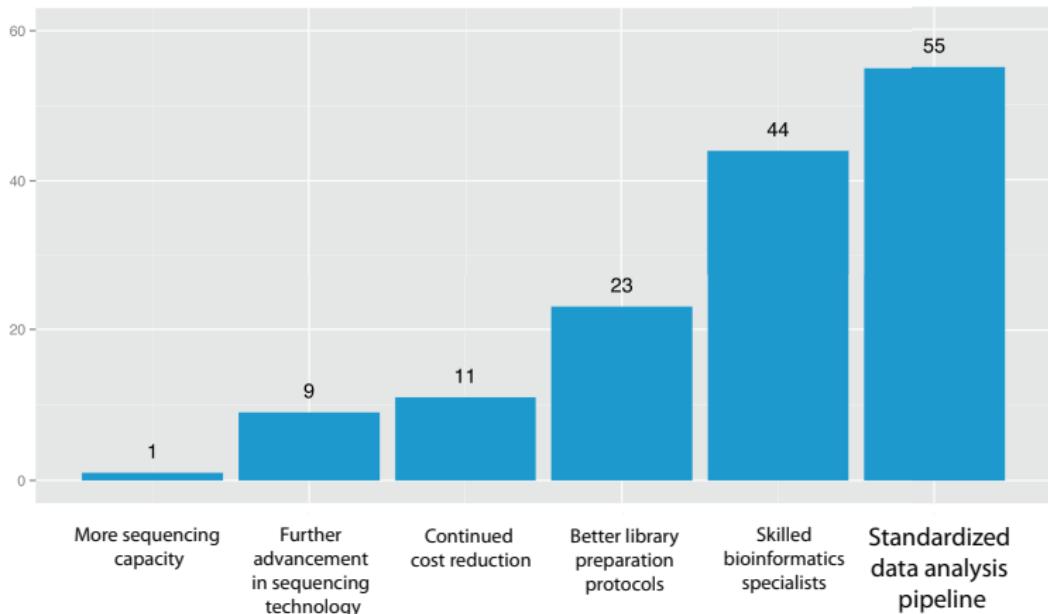


# RNA-Seq Blog Poll Results

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



# Reproducible Quantitative Transcriptome Analysis with oqtans

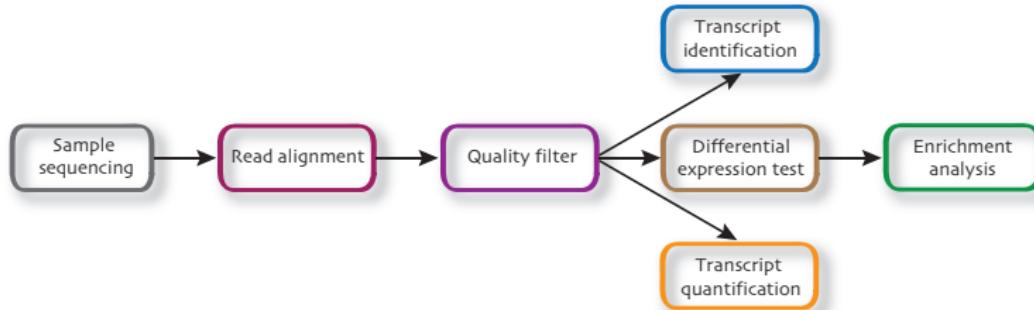
Vipin T. Sreedharan<vipin@cbio.mskcc.org>

