

ORAL PRESENTATION

Open Access

Oqtans: a Galaxy-integrated workflow for quantitative transcriptome analysis from NGS Data

Sebastian J Schultheiss^{1*}, Géraldine Jean¹, Jonas Behr¹, Regina Bohnert¹, Philipp Drewe¹, Nico Görnitz^{1,2}, André Kahles¹, Pramod Mudrakarta¹, Vipin T Sreedharan¹, Georg Zeller^{1,3}, Gunnar Rätsch¹

From Seventh International Society for Computational Biology (ISCB) Student Council Symposium 2011
Vienna, Austria. 15 July 2011

Background

The current revolution in sequencing technologies allows us to obtain a much more detailed picture of transcriptomes via RNA-Sequencing. We have developed the first integrative online platform, oqtans, for quantitatively analyzing RNA-Seq experiments. Our approach of providing a self-contained machine image with the accessible, transparent Galaxy framework [1] minimizes the risk of using a third-party web service for data analysis. These services often disappear a few years after publication and render results irreproducible [2]. With oqtans, bioinformatics becomes reproducible by providing analysis building blocks for a customized workflow of read mapping, transcript reconstruction and quantitation as well as differential expression analysis.

Method

Oqtans includes a comprehensive machine-learning-powered toolsuite developed by the authors for NGS data analysis. PALMapper is a short-read mapper which efficiently computes both unspliced and spliced alignments at high accuracy by taking advantage of base quality information and computational splice site predictions [3]. mTIM is a transcript reconstruction method, which exploits features derived from RNA-seq read alignments and from computational splice site predictions to infer the exon-intron structure of the corresponding transcripts. rQuant is based on quadratic programming. It simultaneously estimates biases

inherent in library preparation, sequencing, and read mapping, and accurately determines the abundances of given transcripts [4]. rDiff is a set of statistical test techniques that determine significant differences between two RNA-seq experiments to find differentially expressed regions with or without knowledge of transcripts.

Results

We compare predictions to the published annotation at the intron and transcript levels. The performance of read aligners is shown in Fig. 1A from *D. melanogaster* data, and transcript segmentation tools in Fig. 1B, on *C. elegans*. Our tools, PALMapper and mTIM, outperform TopHat [5] and Cufflinks [6]. Oqtans is available free and open-source, from <http://oqtans.org> as a virtual machine for cloud computing environments, and ready

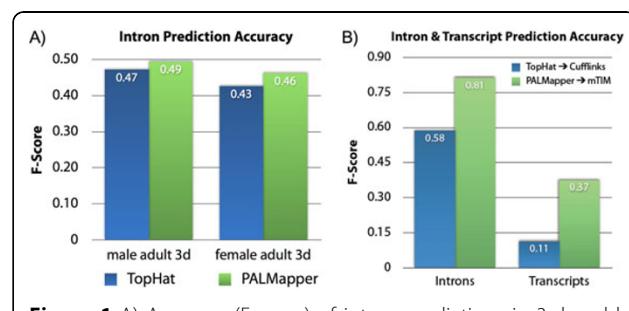


Figure 1 A) Accuracy (F-score) of intron predictions in 3-day-old adults of *D. melanogaster* with aligners PALMapper (green) and TopHat (blue). B) Accuracy of intron predictions with the same aligners and transcript predictions with mTIM (green) and Cufflinks (blue) on *C. elegans* RNA-seq transcriptome data.

* Correspondence: sebastian@bioweb.me

¹Machine Learning in Biology Group, Friedrich Miescher Laboratory of the Max Planck Society, 72076 Tübingen, Germany

Full list of author information is available at the end of the article

to use on our public compute cluster at <http://bioweb.me/mlb-galaxy>.

Author details

¹Machine Learning in Biology Group, Friedrich Miescher Laboratory of the Max Planck Society, 72076 Tübingen, Germany. ²Department of Software Engineering and Theoretical Computer Science, Technical University Berlin, 10578 Berlin, Germany. ³Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany.

Published: 21 November 2011

References

1. Goecks J, Nekrutenko A, Taylor J: Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology* 2010, 11(8):R86.
2. Schultheiss SJ, Münch MC, Andreeva GD, Rätsch G: Persistence and Availability of Web Services in Computational Biology. *PLoS computational biology* 2011, 6(9):e24914.
3. Jean G, Kahles A, Sreedharan VT, De Bona F, Ratsch G: RNA-Seq read alignments with PALMapper. In *Current protocols in bioinformatics* Andreas D Baxevanis [et al] 2010, Chapter 11:Unit 11.16.
4. Bohnert R, Ratsch G: rQuant.web: a tool for RNA-Seq-based transcript quantitation. *Nucleic acids research* 2010, 38(Web Server):W348-351.
5. Trapnell C, Pachter L, Salzberg SL: TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009, 25(9):1105-1111.
6. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L: Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome biology* 2011, 12(3):R22.

doi:10.1186/1471-2105-12-S11-A7

Cite this article as: Schultheiss et al.: Oqtans: a Galaxy-integrated workflow for quantitative transcriptome analysis from NGS Data. *BMC Bioinformatics* 2011 12(Suppl 11):A7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

