Building an open, collaborative, online infrastructure for bioinformatics training

Bérénice Batut

33rd TBI Winterseminar
Februar 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
Building a new open, collaborative and FAIR model for bioinformatics training
Requirements

- Easy to use
- Support for effective training for
  - Individual users
  - Instructors
- Definition of technological infrastructure
- Limited redundancy
An open, collaborative, FAIR, online infrastructure
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)
>   - `GSN461176_untreat_single.deseq.counts`
>   - `GSN461177_untreat_paired.deseq.counts`
>   - `GSN461178_untreat_paired.deseq.counts`
>   - `GSN461179_treat_paired.deseq.counts`
>   - `GSN461180_treat_paired.deseq.counts`
>   - `GSN461181_treat_paired.deseq.counts`
>   - `GSN461182_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}

<table>
<thead>
<tr>
<th>Hands-on: Analysis of the differential gene expression (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Create a new history</strong></td>
</tr>
<tr>
<td><strong>2. Import the seven count files from Zenodo</strong></td>
</tr>
<tr>
<td>- <code>GSN461176_untreat_single.deseq.counts</code></td>
</tr>
<tr>
<td>- <code>GSN461177_untreat_paired.deseq.counts</code></td>
</tr>
<tr>
<td>- <code>GSN461178_untreat_paired.deseq.counts</code></td>
</tr>
<tr>
<td>- <code>GSN461179_treat_paired.deseq.counts</code></td>
</tr>
<tr>
<td>- <code>GSN461180_treat_paired.deseq.counts</code></td>
</tr>
<tr>
<td>- <code>GSN461181_treat_paired.deseq.counts</code></td>
</tr>
<tr>
<td>- <code>GSN461182_untreat_single.deseq.counts</code></td>
</tr>
<tr>
<td><strong>3. DESeq2</strong>: Run DESeq2 with:</td>
</tr>
<tr>
<td>- &quot;Treatment&quot; as first factor with &quot;treated&quot; and &quot;untreated&quot; as levels and selection of count files corresponding to both levels</td>
</tr>
</tbody>
</table>

**Tip**

You can select several files by keeping the CTRL (or COMMAND) key pressed and clicking on the interesting files.
An open, collaborative, FAIR, online infrastructure
Findable material via TeSS

https://tess.elixir-europe.org/
An open, collaborative, FAIR, online infrastructure
Storage of training datasets on Zenodo

https://zenodo.org/
An open, collaborative, FAIR, online infrastructure

Separation between content and form

Population of external bioinformatic training resources

One GitHub repository to collect everything

Findable, accessible, persistent and citable data

Findable
Accessible
Interoperable
Reusable
An open, collaborative, FAIR, online infrastructure
An open, collaborative, FAIR, online infrastructure
Galaxy Training materials

Welcome to Galaxy Training!
Collection of tutorials developed and maintained by the worldwide Galaxy community.

<table>
<thead>
<tr>
<th>Galaxy for Scientists</th>
<th>Tutorials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to Galaxy</td>
<td>13</td>
</tr>
<tr>
<td>Assembly</td>
<td>3</td>
</tr>
<tr>
<td>ChIP-Seq data analysis</td>
<td>2</td>
</tr>
<tr>
<td>Epigenetics</td>
<td>1</td>
</tr>
<tr>
<td>Metagenomics</td>
<td>2</td>
</tr>
<tr>
<td>Proteomics</td>
<td>8</td>
</tr>
<tr>
<td>Sequence analysis</td>
<td>6</td>
</tr>
<tr>
<td>Transcriptomics</td>
<td>5</td>
</tr>
<tr>
<td>Variant Analysis</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Galaxy for Developers and Admins</th>
<th>Tutorials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galaxy Server administration</td>
<td>8</td>
</tr>
<tr>
<td>Development in Galaxy</td>
<td>13</td>
</tr>
<tr>
<td>Train the trainers</td>
<td>6</td>
</tr>
</tbody>
</table>

http://training.galaxyproject.org
An open, collaborative, FAIR, online infrastructure

- training.galaxyproject.org
- github.com/galaxyproject/training-material