

## GenomeLite manual

2011-02-03

Jiyuan An, John Lai, Melanie Lehman, Colleen Nelson: Australian Prostate Cancer Research Center (APCRC-Q) and Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology (QUT), Brisbane, Australia

[j.an@qut.edu.au](mailto:j.an@qut.edu.au)

+61 7 3176 3075

Web page: <http://www.aus-canprostatealliance.org/genomelite>

To cite your use of GenomeLite in your publication:

*Jiyuan An, John Lai, Melanie L Lehman, Colleen Nelson. GenomeLite: A Standalone Genome Browser. Bioinformatics xx, xx-xx (just finish 1<sup>st</sup> draft)*

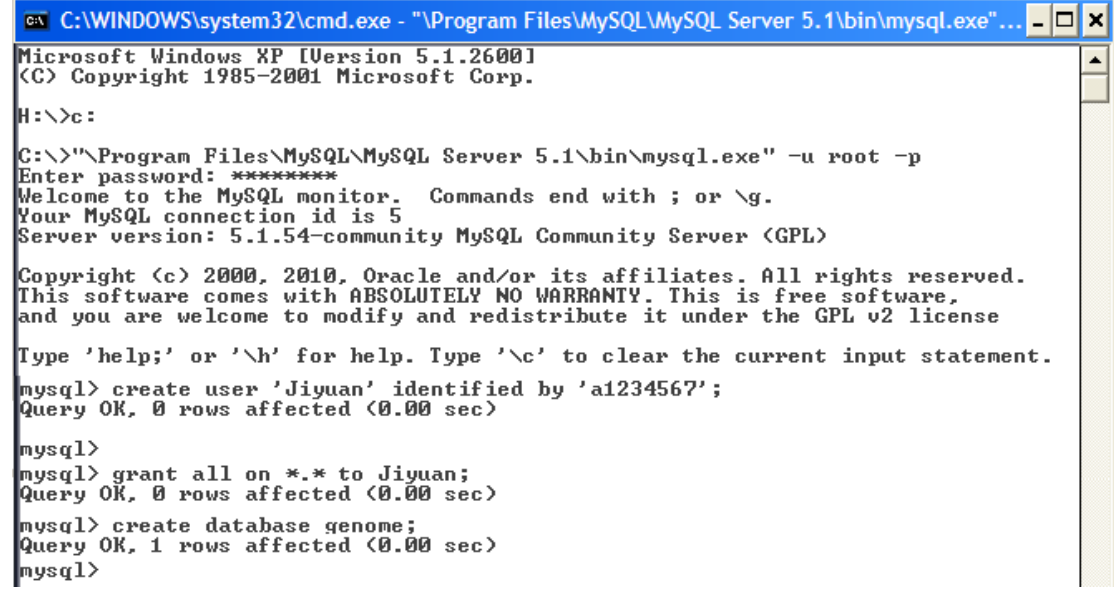
## 1. Installation

### 1.1. Create database

GenomeLite accesses a MySQL database called “genome”. If MySQL server is not installed on the computer, it can be downloaded from: <http://dev.mysql.com/downloads/mysql/>. In the MySQL command line client window, create a database called “genome”:

```
MySQL> create database genome;
```

Figure 1 shows how to create account and a database called genome in windows OS.



```
C:\WINDOWS\system32\cmd.exe - "Program Files\MySQL\MySQL Server 5.1\bin\mysql.exe"...
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

H:\>c:

C:\>"Program Files\MySQL\MySQL Server 5.1\bin\mysql.exe" -u root -p
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 5
Server version: 5.1.54-community MySQL Community Server (GPL)

Copyright (c) 2000, 2010, Oracle and/or its affiliates. All rights reserved.
This software comes with ABSOLUTELY NO WARRANTY. This is free software,
and you are welcome to modify and redistribute it under the GPL v2 license

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> create user 'Jiyuan' identified by 'a1234567';
Query OK, 0 rows affected (0.00 sec)

mysql>
mysql> grant all on *.* to Jiyuan;
Query OK, 0 rows affected (0.00 sec)

mysql> create database genome;
Query OK, 1 rows affected (0.00 sec)

mysql>
```

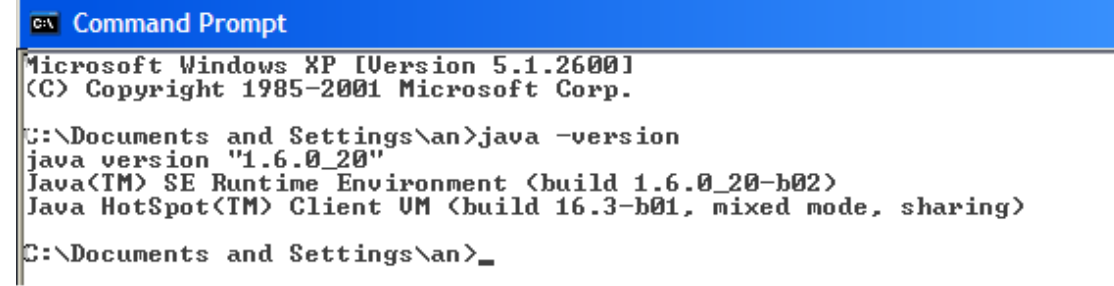
Figure 1 create MySQL database

GenomeLite will add all uploaded tables into the genome database.

If you want to use external MySQL server to show data with other users, it is not necessary to install MySQL in your machine. You also need to create a database called genome in that MySQL server; all the genome browser track data will be stored in the MySQL server. The IP address of the database server will be indicated in track.xml file.

### 1.2. Install JDK

Download JDK 6 (version  $\geq 1.6.0_{19}$ ) to run GenomeLite from <http://jdk6.java.net/>. After installing JDK, the java command will run in a command prompt (Windows) as shown in Figure 2 or any other terminal (Linux). If the version information is not available from the command prompt, then check whether the java path is in the environment variable.



```
C:\> Command Prompt
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\an>java -version
java version "1.6.0_20"
Java(TM) SE Runtime Environment (build 1.6.0_20-b02)
Java HotSpot(TM) Client VM (build 16.3-b01, mixed mode, sharing)

C:\Documents and Settings\an>_
```

Figure 2 - Check whether JDK has been installed in windows.

### 1.3. Download files

GenomeLite related files are zipped in URL <http://www.aus-canprostatealliance.org/genomelite>. The zip file will need to be downloaded into a directory e.g. C:\genomeLite (Windows), or /home/xxx/genomeLite (Linux). Once the file is unzipped, a new directory “genomeLite” will appear. The directory contains four files: track.xml, genomeLite.bat, housekeepingTrack.sql, and genomeLite.jar, and two subdirectories: lib, demo\_data.

