oqtans

Quantitative Transcriptome Analysis in the Cloud — Part II

Géraldine Jean <Geraldine.Jean@univ-nantes.fr>
Department of Computer Science, Combinatorics and Bioinformatics Group, LINA — University of Nantes, France
Sebastian J. Schultheiss <sebastian@bioweb.me>
Computomics Molecular Data Analysis, Center for Bioinformatics, University of Tübingen, Germany

Presented at the Meeting on Advances and Challenges of RNA-Seq Analysis Halle, Germany

UNIVERSITÉ DE NANTES
MAX-PLANCK-GESELLSCHAFT
computomics® molecular data analysis
oqtans online quantitative transcript analysis

NGS data

- Specificities:
  - Rapidity of production
  - Low cost



- Small fragments (reads)



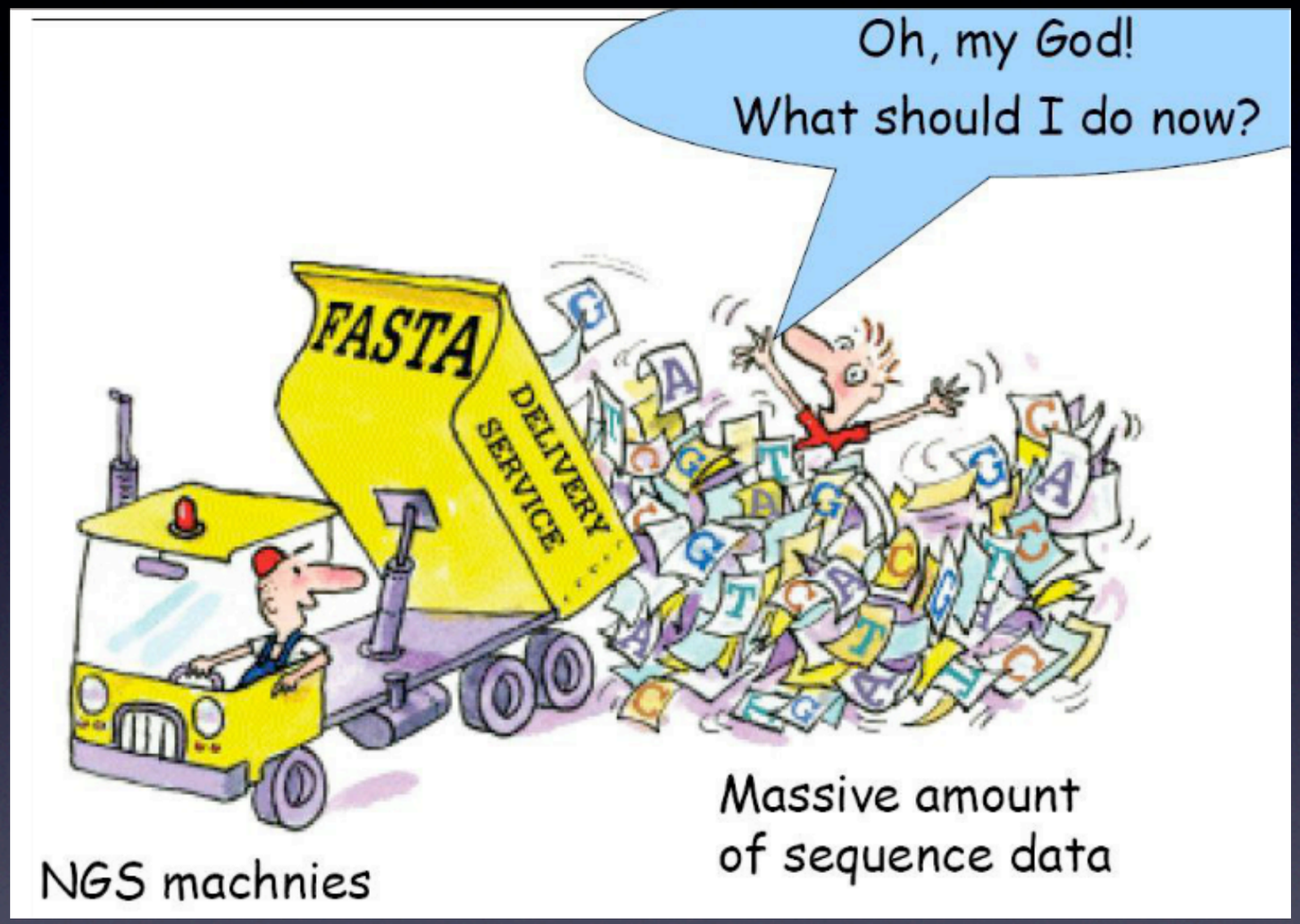






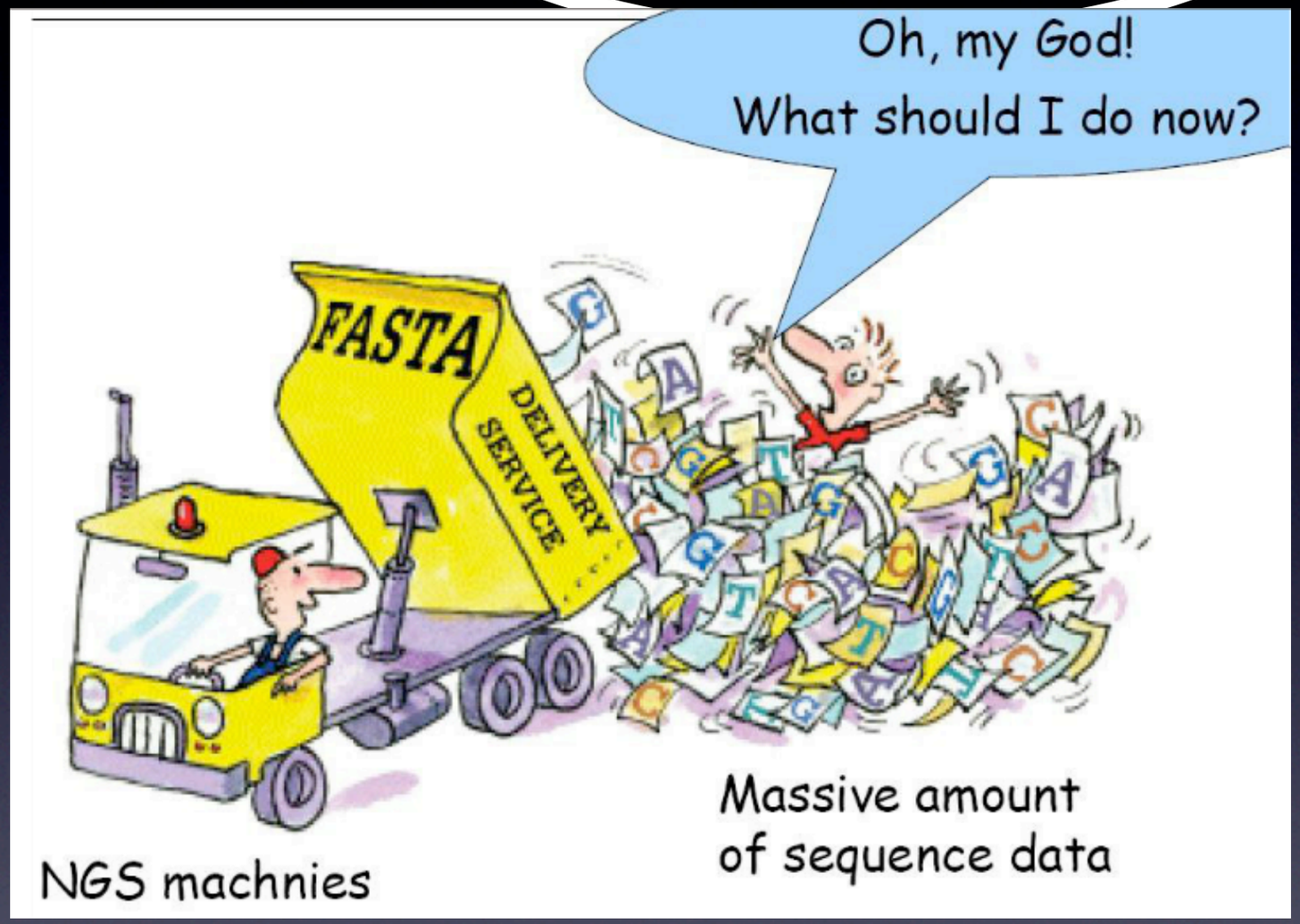


NGS data


Oh, my God!
What should I do now?
FASTA
DELIVERY
SERVICE
NGS machnies
Massive amount
of sequence data

