

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



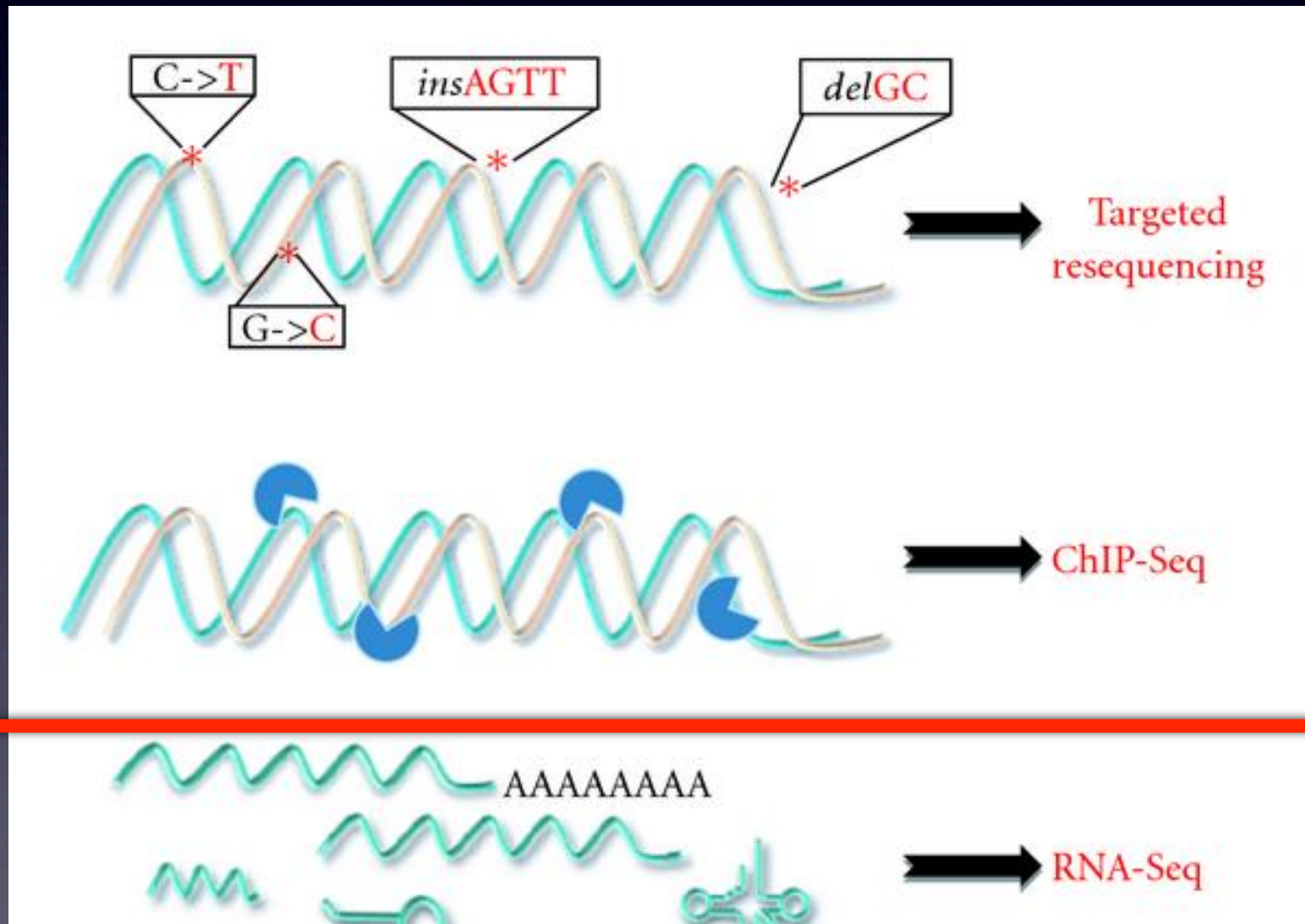
Deep sequencing data

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

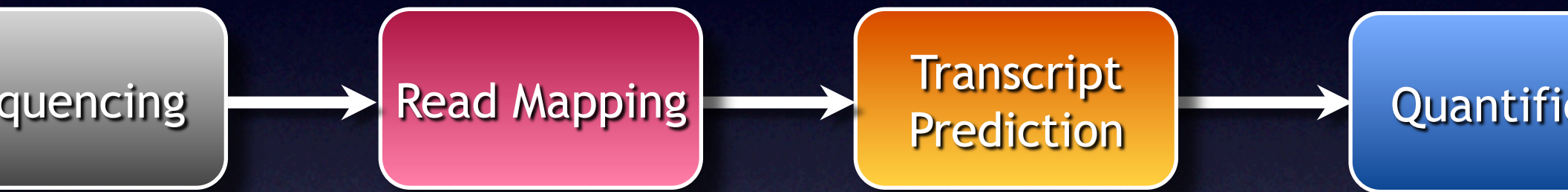
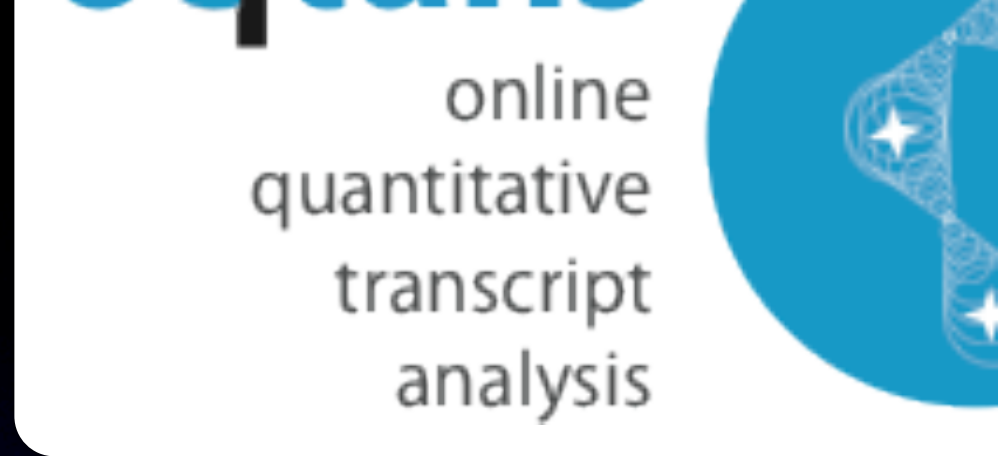


Deep sequencing data

Applied to many scientific contexts:



RNA-seq Analysis



Common analysis tasks

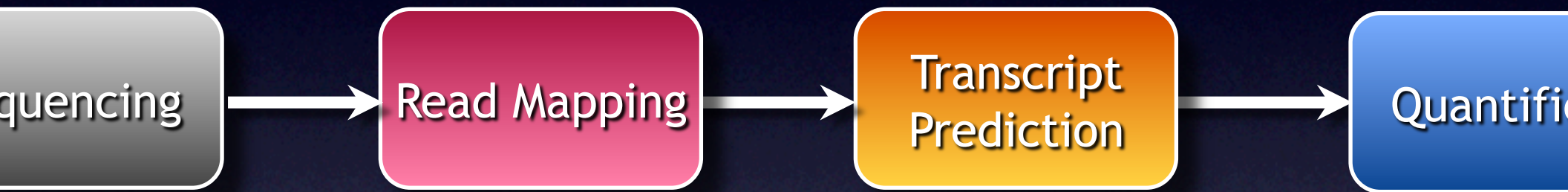
- ▶ compare two samples (wild type, mutant)
- ▶ identify new transcripts

