Customize the Genome Browsers produced by G-OnRamp

Yating Liu

07/2016

# 1. Introduction

In addition to running the G-OnRamp workflow with default settings, one of the key features of Galaxy is the ability to modify an existing workflow (*e.g.*, change tool parameters, add or remove tools) using the Workflow Canvas. This walkthrough will cover topics including how to modify tool parameters, how to add or delete tools, and how to add or remove evidence tracks from the Hub Archive Creator.

Note that this tutorial assumes that the reader is familiar with the basic concepts of G-OnRamp. It will modify the “G-OnRamp:D. biarmipes F element“ workflow that we have created in the “Introduction to G-OnRamp Walkthrough”.

# 2. Modify G-OnRamp workflow

**Log into your account on the G-OnRamp Galaxy instance. Click on “Workflow” in the menu bar to access the list of available workflows. Click on the down** arrow for the “G-OnRamp:D. biarmipes F element” workflow and click on “Edit” to open the Workflow Canvas (Figure 1).

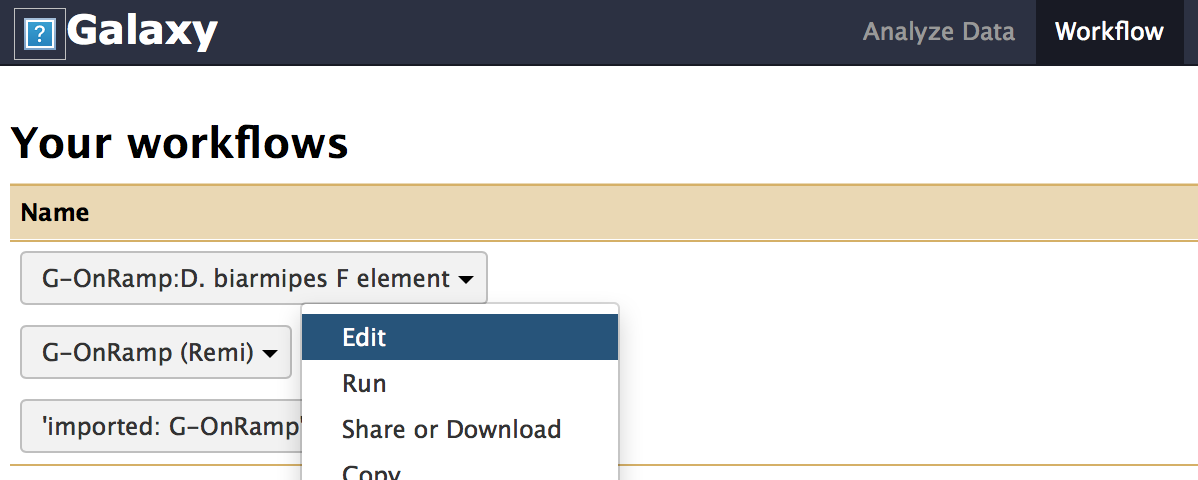


Figure 1: Click on "Edit" to open the Workflow Canvas

## 2.1 Big picture

The entire workflow is shown in Figure 2. Each box represents a tool. The “>” symbol on the left side of the box denotes an input dataset for the tool. The “>” symbol on the right side of the box denotes the output dataset produced by the tool. A tool could have multiple input datasets (*e.g.*, HISAT) and output datasets (*e.g.*, Augustus). The “noodles” between the tools correspond to how data are processed by the different tools within the workflow. Each noodle shows how the output dataset from one tool is used as the input dataset for another tool.

