

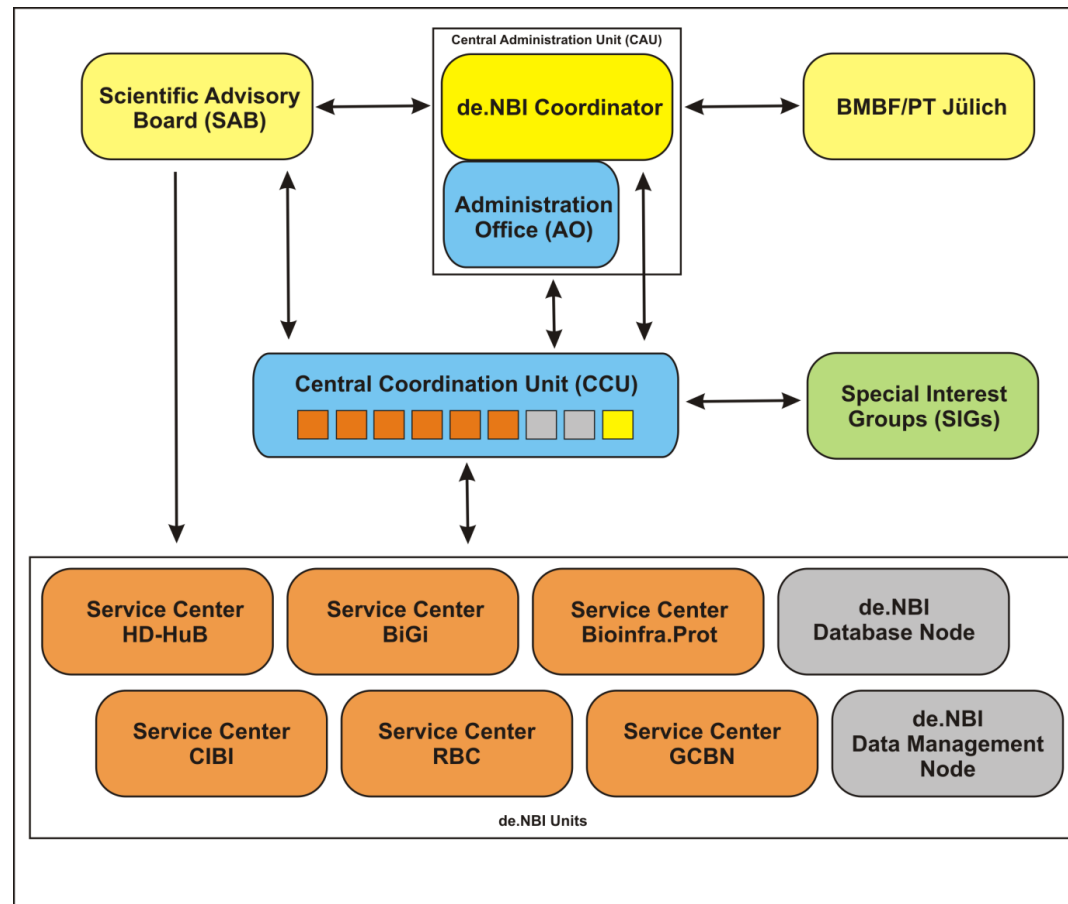
Analyze more, process less:

Galaxy in Freiburg

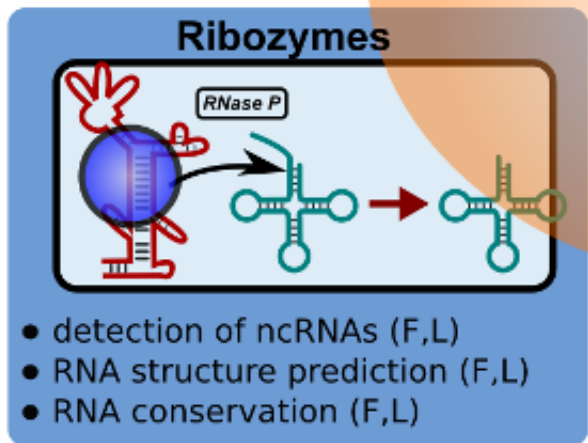
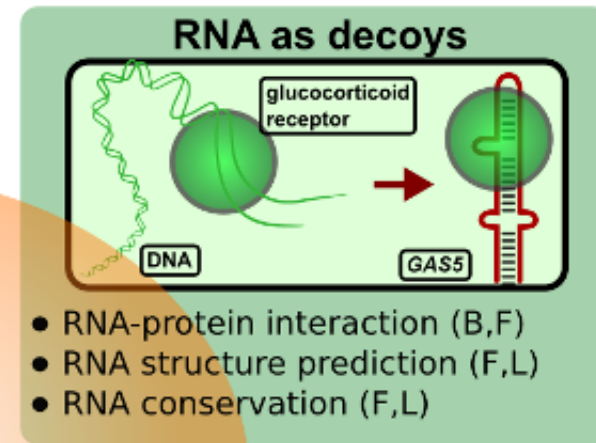
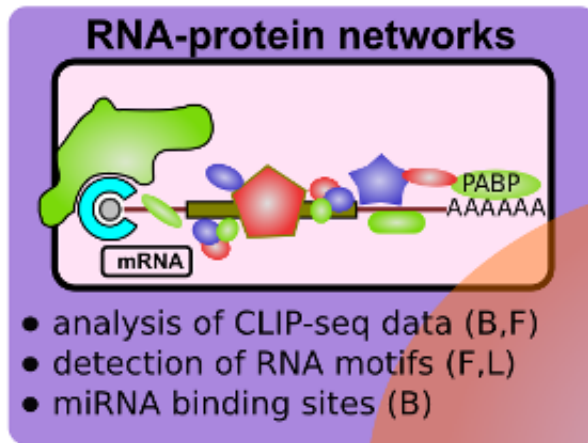
September 2016 – Anika Erxleben – Freiburg Galaxy Team

