

Easier Workflows & Tool comparison with oqtans+

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

Rätsch Lab cBio, Memorial Sloan-Kettering Cancer Center, New York, USA.

MLB AG-Rätsch, FML of the Max Planck Society, Tübingen, Germany.

Center for Bioinformatics, University of Tübingen, Germany.

Galaxy Community Conference, 26 July 2012, Chicago



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



oqtans
online
quantitative
transcript
analysis



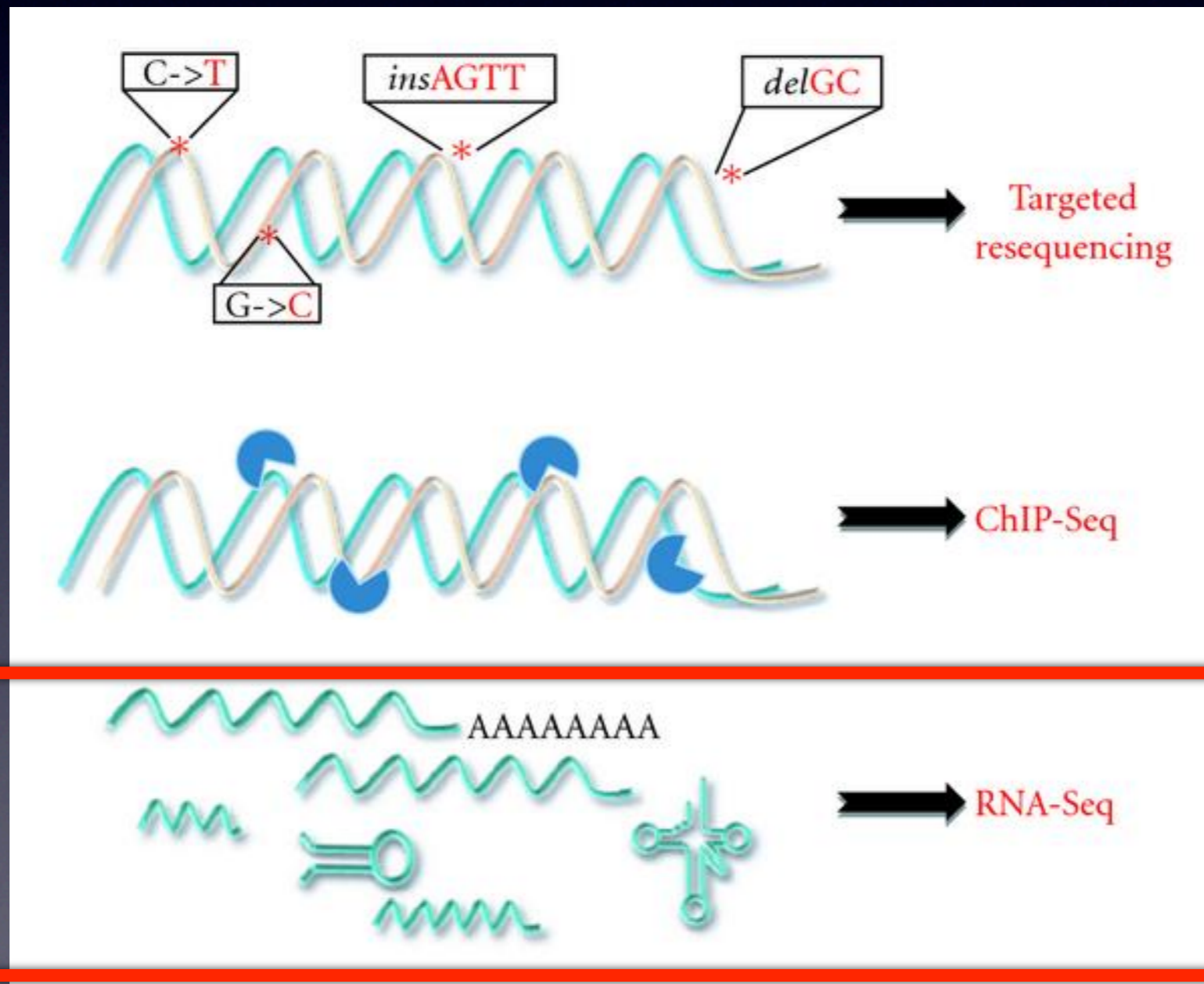
Deep sequencing data

- ▶ Specificities:
 - ▶ Rapidity of production
 - ▶ Low cost
 - ▶ Small fragments (reads)



