Mendelian diseases: gene discovery and diagnostics

Lecture plan

- Disease, syndrome and other definitions
- Establishing the genetic basis of a diseases
- Mendelian diseases: overview and inheritance types
- Penetrance, relative risk, odds ratio

- Mapping disease genes in pre- and post-genome era
- Interpretation of sequence variants in monogenic disease context
- Mendelian disease gene discovery by NGS
- WES diagnostics of Mendelian disorders

The genetic basis of a disease

For **almost all human diseases**, individual susceptibility is, to some degree, influenced by genetic variation

-- Claussnitzer (2020) *Nature*

(1) Some of differences in DNA, alone or in combinations, might render an individual more **susceptible to one disorder** (for example, a type of cancer), but could render the same individual **less susceptible to develop an unrelated disorder** (for example, diabetes).
(2) The environment (including lifestyle) plays a significant role in many conditions (for example, diet and exercise in relation to diabetes), but our cellular and bodily **responses to the environment may differ according to our DNA**.
(3) **The genetics of the immune system**, with enormous variation across the population, determines our response to infection by pathogens.
(4) Most cancers result from an **accumulation of genetic changes that occur through the lifetime** of an individual, which may be influenced by environmental factors.

-- Jackson (2018) *Essays in Biochemistry*

Disease, syndrome and other definitions

Disease (disorder): a medical condition of the body which disrupts the normal functioning and physiological processes. A **genetic disorder** is caused by one or more abnormalities in the genome.

Inherited (hereditary): passed from parents to offspring
Sporadic: a condition that happens by chance (genetic or not)
Genetic: inherited or *de novo*
Congenital (*vs.* acquired): a condition that is present at birth
Phenocopy: a phenotypic variation that resembles the expression of a genotype but is caused by environmental conditions

A **syndrome** is a collection of symptoms which are often associated with a particular disorder.
For genetic cases, syndrome ≈ disorder.
Examples: CHARGE syndrome (*CHD7*), Down syndrome (trisomy 21), Tourette syndrome (unknown). Stockholm syndrome.

Disease, syndrome and other definitions

1. Mendelian (monogenic) disorders depend on the genotype at a single locus, with inheritance following Mendel's laws of segregation (Cystic fibrosis, Haemophilia A)
2. Complex (multifactorial) disorders: the outcome of a complex interplay of multiple genetic and environmental influences (Type II diabetes, coronary heart disease (ИБС) and schizophrenia)

Heritability: the relative contribution of genetic factors to a [disease] phenotype

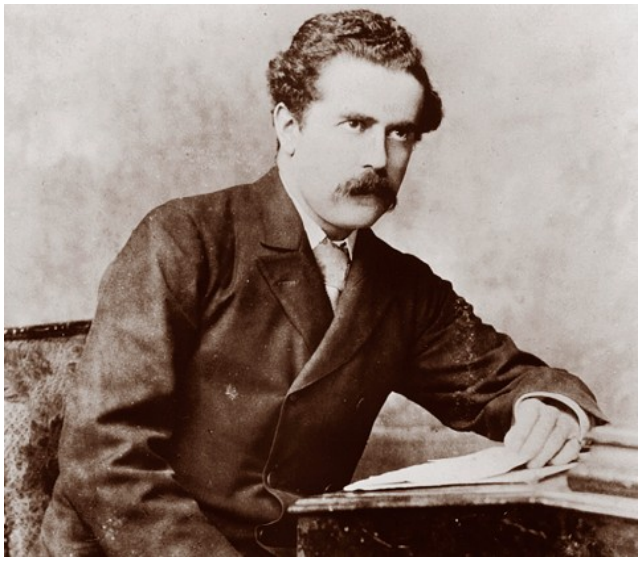
Disease, syndrome and other definitions

1. Mendelian (monogenic) disorders depend on the genotype at a single locus, with inheritance following Mendel's laws of segregation (Cystic fibrosis, Haemophilia A)
2. Complex (multifactorial) disorders: the outcome of a complex interplay of multiple genetic and environmental influences (Type II diabetes, coronary heart disease (ИБС) and schizophrenia)

Heritability: the relative contribution of genetic factors to a [disease] phenotype

3. Mitochondrial disorders result from mutations in mtDNA
4. Chromosomal disorders occur when entire chromosomes or parts of chromosomes are missing or changed.
5. Epigenetic disorders are disorders related to changes in the activity of genes, rather than a mutation in the structure of the DNA

Alkaptonuria: inborn errors of metabolism



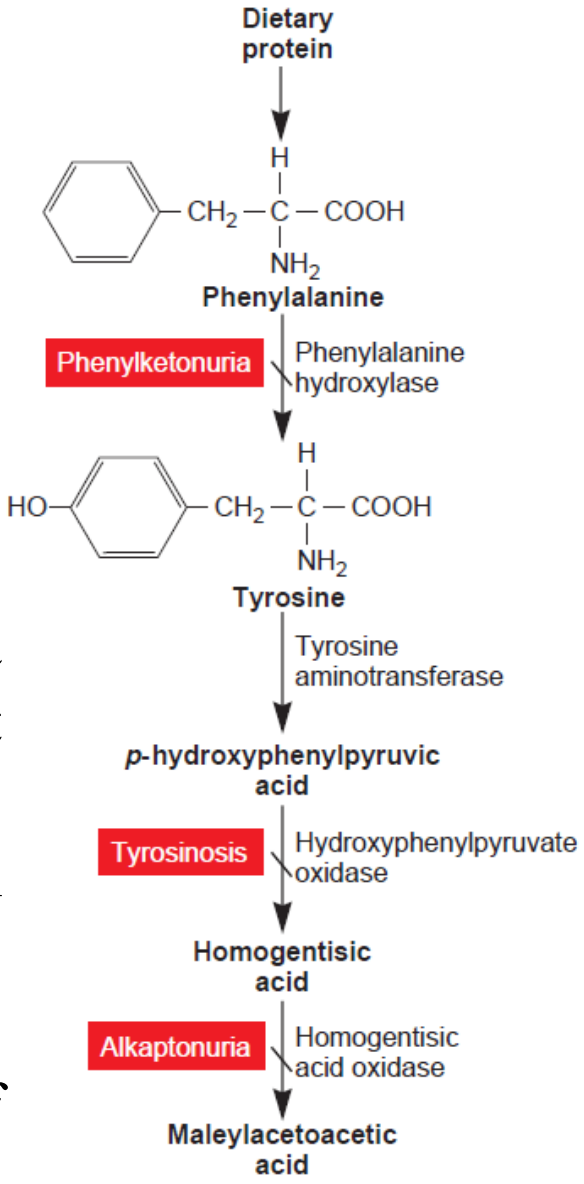
Reprinted from Lancet, vol. ii, 1902, pp. 1616-1620.

THE INCIDENCE OF ALKAPTONURIA:
A STUDY IN CHEMICAL INDIVIDUALITY

ARCHIBALD E. GARROD

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

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



Abnormal levels of homogentisic acid (aka *alkapton*), which is excreted in the urine, causing it to appear black on exposure to air

Alkaptonuria (AKU) is inherited and follows an autosomal recessive pattern.

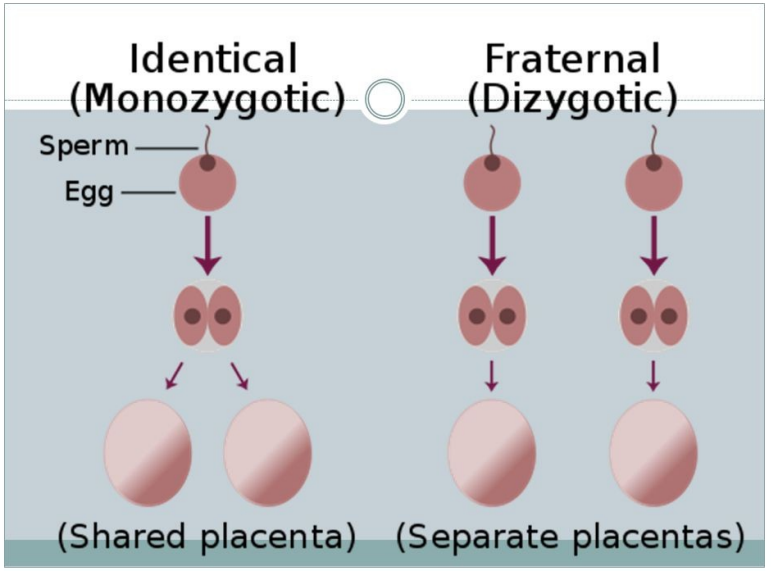
Sir Archibald Garrod (1902): mutation → loss of enzyme → inborn error of metabolism

Brooker – *Genetics, Analysis and Principles*

Establishing the genetic basis of a disease

Monozygotic twins (MZ) develop from one zygote, which splits and forms two embryos.
Dizygotic twins (DZ) develop from separate eggs, each egg is fertilized by its own sperm cell

	Concordance	
Disease type	***MZ***	***DZ***
Monogenic	100%	50%
Complex	70%	25%
Non-genetic	X%	X%



Establishing the genetic basis of a disease

Familial aggregation: does a disease run in families more often than would be expected by chance? Relatives share gene variants, but also share environment (diet, upbringing)

- Segregation patterns (type of inheritance)
- Twin studies (also separated monozygotis twins)
- Adoption studies: affected parents or affected offspring

- **Descriptive [genetic] epidemiology**: international variation in disease risks; migrant studies; admixture studies

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.]

Variant effect: gain- and loss-of-function

Loss-of-function: the product has reduced or no function

Examples: transcription factors; disruption of catalytic function in an enzyme

- Protein-truncating and missense variants?
- Recessive, but in some cases (haploinsufficiency) also dominant

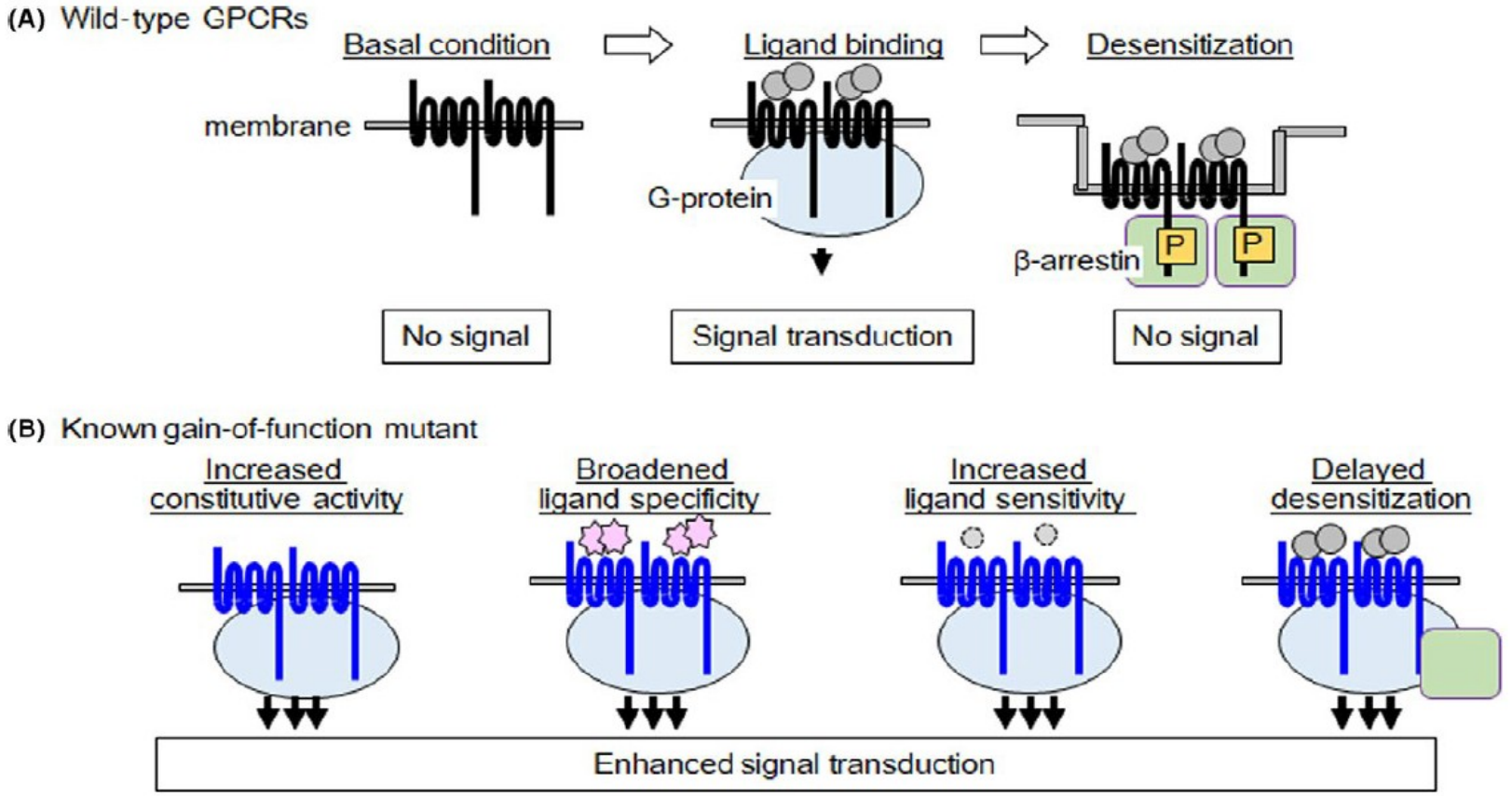
Gain-of-function: the product does "something positively abnormal"

Examples: transcription factors; gain-of-function mutations in G-protein–coupled receptors (GPCRs)

- Mostly missense variants, but also frameshift, inframe deletions
- Presence of a normal allele cannot prevent the mutant allele from behaving abnormally $\Longrightarrow$ dominant?

Variant effect: gain- and loss-of-function

- **G-protein–coupled receptors** are sensors for internal stimuli: hormones, ions and chemokines; light, odour and taste. GPCRs play particularly important roles in the endocrine system.
- Human genome contains >700 GPCRs
- Implicated in various human disorders, including endocrine diseases

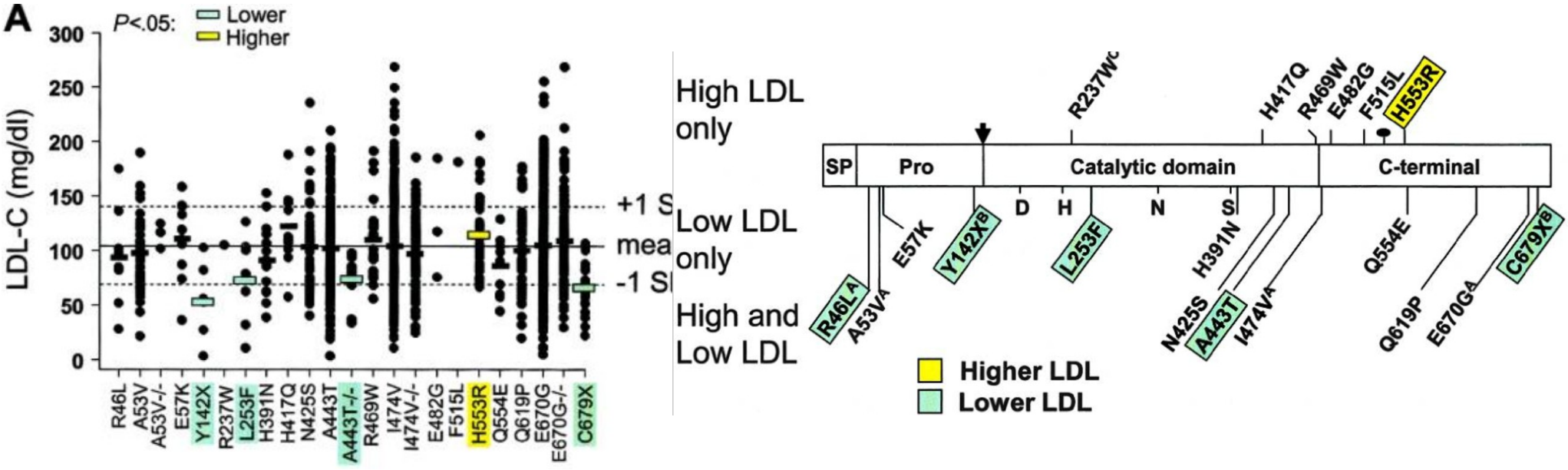


Fukami (2018) *Clin Endocrinol*

Variant effect: gain- and loss-of-function

Serine protease *PCSK9* (Proprotein convertase subtilisin/kexin type 9) regulates low density lipoprotein cholesterol (LDL-C) levels, has both types of variants

