

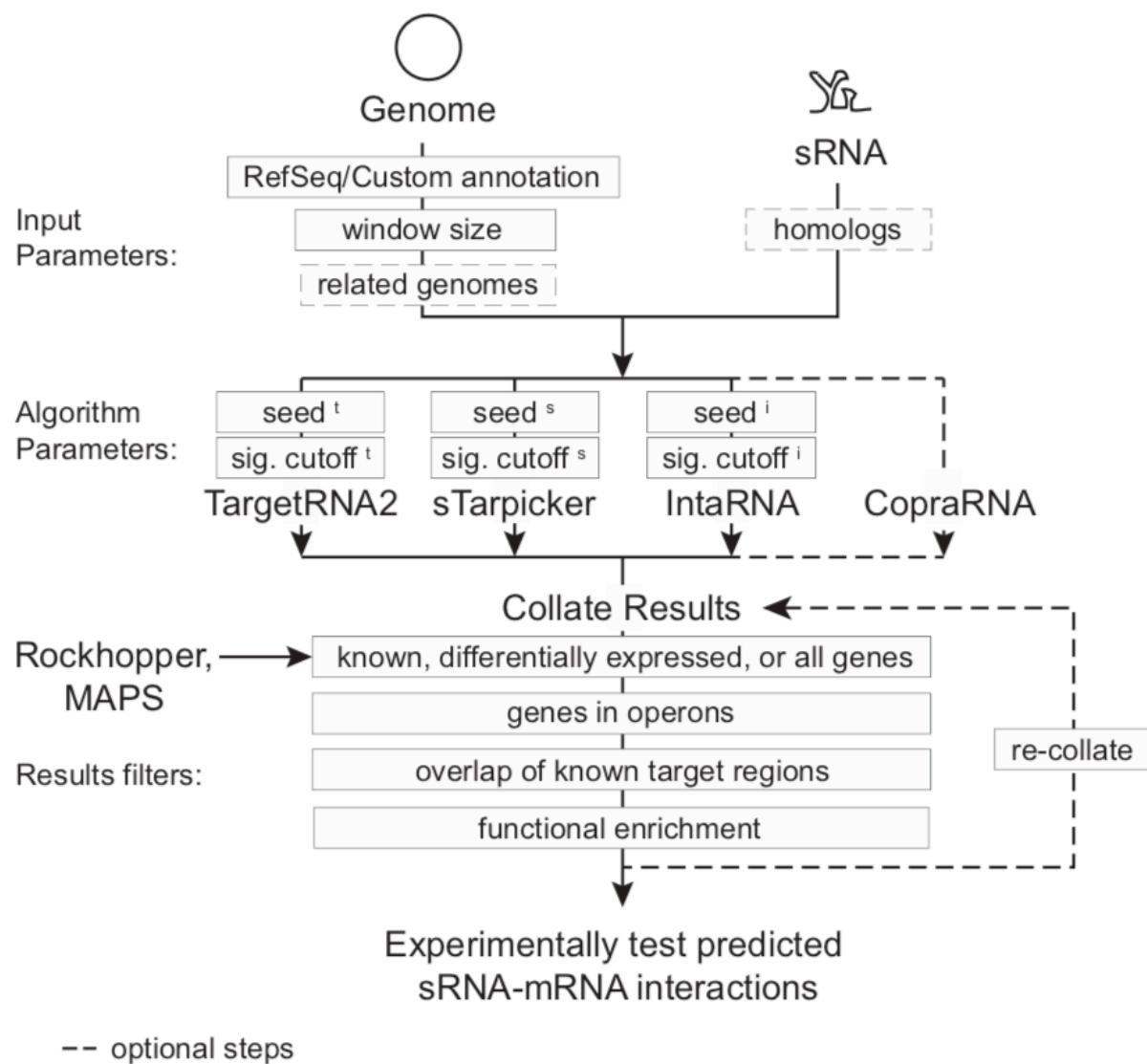


1. Rationale
2. Installing **SPOT**
3. Running **SPOT** – Quick Start
4. Running **SPOT**
5. Data input formats
6. Data output formats
7. Setting up an AWS account
8. Setting up personal AWS interface on your laptop
9. Starting an AWS instance
10. Logging into your AWS instance
11. References

Cite: A.M. King, C.K. Vanderpool, and P.H. Degnan. Systematic evaluation of computational and experimental methods for bacterial small RNA target identification.

1. Rationale

Computational approaches for sRNA target prediction have limitations but are relied upon to generate testable hypotheses for sRNA function. Some algorithms are available online or downloadable (e.g., TargetRNA2, IntaRNA), however these tools frequently yield distinct results, have different data output formats and default search parameters. Therefore, manually compiling results from these disparate tools and integrating the predictions with existing experimental data is not trivial. We have generated an innovative approach to streamline use of multiple existing sRNA target prediction algorithms and integrate predictions with experimental data to generate a unified set of target predictions. To this end, we have developed **SPOT** a flexible software pipeline that searches for sRNA-mRNA binding sites in parallel using separate search tools, collates the predictions, and integrates experimental data using customizable results filters.

Figure 1. Schematic of **SPOT** pipeline analysis (King et al.)

2. Installing SPOT

SPOT is a PERL program that runs TargetRNA2, IntaRNA, StarPicker and CopraRNA in parallel, and collates the results to find consensus sRNA-mRNA targets (Figure 1). Furthermore, additional data types can be utilized to filter the results including expression differences, known binding sites, operon predictions and window size of possible binding sites.

As written the program can run on any Unix/Linux based system, however it has a number of dependencies. To facilitate its use we have set up an [Amazon Web Service \(AWS\)](#) cloud [Amazon Machine Image \(AMI\)](#) with all of the required software installed. Skip to sections 4-7 for setting up your own **SPOT** AMI. However, using the code available [here](#) you can set up and run **SPOT** on a local server.