Deep sequencing data

Applied to many scientific contexts:



oqtans

RNA-Seq Analysis

oqtans
online
quantitative
transcript
analysis



▶ Common analysis tasks

- ▶ compare two samples (wild type, mutant)
- ▶ identify new transcripts

Tang et al. 2011 *Nature Methods*
Grabherr et al. 2011 *Nature Biotech*
Li et al. 2011 *Science*
Gerstein et al. 2010 *Science*

Lee et al. 2011 *Nucleic Acids Res*
Yamashita et al. 2011 *Genome Res*
Daines et al. 2011 *Genome Res*
Ramani et al. 2010 *Genome Res*

RNA-Seq Analysis

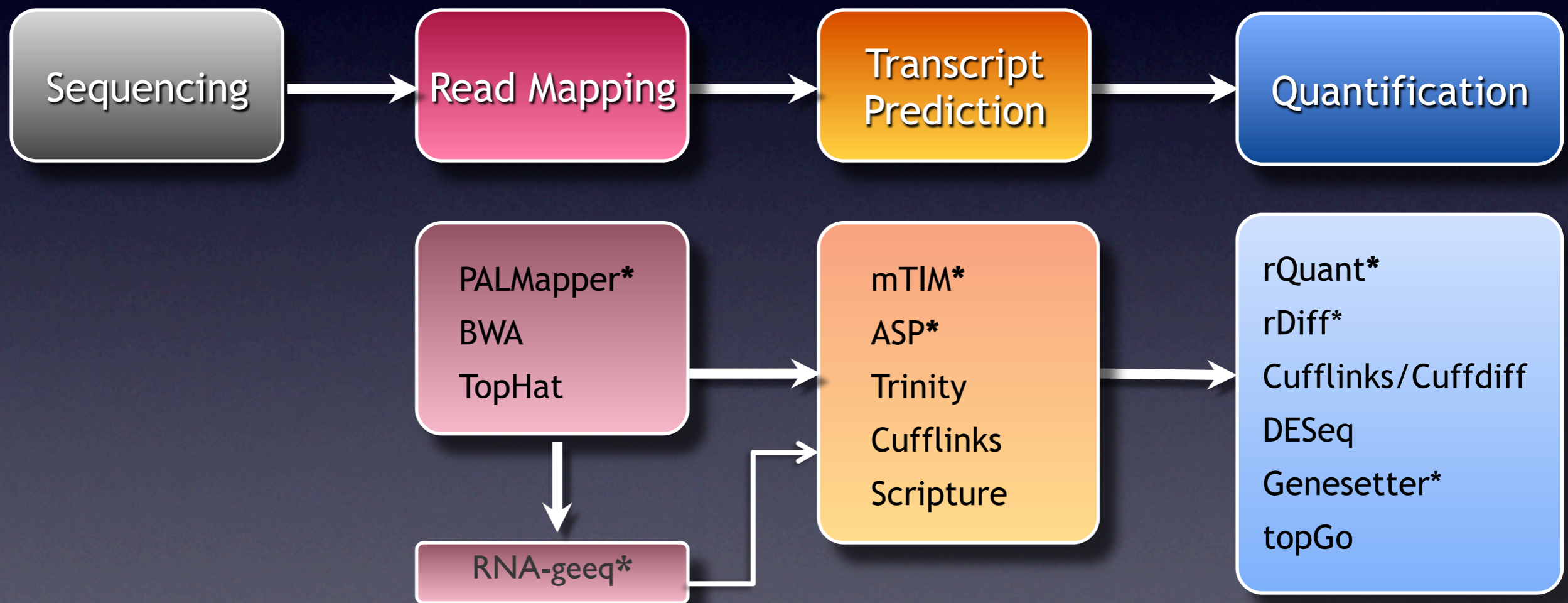
oqtans
online
quantitative
transcript
analysis



