Running SMRT® Analysis on Amazon

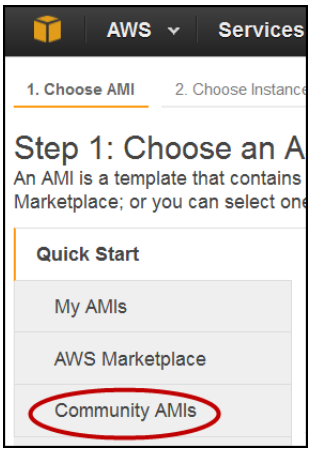
SMRT Portal requires significant hardware and system administration resources for installation and maintenance. This can be a challenge. There is a very practical alternative for those who only need occasional access to SMRT Portal: You can run SMRT Portal on the Amazon Cloud using a public **Amazon Machine Image (AMI)** that Pacific Biosciences maintains and upgrades. You pay **only** for the machine time that you use.

This option is useful for running **small** to **medium-sized** secondary analysis jobs (bacterial genome assembly, bacterial base modification analysis, long amplicon analysis, Iso-Seq™ analysis), and for using SMRT Portal command-line tools. The AMI we provide is **single-node** only, and does **not** support distributed computing.

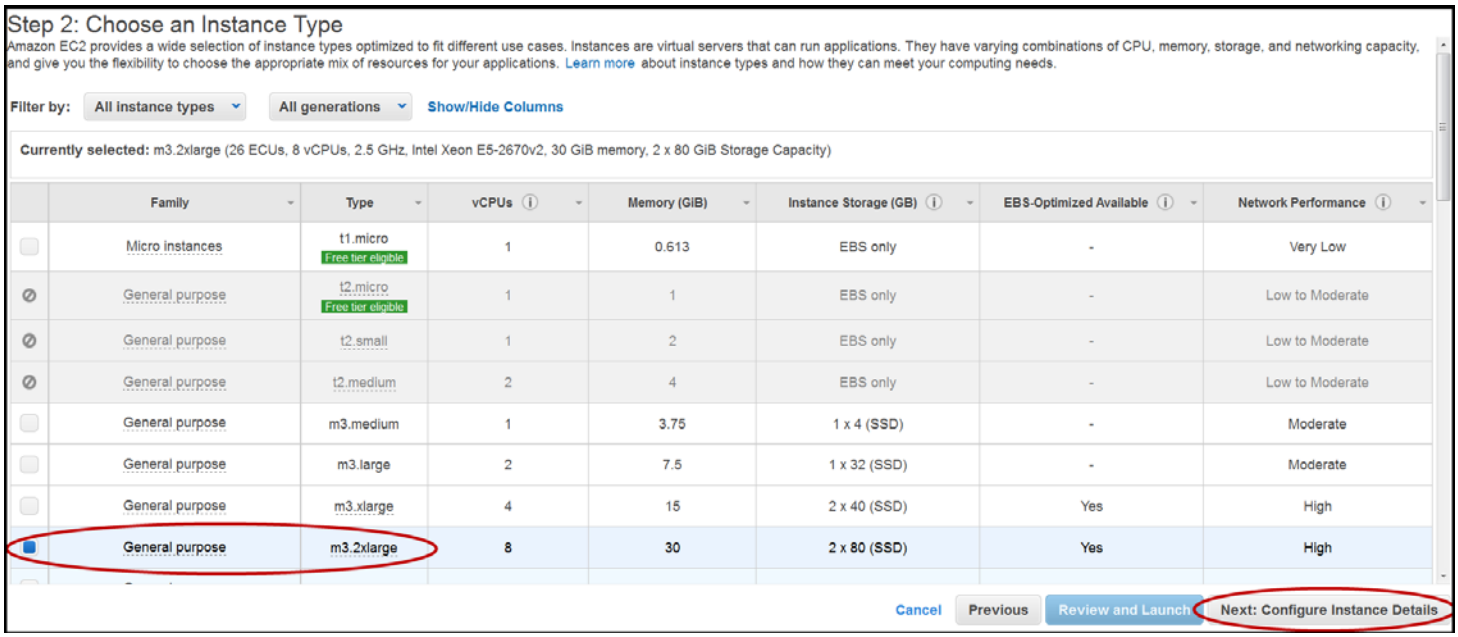
To analyze data with SMRT® Analysis using the AMI, follow these steps:

Step 1: Setting up the Amazon Machine Image

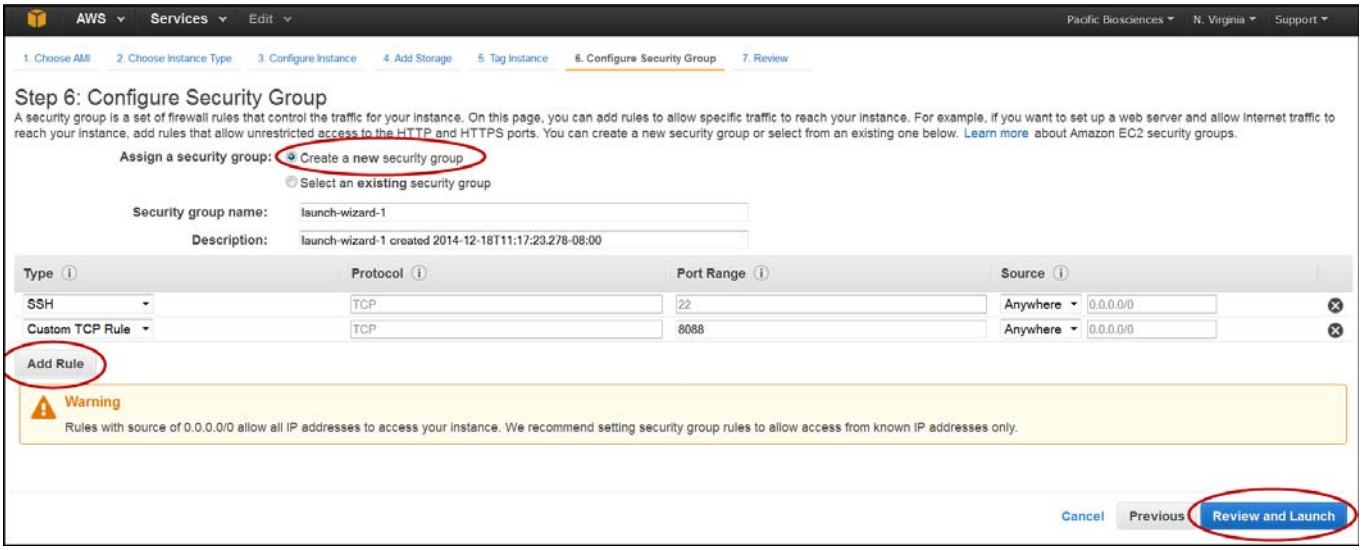
1. Go to `http://aws.amazon.com` and create an Amazon Web Services™ account.
2. Click **My Account**, then choose **AWS Management Console**. (Sign in first if asked.)
3. Set the location to **US East (N. Virginia)** on the upper-right of the page.
4. Click **EC2**.
5. Click **Launch Instance.**
6. Click the **Community AMIs** tab, under **Quick Start**.



