Introduction to Molecular Biology and Genomics

BMI/CS 776

www.biostat.wisc.edu/~craven/776.html

Mark Craven

craven@biostat.wisc.edu

January 2002

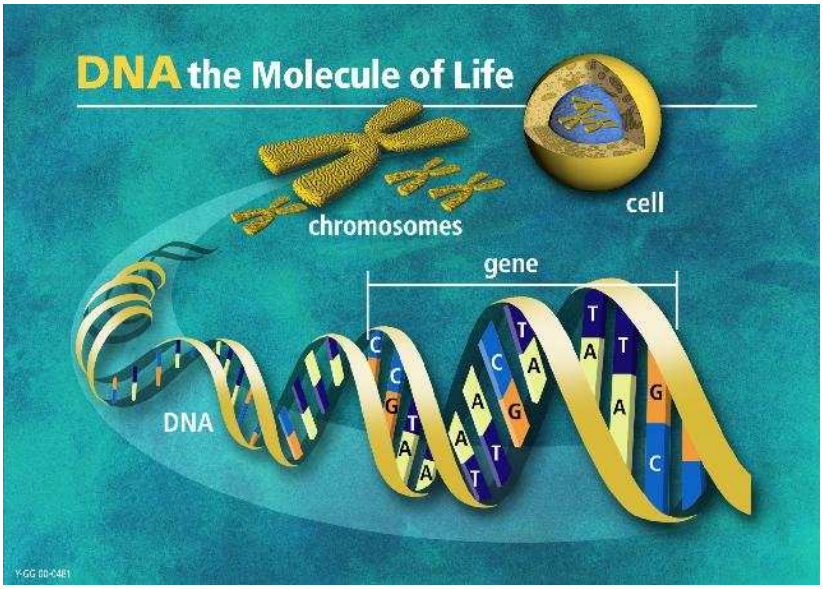


image from the DOE Human Genome Program
http://www.ornl.gov/hgmis

DNA

- can be thought of as the “blueprint” for an organism
- composed of small molecules called *nucleotides*
- four different nucleotides distinguished by the four *bases*: adenine (A), cytosine (C), guanine (G) and thymine (T)
- a polymer: large molecule consisting of similar units (nucleotides in this case)

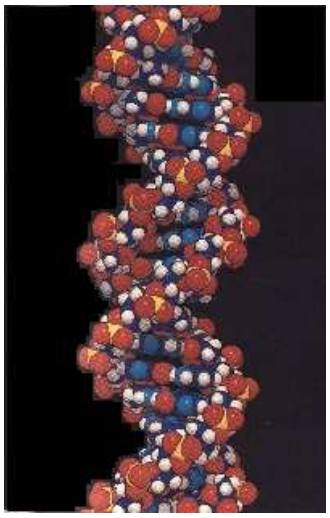
DNA

- a single strand of DNA can be thought of as a string composed of the four letters: A, C, G, T

  ctgctggaccgggtgctaggaccctgactgcc
  cggggccgggggtgcggggcccgctgag…

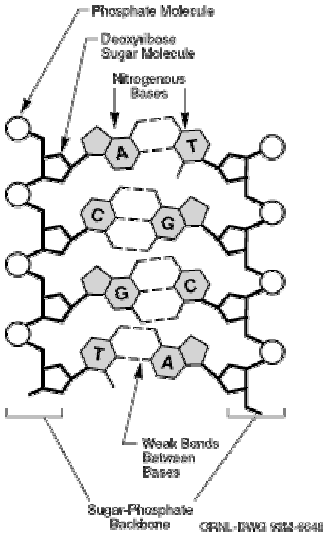
The Double Helix

- DNA molecules usually consist of two strands arranged in the famous double helix



Watson-Crick Base Pairs

- in double-strand DNA
  - A always bonds to T
  - C always bonds to G



The Double Helix

- each strand of DNA has a “direction”
  - at one end, the terminal carbon atom in the backbone is the 5’ carbon atom of the terminal sugar
  - at the other end, the terminal carbon atom is the 3’ carbon atom of the terminal sugar
- therefore we can talk about the *5’* and the *3’* ends of a DNA strand
- in a double helix, the strands are *antiparallel* (arrows drawn from the 5’ end to the 3’ end go in opposite directions)

