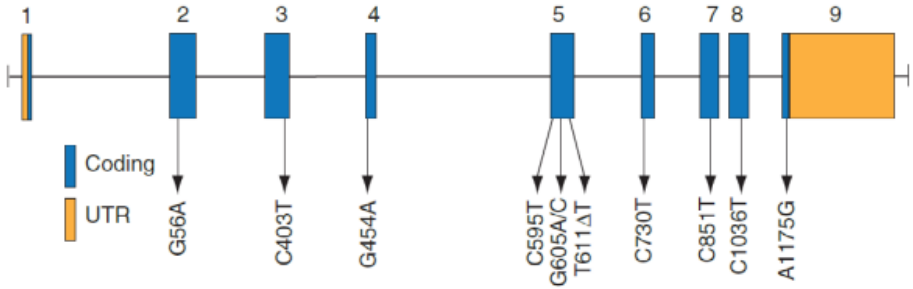


Figure 1. Clinical characteristics of an individual with Miller syndrome and an individual with methotrexate embryopathy.
Figure 2. Genomic structure of the exons encoding the open reading frame of *DHODH*. Arrows indicate the locations of 11 different mutations found in 6 families with Miller syndrome.

По следам первого успеха

Table 2 Mendelian disease-gene identifications by exome sequencing

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

Rabbani (2012) *J Hum Genet*

Приоритизация вариантов: проблема «иголки в стоге сена»

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

Приоритизация вариантов: проблема «иголки в стоге сена»

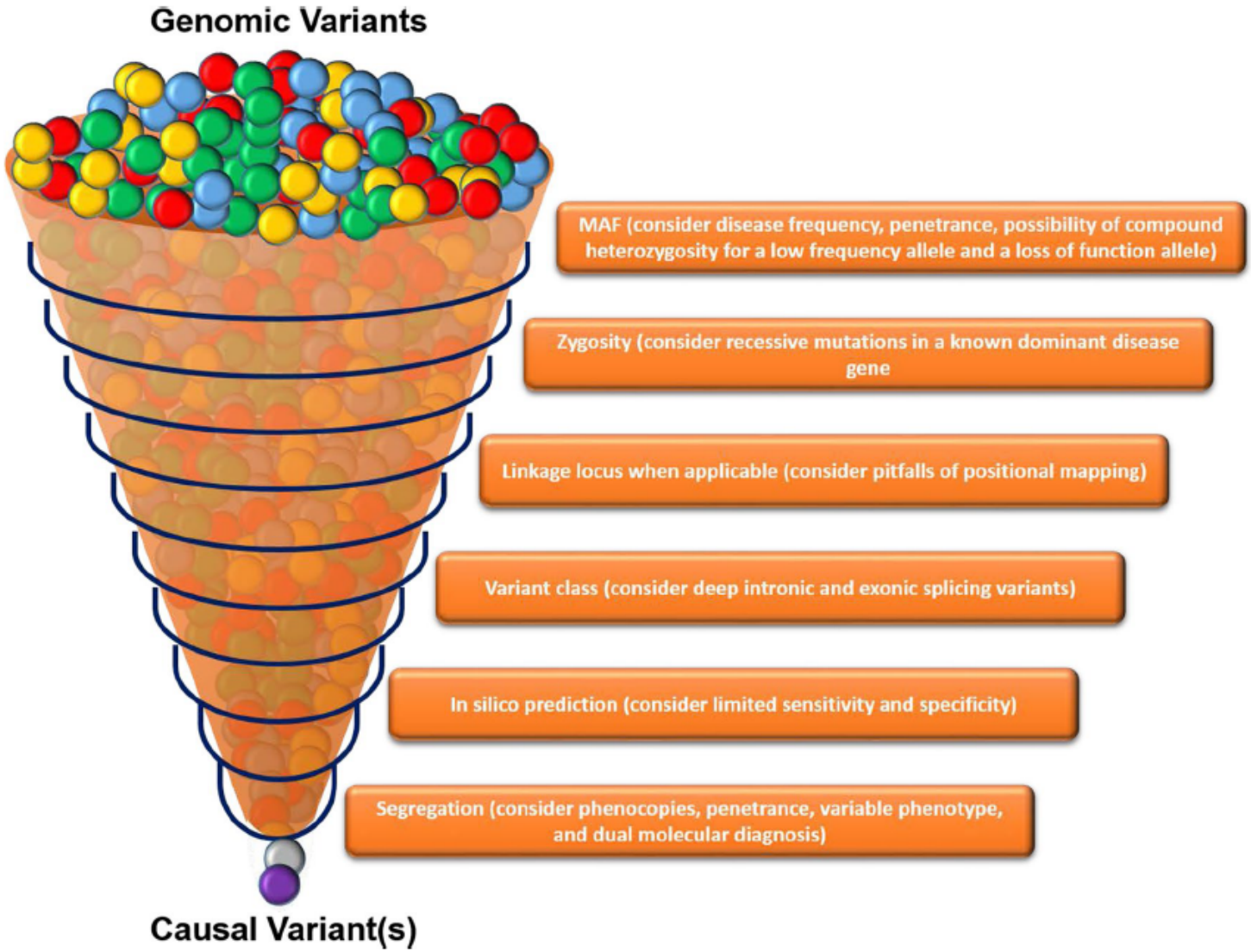
Необходимо определить, какой из выявленных вариантов наиболее вероятно [нарушает функцию гена и] является причиной болезни.

