How many disease-causing variants in a normal person?

Matthew Hurles

Summary

- **What is in a genome?**
- **What is ‘normal’?**
  - **Depends on age**
- **What is a ‘disease-causing’ variant?**
  - **Different classes of variation**
- **Final thoughts**
  - **Comparing screening and diagnostic testing**

The average genome

Types of genetic variation

GGAGAGTCGTGA

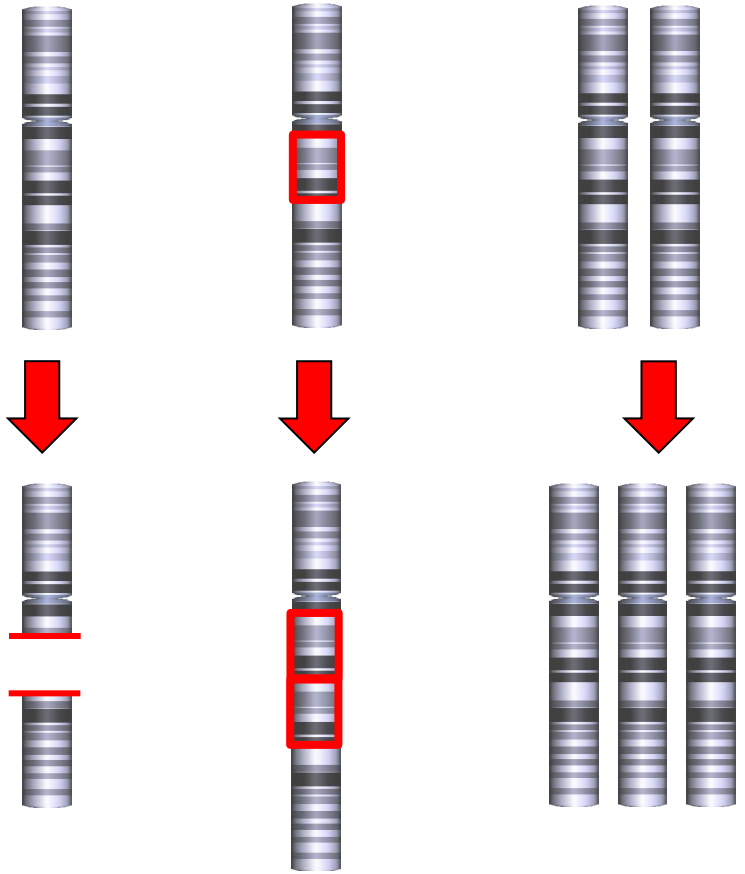
GGAGATTCGTGA

GGAGAGTCGTGA

GGAG----GTGA

OR

GGAGAGGGTCGTGA



Disease-causing variation

Breakdown of disease-causing variation in the Human Gene Mutation Database (HGMD):

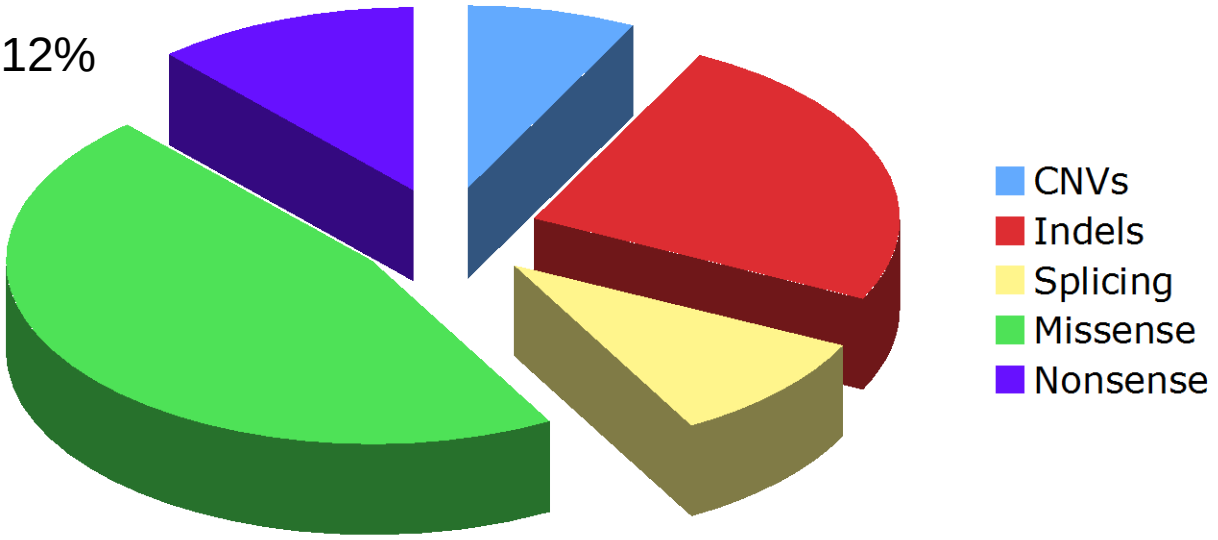
CNVs 8% (likely underascertained)