To use GenomeLite to get a piece of DNA sequence and RNA secondary structure (the functions for Button “DNA” and “RNA” in genome browser window) for a given chromosome region, download genome sequence zip file hg18.zip and unzip it in your working directory. It will generate a folder called hg18. Under this folder, there will be a file for each of the 24 chromosome binary sequences.

**Track.xml** contains the database server and track information. When tracks are added or deleted in genomeLite, the contents of track.xml will be change accordingly. In the track.xml file, there will be tags, <DB> and </DB>, which are used to set the database server where you created a database called “genome”. If the database on your local computer is used, then the contents should be changed as noted below:

```
<IP>localhost</IP>
<account>yourAccount</account>
<password>yourPassword</password>
```

If you are using an external computer, the IP address should be inserted in the first line in place of “localhost”. If the human hg18 assembly is not being used, the genome in the tag <genome> can be replaced with other genome such as hg19, mm9.

```
<genome>
    <assembly>hg18</assembly>
</genome>
```

**genomeLite.bat** is a batch file, which runs in a windows environment. A shortcut can be created on the desktop to allow GenomeLite to run by simply clicking the shortcut. In Linux, you type below to start GenomeLite.

```
>java -Xmx1024m -jar GenomeLite.jar
```

**genomeLite.jar** stores java classes of GenomeLite it is the main file to run this tool.

**housekeepingTrack.sql** is the backup of housekeeping track tables: hg18 assembly refgene, aceview and sno-miR. The next Section will explain how to restore housekeeping track tables.

**lib directory** is to keep MySQL and Java connector; in this version 5.1.15 is used.

**demo\_data directory** provides several demo data files,; this includes a bed file, a wiggle file, and a combinedwiggle file.

## 2. Start GenomeLite

MySQL and JDK should now be installed on the computer. The GenomeLite files should also have been downloaded and unzipped

### 2.1. Pre-processing

**Firstly**, the DB sever should be correctly set in the track.xml file. The default track.xml file has been set to the localhost database server. The account and password will need to be changed. If tracks are being shared with other people, then the remote database server should be set with an IP address, account, and password.

**Secondly**, housekeeping tracks for human genome hg18 need to be created. In the GenomeLite directory, you can find a file called housekeepingTrack.sql, which packed refgene\_hg18, acrvview\_hg18 and sno\_mirna\_hg18 track DB tables into sql file. You can find a video demo in manual directory.

**Error! Reference source not found.**Figure 3 shows how to create housekeep tracks in a database called “genome” in windows. In Linux or MacOS platform, you can input the following command in a terminal:

```
>mysql -u root -p < housekeepingTeacks.sql
```

We will get three house keeping track data tables, refseq hg18, aceview\_hg18 and sno\_mirna\_hg18 in the database genome.

```

C:\WINDOWS\system32\cmd.exe
C:\>cd GenomeLite
C:\GenomeLite>dir
Volume in drive C is SYSTEM
Volume Serial Number is 229C-5FCE

Directory of C:\GenomeLite

07/04/2011  11:44 AM  <DIR>          .
07/04/2011  11:44 AM  <DIR>          ..
31/03/2011  02:09 PM  <DIR>          demo_data
28/03/2011  10:22 PM                36 GenomeLite.bat
30/03/2011  12:06 PM            132,896 GenomeLite.jar
31/03/2011  02:09 PM  <DIR>          hg18
07/04/2011  11:03 AM            84,952,264 housekeepingTracks.sql
31/03/2011  02:09 PM  <DIR>          lib
30/03/2011  09:47 PM            963 track.xml
               4 File(s)      85,086,159 bytes
               5 Dir(s)    10,163,208,192 bytes free

C:\GenomeLite>"\Program Files\MySQL\MySQL Server 5.1\bin\mysql.exe" -uroot -p <
housekeepingTracks.sql
Enter password: *****
C:\GenomeLite>

```

Figure 3 create hg18 refseq housekeeping track data

#### Add table into track.xml

If you have created all three transcript tables, you need to add them into track.xml:

```

<?xml version="1.0"?>
<tracks>
  <DB>
    <IP>localhost or IP address</IP>
    <account>XXX</account>
    <password>XXX</password>
  </DB>
  <genome>
    <assembly>hg18</assembly>
  </genome>
  <browser_size>
    <width>600</width>
    <height>500</height>
  </browser_size>
  <locus>
    <chr>chr19</chr>
    <loci_start>56100543</loci_start>
    <loci_stop>56104930</loci_stop>
  </locus>
  <track>
    <DB_table_name>refgene_hg18_housekeeping </DB_table_name>
    <visible>full</visible>
  </track>
  <track>
    <DB_table_name>aceview_hg18_housekeeping </DB_table_name>
    <visible>full</visible>
  </track>
  <track>
    <DB_table_name>sno_mirna_hg18_housekeeping </DB_table_name>
    <visible>full</visible>
  </track>
</tracks>

```

</tracks>

## Summary

At this point, a database server IP address, account and password in a track.xml file has been set. A directory to keep the refgene, aceview or sno/miR tracks should also now be created. In Windows, double click on the genomeLite.bat file to start GenomeLite, although it is recommended that a shortcut on the Desktop is created for the bat file. In Linux, type:

```
>java -jar -Xms1024m -Xmx1024m GenomeLite.jar
```

GenomeLite has three tabs: one for a genome browser window, a second for a customize window and a third for a add DB table.

## 2.2. Genome browser

After GenomeLite is started, three house-keeping tracks (as indicated in track.xml) are shown in the Genome browser (Figure 4). In the genome browser, most buttons are functionally similar to that of the UCSC genome browser. (1) One track will be highlighted if the mouse pointer moves over on the left margin of the track. (2) Moving across the genome is operated by pressing the buttons “<<<”, “<<”, “<”, “>”, “>>”, “>>>”. “<” and “>” move the browser left or right by 10% of the genomic region currently displayed. “<<” and “>>” moves the browser by 50% of the displayed region, and “<<<”, “>>>” moves the browser by 100% of the displayed region. There are also 3X and 10X zoom in and out buttons. (3) The tracks can be rearranged in genome browser window. To reorder tracks, click and hold the left side of tracks and drag the highlighted track to a new position. The demo video is available in manual.

