RNA-Seq and Alternative Splicing

Matt Gitzendanner: magitz@ufl.edu
Alex Moskalenko: om@hpc.ufl.edu

12/4/12

RNA-Seq

- ▸ De novo assembly of RNA-seq data
  - ◦ SOAPdenovo-Trans (CLI)
  - ◦ Oases (CLI)
  - ◦ Trinity (Galaxy and CLI)
  - ◦ Trans-Abyss

- ▸ Different from genomic de novo assembler
  - ◦ Needs to expect coverage differences

RNA-seq

- ▸ Reference based
  - ◦ TopHat
    - · Align reads to reference genome
    - · Discover transcript splice sites
  - ◦ Cufflinks
    - · Assemble transcript models
    - · Compare different datasets to identify diffenetial expression

---

Trapnell et al., 2012

Nature Protocols 7:562-578

PROTOCOL

Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks

Cole Trapnell[1,2], Adam Roberts[3], Loyal Goff[1,2,4], Geo Pertea[3,6], Daehwan Kim[3,7], David R Kelley[1,2], Harold Pimentel[3], Steven L Salzberg[3,6], John L Rinn[1,2] & Lior Pachter[3,8,9]

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. [3]Department of Computer Science, University of California, Berkeley, California, USA. [4]Computer Science and Artificial Intelligence Lab, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]Department of Medicine, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [6]Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. [7]Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland, USA. [8]Department of Mathematics, University of California, Berkeley, California, USA. [9]Department of Molecular and Cell Biology, University of California, Berkeley, California, USA. Correspondence should be addressed to C.T. (cole@broadinstitute.org).

The Tuxedo Suite

- ▸ Bowtie
- ▸ TopHat (uses Bowtie)
- ▸ Cufflinks
  - ◦ Cuffcompare
  - ◦ Cuffmerge
  - ◦ Cuffdiff
- ▸ CummeRbund

Bowtie
Extremely fast, general purpose short read aligner

TopHat
Aligns RNA-Seq reads to the genome using Bowtie
Discovers splice sites

Cufflinks package

Cufflinks
Assembles transcripts

Cuffcompare
Compares transcript assemblies to annotation

Cuffmerge
Merges two or more transcript assemblies

Cuffdiff
Finds differentially expressed genes and transcripts
Detects differential splicing and promoter use

CummeRbund
Plots abundance and differential expression results from Cuffdiff

---

Trapnell et al. Protocol

Trapnell et al. Protocol cont.

Map reads spanning introns

Isoforms

cummeRbund

CummeRbund is an R package, facilitating visual interpretation of Cufflinks analyses

library(cummeRbund)

csScatter(isoforms(cuff_data), 'q1', 'q2')

