BIOL 206: Introduction to Genetics

Eric T. Rush, MD, FAAP, FACMG
Clinical Geneticist
Children's Mercy Kansas City
Associate Professor of Pediatrics
University of Missouri – Kansas City School of Medicine

Agenda

- Introduction to concepts and terminology
- Description of how the information contained within DNA is used
- Historical context and technology used to understand genetic changes

What is genetics?

- Study of heredity and variation of essential character
  - Heredity – tendency of offspring to resemble their parents
  - Variation – tendency of offspring to vary from their parents
- The term “genetics” was first coined in 1905 to describe the study of heredity
  - Greek “genno”, *γεννώ; “to give birth”*

Genetic Terminology

- **GENOTYPE**: genetic make-up/composition
- **PHENOTYPE**: aspects of a person you can see
- **MUTATION**: alteration in the genetic code from what is expected
  - Does not necessarily communicate a change is pathogenic, so we generally use the term “variant”
- **Benign vs. pathogenic variant**
  - **BENIGN**: not responsible for identified effects
  - **PATHOGENIC**: responsible for identified effects
  - **VARIANT OF UNCERTAIN SIGNIFICANCE:** unclear if responsible for identified effects

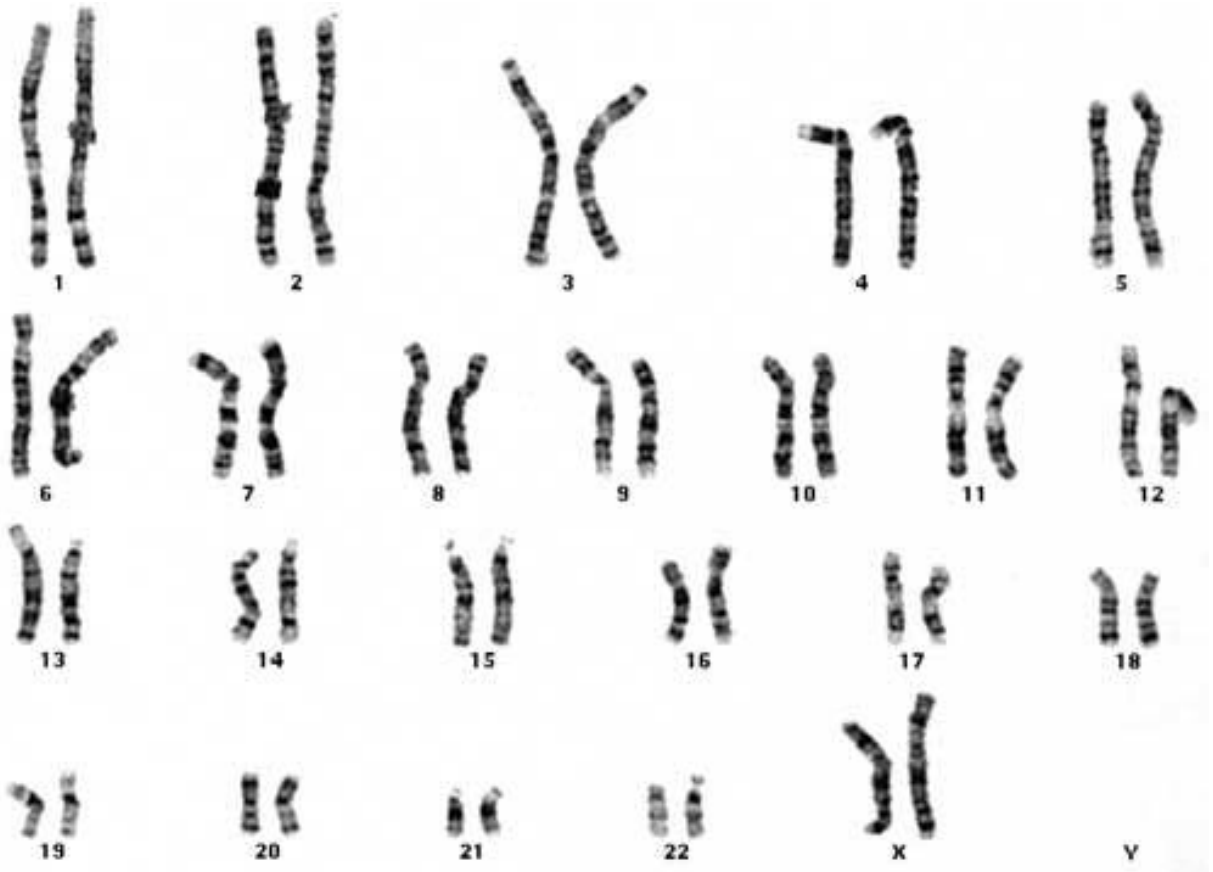
Genetic versus Genomic

- Often used interchangeably in the lay press, but don't make this mistake!
- GENETIC testing refers to specific testing of areas of known function
- GENOMIC testing refers to evaluation of large segments across the entirety of genetic material which may or may not have known function