NGS data

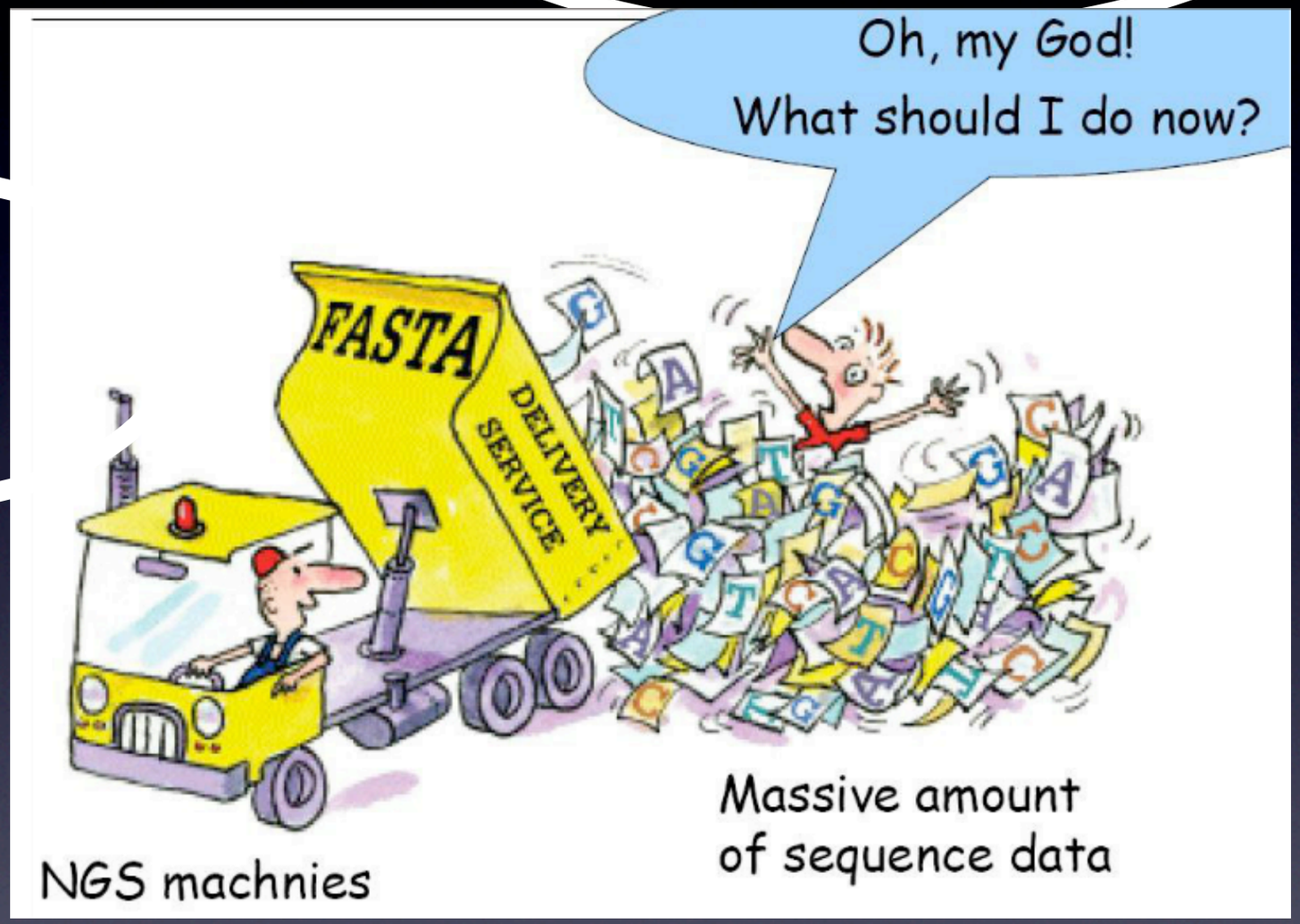
Solve biological problems



NGS data

Solve biological problems

Store data

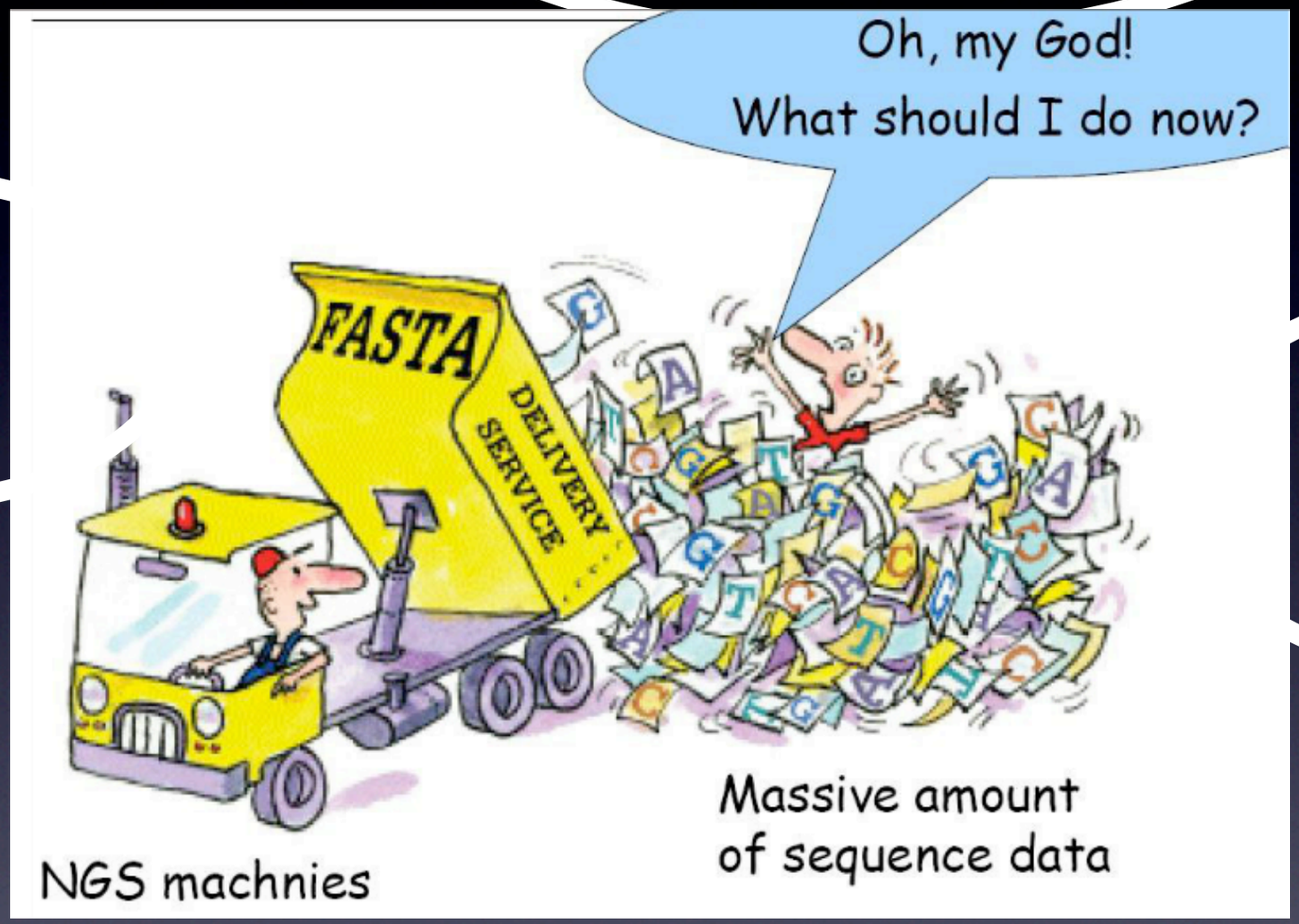


NGS data

Solve biological problems

Store data

Automate analysis

