

Galaxy and the RNA Bioinformatics Center

Cameron Smith
Torsten Houwaart

Albert-Ludwigs-Universität Freiburg



UNI
FREIBURG

German network for bioinformatic infrastructure

Plant Bioinformatics

IPK Gatersleben
Helmholtz-Center München
Forschungszentrum Jülich

Human Bioinformatics

University Heidelberg
DKFZ Heidelberg
EMBL Heidelberg

Data Management

Heidelberg Institute f. theor. studies
University Rostock

Integrative Bioinformatics

University Bremen
Tech. University Braunschweig
DSMZ Braunschweig

RNA Bioinformatics

University Freiburg
University Leipzig
MDC Berlin

Integrative Bioinformatics

University Bochum
ISAS Dortmund

Proteomics

University Bochum
ISAS Dortmund

Microbial Bioinformatics

University Bielefeld
University Gießen



German network for bioinformatic infrastructure

Plant Bioinformatics

IPK Gatersleben
Helmholtz-Center München
Forschungszentrum Jülich

Human Bioinformatics

University Heidelberg
DKFZ Heidelberg
EMBL Heidelberg

Data Management

Heidelberg Institute f. theor. studies
University Rostock

Integrative Bioinformatics

University Bremen
Tech. University Braunschweig
DSMZ Braunschweig

RNA Bioinformatics

University Freiburg
University Leipzig
MDC Berlin

Integrative Bioinformatics

University Bochum
ISAS Dortmund

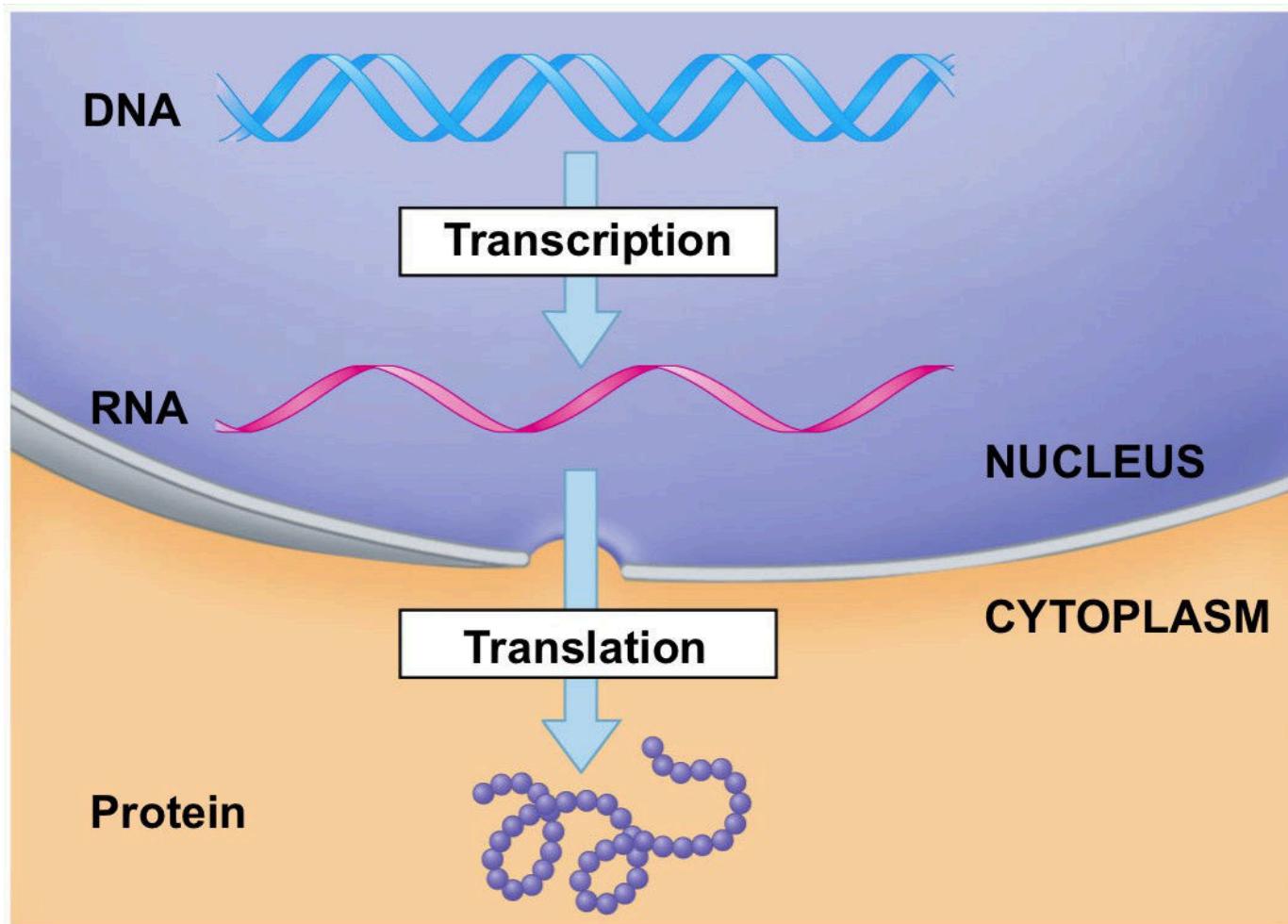
Proteomics

University Bochum
ISAS Dortmund

Microbial Bioinformatics

University Bielefeld
University Gießen

Central Dogma

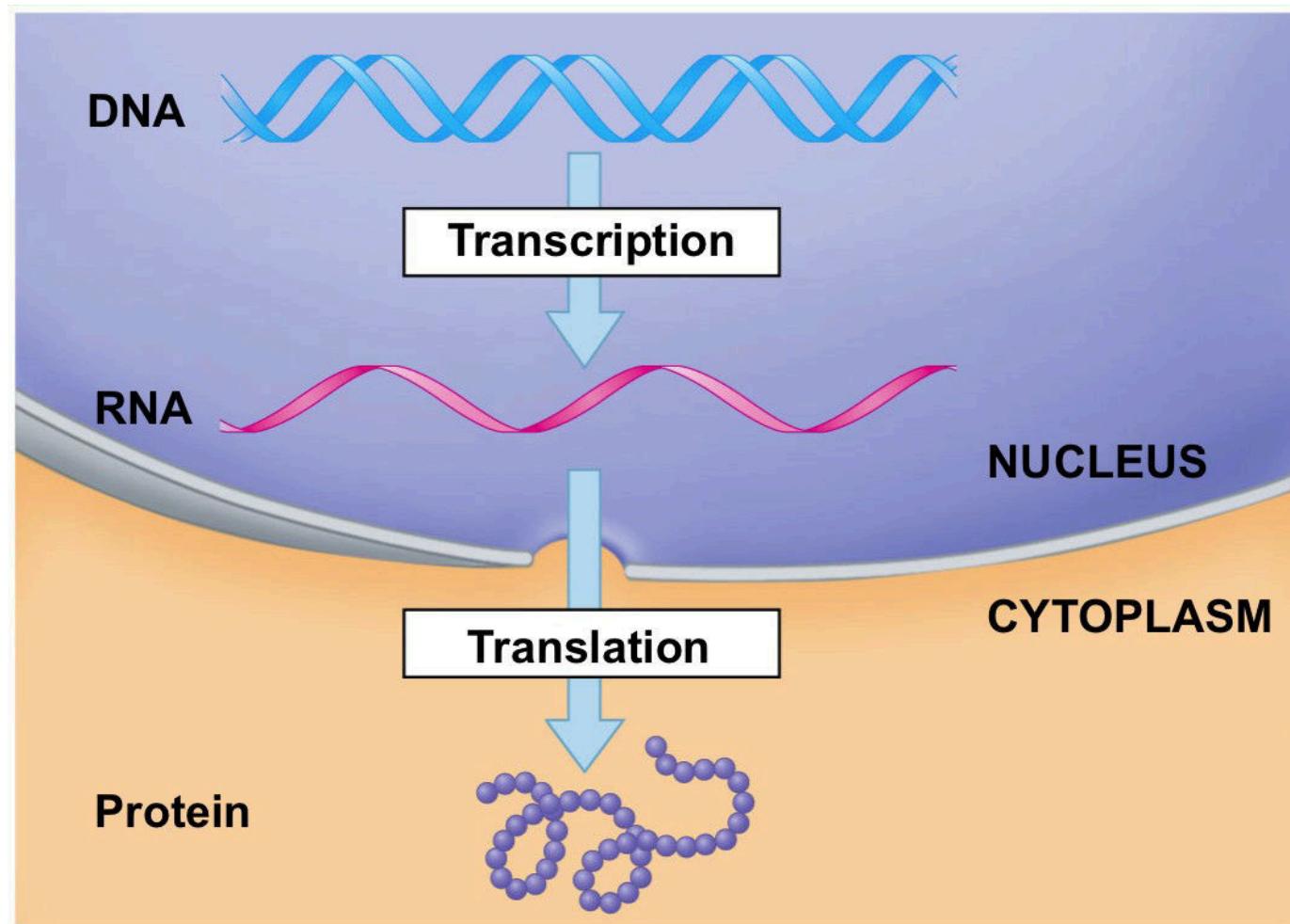


© 2012 Pearson Education, Inc.

Central Methods



Exomseq



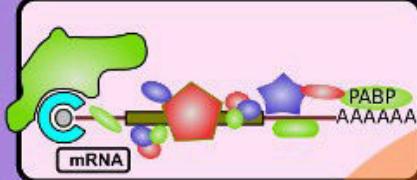
RNAseq

© 2012 Pearson Education, Inc.

Proteomics

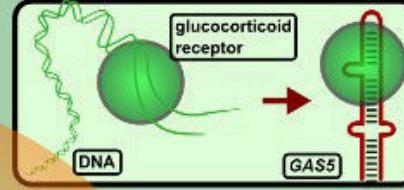
Centre for RNA Bioinformatics

RNA-protein networks



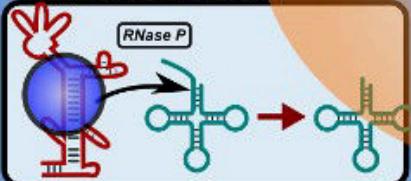
- analysis of CLIP-seq data (B,F)
- detection of RNA motifs (F,L)
- miRNA binding sites (B)

RNA as decoys



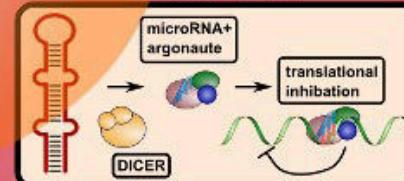
- RNA-protein interaction (B,F)
- RNA structure prediction (F,L)
- RNA conservation (F,L)

Ribozymes



- detection of ncRNAs (F,L)
- RNA structure prediction (F,L)
- RNA conservation (F,L)

Guide-RNA



- miRNA target prediction (B)
- ncRNA target prediction (F,L)
