

## Gene Annotation

### Activity One: When is a Genome Project Finished?

Use the information from the accompanying website to complete the questions on this worksheet.

- Provide two reasons why a 'finished genome' is considered to be an elusive goal for researchers.  
A. \_\_\_\_\_  
B. \_\_\_\_\_
- According to Dr. Stein, why is annotation so important?  
\_\_\_\_\_  
\_\_\_\_\_
- What problem can be solved by teaching a large number of people the basics of how to annotate genes?  
\_\_\_\_\_  
\_\_\_\_\_
- What is the name of the gene prediction method that uses a statistical analysis of the nucleotide base sequence to determine likely locations for genes, including the position of exons and introns?  
\_\_\_\_\_
- What is the name of the gene prediction method that compares the nucleotide base sequence of mRNA transcripts to that of genomic DNA to determine where in the genome the transcript may have originated from?  
\_\_\_\_\_
- Match the terms on the right with the correct descriptions on the left.

a. The ends of a gene that may contain regulatory information	_____	cDNA
b. Sequences created by the reverse transcription of mRNA transcripts	_____	EST
c. A program that searches genomic DNA for alignments	_____	UTR
d. The border between an exon and an intron in a mRNA transcript	_____	Splice Sites
e. Usually the shortest sequences in a community genome database	_____	GeneSeqer
- What is the origin of the sequences represented by a pink map in the Rice Genome Database (OsGDB)?  
\_\_\_\_\_
- If two EST's are generated from the same mRNA transcript, how will this be indicated in the Genome Context View?  
\_\_\_\_\_



Answers:

1. (A)The number of gene structures continues to grow even after a project has been 'completed' and (B) that some gene structures undergo corrections as new information becomes available.
2. "It is the annotation that bridges the gap from the sequence to the biology of the organism."
3. If more people who are working on annotations then it will take less time to compile a set of annotations for a genome sequence.
4. *Ab initio* computation
5. Splice alignment (GeneSeqer)
6. A) UTR B) cDNA C) GeneSeqer D) Splice site E) EST
7. These are existing annotations that are publicly documented in the TIGR data warehouse
8. They will both be contained in a green box on the Evidence Plot