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

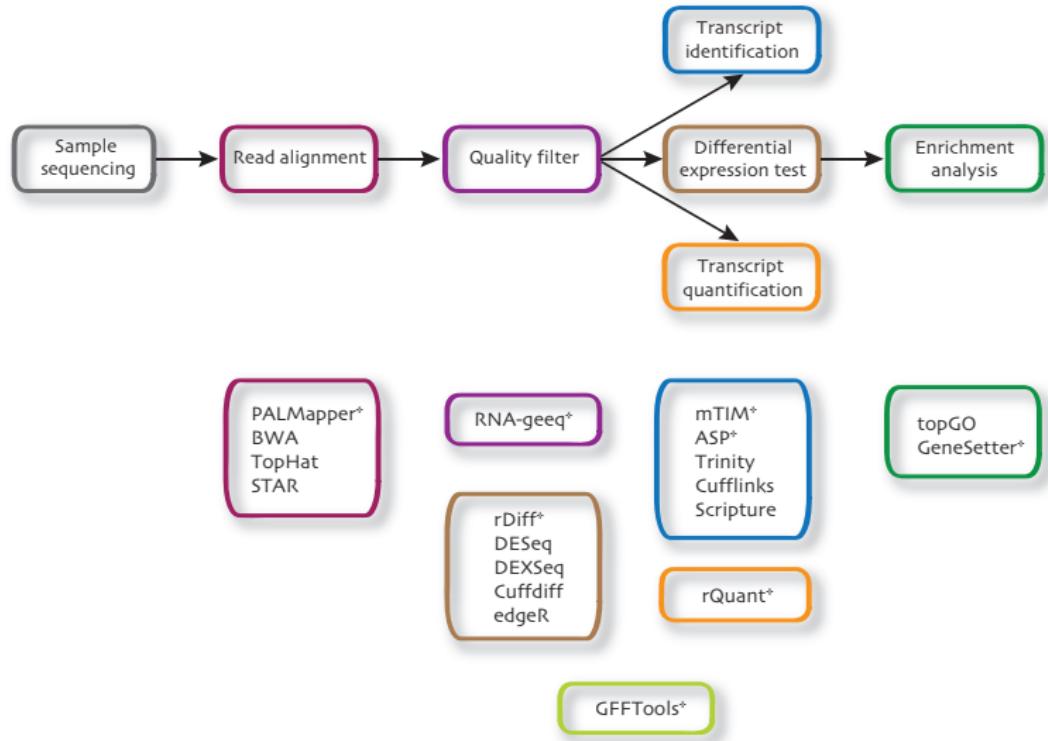
14<sup>th</sup> Bioinformatics Open Source Conference, 19-20 July 2013 at Berlin



# oqtans: online quantitative transcriptome analysis



# oqtans: online quantitative transcriptome analysis



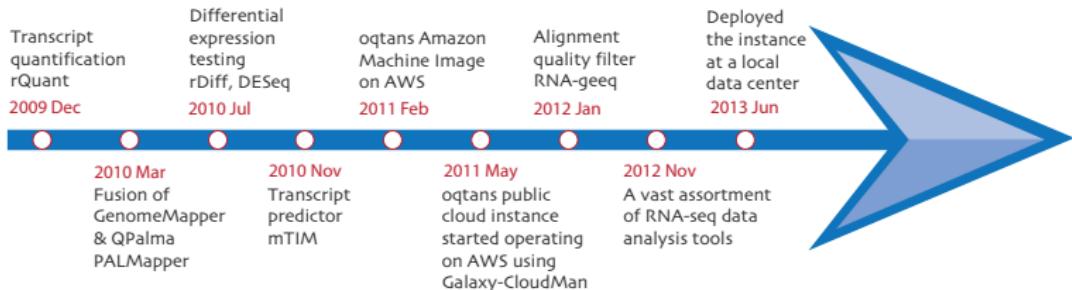
\*developed by Rätsch lab members from cBio MSKCC and FML of the Max Planck Society

# Timeline: Key events in the history of oqtans

- Online since December 2009.

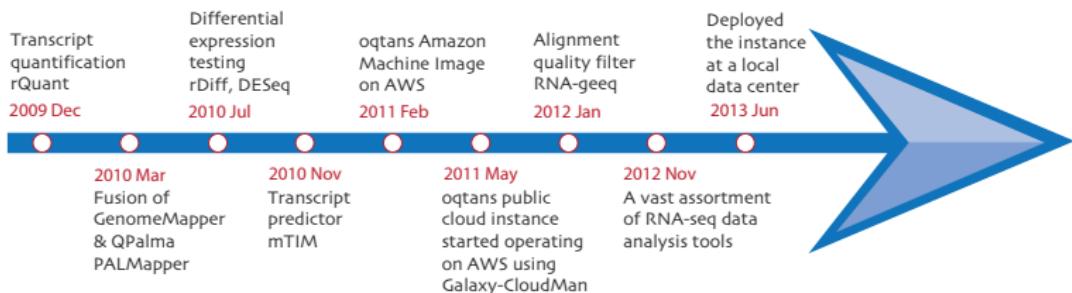
# Timeline: Key events in the history of oqtans

- Online since December 2009.



