

- 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 you AWS instance
- 11. References

Cite: A.M. King, C.K. Vanderpool, and P.H. Degnan. sRNA-target Prediction Organizing Tool (SPOT) integrates computational and experimental data to facilitate functional characterization of bacterial small RNAs

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 <u>innovative approach to streamline use of multiple existing sRNA target prediction</u> <u>algorithms and integrate predictions with experimental data to generate a unified set of</u> <u>target predictions</u>. 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 <u>Amazon Web Service</u> (<u>AWS</u>) cloud <u>Amazon Machine Image (AMI</u>) with all of the required software installed. Skip to sections 4-7 for setting up your own **SPOT** AMI. However, using the code available <u>here</u> 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
- <u>StarPicker</u>
- IntaRNA v1.0.4
- CopraRNA v1.2.9

Several modifications were made to the StarPicker and IntaRNA code to accommodate demands of the pipeline.

Replace the following programs with those provided in the GitHub link. Modifications in the code are marked with ## comments and/or initials (PHD). Descriptions of edits made are listed briefly below.

StarPicker:

sTarPicker_global2.pl: changes made to input of command line arguments

IntaRNA v1.0.4:

add_GI_genename_annotation.pl : distinguish GeneIDs vs GI Nos
get_refseq_from_ftp.pl : Replacement code for get_refseq_from_ftp.sh
IntaRNA_wrapper.pl : Option added to use local GenBank files, use
 get_refseq_from_ftp.pl

rerun_enrichment.pl: code snippet re-running enrichment analysis from IntaRNA_Wrapper.pl

termClusterReport.pl: code modified to handle GeneIDs vs GI Nos

CopraRNA v1.2.9:

Note: D3 Javascript libraries may or may not be accessible using existing framework to generate functional enrichment heatmaps (http://d3js.org/d3.v3.min.js). If problems are encountered, it is possible to edit the master html files in IntaRNA and CopraRNA to use a local version of d3.v3.min.js.

Be sure all programs are added to the user path and all path references in StarPicker, IntaRNA, and CopraRNA match your system installation. The statistics program R is installed as a requirement for IntaRNA and CopraRNA. As such add the following two packages:

- <u>RColorBrewer</u>
- 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

• <u>sendmail</u>

\$ 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.

• <u>edirect</u>

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

• <u>SPOT</u>

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:

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:
           Fasta file of sRNA query
      -r
           RefSeq Accession number (assumes any local files have RefSeq
      -a
           number as their prefix)
            output file prefix (default = TEST)
      -0
           Use local GenBank or PTT&FNA files for all Programs? (default = N
      -g
           use latest from GenBank, CopraRNA cannot use local files)
           Other genome RefSeq ids for CopraRNA listed in quotes '',
      -n
           current max is 5 genomes (default ='')
           Multisequence sRNA file for each genome in CopraRNA list
      -m
             (default ='')
           Email address for job completion notification (default ='')
      -x
Algorithm parameters:
           Number of nt upstream of start site to search (default = 60)
      –u
      -d
           Number of nt downstream of start site to search (default = 60)
            seed sizes for I, T, S e.g., '6 7 6' (defaults TargetRNA = 7,
      -s
            IntaRNA & Starpicker = 6)
           P/Threshold value Cutoff for T, S, I e.g., '0.5 .001 un'
      -c
      (defaults Target = 0.05, Starpicker = 0.5, IntaRNA = top)
Results Filters:
           Number of nt upstream of start site to filter results
      -b
             (default = -20)
           Number of nt downstream of start site to filter results
      -е
            (default = 20)
      Note: -b and -e ignored if using a list (-1) or Rockhopper
            results (-t)
      Note: Set -b and -e to -u and -d to get all possible matches in
            results
           List of up and/or down regulated genes, include binding coord if
      -1
            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)
           Rockhopper q value cutoff (default = 0.01)
      -q
      -k
           Rockhopper Expression cutoff value (default = 100)
           Operon file from DOOR-2 (http://csbl.bmb.uga.edu/DOOR/index.php)
      -p
             (optional)
           Report all genes even if List or Rockhopper provided?
      -w
            (default = No)
            Exclude target predictions by only 1 method? (default = Yes)
      -у
           Note: Does not apply to genes on List or significantly expressed
            from Rockhopper
            Skip sRNA-mRNA detection steps, and just re-analyze data [Yy]es
      -z
            (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 individual program default SEED size and significance 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.