Who?

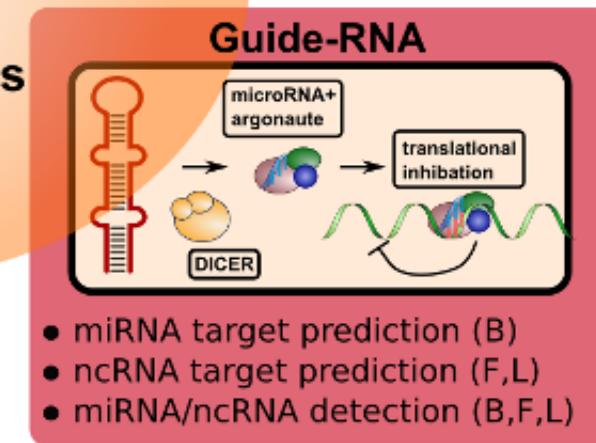
We are part of the German Network for Bioinformatics Infrastructure (de.NBI) & ELIXIR



RBC – RNA Bioinformatics Center



Centre for
RNA
Bioinformatics



What do we offer

- **Reproducible** one-click ready-to-use pipelines
- **Easy access** to data, tools and protocols
- Easy **deployment** of tools and protocols
- **Data** integration and management +SRA, Genbank upload, UniProt, ...
- Data **visualization**
- **Training!**
- Compute resources

What is



?

- An analysis and data integration platform
- A workflow management system
- Open source, community driven software that makes integration of your favourite tools simple

<http://galaxyproject.org>

Our Aim

“Enable accessible, reproducible, and transparent computational research.”

accessible – web interface

Galaxy / Uni Freib

Analyze DataWorkflowShared DataVisualizationAdminHelpUser

Using 45.2 GB

Tools

search tools

Get Data

Send Data

Lift-Over

Text Manipulation

Filter and Sort

Filter data on any column using simple expressions

Sort data in ascending or descending order

Select lines that match an expression

GFF

Extract features from GFF data

Filter GFF data by attribute using simple expressions

Filter GFF data by feature count using simple expressions

Filter GTF data by attribute values list

Select lines by word list

Join, Subtract and Group

Unix Tools

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Bowtie2 (version 0.1)

Is this library mate-paired?:
Single-end

FASTQ file:
13: C0YH1ACXX_TACPCM...equence.txt

Nucleotide-space: Must have Sanger-scaled quality values with ASCII offset 33

Will you select a reference genome from your history or use a built-in index?:
Use a built-in index

Built-ins were indexed using default options

Select a reference genome:
Drosophila melanogaster: dm3

If your genome of interest is not listed, contact the Galaxy team

Parameter Settings:
Use Defaults

You can use the default settings or set custom values for any of Bowtie's parameters.

Specify the read group for this file?:
No

Execute

History

mm9
44.8 GB

91: Plotting on data 69, data 69, and data 75: archive

75: Find differentially methylated region (DMR) on data 72 and data 72: tiling regions
32 regions
format: interval, database: mm9
The much faster fisher library is not installed. Fallback to scipy.
display at Ensembl Current

72: Filtering on data 70 and data 70: affected

71: Filtering on data 70 and data 70: control

70: Merging on data 69: merged

1.Chrom	2.Start	3.End	4	5
chrY	32276	2892117	526	0
chr4	3006166	155629563	177294	0
chr12	3000068	121255743	121795	0
chr15	3002861	103485270	117970	0
chr10	3001261	129991927	140941	0
chr3	3000813	159599241	143455	0

accessible – get your data

Tools
search tools
Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BX table browser
- EBI SRA ENA SRA
- Get Microbial Data
- BioMart Central server
- BioMart Test server
- CBI Rice Mart rice mart
- GrameneMart Central server
- modENCODE fly server
- Flymine server
- Flymine test server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- metabolicMine server
- modENCODE worm server
- WormBase server

