

oqtans

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

Géraldine Jean <gjean@tue.mpg.de>

Sebastian J. Schultheiss <sebi@tue.mpg.de>

*Machine Learning in Biology, Rätsch Lab, FML of the Max Planck Society
Tübingen, Germany*

Presented at ISMB 2011, Vienna, Austria



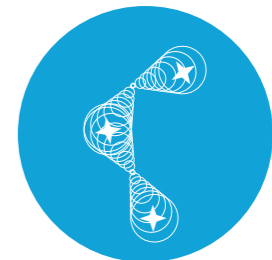
Friedrich Miescher Laboratory
of the Max Planck Society



MAX-PLANCK-GESELLSCHAFT

oqtans

online
quantitative
transcript
analysis



Galaxy Approach

- ▶ Persistent, transparent, reproducible approach to bioinformatics research
- ▶ Integration made simple with xml wrappers of command line tools
- ▶ Accessible: web service, download, cloud

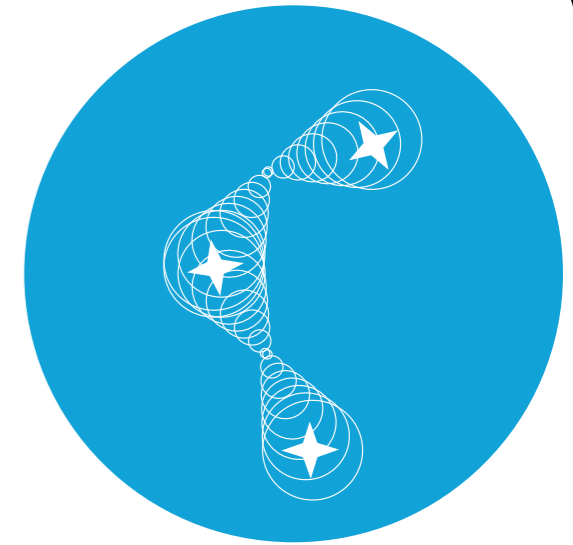
J. Goecks et al. 2010
D. Blankenberg et al. 2010
E. Afgan et al. 2010
S. Koskovsky Pond et al. 2009