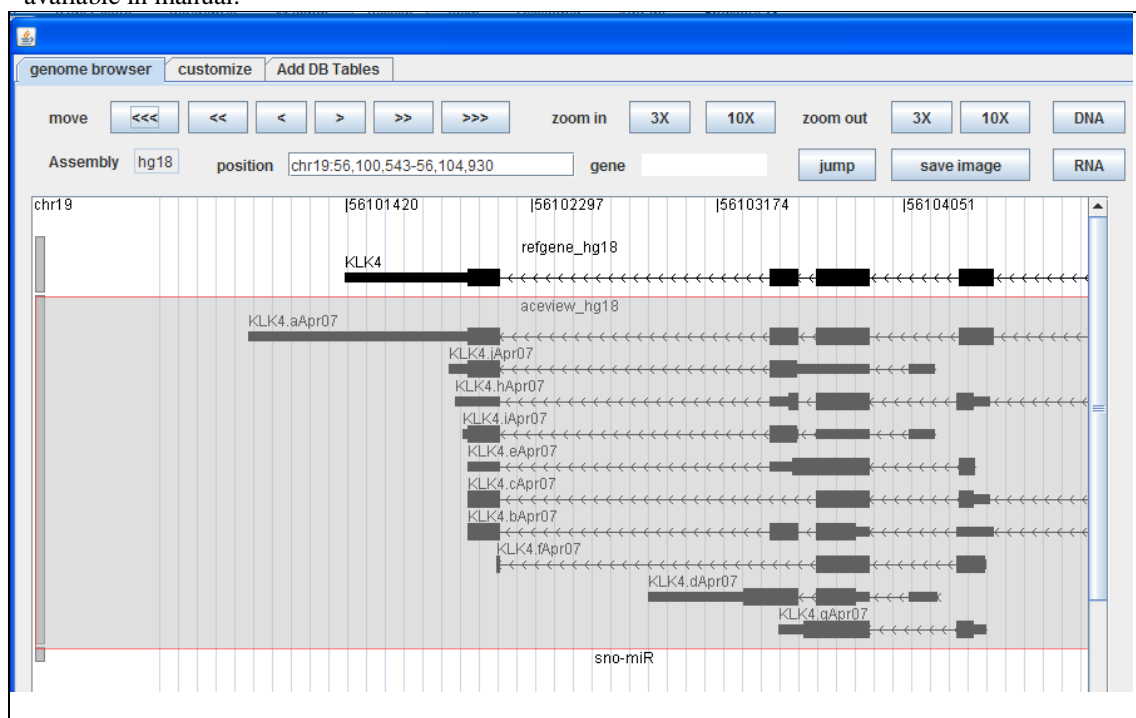


Figure 4 The view of the genome browser in GenomeLite

The most useful function for “zoom in” is to place the mouse pointer at the top of the canvas of the genome browser, click and hold the left mouse button, then drag the pointer to your locus of interest prior to releasing the left mouse button. This will enlarge the area of interest into the full canvas as shown in Figure 5. The highlighted region on the left is shown as the selected full-sized region on the right. (3) A gene or locus can also be inputted in the “position” and “gene” textboxes. By clicking jump the desired locus is represented in full-size on the genome canvas. (4) The button “save image” is used to save the image of the displayed canvas as a png file in your working directory.

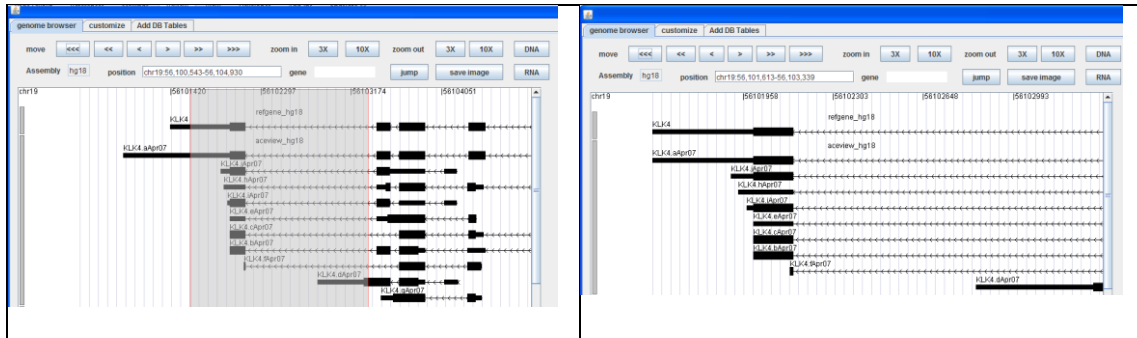


Figure 5 zoom in dragging mouse

GenomeLite also provides functions to obtain DNA sequence and RNA secondary structure in genome browser window. On the up-right corner, DNA and RNA buttons are for the two functions. The result sequence will be shown in a pop up window Figure 6. Secondary structure is represented with brackets as Vienna RNAfold application tool [2]. If you get error saying “assembly hg18 sequence is not available”, you need to download hg18.zip file from GenomeLite webpage as describe in section 1.3.

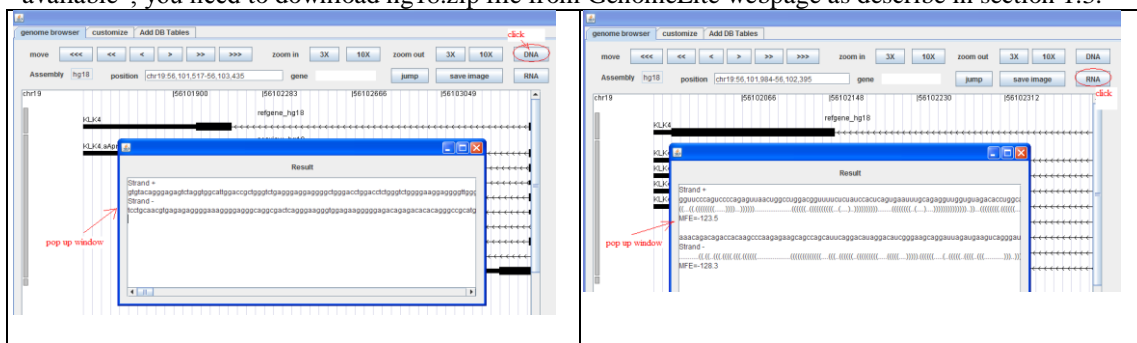


Figure 6 DNA and RNA secondary structure

## 2.3. Add custom tracks

GenomeLite provides four types of custom track inputs: BED, Wiggle, CombinedWiggle and housekeeping track. Below are non-experimental examples for each of these three types of tracks. The demo video called is available in manual.

### 2.3.1. BED track

**BED track:** The UCSC genome browser BED format is adopted. 6, 9, and 12 column formats are supported in GenomeLite. In the demo data, the BED track example is below:

```
chr19 56101200 56101300 cloneA 960 + 56101200 56101300 255,0,0
chr19 56101500 56101670 cloneB 960 + 56101500 56101670 0,0,255
```

The definition of the columns is the same as the UCSC genome browser:

**chrom** - The name of the chromosome (e.g. chr7, chrX, chr2\_random).

**chromStart** - The starting locus of this item. The first base in a chromosome is numbered 0.

**chromEnd** - The ending locus of this item. The chromEnd base is not included in the display of the feature. For example, the first 100 bases of a chromosome is defined as chromStart=0, chromEnd=100, and spans the bases numbered 0-99.