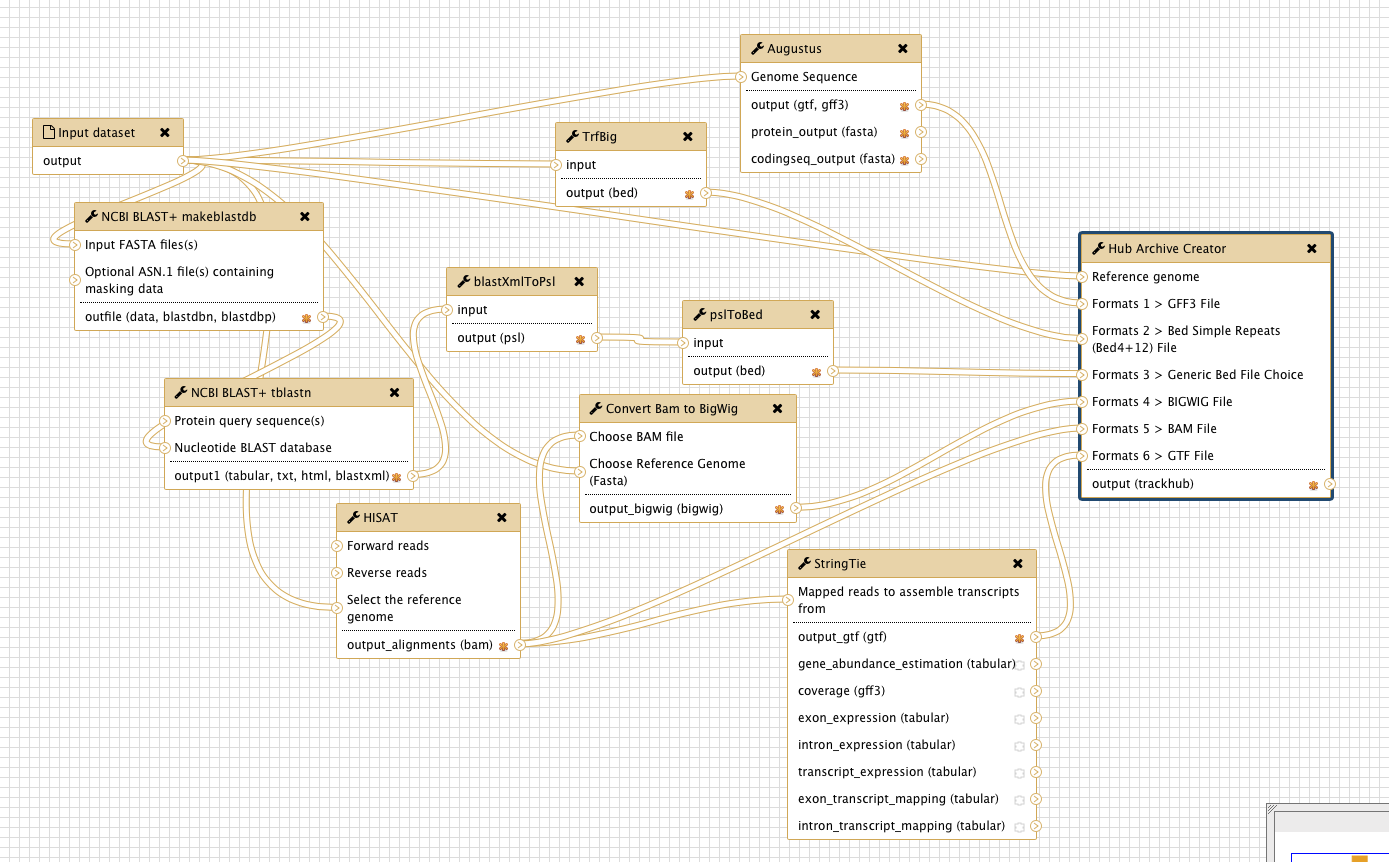


Figure 2: The entire G-OnRamp workflow shown in the Workflow Canvas

Figure 3 shows the connection between the “Input dataset” and the Augustus gene predictor. The output from the “Input dataset” is used as “Genome Sequence” input for Augustus.

