

# Resources for Making Tools Available in the Galaxy Ecosystem

Daniel Blankenberg  
blanked2@ccf.org | @DBlankenberg  
Genomic Medicine Institute of LRI @ Cleveland Clinic

May 12, 2021

# Outline

Standard Galaxy tools

Galaxy Interactive Tools (GxIT)

Data Source Tools

External Display Applications (GEDA)

# Outline

## **Standard Galaxy tools**

Galaxy Interactive Tools (GxIT)

Data Source Tools

External Display Applications (GEDA)

# Where to find this information later?



The screenshot shows the Galaxy Community Hub website. At the top, there is a navigation bar with the Galaxy logo, links for 'Use', 'Learn', 'Community', 'Deploy & Develop', 'Support', and '@jxtx', a search bar, and an 'Edit' button. The main heading is 'Galaxy Community Hub'. Below this, there are two main announcements:

- GCC2021 Fellowship applications due May 17**: Accompanied by a logo for 'GCC 2021 VIRTUAL EDITION' featuring a stylized building.
- Talk abstracts due May 17**: Accompanied by a logo for 'Galaxy Fellowships' featuring the Galaxy logo and a yellow pencil.

Below these announcements, there is a paragraph: 'Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational research.' followed by a bulleted list of features:

- **Accessible:** programming experience is not required to easily upload data, run complex tools and workflows, and visualize results.
- **Reproducible:** Galaxy captures information so that you don't have to; any user can repeat and understand a complete computational analysis, from tool parameters to the dependency tree.
- **Transparent:** Users share and publish their histories, workflows, and visualisations via the web.
- **Community centered:** Our inclusive and diverse users (developers, educators, researchers, clinicians, etc.) are empowered to share their findings.

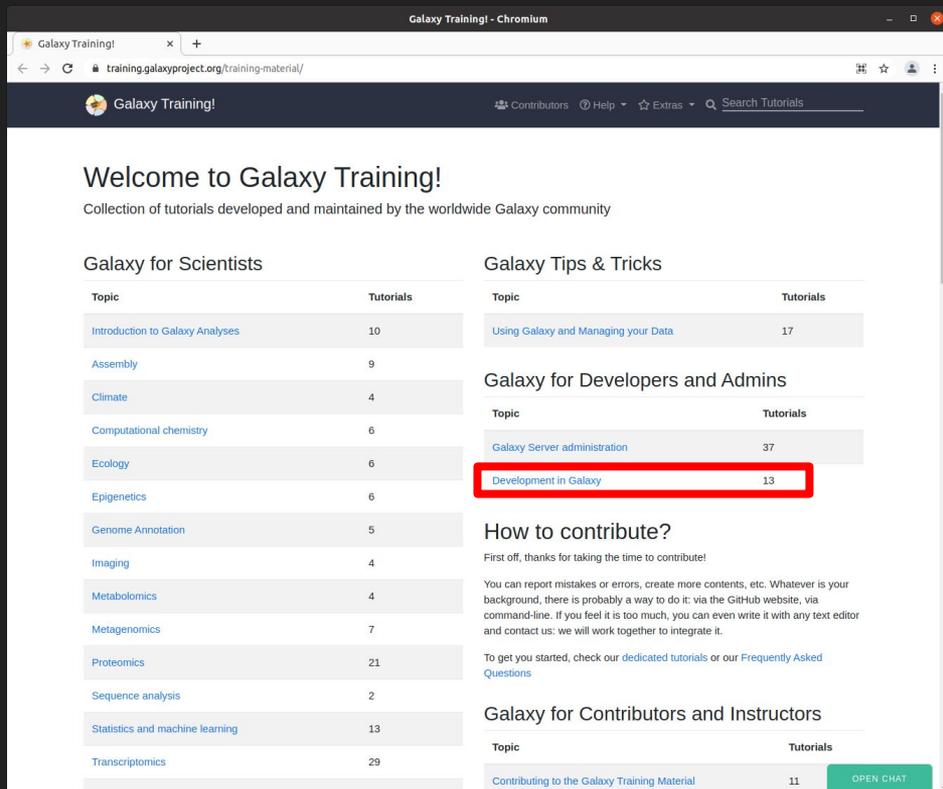
Below the list, it says: 'Welcome to the Galaxy Community Hub, where you'll find community curated documentation of all things Galaxy.'