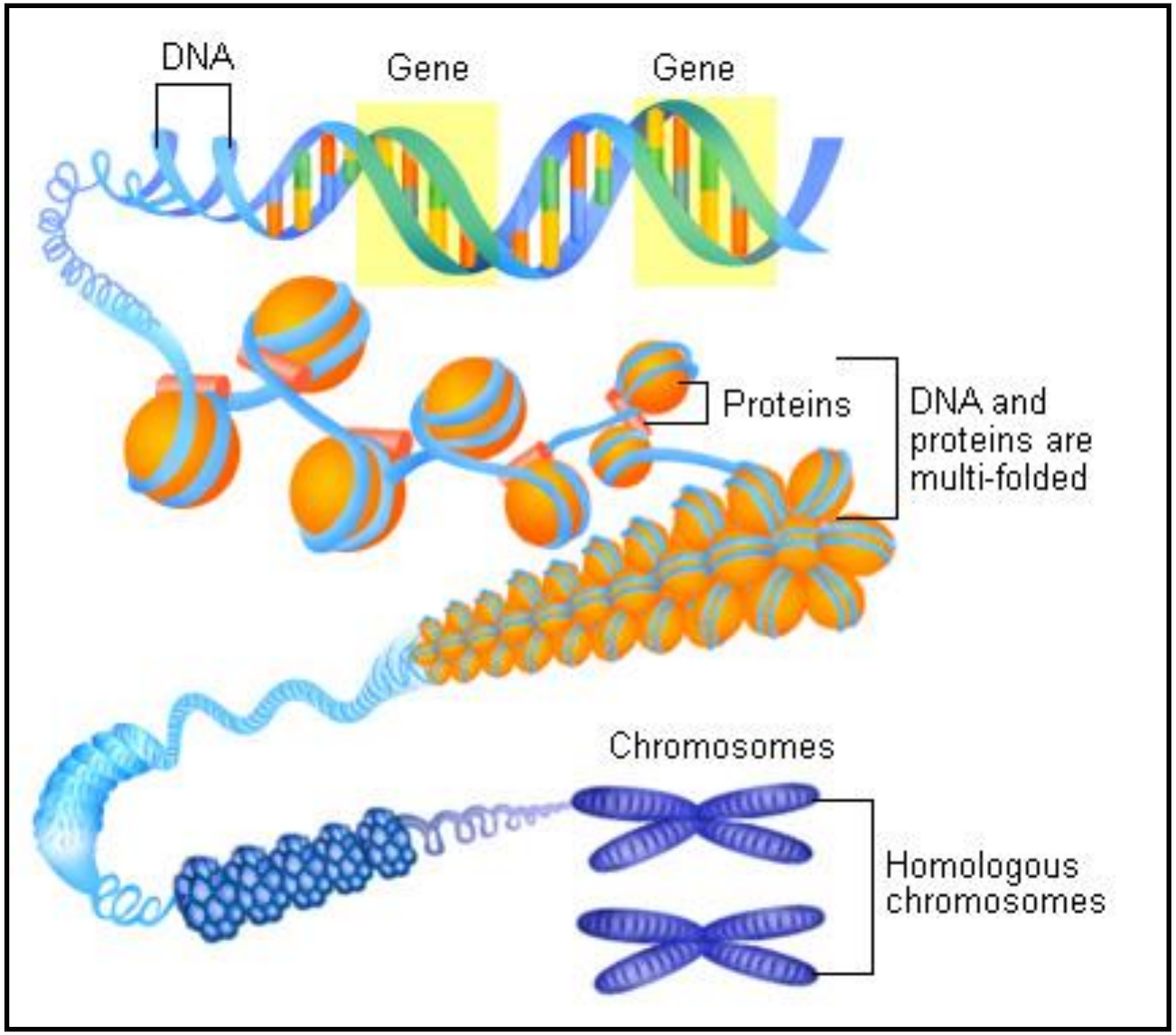
What are Chromosomes?

- Chromosomes are the structures that store genetic information in the form of DNA
- Humans have 46 chromosomes
- 22 pairs = autosomes
- 1 pair = sex chromosomes (Most phenotypic males have one X and one Y chromosomes, while most phenotypic females have two X chromosomes)

Normal Female Karyotype

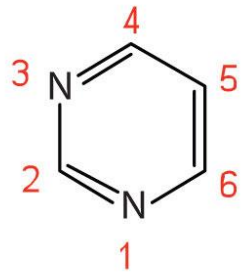


What are DNA and RNA?

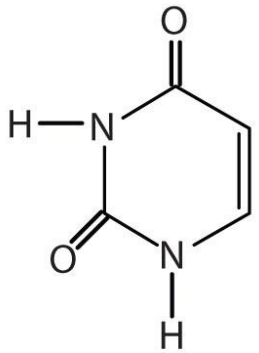


http://youngbloodbiology.wikispaces.com/file/view/chromosome1.jpg/111553413/chromosome1.jpg

