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

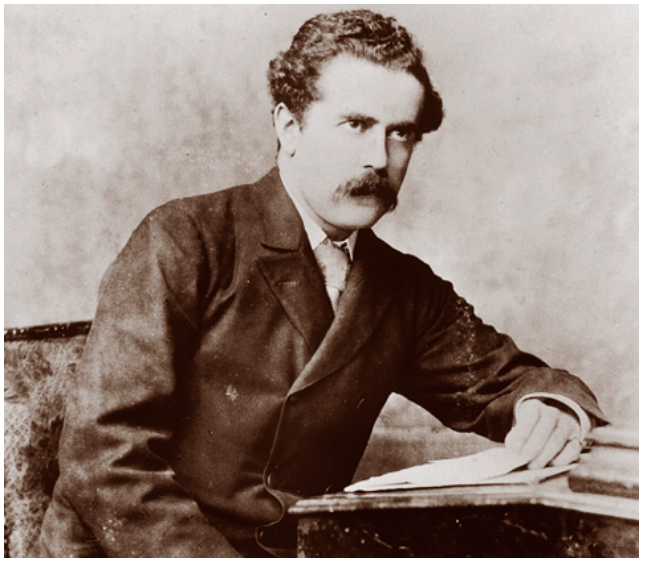
Disease, syndrome and other definitions

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



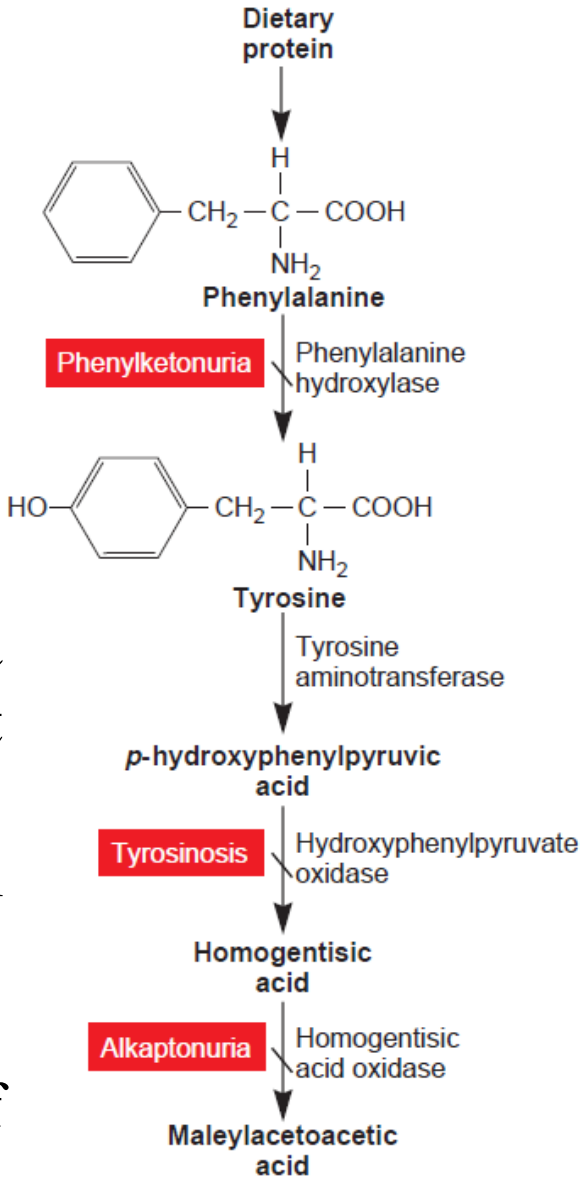
Reprinted from Lancet, vol. ii, 1902, pp. 1616-1620.

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.

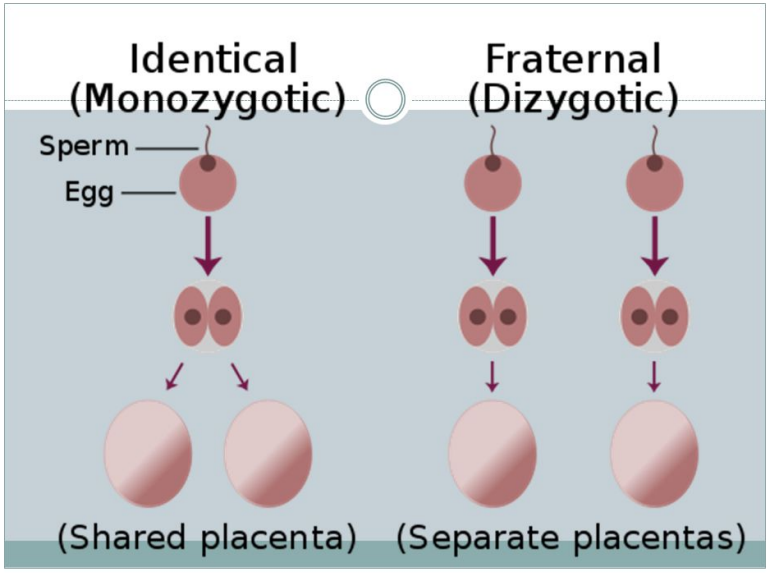
Sir Archibald Garrod (1902): mutation → loss of enzyme → inborn error of metabolism