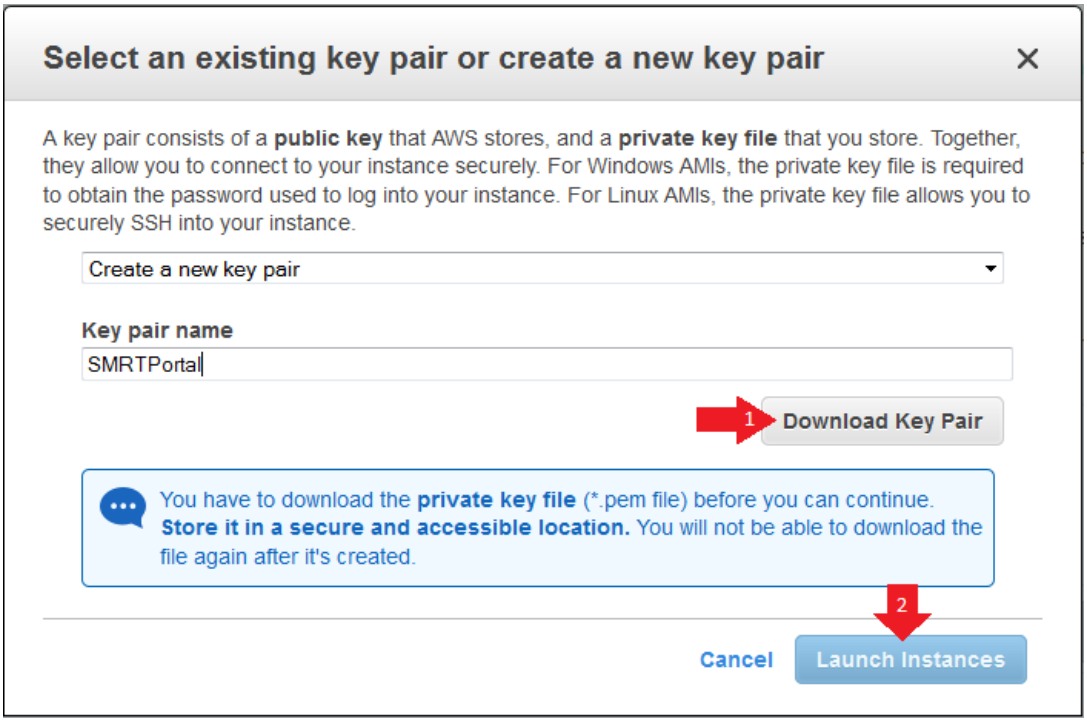
7. Search for **smrt**, choose the latest instance, then click **Select**. (This may take several minutes to load.)
8. Choose an appropriate Instance Type for your analysis, then click **Next: Configure Instance Details**.
   - For **bacterial analysis**, select **m3.2xlarge** from the **General Purpose** family.
   - For **mammalian analysis**, select **r3.8xlarge** from the **Memory Optimized** family.
   - **Note**: A micro instance is **not** sufficient to analyze lambda.
   - Note the number in the **vCPUs** column. After completing the setup, set the `NPROC` setting in `/analysis/etc/smrtpipe.rc` to the correct number selected in this step.



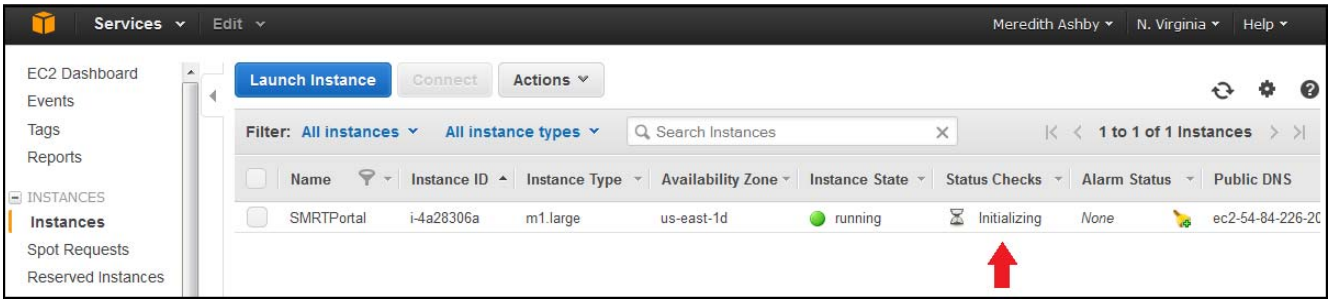
9. Leave **Step 3: Configure Instance Detail** as is and click **Next: Add Storage**.
10. Leave **Step 4: Add Storage** as is and click **Next: Tag Instance**.
11. Enter an arbitrary string for the **Name** value (`smrtportal` is a good default.), then click **Next: Configure Security Group**.
12. Under **Assign a security group**, select **Create a new security group**. If you have launched a SMRT Portal AMI before, you can choose **Select an existing security group** and use one that you previously configured.
13. To access SMRT Portal both via the web and via ssh, you need to open the required ports on the server you are launching. Under **Protocol**, `SSH with Port Range 22` should already be listed.
14. Click **Add Rule**.
15. From the **Type** pull-down menu, select **Custom TCP Rule**. Enter `8080` in the **Port Range** field. (This opens the port necessary for SMRT Portal web access.)
16. Click **Review and Launch**.