W. Miller et al. 2007
J. Taylor et al. 2007
D. Blankenberg et al. 2007
M. Giardine et al. 2005

RNA-Seq Analysis

oqtans

online
quantitative
transcript
analysis



▶ Common analysis tasks

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

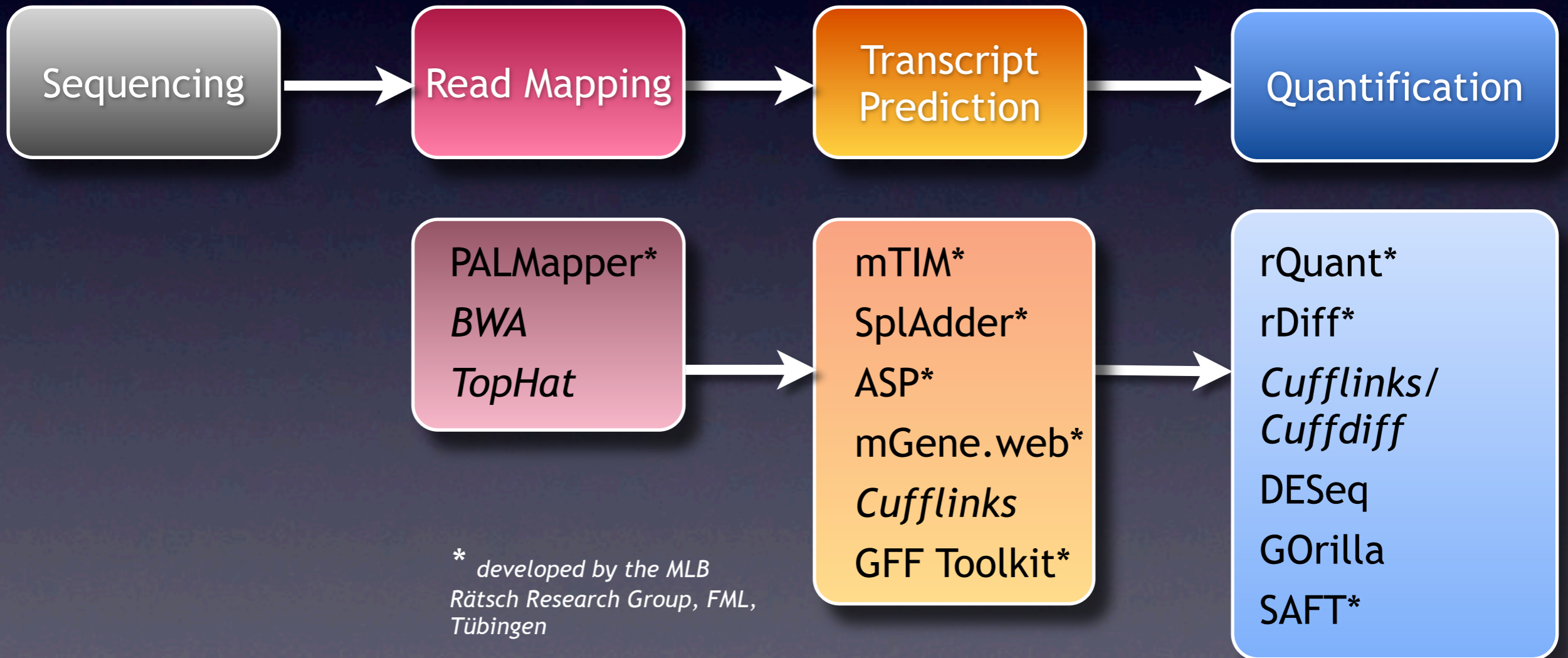
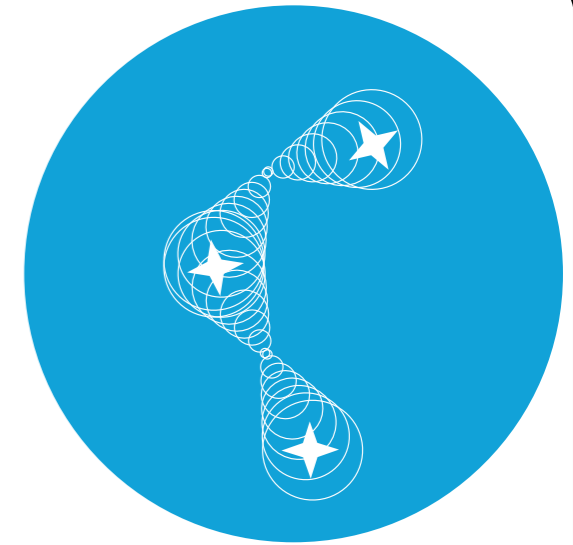
Lee et al. 2011 *Nucleic Acids Res*
Yamashita et al. 2011 *Genome Res*
Daines et al. 2011 *Genome Res*
Ramani et al. 2010 *Genome Res*

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

oqtans Galaxy Tools

oqtans

online
quantitative
transcript
analysis



* developed by the MLB
Rätsch Research Group, FML,
Tübingen

Plus EasySVM, KIRMES and more

oqtans

Transcriptome analysis toolsuite



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

G. Jean et al. 2010 *Curr Protoc Bioinformatics*

oqtans

Transcriptome analysis toolsuite



- ▶ **mTIM**: reconstructs exon-intron structure from alignments and splice site predictions
- ▶ **SplAdder**: adds isoforms to known annotation based on splice graph