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

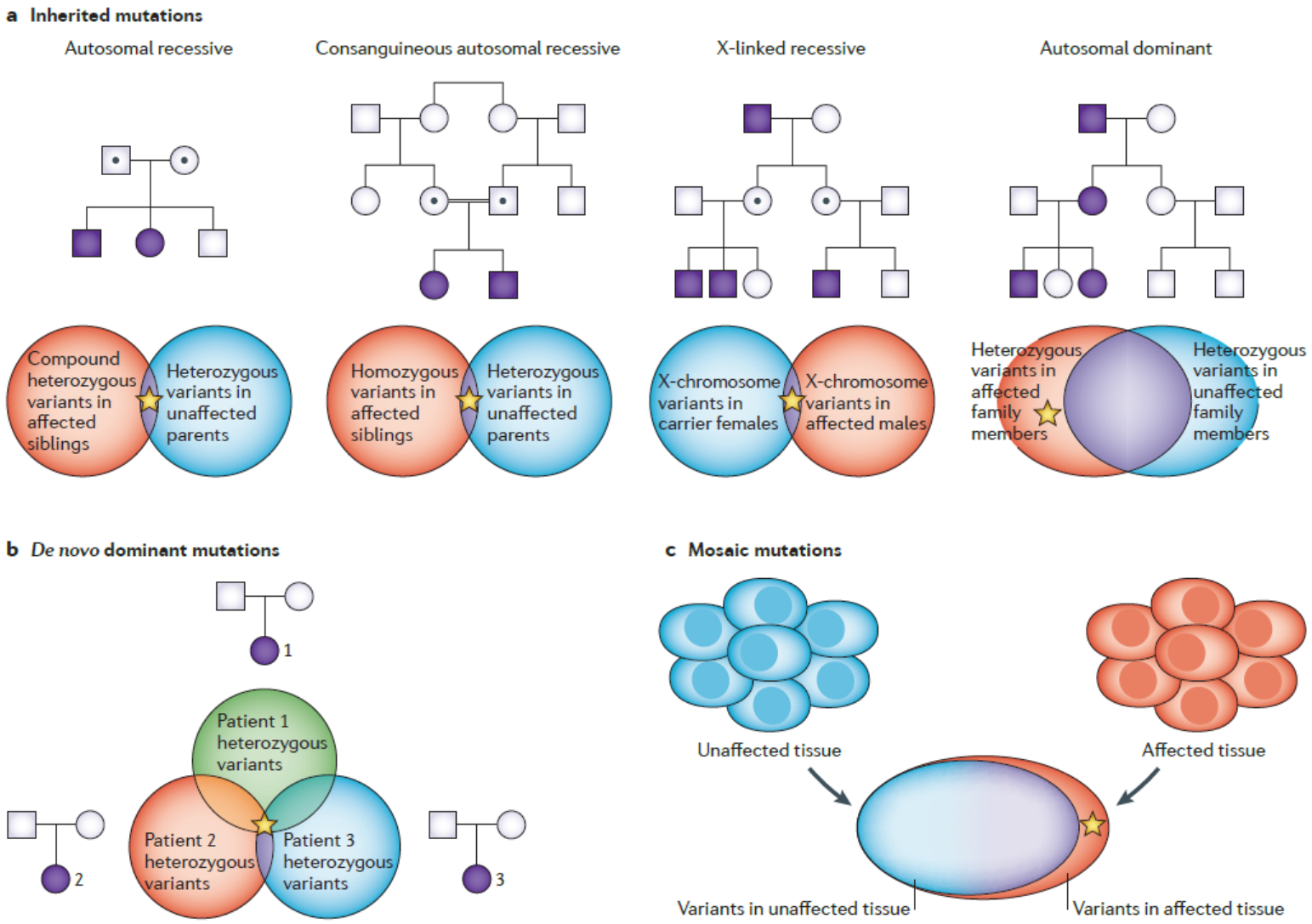
SNVs	Average	Deviation
Singleton	18	13
<0.01%	177	30
0.01-1%	273	23
1-10%	1308	72
>10%	12365	109
Indels		
<=5%	15	5
>5%	151	6

Variants in an individual ExAC exome (Lek (2016) *Nature*).

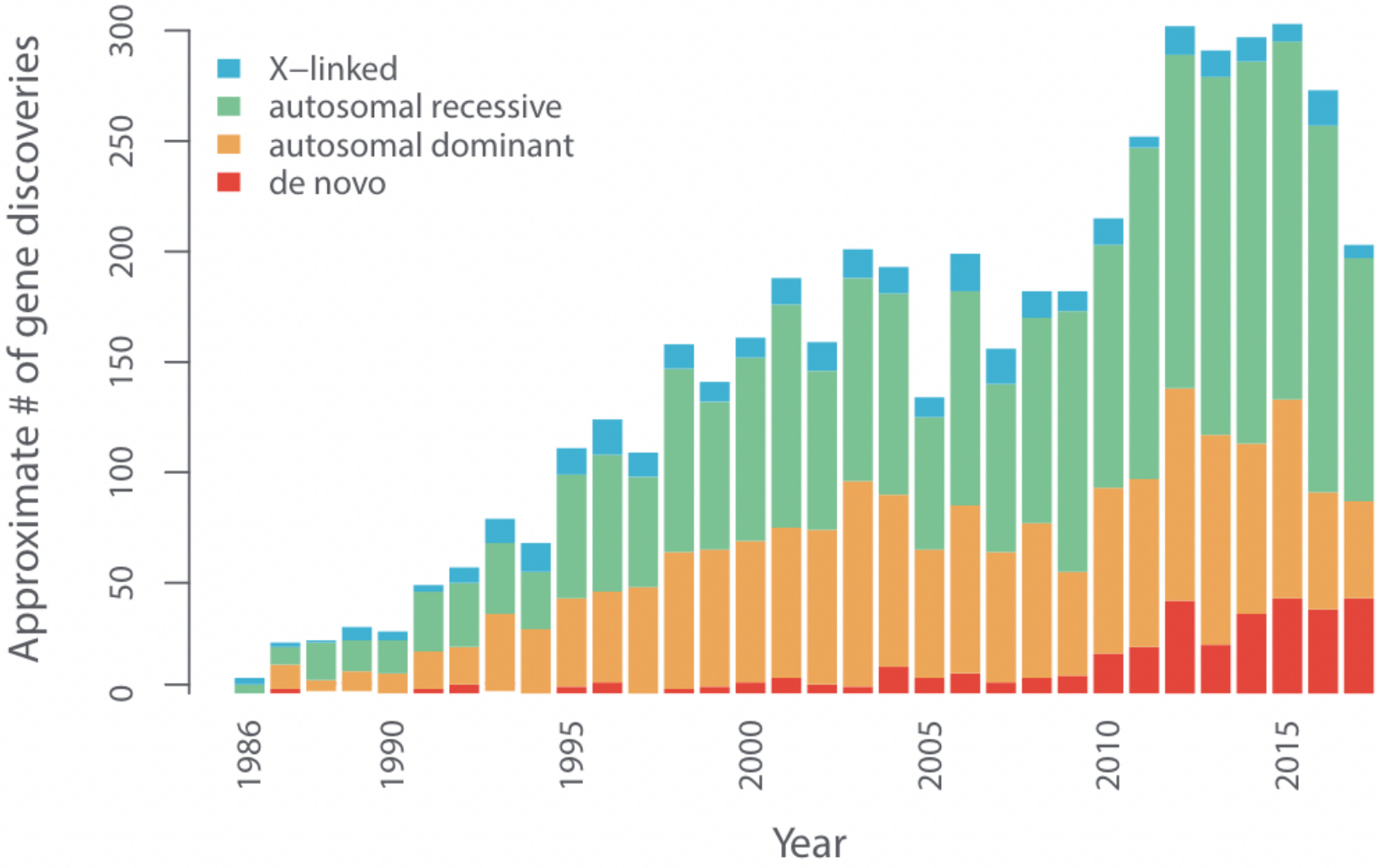
Стратегии поиска // Alkuraya (2016) *Hum Genet*