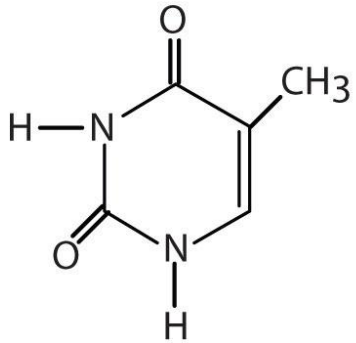
Nucleotides



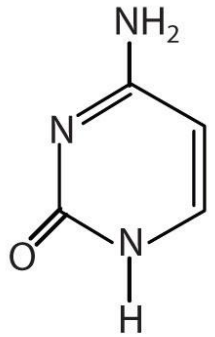
Pyrimidine



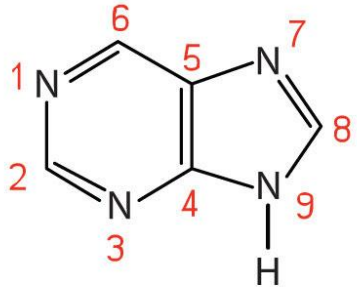
Uracil (U)
RNA only



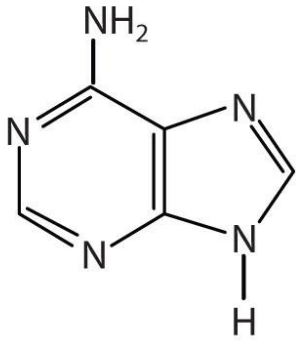
Thymine (T)
DNA only



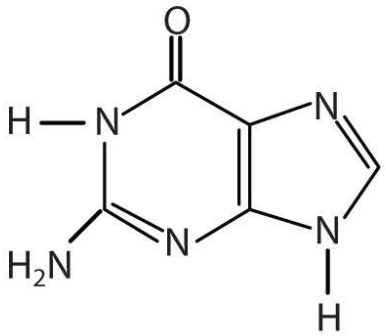
Cytosine (C)
both DNA and RNA



Purine



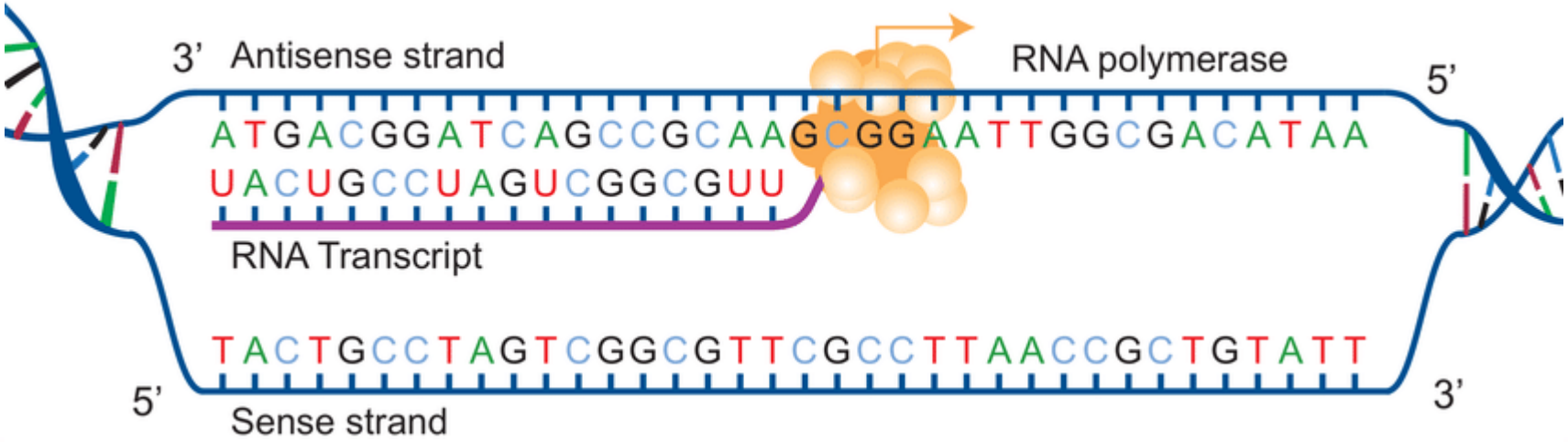
Adenine (A)



Guanine (G)

Transcription

- Messenger RNA (mRNA) is transcribed from a DNA template by RNA polymerase
- 3 phases: initiation, elongation, and termination



Translation

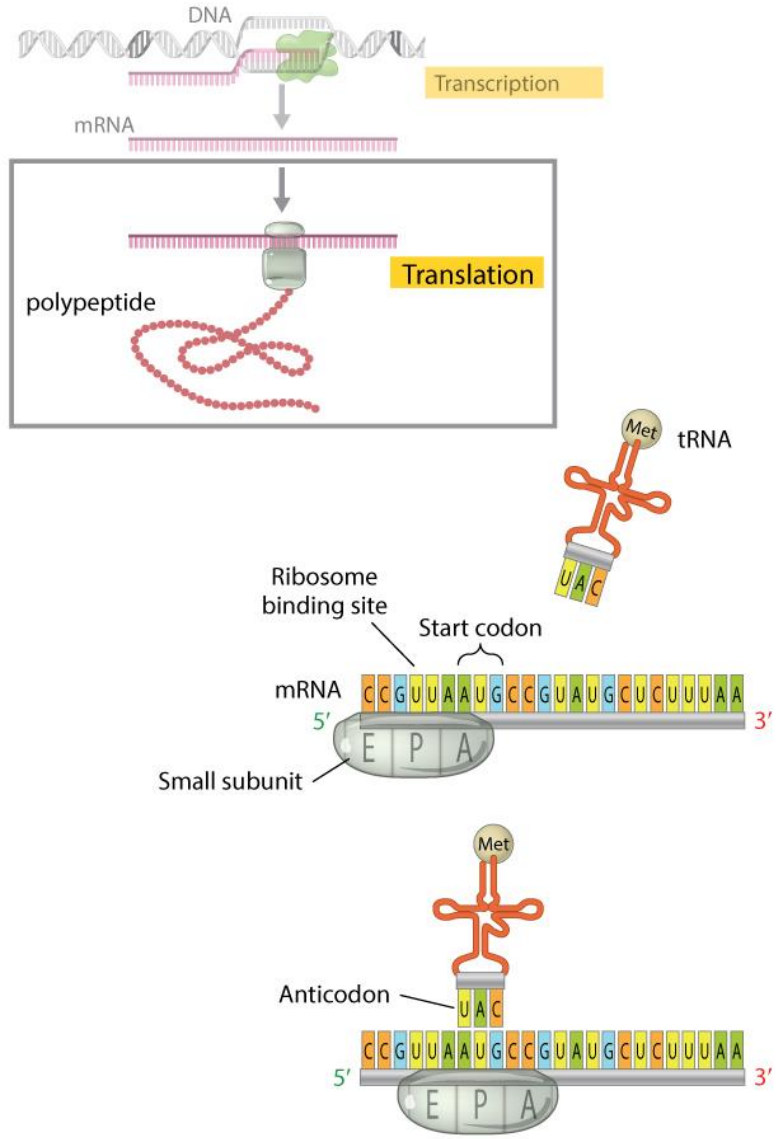
mRNA molecules leaves the nucleus and travels to the ribosome

mRNA sequence attaches to site in the ribosome

tRNA molecules, each with a complementary anti-codon sequence to mRNA also bears a specific amino acid

