

Accel-Amplicon™ Panels: Bioinformatics Guidelines

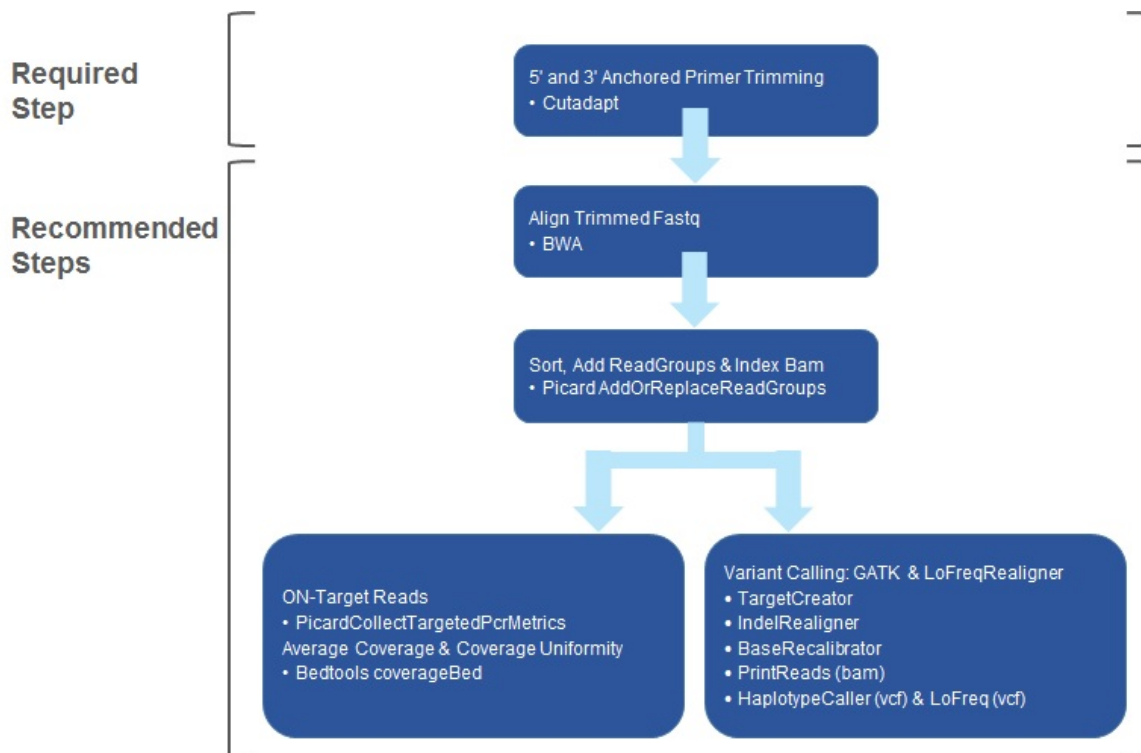
The Accel-Amplicon Panels listed below are designed with overlapping amplicons to allow for contiguous regions of coverage in a single-tube format. Therefore, synthetic primer sequences will be encountered both at the beginning and end of some reads, which must be trimmed during the first step in data analysis.

- **Accel-Amplicon 56G Oncology Panel v2**
- **Accel-Amplicon BRCA1 and BRCA2 Panel**
- **Accel-Amplicon BRCA1, BRCA2, and PALB2 Panel**
- **Accel-Amplicon Comprehensive TP53 Panel**
- **Accel-Amplicon EGFR Pathway Panel**
- **Accel-Amplicon custom panel** (Trimming of custom panels is panel-dependent. Refer to the Protocol to determine if your custom panel requires trimming.)

Pipeline Overview

The suggested data analysis pipeline for Accel-Amplicon Panels begins with mandatory primer trimming using the primer FASTA files and Cutadapt script supplied by Swift Biosciences followed by target enrichment evaluation and variant calling using amplicon panel BED files. **NOTE:** As described in your Accel-Amplicon Panel Protocol, please ensure that adapter trimming is enabled while setting up the sequencing run. Failure to trim adapter sequences will result in incorrect primer trimming and will lead to inaccurate variant calling. To overcome this issue, enable automatic trimming by the sequencer software or use the primer trimming example script supplied by Swift Biosciences, which includes adapter trimming by Trimmomatic as a redundancy check to trim adapter sequences.

Figure 1. Recommended Pipeline for Accel-Amplicon Panels



NOTE: Please note that 5' and 3' anchored primer trimming using Cutadapt is required when performing analysis on sequencing results from Accel-Amplicon Panels, and all subsequent steps may be performed with the tools listed here or the user's preferred pipeline.