Indels 25%

Splicing 10%

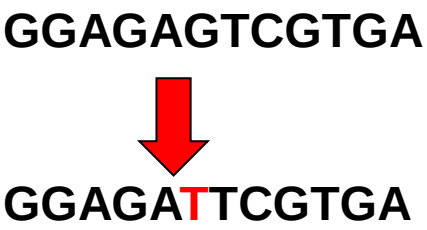
Missense 46%

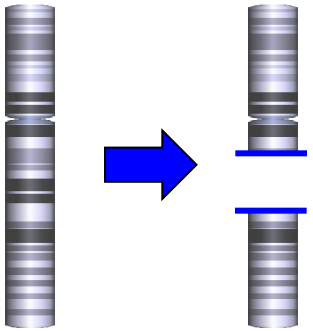
Nonsense 12%



An ‘average’ genome

- **~4,000,000 variants per genome**
  - **>95% in >1% of the population, <200,000 rare**
- **~10,000 protein-altering variants**
  - **>95% in >1% of the population, <500 rare**
- **40-100 new mutations**
  - **90% are single base changes**
  - **1 every 50 million bases**
- **1 in 20 people have a new**





What is ‘normal’?

Defining normal

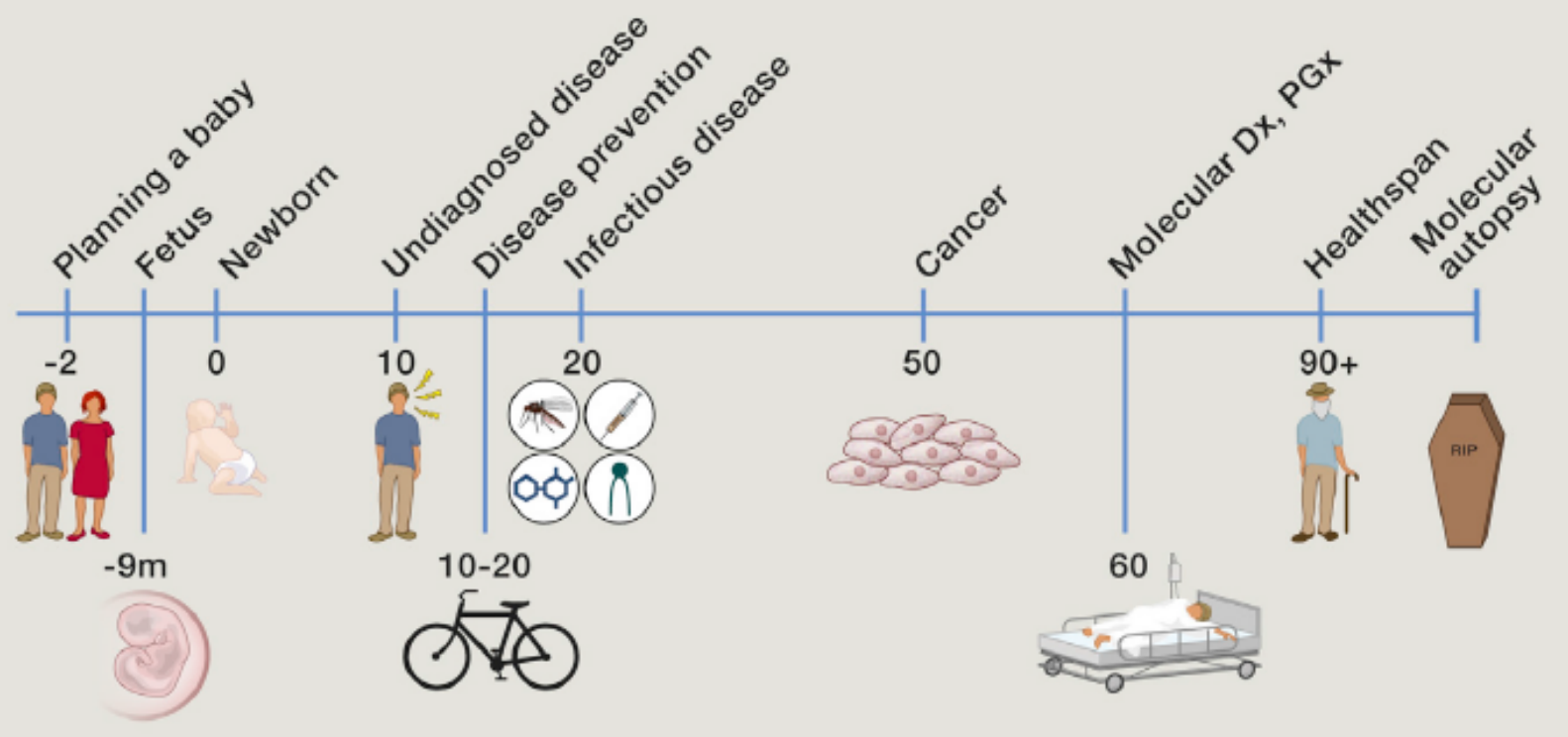
Normal = Not Abnormal

Abnormal = suspected of genetic disease

Abnormal = access to genetic diagnostic testing

Depends on age of onset of genetic disease

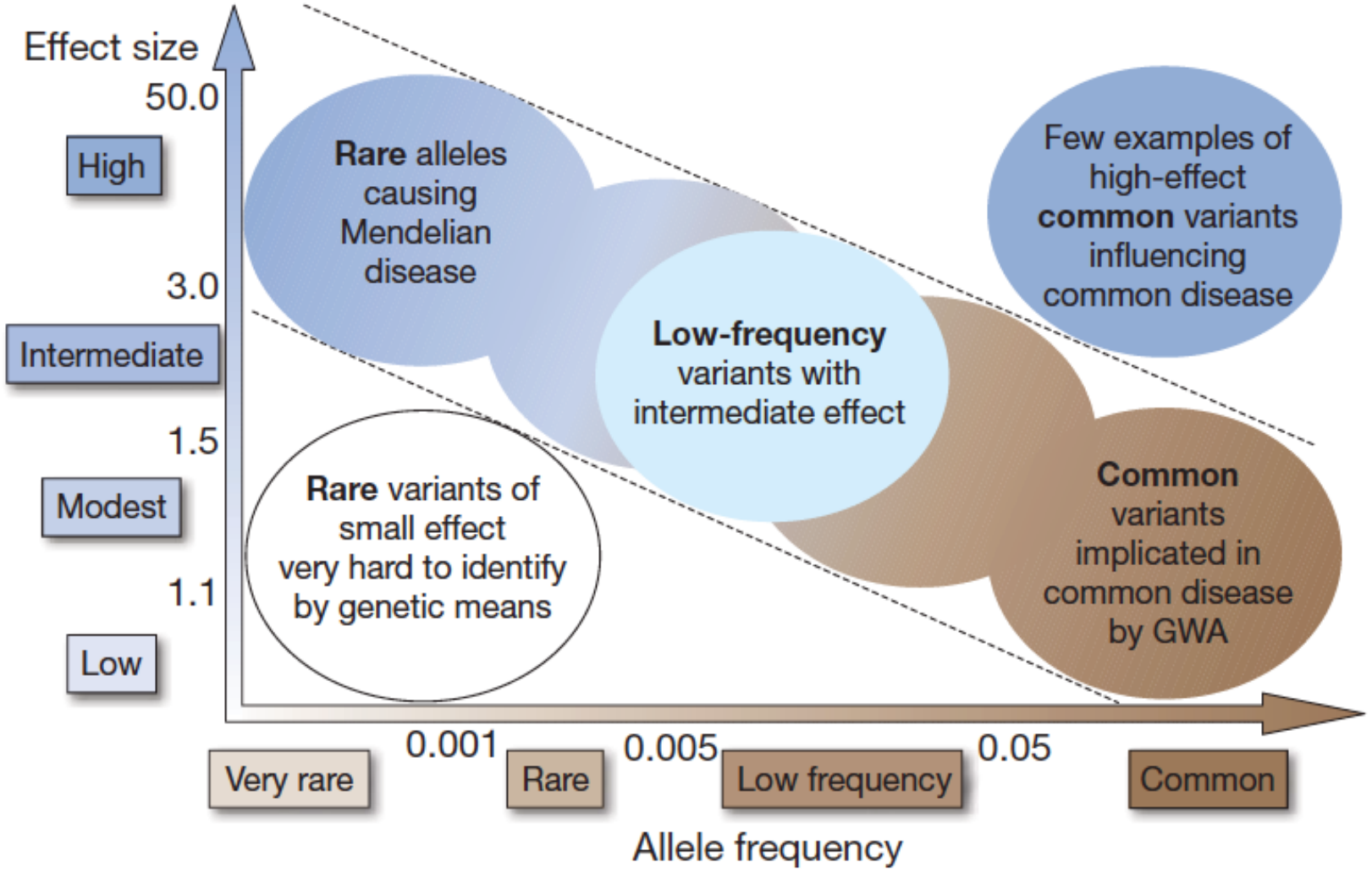
Genetics: from womb to tomb



Source: Topol, *Cell 157*, 2014

What is ‘disease-causing’ variation?