First, download and install the following software tools and all of their dependencies according the authors' instructions:

- [TargetRNA v2](#)
- [StarPicker](#)
- [IntaRNA v1.0.4](#)
- [CopraRNA v1.2.9](#)

Several modifications were made to the StarPicker and IntaRNA code to accommodate demands of the pipeline. See text notes accompanying the **SPOT** program download for code changes for the relevant programs.

The statistics program R is installed as a requirement for IntaRNA and CopraRNA. As such add the following two packages:

- [RColorBrewer](#)
- [gplots](#)

```
$ sudo R  
> install.packages(c('RColorBrewer', 'gplots'))
```

Some of the output from the **SPOT** program will be written in an `xlsx` format using the Excel Writer PERL module:

- [Excel-Writer-XLSX-0.98](#)

```
$ sudo cpan Excel::Writer::XLSX
```

Most existing Unix/Linux installations should have `sendmail` installed. If not, install the appropriate package

- [sendmail](#)

```
$ sudo apt install sendmail-bin
```

SPOT can work with local copies of genomes and annotations. However, to access genomes from NCBI install the `efetch` program from the Entrez Direct (`edirect`) toolkit.

- [edirect](#)

Retrieve and decompress the **SPOT** directory from GitHub containing core pipeline script and its additional required support PERL scripts.

- [SPOT](#)

Make sure **SPOT** and all of the programs are in your user path. Modify the core pipeline script with the absolute path locations for TargetRNA2, IntaRNA, StarPicker and CopraRNA, and other support PERL scripts.

3. Running SPOT - Quick Start

SPOT is a pipeline script that when run without arguments will print all of the possible program options:

```
$ spot.pl
Usage ./spot.pl
Input parameters:
-r      Fasta file of sRNA query
-a      RefSeq Accession number (assumes any local files have RefSeq
        number as their prefix)
...
```

The minimum data required for a **SPOT** search are:

1. A fasta file of the small RNA sequence
2. A RefSeq genome accession number

```
$ spot.pl -r sgrS.fasta -a NC_000913
```

This will initiate a job using the SgrS as the sRNA query and the *E. coli* str. K12 (NC_000913) as the reference genome. Progress of the search will be printed to the screen. Run time will depend on the number of processors available as each search tool is distributed to a separate sub process. By default CopraRNA **is not run** unless specifically requested.

4. Running SPOT

SPOT has an array of actions that control the input, algorithm parameters, and results filtering.

```
$ spot.pl
Usage ./spot.pl
Input parameters:
  -r    Fasta file of sRNA query
  -a    RefSeq Accession number (assumes any local files have RefSeq
        number as their prefix)
  -o    output file prefix (default = TEST)
  -g    Use local GenBank or PTT&FNA files for all Programs? (default = N
        use latest from GenBank, CopraRNA cannot use local files)
  -n    Other genome RefSeq ids for CopraRNA listed in quotes '' ,
        current max is 5 genomes (default = '')
  -m    Multisequence sRNA file for each genome in CopraRNA list
        (default = '')
  -x    Email address for job completion notification (default = '')

Algorithm parameters:
  -u    Number of nt upstream of start site to search (default = 60)
  -d    Number of nt downstream of start site to search (default = 60)
  -s    seed sizes for I, T, S e.g., '6 7 6' (defaults TargetRNA = 7,
        IntaRNA & Starpicker = 6)
  -c    P/Threshold value Cutoff for T, S, I e.g., '0.5 .001 un'
        (defaults Target = 0.05, Starpicker = 0.5, IntaRNA = top)

Results Filters:
  -b    Number of nt upstream of start site to filter results
        (default = -20)
  -e    Number of nt downstream of start site to filter results
        (default = 20)
Note: -b and -e ignored if using a list (-l) or Rockhopper
      results (-t)
Note: Set -b and -e to -u and -d to get all possible matches in
      results
  -l    List of up and/or down regulated genes, include binding coord if
        known e.g.,
b1101\tdown\n
b3826\tup\tsRNA_start\tsRNA_stop\tmRNA_start\tmRNA_stop\n
        OR
  -t    transcriptome expression file from Rockhopper *_transcripts.txt
  -f    Rockhopper fold change cutoff (default = 1.5)
  -q    Rockhopper q value cutoff (default = 0.01)
  -k    Rockhopper Expression cutoff value (default = 100)
  -p    Operon file from DOOR-2 (http://csbl.bmb.uga.edu/DOOR/index.php)
        (optional)
  -w    Report all genes even if List or Rockhopper provided?
        (default = No)
  -y    Exclude target predictions by only 1 method? (default = Yes)
Note: Does not apply to genes on List or significantly expressed
      from Rockhopper
  -z    Skip sRNA-mRNA detection steps, and just re-analyze data [Y]es
        (default = No) (Run in the same directory & requires original
        results files from each program)
```

Given the time **SPOT** runs can take it is recommended to use a queueing tool on large distributed servers (`qsub`, `slurm`). Alternatively, on the AWS server, laptop or other smaller computers it is recommended to use `screen` to ensure that jobs are not prematurely aborted if the user account is logged out of.

```
$ screen -L spot.pl -r sgrS.fasta -a NC_000913
```

Four test datasets and precomputed output files are included in the folder `example_files`. The following examples correspond to the four provided test datasets.