High LDL-C level ⟹ atherosclerosis ⟹ cardiac infarction or stroke

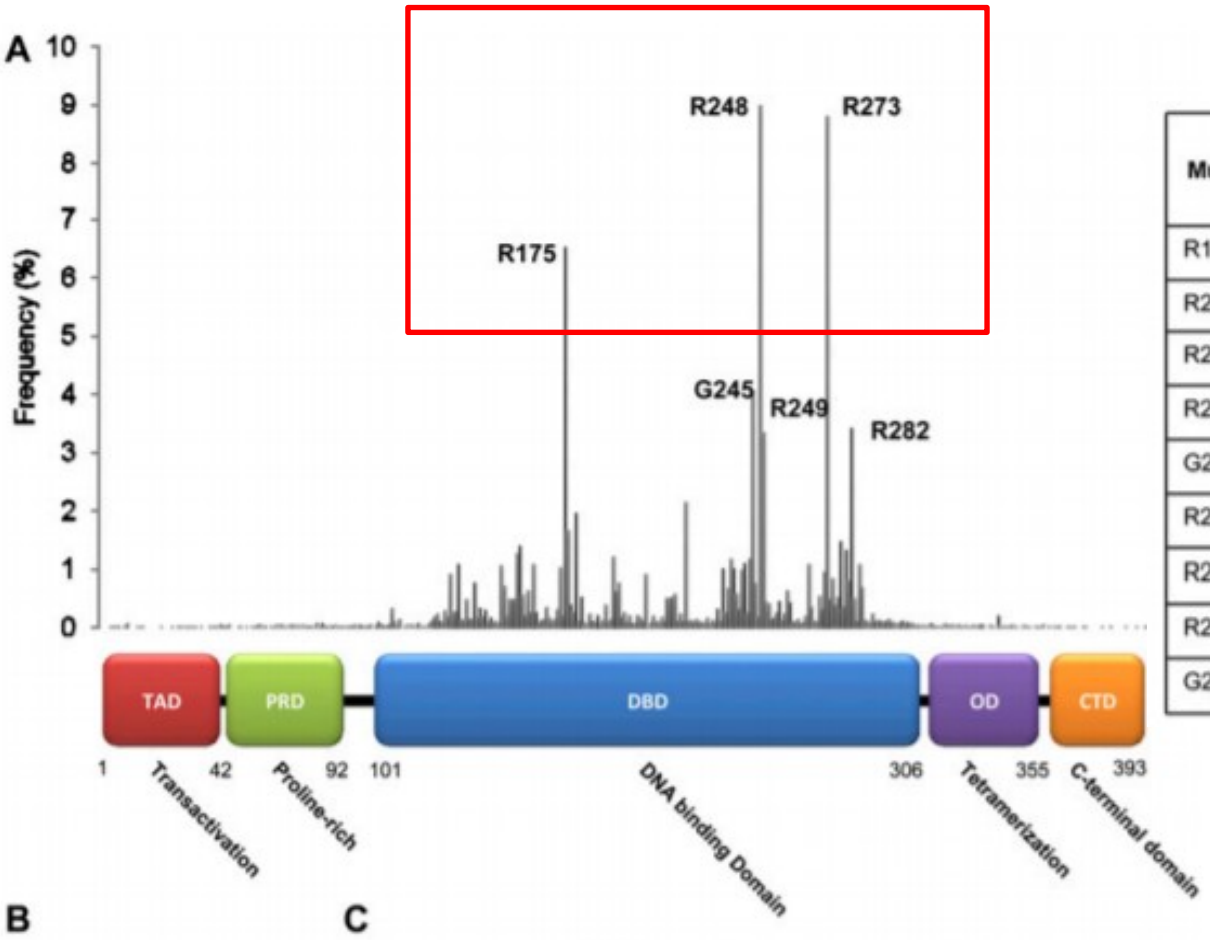


Kotowski (2006) *Am J Hum Genet*

Variant effect: gain- and loss-of-function

TP53 mutational spectrum in human cancers

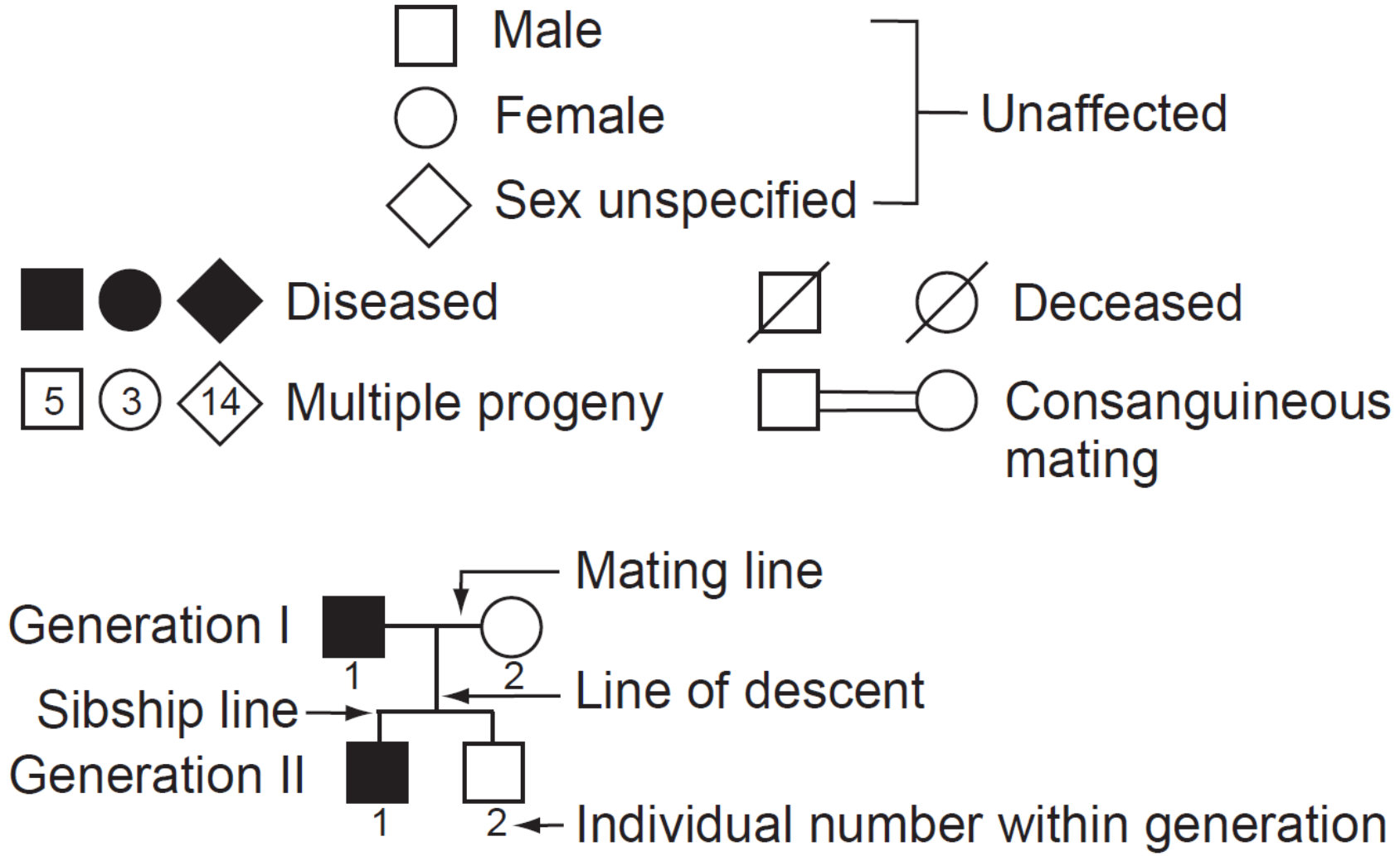
Gain-of-function



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*

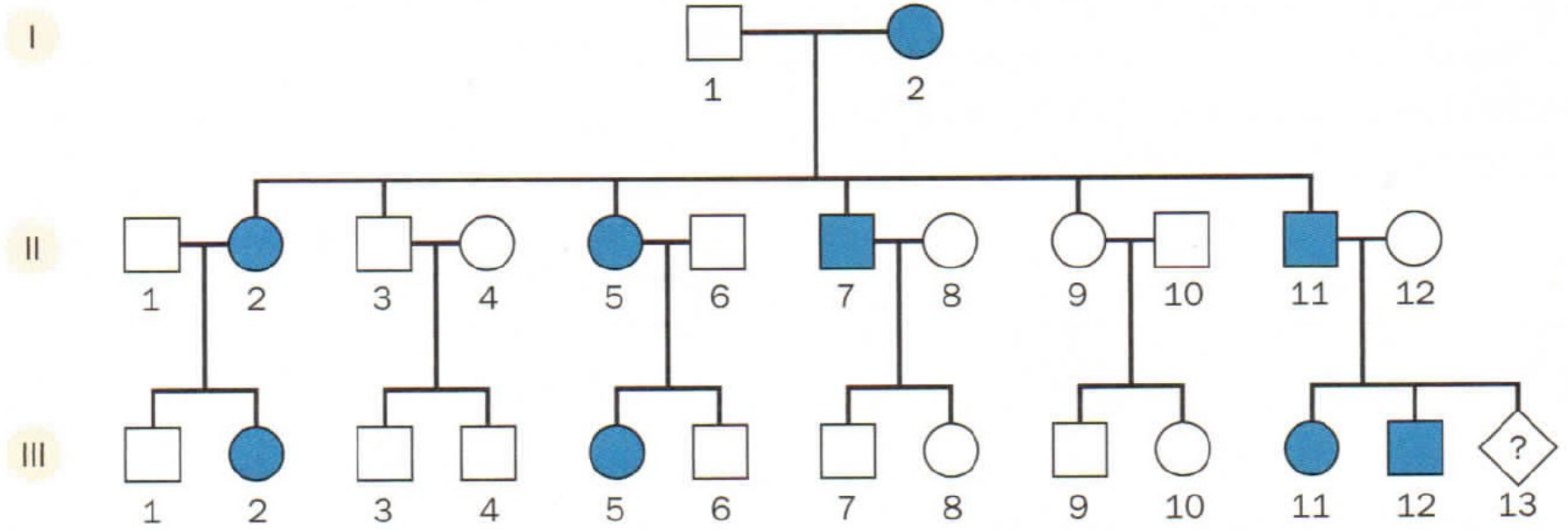
Symbols used in pedigree analysis



Hartwell – *Genetics. From genes to genomes*

Autosomal dominant inheritance

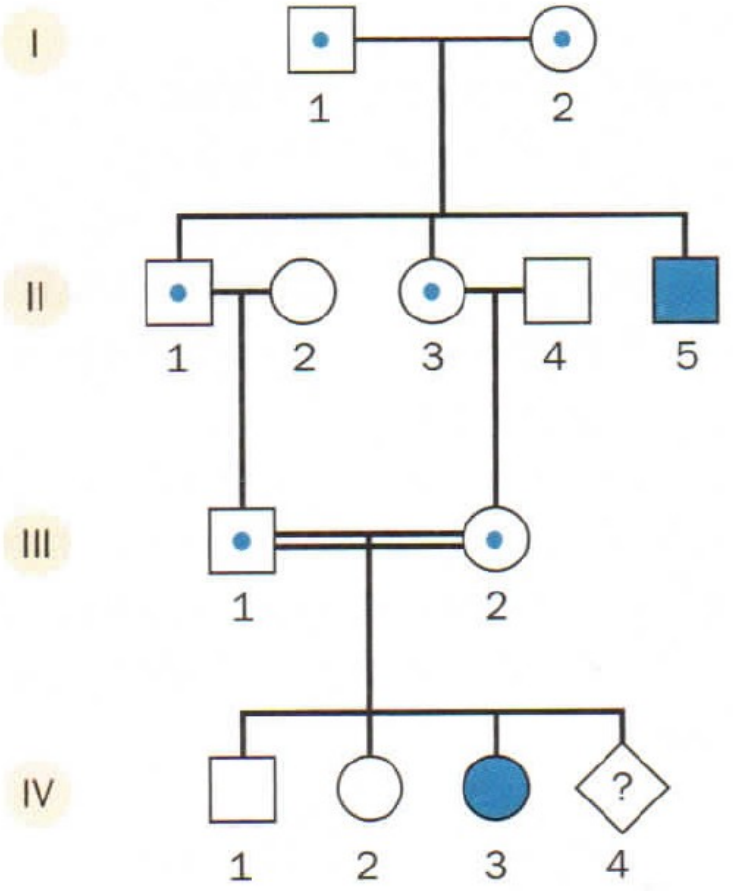
- An affected person (proband) usually has at least one affected parent
- It affects either sex
- A child with one affected and one unaffected parent has a 50% chance of being affected
- Causal variant is gain-of-function or loss-of-function if gene is haploinsufficient; often, *de novo*



Strachan, Read – *Human Molecular Genetics*

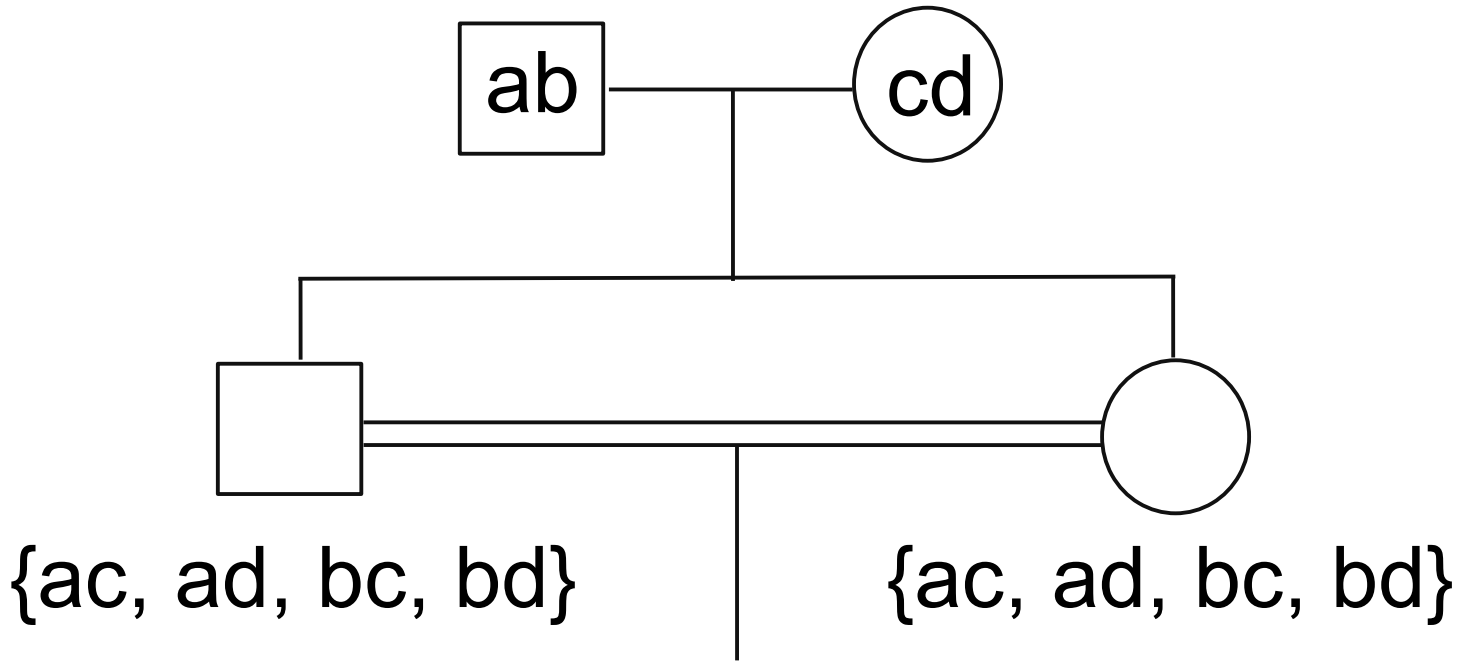
Autosomal recessive inheritance

- Affected people are usually born to unaffected parents, who are usually asymptomatic carriers
- It affects either sex
- A child has a 25% chance of being affected
- Causal variant is loss-of-function
- There is an increased incidence of parental consanguinity



Strachan, Read – *Human Molecular Genetics*

Consanguinity and homozygosity



{…, ba, bd, da, dd, … }

Exercise: list all possible genotypes for the consanguineous offspring and calculate probability of homozygosity, aka the inbreeding coefficient *F*

Consanguinity and homozygosity

Regions of homozygosity (ROH): genome segments showing continuous homozygosity (with no intervening heterozygosity)

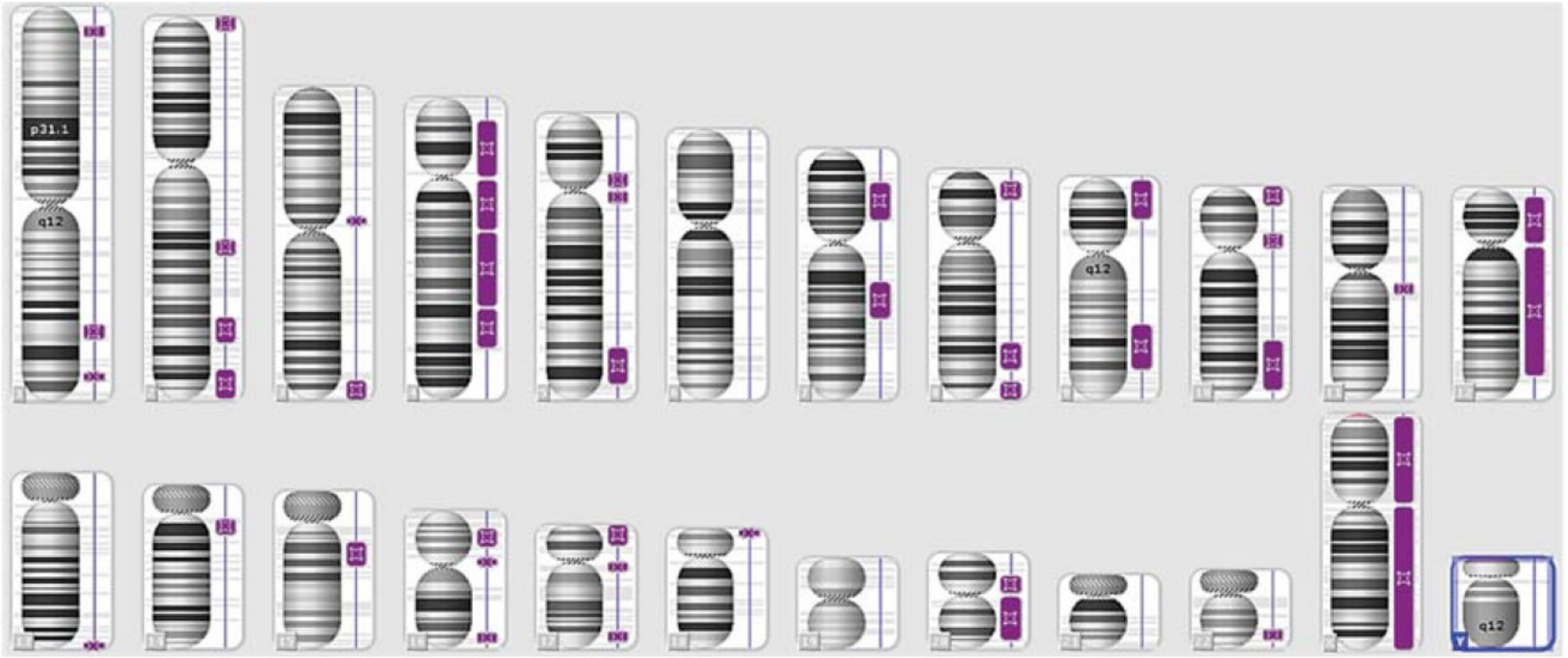


Fig. 1. ROH detected by SNP microarray analysis (Affymetrix Cytoscan HD) in a male child who was the offspring of a brother-sister mating. Each block on the right of the chromosome represents a genomic region at least 3 Mb in size. The laboratory-reported autosomal Froh was >21%.