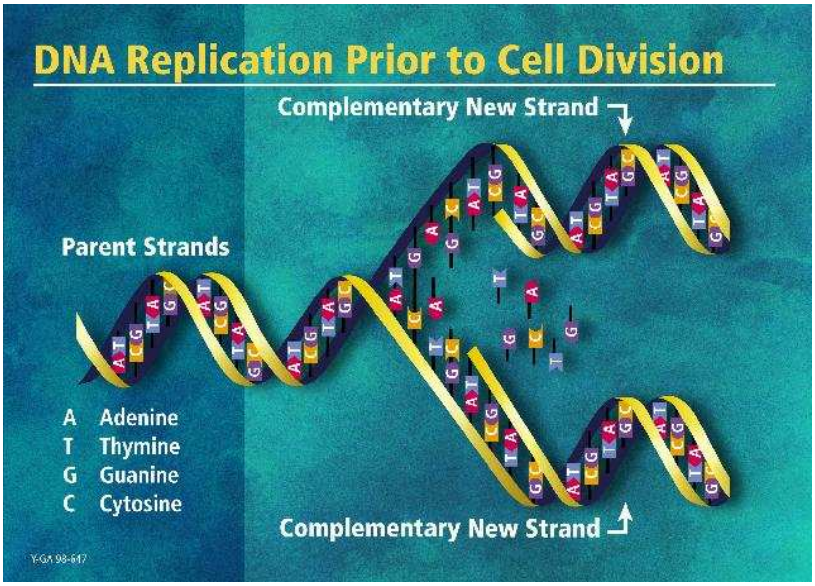
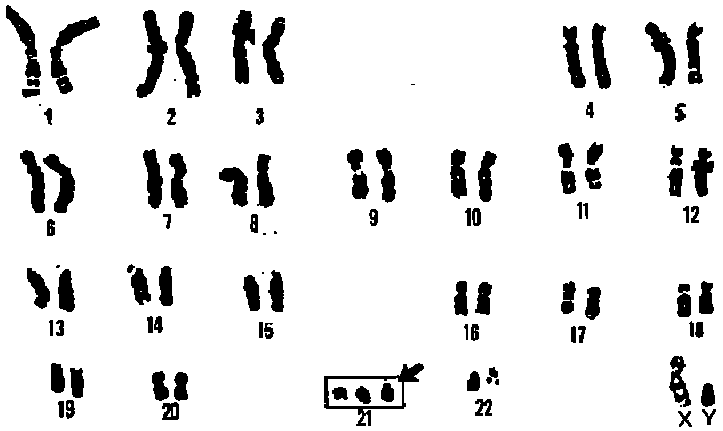


image from the DOE Human Genome Program
http://www.ornl.gov/hgmis

Chromosomes

- DNA is packaged into individual *chromosomes* (along with proteins)
- *prokaryotes* (single-celled organisms lacking nuclei) have a single circular chromosome
- *eukaryotes* (organisms with nuclei) have a species-specific number of linear chromosomes

Human Chromosomes



Genomes

- the term *genome* refers to the complete complement of DNA for a given species
- the human genome consists of 46 chromosomes.
- every cell (except sex cells and mature red blood cells) contains the complete genome of an organism

Proteins

- proteins are molecules composed of one or more *polypeptides*
- a polypeptide is a polymer composed of *amino acids*
- cells build their proteins from 20 different amino acids
- a polypeptide can be thought of as a string composed from a 20-character alphabet

Protein Functions

- structural support
- storage of amino acids
- transport of other substances
- coordination of an organism's activities
- response of cell to chemical stimuli
- movement
- protection against disease
- selective acceleration of chemical reactions