- miRNA/ncRNA detection (B,F,L)



RBC - MDC-Berlin & Leipzig



UNI
FREIBURG

MDC

MDC Berlin - Rajewsky

Webservice: **PicTar**

Databases: **doRiNA, circBase, pSILAC**

Tools: **miReduce, discrover**

MDC Berlin - Ohler

Tools: **RSVP, NASTIseq, S-Peaker, Orthologous enhancers**

MUMMIE, microMUMMIE, TargetThermo, WaRSwap



UNIVERSITÄT LEIPZIG

Uni Leipzig - Stadler

Databases: **tRNAdb** - Transfer RNA Database

Webservices: **DARIO, MITOS**

Tools: ~ **Vienna RNA** toolsuite



RBC - Freiburg



UNI
FREIBURG

RNA Tools Webserver (rna.informatik.uni-freiburg.de)

CopraRNA
CRISPRmap
LocARNA
IntaRNA
CARNA
MARNA
ExpaRNA
INFORNA
antaRNA

Freiburg RNA Tools

Welcome

This web server provides online access to a series of tools developed by the [Freiburg Bioinformatics Group](#). To start using it, please select from the listings below, or use the menu on the left. If you prefer doing a local installation on your machine, please visit our '[Download](#)' section.

If you use our tools for research or education, please cite the corresponding articles from the [Publications](#) section.