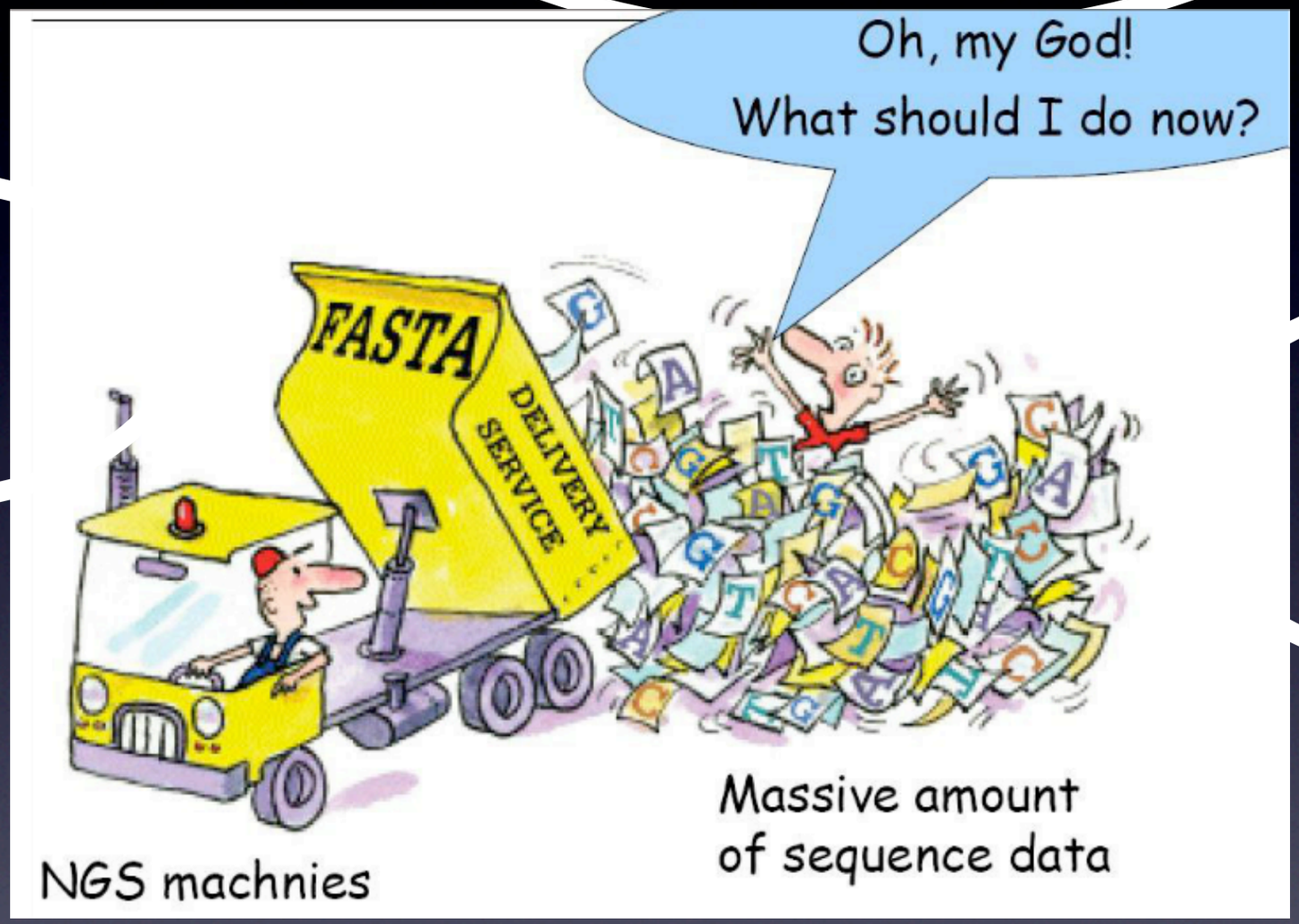


NGS data

Solve biological problems

Store data

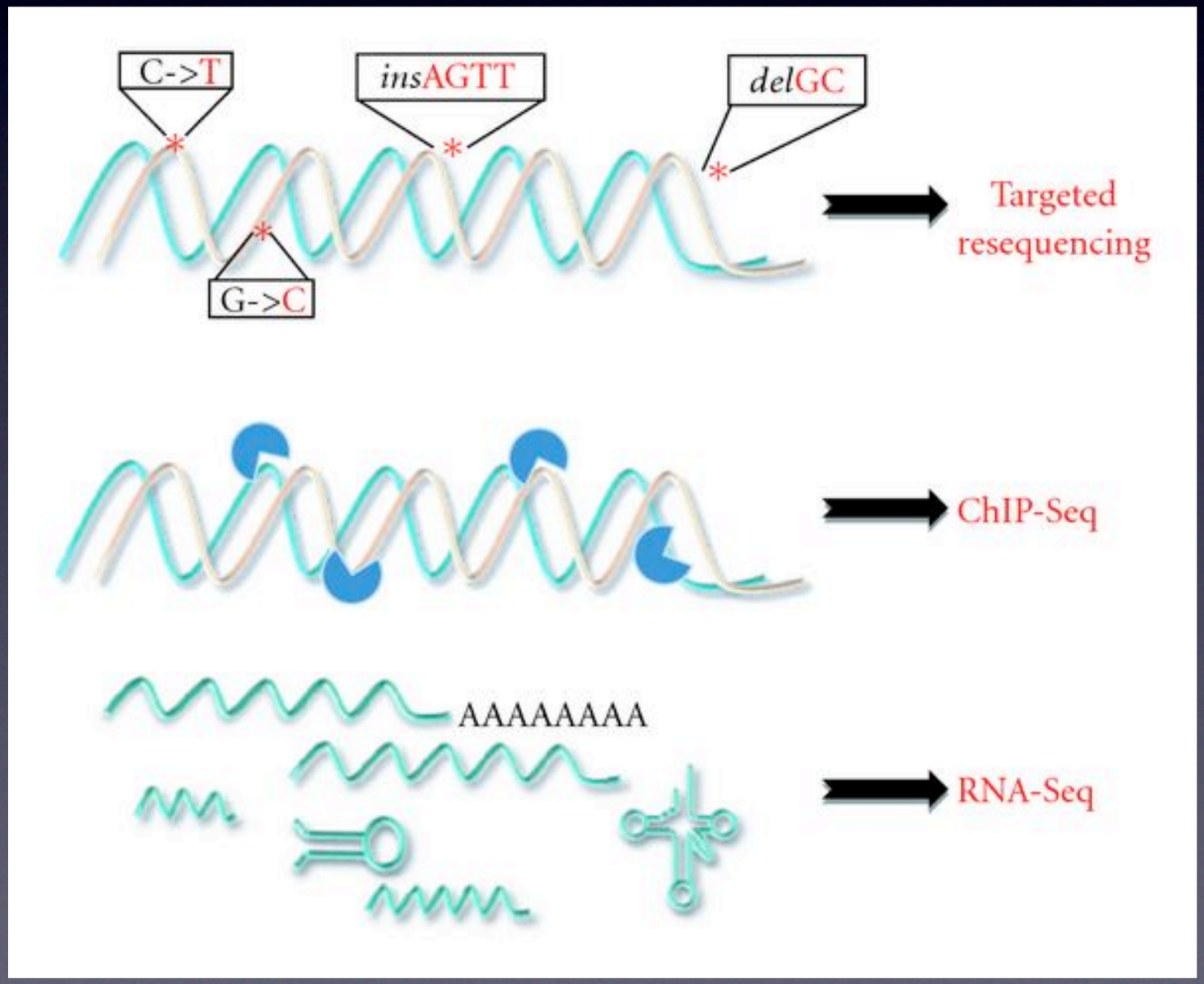
Automate analysis



Need adapted bioinformatics tools to:
store, efficiently and accurately treat data and reproduce analysis