- ▶ Common issues
 - ▶ Reproducibility
 - ▶ Tool availability
 - ▶ Scalability

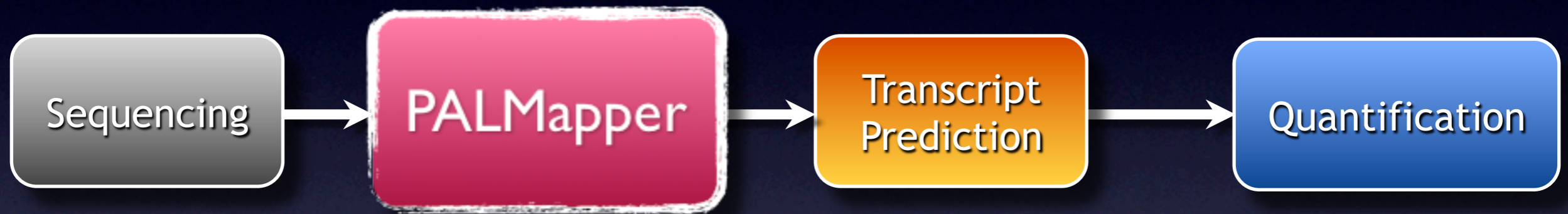
oqtans+ Galaxy Tools

oqtans
online
quantitative
transcript
analysis



oqtans+

Transcriptome analysis toolsuite



- ▶ **PALMapper**: highly accurate short-read mapper using base quality and splice site predictions

oqtans+

Transcriptome analysis toolsuite



- ▶ **mTIM:** reconstructs exon-intron structure from alignments and splice site predictions

J. Behr et al. 2011 *NIPS Machine learning in Comp Bio workshop*,
G. Zeller et al. 2012 i.p.

oqtans+

Transcriptome analysis toolsuite



- ▶ **rQuant:** estimates biases in library prep, sequencing, and read mapping; accurately determines the abundances of transcripts

