

oqtans

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

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

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

Presented at the 7th ISCB Student Council Symposium 2011, Vienna, Austria



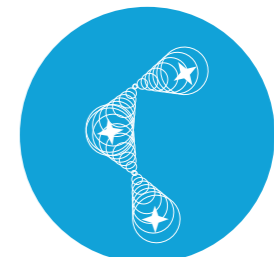
Friedrich Miescher Laboratory
of the Max Planck Society



MAX-PLANCK-GESELLSCHAFT

oqtans

online
quantitative
transcript
analysis



Web Services Availability

Schultheiss 2011, PLoS CB
Schultheiss et al. 2011, PLoS ONE i.r.

- ▶ 927 web services (NAR Web Server Issues)
- ▶ Check availability of every service, 4 times
- ▶ Survey among authors (274 respondents)
- ▶ Problems the authors perceive:
 - ▶ web address change (28%)
 - ▶ missing example data (30%)
 - ▶ program on server not functional (>17%)
 - ▶ undocumented changes (>1%)

Web Service Availability

Schultheiss 2011, PLoS CB
Schultheiss et al. 2011, PLoS ONE i.r.

- ▶ 927 web services (Server Issues)
- ▶ Check availability (4 times)
- ▶ Survey and (Incidents)
- ▶ Problem



