Bioinformatics CSM17 Week 6: DNA, RNA and Proteins

- Transcription (reading the DNA template)
- Translation (RNA -> protein)
- Protein Structure
Transcription - reading the data

- enzyme - transcriptase
- gene opens up and...
- messenger RNA (for each gene) is formed
- *introns* (non-coding regions) are removed
- mRNA moves off through nuclear pores...
- ...to the ribosomes
- Note:
  - DNA sense strand (mRNA has T->U)
  - and antisense strand (complement of mRNA)
Transcription - reading the data

A T G C A T G C  Sense Strand
(Partner)

T A C G T A C G  Antisense
(Template)
Transcription - reading the data

A T G C A T G C  Sense Strand
(Partner)

T A C G T A C G  Antisense
(Template)
Transcription - reading the data

A T G C A T G C  Sense Strand
(Partner)
| | | | | | | |

A  U  G  C  A  U  G  C  Messenger RNA
| | | | | | | |

T  A  C  G  T  A  C  G  Antisense
(Template)
RNA (Ribose Nucleic Acid)

- **single** chain
- sugar-phosphate backbone
- nucleotide bases (nt)
- Adenine (A)
- Guanine (G)
- Cytosine (C)
- **Uracil (U)** [DNA has Thymine (T)]
- plant viruses, retroviruses (e.g. HIV)
Translation - reading the mRNA

• transfer RNA (tRNA) for each amino acid
  – clover-leaf structure
• codons and anticodons (triplets of bases)
• long chains of amino acids are built up
<table>
<thead>
<tr>
<th>First Position</th>
<th>Second Position</th>
<th>Third Position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td><strong>C</strong></td>
<td><strong>G</strong></td>
</tr>
<tr>
<td>UUU</td>
<td>UCU</td>
<td>UAU</td>
</tr>
<tr>
<td>UUC</td>
<td>UCC</td>
<td>UGU</td>
</tr>
<tr>
<td>UUA</td>
<td>UCA</td>
<td>UGA</td>
</tr>
<tr>
<td>UUG</td>
<td>UCG</td>
<td>UGG</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td><strong>C</strong></td>
<td><strong>G</strong></td>
</tr>
<tr>
<td>AUU</td>
<td>ACU</td>
<td>AGU</td>
</tr>
<tr>
<td>AUC</td>
<td>ACC</td>
<td>AGC</td>
</tr>
<tr>
<td>AUA</td>
<td>ACA</td>
<td>AAA</td>
</tr>
<tr>
<td>AUG</td>
<td>Start</td>
<td>Lys (K)</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td><strong>C</strong></td>
<td><strong>G</strong></td>
</tr>
<tr>
<td>GUU</td>
<td>GCU</td>
<td>GAU</td>
</tr>
<tr>
<td>GUC</td>
<td>GCC</td>
<td>GAC</td>
</tr>
<tr>
<td>GUA</td>
<td>GCA</td>
<td>GAG</td>
</tr>
<tr>
<td>GUG</td>
<td>GCG</td>
<td>Gly (G)</td>
</tr>
</tbody>
</table>

- **U** represents the first position of the codon.
- **C** represents the second position of the codon.
- **G** represents the third position of the codon.

The table lists the corresponding amino acids and stop codons for each combination of the three positions.

- **Phe (F)**: UUU, UUC, UUA, UUG
- **Leu (L)**: CUU, CUC, CUA, CUG
- **Ile (I)**: AUU, AUC, AUA, AUG
- **Val (V)**: GUU, GUC, GUA, GUG
- **Asp (D)**: GAU, GAC, GAA, GAG
- **Glu (E)**: GGU, GGC, GGA, GGG
- **His (H)**: CAU, CAC, CAA, CAG
- **Cys (C)**: UGU, UGC
- **Tyr (Y)**: UAU, UAC
- **Stop**: UAA, UAG
- **Leu (L)**: UUA, UCA, UAA, UGA
- **Pro (P)**: CUA, CCA, CAA, CGA
- **Ser (S)**: UCU, UCC, UCA, UCG
- **Thr (T)**: AUA
- **Gln (Q)**: CGU, CGC
- **Asn (N)**: AAC, AAA
- **Arg (R)**: CGA, CGG
- **Glu (E)**: GUC, GCC, GCA, GCG
- **Glu (E)**: UGG, Trp (W)
- **Start**: AUG
- **Lys (K)**: AGA, AGG
- **Asn (N)**: AAG
- **Ser (S)**: AGU, AGC
- **Arg (R)**: GGA, GGG
Amino Acids

- 20 in total
- building blocks of proteins
Proteins & Polypeptides

- polypeptides are short
- proteins are long
- enzymes are proteins that act as catalysts
- haemoglobin (in blood) is a protein
Protein Structure

• strings of amino acids
• primary (first) structure - AA sequence
• secondary (second) structure - e.g. Spiral
• tertiary (third) structure - complex shape
  – has a large influence on function
Mutation

JYC: CSM17
Mutation

• changes of nucleotide bases
• caused by
  – ionizing radiation, mutagenic chemicals
• usually harmful (damaging)
• may be
  – single base (changing one amino acid)
  – frameshift (more serious)
Sequence Comparison
Simple Alignment

match score: 1    mismatch score 0

A A T C T A T A
A A G A T A    4 + 0 = 4 (best)

A A T C T A T A
A A G A T A    1 + 0 = 1 (worst)

A A T C T A T A
A A G A T A    3 + 0 = 3

JYC: CSM17
Sequence Comparison

Simple Alignment with gap penalties

(match score: 1   mismatch score 0   gap penalty -1)

A A T C T A T A
A A G - A T - A  3 + 0 - 2 = 1  (worst)

A A T C T A T A
A A - G - A T A  5 + 0 - 2 = 3  (equal best)

A A T C T A T A
A A - - G A T A  5 + 0 - 2 = 3  (equal best)

A A T C T A T A
- A A G A T A -  1 + 0 - 2 = -1  (worst)

JYC: CSM17
Sequence Comparison
Simple Alignment with origination and length penalties

*match score*: 1   *mismatch score*: 0
*origination penalty*: -2   *length penalty*: -1

A A T C T A T A
A A - G - A T A  5 + 0 - 4 - 2  = -1  (worst)

A A T C T A T A
A A -- G A T A  5 + 0 - 2 - 2  = 1  (best)

Origination penalty is applied for starting a series of gaps
Length penalty is also applied for each gap

JYC: CSM17
Databases

• NCBI Genbank

• EBI European Bioinformatics Institute
  www.ebi.ac.uk

• EMBL European Molecular Biology Lab.
  www.embl.de/

• ...where you can conduct BLAST – Basic Local Alignment Search Tool – searches
Useful Websites

• GeneStudio Sequence Converter  
  – www.genestudio.com/seqverter.htm

• Mitochondrial maps  
  – www.mitomap.org

• European Molecular Biology Laboratory  
  – www.embl.org

• BioEdit  
References & Bibliography