**ANNOTATION**

The annotation of the assembled transcriptome was done using Trinotate pipeline.

**Trinotate:**

Trinotate is a comprehensive annotation suite designed for automatic functional annotation of transcriptomes, particularly de novo assembled transcriptomes, from model or non-model organisms. Trinotate makes use of a number of different well referenced methods for functional annotation including homology search to known sequence data (BLAST+/SwissProt), protein domain identification (HMMER/PFAM), protein signal peptide and transmembrane domain prediction (signalP/tmHMM), and leveraging various annotation databases (eggNOG/GO/Kegg databases). All functional annotation data derived from the analysis of transcripts is integrated into a SQLite database which allows fast efficient searching for terms with specific qualities related to a desired scientific hypothesis or a means to create a whole annotation report for a transcriptome.

Trinotate includes [TrinotateWeb](http://trinotate.github.io/TrinotateWeb.html), which provides a locally-driven web-based graphical interface for navigating transcriptome annotations and analyzing transcript expression and differential expression using the Trinity/RSEM/Bioconductor analysis framework. (<http://trinotate.github.io/>)

**First the programs required for the pipelines were installed. And tested in small test data.**

Programs doanloaded and available at: /mnt/scratch/adam013/programs

Program downloaded from:

Trinotate >  <http://trinotate.github.io>

Trinity > <http://trinityrnaseq.github.io>

TransDecoder > <http://transdecoder.github.io>

Sqlite > <http://www.sqlite.org/>

NCBI BLAST+ > <http://www.ncbi.nlm.nih.gov/books/NBK52640/>

HMMER/PFAM ><http://hmmer.janelia.org/download.html>

**Optional but recommended**

signalP v4 > <http://www.cbs.dtu.dk/cgi-bin/nph-sw_request?signalp>

tmhmm v2 > <http://www.cbs.dtu.dk/cgi-bin/nph-sw_request?tmhmm>

RNAMMER > <http://www.cbs.dtu.dk/cgi-bin/sw_request?rnammer>

**Second, Sequence Databases Required:**

Trinotate **relies heavily on SwissProt and Pfam**, and custom protein files are generated as described below to be specifically used with Trinotate. You can obtain the protein database files by running this Trinotate build process. This step will download several data resources including the latest version of swissprot, pfam, and other companion resources, create and populate a Trinotate boilerplate sqlite database (Trinotate.sqlite), and yield *uniprot\_sprot.pep* file to be used with BLAST, and the *Pfam-A.hmm.gz* file to be used for Pfam searches. Run the build process like so:

$TRINOTATE\_HOME/admin/Build\_Trinotate\_Boilerplate\_SQLite\_db.pl Trinotate

and once it completes, it will provide to you:

Trinotate.sqlite

uniprot\_sprot.pep

Pfam-A.hmm.gz

Prepare the protein database for blast searches by:

makeblastdb -in uniprot\_sprot.pep -dbtype prot

Uncompress and prepare the Pfam database for use with hmmscan like so:

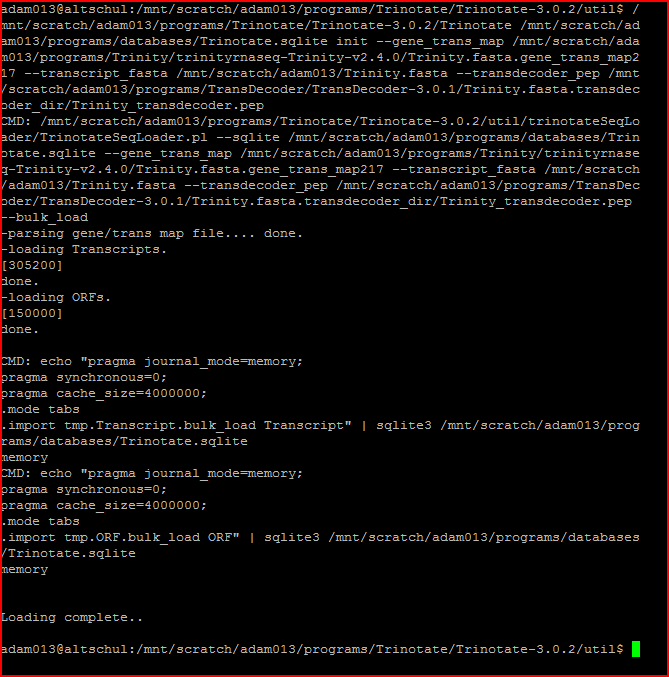
gunzip Pfam-A.hmm.gz

hmmpress Pfam-A.hmm

Command for loading transcripts and coding regions: >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite init --gene\_trans\_map /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Trinity.fasta.gene\_trans\_map217 --transcript\_fasta /mnt/scratch/adam013/Trinity.fasta --transdecoder\_pep /mnt/scratch/adam013/programs/TransDecoder/TransDecoder-3.0.1/Trinity.fasta.transdecoder\_dir/Trinity\_transdecoder.pep

Output of the command:



Loading BLAST homologies:

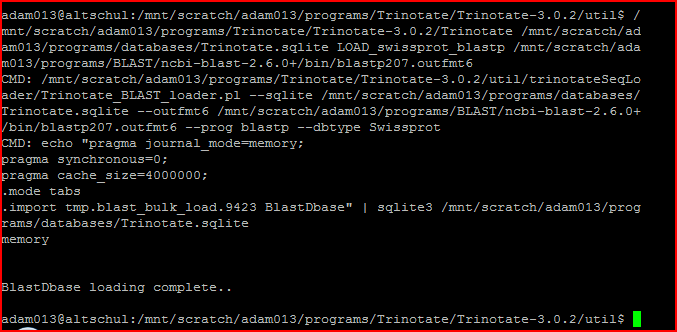
# load protein hits

Trinotate Trinotate.sqlite LOAD\_swissprot\_blastp blastp.outfmt6

Command: >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite LOAD\_swissprot\_blastp /mnt/scratch/adam013/programs/BLAST/ncbi-blast-2.6.0+/bin/blastp207.outfmt6

Command output:



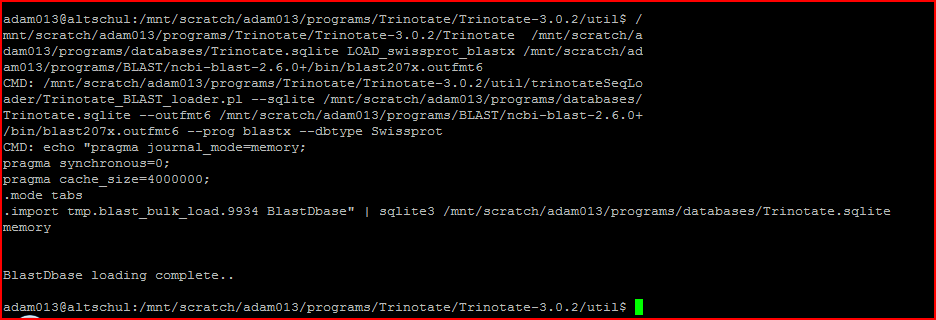
# load transcript hits

Trinotate Trinotate.sqlite LOAD\_swissprot\_blastx blastx.outfmt6

Command : >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite LOAD\_swissprot\_blastx /mnt/scratch/adam013/programs/BLAST/ncbi-blast-2.6.0+/bin/blast207x.outfmt6

Command output :



Optional: load custom database blast hits:

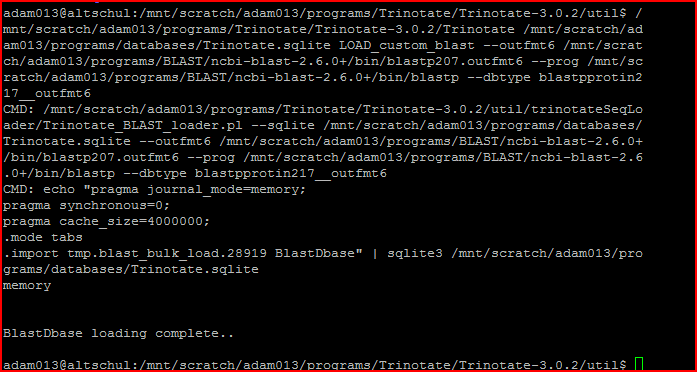
# load protein hits

Trinotate Trinotate.sqlite LOAD\_custom\_blast --outfmt6 custom\_db.blastp.outfmt6 --prog blastp --dbtype custom\_db\_name

Command : >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite LOAD\_custom\_blast --outfmt6 /mnt/scratch/adam013/programs/BLAST/ncbi-blast-2.6.0+/bin/blastp207.outfmt6 --prog /mnt/scratch/adam013/programs/BLAST/ncbi-blast-2.6.0+/bin/blastp --dbtype blastpprotin217\_\_outfmt6

Command output :



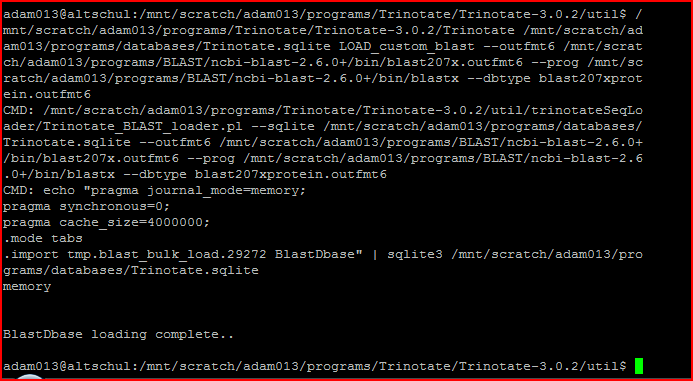
# load transcript hits

Trinotate Trinotate.sqlite LOAD\_custom\_blast --outfmt6 custom\_db.blastx.outfmt6 --prog blastx --dbtype custom\_db\_name

Command : >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite LOAD\_custom\_blast --outfmt6 /mnt/scratch/adam013/programs/BLAST/ncbi-blast-2.6.0+/bin/blast207x.outfmt6 --prog /mnt/scratch/adam013/programs/BLAST/ncbi-blast-2.6.0+/bin/blastx --dbtype blast207xprotein.outfmt6

Command output :



**Load Pfam domain entries**

Trinotate Trinotate.sqlite LOAD\_pfam TrinotatePFAM.out

Command : >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite LOAD\_pfam /mnt/scratch/adam013/programs/Hammer-pfam/TrinotatePFAM227.out

Command output :



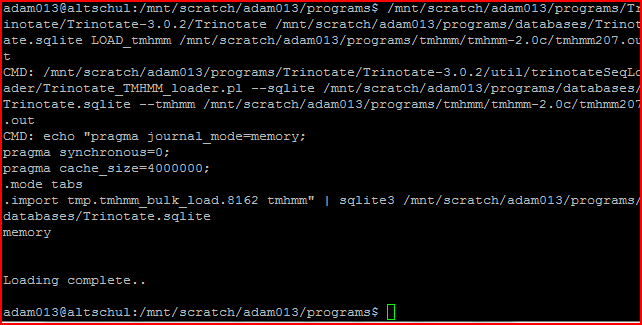
**Load transmembrane domains**

Trinotate Trinotate.sqlite LOAD\_tmhmm tmhmm.out

Command : >>

/mnt/scratch/adam013/programs$ /mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite LOAD\_tmhmm /mnt/scratch/adam013/programs/tmhmm/tmhmm-2.0c/tmhmm207.out

Command output :