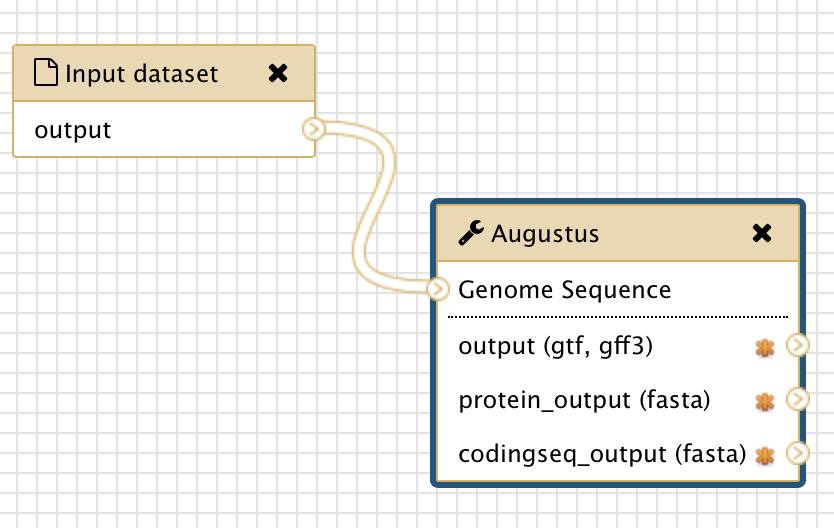


Figure 3: Connection between the “Input dataset” tool and the Augustus tool

## 2.2 Modify tool parameters

You can click on each tool in the Workflow Canvas to learn more about the tool (i.e. what it does and how to use it). You can use the “Details” panel on the right to examine and change the parameter settings for each tool (Figure 4). If the analysis workflow was originally derived from a History, then the settings of each tool within the workflow will reflect the parameters used in that analysis. Galaxy will use default parameters when you add a tool to the Workflow Canvas.

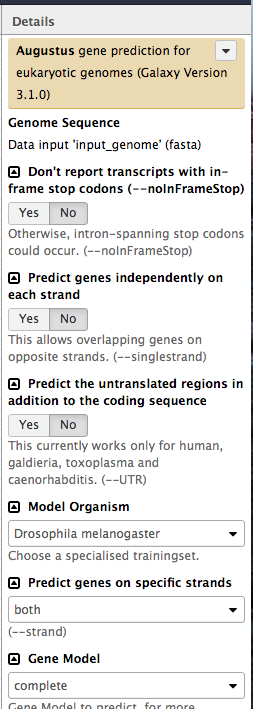


Figure 4: Use the Details panel to edit the parameter settings for the Augustus

Think about which parameter settings you would change if you are going to use this workflow to run your own datasets (*e.g.,* the “Model Organism” parameter for Augustus).

## 2.3 Add an evidence track to the Hub Archive Creator

You can also use the Workflow Canvas to add or delete a tool. For example, if you want to add another gene predictor such as GlimmerHMM to the G-OnRamp workflow, you can use the search field in the “Tools” panel to search for the tool and then click on the link to the tool (Figure 5). The tool will then appear at the Workflow Canvas (Figure 6).

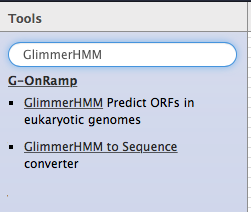


Figure 5: Use the search field in the Tools panel to search for the GlimmerHMM tool

