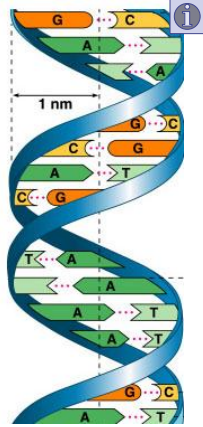
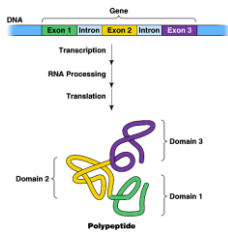


Chapter 14 Control of Eukaryotic Genome



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

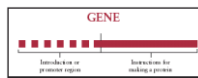
Repetitive DNA

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

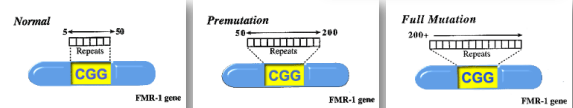
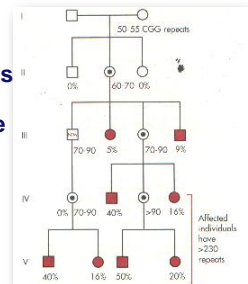
Genetic Disorders of Repeats

- **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

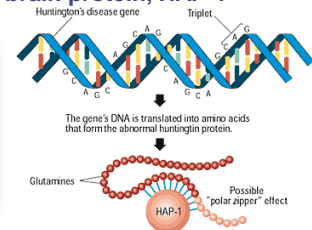
- Rare autosomal dominant degenerative neurological disease
 - ◆ 1st described in 1872 by Dr. Huntington
 - ◆ most common in white Europeans
 - ◆ 1st symptoms at age 30-50
 - death comes ~12 years after onset
- Mutation on chromosome 4
 - ◆ CAG repeats
 - 40-100+ copies
 - normal = 11-30 CAG repeats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



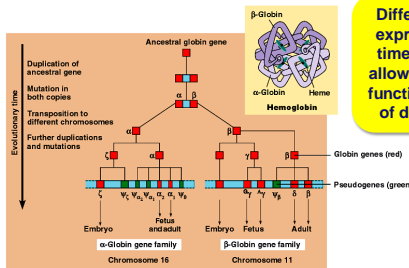
Woody Guthrie



Families of Genes

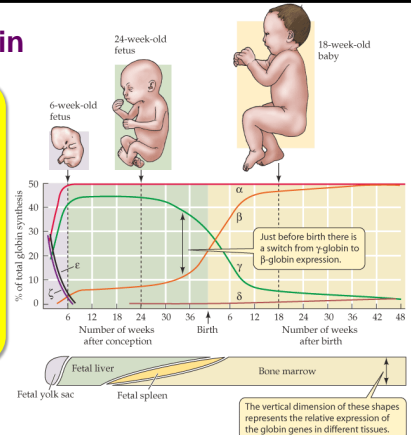
- Human globin gene family
 - ◆ evolved from duplication of common ancestral globin gene

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.

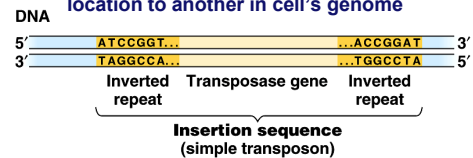


Interspersed Repetitive DNA

- Repetitive DNA is spread throughout genome
 - ◆ interspersed repetitive DNA (SINES Short Interspersed Elements) 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 other. 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.

Transposon at initial site
Target site
1 Transposase
2 Transposase (continuing)
3 DNA polymerase and ligase
Transposon at new site
Inverted repeats
Direct repeats

Transposons

1947 | 1983

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

Barbara McClintock (1984)
ITHACA 1929

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.

DNA of genome Retrotransposon New copy of retrotransposon
1 Transcription
2 Translation
3 Reverse transcription of RNA to DNA
4 Synthesis of second DNA strand
5 Insertion of retrotransposon DNA

Transcription – Another Look...