**name** - The name of this item. This label is displayed to the top of the BED line in the Genome Browser canvas. When the mouse moves over the item, the name appears in several seconds.

**score** - This is not used in the GenomeLite tool.

**strand** - Defines the strand as either '+' or '-'.

**thickStart** - The starting position at which the feature is drawn thickly (for example, the start exon in the gene display).

**thickEnd** - The ending position at which the feature is drawn thickly (for example, the last exon in the gene display).

**itemRgb** - An RGB value of the form Red, Green, Blue (e.g. red color: 255,0,0).

**blockCount** - The number of exons in the BED line.

**blockSizes** - A comma-separated list of the exon sizes. The number of items in this list should be the same as blockCount.

**blockStarts** - A comma-separated list of block starts. All of the blockStart positions should be calculated relative to chromStart. The number of items in this list should be the same as blockCount.

Figure 7 shows how a BED custom track can be added.

1. Select the third tabbed pane “Add DB Tables”
2. Select “bed” in track type
3. Give the track a name such as “demo\_bed”
4. Copy/ paste two lines in the demo\_bed file, or click “file browser” to select the data file.
5. Click “generate custom track”

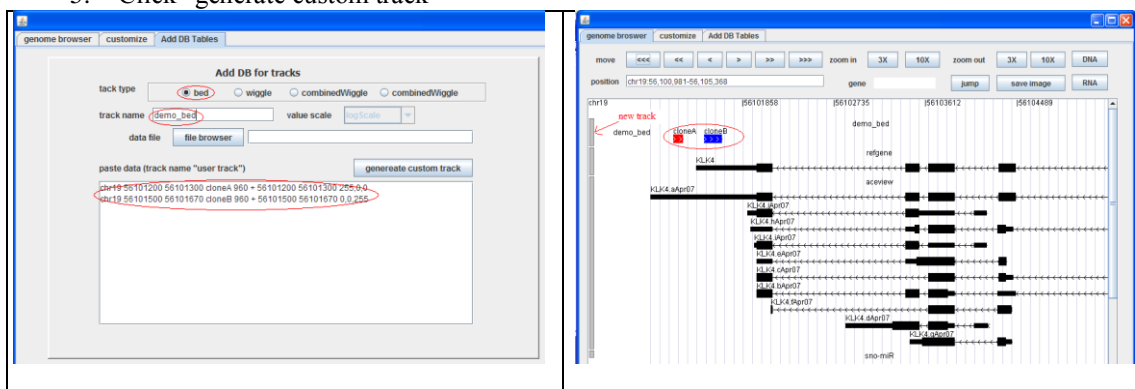


Figure 7 Add BED custom track

The view of the Genome Browser canvas is shown on the right of Figure 7. The new track `demo_bed` has been added on top of the tracks. The tracks can be replaced by dragging the mouse pointer over the left margin of the genome browser canvas as shown in Figure 8; the `demo_bed` track is moved from the top to the second top order.

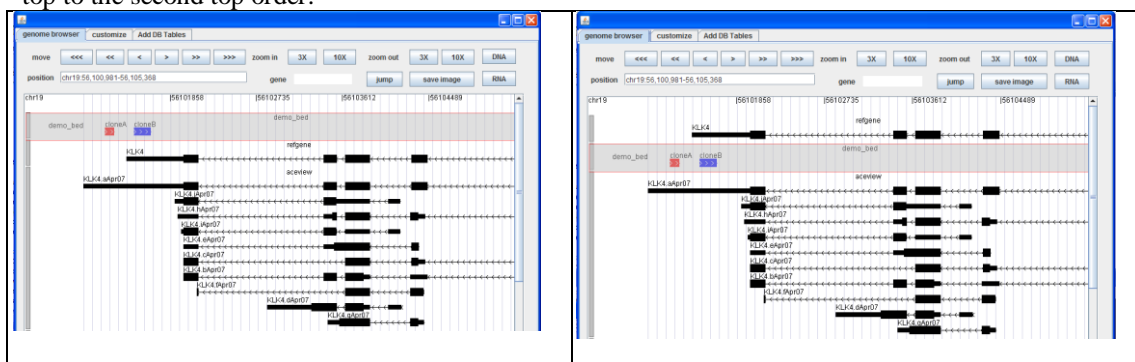


Figure 8 replace tracks in genome browser canvas

### 2.3.2. Wiggle track

**Wiggle track:** We extended the UCSC genome browser wiggle by putting a colour column for each item.

*name1* chr19 + 56101300 56101350 10 244,164,96  
*name2* chr19 + 56101350 56101380 20 244,164,96

The number of columns is extended to 7:

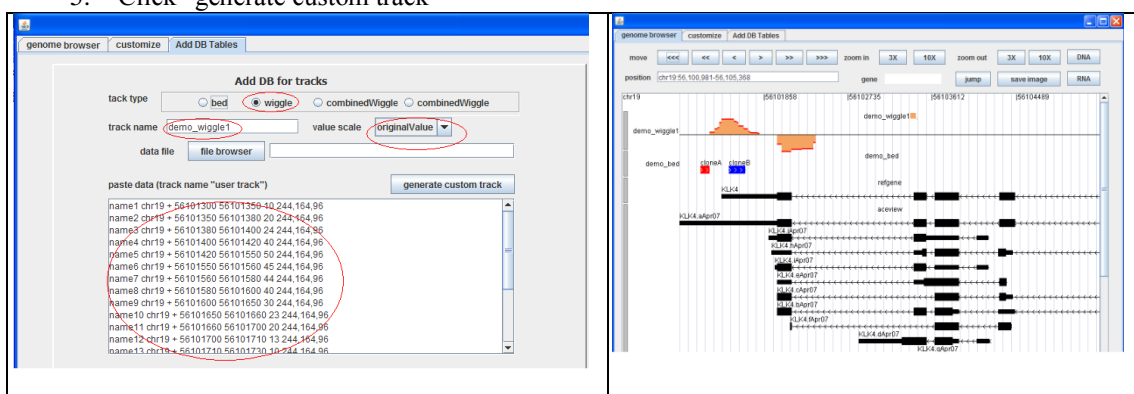
**name** - The name of this item. This label is displayed to the top of the wiggle line in the Genome Browser canvas. When the mouse pointer moves over the item, the name appears after several seconds.