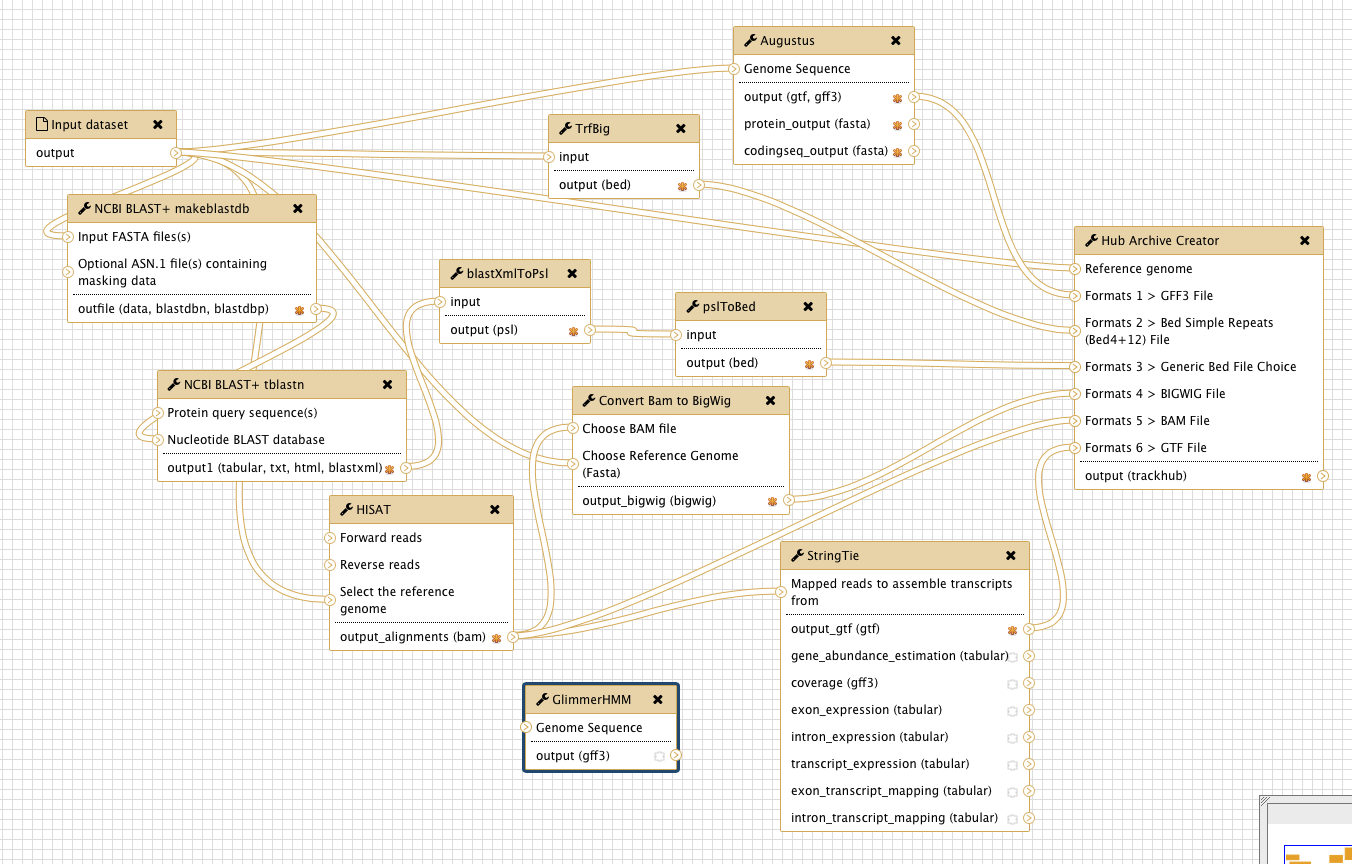


Figure 6: GlimmerHMM is added to the Workflow Canvas

The next step would be to incorporate the GlimmerHMM tool with the rest of the G-OnRamp workflow. This is accomplished by specifying the input and output datasets for the GlimmerHMM tool. There is a “>” symbol on the left side of “Genome Sequence”, which indicates that it requires the genome sequence (i.e. input dataset) as its input. You can create a new connection between the “Input dataset” tool and GlimmerHMM by clicking on the “>” symbol on the right side of the “Input dataset” box and dragging it to the “>” symbol on the left side of the “GlimmerHMM” box next to the “Genome Sequence” field. As you drag the connection, the connection will appear as a green “noodle” (Figure 7).

