GENOMICA APPLICATA

The Human Genome

Intro

- Understanding
  - the organization,
  - variation,
  - and expression of the **human genome** is central to the principles of genomic and precision medicine.

- The comparison of individual genomes underlies the conclusion that **virtually every individual has his or her own unique constitution of gene products**

The «chemical individuality»

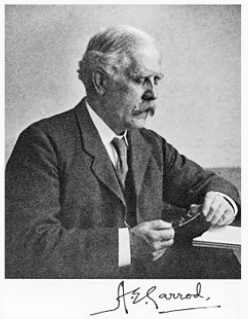
Archibald Garrod

From Wikipedia, the free encyclopedia

Sir Archibald Edward Garrod KCMG FRS[1] (25 November 1857 – 28 March 1936) was an English physician who pioneered the field of inborn errors of metabolism. He also discovered alkaptonuria, understanding its inheritance. He served as Regius Professor of Medicine at the University of Oxford from 1920 to 1927.[2]

Contents [hide]

Archibald Garrod



Born	25 November 1857 London
Died	28 March 1936 (aged 78) Cambridge
Nationality	English
Known for	alkaptonuria
Awards	Fellow of the Royal Society[1]
Scientific career	
Fields	medicine
Institutions	University of Oxford

Education and personal life [edit]

Brilliancy ran in Garrod's family. Archibald was the fourth son of Sir Alfred Baring Garrod, a renowned physician who received his medical degree at the age of 23 and became a professor of medicine at University College, London by the time he was 32. He discovered the abnormal uric acid metabolism associated with gout.[2] Garrod's father also successfully estimated the weight of crystals he obtained from a known quantity of blood, resulting in what Garrod called "the first quantitative biochemical investigation made on the living human body".[3] Garrod's two older brothers also had successful careers.

Charles Keene, a cousin, frequently visited Garrod's childhood home. Keene was a illustrator for the magazine Punch for over 40 years. Influenced by Keene, Garrod wrote an illustrated booklet called *A Handbook of Classical Architecture*. According to Krishna Dronamraju, Garrod displayed an interest in natural history from an early age and was particularly interested in butterflies. At the age of 12, he began collecting them and noted how few female butterflies were present, musing over possible inheritance patterns in mammals.[4]

He was educated at Marlborough College and Christ Church, University of Oxford. He performed poorly at Marlborough, struggling due to his lack of interest in classics, especially Latin prose and grammar. He graduated with a First-class Honours (or a "First") degree in natural science in 1878.[4]

In 1880, he received further medical training at St. Bartholomew's Hospital in London, where he obtained several scholarships including the competitive Brackenbury Scholarship. Garrod graduated in 1884 and then spent a year studying in Vienna at the general hospital, known as the Allgemeines Krankenhaus. His experiences in Vienna formed the basis for his 1886 work, *An Introduction to the Use of the Laryngoscope*, which was very well received. In 1885 he

The «chemical individuality»

Eur J Pediatr. 1986 Apr;145(1-2):2-5.

"Inborn errors of metabolism" and "chemical individuality", two ideas of Sir Archibald Garrod briefly revisited 50 years after his death.

Burgio GR.

Abstract

Two ideas of Sir A. Garrod, "chemical individuality" (1902) and "inborn errors of metabolism" (1908) have proved fundamental for the development of medical knowledge. The latter idea was more fortunate than the former which, however has been extremely heuristic. On the other hand the two ideas are not entirely independent of each other: in fact, a third Garrodian concept, "inborn factors in disease", represents a significant link between them. "Inborn errors of metabolism" revived the laws of genetics and opened the way to interpretation of the molecular diseases with all their inherent practical modern implications (neonatal screening, prenatal diagnosis, and in perspective, genetic engineering). "Chemical individuality" still constitutes a valid premise for knowledge of biological individuality (in other words, the "biological ego") fundamentally programmed for conservation of self and for continuous discrimination of self versus non-self.

Variation in the human genome

- Variation in the human genome has long been the cornerstone of the field of human genetics, and its study led to the establishment of the medical specialty of **medical genetics**.

Genetics or Genomics

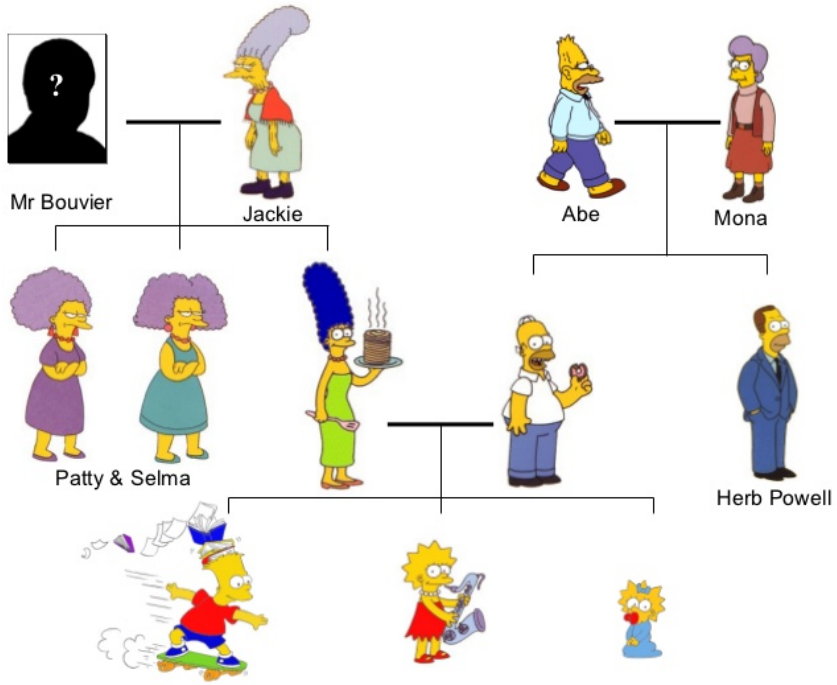
- While these terms seem similar, they in fact describe quite distinct (though frequently overlapping) approaches in biology and in medicine.
- Having said that, there are inconsistencies in the way the terms are used, even by those who work in the field.

Genetics or Genomics

- The field of *genetics* is the scientific study of heredity and of the genes that provide the physical, biological, and conceptual bases for heredity and inheritance.
- To say that something—a trait, a disease, a code, or an information—is “genetic” refers to its basis in genes and in DNA.

Genetics or Genomics

- *Heredity* refers to the familial phenomenon whereby traits (including clinical traits) are transmitted from generation to generation, due to the transmission of ger
child.



Genetics or Genomics

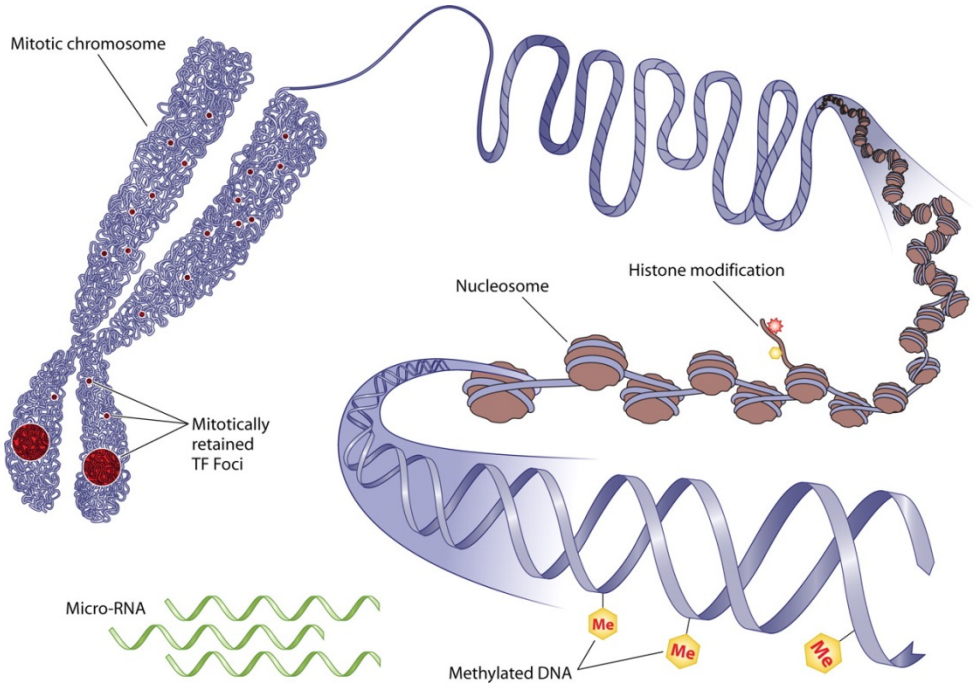
- *Genomics* is the scientific study of a genome or genomes.
- A *genome* is the complete DNA sequence, referring to the entire genetic information of a gamete, an individual, a population, or a species.

