Chapter 2

Transcription, Translation and Regulation of Eukaryotic DNA
Introduction

DNA $\rightarrow$ RNA $\rightarrow$ proteins (Central Dogma)

Process by which information in the DNA is made available to the cell and organism.

Proteins have many functions:
- structural
- contractile
- catalytic (enzymes)
- regulatory
- protective
- storage
Introduction

Much of the genome (up to 80 or 90%) contains DNA that does not code for proteins.

Such DNA has been considered ‘junk’ or parasitic DNA.

Now we know that much of this DNA is transcribed and involved in regulating development and cell function.
Introduction

Development involves coordinated activity of a large number of different proteins coded for by DNA.

DNA also codes for structural elements: rRNA, tRNA, other small RNAs that are used directly without translation into proteins.
RNAs with Multiple Functions

- Small nuclear (snRNA): spice RNA in nucleus
- Small nucleolar (snoRNA): modify rRNAs
- Telomerase RNA: telomere synthesis
- Long noncoding (long ncRNA): regulatory functions
- Micro RNA (miRNA): gene regulation, bind to mRNAs
- Piwi-interacting (piRNA): defend against TEs & ?
- Small interfering (siRNA): gene regulation
Eukaryotic Genes

- Protein-coding genes are split into introns and exons
- Control elements of protein-coding genes are also important: promoters and enhancers
- Boundary elements may also be present to separate active and inactive regions of the chromosomes
Eukaryotic Genomes

• Nuclear membranes separate the processes of transcription in the nucleus and translation in the cytoplasm

• Nuclear membrane controls which molecules enter and leave the nucleus and is complex in structure
RNA Synthesis is Gene Transcription

Transcription is the first step in protein-coding gene expression

Coding strand of DNA serves as template

Sequence of RNA is determined by complementary - base pairing

RNA is a complementary copy of coding strand of DNA
RNA Synthesis is Gene Transcription

Step 1 in process
RNA Synthesis

RNA polymerase does not need a primer
(unlike DNA polymerase)

Synthesis is 5’ to 3’ (like DNA polymerase)

Eukaryotes: 3 types of RNA polymerase

Class I (large rRNA genes)
Class II (proteins & snRNAs)
Class III (tRNA, small rRNAs, some snRNAs)
Transcription

Complex, multistep process

Binding of RNA polymerase to DNA template
Attachment sites called promoters (TATAAT, CAAT)
Enhancers influence efficiency of attachment
Initiation involves ‘melting’ DNA at start site
Elongation involves adding ribonucleotides to produce pre-mRNA
RNA synthesis involves polymerization of free ribonucleotides by RNA polymerase.

DNA template is read in 3’ to 5’ direction.

RNA is synthesized in 5’ to 3’ direction.
Class II eukaryotic protein-coding gene

Noncoding introns are spliced out of pre-mRNA.

RNA polymerase attaches to promoter to initiate RNA synthesis.

Several enhancers affect gene regulation.

Dark segments are exons.
Transcripts are Longer than Protein-coding Genes

Leader sequence at start – length varies

Trailer segment

Termination can occur far down stream
Termination down stream from polyA signal
More details of a eukaryotic class II gene
Processing pre-mRNA

1) Modifying ends of pre-mRNA

Capping 5’ end to enable ribosome to bind with mRNA before protein synthesis starts

Adding polyA tail to 3’ end (40 to 200 As), which may determine how long mRNA survives in cytoplasm
Processing pre–mRNA in nucleus

2) Excising introns in 2 steps

- Occurs in spliceosomes (proteins, snRNAs)
- snRNAs are catalytic agents
- Splicing must be accurate

Introns may have a role in gene regulation, determining when, or in what tissue, the gene will be transcribed

Mutations in introns can be neutral or not
Pre-mRNA becomes mRNA
Translation = Protein Synthesis

- Amino-acid sequence determined by nt sequence of DNA
- Triplet codon in DNA (‘degenerate’ code)
- Requires ribosomes, tRNAs, enzymes to attach each aa to corresponding tRNA, initiation, elongation and termination factors
- Translation occurs in cytoplasm
Ribosomes

Cellular organelles, 2 subunits each of rRNA and proteins

Smaller unit binds mRNA and anticodon of tRNAs

Larger subunit interacts with aa-carrying end of tRNAs and catalyzes formation of peptide bonds