**Load signal peptide predictions**

Trinotate Trinotate.sqlite LOAD\_signalp signalp.out

Command : >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/Trinotate /mnt/scratch/adam013/programs/databases/Trinotate.sqlite LOAD\_signalp /mnt/scratch/adam013/programs/signalp/signalp/signalp-4.1/signalp227.out

Command output :



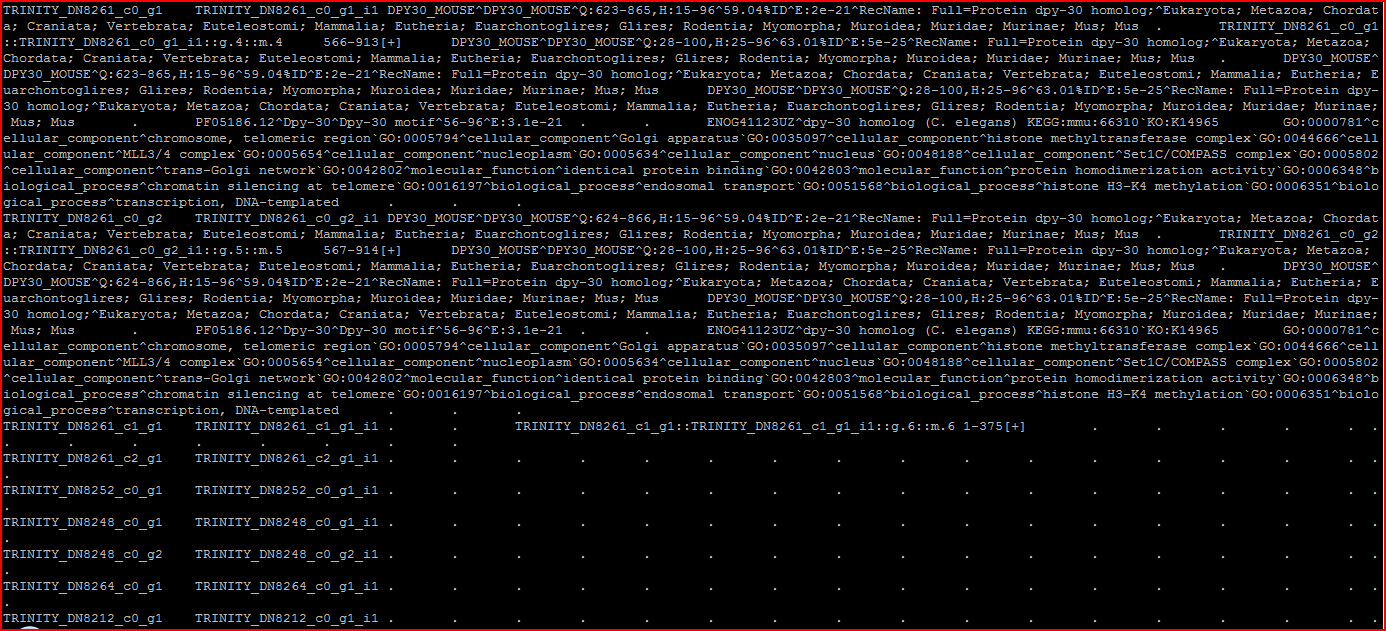
# Trinotate: Output an Annotation Report

Trinotate Trinotate.sqlite report [opts] > trinotate\_annotation\_report.xls

Command : >>

/mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/util$ h ead trinotate\_annotation\_report237.xls

Command output :



0 #gene\_id

1 transcript\_id

2 sprot\_Top\_BLASTX\_hit

3 RNAMMER

4 prot\_id

5 prot\_coords

6 sprot\_Top\_BLASTP\_hit

7 custom\_pombe\_pep\_BLASTX

8 custom\_pombe\_pep\_BLASTP

9 Pfam

10 SignalP

11 TmHMM

12 eggnog

13 Kegg

14 gene\_ontology\_blast

15 gene\_ontology\_pfam

16 transcript

17 peptide

GOseq:

Download the package in gz format.