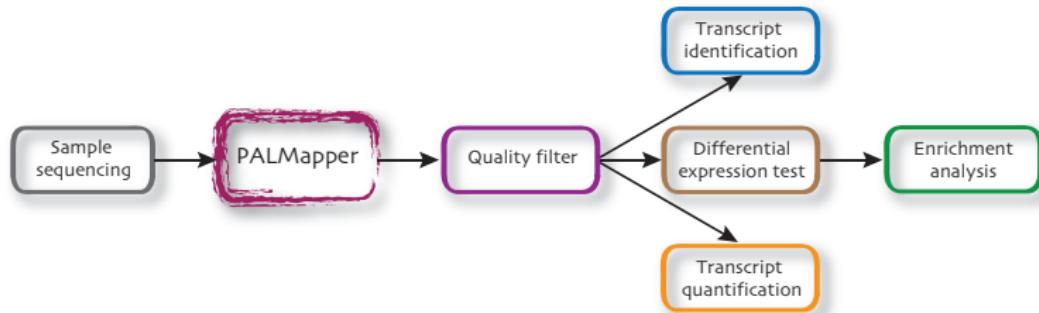
# Timeline: Key events in the history of oqtans

- Online since December 2009.



- 593 registered and 336 active users.
- 12 tools developed from our lab.
- 15 tools from other open-source software development community.

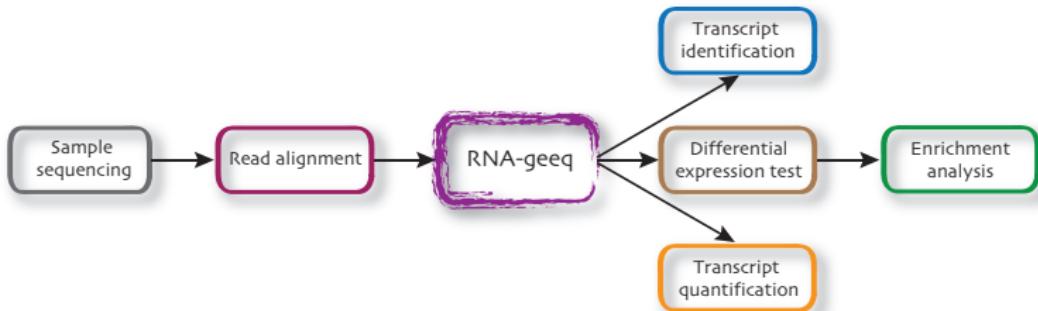
# oqtans: Transcriptome analysis toolsuite



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

<sup>1</sup>G Jean et al. *Curr Protoc Bioinformatics* (2010).

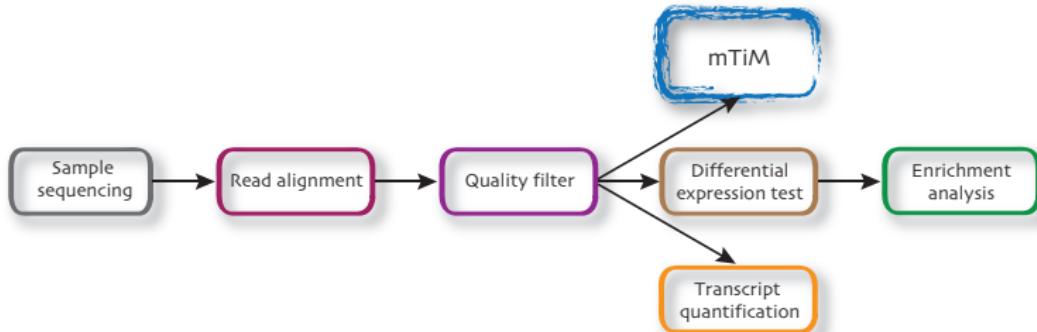
# oqtans: Transcriptome analysis toolsuite



