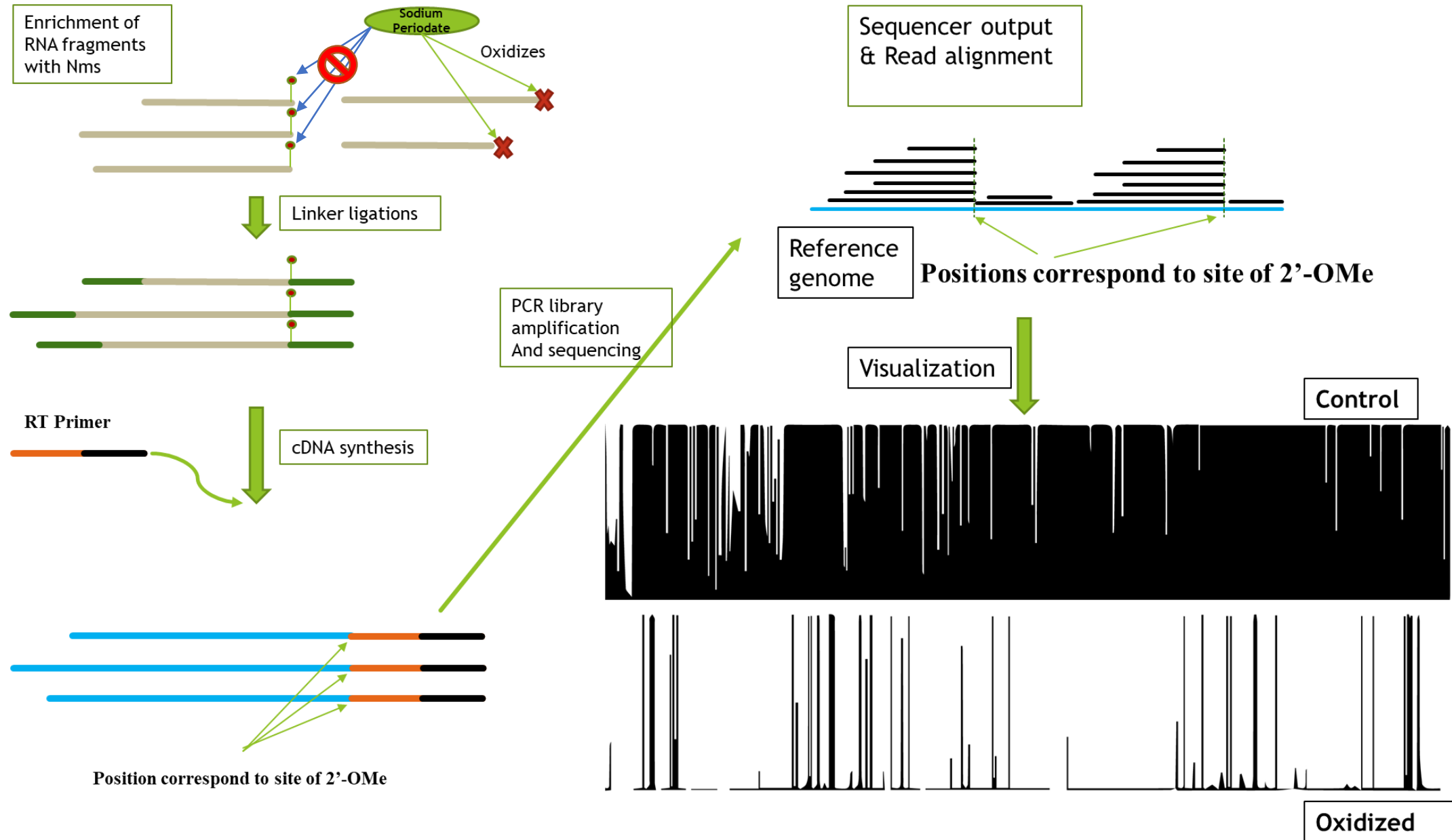


Demonstration of basic RibOxi-seq concept



RNA structure after 3' and 5' linker ligations

5'-/Biosg/ACACUCUUCCCUACACGACGCUCUUCGAUCUNNNN-----insert-----ATCACGCTGTAGGCACCATCAATGACAG/SpC3/- 3'

cDNA structure

Randomer for deduplication 3'-linker sequence

3'-TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNN----insert----TAGTGCGACATCCGTGGTAGTTACTGTCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG-5'

Complementary to partial Illumina i5 PCR primer

In-line barcode for
filtering out RT mispriming

Identical to partial Illumina i7 PCR primer

Sequencing Output (example using 75x75 Paired end sequencing)

Read1:

@ Sequence_identifier+i7_index+i5_index
NNNN_SOME_INSERT_AND_MAYBE_READ_THROUGH_ADAPTER
+
_____QUALITY_OF_THE_READ_____

Read2:

@Sequence_identifier+i7_index+i5_index
GTCATTGATGGTGCCTACAGCGTGAT_SOME_INSERT_AND_READ_THROUGH_ADAPTER
+
_____QUALITY_OF_THE_READ_____

1. *cutadapt* paired-end mode remove read-through adapters:

Read1 output:

```
@ Sequence_identifier+i7_index+i5_index
NNNN_SOME_INSERT
+
_____QUAL
```