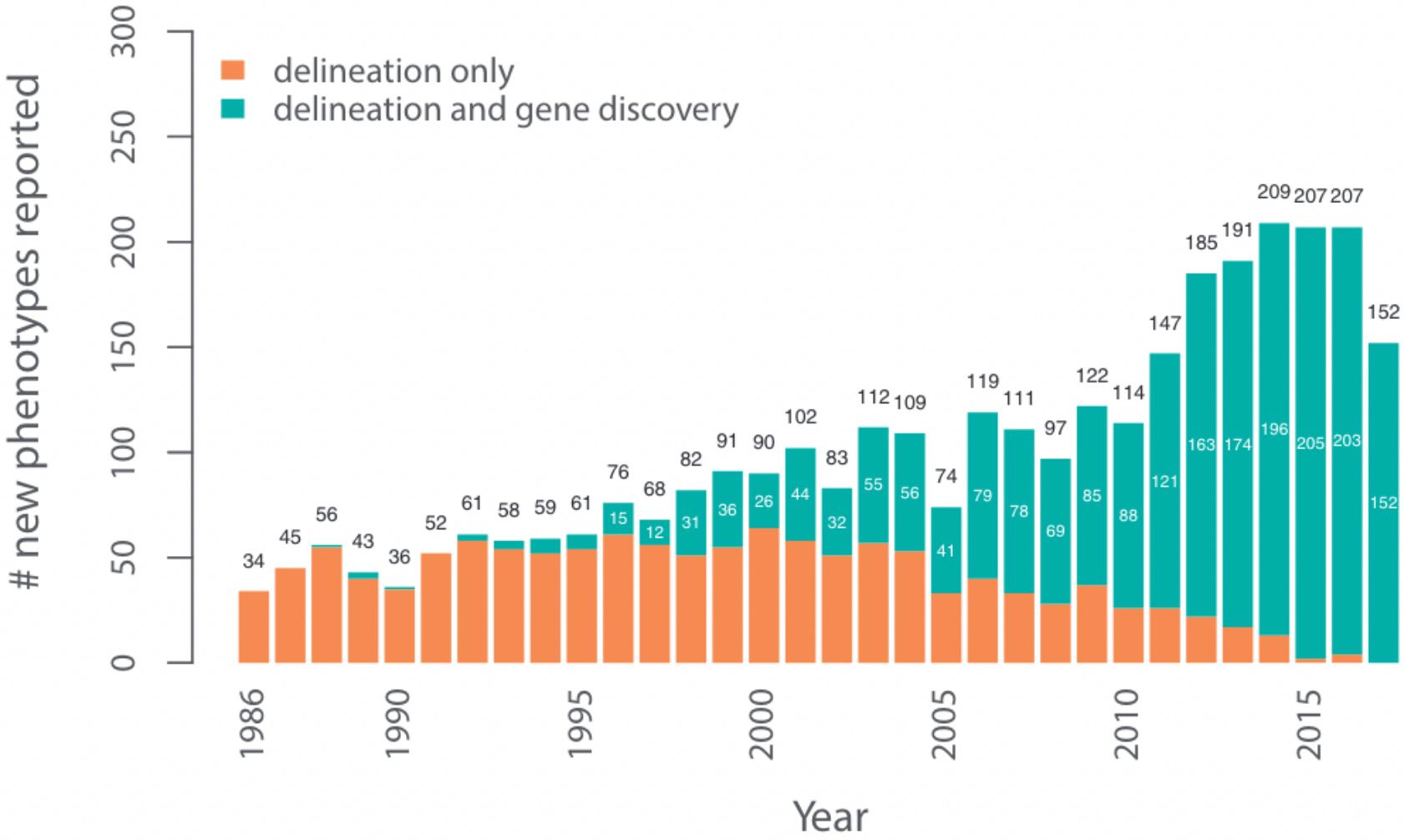
Стратегии поиска // Boycott (2013) *Nat Rev Genet*



Хронология поиска генов // Bamshad (2019) *Am J Hum Genet*



Хронология поиска генов // Bamshad (2019) *Am J Hum Genet*



Данные и биоинформатика

Reference datasets: increasing availability of population genetic variation catalogs (ExAC, gnomAD)

- The confident exclusion of common genetic variants too common
- Addressing the overestimation of disease penetrance arising from multiplex pedigrees
- Efforts to identify the genetic and environmental modifiers responsible

Data sharing

- A more systematic approach to information sharing (Matchmaker Exchange, DECIPHER and GeneMatcher, MyGene2)
- Databases of genes associated with rare disorders (for example, OMIM and ORPHANET)
- Databases of clinically interpreted variants (ClinVar and ClinGen)

Bioinformatics

- *In silico* analysis and prioritization of the discovered genetic variants

Claussnitzer (2020) *Nature*

Сложности и ограничения

Основные сложности:

- Редкие заболевания: малое количество случаев и/или семей
- Вариабельная 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 некодирующих 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

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 болезнетворного до тех, которые скорее всего безвредны.
Система классификации из 5 уровней для менделевских заболевания:

- 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 болезнетворный, 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.[1], Кардымон О.Л.[2], Прохорчук Е.Б.[3], Коновалов Ф.А.[4], Масленников А.Б.[5], Степанов В.А.[6], Афанасьев А.А.[7], Заклязьминская Е.В.[8], Костарева А.А.[9], Павлов А.Е.[10], Голубенко М.В.[6], Поляков А.В.[1], Куцев С.И.[1]

[1] 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 // 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	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product 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. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >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. 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.
Moderate	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

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	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product 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. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >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. 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.
Moderate	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

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 >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

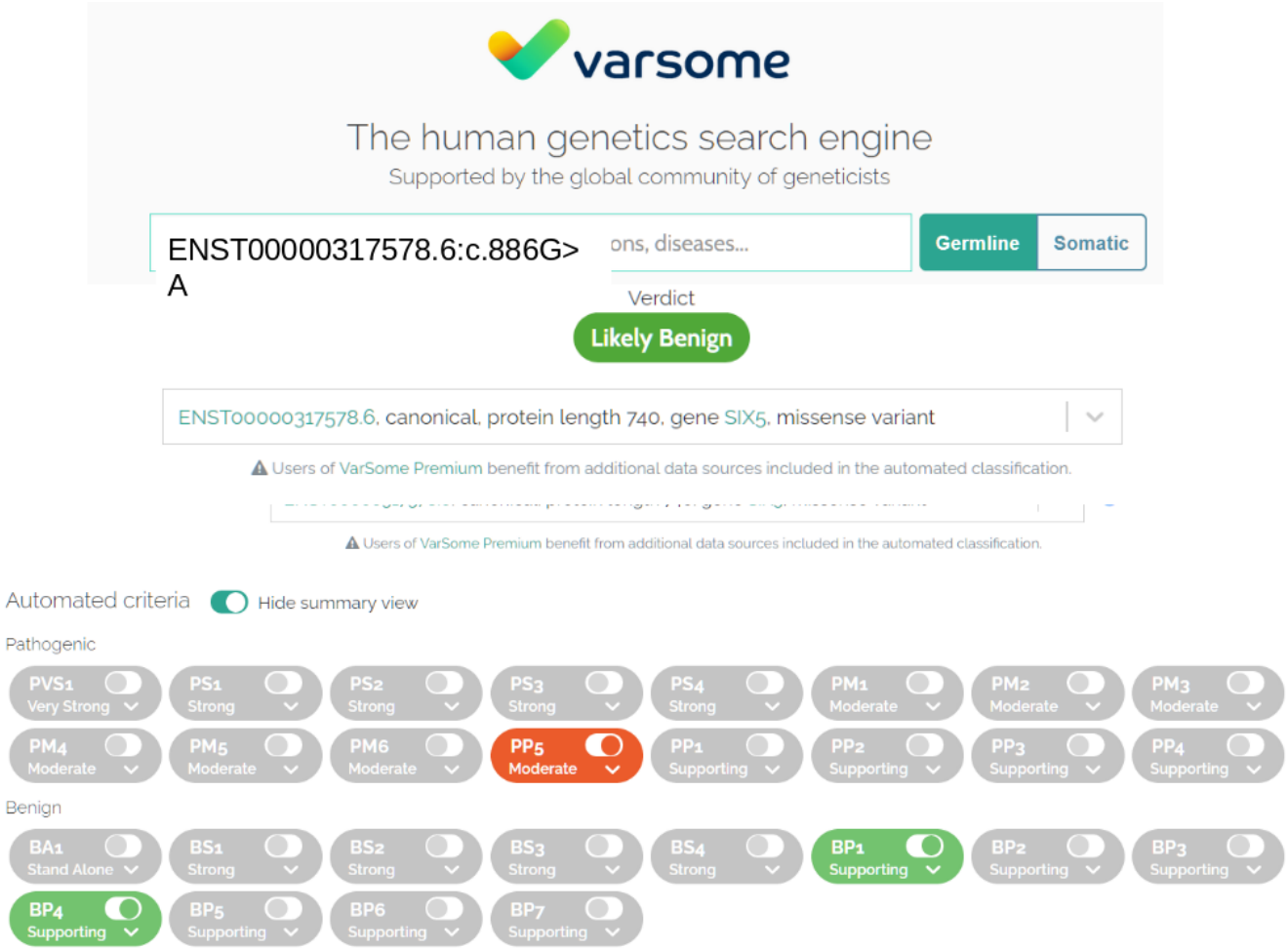
Table 5 Rules for combining criteria to classify sequence variants