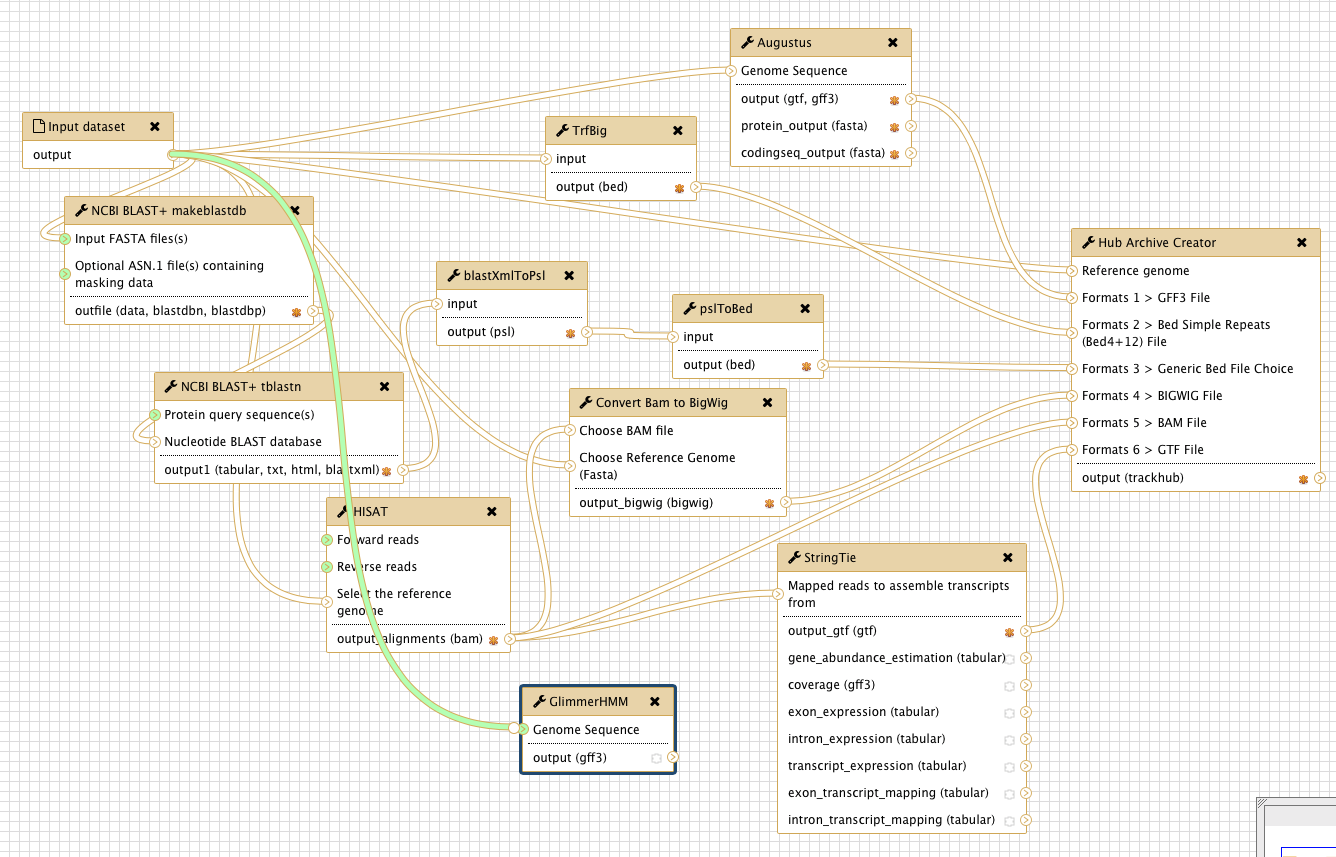


Figure 7: Connect the Input dataset to the GlimmerHMM tool

When you release the mouse next to the “Genome Sequence” field on the left side of the “GlimmerHMM” box, the connection will be created between these two tools.

Using the same approach, you can connect the output of GlimmerHMM to the Hub Archive Creator. However, because all the input connections to the Hub Archive Creator are connected to the output connections from the other tools, we need to add another input connection to the Hub Archive Creator before we can create the connection with GlimmerHMM.

Click on the Hub Archive Creator box in the Workflow Canvas and then examine the Details panel on the right. Scroll down to the bottom of the Details panel and click on the “Insert Formats” button (Figure 8).