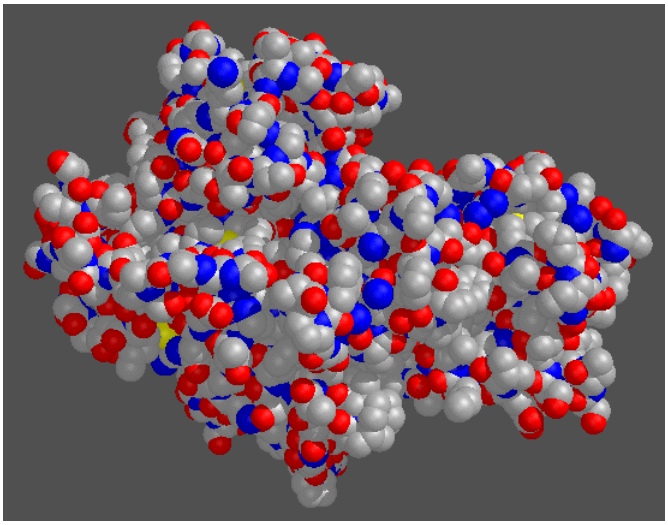
Amino Acids

Alanine	**Ala**	**A**
Arginine	**Arg**	**R**
Aspartic Acid	**Asp**	**D**
Asparagine	**Asn**	**N**
Cysteine	**Cys**	**C**
Glutamic Acid	**Glu**	**E**
Glutamine	**Gln**	**Q**
Glycine	**Gly**	**G**
Histidine	**His**	**H**
Isoleucine	**Ile**	**I**
Leucine	**Leu**	**L**
Lysine	**Lys**	**K**
Methionine	**Met**	**M**
Phenylalanine	**Phe**	**F**
Proline	**Pro**	**P**
Serine	**Ser**	**S**
Threonine	**Thr**	**T**
Tryptophan	**Trp**	**W**
Tyrosine	**Tyr**	**Y**
Valine	**Val**	**V**

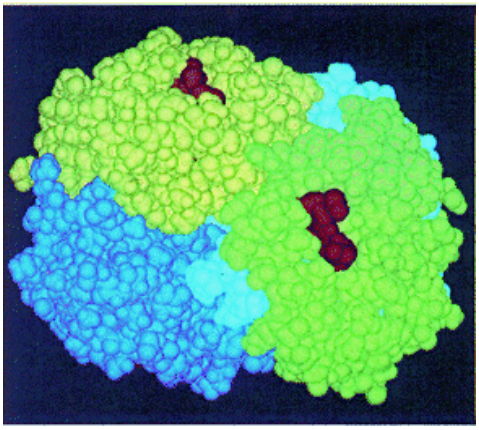
Amino Acid Sequence of Hexokinase

```
        5       10       15       20       25       30
  1 A A S X D X S L V E V H X X V F I V P P X I L Q A V V S I A
 31 T T R X D D X D S A A A S I P M V P G W V L K Q V X G S Q A
 61 G S F L A I V M G G G D L E V I L I X L A G Y Q E S S I X A
 91 S R S L A A S M X T T A I P S D L W G N X A X S N A A F S S
121 X E F S S X A G S V P L G F T F X E A G A K E X V I K G Q I
151 T X Q A X A F S L A X L X K L I S A M X N A X F P A G D X X
181 X X V A D I X D S H G I L X X V N Y T D A X I K M G I I F G
211 S G V N A A Y W C D S T X I A D A A D A G X X G G A G X M X
241 V C C X Q D S F R K A F P S L P Q I X Y X X T L N X X S P X
271 A X K T F E K N S X A K N X G Q S L R D V L M X Y K X X G Q
301 X H X X X A X D F X A A N V E N S S Y P A K I Q K L P H F D
331 L R X X X D L F X G D Q G I A X K T X M K X V V R R X L F L
361 I A A Y A F R L V V C X I X A I C Q K K G Y S S G H I A A X
391 G S X R D Y S G F S X N S A T X N X N I Y G W P Q S A X X S
421 K P I X I T P A I D G E G A A X X V I X S I A S S Q X X X A
451 X X S A X X A
```

Hexokinase



Hemoglobin



- protein built from 4 polypeptides
- responsible for carrying oxygen in red blood cells

Genes

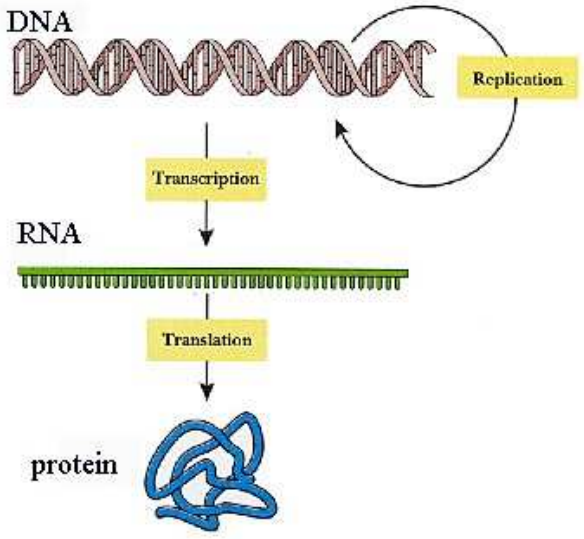
- genes are the basic units of heredity
- a gene is a sequence of bases that carries the information required for constructing a particular protein (polypeptide really)
- a gene is said to *encode* a protein
- the human genome comprises ~ 40,000 genes
  - there is some controversy about this number