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

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

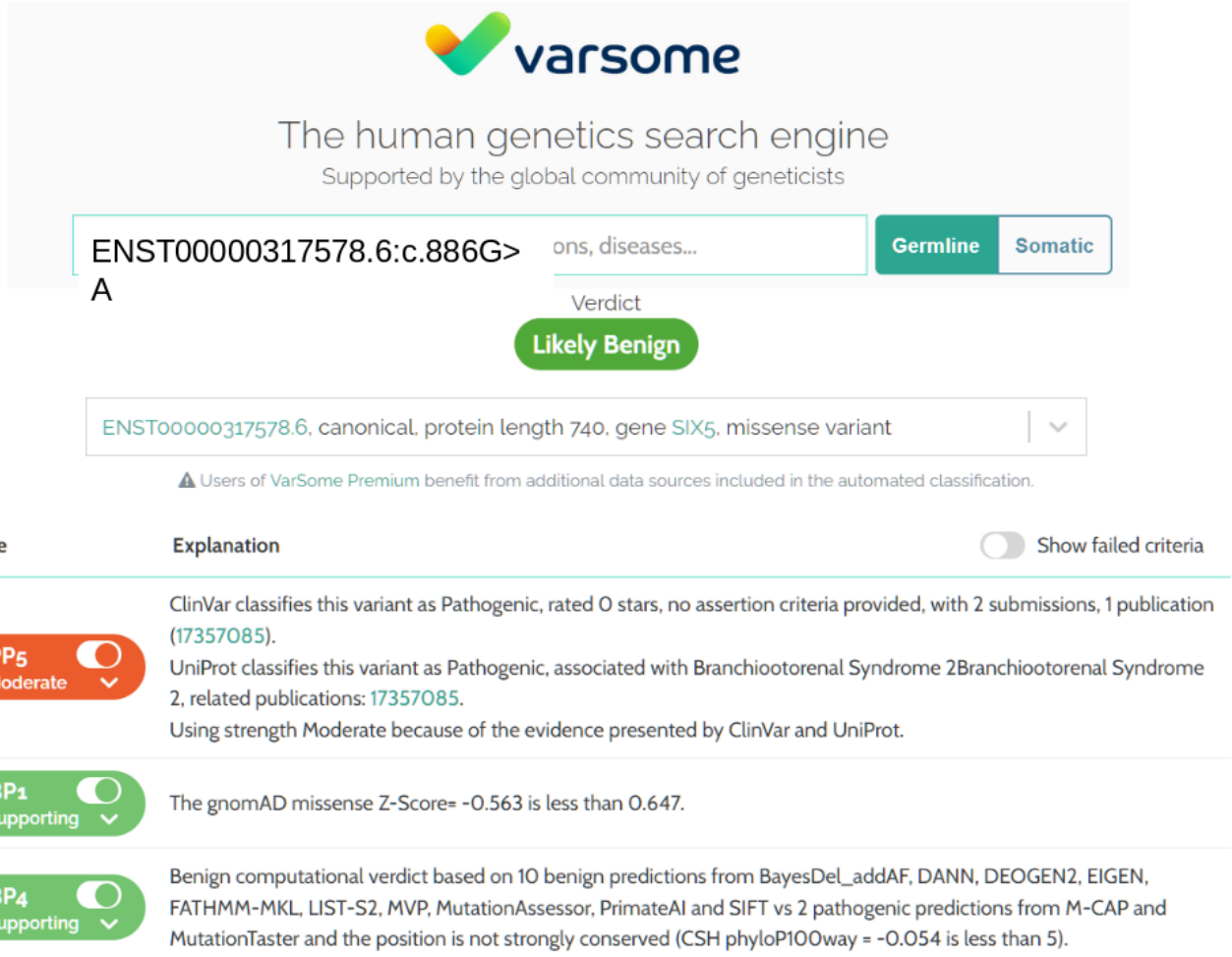
	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		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 BP3	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
Functional data	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	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		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		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Varsome