- **RNA-geeq:** alignment optimization and postprocessing toolbox.<sup>2</sup>
  - SPACE - Alignment evaluation and comparison.
  - SAFT - Simple alignment filtering.
  - MMR - Multiple mapper resolution.

<sup>2</sup>Kahles A, Behr J, and G Rätsch. *In preparation* (2013).

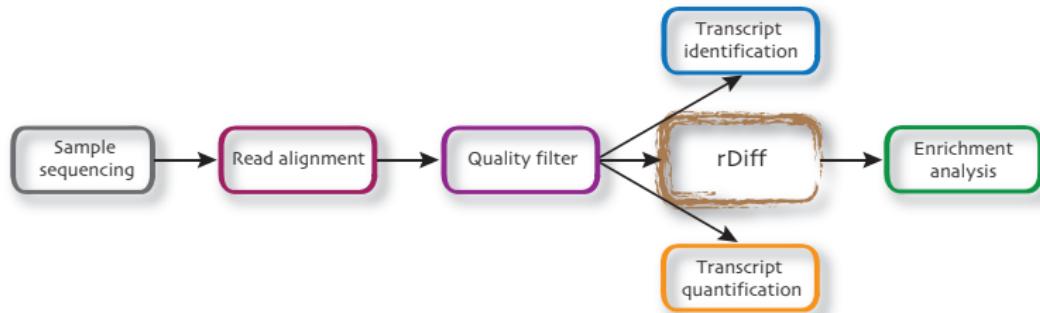
# oqtans: Transcriptome analysis toolsuite



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

<sup>3</sup>Görnitz N et al. *Neural Information Processing Systems* (2011).

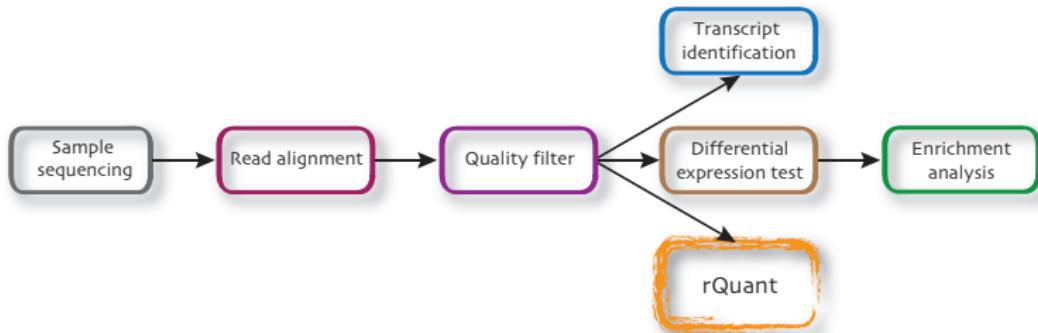
# oqtans: Transcriptome analysis toolsuite



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

<sup>4</sup>P Drewe et al. *Nucleic Acids Res* (2013).

# oqtans: Transcriptome analysis toolsuite



- **rQuant:** estimates bias in library preparations, sequencing and read mapping; accurately determines the abundance of transcripts.<sup>5</sup>

<sup>5</sup>R Bohnert and G Rätsch. *Nucleic Acids Res* (2010).

# 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.<sup>6</sup>

<sup>6</sup>X Gan et al. *Nature* (2011).