At the bottom, there are three columns of content:

- News:** 'COVID updates: B.1.617.2 detected in COG-UK data from the beginning of April - We checked the COG-UK data to see if we'd find representative samples of B.1.617.2 (one of the Indian lineages, and a Variant of Concern).', 'Exploring and monitoring COVID-19 variants with Galaxy - Galaxy Europe has developed a pipeline to automatically create COVID-19 variant information and statistics from raw sequencing data generated by COVID-19 Genomics UK Consortium (COG-UK).', 'GCC2021 Conference Fellowships in honor of James Taylor - Apply by May 17'.
- Events:** 'Galaxy @ 2021 Biology of Genomes - May 17.', 'Galaxy Resources for Tool Developers - May 12. Part of the Galaxy Resources Webinar Series', 'Customizable welcome and introduction page for new Galaxy users - May 13. Giving new users some obvious guidance when they first arrive.', 'Galaxy Training Network CoFest - May 20. Connect with the Galaxy Training Community!', 'Galaxy PaperCuts CoFest Day: May 20 - May 20 A community contribution day'.
- @galaxyproject:** A tweet from Björn Grüning (@bjoerngruning) replying to @bjoerngruning, mentioning insights from MS data and a manuscript on MS-based clinical diagnostics of Coronavirus Disease 2019 (COVID-19).

<https://galaxyproject.org>

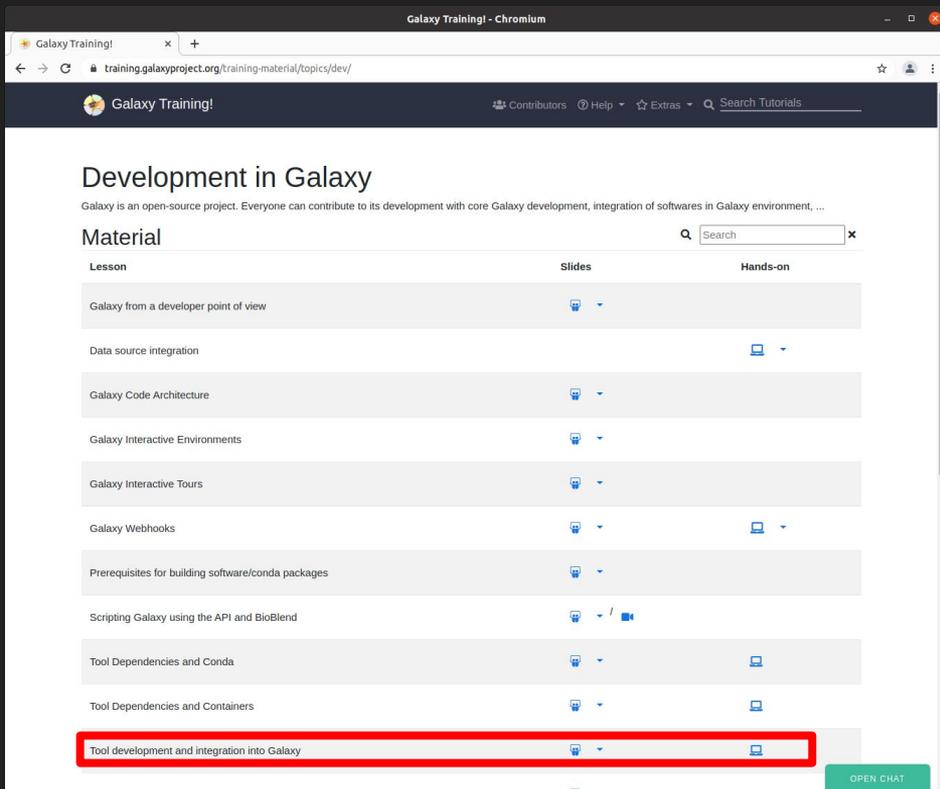
# Where to find this information later?