- The process of transcription includes many points of control
 - when to start reading DNA
 - where to start reading DNA
 - where to stop reading DNA
 - editing the mRNA
 - protecting mRNA as it travels through cell

Primary Transcript

- Processing mRNA
 - protecting RNA from RNase in cytoplasm
 - add 5' cap
 - add polyA tail
 - remove introns

Pre-mRNA
5' Cap Exon Intron Exon Intron Exon 3'
1 30 31 104 105 146 Poly(A) tail
Coding segment
Introns excised and exons spliced together
mRNA 5' Cap AUG UGA Poly(A) tail
Leader 1 146 Trailer

Protecting RNA

- 5' cap added
 - ◆ G trinucleoside (G-P-P-P)
 - ◆ protects mRNA
 - from RNase (hydrolytic enzymes)
- 3' poly-A tail added
 - ◆ 50-250 A's
 - ◆ protects mRNA
 - from RNase (hydrolytic enzymes)
 - ◆ helps export of RNA from nucleus

Dicing & Splicing mRNA

- Pre-mRNA → mRNA
 - ◆ edit out introns
 - intervening sequences
 - ◆ splice together exons
 - expressed sequences
 - ◆ In higher eukaryotes
 - 90% or more of gene can be intron
 - no one knows why...yet
 - ◆ there's a Nobel prize waiting...

Discovery of Split Genes

1977 | 1993

Richard Roberts
NE BioLabs

Philip Sharp
MIT

adenovirus
common cold

Splicing Enzymes

- snRNPs
 - ◆ small nuclear RNA
 - ◆ RNA + proteins
- Spliceosome
 - ◆ several snRNPs
 - ◆ recognize splice site sequence
 - cut & paste
- RNA as ribozyme
 - ◆ some mRNA can splice itself
 - ◆ RNA as enzyme

Ribozyme

1982 | 1989

- RNA as enzyme

DNA → RNA → PROTEINS → CELLS → ORGANISMS

Cech's discovery

Sidney Altman
Yale

Thomas Cech
U of Colorado

Splicing Details

- No room for mistakes!
 - ◆ editing & splicing must be exactly accurate
 - ◆ a single base added or lost throws off the reading frame

AUGCGGCTATGGGUCCGAUAAGGGCCAU
 AUGCGGUCCGAUAAGGGCCAU
 AUG|CGG|UCC|GAU|AAG|GGC|CAU
 Met | Arg | Ser | Asp | Lys | Gly | His

AUGCGGCTATGGGUCCGAUAAGGGCCAU
 AUGCGGGUCCGAUAAGGGCCAU
 AUG|CGG|GUC|CGA|UAA|GGG|CCA|U
 Met | Arg | Val | Arg | STOP

Alternative Splicing

- Alternative mRNAs produced from same gene
 - when is an intron not an intron...
 - different segments treated as exons

Domains

- Modular architecture of many proteins
 - separate functional & structural regions
 - coded by different exons in same "gene"

The Transcriptional Unit (gene?)

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?

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
 - no introns, small amount of non-coding DNA
 - regulatory sequences: promoters, operators

Prokaryote vs. Eukaryote Genome

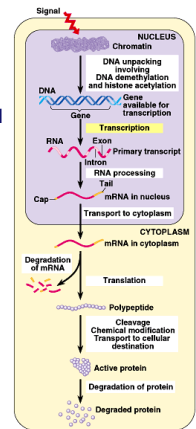
- Eukaryotes
 - much greater size of genome
 - how does all that DNA fit into nucleus?
 - DNA packaged in chromatin fibers
 - regulates access to DNA by RNA polymerase
 - cell specialization
 - need to turn on & off large numbers of genes
 - most of DNA does not code for protein
 - 97% "junk DNA" in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

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**
 - ◆ homeostasis
 - ◆ cells of multicellular organisms must continually turn certain genes on & off in response to signals from their external & internal environment

