#!/usr/bin/perl

# Clint Valentine

# 10/18/2016

use warnings;

use strict;

use diagnostics;

my $blastxFile = "./blastx.outfmt6";

my $edgeFiles = "./edgeR.2013.dir/\*DE\_results";

my %IDHash = ();

unless ( open INFILE, "<", $blastxFile ) {

die "Can't open $blastxFile for reading $!";

}

# Blast Output Format 6 is a tab delimited file with fields:

# Query Label | Target Label | Percent Identiy | Alignment Length |

# Number of Mismatches | Number of Gap Opens | Start Position in Query |

# End Position in Query | Start Position in Target | End Position in Target |

# E-value (Karlin-Altschul Statistics) | Bit Score (Karlin-Altschul Statistics)

while ( <INFILE> ) {

chomp;

# Split line on whitespace and get transcriptID

my @entry = split ' ', $\_;

my $transcriptID = $entry[0];

# Split Target Label on '|' and get swissprotID

my @targetLabel = split '\|', $entry[1];

my $swissprotID = $targetLabel[3];

# If we've seen this transcriptID before, do nothing, otherwise

# add it to the IDhash with the swissprotID as the value

$IDHash{ $transcriptID } = $swissprotID unless

exists $IDHash{ $transcriptID };

}

close INFILE;

for my $edgeFile ( glob $edgeFiles) {

unless ( open EDGEFILE, "<", $edgeFile ) {

die "Can't open $edgeFile for reading $!";

}

# Open an output file [EDGEFILE].annotated in same dir as [EDGEFILE]

open OUTFILE, ">", $edgeFile . '.annotated' or

die "Can't open", $edgeFile . '.annotated', "for reading $!";

# Consume header metadata on first line

my $header = readline(EDGEFILE);

# Write out new header line

# swissprotID | transcriptID | logFC | logCPM | PValue | FDR

print OUTFILE join "\t", "swissprotID", "transcriptID", $header;

while ( <EDGEFILE> ) {

chomp;

# Split line on whitespace and get transcriptID in first location

my @entry = split ' ', $\_;

my $transcriptID = $entry[0];

# Lookup transcriptID's corresponding swissprotID, else, 'None'

my $swissprotID = 'None';

if ( $IDHash{ $transcriptID } ) {

$swissprotID = $IDHash{ $transcriptID };

}

print OUTFILE join "\t", $swissprotID, @entry, "\n";

}

close EDGEFILE;

close OUTFILE;

}