oqtans+

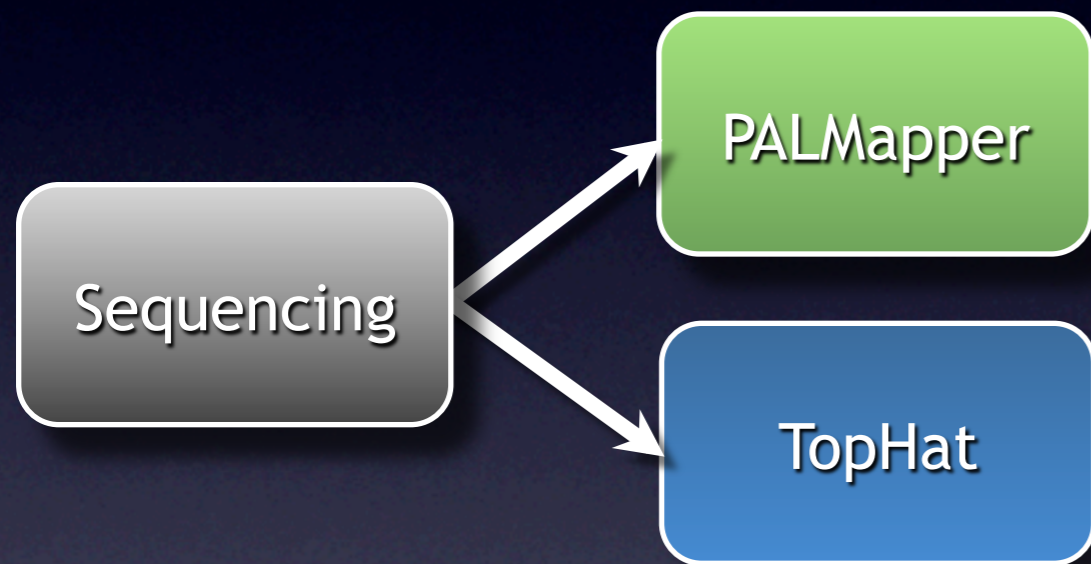
Transcriptome analysis toolsuite



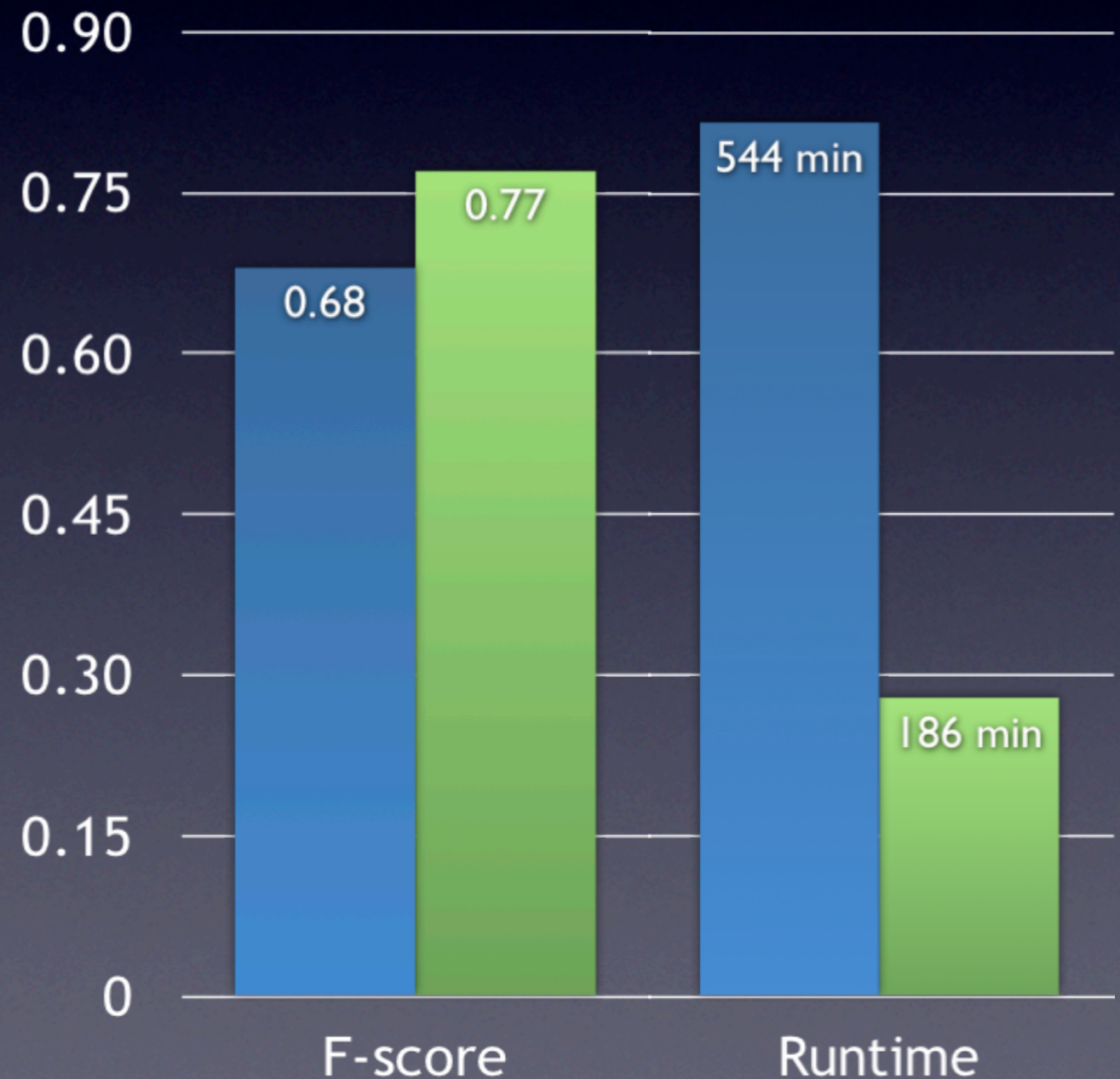
- ▶ **rDiff/DESeq**: determines significant differences in transcript/gene expression between experiments using statistical tests

