

oqtans: A Galaxy-Integrated Workflow for Quantitative Transcriptome Analysis from NGS Data

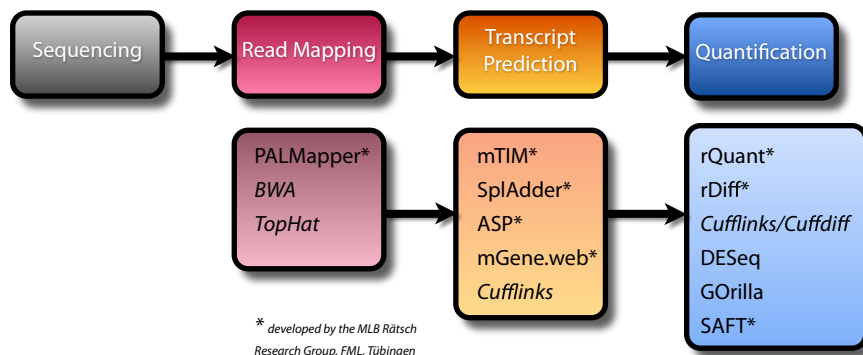
The current revolution in sequencing technologies allows us to obtain a much more detailed picture of transcriptomes via RNA-Sequencing. We have developed the first integrative online platform, **oqtans**, for quantitatively analyzing RNA-Seq experiments. It is based on the Galaxy-framework and provides tools for read mapping, transcript reconstruction and quantitation

as well as differential expression analysis.

Keywords: *next-generation sequencing, transcriptome, RNA-Seq, Galaxy, machine learning, reproducibility, flexibility, cloud computing, analysis pipeline, workflow customization, simplicity, read quantification, transcript prediction, short-read alignment, differential expression analysis*

Infinite possibilities with oqtans Alignment, Annotation, Quantification

NGS data arrives in Galaxy from the sequencing facility and the common transcriptome analysis from RNA-seq is performed in three steps: **Read Alignments, Transcript Prediction and Quantification**. **oqtans** provides a significant number of tools for each of these categories, which are all compatible with tools from other steps, making it possible to design infinitely flexible workflows for quantitatively analyzing RNA-seq experiments.



Our oqtans tools

Oqtans includes a comprehensive machine-learning-powered toolsuite developed on the Max Planck Campus in Tübingen (Germany) 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
- **mTIM** is a transcript reconstruction method, which exploits features derived from RNA-seq read align-

ments 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
- **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.



oqtans.org

Machine-learning-powered tools integrated into Galaxy

<http://oqtans.org>
oqtans@tue.mpg.de
<http://galaxy.fml.mpg.de>

oqtans: Quantitative RNA-Seq Analysis

- Short Read Alignment
- Transcript Identification
- Transcript Quantification
- Differential Expression

oqtans Developers

G. Jean, S. J. Schultheiss,
J. Behr, P. Drewe,
N. Goernitz, A. Kahles,
P. Mudrakarta,
V.T. Sreedharan, G. Zeller,
G. Rättsch

oqtans.org



oqtans.org

Machine-learning-powered tools integrated into Galaxy

<http://oqtans.org>
oqtans@tue.mpg.de
<http://galaxy.fml.mpg.de>

oqtans: Quantitative RNA-Seq Analysis

- Short Read Alignment
- Transcript Identification
- Transcript Quantification
- Differential Expression

oqtans Developers

G. Jean, S. J. Schultheiss, J. Behr, P. Drewe, N. Goernitz, A. Kahles, P. Mudrakarta, V.T. Sreedharan, G. Zeller, G. Rätsch

