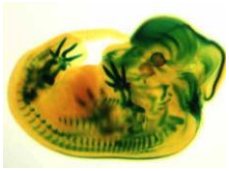
Chapter 19.

Control of Eukaryotic Genome

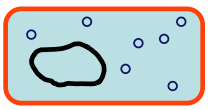


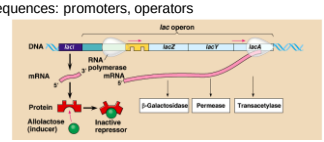
The BIG Questions…

- How are genes turned on & off in eukaryotes?
- How do cells with the same genes differentiate to perform completely different, specialized functions?
- How have new genes evolved?

Prokaryote vs. eukaryote genome

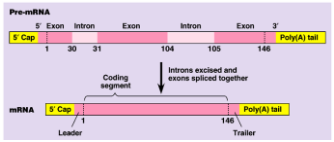
- Prokaryotes
  - small size of genome
  - circular molecule of naked DNA
    - DNA is readily available to RNA polymerase
    - control of transcription by regulatory proteins
      - operon system
  - most of DNA codes for protein or RNA
    - small amount of non-coding DNA
      - regulatory sequences: promoters, operators





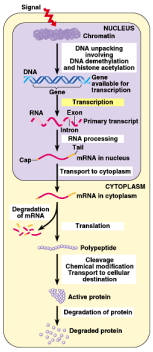
Prokaryote vs. eukaryote genome

- Eukaryotes
  - much greater size of genome
    - how does all that DNA fit into nucleus?
  - DNA packaged in chromatin fibers
    - regulates access to DNA by RNA polymerase
  - cell specialization
    - need to turn on & off large numbers of genes
  - most of DNA does not code for protein
    - 97% "junk DNA" in humans



Points of control

- The control of gene expression can occur at any step in the pathway from gene to functional protein
  - unpacking DNA
  - transcription
  - mRNA processing
  - mRNA degradation
  - translation
  - protein processing
  - protein degradation



Why turn genes on & off?

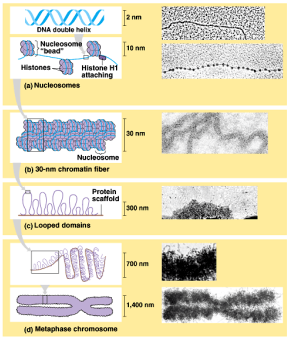
- Specialization
  - each cell of a multicellular eukaryote expresses only a small fraction of its genes
- Development
  - different genes needed at different points in life cycle of an organism
    - afterwards need to be turned off permanently
- Responding to organism's needs
  - cells of multicellular organisms must continually turn certain genes on & off in response to signals from their external & internal environment

DNA packing

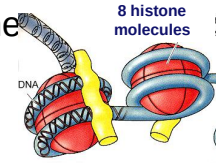
How do you fit all that DNA into nucleus?

- DNA coiling & folding
  - double helix
  - nucleosomes
  - chromatin fiber
  - looped domains
  - chromosome

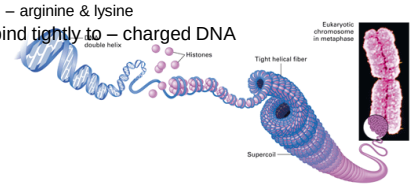
from DNA double helix to condensed metaphase chromosome



Nucleosome

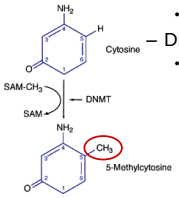


- "Beads on a string"
  - 1st level of DNA packing
  - histone proteins
    - 8 protein molecules
    - many + charged amino acids
      - arginine & lysine
    - bind tightly to – charged DNA



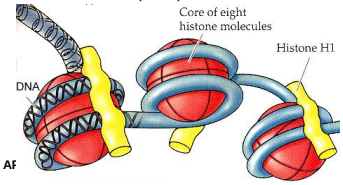
DNA packing & DNA methylation

- Chromatin modifications affect the availability of genes for transcription
  - packing DNA into compact form serves to regulate transcription
    - heterochromatin is compact & not transcribed
    - less compacted euchromatin may be transcribed
  - DNA methylation inactivates genes = off
    - attachment of methyl groups ($–CH_3$) to DNA bases (cytosine) after DNA is synthesized
      - nearly permanent suppression of genes
      - ex. the inactivated mammalian X chromosome

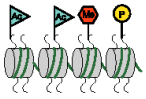


Histone acetylation

- Histone acetylation activates genes = on
  - attachment of acetyl groups ($–COCH_3$) to certain amino acids of histone proteins
    - when histones are acetylated they change shape & grip DNA less tightly = unwinding DNA
    - transcription proteins have easier access to genes