**chrom** - The name of the chromosome (e.g. chr7, chrX, chr2\_random).  
**strand** – for “+” strand rows, the wiggle bars are shown above the center line. For “-” strand rows, the wiggle bars are shown under the center line.  
**chromStart** - The start of the locus for this item. The first base in a chromosome is numbered 0.  
**chromEnd** - The end of the locus for this item. The chromEnd base is not included in the display of the feature. For example, the first 100 bases of a chromosome are defined as chromStart=0, chromEnd=100, and spans the bases numbered 0-99.  
**value** - the expression level. This corresponds to the height of the bar in the genome browser canvas. It will be displayed with the name in the genome browser canvas when the mouse pointer is over this position.  
**Color** - An RGB value of the form Red, Green, Blue (e.g. red colour: 255,0,0).

Figure 9 shows how a wiggle custom track can be added in GenomeLite.

1. Select track type “wiggle”
2. Input track name, which will be shown on the top of the track in the genome browser canvas.
3. For sequencing data, “value scale” is the “originalValue”. For microarray data, “value scale” should usually be the “logscale”
4. Copy/paste the data in demo\_wiggle file, or click “file browser” to select the data file.
5. Click “generate custom track”



**Figure 9 Add wiggle track**

The top track of the genome browser canvas shows the added demo\_wiggle track. Mousingover on the edges of the wiggle shows the name of a wiggle item.

Multiple wiggle tracks can be overlaid semi-transparently in one track. This will be useful for comparison of sequencing data. An example of how to overlay two wiggles is shown below.

We create another wiggle track called demo\_wiggle2 using the same way of demo\_wiggle1. The demo\_wiggle2 data can be found in downloaded files.

The left-top panel in Figure 10 shows the “customize” window, which determines whether tracks are displayed in full or are hidden. The top two tracks/DBs are “demo\_wiggle2\_originalValue\_wiggle” and “demo\_wiggle1\_originalValue\_wiggle”. The top-right in Figure 10 shows the steps on how to put two wiggles together.

1. Select the second row “demo\_wiggle1\_originalValue\_wiggle”
2. Click “+”, dialogue window “select or type a DB name” pop-up
3. Select “demo\_wiggle2\_originalValue\_wiggle” in pop-up dialog window.
4. Click “refresh”, the two wiggle tables go to one line in the Table as shown in the bottom-left of Figure 10
5. Remove “demo\_wiggle2\_originalValue\_wiggle” from the table, then select the row “demo\_wiggle2\_originalValue\_wiggle” and click the “-” button.

The bottom-left of Figure 10 shows the result in the genome browser. The top track is the multiple wiggles for the demo data. If an item is located in the “+” strand, it will be shown above the central line; otherwise, it will be shown on the opposite side. Unlike the UCSC genome browser, GenomeLite allows plus strand items and minus strand items to be displayed even though they overlap.



Note: In the customize window, one row in the table corresponds to one track if it is visible, and is set as “full”. The button “^” and “v” replace the tracks in the genome browser canvas. They have the same function as that demonstrated in Figure 8. To re-position tracks in the genome browser canvas using the “v” and “^” button, (1) select the row in the table of the customize window, (2) click “v” and “^” to move the selected track down and up respectively, and (3) click “refresh” to show the changes in the genome browser canvas. The demo video is available in manual. The size of genome browser can also be adjusted.

GenomeLite accepts UCSC bedgraph and BAM/SAM format. The name column will be added with expression level and the colour column will be automatically produced to let user to change.

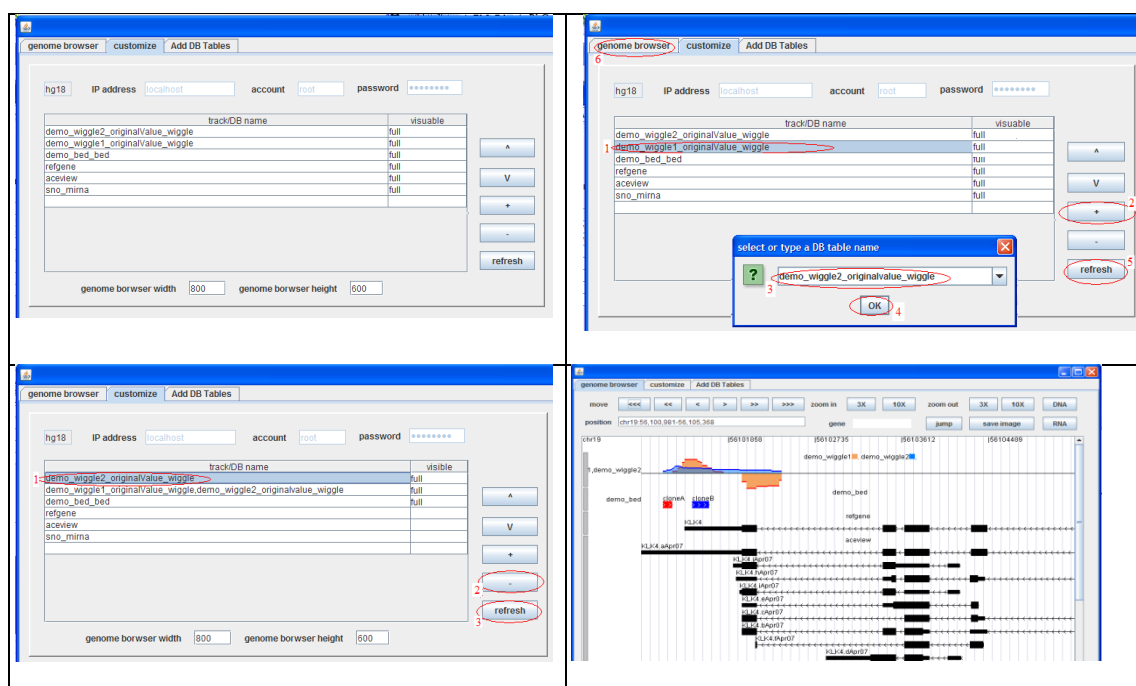


Figure 10 Adding multiple wiggles to one track

### 2.3.3. CombinedWiggle track

The height of bars can be used to illustrate the expression level of genes intuitively. The wiggle track as described in the previous section can be used to carry out this function. Since the genome browser is usually used to see the differential expression of genes, it is important to combine the tracks into one. This will save space in the genome browser canvas. GenomeLite provides a new track type called combinedWiggle, whose one row (or item) combines several related values.

**CombinedWiggle:** it has 6 columns

*probe1 chr19 + 56102250 56102350 11,8,9*  
*probe2 chr19 + 56102600 56102700 5,10,11*

**name** - The name of this item. This label is displayed to the top of the wiggle bar in the Genome Browser canvas. When the mouse pointer moves over the item, the name appears in several seconds.