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

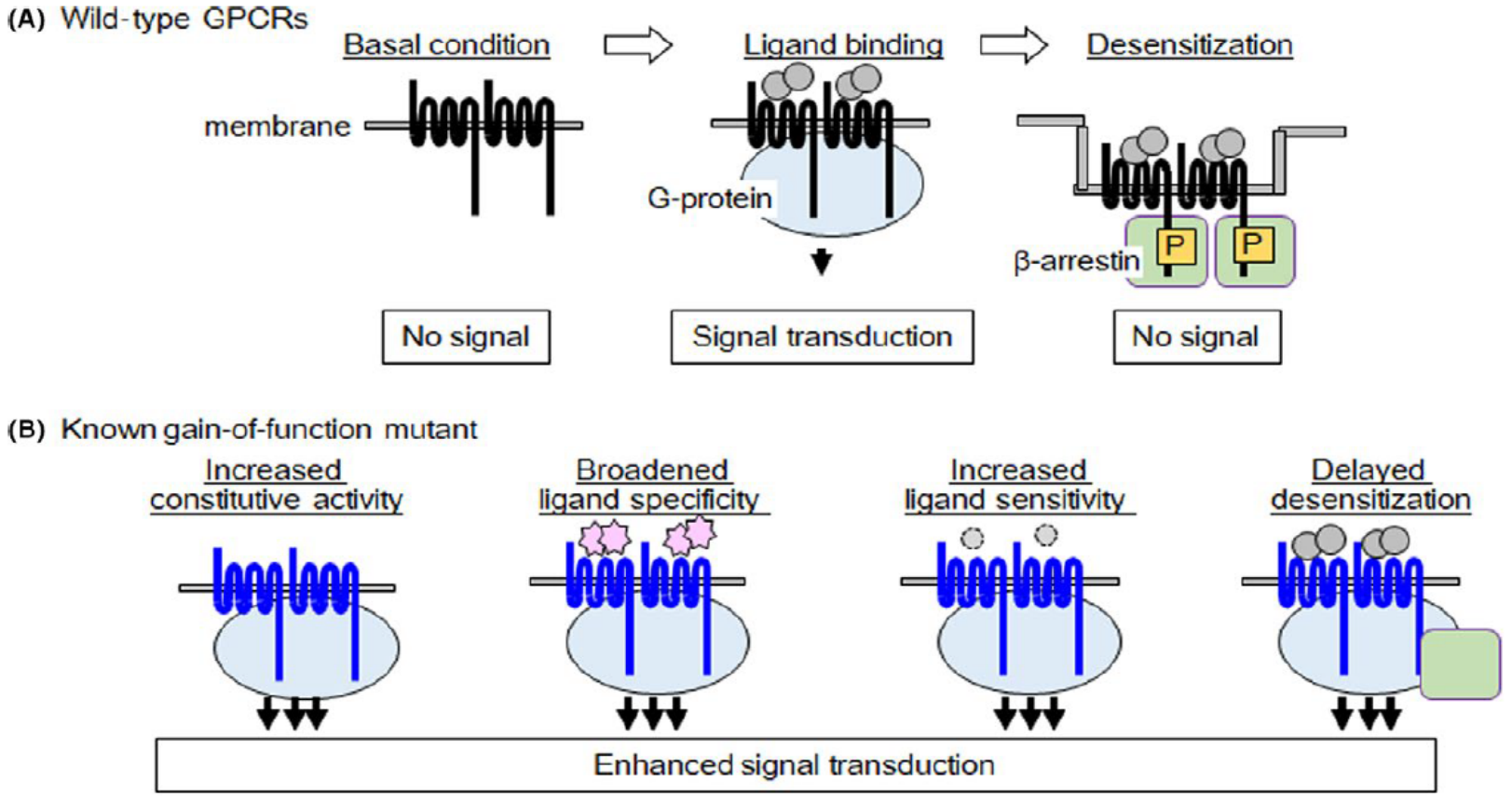
Gain-of-function: the product does "something positively abnormal"