Gene Density

- not all of the DNA in a genome encodes protein:

microbes	90% coding	gene/kb
human	3% coding	gene/35kb

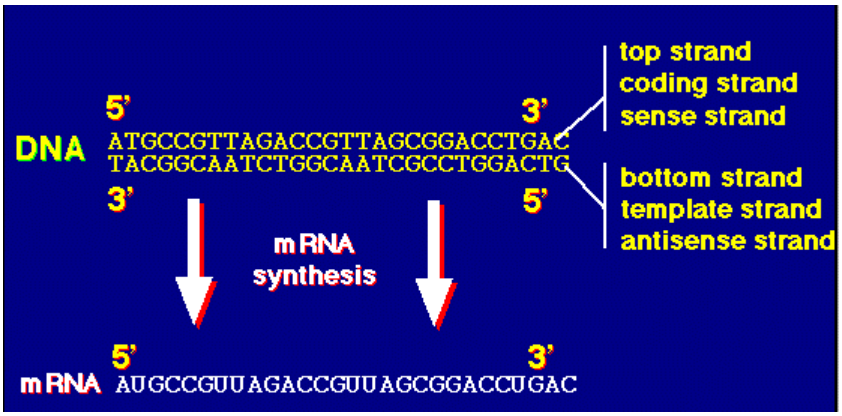
The Central Dogma



RNA

- RNA is like DNA except:
  - backbone is a little different
  - usually single stranded
  - the base uracil (U) is used in place of thymine (T)
- a strand of RNA can be thought of as a string composed of the four letters: A, C, G, U

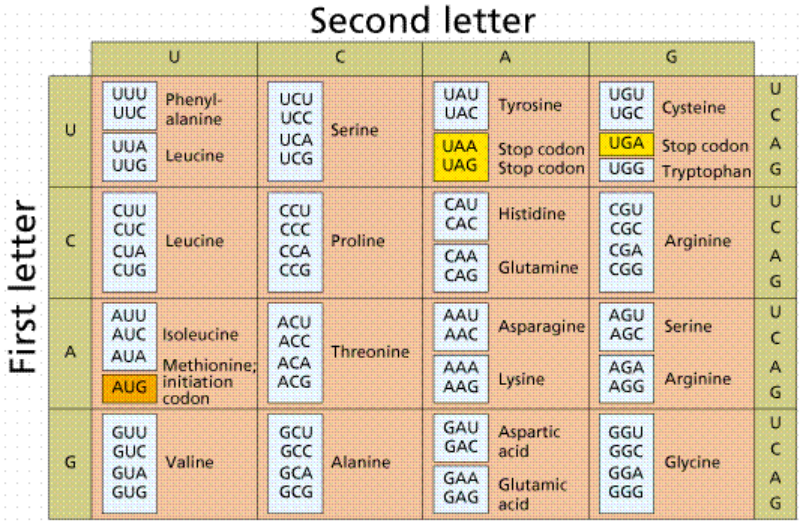
Transcription



Transcription

- *RNA polymerase* is the enzyme that builds an RNA strand from a gene
- RNA that is transcribed from a gene is called *messenger RNA (mRNA)*
  - we'll talk about other varieties of RNA later in the course

