SevenBridges

SCALABLE, REPRODUCIBLE RNA-Seq

SCALABLE, REPRODUCIBLE RNA-Seq

Advances in the RNA sequencing workflow, from sample preparation through data analysis, are enabling deeper and more accurate exploration and measurement of gene expression in the transcriptome.

RNA-SEQ ANALYSIS IS LARGELY DEPENDENT ON THE BIOINFORMATICS TOOLS AVAILABLE TO SUPPORT THE DIFFERENT STEPS OF THE PROCESS.

The Seven Bridges Platform hosts many of the major RNA-Seq tools, in addition to pre-built pipelines for each of the secondary analysis steps. Using these tools and pipelines, you can go from raw sequencing reads, all the way through visualization of differentially expressed genes and transcripts.

The Seven Bridges Platform hosts tools to do the following: quality control, alignment, quantitation, differential expression analysis, fusion gene identification, transcript assembly and visualization.

The most frequently used tools are listed in the figure below:

Quality Control: RSeQC; RNA-SeQ; FastQC	**Aligners:** STAR; TopHat; Bowtie/Bowtie 2
Quantitation: Cuffquant *(from Cufflinks package)*; HTSeq; RSEM; eXpress; MISO	**Differential expression analysis:** Cuffdiff *(from Cufflinks package)*; DESeq/-DESeq 2; MISO; MATS
Fusion genes: ChimeraScan; STAR; TopHat-Fusion; Chimera; Oncofuse	**Transcript assembly:** Trinity; Cufflinks *(from Cufflinks package)*
Visualization: CummeRbund; Spectacle	**Utility toolkits:** Picard; SAMtools; Trimmomatic; FastqMcf; Flexbar

QUALITY CONTROL

Quality control is a critical step in any RNA-Seq experiment. With this in mind, we've ensured that all the needed tools are pre-installed. Whether you are removing low-quality data or validating an experimental design, we've set up each tool to run in the cloud and easily connect with the rest of an experimental pipeline.

Some pre-installed tools include **RSeQC**, **RNA-SeQC**, **FastQC**, **Trimmomatic**, **FastqMcf**, **Flexbar**, **Picard** and **SAMtools**.

SEQUENCE ALIGNMENT

We support each step in the RNA-Seq analysis process, starting with **sequence alignment** to a reference genome or data from a transcriptome database. RNA-Seq pipelines on the Platform use **STAR**, **TopHat** and **Bowtie/Bowtie 2** aligners to map full length RNA sequences and align reads across canonical and non-canonical splice junctions.

QUANTITATION

To assist in transcript and isoform **quantitation**, Seven Bridges has loaded multiple common tools for quick deployment including: **Cuffquant** *(from the Cufflinks package)*, **HTSeq**, **RSEM**, **express** and **MISO** tools.

DIFFERENTIAL EXPRESSION ANALYSIS

Because the identification of differentially expressed genes and transcripts is one of the most common applications of RNA-Seq, the Seven Bridges Platform offers highly optimized tools for the task. It also supports other common differential expression-related work, including detection of alterations in transcription start sites and alternate coding regions, as well as identification of differential splicing between conditions.

Differential expression pipelines on the Seven Bridges Platform use **Cuffdiff** *(from the Cufflinks package)*, **DESeq/DESeq2**, **MISO** and **MATS** tools to find significant differences, accounting for both biological and technical variability in transcript expression, and read count distribution between samples.

FUSION GENE IDENTIFICATION

Detection of known gene fusions, and identification of novel ones can, lead to a better understanding of the triggering mechanism and progression of cancers, where they are frequently found producing unusual genetic modifications.

TRANSCRIPTOME ASSEMBLY

There are several fusion finder tools available on the Platform for use on RNA-Seq data: **ChimeraScan**, **STAR**, **TopHat-Fusion**, **Chimera** and **Oncofuse**. *(See more below in the discussion on the ChimeraScan pipeline)*

Genome-guided methods for transcript assembly use a reference genome as a template to align and assemble reads, whilst for de nova methods reads are assembled directly in transcripts.