Sund & Rehder (2014) *Hum Hered*

Variant effect: recessive and dominant

Dominant:

-- Effect observed both in homozygotes and heterozygotes

-- Variant frequency ~ disease incidence

-- Transmitted from one parent or *de novo*

Examples:

- Trp2332Ter in *CHD7*, CHARGE syndrome
- Arg5179His in *KMT2D* (aka *MLL2*), Kabuki syndrome

Recessive:

-- Effect observed in homozygotes only

-- Variant frequency >> disease incidence

-- Transmitted from both parents

Examples:

- Ex24:p.F508del in *CFTR*, cystic fibrosis
- Ex2:c.35delG in *GJB2*, hearing loss

X-linked recessive inheritance

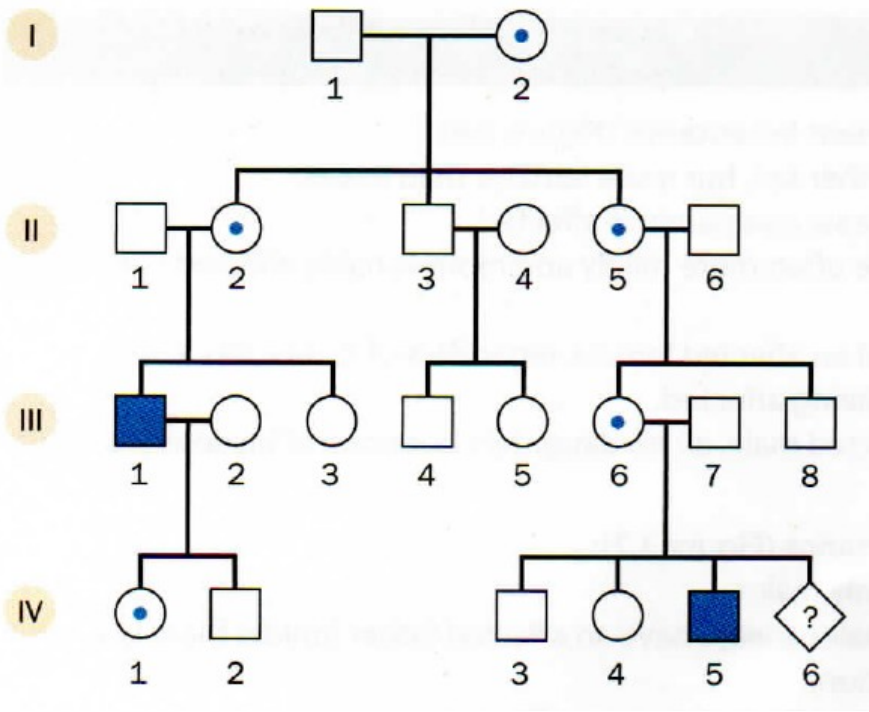
Recall X-chromosome patterns in men and women

- It affects mainly males
- Affected males are usually born to unaffected (carrier) parents
- A mother is normally an asymptomatic carrier
- Females may be affected if

– the father is affected and the mother is a carrier,

– or occasionally as a result of non-random X-inactivation.

- There is no male-to-male transmission in the pedigree

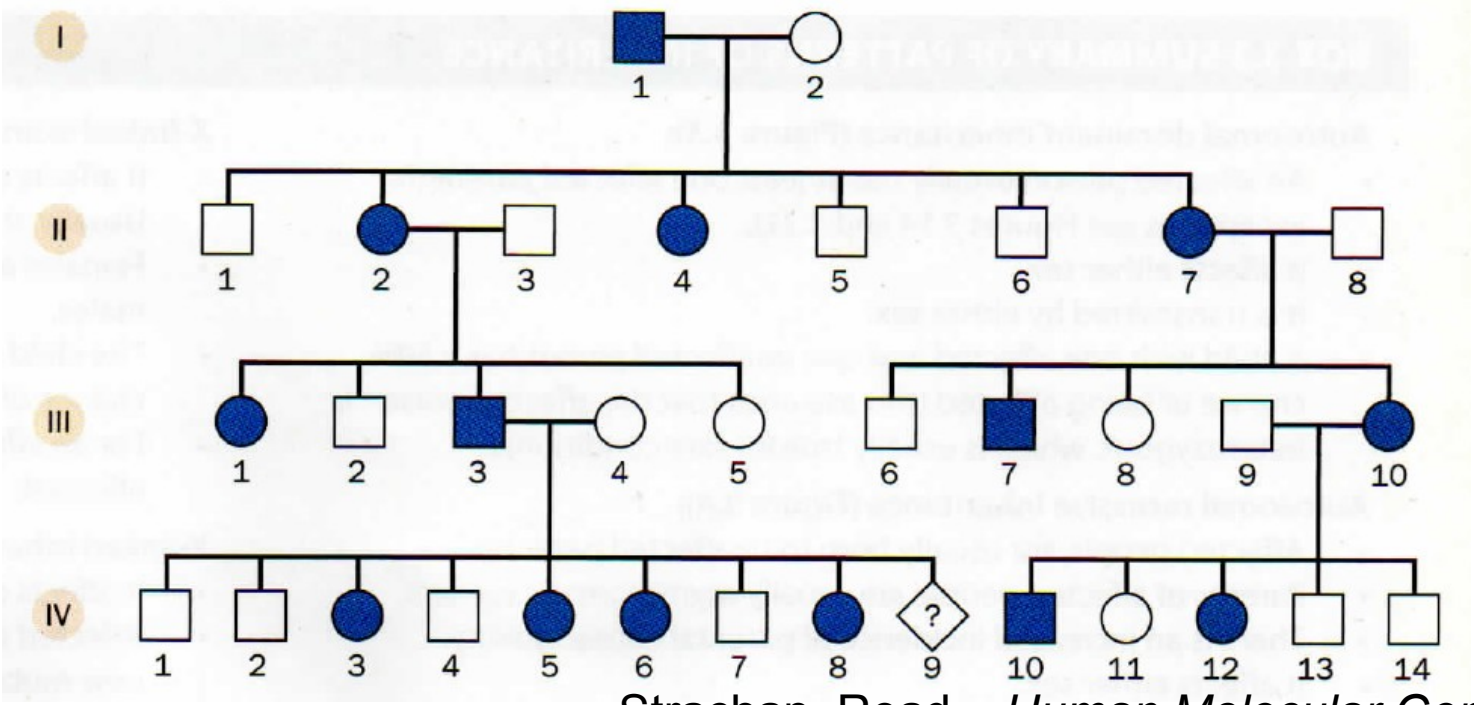


X-linked dominant inheritance

- It affects either sex, but more females than males
- Usually at least one parent is affected
- Females are often more mildly and more variably affected than males
- The child of an affected female, regardless of its sex, has a 50% chance of being affected.
- For an affected male, all his daughters but none of his sons are affected.

Q: why? (See below)

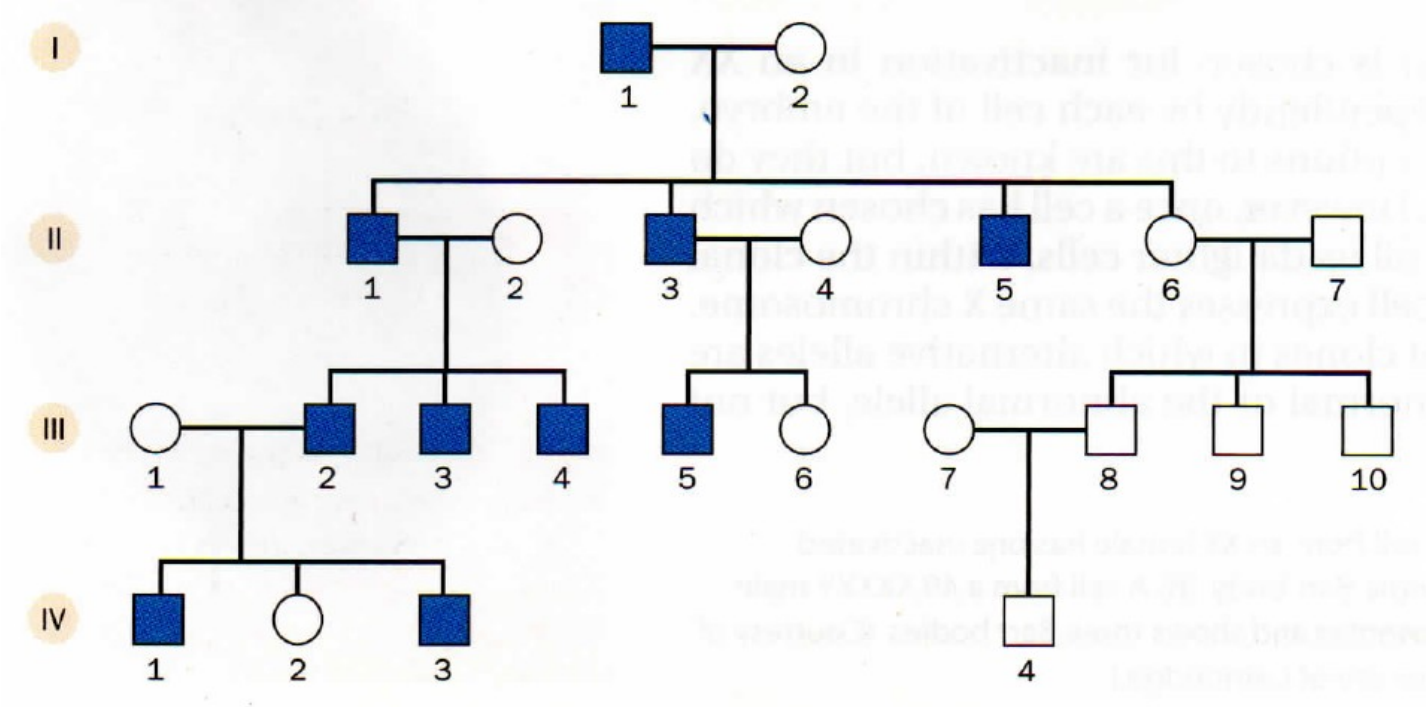
Q: why?



Strachan, Read – *Human Molecular Genetics*

Y-linked inheritance

- It affects only males
- Affected males always have an affected father
  – unless this is a *de novo* mutation
- All sons of an affected man are affected



Exercise

Earlier you found examples of disease-associated mutations for these annotation types:

- Stop-gain
- Synonymous
- Missense
- Splice-site
- Frameshift indel

What is the inheritance mode for each disease mutation? Provide references to the papers explaining the mutation discovery and/or molecular mechanism.

Mendelian diseases: overview

Mendelian (monogenic) diseases depend on the genotype at a single locus (or gene), with inheritance following Mendel's laws of segregation, independent assortment and dominance.

Mendelian inheritance patterns, prevalence per 1,000 births*

- Autosomal dominant 1.40
- Autosomal recessive 1.84 + F×650 (consanguinity-related)
- X-linked recessive 0.05
- X-linked dominant N/A
- Y-linked N/A
- Unknown 1.16

Overall prevalence: **~0.4% of live births**

* Ref: Blencowe (2018) *J Community Genet*

Mendelian diseases: OMIM

Number of Entries in OMIM (Updated May 23rd, 2020) :

MIM Number Prefix	Autosomal	X Linked	Y Linked	Mitochondrial	Totals
Gene description *	15,447	742	51	37	16,277
Gene and phenotype, combined +	35	0	0	0	35
Phenotype description, molecular basis known #	5,431	348	5	33	5,817
Phenotype description or locus, molecular basis unknown %	1,423	118	4	0	1,545
Other, mainly phenotypes with suspected mendelian basis	1,668	103	3	0	1,774
Totals	24,004	1,311	63	70	25,448

Examples of Mendelian diseases

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

Jackson (2018) *Essays in Biochemistry*

Examples of Mendelian diseases

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*

Examples of Mendelian diseases

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*

Examples of Mendelian diseases

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

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

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

Jackson (2018) *Essays in Biochemistry*

Exercise

Use OMIM to find example of a disease for each type of inheritance:

- Autosomal dominant
- Autosomal recessive
- X-linked recessive
- X-linked dominant
- Y-linked
- Mitochondrial

For each case, prepare an example of a related gene and causal mutation in the gene

Complications to the Mendelian inheritance

- **Locus heterogeneity:** the same clinical phenotype can result from mutations at anyone of several different loci.
- **Allelic heterogeneity**: many different mutations within a given gene cause same disease
- **Clinical heterogeneity**: mutations in the same gene produce two or more different diseases in different people. Note: not the same as **pleiotropy**

Example: mutations in the *HPRT* gene can produce either a form of gout (подагра) or Lesch-Nyhan syndrome: severe mental retardation with behavioral problems [OMIM:300322].

- **Incomplete penetrance*:** a person who has the disease genotype does not manifest the disease. In particular, **age-related penetrance** in late-onset diseases.

* **Penetrance** of a disease-causing mutation is the proportion of individuals with the mutation who exhibit clinical symptoms

Complications to the Mendelian inheritance

- **Variable expression**: different family members show different features of the disease
- **Imprinting**: mutation has effect only when inherited from a parent of particular sex.

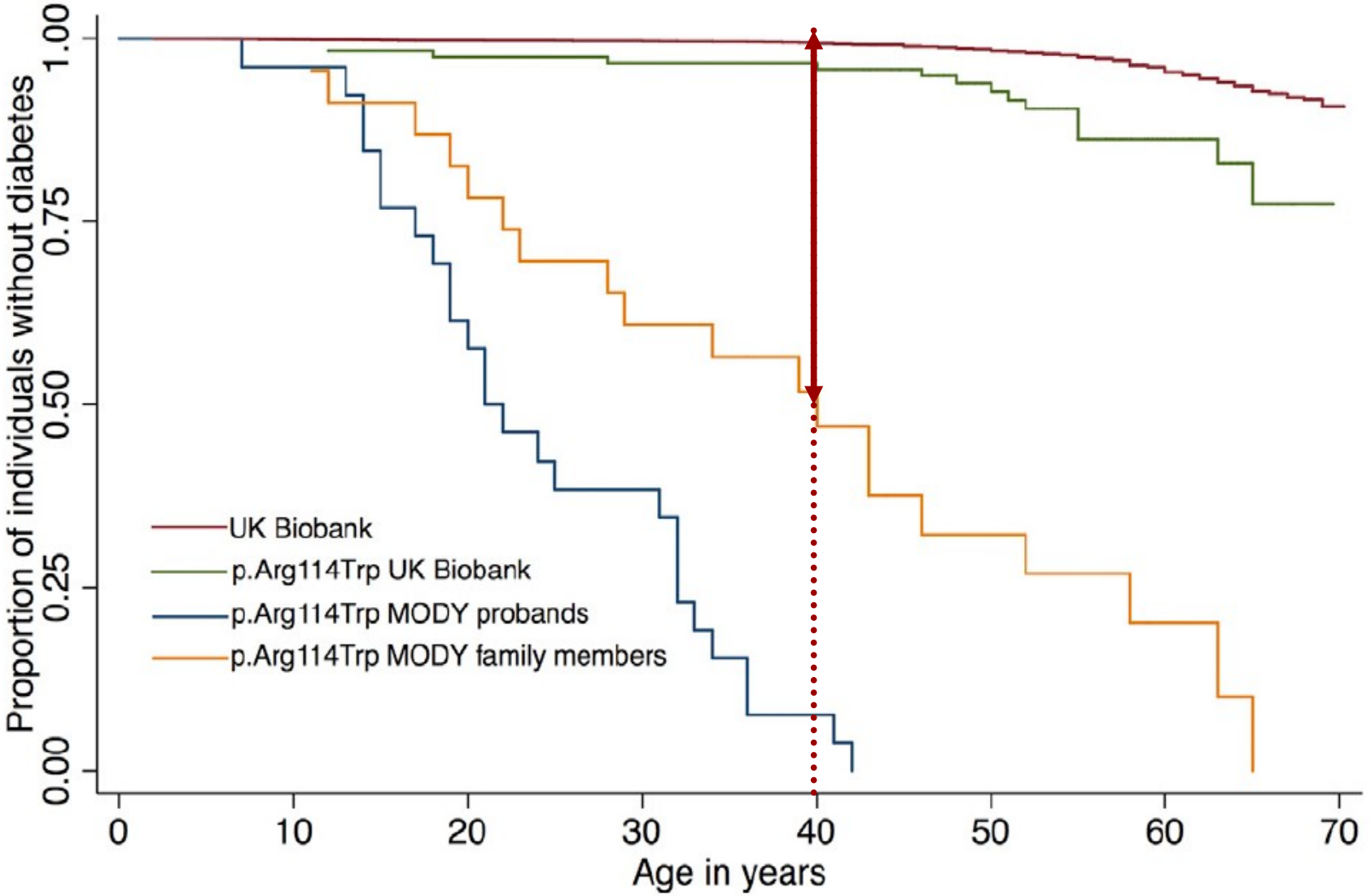
Examples:

– autosomal dominant inheritance of *paragangliomas* [OMIM:168000]; only if inherited from father.

