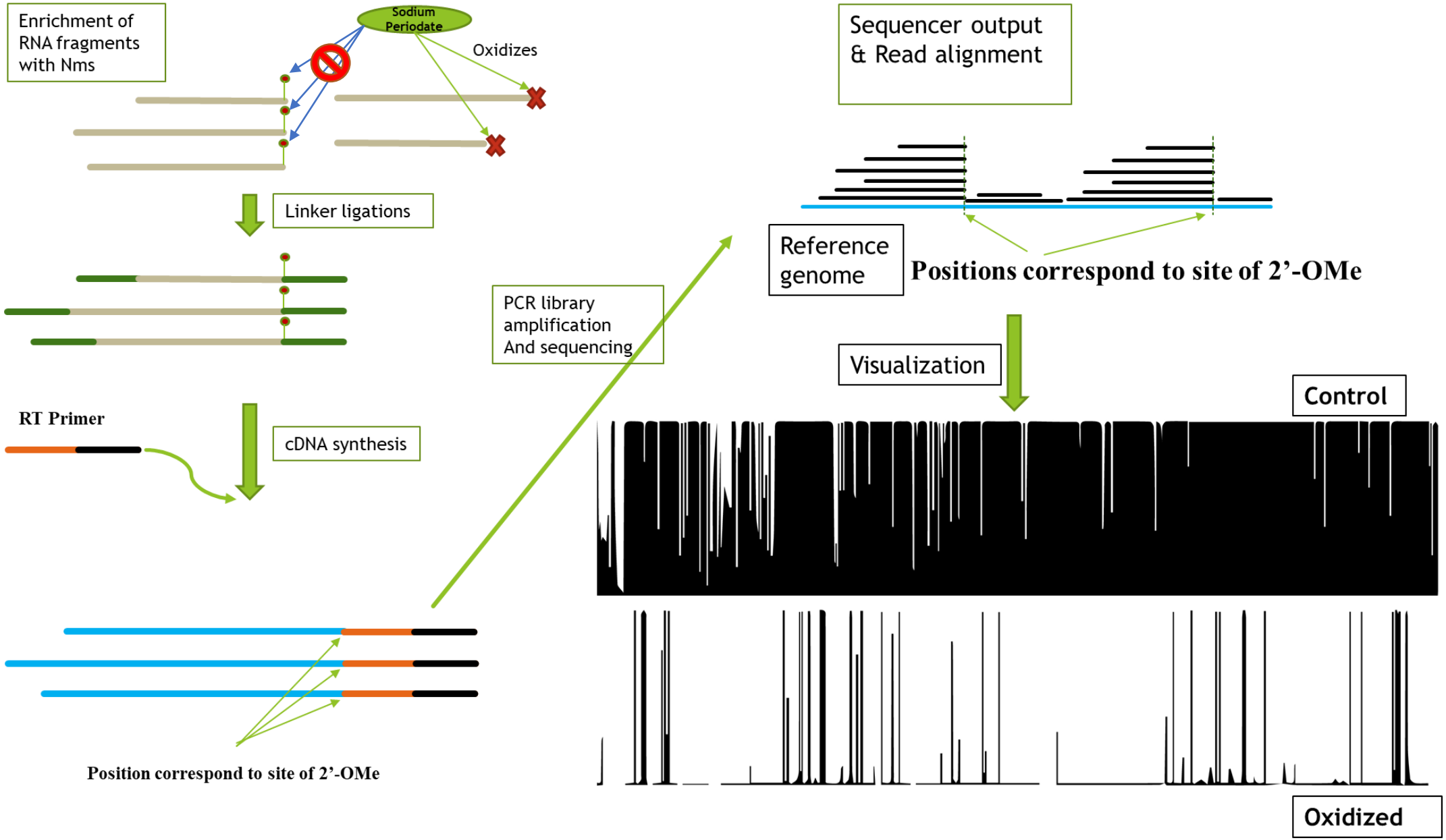
Demonstration of basic RibOxi-seq concept



RNA structure after 3’ and 5’ linker ligations

5'-/Biosg/ACACUCUUUCCCUACACGACGCUCUUCCGAUCUNNNN------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:	Read2:
@ Sequence_identifier+i7_index+i5_index NNNN_SOME_INSERT_AND_MAYBE_READ_THROUGH_ADAPTER + _____QUALITY_OF_THE_READ_____	@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:

	chr	base	gene	seq	ms_wt_186_Ox
2	chr13	97190585	Hexb	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	13263
3	chr9	123462160	Lars2	gttgttgccatggtaatcctgctcagtacga	11174
4	chr2	102829412	Cd44	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	10624
5	chr17	39848136	Rn45s	gaagacggtcgaacttgactatctagaggaa	9501
6	chr17	39847806	Rn45s	tcccccaacttcttagagggacaagtggcgt	4412
7	chr17	39846960	Rn45s	gaggatccattggagggcaagtctggtgcca	3148
8	chr7	19697484	Apoe	CCAAGTCACACAAGAACTGACGTGAGTGTCC	2871
9	chr5	136932984	4933404O12Rik	gtctccaaggtgaacagcctctggcacattg	2850
10	chr1	100180179	Cntnap5b	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	2801
11	chr17	39847687	Rn45s	attccgtgggtggtggtgcatggccgttctt	2405
12	chr11	17221313	Wdr92	gtctccaaggtgaacagcctctggcatgttg	2286

Downstream analysis can then be performed