Pipelines on the Platform use **Trinity** and **Cufflinks** for de nova and genome-guided assembly respectively.

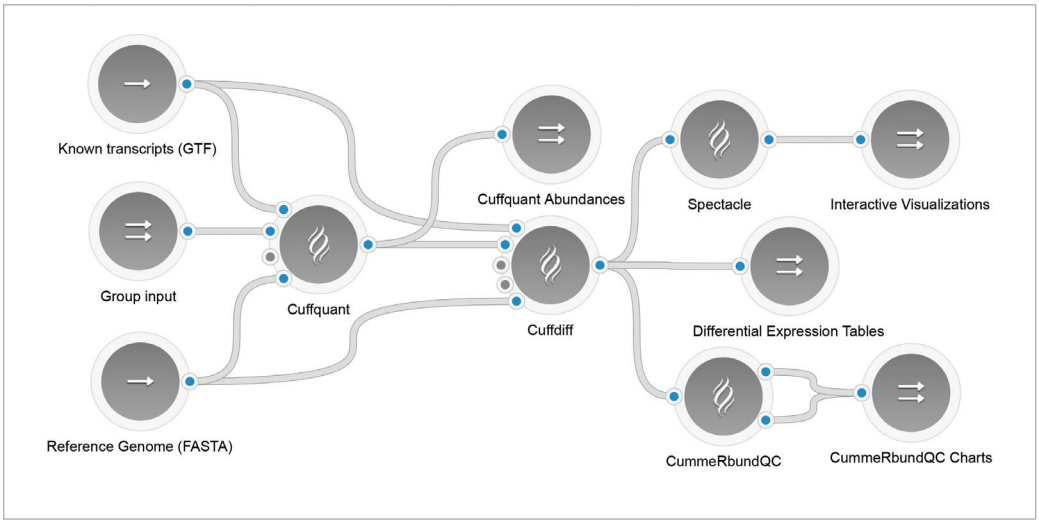
VISUALIZATION

Visualization tools help promote rapid analysis of RNA-Seq data by aggregating and indexing your RNA-Seq data, and allowing you to easily visualize it, while maintaining appropriate relationships between connected data points.

The Seven Bridges Platform incorporates **CummeRbund** and **Spectacle** developed by Seven Bridges *(see the case study below)* as tools for RNA-Seq data visualization.

Seven Bridges Pipelines are ready-to-run bioinformatics workflows. In the following case studies, we outline two pipelines scientists commonly use in the course of RNA-Seq experiments. Both are pre-built, so you can go from raw sequence to interpretable result as quickly as possible.

1. RNA-SEQ DIFFERENTIAL EXPRESSION: CUFFDIFF (WITH VISUALIZATION)



INTRODUCTION

This pipeline uses Cuffdiff to detect differential expression at the gene, isoform, tss and eds levels. The ability to detect true significant changes *(and limit false positive detections)* is determined by the number of replicates included in an experiment and the inter-replicate variability.

