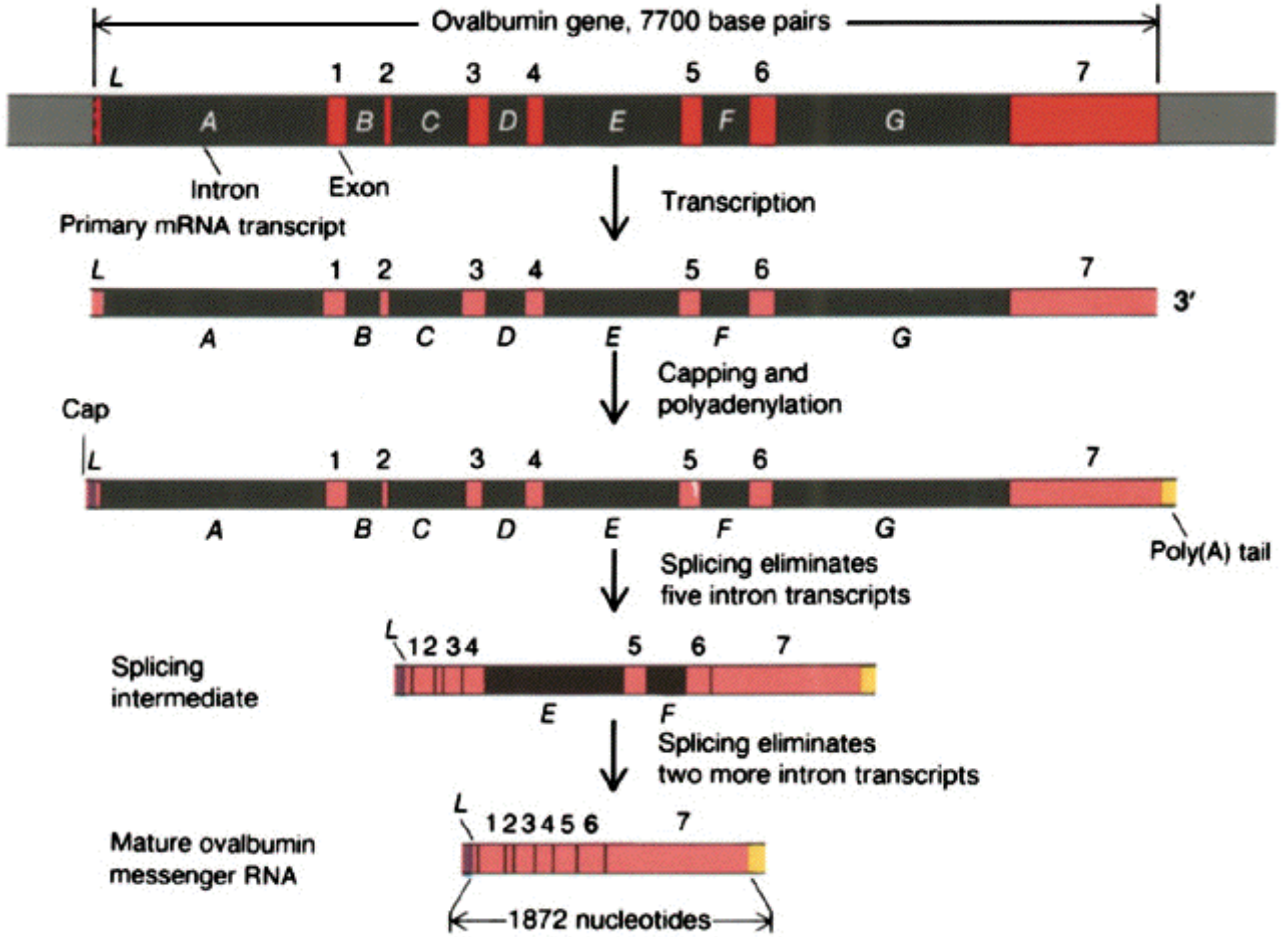
DNA sequences

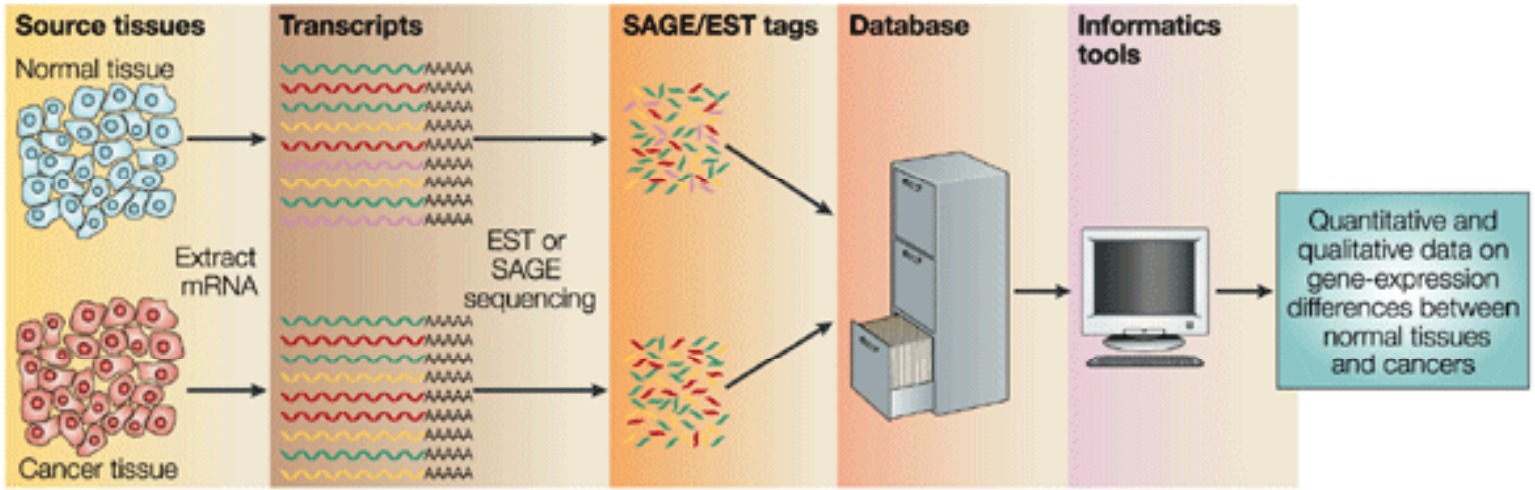
Not all DNA is coding for proteins

- Regulatory regions
- Introns
- Protein-coding region

One protein, many DNA entries

- the primary transcript
- the mature transcript
- the strict protein coding region
- numerous types of partial sequences (ESTs)


Ovalbumin gene, 7700 base pairs
L
1
2
3
4
5
6
7
A
B
C
D
E
F
G
Intron
Exon
Primary mRNA transcript
Transcription
3′
Capping and
polyadenylation
Cap
Poly(A) tail
Splicing eliminates
five intron transcripts
Splicing
intermediate
Splicing eliminates
two more intron transcripts
Mature ovalbumin
messenger RNA
1872 nucleotides



Nature Reviews | Genetics

Retrieving the DNA sequences relevant to my protein

Go to www.expasy.org/sprot/

Enter the accession number of your protein
Click the Cross-Reference button near the top of the form
(different sites concerning the same objects)

