#Supplementary text 1 – Python script for listing disease-associated genes

#Requires installation of Python v2.6 or later, and Biopython v1.55 or later

#Also requires an internet connection, or local version of the NCBI gene database

import sys # import system functions from python

from Bio import Entrez # import Entrez search functions from biopython

Entrez.email = "wconrad@u.washington.edu" # Always tell NCBI who you are when using esearch

disease = sys.argv[1] # The user enters their disease term of interest

term1=disease+'[All Fields] AND (alive[PROP] AND "Homo sapiens"[Organism])'

# This is the term that entrez gene will search

handle = Entrez.esearch(db="gene", term=term1, retmax=800) # the biopython esearch for the above term is stored as the variable “handle”

record = Entrez.read(handle) #the record created by the term search is stored as "record"

hitlist = record["IdList"] #the hitlist variable collects all of the geneIDs from the original search

newfile = open(disease+"\_genes.txt", "w") # a new tab-delimited text file is created that uses the search query as the title and also the word genes.

for id in hitlist:

newfile.write(disease+"\t"+"lit"+"\t"+str(id)+"\n") # this loop enters the query, the designation lit for 'literature interaction' and the geneID as as string.

newfile.close()

#print len(hitlist) #python displays the number of hits