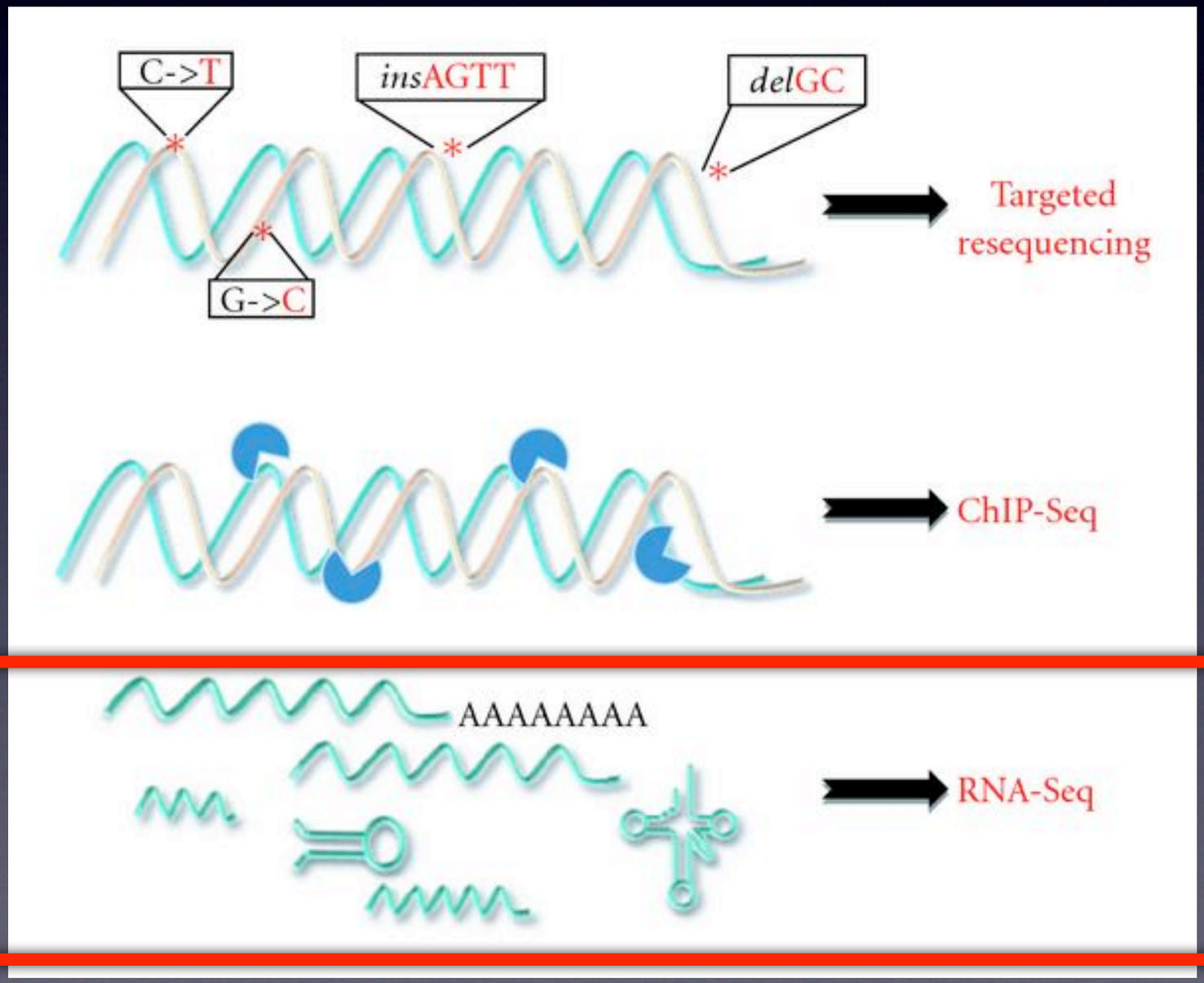
NGS data

Next generation sequencing applied to many scientific contexts:



NGS data

Next generation sequencing applied to many scientific contexts:

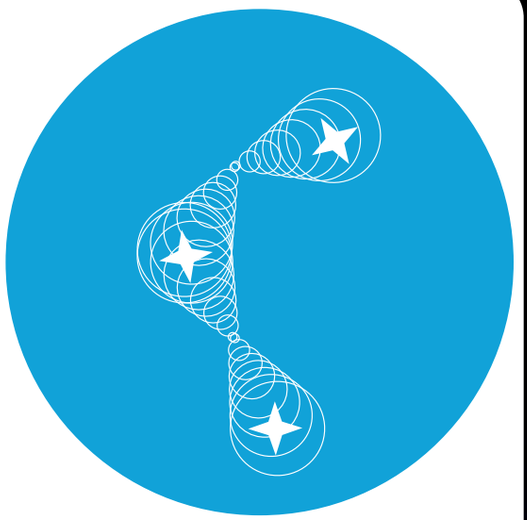


oqtans

RNA-Seq Analysis

oqtans

online
quantitative
transcript
analysis





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

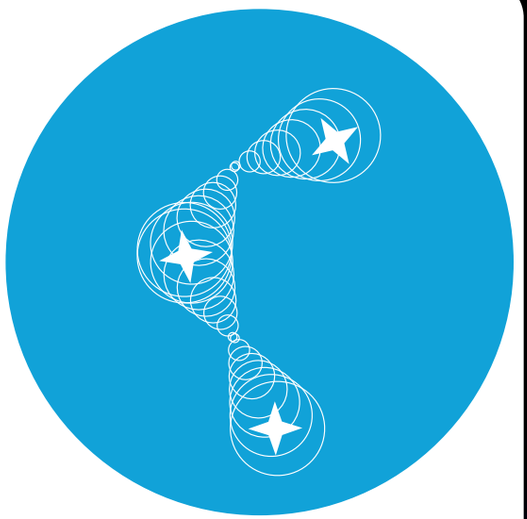
Lee et al. 2011 *Nucleic Acids Res*
Yamashita et al. 2011 *Genome Res*
Daines et al. 2011 *Genome Res*
Ramani et al. 2010 *Genome Res*