oqtans.org

oqtans Availability

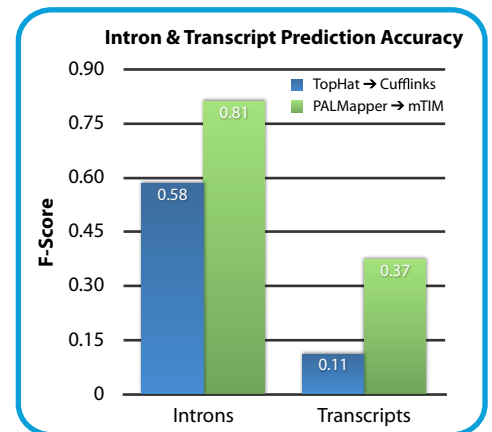
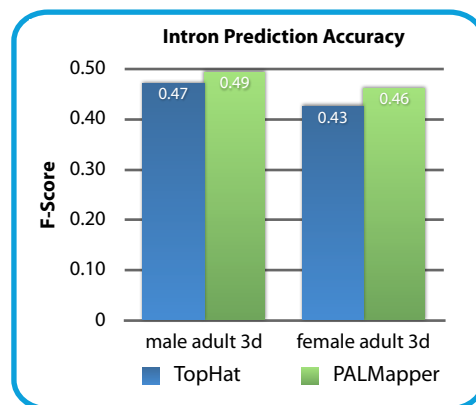
Install our free, open-source tools from released packages at <http://oqtans.org>

Install tools on any Galaxy instance with Python Fabric, downloadable from the Galaxy Community Tool Shed at <http://community.g2.bx.psu.edu>

Give **oqtans** a no-strings-attached trial run on our public instance on a 168 core compute cluster at <http://galaxy.fml.mpg.de>

Performance Comparison

Prediction evaluation comparing to the published annotation, at the intron or transcript level. We show performance of read aligners in the first panel on data from *D. melanogaster*, and transcript segmentation tools in the second panel, on *C. elegans*. Our tools, PALMapper and mTIM outperform the competition.



Reproducibility and Web Service Availability

It is difficult to maintain web services after publication [1]. Our approach of providing a self-contained machine image with the accessible, transparent Galaxy framework [2] minimizes this risk and leads to reproducible analyses in bioinformatics. All the data and how it was produced is kept in a file that will remain accessible and readable for anyone you give it to. The difficulty of installing

Run your own cluster with all **oqtans** tools readily installed and at your disposal as an AMI (Amazon Machine Image) in the Amazon Elastic Compute Cloud. Search for **oqtans** in the AMIs.

We keep an example instance of oqtans running in the cloud. It is reachable via a link at <http://oqtans.org/>

AMI can be downloaded and launched in your own virtualizer (e.g. VirtualBox, Parallels, ...)

and maintaining your own instance of the service is also mitigated. **oqtans** makes science more credible.

Tool References:

PALMapper: G. Jean et al. 2010 *Curr Prot Bioinf*
TopHat: C. Trapnell et al. 2010 *Bioinformatics*
mTIM & SplAdder: G. Zeller et al. 2011 i.p.
Cufflinks: A. Roberts et al. 2011 *Genome Biol*
rQuant: R. Bohnert et al. 2010 *Nucleic Acids Res*
rDiff: O. Stegle et al. 2010 *Nature Preceedings*
DESeq: S. Anders & W. Huber 2011 *Genome Biol*
SAFT: A. Kahles et al. 2011 *Bioinformatics* i.p.
GORilla: E. Eden et al. 2009 *BMC Bioinformatics*
EasySVM: S. Sonnenburg et al. 2010 *JMLR*
ASP: S. Sonnenburg et al. 2009 *BMC Bioinformatics*
KIRMES: S.J. Schultheiss et al. 2009 *Bioinformatics*
GFF Toolkit: V.T. Sreedharan et al. 2011 *BOSC SIG*
mGene.web: G. Schweikert et al. 2009 *NAR*
BWA: Li et al. 2009 *Bioinformatics*

References:

[1] S.J. Schultheiss et al. 2011 *PLoS Comp Biol*
[2] J. Goecks et al. 2010 *Genome Biol*