Introduction to Genomics

Clara Tang

Department of Surgery

HKU Med
LKS Faculty of Medicine
The University of Hong Kong
香港大學李嘉誠醫學院

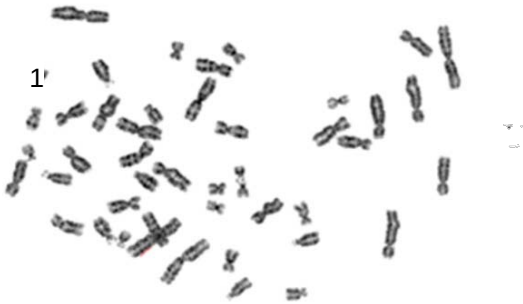
Dr Li Dak-Sum Research Centre
李達三博士研究中心

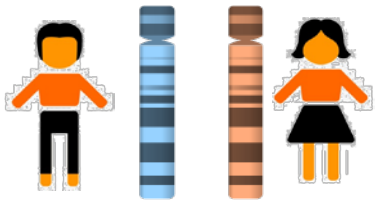
Outline of the talk

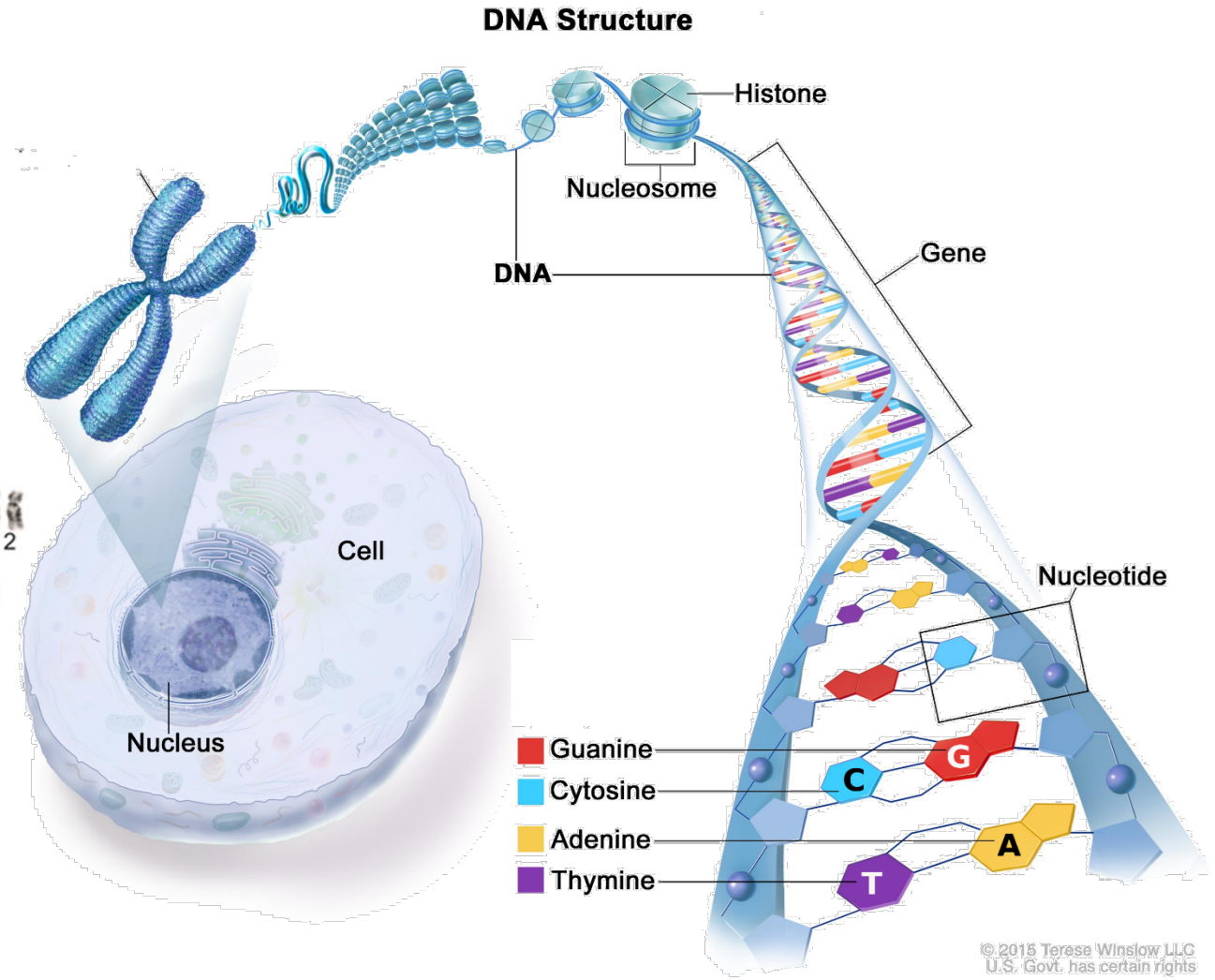
- Genes, genome and genetic variations

- Identification of genetic factors for human diseases

DNA, genes and chromosomes

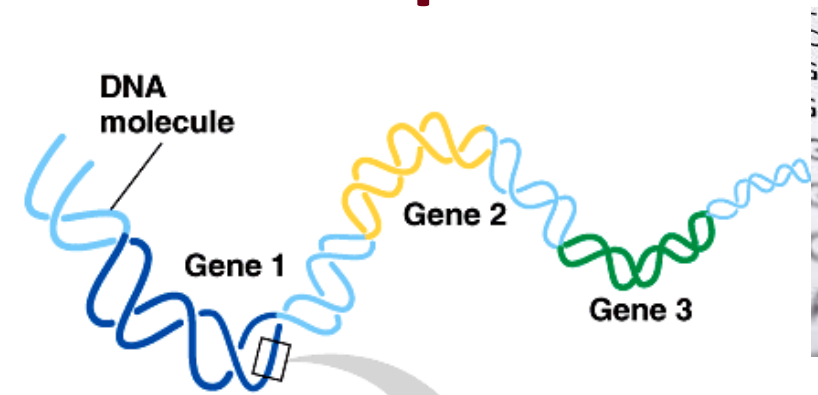


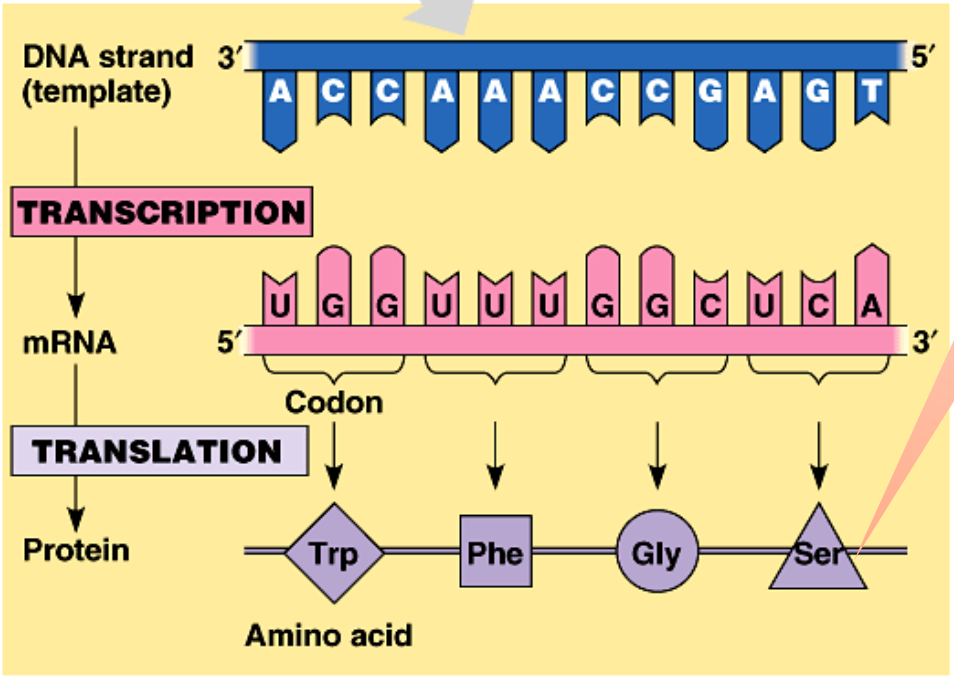


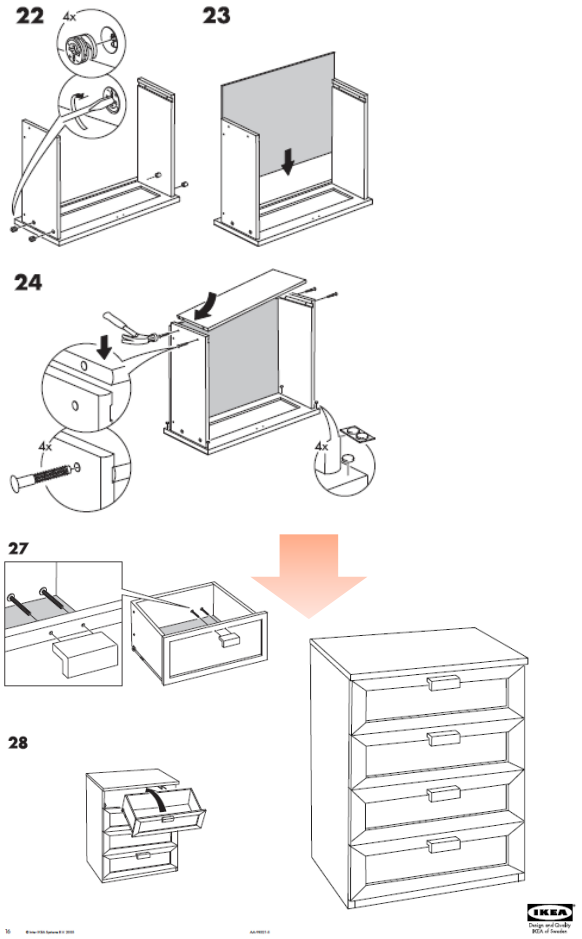


Ref: https://www.nature.com/scitable/topicpage/human-chromosome-number-294#

Genes and proteins







Human genome

1990

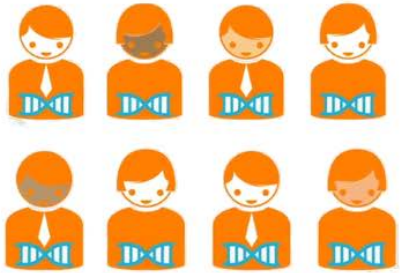
The Human Genome Project began