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

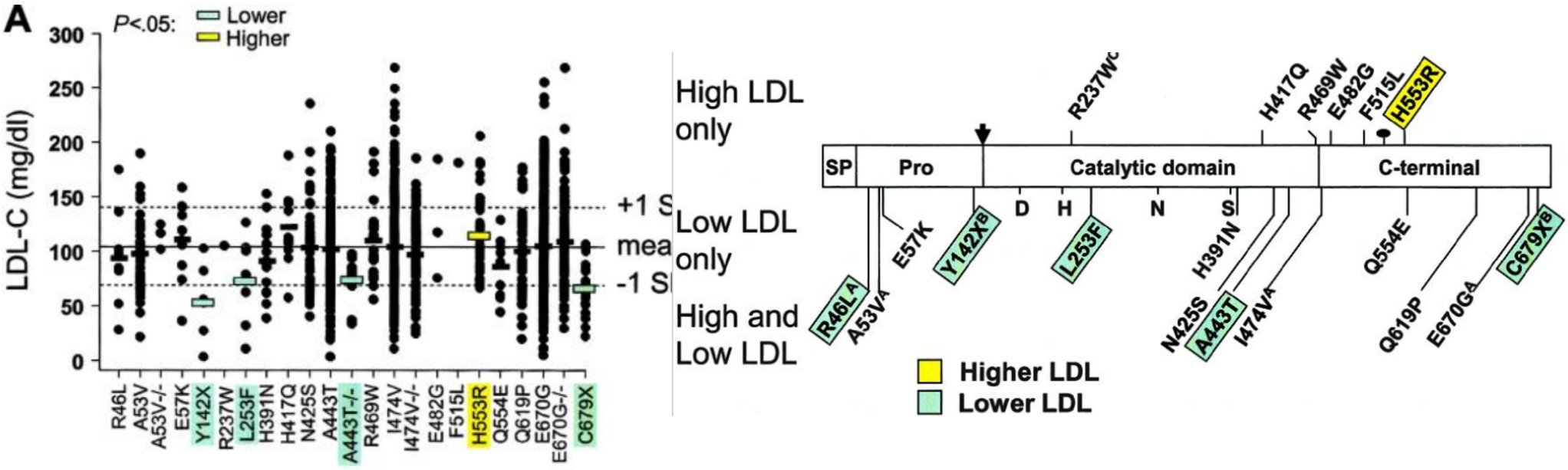


Fukami (2018) *Clin Endocrinol*

Variant effect: gain- and loss-of-function

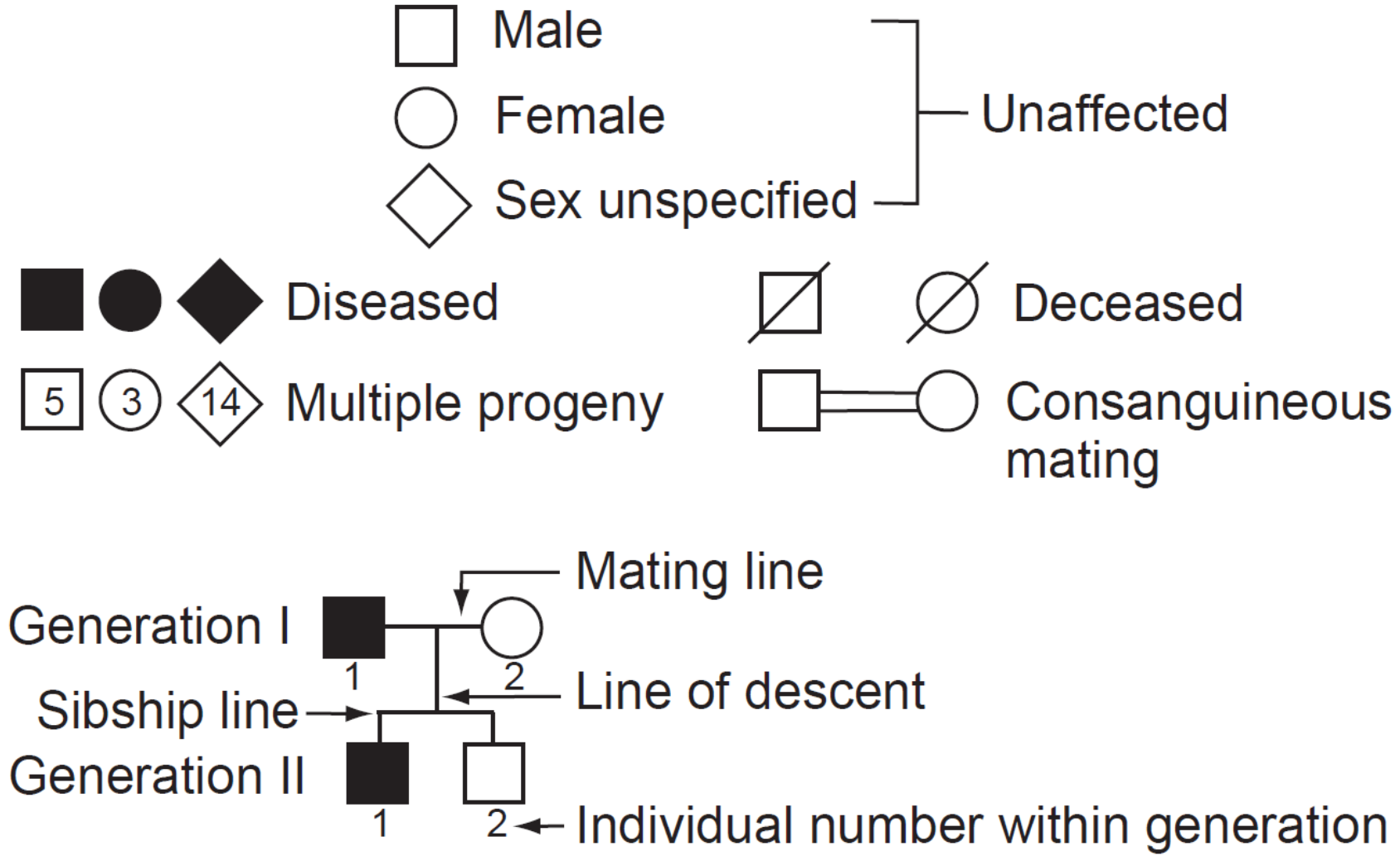
Serine protease *PCSK9* (Proprotein convertase subtilisin/kexin type 9) regulates low density lipoprotein cholesterol (LDL-C) levels, has both types of variants