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

RNA-Seq Analysis

oqtans

online
quantitative
transcript
analysis





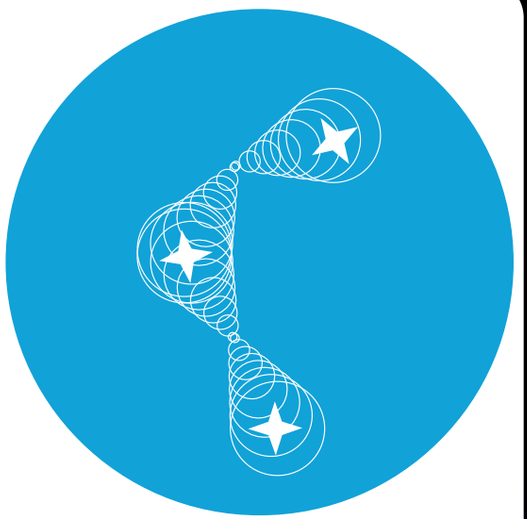
- Common issues
  - Reproducibility
  - Tool availability
  - Scalability

oqtans

Open-Source workbench for Quantitative Transcriptome Analysis

oqtans

online
quantitative
transcript
analysis



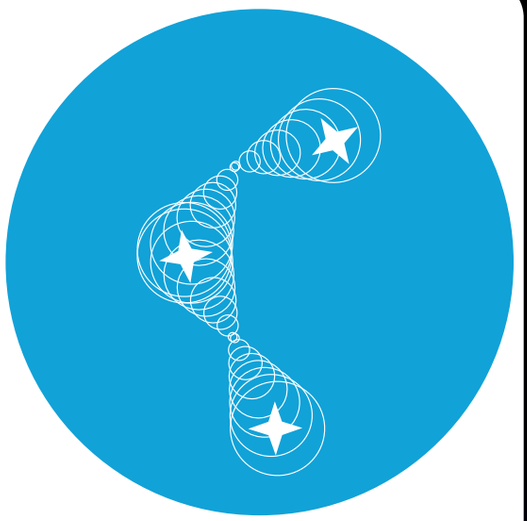


- Based on Galaxy
- Several additional tools on top
- Machine Learning-based tools
- Tool alternatives

oqtans
Galaxy Tools

oqtans

online
quantitative
transcript
analysis



Sequencing → Read Mapping → Transcript Prediction → Quantification

PALMapper*
BWA
TopHat

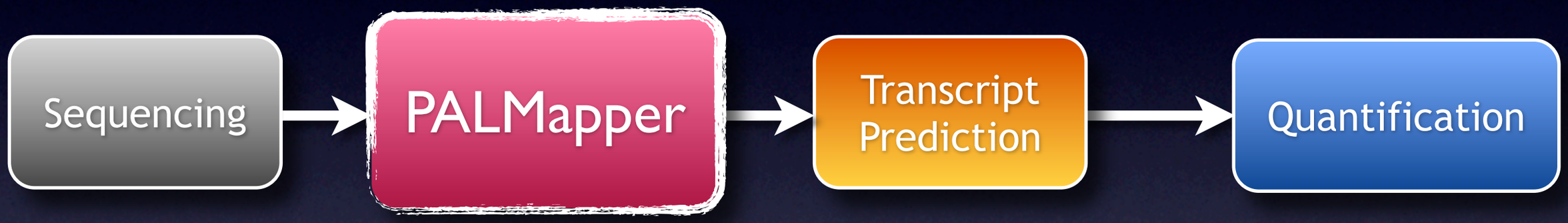
mTIM*
SplAdder*
ASP*
Trinity
Cufflinks
GFF Toolkit*

rQuant*
rDiff*
Cufflinks/ Cuffdiff
DESeq
Genesetter*
topGo
SAFT*

* *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. Rätsch 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.*

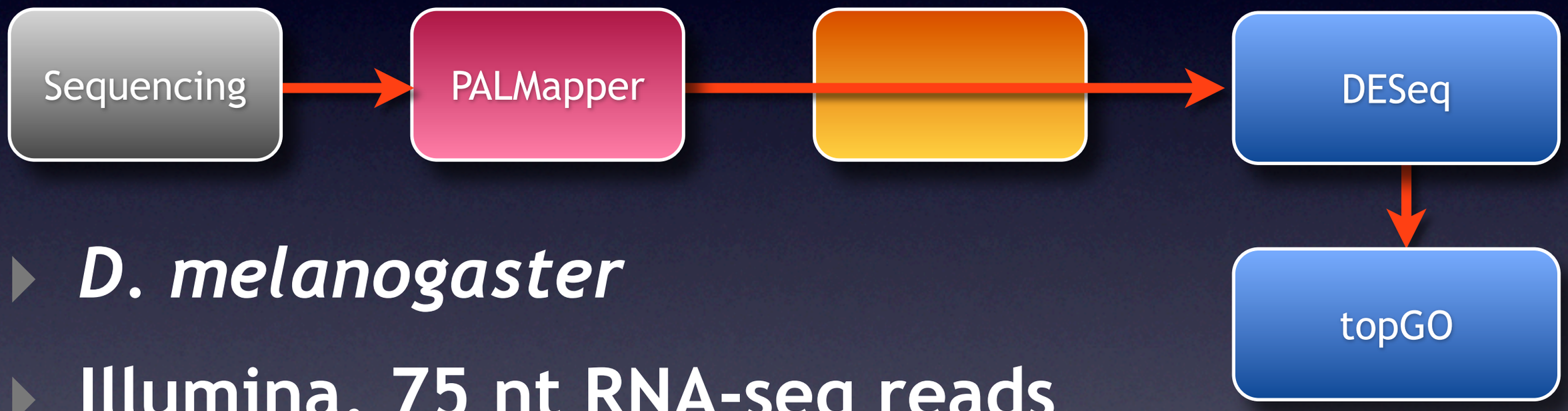
- **topGO**: Gene ontology enrichment analysis and visualization tool

A. Alexa et al. 2006 Bioinformatics

oqtans
Live Demonstration

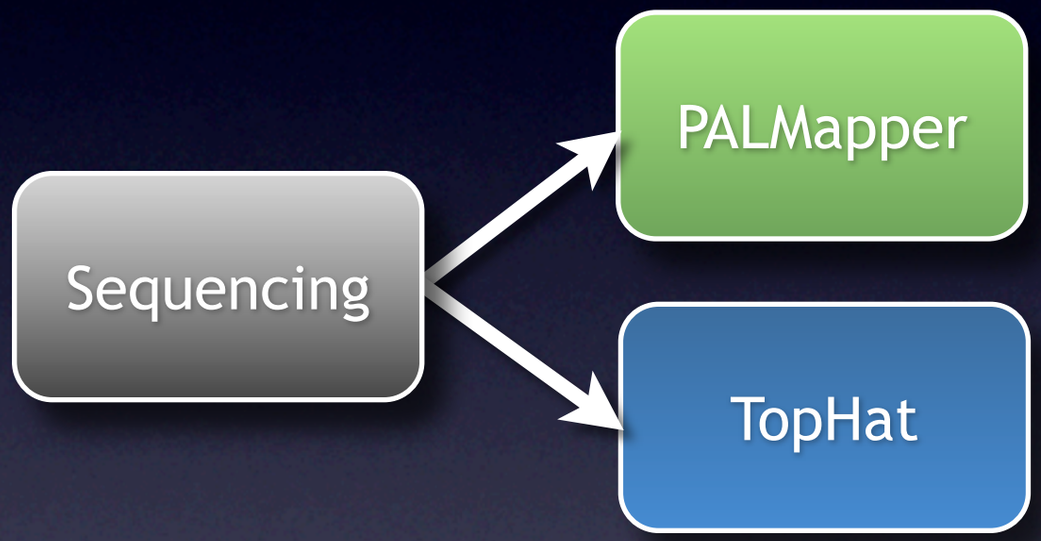


oqtans
Live Demonstration (part I)