Spectrum of disease-associated variation

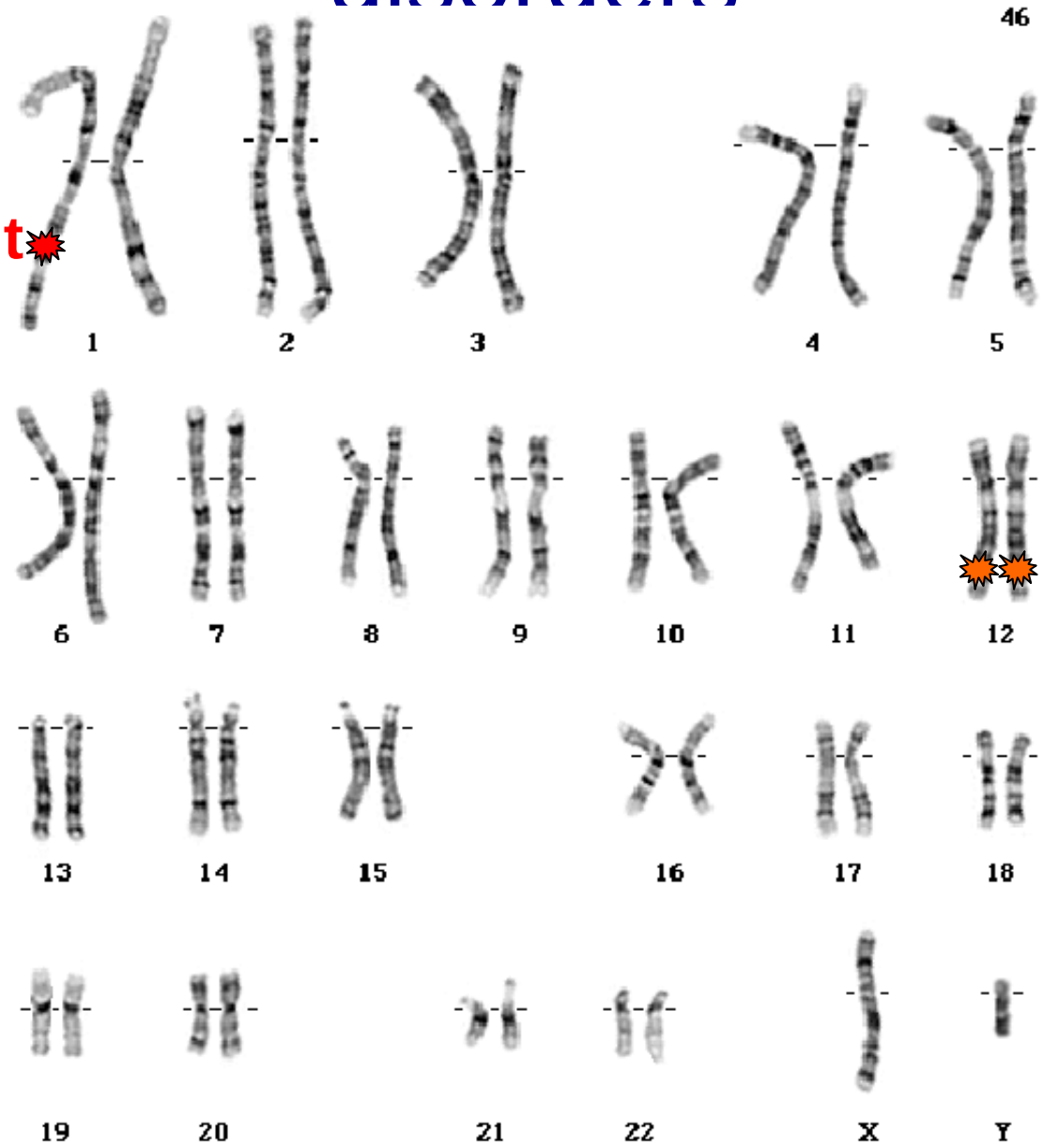


Rare (Mendelian) diseases

- **Affect 1 in 17 (~6%)**
- **>7,000 recognised, likely many more**
- **Most have a genetic component, most childhood onset**
- **Rare genetic diseases are caused by rare genetic variants**
- **Variants in 3,100 (~15%) genes cause >5,000 diseases**
- **Most mutations in the protein-encoding**

