

## Preprocessing

**1** De-deuplication (BBTools Clumpify) (optional)



**2** Split (optional)



**3** Adapter removal (SeqPrep)



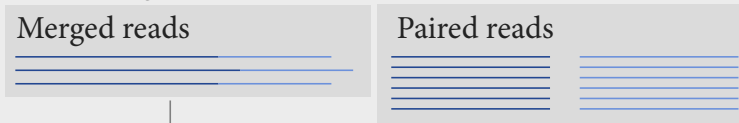
OR

Adapter removal and merge overlapping R1 and R2 (SeqPrep; optional)



## Alignment

**4** Align to viral reference(s) (bwa mem)

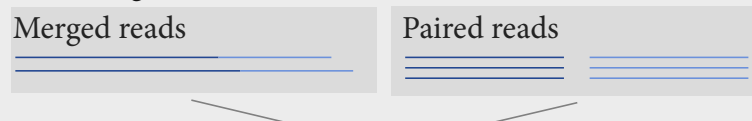


Extract mapped reads

Extract pairs with at least one read mapped

Combined

**5** Align to host reference (bwa mem)



Combined

## Identify integrations

**6**

chimeric reads

integrations

discordant pair

host  
virus



read

read

R1

R2

vector rearrangement

host  
virus



read

ambiguous location

host  
virus



read

## Postprocessing

**7** Filter

pass

Ambiguous location?

fail

Filter failed

unique

host

virus

both

Clustering

Unique, clustered integrations

Integrations with ambiguous location