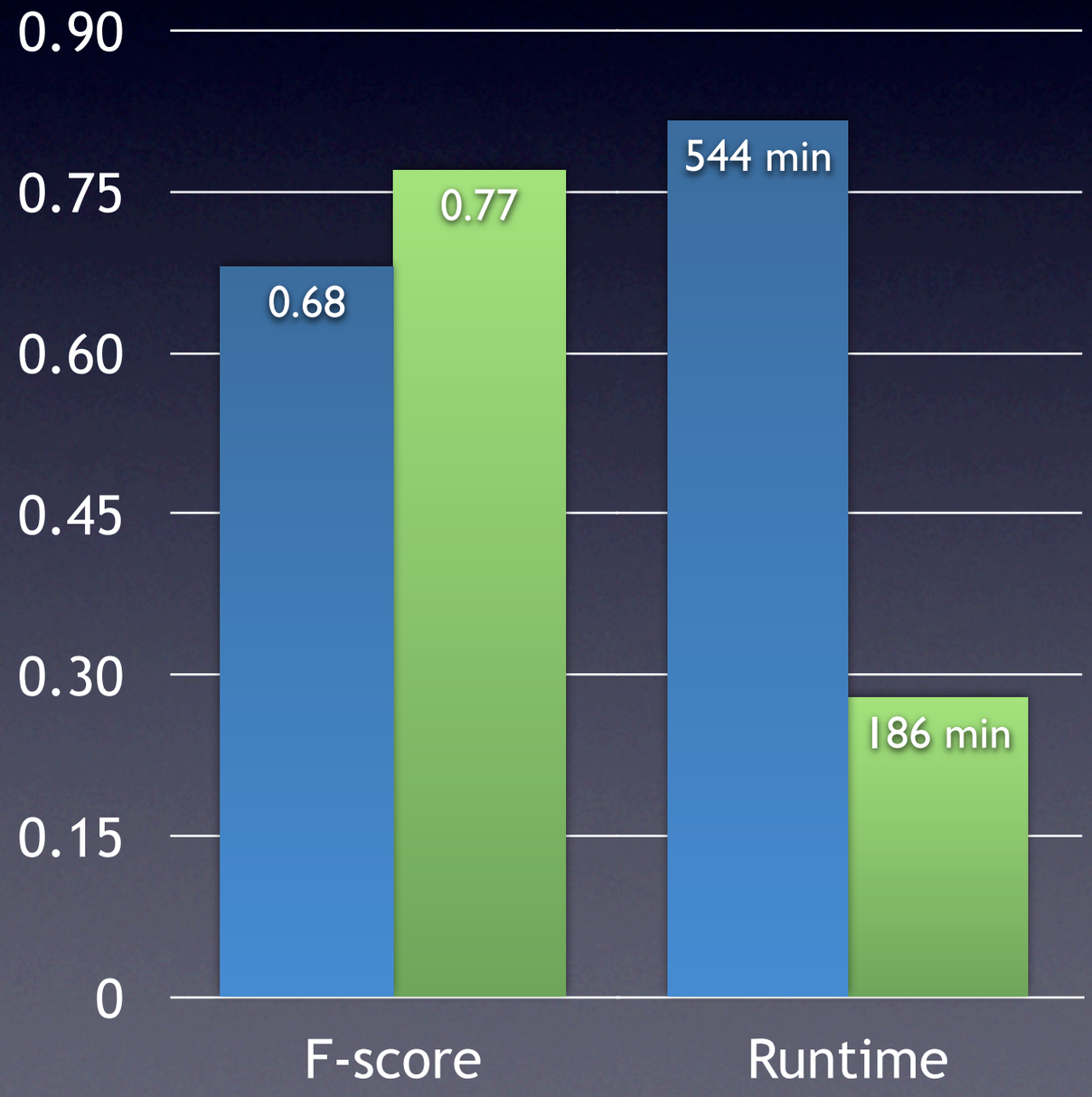


- *D. melanogaster*
- Illumina, 75 nt RNA-seq reads
  - 3-day-old male adult (7.4 million)
  - 3-day-old female adult (12.4 million)

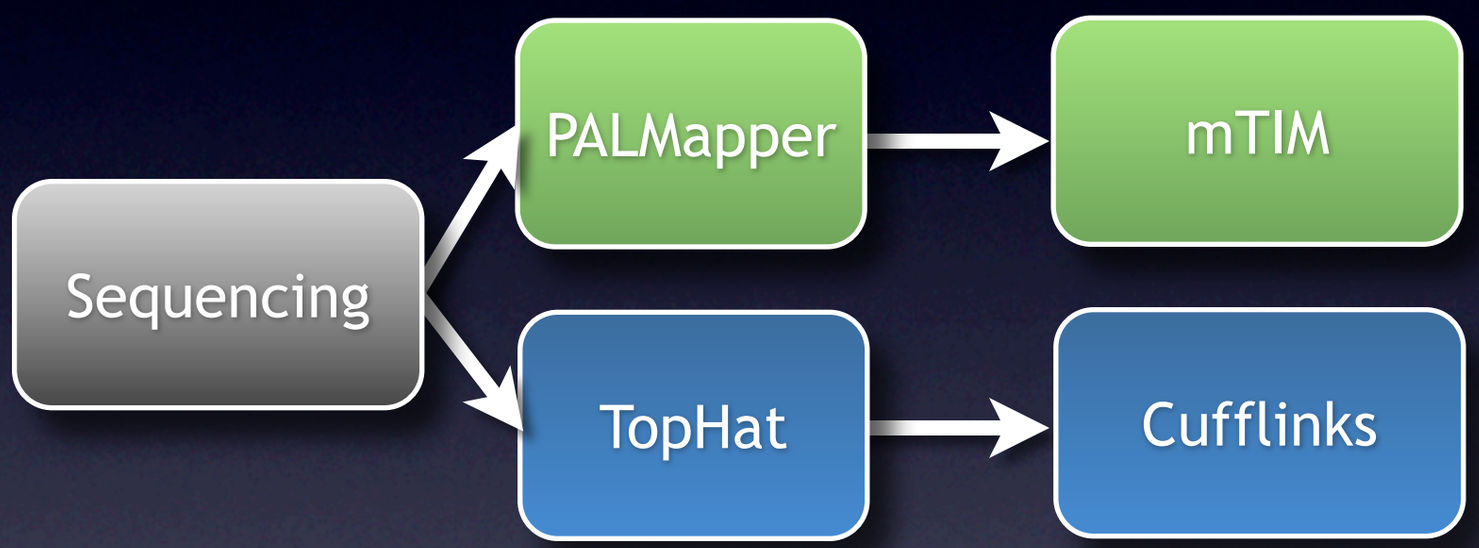
Intron prediction accuracy of read alignments



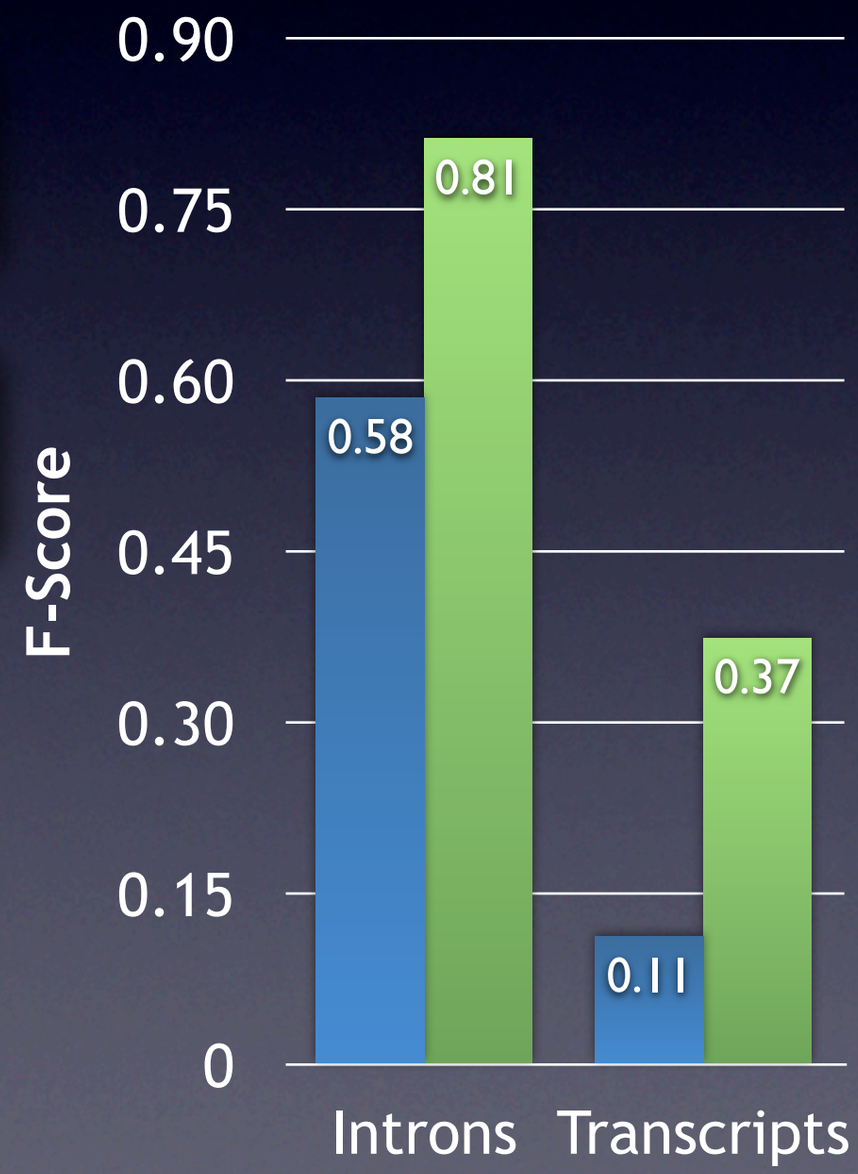
- *C. elegans*
- 75 nt RNA-seq reads (24 million)