O. Stegle et al. 2010 *Nature Preceedings*
S. Anders and W. Huber 2010 *Genome Biology*

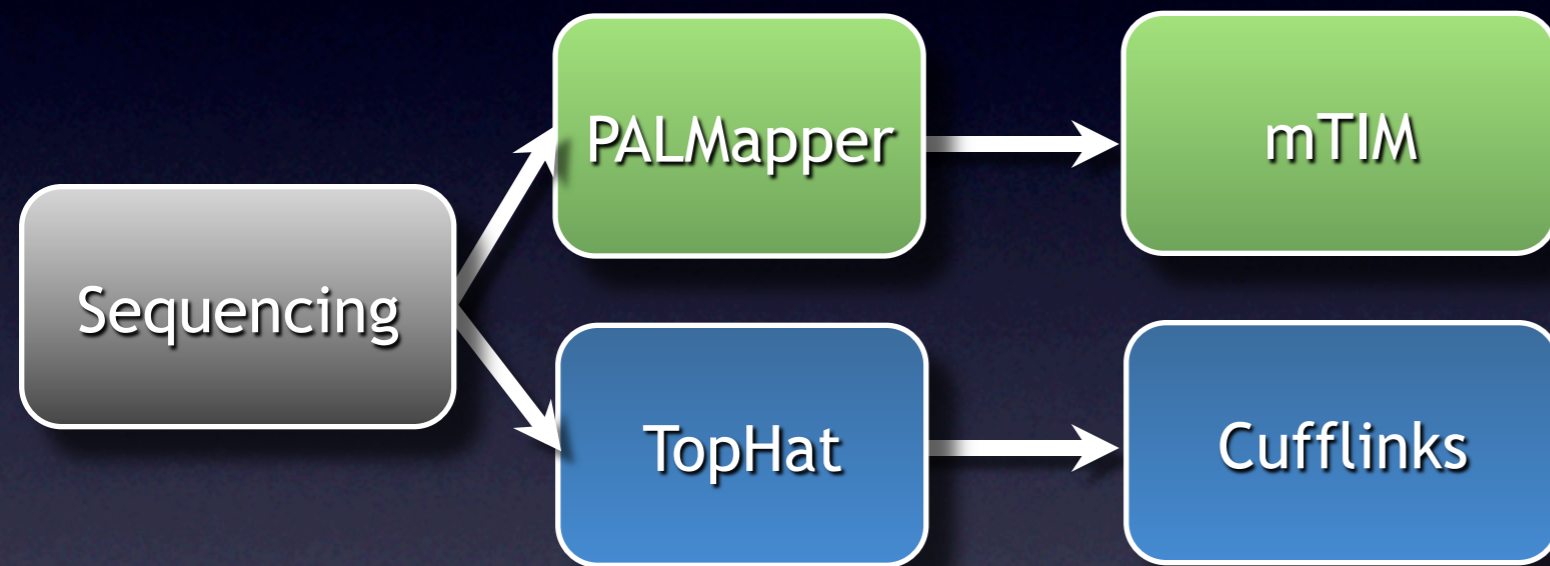
Accuracy of read alignments



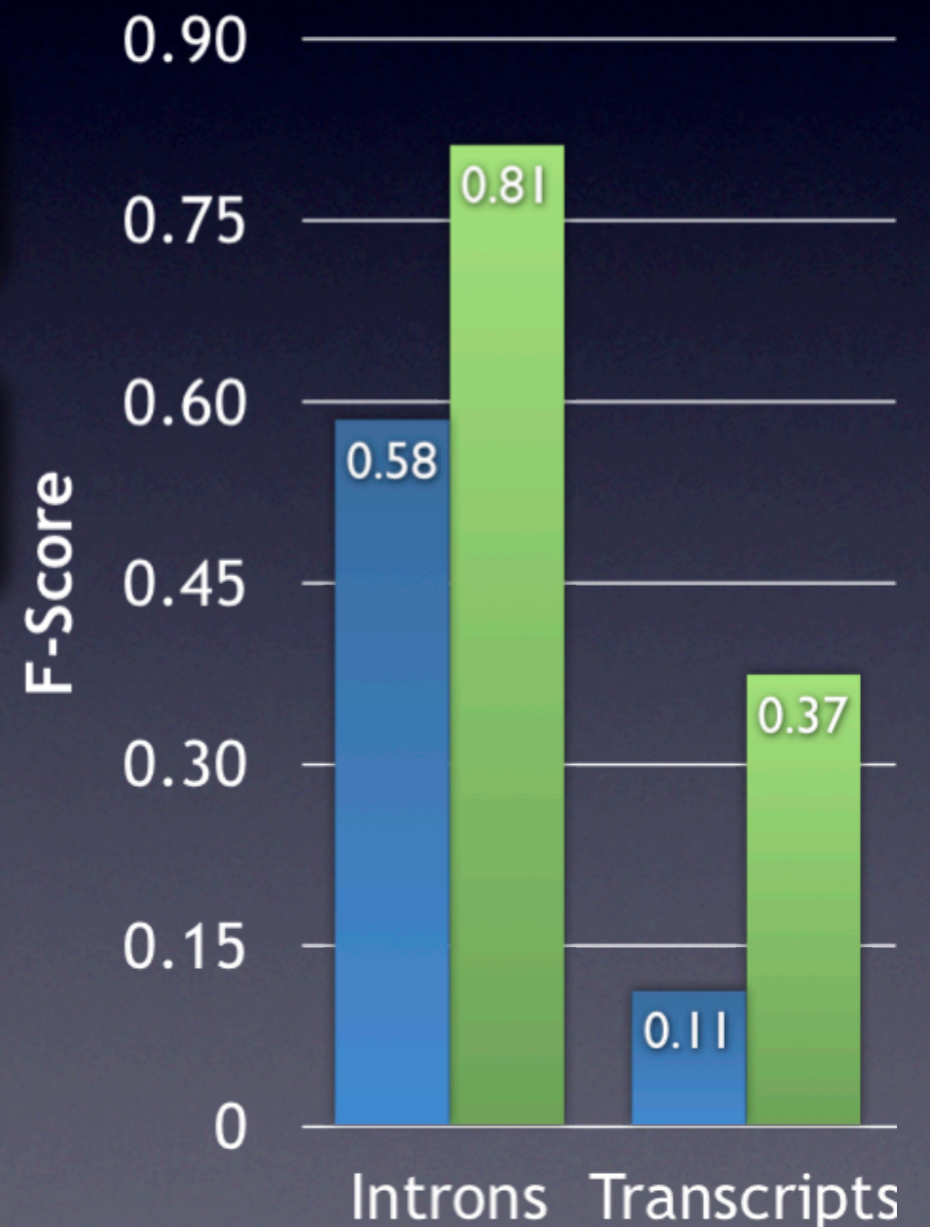