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Insect **genome:** D. melanogaster **assembly:** Apr. 2006 (BDGP R5/dm3)

group: Genes and Gene Prediction Tracks **track:** RefSeq Genes [add custom tracks](#) [track hubs](#)

table: refGene [describe table schema](#)

region: ☒ genome ☐ position chr2L:585265-1649382 [lookup](#) [define regions](#)

Identifiers (names/accessions): [paste list](#) [upload list](#)

filter: [create](#)

Intersection: [create](#)

correlation: [create](#)

output format: BED - browser extensible data [Send output to](#) ☒ [Galaxy](#) ☐ [GREAT](#)

output file: (leave blank to keep output in browser)

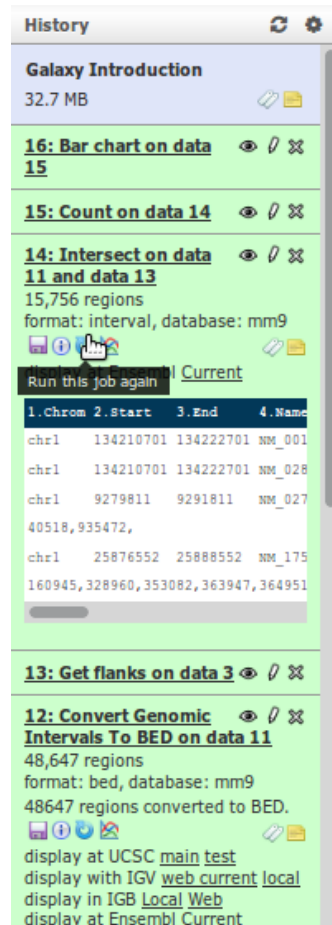
file type returned: ☒ plain text ☐ gzip compressed

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

“Enable accessible, reproducible, and transparent computational research.”

reproducible – reload experiments



History

Galaxy Introduction
32.7 MB

16: Bar chart on data 15

15: Count on data 14

14: Intersect on data 11 and data 13
15,756 regions
format: interval, database: mm9
[Run this job again](#) [Current](#)

1. Chrom	2. Start	3. End	4. Name
chr1	134210701	134222701	NM_001
chr1	134210701	134222701	NM_028
chr1	9279811	9291811	NM_027
	40518,935472,		
chr1	25876552	25888552	NM_175
	160945,328960,353082,363947,364951		

13: Get flanks on data 3

12: Convert Genomic Intervals To BED on data 11
48,647 regions
format: bed, database: mm9
48647 regions converted to BED.
[display at UCSC main test](#)
[display with IGV web current local](#)
[display in IGB Local Web](#)
[display at Ensembl Current](#)

reproducible – reload experiments

History

Galaxy Introduction
32.7 MB

16: Bar chart on data 15

15: Count on data 14

14: Intersect on data 11 and data 13
15,756 regions
format: interval, database: mm9

13: Get flanks on data 3

12: Convert Genomic Intervals To BED on data 11
48,647 regions
format: bed, database: mm9
48647 regions converted to BED.

display at UCSC [main](#) [test](#)
display with IGV [web](#) [current](#) [local](#)
display in IGB [Local](#) [Web](#)
display at Ensembl [Current](#)

Intersect (version 1.0.0)

Return:

Overlapping Intervals

(see figure below)

of:

13: Get flanks on data 3

First dataset

that intersect:

11: Find and Replace on data 10

Second dataset

for at least:

1

(bp)

Execute

TIP: If your dataset does not appear in the pulldown menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.

Screencasts!

See Galaxy Interval Operation [Screencasts](#) (right click to open this link in another window).

Syntax

Where overlap is at least sets the minimum length (in base pairs) of overlap between elements of the two datasets

Overlapping Intervals returns entire intervals from the first dataset that overlap the second dataset. The returned intervals are completely unchanged, and this option only filters out intervals that do not overlap with the second dataset.

Overlapping pieces of Intervals returns intervals that indicate the exact base

reproducible – tools versioning

Cuffdiff

Transcript

99: Cuffmerge on data 89, data 19, and others: merged transcripts

A transcript GFF3 or GTF file produced by cufflinks, cuffcompare, or other source.

✗ This job was initially run with tool id "Filter1", version "1.1.0", which is not currently available. You can rerun the job with this tool, which is a derivation of the original tool.

reproducible - workflows

Tools

search tools

Get Data

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Filter and Sort

- Filter data on any column using simple expressions
- Sort data in ascending or descending order
- Select lines that match an expression

GFF

- Extract features from GFF data
- Filter GFF data by attribute using simple expressions
- Filter GFF data by feature count using simple expressions
- Filter GTF data by attribute values list
- Select lines by word list

Join, Subtract and Group

Unix Tools

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Workflow Canvas | Workflow constructed from history 'Galaxy Introduction'

```
graph LR; A[Input dataset] --> C[Intersect]; B[Input dataset] --> C; C --> D[Count]; D --> E[Bar chart]; E --> F[Dataset];
```

The workflow diagram shows two input datasets feeding into an 'Intersect' tool. The output of the 'Intersect' tool feeds into a 'Count' tool. The output of the 'Count' tool feeds into a 'Bar chart' tool. The 'Bar chart' tool has a 'Dataset' output labeled 'out_file1 (png)'.