Version 3.4.7

Freiburg RNA Tools

CopraRNA

CopraRNA is a tool for sRNA target prediction. It computes whole genome predictions by combination of whole genome IntaRNA predictions using homologous sRNA sequences from distinct organisms.

CRISPRmap

CRISPRmap provides a quick and detailed insight into repeat conservation and diversity of both bacterial and archaeal systems.

MoDPepInt Server



RBC - Freiburg



UNI
FREIBURG



Galaxy Server (galaxy.uni-freiburg.de)

- > 200 registered users
- > 900 bioinformatics tools - a large portion of which are tools for RNA analysis.
- > 2K cores, 1TB RAM.
- > 30 published workflows, histories and data libraries
- > pages/tutorials for training and education

 `galaxytools` github.com/bgruening/galaxytools

- > 300 galaxy wrappers
- ~25 contributors
- new RNA tools into the Freiburg Galaxy Server and `galaxy-rna-workbench`.

Integration of tools

Manually

Cameron Smith, Torsten Houwaart, Björn Grüning

→ Students

→ external contributions

- Nicola Soranzo
- Greg von Kuster
- Youri Hoogstrate
- Eric Rasche

...

(migration from Trello to git issues for better visibility)

Automatic generation of wrappers

Python **argparse** to Galaxy by Eric Rasche

Java **Javadocs** to Galaxy by Sarah Alfeat

XML **CTD** (Command Tool Description) to Galaxy by L. de la Garza and T. Houwaart

