Goal for this tutorial

- Perform comparative expression analysis to identify disease or tissue-specific loci
- Analyze user provided brain samples in hg19, provide regions of interest and identify highly expressed small noncoding RNAs that are limited to these regions
- Compare the results against brain and blood datasets in DASHR
- Note: refer to "SPAR_output_and_report_page.pdf" if you want to under the outputs more

SPAR

analysis, annotation, visualization of small RNA sequencing experiments

discover and chararacterize small RNAs from sequencing



Select genome and upload data



Optional pa	arameters
-------------	-----------

analysis, annotation, visualization of small RNA sequencing experiments	
discover and chararacterize small RNAs from sequencing	
home analyze public datasets analyze your own date sutorial about	
Analyze your own data	
Genome	
Human (Fab. 2009 hg19, GRCh27) 🛛 🔻	
Type of input	
Rew signal (SigWig)	
Provide URLs for signal tracks (BigWig format):	
Positive (4) strend: http://teals.poblugenn.edu/-yyee/Algned.out.fitered.hordCloped.sorted.bem.pos.bigWig	
Negetive (*) strend: http://teals.pobl.upenn.edw'-yyee/Aligned.out.fikered.herdClipped.sorted.bom.neg.bigWig	
Exemple data BigWig	
and click	
Analyze your own BigWig	
Analysis options:	
Minimum read length (ni): 1.5 Maximum read length (ni): 64 Minimum ceak helaht (reads): 1.0 Minimum ceak fold chance: 12	
Specify regions of interest:	
Custom region file (26D/26D5/26D12/GFF/GTF): these regions will be overlapped with the peaks identified from the input data	
45 teating.bed	

Optional (not included in current analyses): Change parameters for SPAR analyses: A) Range of length of peaks to be detected B) how many reads are needed to be called a peak

Step 1: Provide a BED file that contains regions of interest. SPAR analyses results will be limited to these regions if provided.

Landing results page (full view)

SPAR analysis, annotation, visualization of small RNA sequencing experiments descene and descentings amail RNA from apparently									
CUSTOM_ANNOT=uploads1extrop.1.5ed	CLISTOM_AVIIOT=uplosds1tering_118ed Analysis results for job 05126cf								
Ingut files: http://teals.gobi.upenn.edu/yyes/Algned.out.filered.handClgped.sorted.bom.pos.bigHtg http://teals.gobi.upenn.edu/yyes/Algned.out.filered.handClgped.sorted.bom.ngb.bigHtg Linkto results: https://www.lisenwengleb.org/SDARS.SDAR_nut/103128.of (will be kept for 2 weeks) Linkto results: Download results (2019) (will be kept for 2 weeks) Linkto report (PDF): Download report (PDF)									
	RUN SUMMARY								
Reads vs RNA class	Peaks vs RNA class								
INTERACTIVE PEAK BROWSER	•								
VIEW IN GENOME BROWSER	•								
CUSTOM ANNOTATION SUMMARY	•								
PLOTS	•								
DOWNLOAD RESULTS	•								

Landing results / outputs page (top)



Landing results / outputs page (top)



Landing results \rightarrow SPAR report page

Index of /SPAR/SPAR_out/05126cf

	Name	Last modified	<u>Size</u>	<u>Description</u>	
2	Parent Directory		-		
	DASHR_comparison/	04-Feb-2018 19:53	-		
	<pre>ENCODE_comparison/</pre>	04-Feb-2018 19:53	-		
	figures/	04-Feb-2018 19:53	-	_	
	<u>inputs/</u>	04-Feb-2018 19:52	-		
	logs/	04-Feb-2018 19:53	-		Report is available in html / pdf
	<u>results/</u>	04-Feb-2018 19:53	-		
	<u>tracks/</u>	04-Feb-2018 19:53			
Ē	<pre>report.html</pre>	04-Feb-2018 19:53	42K		
E	<u>report.pdf</u>	04-Feb-2018 19:53	1.8M		
E	<u>results.html</u>	04-Feb-2018 19:53	2.4K		
	<pre>run_summary.html</pre>	04-Feb-2018 19:53	585		
	SPAR.hg19.config	04-Feb-2018 19:52	2.1K		