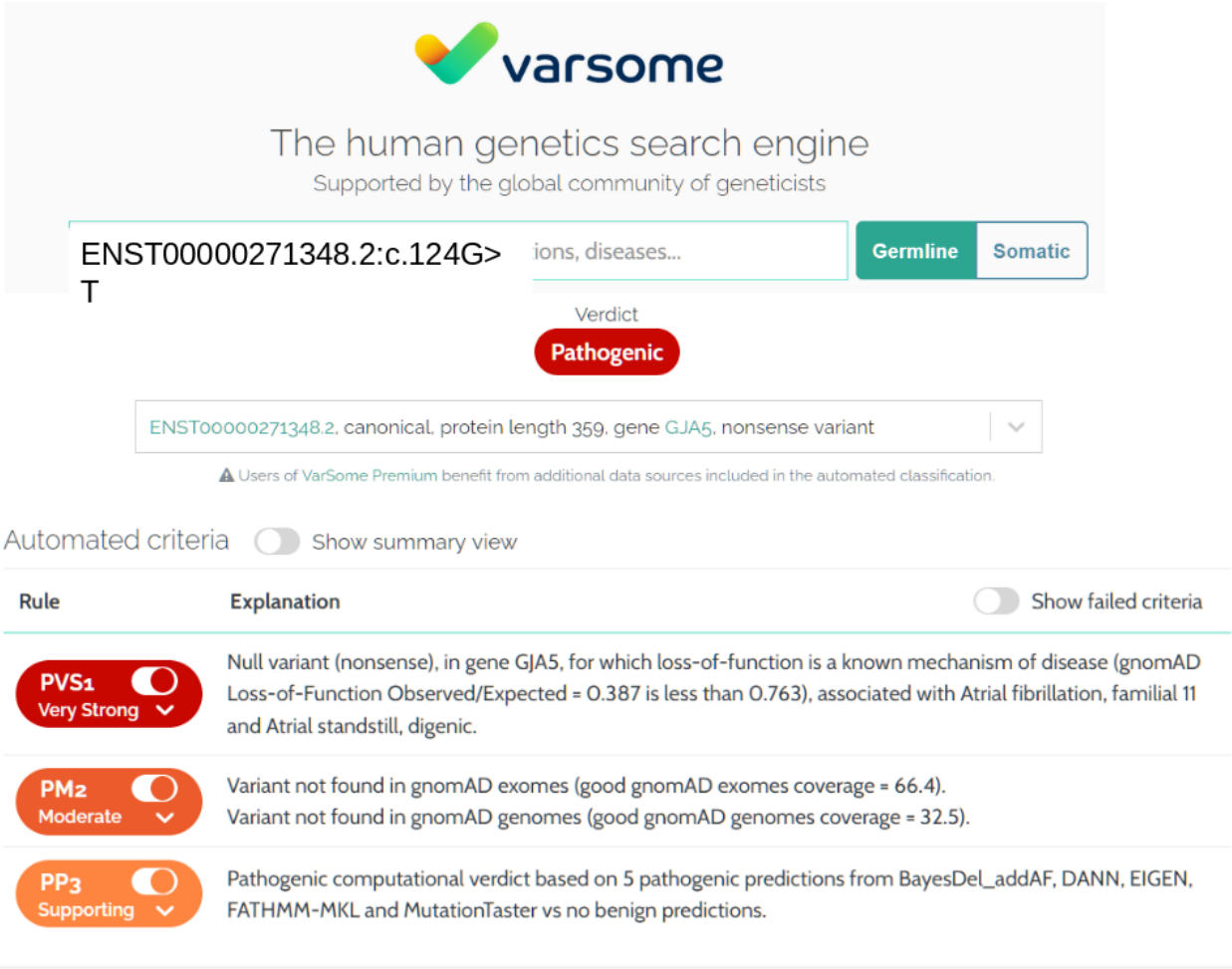
Kopanos (2019) *Bioinformatics*

Varsome



Kopanos (2019) *Bioinformatics*

Varsome



Kopanos (2019) *Bioinformatics*

Sherloc: уточнение критериев клин. 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 клин. классификации вариантов

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

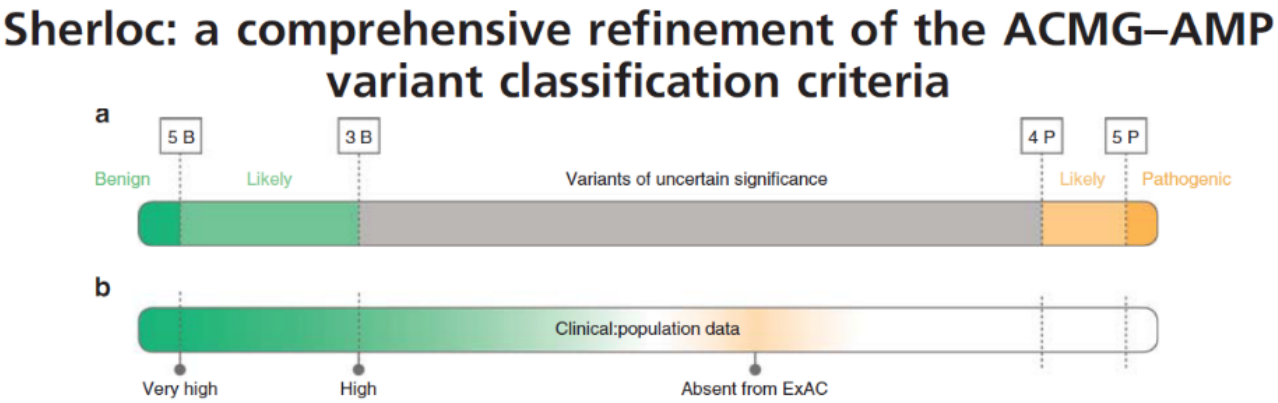
Although both clinical and functional evidence are relevant, they have a **hierarchical relationship**. Clinical data describe human disease directly, whereas functional data are relevant to disease only to the extent to which the measured property correlates with disease physiology. Therefore, when 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 клин. классификации вариантов