Genetics or Genomics

- “Genomics” gave birth to a series of other “-omics” that refer to the comprehensive study of the full complement of genome products
  - proteins (hence, *proteomics*),
  - transcripts (*transcriptomics*), or
  - metabolites (*metabolomics*).

Genetics or Genomics

- By analogy with genetics and genomics, *epigenetics* and *epigenomics* refer to the study of factors that affect gene (or, more globally, genome) function, but without an accompanying change in genes or the genome.



Genetics or Genomics

- *Genomic Medicine* refers to the use of large-scale genomic information and to consideration of the full extent of an individual's genome and other "omes" in the practice of medicine and medical decision making.

Genetics or Genomics

Examples:

- **gene expression profiling** to characterize tumors or to define prognosis in cancer
- **genotyping variants** in the set of genes involved in drug metabolism or action to determine an individual's correct therapeutic dosage

Genetics or Genomics

Examples:

- **scanning the entire genome** for millions of variants that influence one's susceptibility to disease
- analyzing multiple protein or RNA **biomarkers** to detect exposure to potential pathogens

Genetics or Genomics

Examples:

- Monitor therapy
- predictive information in presymptomatic individuals

Characteristics of the Reference Human Genome

- The typical human genome consists of approximately 3 billion (3×10^9) base pairs of DNA,
- 24 types of nuclear chromosomes (22 autosomes, plus the sex chromosomes, X and Y)
- the smaller mitochondrial chromosome

Characteristics of the Reference Human Genome

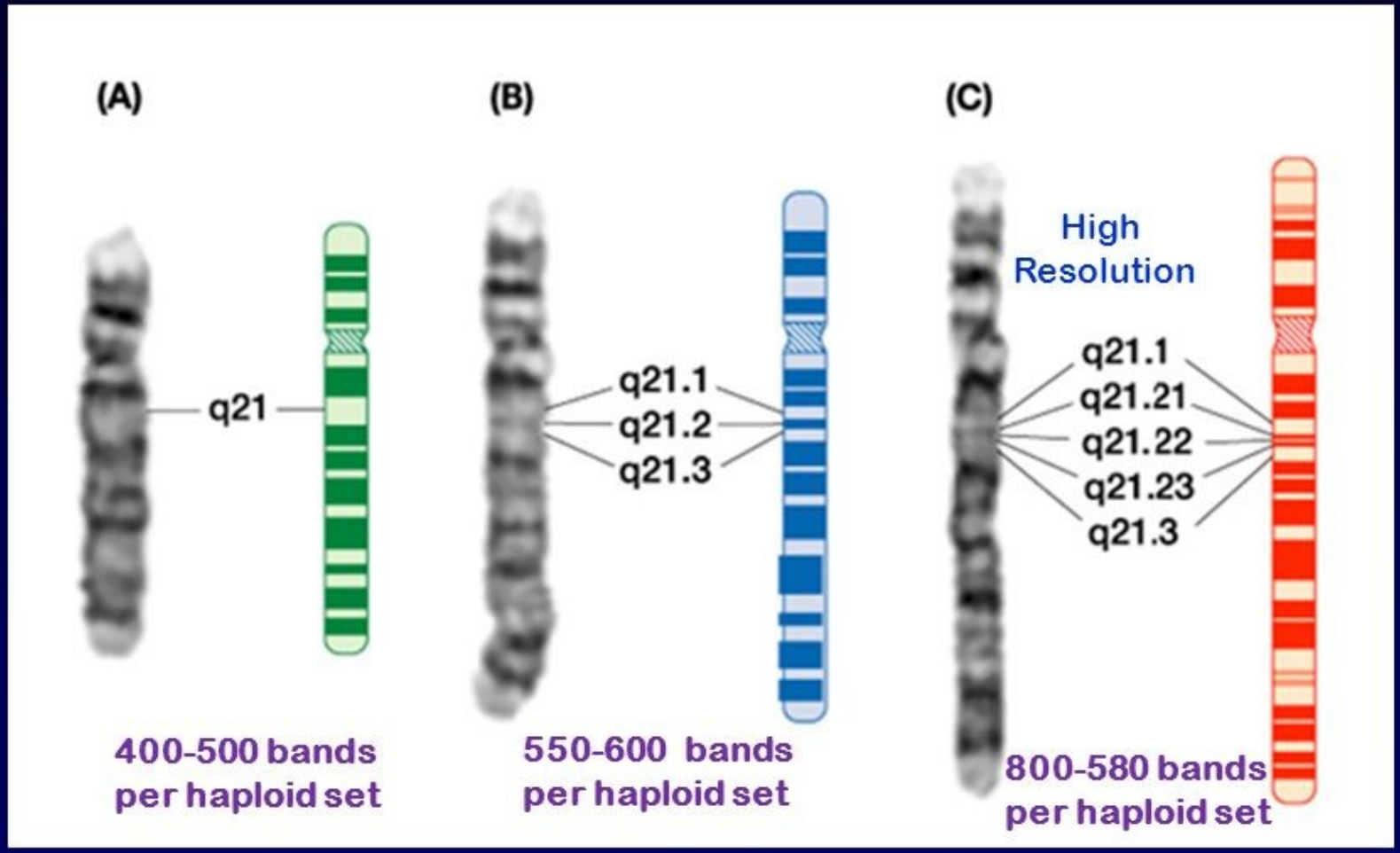
Length of the human genome (base pairs)	3,096,649,726
Number of known protein-coding genes	20,441
Average gene density (number of genes/Mb)	6.6
Number of ncRNA genes	22,219
Number of known short sequence variants	156,148,362
Number of known structural variants	4,485,861

From Ensembl, database GRCh38, version 85.38 (accessed August 2016).

Spectrum of resolution

- Individual chromosomes can best be visualized and studied at metaphase in dividing cells, and **karyotyping of patient chromosomes has been a valuable and routine clinical laboratory procedure for a half century**, albeit at levels of resolution that fall well short of most pathologic DNA variants

High Resolution G banding

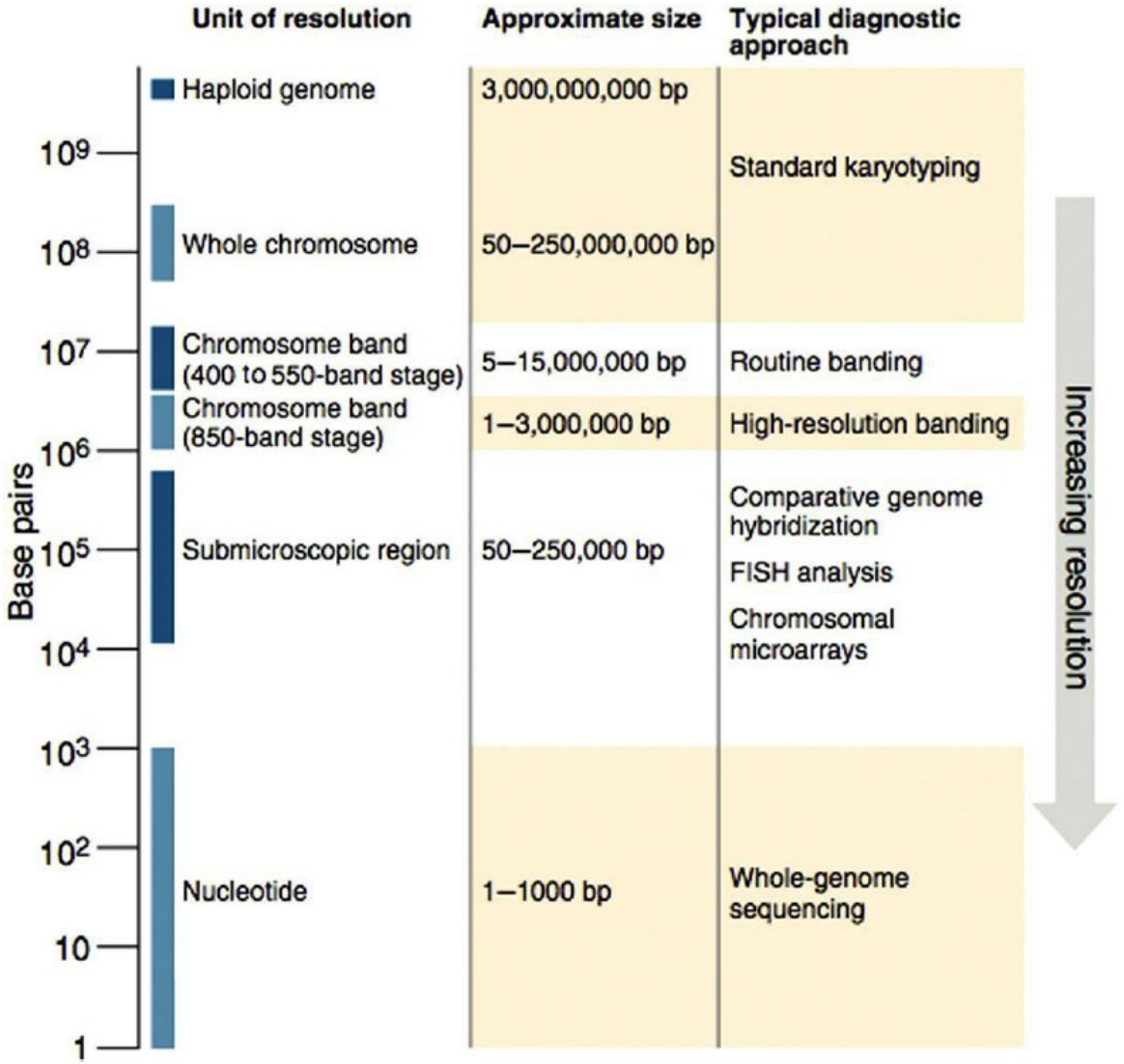


- Human chromosome 4 at varying resolutions due to exact mitotic stage, (or degrees of spreading - squashing - stretching)
- Each band corresponds to about 5000-10000 kb