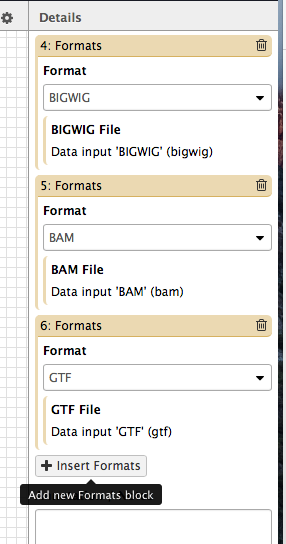


Figure 8: Click on the "Insert Formats" button in the Details panel to add a new input connection to the Hub Archive Creator

In order to establish a connection between two tools, the datatype of the output dataset from the first tool must be the same as the datatype of the input dataset for the second tool. In this case, the output format for GlimmerHMM is GFF3, which means that we need to select the “GFF3” format for the new input connection in the Hub Archive Creator (Figure 9).

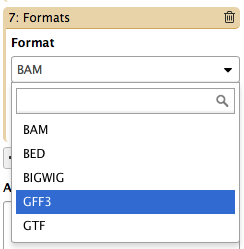


Figure 9: Choose the GFF3 format for the new input connection to the Hub Archive Creator.

A “Formats 7 > GFF3 File” entry will appear in the Hub Archive Creator tool (Figure 10).

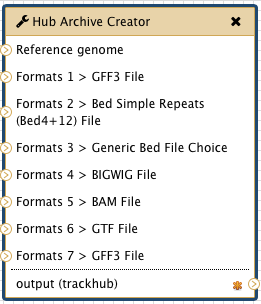


Figure 10: The Hub Archive Creator with a new GFF3 input connection

After you have created the new GFF3 input connection, you can connect the output of the GlimmerHMM tool to the input of the Hub Archive Creator. Click on the “>” symbol next to the “output (gff3)” field in the GlimmerHMM tool, drag it to the “>” symbol next to the “Format 7 > GFF3 File” field in the Hub Archive Creator tool and then release the mouse (Figure 11).

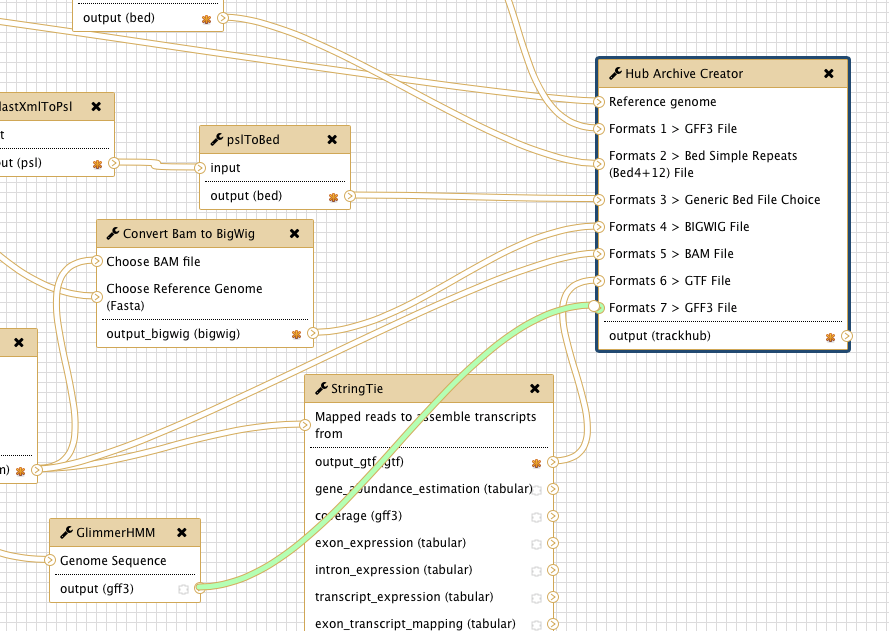


Figure 11: Connect the output of GlimmerHMM to the Hub Archive Creator

## 2.4 Remove an evidence track from the Hub Archive Creator

To delete a tool from the workflow, click on the “x” at the top right corner of that tool. The tool and all its connections will be removed from the workflow. For example, click on the “x” at the top right corner of the GlimmerHMM tool to remove it from the workflow (Figure 12).