Tang et al. 2011 *Nature Methods*
Grabherr et al. 2011 *Nature Biotech*

Lee et al. 2011 *Nucleic Acids*
Yamashita et al. 2011 *Genom*

RNA-seq Analysis

online
quantitative
transcript
analysis

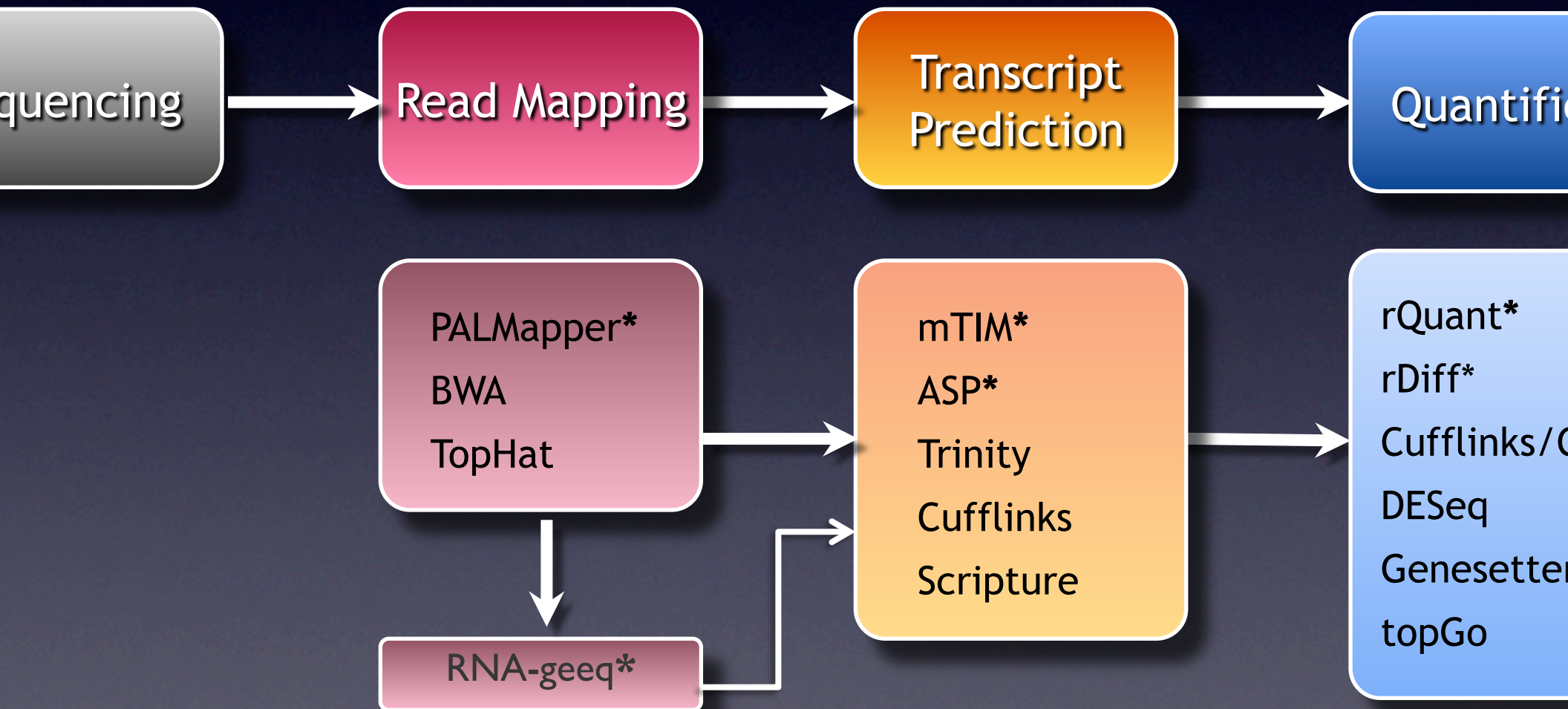


Common issues

- ▶ Reproducibility
- ▶ Tool availability
- ▶ Scalability

Quantis+ Galaxy Tools

online
quantitative
transcript
analysis



trans+

transcriptome analysis toolsuite



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

trans+