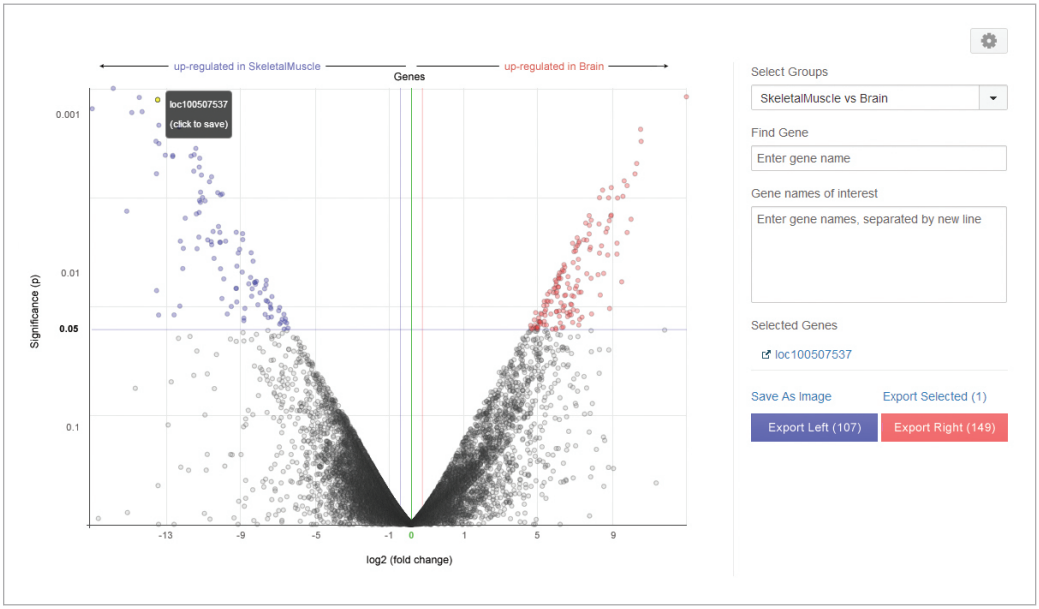
METHOD

While you can supply aligned reads as BAM or SAM files directly to Cuffdiff, using the Cuffquant utility allows you to quantify expression levels of all samples in parallel, and can significantly reduce the total running time of a differential expression analysis. It generates binary files in a CXB format.

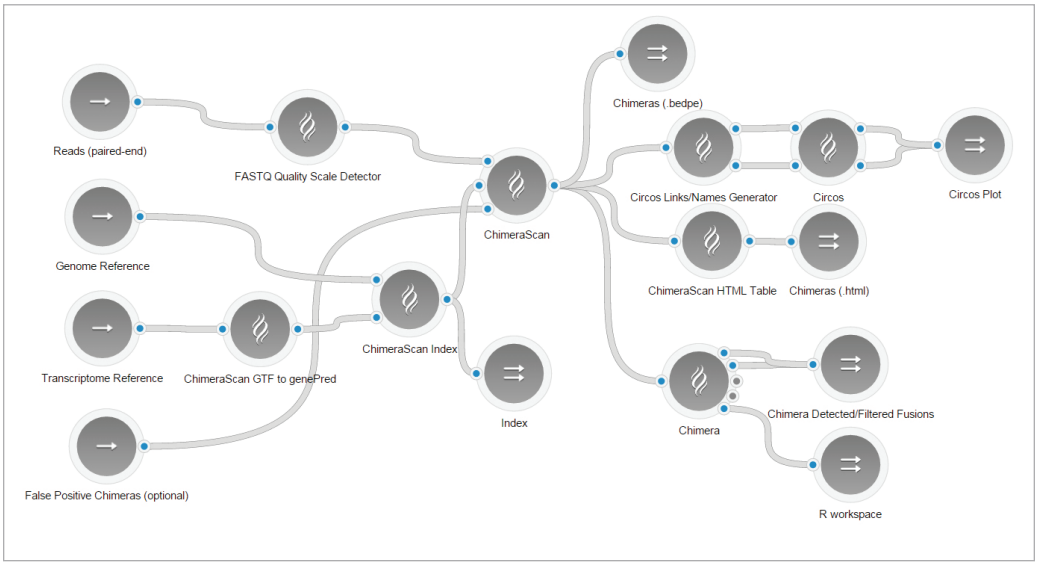
Following quantitation, Cuffdiff performs differential expression tests between groups of samples.[2]

This pipeline also performs basic quality control analysis of your differential expression experiment powered by CummeRbund.[3]

Lastly, this pipeline runs Spectacle, a tool developed by Seven Bridges that renders interactive visualizations from Cuffdiff results. Spectacle allows you to explore differential expression results in the form of interactive Scatter and Volcano plots, and export lists of interesting genes for further analysis.