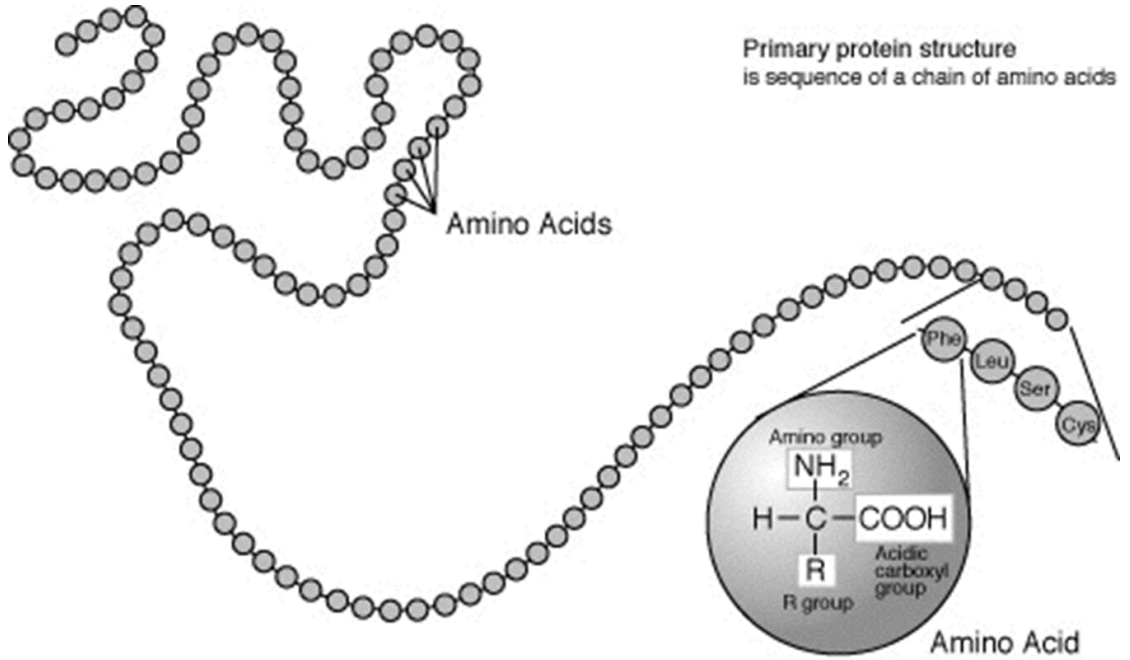
Amino acids are released and incorporated into the growing peptide chain

After peptide completed, process terminates and is released from ribosome



Second base of codon

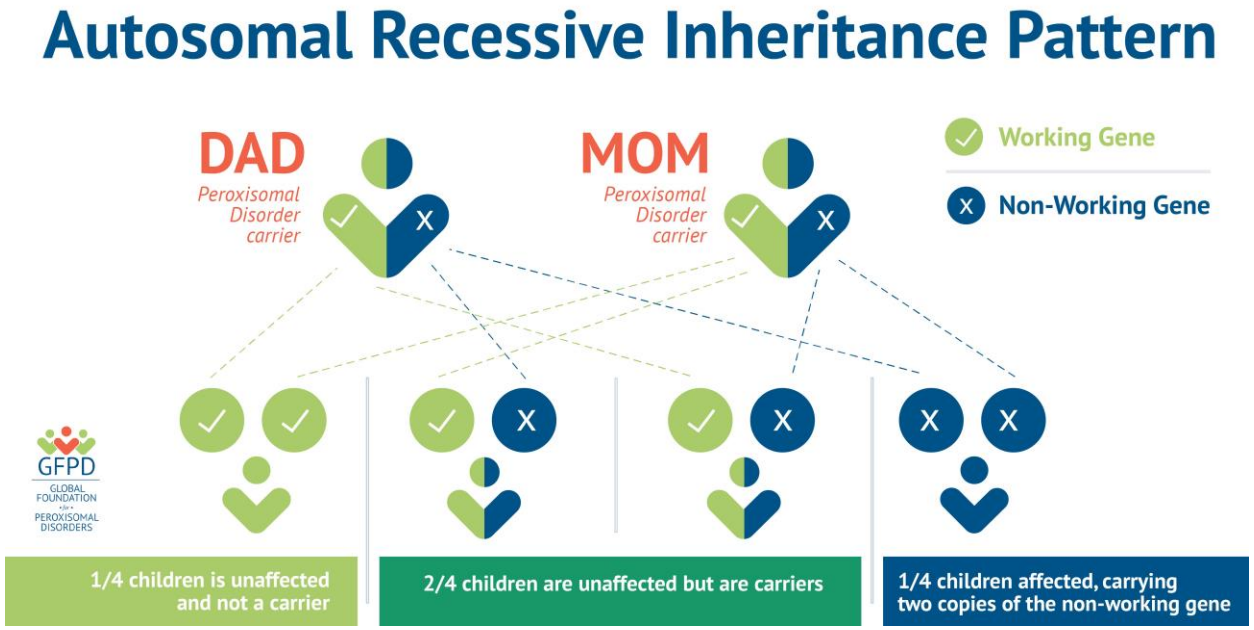
First base of codon	U		C		A		G		Third base of codon
U	UUU	Phenylalanine phe	UCU	Serine ser	UAU	Tyrosine tyr	UGU	Cysteine cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine leu	UCA		UAA	STOP codon	UGA	STOP codon	A
	UUG		UCG		UAG		UGG	Tryptonphan trp	G
C	CUU	Leucine leu	CCU	Proline pro	CAU	Histidine his	CGU	Arginine arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine gin	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine ile	ACU	Threonine thr	AAU	Asparagine asn	AGU	Serine ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine lys	AGA	Arginine arg	A
	AUG	Methionine met (start codon)	ACG		AAG		AGG		G
G	GUU	Valine val	GCU	Alanine ala	GAU	Aspartic acid asp	GGU	Glycine gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid glu	GGA		A
	GUG		GCG		GAG		GGG		G