Read2 output:

```
@Sequence_identifier+i7_index+i5_index
GTCATTGATGGTGCCTACAGCGTGAT_SOME_INSERT
+
_____QUALITY_OF_THE_REA
```

2. *pear* to merge read1 and read2 into a single read:

```
@ Sequence_identifier+i7_index+i5_index
NNNN_SOME_INSERT_ATCACGCTGTAGGCACCATCAATGACAG
+
_____QUALAER_EHT_FO_YTILAUQ_____
```

3. *move_umi.py* moves randomer sequence from read to read identifier and discard reads lacking **ATCACG**:

```
@ Sequence_identifier_NNNN+i7_index+i5_index
SOME_INSERT_ATCACGCTGTAGGCACCATCAATGACAG
+
_____QUALAER_EHT_FO_YTILAUQ_____
```

4. *cutadapt* removes the rest of the linker sequence:

```
@ Sequence_identifier_NNNN+i7_index+i5_index  
SOME_INSERT  
+  
_____QUAL
```

5. *STAR* alignment to genome of interest

6. *samtools* indexing the bam output

7. *umi_tools* deduplication:

Example aligned read #1:

UMI: AATC

Sequence: GGTTACG

Example aligned read #2:

UMI: CGTA

Sequence: GGTTACG

Not duplicates

Example aligned read #1:

UMI: AATC

Sequence: GGTTACG

Example aligned read #2:

UMI: AATC

Sequence: GGTTACG

Duplicates, discard

- 8. *bedtools* converts deduplicated bam to bed file format
- 9. *genomecov* generates genome coverage track of 3'-end only to visualize Nm sites on genome browser
- 9. *riboxi_bed_parsing.py* counts all 3'-end alignments and generates a tab delimited file listing chromosome number, base position, gene name, (+15,base,-15) nucleotide sequence and counts for all samples:

1	chr	base	gene	seq	ms_wt_186_0x	
2	chr13	97190585	Hexb	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		13263
3	chr9	123462160	Lars2	gttggtgcatggtaatcctgctcagtacga		11174
4	chr2	102829412	Cd44	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		10624
5	chr17	39848136	Rn45s	gaagacggtcgaacttgactatctagaggaa		9501
6	chr17	39847806	Rn45s	tcccccaacttcttagagggacaagtggcgt		4412
7	chr17	39846960	Rn45s	gaggatccattggagggcaagtctggtgcca		3148
8	chr7	19697484	Apoe	CCAAGTCACACAAGAACTGACGTGAGTGTC		2871
9	chr5	136932984	4933404012Rik	gtctccaaggtgaacagcctctggcacattg		2850
10	chr1	100180179	Cntnap5b	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		2801
11	chr17	39847687	Rn45s	attccgtgggtggtggtgcatggccgttctt		2405
12	chr11	17221313	Wdr92	gtctccaaggtgaacagcctctggcatgttg		2286

Downstream analysis can then be performed

RibOxi-seq Counts Visualization

Sample:

ms_wt_186_C

Single Base Minimum counts:

50

Gene Symbol:

Lars2

Genome:

mm10

Plot

Show 10 entries

Search:

	gene	chr	base	seq	sample_id	counts
	All	All	All	All	All	All
21686	Hexb	chr13	97190586	GGAGAAGGGTTCCATGTGAACAGCAGTTGAA	ms_wt_186_C	890
110236	Lars2	chr9	123462159	TGTTGTTGCCATGGTAATCCTGCTCAGTACG	ms_wt_186_C	781
58900	Cd44	chr2	102829413	TGTTGTTGCCATGGTAATCCTGCTCAGTATG	ms_wt_186_C	599
38716	Rn45s	chr17	39847805	GTCCCCCAACTTCTTAGAGGGACAAGTGGCG	ms_wt_186_C	586
38155	Rn45s	chr17	39846959	CGAGGATCCATTGGAGGGCAAGTCTGGTGCC	ms_wt_186_C	519
38950	Rn45s	chr17	39848135	AGAAGACGGTCGAAC TTGACTATCTAGAGGA	ms_wt_186_C	510
93454	Apoe	chr7	19697485	CCCAAGTCACACAAGAAGTACGTGAGTGTC	ms_wt_186_C	501
38319	Rn45s	chr17	39847225	CGCAGCTAGGAATAATGGAATAGGACCGCGG	ms_wt_186_C	463
38671	Rn45s	chr17	39847749	GATAACGAACGAGACTCTGGCATGCTAACTA	ms_wt_186_C	427
81265	Alb	chr5	90468666	AGTGTGCAAGAAGTATGCTGAGGCCAAGGAT	ms_wt_186_C	409

Showing 1 to 10 of 244 entries

Previous12345...25Next

Download table

Chromosome

123.37 mb123.38 mb123.39 mb123.4 mb123.41 mb123.42 mb123.43 mb123.44 mb123.45 mb123.46 mb

ms_wt_186_C

Gene Model

Download plot

Download table for this gene

Select a site:

- 123461873.ms_wt_186_C:1
- 123461873.ms_wt_186_C:2
- 123461874.ms_wt_186_C:5
- 123461875.ms_wt_186_C:395
- 123461876.ms_wt_186_C:8

Download plot