Spectrum of resolution

- The ultimate resolution comes from direct sequence analysis, and an increasing number of new technologies have facilitated **comparisons of individual genomes** with the reference human genome sequence

Spectrum of resolution in chromosome and genome analysis

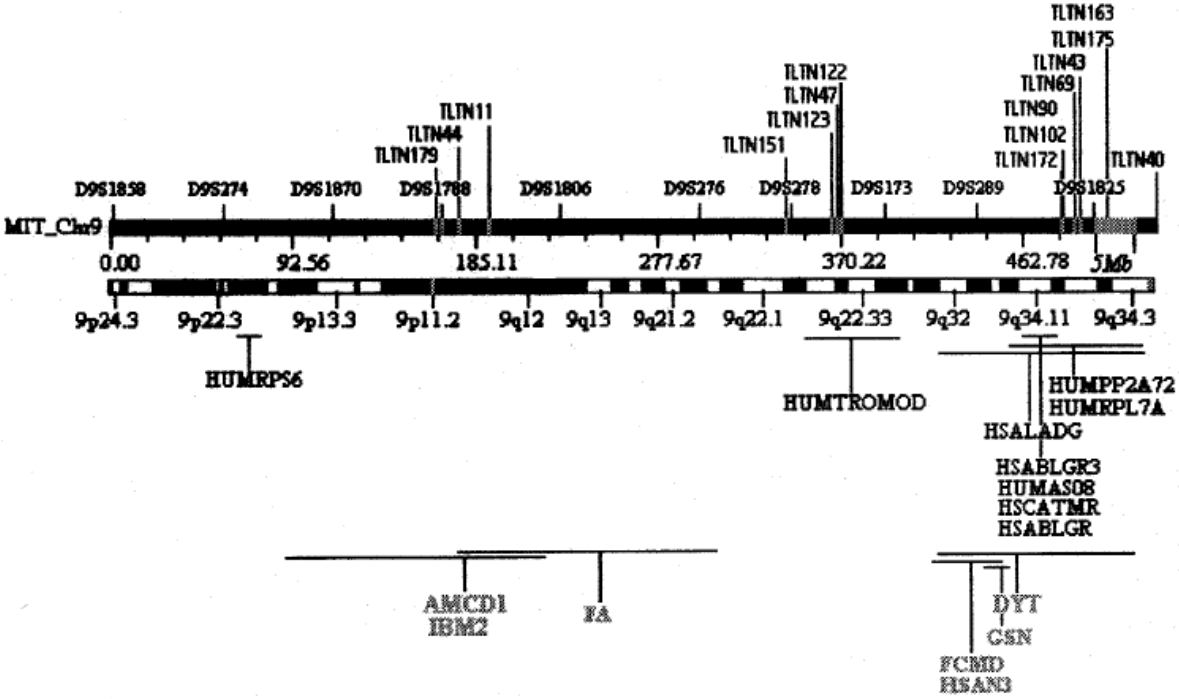


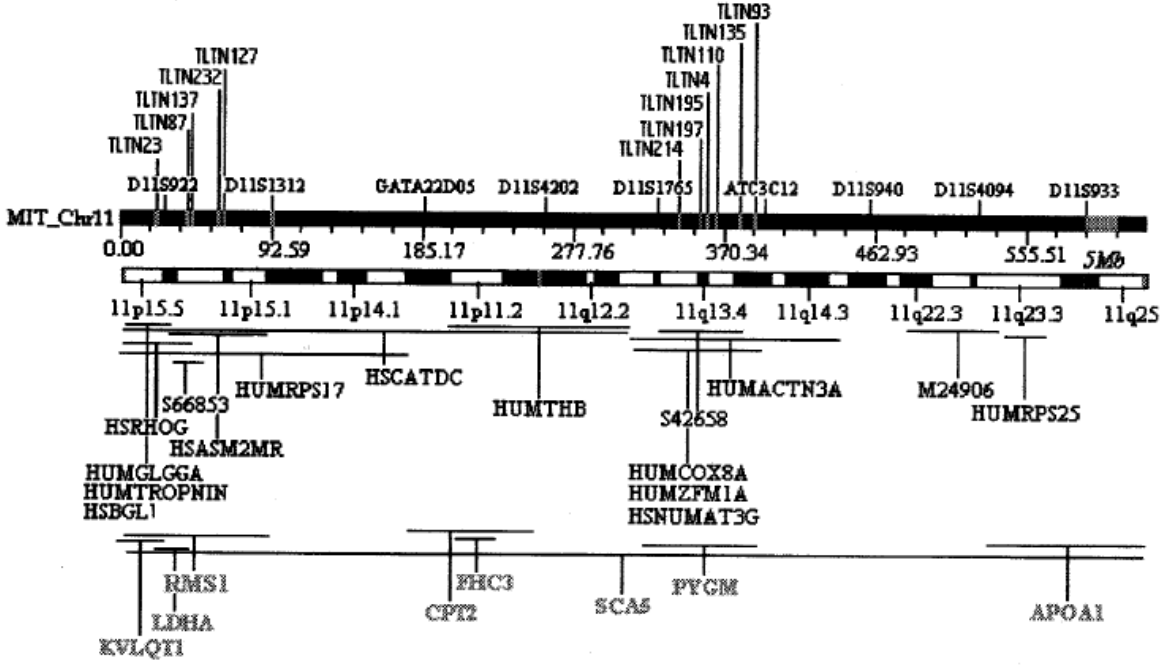
Genes in the Human Genome

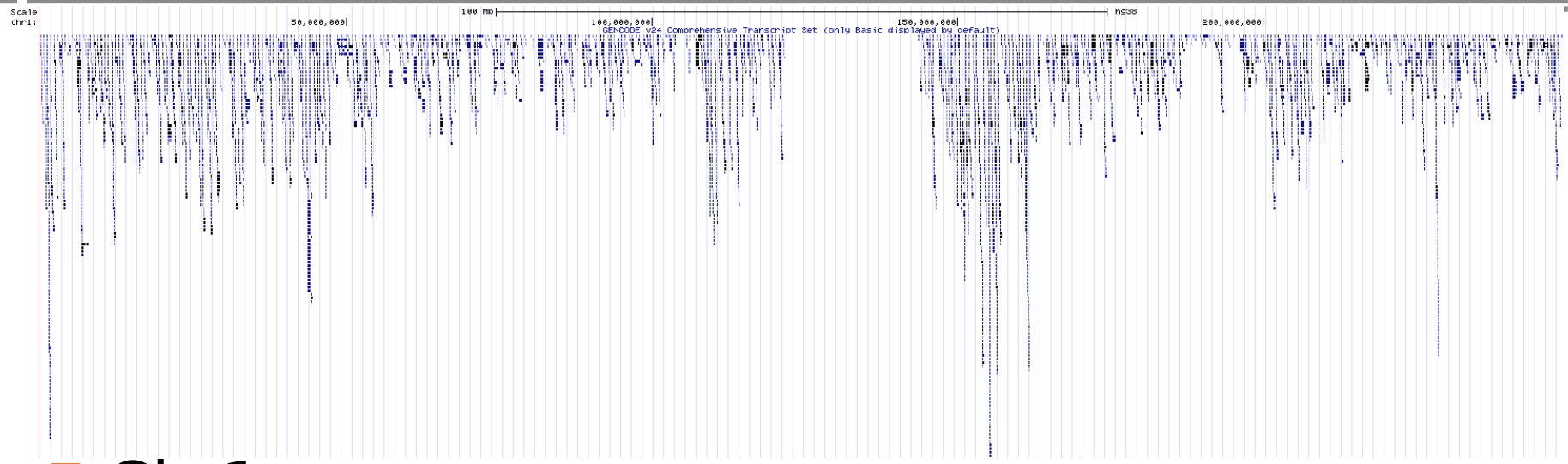
- the human genome contains an estimated 20,000 protein-coding genes
- there are some genes, including clinically relevant genes, that are currently undetected