Intron and transcript accuracy evaluation



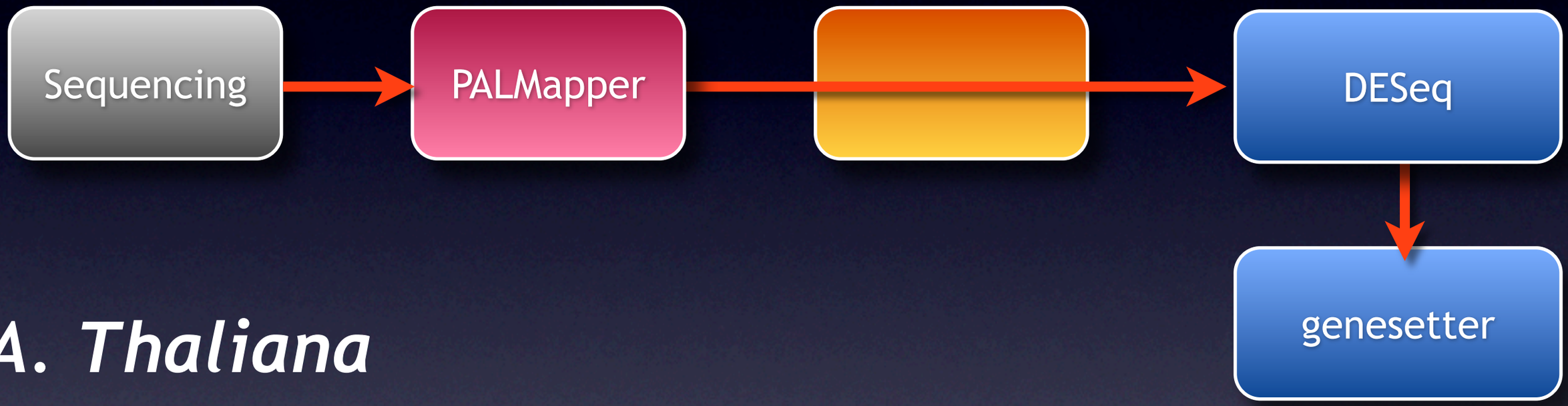
- *C. elegans*
- 75 nt RNA-seq reads (24 million)



oqtans
Live Demonstration



oqtans
Live Demonstration (Part II)

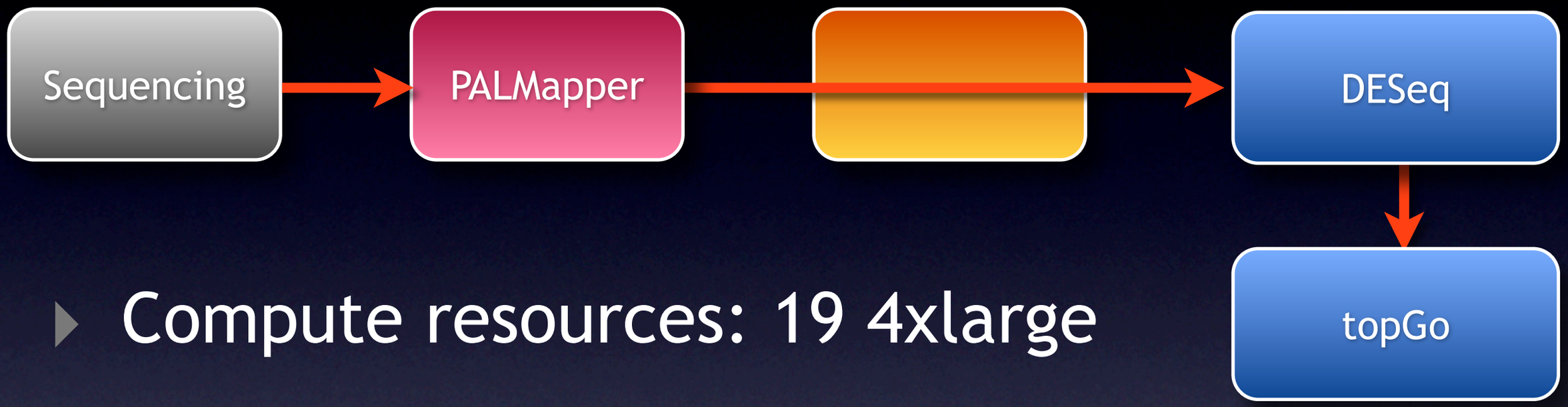


- *A. Thaliana*
- Illumina, 78 nt RNA-seq reads
  - Columbia accession (Col-0) (1.2 millions)
  - Canary Island accession (Can-0) (4.9 millions)

Gan et al. 2011 Nature

Live Demo Results
Part I

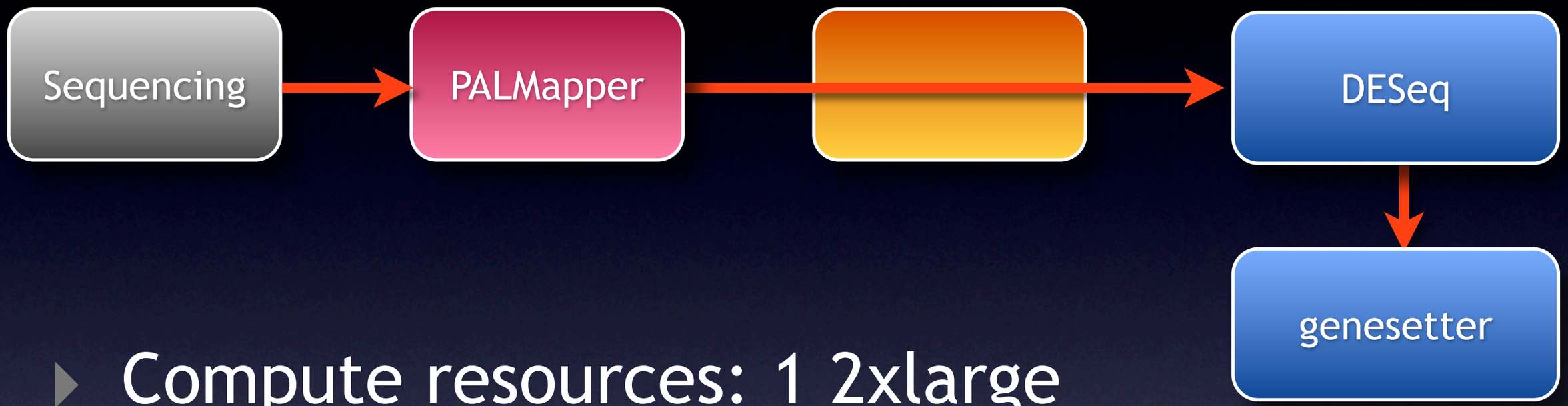
Live Demo Results Part I



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

Live Demo Results
Part II

Live Demo Results Part II



- Compute resources: 1 2xlarge
- Time:
  - Quantitative analysis: under 1 hour
- Cost on Amazon EC2: approx. 0.72€

