Part 1 – Sample Preparation

Goal: extract RNA from phage-infected cultures and prepare libraries for Illumina sequencing.

Overview: Infect host cultures with phage and remove samples at desired time points. Lyse cells, extract RNA, remove DNA, deplete ribosomal RNA, and prepare strand-specific libraries.

Required reagents/equipment:

Item	Vendor	Catalog #	List Price
RNeasy Kit mini (50)	Qiagen	74104	$293
RNAzap (250ml)	Thermo Fisher (Ambion)	AM9780	$67
RNAProtect bacteria reagent (2x100ml)	Qiagen	76506	$253
Lysing Matrix B tubes 100x2ml)	MP Biomedicals	116911100	$271
DynaMag magnetic rack for 1.5ml tubes	Thermo Fisher (formerly life technologies)	12321D	$572
RNase Free tubes	General lab supplier	NA	NA
RNase Out (5000 units)	Thermo Fisher (Invitrogen)	10777019	$168
DNA-free Kit	Thermo Fisher (Ambion)	AM1906	$113
RiboZero bacteria kit (6 rxns)	Illumina	MRZMB126	request quote (approx. $600)
"TruSeq® Stranded mRNA LT - Set A" 48 samples 12 indexes	Illumina	RS-122-2101	request quote (approx. $4000)
RNase free tips	General lab supplier	NA	NA
Bead beater	Biospec products	3110BX	$885
DEPC-treated water (4x1000ml)	Thermo Fisher (ambion)	4387937	$165
Ethanol (200 proof)	General lab supplier	NA	NA
RNase free PCR tubes	General lab supplier	NA	NA
Thermocycler	General lab supplier	NA	NA
Superscript II Reverse Transcriptase (2,000 units)	Thermo Fisher (formerly life tech)	18064022	$88
AMPure XP Beads (5ml)	Beckman Coulter	A63880	$315

Protocol:

Grow 20ml *M. smeg.* to log phase ($O.D._{600}$ = 0.7-1.0) in baffled flask (7H9-ADC, $CaCl_2$, CB, CHX – no tween).
Centrifuge culture in 50ml conical for 3min at 5500xg in large centrifuge, preferably at RT.
Pour off supernatant (BUT SAVE/KEEP STERILE).
Resuspend pellet in phage (normally I do an MOI = 3, see below for details). If you need to, add a bit of the supernatant you saved to resuspend pellet if the phage doesn't do the trick.
Incubate, undisturbed, at RT for 10min.
Pour saved supernatant into pellet and resuspend, put all back into baffled flask.
Put flask in 37^{o}C shaker.

<u>At 30min and 2hrs after you put the culture in the 37C do the following steps:</u>
Add 500ul of culture to 8 microcentrifuge tubes (total will be 4ml).
Add 1ml RNAprotect reagent (Qiagen 1018380) to each tube.
Vortex 5sec.
Incubate at RT for 5min.
Centrifuge 1min at 5000xg.
Decant supernatant and place inverted tube on paper towel for 10sec.
Here you can STORE pellet at -20^{o}C for 2wks or -70 ^{o}C for 4wks.

<u>RNA extraction – RNeasy Mini Kit (Qiagen 74104):</u>
Add 700ul of RLT (1ml of RLT + 10ul of 2-mercaptoethanol) to one 500ul cell pellet, once resuspended take the 700ul and transfer to another tube and resuspend pellet. Do this for 4ml total culture- now have 4ml cells resuspended in 700ul RLT.
Place liquid into 1 matrix lysing tube (matrix B) (MP Biomedicals 6911-100).
Bead beat for 45sec (setting 5) on max speed.
Place on ice for 1min.
Bead beat for 45sec (setting 5) on max speed.
Centrifuge for 10sec at max speed.
Transfer supernatant into new tube.
Add 560ul of 80% EtOH to supernatant.
Shake vigorously.
Transfer 700ul into RNeasy mini spin column.
Centrifuge 15sec at 10,000rpm.
Discard flow-through.
Add the rest of the super to the spin column.
Centrifuge 15sec at 10,000rpm.