**chrom** - The name of the chromosome (e.g. chr7, chrX, chr2\_random).

**strand** - for “+” strand rows, the wiggle bars are shown above the center line. For “-” strand rows, the wiggle bars are shown under the center line.

**chromStart** - The start of the locus for this item. The first base in a chromosome is numbered 0.

**chromEnd** - The end of the locus for this item. The chromEnd base is not included in the display of the feature. For example, the first 100 bases of a chromosome is defined as chromStart=0, chromEnd=100, and spans the bases numbered 0-99.

**value** - the expression level. Unlike the wiggle track, the value in this case represents a string of numbers separated by “,”. All the rows should have the same number of numbers in the value column. Each of these corresponds to the height of the bar in the genome browser canvas. The locus occupied

by the probe is divided by the number of conditions. In the example above, there are 3 values for each probe which represents three expression levels for three experimental conditions.

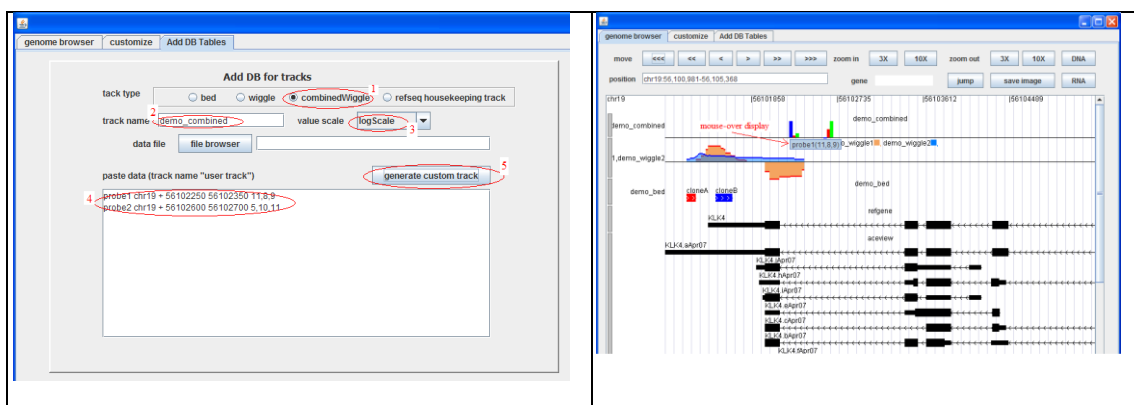


Figure 11 Add a CombinedWiggle track

Figure 11 shows how a CombinedWiggle custom track is added. First, select the third tabbed pane “Add DB Tables”

1. Select “commbinedWiggle” in track type
2. Give the track a name such as “demo\_combined”
3. Select logScale because the data is a microarray type
4. Copy/ paste two lines in demo\_combinedWiggle file, or click “file browser” to select the data file.
5. Click “generate custom track”

The generated CombinedWiggle custom track is shown in the top of the genome browser canvas. When you mouse-over on the bars, the name and expression levels of the conditions are displayed several seconds as shown in red arrow on the right in Figure 11.

### 2.3.4. Refseq housekeeping track

Except human assembly hg18, GenomeLite can visualize any genomes as long as we have their anchor: housekeeping track. You can download refgene annotation from UCSC genome browser (<http://genome.ucsc.edu/cgi-bin/hgTables?command=start>) as shown in Figure 12. The right figure of Figure 12 shows the steps of uploading housekeeping track data. After you upload the housekeeping track data, you can visualize its genome loci after you change the assembly item in track.xml.

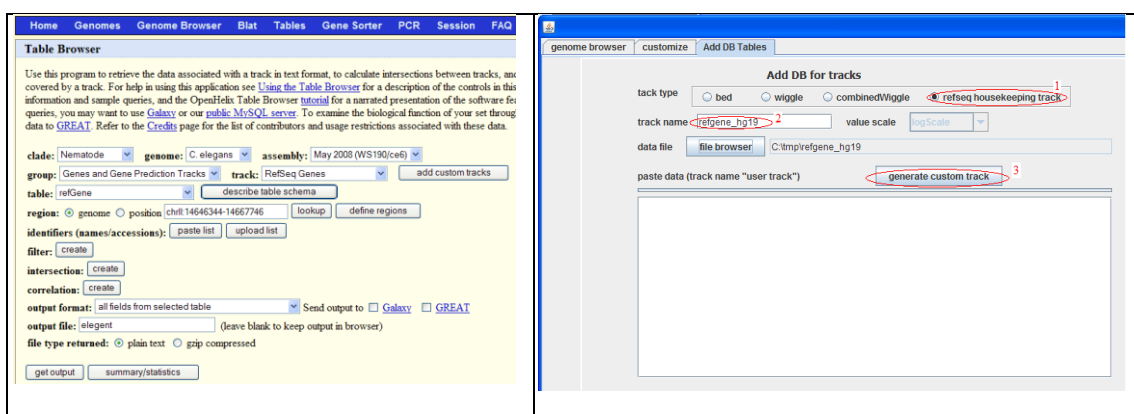
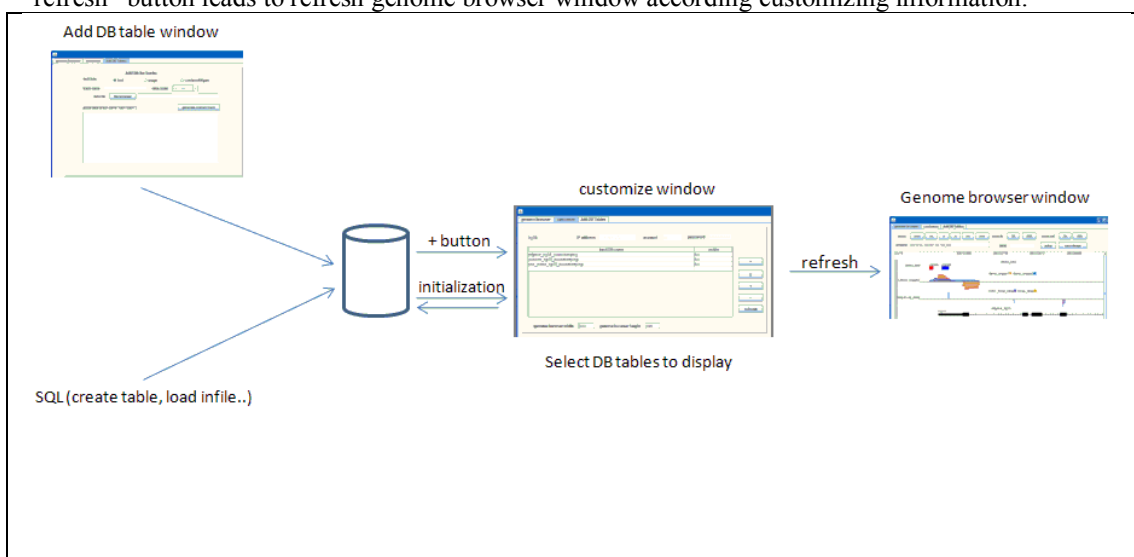


Figure 12 Add housekeeping track

### 2.3.5. Summary

Figure 13 shows the data flow in GenomeLite. Track data are generated by either GenomeLite interface or MySQL command out of GenomeLite that is described in the next section (advanced tips). You can

add/delete (“+”/“-” button) or move up/down (“^” and “V” buttons) tracks in genome browser window. “refresh” button leads to refresh genome browser window according customizing information.