2001

First draft of the human genome released

2003

Human Genome Project completion announced

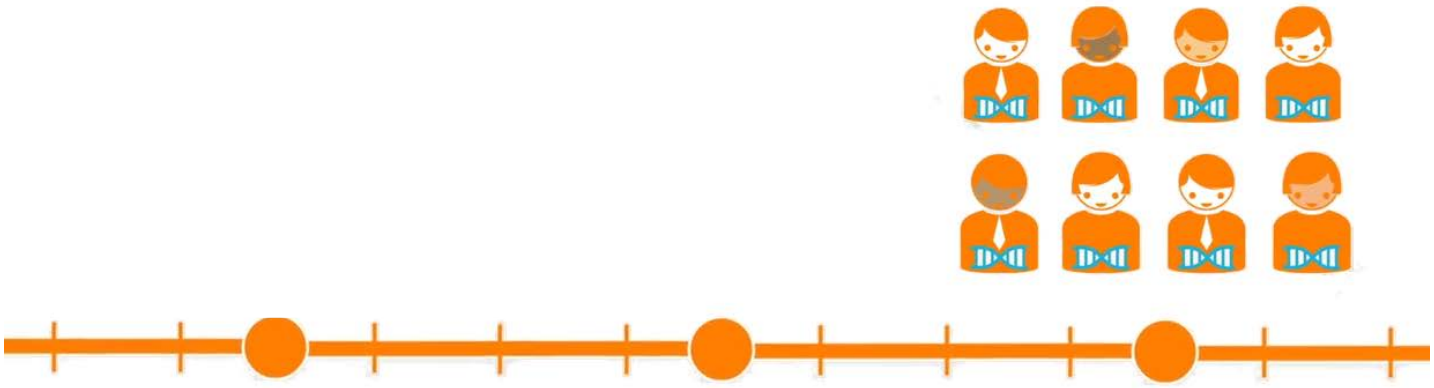


Ref: National Human Genome Research Institute

Human genome

- Human : ~20,000 protein-coding genes
- Mouse : ~22,000 protein-coding genes

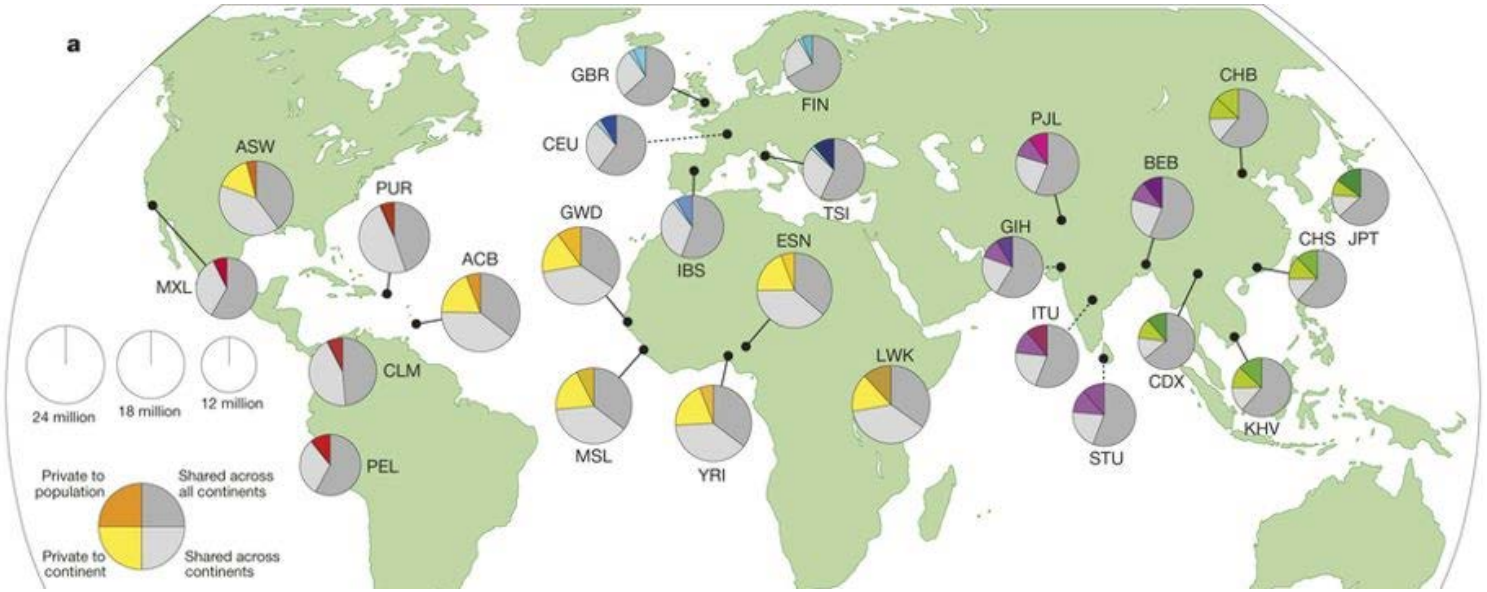
- DNA sequences of any two individuals are **>99% identical**



GENCODE Release 29 (GRCh38.p12) for human and Release M19 (GRCm38.p6) for mouse

Genetic variations

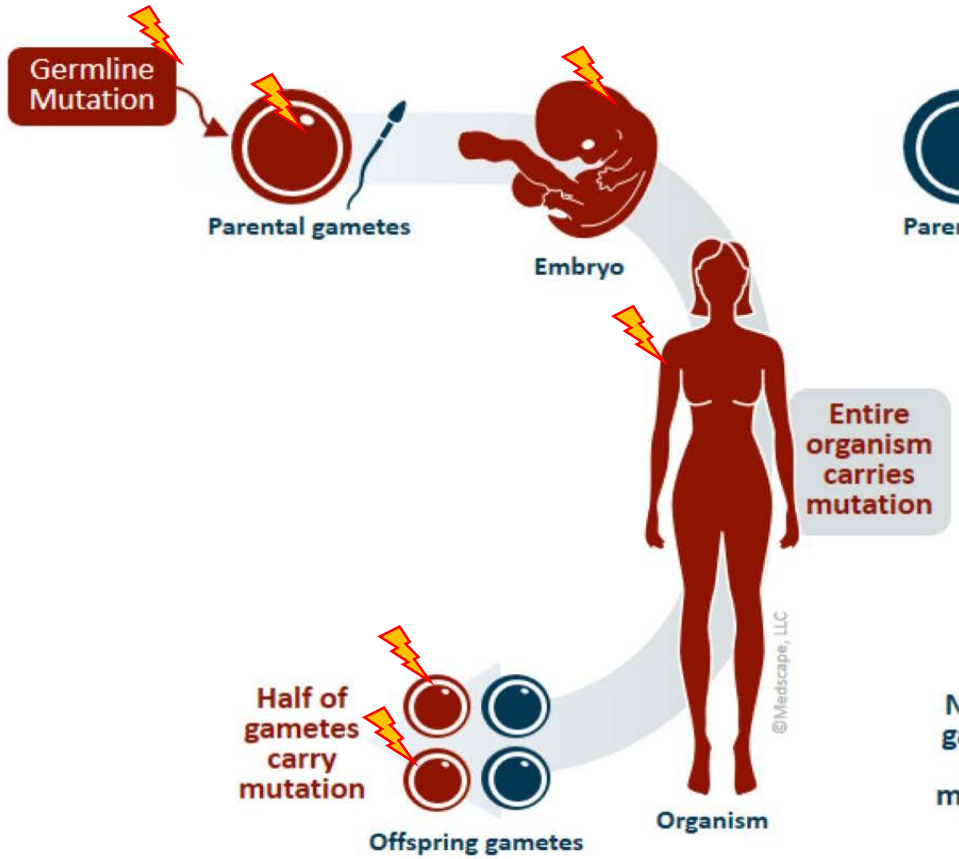
- A typical genome differs from the reference human genome at **4-5 million sites (genetic variations)**
- In a single genome
  - Majority (>95%) of these genetic variants are common
  - 1–4% have a frequency <0.5% in population