Genes in the Human Genome

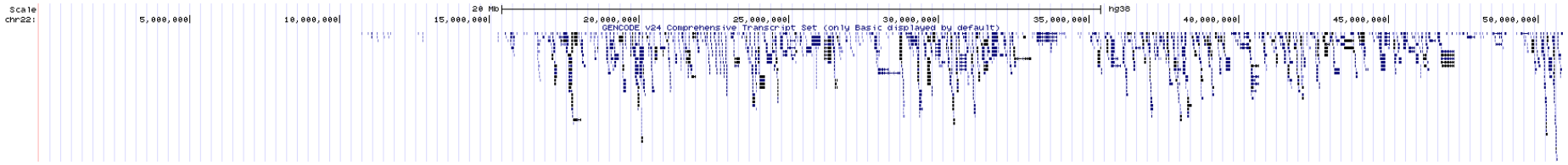
- In addition to being relatively sparse in the genome, genes are distributed quite nonrandomly along the different human chromosomes.
- Some chromosomes are relatively gene-rich, while others are quite gene-poor, ranging from approximately 3 genes/Mb of DNA to more than 20 genes/Mb


TLTN163
TLTN175
TLTN43
TLTN69
TLTN90
TLTN102
TLTN172
TLTN40
TLTN122
TLTN47
TLTN123
TLTN151
TLTN11
TLTN44
TLTN179
D9S1858
D9S274
D9S1870
D9S1788
D9S1806
D9S276
D9S278
D9S173
D9S289
D9S1825
MIT_Chr9
0.00
92.56
185.11
277.67
370.22
462.78
5Mb
9p24.3
9p22.3
9p13.3
9p11.2
9q12
9q13
9q21.2
9q22.1
9q22.33
9q32
9q34.11
9q34.3
HUMRPS6
HUMTROMOD
HUMPP2A72
HUMRPL7A
HSALADG
HSABLGR3
HUMAS08
HSCATMR
HSABLGR
AMCD1
IBM2
FA
DYT
GSN
FCMD
HSAN3


TLTN93
TLTN135
TLTN110
TLTN4
TLTN195
TLTN197
TLTN214
TLTN127
TLTN232
TLTN137
TLTN87
TLTN23
D11S922
D11S1312
GATA22D05
D11S4202
D11S1765
ATC3C12
D11S940
D11S4094
D11S933
MIT_Chr11
0.00
92.59
185.17
277.76
370.34
462.93
555.51
5Mb
11p15.5
11p15.1
11p14.1
11p11.2
11q12.2
11q13.4
11q14.3
11q22.3
11q23.3
11q25
HUMRPS17
HSCATDC
HUMTHB
HUMACTN3A
M24906
HUMRPS25
S66853
HSRHOG
HSASM2MR
S42658
HUMGLGGA
HUMTROPNIN
HSBGL1
HUMCOX8A
HUMZFM1A
HSNUMAT3G
RMS1
LDHA
KVLQT1
FHC3
CPT2
SCA5
PYGM
APOA1


Scale
chr1:
50,000,000
100 Mb
100,000,000
150,000,000
hg38
200,000,000
GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default)

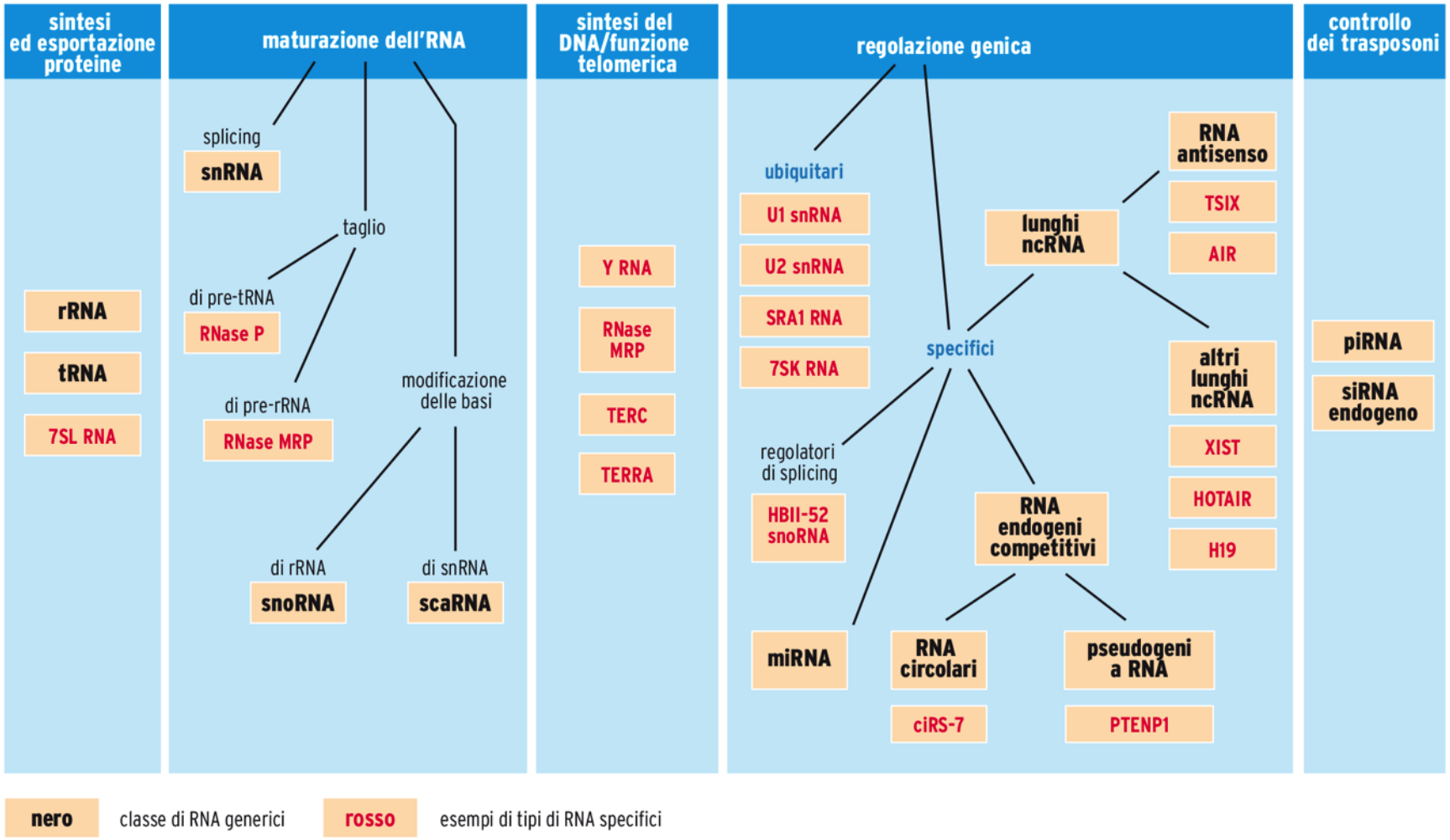
- Chr1
- Chr22


Scale
chr22:
5,000,000
10,000,000
15,000,000
20 Mb
20,000,000
25,000,000
30,000,000
35,000,000
hg38
40,000,000
45,000,000
50,000,000
GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default)

Coding and Noncoding Genes

- There are a number of different types of gene in the human genome.
  - Most genes known or thought to be clinically relevant are protein coding
  - additional genes whose functional product appears to be the RNA itself

ncRNA


sintesi ed esportazione proteine
rRNA
tRNA
7SL RNA
maturazione dell'RNA
splicing
snRNA
taglio
di pre-tRNA
RNase P
di pre-rRNA
RNase MRP
modificazione delle basi
di rRNA
snoRNA
di snRNA
scaRNA
sintesi del DNA/funzione telomerica
Y RNA
RNase MRP
TERC
TERRA
regolazione genica
ubiquitari
U1 snRNA
U2 snRNA
SRA1 RNA
7SK RNA
RNA antisenso
TSIX
AIR
lunghi ncRNA
specifici
altri lunghi ncRNA
XIST
HOTAIR
H19
regolatori di splicing
HBII-52 snoRNA
RNA endogeni competitivi
miRNA
RNA circolari
ciRS-7
pseudogeni a RNA
PTENP1
controllo dei trasposoni
piRNA
siRNA endogeno
nero
classe di RNA generici
rosso
esempi di tipi di RNA specifici

Variation in the Human Genome

- Any given individual carries **4–5 million sequence variants** that are known to exist in multiple forms (i.e., are polymorphic) in our species.

- each and every base pair in the human genome is expected to vary in someone somewhere around the globe.