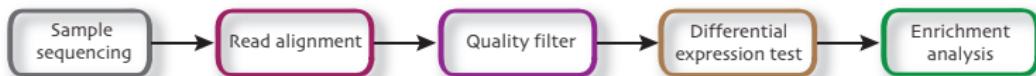
# oqtans on Amazon cloud platform



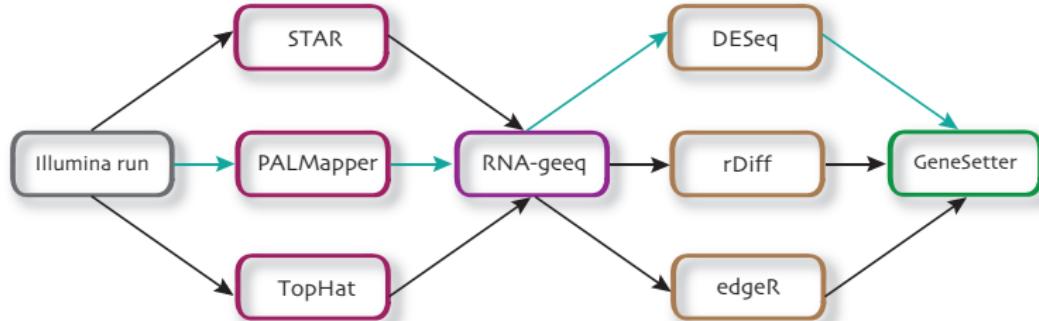
- Compute resources: m3.2xlarge.
  - Alignments: 20 minutes.
  - Quantitative analysis: 10 minutes.
- Cost on Amazon EC2: \$3.82

# oqtans: Tool exchangeability

- General steps:



- Different paths in workflow:



# 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.
  - <http://aws.amazon.com/amis/>  
Search for “**ami-5e389a37**”
- Galaxy Tool Shed.
  - <http://toolshed.g2.bx.psu.edu/>

The screenshot shows the Galaxy web interface with the title "Galaxy / Rätsch Lab". The main area is titled "Tools" with a search bar containing "search tools". Below the search bar is a list of available tools: "Get Data", "OQTANS (V0.1)", "SVM Toolbox (v0.3.3)", "Read Mapping", and "Transcript Prediction/Assembly".

The screenshot shows a GitHub repository page for "ratschlab/oqtans". The repository is marked as "PUBLIC". The page includes the repository name and a link to the GitHub homepage.



# Summary

- oqtans **simplifies** the RNA-seq data analysis workflow.
- Easy to **stantiate** in
  - Cloud service platforms
  - Existing Galaxy installation
  - Command line interface

The screenshot shows the CloudMan from Galaxy application interface. At the top, there are buttons for 'Terminate cluster', 'Add nodes', 'Remove nodes', and 'Access Galaxy'. Below this is a 'Status' section with the following information:

- Cluster name: default
- Disk status: 122M / 100 (2%)
- Worker status: Idle: 3 Available: 3 Requested: 4
- Service status: Applications: 4 Data: 4

To the right, there is a grid of 12 nodes, each represented by a colored square (green, grey, red) indicating its status. A detailed view of one node is shown on the right:

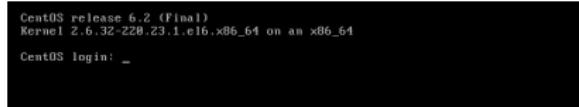
- ID: 000001-MT
- IP: 10.20.79.7.41
- State: Ready
- Architecture: i386
- Type: m1-medium
- Load: 0%



The screenshot shows the Galaxy web interface with the following navigation bar: Analyze Data, Workflow, Shared Data, Help, User. The main area displays a workflow titled "Running workflow 'Test workflow'" with two steps:

- Step 1: Input dataset**: First input is set to "1: galaxy-test-k.txt".
- Step 2: Input dataset**: Second input is set to "1: galaxy-test-k.txt" and "2: galaxy-test-h.txt".

On the left, there are links for Tools (Get Data, Text Manipulation, NGS Manipulation), Workflows, and a history list on the right.



# oqtans Team



Jonas Behr



Regina Bohnert



Philipp Drewe



Nico Görnitz



Géraldine Jean



André Kahles



David Kuo



Sebastian Schultheiss



Georg Zeller



Gunnar Rätsch

<http://oqtans.org>



Would love to hear your experience!