- ▶ *C. elegans*
- ▶ 75 nt RNA-seq reads (24 million)



Intron and transcript accuracy evaluation

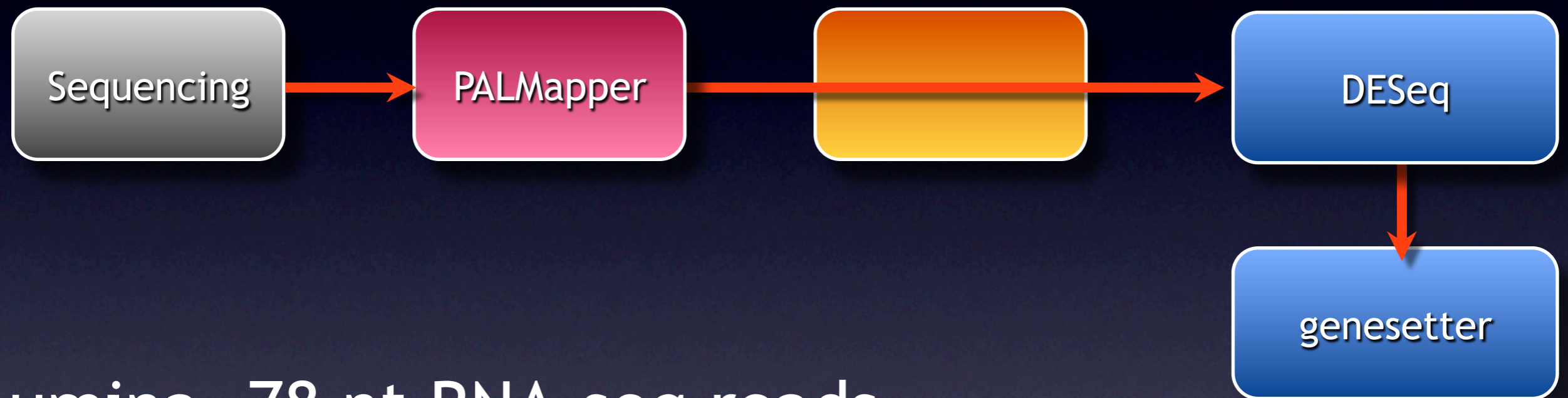


- ▶ *C. elegans*
- ▶ 75 nt RNA-seq reads (24 million)