The screenshot shows the homepage of Galaxy Training! with a navigation bar and a main content area. The main content area is divided into several sections: 'Galaxy for Scientists', 'Galaxy Tips & Tricks', 'Galaxy for Developers and Admins', 'How to contribute?', and 'Galaxy for Contributors and Instructors'. The 'Galaxy for Developers and Admins' section contains a table with two columns: 'Topic' and 'Tutorials'. The row for 'Development in Galaxy' is highlighted with a red box.

Topic	Tutorials
Introduction to Galaxy Analyses	10
Assembly	9
Climate	4
Computational chemistry	6
Ecology	6
Epigenetics	6
Genome Annotation	5
Imaging	4
Metabolomics	4
Metagenomics	7
Proteomics	21
Sequence analysis	2
Statistics and machine learning	13
Transcriptomics	29

<https://training.galaxyproject.org/>

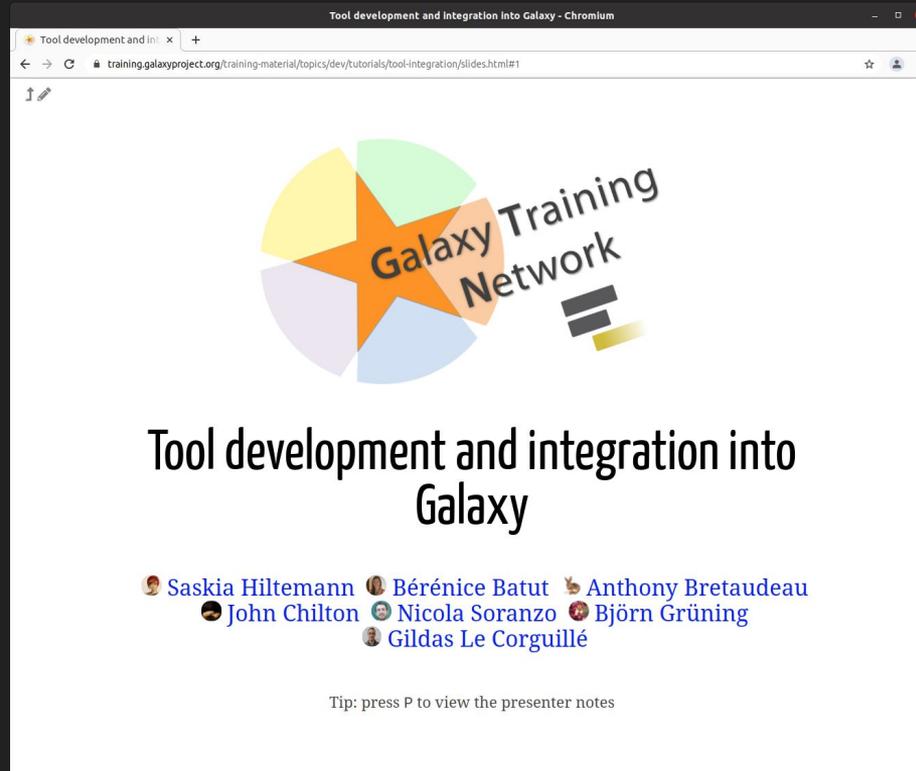


The screenshot shows the 'Development in Galaxy' page, which is a sub-page of the Galaxy Training! website. The page has a search bar and a table with three columns: 'Lesson', 'Slides', and 'Hands-on'. The row for 'Tool development and integration into Galaxy' is highlighted with a red box.

Lesson	Slides	Hands-on
Galaxy from a developer point of view		
Data source integration		
Galaxy Code Architecture		
Galaxy Interactive Environments		
Galaxy Interactive Tours		
Galaxy Webhooks		
Prerequisites for building software/conda packages		
Scripting Galaxy using the API and BioBlend		
Tool Dependencies and Conda		
Tool Dependencies and Containers		
Tool development and integration into Galaxy		

<https://training.galaxyproject.org/training-material/topics/dev/>

# Standard Galaxy Tool Development



Tool development and integration into Galaxy - Chromium

training.galaxyproject.org/training-material/topics/dev/tutorials/tool-integration/slides.html#1



