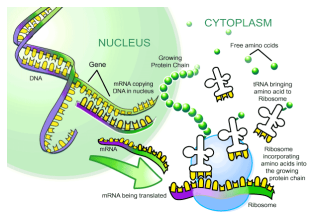## From Gene to Protein
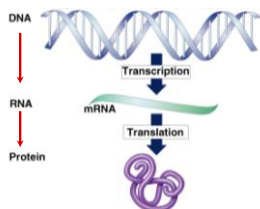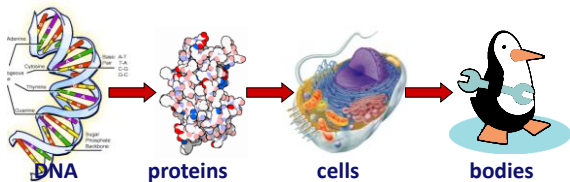
How Genes Work

What do genes code for?
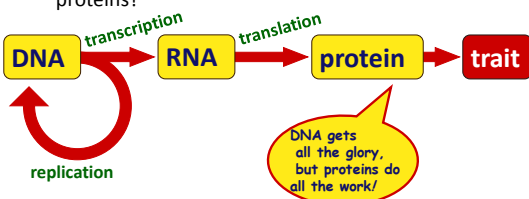
- How does DNA code for cells & bodies?
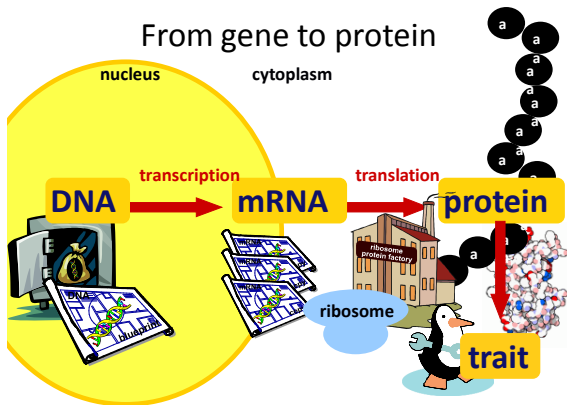 - how are cells and bodies made from the instructions in DNA

DNA → proteins → cells → bodies
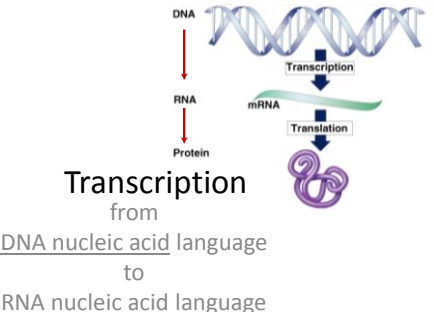
The "Central Dogma"

- Flow of genetic information in a cell
  - How do we move information from DNA to proteins?

DNA — transcription → RNA — translation → protein → trait

replication

DNA gets all the glory, but proteins do all the work!

From gene to protein

nucleus   cytoplasm

DNA —transcription→ **mRNA** —translation→ **protein**

ribosome

trait

DNA

RNA
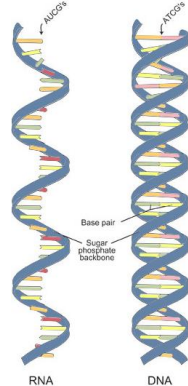
Protein

Transcription

mRNA

Translation

Transcription
from
DNA nucleic acid language
to
RNA nucleic acid language

2007-2008

## RNA

- ribose sugar
- N-bases
  - uracil instead of thymine
  - U : A
  - C : G
- single stranded
- lots of RNAs
  - mRNA, tRNA, rRNA, siRNA...

