Community-driven training for biological data analysis with the Galaxy Training Network

Picture from Bérénice Batut - Icons from the Noun Project and Flaticon

Bérénice Batut

Galaxy Africa - April 2018

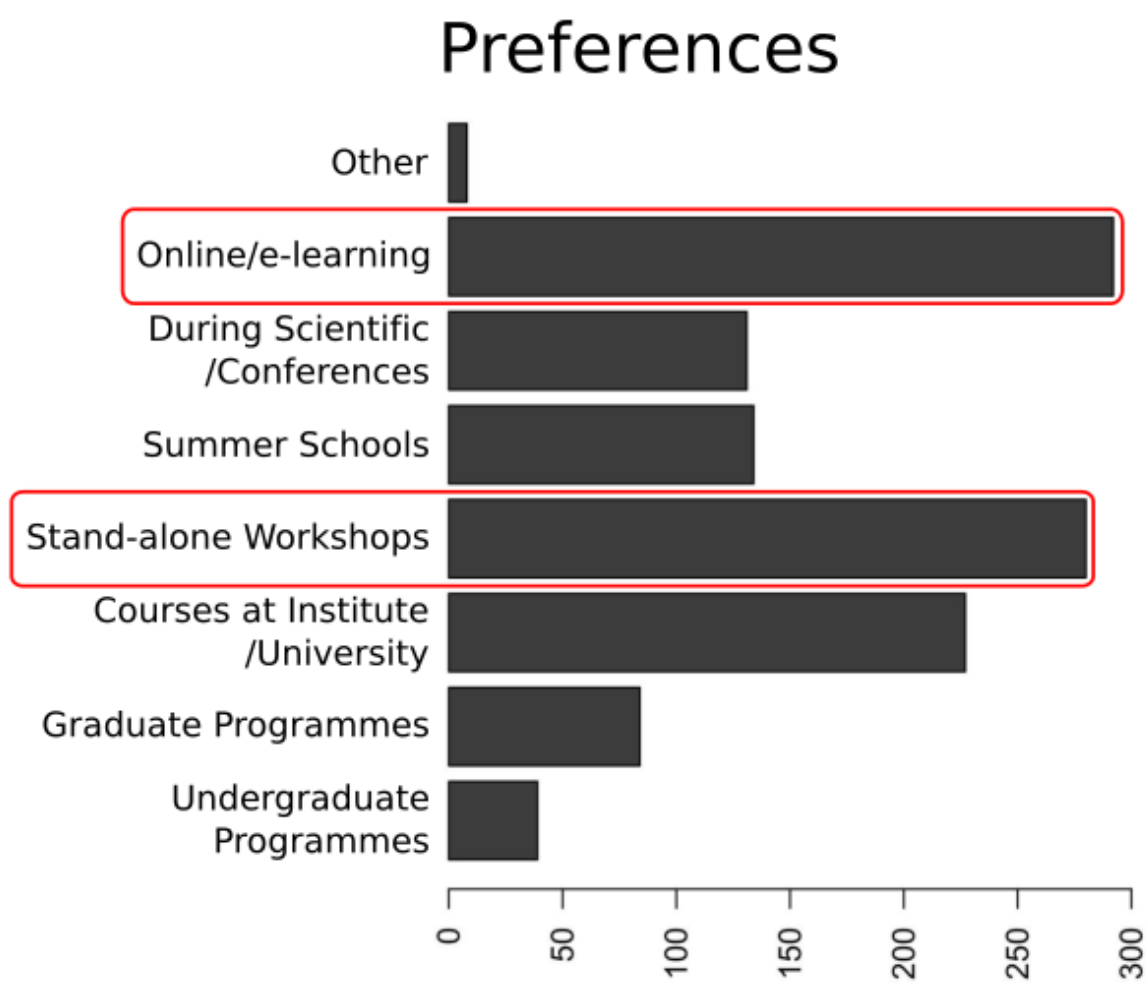
Why caring about bioinformatics training?

Need for bioinformatic training

Bioinformatics has become too central to biology to be left to specialist bioinformaticians

- Explosion of data to analyze
- Access to computational power
- Thousand of possible tools for specialized analyses

An increasing demand for learning bioinformatics

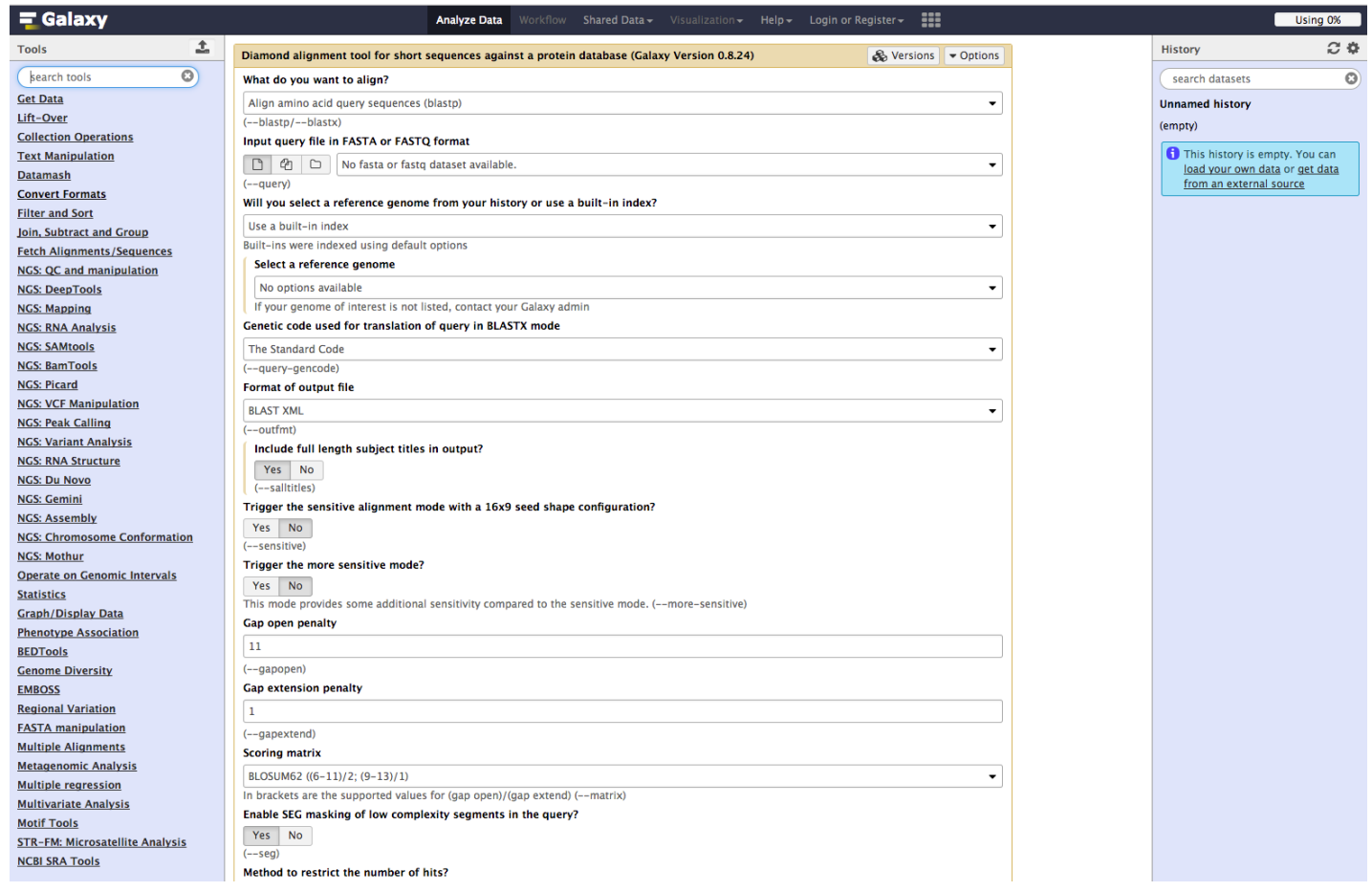


Graphs of Brazas et al, 2017

Galaxy

a great solution !

Computational knowledge: Not required!