Discard flow-through.
Add 700ul of buffer RW1 to spin column.
Centrifuge 15sec at 10,000rpm.
Discard flow-through and collection tube.
Place spin column in new collection tube (provided in kit).
Add 500ul Buffer RPE to spin column.
Centrifuge 15sec at 10,000rpm.
Discard flow-through.
Add 500ul Buffer RPE to spin column.
Centrifuge 2min at 10,000rpm.
Centrifuge 1min at 10,000rpm to get rid of any excess EtOH.
Place column in 1.5ml collection tube (provided in kit).
Add 50ul RNase-free water directly to spin column membrane.
Centrifuge 1min at 10,000rpm to elute RNA.
Place column in **new** 1.5ml collection tube.
Add 50ul RNase-free water directly to spin column membrane.
Centrifuge 1min at 10,000rpm to elute RNA.

Quantify yields – Nanodrop
I normally get around 0.8-1ug/ul for the first elution tube.
Sometimes I don't even do a second elution.

Remove DNA – Complete Turbo DNA-free Kit (Ambion AM1907):
For first elution: Add 0.1 volume of 10x Turbo DNase buffer and 1ul Turbo DNase to 50ul RNA.
Incubate in 37^{o}C heat block for 30min.
Add an additional 1ul Turbo DNase.
Incubate in 37^{o}C heat block for 30min.
Add 0.2 volumes of DNase Inactivation Reagent.
Incubate 5min at RT, mix occasionally (white particles are capturing the DNase and DNA).
Centrifuge 10,000xg for 1.5min and transfer RNA to fresh tube (don't be too greedy, you don't want those white beads in your RNA!).

Remove rRNA – Ribo-Zero Kit (Illumina MRZMB126):
See manufacturer's protocol. Use EtOH ppt. as directed in the protocol to purify mRNA.

Prepare libraries – TruSeq Stranded mRNA Library Prep Kit (Illumina RS-122-2101):
See manufacturer's protocol, start on page 20, step 12.
Provide samples to sequencing facility.

How to calculate Multiplicity of Infection (MOI):

Infection – 2ml *M. smeg.* with Giles at MOI = 0.1

3.5×10^7 cfu/ml X
---------- = --------
0.1 O.D.$_{600}$ 1.0 (my O.D. of *M. smeg.* culture)

$X = 3.5 \times 10^8$ cfu/ml

3.5×10^8 cfu/ml X 2ml culture X MOI 0.1 = 7.0×10^7 pfu needed
(vol. used) (MOI you want)

7×10^7 pfu needed
---------- = 0.00875ml phage lysate added to 2ml culture
8×10^9 pfu/ml (lysate)

Add 8.75ul Giles lysate to 2ml *M. smeg.* culture to get MOI = 0.1

Part 2 – Data Analysis

Goal: process next generation sequencing strand-specific transcription data to analyze with a visualization program.

Overview: Once RNAseq samples have been sequenced, the data needs to be processed in order to visualize genome-wide transcriptional patterns. There are a range of software and data processing pipelines available to process the data, and they fall on a spectrum of cost, ease-of-use, and flexibility. The following summary presents a few examples of the types of processing tools, as well as a general overview of the command line pipeline currently used in the Hatfull lab.

Data Processing Options:

Option	Method	Cost	Pros	Cons	Notes
1	Galaxy (server or locally-installed)	Free	Point-and-click; Online or locally installed	Some users have experienced difficulty/delays with processing and default settings	https://galaxyproject.org/
2	CLC Genomics Workbench	Expensive	Point-and-click	Some tools are propriety	https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/
3	In-house command line pipeline	Free	Maximum user flexibility	Need command line experience; several programs need locally installed	More information below

Command Line Pipeline Details:

Tools/skills needed:

- Mac or Linux operating system
- Sufficient computing power (check software documentation for recommendations)
- Command line experience:
    - Installing software
    - Adding software to shell's path
    - Understanding unique options/flags available to each program
- File format distinctions and specifications (e.g. FastA, tab-delimited, gff3, etc.)

