MOLECULAR GENETICS

Genes ---> Enzyme ---> Metabolism
Central Dogma of Molecular Biology
DNA ---> RNA ---> Protein

GENES = ? DNA is the genetic material...
(what about, retroviruses, as HIV & TMV contain RNA)
a discrete piece of deoxyribonucleic acid is ....
- linear polymer of repeating nucleotide monomers
  nucleotides - A adenine, C cytosine
  T thymidine, G guanine
  letters of the genetic alphabet (code)
- unit of information is CODON = genetic 'word'
  triplet sequence of nucleotide CAT
  3 nucleotides = 1 codon (word) = 1 aa
- definition of word = amino acid

Size of Human Genome:
~ 3,000,000,000 base pairs or 1.5b in single strand genes
~ 500,000,000 codons (words or amino acids)
  average page your textbook = approx 850 words
  thus, human genome is equal to 590,000 pages
  or 470 copies of bio text book
  reading 3 bases/sec = about 47.6 years @ 8h/d - 7d/w
WOW... nanotechnology to the extreme.
Proof DNA is the Genetic Material

1. **Transformation Exp's of F. Griffith....**
   R-strain (benign) absorbs heat killed DNA is transformed

2. **Avery, Macleod, McCarty....** (1940's)
   Transforming substance was active DNA

3. **Alfred Hershey & Martha Chase Experiment 1952** *
   ... VIRAL REPLICATION [ phage infection ]
   a genetically controlled biological activity
   ... novel experiment (1st real use radioisotopes in biology)

CONCLUSION - DNA is genetic material because ----
32^p (nucleic acid) not 35^s (protein) controls viral replication

Structure of DNA... **Double Helix.**
people - Rosy Franklin, Maurice Wilkins, JD Watson,
Francis Crick, Erwin Chargaff, etc....

double stranded, helical, polynucleotide chains, made of
4 nucleotides - A,T,G,C (purine & pyrimidines)
2 polynucleotide strands (polymer chains)
held together via complimentary pairing -
Chargaff's rule ..... A:T  G:C
   A + G / T + C = 1.0
head-tail polarity [5'-----3'] - antiparallel
Replication of DNA  
(Arthur Kornberg - 1959 Nobel)

copying of DNA into DNA is structurally obvious

Patterns of Replication -
conservative, semiconservative, & dispersive
Matt Meselson & Frank Stahl experiment (1958) –
used heavy isotope of N to label DNA - $^{14}$N & $^{15}$N

Enzyme that makes DNA is DNA polymerase
req: deoxy-XTP's and ssDNA template piece
reads template and makes a complimentary copy
reads 3' to 5' and synthesizes in 5" to 3' direction

Replication forks - bidirectional synthesis
Primase activity - RNA polymerase required
Continuous & Discontinuous replication simultaneously
other enzymes required include:
  helicase - untwists DNA
  topoisomerase [DNA gyrase] - removes supercoils,
  binding proteins - stabilize replication fork
  primase (bacterial primosome) - makes RNA primer
  Pol III - synthesizes new DNA strands
  Pol I - remove RNA primer & adds DNA bases
  ligase repairs Okazaki fragment holes

Okazaki fragments
DNA Repair
**GENE EXPRESSION -**

**Transcription** - copying of a DNA sequence into RNA

**Translation** - copying of RNA sequence into protein

Flow of genetic information [Central Dogma Molecular Bio.]

DNA sequence ---> RNA sequence ---> amino acid sequence

TAC     AUG     MET

triplet sequence in DNA --> codon in mRNA ----> amino acid in protein

Information : triplet sequence = genetic word [codon]

Compare Events:

Procaryotes vs. Eucaryotes = Separation of labor

Differences DNA vs. RNA

**Transcription** - RNA polymerase

RNA polymerase binds to promoter DNA region

transcription factors read DNA sequence - make RNA copy

makes a complimentary copy of one of 2 DNA strands

**Kinds of RNA**

* tRNA - small, 80n, anticodon
  function = picks up aa & takes it to ribosome

* rRNA - piece of RNA that make up organelle = ribosome
  RNA Processing -

* mRNA - intermediate size - 100n to 400n
  function - codes for amino acid sequence

* hnRNA - heterogeneous nuclear RNA
  - Primary Transcript
    function - precursor of mRNA in eukaryote
Some other types of RNA molecules:

**small nuclear RNA (snRNP's)** -
plays a structural and catalytic role in spliceosome*

**SRP (signal recognition particle) RNA** -
a component of the protein-RNA complex that recognizes signal sequence of polypeptides targeted to ER

**small nucleolar RNA (snoRNA)** -
aids in processing of pre-rRNA transcripts for ribosome subunit formation in the nucleolus

**small interfering RNA (siRNA)** - also called microRNA;
short (20-24 nucleotide) RNAs present in MODEL eukaryotic organisms as: roundworms, fruit flies, mice, humans, & plants (arabidopsis) helps regulate gene expression by controlling timing of developmental events
also inhibits translation of target mRNAs.
ex: siRNA - c7-fig 19.9*
**INTRONS:**

*What are Introns*? and *What is the Role of Intron DNA*?

is it *DNA Junk* or sophisticated *Genetic Control Elements*?

in 1977 Phillip Sharp & Richard Roberts discovered that DNA contains *introns* ...

*intervening DNA segments that do NOT code for proteins*

**Presence of Introns:**

mostly absent in prokaryotes:

have few non-coding DNA pieces, but as eukaryotic complexity grew, so did a few non-coding DNA pieces.

now makes up greater than 95% of the DNA, i.e., less than 1.5% of human genome encodes proteins 40% of human genome is Transposons & repeat genetic elements.

**Evolutionary Origins?**

may have been self-splicing mobile genetic elements that inserted themselves into host genomes.

**Advent of Spliceosomes:**

a primary RNA transcript is processed by splicing to assemble protein coding exons

spliceosome = catalytic RNA/protein complexes that snip RNAs out of mRNAs, would encourage introns to proliferate, mutate, evolve
TRANSLATION

- process of making a protein in a specific amino acid sequence from a unique mRNA sequence

Sequence of Steps in Translation

1. add an amino acid to tRNA = aa-tRNA - ACTIVATION
2. assemble parts [ribosome, mRNA, aa-tRNA] - INITIATION
3. adding new aa's - peptidyl transferase - ELONGATION
4. stopping the process - TERMINATION

a polypeptide is built on the Ribosome on a polysome review process and figures.

GENETIC CODE

- sequence of nucleotides in mRNA that specifies sequence of amino acids in protein
- Coding Ratio = # n's = 1 aa = 3 nucleotides
- S. Ochoa (1959) - polynucleotide phosphorylase
- M. Nirenberg (1968 Nobel) - synthetic mRNA's
  1 n = U 5'-UUU-3' = phe

CODE is:
universal (but with a few minor some anomalies),
1 initiator codon (AUG),
redundant but non-ambiguous,
and exhibits wobble.
GENETIC CHANGE:

**Mutation** - change in DNA nucleotide sequence results in a different codon = different amino acid sequence

**Point mutation** - single to few nucleotide changes
- deletions, insertions, frame-shift mutations [CAT]
- substitutions -
  - non-sense = change to no amino acid (STOP)
    UCA --> UAA ser to non
  - mis-sense = different amino acid
    UCA --> UUA ser to leu
- effects = no effect, detrimental (lethal), +/- functionality, beneficial
  ex: Sickle Cell Anemia - a mis-sense mutation...

**Recombination** - Recombinant DNA
  genotype change by inserting **NEW DNA** into recipient cell
  1. fertilization n + n = 2n sperm into egg cell
  2. exchange of homologous chromatids (crossing over)
  3. transformation - absorption of DNA by recipient cells
  4. **BACTERIAL CONJUGATION** - plasmids (F+ and R)
      primitive sex-like reproduction
  5. **VIRAL TRANSDUCTION** - via a viral vector
      general transduction - pieces of bacterial DNA are packaged w viral DNA during viral replication
      restricted transduction - a temperate phage goes lytic carrying adjacent bacterial DNA into virus particle
  6. **DESIGNER GENES**
DESIGNER GENES & BIO TECHNOLOGY

- Recombinant DNA Technology
  collection of experimental techniques, allows for isolation, copying, & insertion of new DNA sequences into host-recipient cells by artificial methodologies

**Restriction Endonucleases** - diplotomic cuts at unique DNA sequences, including palindromes

\[
\begin{align*}
5' \text{GAATTC} & \quad 3' \\
3' \text{CTTAAG} & \quad 5'
\end{align*}
\]

DNA cut as such can be recombined (reannealed) or spliced w other DNA molecules to produce new genes

**What's Being Done?**

**Cloning of Genes...**
1. Plasmids... libraries
2. Probes via cDNA & reverse transcriptase
3. Polymerase Chain Reaction.... O.J. & Jurassic Park

**Locating Genes** - electrophoresis & restriction maps
DNA fingerprints, hybridization,

**Southern Blots** - a technique for detecting specific DNA sequences separated by gel electrophoresis via hybridization to a previously radioactively labeled nucleic acid probe.