Ribosome = a ribozyme (acts as an enzyme)
Transfer RNAs

Carry an aa to the ribosome where they bind to the mRNA attached to the ribosome

Have a 3-base sequence, anticodon, complementary to a specific codon in mRNA

Binding between mRNA and tRNA is by hydrogen bonds
Transfer RNA has a complex shape

Anticodons determine which aa is inserted into growing polypeptide

tRNAs are ss RNA molecules 70 to 90 nt long
Peptide Bonds

Made between aligned aas until STOP codon is reached at end of mRNA and completed protein is released

Synthesis occurs from the amino end toward the carboxyl end
A polysome from *Chironomus* salivary gland cell

Ribosomes are attached along the mRNA
Extended protein chains are protruding from the ribosomes

This end shows the earlier stages of protein production
Proteins

Three-dimensional structure is important to function

AA sequence contains sufficient information to specify 3-D structure usually

Protein folds are fundamental to function

Most small proteins fold spontaneously into correct form

Genome data and protein structure under intense study (genomics and proteomics)
mRNA Surveillance

Damaged or truncated mRNAs are monitored

Aberrant mRNAs are degraded

Quality control is essential
Key Points

- **Transcription**
  - Protein-coding genes are transcribed into pre-mRNA that must be processed in the nucleus into mRNA before moving to the cytoplasm for translation.
  - Transcription requires RNA polymerase, but no primer; transcription of coding strand of DNA occurs from 5’ to 3’ end.
  - Ribonucleotides include A, C, G, and U.
Key Points

• Transcription

• Processing of mRNA involves splicing out the introns in a spliceosome, capping, and methylation of the 5’ end and adding a poly(A) tail to the 3’ end

• mRNA for protein-coding genes includes a leader sequence and a trailer segment

• genes for rRNA, tRNA and other small RNAs are used directly without translation into proteins
Key Points

- **Translation**
  - Protein-coding genes are translated in the cytoplasm on ribosomes (which consist of rRNAs and proteins (ribozymes))
  - tRNAs have an anticodon complementary to the codon of the mRNA
  - tRNAs have a 3D structure and contain an amino-acid attachment site where the aa is attached by specific enzymes
Key Points

• Translation
  • tRNAs carry a specific aa to the ribosome where they bind to the mRNA molecule based on complementary base pairing
  • mRNA is monitored and destroyed if damaged
  • stability of mRNA is regulated so specific proteins are produced in appropriate amounts at the appropriate location to ensure normal development and metabolic activities
The Nucleus

The nucleus distinguishes prokaryotes from eukaryotes: eukaryotic nucleus is where DNA replication, transcription and RNA processing occur.

Nuclear membranes prevent free movement between nucleus and cytoplasm.

Nucleus a distinct compartment.
Import and Export from the Nucleus

Nuclear pore complex has a diameter of 120 nm and consists of > 50 proteins

Ca. 30 X size of a ribosome

mRNA is transported from nucleus to cytoplasm after synthesis and processing
Import and Export from the Nucleus

Proteins (histones, RNA and DNA polymerases, splicing factors) manufactured in cytoplasm are transported into nucleus via pore complex.

Transport is active (most) or passive (small molecules < 50 kDa in size).
Import and Export from the Nucleus

Many proteins contain specific signals (exportins) that ensure delivery to correct location in the cell.
Protein Transport in Cytoplasm

Proteins are transferred to endoplasmic reticulum (ER) while being translated on membrane-bound ribosome.

ER is a network of membrane-enclosed tubules and sacs extending throughout the cytoplasm.
Protein Transport in Cytoplasm

Proteins destined for cytosol (fluid part of cytoplasm), nucleus or mt are synthesized on free ribosomes and released into cytosol.

Rough and smooth endoplasmic reticulums (ER) perform different functions:
- Rough ER involved in protein processing
- Smooth ER functions in lipid metabolism
Protein Transport in Cytoplasm

Proteins received from ER are processed and sorted for transport to Golgi complex.

The Golgi specializes in processing and sorting proteins and lipids prior to transport to lysosomes, plasma membrane, or secretion.

Protein modifications in Golgi include glycosylation.
Protein Transport in Cytoplasm

Glycosylation involves adding carbohydrates to proteins

Bacteria such as *E. coli* can’t glycosylate, so can’t produce proteins with proper modifications for eukaryotes

• This can be important in biotechnology projects
Chaperones and the Proteosome

Protein function depends on correct protein folding into 3-dimensional forms

**Chaperones** are proteins that catalyze proper protein folding

**Proteosomes** are large protein complexes where proteins are destroyed after **ubiquitin** is attached to the protein
Secondary Structures of Proteins

Most common types

alpha helix

beta sheet
RNA Silencing or Interference

A mechanism to defend against invasion by mobile DNA elements (TEs) that cause mutations when they insert into or near a gene.