**Caution: Published results
may not be reproducible**

Galaxy Approach



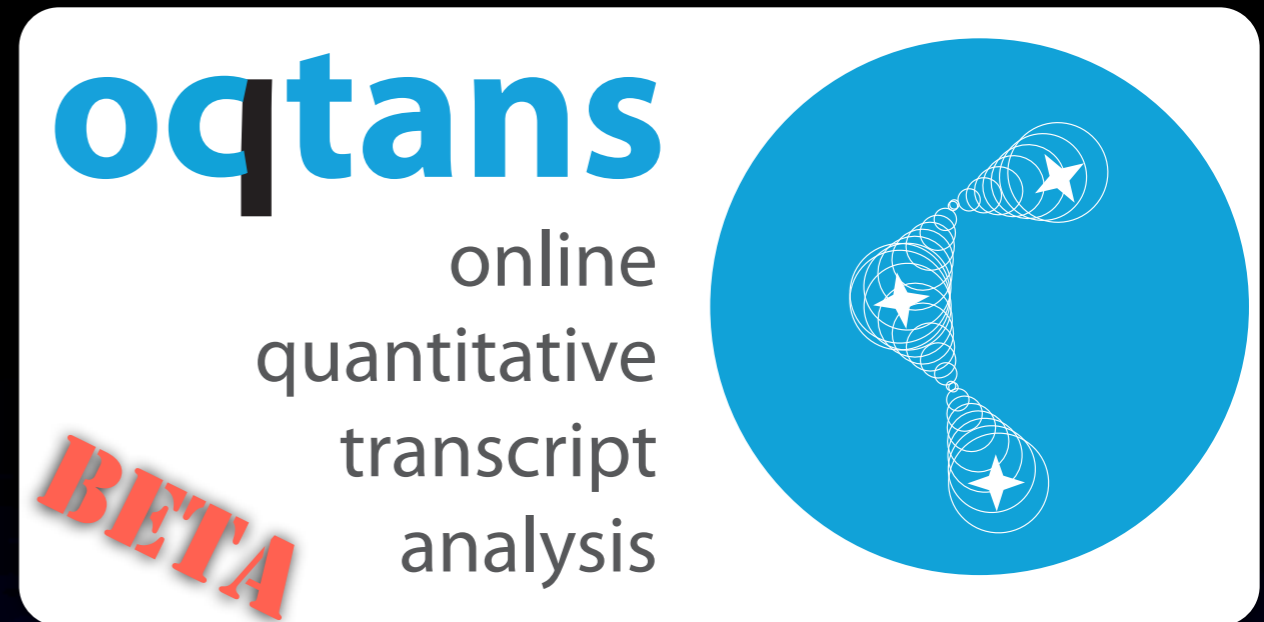
Workshop
WK06

- ▶ 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

