Name _____ Class Period _____ Date _____

Gene Annotation
Activity Four: Evaluating Existing Annotations with GAEVAL
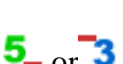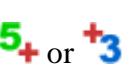
Use the information from the corresponding website to complete the questions on this worksheet.

1. Briefly describe what each flag in indicates.

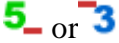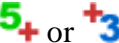| GAEVAL Flag	Description of the problem between the existing annotation and the EST, cDNA alignment data.
⊠	
◕	
◑	
◙	
●	
⇥	
⇥	
5 or 3	
5+ or +3	
AS	

2. Which flags would you use to describe the following? Suggest a way to improve the annotation.
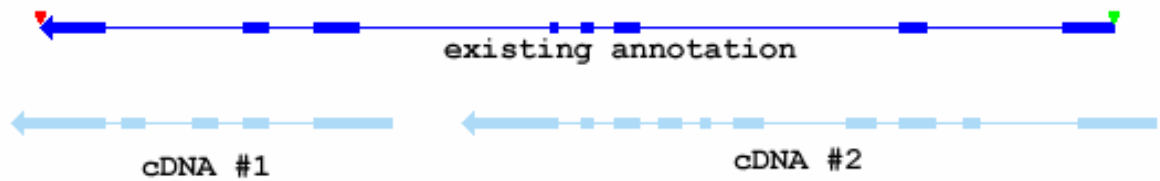
existing annotation

cDNA #1   cDNA #2

_____

_____

_____

_____

GAEVAL Flag	Description of the problem between the existing annotation and the EST, cDNA alignment data.
⊠	no known EST of cDNA have been assigned to a predicted gene by GeneSeqer
◕	expressed sequence exists in the region of the gene annotation, however no splice sites were confirmed
◐	at least one splice site is confirmed
◕	all splice sites are aligned, however there are some gaps in the coverage between splice sites
●	all splice sites are aligned and also that the annotation is completely covered
⇶	incorrectly split a single gene into multiple adjacent gene annotations
⇶	incorrectly merge multiple adjacent gene structures into a single gene annotation
5− or **−3**	the 5' or 3' end of the annotation does not have splice alignment support and should there fore be shortened to match the evidence
5+ or **+3**	the 5' or 3' end of the annotation should be extended (added) to reflect the cDNA evidence
⊸AS⊸	a second cDNA or EST contains a region that supports an alternative gene model with a different number of exons

2. ◐**5**+**+3**⇶ There are several exons that are present in the cDNA that are not present in the existing annotation. Also there are two separate cDNA, suggesting two gene models (fission or the existing annotation) and both ends of the existing annotation need to be extended to match the cDNA evidence.

1. The Evidence Plot and the Evidence Table
2. The Evidence Plot
3. The entire structure (all of the exons) are added to the (green) gene model
4. First click and it is selected (not the other exons on the same structure) on the second click it is deselected.
5. The coordinates of the exon correspond to nucleotide bases in the genome, basically defining where the exon starts and stops in the sequence a chromosome
6. The Max Sim Score and the number of cDNA and EST structures that contain the exons in question
7. A color block represents a groups of variants for the same exon, so only one should be selected for each gene model
8. Exons