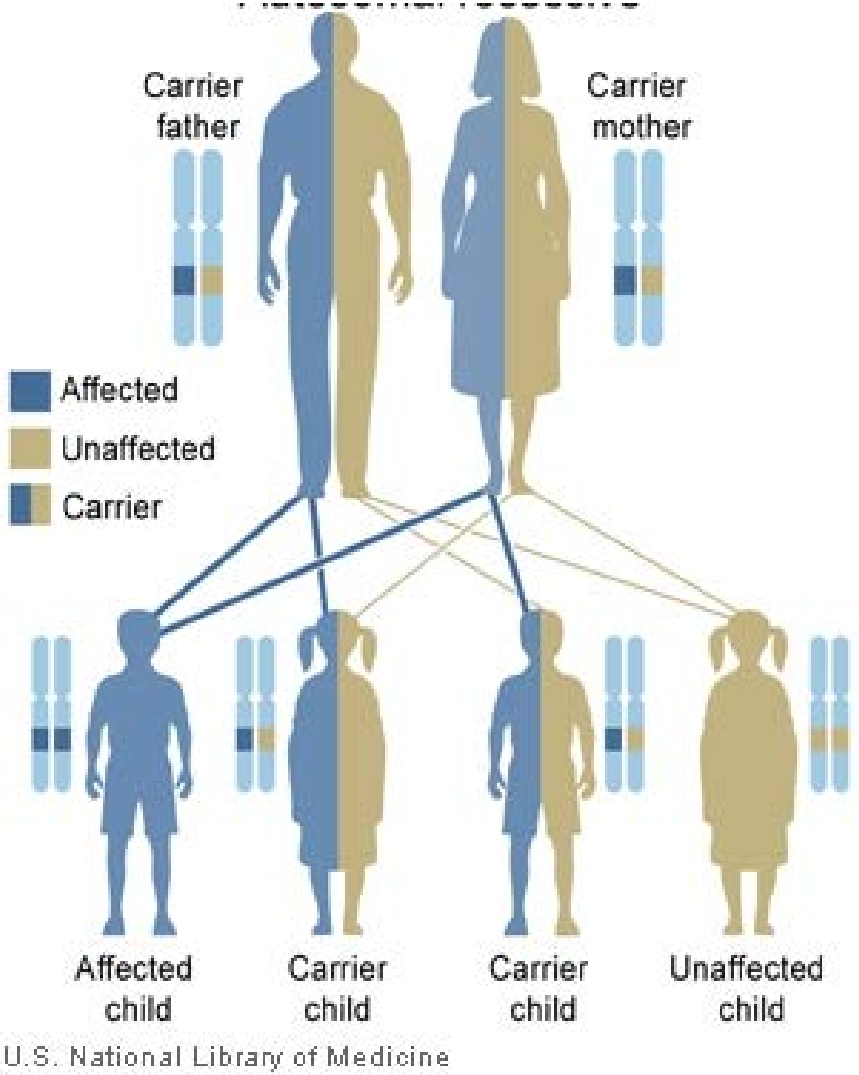
Two types of rare genetic disorders

Dominant

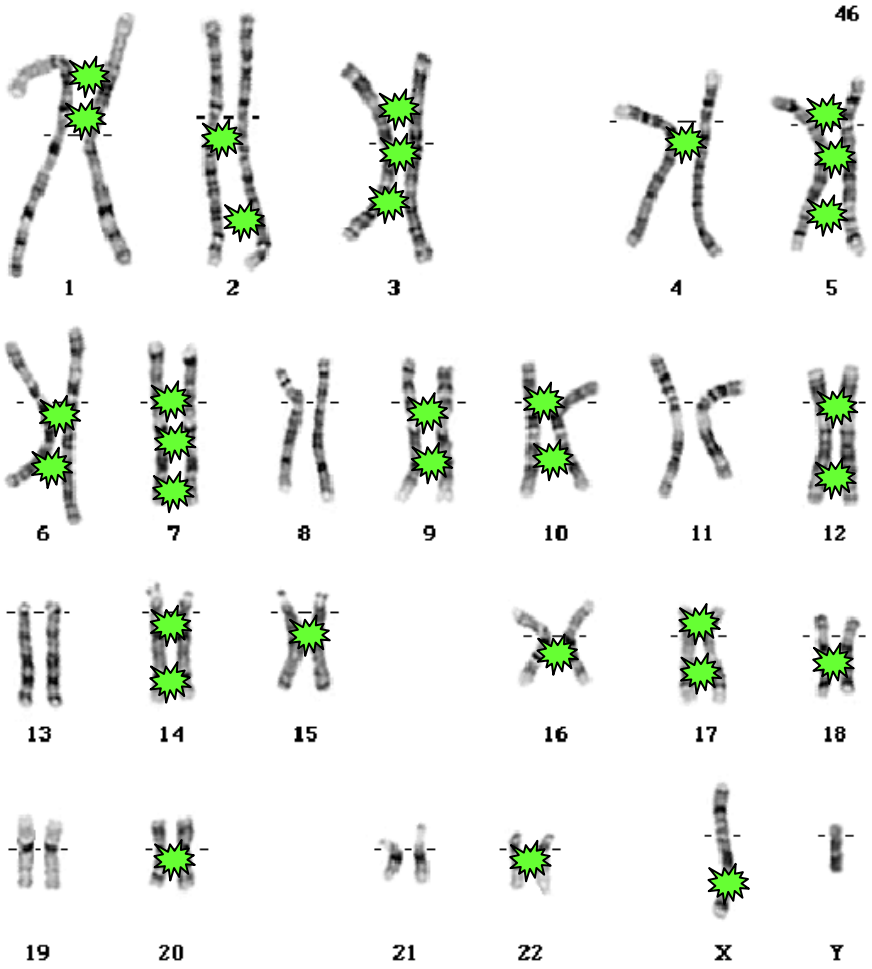


Recessive

Recessive disorders



Genetics of common diseases



+

Rare and common diseases

- **Most known variants associated with monogenic diseases are highly penetrant (OR>10) rare variants in protein-coding sequences (N=100,000's described)**

- **Most known variants associated with complex diseases are weakly associated (OR<1.5) common variants outside of protein-coding sequences (N=~1,000's described)**

Different classes

- **Rare variants causing early-onset rare genetic disease**
- **Rare variants causing late-onset rare genetic disease**
- **Rare variants potentially causing disease in offspring**
  - **E.g. Recessive carrier status**
- **Cumulative common variants influencing disease susceptibility**
- **Variants influencing drug response**

Rare variants causing early-onset rare genetic disease

- **Perhaps 3-5% of individuals suffer from an early-onset rare genetic disease**
- **By late childhood, most of these diseases will be apparent**
- **Every individual has ~50 rare, protein-altering variants in the ~2,500 genes known to cause early-onset rare genetic disease**

  ~10-20 of these variants have genotype compatible with disease
- **>99% of rare, protein-altering variants in 'normal individuals' in these genes DO NOT cause early-onset rare genetic disease**