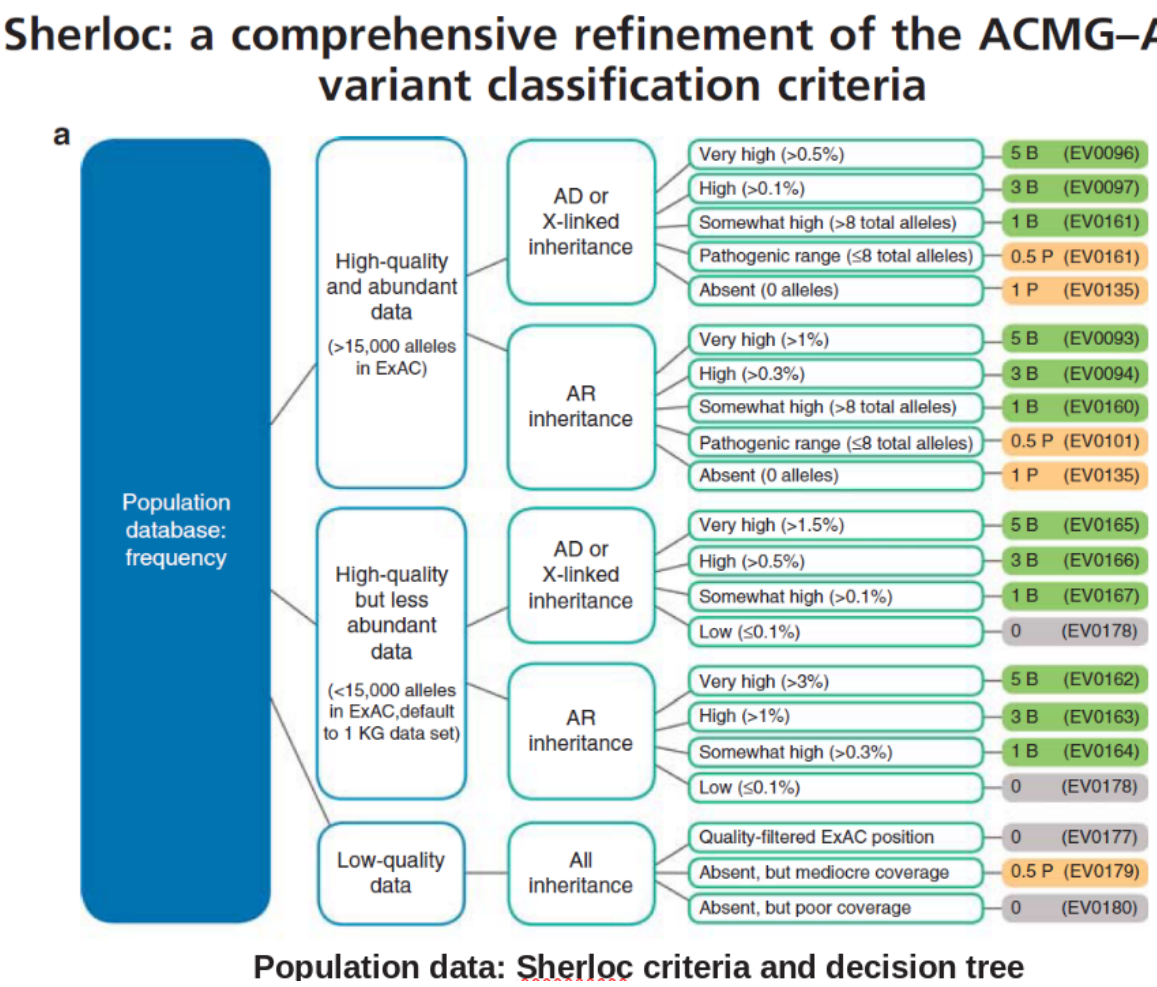
Nykamp (2017) *Genetics in Medicine*

Sherloc: уточнение критериев клин. классификации вариантов



Nykamp (2017) *Genetics in Medicine*

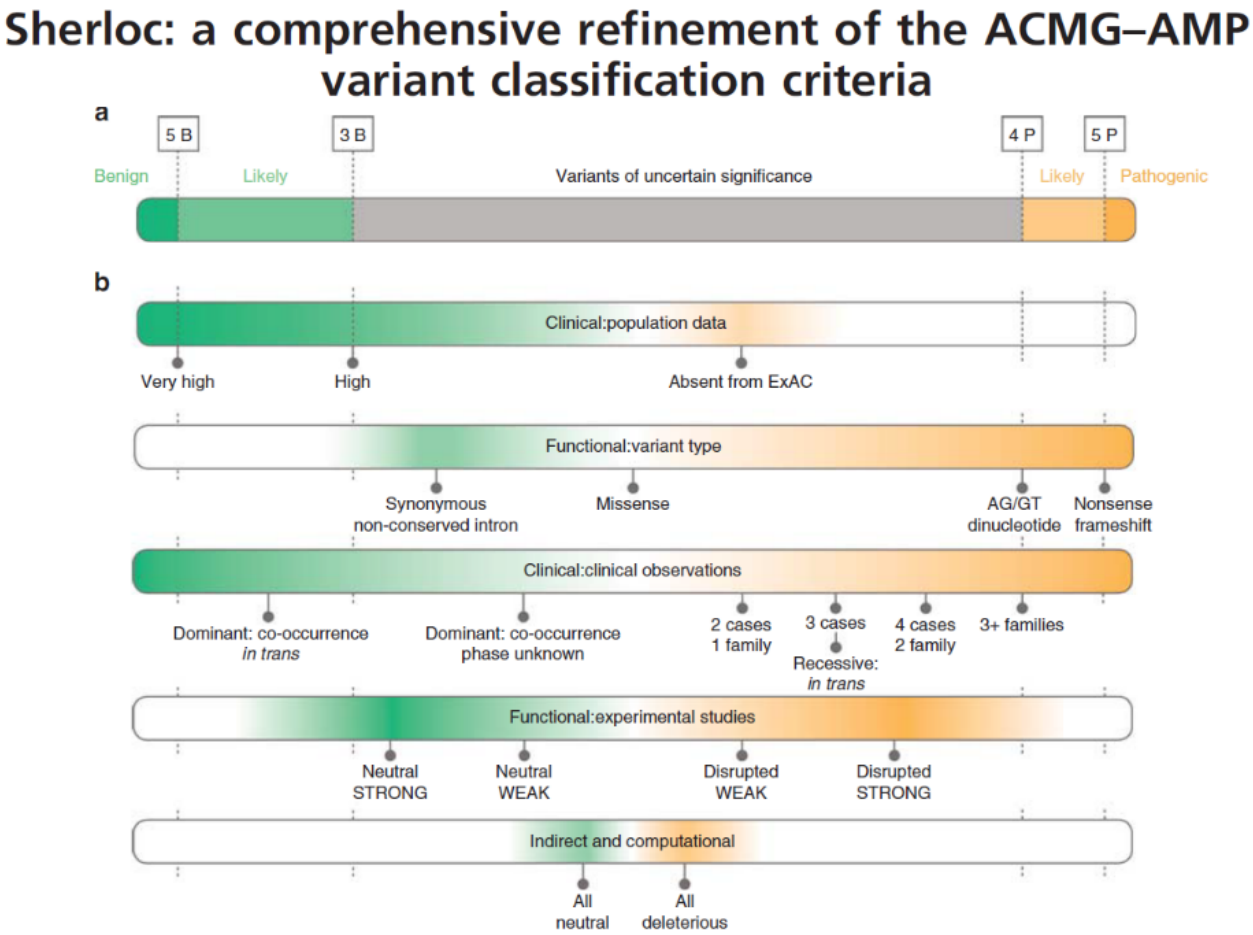
Sherloc: уточнение критериев клин. классификации вариантов



Population data: Sherloc criteria and decision tree

Nykamp (2017) *Genetics in Medicine*

Sherloc: уточнение критериев клин. классификации вариантов



Nykamp (2017) *Genetics in Medicine*

Sherloc: уточнение критериев клин. классификации вариантов

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

Пример 1: *TTC8* c.459G>A (p.Thr153=)

- Очень редкая тихая 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
- В нашей лаборатории наблюдается в гомозиготном состоянии у неродственных пациентах и теперь классифицируется как **патогенная**

Пример 2: *CDH1* c.1118C>T (p.Pro373Leu) вариант в гене, ассоциированным с наследственным диффузным раком желудка и дольным раком груди

- 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
- Компбютерные 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**

Sherloc: уточнение критериев клин. классификации вариантов

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

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 немкольких 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**

WES-диагностика: постановка задачи

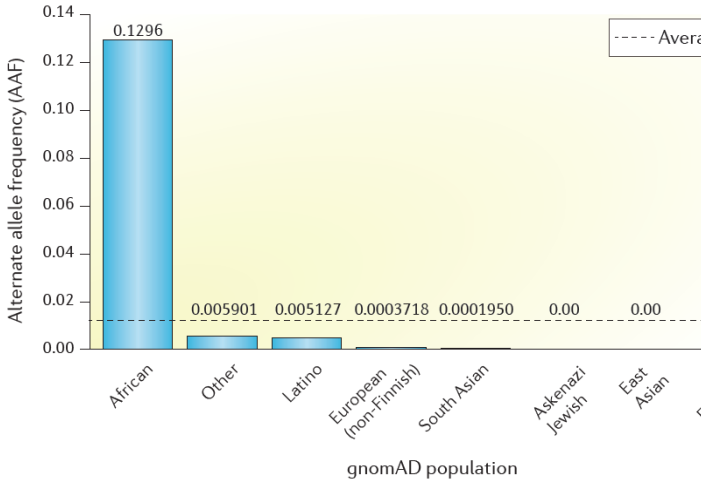
Применимость: атипичное проявление; симптомы, схожие у нескольких нарушений; сложность подтверждения клиническими или лабораторными критериями

