

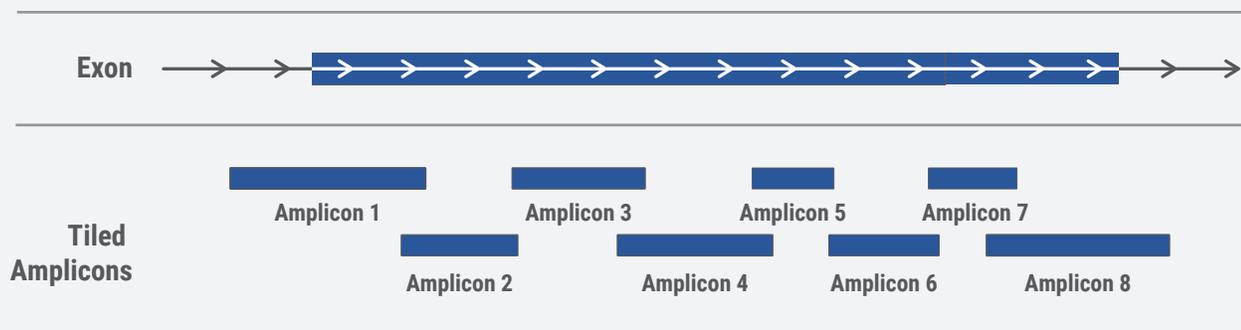
## PRIMERCLIP™: A TOOL FOR TRIMMING PRIMER SEQUENCES USING COMMAND LINE OR GALAXY\*

### Importance of Soft-Clipping Primer sequences for Accel-Amplicon™ and Accel-Amplicon Plus Panels™

Accel-Amplicon and Accel-Amplicon Plus Panels leverage a short overlapping (tiled) amplicon design (Figure 1) to provide contiguous coverage of regions of interest in a single-tube format, yet maintain compatibility with short length of Cell-Free DNA (cfDNA) and damaged/degraded Formalin-Fixed Paraffin Embedded (FFPE) samples.

Due to the recommended 2x151 read length and short amplicon design, synthetic primer sequences introduced during PCR will be encountered both at the beginning and end of reads, which must be trimmed to avoid variant calling from these artificial sequences. To accomplish this, Swift has developed and validated the “Primerclip” tool to soft-clip primer bases after alignment. Primerclip bioinformatically clips the 5’ and 3’ primer bases so they are no longer a part of the alignment and hence no variants are being called from these synthetic sequences. In addition to the speed, Primerclip also has the advantage of improving variant calling at ends of alignments which may be otherwise compromised due to edge effect. Variants present at edges/ends of the amplicons will have a better chance of getting called if the primer bases were present during the alignment and were trimmed/clipped just before variant calling.