Rare variants causing late-onset rare genetic disease

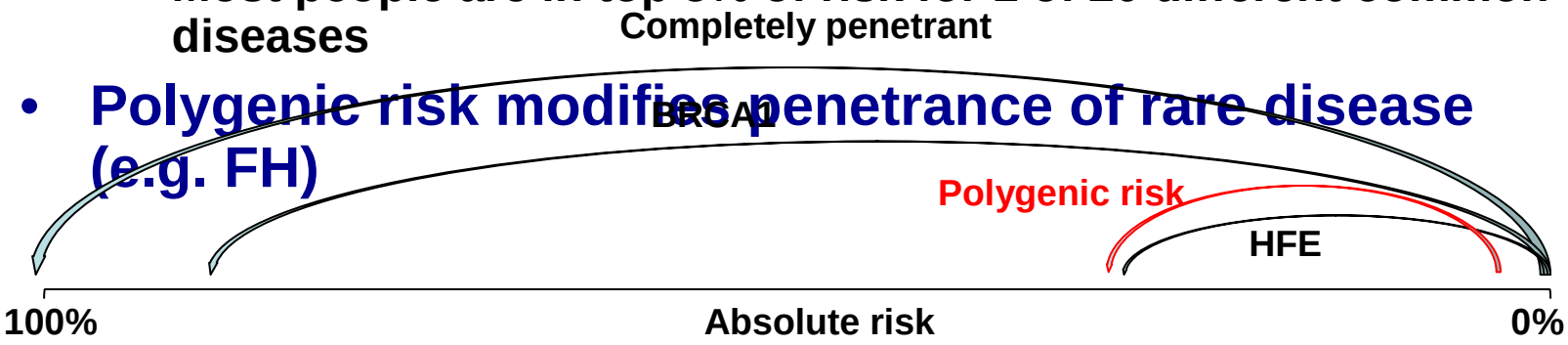
- **Most attention focused on 56 'actionable' genes (ACMG)**
  - **Only 'known pathogenic' and 'expected pathogenic' variants**
  - **Expect ~1% of unselected individuals**
- **Every individual has 0-6 rare, protein-altering variants in these genes (mean of 1-3 variants)**
- **Various studies have estimated the prevalence of disease-causing variants in these 56 genes to be ~2% of individuals (e.g. UK10K 2015)**
- **>90% of rare, protein-altering variants in 'normal individuals' in these genes DO NOT cause late-onset rare genetic disease**

Rare variants causing disease in offspring (e.g. recessive)

- **Risk of disease depends primarily on the intersection of damaging variants carried by mother and father**
- **Every individual has ~30-40 rare, protein altering variants in known recessive genes**
- **Genotyping of known pathogenic variants in 108 disorders in 23,000 individuals found 0.24 variants per individual (Lazarin *et al* 2013)**
  - **~1/500 pregnancies at risk of 88 moderate/severe/profound recessive disorders (1/125 couples, Haque et al ASHG)**
- **Curation of 'known' (HGMD) pathogenic variants in known recessive genes found 5.5 variants per individual (Berg *et al* 2013)**
- **Estimating from excess disease risk from parental relatedness suggests 0.58 variants per individual for post-**

Cumulative common variants influencing common disease susceptibility

- **The combination of tens of risk alleles, each with an odds ratio of ~1.1, can confer a substantial relative risk**
- **Can rank individuals according to their 'polygenic risk'**
- **Currently of little additional clinical predictive value:**
  - **Top 10% have relative risk of 1.4 to 2.2X (Kong *et al* 2014)**
  - **Often ignored by medical genetics community**
- **Potential for greater clinical value:**
  - **5% of population at greatest risk could have relative risk of 3-7X (Wray and Goddard 2010)**
  - **Most people are in top 5% of risk for 1 of 20 different common diseases**
- **Polygenic risk modifies penetrance of rare disease (e.g. FH)**



Variants influencing drug (and dietary) response

- **Example of gene-environment interaction**
- **Variation in 11-20 genes affect 44-80 medications**
  - **~7% of FDA approved drugs**
  - **~18% of prescriptions**
- **Most clinically relevant alleles have been well defined**
- **~40% of individuals have high risk genotypes in 1 of 3 genes (*CYP2D6, CYP2C19, TPTM*)**

Summary table

Class	% of individuals with disease-causing variants	Rare, protein-altering variants per individual	PPV
Early-onset rare disease	3 – 5%	~20	<0.01
Late-onset rare disease	1 – 2%	1-3	<0.01
Carrier frequency	50 – 100%	~35	<0.1
Cumulative polygenic risk	NA	>100	NA
Drug response	40%	0.4	~1