Details

Tool: Bar chart

Version: 1.0.0

Dataset

Data input 'input' (tabular)

Use X Tick labels?: Yes

Use this column for X Tick labels: 2

Numerical columns: 1

Plot title: My first galaxy bar plot

Label for Y axis: number of genes

Minimal value on Y axis: 0

Maximal value on Y axis: 0

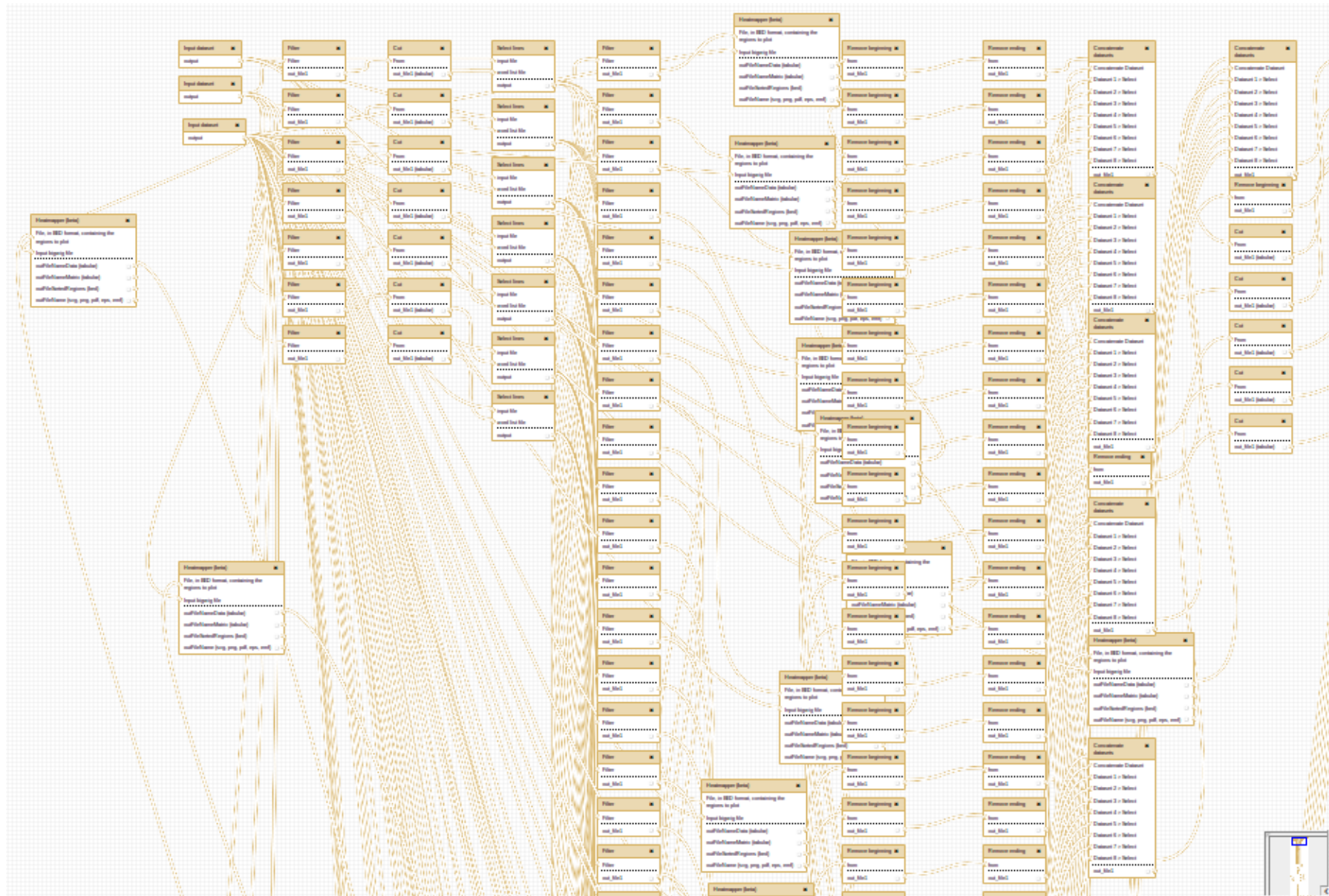
Choose chart size (pixels): Normal: 800 by 600

Edit Step Actions

Rename Dataset

out_file1 Create

reproducible - workflows



“Enable accessible, reproducible, and
transparent computational research.”


transparent – sharing everything

Share or Publish History 'Aspergillus NCBI Submission'

Making History Accessible via Link and Publishing It

This history is currently **accessible via link**.

Anyone can view and import this history by visiting the following URL:

<http://galaxy.pharmaceutical-bioinformatics.org/u/bjoern/h/aspergillus-ncbi-submission> 

You can:

[Disable Access to History Link](#)

Disables history's link so that it is not accessible.