Inheritance Patterns

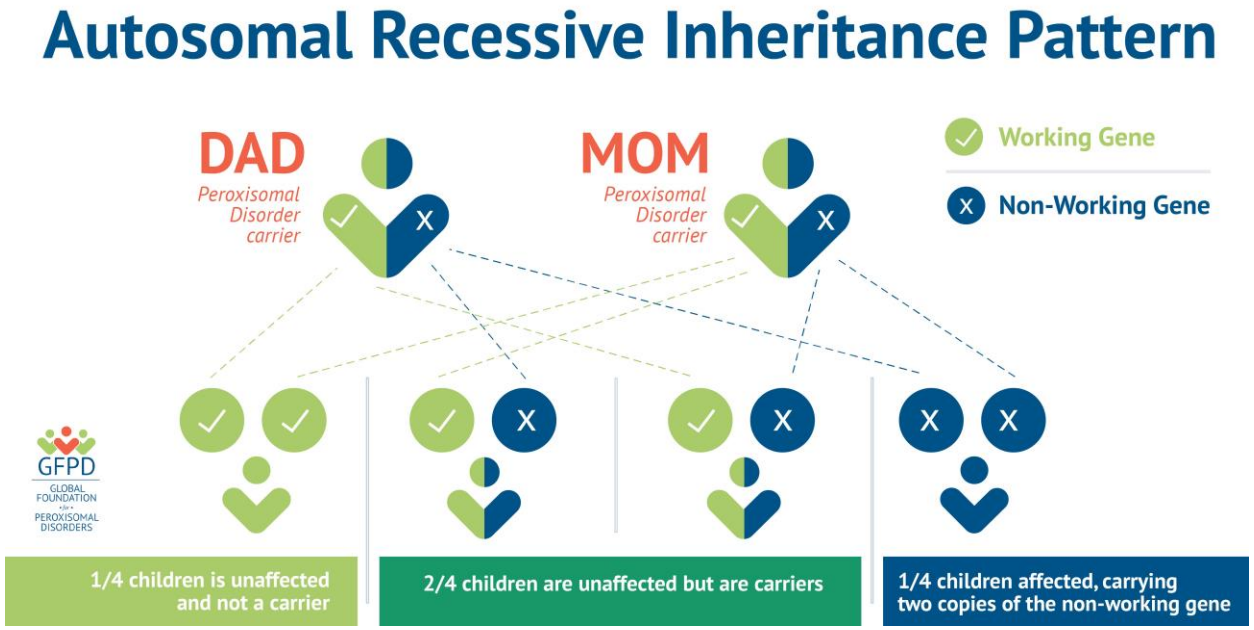
- Autosomal Recessive
  - The gene in question is located on an autosome (not X-chromosome)
  - Both copies of a gene (from mother and father) must have a variant (mutation) that reduces or eliminates normal function to have a disease
  - A person with such a change to one copy of a gene will typically be healthy, but can pass on to a child
- Autosomal Dominant
  - The gene in question is located on an autosome
  - Only one copy of a gene must have a variant that alters normal function to have a disease

Recurrence Risk and Epidemiology



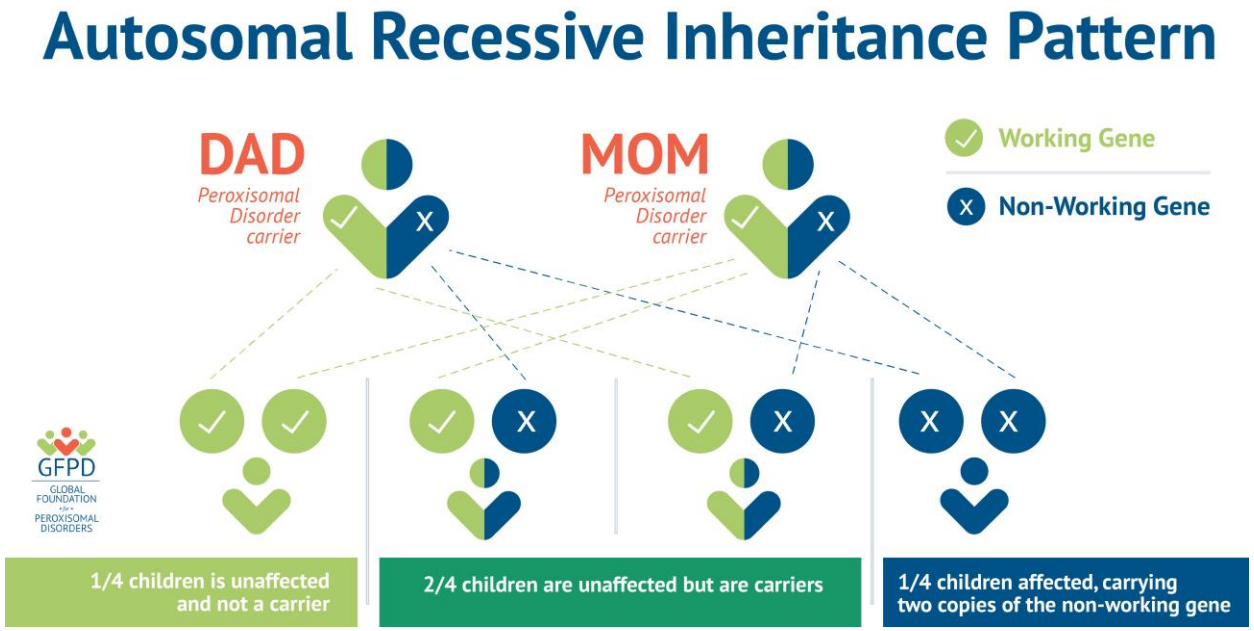
- When both parents are carriers for a disease, the risk of a child inheriting the copy of the altered gene from both parents is 1 in 4 or 25%