When ds RNA is injected into eukaryotic cells by TEs or viruses, the ds RNA appears to function as a signal that the cell has been invaded.
ds RNA is cut into smaller chunks of about 22 nt, which are degraded in a second reaction by RNase.

The 2-step reaction is efficient because each molecule of ds RNA primes several RNase molecules.

RNA silencing can be used to knock out genes to study their function.
RNA Silencing or Interference

double-stranded RNA (ds RNA) is introduced into the organism or produced naturally.

Dicer, an RNase III enzyme, cleaves the ds RNA into fragments called small interfering RNAs (siRNAs) that are 21-23 nt in length.

The siRNAs unwind and the antisense strand binds to the RNA-induced silencing complex (RISC), which includes Argonaute proteins. Argonaute cleaves and eliminates the sense strand of the siRNA duplex, leading to an active RISC.

The complex of small interfering RNAs (siRNAs) and RISC couples to the target homologous mRNA.

The mRNA is cleaved and is unable to be translated.
Gene Regulation

How can a single genome code for so many cell and tissue types?

Multiple mechanisms regulate genes

- Regulate transcription
- Alternative splicing of mRNA transcripts
- DNA amplification
- Methylation of cytosine bases
- Translational control
Regulating Transcription Levels

Involves transcriptional activator proteins that bind with upstream DNA sequences to prepare a gene for transcription.

Some have a helix-turn-helix structure, which is a sequence of aa that forms a pair of alpha helices separated by a bend, which fits into the grooves of ds DNA and allows proteins to bind to the DNA.

Ex: homeo domain
Regulating Transcription Levels

Zinc finger proteins a second type of activator protein

Have loops of repeating aa each associated with a zinc atom

Can bind with major groove of DNA
Many genes are active only a short time during development in precisely defined domains of body

Gene regulation occurs in multiple ways, including regulating mRNA stability

- Removal of polyA tail
- Premature termination of translation due to premature termination codes
- mRNA relocalization
Regulating Transcription Levels

Leucine zippers a third type of DNA-binding proteins

- Contain 4 to 5 leucines separated by 6 aa
- Can interdigitate and dimerize in a specific interaction with DNA

Hormones (such as ecdysone) also act directly on genome to activate or repress genes
Gene regulation with alternative promoters

A) Gene structure

B) Larval transcript and processing

C) Adult transcript and processing

Larval transcript

Adult transcript
Epigenetic Regulation

Some inheritance is non-Mendelian (not related to DNA sequence in the genome)

- DNA may be reversibly modified by histone modifications, methylation of cytosines, chromatin remodeling, and targeting of small ncRNAs

- A fertilized egg develops into an adult because the genome becomes epigenetically programmed to generate distinct epigenomes in different cell types
Insulators and Boundaries

Genes have ‘neighborhoods’ that can affect transcription

**Insulators**: specialized DNA sequences provide a barrier against influences of surrounding DNA sequences by blocking enhancer-promoter interactions when placed between the enhancer and promoter

**Transgenes** vary in expression depending on where they are located: insulators may be useful to enhance expression
Imprinting

**Reversible**, differential marking of genes or chromosomes determined by sex of parent from whom genetic material is inherited.

**Methylation** of cytosines at carbon 5 position of CpG dinucleotides is common in many prokaryotes and eukaryotes.

**Methylation** a defense mechanism against invading DNA parasites.
Imprinting

Methylation role in insects is controversial

**Hypomethylation** of chromosomes in citrus mealybugs is associated with chromosome heterochromatization

Methylation amount in insects varies

*D. melanogaster* especially controversial
**Genetics and Epigenetics**

**Genetics:** mutations in DNA are inherited in the soma and germ line.

**Epigenetics:** changes in chromatin structure, including histone modifications, effects of small noncoding RNAs, that are reversible. Some are limited to the soma but others may be transmitted through the germ line. Inheritance that is NOT based on differences in DNA sequences.
Genome Evolution

Split genes and RNA splicing crucial to understanding eukaryotic genome evolution

Alternative RNA splicing produces multiple mRNAs → different proteins

‘junkyard evolution’ = some solutions can occur with assembly of old components into new combinations
Genome Evolution

Unlike prokaryotes, eukaryotes make RNA that differs from the DNA in the genome. This allows genomic information to be influenced by the environment.