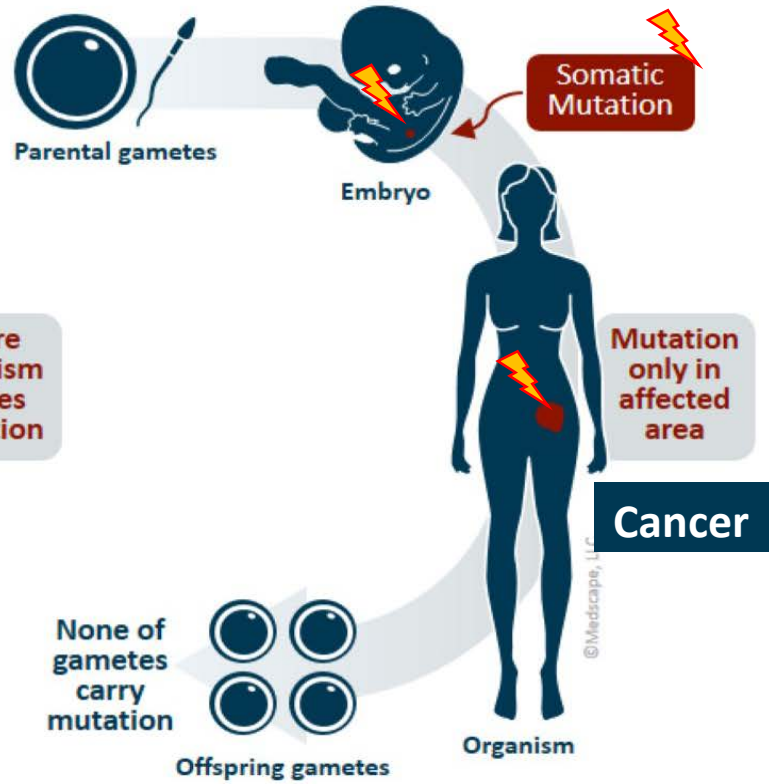


Ref: A global reference for human genetic variation. *Nature* **526**, 68–74

Germline and somatic mutations

- Mutation is one of the sources of genetic variations





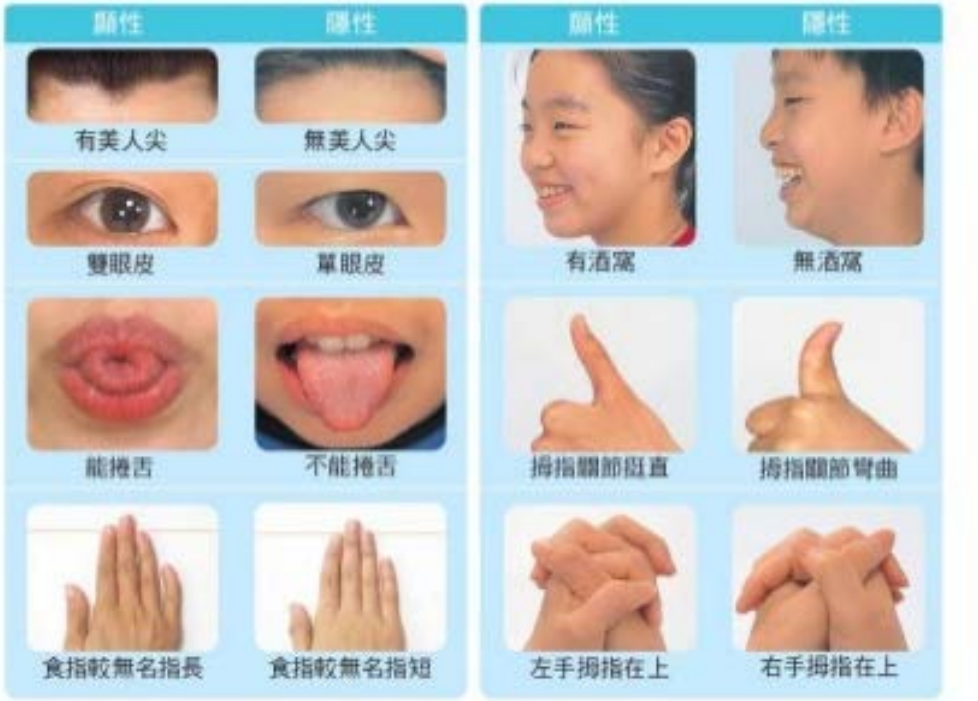
Germline mutation is **heritable**

Somatic mutation is **NOT inherited**

Source: Medscape education

Understanding genetic variations

- Many genetic variations have little or no effect, but some affect the expression and/or functions of genes
- Most of the genetic variations are neutral on fitness
  - Do not affect our ability to survive, adapt or reproduce

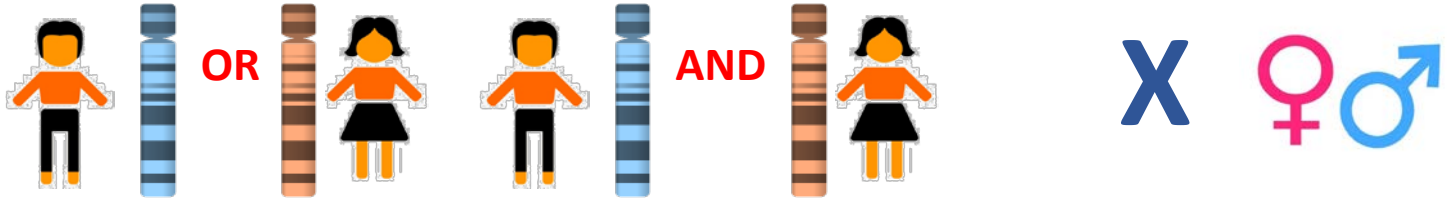


Understanding genetic variations

- Some variations lead to or increase risk of diseases

Monogenic diseases	Polygenic diseases

- **Monogenic diseases**
  - result from variations in a single gene



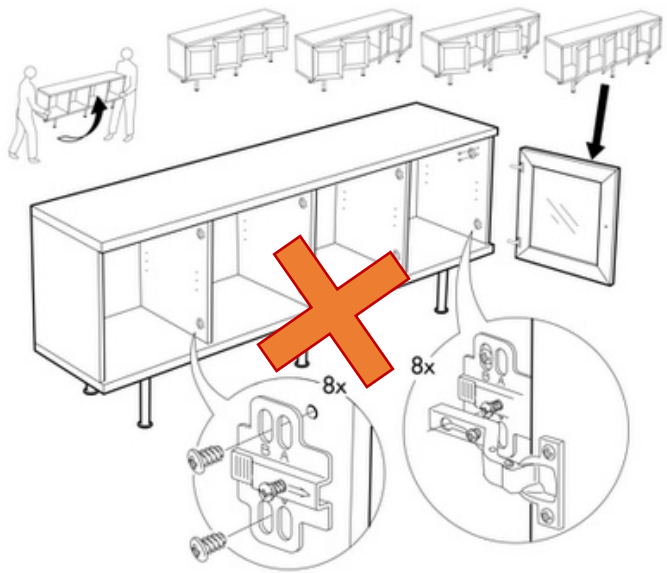
Dominant **Recessive** **X-linked**

- **Polygenic diseases**
  - result from variations in more than one genes

Understanding genetic variations

- Some variations lead to or increase risk of diseases

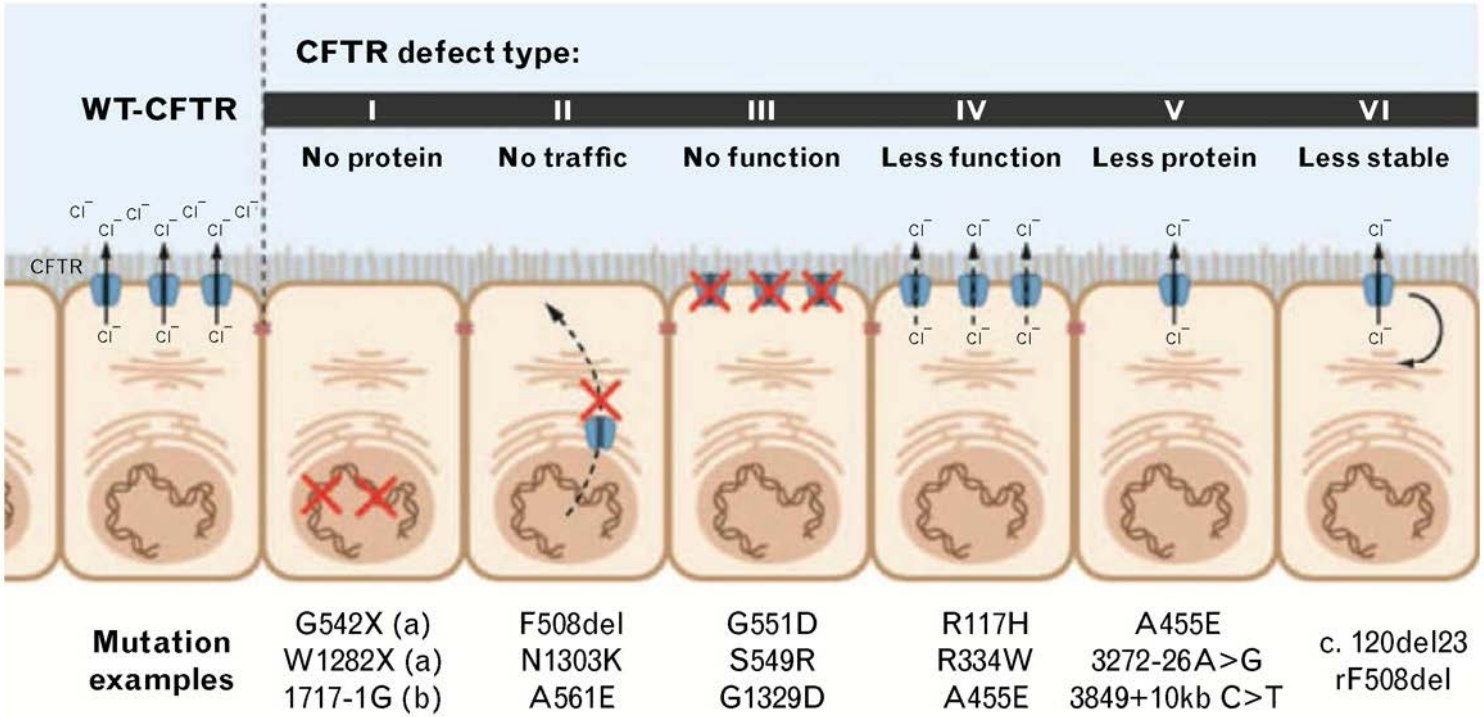
Monogenic diseases



Polygenic diseases

Understanding genetic variations

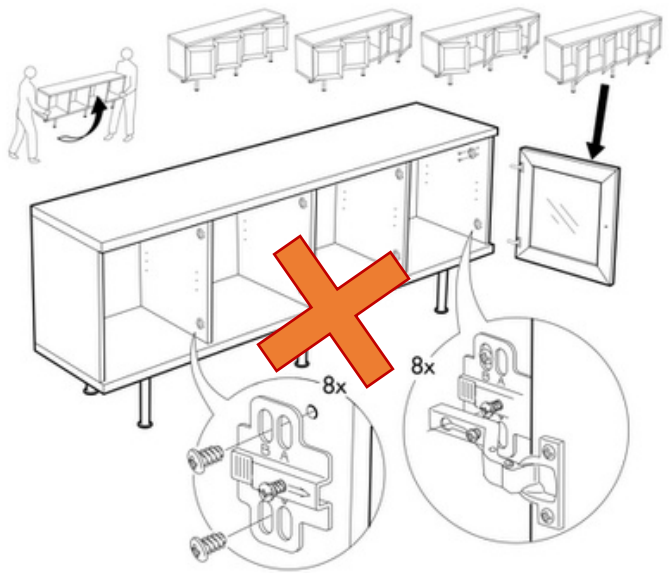
- Rare **monogenic diseases**
  - It is estimated that over 10,000 of human diseases are known to be monogenic
  - Rare variations in *CFTR* cause cystic fibrosis with various severity



