Metabolism Teaches Us About Genes
- Metabolic defects
  - studying metabolic diseases suggested that genes specified proteins
    - alkaptonuria (black urine from alkapton, a.k.a. homogentisic acid)
    - PKU (phenylketonuria)
  - each disease is caused by non-functional enzyme

1 Gene – 1 Enzyme Hypothesis
- Beadle & Tatum
  - Compared mutants of bread mold, Neurospora fungus
    - created mutations by X-ray treatments
      - X-rays break DNA
      - inactivate a gene
    - wild type grows on “minimal” media
      - sugars + required precursor nutrient to synthesize essential amino acids
    - mutants require added amino acids
      - each type of mutant lacks a certain enzyme needed to produce a certain amino acid
      - non-functional enzyme = broken gene
So... What is a Gene?

- One gene – one enzyme
- All genes code for enzymes
- But there is code for proteins that are not enzymes
- One gene – one protein
- A gene codes for a single protein
- But many proteins are composed of several polypeptides chains each coded by a different gene
- One gene – one polypeptide
- But many genes have the code for only RNA
- One gene – one product
- But many genes can code for more than one product ...

Defining a Gene...

“Defining a gene is problematic because... one gene can code for several protein products, some genes code only for RNA, two genes can overlap, and there are many other complications.”

— Elizabeth Pennisi, Science 2003

The “Central Dogma”

- How do we move information from DNA to proteins?

From nucleus to cytoplasm...

- Where are the genes?
  - Genes are on chromosomes in nucleus
- Where are proteins synthesized?
  - Proteins made in cytoplasm by ribosomes
- How does the information get from nucleus to cytoplasm?
  - Messenger RNA

RNA

- Ribose sugar
- N-bases
  - Uracil instead of thymine
  - U : A
  - C : G
- Single stranded
- mRNA, rRNA, tRNA, siRNA...

Transcription

DNA transcription RNA transcription
Transcription

- Transcribed DNA strand = **template strand**
  - untranscribed DNA strand = **coding strand**
- Synthesis of complementary RNA strand
  - transcription bubble
- Enzyme that facilitates the building of RNA:
  - RNA polymerase

Role of promoter

1. Where to start reading = starting point
2. Which strand to read = template strand
3. Direction on DNA = always **reads** DNA 3’→5’

Transcription

- Initiation
  - RNA polymerase binds to **promoter** sequence on DNA

Transcription

- Elongation
  - RNA polymerase unwinds DNA ~20 base pairs at a time
  - reads DNA 3’→5’
  - builds RNA 5’→3’ (the enzyme governs the synthesis!)

No proofreading

- 1 error/10^6 bases
- many copies
- short life
- not worth it!

Transcription

- Termination
  - RNA polymerase stops at **termination** sequence
  - mRNA leaves nucleus through pores

Prokaryote vs. Eukaryote Genetics

- Differences between prokaryotes & eukaryotes
  - time & physical separation between processes
  - RNA processing

Eukaryotic Post-transcriptional Processing

- Primary transcript
  - eukaryotic mRNA needs work after transcription
- Protect mRNA
  - from RNA-ase enzymes in cytoplasm
    - add 5’ G cap
    - add polyA tail
- Edit out introns
Protecting RNA
- 5' methylated G cap added
  - G trinucleoside (G-P-P-P)
  - protects mRNA from RNase (hydrolytic enzymes)
- 3' poly-A tail added
  - 50-250 A's
  - helps export of RNA from nucleus
  - protects mRNA from RNase (hydrolytic enzymes)

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 introns!!

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

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

Ribozyme
- RNA as enzyme
  - DNA → RNA → PROTEINS → CELLS → ORGANISMS
  - RNA as enzyme

Splicing Enzymes
- snRNPs
  - small nuclear RNA
  - RNA + proteins
- Spliceosome
  - several snRNPs
  - recognize splice site sequence
    - cut & paste
- RNA as a ribozyme
  - some mRNA can splice itself
  - RNA as enzyme
Discovery of Split Genes

1977 | 1993

Richard Roberts
NE BioLabs

Philip Sharp
MIT

adenovirus
common cold

Structure of Antibodies

antigen-binding site
variable region

How do vertebrates produce millions of antibody proteins, if they only have a few hundred genes coding for those proteins?