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 to search for binding sites.

```
$ cd test02
$ ls
sgrS.fasta
sgrS_diff.txt
NC_000913.fna
NC_000913.ptt
$ spot.pl -r sgrS.fasta -1 sgrS_diff.txt -u 150 -d 100 -c '0.5 0.001 un' -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 fnaptt2gbk.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.

test03 - SPOT was designed to allow re-analysis of existing results. This example code block is rung 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. This search also does not use the list of differentially expressed genes.

```
$ cd test03
$ ls
sgrS.fasta
sgrS_diff.txt
NC_000913.fna
NC_000913.ptt
...
$ spot.pl -r sgrS.fasta -u 150 -d 100 -c '0.5 0.001 un' -o changed_50_30 -a
NC_000913 -g Y -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' -u 150 -d 100
```

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

5. Data input formats

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

<u>RefSeq ID</u> – 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.

<u>Local Files</u> – 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 include:

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

<u>.ptt Files</u> – 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							
4141 proteins								
Location	Strand	Length	PID	Gene	Synonym	Code	COG	Product
190255	+	21		thrL	b0001	-	-	thr operon leader peptide
3372799	+	820		thrA	b0002	-	-	Bifunctional aspartokinase
28013733	+	310		thrB	b0003	-	-	homoserine kinase
37345020	+	428		thrC	b0004	-	-	L-threonine synthase

Note: As indicated above, customization of PTT files allows users to correct or change annotations based on new data. Furthermore, by modifying PTT files **RNAs can be included** in the annotation. First, this allows for sRNA – RNA interactions to be identified. Second, this approach was used in the manuscript to perform a 'reverse' search. For a 'reverse' search the PTT file is edited to ONLY include the known sRNAs. Then, the user supplies the UTR or putative sRNA binding region to **SPOT** as a fasta file if it were the sRNA. 'Reverse' searches cannot use CopraRNA and as sRNAs do not have GI numbers and may not have GeneIDs - no functional enrichment plots will be produced. This may result in several warnings when the **SPOT** is run, however it should not influence the final composite predictions. <u>Differentially expressed genes</u> – Lists of differential genes can be formatted as tab separated files one of two ways. **Do not include a 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	ир	168	187	-96	-76

<u>Rockhopper *transcript.txt files</u> – **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 *q* Values and the estimated fold-change.

<u>sRNA Multisequence Fasta Files</u> – 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.

TargetRNA2_*txt = TargetRNA2 Primary report *.output = Starpicker Primary report intarna_websrv_table_truncated.csv = IntaRNA Primary report *_hIntaRNA.csv = CopraRNA Primary report

SPOT generates several output files for further analysis:

<u>XLSX file</u> – 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.

	A	в	с	D	E	F	G	н		1	к	L	м	Ν	0	
1	Locus	Gene	T-	T- Pvalue	T- sRNA_ start	T- sRNA_ stop	T- mRNA_ start	T- mRNA_ stop	T-Structure	S-	S- Pvalue	S- sRNA_ start	S- sRNA_ stop	S- mRNA_ start	S- mRNA_ stop	S-Structure
2	b1101	ptsG	-10.51	0.018	167	187	-29	.9	<pre>Ec_sgrS 187 3' U-GUGGUUAUGAGUCAGUGUGU 5' 167</pre>	-20.8	0.004	169	187	-28	-8	sRNA(Ec_sgr5) 169uguga . Target(ptsG_b1101) -8 ucacgag