G. Zeller et al. 2011 i.p.

oqtans

Transcriptome analysis toolsuite



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

R. Bohnert & G. Rättsch 2010 *Nucleic Acids Res*

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*

oqtans

Transcriptome analysis toolsuite



- ▶ **SAFT:** The simple alignment filtering tool evaluates alignment accuracy

A. Kahles et al. 2011 *i.p.*

- ▶ **GOrilla:** Gene ontology enrichment analysis and visualization tool

E. Eden et al. 2009 *BMC Bioinformatics*

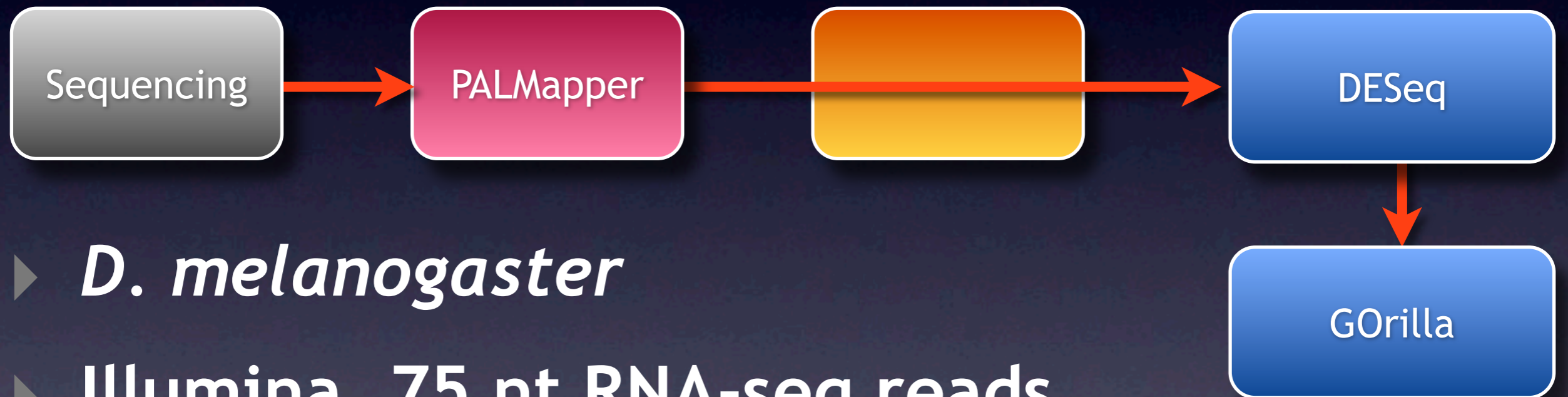
oqtans

Live Demonstration



oqtans

Live Demonstration



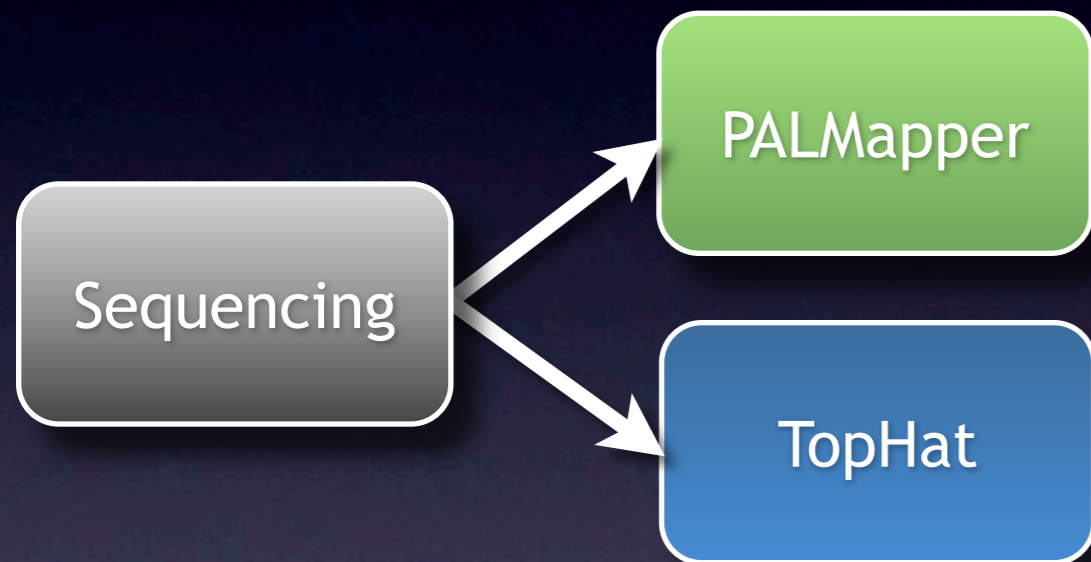
▶ *D. melanogaster*

▶ Illumina, 75 nt RNA-seq reads

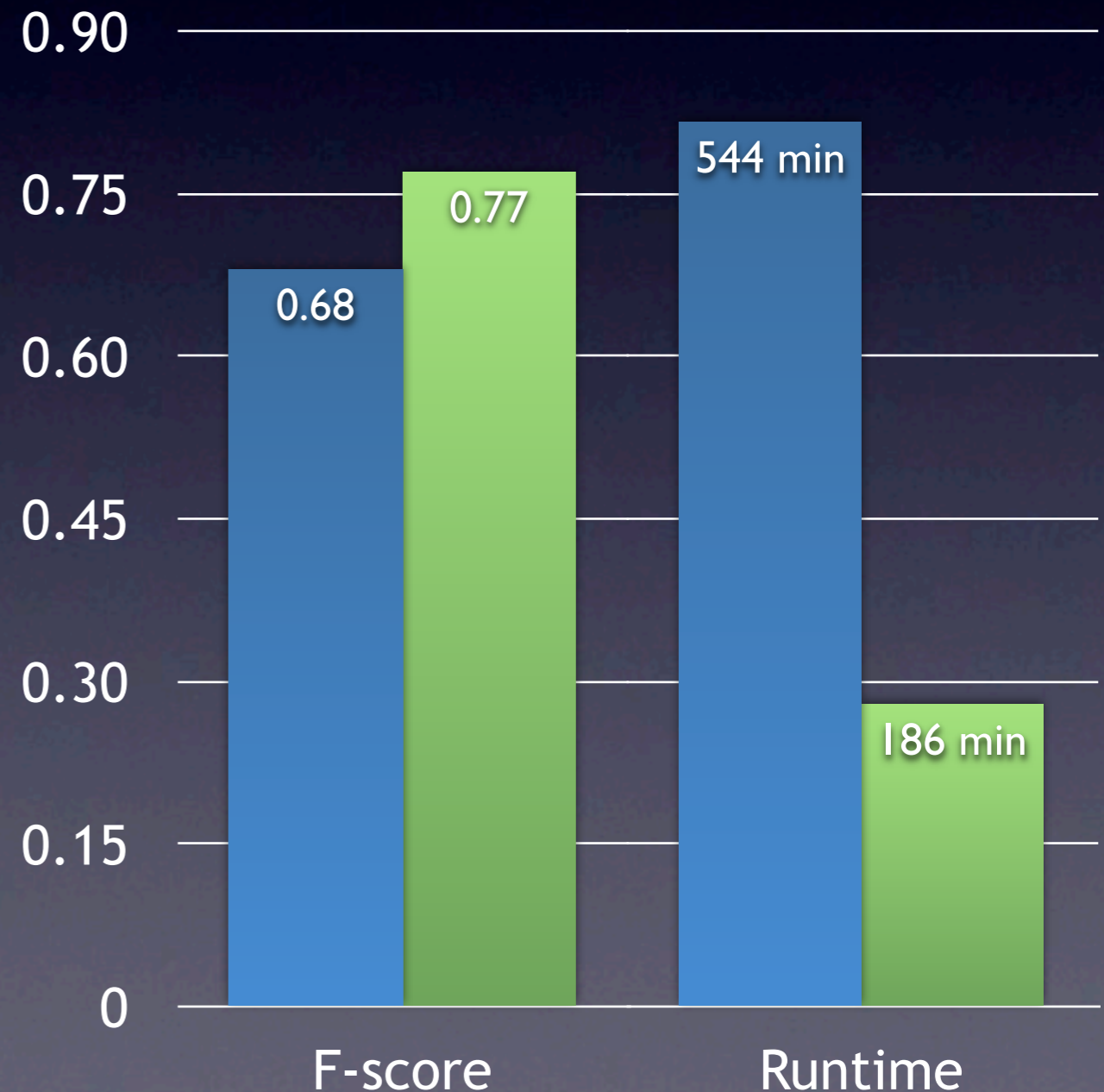
▶ 3-day-old male adult (7.4 million)