Apache/2.2.3 (CentOS) Server at www.lisanwanglab.org Port 443

SPAR report.pdf

A 15 page report containing tables and figures for download and sharing

SPAR - Sequencing-based Pipeline for Ana	alysis of small RNAs	1 / 15		¢	Ŧ	ē	□-		
	S analysis, annotation, visualiz discover and cha Link to results: https://www.lisanwanglab Link to results: Download results (ZIP) (w Link to report (PDF): Download report (PDF)	SPAR analysis, annotation, visualization of small RNA sequencing experiments discover and chararacterize small RNAs from sequencing uk to results: https://www.lisanwanglab.org/SPAR/SPAR_out/05126cf (will be kept for 2 weeks) uk to results: https://www.lisanwanglab.org/SPAR/SPAR_out/05126cf (will be kept for 2 weeks) uk to results: https://www.lisanwanglab.org/SPAR/SPAR_out/05126cf (will be kept for 2 weeks) uk to results: https://www.lisanwanglab.org/SPAR/SPAR_out/05126cf (will be kept for 2 weeks) uk to results: https://www.lisanwanglab.org/SPAR/SPAR_out/05126cf (will be kept for 2 weeks) uk to results: https://www.lisanwanglab.org/SPAR/SPAR_out/05126cf (will be kept for 2 weeks)							
		Run Summary							
	Reads	13,756,091	- 11				11		
	Expressed small RNA loci	6,661	- 11						
	Reads (annotated)	12,801,044					+		
	Reads (unannotated)	955,047	- 11						
	Genes (annotated)	1,982							
		0.001							

Landing results / outputs page (bottom)

After clicking show more at the bottom of the "Run Summary"

Summary
information on
reads

Summary information on peaks

Reads	13,756,091
Expressed small RNA loci	6,661
Reads (annotated)	12,801,044
Reads (unannotated)	955,047
Genes (annotated)	1,982
Called peaks (annotated)	2,201
Called peaks (unannotated)	4,460
Expressed loci length (average)	21.10
Genome coverage (nucleotides)	140,575
RPM (average)	150.13

۲ Show less **Step 1: Expand this to** view the interactive peak browser

INTERACTIVE PEAK BROWSER

VIEW IN GENOME BROWSER CUSTOM ANNOTATION SUMMARY PLOTS

DOWNLOAD RESULTS



VIEW IN GENOME BROWSER



•

Filter sncRNA loci with the top 5% of expression

•

INTER/	ACTIVE P	EAK BROV	VSER			Step 1: Type "95." in this text box and click enter						
Browse peak Download Ta	s (expressed ble	d small RNAs)	l pa	Numbe assed t	er of loc his crite	i that eria: 60						-
Loci: 1-10	/ 60				н	• Pap	1 ▼ of 6 ▶ ▶		Result	s per page 10	• ? }	🕅 Columnsv
chr	peakID	Expression (raw)	strand	Same 5' end read %	Expression (RPM)	Expression (percentile	Annotation	Annotated class	Conservatio	mRNA?	lncRNA?	repeat?
Clear 🔻			Clear 🔻			95.		🗌 Clear		🗌 Clear	🗌 Clear	🗌 Clear
chr1	P131	1358.61	+	0.587891	130.7399	95.13	chr1:145157156- 145157229:+:tRNA-Val- CAC-chr1-16	tRNA	0.3593	Intergenic	Non- lncRNA	Non- repeat
chr1	P150	1400.25	+	0.656677	134.747	95.21	chr1:153671250- 153671282:+:piR-35176	piRNA	0.8856	Intergenic	Non- lncRNA	Non- repeat
chr1	P109	1818.27	-	0.590711	174.9733	95.94	chr1:156420353- 156420375:-∶hsa-miR- 9-3p	mir-3p	1	3' UTR Intron	lncRNA Intron	Non- repeat
chr1	P104	1709.2	-	0.809757	164.4774	95.71	chr1:159141610- 159141684:-:tRNA-Arg- TCT-4-1	tRNA	1	Intergenic	Non- lncRNA	Non- repeat
chr1	P103	1303	-	0.590763	125.3885	95.02	chr1:159141610- 159141684:-:tRNA-Arg- TCT-4-1	tRNA	1	Intergenic	Non- lncRNA	Non- repeat
chr1	P218	1779.94	+	0.574177	171.2848	95.84	chr1:204506526- 204506599:+:tRNA-Lys-	tRNA	0.8393	Intergenic	lncRNA Exon	Non- repeat

Further filter sncRNA loci that only present in Promoter, 3' or 5' UTR exonic regions



P993 in UCSC genome browser



View in Genome Browser (Output)



View in Genome Browser (Output)



Peaks without annotation

Peaks overlapping custom intervals

Un-annotated SPAR peaks

Intervals with peaks

#Custom intervals:					
#Peaks within the custom intervals:	4460				
#Peaks outside the custom intervals:	0				
#Custom intervals with at least one peak:	2230				

Plots (Output)



Rpm distribution of identified small rna loci

Plots (Output)

4 different plot sessions

SEGMENTATION

Length distribution of rpm of segmented loci Genomic length distribution of identified small rna loci

LENGTHEXPRESSION

Rpm distribution of identified small rna loci

Percentile distribution of identified small ma loci across different lengths

PROCESSING

5p end positional offset between identified loci and small ma gene

Processing specificity at 5p end of identified small rna loci

Read pipeup at 5p end of identified small rna loci

GENOMEWIDE

Proportion of expre Genomewide distril Genomewide distril Proportion of mapp Step 1: Click to display, corresponding plot will be shown in the right



5p end positional offset between identified loci and small rna gene

Finish viewing the plots, want to perform own analyses with SPAR results – visit download



Download results (Top) – bulk download



Download results (Bottom) – download by sncRNA class

Step 1: Click to download, e.g. all annotated loci by SPAR that falls within custom annotation region

Step 2: Click to download, e.g. all genomewide unannotated loci from SPAR

bash an oracle_custom.sh SPAR_out/05126cf/results/peaks_annot.xls uploads/t Annotated SPAR peaks (Custom annotation) Un-annotated SPAR peaks (Custom annotation)

Total number of reads falling onto this RNA class

DOWNLOAD PEAK TABLES BY RNA C

RNA class	Peaks	Genes	Reads	Percentage of reads
Annotated	2,201	1,982	12,801,044	93.06
Unannotated	4,460	4,460	955,047	6.94
miRNAprimary	45	45	182,566	1.33
mir-3p	242	242	3,721,158	27.05
mir-5p	258	258	5,349,913	38.89
mir-5p3pno	139	139	452,096	3.29
piRNA	215	203	724,936	5.27
rRNA	190	126	196,397	1.43
scRNA	151	146	193,519	1.41
snRNA	192	167	151,364	1.10
snoRNA	302	223	198,124	1.44
tRF-3-trfdb	89	89	67,183	0.49
tRF-5-trfdb	158	158	1,007,724	7.33
tRF3	2	2	461	0.00
tRNA	218	184	555,597	4.04

Step 3: Click to download, e.g. tRNA only loci results from SPAR

Download results (Output)

We found earlier that P1032, P993, P997, P1479 are of interest, not sure if these peaks are found in any of the DASHR brain tissues.

First excel file downloaded:

1) DASHR_comparison_all.xls – if any of the above peaks are found in this file, meaning they can be present in any DASHR tissue; not necessarily expressed in brain.



P993 and P997 were found in this table. We extracting the four brain related DASHR tissues (now in column I to L). Only column J (a prefronal cortex brain data) have both peaks expressed.

Download results (Output)

Second excel files downloaded:

2) not_in_DASHR.xls (for sanity check only)

	not_in_DASHR - Excel																		
File	Home	Insert	Page Lay	out	Formulas [Data	Review	Vi	ew Ç⊺ellı	ne what yo	ou want to do								
Get External Data ▼	New Query *	Show Qu From Tal	ueries ble Fources	Refresh All •	Connections Properties	2↓ ∡↓	Z A A Z Sort	Filter	Clear	Text to Columns	IFIash Fill I™ Remove D Solution For the second	Duplicates lation ×	I+□ Conso ¤ [©] Relatio ∭ Manag	lidate onships ge Data Mo	W del An	hat-If F alysis ▼	orecast Sheet		öroup ▼ Ingroup ▼ Subtotal
	Get	& Transfor	m	Cor	nections		S	ort & Fi	ter			Data Tool	s			Foreca	st		Outline
E 5	E ST CT																		
A1	A1 \bullet : $\times \checkmark f_{\infty}$ #peakChr																		
A	В	с	D	E	F			G		н	I	J	к	L	М	N	0		P
1 #peakC	hr peakChrS	peakChrE	peakID	peakExpr	ess peakExprPe	annotl	D			annotRNA	class brainog1	l_brainpfc	1 brainpfc2	braintgm	overlap	peakOv	er DASHI	Rpeak	OverlapPct
2 chr16	636771	636809	P1032	1736	.59 95.76	chr16:6	36735:636	6806:-:tF	NA-Gly-CCC-2-2	tRNA					0		0	0	
3 chr2	70249026	70249065	P1479	1745	.09 95.77	chr2:70	248990:70	0249061	:-:tRNA-GIy-CCC-	tRNA					0		0	0	

P1032 and P1479 were found in this table, i.e. not expressed in any of the DASHR tissues, but present in the ENCODE prefrontal cortex data.

