RNA-Seq Blog Poll Results

What is the greatest immediate need facing the RNA Sequencing community?

- More sequencing capacity: 1
- Further advancement in sequencing technology: 9
- Continued cost reduction: 11
- Better library preparation protocols: 23
- Skilled bioinformatics specialists: 44
- Standardized data analysis pipeline: 55
Reproducible Quantitative Transcriptome Analysis with oqtans

Vipin T. Sreedharan<vipin@cbio.mskcc.org>

Rätsch Laboratory, Computational Biology Center
Memorial Sloan-Kettering Cancer Center, USA
University of Tübingen, Germany

14th Bioinformatics Open Source Conference, 19-20 July 2013 at Berlin
**oqtans**: online quantitative transcriptome analysis

- Sample sequencing
- Read alignment
- Quality filter
- Transcript identification
- Transcript quantification
- Differential expression test
- Enrichment analysis
**oqtans:** online quantitative transcriptome analysis

- Sample sequencing
- Read alignment
- Quality filter
- Transcript identification
- Transcript quantification
- Differential expression test
- Enrichment analysis

Tools:
- PALMapper*
- BWA
- TopHat
- STAR
- RNA-geeq*
- mTIM*
- ASP*
- Trinity
- Cufflinks
- Scripture
- rQuant*
- rDiff*
- DESeq
- DEXSeq
- Cuffdiff
- edgeR
- mTIN
- ASP
- Trinity
- Cufflinks
- Scripture
- rQuant
- topGO
- GeneSetter*
- mGene*
- KIRMES*
- ARTS*
- EasySVM*

*developed by Rätsch lab members from cBio MSKCC and FML of the Max Planck Society

Vipin @ Rätsch Laboratory

BOSC 2013

July 20, 2013
## Timeline: Key events in the history of oqtans

- Online since December 2009.
Timeline: Key events in the history of oqtans

- Online since December 2009.

- 2009 Dec: Transcript quantification (rQuant)
- 2010 Mar: Fusion of GenomeMapper & QPalma (PALMapper)
- 2010 Jul: Differential expression testing (rDiff, DESeq)
- 2010 Nov: Transcript predictor (mTIM)
- 2011 Feb: oqtans Amazon Machine Image on AWS
- 2011 May: oqtans public cloud instance started operating on AWS using Galaxy-CloudMan
- 2012 Jan: Alignment quality filter (RNA-geeq)
- 2012 Nov: Deployed the instance at a local data center
- 2013 Jun: A vast assortment of RNA-seq data analysis tools
Timeline: Key events in the history of oqtans

- Online since December 2009.

- Transcript quantification rQuant 2009 Dec

- Differential expression testing rDiff, DESeq 2010 Jul

- Fusion of GenomeMapper & QPalma PALMapper 2010 Mar

- oqtans Amazon Machine Image on AWS 2011 Feb

- Alignment quality filter RNA-geeq 2012 Jan

- Deployed the instance at a local data center 2013 Jun

- 593 registered and 336 active users.

- 12 tools developed from our lab.

- 15 tools from other open-source software development community.
oqtans: Transcriptome analysis toolsuite

**PALMapper**: highly accurate, variation-aware sequencing read mapper using base quality and splice site predictions.¹

**oqtans**: Transcriptome analysis toolsuite

- **RNA-geeQ**: alignment optimization and postprocessing toolbox.\(^2\)
  - SPACE - Alignment evaluation and comparison.
  - SAFT - Simple alignment filtering.
  - MMR - Multiple mapper resolution.

---

**oqtans**: Transcriptome analysis toolsuite

- **mTiM**: reconstructs exon-intron structure from read alignments and splice site predictions.³

---

**oqtans**: Transcriptome analysis toolsuite

- **rDiff**: accurate detection of differential RNA processing from RNA-seq data.\(^4\)
  - Parametric test (complete genome annotation).
  - Non Parametric test (incomplete genome annotation).

---

**oqtans**: Transcriptome analysis toolsuite

- **rQuant**: estimates bias in library preparations, sequencing and read mapping; accurately determines the abundance of transcripts.\(^5\)

---

Workflow

Sequencing experiments to biological insight

- Compute the fraction of unexpressed, expressed and differentially expressed family of genes from two strains of *A. thaliana*.

- Illumina, 78 nt RNA-seq reads
  - Columbia accession (Col-0) 13 million reads.
  - Canary Island accession (Can-0) 11.8 million reads.

- Two replicates per accession.\(^6\)

**oqtans on Amazon cloud platform**

- **Compute resources**: m3.2xlarge.
  - Alignments: 20 minutes.
  - Quantitative analysis: 10 minutes.
- **Cost on Amazon EC2**: $3.82
**oqtans: Tool exchangeability**

- **General steps:**
  - Sample sequencing
  - Read alignment
  - Quality filter
  - Differential expression test
  - Enrichment analysis

- **Different paths in workflow:**
  - Illumina run
  - PALMapper
  - TopHat
  - RNA-geeq
  - rDiff
  - edgeR
  - STAR
  - DESeq
  - GeneSetter
oqtans Availability

- Our public Galaxy instance.
  - http://galaxy.cbio.mskcc.org

- Public git repository.
  - http://github.com/ratschlab/oqtans

- Machine Image on Amazon Web Service.
    Search for "ami-5e389a37"

- Galaxy Tool Shed.
  - http://toolshed.g2.bx.psu.edu/
Summary

- oqtans simplifies the RNA-seq data analysis workflow.
- Easy to instantiate in
  - Cloud service platforms
  - Existing Galaxy installation
  - Command line interface
http://oqtans.org

Would love to hear your experience!