Pathways Project: Annotation Notebook Tables

Katie Sandlin and Alexa Sawa

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			

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)

		2nd Closest Upstream	Closest Upstream	Nested[1] Gene	Target Gene	Closest Downstream	2nd Closest Downstream
D. melanogaster	Gene Symbol						
	Strand (+/-)						
Target Species	NCBI RefSeq Gene (mRNA) Accession						
	NCBI RefSeq Protein Accession						
	Strand (+/-)						
Best *blastp* Result	Accession						
	D. melanogaster Gene Symbol[2]						
	E-Value						
	Percent Identity						
Are the genes in the two species orthologs? (yes/no)							

[1] Leave column blank if target gene is not nested within another gene or another gene is not nested within your target gene.

[2] If the *D. melanogaster* gene symbol of the best match is not provided in the "Description" column of the *blastp* results, you can identify it by clicking on the Accession Number (listed in the "Accession" column). In the new internet browser window that opens, scroll to the "FEATURES" section near the bottom. Next to the "CDS" feature you'll see the gene symbol listed after "/gene=".

TABLE 3: Isoforms with unique coding sequences in *D. melanogaster*	
Unique isoform(s) based on coding sequence	**Other isoforms with identical coding sequences**

TABLE 4: Summary of *tblastn* CDS-by-CDS search results

CDS	FlyBase ID	Query Length Size (aa)	*D. melanogaster*		Target Species		Subject Frame
			Query Start	Query End	Subject Start	Subject End	

TABLE 5: Gene Model for [insert isoform name here] in target species

CDS	FlyBase ID	Frame	Splice Acceptor Phase	Coordinates		Splice Donor Phase
				Start	End	
			0			