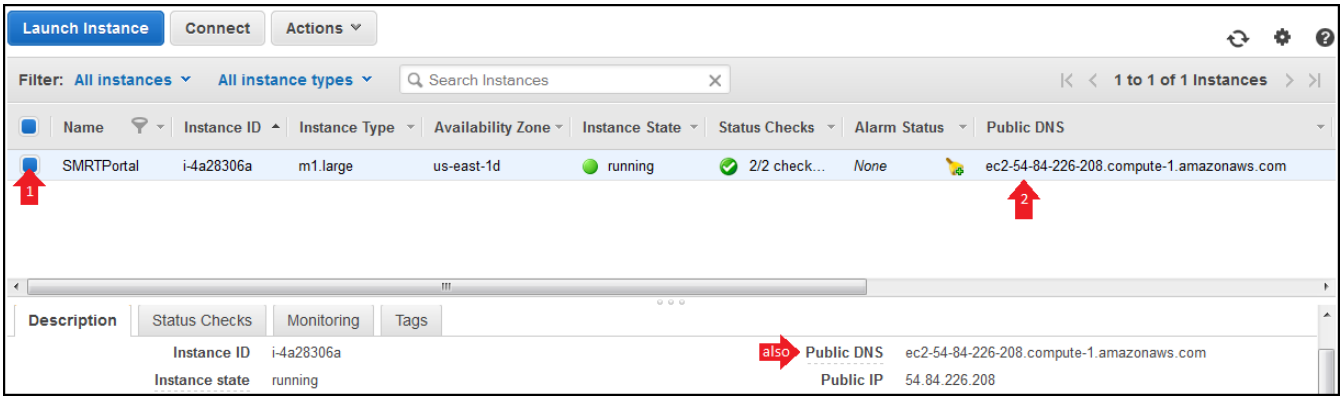
17. Disregard the security group and free usage tier warning and click **Launch**.
18. In the pop-up window, select **Create a new key pair** and name the key pair. Click **Download Key Pair** and save the file to a safe location on your local machine. (The key pair file has a `.pem` extension.)
19. Click **Launch Instances**.



20. To view your launched instance, click **View Instances** in the bottom right corner of the page. This will take you back to the Instances page of the EC2 Management Console (where we started). Your instance will take a few minutes to boot up and run through a status check.



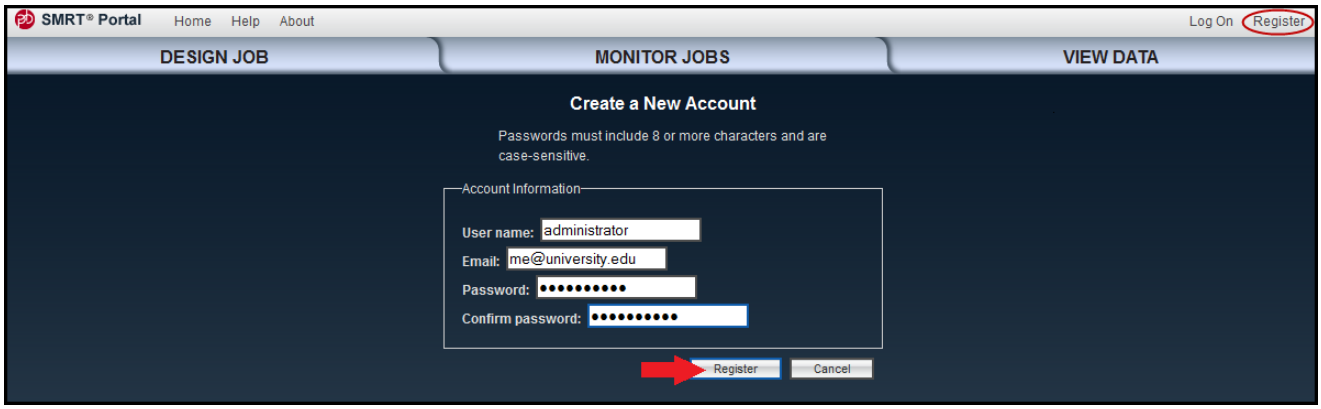
21. Select the instance you just created and note the Public DNS address in the **Description** tab.
    Note: You will need this address later on.



Step 2: Setting up SMRT® Portal on the Amazon Machine Image

Create a SMRT® Portal administrative user. (You only need to do this **once**.)

1. Use your web browser to access the cloud SMRT® Portal instance by entering `http://<public_dns>:8080/smrtportal`
2. Click **Register** at the top right.
3. Create a user named `administrator` (all lowercase). This user is special, as it is the only user that does **not** require activation on creation.
4. Enter the user name `administrator`.
5. Enter an email address. All administrative emails, such as new user registrations, will be sent to this address. **Note**: You won't actually be able to receive email unless you specify an email server that your instance of AMI can see.
6. Enter the password and confirm the password.
7. Click **Register**.



Step 3: Using SSH to Access your Instance

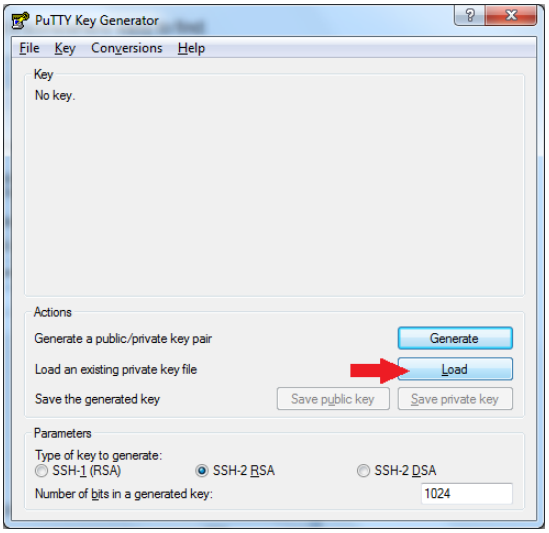
For Mac OS or Linux

1. Ensure that the permissions for the key pair file (with a `.pem` extension) you saved in Step 18 are correct. From a terminal window, enter the following:

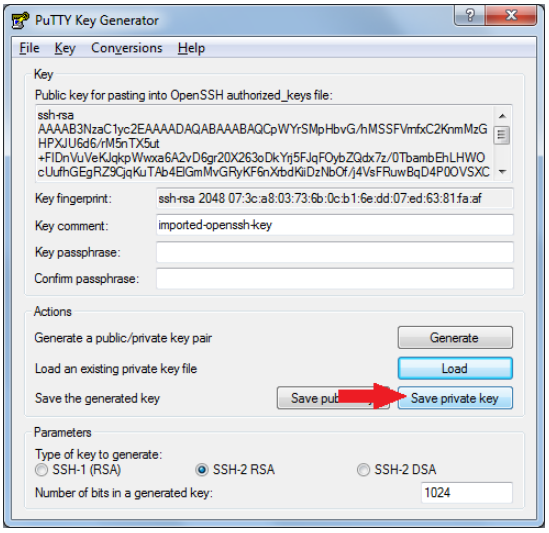
```
$ chmod 400 KEYPAIR.pem
$ ssh -i KEYPAIR.pem ubuntu@{amazon-dns}
```

For Windows

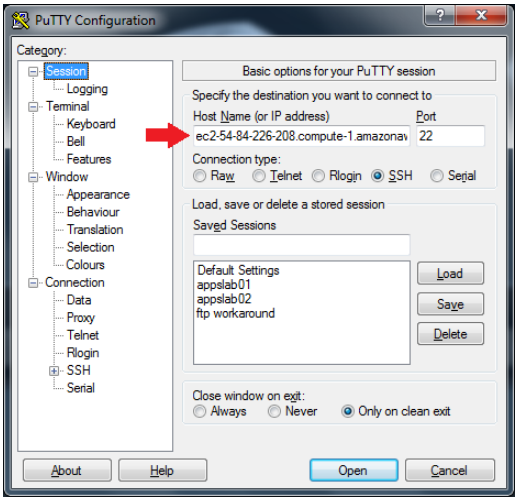
1. Use **PuTTY** and **PuTTYgen** to convert your `.pem` to a `.ppk` file (once per key) then access the AMI. Open **PuTTYgen** and click **Load**. You need to view all files to see and select the `.pem` file.