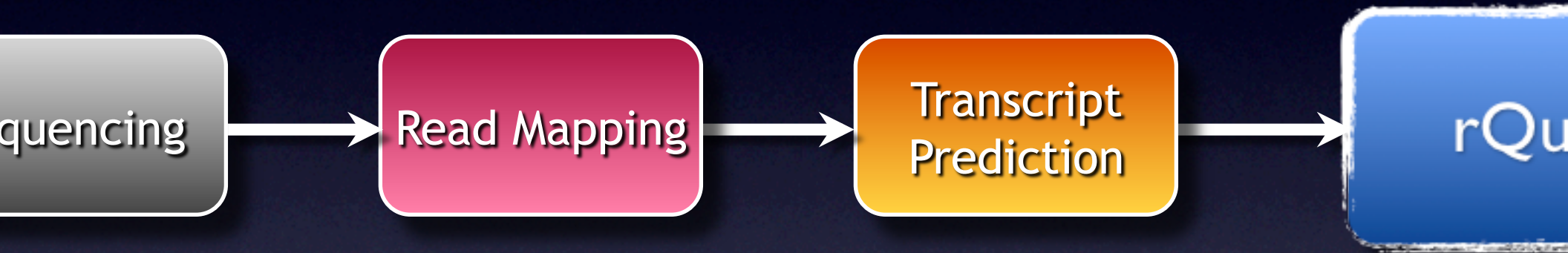
transcriptome analysis toolsuite



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

trans+

transcriptome analysis toolsuite



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

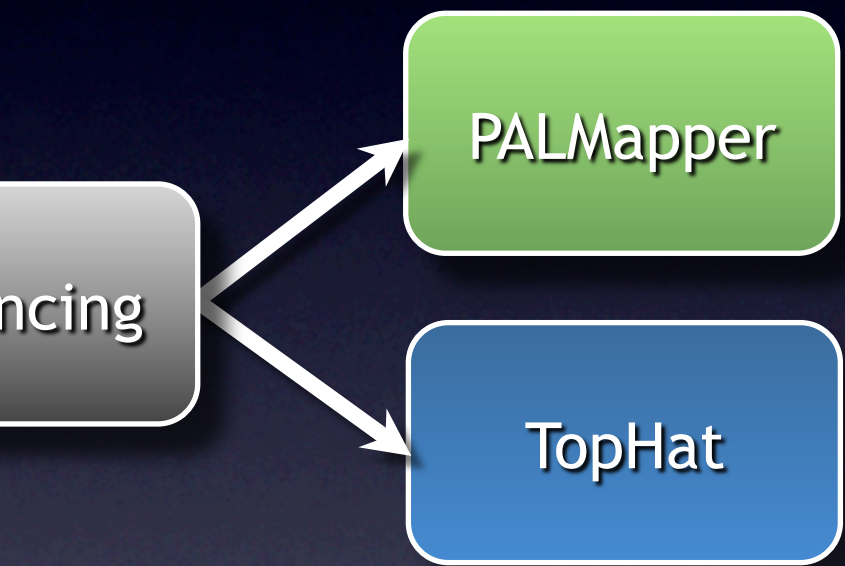
trans+

transcriptome analysis toolsuite



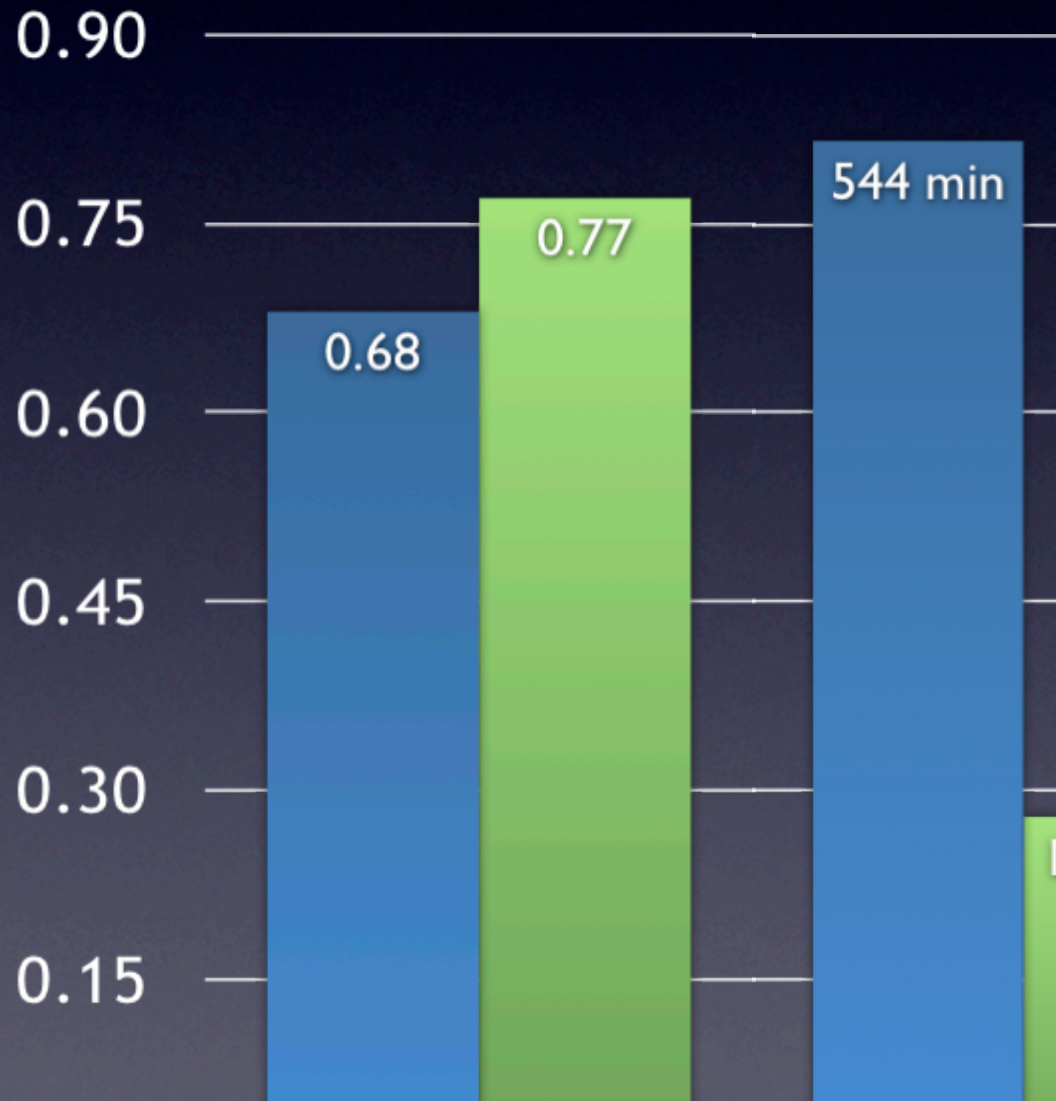
rDiff/DESeq: determines significant differences in transcript/gene expression between experiments using statistical tests.

