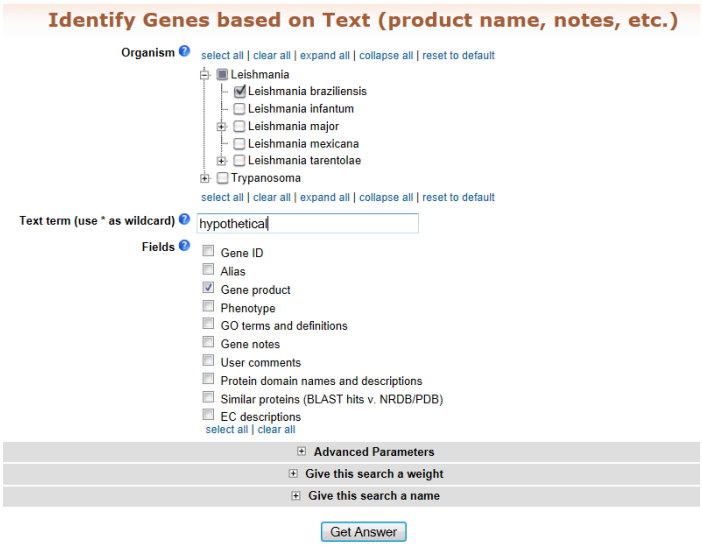
Motif Searches and Regular Expressions
Exercise 5

5.1 Using InterPro domain searches to identify unannotated kinesin motor proteins.
For this exercise use http://tritrypdb.org

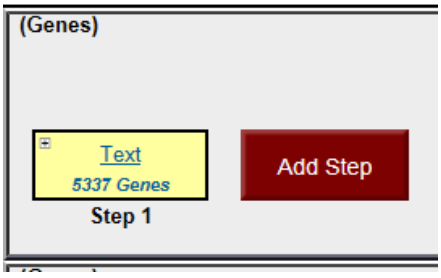
a. Identify all genes annotated as hypothetical in *L. braziliensis.*

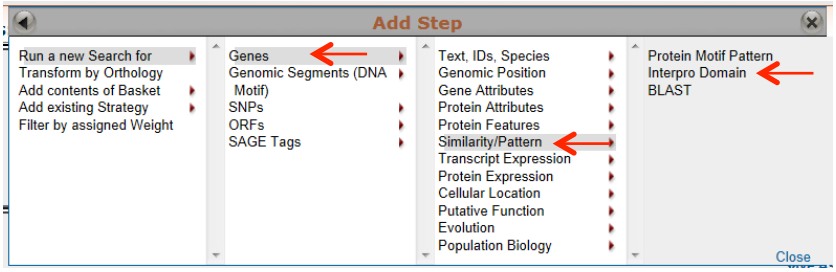
Hint: use the full text search and look for genes with the word “hypothetical” in their product names.

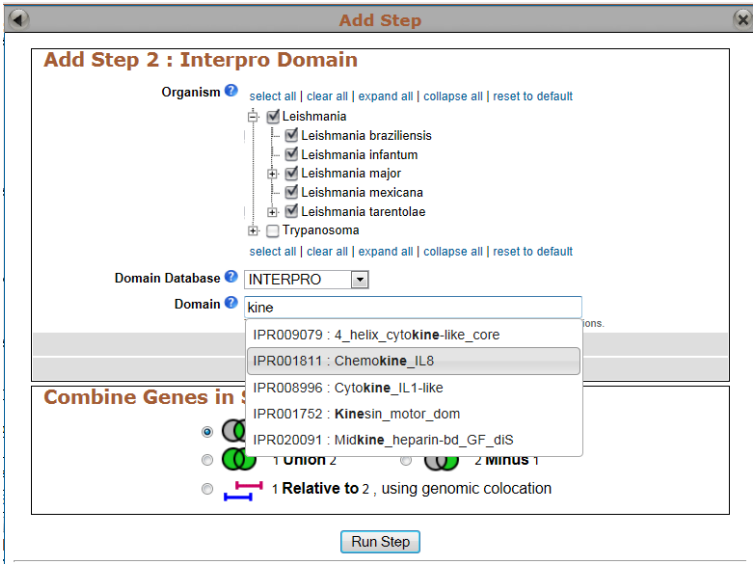


b. How many of these hypothetical genes have a kinesin-motor protein InterPro domain?

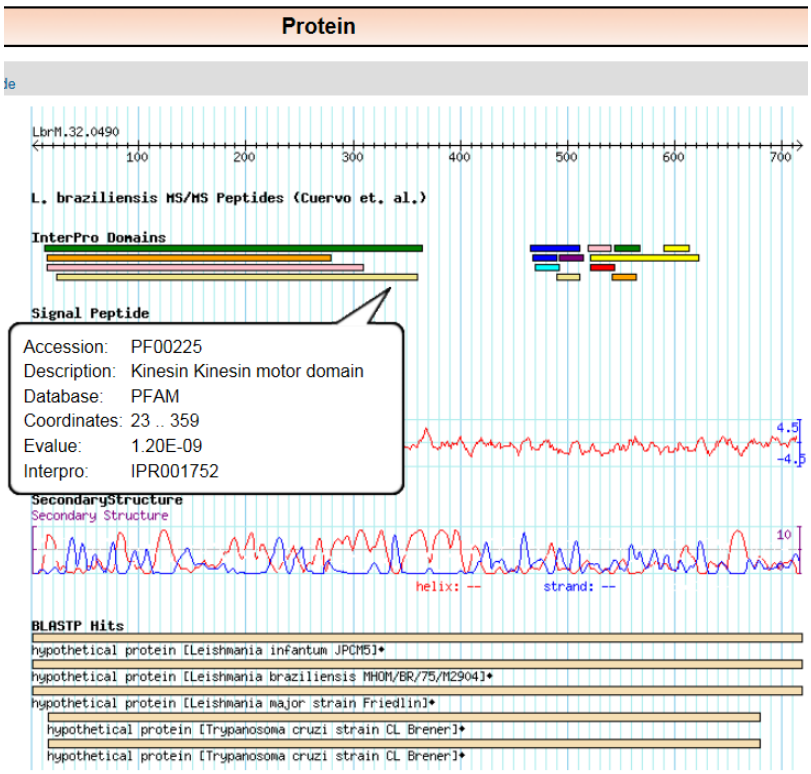
Hint: add a step to the strategy. Go to the “Interpro Domain” search under similarity/pattern, start typing the work kinesin and it should autocomplete.







c. Go to the gene page for LbrM.32.0490 and look at the protein feature section. Does this look like a possible motor protein?



5.2 Using regular expressions to find motifs in TriTypDB: finding active trans-sialidases in *T. cruzi.*

a. *T. cruzi* has an expanded family of trans-sialidases. In fact, if you run a text search for any gene with the word "trans-sialidase", you return over 1400 genes!!! Try this and see what you get.

b. However, not all of these are predicted to be active. It is known that active trans-sialidases have a signature tyrosine (Y) at position 342 in their amino acid sequence. Add a motif search step to the text search in 'a' to identify only the active trans-sialidases.
   - Hint: for your regular expression, remember that you want the first amino acid to be a methionine, followed by 340 of any amino acid, followed by a tyrosine 'Y'. Refer to regular expression tutorial if you need to.

If you need help, you can go to this sample strategy below to see the answer:
http://tritrypdb.org/tritrypdb/im.do?s=a905e36f634f7b42