- Web interface for numerous bioinformatics tools
- Scalable
- No issue with computer configuration during training

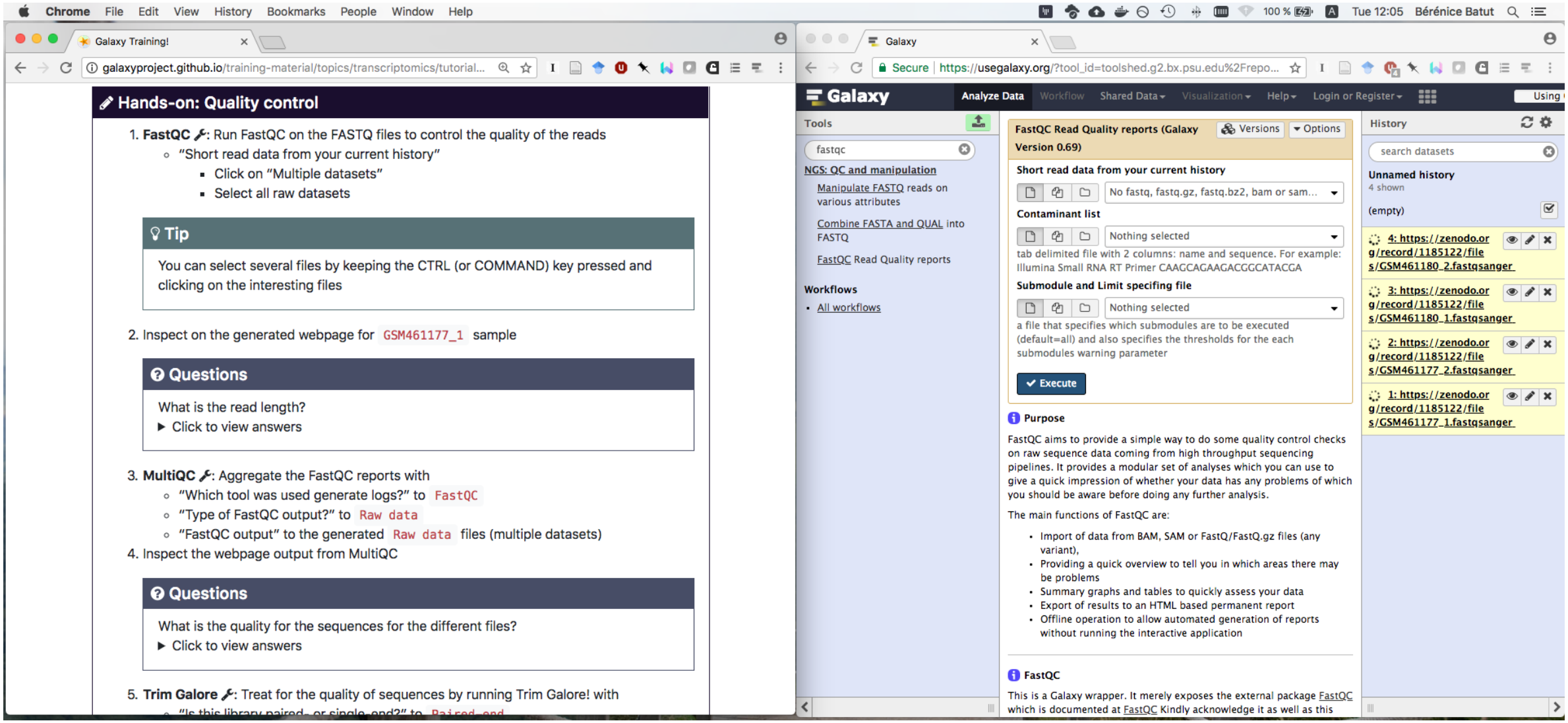

Training
ressources
Best
Practices
Galaxy
Training
Network
Platform
recommendations
Trainer
Directory

Building an infrastructure facilitating data analysis training in life sciences

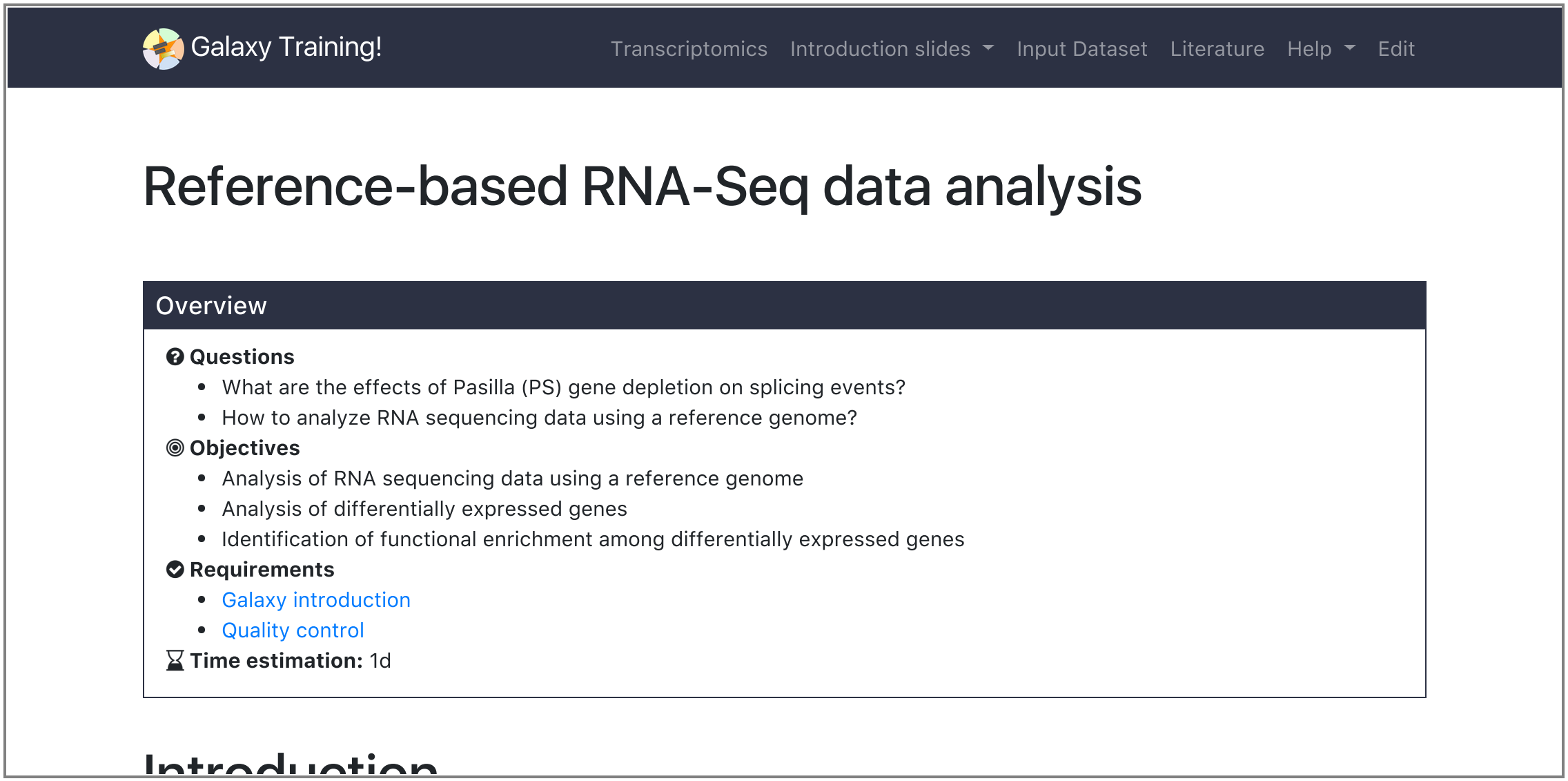
Requirements for a training infrastructure

- Interactive learning platform
- Support for current research problems
- Usable for effective training for individual users & instructors
- Community driven (content creation and maintenance)
- FAIR: Findable, Accessible, Interoperable, Reusable
- Open