[Publish History](#)

Publishes the history to Galaxy's [Published Histories](#) section, where it is publicly listed and searchable.

Sharing History with Specific Users

The following users will see this history in their history list and will be able to view, import, and run it.

Email

[wolfgang.huettel@pharmazie.uni-freiburg.de](#) ▾

[loubna.youssar@pharmazie.uni-freiburg.de](#) ▾

[Share with another user](#)

[Back to Histories List](#)

transparent – for other researchers

Windshield splatter analysis with the Galaxy metagenomic pipeline

Sergei Kosakovsky Pond^{1,2,6,9}, Samir Wadhawan^{3,6,7}, Francesca Chiaromonte⁴,
Guruprasad Ananda^{1,3}, Wen-Yu Chung^{1,3,8}, James Taylor^{1,5,9},
Anton Nekrutenko^{1,3,9} and The Galaxy Team¹

✚ Author Affiliations

✚ Author Notes

^{✚6} These authors contributed equally to this work.

Abstract

Footnotes

[Supplemental material is available online at <http://www.genome.org>. All data and tools described in this manuscript can be downloaded or used directly at <http://galaxyproject.org>. Exact analyses and workflows used in this paper are available at <http://usegalaxy.org/u/aun1/p/windshield-splatter>.]

Article published online before print. Article and publication date are at first publication that features a live online supplement providing access to exact analyses and workflows used in the article.

Footnotes

[Supplemental material is available online at <http://www.genome.org>. All data and tools described in this manuscript can be downloaded or used directly at <http://galaxyproject.org>. Exact analyses and workflows used in this paper are available at <http://usegalaxy.org/u/aun1/p/windshield-splatter>.]

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/dol/10.1101/gr.094508.109>.

“Enable accessible, reproducible, and transparent computational research.”

Tools, tools, tools



Freiburg Galaxy instance

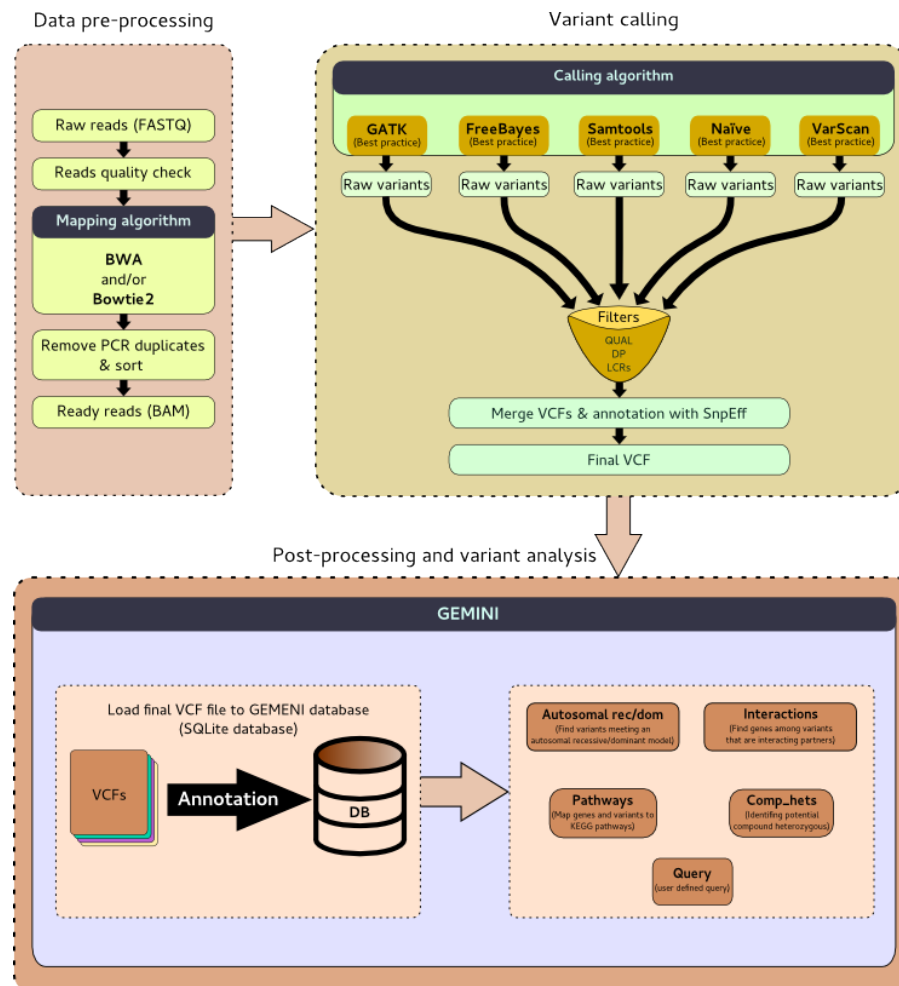
- RNA-seq
- ChIP-seq
- MethylC-seq
- Exome-seq
- Genome annotation
- Metagenomics
- Proteomics
- ... and more