▶ 3-day-old female adult (12.4 million)

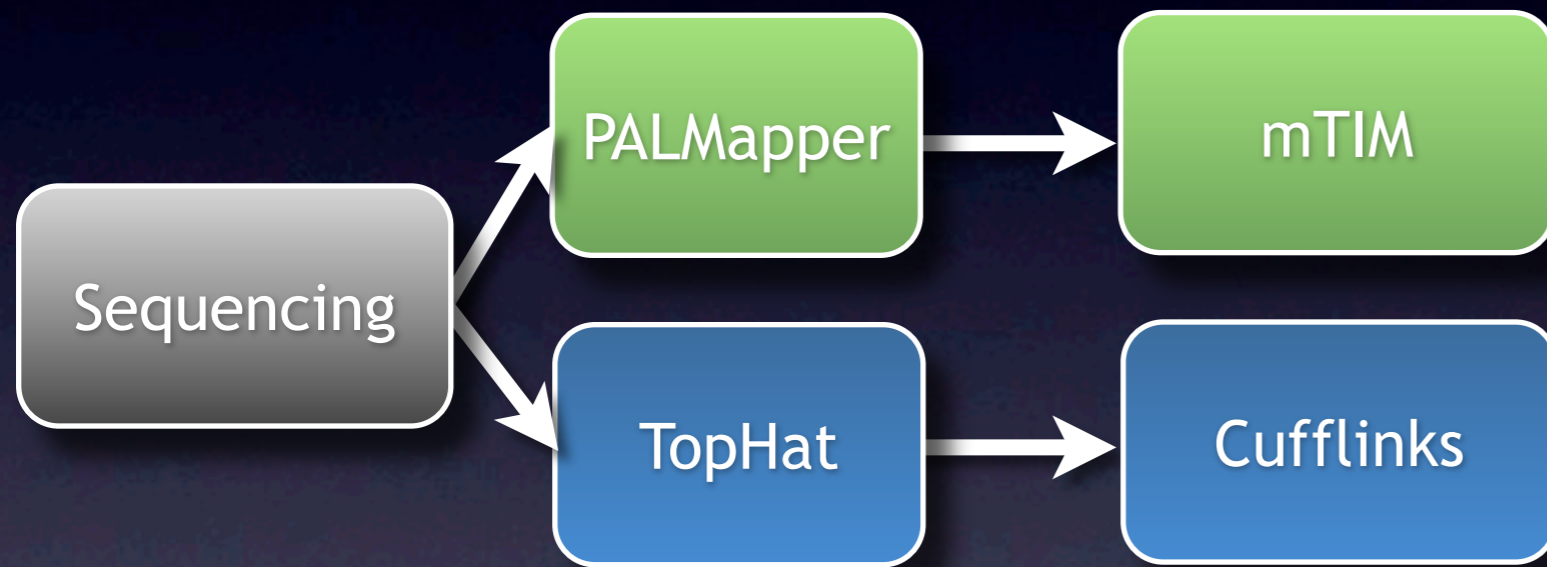
Intron prediction accuracy of read alignments