Perl **Getopt::Long** to Galaxy

CommonWorkflowLanguage

workflows

Build, Test, Use, Share

RNAseq

- Quality Control
- Differential Expression Analysis

Exomseq

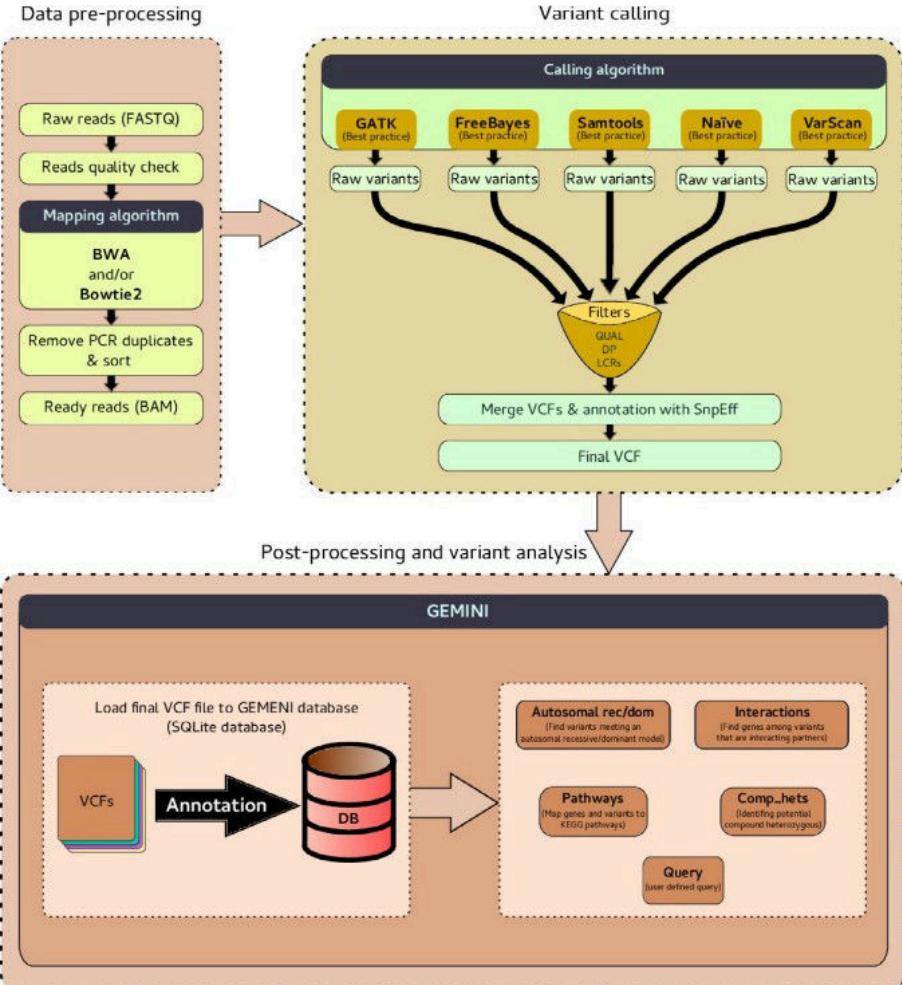
- Variant calling

Proteomics

- OpenMS toolsuite

→ RNAseq + Proteomics

- individual workflows /
iterative process



RBC-services: workflows

workflows

Build, Test, Use, Share

RNAseq

- Quality Control
- Differential Expression Analysis

Exomseq

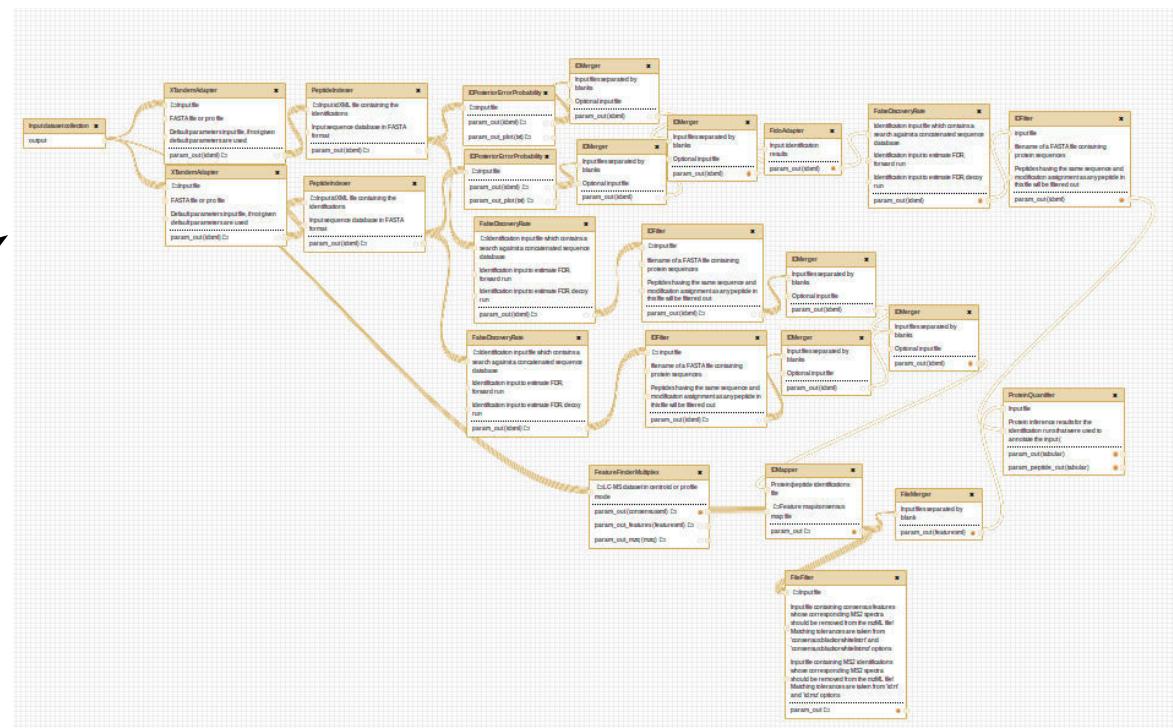
- Variant calling:

Proteomics

- OpenMS toolsuite

→ RNAseq + Proteomics

- individual workflows /
iterative process

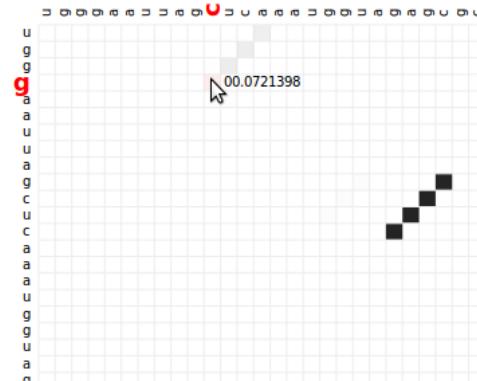
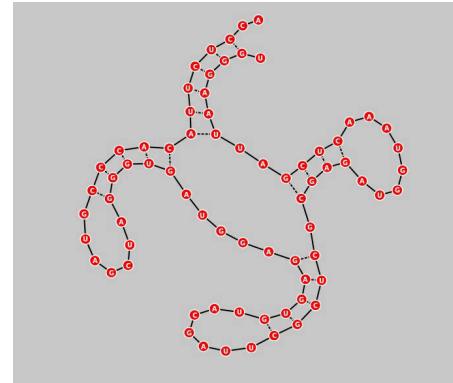
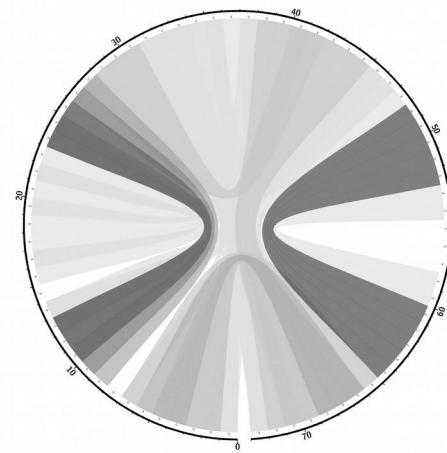


visualization

RNA secondary structure

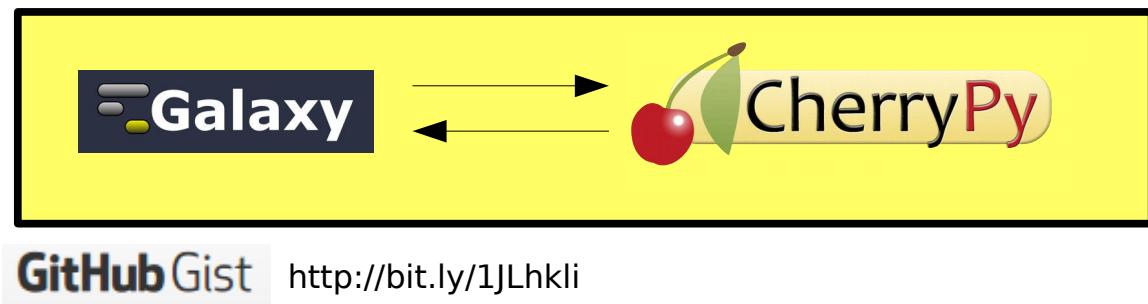
Graph (within a force field, D3 - js library) DotPlot Matrix interactive

→ Circos to see competing RNA basepair motifs



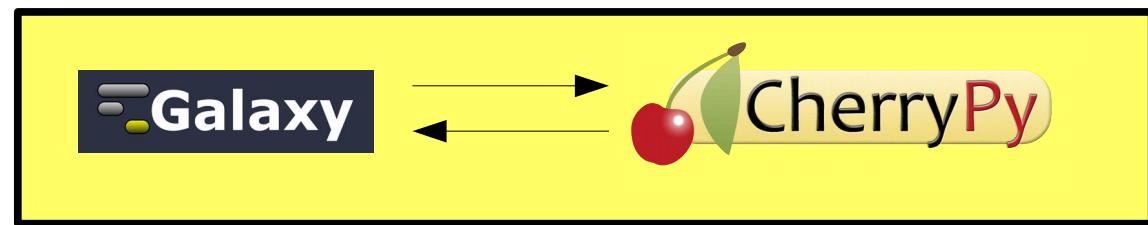
integration

How to integrate a database into Galaxy?



integration

How to integrate a database into Galaxy?