Understanding genetic variations

- Some variations lead to or increase risk of diseases

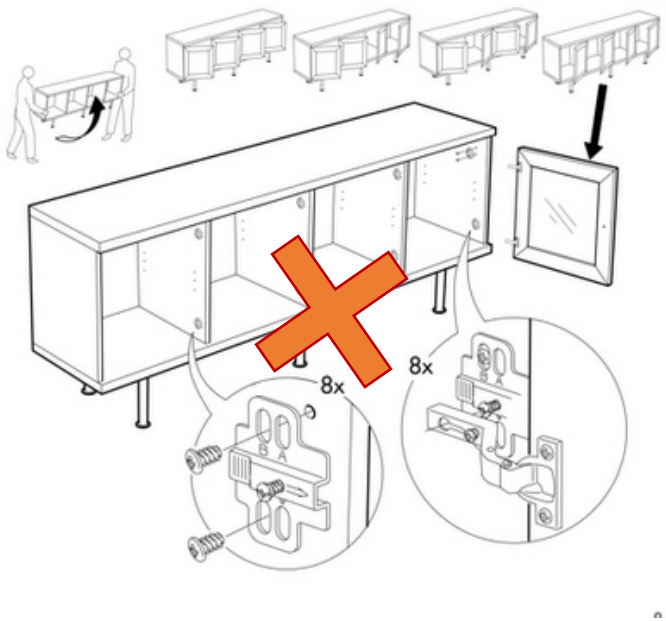
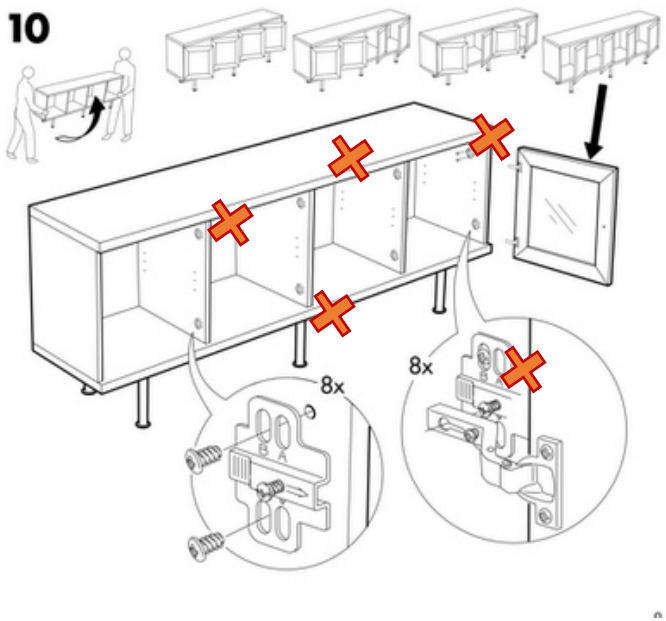
Monogenic diseases



- Beta thalassemia
- Huntington's disease
- Marfan syndrome
- Rett syndrome
- Tay-Sachs disease

Understanding genetic variations

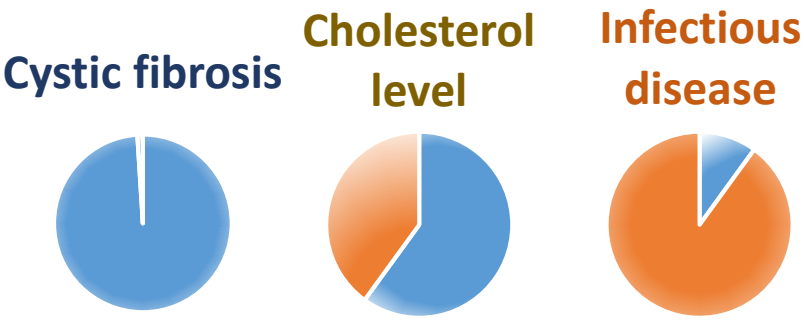
- Some variations lead to or increase risk of diseases

Monogenic diseases	Polygenic diseases
	

Understanding genetic variations

Genetic factors

Environmental factors
- diet, exercise, etc.



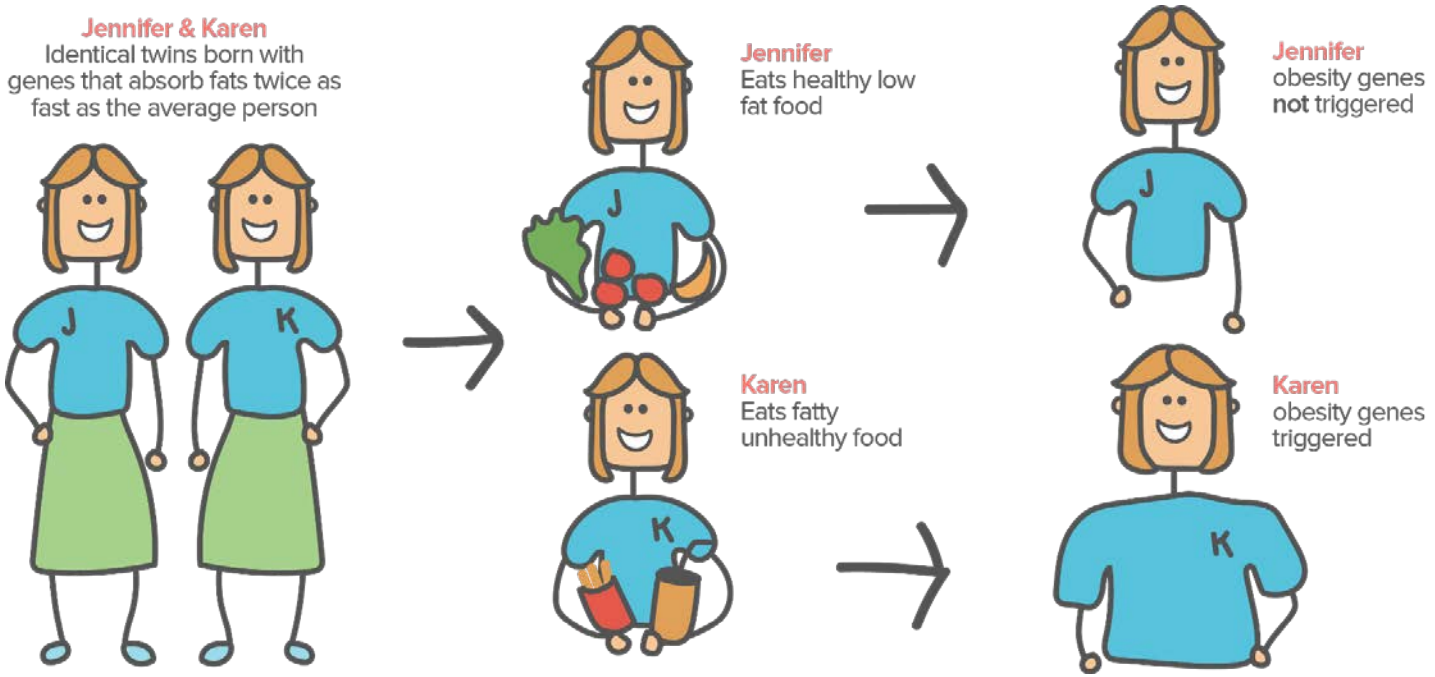


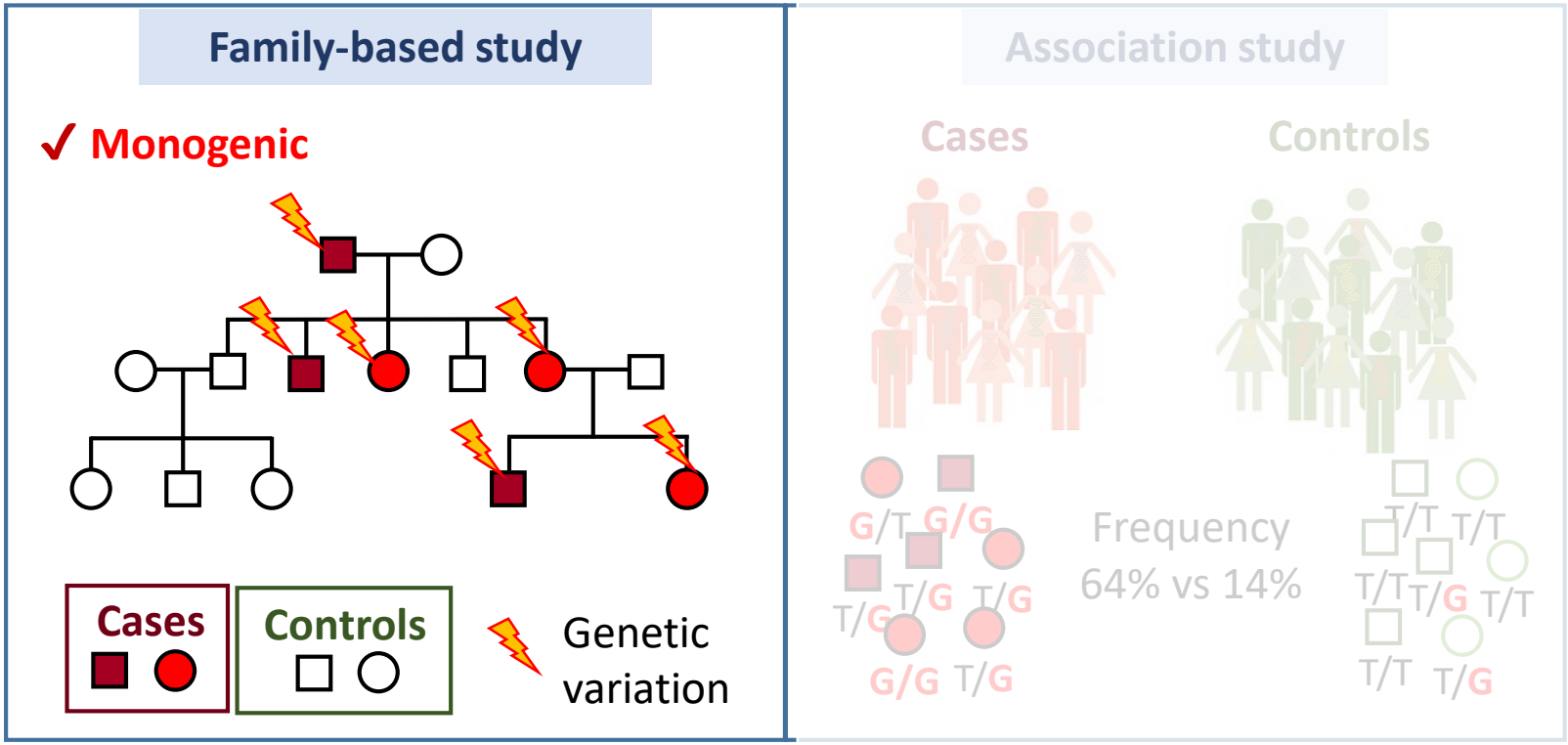
Figure source: https://www.khanacademy.org/