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



Kotowski (2006) *Am J Hum Genet*

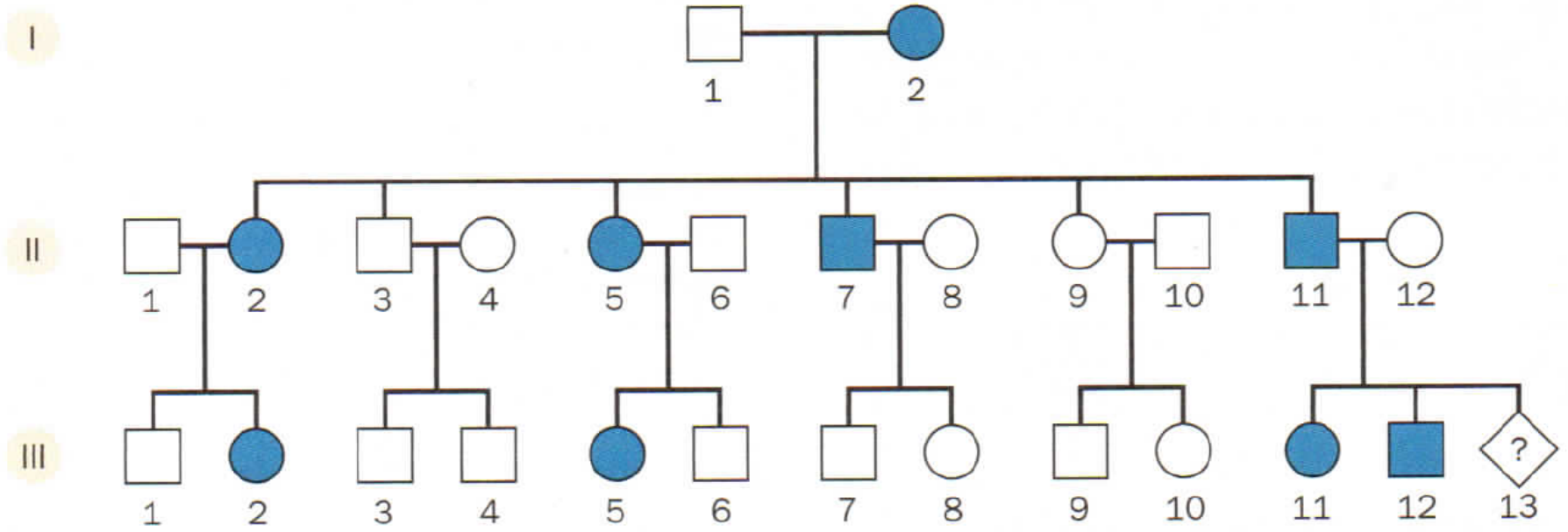
Symbols used in pedigree analysis



Hartwell – *Genetics. From genes to genomes*

Autosomal dominant inheritance

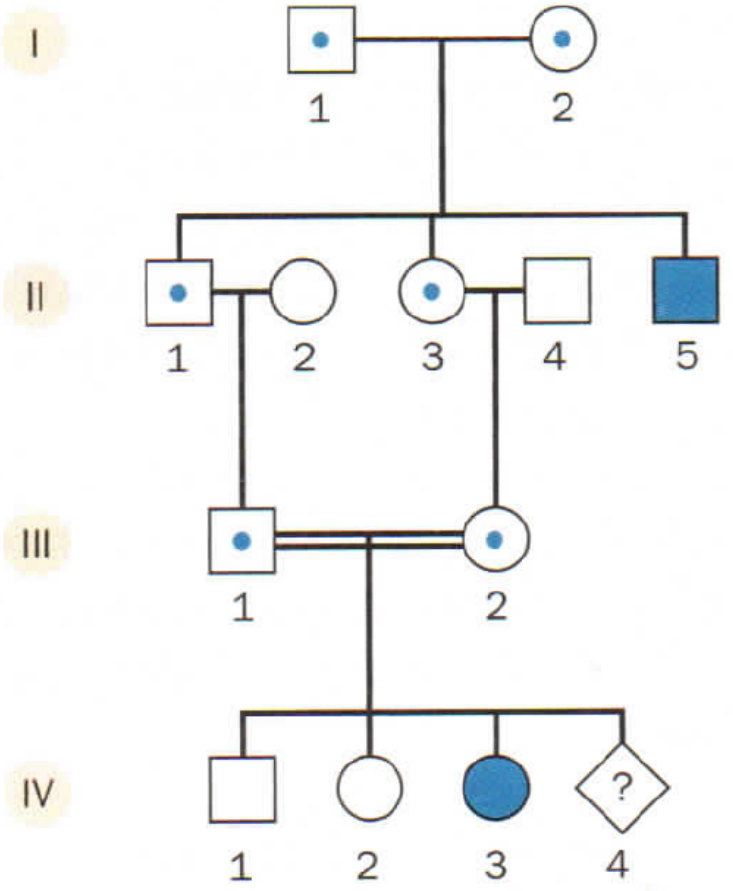
- An affected person (proband) usually has at least one affected parent
- It affects either sex
- A child with one affected and one unaffected parent has a 50% chance of being affected
- Causal variant is gain-of-function or loss-of-function if gene is haploinsufficient; often, *de novo*