– *Beckwith-Wiedemann syndrome* [OM1M:130650], only in babies who inherit it from their mother

- **Phenocopy**: disease without causal genotype. Example: deafness
- ***De novo* mutations** complicate Mendelian inheritance
- **Mosaicism** in germ-line of somatic cells

Penetrance: examples



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

Wright (2019) AJHG

Penetrance: ClinVar examples

Gene, variant, ClinVar ID	*Disease*	*Penetrance*
BRCA1 DNA Repair Associated *BRCA1* p.Arg1699Gln SCV000210198.11	Breast cancer, ovarian cancer	A study of 4,024 individuals from 129 families (Moghadasi 2017): a **20% risk of breast cancer** and a **6% risk of ovarian cancer** by age 70. Lifetime risks associated with typical BRCA1 variants are estimated to be **57 to 87% for female breast cancer and 24 to 54% for ovarian cancer** (Claus 1996, Antoniou 2003, King 2003, Risch 2006, Chen 2007)
Homeostatic Iron Regulator *HFE* p.Cys282Tyr SCV000221190.3	Hemochromatosis	Biochemically, 82% of p.Cys282Tyr homozygotes were shown to have elevated transferrin saturation (Pederson 2009); however, **<5% of individuals** with biallelic pathogenic HFE variants exhibit clinical symptoms of HH (Beutler 2002 , Gurrin 2009)
Leucine Rich Repeat Kinase 2 *LRRK2* 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 **from 20% at age 50 years to 80% at age 70 years** (PMID: 18986508, 15726496).

Penetrance, relative risk, odds ratio

	Diseased	*Healthy*
Mutation	D_m	H_m
No mutation	D_0	H_0

Disease risk: probability of disease with mutation: $\frac{D_m}{D_m + H_m}$

– Similar to penetrance

– Does not account for risk without mutation

Risk ratio: $RR = \frac{D_m(D_0 + H_o)}{D_0(D_m + H_m)}$

Exercise: When *OR* ≈ *RR* ?

Odds ratio: $OR = \frac{D_m/H_m}{D_0/H_0} = \frac{D_m H_0}{D_0 H_m}$ // Odds of an event: $p/(1\text{-}p)$

Exercise: calculate OR, RR values for $D_m = 60$, $H_m = 40$, $D_0 = 2$, $H_0 = 48$

MAF and OR: UK Biobank examples

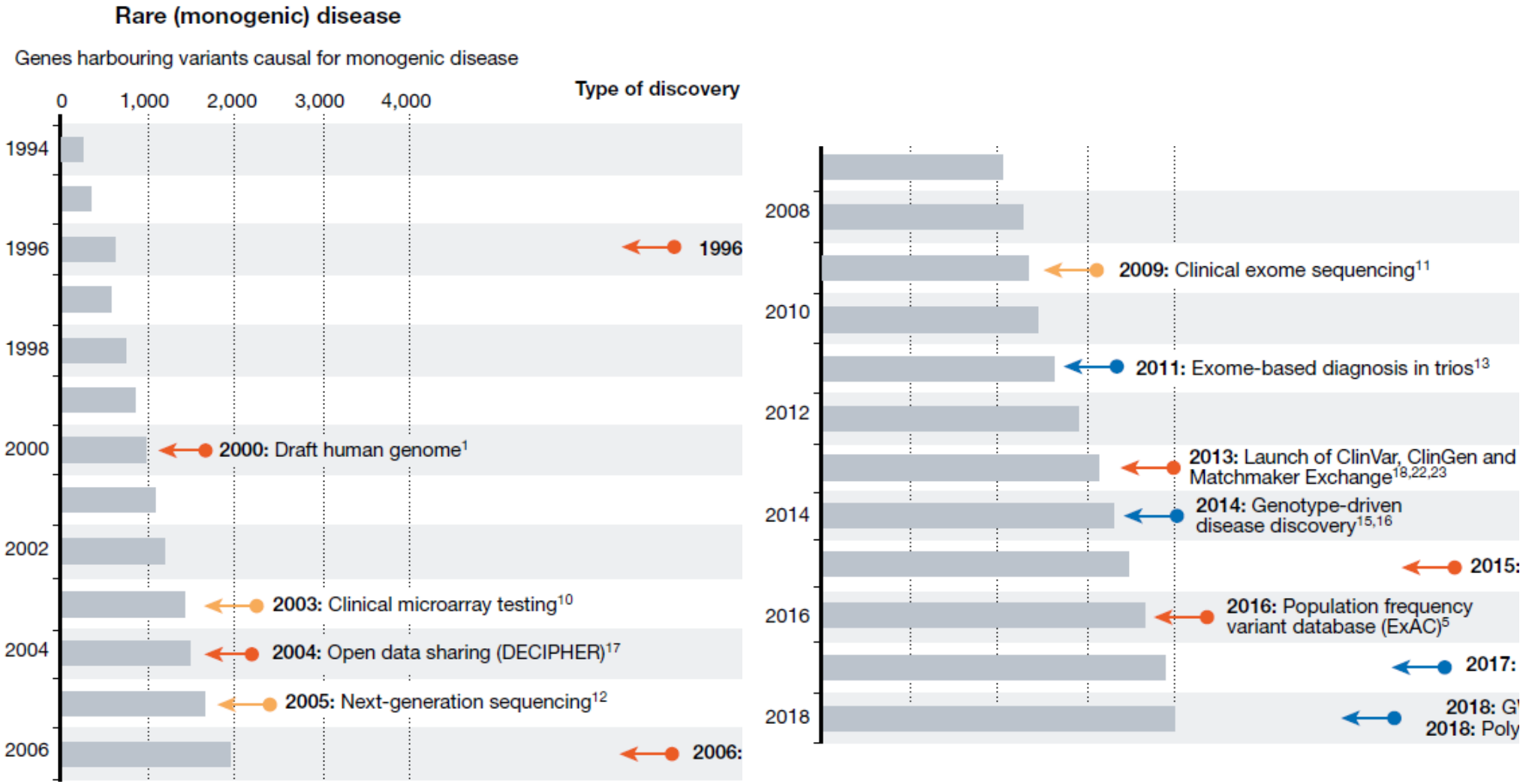
Gene	UKB ID	Position (GRCh37)	HGVS	MAF White British (%)	Significantly Associated Trait(s) in UKB (Units)	Odds Ratio or Beta [95% CI]	p value	Linked Disease (Mode of Inheritance)
ACSF3	dbSNP: rs141090143	chr16: 89220556 C>T	GenBank: NM_174917: c.C1672T:p.R558W	0.632	ease of sunburn (number of episodes)	0.31 [0.20, 0.42]	4×10^{-10}	combined malonic and methylmalonic aciduria (AR)
AR	dbSNP: rs137852591	chrX: 66941751 C>G	GenBank: NM_000044: c.C2395G:p.Q799E	0.129	skeletal mass (SD)	−0.16 [−0.21, −0.11]	1×10^{-10}	partial androgen insensitivity syndrome (XLR)
					height (cm)	−0.85 [−1.27, −0.43]	1×10^{-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]	1×10^{-8}	partial androgen insensitivity syndrome (XLR)
ERCC4	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]	2×10^{-8}	xeroderma pigmentosum (AR)
FLG	dbSNP: rs150597413	chr1: 152277622 G>T	GenBank: NM_002016: c.C9740A:p.S3247X	0.369	eczema	1.66 [1.40, 1.98]	9×10^{-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]	5×10^{-16}	ichthyosis vulgaris (AD)
GCK	dbSNP: rs104894006	chr7: 44189591 G>A	GenBank: NM_000162: c.C556T:p.R186X	0.001	maturity-onset diabetes of the young	68 [14, 325]	2×10^{-8}	diabetes mellitus (AD)

Wright (2019) AJHG

Disease gene discovery

- Rare clinical observations: difficult to observe recurrence required for Mendelian patterns
- *De novo* mutations: no segregation in families // dominant or compound heteterozygotes in case of recessive
- Causal mutations, not genes, are needed: functional assays needed; bias towards obvious variants
- All abovementioned complications to Mendelian inheritance: locus, allelic and clinical heterogeneity; incomplete penetrance and variable expression; imprinting, phenocopies and mosaicism

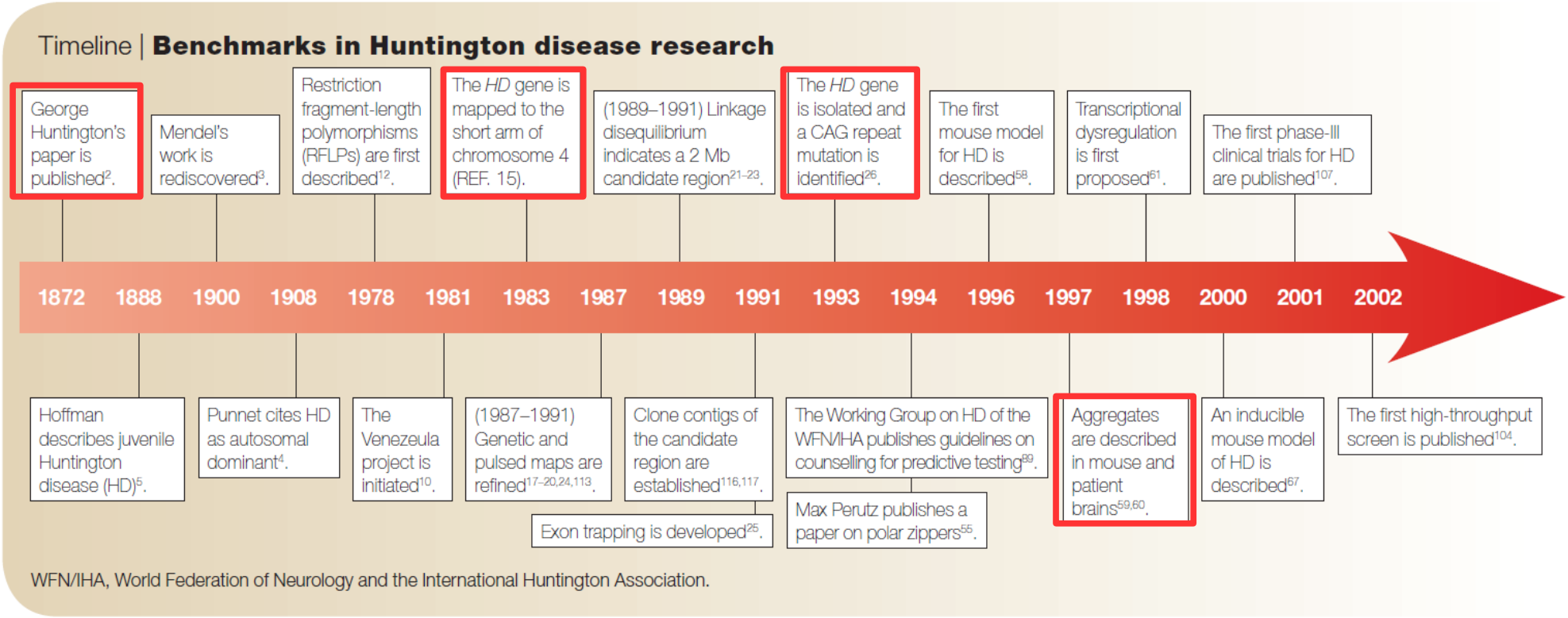
Disease gene discovery



Claussnitzer (2020) *Nature*

Huntington disease gene discovery

- Late-onset (30-45 years old) neurodegenerative, progress ~15-20 years
- Psychiatric disturbances, motor impairments and a cognitive decline
- Dominant inheritance, no sporadic forms
- First genetic disease locus to be mapped to a chromosome (1983)
- Still (2018) no treatment besides symptomatic



Bates (2005) *Nat Rev Genet*

Huntington disease gene discovery

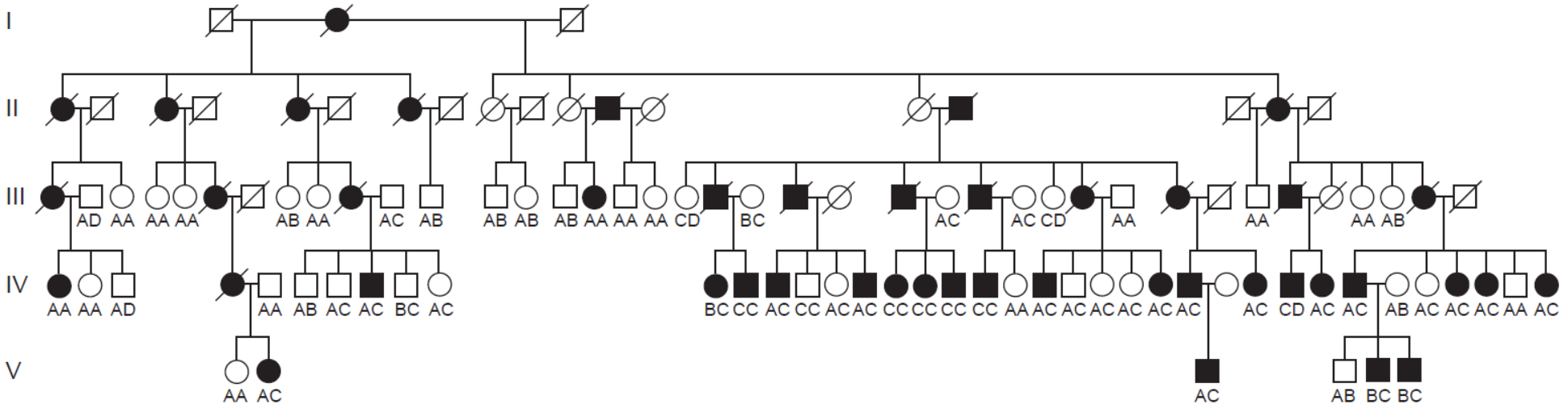
- High incidence of HD in Venezuela, single founder
- 1983, James Gusella Lab: eighth polymorphic marker studied mapped the HD to the ~4cM telomeric fragment of chr4
- No technology to "walk" along a chromosome for >100-200kbp
- Collaboration to map and clone the HD gene
- 1993: the *HTT* (huntingtin) gene cloned by the joint effort of 9 labs; 10,366 transcript with a **CAG-triplet (Gln) repeat** in exon 1 that was polymorphic on normal chromosomes and expanded in HD
- 1993–1996: **The $(CAG)_N$ ranges**: 6–35 benign; >40 are fully penetrant and cause HD within a normal lifespan; >70: juvenile offset
- **Paternal only anticipation**: $(CAG)_N$ expands during transmission
- Poly-Gln repeats in the pathogenic range spontaneously aggregate into amyloid fibrils $\Longrightarrow$ neuronal degeneration
- Testing in childhood for adult-onset untreatable disorders holds the potential of more harm than benefit

Bates (2005) *Nat Rev Genet*

Mendelian disease gene discovery by linkage

- Multiple pedigrees with affected and unaffected members
- Map of polymorphic DNA markers with known genetic distances