**Figure 13 The relationship among the three windows**

### 3. Advanced Tips

Here we explain the usage for the “initialization” button in the customize window. This button is used to show the custom tracks that are created outside of GenomeLite. The database tables can be divided into two types: one for house-keep tracks (refSeq, aceview and sno/miRNA for human, mouse and other species), and one for custom tracks.

#### 3.1. The name of track database tables

In the default housekeeping tracks, there are “refgene\_h18\_housekeeping”, “aceview\_h18\_housekeeping” and “sno\_mirna\_h18\_housekeeping” for hg18. Users can download other assembled human transcript/genome annotations such as hg19. Users can also download transcript/genome annotations from other species such as mouse or chicken.

The BED table has the suffix “\_bed”. GenomeLite recognises tables with “\_bed” suffixes as custom BED tracks. The wiggle table has the suffix “\_logscale\_wiggle” or “\_originalvalue\_wiggle”, which depend on the values in the item. Traditionally, microarray data is represented as logscale and its range is around [2,15], while sequencing data is usually measured as the number of reads. Therefore, sequencing data is usually the originalValue. Similarly, the combined Wiggle table has the suffix “\_logscale\_combinedWiggle” or “\_originalvalue\_combinedwiggle”.

Follow Sections 2.3.1, 2.3.2 and 2.3.3 for instructions for creating tables in GenomeLite.

#### Create BED table

The full set of columns for BED files is 12 columns. If the user takes the first 6 or 9 columns, the 6-column, 9-column BED tables will be generated accordingly.

```

create table demo_bed(
chrom varchar(255),
chromStart int,
chromEnd int,
name varchar(255),
score int,

```

```
strand varchar(10),
thickStart int,
thickEnd int,
itemRgb varchar(20),
blockCount int,
blockSizes varchar(255),
blockStarts varchar(255),
index chrom_start (chrom,chromStart),index chrom_stop (chrom,chromEnd))
)engine=MyISAM"
```

### Create wiggle table

The wiggle table has 7 columns.

```
create table demo_wiggle_originalValue_wiggle(
name varchar(200),
chrom varchar(255),
strand varchar(1),
chromStart int,
chromEnd int,
value double,
color varchar(20),
index chrom_strand_start (chrom,strand,chromStart),index chrom_strand_stop
(chrom,strand,chromEnd))
engine=MyISAM";
```

```
LOAD DATA LOCAL INFILE 'E:\\ENCODE_wig\\H3k4me1\\H1hesc.wig'
INTO TABLE H3k4me1_H1hesc_originalValue_wiggle
fields terminated by "\\t"
lines terminated by "\\n"
ignore 1 lines;
```

### Create CombinedWiggle table

The CombinedWiggle table has 6 columns.

```
create table demo_combined_logScale_combinedWiggle(
name varchar(200),
chrom varchar(255),
strand varchar(1),
chromStart int,
chromEnd int,
multi_expr varchar(200),
index chrom_start (chrom,chromStart),index chrom_stop (chrom,chromEnd)
)engine=MyISAM
```

After the table is created, users can load data into the table using the SQL command “LOAD DATA LOCAL INFILE”, which is a quick way to load data into a database table.

Finally, you can change “full” to “hide” to hide tracks that are of no interest. You can also use “-” to remove tracks from the table, but it will not affect the tables in the database “gnome”. Click “+” to overlay two or more tables into one track. If you want to add a new track, select the empty row (bottom) in the table, and click “+”.

## 3.2. Generate coverage of wiggle data from a BAM/SAM file

Checking coverage is one of the main methods to determine biological functions from sequencing data such as SAM/BAM files [6]. Wiggle tracks can be used to show coverage across the genomic locus. Wiggle tracks in the UCSC genome browser has three limitations. (1) Wiggles cannot be overlaid for “+” and “-” strands. (2) No colours are shown in wiggles. (3) There is no mouse over display function for Wiggles tracks. GenomeLite addresses all three limitations, but an extended wiggle table format as

shown in Section 3.1 is needed. Listed below, is a detailed protocol for generating extended wiggle data:

1. Use bowtie, soap2 or other alignment tools to generate a SAM file, say sample.sam
2. Use samtools to convert SAM to BAM  
`>samtools view -S -b -o sample.bam sample.sam`
3. Sort the BAM file  
`>samtools sort sample.bam sample.sorted`
4. Get the chromosome size  
`>samtools view -H sample.nodup.bam | awk '/@SQ/ {OFS="\t"; gsub("SN:", "", $2); gsub("LN:", "", $3); print $2, $3}' > chromsizes.txt`
5. Create “+” strand coverage  
`>genomeCoverageBed -bg -ibam sample.sorted.bam -g chromsizes.txt -strand + > sample.plus`
6. Create “-” strand coverage  
`genomeCoverageBed -bg -ibam sample.sorted.bam -g chromsizes.txt -strand - > sample.minus`
7. Add extend columns  
`>gawk '{print "Name " $1 " " $2 " " $3 " " $4 " 244,164,96"}' sample.plus > sample.ext`  
`>gawk '{print "Name " $1 " " $2 " " $3 " " $4 " 244,164,96"}' sample.minus >> sample.ext`  
 Note, different colours for different wiggle files will need to be set in order to have clear distinction of overlapping transcripts. “middle” colours are suggested, instead of the pure colours “255,0,0” or “0,255,0”.

### 3.3. Create other house-keeping track data

In GenomeLite you can use different version of genomes, such as we can use hg19 instead of hg18. You can also use other genomes such as mouse, rat. Except upload housekeeping track data from “Add DB” window in Section 2.3.4, here we give an example for upload mouse housekeeping track data. We have downloaded “refgene\_mm9” from the UCSC genome browser <http://genome.ucsc.edu/cgi-bin/hgTables?command=start>. You can find the file in “housekeeping\_tracks” subdirectory. Other house keeping track data can be downloaded in the same manner.

The other transcript annotation such as ENSEMBL can also be done as above. Please note that the file name has to be end with assembly, for example ensemble\_hg18 or ensemble\_hg19.