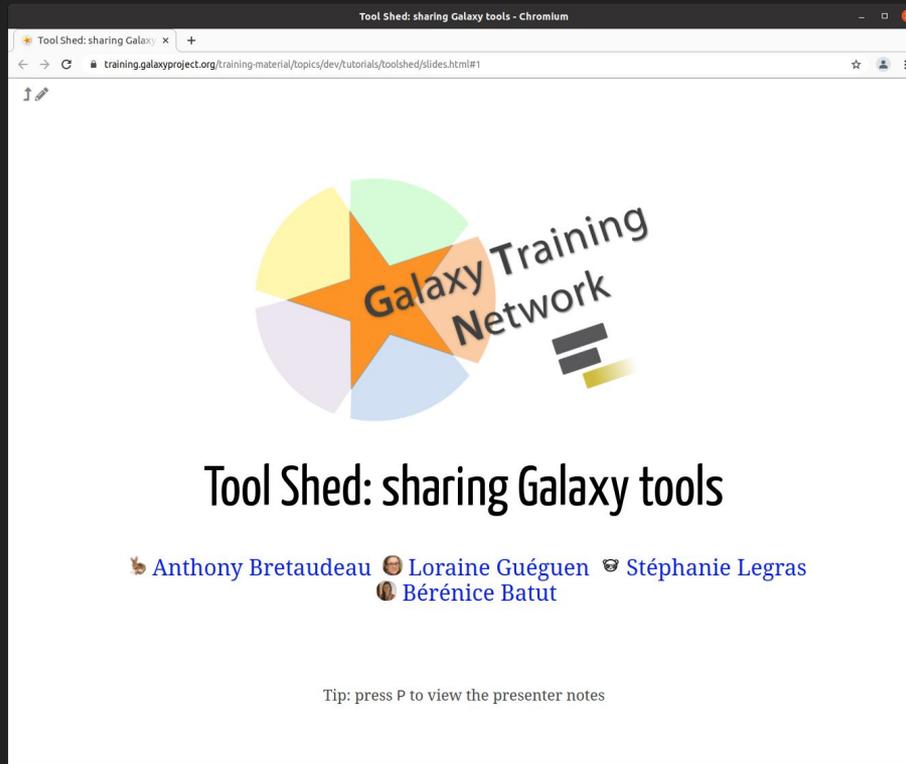
## Tool development and integration into Galaxy

 Saskia Hiltemann  Bérénice Batut  Anthony Bretaudeau  
 John Chilton  Nicola Soranzo  Björn Grüning  
 Gildas Le Corguillé

Tip: press P to view the presenter notes

<https://training.galaxyproject.org/training-material/topics/dev/tutorials/tool-integration/slides.html#1>

# Publishing to the Tool Shed



The image shows a web browser window displaying a slide from a presentation. The browser's address bar shows the URL: <https://training.galaxyproject.org/training-material/topics/dev/tutorials/toolshed/slides.html#1>. The slide features the Galaxy Training Network logo, which consists of a circular arrangement of five colored segments (yellow, green, orange, purple, blue) with the text 'Galaxy Training Network' overlaid. Below the logo, the title 'Tool Shed: sharing Galaxy tools' is displayed in a large, bold, black font. Underneath the title, the names of the presenters are listed: Anthony Bretaudeau, Loraine Guéguen, Stéphanie Legras, and Bérénice Batut, each accompanied by a small circular profile icon. At the bottom of the slide, a tip is provided: 'Tip: press P to view the presenter notes'.

<https://training.galaxyproject.org/training-material/topics/dev/tutorials/toolshed/slides.html#1>

# Outline

Standard Galaxy tools

**Galaxy Interactive Tools (GxIT)**

Data Source Tools

External Display Applications (GEDA)

# Galaxy Interactive Tools (GxIT)

Software that has its own Graphical User Interface (GUI)