DNA —transcription→ **RNA**

AUCG's   ATCG's

Base pair

Sugar
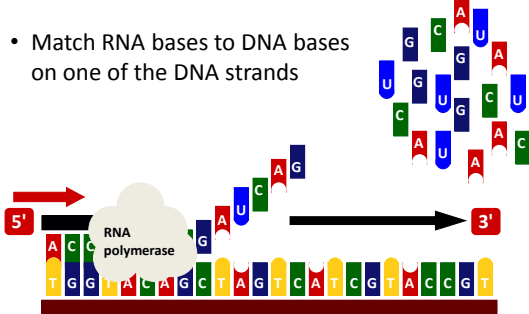phosphate
backbone

RNA   DNA

Transcription

- Making mRNA
  - transcribed DNA strand = template strand
  - untranscribed DNA strand = coding strand
    - same sequence as RNA
  - synthesis of complementary RNA strand
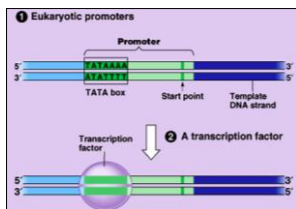   - transcription bubble
  - enzyme
    - RNA polymerase
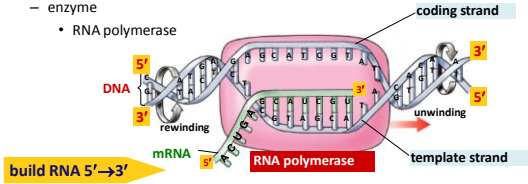
Matching bases of DNA & RNA

- Match RNA bases to DNA bases on one of the DNA strands


Which gene is read?

- Promoter region
  - binding site before beginning of gene
  - TATA box binding site
  - binding site for RNA polymerase & transcription factors
- Enhancer region
  - binding site far upstream of gene
   - turns transcription on HIGH

Transcription Factors

- Initiation complex
  - transcription factors bind to promoter region
    - suite of proteins which bind to DNA
    - turn on or off transcription
  - trigger the binding of RNA polymerase to DNA
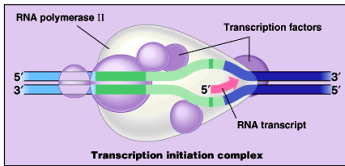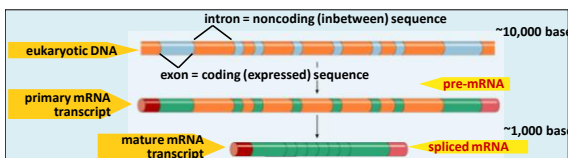

Eukaryotic genes have junk!

- Eukaryotic genes are not continuous
  - exons = the real gene
    - expressed / coding DNA
  - introns = the junk
    - inbetween sequence
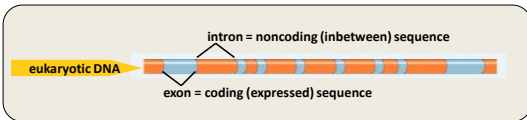
mRNA splicing

- Post-transcriptional processing
  - eukaryotic mRNA needs work after transcription
  - primary transcript = pre-mRNA
  - mRNA splicing
    - edit out introns
 - make mature mRNA transcript

