Instructions:

your goal will be to refactor a program so that it more modular and easier to maintain.

four input files are:

* [blastp.outfmt6](https://canvas.instructure.com/courses/1688943/files/83877424/download?wrap=1)  
  BLAST output file (format is 6, meaning it's tab-separated)
* [diffExpr.P1e-3\_C2.matrix](https://canvas.instructure.com/courses/1688943/files/83877433/download?wrap=1)  
  Trinity output of differentially expressed genes under 4 stress conditions
* [gene\_association\_subset.gaf](https://canvas.instructure.com/courses/1688943/files/83877427/download?wrap=1)  
  GO annotations file
* [go-basic.obo](https://canvas.instructure.com/courses/1688943/files/83877431/download?wrap=1)  
  GO terms

**Part 1: Refactor**

Refactor the program so that nothing is of global scope, and most importantly, break down that single chunk of code into smaller, simpler functions.  Feel free to refactor any code that is overly complex and/or is serving no purpose.

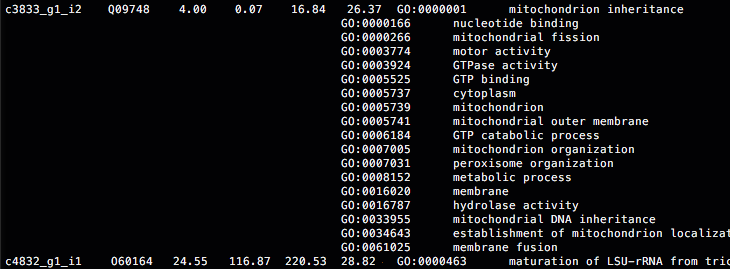
For every function created, remember to give it a docstring.  You may use any current comments in the program as part of your docstrings and give your functions useful meaningful names.

Utilize main() and if \_\_name\_\_ == "\_\_main\_\_" so that the script can also be used as a module.

**Part 2: Update**

Now that the program is modularized, it should be easier to implement the following changes:

* a transcript can have more than one possible predicted match - for now, we are only going to consider the best matches.  Update the code so that the transcripts are mapped to their *first* corresponding BLAST hit with identity over 99 - I encourage you to look at the input file at this point so that you understand why "first" is emphasized.
* currently, the program is mapping protein to its GO annotations in a 1:1 fashion, but a protein can be annotated with more than one GO ID.  Update the mapping so that all (unique) GO annotations are included in the mapping.
* change the format of the output file so that only new information is printed per line, that is, the format should look something like this (GO terms have been sorted):



Similar to before, keep the values in their respective columns, e.g. when opening the file in Excel, all the GO IDs should line up in the same column.

By the end, your output should match with [this](https://canvas.instructure.com/courses/1688943/files/84016688/download?wrap=1).

**Part 3: Test**

Once you are done making all the updates needed, create a testing script for the program.  Try to think of as many test cases as you can and implement them into your testing script.  The more test cases you can provide, the more robust your program will most likely be.