Replication

DNA must be duplicated in a process called replication before a cell divides. The replication of DNA allows each daughter cell to contain a full complement of chromosomes.

Transcription

For a given gene, only one strand of the DNA serves as the template for transcription. An example is shown below. The bottom (blue) strand in this example is the template strand, which is also called the minus (-) strand, or the sense strand. It is this strand that serves as a template for the mRNA synthesis. The enzyme RNA polymerase synthesizes an mRNA in the 5’ to 3’ direction complementary to this template strand. The opposite DNA strand (red) is called the coding strand, the non-template strand, the plus (+) strand, or the antisense strand.

The easiest way to find the corresponding mRNA sequence (shown in green below) is to read the coding, non-template, plus (+), or antisense strand directly in the 5’ to 3’ direction substituting U for T.

```
5’ T G A C C T T C G A A C G G A T G G A A A G G 3’
3’ A C T G G A A G C T T G C C C T A C C T T T C C 5’

5’ U G A C C U U C G A A C G G A U G G A A A G G 3’
```

As we’ve learned more about the structure of DNA, RNA, and proteins, it become clear that transcription and translation differ in eukaryotes and prokaryotes. Specifically, eukaryotes have intervening sequences of DNA (introns) within a given gene that separating coding fragments of DNA (exons). A primary transcript is made from the DNA, and the introns are sliced out and exons joined in a contiguous stretch to form messenger RNA which leaves the nucleus. Translation occurs in the cytoplasm. Remember, prokaryotes do not have a nucleus.

Translation

If information in a mRNA sequence is decoded to form a protein. In this process a triplet of nucleotides (a codon) in the RNA has the information of a single amino acid. Translation occurs on a large RNA-protein complex called the ribosome. An intermediary transfer RNA (tRNA) molecule becomes covalently linked to a single amino acid by the
enzyme tRNA-acyl synthetase. This "charged" tRNA binds through a complementary anticodon region to the triplet codon in the tRNA. The ribosome/tRNA complex ratchets down the mRNA allowing a new "charged" tRNA complex to bind at an adjacent site. The two adjacent amino acids form a peptide bond in a process driven by ATP cleavage. This process repeats until a "stop" codon appears in the mRNA sequence. The genetic code shows the relationship between the triplet mRNA codon and the amino acid which corresponds to it in the growing peptide chain.

Figure: Codon:Anticodon interactions between mRNA and tRNA

![Codon:Anticodon interactions between mRNA and tRNA](image)

<table>
<thead>
<tr>
<th>2nd base in codon</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st base in codon</td>
<td>U</td>
<td>Leu</td>
<td>Ile</td>
<td>Val</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Leu</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Ser</td>
<td>Pro</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Ser</td>
<td>Pro</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>His</td>
<td>Asp</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>His</td>
<td>Asp</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>STOP</td>
<td>Gly</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STOP</td>
<td>Gly</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STOP</td>
<td>Gly</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Figure: Central Dogma Differences in Eukaryotes and Prokaryotes

![Central Dogma Differences in Eukaryotes and Prokaryotes](image)
As was mentioned in the Protein Chapter (amino acid section) two other amino acids occasionally appear in proteins (excluding amino acids altered through post-translational modification. One is selenocysteine, which is found in Archea, eubacteria, and animals. The other is just recently found is pyrrolysine, found on Archea. These new amino acids derive from modification of Ser-tRNA and probably Lys-tRNA after the tRNA is charged with the normal amino acid, to produce selenocys-tRNA and pyrrolys-tRNA, respectively. The pyrrolys-tRNA recognizes the mRNA codon UAG which is usually a stop codon, while selenocys-tRNA recognizes UGA, also a stop codon. Clearly only a small fraction of stop codons in mRNA sequences would be recognized by this usual tRNA complex. What determines that recognition is unclear.

What is a gene?

The definition of a gene can differ depending on whom you ask. The world gene has literally become a cultural icon of our day. Can our genes explain what it is to be human? The definition of a gene has changed with time.

Eukaryotic genes contain exons (coding regions) and introns (intervening sequencings) that are transcribed to produce a primary transcript. In a post-transcriptional process, introns are spliced out by the splicesome, to produce a messenger RNA, mRNA, which is translated into a protein sequence. (See diagram above).

Over the last 100 years, as our understanding of biochemistry has increased, the definition of a gene has evolved from

- the basis of inheritable traits
- certain regions of chromosomes
- a segment of a chromosome that produces one enzyme
- a segment of a chromosome that produces one protein
- a segment of a chromosome that produces a functional product
The last definition was necessary since some gene products that have function (structural and catalytic) are RNA molecules. The last definition also includes regulatory regions of the chromosome involved in transcriptional control. Snyder and Gerstein have developed five criteria that can be used in gene identification which is important as the complete genomes of organisms are analyzed for genes.