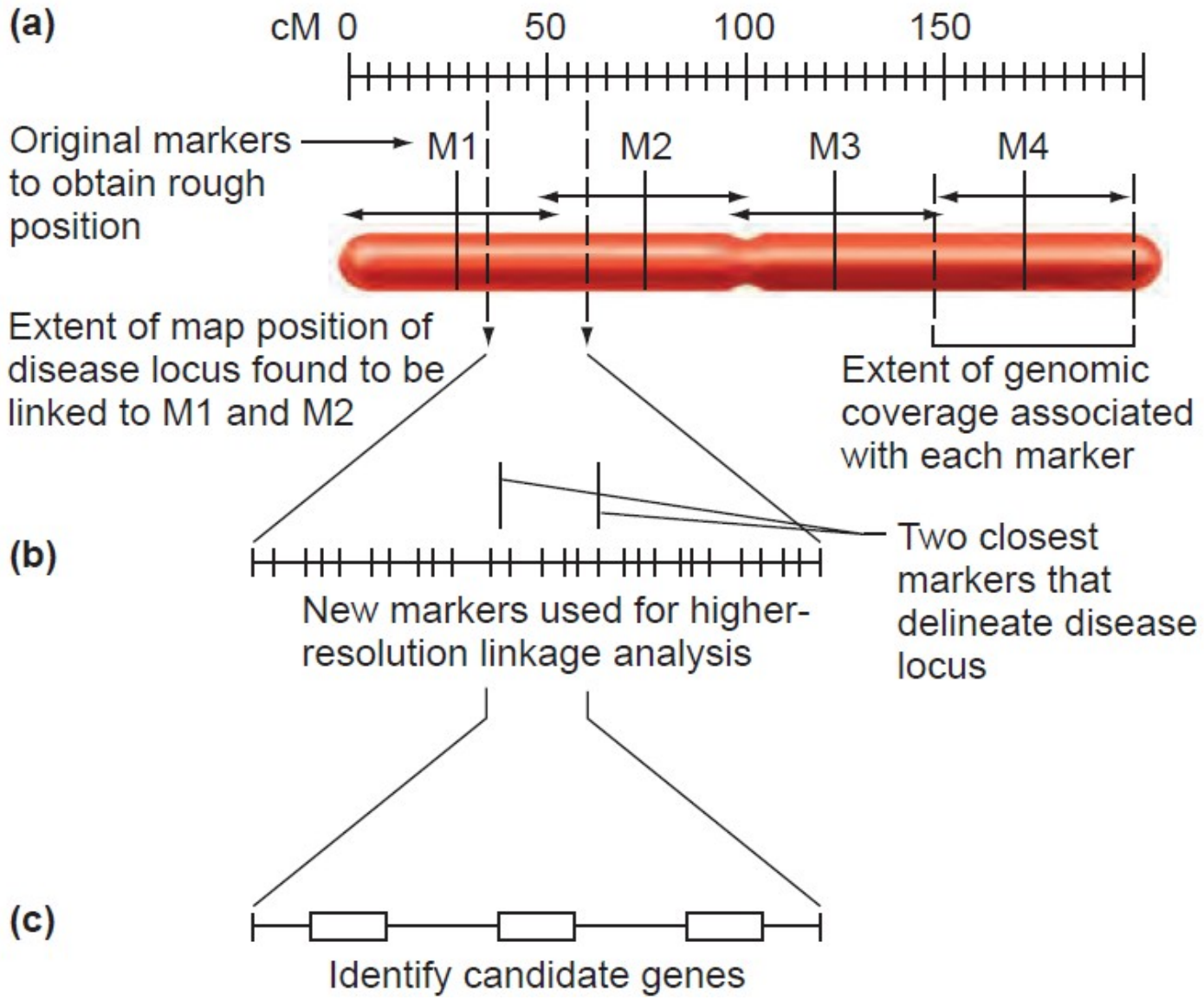
1. Find DNA markers that cosegregate (are in linkage) to the disease trait in pedigrees, identify putative region of the disease gene
2. Sequence the genes within the linked locus to identify disease-causing alleles, check alleles in healthy individuals
3. Conduct confirmatory functional studies in cellular and animal models



Detection of locus responsible for Huntington disease

Hartwell – *Genetics. From genes to genomes*

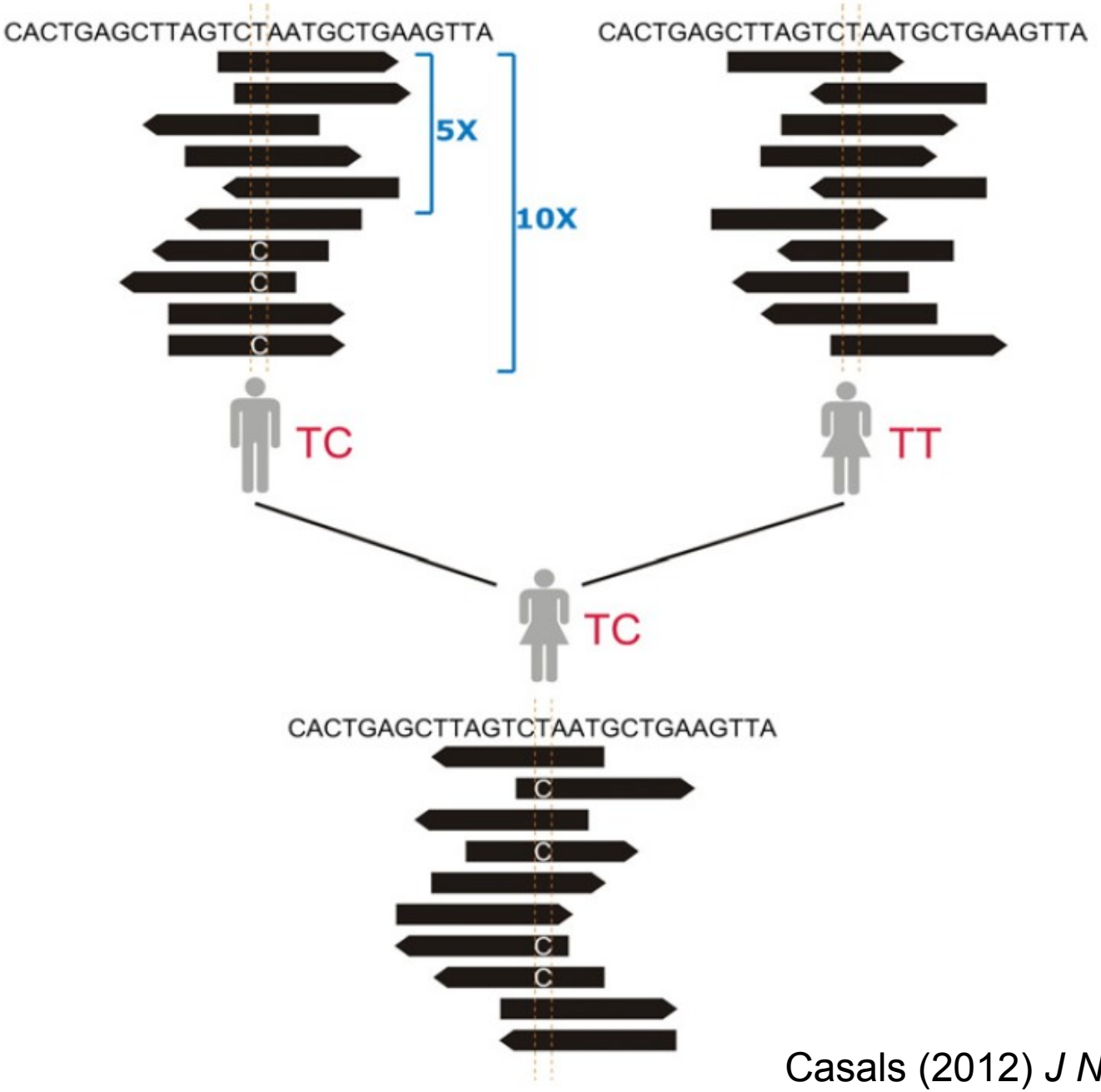
Mendelian disease gene discovery by linkage



From phenotype to chromosomal location to guilty gene

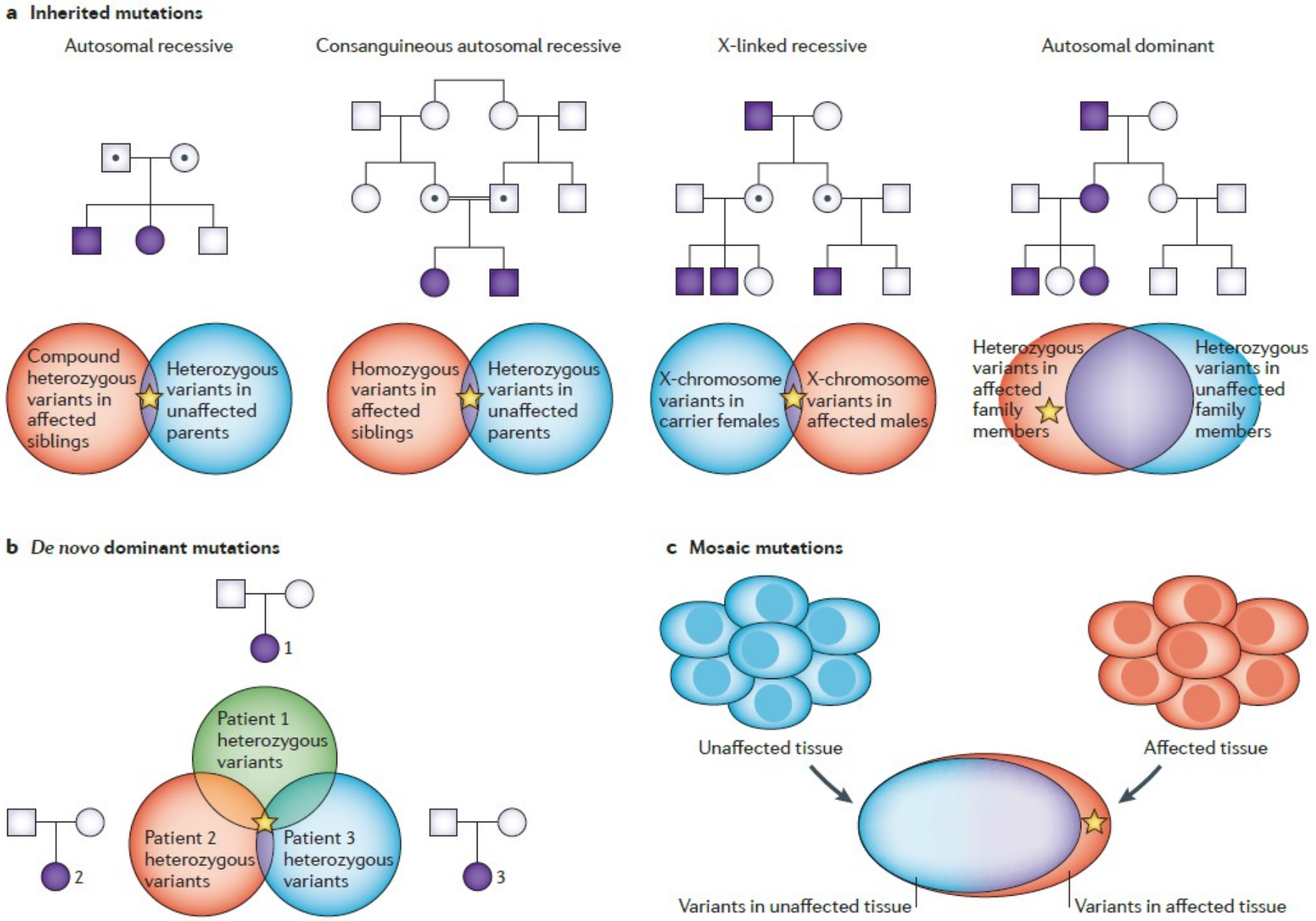
Hartwell – *Genetics. From genes to genomes*

Mendelian disease gene discovery by NGS



Casals (2012) *J Neuroimmunology*

Mendelian disease gene discovery by NGS



Boycott (2013) *Nat Rev Genet*

Mendelian disease gene discovery by NGS

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*

Mendelian disease gene discovery by NGS

Variant prioritization: «needle in the haystack» problem, determining which variants identified in the course of whole-exome or whole-genome sequencing are most likely to damage gene function and underlie the disease phenotype.

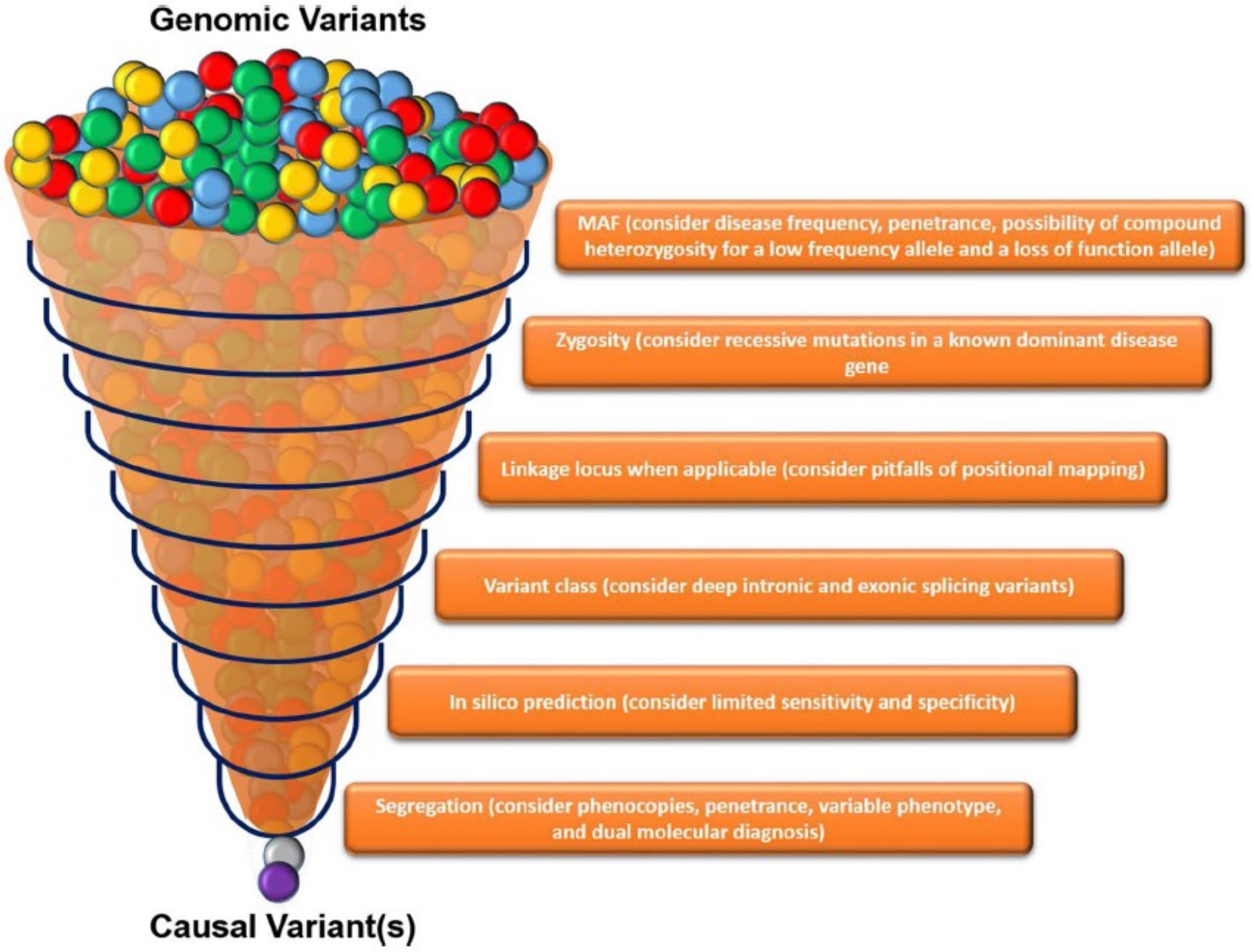
SNVs	***Average***	***Deviation***
PTV *HIGH*	97	6
Missense *MODERATE*	6291	139
Synonymous *LOW*	7192	88
Other *MODIFIER*	561	13
Indels		
Frameshift	69	3
Other	41	3

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*

Mendelian disease gene discovery by NGS



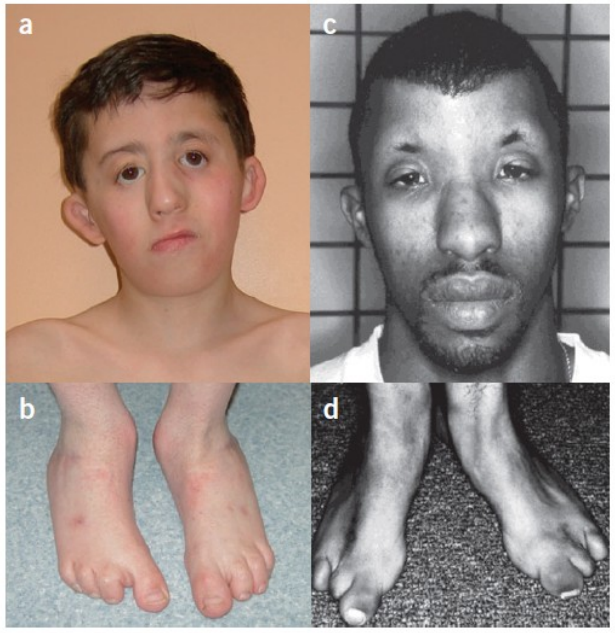
Alkuraya (2016) *Hum Genet*

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



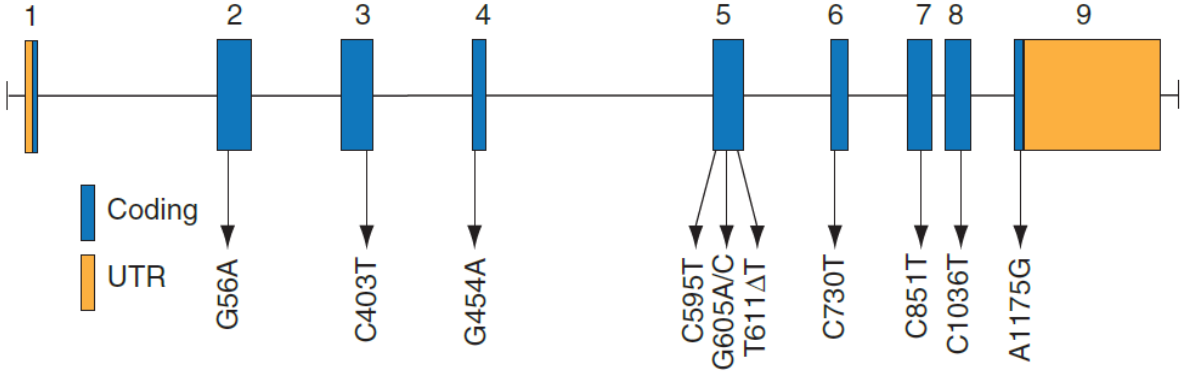


Figure 1. Clinical characteristics of an individual with Miller syndrome and an individual with methotrexate embryopathy.