3	b0060	polB	-9.86	0.025	151	169	-2	17	Ec_sgrS 169 3' UGUACUACGUCCGUUCAGU 5' 151 ; ; ; ; ; polB -2 5' occuugGCGCAGGCAGGUU 3' 17 gene product:DNA polymerase II							

- 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 one 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.

	A	В	С	D	E	F	G	н	1	J	к	L	м	Ν
			T			C	T	64-r		Count	T			
			larget			Count	TargetRNA2_	Star	IntakNA	count	TargetKNA		INTAKNA_	
1	Locus	Gene	RNA2	Star	IntaRNA	_ALL	20_20	20_20	20_20	20_20	2_M	Star_M	м	Rank
2	b1101	ptsG	1	1	1	3	1	1	1	3	F	F	F	4
3	b0060	polB	1		1	2	1		1	2	F		F	4

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



<u>COLLATED RESULTS folder</u> – 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, a functional enrichment heatmap is included as a *pdf file similar to those individually provided by IntaRNA and CopraRNA - however it represents the collated results.



Predicted interaction regions sRNA

Example result files for the sRNA SgrS and corresponding test datasets are available with the **SPOT** software distribution.

7. Setup an AWS account

Navigate to the new account setup page: https://portal.aws.amazon.com/billing/signup#/start										
	Create an AWS account									
AWS Accounts Include 12 Months of Free Tier Access	Email address Password									
Including use of Amazon EC2, Amazon S3, and Amazon DynamoDB Visit aws.amazon.com/free for full offer terms	Confirm password									
	AWS account name									
	Continue Sign in to an existing AWS account									
	© 2018 Amazon Web Services, No. or Its affiliates. All rights reserved. Privacy Policy Tame 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/

Join AWS Educate

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
<u>http://xquartz.macosforge.org/landing/</u>
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 <u>http://sourceforge.net/project/downloading.php?group_id=156984&filename=Xming-6-9-0-31-</u> <u>setup.exe</u> How to setup xMing : http://www.geo.mtu.edu/geoschem/docs/putty_install.html

Download & Install WinSCP http://winscp.net/eng/download.php or Oownload & 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)

sRNA-target Prediction Organizing Tool v1 – Manual

3. On the right-hand side bar under "IMAGES" select "AMIs"



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

aws Service	s 🗸 Resource Groups 🗸 🍾	∆ ● degnar	n 👻 N. Virginia 👻	Support 👻
EC2 Dashboard 5 Events Tags	Launch Actions V Owned by me V Q search : SPOTv1 Add		<u>∆</u> ⊘ K < 1 to	근 후 ② 1 of 1 > 거
Reports Limits	Name AMI Name AMI ID Source	Owner	- Visibility -	Status -
 INSTANCES Instances Launch Templates Spot Requests Reserved Instances Dedicated Hosts Scheduled Instances 	SPOTv1 ami-0ac524786ce503ae1 243768595940/	2437685959	40 Private	available
IMAGES AMIs Bundle Tasks	Image: ami-0ac524786ce503ae1			
 ELASTIC BLOCK STORE Volumes Snapshots Lifecycle Manager 	AMI ID ami-0ac524786ce503ae1 Owner 243768595940 Status available St Croation data September 10, 2018 at 10:31:20	AMI Name S Source 2 tate Reason -	SPOTv1 243768595940/SPOTv1	Edit

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

	ΠΙΟ.ΖΛΙαΙζ		a proces	55015, 04				
a	WS Services ~	Resource Group	os v 🍾			∴ de	gnan 👻 N. Virginia 👻	Support 👻
1. Choos	se AMI 2. Choose Instance	Type 3. Configure	Instance 4. Add St	orage 5. Add Tags	6. Configure Security Group	7. Review		
Step	2: Choose an Ins	stance Type						
Curren	tly selected: t2.micro (Varial	ble ECUs, 1 vCPUs,	2.5 GHz, Intel Xeon	Family, 1 GiB memory,	EBS only)			