Accuracy of read alignments

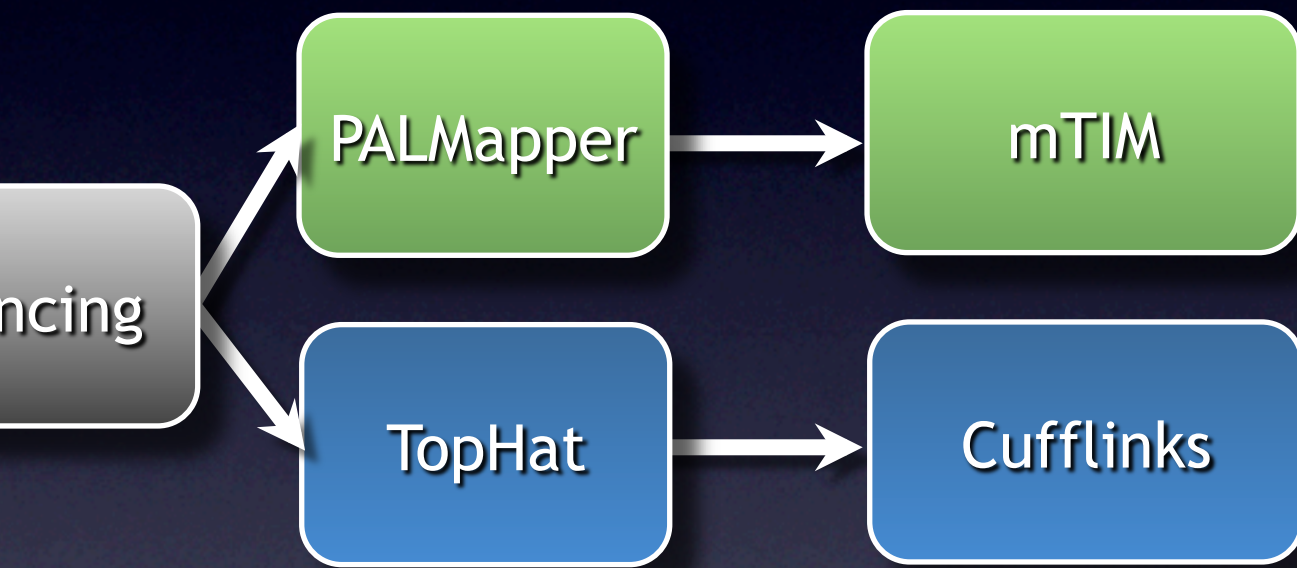


C. elegans

100 million RNA-seq reads
(10 million)