Recurrence Risk and Epidemiology



- When both parents are carriers for a disease, the risk of a child inheriting a copy of the altered gene from one, but not both parents is 1 in 2 or 50%

Recurrence Risk and Epidemiology



Q: Why is it that most families with ZSD don't have a known family history if the risk is 1 in 4?

A: Because the 1 in 4 risk only comes into play when two carriers have already formed a mating pair. The odds are against that pair meeting in the first place.

Time out for Terminology

- Homozygous – both copies of a gene have the same genetic change (e.g. *PEX1* p.Gly843Asp/Gly843Asp)
- Compound Heterozygous – both copies of a gene are altered, but the change itself is different (e.g. *PEX1* p.Gly843Asp/p.Arg135Ter)
- Heterozygous – one copy of a gene is altered (e.g. *PEX1* p.Gly843Asp/WT)
  - "Carrier" is a term often used to describe heterozygous state in recessive disease

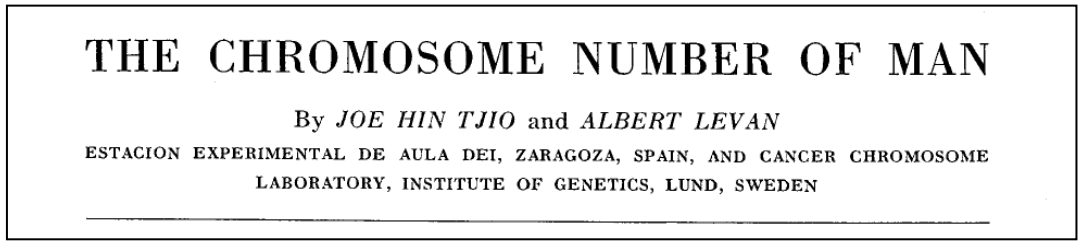
DNA Sequencing

- The process of reading the order of the nucleotide bases in a DNA molecule

- Sequence alteration: Changes in the genetic code at the **base pair** level

Control: AAAGGCGCG**T**ATAGGC
Patient: AAAGGCGCG**C**ATAGGC

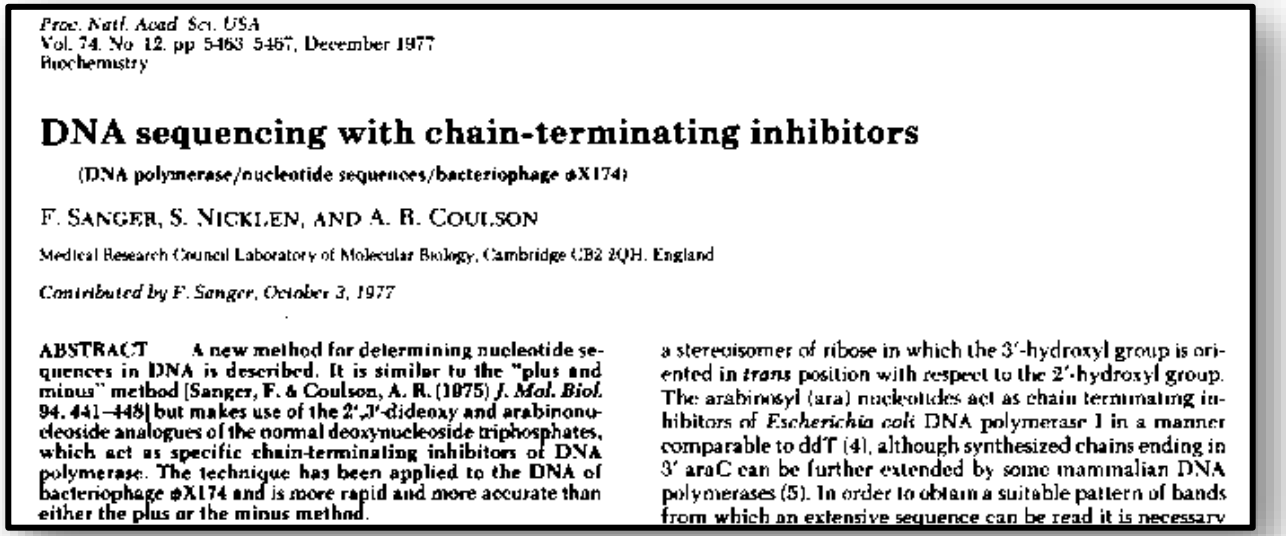
The Humble Beginning



THE CHROMOSOME NUMBER OF MAN

By *JOE HIN TJIO* and *ALBERT LEVAN*

ESTACION EXPERIMENTAL DE AULA DEI, ZARAGOZA, SPAIN, AND CANCER CHROMOSOME LABORATORY, INSTITUTE OF GENETICS, LUND, SWEDEN



Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage φX174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977

ABSTRACT A new method for determining nucleotide sequences in DNA is described. It is similar to the "plus and minus" method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 94, 441–448] but makes use of the 2′,3′-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage φX174 and is more rapid and more accurate than either the plus or the minus method.

a stereoisomer of ribose in which the 3′-hydroxyl group is oriented in *trans* position with respect to the 2′-hydroxyl group. The arabinosyl (ara) nucleotides act as chain terminating inhibitors of *Escherichia coli* DNA polymerase I in a manner comparable to ddT (4), although synthesized chains ending in 3′ araC can be further extended by some mammalian DNA polymerases (5). In order to obtain a suitable pattern of bands from which an extensive sequence can be read it is necessary

Human Genome Project

- “Eric, I need you to be as specific as possible about the ROI. We pay by the base pair!”
  - *Email received by me Fall 1999*

Approximate cost per megabase in 1999: **$ 13,000**

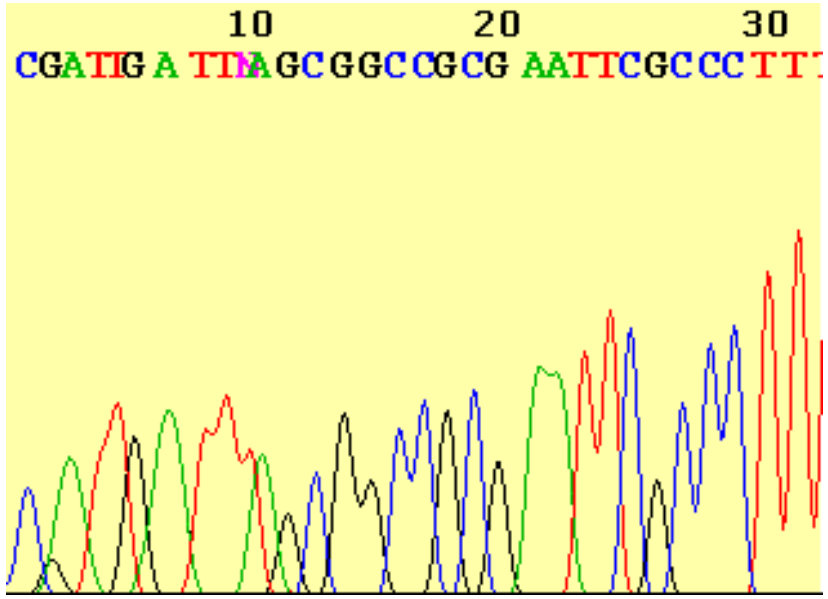
Chromosome 22, the smallest human chromosome, successfully sequenced

Sanger Sequencing

- Also known as “chain terminator sequencing”
- Oligonucleotide primers complementary to a template are used with each of the four nucleotides
- DNA polymerase is used to extend the sequence
- Chain terminating nucleotides are placed in low concentration as well
  - Usually dideoxynucleotides with a fluorescent dye tag

Sanger Sequencing

- Differently sized fragments result and are terminated at the position that the dideoxynucleotide is incorporated
- Fluorescent dye can be detected



Sanger Sequencing

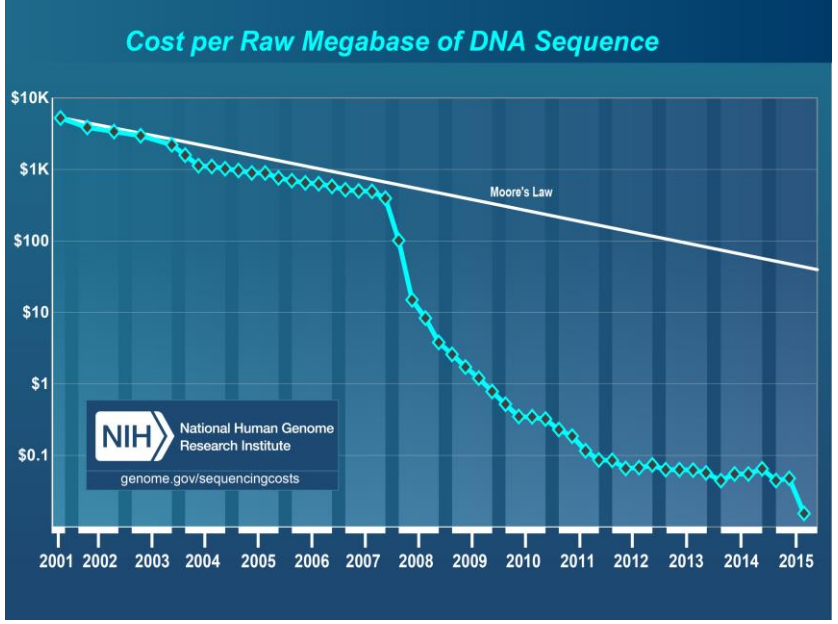
- Identification of changes at the base pair (DNA) level in one gene of interest
    1. Useful when a specific genetic syndrome or disorder is suspected based on clinical findings and/or family history
    2. Most diagnostically useful and cost-effective when a familial mutation is already known, permitting targeted mutation analysis