>2000 tools

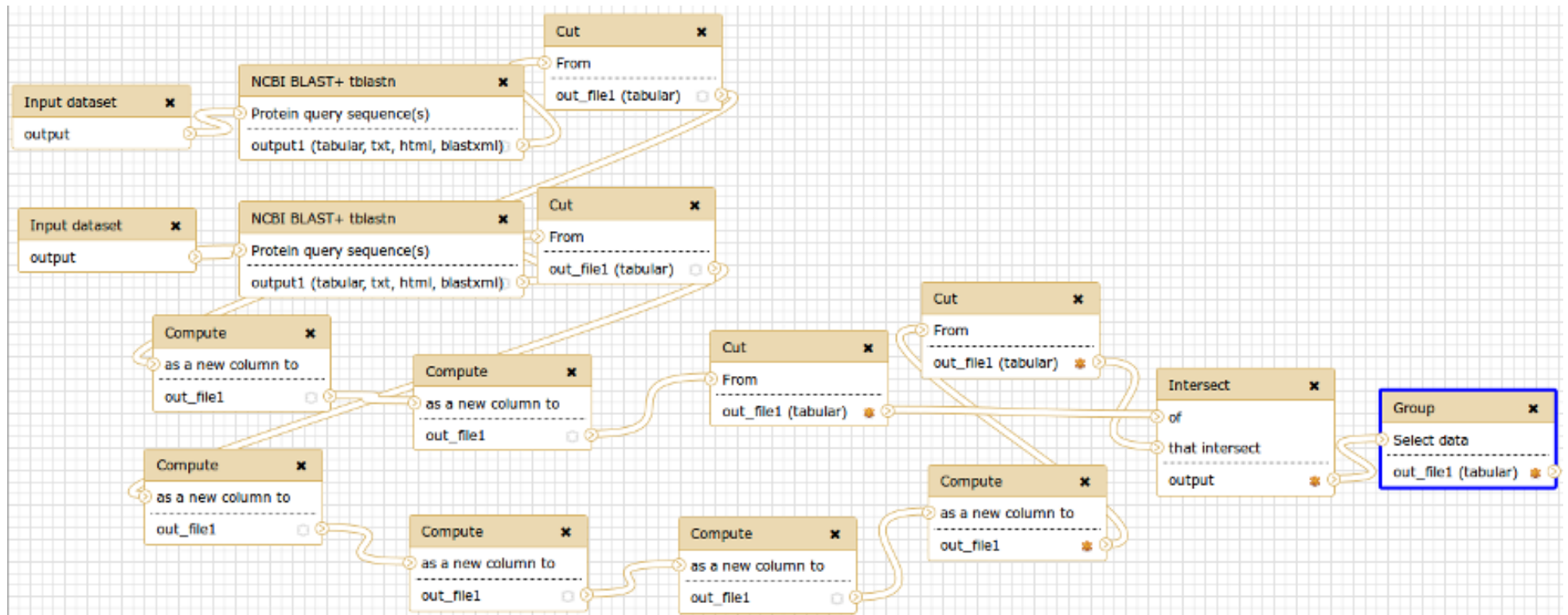
- Text manipulation
- Format converters
- Filtering and sorting
- Join, subtract, group
- Unix tools
- Alignment tools
- Mapping
- Statistics
- Plotting
- EMBOSS
- WES/WGAS
- NCBI Blast+
- Blast2GO
- DAVID
- ChemicalToolBoX
- ... and many more

genVAST variant calling (SNPs)