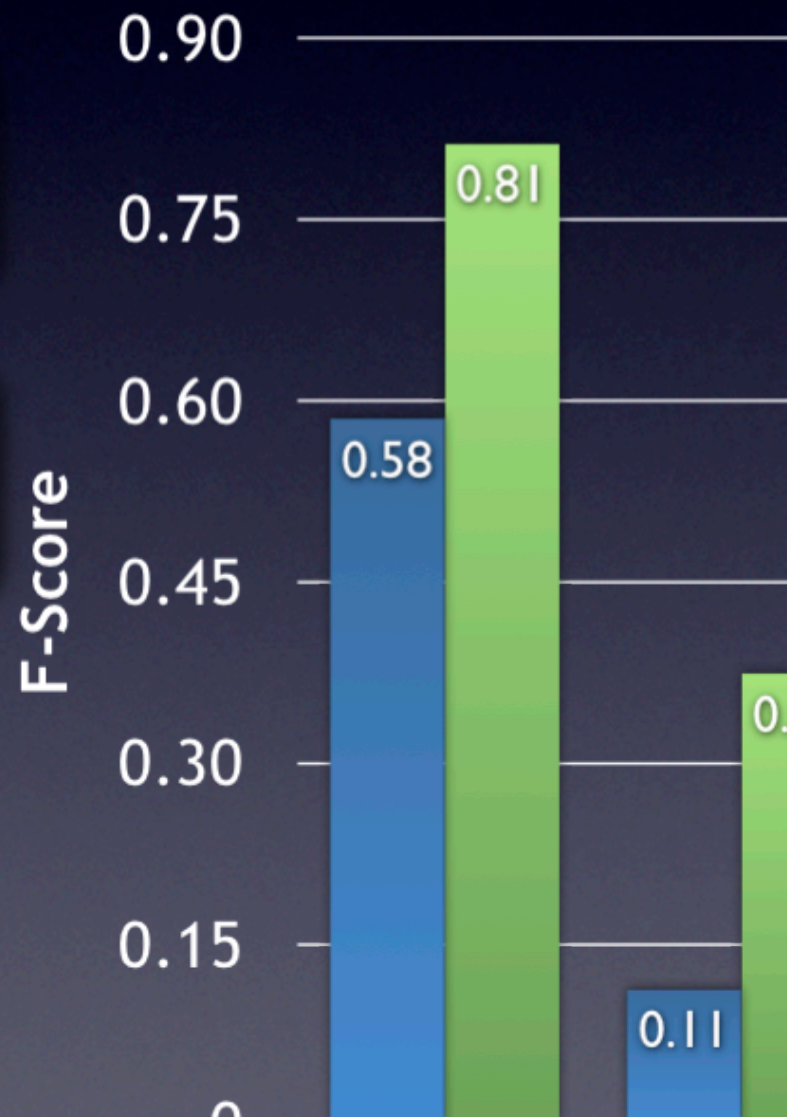


Intron and transcript accuracy evaluation

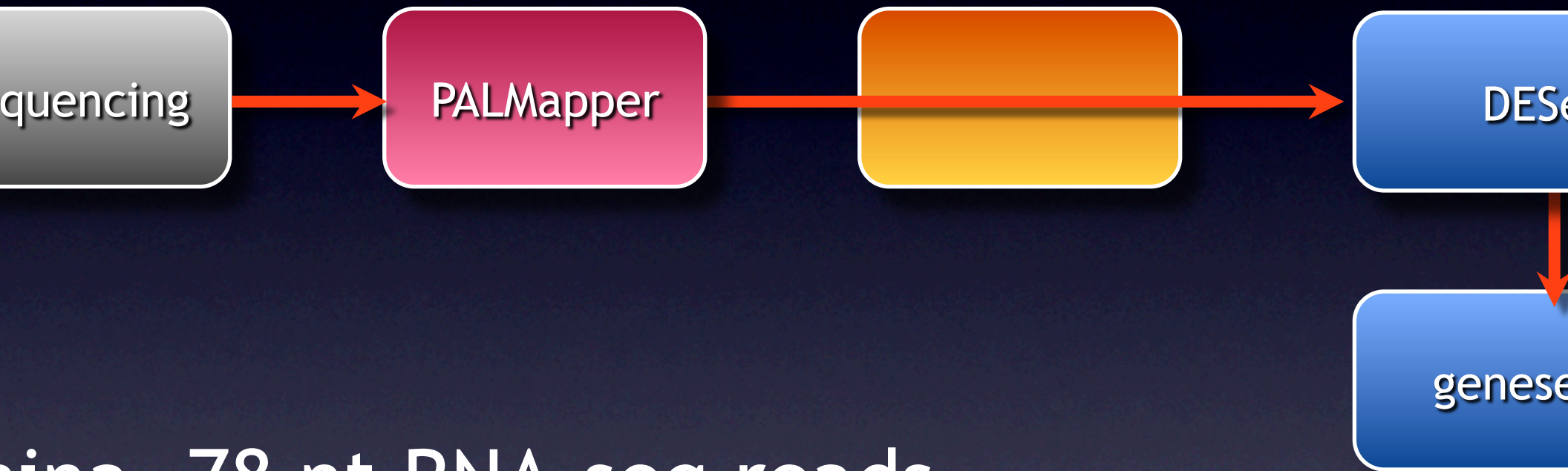


C. elegans

100 million RNA-seq reads
(4 million)