**Microarrays** - passes cDNA of the cell's mRNA over slide with ssDNA copies of all a cell's genes;
DNA microchips are made by high speed robotics akin to Intel chip making; cDNA (mRNA's) are fluoresecently tagged so easy to see in slide's wells.

**Gene Sequencing** - Human Genome Project & dideoxy DNA.
Some Practical Applications of DNA Technology
- What's been Done...

1. **Medical...** disease may involve changes in gene expression
   a. **disease/infection diagnosis:**
      PCR & labeled probes from pathogens help identify microbe types:
      [RT-PCR] - HIV RNA -RT-> cDNA -PCR->
      probe can ID AIDS infection
   b. **RLFP** - Restriction Length Fragment Analysis - markers often inherited with disease
      what is RFLP* genetic testing & polymorphism --->
      RFLP markers & disease* & MST II cuts Sickle Cell*
      & Dde-I cuts Sickle Cell*
      paternity testing via DNA fingerprinting
   c. **Gene Therapy...** idea is to replace defective genes
      - microinjection of DNA* & fig 20.16*
        & (ADA Deficiency & Ashanti DeSilva update)
      - **SCID** (Severe combined immunodeficiency - a single gene enzyme defect):
        clinical trials in 2000 resulted in 2 of 9 cured, but developing lukemia: the retroviral vector inserted repair gene near bone marrow cell genes involved in blood cell division, thus lukemia. trials stopped.

2. **Pharmaceutical Products...** manufactured drugs
   protropin (an ethical dilemma)* &
   Recombinant bacteria* = Humulin
3. Forensics...

- **DNA fingerprinting** - example judicial modus operandi
  a murder case* & a rape case*
  DNA fingerprinting usually looks a 5 RFLP markers and
  blood is tested via Southern Blotting (20.10) using probes
  for these alleles

- **Simple Tandem Repeats** (short- 5n to 6n) - trinucleotide
  (3n) repeats can undergo an increase in copy number
  by a process of dynamic mutation; # of tandem
  repeats is unique to a genetic indiv. Variation in the
  length of these repeats is polymorphic. Individual A
  has ACA repeated 65 times @ loci 121, 118, & 129
  individual B has different repeat pattern at these loci.

**STR's** can cause genetic diseases as well:
  CCG trinucleotide occur in fragile sites on human
  chromosomes (folate-sensitive group).
  fragile X (FRAXA)- responsible for familial mental
  retardation. Another FRAXE is responsible for a rarer
  mild form of mental retardation.
  mutations of AGC repeats gives a #- of neurological
  disorders.

4. Environmental Clean-up...

  bacteria can extract heavy metals (Cu, Pb, Ni) from the
  environment & convert them into not toxic compounds
  genetically modified bacteria may be the "miner's" of
  the future.
5. **Franken Food...**
   genetically modified (GM) animals & agricultural crops

**Transgenics** –
organisms with inserted foreign DNA in their genomes

**Animals**
- GFP novelties + Dolly
  - animal cloning companies --->
  - "pharm" animals (20.18*) - sheep carry human human blood protein gene that inhibits enzyme in cystic fibrosis patients

**Plants**
- genetically modified crop plants - fig 20.19*
  - to get Ti plasmids in = a DNA gun*
  - Frankenfood & Edible Vaccines
  - National Plant Genome Initiative Plan update

**An overview of biotechnology**
- History of Biotechnology
- Human Genome Project & Biotech Companies
Control of Gene Expression -

How do we know a gene has been active (turned on)
look for gene's product, i.e., protein
increase in enzyme activity = gene action?
no enzyme activity = no gene action
but, pre-existing inactive enzymes --> active forms
ZYMOGENS - pepsionogen ---> pepsin
- tyrpsinogen ----> trypsin

2 circumstances  1) pre-existing inactive enzyme
  2) de novo (new) enzyme synthesis

Mechanism of Gene Action  (turning on/off genes)

Jacob and Monod - prokaryotes  -  Lactose Operon

E. coli

glucose  \[\text{NO beta-galactosidase}\]