Strachan, Read – *Human Molecular Genetics*

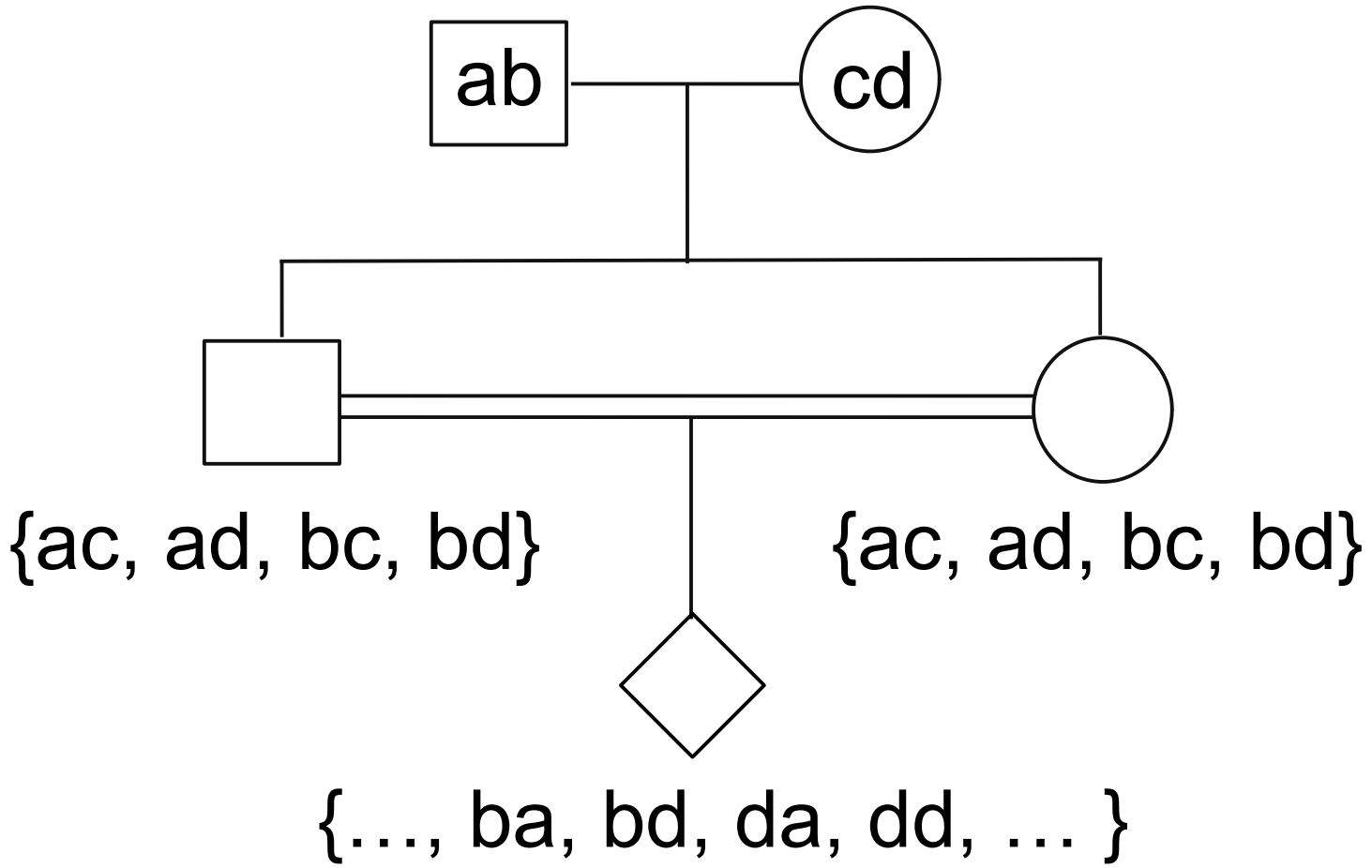
Autosomal recessive inheritance

- 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



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)

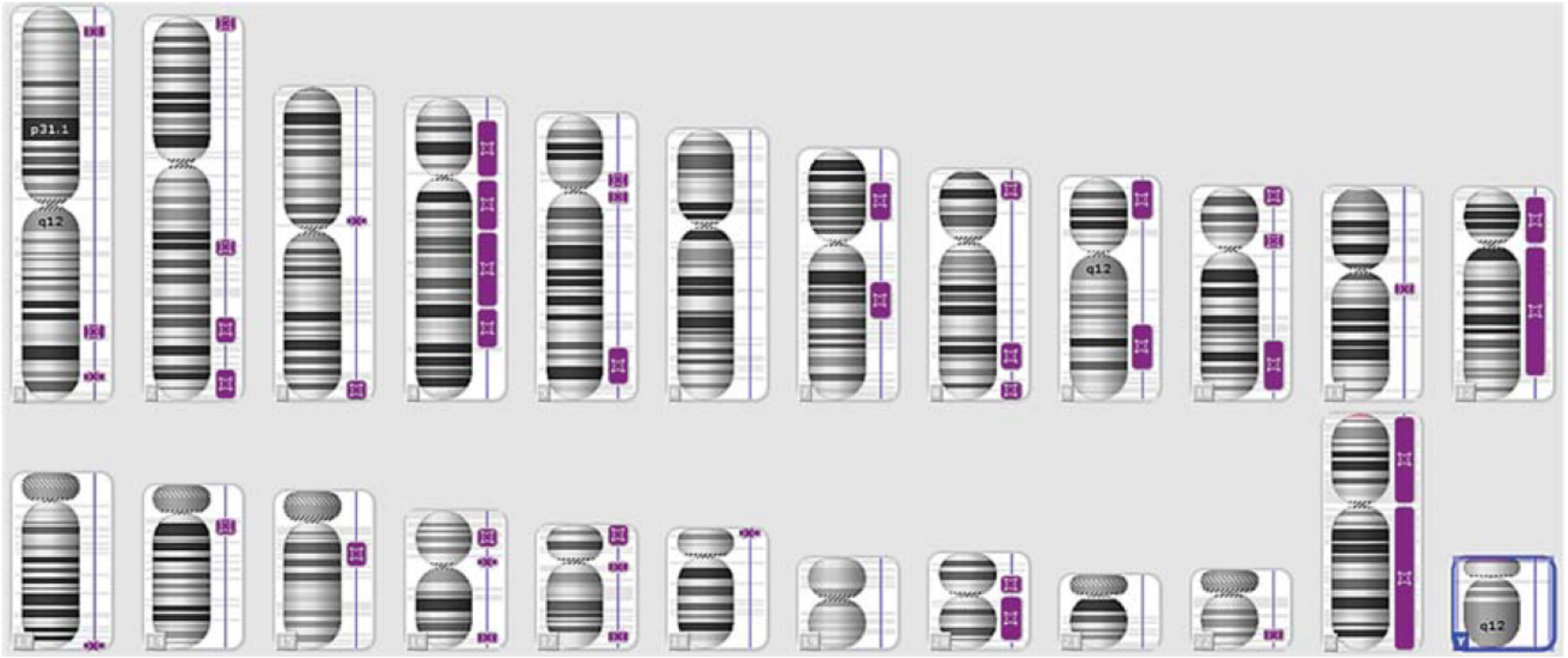


Fig. 1. ROH detected by SNP microarray analysis (Affymetrix Cytoscan HD) in a male child who was the offspring of a brother-sister mating. Each block on the right of the chromosome represents a genomic region at least 3 Mb in size. The laboratory-reported autosomal Froh was >21%.

