Exercises for Galaxy
(Version 2, 2010)

You can choose to do these using the questions as your only guide—or see the following pages for the step-by-step checklist to finding these answers.

Exercises 2 and 3 flow from the previous exercise. It will be best to do all three exercises at one sitting, or save the previous exercise in the Galaxy history to continue with later. A history is automatically saved. If you are registered and logged in, you can return to that history from any computer and work on it from your list. If you are not logged in, the history is still saved, but can only be accessed from the computer from which you created it.

1. Obtain UCSC Genome Browser data using the Galaxy interface. Examine the data, and begin to prepare the data for subsequent analyses by editing attributes.

Skills: Obtaining data, editing annotations, naming and tagging history

2. Obtain 50bps of flanking sequence from SNPs obtained in exercise 1

Skills: manipulating and analyzing data.

3. View history list and then create a workflow from exercises 1 and 2.

Skills: viewing histories and creating workflows from histories.

Galaxy, version 2.0.
Correspond to the data available in February 2010.

The materials and slides offered are for non-commercial use only. Reproduction, distribution and/or use for commercial purposes are strictly prohibited.
Copyright 2010, OpenHelix, LLC.

Seattle
12600 SE 38th Street, Suite 230
Bellevue, WA 98006
(425) 401-1400

Boston
65 Main Street
Somerville, MA 02145
(617) 627-9398

San Francisco
193 Haight Street
San Francisco, CA 94012
(415) 252-1519

Step-by-Step checklist/instructions for Galaxy exercises

1. Obtain UCSC Genome Browser data using the Galaxy interface. Examine the data, and begin to prepare the data for subsequent analyses by editing attributes.

Step	Action	✓
1	Access the **Galaxy site** (http://galaxy.psu.edu), and click on **Go to our public site** in the top green box area.	
2	**Click on the "Get Data" link in the left Tools area** to expand that topic**. Click UCSC Main.**	
3	On the right frame the UCSC Table Browser interface will be displayed. Make these changes to the UCSC interface: **In clade choose "Mammal," "Human" and "Mar. 2006" for assembly.** **In group**: **choose Variation and Repeats** and for Track choose "SNPS (130)." *(primary data table, "snp130", will be set automatically for you)* **In region**: move radio button to **position**, and clear out the default location. **Enter BRCA1** in the text box. Click **lookup**. From the results list **choose the first BRCA1 item (uc010cza.1)** at the top by clicking on it. Its position will display back in the position box. **Leave any other menu settings as default at this time.**	
4	Leave the output format menu set to "BED – Browser extensible Data". **Check the box "Send output to Galaxy".**	
5	**Click the "get output" button** near the bottom of the form.	
6	The next page to load offers a **"Send query to Galaxy" button**. Leaving all as default, **click that button.**	
7	You should now have an item in your **history** on the right side. It starts of grey or yellow, and will turn green when complete.	
8	**Click the title hyperlink** to examine the summary data.	
9	**Click the eyeball icon** to view the data in the center panel area.	
10	**Click the pencil icon** to be able to prepare the data set for subsequent analyses.	
11	**Change the name to: BRCA1 SNPs.**	
12	**Click the "Save" button** at the lower area of the center panel.	
13	Your file has been renamed and the interface will indicate that your attributes have been updated. Change the history name **to "BRCA1 SNPs Flanking" by clicking the pencil icon next to the words "unnamed history" and typing those words, hit enter.** If you are registered and signed in, you can **add tags by clicking the text "0 tags" and start typing. Add "snps" and then a space, and "brca1" and then a space.**	

2. Obtain 50bps of flanking sequence from SNPs obtained in exercise 1.

Step	Action	✓
1	You will need to obtain the flanking region and then sequence of the SNPs you downloaded in exercise 1. If you are doing this exercise immediately after exercise 1, the data will be in your history. **If not, go to the "Options" pull down menu** at the top of the history column and **click "Saved Histories**." Your previously created history "BRCA1 SNPs Flanking" will be there. **Click on the title and pulldown menu, click "Switch"** to add history to your current analysis.	
2	First, get the flanking region of these SNPs. In the left Tools column, **click the heading "Operate on Genomic Intervals." Then click on the tool "Get Flanks...".**	
3	The options for getting the flanking regions will appear in the middle column**. Change the "Location of the flanking region/s" menu to "both"** (alternately you could have chosen either upstream or downstream). **Leave all the rest of the options as default** (Select Data: BRCA1 SNPs, Region: Whole Feature, **Location of Flanking region: both**, Offset: 0, Length of flanking region: 50). **Click "Execute."**	
4	This will result in a second dataset with locations of 50bp upstream of each SNP. Now obtain the sequence of these locations. In the left column, **click the heading "Fetch Sequences." Click the tool "Extract Genomic DNA."**	
5	The options appear in the middle column. **Make sure "2: Get Flanks on data 1" is chosen in the first pull down menu "Fetch sequences…" Then change the second pull down menu (output data type) to "Interval.**" FASTA will return a FASTA formatted raw sequence. Interval will return a tab-delineated file with the sequence, but also the SNP IDs. **Click "Execute."**	
6	**Open the latest dataset "Extract Genomic DNA on data 2" by clicking the linked title**. You'll notice the first three columns are locations, the 4th is SNP ID, the 5th column is unused (0), the 6th column is strand and 7th is the flanking squence. **You can now view the data (click eye icon) or save the tab-delineated text file to your computer by clicking the "save" link in the dataset description. Save this analysis history as you did in exercise 1.**	

3. View history list and then create a workflow from exercises 1 and 2, run workflow using a different dataset. Note: *for workflows in this exercise you will need to be registered and logged in. This is free and very simple. We recommend you register.*

Step	Action	✓
1	If you are doing this exercise immediately after exercise 1, the data will be in your history. **If not, go to the "Options" pull down menu** at the top of the history column and **click "Saved Histories**." Your previously created history from exercise 2 will be there. **Click on the title and pull down menu, click "Switch"** to add history to your current analysis.	
2	**Click "Options" in the right hand History column. Click "Extract Workflow."** Make sure there are 3 steps and that all check boxes ("Treat as input dataset,"Get flanks," and Extract Genomic DNA) are checked**. Type in a workflow name such as "Get SNP flanking" in the workflow name text box and click "Create Workflow."**	
3	Your new workflow is now created and you can run this analysis on any similar dataset. Before we look at this workflow and run it, let's clear the history. **Click the Options menu in the right hand History column and click "Create New."** Your history is now cleared. Let's get some new data to run this analysis on.	
4	Repeat the steps 1-12 of exercise 1, but this time instead of typing in BRCA1, type in CLOCK and click the first linked result (uc003hba.1) and when naming, name it CLOCK SNPs or something descriptive.	
5	Once the CLOCK SNP data is uploaded to Galaxy**, Click the "Workflow" tab** at the top of the screen.	
6	You will now be at a list of any workflows you've created. **Click the workflow you just created, "Get SNP Flanking" and choose "Run"** in the resulting pull down menu.	
7	The workflow steps will now appear in the middle column. In "Step 1: Input dataset," there will be a pull down menu with your current datasets. Your newly uploaded **dataset "Clock SNPs" should be there, choose it if it already isn't and click "Run Workflow" at the bottom.**	
8	The workflow will then proceed to do the steps of the analysis you did in exercise 2 automatically on the new dataset! When completed, **click the heading of the third dataset "Extract Genomic DNA on data 2"** and view the data. There should be 816 regions and you'll notice the locations, IDs and sequence. Workflows are great a great method to redo analyses automatically and simply on many datasets.	