NO beta-galactosidase  \[\beta\text{-galactosidase}\]
lactose  \[\beta\text{-galactosidase}\]

\[\text{lactose } \longrightarrow \text{ glucose + galactose}\]

Operon = series of mapable-linked genes controlling
synthesis of protein
promoter - binds RNA polymerase
operator - binds repressor protein
regulator - makes repressor protein
structural - make enzyme proteins

\[
\begin{array}{ccccccc}
p & Rg & cap & p & o & Sg1 & Sg2 & Sg3 \\
\end{array}
\]
Mechanism of Gene Action  (turning on/off genes)

Eukaryotes -  
more complex  (more DNA - nuclear compartment)

promoter - site where RNA polymerase binds
enhancer - sites where enhancers/transcription factors bind

transcription factors - proteins that help transcription

Some examples: Eukaryotic gene expression controls

Differential Gene Activity...  selective expression of genes
  i.e., different cell types express different genes
  liver vs. lens cell
1. role of activators in selective gene expression (DGA)

2. molecular turnover - ½ life mRNA's*
   &  longevity of some proteins*

3. steroid hormones  (figure*)

4. Processing of RNA transcript  (figure*)
   cut/spliced in nucleus and capped for transport
   intron - pieces cut out (non gene-proteins)
   exons - pieces transported to cytoplasm
   alternative splicing (next page)*
ALTERNATIVE SPLICING

at the beginning of the 3rd millennium, the estimates of the number of human genes was 153,000 making about 90,000 proteins;

by the first draft of the Human Genome Sequence (summer of 2005) the number had shrunk to ~ 20-25,000 protein coding genes.

The current estimates of the NHGRI puts the number of human genes at less than 25,000. (& there is actually a betting pool)

But, humans still make about 90,000 proteins. How from only 25,000 genes????

In 2004 the mouse genome was sequenced and we learned it also has 25K genes (the same as man) and we both share many of the same exons and introns.

If Mice and Men are so genomically similar, what makes so vastly different?

alternative splicing --->
1977 - Philip Sharp (MIT) & Richard Roberts (NE Biolabs) discover split genes & presence of introns via DNA-RNA hybridizations and an excision splicing process*.

1980 - Randolph Wall (UCLA) discovers alternate splicing; i.e., some introns left in or some exons cut out

1984 - Tom Maniatis & Mike Green (Harvard) describe a Splicing Machine = Splicesome* a highly conserved complex of:
- 5 small nuclear uridine rich RNA molecules (snRNA - U1, U2, U4, U5, & U6)
- 150 proteins - can recognize exon/intron interfaces & excise the introns.

Splicesomes cut @ exon/intron interfaces:
- at the 5' splice site and 3' end
  a) branch site  b) polypyrimidine site  c) 3' splice site
cutting involves Splicing Regulator proteins (SR):
  - there are some 10 known different SR proteins identified so far
  - they bind to an Exon nucleotide sequence called Exonic Splicing Enhancers [ESE]
  - SR binding recruits Spliceosome Basal Machinery to 5' Splice Site
  - Exons also hold repressor nucleotide sequences [ESS] - Exonic Splicing Suppressor
when an SR binds @ an ESS- prevents Spliceosome machinery from splicing
Some examples of alternative splicing:

1. **Bcl-x** gene makes a regulator protein for programmed cell death apoptosis:
   
   gene actually makes **2 proteins via alternative splicing**
   
   1. **Bcl-x (L) [larger]**
   2. **Bcl-x (S) [smaller]**

   suppresses apoptosis promotes apoptosis

2. **sxl** (sex lethal gene of Drosophila)

   when male sxl (exon-2) is skipped during splicing .
   
   gene makes a female specific sxl-protein (figure*) –
   
   this protein binds to all subsequent premRNA of same gene resulting in excision of the male exon form all mRNA's -> female flies
   
   if male exon is retained in 1st round of splicing --> male sxl protein --> male flies

3. **tropomyosin** - exon skipping = different versions of tropomyosin are produced for skeletal muscle, smooth muslce, fibroblast cells, live and brain cells. figure*

**EXON SKIPPING** - is most common form of Alternate Splicing in mammals

**INTRON RETENTION** - changing the length of a processed mRNA is the most common form of alternate splicing in plants & lower multicelld organisms .
Cancer & Gene Expression

cancer often results from gene changes affecting cell cycle control cancer genes, such as adenomatous polyposis coli causes 15% of colorectal cancers is a tumor suppressor gene, a type of Oncogenes*