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

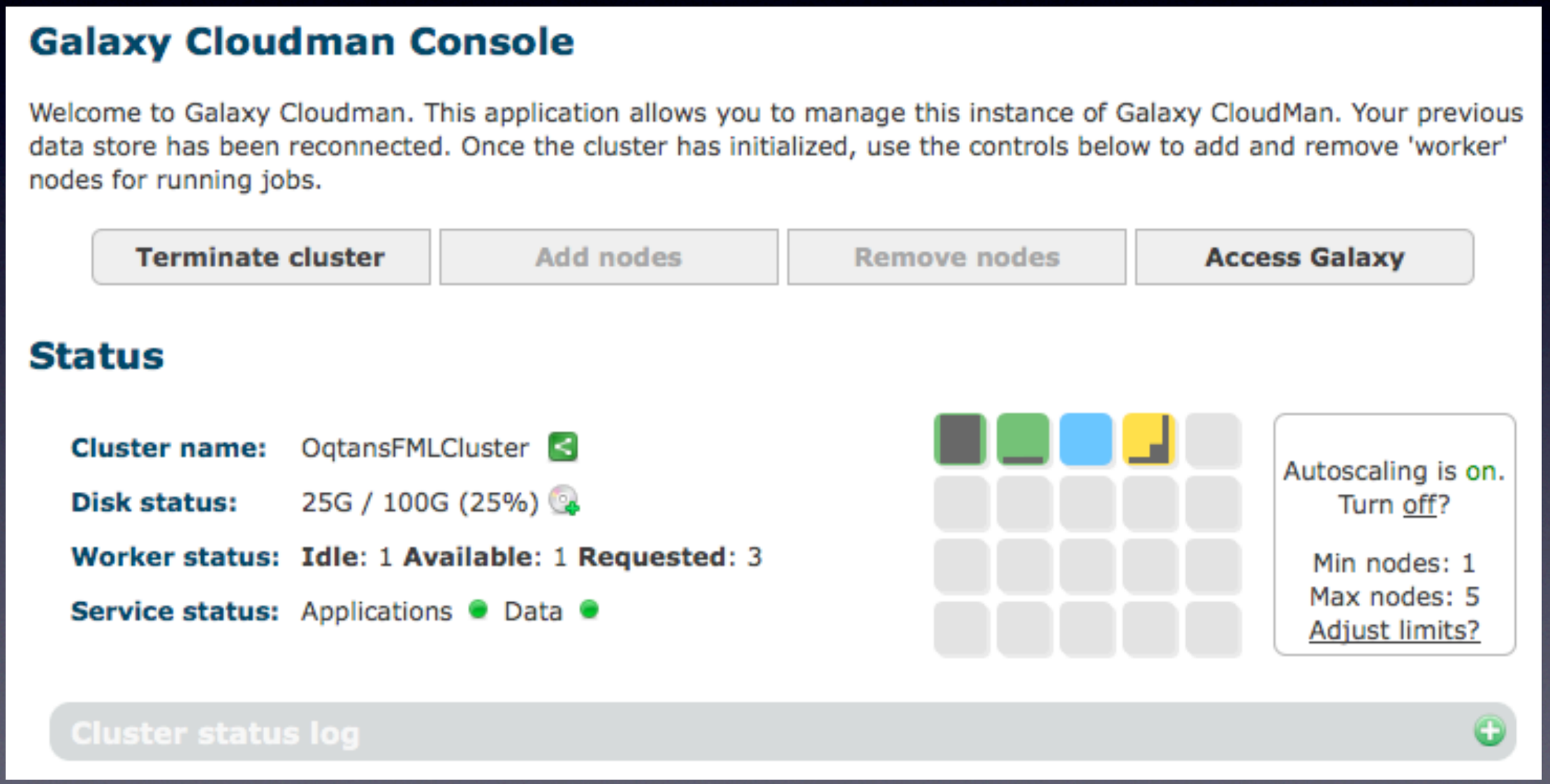
oqtans Availability: Source Code

- 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://toolshed.g2.bx.psu.edu/

oqtans Availability: Cloud Computing

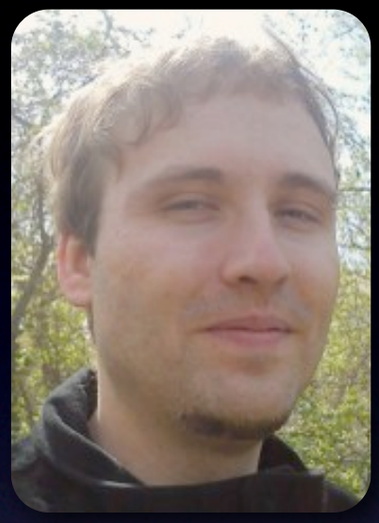
- Demo cloud instance with all **oqtans** tools
  - http://cloud.oqtans.org
- AMI at Amazon Web Services for EC2
  - Cloudman to launch any number of instances as a compute cluster

oqtans Availability: Cloud Computing

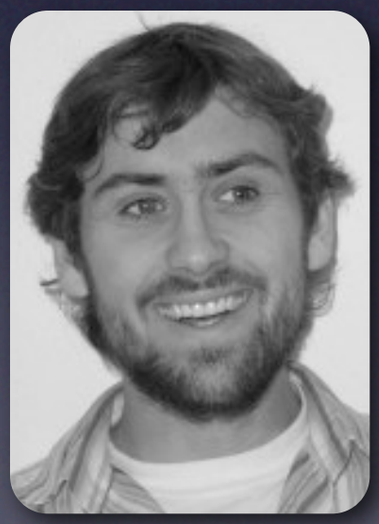


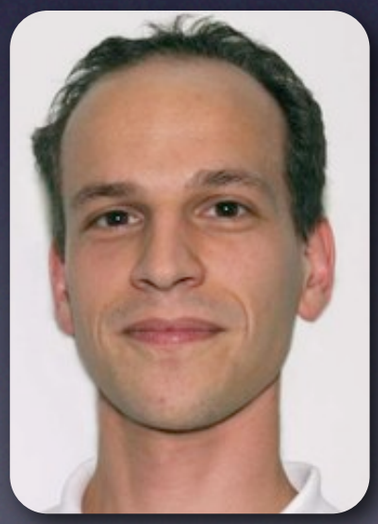
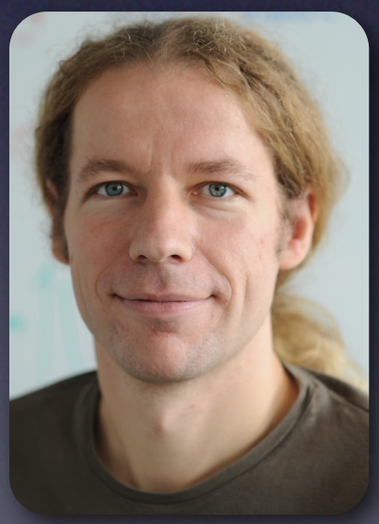
http://oqtans.org



Jonas Behr, Regina Bohnert, Philipp Drewe, Nico Görnitz



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