Interactive learning via hands-on tutorials

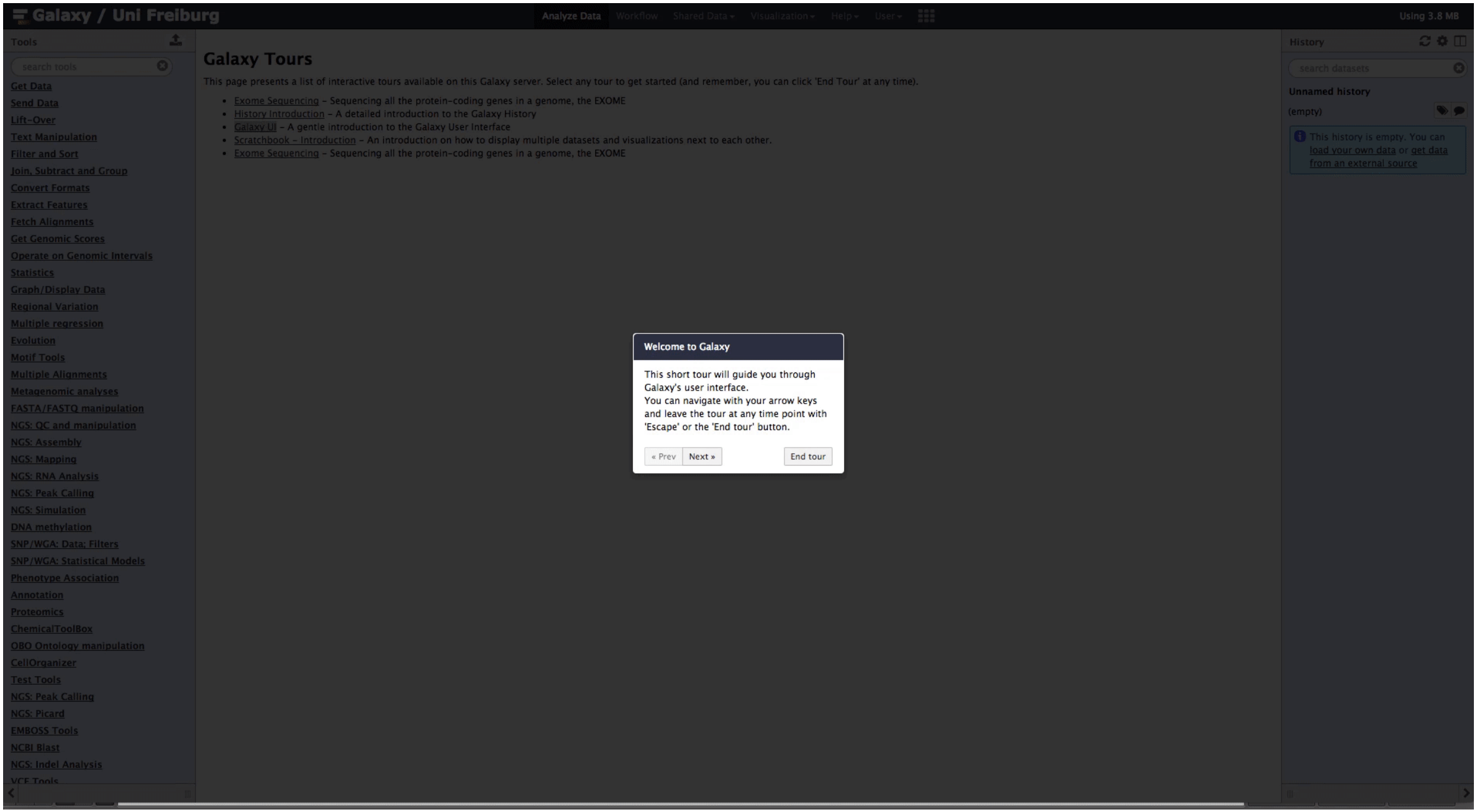


Hands-on tutorials built around a ”research story”

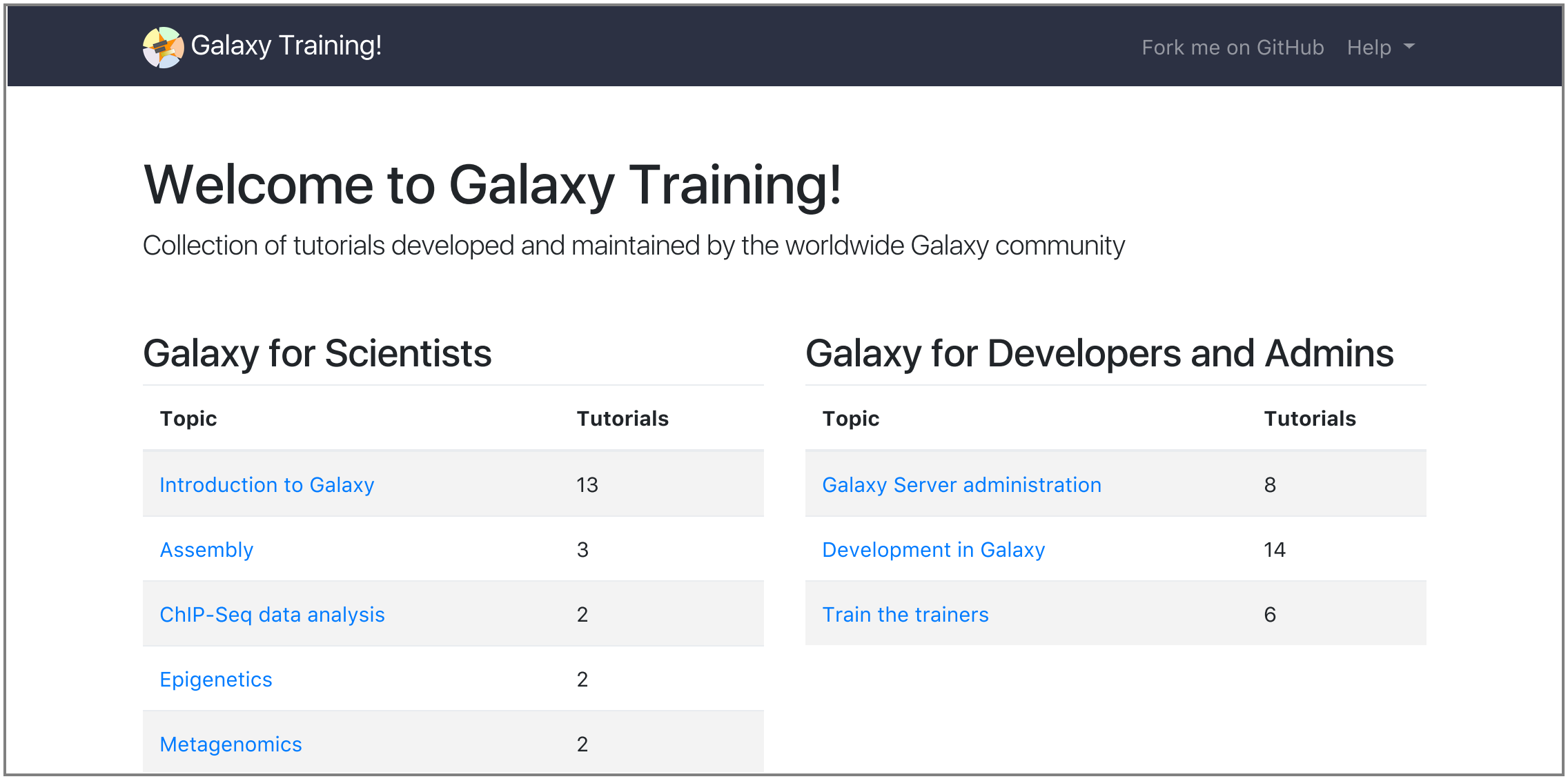


Transcriptomics - Reference-based RNA-Seq data analysis

Hands-on also supported by Interative Tours

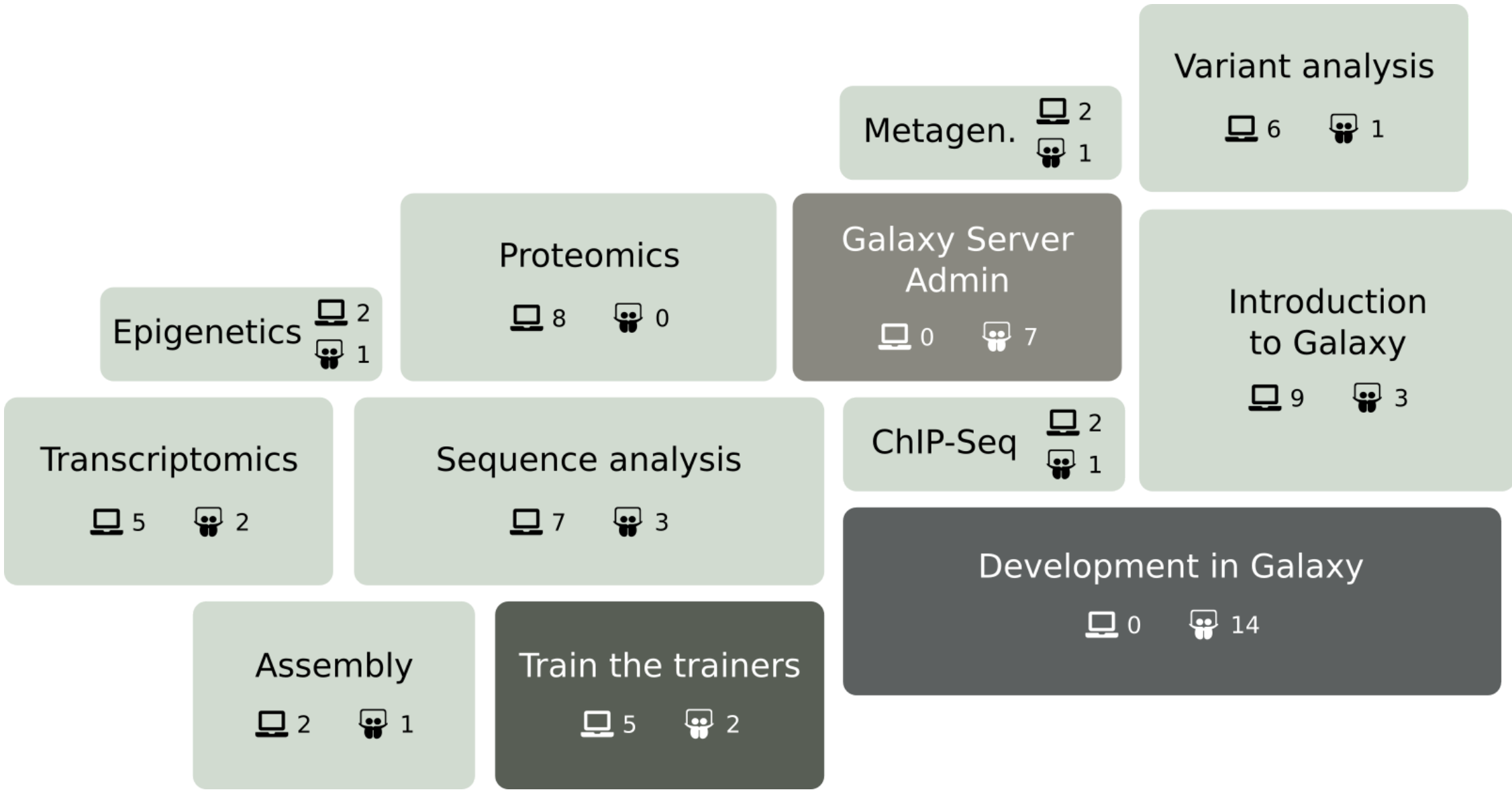


A collection of materials covering many topics



https://training.galaxyproject.org

A collection of materials covering many topics



More than 80 tutorials!