Sund & Rehder (2014) *Hum Hered*

Variant effect: recessive and dominant

Dominant: effect observed both in homozygotes and heterozygotes

Variant frequency ~ disease incidence; transmitted or *de novo*

Examples:

-- Trp2332Ter in *CHD7*, CHARGE syndrome

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

Recessive: effect observed in homozygotes only

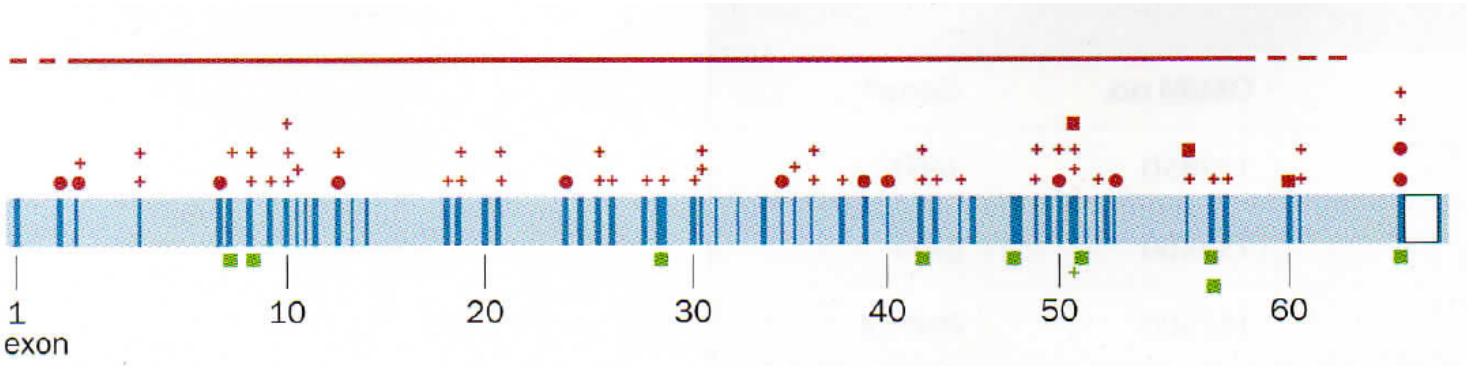
Variant frequency >> disease incidence; transmitted from both parents

Examples:

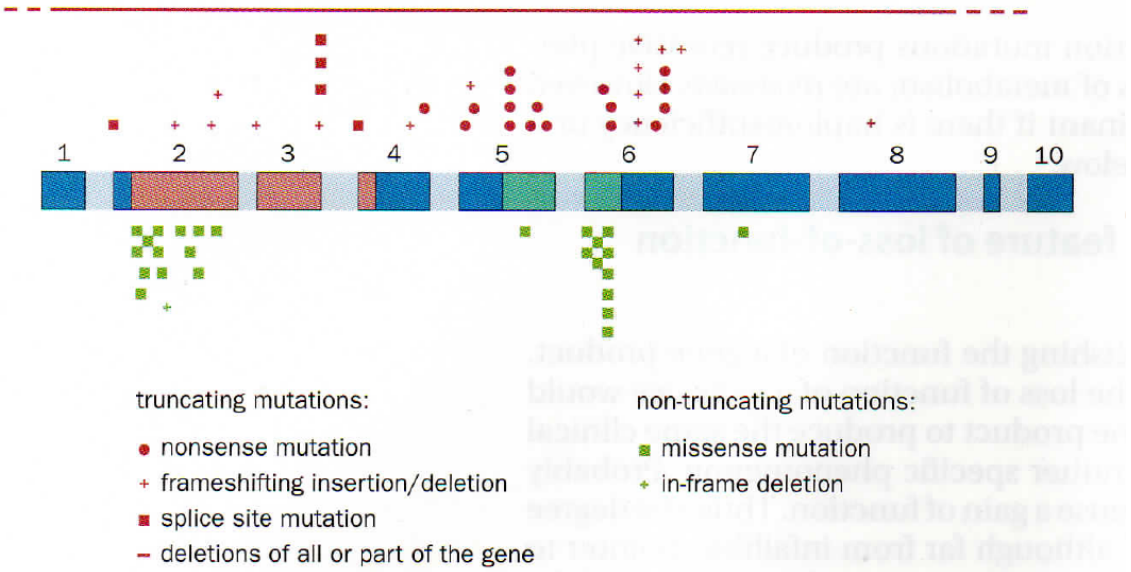
-- Ex24:p.F508del in *CFTR*, cystic fibrosis

-- Ex2:c.35delG in *GJB2*, hearing loss

Variant effect: recessive and dominant



Recessive: *ATM* gene, patients with ataxia-telangiectasia (OMIM:208900, aka Louis-Bar syndrome)



Dominant: *PAX3* gene, patients with Waardenburg syndrome type I (OMIM:193500, hearing loss and pigmentary abnormalities)

Strachan, Read – *Human Molecular Genetics*

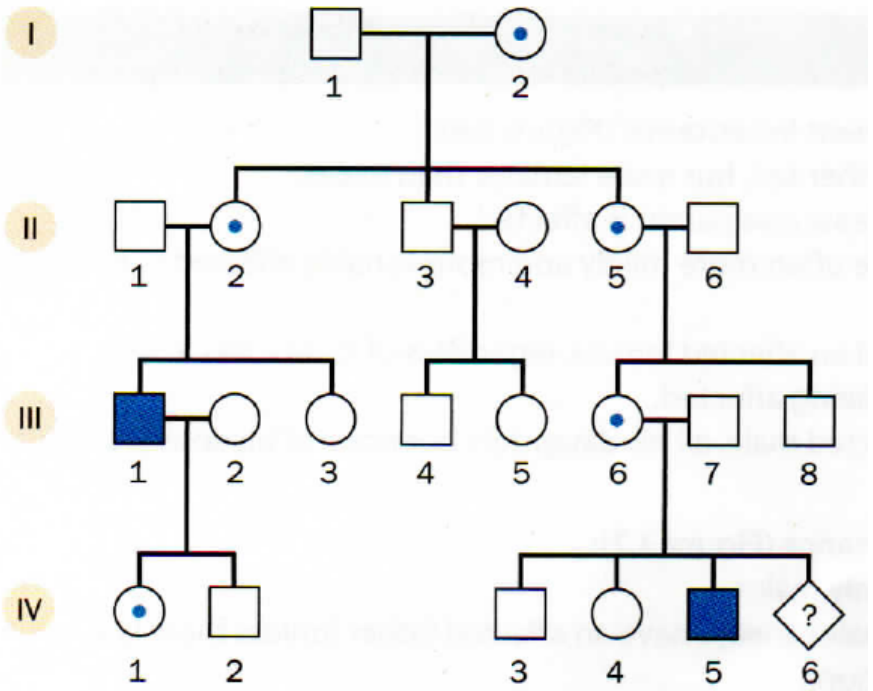
X-linked recessive inheritance

- It affects mainly males
- Affected males are usually born to unaffected parents
- The 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.