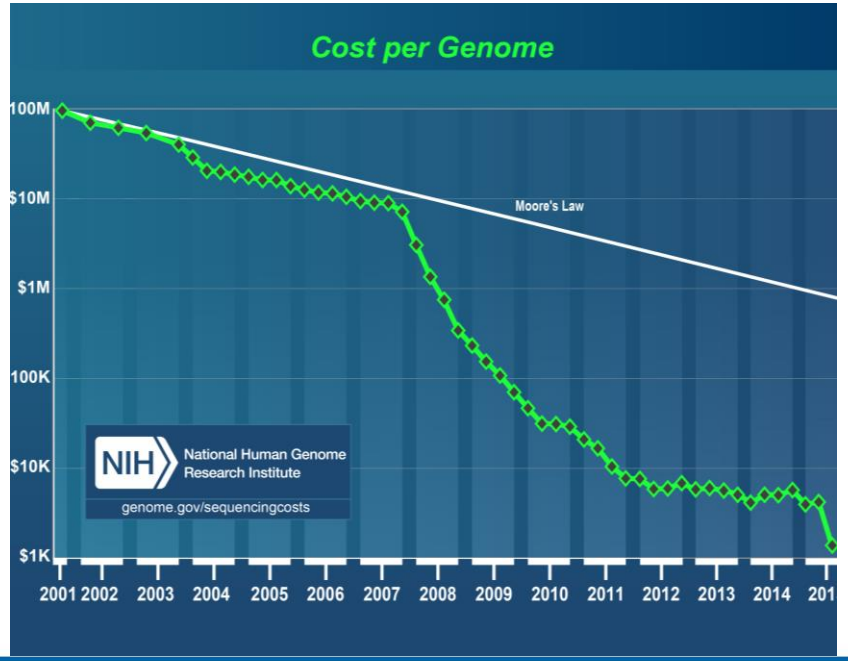
Moore's Law and Genetic Testing

In 1965, Gordon Moore observed that the number of components in an integrated circuit doubled every year, which was amended in 1975 to every two years, and held true until about 2012

Cost of sequencing mirrored Moore's law until 2008, and since then has progressively exceeded it

Why?





Next Generation Sequencing

1980

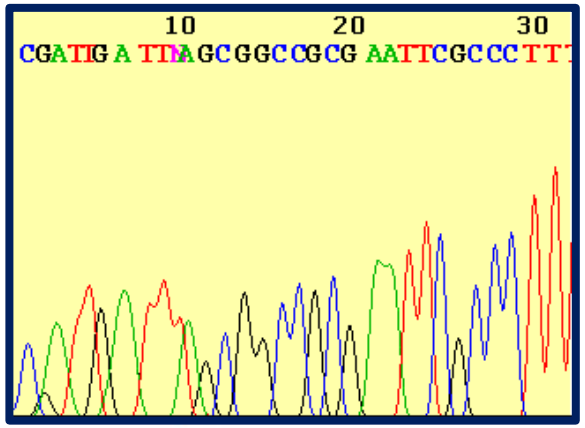
Automated sequencers have been available since the mid-1980's using Sanger technology

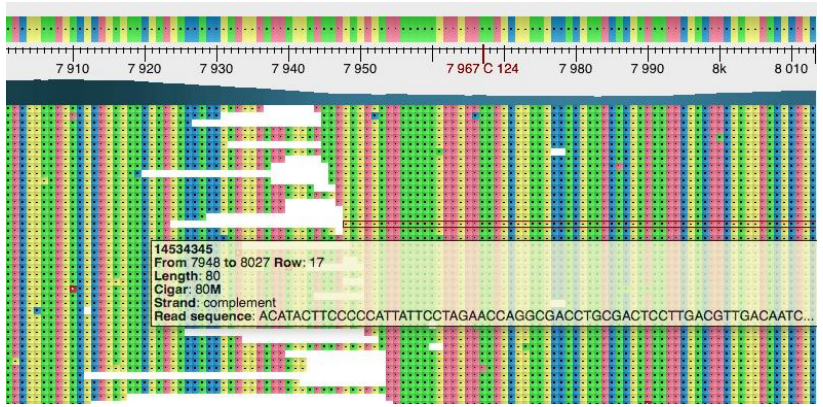
1990

In the mid-1990's, new technology such as pyrosequencing became available

2000

In 2004, 454 Life Sciences marketed a version of this for clinical use





Exome Sequencing

- Sequencing of all coding regions of all genes
  - Estimated that 85% of all mutations will be in the exome
- Cost of exome sequencing:
  - June 2010: $2,500 through a few centers
  - June 2012: $999
  - January 2013: $698
  - February 2016: $399
- What's the catch?
  - Whole exome isn't really whole exome
  - You get what you pay for

Why do we use exome?

- Patient's history and physical examination strongly suggests genetic etiology, but does not identify the specific etiology
- The patient has had extensive evaluation, usually including several rounds of genetic testing without forthcoming diagnosis
- Preferably, both parental specimens will be collected at the same time as the proband

Limitations of Exome Sequencing

- Whole exome is really only about 85-90% of the total coding region
- Mutations in non-coding region still can be important for disease
- Human genome is massively variable and much we don’t understand
  - Of the 20,687 presently understood protein-coding genes, human disease is only presently described in 3,834

Whole Genome Sequencing (minus interpretation)

Cost of whole genome sequencing	
1990-2003:	$ 2,700,000,000 (Human Genome Project)
2004-2007:	$ 100,000,000 (J. Craig Venter)
January-April 2008:	“less than $1.5 million” (James Watson)
June 2009:	$ 48,000
June 2010:	$ 19,500
February 2013:	$ 8,500
April 2016:	$1,500
June 2020:	$599


Children's Mercy
KANSAS CITY