Readme
IsoEM: Isoform expression level estimation through Expectation Maximization

Input

The IsoEM plugin requires as input the fragment length distribution mean and standard deviation, and the reference genome and transcript library. The input can be provided through the IsoEM input form. The genome/transcript library is selected through a drop down list that refers to bundled genome fasta, transcript library GTF, and corresponding tmap index that are included within the plugin.

Output

Three outputs are provided by IsoEM

isoform_estimates.html has the Fragment Per Kilobase per Million reads (FPKM) estimates for the isoforms in the selected transcript library. It also includes a link for a downloadable text file with the isoform FPKM estimates.

gene_estimates.html has the FPKM estimates for the genes in the selected transcript library. It also includes a link for a downloadable text file with the gene FPKM estimates.

UCSC_tracks.html: has links to two downloadable UCSC track files. The coverage tracks file is a bed file that includes two UCSC tracks which has weighted coverages of the first and second most abundant isoforms of every gene. The FPKM labeled transcript track is a GTF that includes expressed isoforms labeled with their estimated FPKM values and color coded with shades of gray according to their expression levels on a $\log_2$ scale, where an isoform with FPKM value greater than or equal to 1000 is displayed in black. The intensity (shade) of each isoform is calculated as follows:

$$\mathtt{intensity = min(1000,\ 1\ +\ c\ *\ log_2\,max(1,fpkm)),}$$
$$\mathtt{where\quad c\ =\ 999/log_2\,1000}$$

Displaying UCSC tracks

Once downloaded, the UCSC tracks can be uploaded to the UCSC genome browser at http://genome.ucsc.edu/cgi-bin/hgGateway as follows:

- Select the genome.
- Select *add custom tracks*.
- Upload the bed file and GTF file, and select *submit*. Alternatively, paste the URLs of the track files in the *Paste URLs or data* box, and select *submit*.
  Three tracks should appear in the Manage custom track table.
- Select *go to genome browser*.
- In the genome browser, right click on each of the three uploaded tracks and select *full* from the pop-up menu.
- If the isoform values are not fully displayed, select *configure*, and on the *Configure Image* page, increase the *label area width*.

- The display of each track can be additionally controlled through settings available in the *Configure* option in the track pop-up menu.

Figure 1 shows a display of the IsoEM generated coverage tracks and FPKM labeled isoforms track for RPL15 (ENSG00000174748) in the MAQC UHR sample generated from Ion Torrent Auto_GOG-140_284 run reads.

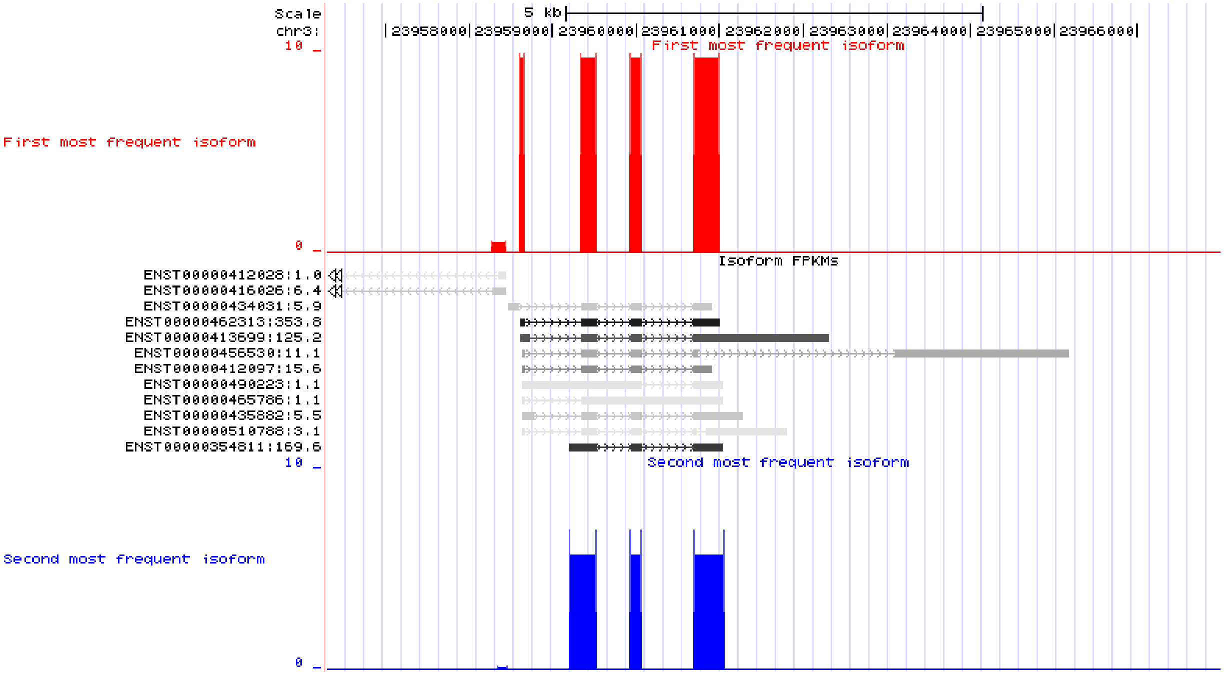

Figure 1: UCSC custom tracks generated by IsoEM

Contact Information

sahar@engr.uconn.edu

ion@engr.uconn.edu

alexz@cs.gsu.edu