cummeRbund

csVolcano(genes(cuff_data), 'q1', 'q2')
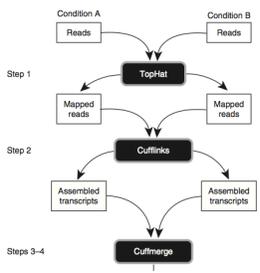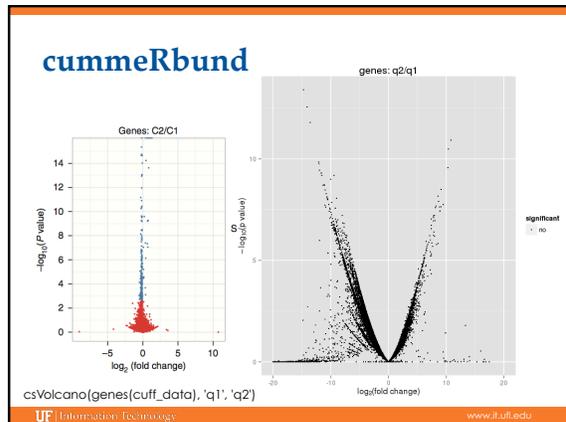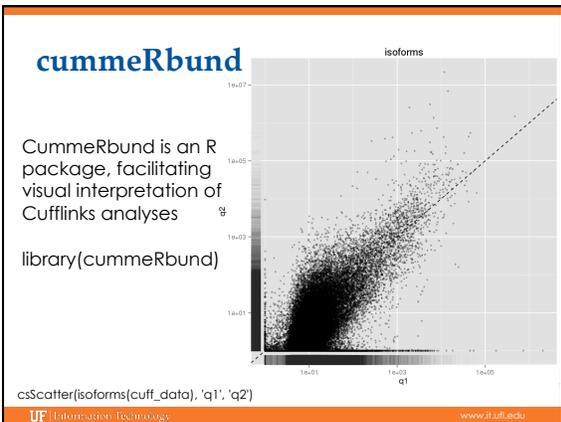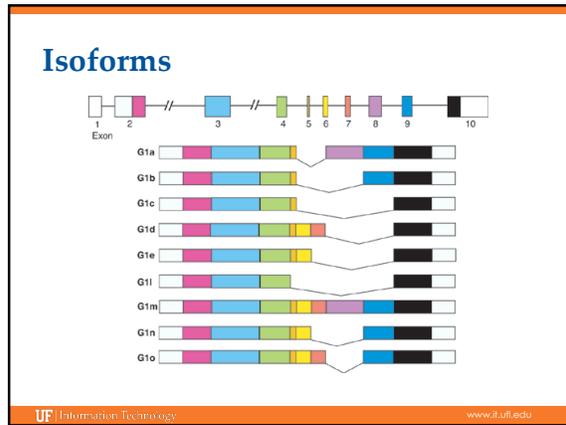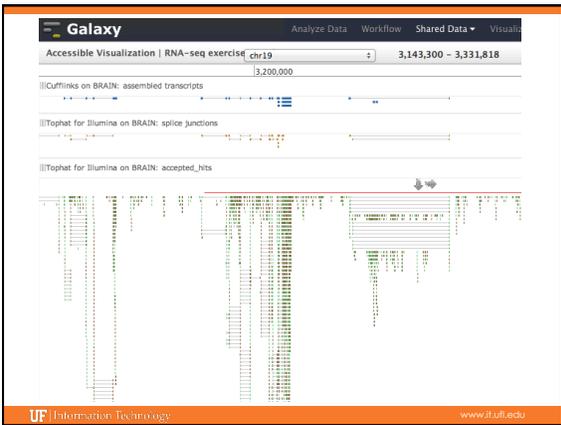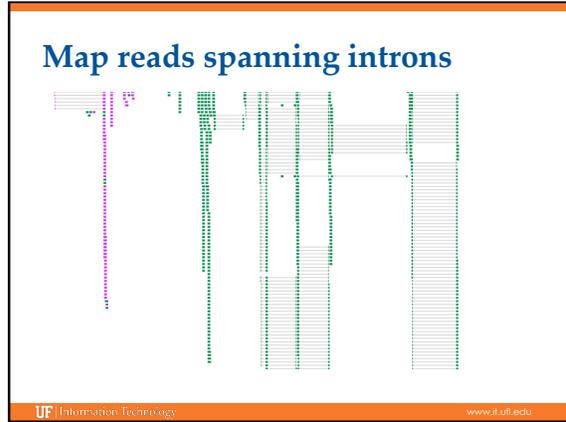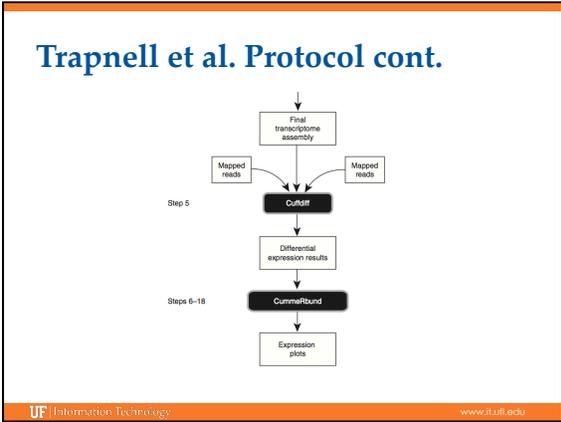
Miso

MISO / Probabilistic analysis and design of RNA-Seq experiments for identifying isoform regulation

Home | Paper | Software | Documentation | Datasets | Contact

- Improved accuracy in isoform frequencies
- Takes TopHat output (or other SAM file)

chr17:45816186:45816265:-@chr17:45815912:45815950:-@chr17:45814875:45814965:-

Sashimi plot
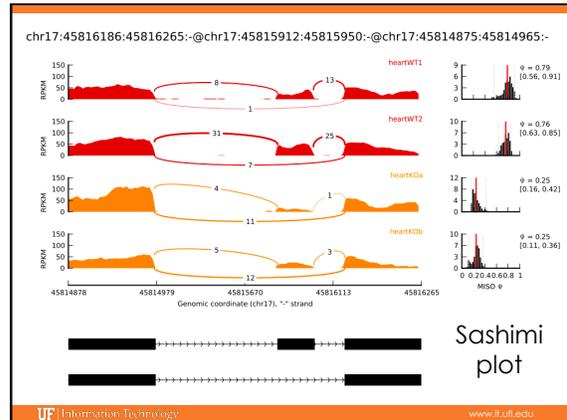
UF HPC Training: Spring 2013

- Mondays 11:00-noon
- Starts January 14th
- Session topics needed

UF Research Computing

- Help and Support
  - Help Request Tickets
    - https://support.hpc.ufl.edu
    - For any kind of question or help requests
    - Searchable database of solutions
  - We are here to help!
    - support@hpc.ufl.edu

UF Research Computing

- Help and Support (Continued)
  - http://wiki.hpc.ufl.edu
    - Documents on hardware and software resources
    - Various user guides
    - Many sample submission scripts
  - http://hpc.ufl.edu/support
    - Frequently Asked Questions
    - Account set up and maintenance