Genetic risk factors

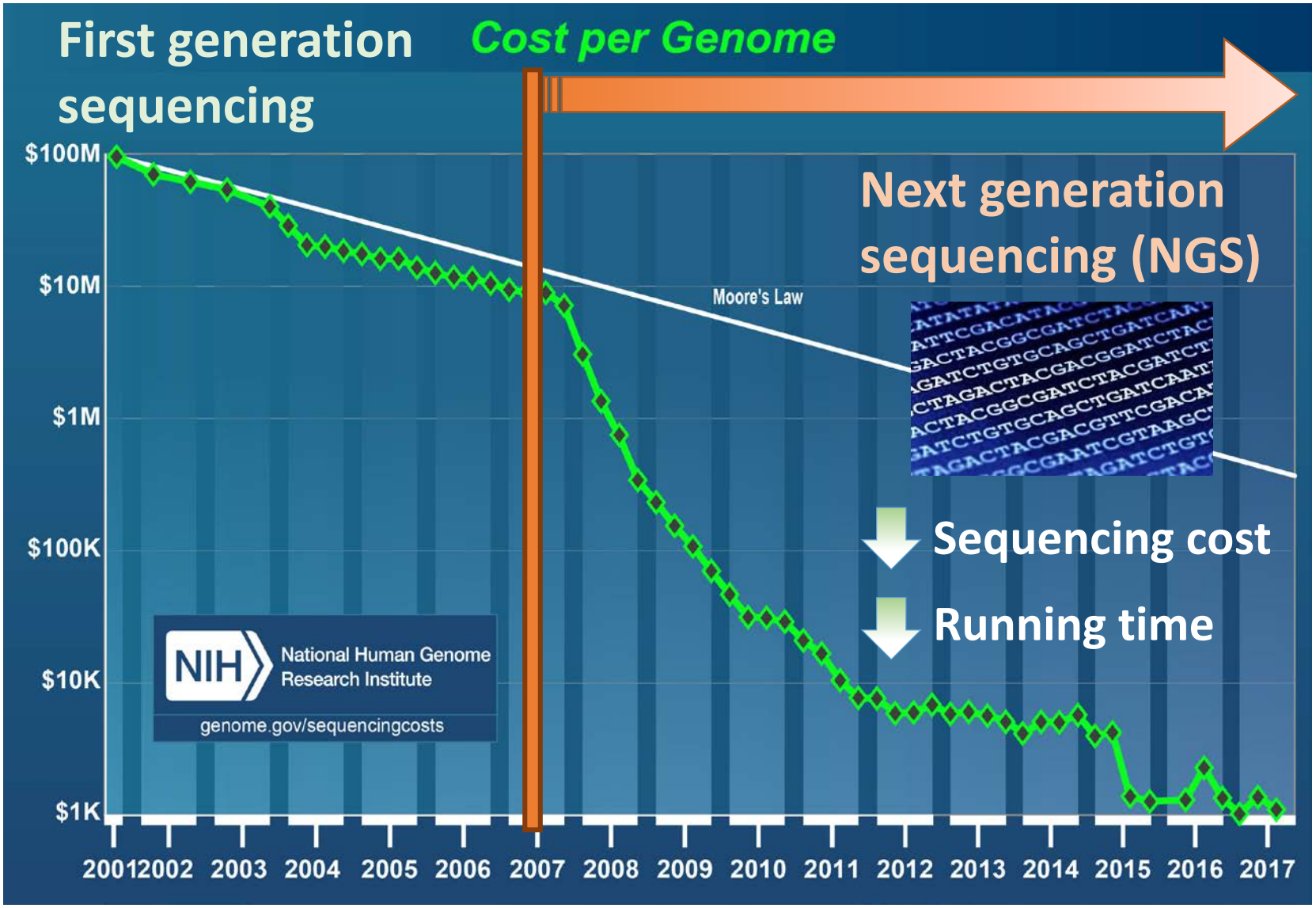
Figure source: http://twitter.com/SynBiologue

Genetic studies for genetic risk factors

- *Genetics* is the study of inherited genetic variations
  - To identify genes associated with diseases and traits
  - To improve our health and well-being



The revolution in sequencing technology

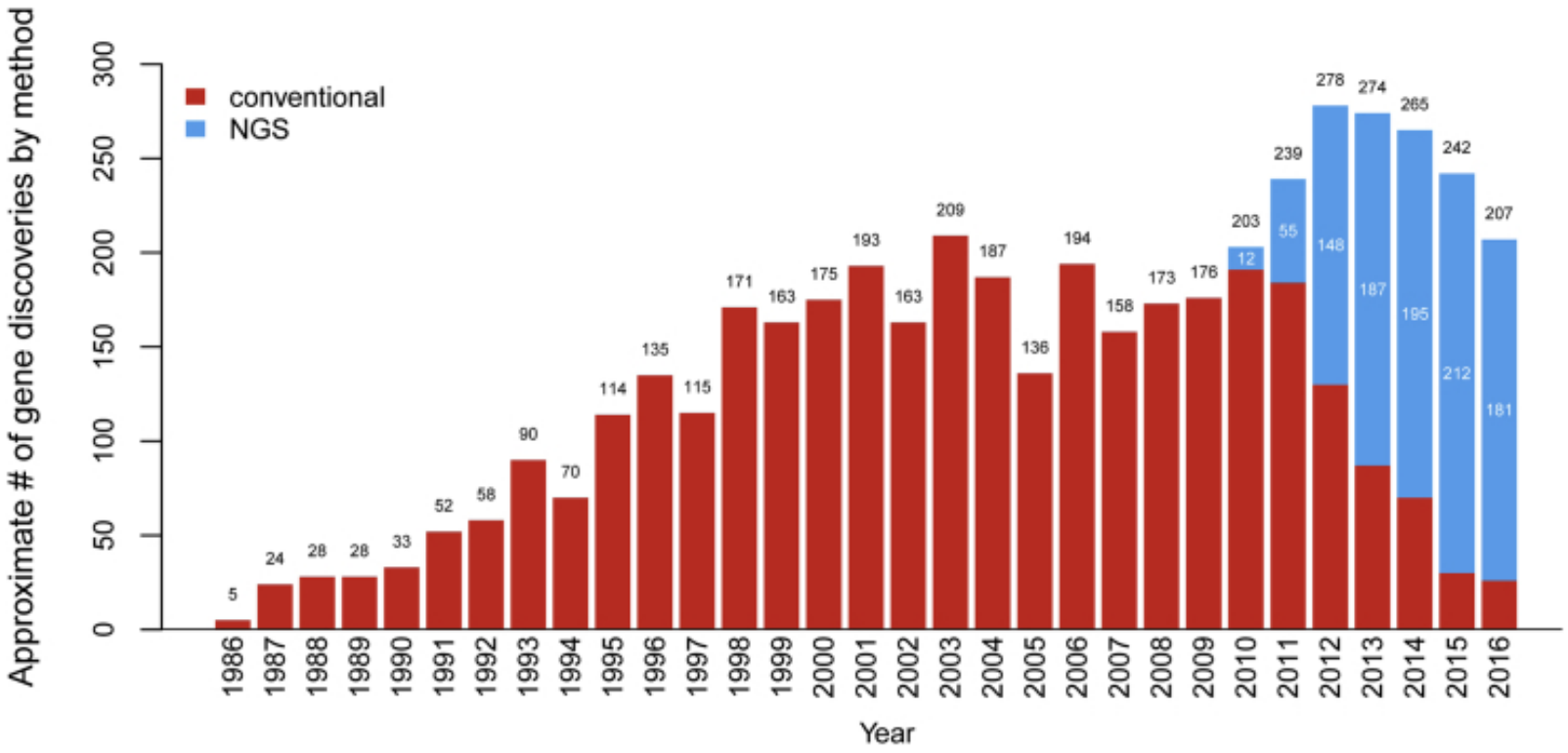


The revolution in sequencing technology

First generation sequencing

Next generation sequencing (NGS)

Gene discoveries on rare genetic diseases by new NGS method

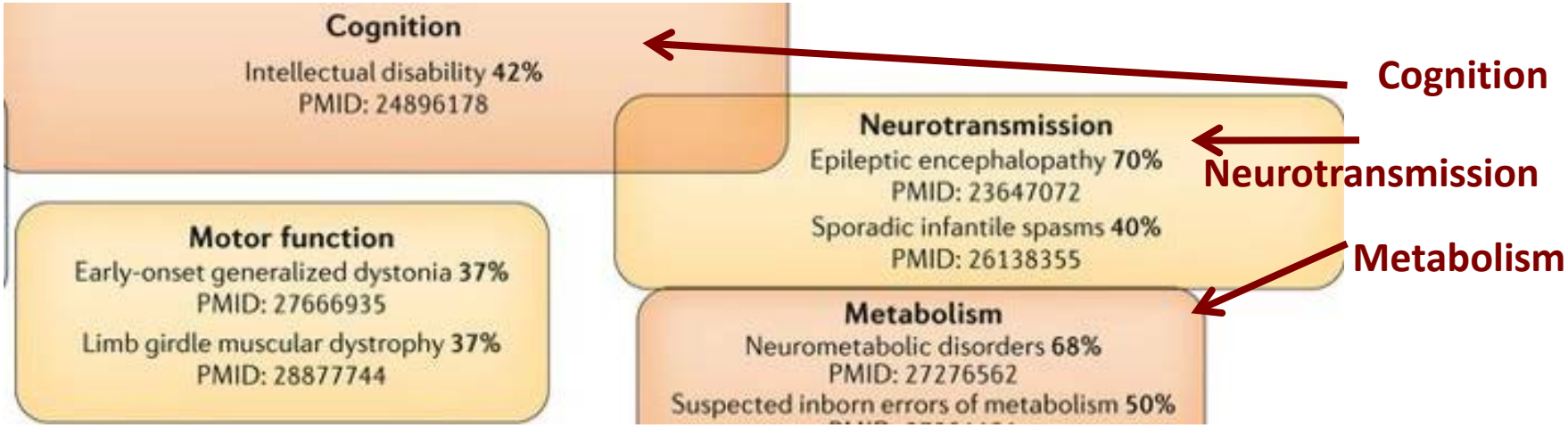


Since 2013, NGS have discovered nearly three times as many genes as the conventional approaches for rare genetic diseases

Ref: Am J Hum Genet. 2017; 100(5): 695–705

Promise of next generation sequencing (NGS) in healthcare

- Can detect **rare** variants that are deleterious
- Diagnostic yield for **finding the causal variants for rare pediatric diseases** is around 30-70%



Adapted from Nature Reviews Genetics 19, 253–268 (2018)