	Family	- Type -	vCPUs (j) -	Memory (GiB) 🗸	Instance Storage (GB)	EBS-Optimized Available	Network Performance	IPv6 • Support •
	General purpose	t2.nano	1	0.5	EBS only	-	Low to Moderate	Yes
	General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Moderate	Yes
	General purpose	t2.small	1	2	EBS only	-	Low to Moderate	Yes
	General purpose	t2.medium	2	4	EBS only	-	Low to Moderate	Yes
	General purpose	t2.large	2	8	EBS only	-	Low to Moderate	Yes
	General purpose	t2.xlarge	4	16	EBS only	-	Moderate	Yes
	General purpose	t2.2xlarge	8	32	EBS only	-	Moderate	Yes
	General purpose	m5d.large	2	8	1 x 75 (SSD)	Yes	Up to 10 Gigabit	Yes
	General purpose	m5d.xlarge	4	16	1 x 150 (SSD)	Yes	Up to 10 Gigabit	Yes
					Cancel	revious Review and Lau	nch Next: Configure In	nstance Details
	General purpose	m5.4xlarge	16	64	EBS only	Yes	Up to 10 Gigabit	Yes

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

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

aws Services - Resource	ce Groups 👻 🍾			¢	degnan 👻	N. Virginia 👻	Support 👻						
1. Choose AMI 2. Choose Instance Type 3. C	onfigure Instance 4. Add Storage 5	5. Add Tags 6. Configure	e Security Group 7. Review	1									
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 (j)	Number of instances (i) 1 Launch into Auto Scaling Group (i)												
Purchasing option (j)	Request Spot instances												
Network (j)	vpc-f517f78f (default)	÷ C											
Subnet (j	No preference (default subnet in any	ny Availability Zon \$	Create new subnet										
Auto-assign Public IP 🧃	Use subnet setting (Enable)	\$											
Placement group ()	Add instance to placement group.												
IAM role (j)	None	÷ C	Create new IAM role										
Shutdown behavior (i)	Stop	\$											
Enable termination protection (j)	Protect against accidental termina	ation											
Monitoring (j)	Enable CloudWatch detailed moni Additional charges apply.	itoring											
EBS-optimized instance (j)	Launch as EBS-optimized instance	ce											
			Cancel	Previous	Review and	Launch No	ext: 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

aws	Services 🗸	Resource Groups 🐱	*				¢ د	degnan 👻 N. Virg	inia 👻 Support 👻
1. Choose AMI 2.	. Choose Instance Ty	pe 3. Configure Instance	4. Add Storage	5. Add Tags	6. Configure Security Gr	oup 7. Review	/		
Step 4: Add Your instance will be edit the settings of the storage options in Ar	Storage launched with the he root volume. Yo mazon EC2.	e following storage device s ou can also attach additiona	ettings. You can at I EBS volumes afte	tach additional r launching an	EBS volumes and instar nstance, but not instand	nce store volume ce store volumes	es to your instance a. Learn more abo	∍, or out	
Volume Type (i)	Device (j)	Snapshot (i)	Size (GiB) (i)	Volume Ty	be (j)	IOPS~(i)	Throughput (MB/s) (i)	Delete on Termination (i)	Encrypted (i)
Root	/dev/sda1	snap-0ea8cfee9142df0b9	25	General Pu	pose SSD (GP2)	\$ 100 / 3000	N/A		Not Encrypted
Add New Volume	customers can ge	at up to 30 GB of EBS Gene	ral Purpose (SSD) i	or Magnetic sto	rage. Learn more abou	t free usage tier	eligibility and		
						Can	cel Previous	Review and Lau	nch 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



13. For example Add a key = "Name" and value = "my-SPOT" or "SPOT-server"

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

aws	Services	 Resource 	e Groups 🗸	*			1	🗘 de	gnan 👻	N. Virginia 👻	Support 👻	
1. Choose AMI	2. Choose Instar	nce Type 3. Co	onfigure Instance	4. Add Storage	5. Add Tags	6. Configure Security Group	7. Review					