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 transport
 - out of nucleus
 - through cytoplasm
 - protection from degradation
 - ◆ translation
 - ◆ protein processing
 - ◆ protein degradation

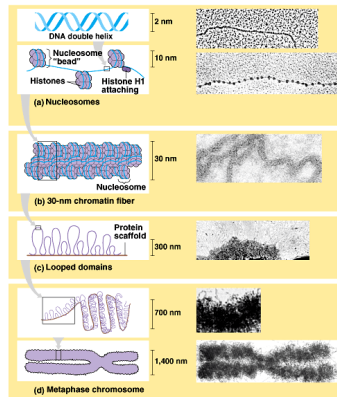


DNA Packing

How do you fit all that DNA into nucleus?

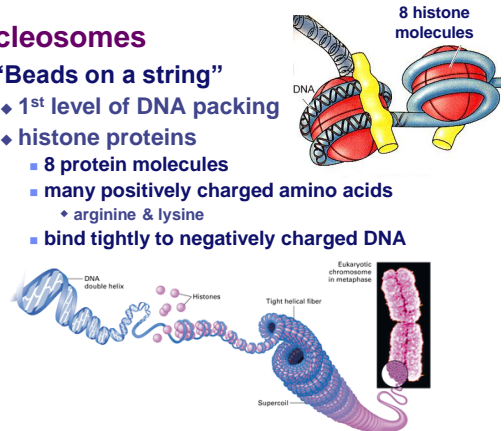
- ◆ DNA coiling & folding
 - double helix
 - nucleosomes
 - chromatin fiber
 - looped domains
 - chromosome

from DNA double helix to condensed chromosome



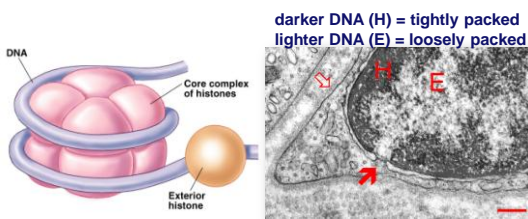
Nucleosomes

- “Beads on a string”
 - ◆ 1st level of DNA packing
 - ◆ histone proteins
 - 8 protein molecules
 - many positively charged amino acids
 - ◆ arginine & lysine
 - bind tightly to negatively charged DNA



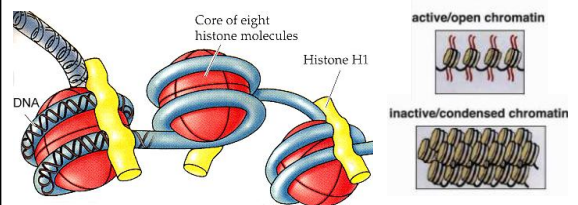
DNA Packing

- Degree of packing of DNA regulates transcription
 - ◆ **tightly** packed = no transcription
 - ◆ **loosely** packed = transcription
 - ◆ **genes turned off**
 - ◆ **genes turned on**



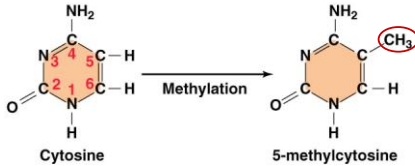
Histone Acetylation

- **Acetylation of histones** unwinds DNA
 - ◆ **loosely** packed = transcription
 - ◆ **genes turned on**
 - attachment of acetyl groups ($-COCH_3$) to histones
 - ◆ conformational change in histone proteins
 - ◆ transcription factors have easier access to genes



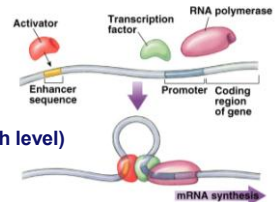
DNA Methylation

- **Methylation of DNA** blocks transcription factors
 - ◆ no transcription = **genes turned off**
 - ◆ attachment of methyl groups (-CH₃) to cytosine
 - C = cytosine
 - ◆ nearly permanent inactivation of genes
 - ex. inactivated mammalian X chromosome