Used both by individual users

[redacted] 11. March 2018 at 19:45

question about tutorial "Analyses of metagenomics data - The global picture"

To: berenice.batut@gmail.com

Hi

I am doing the tutorial "Analyses of metagenomics data - The global picture" and I have a question.

How I execute the following instruction in galaxy:

Hands-on: Krona

1. **Visualize with Krona** with the following parameters
   - "Input file" to taxonomy output from `Classify.otu` (collection)
   - Set **Is this output from mothur?** to `Yes`

I can´t found it in menu.

Thanks.

--
[redacted]

[redacted] 29. November 2017 at 12:55

Query regarding sequence analysis tutorial on Galaxy

To: berenice.batut@gmail.com

Dear sir/mam,
I am following the RNA sequence analysis tutorial on galaxy (https://galaxyproject.github.io/training-material/topics/sequence-analysis/). I am unable to find **Augustus** in the Gene predication tutorial.
Please help

-- Thanks and Regards
[redacted]
Near Sola Bridge,Thaltej
Ahmedabad-380054
Gujrat,INDIA

[redacted] Training 7. June 2017 at 06:54 NM

Having trouble with the 16S Microbial Analysis with Mothur tutorial

To: berenice.batut@gmail.com

Dear Bérénice,

I recently found this tutorial on the galaxy site https://galaxyproject.github.io/training-material//Metagenomics/tutorials/mothur-miseq-sop and decided to contact you. I went through it till the "Hands-on: Combine forward and reverse reads into contigs." Unfortunately I'm not sure what dataset to specify on the **rfastq - Reverse Fastq Sequence file**. I made the previous dataset pair but i haven't successfully made any progress going through the workflow. I tried selecting the files manually for the forward and reverse using the multiple datasets icon but it didn't yield the required 6 new collections, `trim.contig.fasta and scrap.contig.fasta.` Instead i got a `log file, contigs.matched file, contigs.qual and contigs.fasta file.` In addition, i tried specifying the dataset collection with the paired ataset as input for both **ffastq** and **rfastq** and i got a lot of errors. What could i be doing wrong?

Please advice.

Regards,
[redacted]e.

Make.contigs Aligns paired forward and reverse fastq files to contigs as fasta and quality (Galaxy Version 1.27.0) Options

ffastq - Forward Fastq Sequence file

40: Mock_R2.fastq

rfastq - Reverse Fastq Sequence file

40: Mock_R2.fastq

[redacted] 28. February 2018 at 19:43 SC

galaxy reference based rnaseq analysis

To: berenice.batut@gmail.com

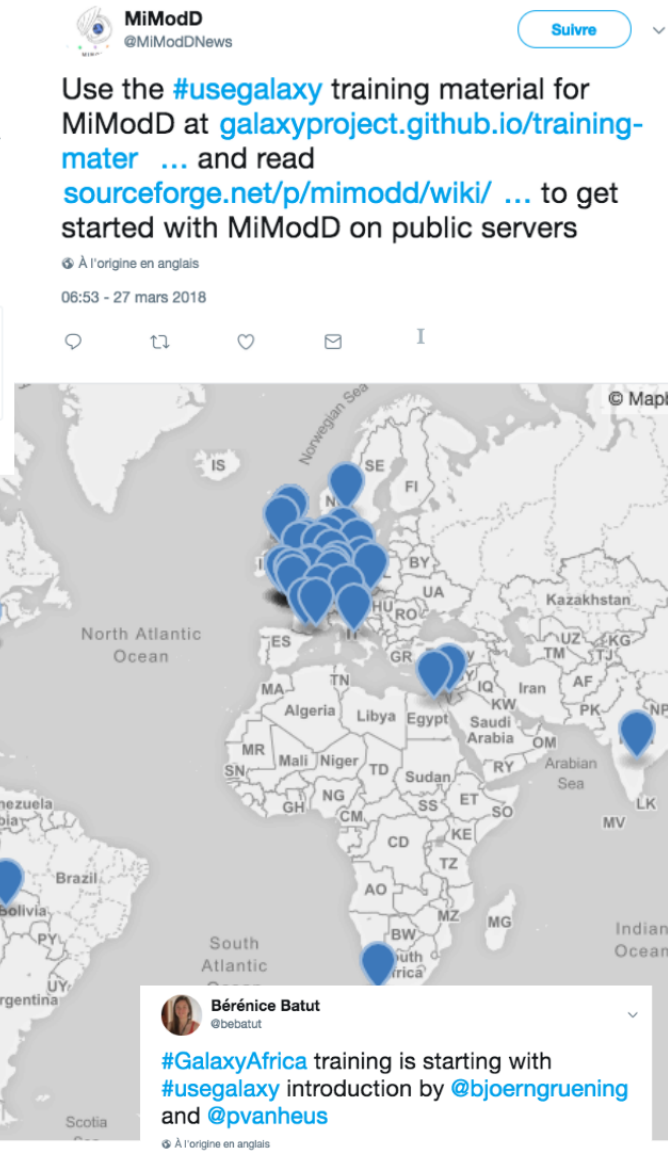
Hi,

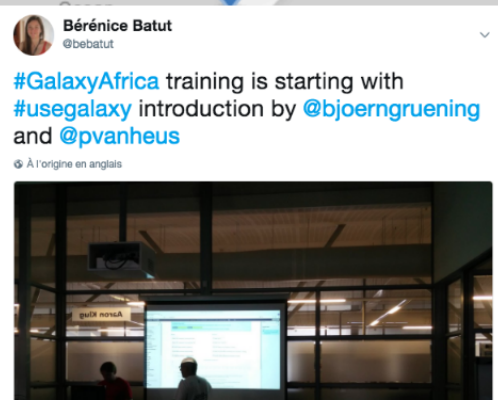
Last week, I analyze my data following your reference based rnaseq analysis using HISAT2 with HTseq. But now it seems like it changed to STAR alignment based analysis. Is there a way that I can find other tutorial. If so please provide me the link.

Thanks
Sri

Used both by individual users & instructors

Used both by individual users & instructors





Requirements for a training infrastructure

- [x] Interactive learning platform
- [x] Support for current research problems
- [x] Usable for effective training for individual users & instructors
- [ ] Community driven (content creation and maintenance)
- [ ] FAIR: Findable, Accessible, Interoperable, Reusable
- [ ] Open

Building an infrastructure to facilitate community-led content development

- Makes tutorial creation a convenient, hassle-free process
- Enables transparent peer-review and curation to guarantee high-quality and current content

Separation between content and format

```
Here treatment is the primary factor which we are interested in. The sequencing
type is some further information that we know about the data that might affect the
analysis. This particular multi-factor analysis allows us to assess the effect of
the treatment, while taking the sequencing type into account, too.

> ### {% icon comment %} Comment
>
> We recommend you to add as many factors as you think may affect gene expression
in your experiment. It can be the sequencing type like here, but it can also be the
manipulation (if different persons are involved in the library preparation), ...
{: .comment}

> ### {% icon hands_on %} Hands-on: Analysis of the differential gene expression
(1)
>
> 1. Create a new history
> 2. Import the seven count files from [Zenodo]
(https://dx.doi.org/10.5281/zenodo.290221)
>    - `GSM461176_untreat_single.deseq.counts`
>    - `GSM461177_untreat_paired.deseq.counts`
>    - `GSM461178_untreat_paired.deseq.counts`
>    - `GSM461179_treat_single.deseq.counts`
>    - `GSM461180_treat_paired.deseq.counts`
>    - `GSM461181_treat_paired.deseq.counts`
>    - `GSM461182_untreat_single.deseq.counts`
>
> 3. **DESeq2** {% icon tool %}: Run **DESeq2** with:
>    - "Treatment" as first factor with "treated" and "untreated" as levels and
selection of count files corresponding to both levels
>
>       > ### {% icon tip %} Tip
>       >
>       > You can select several files by keeping the CTRL (or COMMAND) key pressed
and clicking on the interesting files
>       {: .tip}
>
```

Markdown

Here treatment is the primary factor which we are interested in. The sequencing type is some further information that we know about the data that might affect the analysis. This particular multi-factor analysis allows us to assess the effect of the treatment, while taking the sequencing type into account, too.

Comment

We recommend you to add as many factors as you think may affect gene expression in your experiment. It can be the sequencing type like here, but it can also be the manipulation (if different persons are involved in the library preparation), ...