Вход: Клинические симптомы (HPO), медицинские записи

Аннотация: VEP; gnomAD; ClinVar, OMIM

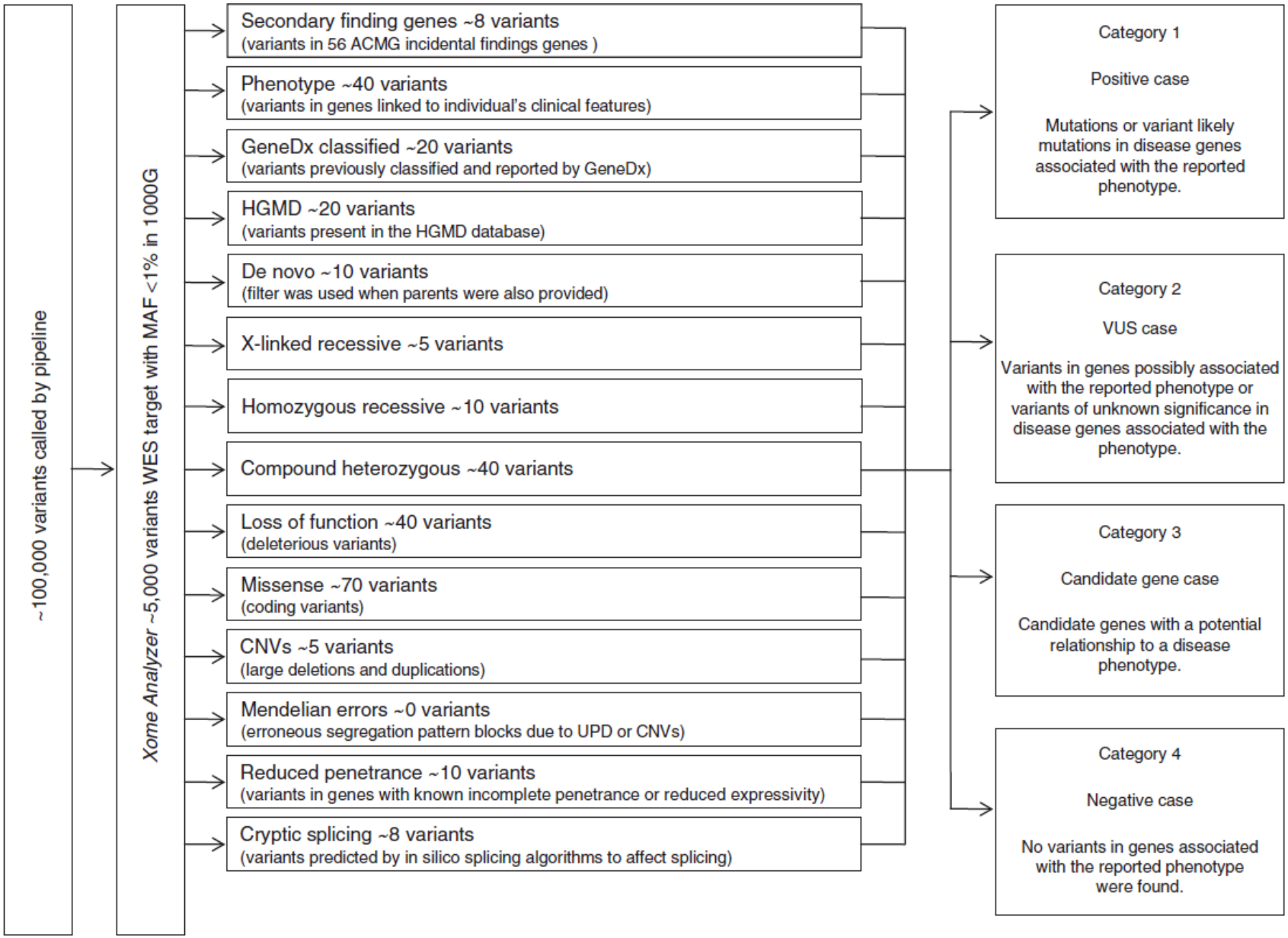
Фильтрация вариантов и приоритизация:
20,000-100,000 $\rightarrow$ 50 - 1,000
известные патогенные варианты; редкие (MAF $<$ 0.5%) или новые PTV; другие варианты, гены с ассоциированными фенотипами (ClinVar, OMIM, HPO)

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



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

WES-диагностика: схема // Retterer (2016) *Genet Me*



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

1. Новый гомозиготный вариант (Asp652Asn) в транспортере 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)**, но острый колит является нетипичным симптомом XLP2. Диагноз XLP2 предоставил специальный путь лечения.

Промежуточный обзор эффективности WES-диагностики

Clinical exome sequencing: results from 2819 samples reflecting 1000 families

Daniel Trujillano[*,1,10], Aida M Bertoli-Avella[1,10], Krishna Kumar Kandaswamy[1,10], Maximilian ER Weiss[1], Julia Köster[1], Anett Marais[1], Omid Paknia[1], Rolf Schröder[1], Jose Maria Garcia-Aznar[1], Martin Werber[1], Oliver Brandau[1], Maria Calvo del Castillo[1], Caterina Baldi[1], Karen Wessel[1], Shivendra Kishore[1]

We report our results of 1000 diagnostic WES cases based on 2819 sequenced samples from 54 countries with a wide phenotypic spectrum. Clinical information given by the requesting physicians was translated to HPO terms. WES processes were performed according to standardized settings. We identified the underlying pathogenic or likely pathogenic variants in 307 families (30.7%). In further 253 families (25.3%) a variant of unknown significance, possibly explaining the clinical symptoms of the index patient was identified. WES enabled timely diagnosing of genetic diseases, validation of causality of specific genetic disorders of ***PTPN23***, ***KCTD3***, ***SCN3A***, ***PPOX***, ***FRMPD4***, and ***SCN1B***, and setting dual diagnoses by detecting two causative variants in distinct genes in the same patient. We observed a better diagnostic yield in consanguineous families, in severe and in syndromic phenotypes. Our results suggest that WES has a better yield in patients that present with several symptoms, rather than an isolated abnormality. We also validate the clinical benefit of WES as an effective diagnostic tool, particularly in nonspecific or heterogeneous phenotypes. We recommend WES as a first-line diagnostic in all cases without a clear differential diagnosis, to facilitate personal medical care.

European Journal of Human Genetics (2017) **25**, 176–182; doi:10.1038/ejhg.2016.146; published online 16 November 2016

Clinical application of whole-exome sequencing across clinical indications

Kyle Retterer, MS[1], Jane Juusola, PhD[1], Megan T. Cho, ScM[1], Patrik Vitazka, MD, PhD[1], Francisca Millan, MD[1], Federica Gibellini, PhD[1], Annette Vertino-Bell, MS[1], Nizar Smaoui, MD[1,2]

Purpose: We report the diagnostic yield of whole-exome sequencing (WES) in 3,040 consecutive cases at a single clinical laboratory.

Methods: WES was performed for many different clinical indications and included the proband plus two or more family members in 76% of cases.