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.

Mendelian disease gene discovery by NGS

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*

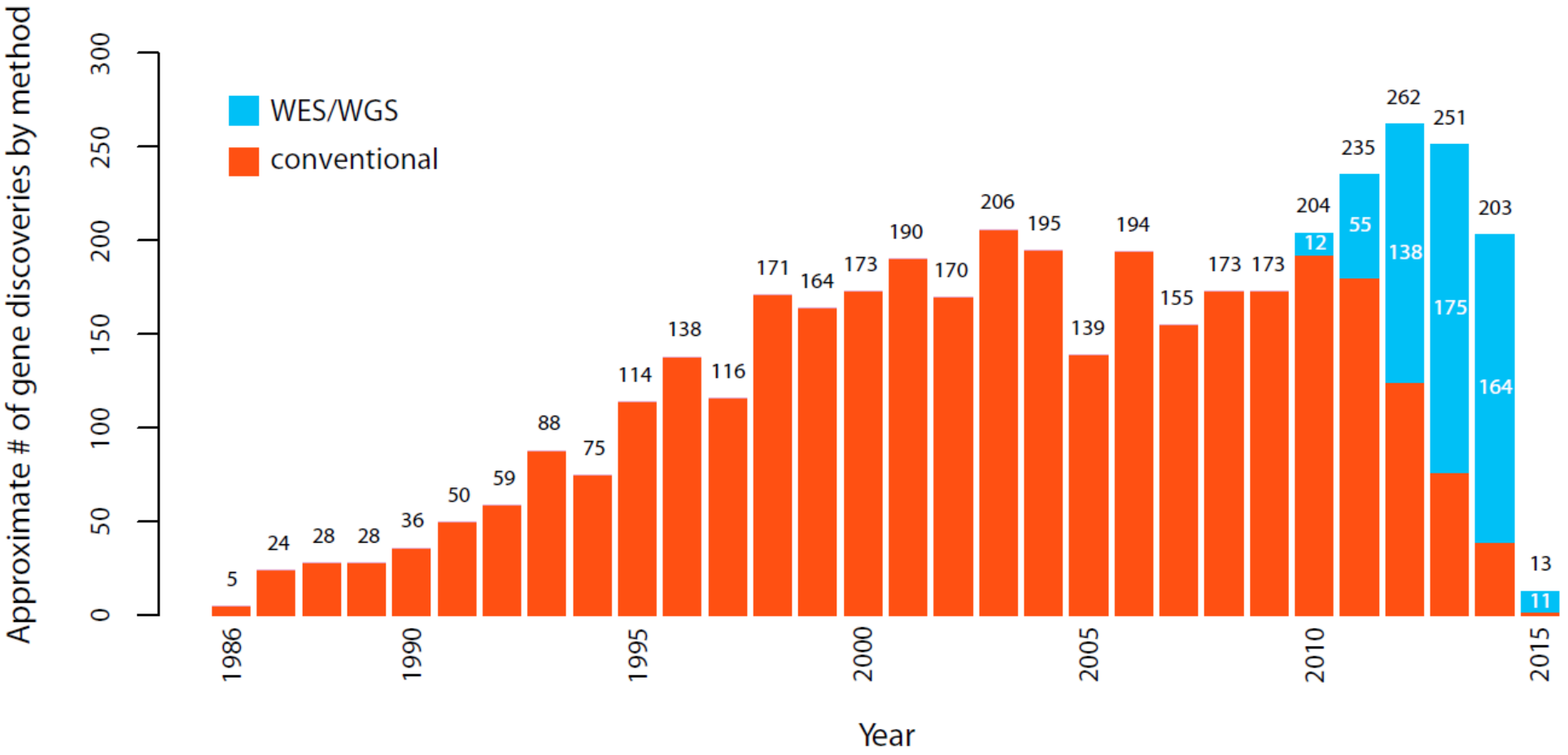
Mendelian disease gene discovery by NGS

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]

Rabbani (2012) *J Hum Genet*

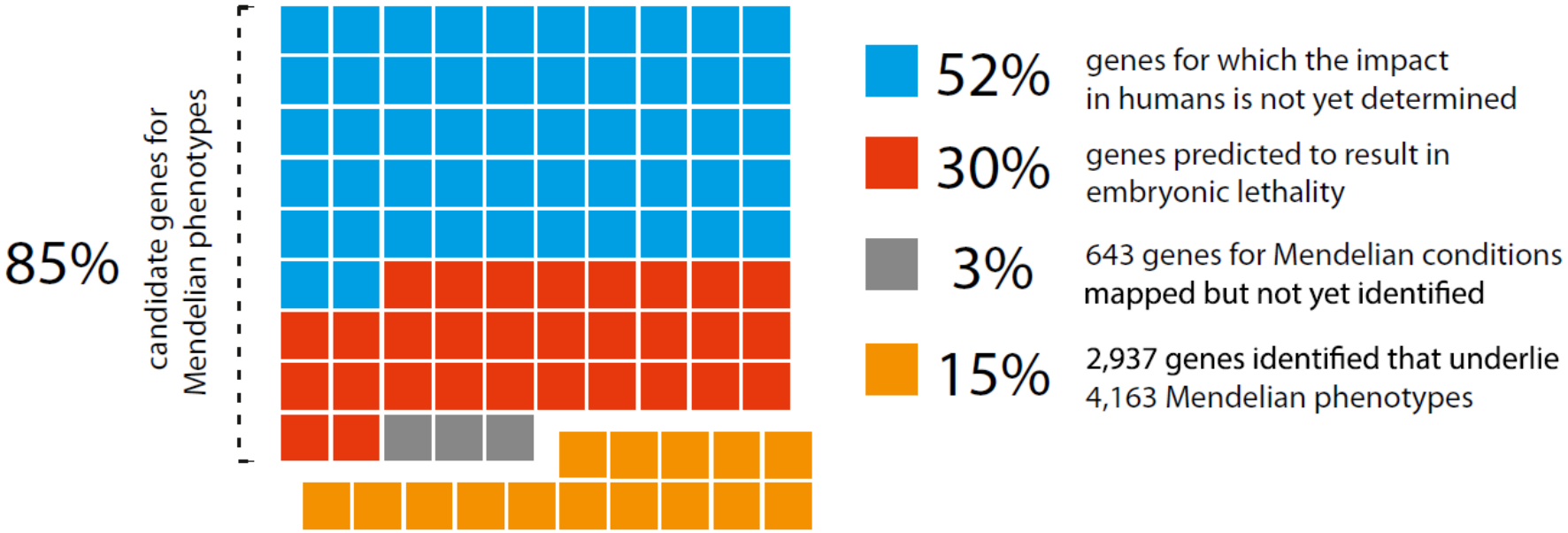
Mendelian disease gene discovery by NGS



Approximate Number of Gene Discoveries Made by WES and WGS versus Conventional Approaches since 2010

Chong (2015) *Am J Hum Genet*

Mendelian disease gene discovery by NGS



Relationship between human protein-coding genes and Mendelian phenotypes
Of approximately ~19,000 protein-coding genes predicted to exist in the human genome, variants that cause Mendelian phenotypes have been identified in ~2,937 (~15.5%; orange squares). Genes underlying ~643 Mendelian phenotypes (~3.38%; gray squares) have been mapped but not yet identified. On the basis of analysis of knockout mouse models, LOF variants in up to ~30% of genes (~5,960; red squares) could result in embryonic lethality in humans. Note that the consequences of missense variants in these genes could be different. For a minimum of ~52% of genes (~10,330; blue squares), the impact in humans has not yet been determined. Collectively, ~16,063 genes remain candidates for Mendelian phenotypes.

Chong (2015) *Am J Hum Genet*

Disease genetics in the post-genome era

Microarrays and NGS

- Detection of structural variation
- NGS has enabled the full range of causal genetic variation
- Reduced reliance on multiplex pedigrees in favour of collections of affected cases, often with parents, has proven decisive in identifying new dominant disorders

Functional assays

- Highly parallelized in vitro cellular assays that allow assessment of the functional effects of all variants in a disease-associated gene can transform interpretation of novel variants
- Functional analysis of disease-relevant tissues from patients using RNA sequencing and DNA methylation assays can identify previously cryptic causal genetic variants outside of protein-coding genes

Claussnitzer (2020)

Disease genetics in the post-genome era

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)

Disease genetics in the post-genome era

General complications:

- Rare diseases: small number of cases and/or families
- Variable penetrance
- Unknown mode of inheritance
- Locus heterogeneity, phenotypic heterogeneity
- *De novo* or inherited variants (sporadic vs family cases)
- Allele frequency can be deceiving
- *In silico* prediction algorithms are limited

Whole exome sequencing (WES) limitations:

- Many non-coding variants not detected
- Difficulties in detecting structural variants and CNVs
- False negative (coverage) and false positive variant calls
- Large number of candidate variants, filtering required

Whole genome sequencing (WGS) limitations:

- Too much data, even more filtering required
- Sequencing and processing costs

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

The clinical significance of any given sequence variant falls along a gradient, ranging from those in which the variant is almost certainly pathogenic for a disorder to those that are almost certainly benign.

A five-tier system of classification for variants relevant to Mendelian disease:

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

We propose that the terms *likely pathogenic* and *likely benign* be used to mean greater than 90% certainty of a variant either being disease causing or benign in order to provide laboratories with a **common, albeit arbitrary, definition**

Richards (2015) *Genetics in Medicine*

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).

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 PM2 Absent from controls (or at extremely low frequency if recessive) (**Table 6**) in Exome Sequencing Project,

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 PM2 Absent from controls (or at extremely low frequency if recessive) (**Table 6**) in Exome Sequencing Project,

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 sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

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

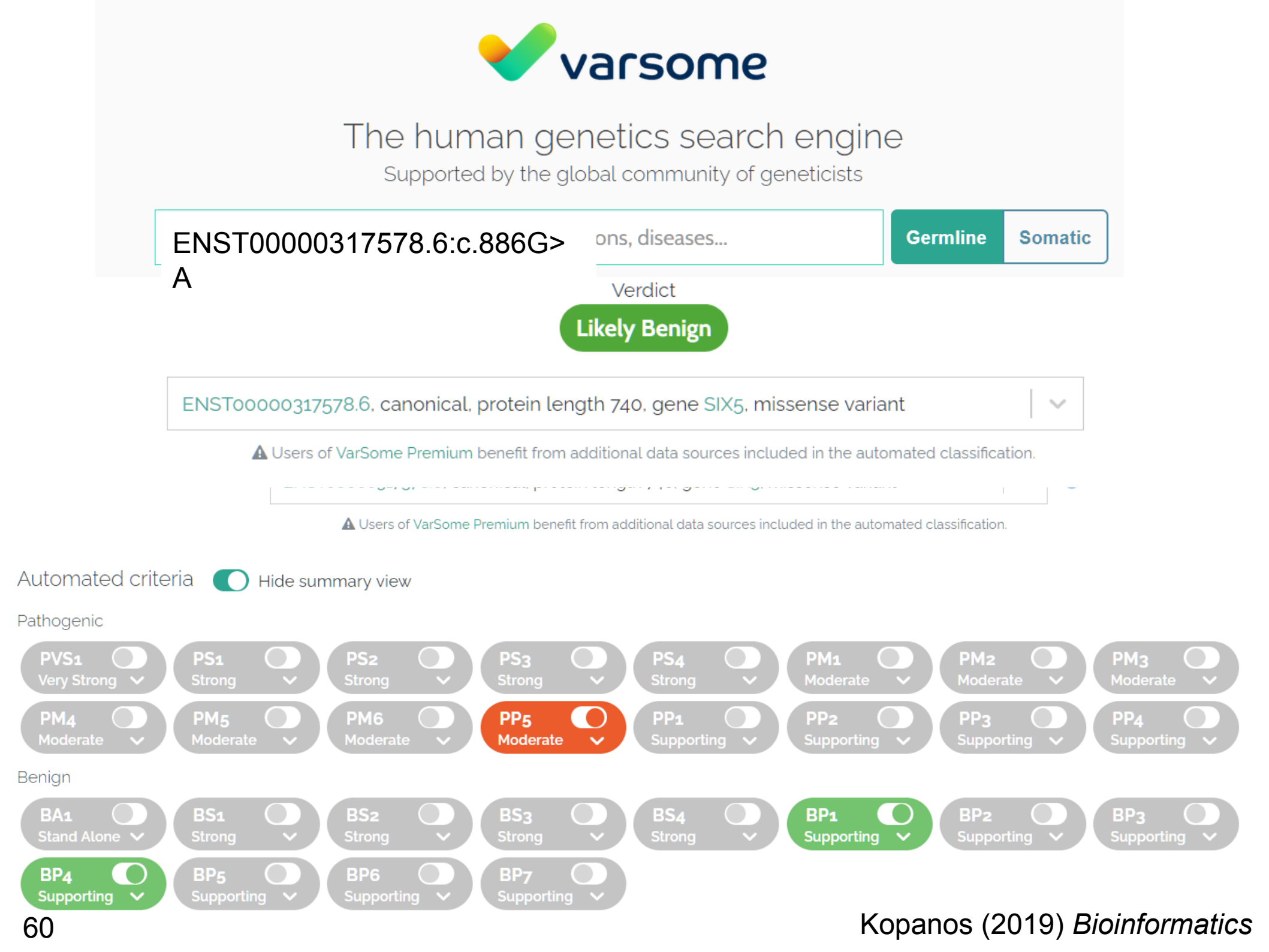
Richards (2015) *Genetics in Medicine*

ACMG STANDARDS AND GUIDELINES

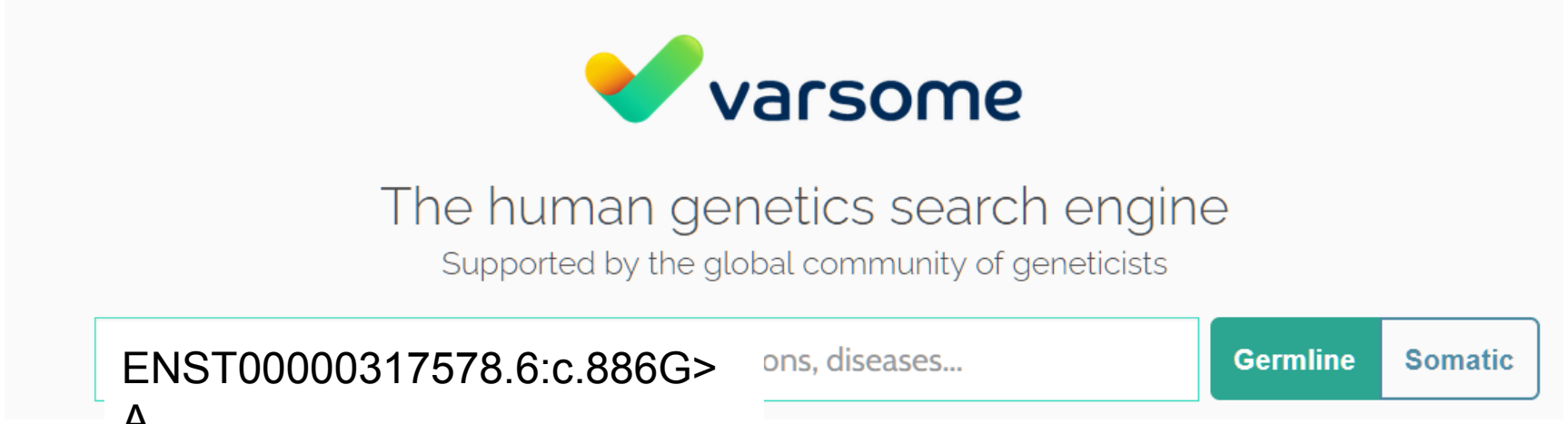
Genetics in Medicine

	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			

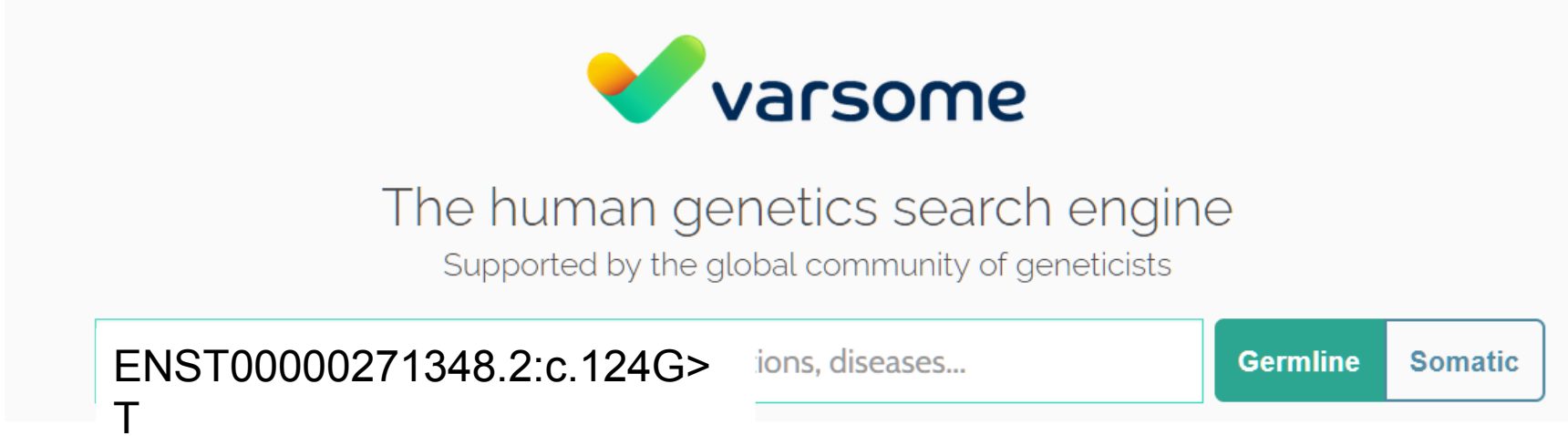
Richards (2015) *Genetics in Medicine*



Kopanos (2019) *Bioinformatics*



Kopanos (2019) *Bioinformatics*



Kopanos (2019) *Bioinformatics*

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*

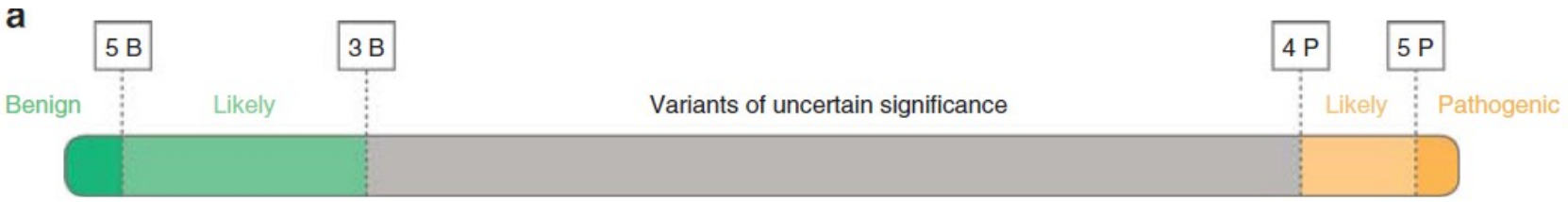
Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria

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).