Splicing must be accurate

- No room for mistakes!
  - 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|
```
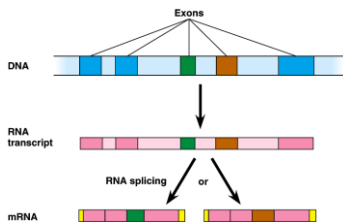
RNA splicing enzymes

- snRNPs
  - small nuclear RNA
- Spliceosome
  - several snRNPs
  - recognize splice site sequenc
    - cut & paste gene

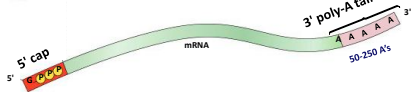No, not smurfs! "snurps"
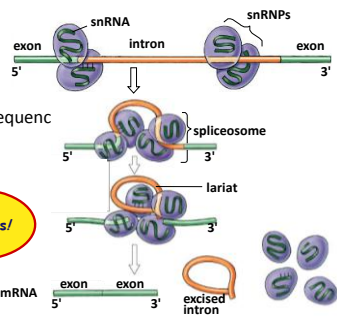
Alternative splicing

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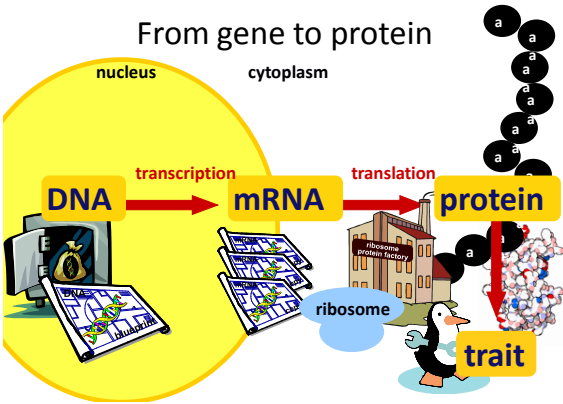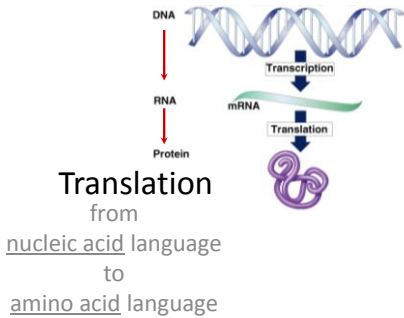 - different segments treated as exons

More post-transcriptional processing

- Need to protect mRNA on its trip from nucleus to cytoplasm
  - enzymes in cytoplasm attack mRNA
    - protect the ends of the molecule
    - add 5' GTP/Methylated cap
    - add poly-A tail
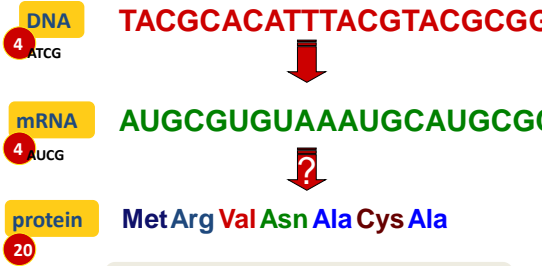   - longer tail, mRNA lasts longer: produces more protein
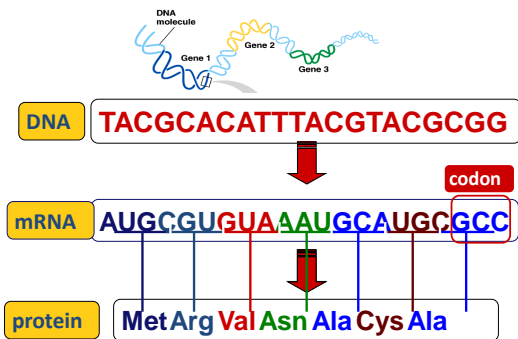

From gene to protein

Translation
from
nucleic acid language
to
amino acid language

2007-2008

How does mRNA code for proteins?

DNA **4** ATCG
TACGCACATTTACGTACGCGG

mRNA **4** AUCG
AUGCGUGUAAAUGCAUGCGG
?

protein **20**
Met Arg Val Asn Ala Cys Ala

How can you code for 20 amino acids with only 4 nucleotide bases (A,U,G,C)?

mRNA codes for proteins in triplets

DNA TACGCACATTTACGTACGCGG

codon

mRNA AUGCGUGUAAAUGCAUGCGCC

protein Met Arg Val Asn Ala Cys Ala

Cracking the code

1960 | 1968
Nirenberg & Khorana

• Crick
 – determined 3-letter (triplet) codon system

WHYDIDTHEREDBATEATTHEFATRAT

■ Nirenberg (47) & Khorana (17)
 ◆ determined mRNA–amino acid match
 ◆ added fabricated mRNA to test tube of ribosomes, tRNA & amino acids
  ▪ created artificial UUUUU... mRNA
