**From:** Krista Bean
**To:** Mikhail Dozmorov, PhD (Mikhail-Dozmorov@omrf.org)
**Subject:** MYSQL Query Script
**Date:** Wednesday, February 20, 2013 11:53:07 AM
**Attachments:** genes.txt
refgene.pl
trimmer.pl

Mikhail,

Attached are the perl scripts I put together to run the mysql query.

The command line arguments to run the pipeline are:

perl refgene.pl genes.txt hg19 | sort –k2,2 –k 3,3n | perl trimmer.pl | uniq –u | (bedtools commands)

The first script reads in the list of genes (a sample is attached) and the database to query, and then performs the query, pulling out all of the data matching any of the gene names in the text file. It also makes sure that the gene is on the accepted list of chromosomes (which is programmed into the file). This output is directed to the sort command to get the data sorted by chromosome and start position for the bedtools. The next perl script takes that output and trims it up to create a clean bed file (basically in the first script I had to parse out the chromosome from the chr that is returned and use that to sort. This perl file removes the chromosome number from the file so it is back in true bed format). Finally, the uniq command removes all duplicate lines from the file. Because of the way I did the query, there were duplicate entries returned, but this command removes them. You can then pipe this output to the bedtool commands that you need.

To get this to run, you must have perl and the DBI and DBD::mysql perl packages installed. If you are working on a windows machine, I would recommend installing active perl. You can then go to the start menu, and under the active perl folder is a perl package manager. From there you can easily install DBI and DBD::mysql. If you are running a linux environment, you can use cpan from the command line to install these packages. (To install DBD::mysql on the linux environment you must first have mysql installed).

I didn't really have a test file to check against to make sure the mysql output returned was correct, but feel confident in it. The mysql query should be working right and the rest of the pipeline is good. But if you run into problems, let me know and I'll correct them.

If you need help getting these scripts to run on your machine or have any questions, just let me know!

Thanks,
Krista