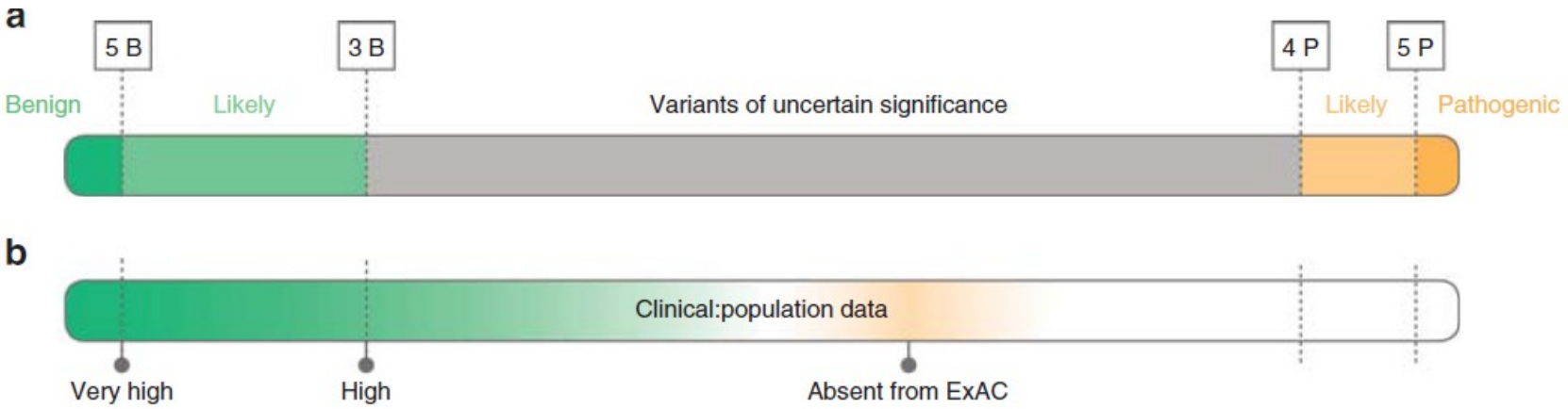
Nykamp (2017) *Genetics in Medicine*

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



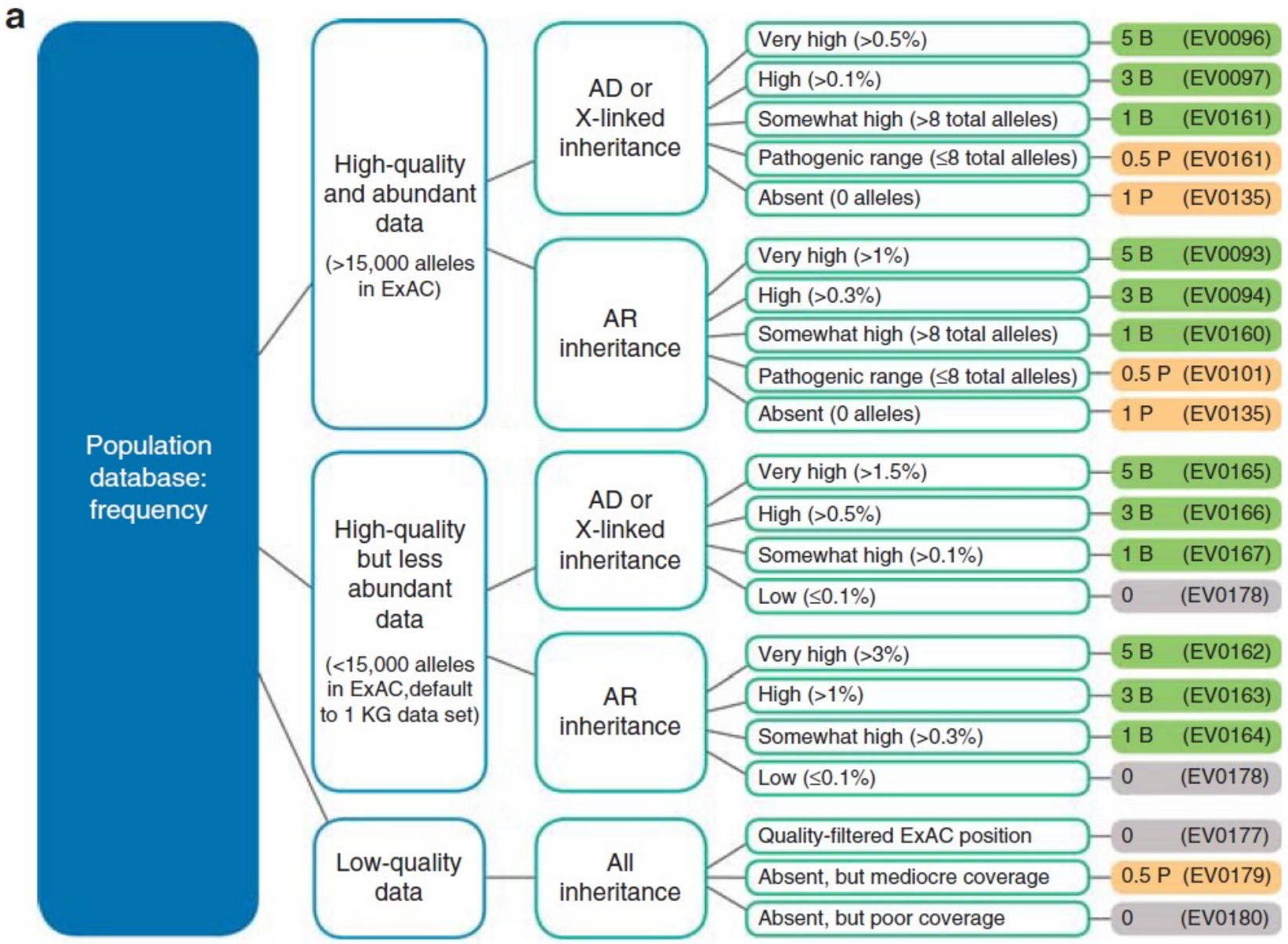
Nykamp (2017) *Genetics in Medicine*

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



Nykamp (2017) *Genetics in Medicine*

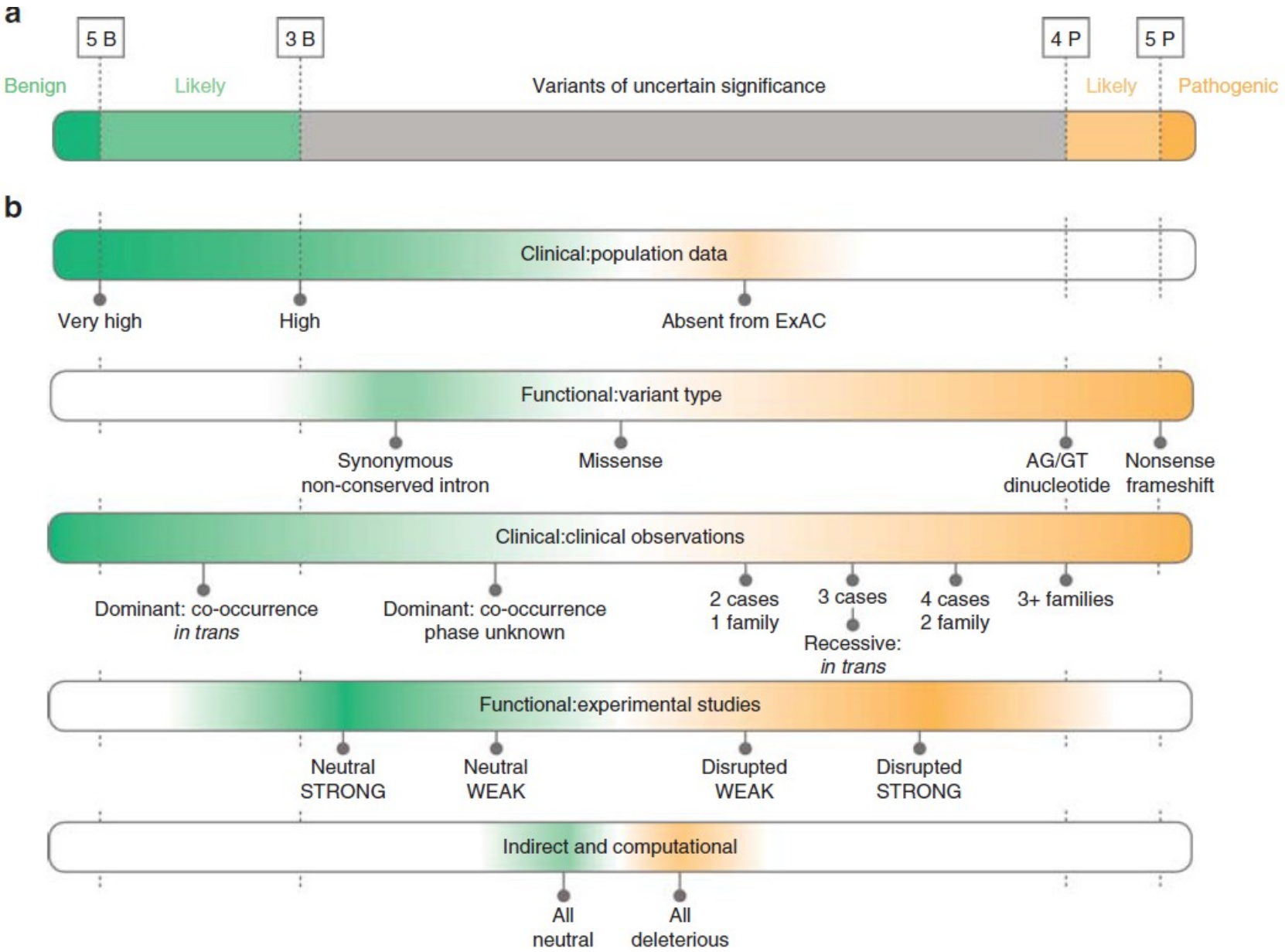
Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria



Population data: Sherloc criteria and decision tree

Nykamp (2017) *Genetics in Medicine*

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



Classification scoring thresholds and evidence categories

Nykamp (2017) *Genetics in Medicine*

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

Example 1: *TTC8* c.459G>A (p.Thr153=)

- A very rare silent change (0.02% in ExAC) in a gene that can cause Bardet–Biedl syndrome
- Predicted to disrupt normal splicing
- Observed in the homozygous state in three affected siblings in a single family
- Observed in our laboratory in the homozygous state in an unrelated affected individual and is now classified as **pathogenic**

Example 2: *CDH1* c.1118C>T (p.Pro373Leu) is a variant in a gene associated with hereditary diffuse gastric cancer and lobular breast cancer.

- absent from ExAC
- supported by strong functional studies: impairs cell–cell adhesion and leads to increased cellular motility and activation of *EGFR*, mitogen-activated protein kinase, and Src kinase.
- Computational predictors recapitulate this conclusion.
- Clinical observations, however, are inconclusive: the variant has been found in affected and unaffected individuals in the same family.

Without supporting clinical observations, **likely pathogenic seems premature**

Nykamp (2017) *Genetics in Medicine*

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

Example 3: *CDKN2A* c.9_32dup24

- In-frame duplication
- Predicted to have no effect on protein function
- Demonstrated not to affect CDK4 or CDK6 binding
- Identified in several individuals affected with melanoma
- Segregate with disease (incomplete penetrance) in several melanoma families

The **abundance of positive clinical evidence trumps the negative functional evidence** (CDK4/6 binding efficiency is not the relevant molecular consequence)

Example 4: *SCN5A* c.3578G>A (p.Arg1193Gln)

- Missense change in the voltage-gated cardiac sodium channel.
- Demonstrated to destabilize inactivation gating and to lead to a persistent current in vitro.23
- Glycine is present at the equivalent position in the horse ortholog,
- >7% of the East Asian population, with 17 homozygotes reported in ExAC.

The abundance of **negative clinical evidence outweighs the positive functional evidence**

Nykamp (2017) *Genetics in Medicine*

Exercise

Earlier you found examples of disease-associated mutations for these annotation types:

- Stop-gain
- Synonymous
- Missense
- Splice-site
- Frameshift indel

Use `submission_summary.txt` available at ClinVar FTP to explain which criteria were used to classify each variant as pathogenic

ClinVar: open database of disease mutations

Category of analysis	Current total (May 13, 2020)
Records submitted	1141302
Records with assertion criteria	969361
Records with an interpretation	1119301
Total genes represented	32838
Unique variation records	745458
Unique variation records with interpretations	733504
Unique variation records with assertion criteria	635153
Unique variation records with practice guidelines (4 stars)	656
Unique variation records from expert panels (3 stars)	10911
Unique variation records with assertion criteria, multiple submitters, and no conflicts (2 stars)	101805
Unique variation records with assertion criteria (1 star)	488040
Unique variation records with assertion criteria and a conflict (1 star)	33741
Unique variation records with conflicting interpretations	34051
Genes with variants specific to one gene	11064
Genes with variants specific to one protein-coding gene	10971
Genes included in a variant spanning more than one gene	33087
Variants affecting overlapping genes	27744
Total submitters	1565

WES diagnostics of Mendelian disorders

Applicability

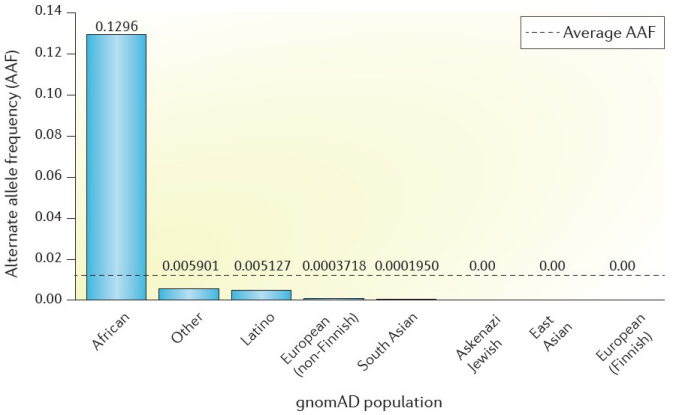
- Atypical manifestation
- Symptoms shared among multiple disorders
- Difficult to confirm by clinical or laboratory criteria