The Genetic Code



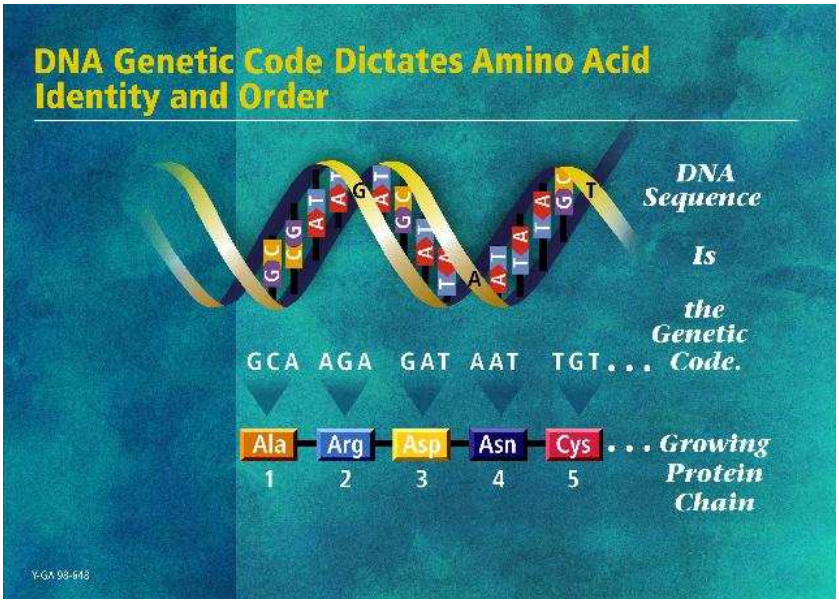
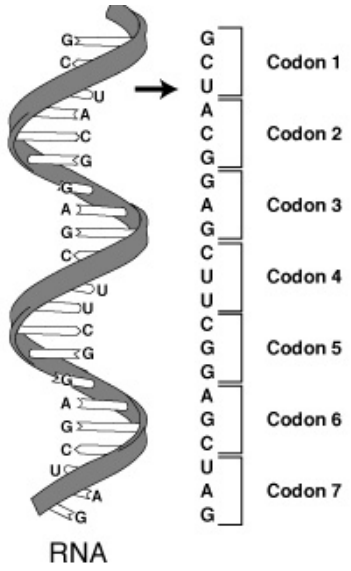


image from the DOE Human Genome Program
http://www.ornl.gov/hgmis

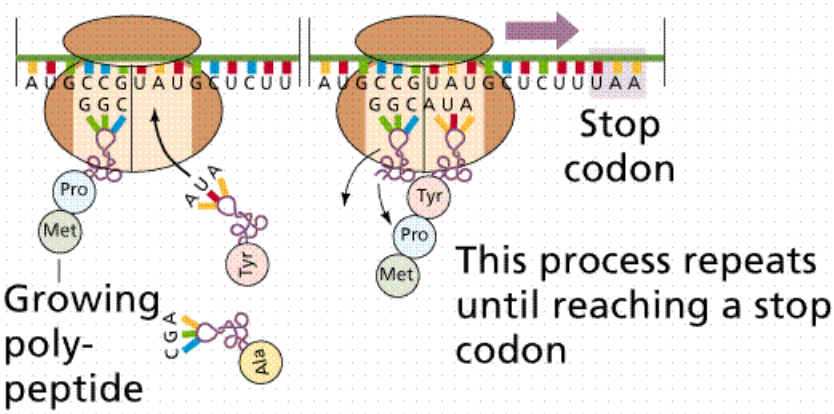
Translation

- *ribosomes* are the machines that synthesize proteins from mRNA
- the grouping of codons is called the *reading frame*
- translation begins with the *start codon*
- translation ends with the *stop codon*