Step 6: Co A security group is server and allow In Learn more about	nfigure S a set of firewanternet traffic to Amazon EC2	Security G all rules that cont o reach your insta security groups.	roup rol the traffic for ance, add rules t	your instance. On hat allow unrestric	this page, you o ted access to th	can add rules to allow specific to allow specific to the HTTP and HTTPS ports. You	raffic to reach your i can create a new se	nstance. ecurity gr	. For exam roup or se	ple, if you want lect from an exi	to set up a w sting one belo	veb ow.
	Assign a se	curity group:	Create a new se	ecurity group								
		C	Select an existi	ng security group								
	Security g	group name:	launch-wizard	-6								
	1	Description:	launch-wizard	-6 created 2018-0	7-10T10:04:47.0	063-04:00						
Type (i)		Protocol (i)		Port Range (i)		Source (i)			Descrip	tion (i)		
SSH \$	•	TCP		22		Custom \$ 0.0.0.0/0			e.g. SS	H for Admin De	sktop	⊗
Add Rule												
A Warni Rules v	ing with source of 0	0.0.0/0 allow al	I IP addresses to	o access your insta	ance. We recom	mend setting security group rule	es to allow access fr	rom knov	wn IP addr	esses only.		
								Car	ncel P	revious	view and Lau	unch

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

aws	Services - Re	source Groups	~ *		Δ •	degnan 👻	N. Virginia 👻	Support 👻
1. Choose AMI	2. Choose Instance Type	3. Configure Instan	4. Add Storage	5. Add Tags 6. Configure Se	curity Group 7. Review			
Step 7: Re Please review your	view Instance	Launch You can go back	to edit changes for e	ach section. Click Launch to assi	gn a key pair to your instance	and complete	e the launch proc	cess.
A Impro Your in: You car	ve your instances' stances may be accessib n also open additional po	security. Your ole from any IP add orts in your security	security group, dress. We recommer y group to facilitate a	launch-wizard-6, is open to d that you update your security g access to the application or servic	D the world. roup rules to allow access fro e you're running, e.g., HTTP (m known IP a 80) for web se	ddresses only. ervers. Edit secu	rity groups
 AMI Details 	3							Edit AMI
▲ SR SR Ro ✓ Instance Ty	NA-target Prediction Orga ot Device Type: ebs Virtualiz	nizing Tool v1 ation type: hvm					Edit	instance type
Instance Ty	rpe ECUs	vCPUs M	lemory (GiB)	Instance Storage (GB)	EBS-Optimized Available		Network Perfor	mance
t2.micro	Variable	1 1		EBS only	-		Low to Moderate	•
✓ Security G	roups						Edit se	ecurity groups
Security Gr	oup ID	Name		Des	cription			
sg-02ba9bc17	7710fb86d	launch-wiz	zard-6	launc	h-wizard-6 created 2018-07-1	0T11:43:28.04	18-04:00	
							Cancel Prev	ious 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.



19. You will need to download the key and save it to a private location on your computer (e.g., the folder \sim /.ssh/).

a	WS Services ~	Resource Grou	ups ~ 🍾			¢	degnan 👻	N. Virginia 👻	Support 👻
Launo	ch Status								
٢	Your instances are n The following instance lau	ow launching nches have been ir	nitiated: i-0f99c725900b146c	e View launch log					
0	Get notified of estim Create billing alerts to get	ated charges t an email notification	on when estimated charges o	n your AWS bill exceed an amou	nt you define (for example, if	you exc	ceed the free u	sage tier).	
How to	connect to your insta	nces							
Your insta	ances are launching, and it m to accrue until you stop or te	nay take a few mini erminate your insta	utes until they are in the runn i nces.	ng state, when they will be read	/ for you to use. Usage hours	s on you	r new instance	es will start imme	ediately and