Variation in the Human Genome

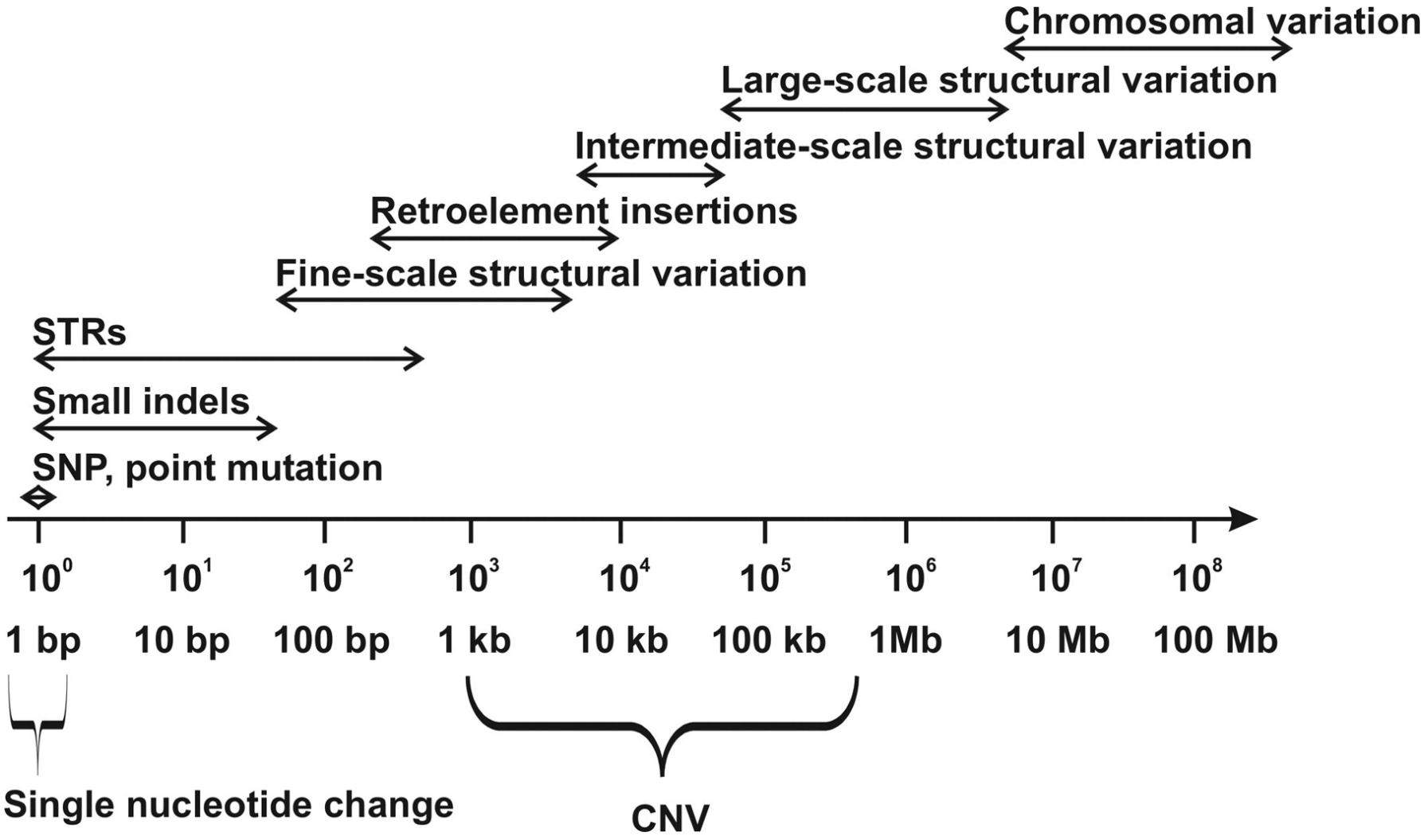
- **Types of Variation**
  - any two randomly selected individuals have sequences that are 99.9% identical

- The majority of these differences involve simply a single unit in the DNA code and are referred to as single-nucleotide polymorphisms (SNPs)

Variation in the Human Genome

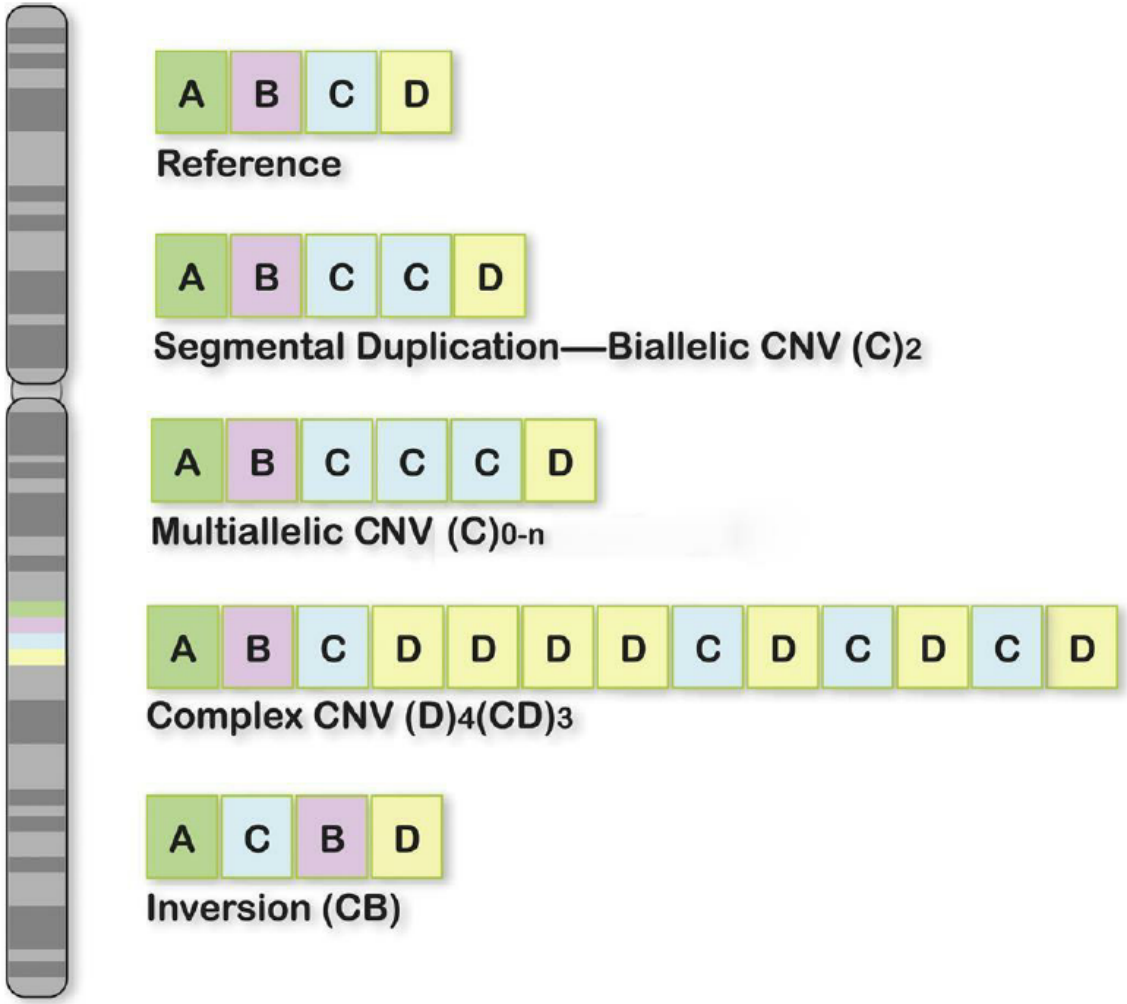
- The remaining variation consists of insertions or deletions (**in/dels**) of (usually) short sequence stretches, variation in the number of copies of repeated elements or inversions in the order of sequences at a particular locus in the genome

Variation in the Human Genome



Variation in the Human Genome

- Schematic representation of different types of structural polymorphism in the human genome, leading to deletions, duplications, inversions, and CNV changes relative to the reference arrangement.



Common Variation in the Human Genome

Type of Variation	Size Range (approx.)[a]	Effect(s) in Biology and Medicine
Single-nucleotide polymorphisms	1 bp	Nonsynonymous → functional change in encoded protein?
		Others → potential regulatory variants?
		Most → no effect? ("neutral")
Copy number variants (CNVs)	10 kb to 1 Mb	Gene dosage variation → functional consequences?Most → no effect or uncertain effect
Insertion/deletion polymorphisms (in/dels)	1 bp to 1 Mb	In coding sequence: frameshift mutation? → functional change
		Most → uncertain effects
Inversions	Few bp to 100 kb	? break in gene sequence
		? long-range effect on gene expression
		? indirect effects on reproductive fitness
		Most → no effect? ("neutral")
Segmental duplications	10 kb to >1 Mb	Hotspots for recombination → polymorphism (CNVs)

Variation in the Human Genome

- http://www.internationalgenome.org/abou

Populations: - African; - American; - East Asian; - European; - South Asian;

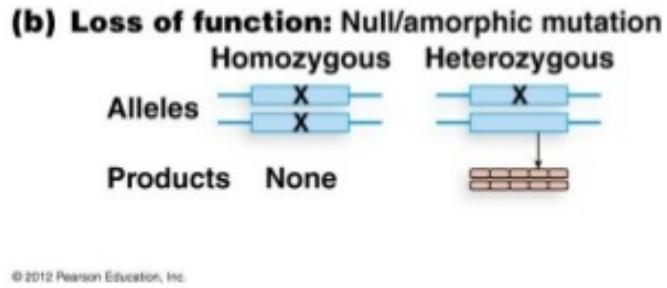
Variation in Individual Genomes

the 1000 Genomes Project concluded that each genome carries

- 100 or more likely loss-of-function mutations
- 10,000 nonsynonymous changes
- 500,000 variants that overlap known gene regulatory regions.