Requires dependency to be in a container

Specify one or more entry points

port

optional URL

Examples

<https://github.com/galaxyproject/galaxy/tree/dev/tools/interactive>

# Galaxy Interactive Tools (GxIT)

```
1 <tool id="interactive_tool_cellxgene" tool_type="interactive" name="Interactiv
2 <requirements>
3   <container type="docker">quay.io/galaxy/cellxgene-galaxy-ie:ie2</conta
4 </requirements>
5 <entry_points>
6   <entry_point name="Cellxgene Single Cell Visualisation on $infile.disp
7     <port>80</port>
8   </entry_point>
9 </entry_points>
10 <command><![CDATA[
11   #import re
12   #set $fancy_name = '/tmp/galaxy_cellxgene_' + re.sub('[^\w\_]', '_',
13   cp '${infile}' '${fancy_name}'
14   &&
15   cellxgene launch --host 0.0.0.0 --port 80 '${fancy_name}'
16 ]]>
17 </command>
18 <inputs>
19   <param name="infile" type="data" form
20 </inputs>
21 <outputs>
22   <data name="out_file1" format="txt" /
23 </outputs>
24 <tests>
25 </tests>
```

Interactive CellXgene Environment (Galaxy Version 0.1) Favorite Options

AnnData Dataset  
80: Manipulate AnnData (rename\_categories) on data 54

Requires layout

On demand differential expression  
Enable

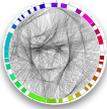
threshold for log fold change  
0.01

Genes are only returned in differential expression if the effect size is above the specified threshold for log fold change. Defaults to 0.01 (--diffexp-lfc-cutoff)