DNA of differentiated B cell
chromosome of undifferentiated B cell

From Gene to Protein

DNA → mRNA → protein

mRNA leaves nucleus through nuclear pores

proteins synthesized by ribosomes using instructions on mRNA

The Transcriptional Unit (gene?)

DNA

promoter
transcription start
transcriptional unit
transcription stop
introns
exons

Translation in Prokaryotes

- Transcription & translation are simultaneous in bacteria
  - DNA is in cytoplasm
  - no mRNA editing needed

DNA polymerase
RNA polymerase
DNA
RNA
Ribosomes
mRNA
Polyribosome
Prokaryote vs. Eukaryote Genetics

- **Prokaryotes**
  - DNA in cytoplasm
  - circular chromosome
  - naked DNA
  - no introns

- **Eukaryotes**
  - DNA in nucleus
  - linear chromosomes
  - DNA wound on histone proteins
  - introns vs. exons

How Does DNA Code for Proteins

- **DNA**
  \[ \text{TACGCACATTACGTACGCGG} \]

- **mRNA**
  \[ \text{AUGCGUGUAUAUGCAUGCGCC} \]

- **Protein**
  \[ \text{MetArgValAsnAlaCysAla} \]

Cracking the Code

- **Nirenberg & Matthaai**
  - determined 1st codon–amino acid match
    - UUU coded for phenylalanine
  - created artificial poly(U) mRNA
  - added mRNA to test tube of ribosomes, tRNA & amino acids
    - mRNA synthesized single amino acid polypeptide chain

Translation

- **Codons**
  - blocks of 3 nucleotides decoded into the sequence of amino acids

mRNA Codes for Proteins in Triplets

- **DNA**
  \[ \text{TACGCACATTACGTACGCGG} \]

- **mRNA**
  \[ \text{AUGCGUGUAUAUGCAUGCGCC} \]

- **Protein**
  \[ \text{MetArgValAsnAlaCysAla} \]

mRNA codes for proteins in triplets... CODONS!
The Code!

- For ALL life!
  - strongest support for a common origin for all life
- Code is redundant
  - several codons for each amino acid

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

How are Codons Matched to Amino Acids?

- DNA
  - TACGCACTTTAGTGACGCGG
- mRNA
  - AUGGUUGUAAAUGCAUGCCCGC
- tRNA
  - UAC
- amino acid
  - Met
- anti-codon
  - CAU

tRNA Structure

- “Clover leaf” structure
  - anticodon on “clover leaf” end
  - amino acid attached on 3’ end

Loading tRNA

- Aminoacyl tRNA synthetase
  - enzyme which bonds amino acid to tRNA
  - endergonic reaction
    - ATP → AMP
  - energy stored in tRNA-amino acid bond
    - unstable
  - so it can release amino acid at ribosome

Ribosomes

- Facilitate coupling of tRNA anticodon to mRNA codon
  - organelle or enzyme?
- Structure
  - ribosomal RNA (rRNA) & proteins
  - 2 subunits
    - large
    - small
Ribosomes
- **P** site (peptidyl-tRNA site)
  - holds tRNA carrying growing polypeptide chain
- **A** site (aminoacyl-tRNA site)
  - holds tRNA carrying next amino acid to be added to chain
- **E** site (exit site)
  - empty tRNA leaves ribosome from exit site

Building a Polypeptide
- **Initiation**
  - brings together mRNA, ribosome subunits, proteins & initiator tRNA
- **Elongation**
- **Termination**

Elongation: Growing a Polypeptide

Termination: Release Polypeptide
- **Release factor**
  - “release protein” bonds to A site
  - releases ribosome subunits, mRNA, and polypeptide

So now what happens to the polypeptide?

Protein Targeting
- **Signal peptide**
  - address label

Destinations:
- secretion
- nucleus
- mitochondria
- chloroplasts
- cell membrane
- cytoplasm

Can you tell the eukaryotic story?