Week 7 Lab: Pathways Parts A-C Worksheet

Name:

Instructions: A copy of this worksheet will be randomly selected from your group at the end of Week 7's lab. Answer questions completely and clearly, fully indicating your reasoning/ thoughts (often writing in complete sentences is helpful). I expect the responses to <u>have no typos and minimal grammatical error</u>. You are allowed/encouraged to discuss questions with your labmates.

A1. Fill out the following information. Note that gene symbols and species names are italicized if typed, and underlined if hand-written. FB indicates information you gather from FlyBase.

Target Gene Symbol		**RefSeq locus (FB)**	
Target Species		**Team Name (come up with one with your group)**	
Target Gene Name in *D. melanogaster* (FB)		**Names of your classmates on the team**	

A2. Briefly describe the function of your assigned gene in the insulin signaling pathway using the information you read on FB (do your best to write in your own words, and if necessary to copy, use quotes!).

A3. What is the scaffold your gene is on in *D. melanogaster*?

A4. Save the image of the genomic neighborhood of your target gene in *D. melanogaster* from the GEP UCSC Genome Browser including both nearest two upstream and two downstream genes and nested/nesting gene(s) (if present) in your lab notebook and/or a folder on your laptop. Title your file "TeamName_GeneSymbol_Dmel_GenomicNeighborhood.png". Have one groupmate per research project upload the image file with the appropriate filename onto Box (using BB's 'Assignment Upload' link in left sidebar).

A5. Sketch the genomic neighborhood of your target gene in *D. melanogaster* (using the example in step A9). Be sure your sketch includes the gene symbols of the surrounding genes, indicates their orientation, and the direction of transcription. Also, note which genes are upstream, and which are downstream. See example figure from protocol.

A6. Does your target gene have multiple isoforms in *D. melanogaster*? If so, how many and which is the longest isoform?
Note: Be sure to use the longest isoform for your tblastn against the entire genome of your target species.

B1. What is the number of amino acids in the translated protein?

B2. What is the assembly version you are using? (Look at the column titled 'Genome Browsers' under the Pathways Project Genome Assemblies webpage, http://tiny.cc/genome_assemblies).

B3. Save the screenshot of the "Descriptions" panel of your *tblastn* results of the amino acid sequence of the *D. melanogaster* protein coding isoform for your target gene against the genome assembly in your target species in your lab notebook and/or a folder on your laptop. Title your file "TeamName_GeneSymbol_TargetSpecies_tblastn.png". Have a teammate who hasn't done so already upload the image file with the appropriate filename onto Box (using BB's 'Assignment Upload' link in left sidebar).

B4. Summarize your *tblastn* search results. What scaffold (i.e. name the accession number) do you think your target gene is on in the target species? Is there only one very good match, or is there some ambiguity about where your assigned gene is located in the target species? Use data in your screenshot from B3 to help justify your answer.

B5. After sorting the *tblastn* search results for your <u>best</u> match by "Subject start position", fill in Table 1 below to help you identify the best collinear set of alignments to the protein of your target gene in the target species' genome.

TABLE 1: **Summary of the *tblastn* search results for the best scaffold match**

| Range | D. melanogaster | | Target Species | | E-Value | Identities (%) | Subject Frame |
	Query Start	Query End	Subject Start	Subject End			
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

Note: If you have >10 ranges, ask your instructor or TA for a longer table. If you have more than one very good match in your *tblastn* search, also ask for an additional table.

B6. Summarize your findings from Part B by recording the following information for your target gene in the target species:

- **Scaffold:**

- **Scaffold accession number:**

- **Coordinates spanning region:**

B7. Write a hypothesis on where the orthologous region to your target gene is located in your target species (based on the information from B6).

B8. Explain briefly the evidence that supports your hypothesis that your coordinates are where in the target species the ortholog of your target gene is located.
Note: Remember to be specific, cite the related table, and if needed, describe discrepancies to keep an eye on.

C1. Record the RefSeq accession number (i.e. beginning with XM) for your putative ortholog.

C2. Save the image of the genomic neighborhood of your target gene in the <u>target species</u> from the GEP UCSC Genome Browser including both nearest two upstream and two downstream genes in your lab notebook and/or a folder on your laptop. Title the file "TeamName_GeneSymbol_TargetSpecies_GenomicNeighborhood.png". Have one groupmate per research project upload the image with the appropriate filename onto Box. If someone has not submitted one yet, have them do the upload.

C3. Fill in Table 2 below to summarize your results for Parts A-C. Any entry that should not have any data should be marked with n/a.

TABLE 2: *blastp* search results for the protein sequences of the genomic neighborhood of the target gene in the target species against the *D. melanogaster* reference protein database (refseq_protein)

From Part A examination of *D. melanogaster assembly*		2nd Closest Upstream	Closest Upstream	Nested Gene	Target Gene	Closest Downstream	2nd Closest Downstream
a. Dmel	Gene Symbol						
	Strand (+/-)						
From Part C comparison of NCBI transcripts to *D. melanogaster* proteins		2nd Closest Upstream	Closest Upstream	Nested Gene	Target Gene	Closest Downstream	2nd Closest Downstream
b. Target Species	NCBI RefSeq Gene (mRNA) Accession						
	NCBI RefSeq Protein Accession						
	Strand (+/-)						
c. Best *blastp* Result	Accession						
	D. melanogaster Gene Symbol						
	E-Value						
	Percent Identity						
d. Are the genes in the two species orthologs? (yes/no)							

C4. Sketch the genomic neighborhood of your target gene in the target species following example in step C20. Be sure your sketch includes the gene symbols of the surrounding genes, indicates their orientation, and the direction of transcription. Also, note which genes are upstream, and which are downstream.

C5. Compare your sketches of the genomic neighborhoods of your target gene in both *D. melanogaster* (Part A) and the target species (Part C). You may find it useful to draw your gene models side by side on the board or a separate sheet of paper. Note if your target gene is on different strands, you may find it useful to re-orient the *D. melanogaster* genomic neighborhood as if it were on the opposite strand. Are there any differences between the two genomic neighborhoods? If so, explain.