Click View	w Instances to monitor your	r instances' status.	Once your instances are in th	e running state, you can conne	ct to them from the Instance	s screer	n. Find out ho	v to connect to y	our instances.
▼ Here	e are some helpful res	sources to get	t you started						
How t	to connect to your Linux inst	ance	Amazon EC2: User Guide						
Learn	about AWS Free Usage Tier		Amazon EC2: Discussion F	Forum					
While yo	ur instances are launching	you can also							
Create	e status check alarms to be r	notified when these	e instances fail status checks.	(Additional charges may apply)					
Create	e and attach additional EBS	volumes (Addition	al charges may apply)						

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

Manage security groups

aws Services - Res	ource Groups 🗸	*			Д •	degnan 🖌 N. Vir	ginia 🗸	Suppor	t •	
EC2 Dashboard	stance Conne	ct Actions *					Д	Ð	¢	0
Events	w tage and attributes or s	oarch by knaword				a	/ / 1+	- E of E		
Tags	by tags and attributes or a	earch by Keyword						0000		
Reports	ne – Inst	ance ID 👻	Instance Type	- Availability Zone -	Instance State	Status Checks	Alarm	Status	Р	ublic
Limits	T-server i-Off	b62b4a7db07be	t2 micro	us-east-1a	running	2/2 checks	None		er	c2-18
INSTANCES md2	018 i-0ec	fd56d143aa83ac	t2 micro	ue-east-1c	etonned	• E/E 0/100/10 ///	None			
Instances		40-45075-4705		us-bast-10			None		•	
Launch Templates Instance:	i-0ffbb62b4a7db07be	(SPOT-server)	Public DNS: e	c2-18-206-100-156.com	pute-1.amazona	ws.com				
Spot Requests		,								
Reserved Instances Descript	ion Status Checks	Monitoring	Tags							
Dedicated Hosts	Instance ID	i-Offbb62b4a7dt	07be		Public DNS (IPv4)	ec2-i XX XX XX XX	compute			
Scheduled Instances	instance ib	1 ONDOLD AN AL				1.amazonaws.com	compute			
MAGES	Instance state	running			IPv4 Public IP	XX.XX.XX.XX				
AMIE	Instance type	t2.micro			IPv6 IPs	-				
Aivis	Elastic IPs				Private DNS	ip-".e	c2.internal			
Bundle Tasks	Availability zone	us-east-1a			Private IPs					
ELASTIC BLOCK STORE	Security groups	launch-wizard-5	5. view inbound	Sec	ondary private IPs					
Volumes	Pahadulad aventa	rules. view outb	oound rules			VDC_f517f79f				
Snapshots	AMI ID	ubuntu/images/	hvm-ssd/ubuntu-		Subpet ID	subnet-1bcd1647				
Lifecycle Manager		xenial-16.04-am	id64-server-		Subilet ID	Subnet Tbourtown				
NETWORK & SECURITY		20180814 (ami- 04169656fea786	6776)							
Security Groups	Platform	-		Ν	Network interfaces	eth0				
Elastic IPs	IAM role	-		S	ource/dest. check	True				
Placement Groups	Key pair name	MyKeyPair			T2/T3 Unlimited	Disabled				
Key Pairs	Owner	243768595940			EBS-optimized	False				
		A 100 0040	10 50 14 114							

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 "IPv4 Public IP" address for your instance. You will use this to login to your server.

IPv4 Public IP:	

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 IPv4 Public IP from above

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

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

Login from **Windows** using PuTTY

a. Open PuTTY

- b. Under Category, click on SSH > Auth
- c. Click browse

sRNA-target Prediction Organizing Tool v1 – Manual

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 "SPOT" 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. sRNA-target Prediction Organizing Tool (SPOT) integrates computational and experimental data to facilitate functional characterization of bacterial small RNAs.
- 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.