Multiple reference genomes and transcriptomes
A. thaliana

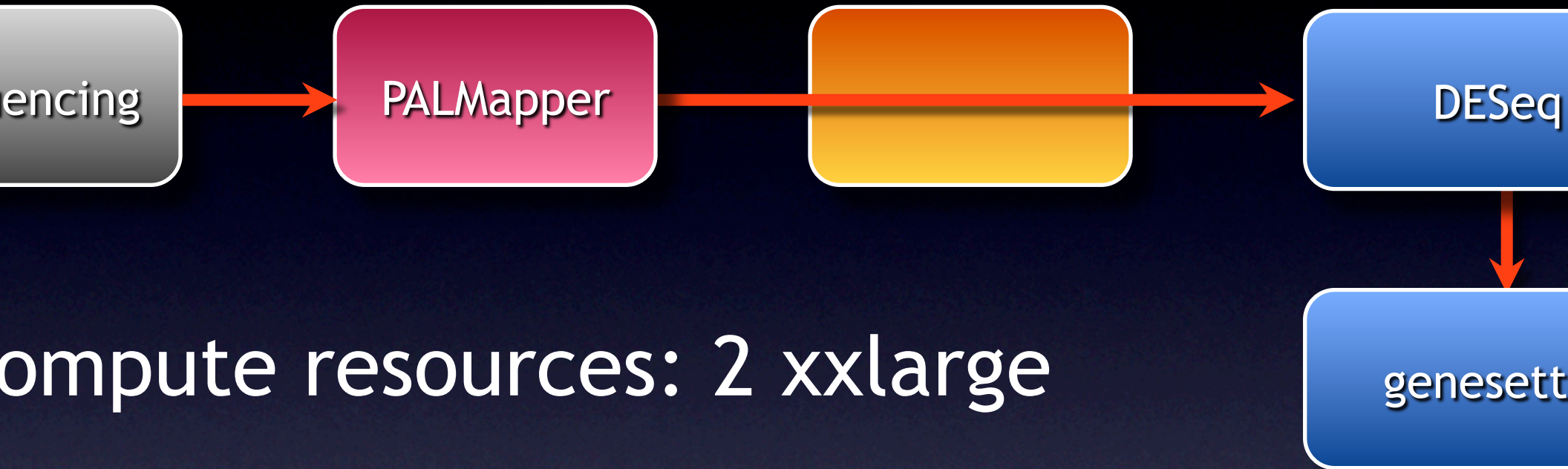


...mina, 78 nt RNA-seq reads

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
Mapper	0.4
	0.5.7
at	1.5.0

Read Alignment Filtering	Version
SAFT	0.2
Multi-Mapper Resolution	0.1

Transcript Prediction	Version
M	0.2
links	1.3.0
ty	r2012-06-08
ature	Beta-2
	0.3

Quantification	Version
rQuant	2.2
rDiff	0.2
Cufflinks/Cuffdiff	1.3.
DESeq	1.6.
topGO	0.1
Genesetter	0.1