test01 - Examine entire *E. coli* str. K12 genome for SgrS sRNA target mRNAs. This folder only has the sRNA sequence in a `fasta` file, uses the default settings and retrieves the genome sequence for *E. coli* from GenBank. The final option is to have an email sent to the user after the job has completed.

```
$ cd test01
$ ls
sgrS.fasta
$ spot.pl -r sgrS.fasta -o stringent -a NC_000913 -x username@email.edu

=====Prepping RefSeq Files=====
[Thu Aug 23 23:14:27 UTC 2018]
...
```

test02 - Examine *E. coli* str. K12 genome for SgrS sRNA target mRNA matches among a set of defined differentially expressed genes (`sgrS_diff.txt`). In this case the user has a `fasta` file and a traditional GenBank protein translation file (`PTT`). The user also indicates a larger window size 150 nt upstream of the CDS start position and 100 nt downstream.

```
$ cd test02
$ ls
sgrS.fasta
sgrS_diff.txt
NC_000913.fna
NC_000913.ptt
$ spot.pl -r sgrS.fasta -l sgrS_diff.txt -u 150 -d 100 -o relaxed -a
NC_000913 -g Y
```

Note: PTT files can be easily generated in Excel. Allowing for *customization* of gene annotations and subsequent analyses. A script included with **SPOT** is `fnappt2gbk.pl` which can be used to generate GenBank files using the genome PTT and `fasta` files as inputs. However, always make sure that MAC or DOS line breaks are converted into UNIX line breaks.

Note

test03 - SPOT was designed to allow re-analysis of existing results. This example code block is run in a folder containing the results of **test02**'s search. In this case even though the upstream/downstream region searched was 150nt and 100nt, the reanalysis eliminates any binding sites found outside of 50nt upstream and 30nt downstream.

```
$ cd test03
$ ls
sgrS.fasta
sgrS_diff.txt
NC_000913.fna
NC_000913.ptt
...
$ spot.pl -r sgrS.fasta -l sgrS_diff.txt -a NC_000913 -o changed_50_30 -b 50
-e 30 -z Y
```

test04 – SPOT can also be run using a *transcript.txt file generated by the RNAseq analysis program Rockhopper directly (instead of list as in example **test02**). In this example default expression cutoffs are used, however these can be specified by the user. In addition, when provided a set of sRNA homologs and target genomes CopraRNA can be run. In these instances only genomes in RefSeq can be used. Custom genome annotations cannot be utilized.

```
$ cd test04
$ ls
NC_000913_SgrS_transcripts.txt
sgrS.fasta
sgrS_homologs.fasta
$ spot.pl -r sgrS.fasta -t NC_000913_SgrS_transcripts.txt -o express -a
NC_000913 -m sgrS_homologs.fasta -n 'NC_002695 NC_011740'
```

When the jobs have completed compare your results to the files in the corresponding _results folder.

5. Data input formats

sRNA fasta file – DNA sequence of sRNA in a standard fasta file. File extension does not matter (.fasta, .fa, .fna, .frn, .ffn)

RefSeq ID – Standard RefSeq IDs can be used and GenBank files (.gbff) will be retrieved using efetch. Program will retrieve additional replicons (e.g., plasmids) or scaffolds associated with the provided RefSeq IDs, however, the search will only be carried out on the file with a name corresponding to the input RefSeq ID. By default the .gbff is renamed to a .gb file, and .fna and .ptt files are generated.

Local Files – Different combinations of local files can be used. They all **must** have the same prefix and end in the following suffixes:

.fna	Genome fasta sequence
.ptt	Protein translation table – gene annotation
.gb or .gbk	Genbank file

Files **without** these suffixes will be ignored. All must have unix linebreaks and the .ptt file must be tab separated. Allowed input combinations:

	.fna	.ptt	.gb or .gbk	Status	Action
1.	✓	✓	✓	okay	Start run
2.	✓	✓		okay	Make .gb file, start run
3.	✓		✓	okay	Make .ptt file, start run
4.		✓	✓	okay	Make .fna file, start run
5.			✓	okay	Make .fna and .ptt file, start run
6.	✓			bad	Abort run
7.		✓		bad	Abort run

.ptt Files – This is a legacy GenBank annotation format. However, the StarPicker algorithm used here requires this format. This format is very easy to generate in Excel and can allow users of **SPOT** to customize their annotations. See example:

Escherichia coli str. K-12 substr. MG1655, complete genome - 1..4641652								
4141 proteins								
Location	Strand	Length	PID	Gene	Synonym	Code	COG	Product
190..255	+	21		thrL	b0001	-	-	thr operon leader peptide
337..2799	+	820		thrA	b0002	-	-	Bifunctional aspartokinase
2801..3733	+	310		thrB	b0003	-	-	homoserine kinase
3734..5020	+	428		thrC	b0004	-	-	L-threonine synthase

Differentially expressed genes – Lists of differential genes can be formatted as tab separated files one of two ways. **Do not include the header line**

Simple:

Locus	Expression
b1101	down
b3826	up

With known binding sites:

Locus	Expression	sRNA_start	sRNA_stop	mRNA_start	mRNA_stop
b1101	down	168	187	-30	-9
b3826	up	168	187	-96	-76

Rockhopper *transcript.txt files – **SPOT** can read default output files of Rockhopper from simple pairwise RNAseq experiments. Files generated with the verbose output option in Rockhopper cannot be read. Files should have 12 columns including the normalized expression values for the treatment and control, the qValues and the estimated fold-change.

sRNA Multisequence Fasta Files – If running CopraRNA sRNA files must conform to expectations of the CopraRNA program:

1. RNA sequence must have Us instead of Ts
2. The sequence names must correspond to the individual genome RefSeq IDs
3. Must include the focal genome sRNA sequence as well

6. Data output formats

Data from each individual algorithm is preserved in the output folder for manual investigation.

SPOT generates several output files for further analysis:

XLSX file – Primary file containing consensus table of sRNA-mRNA predictions from the 3 or 4 tools used in the run. File name prefix corresponds to run output prefix that was assigned (-o , default= TEST). Sheet 1 (complete.txt) shows the aligned predictions, p values, and coordinates for the predicted interaction for each gene. Sheet 2 (summary.txt) has the counts predicted by each gene, and a summary letter and ranking based location and on the number of algorithms that found the same prediction.