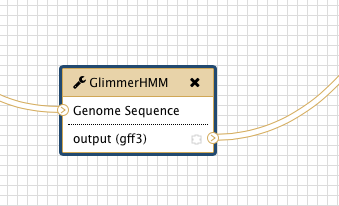


Figure 12: Delete GlimmerHMM from the workflow

When you delete the GlimmerHMM tool, the connections between GlimmerHMM and the Hub Archive Creator will be removed. However, the “Format 7 > GFF3 File” input connection will need to be removed manually.

Click on the Hub Archive Creator tool and scroll down to the last “Formats” entry in the Details panel. Click on the “Trash” icon at the top right corner to delete the entry (Figure 13).

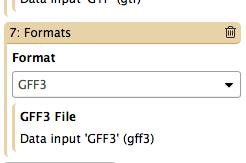


Figure 13: Remove an evidence track from the Hub Archive Creator

## 2.5 Show or hide a result dataset

To show a dataset in the History, you can mark the dataset as a workflow output by clicking on the “\*” symbol. All unmarked datasets will be hidden from your History. For example, StringTie will produce eight output files. However, because only the “output\_gtf (gtf)” output is marked, only the gtf file will appear in your history after you run the workflow (Figure 14). The other seven output datasets will be hidden from the History. This feature is particularly useful when you are working with large workflows that produce many temporary datasets.

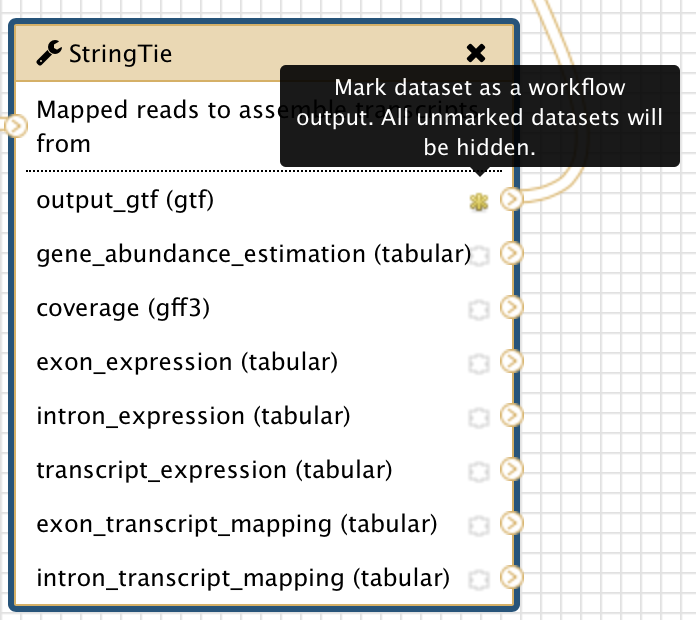


Figure 14: The yellow star indicates that the dataset has been marked as a workflow output

## 2.6 Save the changes to the workflow

Don’t forget to save your changes before you leave the page. Click on the settings icon at the top right corner of Workflow Canvas and then click on “Save” (Figure 15).

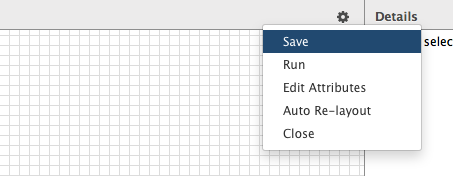


Figure 15: Save the changes that you have made to the workflow

# 3. Exercise: edit the workflow to run RNA-Seq analysis on three different RNA-Seq samples

The three RNA-Seq test datasets that we will use in this exercise are available in the “Dbia3” folder within the Intro\_G-OnRamp folder on your Desktop.

## 3.1 Make a copy of the current workflow

Make a copy of “G-OnRamp:D. biarmipes F element” workflow and rename the new workflow as “G-OnRamp:Dbia3\_RNA-Seq” (Figure 16). Right click on the new workflow and click on Edit to go to the Workflow Canvas.

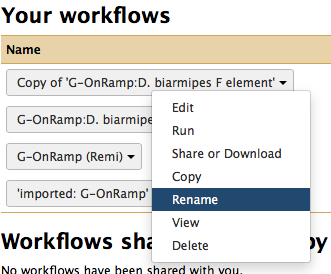


Figure 16: Copy the current workflow and rename it as “G-OnRamp:Dbia3\_RNA-Seq”

## 3.2 Edit the “G-OnRamp:Dbia3\_RNA-Seq” workflow

Given that there are three paired-end RNA-Seq datasets, you can run the HISAT and StringTie analyses on each dataset within a single workflow and then view the evidence tracks for all three samples on the UCSC genome browser.