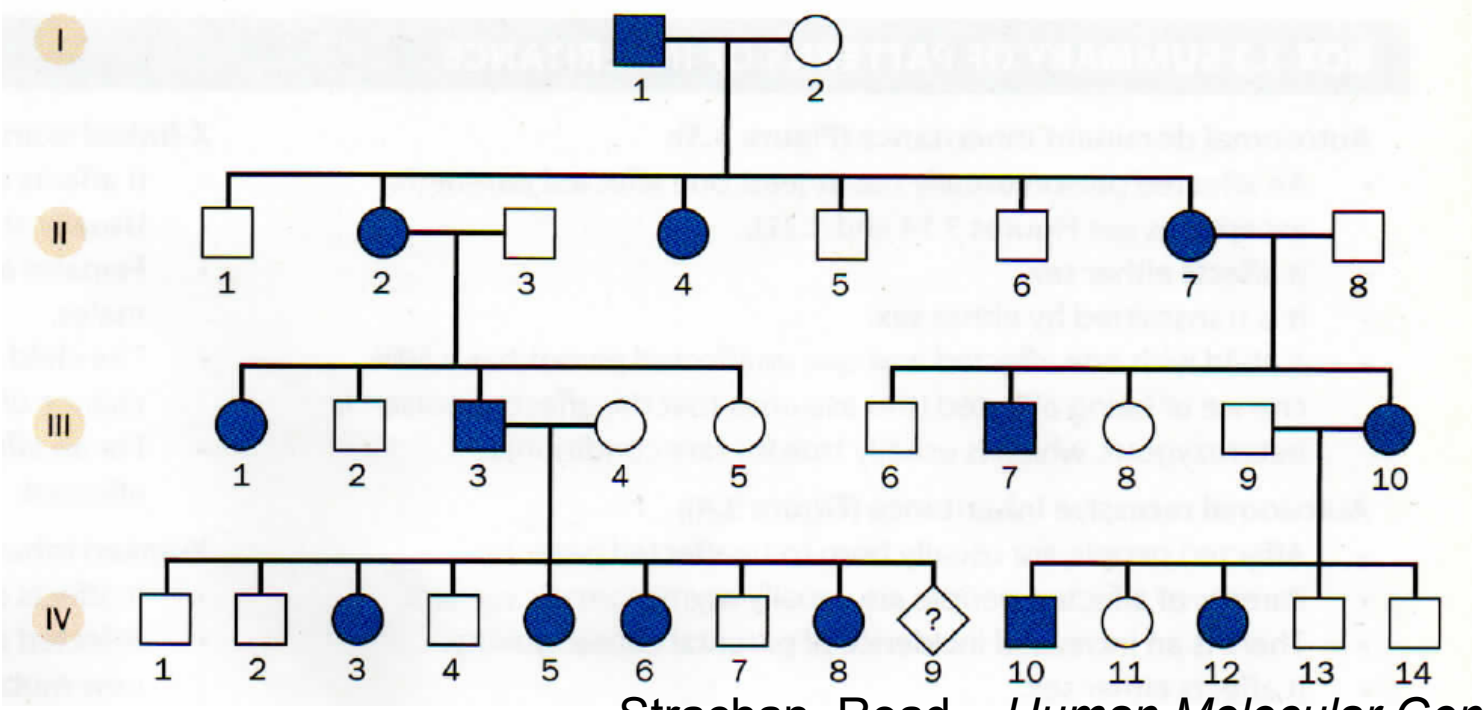
- There is no male-to-male transmission in the pedigree