- A** Prediction overlaps a known binding site
B→E Predictions that are not coincident with a known binding site when was provided for that gene. Shared letters overlap the same site.
F→I Predictions when no known binding site was provided. Shared letters overlap the same site.

summary.pdf file – This file has a R generated plot that corresponds to Sheet 2 (summary.txt) which can be imported to Illustrator.

COLLATED_RESULTS folder – This folder contains plots generated based on IntaRNA tools showing the localization of binding sites of the mRNA and sRNA as *pdf, *png and *ps files. In addition, if CopraRNA is run, functional heatmap data are included here as well.

7. Setup an AWS account

Navigate to the new account setup page:

<https://portal.aws.amazon.com/billing/signup#/start>

AWS Accounts Include
12 Months of Free Tier Access
Including use of Amazon EC2, Amazon S3, and Amazon DynamoDB
Visit [aws.amazon.com/free](#) for full offer terms

Create an AWS account

Email address

Password

Confirm password

AWS account name

Continue

[Sign in to an existing AWS account](#)

© 2018 Amazon Web Services, Inc. or its affiliates.
All rights reserved.
[Privacy Policy](#) | [Terms of Use](#)

For now, set up your home region as “**U.S. East (W. Virginia)**” later you can switch this as necessary.

Unfortunately, when setting up an account you will need a credit card number

Input Education credit - Depending on your application it may be possible to apply for education credits to defray the cost of the AWS server time:

<https://aws.amazon.com/education/awseducate/>



8. Setup personal AWS interface on your laptop

People with MACs:

✓ Terminal will already be installed /Applications/Utilities

✗ Download & Install **XQuartz** if not already installed

<http://xquartz.macosforge.org/landing/>

duck Download & Install **Cyberduck** <https://cyberduck.io/?l=en>

People with PCs:

✗ Download & Install **PuTTY**

<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>

✗ Download & Install **xMing**

http://sourceforge.net/project/downloading.php?group_id=156984&filename=Xming-6-9-0-31-setup.exe

How to setup **xMing** : http://www.geo.mtu.edu/geoschem/docs/putty_install.html

 Download & Install **WinSCP** <http://winscp.net/eng/download.php>
or

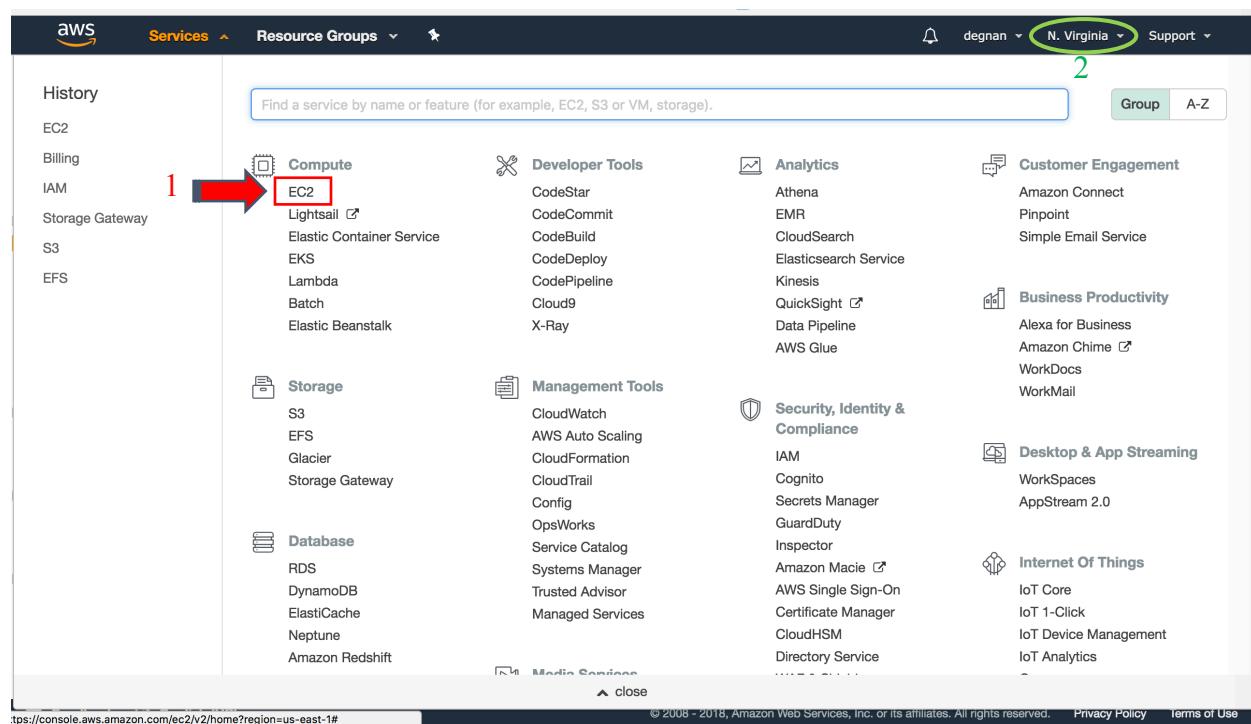
 Download & Install **Cyberduck** <https://cyberduck.io/?l=en>

9. Starting an AWS instance

For in-depth instructions regarding starting an AWS instance please see:

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/launching-instance.html>

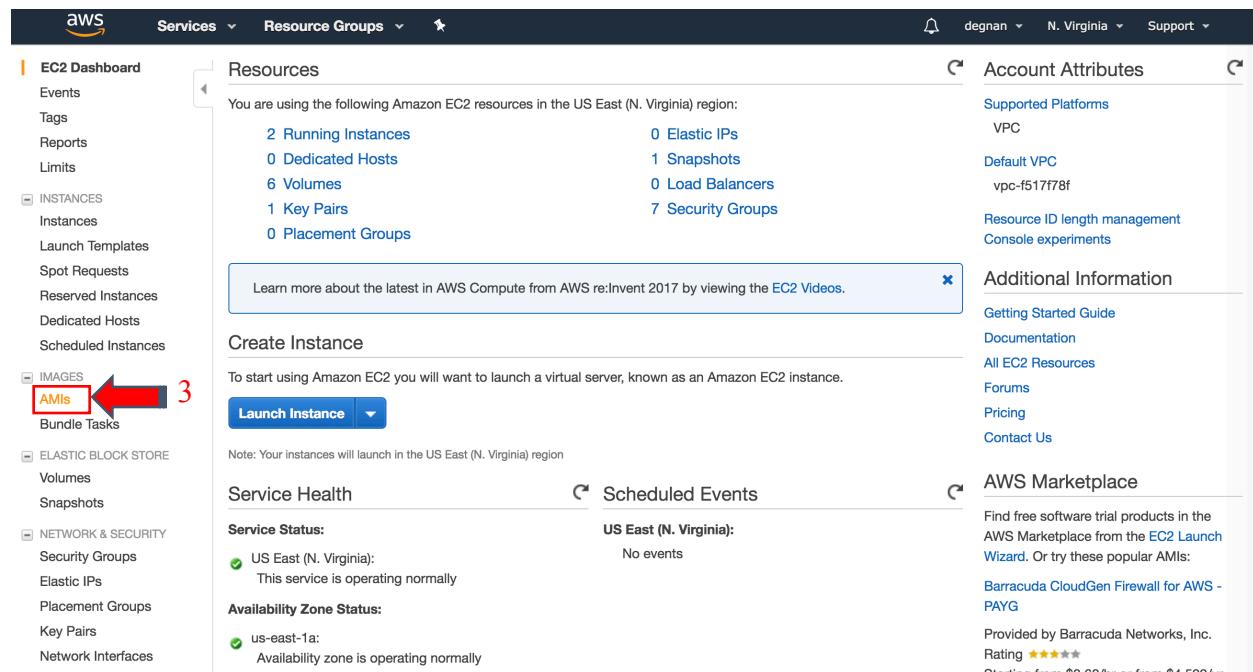
1. After making and logging into your AWS account find your way to the EC2 (Elastic Computing Cloud) page. You can find it under “Services” menu on the upper left-hand corner of the page:



<https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#Home>:

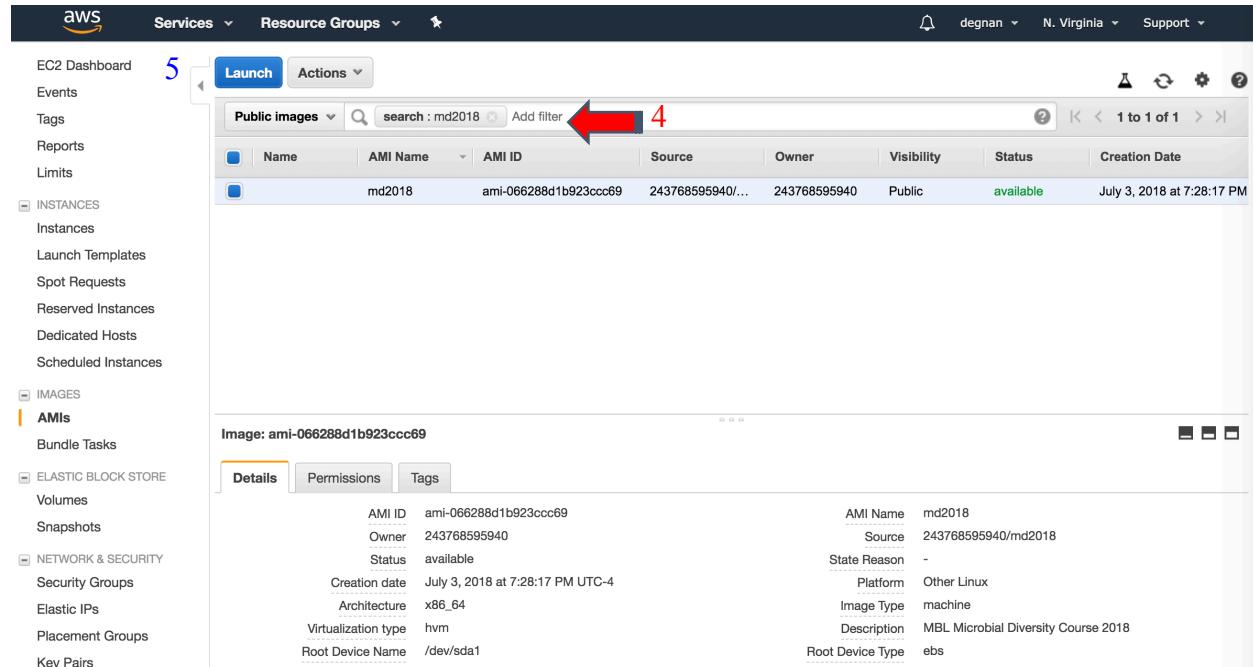
2. Make sure your home region as “**U.S. East (W. Virginia)**”. Your region is indicated in the upper right-hand corner of the page (circled above)

3. On the right-hand side bar under “**IMAGES**” select “**AMIs**”



The screenshot shows the AWS EC2 Dashboard. On the left sidebar, under the 'IMAGES' section, the 'AMIs' option is highlighted with a red box and a red arrow pointing to it. The main content area displays EC2 resource statistics: 2 Running Instances, 0 Dedicated Hosts, 6 Volumes, 1 Key Pairs, 0 Placement Groups, 0 Elastic IPs, 1 Snapshots, 0 Load Balancers, and 7 Security Groups. Below this, there's a 'Create Instance' section with a 'Launch Instance' button. To the right, there's an 'Account Attributes' panel and an 'Additional Information' panel.