1. identification of an open reading frame (ORF) - this would include a series of codons bounded by start and stop codons. This gets progressively harder to do if the gene has a large number of exons imbedded in long introns.
2. specific DNA features within genes - these would include a bias towards certain codons found in genes or splice sites (to remove intron RNA)
3. comparing putative gene sequences for homology with known genes from different organisms, but avoiding sequences that might be conserved regulatory regions.
4. identification of RNA transcripts or expressed protein (which does not require DNA sequence analysis as the top three steps do) -
5. inactivating (chemically or through specific mutagenesis) a gene product (RNA or protein).

New findings make it even more complicated to define a gene, especially if the transcripts of a "gene region" are studied. Cheng et al studied all transcripts from 10 different human chromosomes and 8 different cell lines. They found a large number of different transcripts, many of which overlapped. Splicing often occur between nonadjacent introns. Transcripts were found from both strands and were from regions containing introns and exons. Other studies found up to 5% of transcripts continued through the end of "gene" into other genes. 63% of the entire mouse genome, which is comprised of only 2% exons, is transcribed.

New Functions for Introns

Recent studies have shown that not all of the DNA interspersed between coding sections of eukaryotic DNA (introns) are devoid of information content. Some of the intron DNA information ultimately gets transcribed and ultimately translated into protein structure. Introns with this property can be classified in two ways:

Group 1 introns: found commonly in yeast, algae, viruses and plant mitochondria and chloroplasts. After they have been spliced out of transcribed RNA, these RNA sequences ultimately get translated into proteins with endonuclease activity. These translated proteins differ from typical protein restriction enzymes which bind and cleave DNA at specific DNA sites of about six base pairs (leading to many cleavages of chromosomal DNA given the high probability of finding such sites in a large genome). Group 1 protein introns recognize specific sequences of 15-40 base pairs, which would be found randomly distributed in the genome only once in about 1 billion base pairs. These Group 1 introns are prime candidates for genetic engineering to produce mutant enzymes which could cut only one specific site in the human genome, allowing insertion of novel therapeutic DNA sequences into a specific location. This would alleviate the problem faced in recent DNA therapy regimes in which a therapeutic DNA is inserted into a virus which inserts randomly into the host DNA, potentially leading to mutagenesis of host DNA. These endonucleases are often called homing endonucleases.

Group 2 Intron also form endonucleases with DNA sequence recognition sites of 30-35 base pairs. Both the endonuclease and its RNA coding sequences are involved in ultimate insertion of intron DNA (carried on a plasmid) into a specific site in host DNA. The RNA, which is easier to engineer than proteins, facilitates specific insertion of intron DNA into host DNA. Such insertions can be used to knockout specific genes with high specificity. A DNA sequence can
be added by linking a gene of interest to the intron DNA of the plasmid.

Another form of novel "junk" DNA involves the transcription and ultimate translation of a gene whose structure predicts should be larger than the actual protein which results. Post-translational modification of proteins through select proteolysis is common. A subset of some mature proteins result from selective removal of an internal piece of the nascent protein (an intein) from the N terminal (N-extein) and C terminal (C-extein) of the protein and subsequent religation of the N-extein and C-extein through peptide bond formation. Intein sequence can even be engineered to connect two different discrete proteins, which is stable at low temperatures, but in which the intein is spliced out at higher temperatures. The splicing of the N and C-exteins occur before the intein is removed. This process occurs during protein folding and requires no enzymes. The biological role, if any, of inteins is not clear. However, they can be used in the laboratory in the production and purification of eukaryotic proteins in bacterial whose expression in the absence of an intein produces a protein that is toxic to the bacterial cell or difficult to purify.

Figure: Mechanism of Intein Splicing

http://www.neb.com/inteins/int_mech.html
Mutations

The figures below show the types of mutations and how point mutations can occur from mismatch pairing, incorporation of base analogs and from chemical mutagenesis.

Figure: Another View of Mutations

Figure: Point Mutations from: Mismatch pairing and Incorporation of Base Analogs
Figure: Point Mutations from Alkylating Agents

3. from chemical mutagens
   - oxidative deamination of xanthine and xanthosine and (HNO₂) or xanthosines
   - alkylating agents
References

11. Creation's 7th day. (new chemistry with more than 20 aa and more than 4 nucleotides). Science. 289, pg 232 (2000)

Contributors

- Prof. Henry Jakubowski (College of St. Benedict/St. John's University)