oqtans+ Workflow

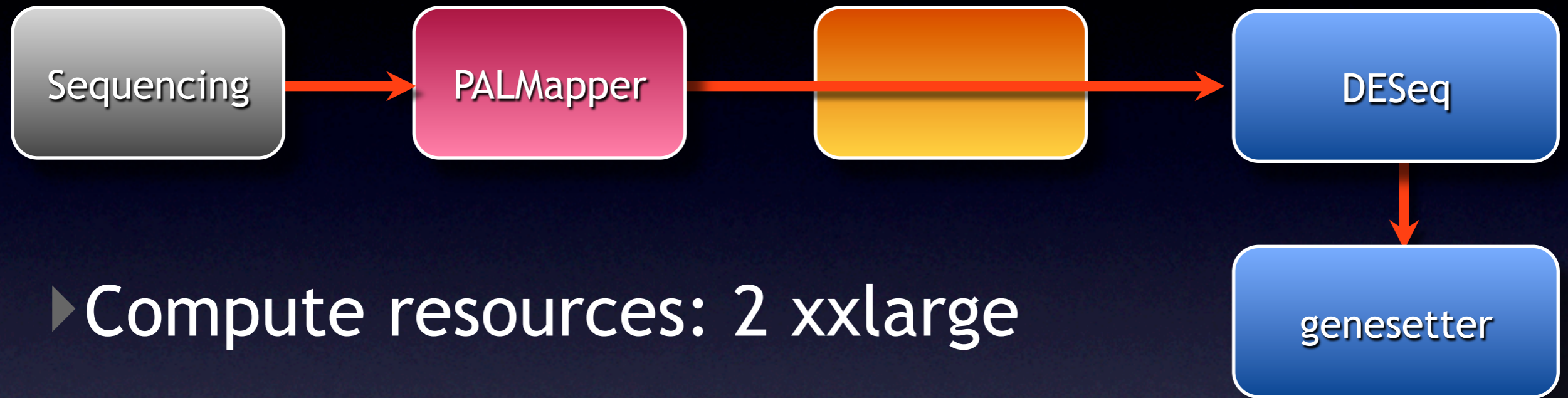
- ▶ Multiple reference genomes and transcriptomes for *A. thaliana*



- ▶ Illumina, 78 nt RNA-seq reads

- ▶ Columbia accession (Col-0) (1.2 million)
- ▶ Canary Island accession (Can-0) (4.9 million)

oqtans+ on AWS cloud



- ▶ Compute resources: 2 xxlarge

- ▶ Time:

- ▶ Alignments: 20 minutes

- ▶ Quantitative analysis: 10 minutes

- ▶ Cost on Amazon EC2: approx. \$2.82

oqtans+: Package contents

Read Mapping	Version
PALMapper	0.4
BWA	0.5.7
TopHat	1.5.0

Read Alignment Filtering	Version
SAFT	0.2
Multi-Mapper Resolution	0.1

Transcript Prediction	Version
mTIM	0.2
Cufflinks	1.3.0
Trinity	r2012-06-08
Scripture	Beta-2
ASP	0.3

Quantification	Version
rQuant	2.2
rDiff	0.2
Cufflinks/Cuffdiff	1.3.0
DESeq	1.6.1
topGO	0.1
Genesetter	0.1