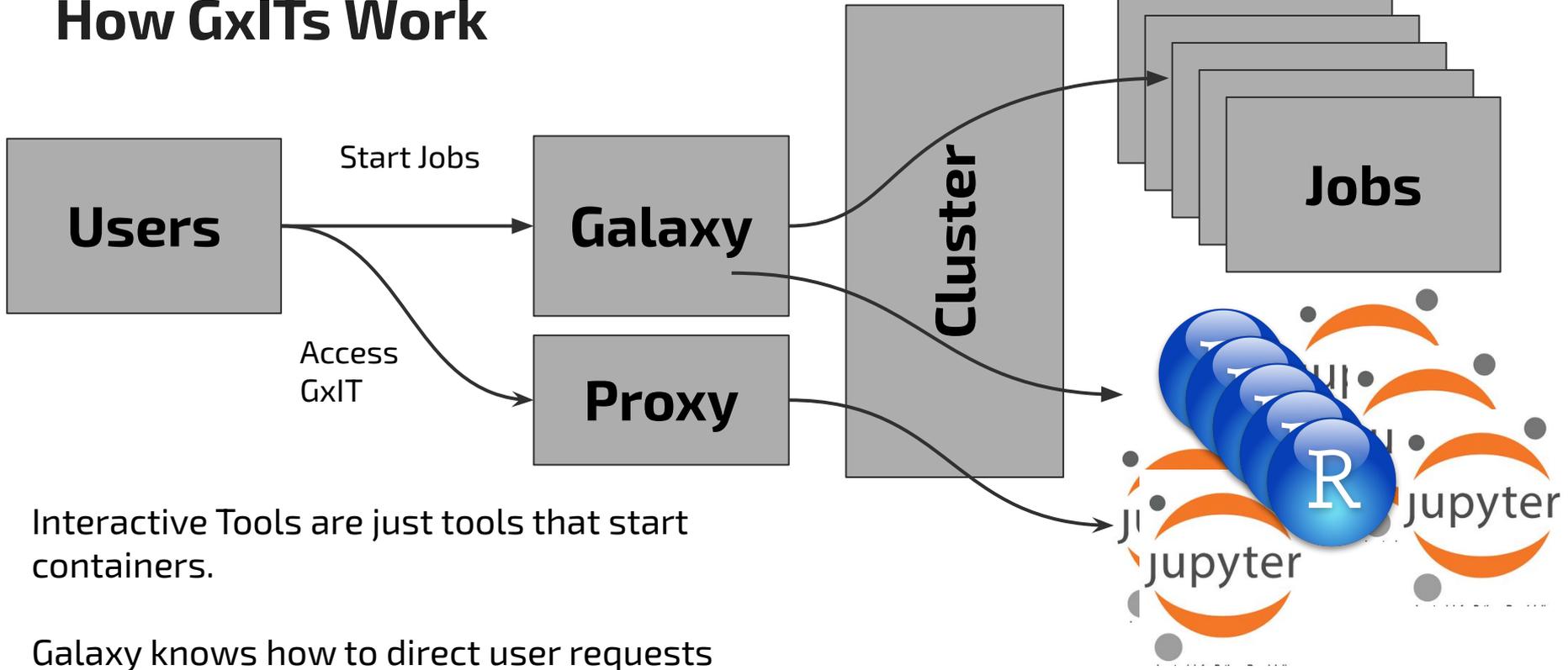
Embeddings  
+ Insert Embeddings

Enable user annotation of data (experimental)  
Enable

Annotations file



# How GxITs Work



Interactive Tools are just tools that start containers.

Galaxy knows how to direct user requests to cluster where the tools are running

# Outline

Standard Galaxy tools

Galaxy Interactive Tools (GxIT)

## **Data Source Tools**

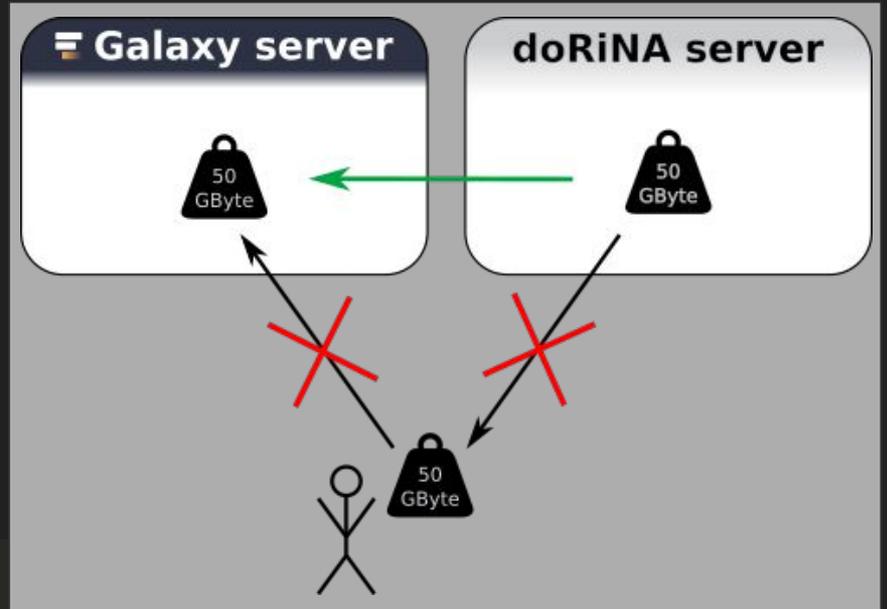
External Display Applications (GEDA)

# Data Source Tools

Leverage existing Data Warehouse query interfaces

Uses a URL callback methodology

<https://galaxyproject.org/admin/internals/data-sources/>



<https://training.galaxyproject.org/training-material/topics/dev/tutorials/data-source-integration/tutorial.html>

Output knownGene as BED - ChromIum

genome.ucsc.edu/cgi-bin/hgTables

Output knownGene as BED

Include custom track header:

name=

description=

visibility=

url=

Create one BED record per:

- Whole Gene
- Upstream by  bases
- Exons plus  bases at each end
- Introns plus  bases at each end
- 5' UTR Exons
- Coding Exons
- 3' UTR Exons
- Downstream by  bases