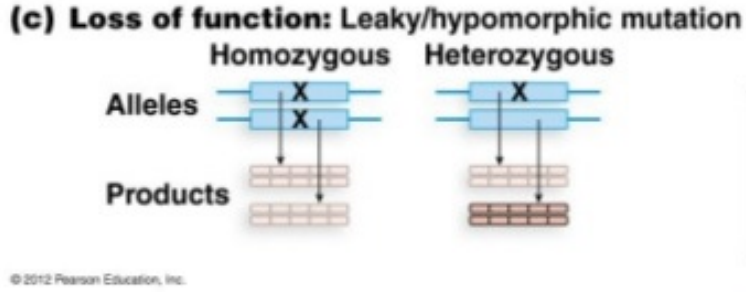
Variation in Individual Genomes

thousands of genes in the human genome are highly tolerant to many mutations that appear likely to result in a loss of function



Null alleles produce no functional product. Homozygous null organisms have mutant (amorphic) phenotype due to absence of the gene product. Heterozygous organisms produce less functional gene product than homozygous wild-type organisms and may have mutant phenotype. See text for discussion of dominant versus recessive mutations.

Amorphic = no function



Leaky mutant alleles produce a small amount of wild-type gene product. Homozygous organisms have a mutant (hypomorphic) phenotype. Heterozygous organisms may also be mutant.

Hypomorphic = less function

De NOVO MUTATION

- studies have shown that any individual carries an estimated 30–70 new mutations per genome that were not present in the genomes of his or her parents.
- generation of a new length variant depends on recombination, rather than on errors in DNA synthesis to generate a new base pair
- the measured rate of formation of new CNVs is orders of magnitude higher than that of base substitutions

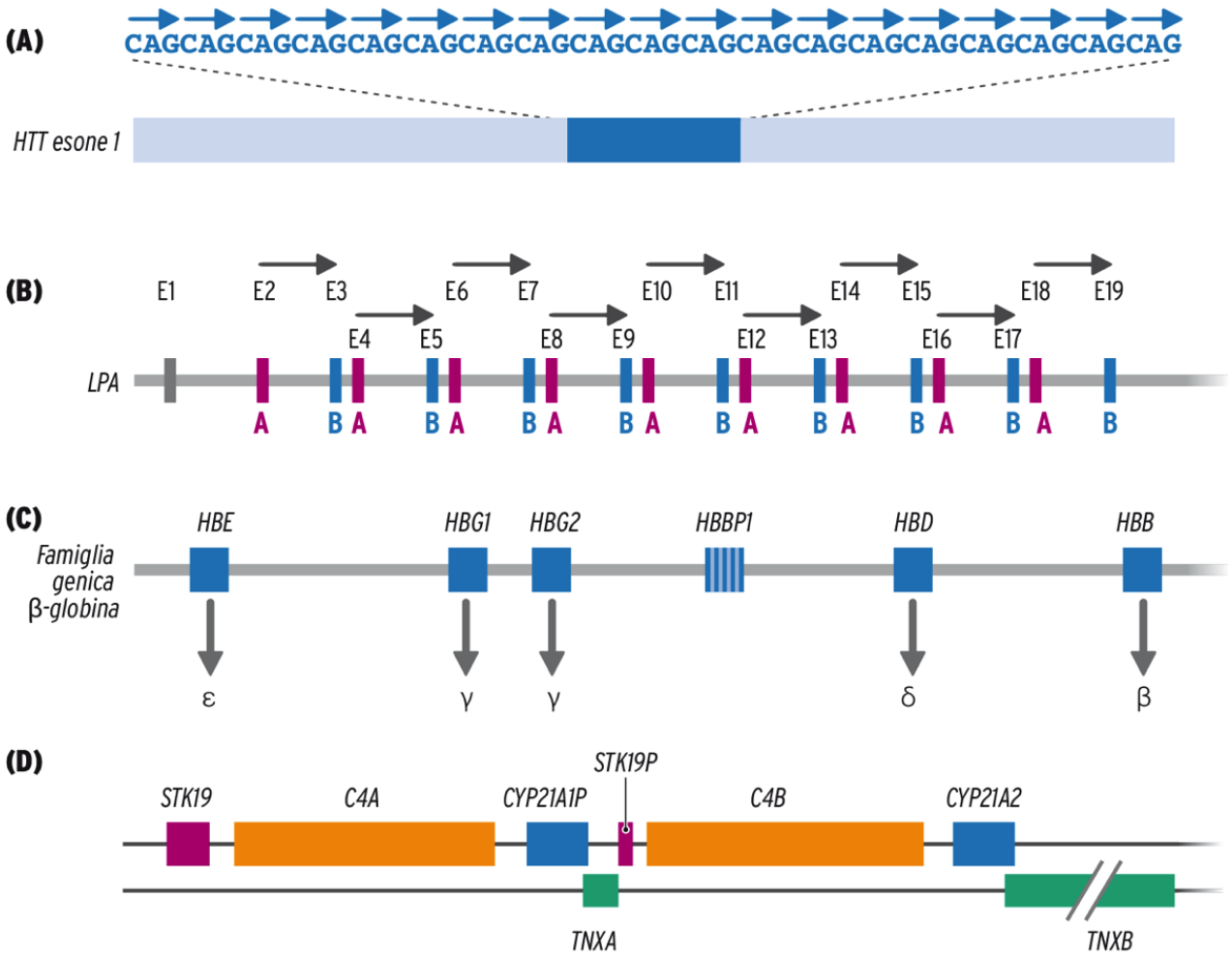
Variation in Individual Genomes

the number of SNPs described for our species is still incomplete

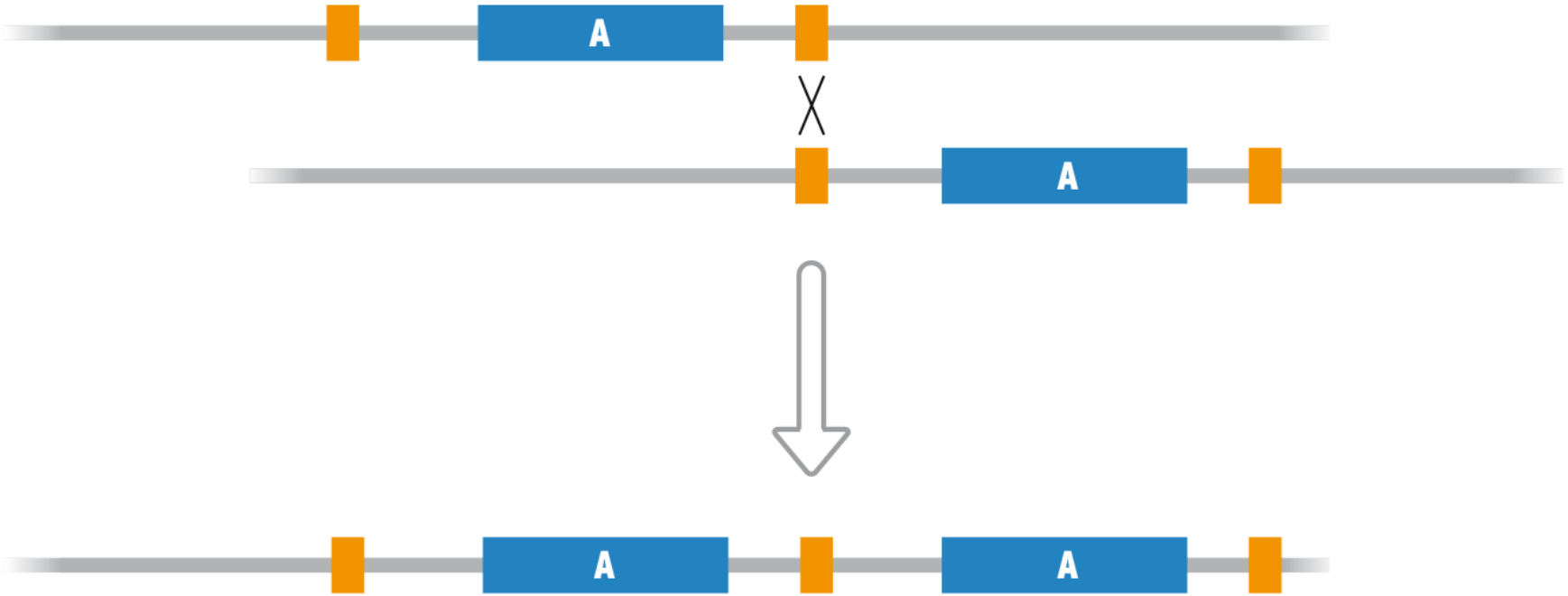
each genome carries thousands of nonsynonymous SNPs