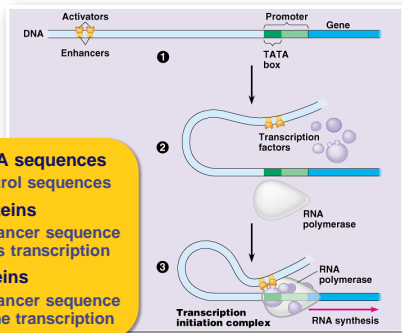
Transcription Initiation

- **Control regions on DNA**
 - ◆ promoter
 - nearby control sequence on DNA
 - binding of RNA polymerase & transcription factors
 - “base” rate of transcription
 - ◆ enhancers
 - distant control sequences on DNA
 - binding of activator proteins
 - “enhanced” rate (high level) of transcription



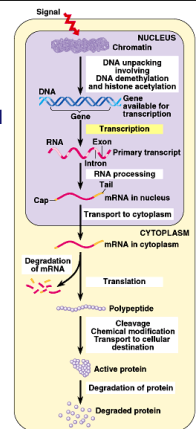
Model for Enhancer Action

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



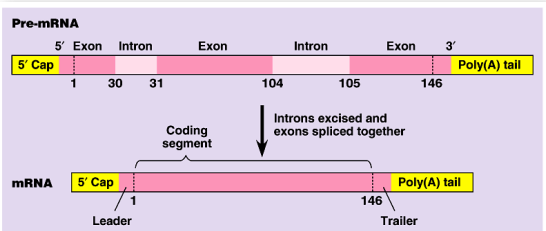
Points of Control

- **The control of gene expression can occur at any step in the pathway from gene to functional protein**
 - ◆ unpacking DNA
 - ◆ transcription
 - ◆ mRNA processing
 - ◆ mRNA transport
 - out of nucleus
 - through cytoplasm
 - protection from degradation
 - ◆ translation
 - ◆ protein processing
 - ◆ protein degradation



Regulation of mRNA Degradation

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



RNA Interference

- **Small RNAs (sRNA, iRNA, RNAi)**
 - ◆ short segments of RNA (21-28 bases)
 - bind to mRNA
 - create sections of double-stranded mRNA
 - “death” tag for mRNA
 - ◆ triggers degradation of mRNA
 - ◆ cause gene “silencing”
 - even though post-transcriptional control, still turns off a gene
 - siRNA

RNA Interference

Small RNAs

mRNA

double-stranded RNA
sRNA + mRNA

mRNA degraded

functionally turns gene off

A hot new topic in biology

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

Initiator tRNA

Met

mRNA

Start codon

mRNA binding site

Small ribosomal subunit

Large ribosomal subunit

P site

E site

A site

Met

Translation initiation complex

GTP → GDP + Pi

Protein Processing & Degradation

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

Ubiquitin

Protein to be degraded

Ubiquitinated protein

Proteasome

Protein entering a proteasome

Degraded protein (peptides)

Proteasome and ubiquitin to be recycled

Ubiquitin

1980s | 2004

- "Death tag"
 - mark unwanted proteins with a label
 - 76 amino acid polypeptide, **ubiquitin**
 - labeled proteins are broken down rapidly in "waste disposers"
 - proteasomes

Aaron Ciechanover
Israel

Avram Hershko
Israel

Irwin Rose
UC Riverside

Proteasome

- Protein-degrading "machine"
 - cell's waste disposer
 - can breakdown all proteins into 7-9 amino acid fragments

- transcription
 - DNA packing
 - transcription factors
- mRNA processing
 - splicing
- mRNA transport out of nucleus
 - breakdown by sRNA
- mRNA transport in cytoplasm
 - protection by 3' cap & poly-A tail
- translation
 - factors which block start of translation
- post-translation
 - protein processing
 - protein degradation
 - ubiquitin, proteasome