Figure 14 shows how to create mouse mm9 refseq housekeeping track data in windows.

- refgene\_hg19 and refgene\_mm9 See red ellipse mark in the figure.
- Login MySQL and goto genome database see mark (1) and (2).
- Go to database genome. Type “create database genome”, if genome database has not been created. See mark (3).
- Create table refgene\_mm9\_housekeeping table. Copy and paste the following contents.

```
create table refgene_mm9_housekeeping (
  bin      smallint(5) unsigned,
  name     varchar(255),
  chrom    varchar(255),
  strand   char(1),
  txStart  int(10) unsigned,
  txEnd    int(10) unsigned,
  cdsStart int(10) unsigned,
  cdsEnd   int(10) unsigned,
  exonCount int(10) unsigned,
  exonStarts longblob,
  exonEnds  longblob,
  id       int(10) unsigned,
  name2    varchar(255),
  cdsStartStat enum('none','unk','incmpl','cmpl'),
  cdsEndStat  enum('none','unk','incmpl','cmpl'),
  exonFrames  longblob,
```

```

index chrom_start (chrom,chromStart),index chrom_stop (chrom,chromEnd)
) engine=MyISAM;
See red Mark (3).

```

- Load data into created table. Type  
LOAD DATA LOCAL INFILE 'refgene\_mm9' INTO TABLE refgene\_mm9\_housekeeping  
lines terminated by "\n" ignore 1 lines  
See read Mark (4).

```

C:\WINDOWS\system32\cmd.exe - "\Program Files\MySQL\MySQL Server 5.1\bin\mysql.exe" ...
C:\GenomeLite\housekeeping_tracks>
C:\GenomeLite\housekeeping_tracks>
C:\GenomeLite\housekeeping_tracks>dir
Volume in drive C is SYSTEM
Volume Serial Number is 229C-5FCE

Directory of C:\GenomeLite\housekeeping_tracks

07/04/2011  01:45 PM    <DIR>          .
07/04/2011  01:45 PM    <DIR>          ..
14/02/2011  08:54 AM             10,869,044 refgene_hg19
08/03/2011  10:39 AM             8,031,558 refgene_mm9
                2 File(s)      18,900,602 bytes
                2 Dir(s)   10,102,792,192 bytes free

C:\GenomeLite\housekeeping_tracks>"\Program Files\MySQL\MySQL Server 5.1\bin\mys
ql.exe" -uroot -p
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 47
Server version: 5.1.54-community MySQL Community Server (GPL)

Copyright (c) 2000, 2010, Oracle and/or its affiliates. All rights reserved.
This software comes with ABSOLUTELY NO WARRANTY. This is free software,
and you are welcome to modify and redistribute it under the GPL v2 license

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use genome
Database changed
mysql> create table refgene_mm9_housekeeping (
-> bin          smallint(5) unsigned,
-> name         varchar(255),
-> chrom        varchar(255),
-> strand       char(1),
-> txStart      int(10) unsigned,
-> txEnd        int(10) unsigned,
-> cdsStart     int(10) unsigned,
-> cdsEnd       int(10) unsigned,
-> exonCount    int(10) unsigned,
-> exonStarts   longblob,
-> exonEnds     longblob,
-> id           int(10) unsigned,
-> name2        varchar(255),
-> cdsStartStat enum('none','unk','incmpl','cmpl'),
-> cdsEndStat   enum('none','unk','incmpl','cmpl'),
-> exonFrames   longblob,
-> index(chrom,txEnd),index(chrom,txStart)
-> ) engine=MyISAM;
Query OK, 0 rows affected (0.08 sec)

mysql> LOAD DATA LOCAL INFILE 'refgene_mm9' INTO TABLE refgene_mm9_housekeeping l
ines terminated by "\n" ignore 1 lines;
Query OK, 28190 rows affected (0.23 sec)
Records: 28190 Deleted: 0 Skipped: 0 Warnings: 0

mysql>

```

Figure 14 create mm9 refseq housekeeping track data

Following the same manner, you can create tables for other housekeeping tracks below:

**create aceview database and upload**

```

create table aceview_mm9_housekeeping (
bin          smallint(5) unsigned,
name         varchar(255),
chrom        varchar(255),
strand       char(1),
txStart      int(10) unsigned,
txEnd        int(10) unsigned,
cdsStart     int(10) unsigned,

```

```

cdsEnd int(10) unsigned,
exonCount int(10) unsigned,
exonStarts longblob,
exonEnds longblob,
index (chrom,txStart), index (chrom,txEnd)
) engine=MyISAM;
LOAD DATA LOCAL INFILE 'aceview_mm9' INTO TABLE aceview_mm9_housekeeping lines
terminated by "\n" ignore 1 lines ;

```

#### **create sno/miRNA database and upload**

```

create table sno_miRNA_mm9_housekeeping (
bin int,
chrom varchar(10),
chromStart int,
chromEnd int,
name varchar(20),
score int,
strand char(1),
thickStart int,
thickEnd int,
type varchar(20),
index (chrom,txStart), index (chrom,txEnd)
) engine=MyISAM;
LOAD DATA LOCAL INFILE 'sno_mirna_mm9' INTO TABLE sno_miRNA_mm9_housekeeping
lines terminated by "\n" ignore 1 lines ;

```

## **References**

1. Fujita P.A., Rhead B., Zweig A.S., Hinrichs A.S., Karolchik D., Cline M.S., Goldman M., Barber G.P., Clawson H., Coelho A., et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* 2011;39:D876-D882.
2. I.L. Hofacker, W. Fontana, P.F. Stadler, S. Bonhoeffer, M. Tacker, P. Schuster (1994)
3. Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte f. Chemie* 125: 167-188.
4. Karolchik,D., Baertsch,R., Diekhans,M., Furey,T.S., Hinrichs,A., Lu,Y.T., Roskin,K.M., Schwartz,M., Sugnet,C.W., Thomas,D.J. et al. (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, 31, 51–54.
5. Rhead,B., Karolchik,D., Kuhn,R.M., Hinrichs,A.S., Zweig,A.S., Fujita,P.A., Diekhans,M., Smith,K.E., Rosenbloom,K.R., Raney,B.J. et al. (2010) The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.*, 38, D613–D619.
6. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078-9.
7. Lai J, Lehman ML, Dinger ME, Hendy SC, Mercer TR, Seim I, Lawrence MG, Mattick JS, Clements JA, Nelson CC. A variant of the KLK4 gene is expressed as a cis sense-antisense chimeric transcript in prostate cancer cells. *RNA* 2010, 16:1156–1166.
8. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. *Nature Biotechnology* 29, 24–26 (2011).
9. <http://wis.cs.ucla.edu/~hxwang/proj/delta.html>