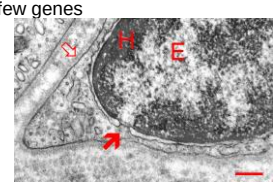


The Histone Code



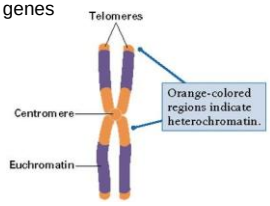
Chromatin structure

- Chromatin structure is part of gene regulation
  - heterochromatin
    - highly compact
    - not transcribed
    - found in regions of few genes
  - euchromatin
    - "true" chromatin
    - transcribed
    - found in regions of many genes



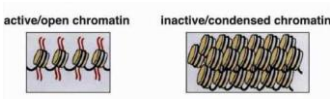
Heterochromatin

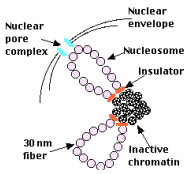
- Heterochromatin
  - highly compact
  - not transcribed (inactive genes)
    - increased methylation of histones & cytosines
    - decreased acetylation of histones
  - found in regions of few genes
    - centromeres
    - telomeres
    - other "junk" DNA



Euchromatin

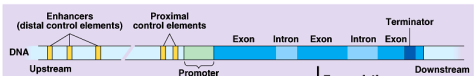
- Euchromatin
  - loosely packed in loops of 30nm fibers
    - near nuclear pores
  - transcribed "active" genes
    - decreased methylation of histones & cytosines
    - increased acetylation of histones
  - found in regions of many genes





Transcription initiation

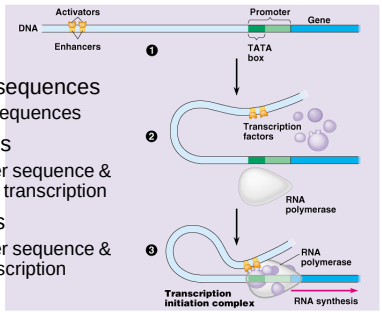
- Control regions on DNA
  - promoter
    - initial binding of RNA polymerase
  - proximal control sequences
  - enhancers = distal control sequences
- Regulatory proteins
  - transcription factors
    - control proteins which interact with DNA & with each other
  - initiation complex
    - only when complete complex has assembled can RNA polymerase move along DNA template to make mRNA



AP Bi

Model for Enhancer action

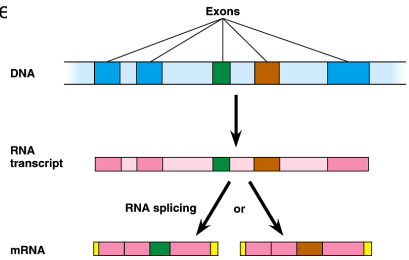
- Enhancer DNA sequences
  - distant control sequences
- Activator proteins
  - bind to enhancer sequence & stimulates gene transcription
- Silencer proteins
  - bind to enhancer sequence & block gene transcription



Turning on Gene movie

Post transcriptional control

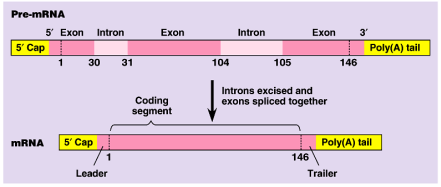
- Alternative RNA splicing
  - variable processing of exons creates a family of prote



AP Biology

Regulation of mRNA degradation

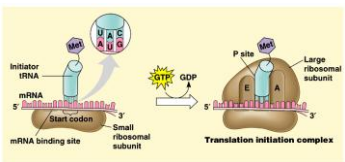
- Life span of mRNA determines pattern of protein synthesis
  - mRNA can last from hours to weeks



RNA processing movie

Control of translation

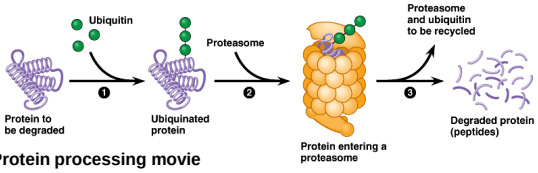
- Block initiation stage
  - regulatory proteins attach to 5' end of mRNA
    - prevent attachment of ribosomal subunits & initiator tRNA
    - block translation of mRNA to protein



AP Biology

Protein processing & degradation

- Protein processing
  - folding, cleaving, adding sugar groups, targeting for transport
- Protein degradation
  - ubiquitin tagging & proteosome degradation



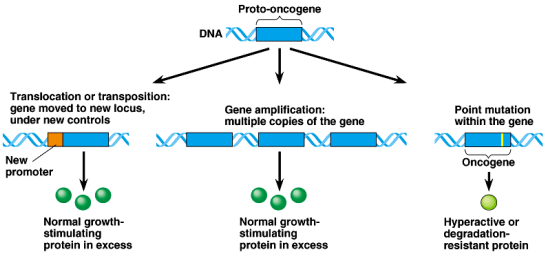
Protein processing movie

Cancers = cell cycle genes

- Cancer results from genetic changes that affect the cell cycle
  - proto-oncogenes
    - normal cellular genes code for proteins that stimulate normal cell growth & division
  - oncogenes
    - mutations that alter proto-oncogenes cause them to become cancer-causing genes