Below are some hints on how to construct this workflow:

First, you need to add two additional instances of the HISAT and StringTie tools to the workflow in order to perform the RNA-Seq read alignment and the transcript assembly for the three samples.

Second, you need to modify the parameter settings for HISAT. Select the “Individual paired reads” option under the “Single end or paired reads?” field (Figure 17). Select the “Use a genome from history” option under the “Source for the reference genome to align against” field (Figure 18).

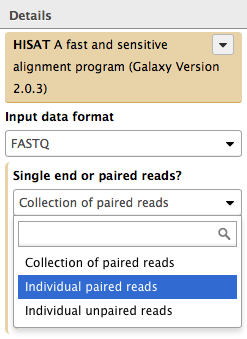


Figure 17: Choose the “Individual paired reads” option under the “Single end or paired reads” field

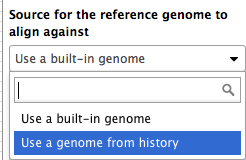


Figure 18: Choose “Use a genome from history” for “Source for the reference genome to align against”

Third, you need to incorporate the new instances of the HISAT and StringTie tools with the rest of the G-OnRamp workflow. (Hint: you will need to add input connections to the Hub Archive Creator in order to connect the outputs from the HISAT and StringTie tools to the Hub Archive Creator.)

## 3.3 Upload your datasets and run the workflow

The test datasets that we will use in this walkthrough are available in the “Dbia3” folder within the Intro\_G-OnRamp folder on your Desktop. There are eight datasets in the folder: the reference genome sequence from the *Drosophila biarmipes* Muller F element, a collection of *Drosophila melanogaster* protein sequences, and the forward and reverse paired-end reads from three Dbia3 RNA-Seq samples (i.e. adult females, adult males and mixed embryos). Upload your datasets to your Galaxy account and run the “G-OnRamp: Dbia3\_RNA-Seq” workflow. Finally, you can then view the RNA-Seq results produced by the modified G-OnRamp workflow as different evidence tracks on the UCSC genome browser (Figure 19).

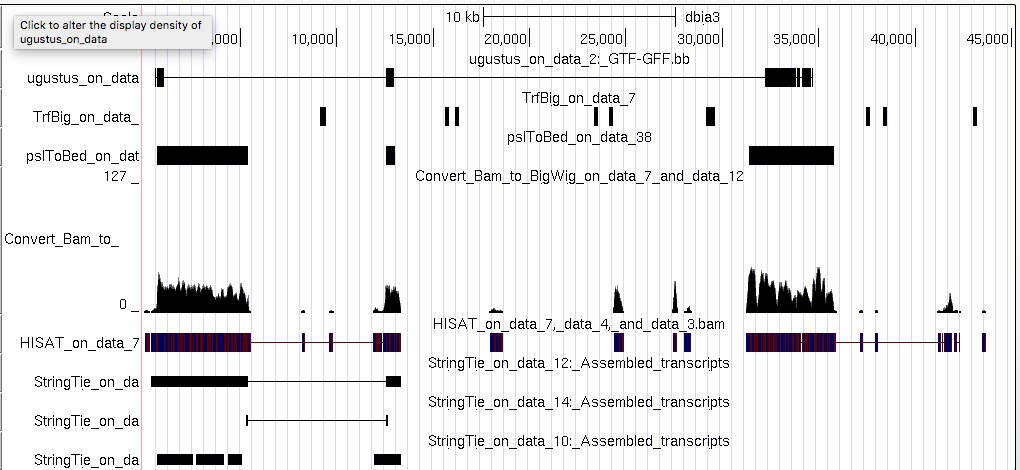


Figure 19: View the Dbia3 genome assembly and the evidence tracks produced by the Hub Archive Creator on the UCSC genome browser