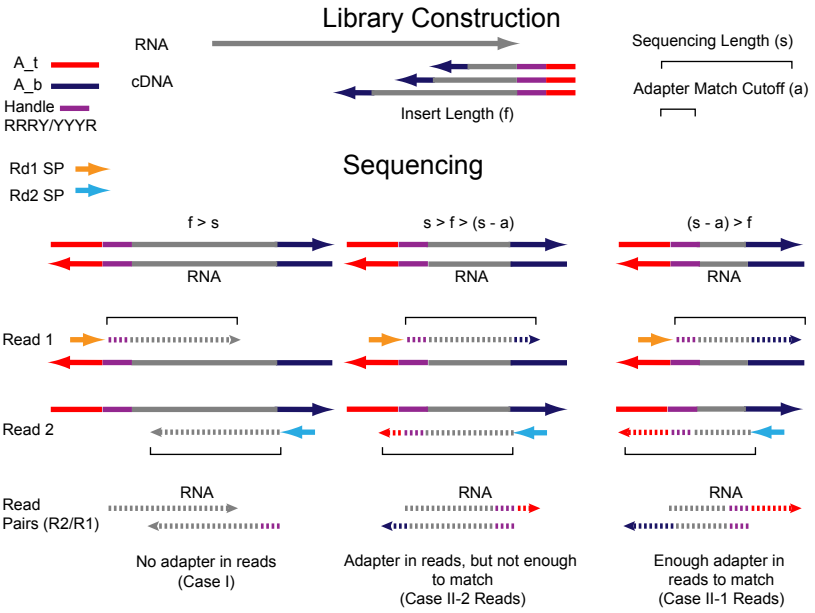
Adapter Trimmer Algorithm



Adapter Clipping Step 1: Use bowtie to separate Case I and Case II Reads

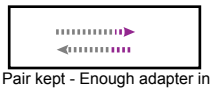
Trim 4nts from the 5' end of the reads to remove handle in Read 1, and 4nts from the 3' end of reads to remove special case for read 2 extending into the handle but not the adapter.

Use bowtie to match against targets files.

Adapter Clipping Step 2: Use cutadapt on Case II Reads

Starting with Reads that were thrown away in Step 1 (Case II Reads)

Apply cutadapt to original files using adapter match cutoff to match adapter.



Adapter Clipping Step 3: Trim Search, Look for Revcomp on remaining Case II Reads

For base = s-1, s-2, ..., s-a+1

ex: base = s-1 : No reads kept since haven't trimmed off enough adapter

Trim down each read to length = base

Revcomp Read 1

Align both reads to targets in forward direction (do not search revcomp strand of targets).

(Trim off 4nts on 3' end before align to remove handle).

Revcomp Read 1 Back and Store (before removing 4nts off 3' end)

Nothing Stored

ex: base = s-2 : Keep middle read since trim off exactly correct amount of adapter

Trim down each read to length = base

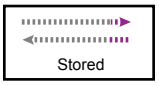
Revcomp Read 1

Align both reads to targets in forward direction (do not search revcomp strand of targets).

(Trim off 4nts on 3' end before align to remove handle).

Should not align | Aligned | Do not align (contains adapter)

Revcomp Read 1 Back and Store (before removing 4nts off 3' end)



ex: base = s-a+1 : No reads kept since have trimmed off too much adapter (go 2nts into handle)

Trim down each read to length = base

Revcomp Read 1

Align both reads to targets in forward direction (do not search revcomp strand of targets).

(Trim off 4nts on 3' end before align to remove handle).

Should not align | Do not align (seqs don't match) | Do not align (contains adapter)

Revcomp Read 1 Back and Store (before removing 4nts off 3' end)

Nothing Stored

At each step temporarily store reads that do not align so can search through those for speedups.

Final Processing

Combine all reads in single file including those with no matches in case they are due to mutations that spats can handle.
Remove 4nts from 3' end of R2 to remove handle sequence in R2 in short reads.

Store

From Step 1:

From Step 2:

From Step 3: