

UW/Sage workshop tutorial: Accessing the brain-derived snRNA-seq data from an Alzheimer's Disease case control study

IMPORTANT: You will not be able to access the data in Synapse or Sage's AWS services without completing the following steps **BEFORE** the workshop ([see Anna's workshop notification emails for details](#)):

- A. Register for a Synapse account
- B. Apply for access to ROSMAP data by filing a data use certificate request form

WORKSHOP TUTORIAL:

Setting up your Rstudio notebook for the workshop, using Sage's AWS Service catalog.

1. Open a web browser and go to 'sc.sageit.org'. Use your Synapse credentials to log in; click "Allow" when prompted for AWS Service Catalog permissions.

The image shows a web browser window with the URL `signin.synapse.org/?response_type=code&client_id=100053&redirect_uri=https%3A%2F%2Fsynapse-login-scipoolprod.scipool...`. The browser's address bar and tabs are visible at the top. Below the browser window, the SageBionetworks logo is displayed. The main content area shows a login form with two options: "Sign in with Google" and "Sign in with your Sage Bionetworks Synapse account". The second option is selected, and the form contains a text input field with the value "lhealth", a password field with masked characters, and a "Sign in" button. Below the form are links for "Forgot password?" and "Register It's free!".

Below the login form, the SageBionetworks logo is displayed again. The main content area shows a permission dialog box titled "Synapse AWS Service Catalog requests permission:". The dialog lists the following permissions:

- To see your Synapse user ID, which can be used to access your public profile
- To see your Synapse username
- To see your company, if you share it with Synapse
- To see your first name, if you share it with Synapse
- To see your last name, if you share it with Synapse
- To see your team membership

Below the list, the text states: "By clicking 'Allow', you allow this app to use your information in accordance with their [terms of service](#) and [privacy policy](#)." At the bottom of the dialog are two buttons: "Deny" and "Allow".

- After logging in you will land on the “Products” page, listing the available services. Click the link to “EC2: Ubuntu Linux with Notebook Software.” Launch Product.

The screenshot displays the AWS Service Catalog console. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a user profile. The main header reads 'Products list'. Below this is a search bar and a table of products. The table has columns for Product name, Vendor, Owner, and Description. The product 'EC2: Ubuntu Linux with Notebook Software' is circled in blue. Below the table, the 'Product details' page for the selected product is shown. It includes a description, owner, distributor, email contact, support link, and support description. A 'LAUNCH PRODUCT' button is at the bottom.

Product name	Vendor	Owner	Description
EC2: Linux	Sage Bionetworks	Sage Bionetworks	This product builds one Linux EC2 instance, either Amazon Linux or Ubuntu
EC2: Ubuntu Linux with Notebook Software	Sage Bionetworks	Sage Bionetworks	This product builds one Linux EC2 instance using an Rstudio AMI based on Ubuntu and maintained by Sage Bionetworks.
EC2: Ubuntu Linux with Workflow Software	Sage Bionetworks	Sage Bionetworks	This product builds one Linux EC2 instance using an Ubuntu AMI with workflows software installed. Maintained by Sage Bionetworks.
EC2: Windows	Sage Bionetworks	Sage Bionetworks	This product builds one Microsoft Windows EC2 instance with Jumpcloud integration.
S3: Private Encrypted Bucket	Sage Bionetworks	Sage Bionetworks	This product builds an AWS S3 bucket encrypted with private access accessible from any source.

EC2: Ubuntu Linux with Notebook Software

This product builds one Linux EC2 instance using an Rstudio AMI based on Ubuntu and maintained by Sage Bionetworks.

Owner: Sage Bionetworks

Distributor: Sage Bionetworks

Email contact: sc-support@sagebase.org

Support link: <https://www.synapse.org/#!Synapse:syn2580853/discussion/default>

Support description: Sage Bionetworks Support

LAUNCH PRODUCT

3. You will land on the Product Version page:
 - a. Name your instance in the Name* blank
 - b. Click button next to the latest version
 - c. Click “NEXT”

Launch - EC2: Ubuntu Linux with Notebook Software

Product version
Parameters
TagOptions
Notifications
Review

Product Version

Specify a provisioned product name and then select the version that describes the provisioned product that you want to create.

Provisioned product

A provisioned product is a collection of related resources that you provision and update as a single unit.

Name*

Product Version

Version*

By name <input type="text"/>				
	Versions ▾	Provided by ▾	Created time ▾	Description ▾
<input checked="" type="radio"/>	v1.1.8	Sage Bionetworks	Oct 1st 2020 13:15:36 ...	Remove VpcName input. Last updated at 2020-10-...
<input type="radio"/>	v1.1.7	Sage Bionetworks	Sep 18th 2020 15:56:13...	Use SynapseTagger
<input type="radio"/>	v1.1.3	Sage Bionetworks	Aug 13th 2020 19:33:41...	Update Cloudformation hooks
<input type="radio"/>	v1.1.2	Sage Bionetworks	Jul 16th 2020 14:40:02 ...	Remove Jumpcloud Integration.
<input type="radio"/>	v1.1.1	Sage Bionetworks	Jul 13th 2020 09:55:42 ...	adds alb-listener-rule custom resource, enabling co...

*Required

CANCEL NEXT

4. Select parameters
 - a. For EC2 Instance Type, scroll to “r5.4xlarge.”
 - b. Notebook type = Rstudio.
 - c. Disk size = 32.
 - d. Click “NEXT”

Launch - EC2: Ubuntu Linux with Notebook Software

Product version
Parameters
TagOptions
Notifications
Review


Parameters

Specify values or use the default values for the parameters.

Linux Instance Configuration

EC2 Instance Type ▾
Amazon EC2 Instance Type

Notebook Type ▾
Type of notebook software to install

Disk Size 
The EC2 volume size (in GB)

CANCEL PREVIOUS NEXT

5. Use drop-down menu to fill in TagOptions
 - a. Project = amp-ad
 - b. Department = NDR.
 - c. Leave new key field blank
 - d. Click “NEXT”

aws

Services

Resource Groups

ServiceCatalogEndusers/3403...N. VirginiaSupport

Products list - Product details - Launch product

We've redesigned the AWS Service Catalog end user console to make it easier to use. [Try out the beta console today.](#)

Launch - EC2: Ubuntu Linux with Notebook Software

Product version

Parameters

TagOptions

Notifications

Review

TagOptions

The following TagOptions are added by the administrator to either Product or Portfolio. Some TagOptions may have multiple values and you are required to select the most applicable value. [Learn more.](#)

Key (127 characters maximum)	Value (255 characters maximum)	
Project	amp-ad	
Department	NDR	
New key	New value	+

CANCEL

PREVIOUS

NEXT

6. Ignore the Notifications page, click “NEXT”

Launch - EC2: Ubuntu Linux with Notebook Software

Product version

Parameters

TagOptions

Notifications

Review

Notifications

☐Enable provisioned product event notifications to be streamed to an Amazon SNS topic.

CANCEL

PREVIOUS

NEXT

7. Review and launch your instance.

Launch - EC2: Ubuntu Linux with Notebook Software

Product version

Parameters

TagOptions

Notifications

Review

Review

Product version EDIT

Name: lhealth-scrnaseq-wkshop5
Product: EC2: Ubuntu Linux with Notebook Software
Product version: v1.1.8
Description: Remove VpcName input. Last updated at 2020-10-07 16:39:45.
Provider: Sage Bionetworks
Launch as: ("description":"null","type":"LAUNCH")

Parameters EDIT

Linux Instance Configuration
EC2InstanceType: r5.4xlarge
NotebookType: Rstudio
VolumeSize: 32

Tags EDIT

Project: amp-ad
Department: NDR

Notifications EDIT

No SNS topics provided

Resource changes CREATE PLAN

CANCEL PREVIOUS LAUNCH

8. Go to the “Provisioned products list” menu (in the menu icon, upper left corner) and wait for your notebook to change status.

us-east-1.console.aws.amazon.com/servicecatalog/home?region=us-east-1&isSceuc=true#/stacks

Apps AWS_instance Synapse GitHub Sage Intranet JumpCloud Reprints Desk AWS Console Paychex Confluence Jupyter Zoom

aws Services Resource Groups ServiceCatalogEndusers/3403... N. Virginia Support

Provisioned products list

Search

Ex. Name OR status:available

Filter by Sort by Count

User Relevance 100

Showing 1-3

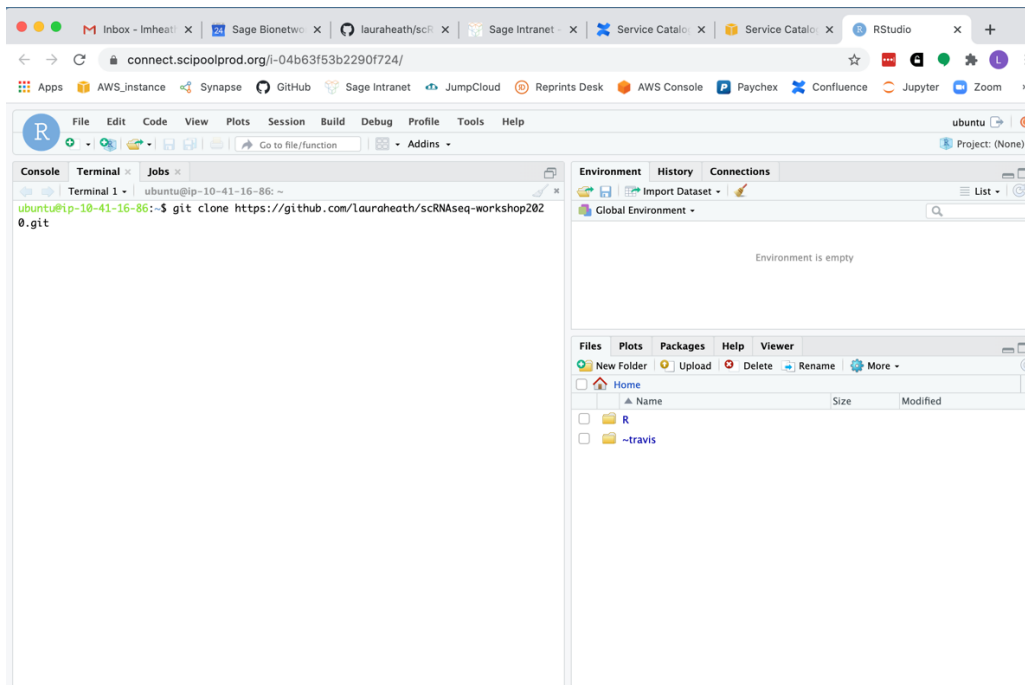
Provisioned Product name	Type	ID	Created	Status	Status message
lhealth-workshoptest3	CFN_STACK	pp-dmmrmsrh24cai	2020-08-24 10:30:50 UTC-0700	UNDER_CHANGE	

9. When the Status says **Available**, click on the provisioned product name (in this example, [lhealth-workshoptest3](#)). Scroll down to the Outputs list and click on the link next to NotebookConnectionURI. This will open an Rstudio notebook in a new tab in your browser.

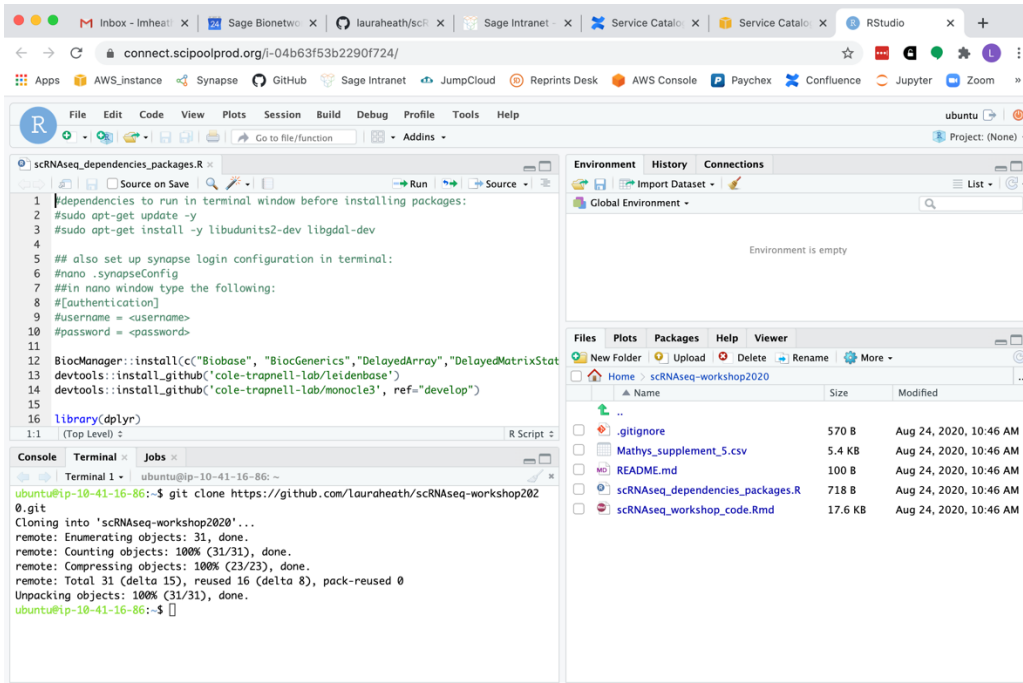
▼ Outputs:

Key	Value	Description
CloudformationStackARN	arn:aws:cloudformation:us-east-1:237179673806:stack/SC-237179673806-pp-dmmmsrh24cai/8e600410-e62f-11ea-8a70-12695902b27f	The ARN of the launched Cloudformation Stack
EC2ConsoleURI	https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#Instances:search=i-04b63f53b2290f724	Check your instance status with this link to the ...
LinuxInstanceId	i-04b63f53b2290f724	The ID of the EC2 instance
LinuxInstancePrivateIpAddress	10.41.16.86	The IP Address of the EC2 instance
ConnectionInstructions	https://sagebionetworks.jira.com/wiki/spaces/IT/pages/996376579/Connect+to+Provisioned+Instances	Guidelines on connecting to instances
NotebookConnectionURI	https://connect.scipoolprod.org/i-04b63f53b2290f724/	Notebook server login page
EC2InstanceType	r4.2xlarge	The EC2 instance type
ConnectionURI	https://us-east-1.console.aws.amazon.com/systems-manager/session-manager/i-04b63f53b2290f724?region=us-east-1	Starts a shell session in the AWS Console

10. Go to the Rstudio Terminal (next to the Console tab, upper left) and clone the scRNA seq workshop repository with the following command:
`git clone https://github.com/lauraheath/scRNAseq-workshop2020.git`



11. You will see the `scRNAseq-workshop2020` folder added to your files (lower right panel). Click on it, then click on “`scRNAseq_dependencies_packages.R`”.

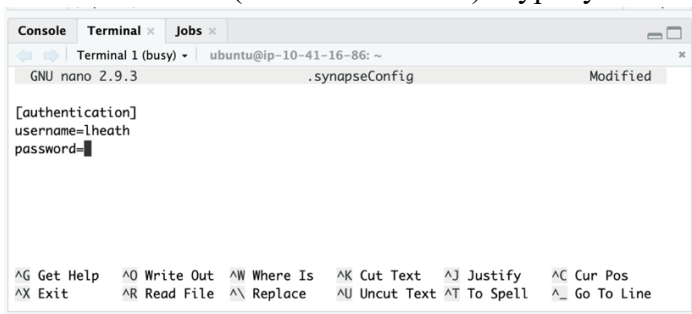


12. First, we need to install some dependencies. Copy line 2 (“`sudo apt-get update -y`” without the hashtag), paste into the terminal window prompt, and hit return; then do the same for line 3.

13. To enable easy Synapse login from your notebook, set up your login configuration as follows: after typing “`nano .synapseConfig`” in Terminal, a nano environment will appear. Type the following, line by line (no hashtags, don’t forget the brackets, and use your own Synapse login information:

```
[authentication]
username = yourSynapseUsername
password = yourSynapsePassword
```

Exit nano (control-X on Mac). Type ‘y’ when prompted to save the `.synapseConfig` file.



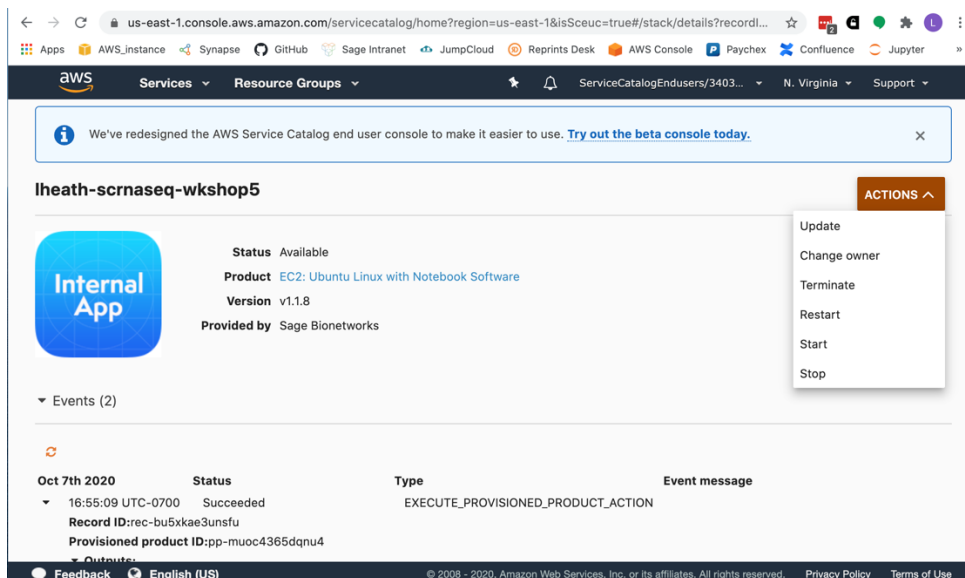
14. Now install the necessary packages one line at a time (Lines 12-14 of `scRNAseq_dependencies_packages.R`) and install libraries (lines 16-20). Installing the packages will take ~35 minutes. Also, the `monocle3` install will immediately prompt you to update existing packages (do update the existing packages by typing “1” when prompted).

NOTE: if errors appear when attempting to load libraries (lines 16-20), restart R session and rerun lines 16-20

15. Now you are ready to upload scRNA seq data from Synapse and use Monocle3 for trajectory analyses. Launch “`scRNAseq_workshop_code.Rmd`” and you are on your way!

POST WORKSHOP CLEANUP: When finished working within your notebook, please return to the AWS console and stop or terminate your instance (under the “Actions” button). You can return later and start your instance back up again if you do not terminate it. If you have a github account, you can clone the workshop code and any changes you made today into your own branch. **DO NOT store ROSMAP data in your github repo or outside of the EC2 instance/Synapse environment, per the DUC agreement for ROSMAP.** We encourage users to avoid storing data on their instances in the long term.

Stopping, starting, and terminating your instance:



Finding the Rstudio notebook URL after restarting an instance (go to the most recent “Succeeded” tab, then “Outputs”):

▼	11:53:47 UTC-0700	Succeeded	PROVISION_PRODUCT
	Record ID:rec-i4xz3wyhtuh3w		
	Provisioned product ID:pp-muoc4365dquu4		
	▼ Outputs:		
Key	Value	Description	
CloudformationStackARN	arn:aws:cloudformation:us-east-1:237179673806:stack/SC-237179673806-pp-muoc4365dquu4/94d113a0-08cd-11eb-91e4-0acaf3694d17	The ARN of the launched Cloudformation Stack	
LinuxInstancePrivateIpAddress	10.41.26.184	The IP Address of the EC2 instance	
LinuxInstanceId	i-0ce3e6da467ae8b5a	The ID of the EC2 instance	
EC2ConsoleURI	https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#instances:search=i-0ce3e6da467ae8b5a	Check your instance status with this link to the AWS Conso...	
ConnectionURI	https://us-east-1.console.aws.amazon.com/systems-manager/session-manager/i-0ce3e6da467ae8b5a?region=us-east-1	Starts a shell session in the AWS Console	
EC2InstanceType	r5.4xlarge	The EC2 instance type	
NotebookConnectionURI	https://connect.scpoolprod.org/i-0ce3e6da467ae8b5a/	Notebook server login page	
ConnectionInstructions	https://sagebionetworks.jira.com/wiki/spaces/IT/pages/996376579/Connect+to+Provisioned+Instances	Guidelines on connecting to instances	