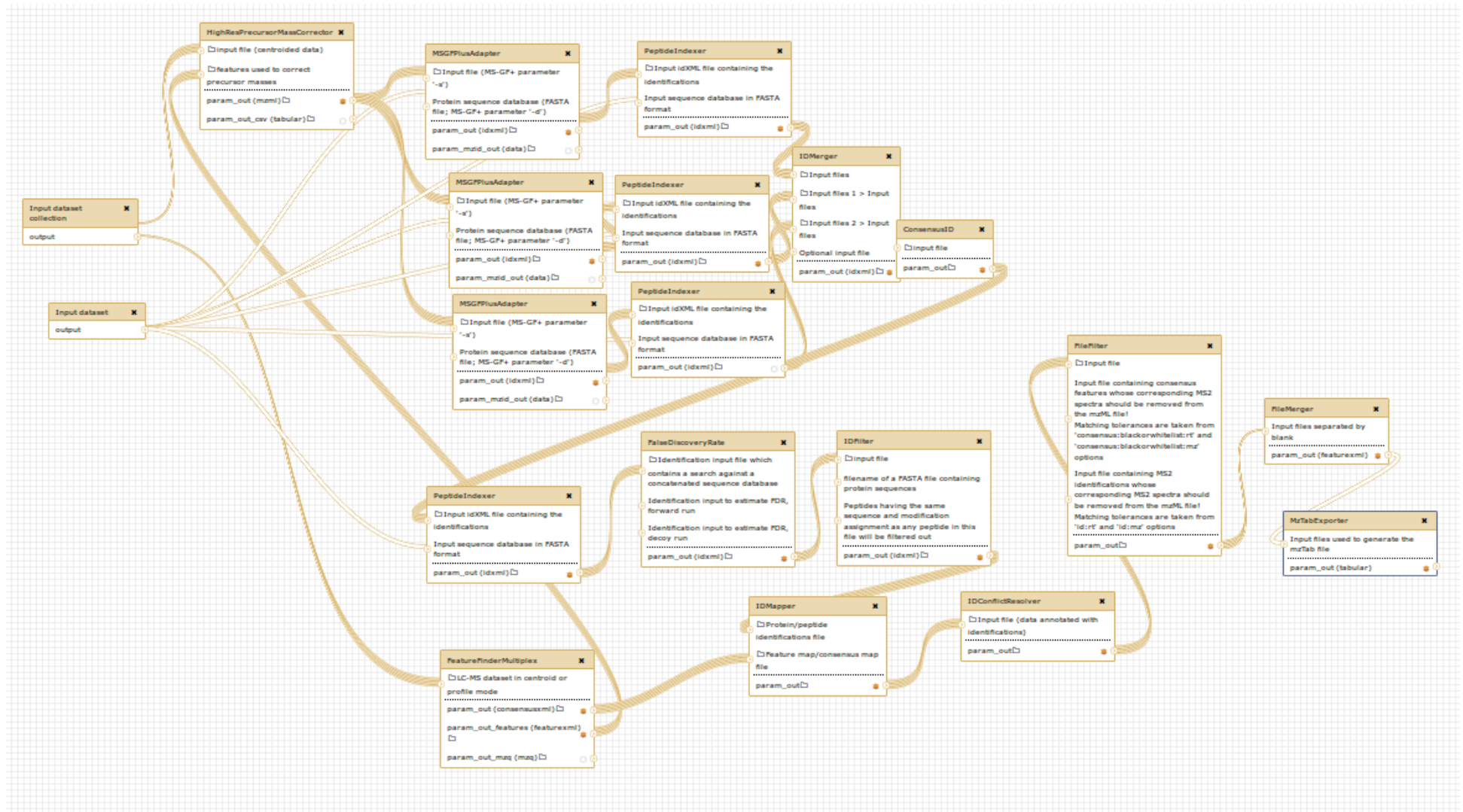
- Pipeline for exome-seq data analysis



BLAST+ and genome annotation



Proteomics

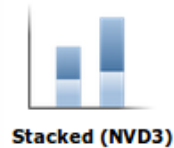
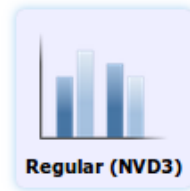


Galaxy / Uni Freiburg Analyze Data Workflow Shared Data Visualization Admin Help User Using 1021.1 GB

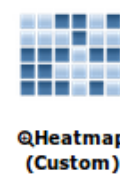
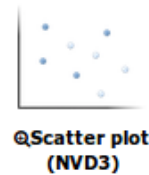


visualization - charts

- Bar diagrams



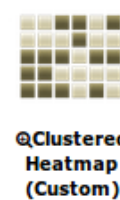
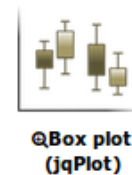
- Others