Codons and Reading Frames



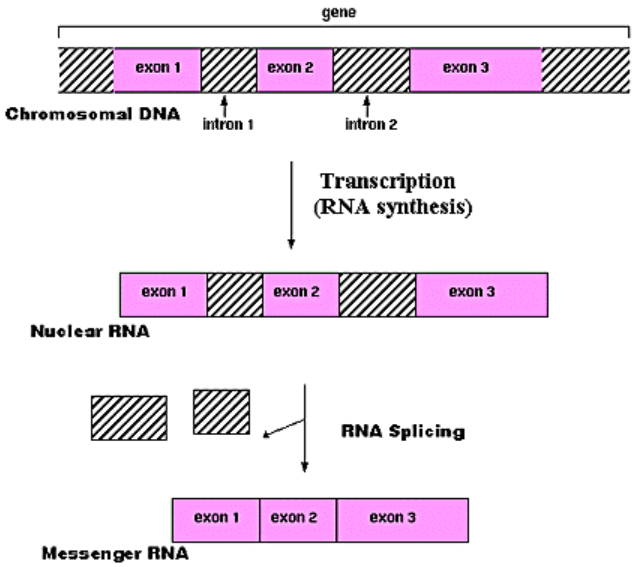
Translation



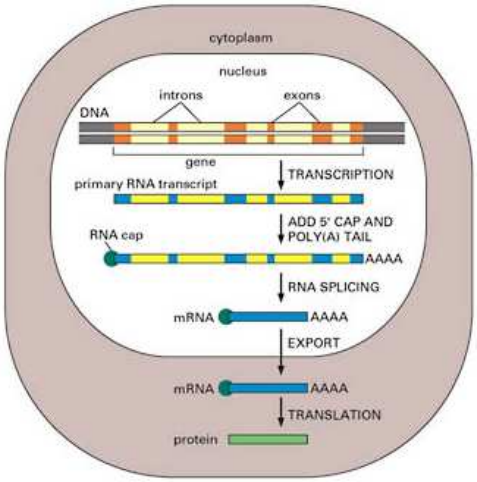
RNA Processing in Eukaryotes

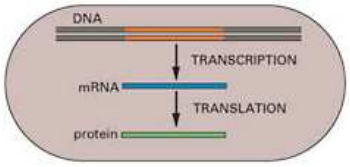
- *eukaryotes* are organisms that have enclosed nuclei in their cells
- in eukaryotes, mRNA consists of alternating *exon/intron* segments
- exons are the coding parts
- introns are spliced out before translation

RNA Splicing



Protein Synthesis in Eukaryotes vs. Prokaryotes





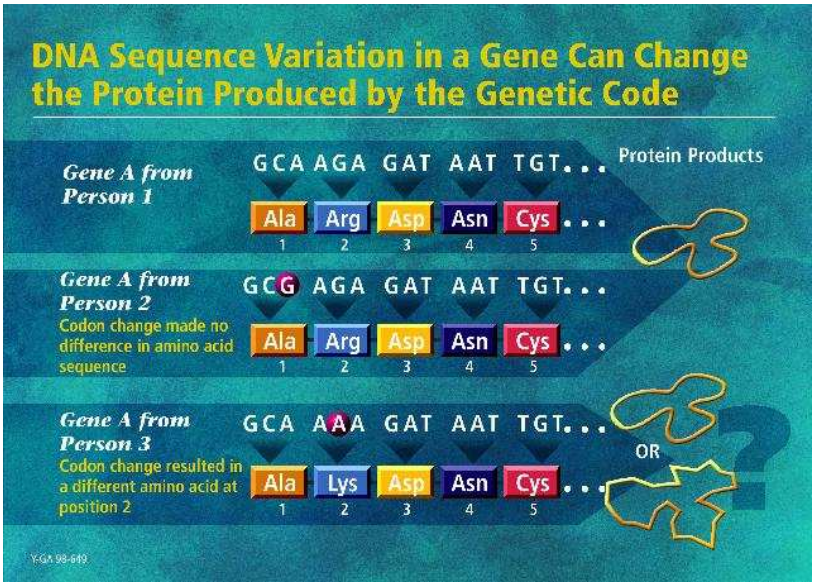
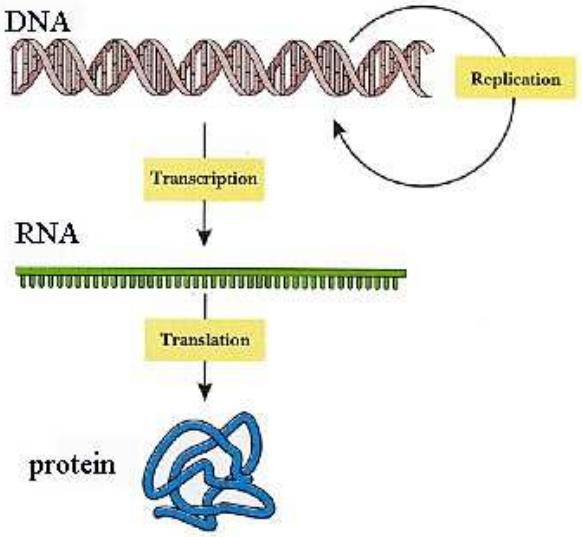


image from the DOE Human Genome Program
http://www.ornl.gov/hgmis

Summary