Database preparation

This document describes the creation of a concatenated target-decoy database for use in proteomics data analysis. This walkthrough assumes the use of the command line throughout.

Required software

1. Download SearchGUI from http://compomics.github.io/projects/searchgui.html if not already done.

Protocol

1. Navigate to the directory where you would like to store your database.
2. Download the appropriate database using wget.

```
wget ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/Eukaryota/UP000005640_9606.fasta.gz
```

3. Extract the database.

```
gunzip UP000005640_9606.fasta.gz
```

4. Rename the database.

```
mv UP000005640_9606.fasta ./uniprot_human_mar2018_fwd.fasta
```

5. Add contaminants to the database. We typically modify the identifiers of the contaminant proteins using Notepad to contain the text "CONT" so we can identify them later. In addiiton, we remove all human proteins from the contaminant database.

```
wget ftp://ftp.thegpm.org/fasta/cRAP/crap.fasta

cat uniprot_human_mar2018_fwd.fasta crap.fasta > uniprot_human-crap_mar2018_fwd.fasta
```

6. Prepare the decoy database. Note, I am using my own directories here which you should change the appropriate directories for your specific system.

```
java -cp /projects/ptx_analysis/chughes/software/software_searchGUI/SearchGUI-3.2
.20/SearchGUI-3.2.20.jar eu.isas.searchgui.cmd.FastaCLI -in /projects/ptx_analysi
s/chughes/databases/mar2018/uniprot_human-crap_mar2018_fwd.fasta -decoy
```

Note the number of forward, reverse, and total entries for later use in publications. Your database is now ready for use.