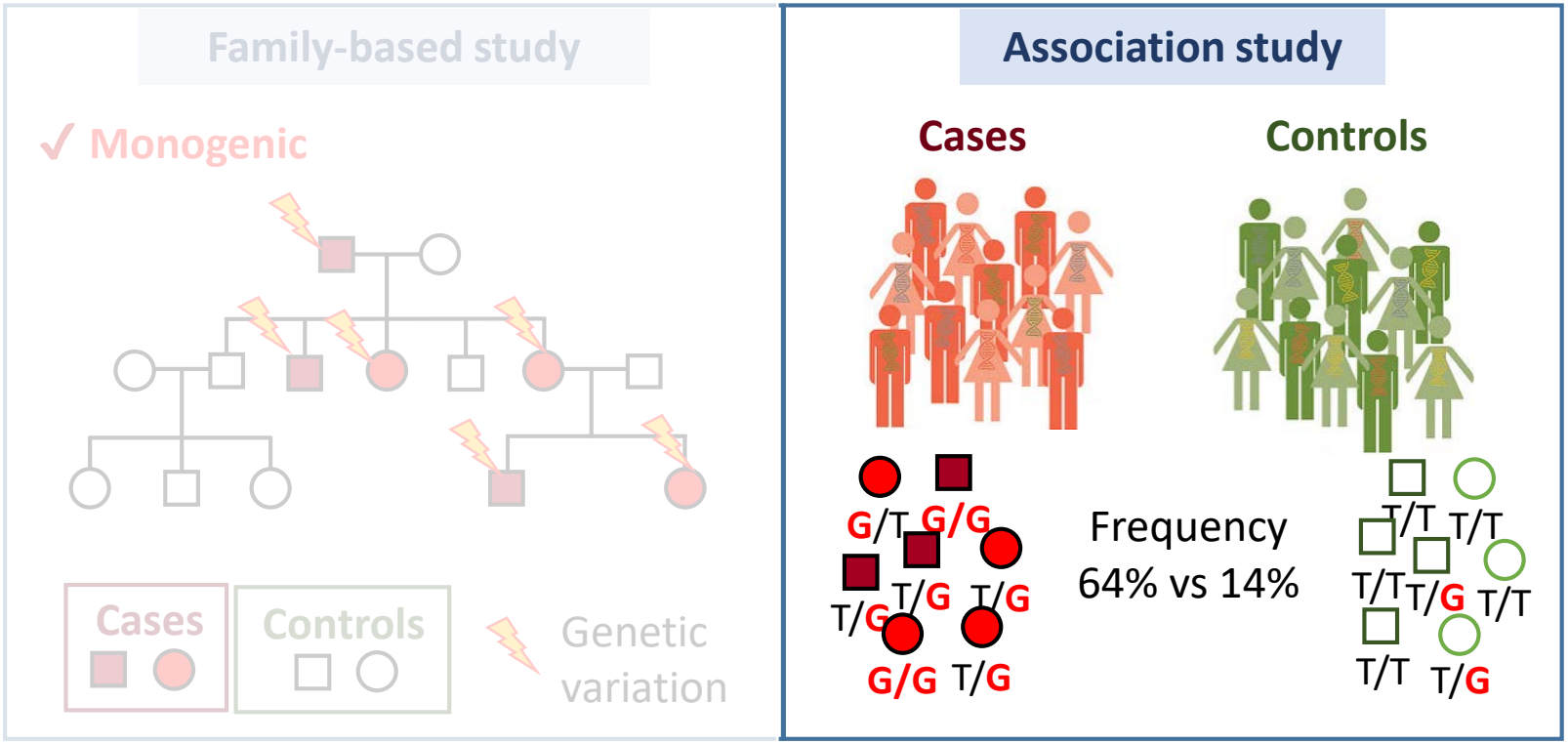
Promise of next generation sequencing (NGS) in healthcare

Early diagnosis | **Prevention** | **Rapid intervention** | **Precision treatment** | **Continuous assessment**

- Studies reported that the molecular diagnosis **affects clinical management** in ~25-50% of these diagnosed patients
  - dietary management, supplement or medications

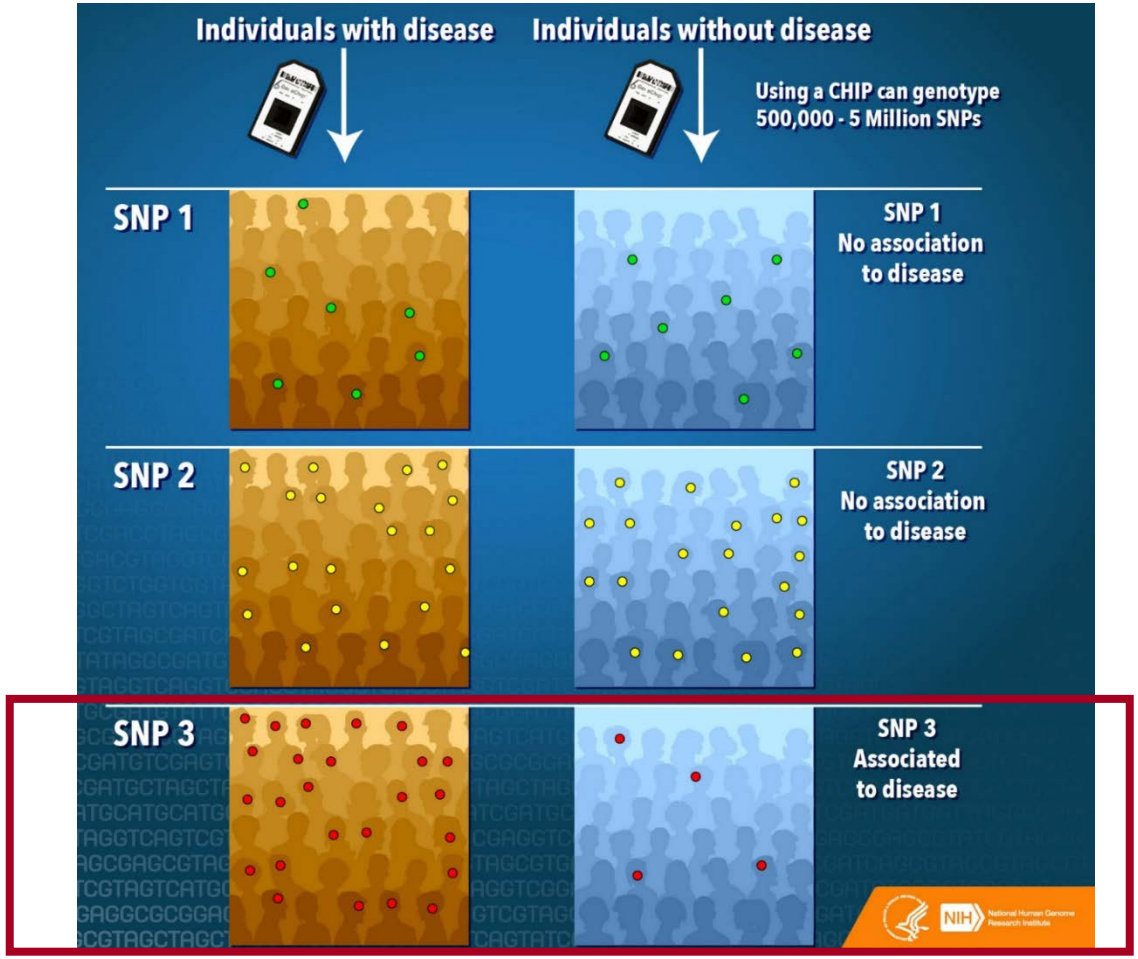
Genetic studies for genetic risk factors

- *Genetics* is the study of inherited genetic variations
  - To identify genes associated with diseases and traits
  - To improve our health and well-being



Finding genetic risk factors in large scale

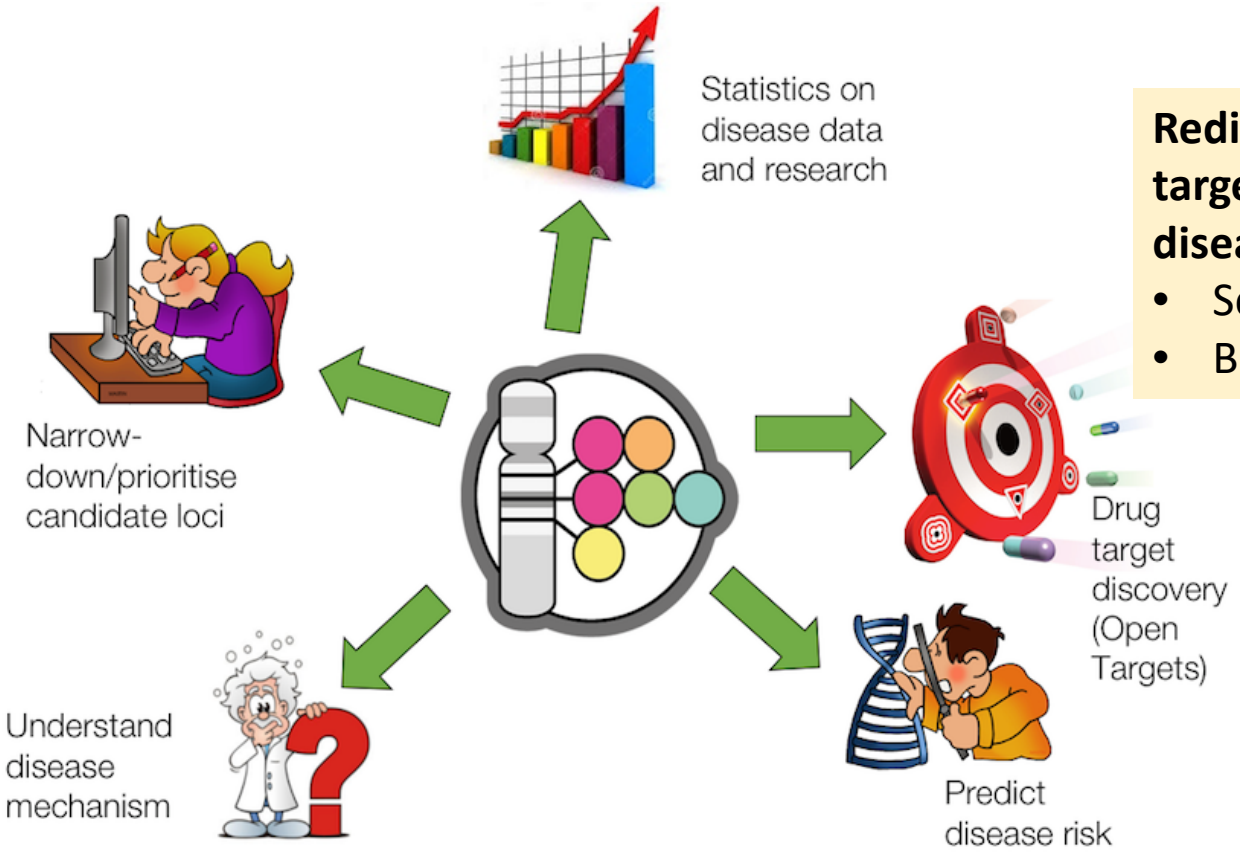
- Genomewide association study **(GWAS)**



500,000 to 5 millions variants

Genetic risk factors identified by GWAS

- As of 2018-10-29, there are >3644 publications
  - reporting ~78236 robust genetic associations
  - with >300 diseases and traits