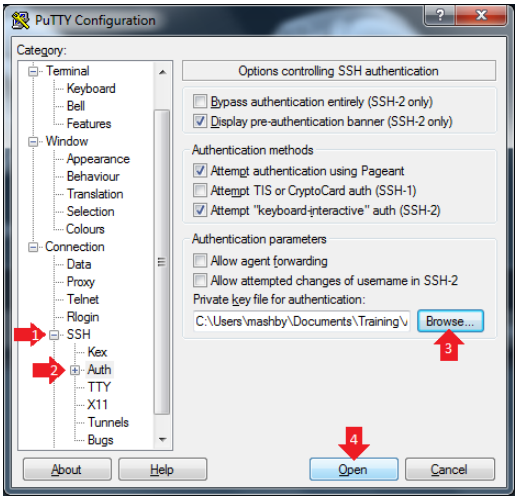
2. Click **Save private key**.



3. Close **PuTTYgen** and open **PuTTY**. Enter the public DNS address of your AMI in the **Host Name** field.



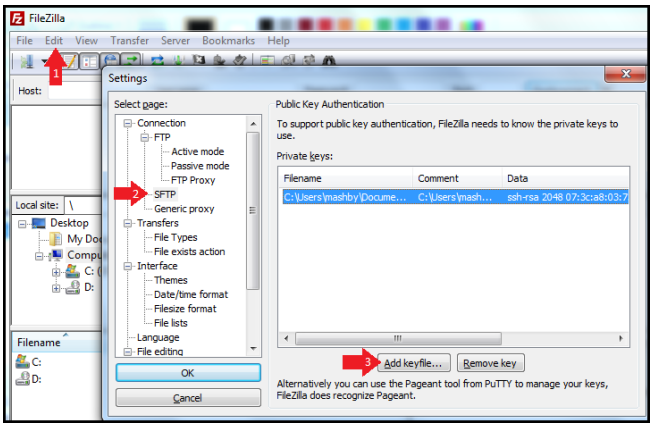
4. In the left menu, expand **SSH** and click **Auth**. Click **Browse** to select the private key .ppk file you just made, then click **Open** to open the connection to your AMI. When prompted, log in as `ubuntu`.



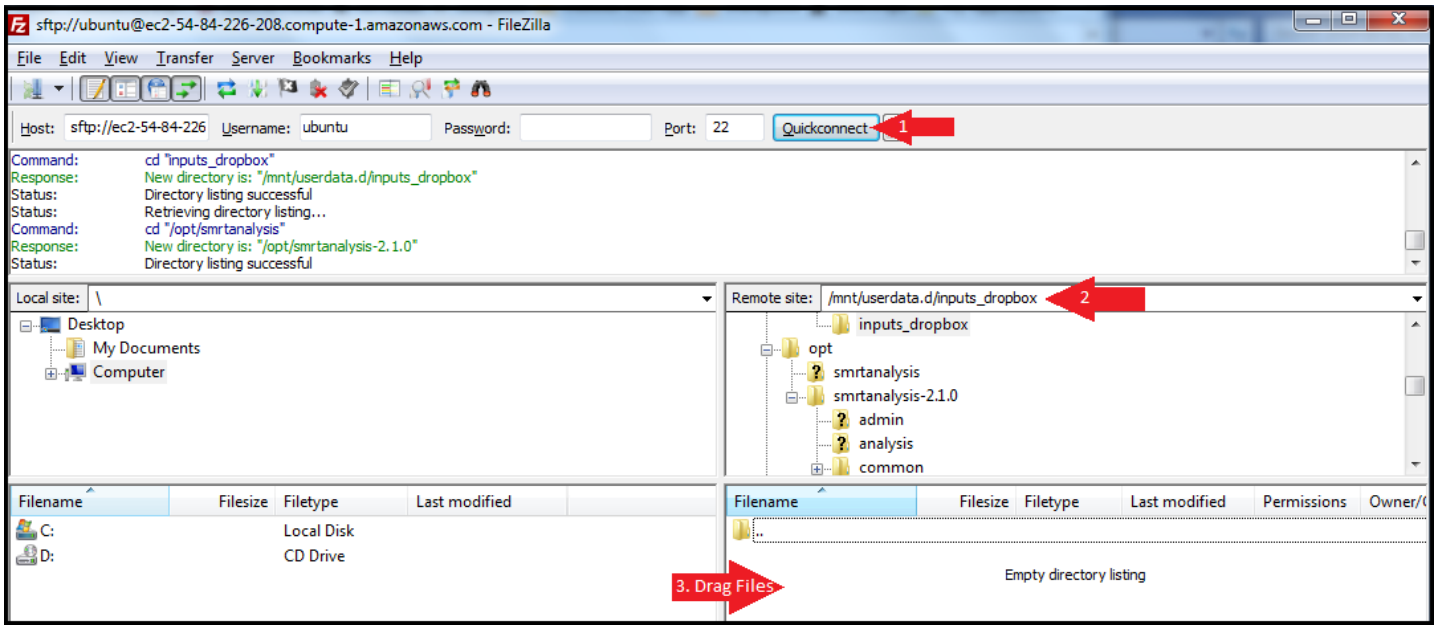
For the next step, you can upload data to SMRT Portal using one of three methods: FileZilla, scp, or mounting an existing S3 bucket.