Hands-on: Analysis of the differential gene expression (1)

1. Create a new history
2. Import the seven count files from Zenodo
   - `GSM461176_untreat_single.deseq.counts`
   - `GSM461177_untreat_paired.deseq.counts`
   - `GSM461178_untreat_paired.deseq.counts`
   - `GSM461179_treat_single.deseq.counts`
   - `GSM461180_treat_paired.deseq.counts`
   - `GSM461181_treat_paired.deseq.counts`
   - `GSM461182_untreat_single.deseq.counts`
3. **DESeq2**: Run **DESeq2** with:
   - "Treatment" as first factor with "treated" and "untreated" as levels and selection of count files corresponding to both levels

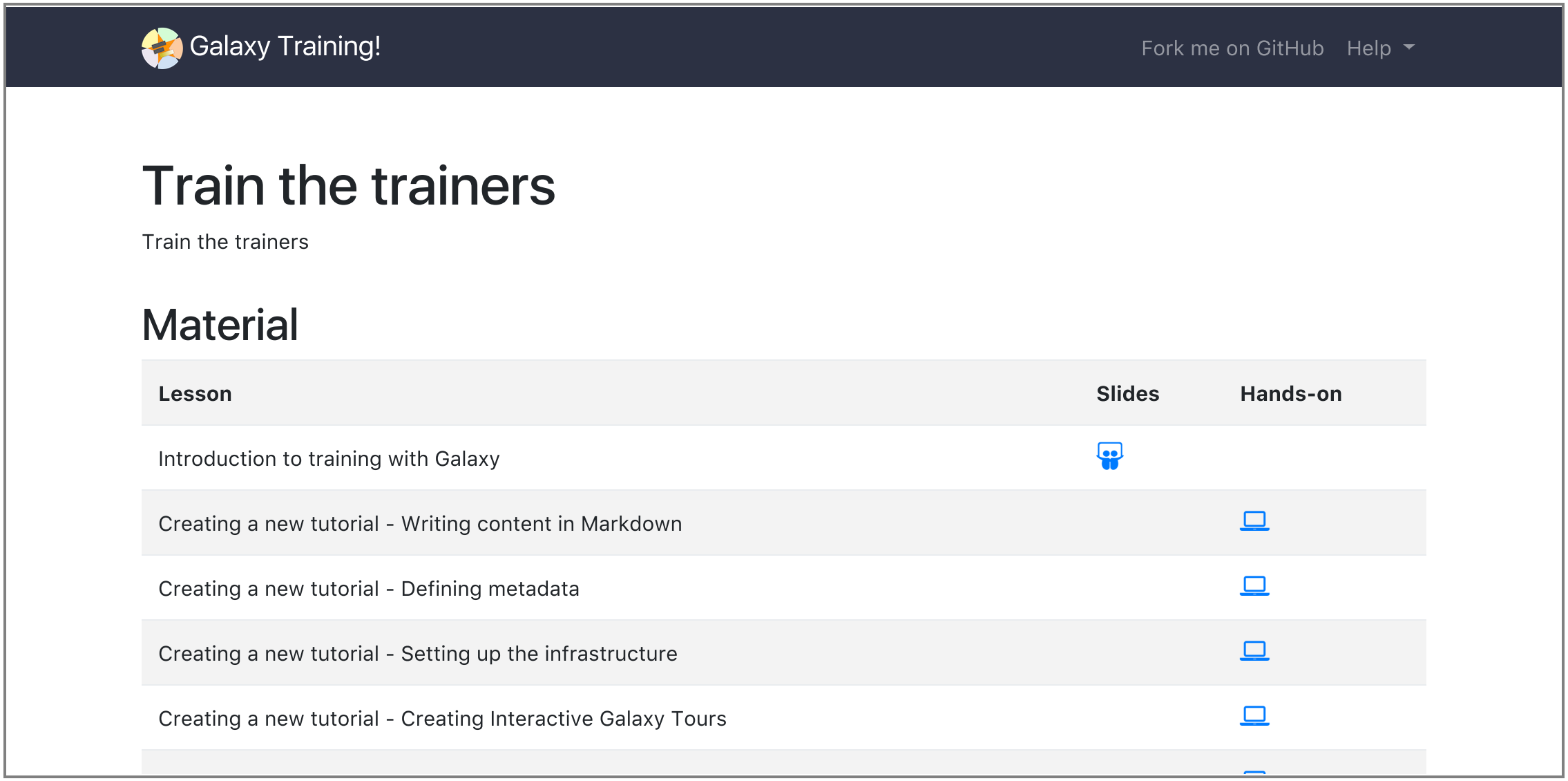
   Tip

   You can select several files by keeping the CTRL (or COMMAND) key pressed and clicking on the interesting files

User-friendly HTML

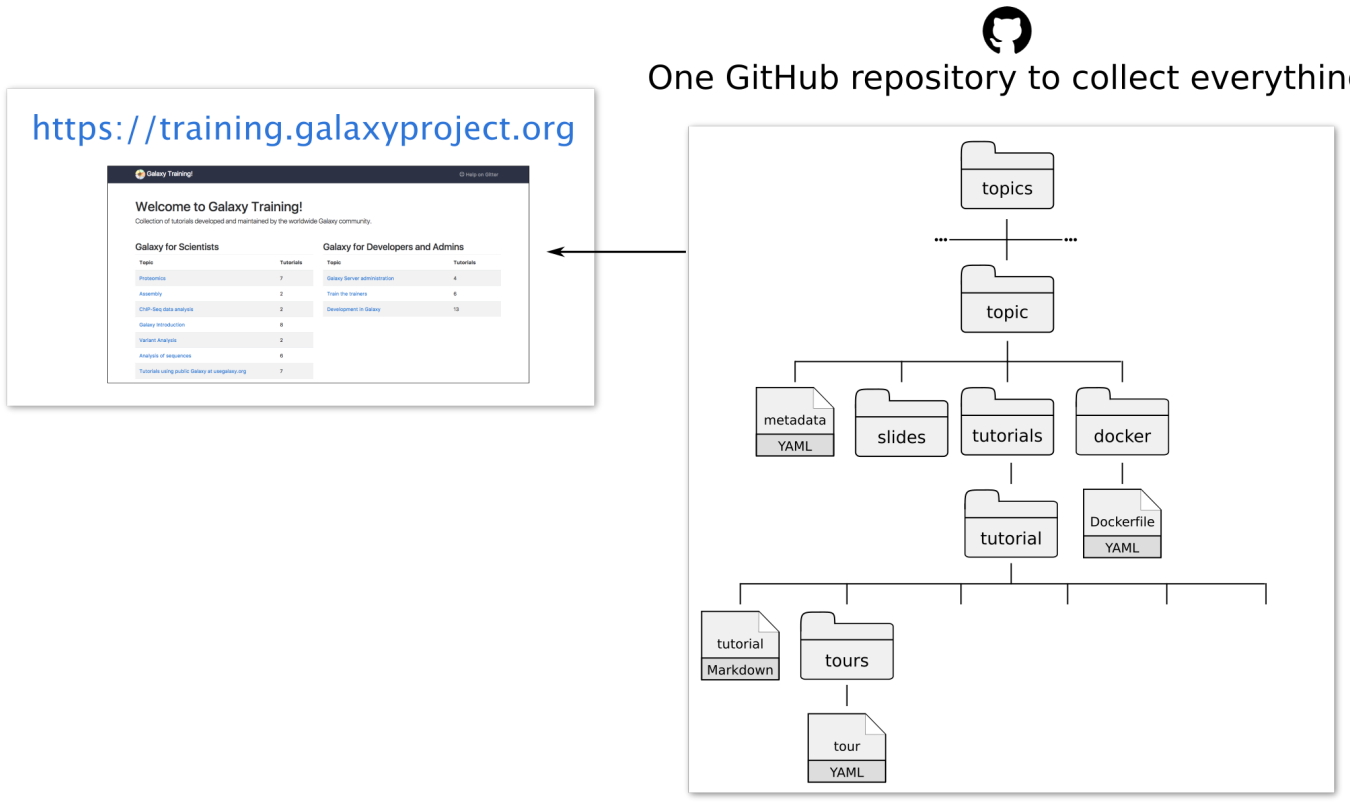
https://training.galaxyproject.org

Creating a tutorial



http://galaxyproject.github.io/training-material/topics/training/

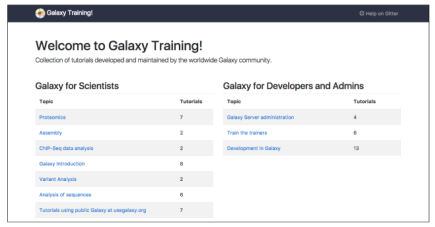
One GitHub repository to collect everything