Note: if a feature is close to the beginning or end of a chromosome, extending past the edge of the chromosome.

```
1 <?xml version="1.0"?>
2
3 If the value of 'URL method' is 'get', the request will consist of the value of 'URL' coming back in
4 the initial response. If value of 'URL method' is 'post', any additional params coming back in the
5 initial response ( in addition to 'URL' ) will be encoded and appended to URL and a post will be performed.
6
7 <tool name="UCSC Main" id="ucsc_table_direct1" tool_type="data_source" version="1.0.0">
8 <description>table browser</description>
9 <cmd_line_operation>
10 <cmd_line_operation operation="0224-c/edam_operation">
11 </cmd_line_operation>
12 </cmd_line_operation>
13 <command>[[CDATA[
14 python '$_tool_directory'/_data_source.py: '$output' '$_app' '_config.output_size_limit'
15 ]]</command>
16 <inputs action="https://genome.ucsc.edu/cgi-bin/hgTables? check_values="false" method="get">
17 <input type="get" id="UCSC Table Browser" $GALAXY_URL"/>
18 <param name="GALAXY_URL" type="baseurl" value="" tool_runner"/>
19 <param name="tool_id" type="hidden" value="ucsc_table_direct1"/>
20 <param name="genome" type="hidden" value="hg19"/>
21 <param name="hgta_compressType" type="hidden" value="none"/>
22 <param name="hgta_outputType" type="hidden" value="bed"/>
23 </inputs>
24 <request_param translation>
25 <request_param galaxy name="URL" method="name" remote name="URL" missing="post"/>
26 <request_param galaxy name="dbkey" remote name="db" missing="" />
27 <request_param galaxy name="organism" remote name="org" missing="unknown species"/>
28 <request_param galaxy name="table" remote name="hgta_table" missing="unknown table"/>
29 <request_param galaxy name="description" remote name="hgta_regionType" missing="no description"/>
30 <request_param galaxy name="data_type" remote name="hgta_outputType" missing="auto">
31 <value translation>
32 <value galaxy value="auto" remote value="primaryTable"/>
33 <value galaxy value="auto" remote value="selectedFields"/>
34 <value galaxy value="hg19" remote value="hg19data"/>
35 <value galaxy value="interval" remote value="tab"/>
36 <value galaxy value="hgta" remote value="hyperlinks"/>
37 <value galaxy value="fasta" remote value="sequence"/>
38 <value galaxy value="gff" remote value="gff"/>
39 </value translation>
40 </request_param>
41 <request_param translation>
42 <shits minwidth="800"/>
43 </request_param>
44 <output>
45 <data name="output" format="tabular" label="${tool.name} on ${organism}: ${table} (#if $description == 'range' then $getVar('position', 'unknown position') ) else $description"/>
46 </output>
47 <options sanitize="False" refresh="True"/>
48 </citation>
49 <citation type="doi">10.1093/database/baf011/citation>
50 <citation type="doi">10.1101/gr.229102/citation>
51 </citation>
52 </tool>
```

# Outline

Standard Galaxy tools

Galaxy Interactive Tools (GxIT)

Data Source Tools

**External Display Applications (GEDA)**

# External Display Applications (GEDA)

**A**

**History**

search datasets

**ERR091571 - PE WGS HiSeq**

1 shown, 2 hidden

132.39 GB

**3: Map with BWA-MEM on data 2 and data 1 (mapped reads in BAM format)**

30.6 GB

format: **bam**, database: **hg38**

display at UCSC main  
display with IGV local Human hg38  
display in IGB View  
display at bam.iobio bam.iobio.io

Binary bam alignments file

## Tutorial

<https://galaxyproject.org/admin/tools/external-display-applications-tutorial/>

## More examples:

[https://static-content.springer.com/esm/art%3A10.1038%2Fs41592-019-0727-x/MediaObjects/41592\\_2019\\_727\\_MOESM1\\_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1038%2Fs41592-019-0727-x/MediaObjects/41592_2019_727_MOESM1_ESM.pdf)

**B**

Filename: \$GALAXY\_ROOT/display\_applications/iobio/bam.xml

```

<display id="iobio_bam" version="1.0.0" name="display at bam.iobio">
  <link id="bam_iobio" name="bam.iobio.io">
    <url>http://bam.iobio.io/?bam=${bam_file.qp}</url>
    <param type="data" name="bam_file" url="galaxy.bam" />
    <param type="data" name="bai_file" url="galaxy.bam.bai" metadata="bam_index" />
  </link>
</display>

```

**C**

Filename: \$GALAXY\_ROOT/config/datatypes\_conf.xml

```

<datatype extension="bam" type="galaxy.datatypes.binary:Bam">
  <display file="iobio/bam.xml" />
</datatype>

```

Blankenberg D\*, Coraor N, Chilton J. Galaxy External Display Applications: Closing a dataflow interoperability loop. *Nature Methods*. Feb. 2020. <https://rdcu.be/b0xnC>

# Join the Galaxy Community!