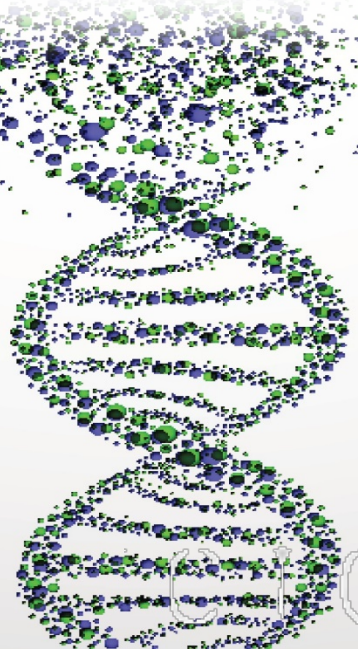
Rediscovery of drug targets for common diseases

- Schizophrenia
- Blood lipids

Summary

- With the recent technological advances, we are now learning more about how each gene works and what it does
- Genetic studies have and will continue to improve our knowledge on the genetic causes of various diseases, which
  - will lead to advances in disease prevention
  - will shed light on healthcare management, treatment and drug development

Thank you



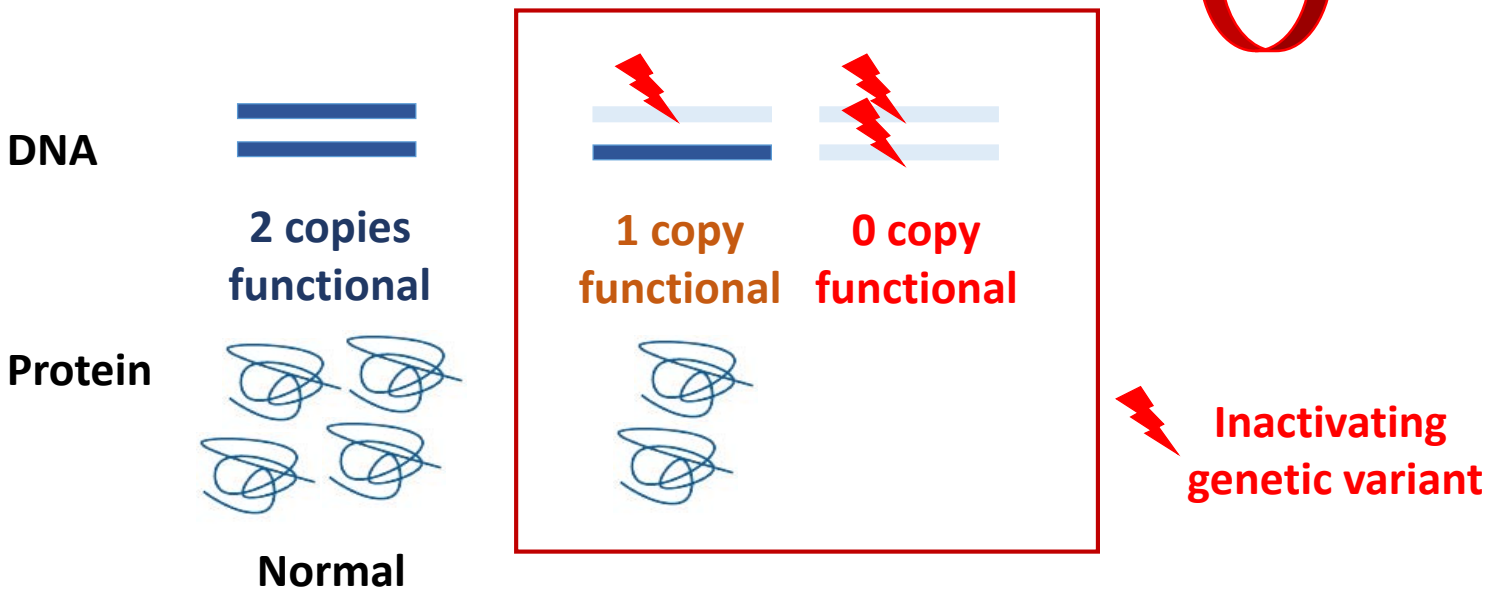
Prof Paul KH Tam
Prof Pak C Sham
Dr Merce Garcia-Barcelo
All labmates

Department of Surgery
HKU Med LKS Faculty of Medicine The University of Hong Kong 香港大學李嘉誠醫學院
Dr Li Dak-Sum Research Centre 李達三博士研究中心

Natural human "knockout" by inactivating variations

- Natural human "knockout" can be recognized as **"experiments of nature"**
- ExAC database involving> 60,000 individuals found an average of 120 inactivating genetic variations
  - 85 variations resulting in **1 functional** copy of gene
  - 35 variations resulting in **0 functional** copy of gene

Human "knockout"



Human “knockout” project

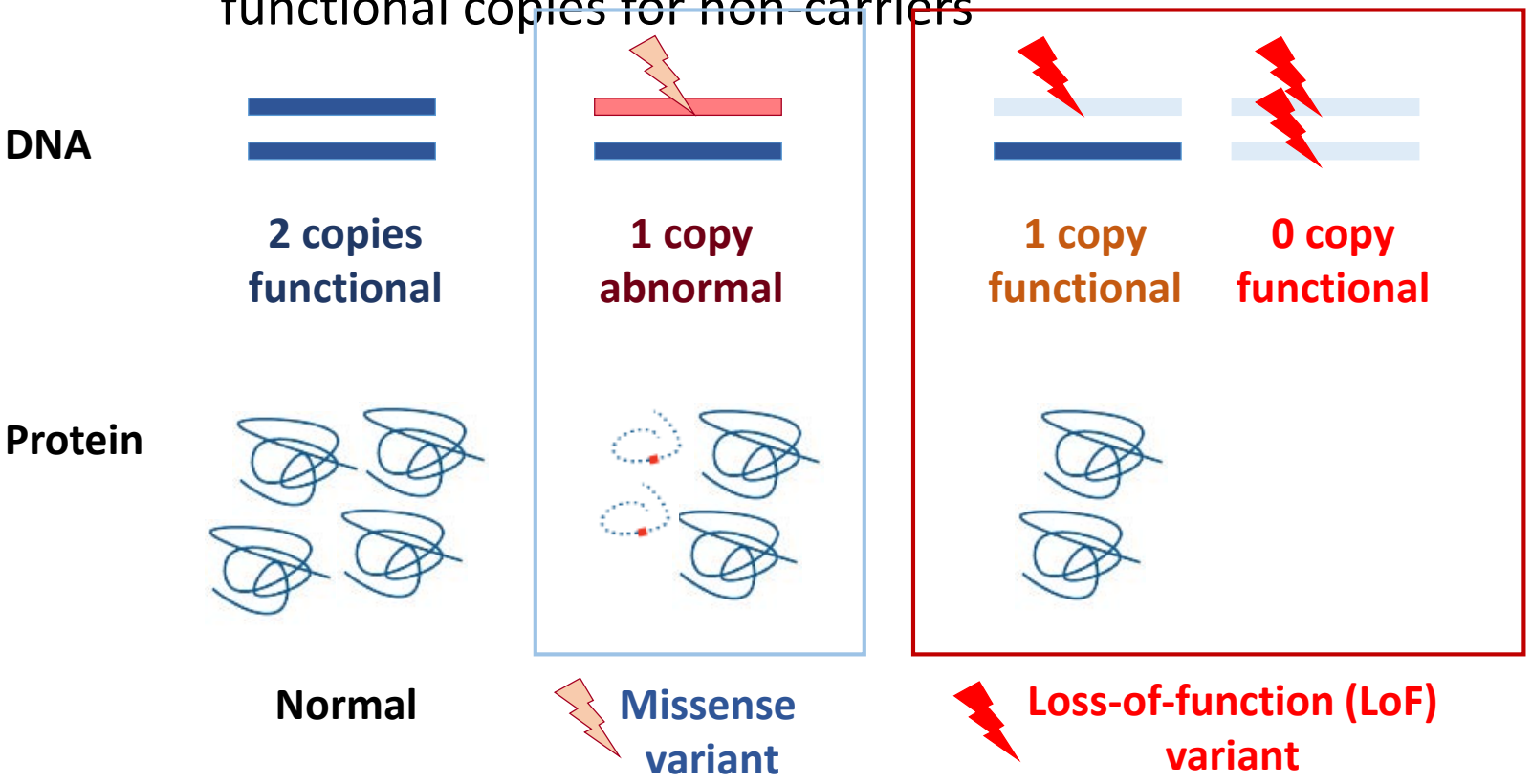
- Studying the human “knockout” can inform
  - gene-level constraint against gene inactivation
  - the functional impact of gene inactivation

- Implication for drug development
  - To **rule-in drug targets**
    - Elimination of *CCR5* protects against HIV infection
    - Elimination of *APOC3* protects against coronary heart disease
  - To **rule-out drug targets**
    - Elimination of *PLA2G7*, that associated with coronary atherosclerosis, demonstrates no protection from cardiovascular disease
    - Such observation matches with result from recent clinical trial

“Genetic superheros”

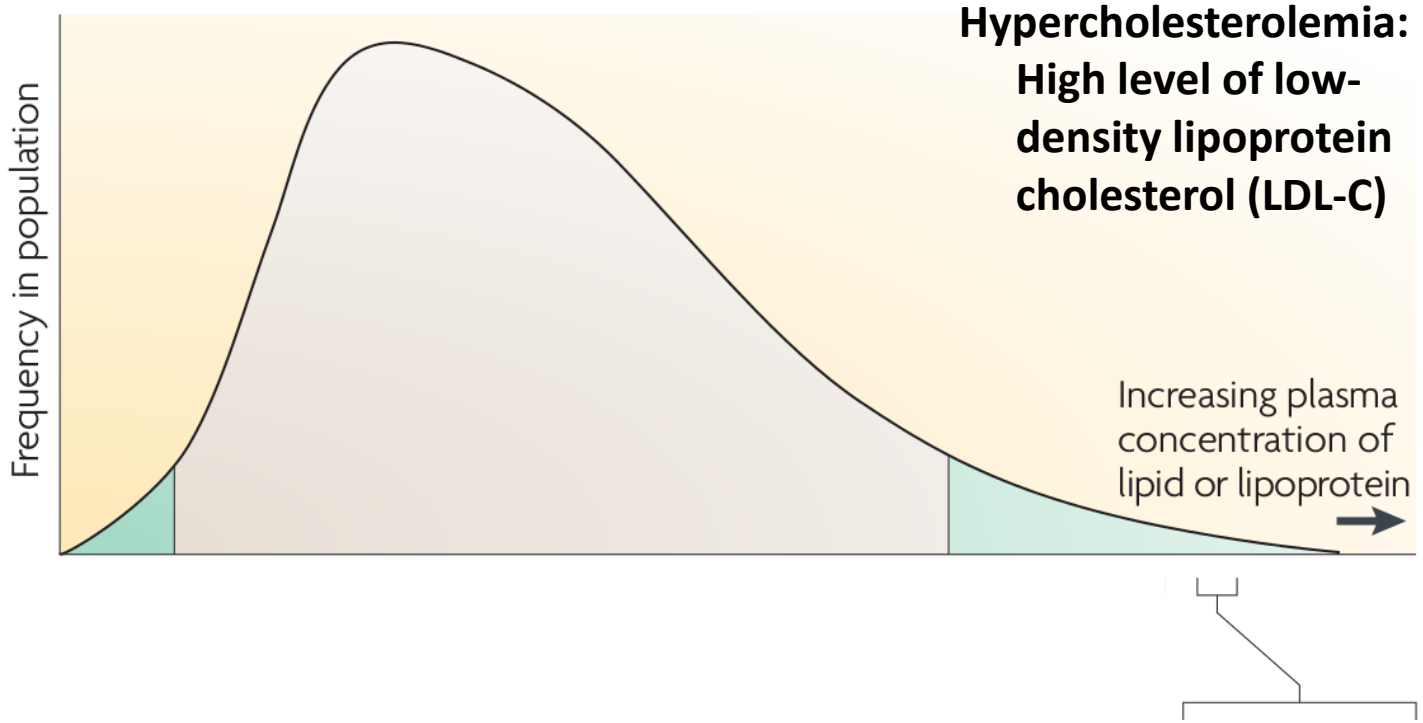
Rare genetic variants associated with diseases