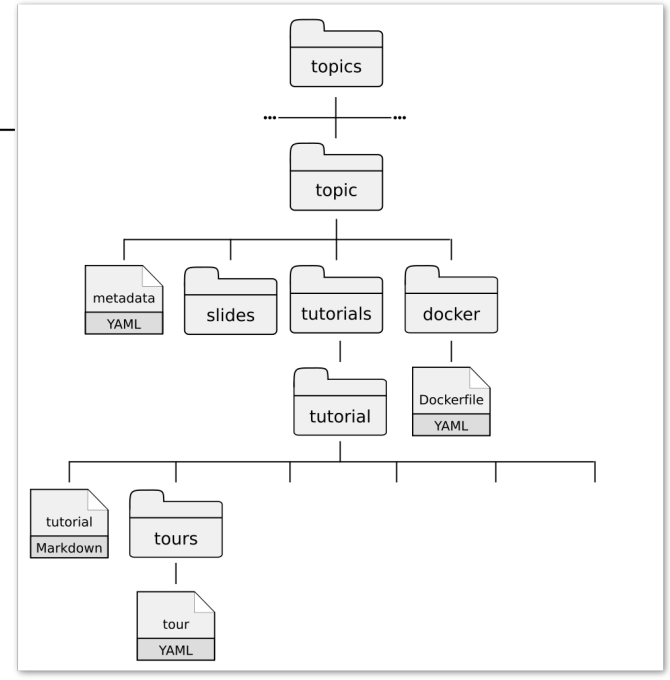


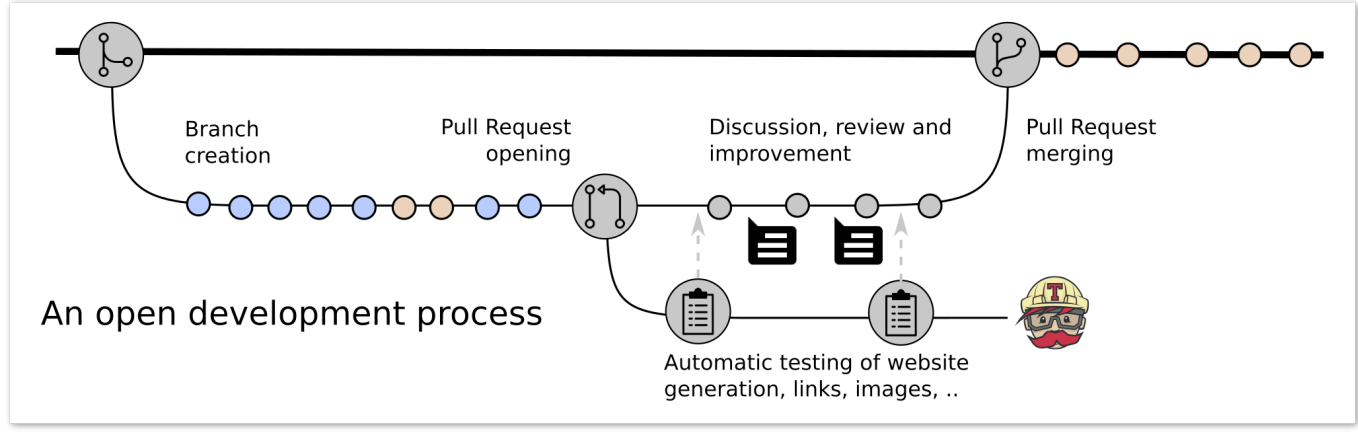
An open and accessible development process