GitHub Gist <http://bit.ly/1JLhkli>

MDC Berlin

doRiNA

05.05.2015 I talked to Ricardo [...] he is on it right now. We will contact you if we are stuck at some point.

08.05.2015 dorina repository on the toolshed

training

1-week Galaxy NGS workshop - every 6 months
25+ participants from Freiburg and Germany wide

- Introduction to Galaxy
- Chipseq
- Exomseq
- RNAseq
- Bring your own data/problems

On demand training

Bled Slovenia, Galaxy DevDay, Genome annotation workshop, iClip RNA workshop,
Swiss-German Galaxy workshop

Training Material

Docker bundled „training benches“, video tutorials, Galaxy pages, published histories and workflows, data libraries

In a box



Docker Image - automated build

<https://registry.hub.docker.com/u/bgruening/galaxy-rna-workbench/>



Github Repository - push triggers docker build

<https://github.com/bgruening/galaxy-rna-workbench>



Ubuntu 14.04 based

https://registry.hub.docker.com/_/ubuntu/



Galaxy (stable) - pulled and pimped

<https://github.com/galaxyproject/galaxy>



galaxy-rna-workbench



UNI
FREIBURG

Use:



```
$ docker run -p 8080:80 bgruening/galaxy-rna-workbench
```



Browse to <http://localhost:8080>, ready to galaxy!



UNI
FREIBURG

galaxy-rna-workbench

Why?

- Automated build
open and reliable
- Scalable
from netbook to cluster
- Provides security
spin it up, do your work, shut it down
- Actively developed
released concurrently with Galaxy