2 kinds of human cancer genes:
Ras (proto-oncogene gene*) = 30% human cancers is a G-protein that promotes cell division proteins a Ras mutation --> hyperactive Ras protein --> division fig 19.12a*
p53 (tumor suppressor gene* = 50% human cancers) fig 19.12b*
p53 is a transcription factor promoting cell cycle inhibiting proteins. [DNA damage --> active p53 --> p51 gene --> protein binds to cyclin dependent kinase stops cell division] thus a p53 mutation --> excess cell division (cancer).

other cancer genes can lead to new gene actions as: BRCA1 and BRCA2 (tumor suppressor genes) are involved in 50% of breast cancers in humans
Organization of the Human genome:

highly condensed DNA (heterochromatin) is often not expressed and chemical modification of histones in genomic DNA often regulates DNA's action

ACYLATION of histones (lys) unfolds chromatin c7-fig 19.4*

METHYLATION of nucleotides leads (often to C) favors condensation & leads to inactive DNA... once CH₃ passed generationally.

Non-Coding Genes and Genome Structure:

in prokaryotes: DNA codes for proteins, tRNA, or rRNA
    avg. gene = 1,000np)
    non-coding DNA is mostly promoters; generally no introns

in eukaryotes: most DNA is non-coding (mostly introns)
    avg. gene is 27,000np c7 fig 19.14*

Transposable Genetic Elements

- largest portion of human genome

2 types: 1. transposons (cut & paste) - c7-fig 19.16a*
    genetic element that moves by means of a DNA intermediate

2. retro-transposon (RNA intermediate)
    - c7-fig 19.16b*
    move via a transcript of retrotransposon DNA
Non-coding genome elements - ALU elements

in humans one group of transposable genes is Alu elements
~10% human genome (1.4 million copies - some 500,000)
each Alu element is about 300 n long with a poly-A tail,
non-coding; can copy themselves & reinsert randomly
(like a genomic parasite)
1 new insert per every 100-200 births
5% of human alternately splice genes hold Alu pieces
origins of Alu ? -> piece jumped into intron --mutation-->
created new 5' or 3' splice site resulting in exon that
was alternately spliced into mRNA.
Alu pieces are transcribed into RNA & function is unknown

next largest portion of Human Genome is Repetitive DNA...
15% of human genome is Repetitive DNA.
large piece repeats (10K-300K np) often copied from
one chromosome
3% is single DNA short sequence repeats of --GTTAC--
repeats sequences can be from 5n (short) up to 500n
most are located at telomers and centromeres

next largest portion of Human Genome is Intronic DNA...
1.5% of Human genome is exonic DNA (i.e., 25,000 genes)
½ of exonic DNA is "traditional" single gene unique DNA
24.0% is intronic multigene family DNA... (fig 19.17a & b)
ex: identical repeats- rDNA - tandem repeats for rRNA
advantage: makes millions of ribosomes
mass production: primary transcript is processed
EXON Shuffling...

can lead to significant genome rearrangement
(may be evolutionarily significant)

exons are mixed & matched due to meiotic division errors
occurs via unequal crossing over at non-sister transposable
element sites  c7 fig 19.18*

recall that proteins often have a modular architecture
made up of discrete structural & functional regions called
domains

different exons code for different domains  fig 17.12

ex:  TPA (tissue plasmogen activator)
an extracellular protein that prevents blood clotting
protein has 4 domains - 3 each coded by different
exon & 1 duplicate

each exon is also found in other proteins
origin may have been by exon shuffling - fig 19.20*
Definitions of a Gene:

Mendel's Particles... unit of heredity responsible for phenotype
Morgan's Loci... placed gene on a chromosome, i.e.,
               it's a cellular entity, that is part of chromosome - mapable
Watson & Crick... is sequence of specific nucleotides along
               the length of DNA molecule

Molecular Definition -
1 nucleotide = 0.34 nm thus tRNA = 81n x 0.34 = 27.5 nm
1 nucleotide = 340 amu thus tRNA = 81 x 340 = 27,540 amu

Modern Molecular - a functional definition:
Biological... DNA sequence coding for a specific polypeptide
Split Genes... Introns & Exons:
euc genes contain non-coding segments
               (introns) with no corresponding protein & coding segments (exons) = proteins
Others pieces ...
               any definition should also include:
               promoter sequences, enhancers,
               regulator gene, operators, CRP...
               segments that code for rRNA, tRNA, & snRNP's

"A GENE is a region of DNA that CODES for an RNA"