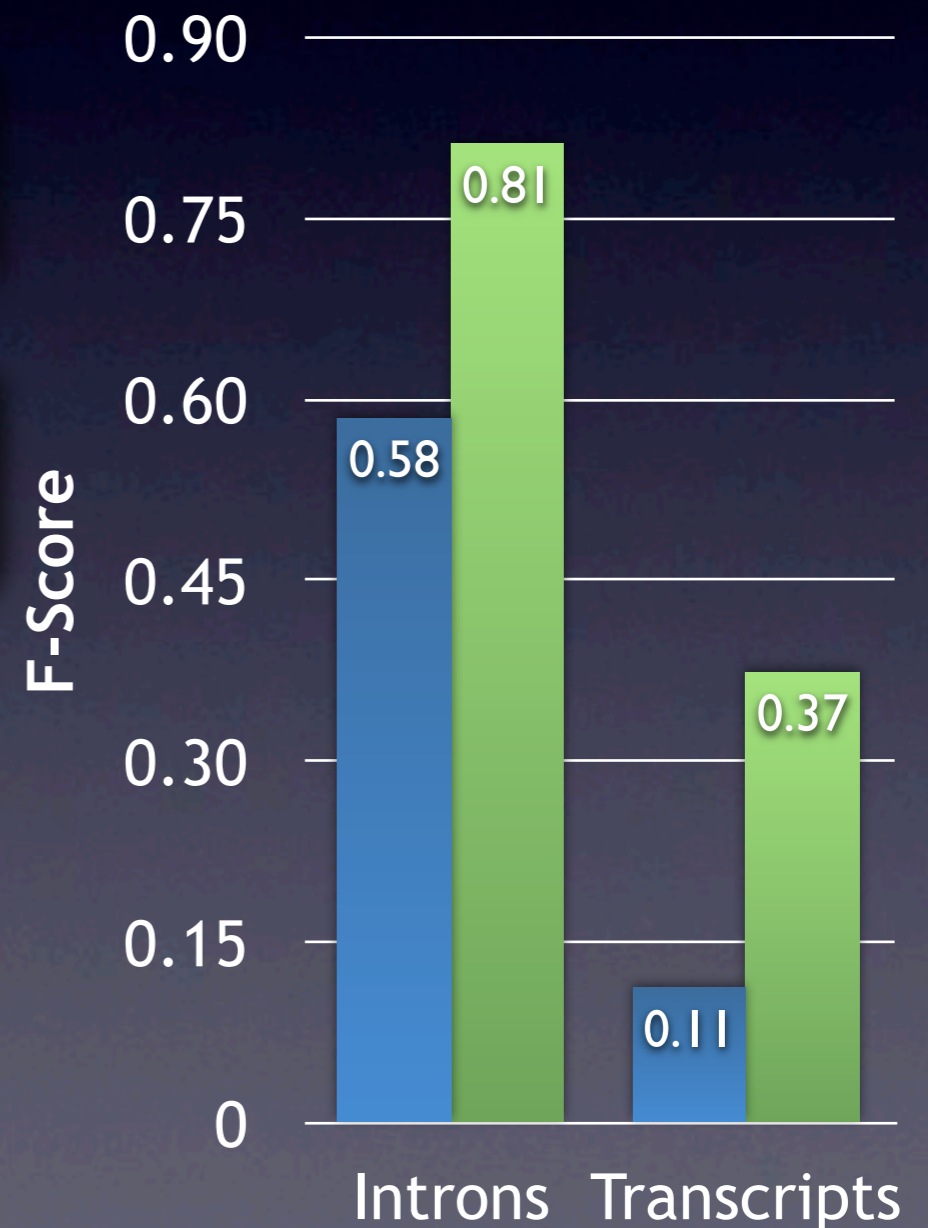
- ▶ *C. elegans*
- ▶ 75 nt RNA-seq reads (24 million)