Step 4a: Uploading Your Data to SMRT® Portal using FileZilla (Windows)

1. Download and install **FileZilla**, a free FTP client. (`http://filezilla-project.org/download.php`)
2. In **FileZilla**, choose **Edit > Settings,** then click **Connection > SFTP**.



3. Click **Add keyfile...** and select **either** the `.ppk` file you generated with PuTTYgen (if you used that for SSH access) **or** the `.pem` file you downloaded in Step 18.
4. If you selected the `.pem` file, you see a dialog box asking for permission to convert the file. Click **Yes** and save the file with a different name, such as `smrtportal_filezilla.ppk`.
5. Click **OK.**
6. Enter the Public DNS address (from Step 21) in the **Host** field.
7. Enter `ubuntu` in the **Username** field.
8. Enter `22` in the **Port** field.
9. Click **Quickconnect** to log on to the remote SMRT Portal instance.
10. In the **Remote Site** box, enter `/opt/smrtanalysis/common/inputs_dropbox`. You see the path name change as FileZilla connects to the folder, as the name is a symlink.
11. In the **Local Site** box, locate the files to transfer and drag them over to the **Remote Site** box.



You can now use SMRT® Portal to scan inputs and run analysis jobs.

Step 4b: Uploading Your Data to SMRT® Portal using scp

Open a terminal session and use `scp` to push, from your local machine, input data into the SMRT® Portal dropbox:

```
scp -r -i customer.pem <path to local data directory> ubuntu@{amazon-dns}:/opt/smrtanalysis/userdata/inputs_dropbox
```

Step 4c: Copying data from an existing S3 bucket to the EC2 instance

The simplest way to share data from an amazon S3 instance is to have the S3 owner make the data public through their S3 Console. Use this command in your AMI to copy the data into your current directory:

```
$ wget http://s3.amazonaws.com/public_S3/folder/file.ext
```

If the data is **private**, you can use Amazon Web Services Command Line Interface (AWS CLI) to copy data from a S3 volume whose owner has shared the security key with you.

To import SMRT Cells you have copied into your server, do the following **within SMRT Portal**:

1. Click **Import and Manage**.
2. Click **Import SMRT Cells**.
3. Click **Add** and enter the path to the data: `/home/username/my_data.`
4. Click **OK**, then click **Scan**.

Step 5: Stopping or Terminating the Amazon Machine Instance

Note: After you are **finished** working with the AMI, please remember to **stop** the instance, as you are charged an hourly rate.

- When the instance is **stopped**, you are **not** charged hourly, but you **are** charged for storage.
- When the instance is **terminated**, you are **not** charged hourly, nor are you charged for storage.

1. Click **Instances**, then right-click the running instance and choose **Instance State > Stop**.
2. Click **Instances**, then right-click a non-running instance and choose **Instance State > Terminate**.

Pricing Information

http://aws.amazon.com/ec2/pricing/

P/N 100-184-200-07