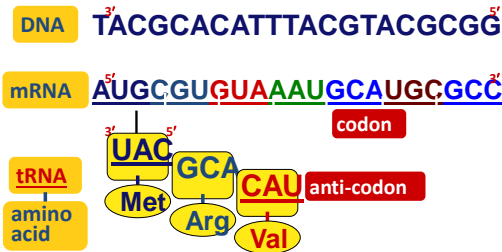 ▪ found that UUU coded for phenylalanine

The code

- Code for <u>ALL</u> life!
  - strongest support for a common origin for all life
- Code is redundant
  - several codons for each amino acid
  - 3rd base "wobble"

- **Start codon**
  - ◆ AUG
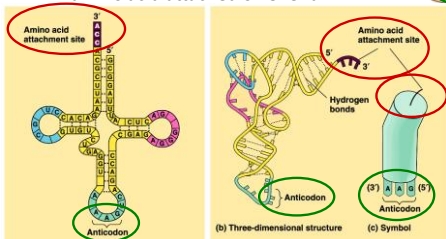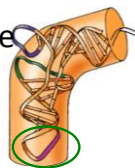 - ◆ methionine
- **Stop codons**
  - ◆ UGA, UAA, UAG

How are the codons matched to amino acids?

DNA 3' TACGCACATTTACGTACGCGG 5'

mRNA 5' AUGCGUGUAAAUGCAUGCGCC 3'

codon

3' UAC 5' GCA CAU anti-codon

tRNA
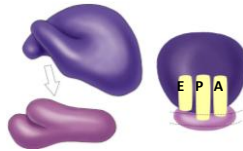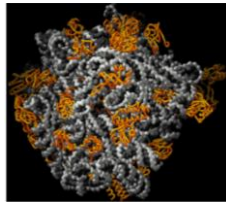**amino acid**

Met Arg Val

Transfer RNA structure

- "Clover leaf" structure
  - anticodon on "clover leaf" end
  - amino acid attached on 3' end
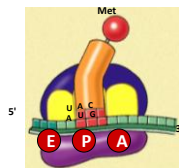
Ribosomes

- Facilitate coupling of tRNA anticodon to mRNA codon
- Structure
  - ribosomal RNA (rRNA) & proteins
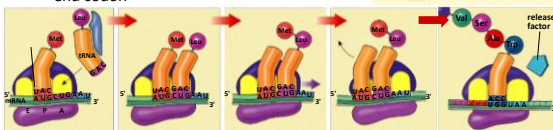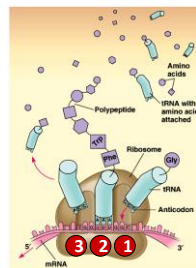 - 2 subunits
    - large
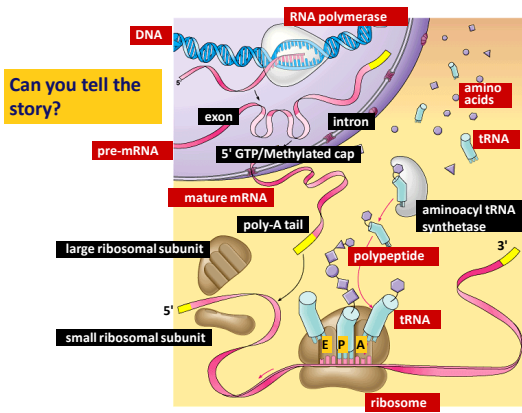   - small

Ribosomes

- A site (aminoacyl-tRNA site)
  - holds tRNA carrying next amino acid to be added to chain
- P site (peptidyl-tRNA site)
  - holds tRNA carrying growing polypeptide chain
- E site (exit site)
  - empty tRNA leaves ribosome from exit site

Building a polypeptide

- Initiation
  - brings together mRNA, ribosome subunits, initiator tRNA
- Elongation
  - adding amino acids based on codon sequence
- Termination
  - end codon

Can you tell the story?

DNA

RNA polymerase

exon

intron

pre-mRNA

5' GTP/Methylated cap

mature mRNA

poly-A tail

large ribosomal subunit

polypeptide

aminoacyl tRNA synthetase

amino acids

tRNA

tRNA

3'

5'

small ribosomal subunit

E P A

ribosome