oqtans Galaxy Tools



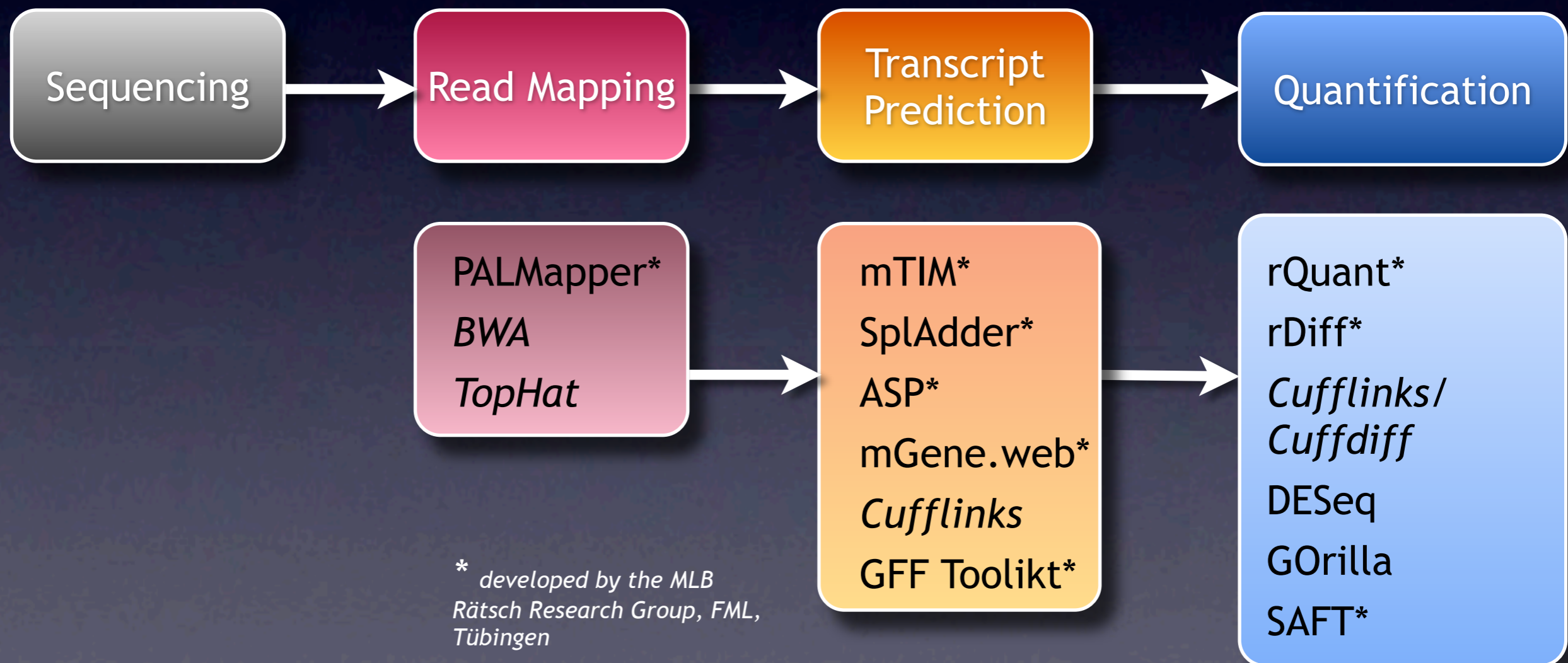
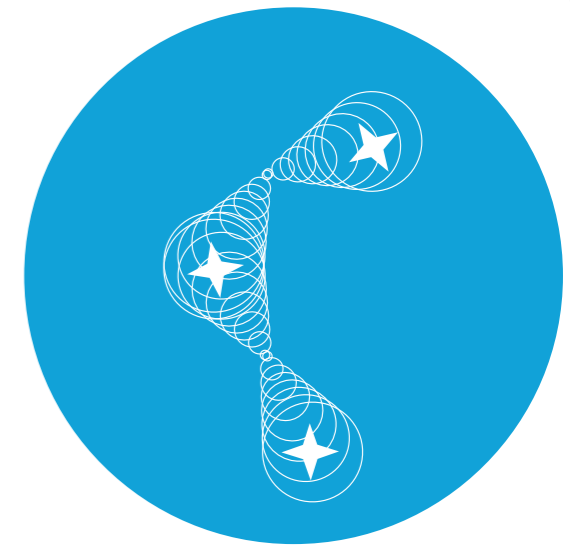
- ▶ Common analysis tasks
 - ▶ compare two samples (wildtype, mutant)
 - ▶ identify new transcripts