Software used:

For reference file preparation:

- General text editor like Text Wrangler (NOT a word processor like MS Word)
- EMBOSS Seqret: http://www.ebi.ac.uk/Tools/sfc/emboss_seqret/

Data quality assessment:

- FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Command line tools:

- cutadapt: https://cutadapt.readthedocs.io/en/stable/index.html
- bowtie2: http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
- sed: pre-installed command line utility on Linux/Unix systems
- samtools: http://www.htslib.org/
- bedtools: http://bedtools.readthedocs.io/en/latest/

Data visualization:

- Integrative Genomics Viewer (IGV): https://www.broadinstitute.org/igv/

Input file(s) needed:

	File example	Format	Description	Software	Notes
1	raw_data.fastq	Fastq	Raw sequencing data	FastQC; cutadapt	Obtained from Illumina
2	reference_seq.fasta	FastA	Reference genome sequences	bowtie2; IGV	Contains sequences of host and phage genomes
3	bowtie_index.bt2		Index files	bowtie2	Generated by bowtie2 using reference_seq.fasta; Comprised of 6 different files
4	samtools_index.fai		Index files	samtools	Generated by samtools using reference_seq.fasta
5	reference_seq.genome	Tab-delimited	Reference genome sizes	bedtools	Contains sequence length of host and phage genomes
6	gene_annotations.gff3	Gff3	Gene annotation coordinates	IGV	Can be tricky to create

Step-by-step preparation of reference files for data processing:

Step	Goal	Software	Input file(s)	Output file(s)	Example command or Notes
1	Create reference_seq.fasta	text editor	FastA files from phagesdb and GenBank	reference_seq.fasta	Manually create a multi-FastA file by copying and pasting sequences in a text editor.
2	Create bowtie_index.bt2	bowtie2-build	reference_seq.fasta	bowtie_index.bt2	bowtie2-build reference_seq.fasta bowtie_index
3	Create samtools_index.fai	samtools faidx	reference_seq.fasta	samtools_index.fai	samtools faidx reference_seq.fasta
4	Create reference_seq.genome	text editor		reference_seq.genome	Manually create a two-column tab-delimited list of genome names and sizes
5	Create gene_annotations.gff3	EMBOSS; Excel; text editor;	GenBank files of phage and host genomes	gene_annotations.gff3	Use tools like EMBOSS or CLC Workbench to create gff3 file for each genome. Manually combine host and phage genomes in a single text file using a text editor.

Step-by-step data processing:

Step	Goal	Software	Input file(s)	Output file(s)	Example command
1	Assess raw data quality	FastQC	raw_data.fastq	raw_fastqc.html	
2	Trim reads	cutadapt	raw.fastq	trimmed.fastq	cutadapt -q 30,30 -o trimmed.fastq raw.fastq
3	Assess trimmed data quality	FastQC	trimmed.fastq	trimmed_fastqc.html	
4	Map to genome	bowtie2	trimmed.fastq; bowtie_index.bt2	mapped.sam	bowtie2 -x bowtie_index -U trimmed.fastq -S mapped.sam
5	Remove non-uniquely mapped reads	sed	mapped.sam	unique.sam	sed '/XS:i/d' mapped.sam > unique.sam
6	Convert to BAM	samtools view	unique.sam	unique.bam	samtools view -b -o unique.bam unique.sam
7	Sort data	samtools sort	unique.bam	sorted.bam	samtools sort -o sorted.bam -T temp unique.bam
8	Index data	samtools index	sorted.bam	sorted.bai	samtools index sorted.bam
9	Convert to BED	bedtools bamtobed	sorted.bam	sorted.bed	bedtools bamtobed -i sorted.bam > sorted.bed
10	Compute top strand coverage	bedtools genomecov	sorted.bed; reference_seq.genome	plus.bedgraph	bedtools genomecov -bga -strand + -i sorted.bed -g reference_seq.genome > plus.bedgraph
11	Compute bottom strand coverage	bedtools genomecov	sorted.bed; reference_seq.genome	minus.bedgraph	bedtools genomecov -bga -strand - -i sorted.bed -g reference_seq.genome > minus.bedgraph
12	Visualize data	IGV	plus.bedgraph; minus.bedgraph; reference_seq.fasta; gene_annotations.gff3		