2. A TECHNICAL SHOWCASE: FUSION TRANSCRIPT DETECTION - CHIMERASCAN



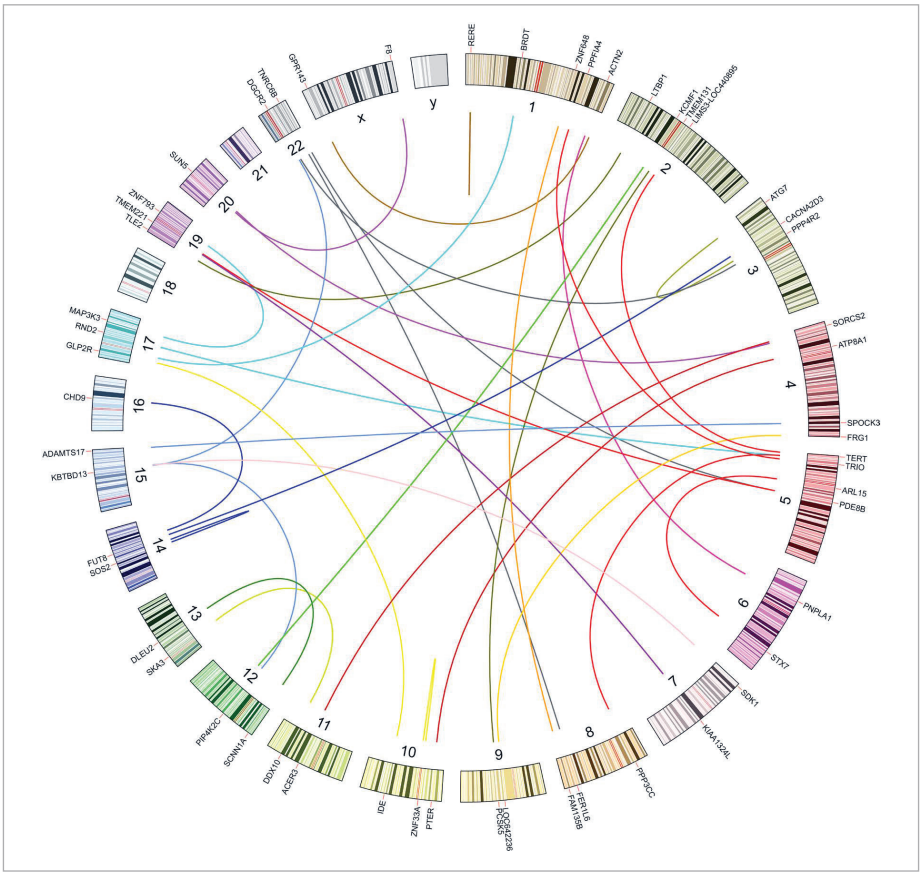
INTRODUCTION

In this pipeline, we have used the ChimeraScan software package that is aimed at detecting fusion genes.[1] While ChimeraScan is found to perform strongly with few false positive fusion detections, this pipeline only accepts paired-end RNA-Seq data. With this in mind, we have developed a Fusion Transcript Detection - STAR+ Chimera pipeline that can provide fast performance with both single-, and paired-end reads.

METHOD

Fusion transcripts detection is performed using the ChimeraScan software package *(including the primary program that detects fusion genes)*. The procedure also makes use of an accessory tool that prepares references for proper indexing in upstream analysis, and tools for preparing output files in HTML table format as well as a format suitable for visual representation.[1]

Additionally, Chimera has been added to this pipeline to provide additional control of detected fusion genes.[2] Finally, graphical representation of identified fusion genes is provided by the genomic coordinate visualization tool, Circos.[3]



REFERENCES

1. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. Jan, 31 (1):46-53. (2013)
2. a. Carrara, M. et al. State-of-the-Art Fusion-Finder Algorithms Sensitivity and Specificity. BioMed Research International, vol. 2013, Article ID 340620, 6 pages, 2013.

   b. Mitelman, F. et al. The impact of translocations and gene fusions on cancer causation. Nat. Rev. Cancer 7, 233–245 (2007).

   c. Iyer, M. K. et al. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 27, 2903–2904 (2011).

   d. Calogero, R. A. et al. chimera: A package for secondary analysis of fusion products. (2014). at bioconductor.org

   e. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009).
3. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. Mar, 7(3): 562-578. (2012)

TEAM@SBGENOMICS.COM
SBGENOMICS.COM

SevenBridges