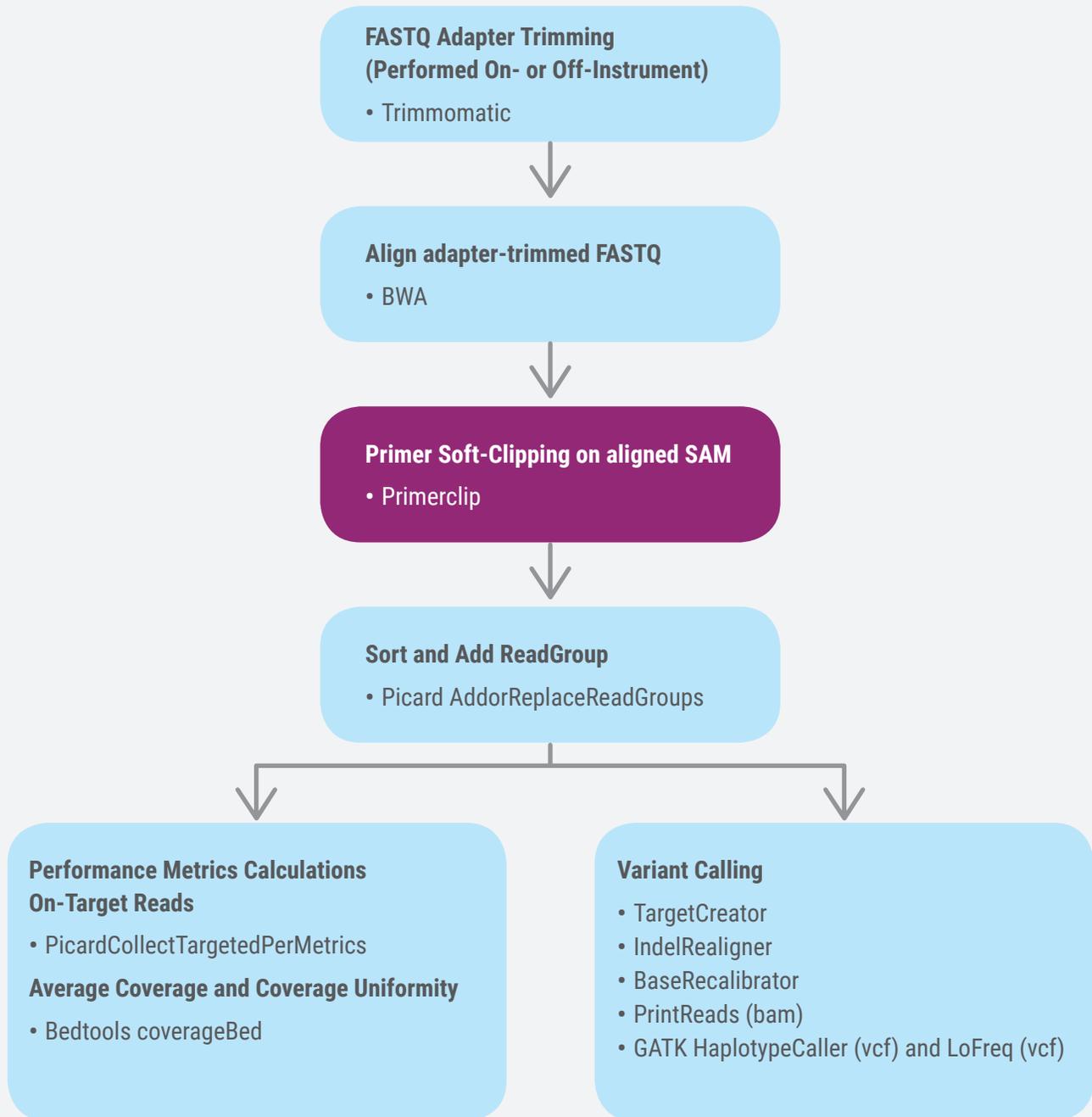
**NOTE:** Soft-clipping by Primerclip is only performed at the ends of reads and not when the primer sequence is present internally (within a read), such that when amplicons are tiled contiguously across a region, coverage of the trimmed primer sequence is provided by true sequence from an adjacent amplicon within that region.



**Figure 1:** Accel-Amplicon Panels leverage overlapping amplicons to allow for contiguous regions of coverage in a single-tube format. For example, this schematic shows eight overlapping amplicons across the region of interest to provide contiguous coverage of the entire exon in a single-tube format.

## Pipeline Overview

Trimming primer sequences is mandatory for accurate data output and variant calling from Accel-Amplicon panels. A suggested data analysis pipeline for Accel-Amplicon Panels includes adapter trimming, alignment and post-alignment primer soft-clipping using the panel-specific masterfile and the Primerclip pre-compiled binary supplied by Swift Biosciences, followed by target enrichment evaluation and variant calling using the appropriate amplicon panel's target BED file. Below is an illustration of an example pipeline for data analysis of Accel-Amplicon Panels.



*Figure 2. Recommended Bioinformatics Pipeline for Accel-Amplicon Panels. Post-alignment soft clipping of primer sequences using the Primerclip tool (highlighted in purple) is the mandatory step when performing analysis on sequencing results from Accel-Amplicon Panels, and all other steps may be performed as listed here or the user's preferred pipeline.*

## Adapter Trimming

Primerclip is designed to be used after the Illumina adapter trimming. Adapter trimming could be done on- or off-instrument. For on-instrument use, please ensure that adapter-trimming setting is enabled while setting up the sequencing run. Alternatively, this step could be performed bioinformatically prior to data analysis.

## Alignment

After adapter sequences have been trimmed, an alignment tool such as BWA (Li H. and Durbin R. 2010) should be utilized to align sequenced reads to the reference genome.

## Primer Trimming

After performing the alignment step, Primerclip should be used to trim primer sequences, unique to the Accel-Amplicon panels. To execute Primerclip, use the panel-specific masterfile, provided by Swift, and your SAM alignment file as input to the tool. The output will be a primer-clipped SAM file, which can be converted to BAM and used for regular downstream analysis. Please refer to Primerclip's README document for comprehensive instructions to install and use this tool.