bgruening / galaxy-stable

Galaxy Docker Image

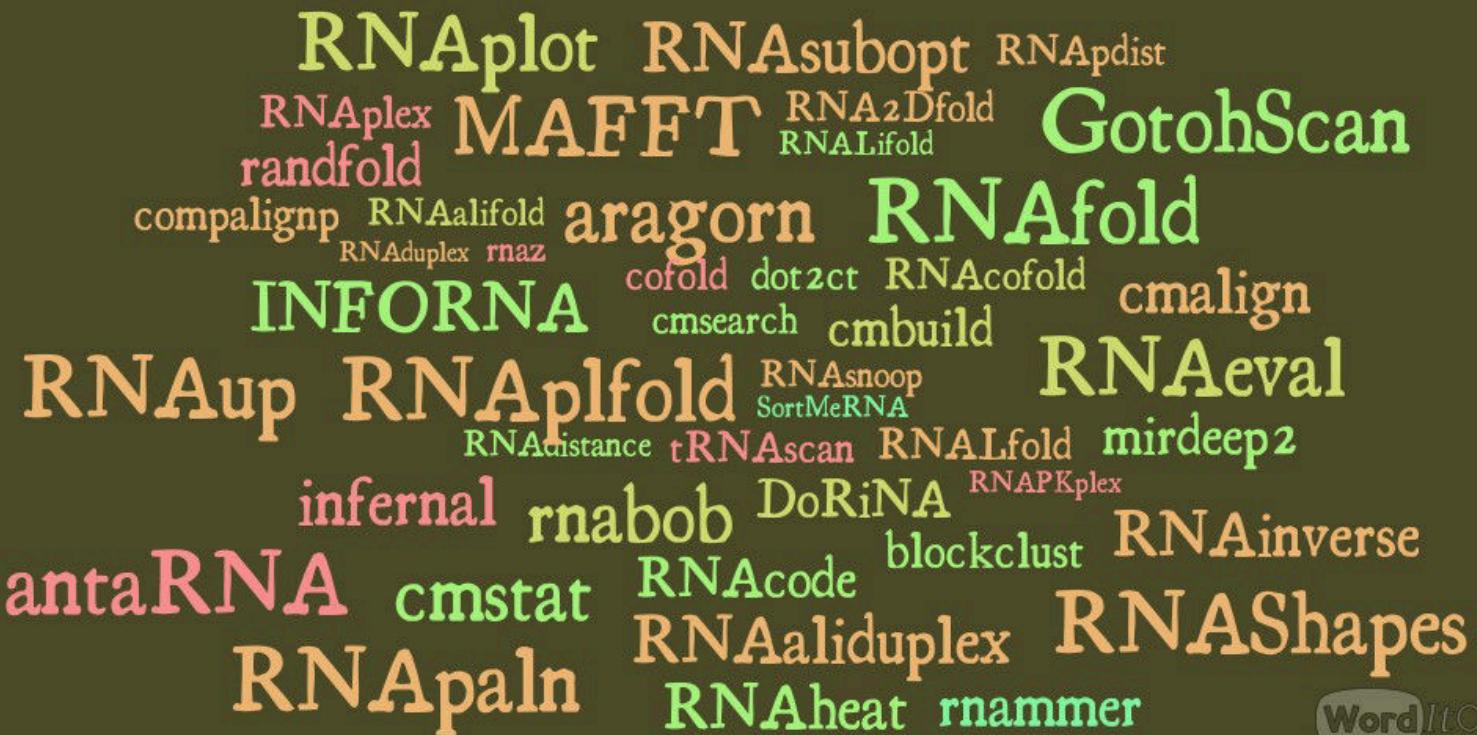
★ 8 0 2326

bgruening / galaxy-rna-workbench

No description set

★ 2 0 124

Tools: 50+



The word cloud displays approximately 50 RNA-related tools, including:

- RNAplot, RNAsubopt, RNApdist
- RNAplex, MAFFT, RNA2Dfold, RNALifold, GotohScan
- randfold
- compalignp, RNAalifold, aragorn, RNAfold
- RNAAduplex, rnaz
- INFORNA, cofold, dot2ct, RNAcofold, cmsearch, cmbuild, cmalign
- RNAup, RNAlfold, RNAsnoop, SortMeRNA, RNAsnooper
- RNAeval, RNAdistance, tRNAscan, RNALifold, mirdeep2
- infernal, rnabob, DoRiNA, RNAPKplex
- antaRNA, cmstat, RNACode, blockclust, RNAinverse
- RNAliduplex, RNAShapes
- RNApaln, RNAheat, rnammer

WordItOut



galaxy-rna-workbench



UNI
FREIBURG

Utilities! via the bioblend API [doc++](#)

Fetch and index

Specify a set of reference genomes and indexers,
the genomes will be fetched and indexed during the image
build process - and immediately available at runtime

Populate data libraries

Functional test data from workbench tools and arbitrary
folders, auto-bundled as galaxy data libraries

Web-based image configuration

Allowing a user to specify tools, genomes, indexers and
data to be included in a custom workbench
→ exporting a docker image

Interactive Environments

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 54.4 MB

IP[y]: Notebook ipython_galaxy_notebook (unsaved changes)

File Edit View Insert Cell Kernel Help

Cell Toolbar: None

Welcome to the interactive Galaxy IPython Notebook.

You can access your data via the dataset number. For example, handle = get(42). To save data, write your data to a file, and then call put('filename.txt'). The dataset will then be available in your galaxy history. To save your notebook to galaxy, click the large green button at the top right of the IPython interface.

In [2]:

```
import numpy as np
import matplotlib.pyplot as plt
```

In [4]:

```
data = np.loadtxt( get(44), delimiter='\t', skiprows=1, usecols=[5])
plt.plot(values)
plt.savefig('./figure.png')
plt.show()
```

In [1]:

```
import csv

values = list()
with open( get(44) ) as handle:
    for row in csv.DictReader( handle, delimiter='\t', quoting=csv.QUOTE_NONE ):
        if row['logFC'].isdigit():
            values.append( row['logFC'] )
values.sort()
plt.plot(values)
plt.savefig('./logFC_low.png')
plt.show()
```

History

ipython_galaxy_notebook.ipynb

figure.png

example data

4,001 lines format: tabular, database: ?

Charts degust IPython Scatterplot Trackster Venn Diagramm

41: example-data.csv

39: Sort on data 38

38: Group on data 37

37: Motif finder on data 19

35: Select on data 33

34: Select on data 33

33: Sort on data 30

32: Sort on data 31

31: Group on data 30

30: Motif finder on

localhost:8080/visualization/show/python?dataset_id=c385e49b9fe1853c



Galaxy powered RNA research



UNI
FREIBURG

We are building an RNA focused
research **community**

Join Us! and contribute:

- Tools
- Workflows
- Data
- Documentation

rbc@informatik.uni-freiburg.de - admin

galaxy@informatik.uni-freiburg.de - development

acknowledge:

German Federal Ministry of Education and Research - BMBF grant 031 A538A RBC (de.NBI)