Proto-oncogenes & Oncogenes

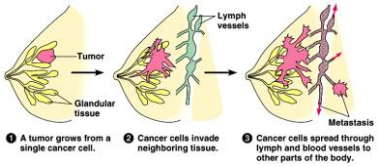
- Genetic changes that can turn proto-oncogenes into oncogenes
  - removing repression of genes



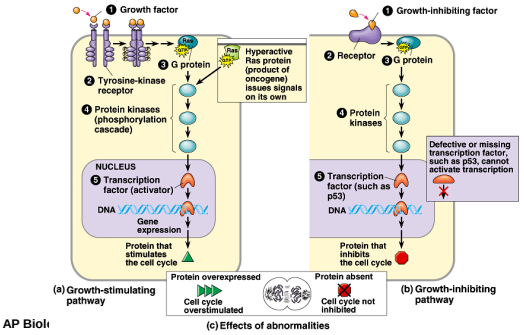
Cancers = failures of regulation

- Cancer cells have escaped cell cycle controls
  - do not respond normally to the body's control mechanisms
    - divide excessively & invade other tissues
    - if unchecked, they can kill the organism

growth & metastasis of malignant breast tumor

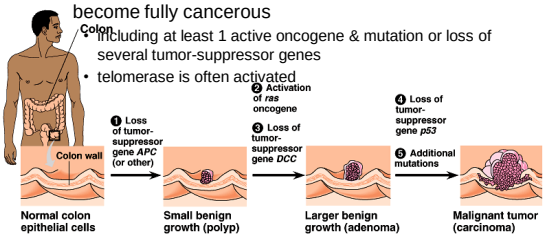


Effects of signal pathways



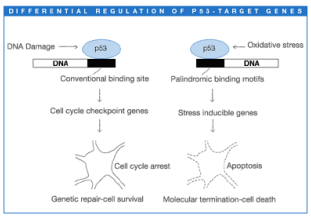
Multi-step model for cancer

- Multiple mutations underlie the development of cancer
  - several changes must occur at DNA level for cell to become fully cancerous
    - including at least 1 active oncogene & mutation or loss of several tumor-suppressor genes
    - telomerase is often activated

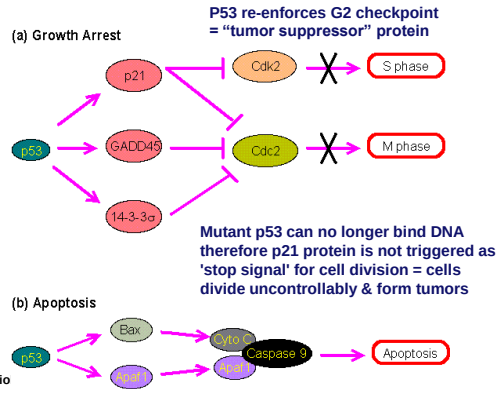


p53 gene

- "Guardian of the Genome"
  - the "anti-cancer gene"
  - after DNA damage is detected, p53 initiates:
    - DNA repair
    - growth arrest
    - apoptosis



almost all cancers have mutations in p53



Structure of the Eukaryotic Genome

How many genes?

- Genes
  - only ~3% of human genome
  - protein-coding sequences
    - 1% of human genome
  - non-protein coding genes
    - 2% of human genome
    - tRNA
    - ribosomal RNAs
    - siRNAs

What about the rest of the DNA?

- Non-coding DNA sequences
  - regulatory sequences
    - promoters, enhancers
    - terminators
  - "junk" DNA
    - introns
    - repetitive DNA
      - centromeres
      - telomeres
      - tandem & interspersed repeats
    - transposons & retrotransposons
      - Alu in humans

Repetitive DNA

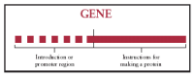
Repetitive DNA & other non-coding sequences account for most of eukaryotic DNA

Table 19.1 Types of Repetitive DNA

Tandemly Repetitive DNA (Satellite DNA)	
Repeated units at a site are usually identical	
Proportion of mammalian DNA:	10–15%
Length of each repeated unit:	1–10 base pairs
Total length of repetitive DNA per site, in base pairs:	
Regular satellite DNA	100,000–10 million
Minisatellite DNA	100–100,000
Microsatellite DNA	10–100
Interspersed Repetitive DNA	
"Copies" are very similar but not identical	
Proportion of mammalian DNA:	25–40%
Length of each repeated unit:	100–10,000 base pairs
Number of repetitions per genome:	10–1 million