To create a gene lengths file, first create a file containing the transcript lengths:

% ${TRINITY\_HOME}/util/misc/fasta\_seq\_length.pl Trinity.fasta > Trinity.fasta.seq\_lens

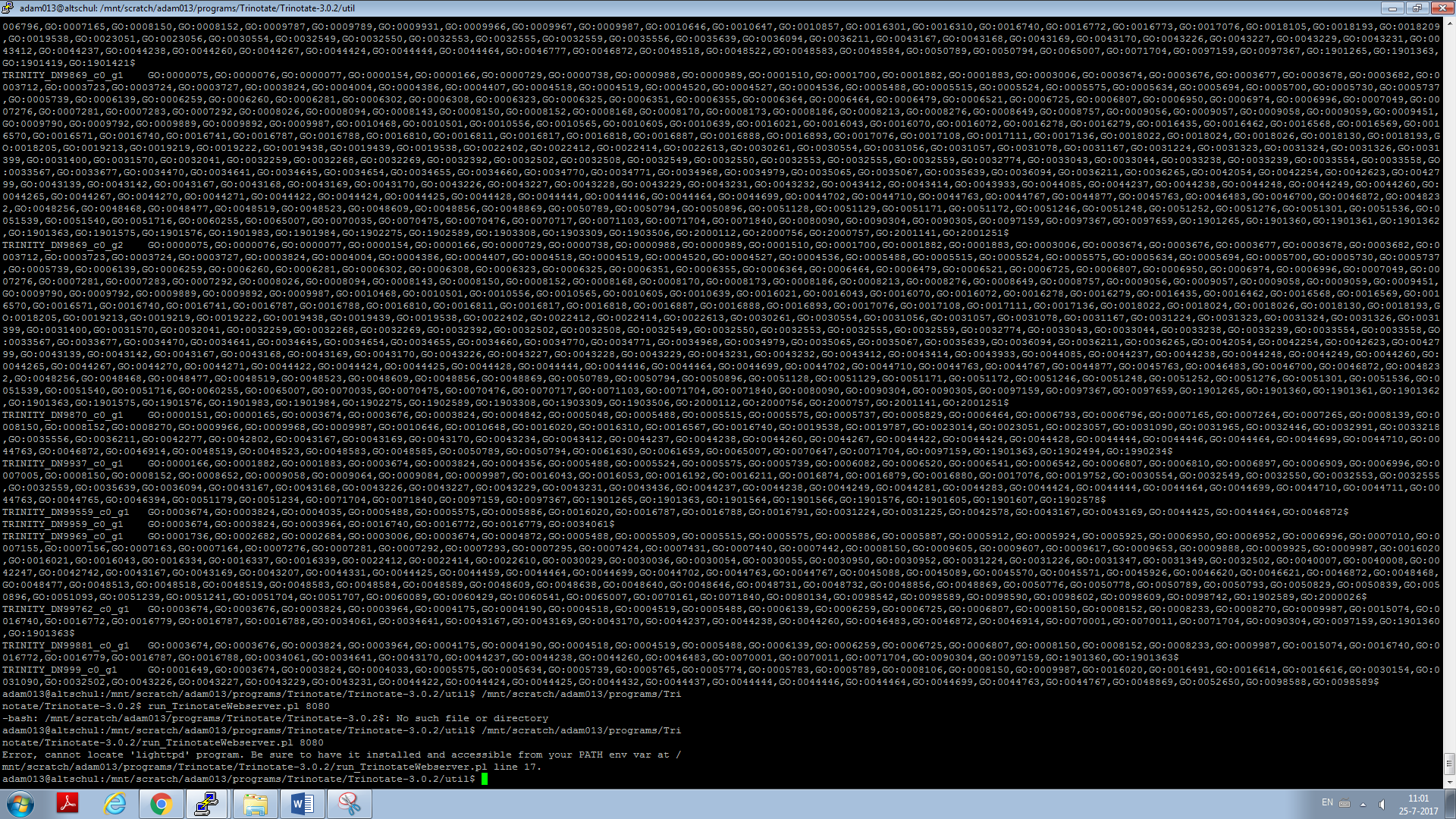
and then use that file when creating the gene lengths file:

% ${TRINITY\_HOME}/util/misc/TPM\_weighted\_gene\_length.py \

--gene\_trans\_map trinity\_out\_dir/Trinity.fasta.gene\_trans\_map \

--trans\_lengths Trinity.fasta.seq\_lens \

--TPM\_matrix isoforms.TMM.EXPR.matrix > Trinity.gene\_lengths.txt



**Go enrichment analysis:**

**Steps:**

1. Extract all GO assignments for each gene feature

${TRINOTATE\_HOME}/util/extract\_GO\_assignments\_from\_Trinotate\_xls.pl \

--Trinotate\_xls trinotate.xls \

-G --include\_ancestral\_terms \

> go\_annotations.txt

1. Create 'factor\_labeling.txt' file

factor (tab) gene\_id >>> have being created using Rscript and imported with filzella

1. Create a file containing the transcript lengths

% ${TRINITY\_HOME}/util/misc/fasta\_seq\_length.pl Trinity.fasta > Trinity.fasta.seq\_lens

1. Create the gene lengths file

% ${TRINITY\_HOME}/util/misc/TPM\_weighted\_gene\_length.py \

--gene\_trans\_map trinity\_out\_dir/Trinity.fasta.gene\_trans\_map \

--trans\_lengths Trinity.fasta.seq\_lens \

--TPM\_matrix isoforms.TMM.EXPR.matrix > Trinity.gene\_lengths.txt

1. Run GOseq

${TRINITY\_HOME}/Analysis/DifferentialExpression/run\_GOseq.pl \

--factor\_labeling factor\_labeling.txt \

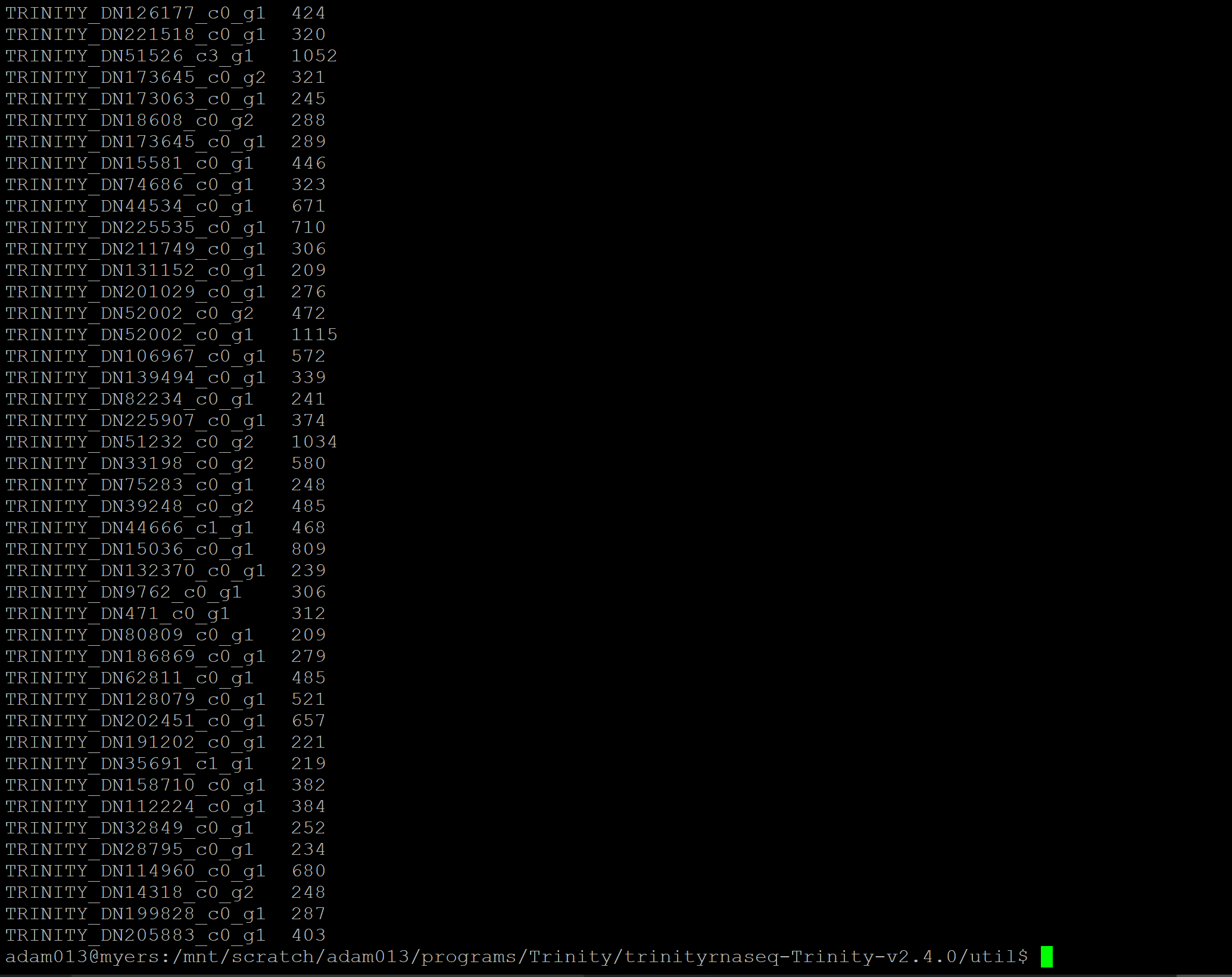
--GO\_assignments go\_annotations.txt \

--lengths gene.lengths.txt

**Script for performing gene length file >>>>**

/mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/TPM\_weighted\_gene\_length.py --gene\_trans\_map /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Trinity.fasta.gene\_trans\_map217 --trans\_lengths /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/Trinity.fasta.seq\_247lens --TPM\_matrix /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/trans\_count.TMM.EXPR.matrix > Trinity.gene\_lengths298.txt

**Done**



**GO enrichment analysis for result of control vs infected sampls**

ctrl\_L\_vs\_trig\_L

/mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/run\_GOseq.pl --factor\_labeling /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/factor\_label\_newyear/ctrl\_L\_vs\_trig\_L\_unique.txt --GO\_assignments /mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/util/go\_annotations247.txt --lengths /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/Trinity.gene\_lengths298.txt --background /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/background\_gene\_list.txt

ctrl\_D\_vs\_trig\_D

/mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/run\_GOseq.pl --factor\_labeling /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/factor\_label\_newyear/ctrl\_D\_vs\_trig\_unique.txt --GO\_assignments /mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/util/go\_annotations247.txt --lengths /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/Trinity.gene\_lengths298.txt --background /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/background\_gene\_list.txt

ctrl\_L\_vs\_climb\_L

/mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/run\_GOseq.pl --factor\_labeling /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/factor\_label\_newyear/ctrl\_L\_vs\_climb\_L\_unique.txt --GO\_assignments /mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/util/go\_annotations247.txt --lengths /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/Trinity.gene\_lengths298.txt --background /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/background\_gene\_list.txt

ctrl\_D\_vs\_climb\_D

/mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/run\_GOseq.pl --factor\_labeling /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/factor\_label\_newyear/ctrl\_D\_vs\_climb\_D\_unique.txt --GO\_assignments /mnt/scratch/adam013/programs/Trinotate/Trinotate-3.0.2/util/go\_annotations247.txt --lengths /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/util/misc/Trinity.gene\_lengths298.txt --background /mnt/scratch/adam013/programs/Trinity/trinityrnaseq-Trinity-v2.4.0/Analysis/DifferentialExpression/background\_gene\_list.txt