Deployment

Shell script

```
#!/bin/bash
set -e

# Import urlopen
if ! command -v curl >>/dev/null; then
    echo "curl is not installed. Installing..."
    sudo apt-get install curl
fi

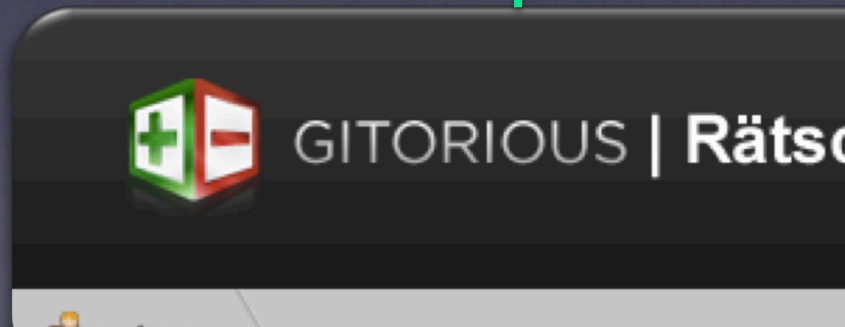
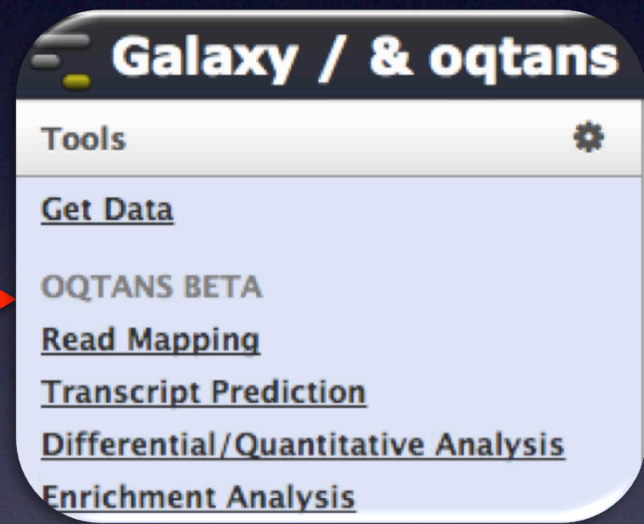
# Function to get the latest commit
get_latest_commit() {
    curl -s http://commits.server.com/latest | jq -r '.commit'
}

# Check if local changes have been committed
if ! git diff --quiet; then
    echo "Local changes have not been committed. Committing..."
    git commit -m "Deploying latest version"
fi

# Get the latest version of the app
LATEST_COMMIT=$(get_latest_commit)

# Update the app
echo "Updating app to latest version: $LATEST_COMMIT"
git checkout $LATEST_COMMIT

# Restart the service
sudo systemctl restart apache2
```



How to resolve the requirements for 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.oxford-journals.org/>

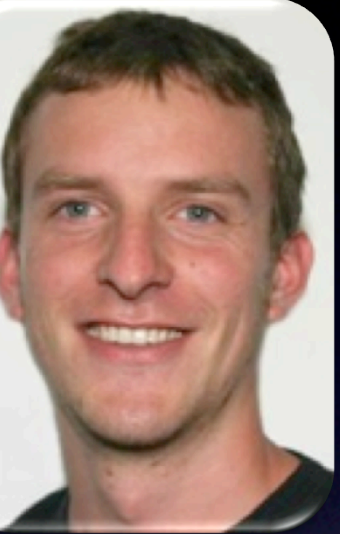
oqtans+ Availability: Cloud Computing

Demo cloud instance with all oqtans+ tools

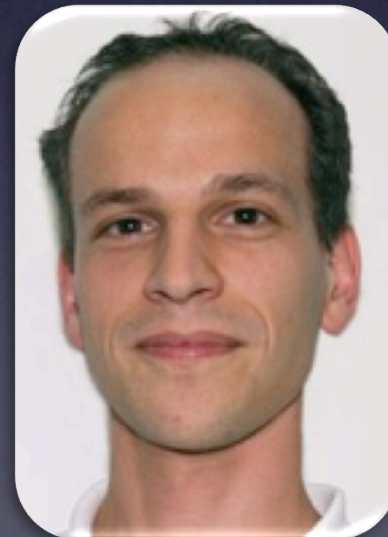
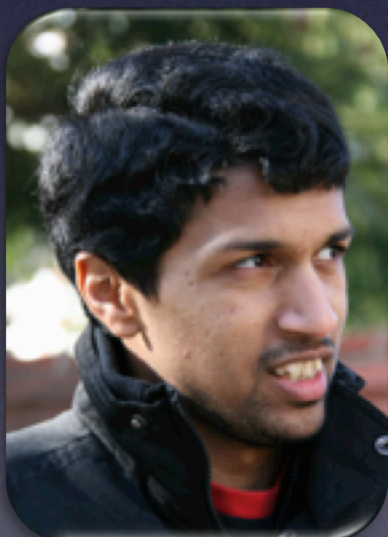
- ▶ <http://cloud.oqtans.org>

AMI at Amazon Web Services for EC2

- ▶ Cloudman to launch any number of instances as a compute cluster



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



Ré Kahles Pramod Mudra Sebastian Schultheiss Georg Zeller Gunnar Rää