Please visit <http://github.com/swiftbiosciences/primerclip> to download Primerclip, panel-specific BED(s) and Masterfile.

After primer trimming is complete, use Picard "AddOrReplaceReadGroups" tool (Picard, Broad Institute) to Sort and Add ReadGroups to the SAM file. The processed SAM file could then be converted to BAM for downstream analysis.

## Performance Metrics

To assess data quality, calculate percent on-target and coverage uniformity using the following equations.

### On-Target

The on-target metric is a measure of the proportion of the total aligned reads sequenced from an NGS library are aligned to the intended target regions on the reference genome. On-target is calculated as ratio of "total bases that map/align to the intended target region" and "total aligned bases".

$$OT\% = \frac{B^{OT}}{B^{Tot}} * 100.0$$

Where,

$B^{OT}$  = number of aligned bases which coincide with or are adjacent to a target base

$B^{Tot}$  = total number of aligned bases

### Coverage Uniformity

Coverage uniformity is calculated as ratio of "number of target bases that have coverage at or above 20%" and "total number of target bases."

$$CU\% = \frac{B^{0.2\mu}}{B^{OT}} * 100.0$$

Where,

$B^{0.2\mu}$  = total number of bases with coverage  $\geq 20\%$  of mean

$B^{OT}$  = total number of target bases

## Additional Consideration

Please note that while use of picard tools MarkDuplicates is a common quality control step to detect low-complexity libraries, the use of MarkDuplicates is not applicable to targeted panels which use PCR-based target enrichment methods to create targeted sequencing libraries, such as Accel-Amplicon panels. These targeted panels contain high numbers of identical (particularly with regard to alignment start position) library fragments by design, and we recommend MarkDuplicates not be included in the workflow for analyses of Accel-Amplicon libraries.

## Panels Containing Sample\_ID as a Spike-in

**IMPORTANT:** Consider the following information in order to achieve the most accurate sequencing results from libraries prepared from Accel-Amplicon 56G Oncology Panel v2 and Accel-Amplicon 57G Pan-Cancer Profiling Panels:

- **On-target Percentage (% OT):** Use the combined BED file (available at [www.swiftbiosci.com](http://www.swiftbiosci.com) or by emailing us at [TechSupport@swiftbiosci.com](mailto:TechSupport@swiftbiosci.com)) to calculate overall % OT for the entire panel. Calculating % OT for the entire panel based on a BED file specific to main panel content only or Sample\_ID only will result in incorrect assessment of % OT, as the reads from the excluded panel will be incorrectly reported as off-target.
- **Coverage Uniformity Percentage (% CU):** Use the BED file specific to main panel content or Sample\_ID to calculate % CU for each subset of the data. By design, Sample\_ID targets in the combined panel have lower coverage than main panel targets because they are for germline variants; therefore, analyzing % CU across the combined panel will inaccurately reflect Sample\_ID targets as low coverage dropouts.

### For Other Accel-Amplicon Panels

Use the panel-specific BED file, supplied with each panel, to calculate % OT and % CU for the sequenced reads.

## Suggestions for Variant Calling

You may use various publicly available tools such as LoFreq (Genome Institute of Singapore), LoFreq-somatic, Varscan, FreeBayes, Strelka, or other commonly used tools such as Mutect and HaplotypeCaller (GATK, Broad Institute), to call variants in your samples. A Q-Seq HDx™ quantitative multiplex DNA reference standard (HD701) is included in your kit for use as a positive control. This standard contains precise allelic frequencies for major oncology targets verified by digital PCR and may be utilized to assess your pipeline's accuracy of variant calling. For more information, please refer to the section "DNA Standards for Accel-Amplicon Panels" in your Accel-Amplicon Panel Protocol.



### Swift Biosciences, Inc.

674 S. Wagner Road, Suite 100 • Ann Arbor, MI 48103 • 734.330.2568 • [www.swiftbiosci.com](http://www.swiftbiosci.com)

© 2018, Swift Biosciences, Inc. The Swift logo, Primerclip, Accel-Amplicon, and Accel-Amplicon Plus are trademarks of Swift Biosciences. This product is for Research Use Only. Not for use in diagnostic procedures. Illumina is a registered trademark of Illumina, Inc. Q-Seq HDx is a trademark of Horizon Discovery Group plc. 18-2047, 07/18