See what happens when you use each of them

Click the GenBank link

EMBL, GenBank, DDBJ:
unified DNA sequence databases but in different formats

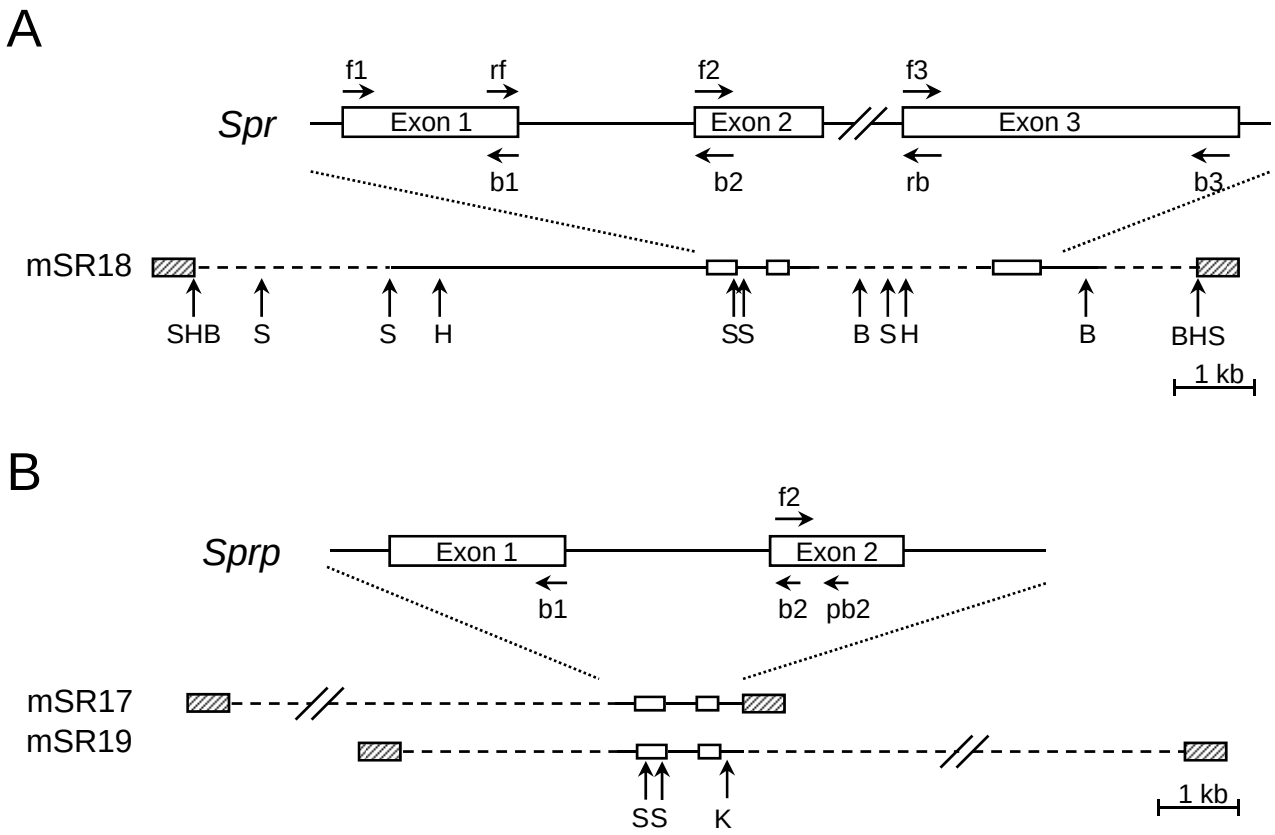
GenBank

Consists of 4 parts

The locus name: write down the accession number
The reference section:
The features section
- promoter elements
- ribosome binding sites (RBS)
- protein coding segments (CDS)
- poly-A site

The sequence section: human mRNA sequence
- actual nucleotide sequence submitted
- save FASTA format of the sequence as a text file

STS: Sequence-tagged site
INSDC: International Nucleotide Sequence Database Collaboration
RefSeq: Reference sequence
MGC: Mammalian gene collection
IMAGE: I.M.A.G.E. Consortium


A
f1
rf
f2
f3
Spr
Exon 1
Exon 2
Exon 3
b1
b2
rb
b3
mSR18
SHB
S
S
H
SS
B
S
H
B
BHS
1 kb
B
f2
Sprp
Exon 1
Exon 2
b1
b2
pb2
mSR17
mSR19
SS
K
1 kb

BLAST (Basic Local Alignment Search Tool)

Compare protein sequence to others

:The first step of sequence analysis
Prediction of protein function
3-D structure
domain organization
identification of homologues

Go to www.ncbi.nlm.nih.gov/BLAST/
Click protein-protein BLAST (blastp)
Copy & paste your protein sequence in the blank window

Change Algorithm parameters

BLAST (Basic Local Alignment Search Tool)

What the resulting page shows?

- sequences
- score
- E value: an assessment of the statistical significance of the score

Click one of the score button

- query sequence
- matching residues
- database sequence

The searched sequences do not show complete ORF residues

Home assignment

DNA sequence of your selected protein
Describe the features of the DNA sequence
Compare the BLAST result with your protein sequences