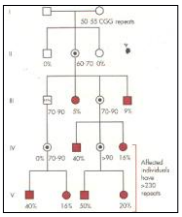
Genetic disorders of repeats

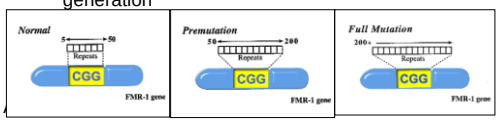
- Fragile X syndrome
  - most common form of inherited mental retardation
  - defect in X chromosome
    - mutation of FMR1 gene causing many repeats of CGG triplet in promoter region
      - 200+ copies
      - normal = 6-40 CGG repeats
    - FMR1 gene not expressed & protein (FMRP) not produced
      - function of FMR1 protein unknown
      - binds RNA



Fragile X syndrome

- The more triplet repeats there are on the X chromosome, the more severely affected the individual will be
  - mutation causes increased number of repeats (expansion) with each generation





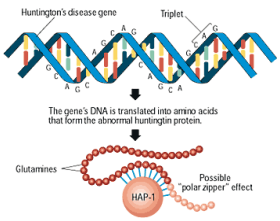
Huntington's Disease

- Rare autosomal dominant degenerative neurological disease
  - 1st described in 1872 by Dr. Huntington
  - most common in white Europeans
  - 1st symptoms at age 30-50
    - death comes ~12 years after onset
- Mutation on chromosome 4
  - CAG repeats
    - 40-100+ copies
    - normal = 11-30 CAG repeats
    - CAG codes for glutamine amino acid

Huntington's disease

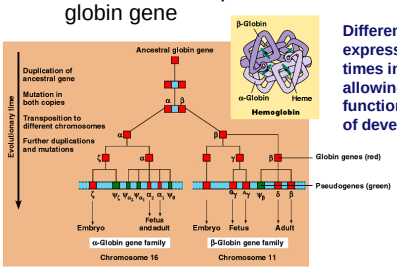
- Abnormal (huntingtin) protein produced
  - chain of charged glutamines in protein
  - bonds tightly to brain protein, HAP-1

Woody Guthrie



Families of genes

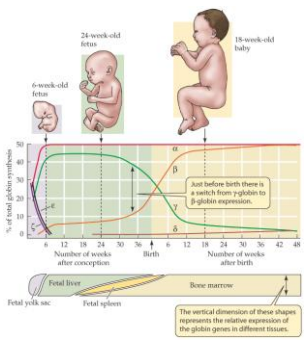
- Human globin gene family
  - evolved from duplication of common ancestral globin gene



Different versions are expressed at different times in development allowing hemoglobin to function throughout life of developing animal

Hemoglobin

differential expression of different beta globin genes ensures important physiological changes during human development

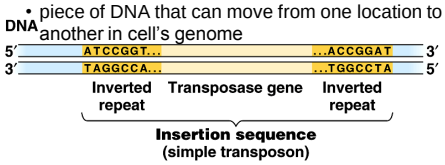


Interspersed repetitive DNA

- Repetitive DNA is spread throughout genome
  - interspersed repetitive DNA make up 25-40% of mammalian genome
  - in humans, at least 5% of genome is made of a family of similar sequences called, *Alu* elements
    - 300 bases long
    - *Alu* is an example of a "jumping gene" – a transposon DNA sequence that "reproduces" by copying itself & inserting into new chromosome locations

Rearrangements in the genome

- Transposons
  - transposable genetic element
    - piece of DNA that can move from one location to another in cell's genome

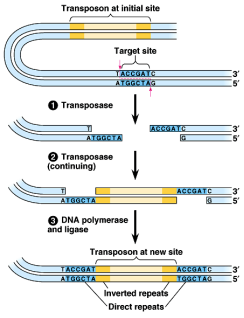


One gene of an insertion sequence codes for transposase, which catalyzes the transposon's movement. The inverted repeats, about 20 to 40 nucleotide pairs long, are backward, upside-down versions of each oth. In transposition, transposase molecules bind to the inverted repeats & catalyze the cutting & resealing of DNA required for insertion of the transposon at a target site.

Transposons

Insertion of transposon sequence in new position in genome

insertion sequences cause mutations when they happen to land within the coding sequence of a gene or within a DNA region that regulates gene expression



Transposons 1947|1983

- Barbara McClintock
  - discovered 1st transposons in *Zea mays* (corn) in 1947

Barbara McClintock (L) and Jacques Monod (R) at a Cold Spring Harbor meeting, 1946.

Courtesy of Cold Spring Harbor Laboratory Archives.

Retrotransposons

- Transposons actually make up over 50% of the corn (maize) genome & 10% of the human genome.

Most of these transposons are retrotransposons, transposable elements that move within a genome by means of RNA intermediate, transcript of the retrotransposon DNA