One GitHub repository to collect everything

https://training.galaxyproject.org



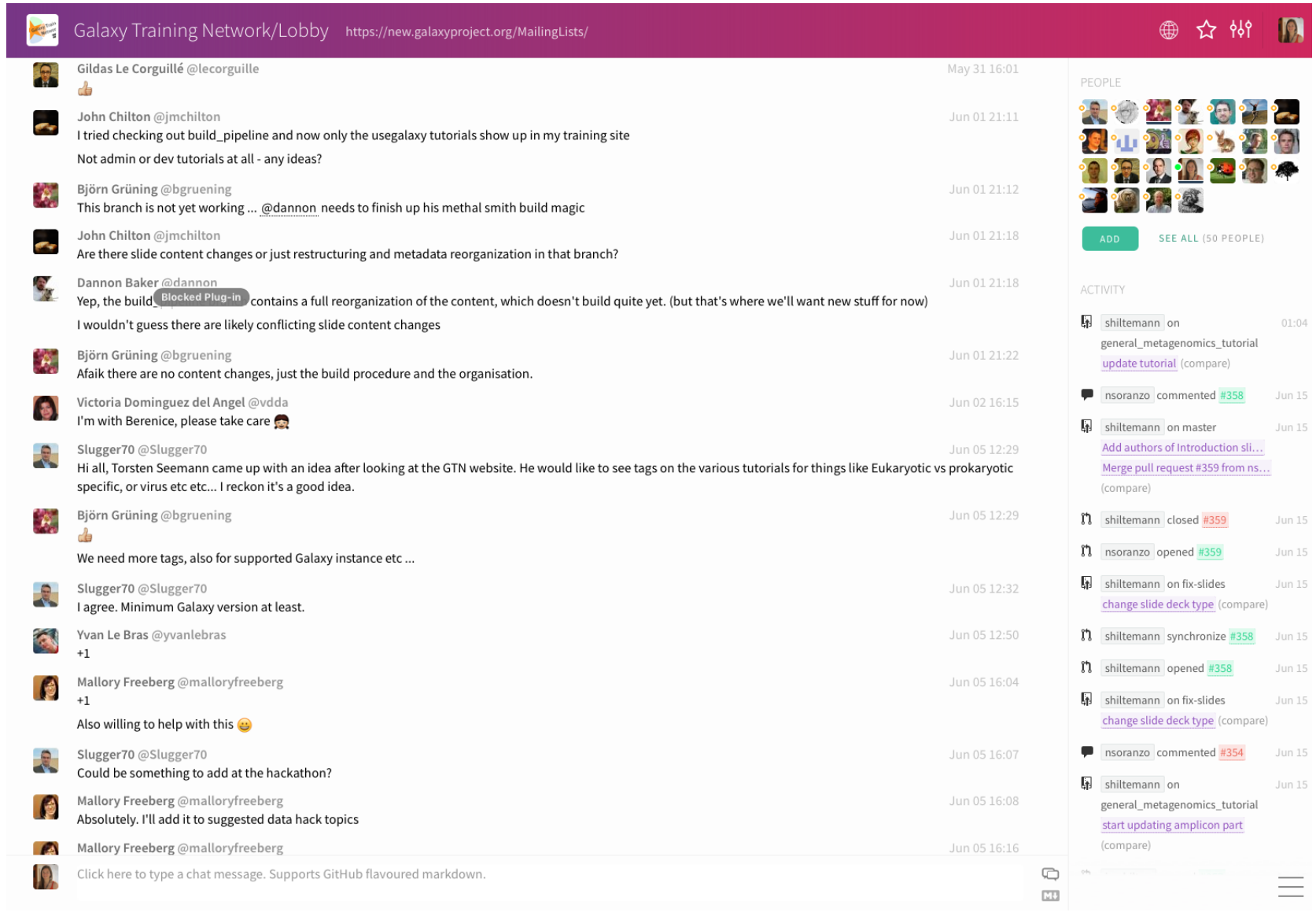




Community-driven

Community-driven

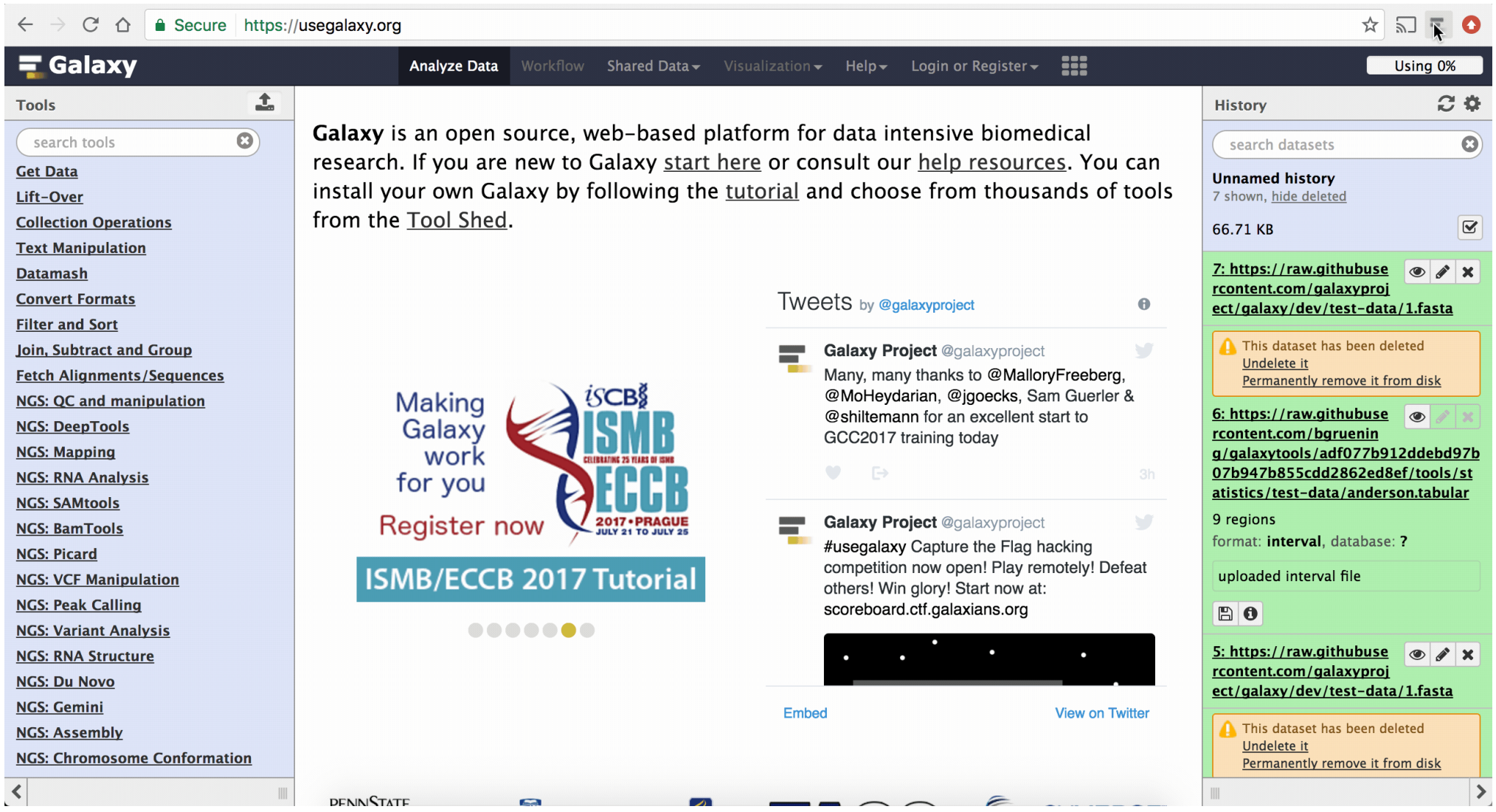
A constant support



Gitter: Galaxy-Training-Network/Lobby

Galaxy Tour Builder

A web extension to develop interactive tours

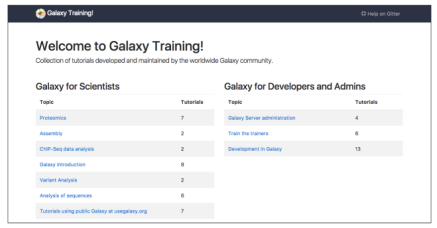


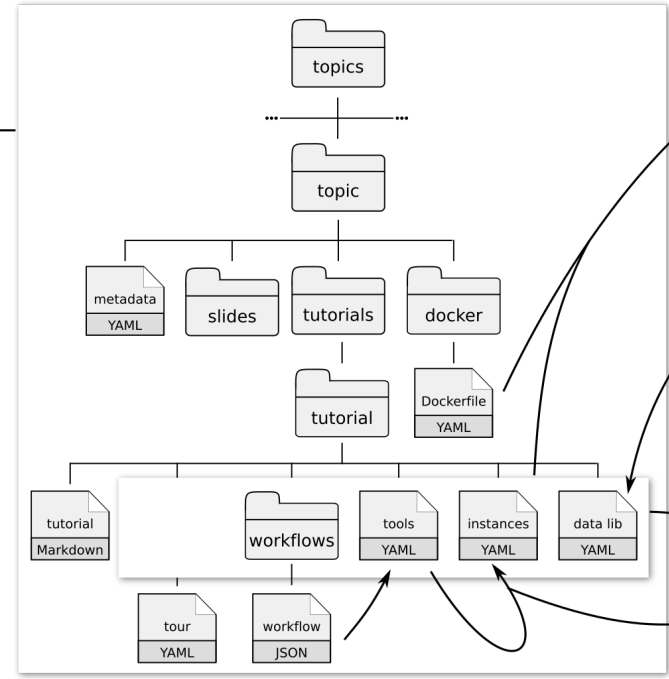
https://github.com/TailorDev/galaxy-tourbuilder