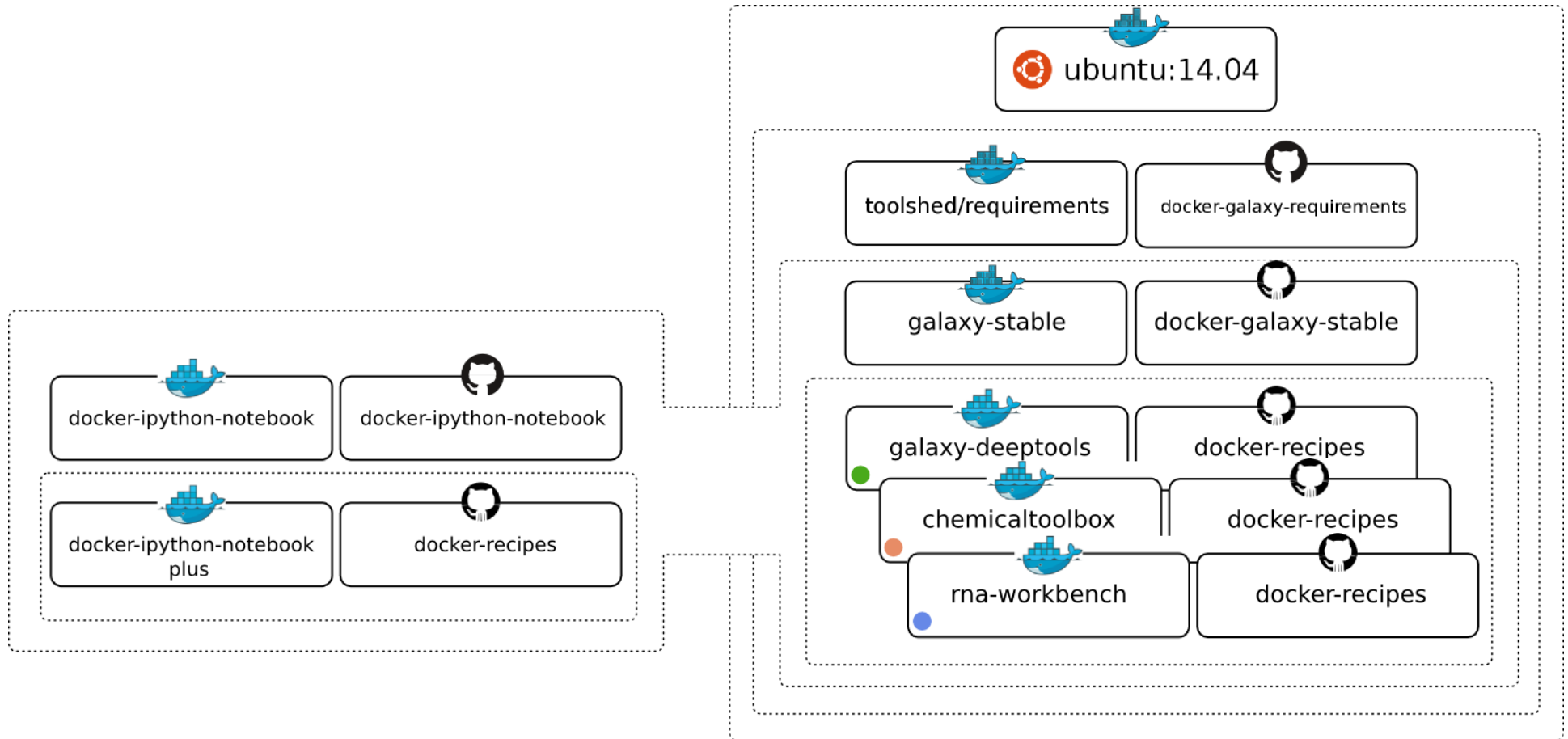
- Area charts



- Data processing (requires 'charts' tool from Toolshed)



Virtualized Galaxy flavours

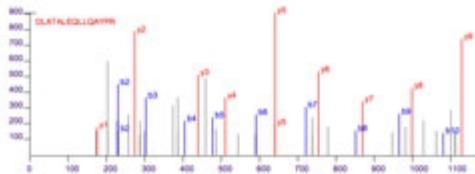
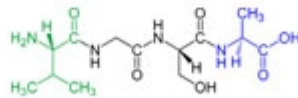
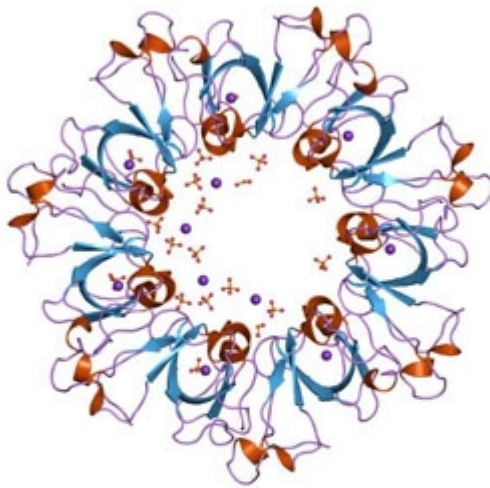
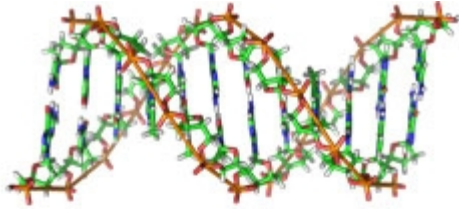


<https://www.docker.io>

Freiburg Galaxy stats

- 360 registered users
- 48 different groups
- ~1000 different tools
- 110,000 jobs per year
- 10 workshops with ~240 participants

It's there!



- Genomics
- Proteomics
- Metabolomics
- Imaging, literature sciences, engineering, ...

Training!



- Galaxy HTS data analysis workshops
 - Developers day & international workshops
 - GTN (Galaxy Training Networks)
 - iPython for teaching
 - Galaxy for lecturing
-
- Training material, protocols & slides online on github!
<https://github.com/bgruening/trainings-material>
 - Galaxy interactive tours

Community is the key

- ~2900 publications since 2010, 1600 articles from 2015-2016
- 75 known public instances
- <https://biostar.usegalaxy.org> (2016: 4600 active users)
- ~1200 registered developers (galaxy-dev mailing list)
- ~400 mails per month
- Annual international community conference
- 2 positions for community-related questions

Thank you!

Freiburg Galaxy Team

Dr. Björn Grüning

- Dr. Torsten Houwaart
 - Dr. Berenice Batut
 - Clemens Blank
 - Joachim Wolff
 - Prof. Rolf Backofen
-
- galaxy@informatik.uni-freiburg.de
 - <http://www.bioinf.uni-freiburg.de>