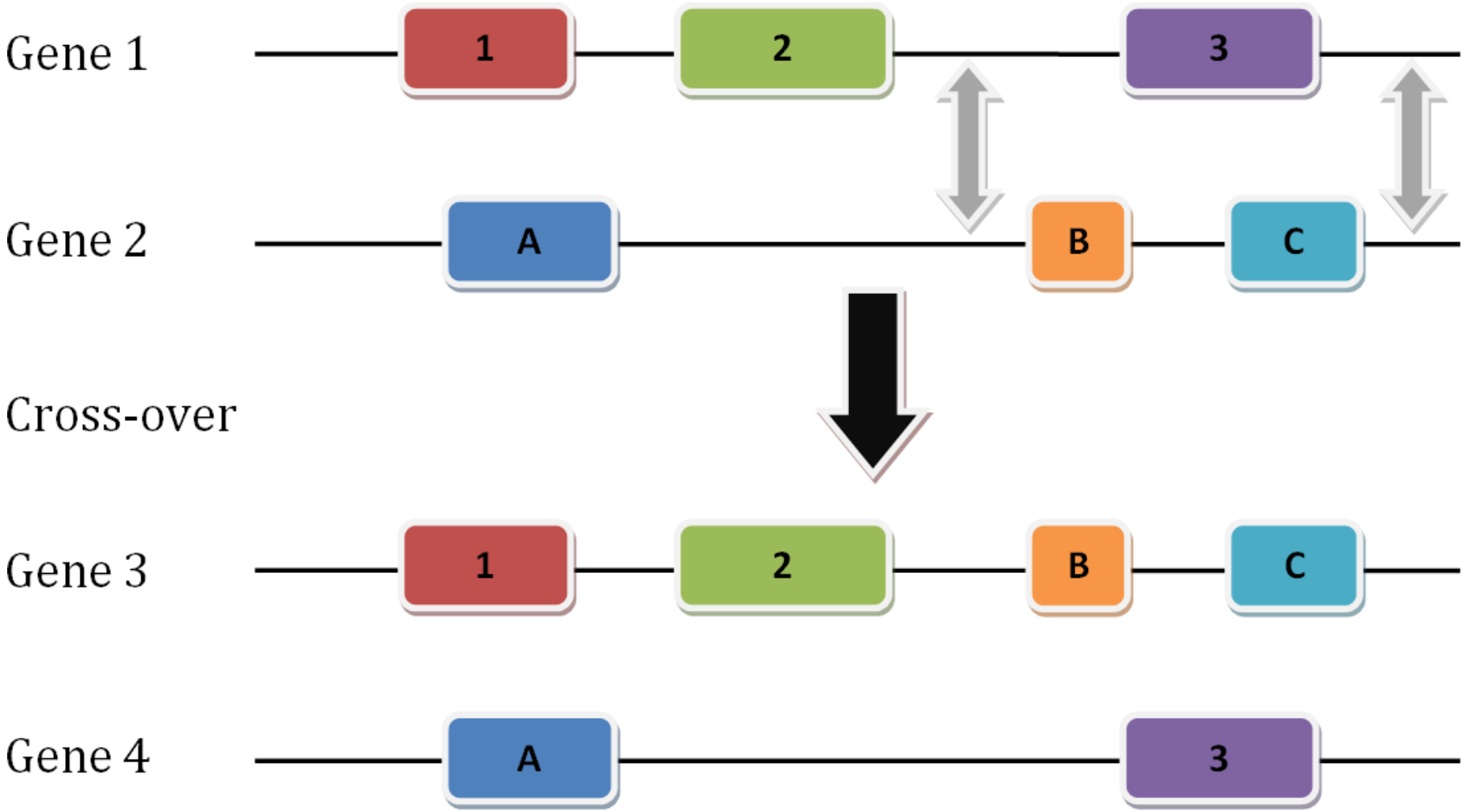
These measurements underscore the potential impact of gene and genome variation on human biology and on medicine.

Variation in Individual Genomes

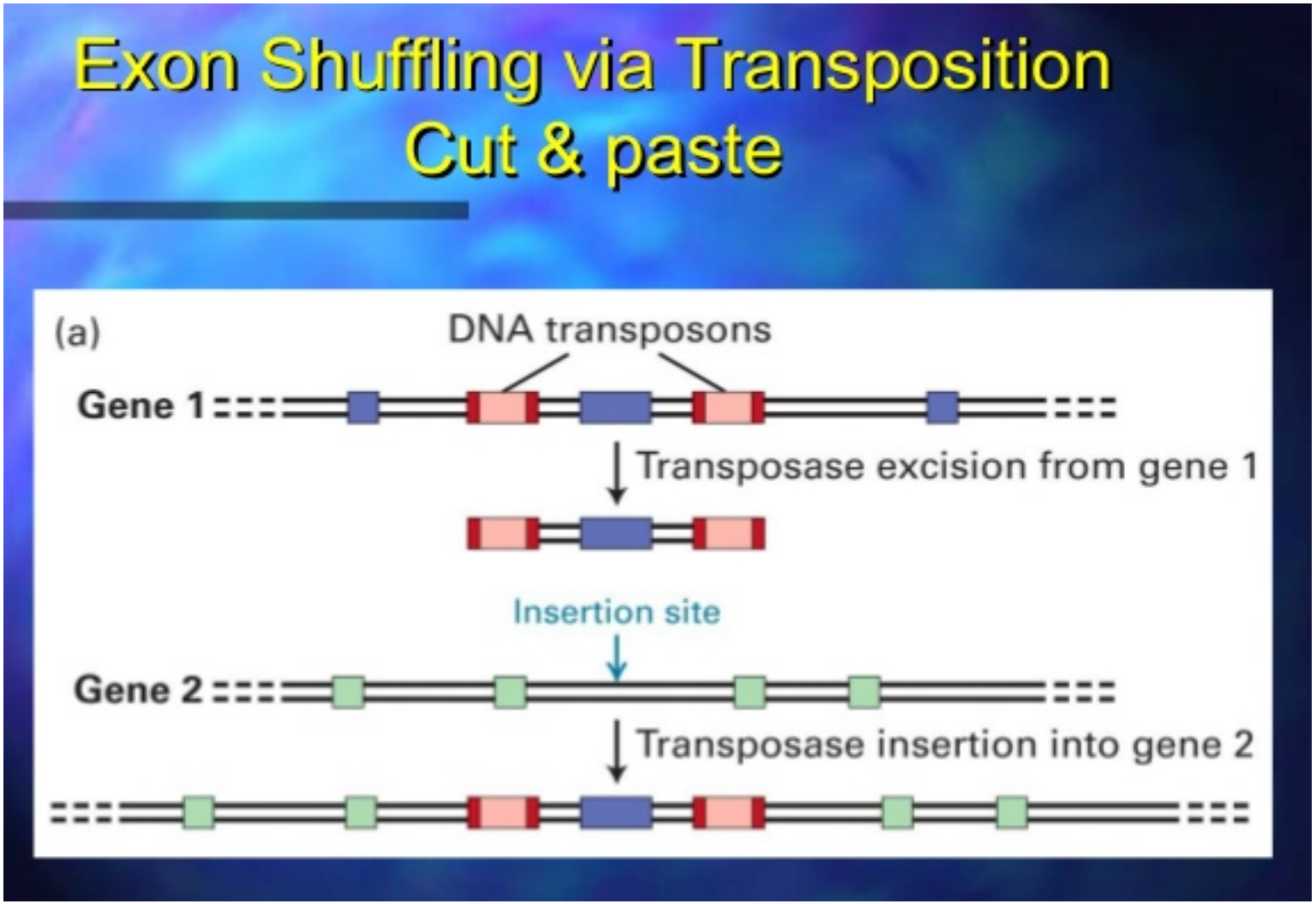


Variation in Individual Genomes




Gene 1
1
2
3
Gene 2
A
B
C
Cross-over
Gene 3
1
2
B
C
Gene 4
A
3

Variation in Individual Genomes



Copy Number Variation

Over the past decade, a number of important studies have focused on the prevalence of structural variants in the genome,

collectively account for far more variation in genome sequence (expressed in terms of the amount of genomic DNA affected) than do SNPs

