



Visualizing RNA-Seq data with RNAseqViewer



Xavier Rogé, Xuegong Zhang

Bioinformatics Division / CSSB, TNLIST and MOE Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University
rogex10@mails.tsinghua.edu.cn, zhangxg@tsinghua.edu.cn

Motivation

Large datasets

New advances in RNA sequencing have opened up new horizons in the field of transcriptomics and given access to new extensive data.

A complement to automated tools

The analysis of these data needs effective visualization tools, so as scientists can gain an insight into the data and are able to review the results of the computational tools.

Result

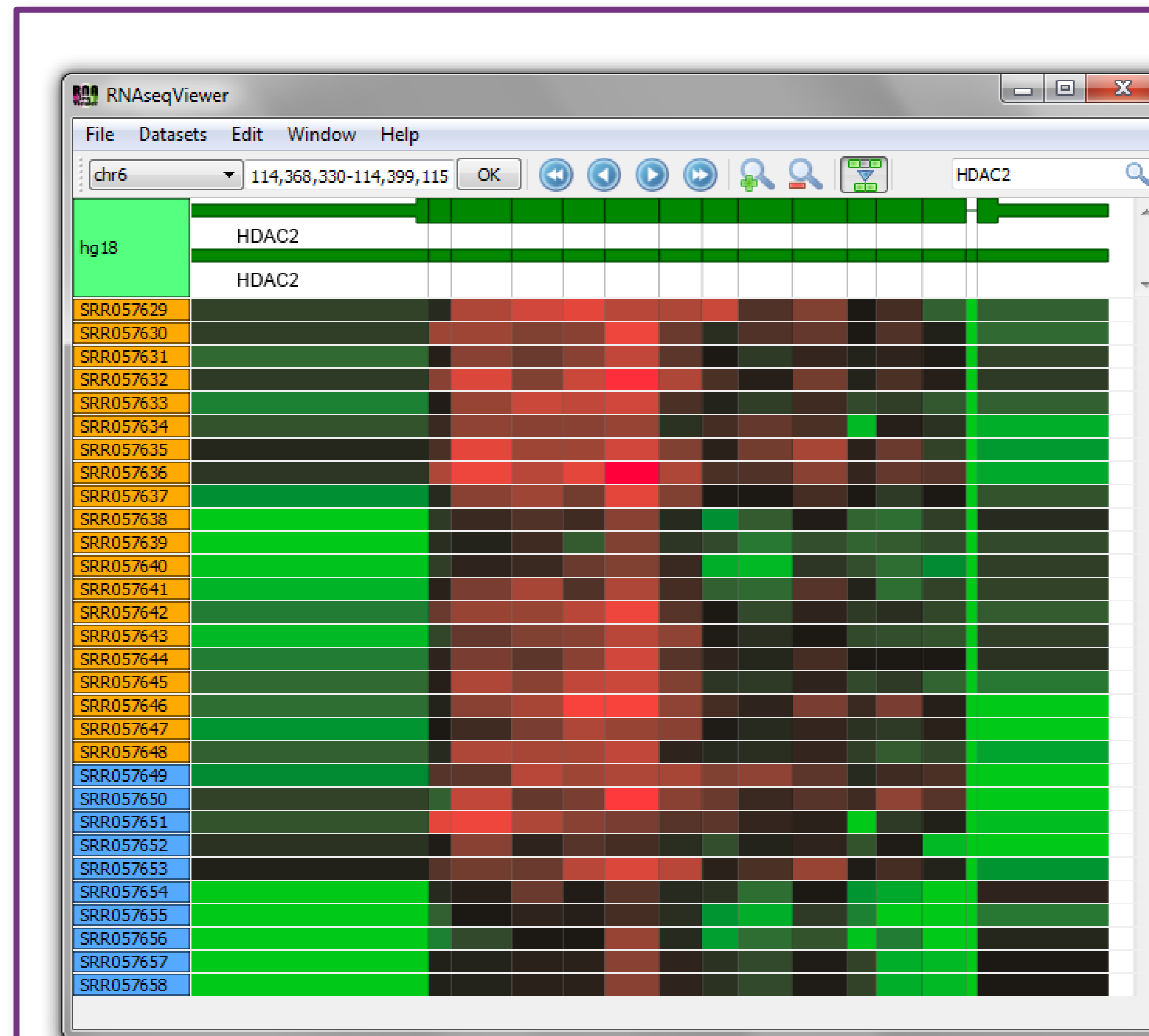
A visualization tool dedicated to RNA-Seq