oqtans Galaxy Tools

oqtans

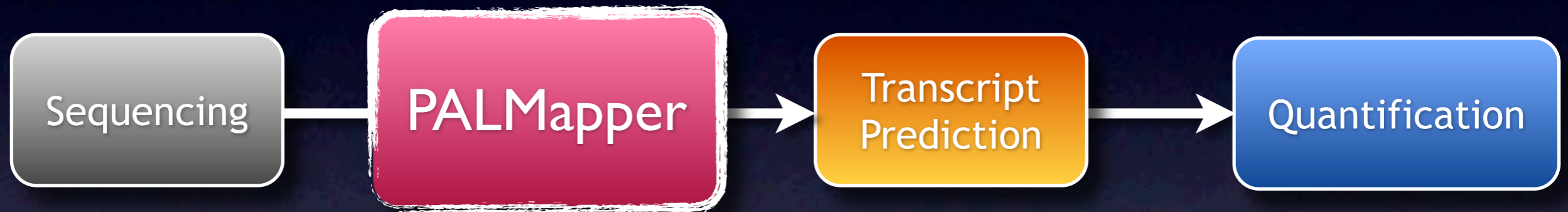
online
quantitative
transcript
analysis

BETA



oqtans Machine Learning Tools

Poster
U65



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

oqtans Machine Learning Tools

Poster
M20



- ▶ **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

Machine Learning Tools

HiTSEQ
SIG



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

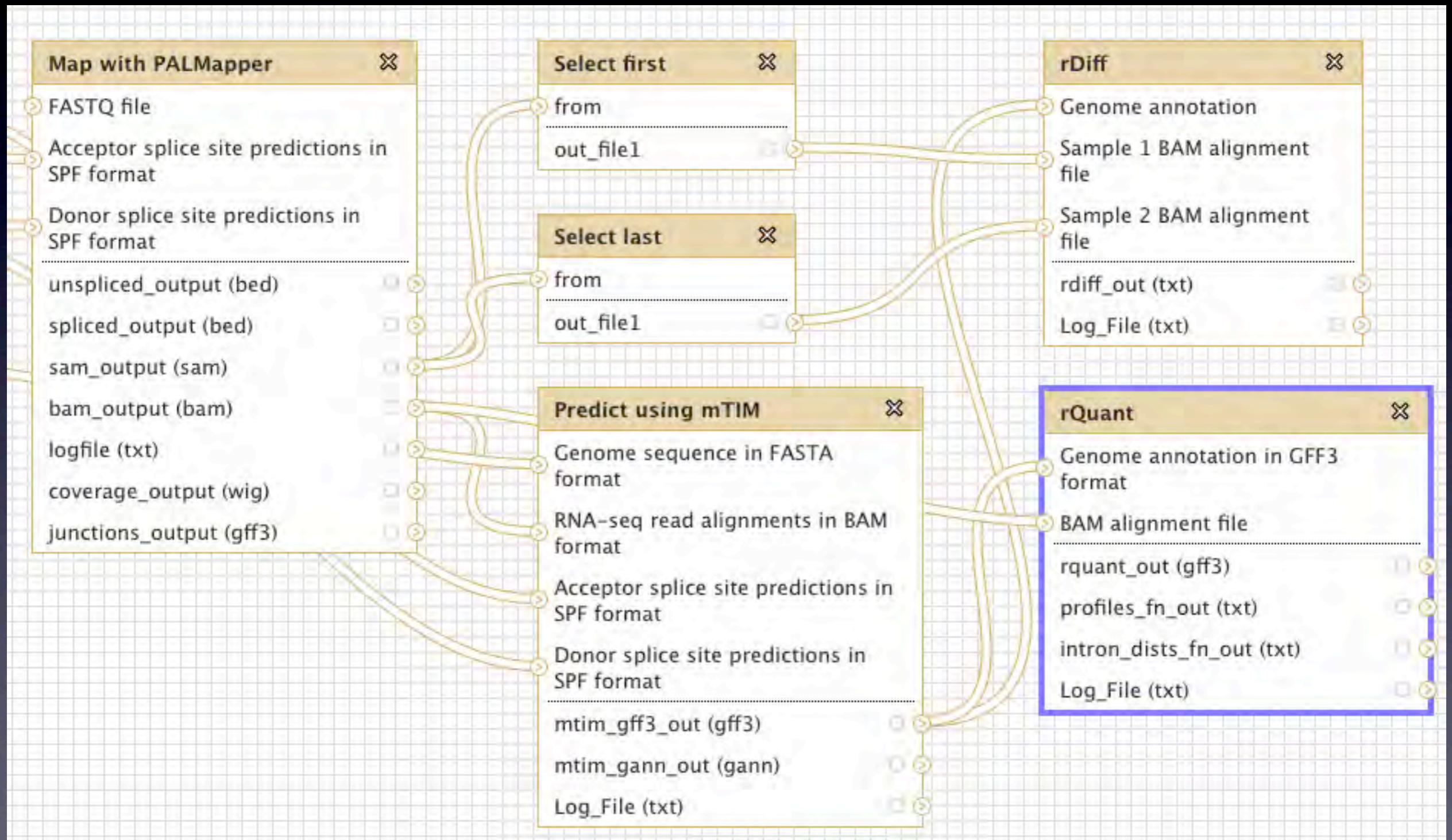
oqtans

Machine Learning Tools



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

oqtans Workflow in Galaxy



oqtans Availability: Our Server

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

oqtans Availability: Source Code

- ▶ Packaged releases 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/>

Galaxy Tool Installers

Joint work: Enis Afgan & Galaxy Team
James Taylor, Anton Nekrutenko
AG Rätsch

- ▶ MLB group tools into any Galaxy installation
- ▶ Python Fabric scripts: used to manage automation of a remote server

```
install_cmd("wget %s" % self.tool_env['url'])  
install_cmd('chown -R %s %s' % (env.user, install_dir_root))
```

- ▶ Adjusted to Ubuntu on Galaxy Cloud Image

oqtans Availability: Cloud Computing

Tech
Track Talk
Tues 2:30

- ▶ Demo cloud instance with all oqtans tools
- ▶ 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

Tech
Track Talk
Tues 2:30

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: 2.7G / 20G (14%) 

Worker status: Idle: 0 Available: 0 Requested: 0

Service status: Applications  Data 



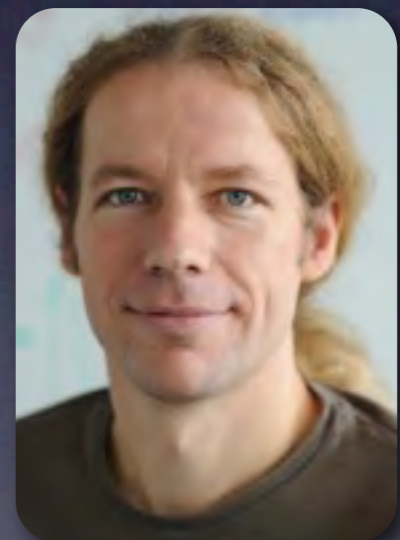
Autoscaling is **off**.
Turn on?

<http://oqtans.org>

Tech
Track Talk
Tues 2:30



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



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