Copy number variation

Global variation in copy number in the human genome

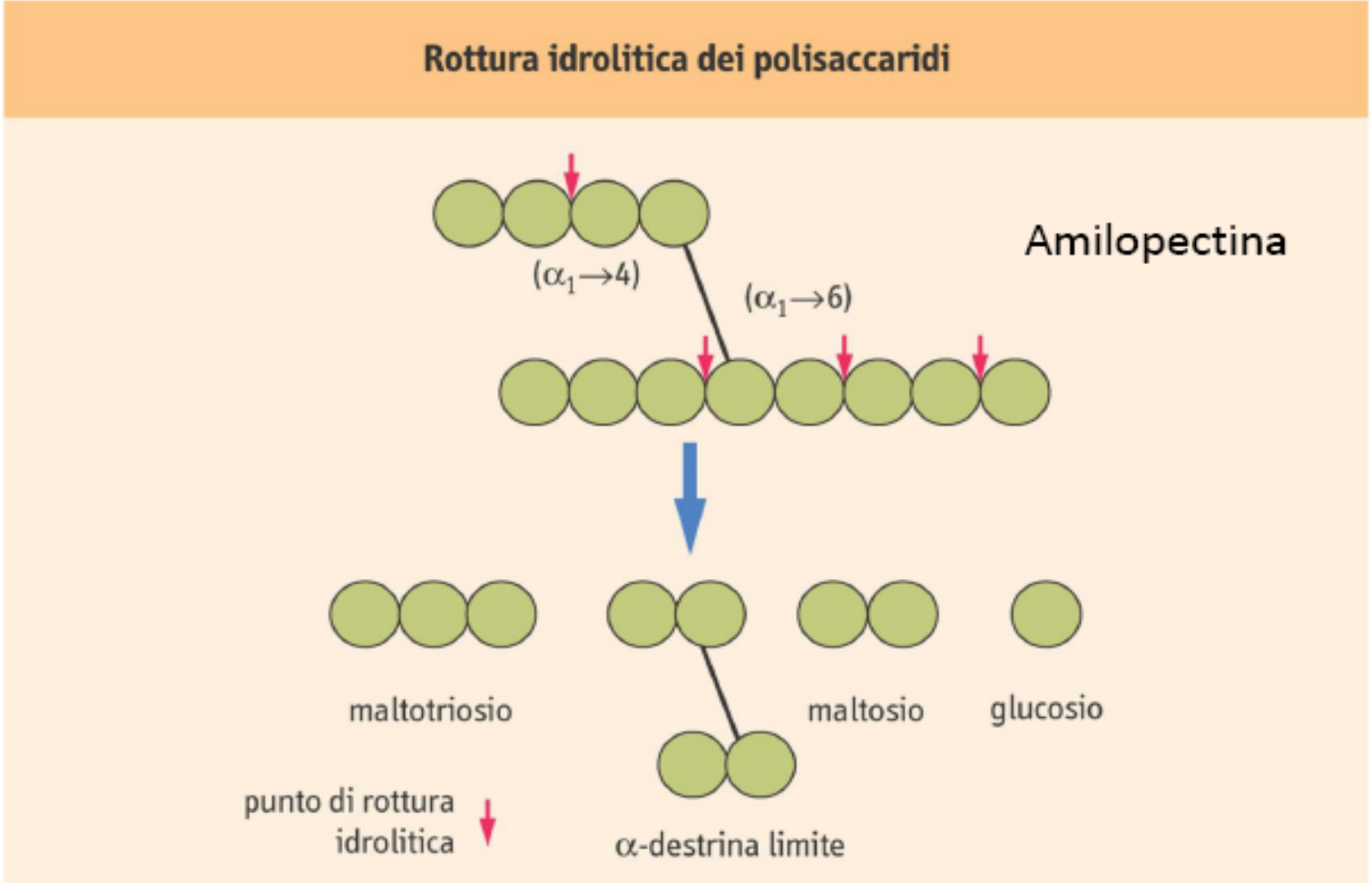
Richard Redon[1], Shumpei Ishikawa[2,3], Karen R. Fitch[4], Lars Feuk[5,6], George H. Perry[7], T. Daniel Andrews[1], Heike Fiegler[1], Michael H. Shapero[4], Andrew R. Carson[5,6], Wenwei Chen[4], Eun Kyung Cho[7], Stephanie Dallaire[7], Jennifer L. Freeman[7], Juan R. González[8], Mònica Gratacòs[8], Jing Huang[4], Dimitrios Kalaitzopoulos[1], Daisuke Komura[3], Jeffrey R. MacDonald[5], Christian R. Marshall[5,6], Rui Mei[4], Lyndal Montgomery[1], Kunihiro Nishimura[2], Kohji Okamura[5,6], Fan Shen[4], Martin J. Somerville[9], Joelle Tchinda[7], Armand Valsesia[1], Cara Woodwark[1], Fengtang Yang[1], Junjun Zhang[5], Tatiana Zerjal[1], Jane Zhang[4], Lluis Armengol[8], Donald F. Conrad[10], Xavier Estivill[8,11], Chris Tyler-Smith[1], Nigel P. Carter[1], Hiroyuki Aburatani[2,12], Charles Lee[7,13], Keith W. Jones[4], Stephen W. Scherer[5,6] & Matthew E. Hurles[1]

Copy number variation (CNV) of DNA sequences is functionally significant but has yet to be fully ascertained. We have constructed a first-generation CNV map of the human genome through the study of 270 individuals from four populations with ancestry in Europe, Africa or Asia (the HapMap collection). DNA from these individuals was screened for CNV using two complementary technologies: single-nucleotide polymorphism (SNP) genotyping arrays, and clone-based comparative genomic hybridization. A total of 1,447 copy number variable regions (CNVRs), which can encompass overlapping or adjacent gains or losses, covering 360 megabases (12% of the genome) were identified in these populations. These CNVRs contained hundreds of genes, disease loci, functional elements and segmental duplications. Notably, the CNVRs encompassed more nucleotide content per genome than SNPs, underscoring the importance of CNV in genetic diversity and evolution. The data obtained delineate linkage disequilibrium patterns for many CNVs, and reveal marked variation in copy number among populations. We also demonstrate the utility of this resource for genetic disease studies.

http://www.nature.com/nature/journal/v444/n7118/abs/nature05329.html

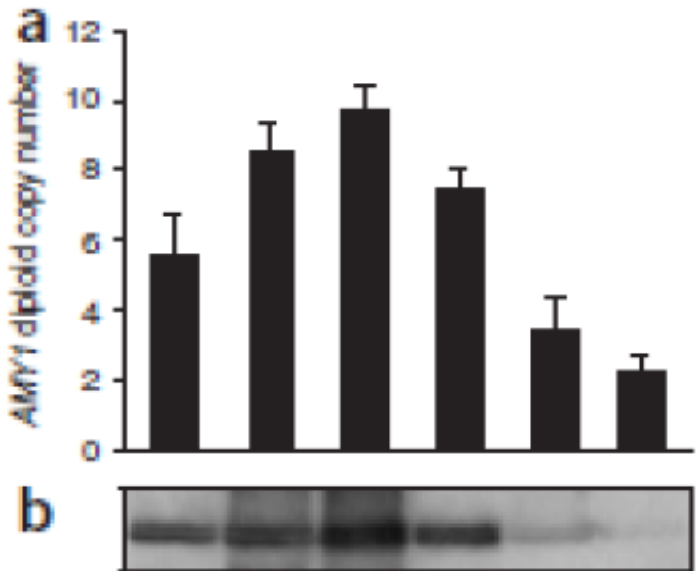
Amilasi

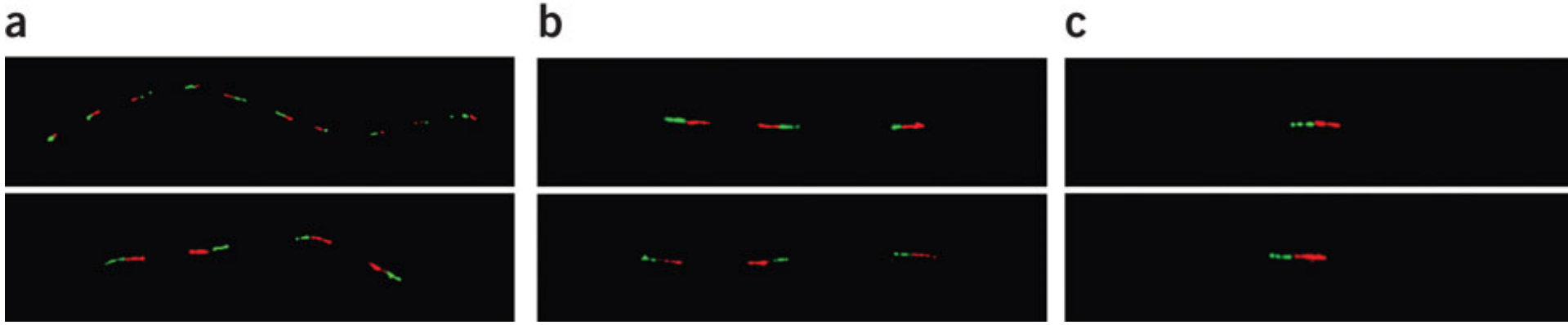
2 forme isoenzimatiche
alfa-amilasi salivare o ptialina
alfa-amilasi pancreatica



•Il gene per l'amilasi salivare *AMY1* è presente in copia multipla nel genoma
• il numero di copie varia tra gli individui e tra le popolazioni
e corrisponde all'espressione della proteina nella saliva

a) Numero di copie del gene in diversi individui





http://www.nature.com/ng/journal/v39/n10/abs/ng2123.html