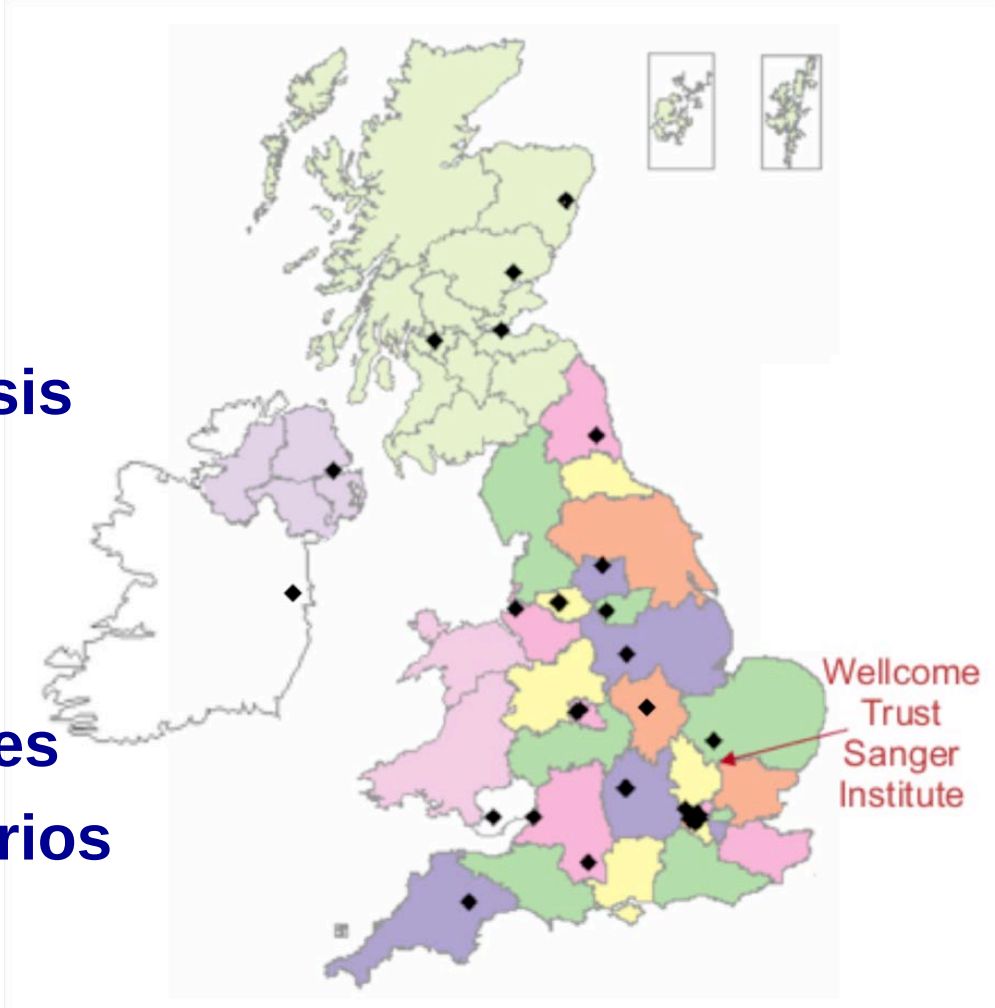
Reasons for low Positive Predictive Value:

- **Inability to distinguish pathogenic and benign variation**
- **Inaccurate annotation of likely molecular impact**
- **Sequencing errors**

Final thoughts

Deciphering Developmental Disorders

- **Collaboration:**
  - **Families + NHS + WTSI**
- **Objectives:**
  - **Understand genetic basis**
  - **Catalyse improved diagnosis**
- **Strategy:**
  - **Recruited 13,958 families**
  - **Exome sequencing of trios**
  - **Systematic clinical phenotyping**
  - **Feedback likely genetic diagnoses**



~20,000 variants in genes

Focus on rare variants predicted to alter proteins

~400 rare & protein-altering

Focus only on 1,200 genes known to cause DDs

10-20 in relevant disease genes

Compare to DNA from (typically unaffected) parents

0-2 with relevant inheritance

Clinical review, phenotype matching

0-2 diagnoses

- Diagnoses for 30-35% of children
- ~80% of diagnostic variants are novel

Final thoughts

- **Sensitivity of 'known' pathogenic variants**
  - **Good for pharmacogenetics, polygenic risk and some recessive disorders**
  - **Currently poor for most dominant disorders and other recessive disorders**

- **Specificity of 'known' pathogenic variants (rare HGMD: ~10 per individual)**
  - **Residual contamination of databases used in clinical practice**
  - **Extrapolation of penetrance from clinical studies of selected individuals to the general population**
  - **Resolving penetrance vs uncertain pathogenicity is challenging**

- **Screening prenatally: fetus or parents**
  - **Risk of known pathogenic variants in 88 recessive diseases (~1/500)**
  - **Risk of dominant *de novo* disease (1/500-1/1000)**

Questions/Comments