4. In the search bar switch from “Owned by me” to “Public images” and search for “SPOTv1”

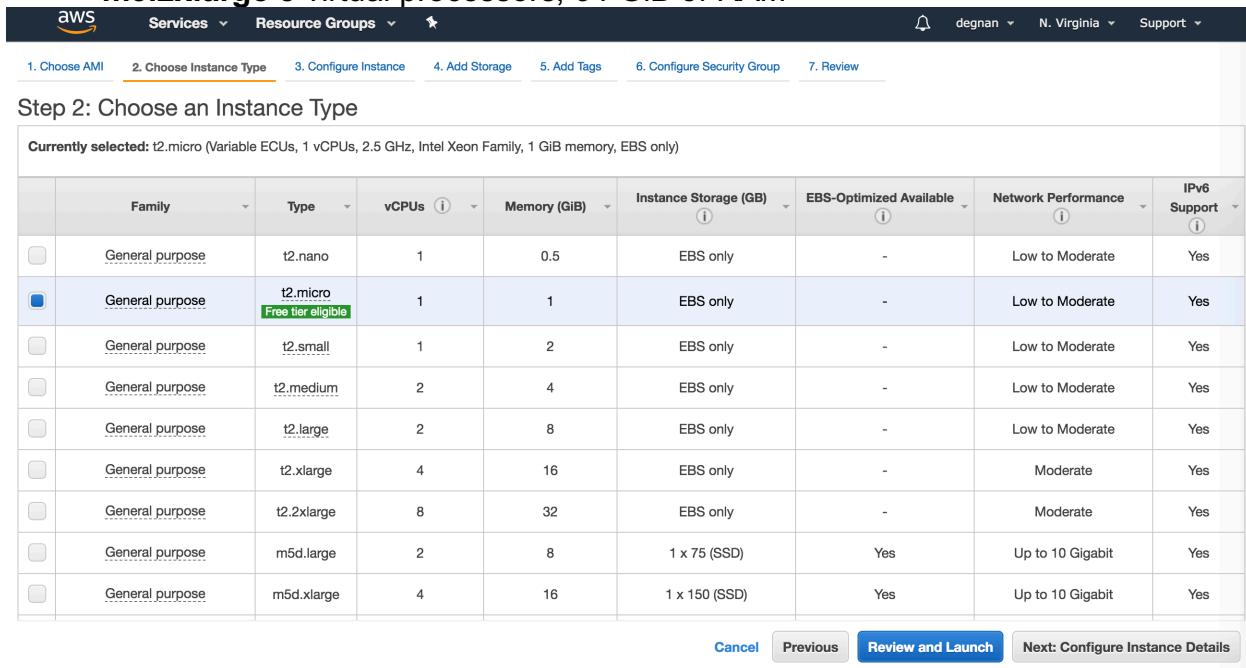


The screenshot shows the AWS AMI search results. The search bar at the top is highlighted with a red box and a red arrow pointing to it. The search term 'md2018' is entered. The results table shows one item: 'Image: ami-066288d1b923ccc69'. The table has columns for Name, AMI Name, AMI ID, Source, Owner, Visibility, Status, and Creation Date. The details for this AMI are shown below the table, including its ID, owner, creation date, and various specifications like Architecture (x86_64), Virtualization type (hvm), and Root Device Name (/dev/sda1).

5. Select the blue “Launch” button

6. Now you are on AWS “Step 2: Choose and Instance Type” – Select your computer:

t2.micro is the only free option, however it is maxed out at 1GiB of RAM, 1 processor and 30GiB of storage. Very slow
m5.2xlarge 8 virtual processors, 64 GiB of RAM

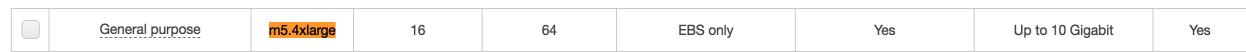


Step 2: Choose an Instance Type

Currently selected: t2.micro (Variable ECUs, 1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB memory, EBS only)