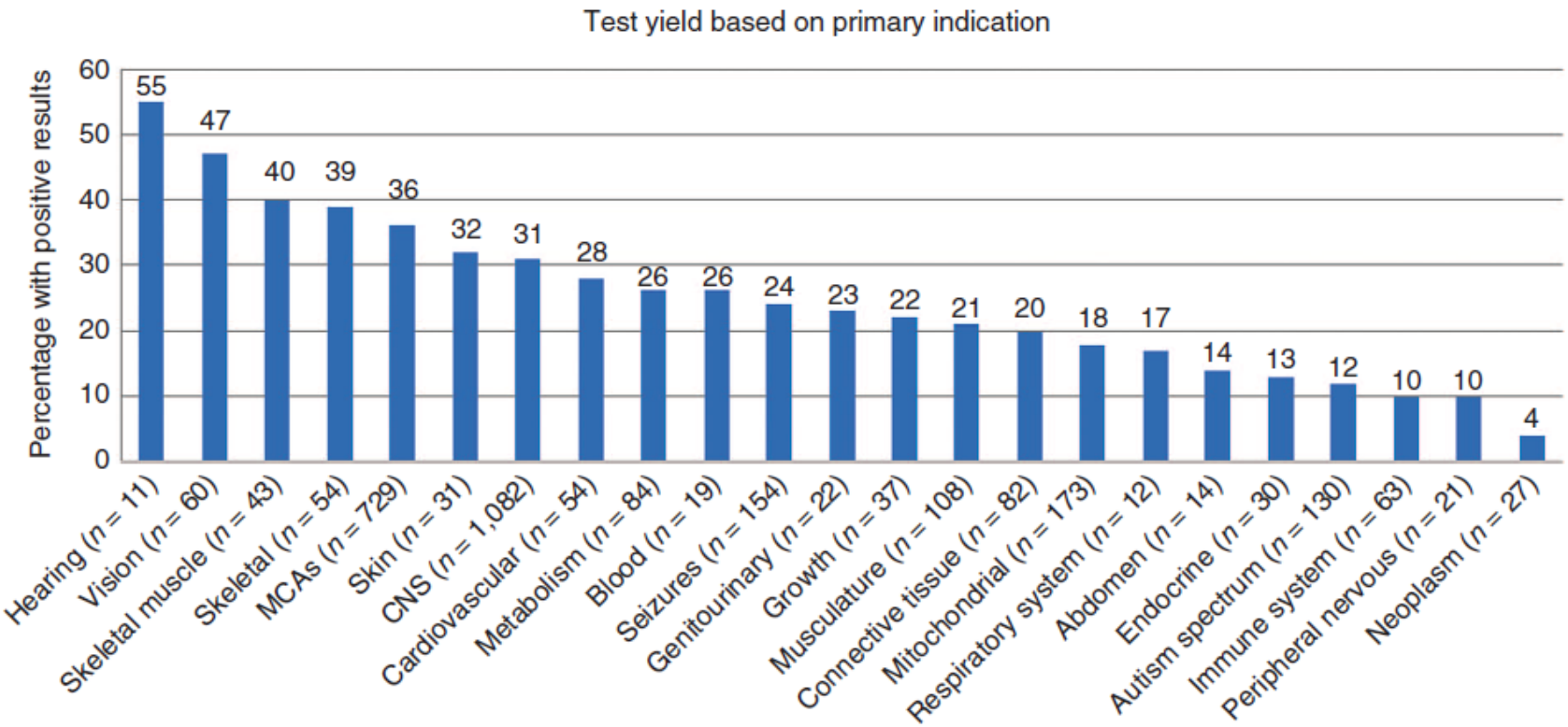
Results: The overall diagnostic yield of WES was 28.8%. The diagnostic yield was 23.6% in proband-only cases and 31.0% when three family members were analyzed. The highest yield was for patients who had disorders involving hearing (55%, $N = 11$), vision (47%, $N = 60$), the skeletal muscle system (40%, $N = 43$), the skeletal system (39%, $N = 54$), multiple congenital anomalies (36%, $N = 729$), skin (32%, $N = 31$), the central nervous system (31%, $N = 1,082$), and the cardiovascular system (28%, $N = 54$). Of 2,091 cases in which secondary findings were analyzed for 56 American College of Medical Genetics and Genomics–recommended genes, 6.2% ($N = 129$) had reportable pathogenic variants. In addition to cases with a definitive diagnosis, in 24.2% of cases a candidate gene was reported that may later be reclassified as being associated with a definitive diagnosis.

Conclusion: Our experience with our first 3,040 WES cases suggests that analysis of trios significantly improves the diagnostic yield compared with proband-only testing for genetically heterogeneous disorders and facilitates identification of novel candidate genes.

Genet Med advance online publication 3 December 2015

Промежуточный обзор эффективности WES-диагностики

- Общий диагностический выход WES составил 28.8% в **3,040** случаях; 23.6% в случаях анализа только пробанда и 31.0% в случаях анализа трио
- В 24.2% случаев были указаны гены-кандидаты, которые могут быть реклассифицированны как ассоциированные с определенным диагнозом
- Из 2,091 случая, в которых были проанализированы вторичные находки в 56 рекомендованных ACMG генах, 6.2% (129) случаях был найден патогенный вариант, выносимый в заключение



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 симтомов**

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[a]	Variants to report[b]
Hereditary breast and ovarian cancer	604370 612555	20301425	Adult	*BRCA1* *BRCA2*	113705 600185	AD	KP and EP
Li-Fraumeni syndrome	151623	20301488	Child/adult	*TP53*	191170	AD	KP and EP
Peutz-Jeghers syndrome	175200	20301443	Child/adult	*STK11*	602216	AD	KP and EP
Lynch syndrome	120435	20301390	Adult	*MLH1* *MSH2* *MSH6* *PMS2*	120436 609309 600678 600259	AD	KP and EP
Familial adenomatous polyposis	175100	20301519	Child/adult	*APC*	611731	AD	KP and EP
MYH-associated polyposis; adenomas, multiple colorectal, *FAP* type 2; colorectal adenomatous polyposis, autosomal recessive, with pilomatricomas	608456 132600	23035301	Adult	*MUTYH*	604933	AR[c]	KP and EP
Juvenile polyposis	174900	20301642	Child/adult	*BMPR1A* *SMAD4*	601299 600993	AD	KP and EP
Von Hippel–Lindau syndrome	193300	20301636	Child/adult	*VHL*	608537	AD	KP and EP
Multiple endocrine neoplasia type 1	131100	20301710	Child/adult	*MEN1*	613733	AD	KP and EP
Multiple endocrine neoplasia type 2	171400 162300	20301434	Child/adult	*RET*	164761	AD	KP
Familial medullary thyroid cancer[d]	1552401	20301434	Child/adult	*RET*	164761	AD	KP
PTEN hamartoma tumor syndrome	153480	20301661	Child/adult	*PTEN*	601728	AD	KP and EP
Retinoblastoma	180200	20301625	Child	*RB1*	614041	AD	KP and EP
Hereditary paraganglioma-pheochromocytoma syndrome	168000 (PGL1) 601650 (PGL2) 605373 (PGL3) 115310 (PGL4)	20301715	Child/adult	*SDHD* *SDHAF2* *SDHC* *SDHB*	602690 613019 602413 185470	AD	KP and EP KP KP and EP
Tuberous sclerosis complex	191100 613254	20301399	Child	*TSC1* *TSC2*	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,[1] Ying-Chen Claire Hou,[1] Hung-Chun Yu,[1] Rachana Sainger,[1] C. Thomas Caskey,[2] J. Craig Venter,[1,3,*] and Amalio Telenti[3,*]

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 × 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[1], Amber Hildreth[1,2,3], Sergey Batalov[1], Yan Ding[1], Shimul Chowdhury[1], Kelly Watkins[1], Katarzyna Ellsworth[1], Brandon Camp[1], Cyrielle L. Kint[4], Calum Yacoubian[5]

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		Rhabdo-myolysis	Dystonia, dev. delay	Hypoglycemia, seizures		Pulmonary hemorrhage, PPHN		Diabetic ketoacidosis		Neonatal seizures		HIE, anemia		Pseudomonal septic shock		Neonatal seizures	
Method	**Auto.**	**Auto.**	**Auto.**	**Auto.**	**Auto.**	**Std.**	**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	1
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 agamma-globulinemia 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.	*BTK* 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†**	**34:38**	**22:04**	**38:37**	**20:53**	**48:23**

Выводы

- Менделевские (моногенные) заболевания зависят от генотипа в одном единственном локусе, наследование подчиняется законам Менделя
- Однако существует много отклонений от простого паттерна наследования
- Семейная агрегация и описательная эпидемиология помогают установить генетические основы заболеваний
- Важные паттерны наследования менделевских заболеваний: аутосомно-доминантные, аутосомно-рецессивные, X-сцепленное рецессивное, X-сцепленное доминантное, Y-сцепленное
- Пенетрантность, относительный риск и отношение шансов являются различными, но близкими мерами. Пенетрантность вариантов зачастую неизвестна или завышена!
- Поиск генов, ассоциированных с заболеваниями, значительно изменился и ускорился за счет NGS
- Появляются стандарты и руководства в области: 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). Sherloc: 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., Mahdieh, 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.