Input

- Clinical symptoms (HPO)
- Medical record

Annotation

- VEP
- gnomAD
- ClinVar, OMIM



Bamshad (2011) *Nat Rev Genet*

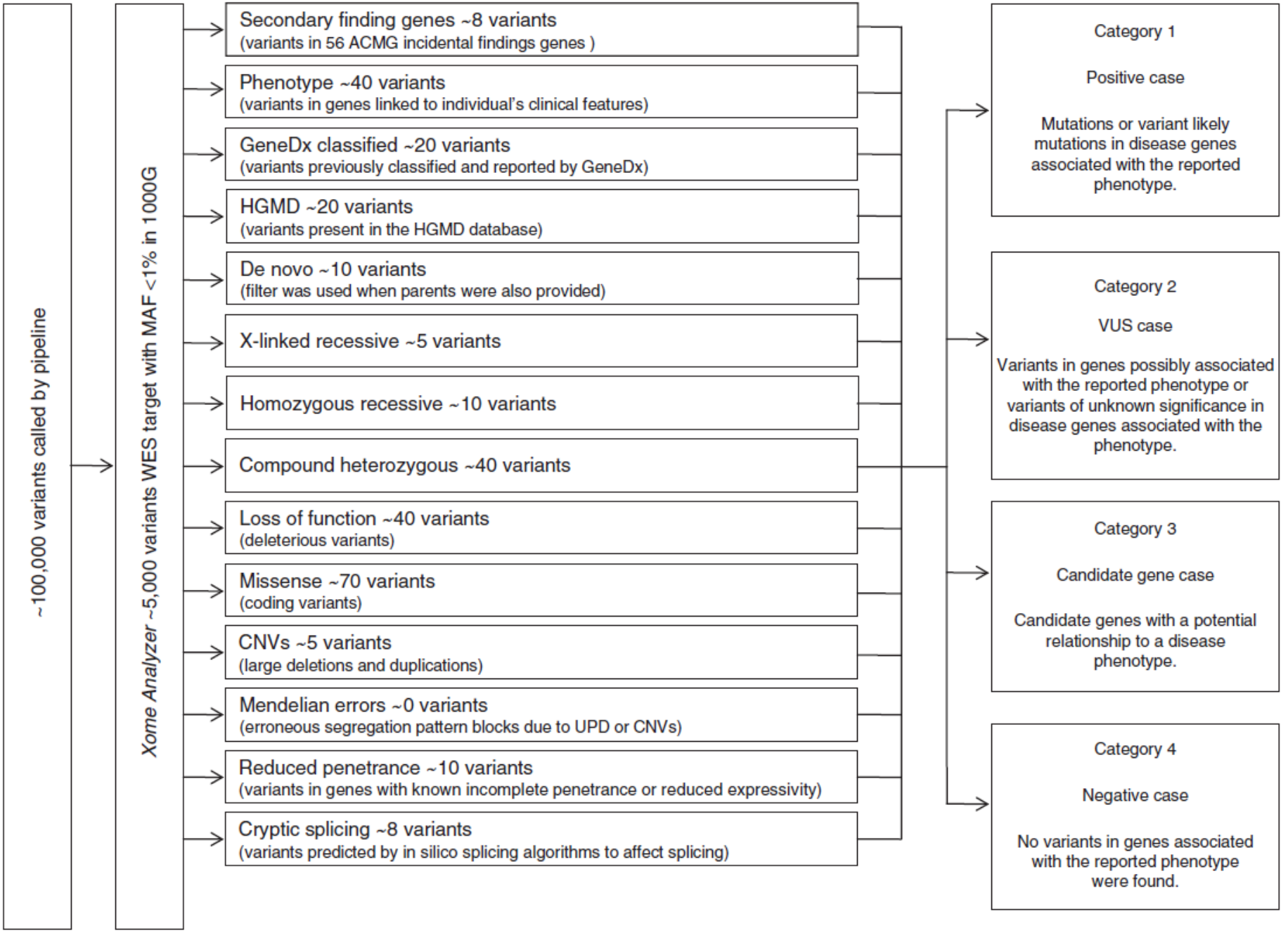
Variant filtering and prioritization: 20,000–100,000 → 50–1,000

- Known pathogenic variants
- Rare (MAF<0.5%) or novel PTVs
- Other variants/genes with associated phenotypes (ClinVar, OMIM, HPO)

Output

- Clinical report with diagnosis, candidate gene/variant
- Referrals
- Sanger sequencing requested

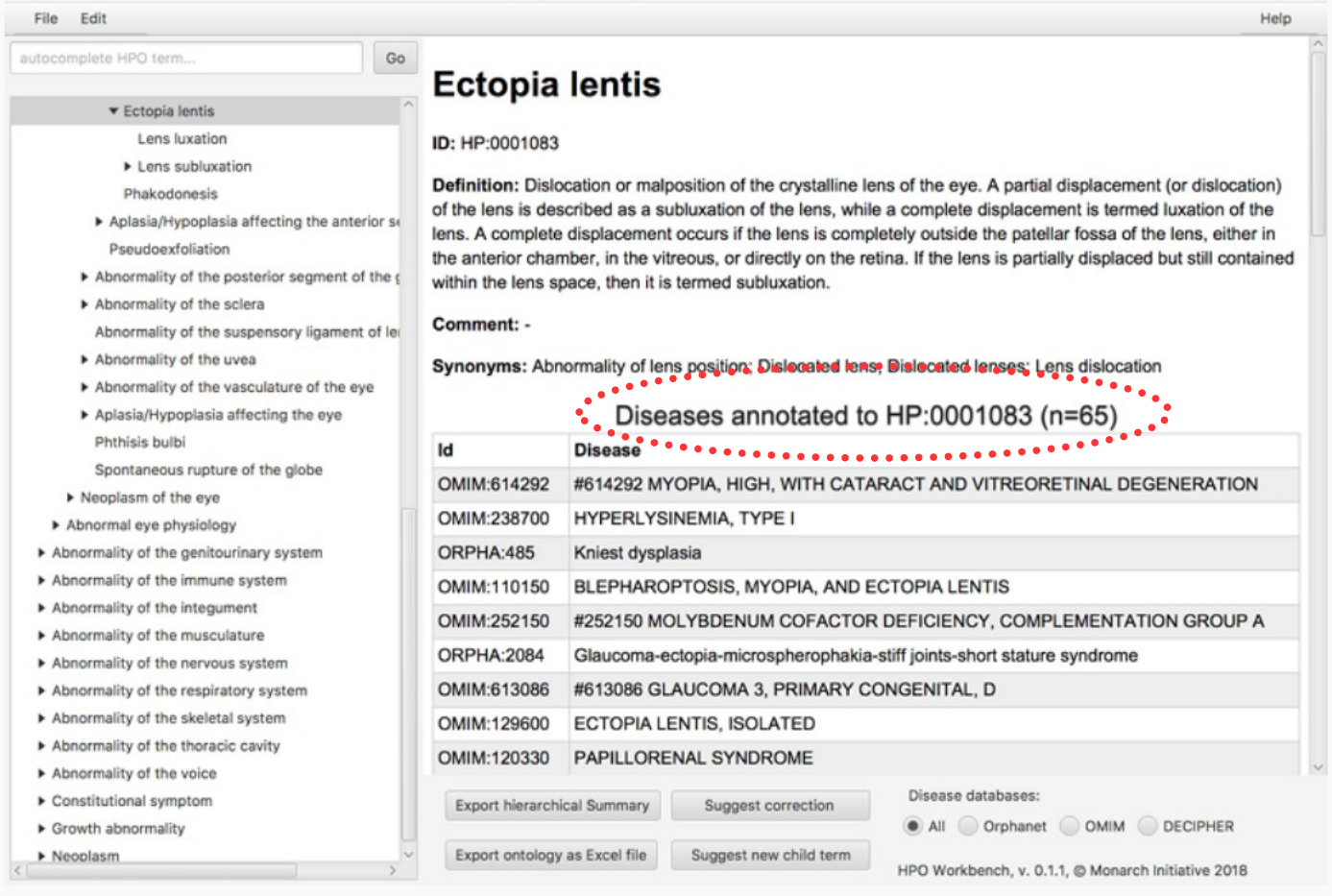
WES diagnostics of Mendelian disorders



Retterer (2016) *Genet Med*

Human Phenotype Ontology

- The Human Phenotype Ontology (HPO) project provides an ontology of **medically relevant phenotypes, disease-phenotype annotations**, and the algorithms
- The HPO can be used to support differential diagnostics, translational research, and … **the means to compute** over the clinical phenotype
- The HPO currently contains **over 13,000 terms**. All relationships in the HPO are simple class-subclass relationships



Human Phenotype Ontology

Seizure HP:0001250

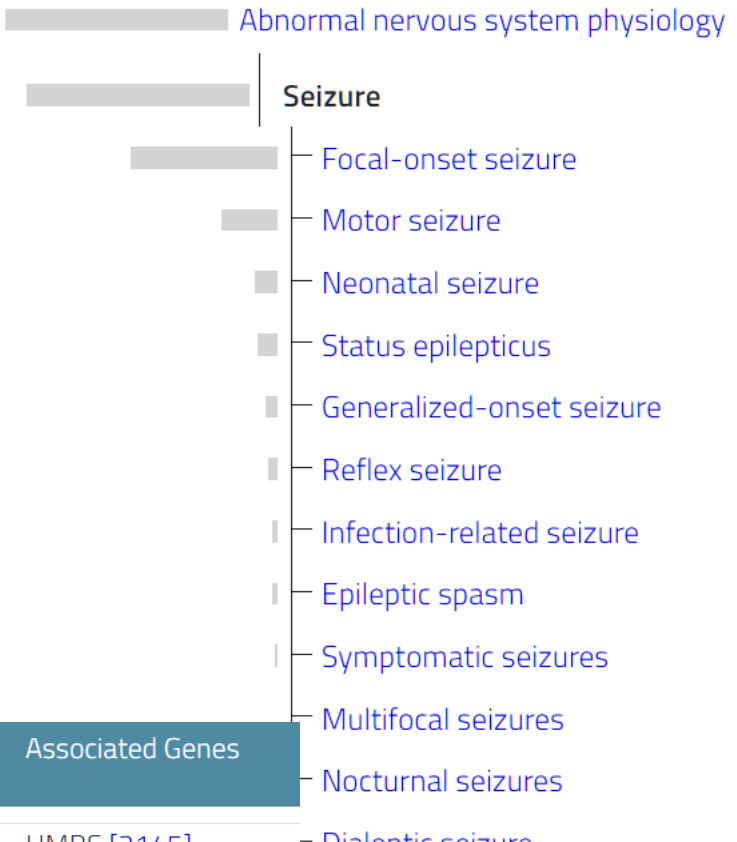
A seizure is an intermittent abnormality of nervous system physiology characterised by a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain.

Synonyms: *Epileptic seizure, Epilepsy, Seizures*

Comment: A type of electrographic seizure has been proposed in neonates which does not have a clinical correlate, it is electrographic only. The term epilepsy is not used to describe recurrent febrile seizures. Epilepsy presumably reflects an abnormally reduced seizure threshold.

Pubmed References: *PMID:15816939*

Cross References: *SNOMEDCT_US:246545002, UMLS:C0036572, SNOMEDCT_US:128613002, SNOMEDCT_US:313307000, SNOMEDCT_US:84757009, UMLS:C0014544, SNOMEDCT_US:91175000, MSH:D004827, MSH:D012640*



Disease Id	Disease Name	Associated Genes
ORPHA:79276	Acute Intermittent Porphyria	HMBS [3145]
OMIM:201450	Acyl-coa Dehydrogenase, Medium-chain, Deficiency Of	ACADM [34]
OMIM:201470	Acyl-coa Dehydrogenase, Short-chain, Deficiency Of	ACADS [35]
ORPHA:974	Adams-oliver Syndrome	ARHGAP31 [57514] DOCK6 [57572] RBPJ [3516] DLL4 [54567]

Displaying 20 out of 2335. View all

WES diagnostics of Mendelian disorders

Examples:

1. A novel homozygous variant (Asp652Asn) in solute carrier family 26, member 3 *SLC26A3* — a gene that is known to cause a **congenital chloride-losing diarrhea** — was identified in a child originally suspected to have a different diagnosis of Bartter syndrome.

2. A novel Cys203Tyr variant in X□linked inhibitor of apoptosis (*XIAP*) in a young boy with severe inflammatory bowel disease in whom a definitive diagnosis was elusive. Mutations in *XIAP* are a known cause of **X□linked lymphoproliferative syndrome type 2** (XLP2), but severe colitis is an unusual symptom of XLP2. The diagnosis of XLP2 suggested a specific course of treatment.

Bamshad (2011) *Nat Rev Genet*

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.

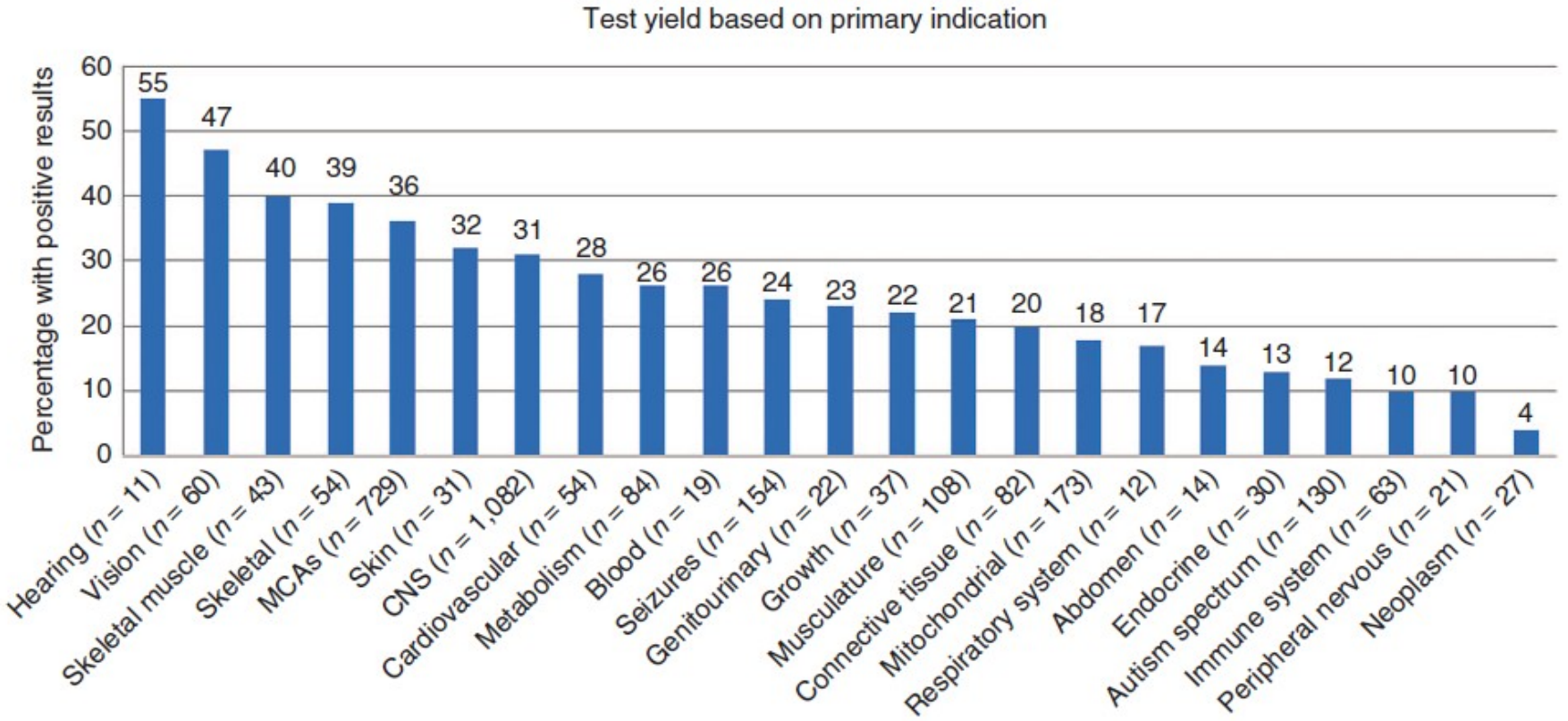
Results: The overall diagnostic yield of WES was 28.8%. The diagnostic yield was 23.6% in proband-only cases and 31.0% when three family members were analyzed. The highest yield was for patients who had disorders involving hearing (55%, $N = 11$), vision (47%, $N = 60$), the skeletal muscle system (40%, $N = 43$), 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

WES diagnostics of Mendelian disorders

- The overall diagnostic yield of WES was 28.8% in **3,040 cases**; 23.6% in proband-only cases and 31.0% when three family members were analyzed
- In 24.2% of cases a candidate gene was reported that may later be reclassified as being associated with a definitive diagnosis
- Of 2,091 cases in which secondary findings were analyzed for 56 ACMG–recommended genes, 6.2% (N = 129) had reportable pathogenic variants



Retterer (2016) *Genet Med*

ACMG-56 2.0: secondary findings genes

- **ACMG-56**: a list of genes to be reported as incidental or secondary findings
- **The goal**: to identify and manage risks for selected highly penetrant genetic disorders through established interventions aimed at preventing or significantly reducing morbidity and mortality.
- **Updates:** 2013: started; 2017: -1 gene, +4 genes

- **Example:** *ATP7B* is associated with autosomal-recessive Wilson disease (MIM 277900). Morbidity among homozygotes directly correlates with copper deposition in the liver, brain, and eye. The disease is progressive, and, if left untreated, premature death is likely. In some cases, liver failure may be the presenting sign. <...> Treatment for Wilson disease involves administration of copper chelating agents and/or zinc to block intestinal absorption of copper; **treatment is extremely effective when administered prior to the onset of symptoms**.

Kalia (2017) *Genetics in Medicine*

ACMG-56 2.0: secondary findings genes

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)	20301715	Child/adult	*SDHD* *SDHAF2* *SDHC*	602690 613019 602413	AD	KP and EP KP KP and EP

Exercise: give an example of ACMG-56 gene and its pathogenic mutation

Kalia (2017) *Genetics in Medicine*

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**

Shah (2018) *Am J Hum Genet*

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**

Summary

- Mendelian (monogenic) disorders depend on the genotype at a single locus, with inheritance following Mendel's laws of segregation
- ...However, this is rather an exception than a rule, because of many complications
- Familial aggregation and descriptive epidemiology help establish the genetic basis of a disease
- Major Mendelian disease inheritance patterns: autosomal dominant, autosomal recessive, X-linked recessive, X-linked dominant, Y-linked
- OMIM and ClinVar are invaluable sources of information on Mendelian diseases
- Penetrance, relative risk and odds ratio measures related yet different aspects of disease risks. Variant penetrance are often unknown or inflated!
- Disease gene discovery has been dramatically transformed and accelerated by next-generation sequencing
- There are emerging standards and guidelines in the field: from variant interpretation to secondary findings reporting

Further reading

- 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.

Further reading

- 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