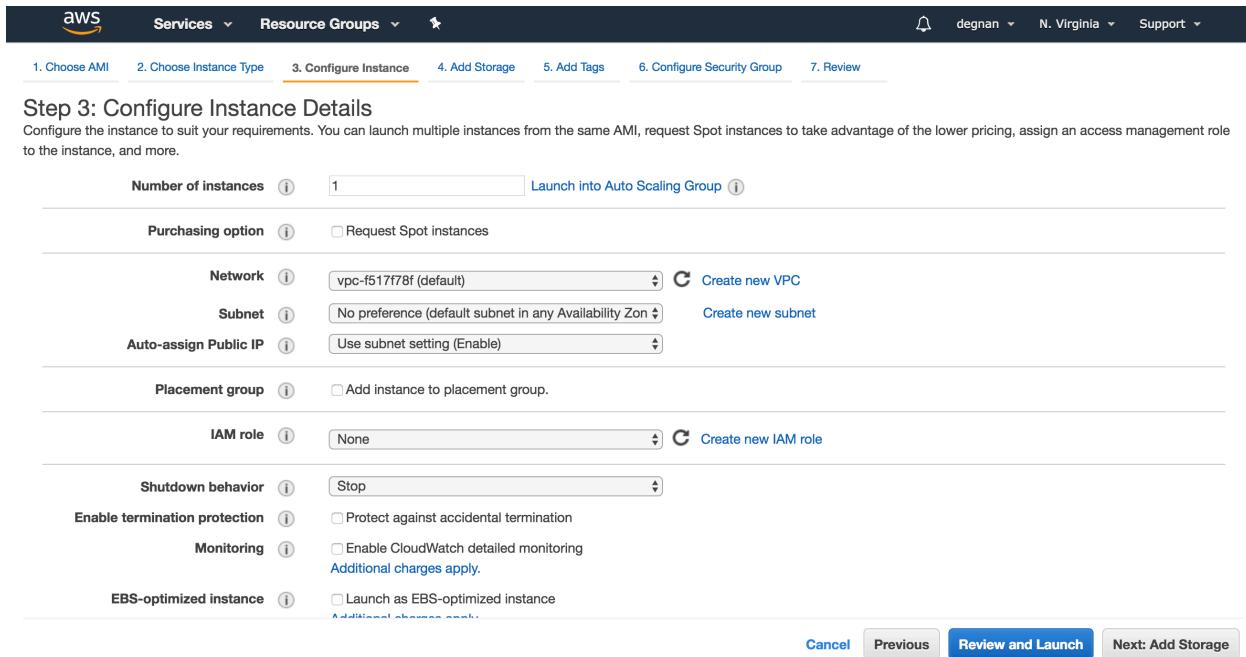
	Family	Type	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance	IPv6 Support
<input type="checkbox"/>	General purpose	t2.nano	1	0.5	EBS only	-	Low to Moderate	Yes
<input checked="" type="checkbox"/>	General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.small	1	2	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.medium	2	4	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.large	2	8	EBS only	-	Low to Moderate	Yes
<input type="checkbox"/>	General purpose	t2.xlarge	4	16	EBS only	-	Moderate	Yes
<input type="checkbox"/>	General purpose	t2.2xlarge	8	32	EBS only	-	Moderate	Yes
<input type="checkbox"/>	General purpose	m5d.large	2	8	1 x 75 (SSD)	Yes	Up to 10 Gigabit	Yes
<input type="checkbox"/>	General purpose	m5d.xlarge	4	16	1 x 150 (SSD)	Yes	Up to 10 Gigabit	Yes

Cancel Previous Review and Launch Next: Configure Instance Details



7. Select “Next: Configure Instance Details” button on bottom-right

8. On “Step 3: Configure Instance Details” page – *leave defaults as-is*



Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of instances	1	Launch into Auto Scaling Group
Purchasing option	<input type="checkbox"/> Request Spot instances	
Network	vpc-f517f78f (default)	<input type="button"/> Create new VPC
Subnet	No preference (default subnet in any Availability Zone)	<input type="button"/> Create new subnet
Auto-assign Public IP	Use subnet setting (Enable)	
Placement group	<input type="checkbox"/> Add instance to placement group.	
IAM role	None	<input type="button"/> Create new IAM role
Shutdown behavior	Stop	
Enable termination protection	<input type="checkbox"/> Protect against accidental termination	
Monitoring	<input type="checkbox"/> Enable CloudWatch detailed monitoring <small>Additional charges apply.</small>	
EBS-optimized instance	<input type="checkbox"/> Launch as EBS-optimized instance <small>Additional charges apply.</small>	

Cancel Previous Review and Launch Next: Add Storage

9. Select the “Next: Add Storage” button on bottom-right, to move to the next step

10. On the “Step 4: Add Storage” adjust local disk size to 30 GiB

Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more](#) about storage options in Amazon EC2.

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encrypted
Root	/dev/sda1	snap-0ea8cfee9142df0b9	25	General Purpose SSD (GP2)	100 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted

Add New Volume

Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage. [Learn more](#) about free usage tier eligibility and usage restrictions.

Cancel Previous Review and Launch Next: Add Tags

11. Select the “Next: Add Tags” button on bottom-right, to move to the next step

12. On the “Step 5: Add Tags” *optionally* hit the “Add Tag” button OR skip to step 14

Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver. A copy of a tag can be applied to volumes, instances or both. Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

Key	(127 characters maximum)	Value	(255 characters maximum)	Instances	Volumes
<i>This resource currently has no tags</i>					

Choose the Add tag button or [click to add a Name tag](#). Make sure your [IAM policy](#) includes permissions to create tags.

Add Tag (Up to 50 tags maximum)

13. For example Add a *key* = “Name” and *value* = “my-SPOT”

Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver.

A copy of a tag can be applied to volumes, instances or both.

Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

Key	(127 characters maximum)	Value	(255 characters maximum)	Instances	Volumes
Name		my-md2018		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Add another tag (Up to 50 tags maximum)

Cancel Previous Review and Launch Next: Configure Security Group

14. Select the “Next: Configure Security Group” button on bottom-right, to move to the next step

15. On “Step 6: Configure Security Group” page – **leave defaults as-is**

Step 6: Configure Security Group

Security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below.

[Learn more](#) about Amazon EC2 security groups.

Assign a security group: Create a new security group Select an existing security group

Security group name: launch-wizard-6

Description: launch-wizard-6 created 2018-07-10T10:04:47.063-04:00

Type	Protocol	Port Range	Source	Description
SSH	TCP	22	Custom 0.0.0.0/0	e.g. SSH for Admin Desktop

Add Rule

Warning
Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

Cancel Previous Review and Launch Next: Launch Instance

Note: For now, we will ignore the *Warning*. In the future consider making your instances harder to access by non-users in your lab/group

16. Select the “Review and Launch” button on bottom-right, to move to the next step

Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

AMI Details

Instance Type

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
m2.4xlarge	26	8	68.4	2 x 840	Yes	High

Launch

17. You can inspect the settings before hitting the “**Launch**” button. As before ignore warnings.

18. Now it asks you to select or create a key pair.

Select an existing key pair or create a new key pair

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about [removing existing key pairs from a public AMI](#).

Proceed without a key pair

I acknowledge that I will not be able to connect to this instance unless I already know the password built into this AMI.