Strachan, Read – *Human Molecular Genetics*

X-linked dominant inheritance

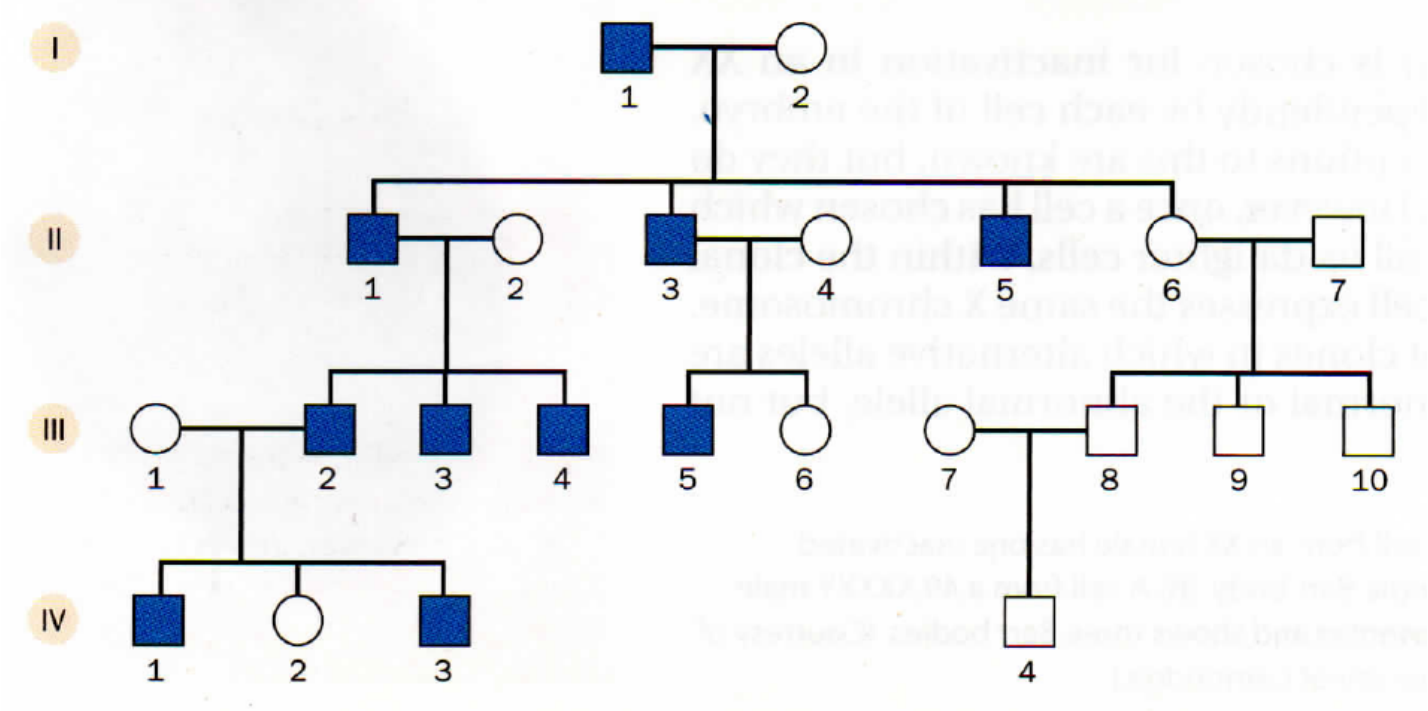
- It affects either sex, but more females than males *Q: why?*
- Usually at least one parent is affected
- Females are often more mildly and more variably affected than males *Q: why?*
- The child of an affected female, regardless of its sex, has a 50% chance of being affected.
- For an affected male,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



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.

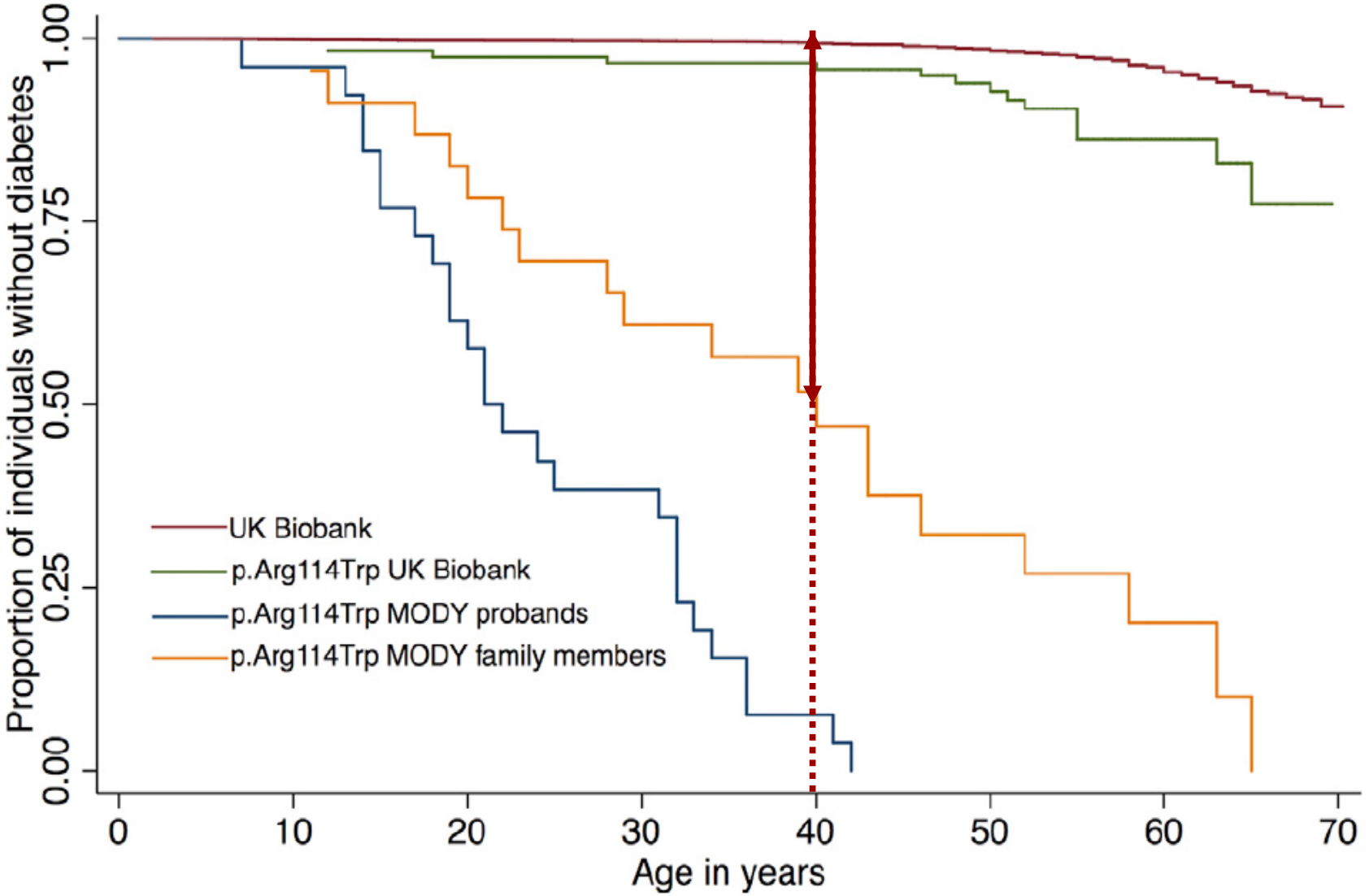
Examples:

– 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