- These causal variants can be
  - **Missense** : affecting protein structure
  - **Protein truncating/Loss-of-function (LoF)**
    - resulting in 0 or 1 functional copy of genes compared to 2 functional copies for non-carriers



Understanding genetic variations

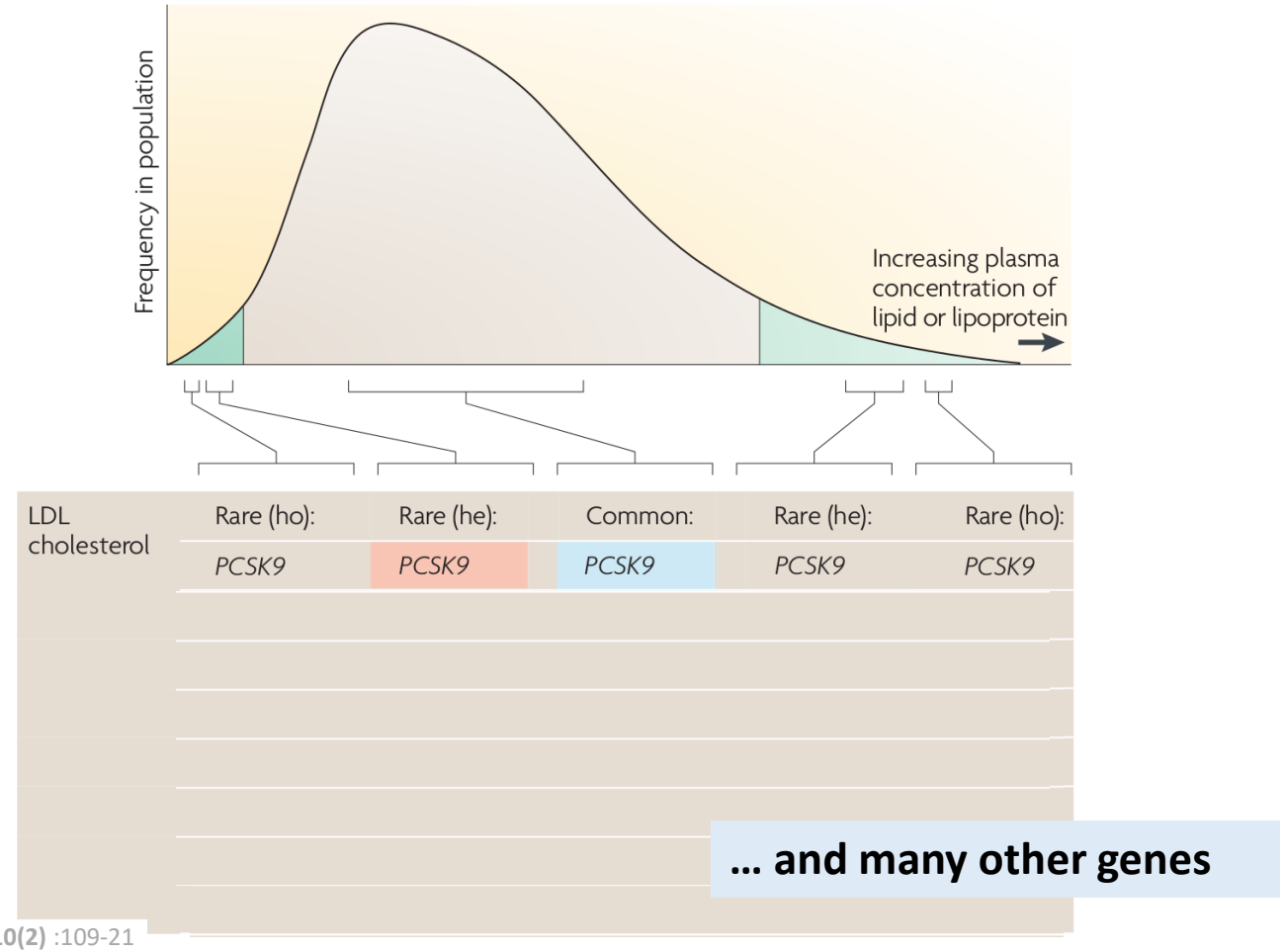
- Some variations lead to or increase risk of diseases
- **Polygenic diseases/traits**



Adapted from *Nat Rev Genet.* **10(2)** :109-21

Understanding genetic variations

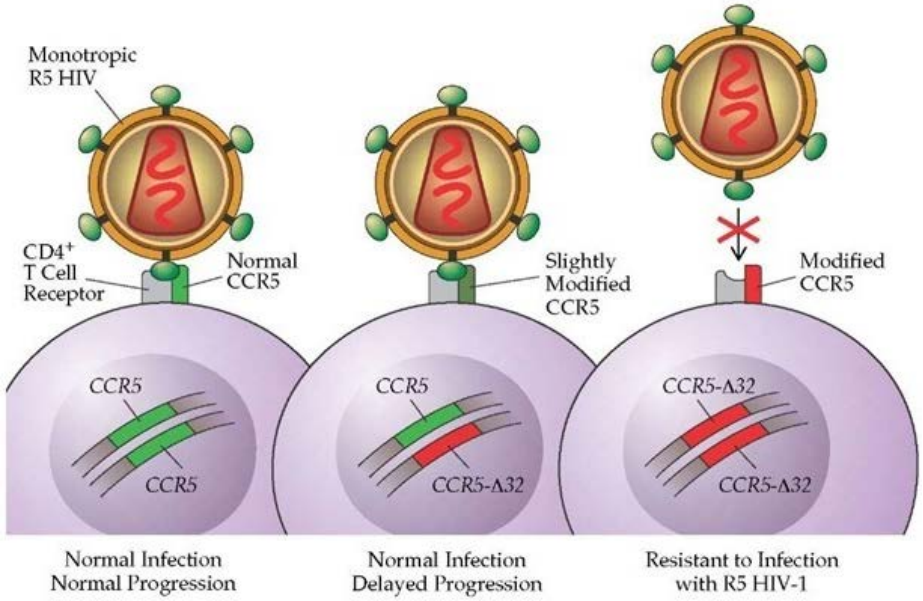
- Some variations lead to or increase risk of diseases
- **Polygenic diseases/traits**



Adapted from *Nat Rev Genet.* **10(2)** :109-21

Understanding genetic variations

- Some variations may improve our survivability
- Inactivating variations in *CCR5* gene may delay the progression of AIDS and in some cases bring about immunity, e.g.
  - *CCR5*-Δ32 mutation makes it harder for HIV to bind to the surface of the immune cells and infiltrate them



- CCR5 deficiency may increase risk of symptomatic West Nile virus infection (Glass *et al.* JEM 203(1):35-40)

Figure source: http://what-when-how.com/acp-medicine/hiv-and-aids-part-2/

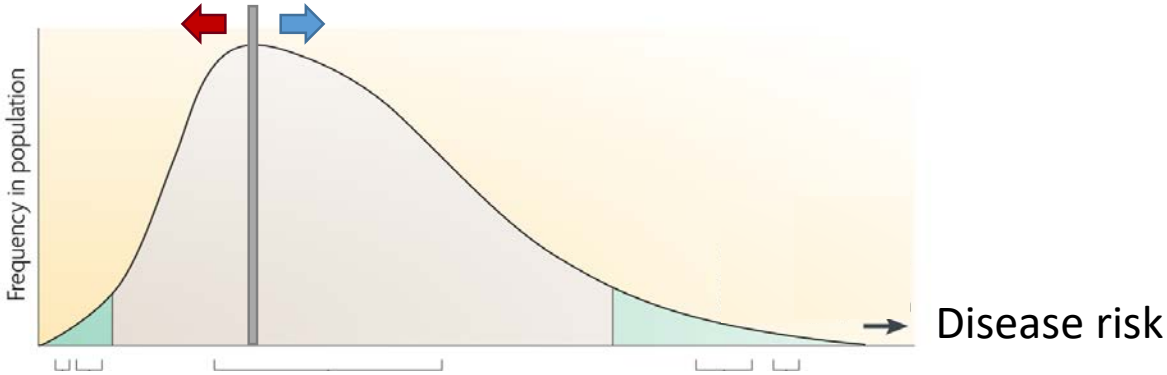
Genetic risk factors identified by GWAS

- As of 2018-10-29, there are >3644 publications
  - reporting ~78236 robust genetic associations
  - with >300 complex diseases and traits

- Biological insights from discoveries of GWAS
  - Identification of disease-susceptibility genes
  - Implication of new disease-relevant biological pathways
  - Rediscovery of drug targets
    - *DRD2* encoding target for anti-psychotic drug—dopaminergic receptor—is associated with schizophrenia
    - *HMGCR* encoding target for statin is strongly associated with blood lipids including total cholesterol and low-density lipoprotein cholesterol (LDL-C)

Drug targets implicated from GWAS

- Most of the genetic associations detected by GWAS
  - involve **common genetic variants** and
  - the **risks on diseases (effect size)** are **small**



- Effect size does not affect the efficacy of being a drug target
  - **LDL-associated genetic variant** near *HMGCR* decreases LDL-C levels by a modest **2.5 mg/dL**
  - Use of **statins**, which inhibit function of the enzyme encoded by *HMGCR*, typically decreases LDL-C levels by 20–40%, or **~14–70 mg/dL**