Launch Instances

19. You will need to download the key and save it to a private location on your computer.

Your instances are now launching
The following instance launches have been initiated: i-0f99c725900b146ce View launch log

Get notified of estimated charges
Create billing alerts to get an email notification when estimated charges on your AWS bill exceed an amount you define (for example, if you exceed the free usage tier).

How to connect to your instances

Your instances are launching, and it may take a few minutes until they are in the **running** state, when they will be ready for you to use. Usage hours on your new instances will start immediately and continue to accrue until you stop or terminate your instances.

Click [View Instances](#) to monitor your instances' status. Once your instances are in the **running** state, you can [connect](#) to them from the Instances screen. [Find out](#) how to connect to your instances.

Here are some helpful resources to get you started

- [How to connect to your Linux instance](#)
- [Amazon EC2: User Guide](#)
- [Learn about AWS Free Usage Tier](#)
- [Amazon EC2: Discussion Forum](#)

While your instances are launching you can also

[Create status check alarms](#) to be notified when these instances fail status checks. (Additional charges may apply)

[Create and attach additional EBS volumes](#) (Additional charges may apply)

[Manage security groups](#)

20. From here you can navigate using the left-hand side bar to your “Instances”

EC2 Dashboard

Instances

- Instances
- Launch Templates
- Spot Requests
- Reserved Instances
- Dedicated Hosts
- Scheduled Instances

IMAGES

- AMIs
- Bundle Tasks

ELASTIC BLOCK STORE

- Volumes
- Snapshots

NETWORK & SECURITY

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs

Launch Instance | Connect | Actions

Filter by tags and attributes or search by keyword

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)
i-0c7adbc18dc62c77c	m4.xlarge	us-east-1d	running	2/2 checks ...	None	ec2-54-209-226-246.co..	
md2018	i-0ecfd56d143aa83ac	t2.micro	us-east-1c	running	2/2 checks ...	None	ec2-52-207-221-236.co..

Instance: i-0ecfd56d143aa83ac (md2018) Public DNS: ec2-52-207-221-236.compute-1.amazonaws.com

Description | Status Checks | Monitoring | Tags

Instance ID	i-0ecfd56d143aa83ac	Public DNS (IPv4)	ec2-52-207-221-236.compute-1.amazonaws.com
Instance state	running	IPv4 Public IP	52.207.221.236
Instance type	t2.micro	IPv6 IPs	-
Elastic IPs		Private DNS	ip-172-31-89-32.ec2.internal
Availability zone	us-east-1c	Private IPs	172.31.89.32
Security groups	launch-wizard-2. view inbound rules. view outbound rules	Secondary private IPs	
Scheduled events	No scheduled events	VPC ID	vpc-f517f78f
AMI ID	ubuntu/images/hvm-ssd/ubuntu-xenial-16.04-amd64-server-20180522 (ami-a4dc46db)	Subnet ID	subnet-6394314d
Platform	-	Network interfaces	eth0
IAM role	-	Source/dest. check	True

21. Instance state will be “*Initializing*” until the computer has “booted” up.

22. Once the Instance state switches to “*running*” and you select the instance, details of the instance will be shown below.

23. Find and copy the “Public DNS (IPv4)” address for your instance. You will use this to login to your server.

Public DNS (IPv4):	
-----------------------	--

10. Logging into you AWS instance

To log into the server you will need your:

1. Private ssh key [yourid_key.pem](#)
2. [username](#) = first name and last initial as one word (e.g., Jane Doe = janed)
3. [XX-XX-XX-XX](#) = Your specific Public DNS (IPv4) from above

Login using Terminal on a **MAC or UNIX**.

```
$ ssh -Y -i ~/.ssh/yourid_key.pem username@XX-XX-XX-XX
```

Login from **Windows** using PuTTY

- a. Open PuTTY
- b. Under Category, click on SSH > Auth
- c. Click browse
- d. Find your private key ([yourid_key.pem](#)) and select it
- e. Under Category, click Session and input address of your EC2 instance ([XX-XX-XX-XX](#)) in the "host name" box
- f. Type "md2018" in the box under saved sessions and click save.
- g. Double-click on the "SPOT" that appears under saved sessions.
- h. Log in with your username. Your key should be used automatically.
- i. For future logins, just double-click the "SPOT" saved session.

Once entered you will find yourself on the command line interface:

```
delta7:Desktop degnan$ ssh -Y -i ~/.ssh/MyKeyPair.pem ubuntu@18.206.100.156
Warning: No xauth data; using fake authentication data for X11 forwarding.
Welcome to Ubuntu 16.04.5 LTS (GNU/Linux 4.4.0-1065-aws x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

 Get cloud support with Ubuntu Advantage Cloud Guest:
   http://www.ubuntu.com/business/services/cloud

2 packages can be updated.
0 updates are security updates.

New release '18.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Thu Aug 23 21:50:22 2018 from 138.23.161.215
ubuntu@ip-172-31-38-185:~$ █
```

11. References

- Busch A, Richter AS, Backofen R. 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 24:2849-2856.
- Kery MB, Feldman M, Livny J, Tjaden B. 2014. TargetRNA2: identifying targets of small regulatory RNAs in Bacteria. Nucleic Acids Research 42:W124-129.
- King AM, Vanderpool CK, and Degnan PH. Systematic evaluation of computational and experimental methods for bacterial small RNA target identification.
- McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, Vanderpool CK, Tjaden B. 2013. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Research 41:e140.
- Wright PR, Georg J, Martin M, Sorescu DA, Richter, AS, Lott S, Kleinkauf R, Hess WR, Backofen R. 2014. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Research 42:W119-W123.
- Ying X, Cao Y, Wu J, Liu Q, Cha L, Li W. 2011. sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. PloS One 6:e22705.