

Outline

Riboprofiling samples

Trimming reads

Quality filtering

Removing contaminants

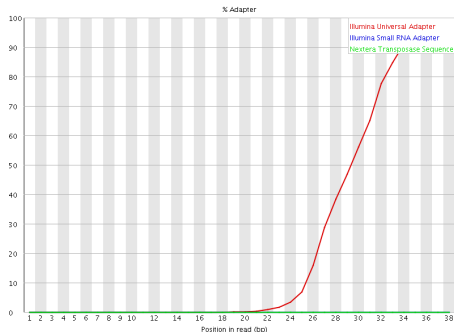
Mapping to transcriptome: protein-coding + lncRNA

Riboprofiling samples

	6h	18h	120h
Reads	42130444	45982977	44077266
Sequence length	50	50	50
%GC	59	55	56

Trimming

✖ Adapter Content



- ▶ min length: 25
- ▶ min adapter alignment length: 5
- ▶ unclipped discarded
- ▶ first base discarded

Trimming

FASTX-Toolkit

id	input	output	discarded TooShort	discarded AdapterOnly	discarded NonClipped
H006	42,130,444	38,531,976	3,066,153	14,044	518,271
H018	45,982,977	42,037,165	3,233,491	70,200	642,121
H120	44,077,266	36,994,281	5,778,075	701,398	603,512

Cutadapt

`-quality-cutoff=10` Trim low-quality bases from 3' ends of each read before adapter removal.

id	input	output	discarded TooShort	reads with Adapter
H006	42,130,444	34,439,444 (81.7%)	6,933,352 (16.5%)	41,271,051 (98.0%)
H018	45,982,977	39,441,134 (85.8%)	5,682,240 (12.4%)	44,907,869 (97.7%)
H120	44,077,266	32,501,482 (73.7%)	10,638,847 (24.1%)	42,985,050 (97.5%)

Quality filtering

Codetri

- ▶ hq = 30
- ▶ lq = 26
- ▶ min
length=25

id	readsInput	unpairedReadsOutput
H006	38,531,976	30,125,320
H018	42,037,165	35,753,745
H120	36,994,281	29,668,867

Cutadapt

–quality-cutoff=10 Trim low-quality bases from 3' ends of each read before adapter removal.

id	input	output	discarded TooShort	reads with Adapter
H006	42,130,444	34,439,444 (81.7%)	6,933,352 (16.5%)	41,271,051 (98.0%)
H018	45,982,977	39,441,134 (85.8%)	5,682,240 (12.4%)	44,907,869 (97.7%)
H120	44,077,266	32,501,482 (73.7%)	10,638,847 (24.1%)	42,985,050 (97.5%)

Removing contaminants - rRNA

- ▶ Bowtie: default parameters; seed length=20

id	reads processed	reads with at least one alignment	reads that failed to align
H006	30,125,320	25,292,691 (83.96%)	4,832,629 (16.04%)
H018	35,753,745	10,480,722 (29.31%)	25,273,023 (70.69%)
H120	29,668,867	11,498,713 (38.76%)	18,170,154 (61.24%)

- ▶ STAR: without split mapping; max 10 multimaps; min read alignment 2/3 read length

id	reads processed	uniquely mapped	multiple loci	too many loci	discarded too short
H006	34,439,444	29,084,185 (84.45%)	4,775 (0.04%)	158 (0.00%)	15.51%
H018	39,441,134	11,902,992 (30.18%)	73,911 (0.19%)	4,052 (0.01%)	69.62%
H120	32,501,482	12,525,154 (38.54%)	39,688 (0.12%)	10,985 (0.03%)	61.29%

Mapping to transcriptome: protein-coding + lncRNA

- ▶ default parameters (seed length=28)
- ▶ -k 10
report up to $\langle int \rangle$ good alignments per read (default: 1)
- ▶ -m 10
suppress all alignments if $\langle int \rangle$ exist (def: no limit)
- ▶ -n 1
max mismatches in seed (can be 0-3, default: -n 2)

id	reads processed	reads with at least one alignment	reads that failed to align	reads with alignments suppressed due to -m:
H006	4,832,629	3,705,610 (76.68%)	985,545 (20.39%)	141,474 (2.93%)
H018	25,273,023	20,743,108 (82.08%)	3,626,492 (14.35%)	903,423 (3.57%)
H120	18,170,154	11,236,990 (61.84%)	5,810,052 (31.98%)	1,123,112 (6.18%)

Mapping to transcriptome: protein-coding + lncRNA

Bowtie - 1st pipeline test

id	reads processed	reads with at least one alignment	reads that failed to align	reads with alignments suppressed due to -m:
H006	4,832,629	3,705,610 (76.68%)	985,545 (20.39%)	141,474 (2.93%)
H018	25,273,023	20,743,108 (82.08%)	3,626,492 (14.35%)	903,423 (3.57%)
H120	18,170,154	11,236,990 (61.84%)	5,810,052 (31.98%)	1,123,112 (6.18%)

STAR - 2nd pipeline test

id	reads processed	uniquely mapped	multiple loci	too many loci	discarded too short
H006	5,340,484	727,057 (13.61%)	979,375 (18.34%)	51,635 (0.97%)	66.99%
H018	27,464,231	2,257,427 (8.22%)	3,443,662 (12.54%)	207,718 (0.76%)	78.35%
H120	19,936,640	2,312,722 (11.60%)	3,927,984 (19.70%)	239,446 (1.20%)	67.37%

Mapping to transcriptome: protein-coding + lncRNA

STAR - 2nd pipeline test

without split mapping; max 10 multimaps; min read alignment 2/3 read length

id	reads processed	uniquely mapped	multiple loci	too many loci	discarded too short
H006	5,340,484	727,057 (13.61%)	979,375 (18.34%)	51,635 (0.97%)	66.99%
H018	27,464,231	2,257,427 (8.22%)	3,443,662 (12.54%)	207,718 (0.76%)	78.35%
H120	19,936,640	2,312,722 (11.60%)	3,927,984 (19.70%)	239,446 (1.20%)	67.37%

STAR - 2nd pipeline test

without split mapping; max 10 multimaps; min read alignment 0.4 read length

id	reads processed	uniquely mapped	multiple loci	too many loci	discarded too short
H006	5,340,484	819,649 (15.35%)	2,825,198 (52.90%)	1,687,294 (31.59%)	0.07%
H018	27,464,231	2,535,749 (9.23%)	15,238,915 (55.49%)	9,583,448 (34.89%)	0.26%
H120	19,936,640	2,716,101 (13.62%)	8,081,281 (40.53%)	9,027,174 (45.28%)	0.43%