We developed a new software program, RNAseqViewer, to visualize the various data from the RNA-Seq analyzing process for single or multiple samples. By focusing on expression of genes and transcript isoforms, the program offers innovative ways to present the transcriptome data in a quantitative and interactive manner.

Availability

Source code and binaries are freely available for academic use at:

<http://bioinfo.au.tsinghua.edu.cn/software/RNAseqViewer/>



Analysing multiple samples

The screenshot shows the read expression of 30 samples and one gene annotation track.

The figure was obtained by first mapping the reads from the study GSE22260^[1] to the reference genome hg18 using TopHat. Resulting BAM files were loaded in RNAseqViewer and the "heatmap view" was chosen. Intronic regions have been hidden for better visualization of exon expression.

The first 20 tracks are samples from 20 prostate cancer tumors, and the 10 following tracks come from 10 matched normal tissues. The color depends on the FPKM of each sample (represented by the rows) for each exon (represented by the columns).

The resulting heatmap highlights the differential gene and isoform expression among the samples.

[1] Kannan K, Wang L, Wang J, Ittmann MM et al. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proc Natl Acad Sci USA* 2011 May 31;108(22):9172-7

Supported data (with file formats and types of graphs)

Mapped reads	Junction reads	Transcriptomes	DNA sequences	Gene annotations	Numeric data	Othe tracks
SAM BAM	BED	GTF	FASTA	RefFlat GTF	Wiggle	BED
• Aligned reads • Coverage plot • Heatmap	• Junction lines • Junction reads	• Compact view • Expanded view	• Nucleotide sequence	• Compact view • Expanded view	• Histogram	• Compact view • Expanded view

Features

Intuitive views

Different types of view for each data set allow the visualization of different levels of information, including heatmap-like views for informative and yet very compact tracks, making possible to visualize dozens of samples simultaneously.

Dynamic user interface

Special attention has been given to the user interface, so that the data can be explored in a fast and intuitive way. Users can drag the tracks and use the mouse wheel for panning and zooming, or use the tool bar's button or keyboard shortcut. Access to a specific region can be made by entering the region's coordinates or by searching a gene by its name.