oqtans+ Automated Tool Deployment

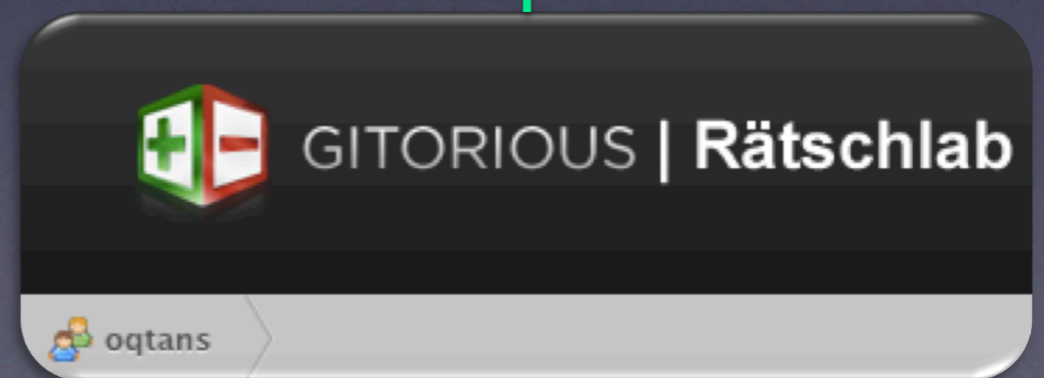
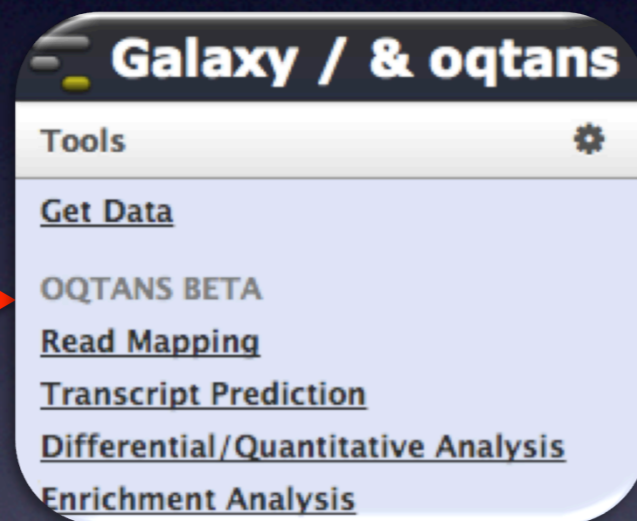
fabric script

```
from urllib import urlopen
from fabric.api import *

def get_latest_commit():
    return urlopen('http://commits.server.com/latest').read()

@task
def check():
    """check if local changes have been committed"""
    local_version = local('git rev-parse HEAD')
    if local_version != get_latest_commit():
        abort("!! Local changes haven't been committed !!")

@task
def deploy():
    """publish the latest version of the app"""
    with cd('/var/app'):
        run('git remote update')
        run('git checkout %s' % get_latest_commit())
    sudo("/etc/init.d/apache2 graceful")
```



How to resolve the requirements of OS specific packages ?

oqtans+ Availability: Our Server

- ▶ Public compute cluster
 - ▶ 12 nodes, 112 CPUs
 - ▶ our Galaxy test instance
 - ▶ All tools described here, and more!
- ▶ <http://bioweb.me/mlb-galaxy>

oqtans+ Availability: Source Code

- ▶ Free, open-source packages of our own tools
 - ▶ Including Galaxy Tool Wrappers
 - ▶ <http://oqtans.org>
- ▶ Fabric scripts to install on any Galaxy instance
- ▶ Community Tool Shed
<http://toolshed.g2.bx.psu.edu>

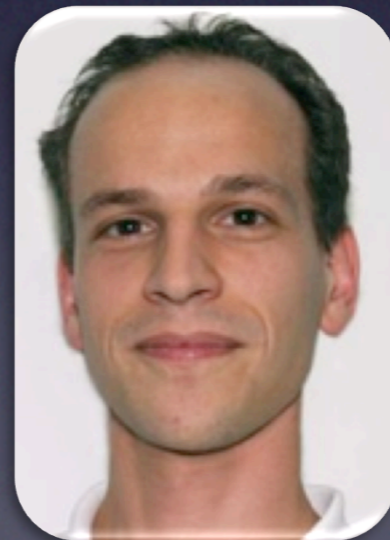
oqtans+ Availability: Cloud Computing

- ▶ Demo cloud instance with all oqtans+ tools
 - ▶ <http://cloud.oqtans.org>
- ▶ AMI at Amazon Web Services for EC2
 - ▶ Cloudman to launch any number of instances as a compute cluster

<http://oqtans.org>



Jonas Behr, Regina Bohnert, Philipp Drewe, Nico Görnitz, Géraldine Jean



André Kahles, Pramod Mudra, Sebastian Schultheiss, Georg Zeller, Gunnar Rätsch