Intron and transcript accuracy evaluation



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



oqtans Availability: Our Server

- ▶ External compute cluster
 - ▶ 21 nodes, 168 CPUs
 - ▶ our Galaxy test instance
 - ▶ All tools described here, and more!
- ▶ <http://galaxy.fml.mpg.de>

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://community.g2.bx.psu.edu/>

oqtans Availability: Cloud Computing

- ▶ Demo cloud instance with all oqtans tools
 - ▶ <http://tinyurl.com/oqtans-ismb>
- ▶ AMI at Amazon Web Services for EC2
 - ▶ Cloudman to launch any number of instances as a compute cluster
- ▶ AMI in your own virtualizer (e.g. Virtual Box)
 - ▶ search for “oqtans”

oqtans Availability: Cloud Computing

Galaxy Cloudman Console

Welcome to Galaxy Cloudman. This application allows you to manage this instance of Galaxy CloudMan. Your previous data store has been reconnected. Once the cluster has initialized, use the controls below to add and remove 'worker' nodes for running jobs.


Terminate cluster

Add nodes

Remove nodes



Access Galaxy

Status

Cluster name: OqtansFMLCluster 

Disk status: 25G / 100G (25%) 

Worker status: Idle: 1 Available: 1 Requested: 3

Service status: Applications  Data 



Autoscaling is **on**.
Turn off?

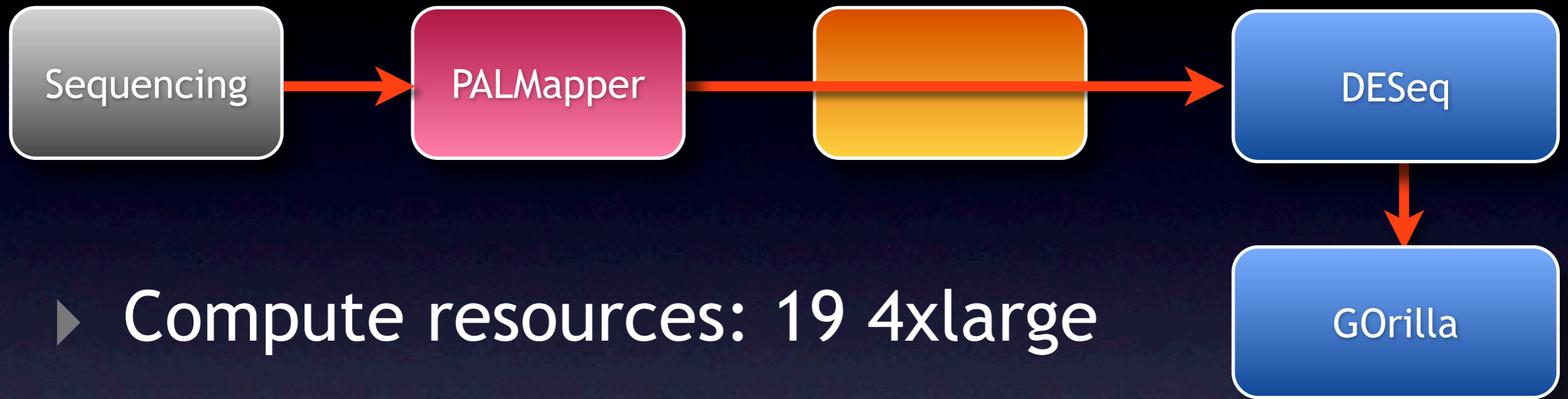
Min nodes: 1
Max nodes: 5
Adjust limits?

Cluster status log



Live Demo Results

Live Demo Results



- ▶ Compute resources: 19 4xlarge
- ▶ Time:
 - ▶ Alignments: 20 minutes
 - ▶ Quantitative analysis: 10 minutes
- ▶ Cost on Amazon EC2: approx. 7.82€

<http://oqtans.org>



Jonas Behr, Regina Bohnert, Philipp Drewe, Nico Görnitz



André Kahles, Pramod Mudrakarta, Vipin T. Sreedharan, Georg Zeller, Gunnar Rätsch