Small memory footprint

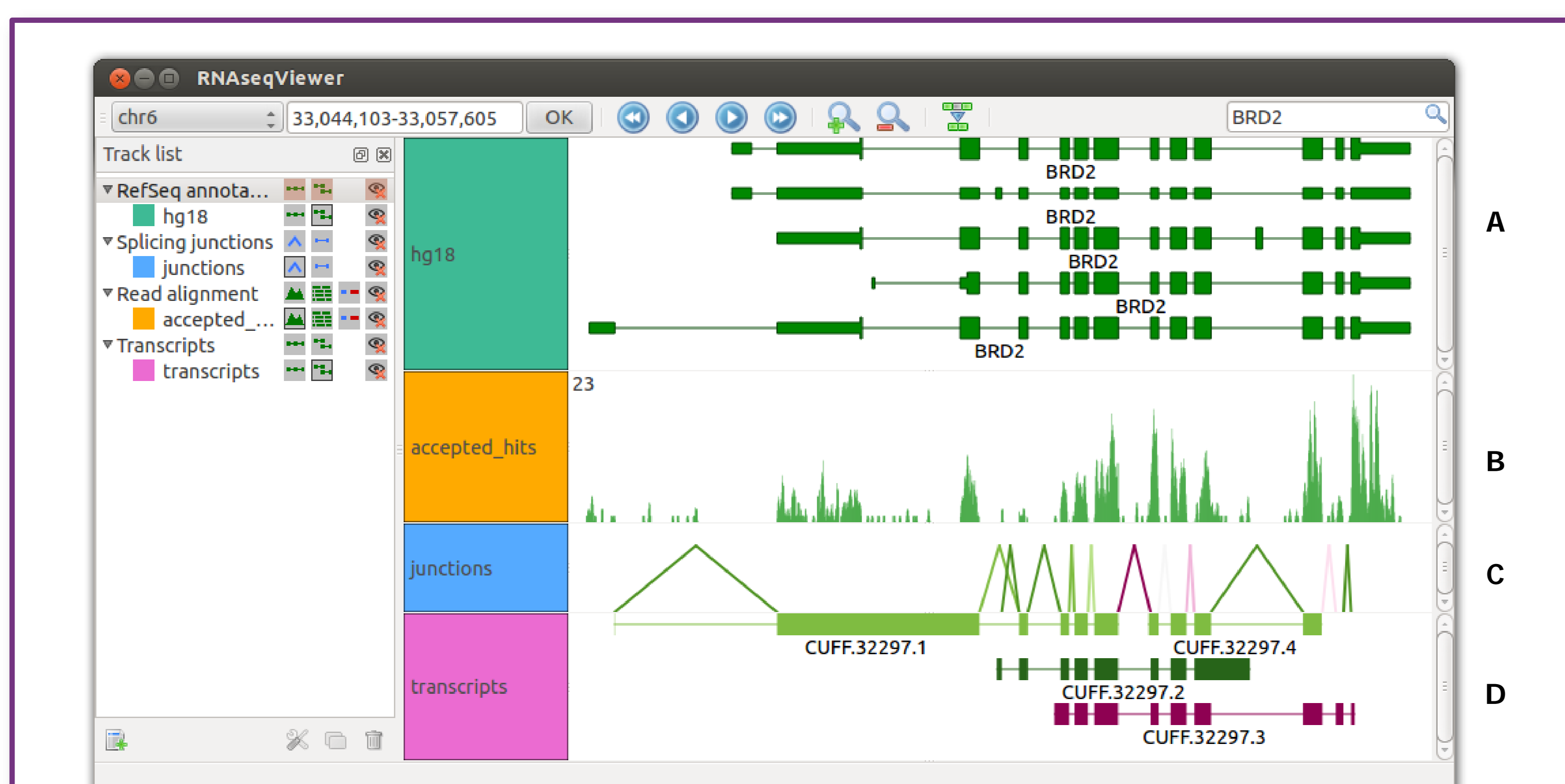
The management of the memory has been designed so that very large data sets can be visualized without exceeding memory limits nor affecting the fluidity of the user interface.

Feature table

Navigation	Dynamic zooming and panning (mouse / clickable buttons / keyboard) Search by position or by gene name
Supported formats	SAM/BAM, GTF, BED, Wiggle, RefFlat, FASTA
Graphs	Multiple types of view for each format Extensive use of colors to indicate RPKM or scores
Language	Python
Distributions	Linux and Windows executables Python source code (needs Python 3, PySide and Samtools)
Internationalization	English, Chinese, French
Other features	Hide intronic regions Print out tracks Export tracks as PDF or image (PNG, SVG, JPEG, TIF, etc.)

Conclusion

RNAseqViewer is a handy tool for scientists who use RNA-Seq data to compare multiple transcriptomes, for example, to compare gene expression and alternative splicing of cancer samples or of different development stages.



Analysing a single sample

RNAseqViewer can combine different types of graphs for a detailed analysis of a given sample. The left part of the screenshot shows the list of tracks and the options to choose the type of view. The right part shows 4 tracks:

- Referene genome annotation (human genome, hg18 assembly) in *expanded* view. The view has been positioned by searching for the gene "BRD2" in the top-right box.
- Coverage graph, obtained by loading TopHat's output 'accepted_hits.bam' and selecting the *coverage* view.
- Splicing junctions, obtained by loading TopHat's output 'junctions.bed' and selecting the *broken lines* view. Colors depend on the number of junction reads which map to each junction.
- Transcriptome assembly, obtained by loading Cufflinks' output 'transcripts.gtf'. The program shows 4 isoforms, which color depends on their respective FPKM.