Ensuring accessibility of tutorials

One GitHub repository to collect everything

https://training.galaxyproject.org





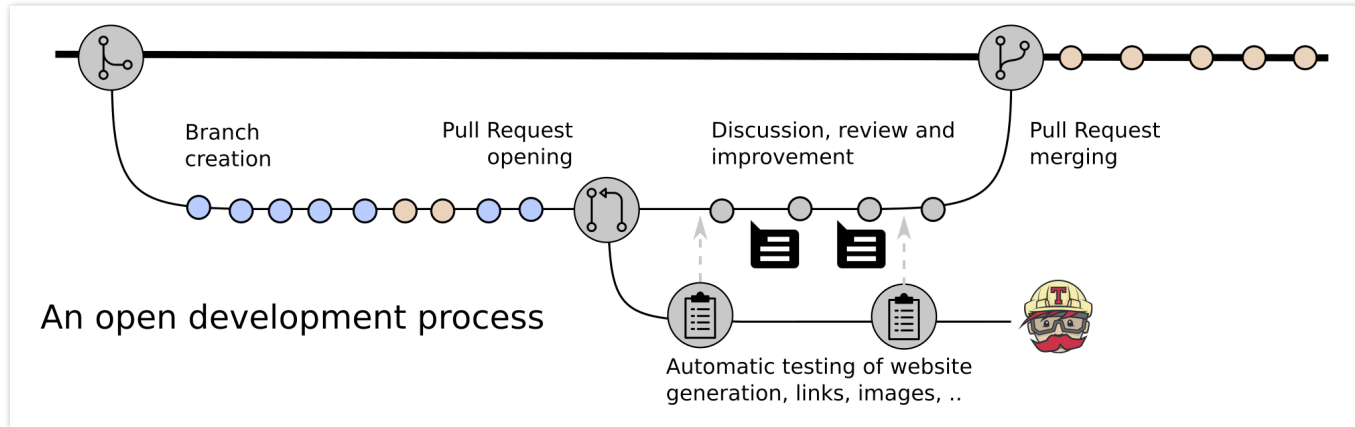
Self-training boxes

Findable, accessible, persistent and citable data

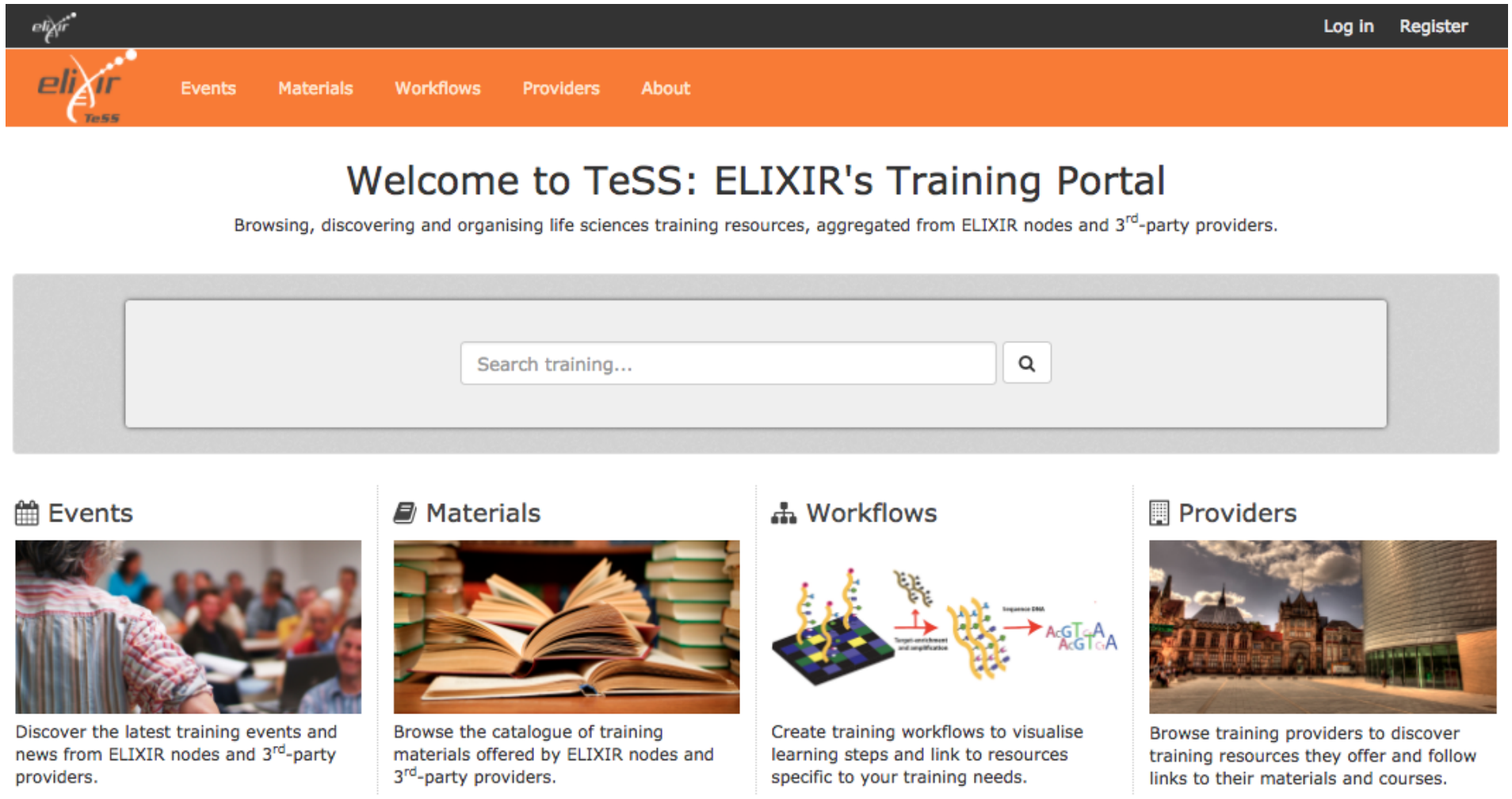
Input file

Public Galaxy Servers

Annotation & configuration



TeSS: the ELIXIR's training portal

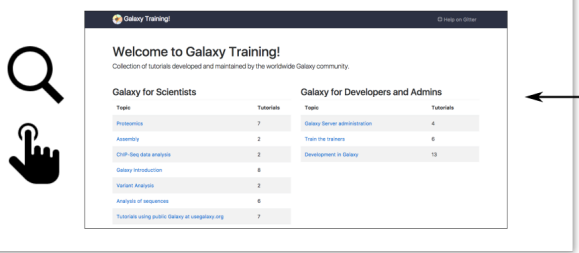


https://tess.elixir-europe.org/

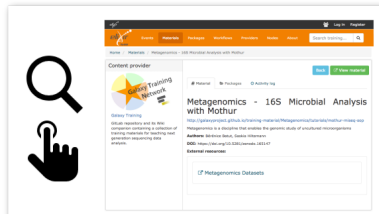
Findable, Accessible, Interoperable, Reusable

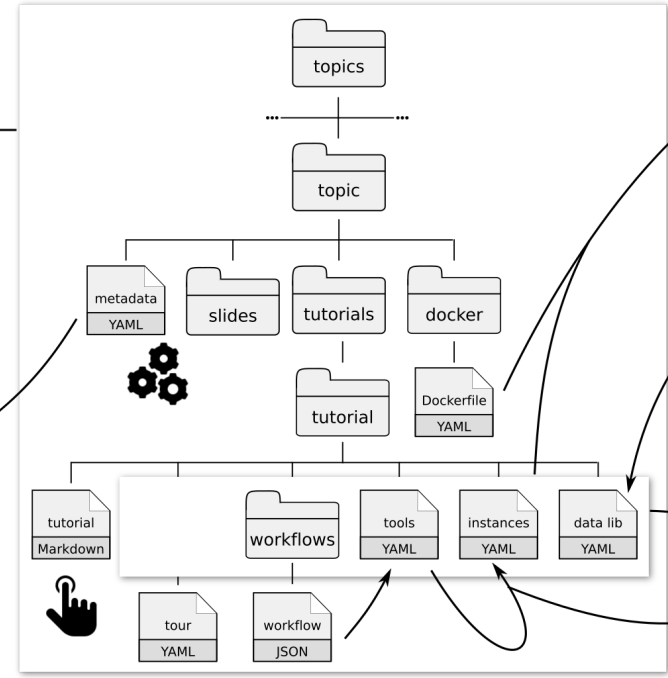
One GitHub repository to collect everything

https://training.galaxyproject.org



Population of external bioinformatic training resources

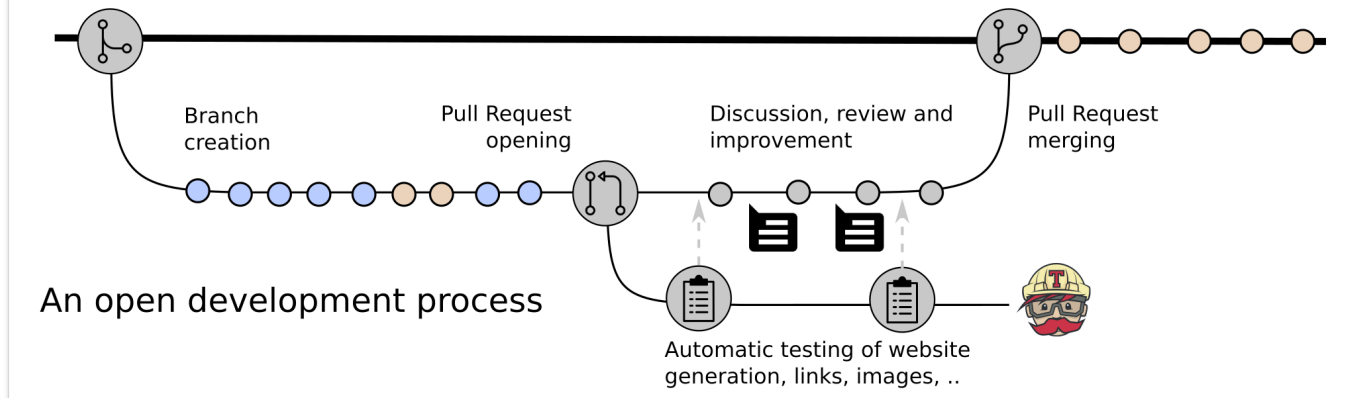




Self-training boxes

Findable, accessible, persistent and citable data

Annotation & configuration



Findable

Accessible

Interoperable

Reusable

Requirements for a training infrastructure

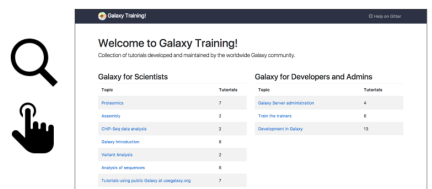
- ☑ Interactive learning platform
- ☑ Support for current research problems
- ☑ Usable for effective training for individual users & instructors
- ☑ Community driven (content creation and maintenance)
- ☑ FAIR: Findable, Accessible, Interoperable, Reusable
- ☑ Open

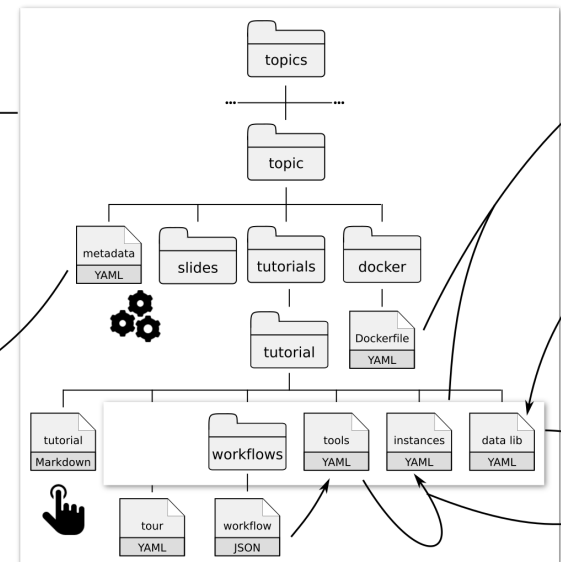
Thank you!

Sponsors


One GitHub repository to collect everything

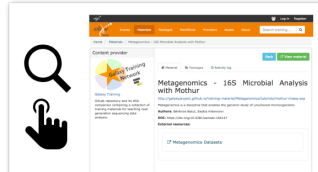
https://training.galaxyproject.org





Self-training boxes

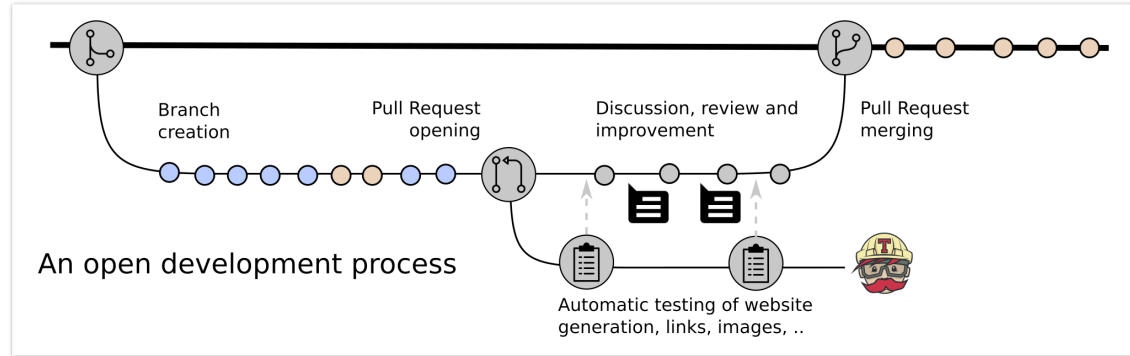
Findable, accessible, persistent and citable data



Population of external bioinformatic training resources

Public Galaxy Servers

Annotation & configuration



Findable
Accessible
Interoperable
Reusable

Community-driven data analysis training for biology

training.galaxyproject.org

github.com/galaxyproject/training-material