Conclusion – out of the 4 sncRNAs we identified, two were in silico confirmed to be highly expressed across different brain datasets in both DASHR and ENCODE.

Run log (Output)

Step 1: click this to minimize "Download results"

DOWNLOAD RESULTS

Content	Description	Download
Peak tracks	Genome-wide tracks	plus strand [bigBED] [382.36 KB] minus strand [bigBED] [384.94 KB]
Raw signal tracks	Genome-wide tracks	plus strand [bigWig] [5.00 MB] minus strand [bigWig] [3.60 MB]
Peak tables	Tables with detailed information: 1. Annotated peaks: peaks overlapping sncRNA genes or mature products with annotations. 2. Unannotated peaks: peaks that do not overlap with annotated sncRNA genes or mature products.	Annotated peaks [xls] Unannotated peaks [xls] All called peaks [xls] [bed]
Gene expression	Small RNA gene expression	Gene expression table (xls) [3.31 MB]
Comparison with DASHR	Integrated expression tables	Annotated peaks [xls] Umannotated peaks [xls] All (Annotated + Umannotated) peaks [xls] [bed] Peaks not in DASHR [xls]
Comparison with ENCODE	Integrated expression tables	Annotated peaks [x1s] Umannotated peaks [x1s] All (Annotated + Unannotated) peaks [x1s] [bed] Peaks not in ENCODE [x1s]

DOWNLOAD PEAK TABLES BY RNA CLASS

RNA class	Peaks	Genes	Reads	Percentage of reads
Annotated	1,282	1,214	9,647,188	92.84
Unannotated	4,799	4,799	744,547	7.16
miRNAprimary	18	18	040,098	6.22
mir-3p	224	224	1,779,736	17.13
mir-5p	285	285	5,786,729	54.92
mir-5p3pno	00	00	38,553	0.29
piRNA	99	99	747,688	7.19
rrina	98	81	42,675	0.41
SCRNA	225	215	227,677	2.19
snRNA	137	132	17,890	0.17
SNORNA	34	29	341, 572	3.29
tRF3	6	6	1,323	0.01
tRF5	9	9	998	0.01
trna.	109	138	183,725	1.88

Step 2: click this to maximize "Run log"

Run log (Output)

Run log displayed agin

RUN LOG

Genome: hg38

Fri, 26 Jan 2018 15:37:49 -0500 Loading PLUS strand: http://dashr2.lisanwanglab.org/DASHRv2/tracks/hg38/ENCODE_dataportal_hg38/fron Fri, 26 Jan 2018 15:37:49 -0500 Loading MINUS strand: http://dashr2.lisanwanglab.org/DASHRv2/tracks/hg38/ENCODE_dataportal_hg38/fro Fri, 26 Jan 2018 15:37:49 -0500 Starting SPAR run Fri, 26 Jan 2018 15:37:49 -0500 Analysis parameters: min. fold change=2; min. coverage=10; min. size=14; max. size=44 bash run_SPAR.sh SPAR_out/ca8ac37/tracks/raw.pos.bigWig SPAR_out/ca8ac37 pos SPAR_out/ca8ac37/config.hg38.sh 2>&1 Fri, 26 Jan 2018 15:37:49 -0500 Processing job ca8ac37..... Fri, 26 Jan 2018 15:37:55 -0500 DONE.

Fri, 26 Jan 2018 15:37:55 -0500 Sending results to the web server..... Fri, 26 Jan 2018 15:39:38 -0500 Processed 10,391,735 reads in 109 seconds (95,337 reads / second)

Fri, 26 Jan 2018 15:39:38 -0500 Creating plots Fri, 26 Jan 2018 15:39:46 -0500 DONE creating plots

SPAR run log

 Questions?

 SPAR@lisanwanglab.org

 Wang Lab
 Penn Neurodegeneration inomics Center
 University of Pennsylvania

 Send questions to this email: SPAR@lisanwanglab.org

END