Primer Trimming

To process data obtained with Accel-Amplicon Panels, primer trimming is required as the first step following adapter trimming, which may be achieved on or off the instrument as described in your Accel-Amplicon Protocol.

Suggested Dependency List of Tools

The following list is an example of tools required to run this pipeline. However, other tools may be utilized to perform similar analysis. **NOTE:** If you are new to this process and would like more information to get started, please contact us for example scripts to run complete data analysis pipelines.

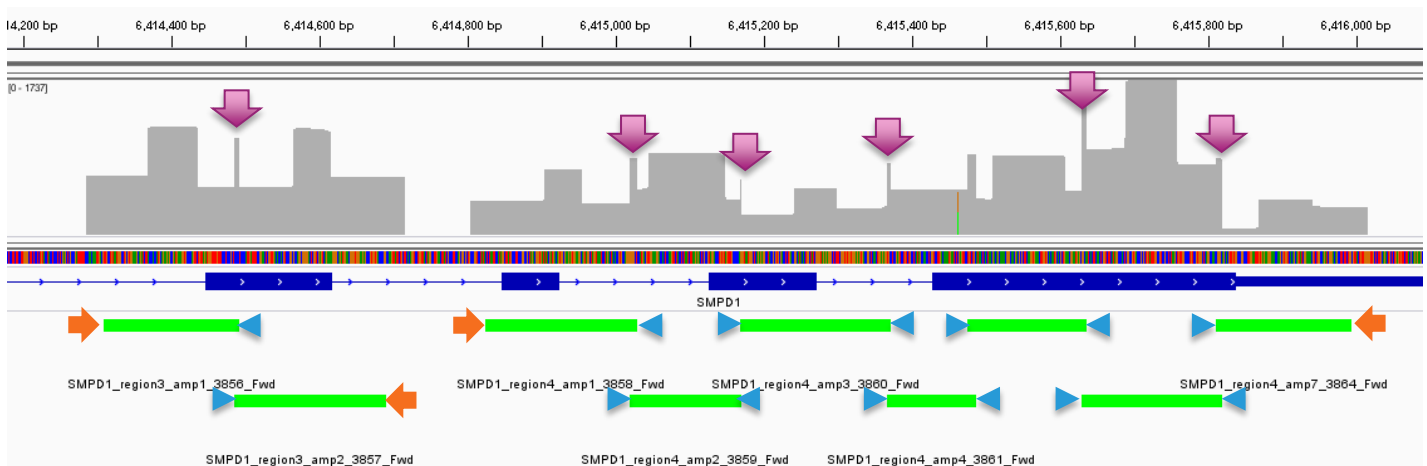
- **GNU Linux/BSD/*nix system** – computer operating system (Unix Operating System)
- **GNU parallel** – shell tool for executing jobs in parallel (GNU Operating System)
- **Trimmomatic (requires Java)** – tool to trim Illumina® adapter sequences (USADELLAB and Java)
- **Cutadapt** – tool to trim Swift Biosciences primer sequences (Python)

The latest version of Cutadapt (Martin, et al. 2011) is recommended for anchored 5' and 3' trimming of primer sequences from paired-end (PE) FASTQ files. If Cutadapt is being used via Galaxy, edits to the wrapper file may be required to enable reading multiple primer sequences from a FASTA file. Primer sequences in the primer trim files have special characters before the 5' and after the 3' bases to indicate the start and end of the sequence to be trimmed. Due to the recommended 2x151 read length and short amplicon design, synthetic primer sequences will be encountered both at the beginning and end of reads and must be trimmed. However, since many Accel-Amplicon Panels have overlapping amplicons for comprehensive coverage, trimming must not be performed when the sequence is present internally (within a read), as each primer sequence will be separately covered by a different amplicon.

Tiled Amplicons

In Accel-Amplicon Panels with overlapping amplicons, amplicons overlap to cover the same region (purple arrows in Figure 2). Therefore, the synthetic primer sequences indicated as blue arrows in Figure 2 should be removed from the sequencing data before performing variant calling. Since the target BED file only contains the “target” regions (sequences between 2 primers, green in Figure 2), the orange primer sequences do not have to be trimmed from the data. **NOTE:** Although the orange sequences are primer sequences, they are not included in the target BED file; therefore, they do not have to be trimmed.

Figure 2. Representation of Tiled Amplicons in Accel-Amplicon Panels



Primer Trimming Example Script Using Cutadapt

Swift Biosciences supplies an example script for running Cutadapt to perform primer trimming. Please contact TechSupport@swiftbiosci.com or visit us at www.swiftbiosci.com to download this script along with the 5' and 3' primer trimming files appropriate for your panel. The Cutadapt script is written to take advantage of multiple CPUs.

These commands may be directly included in your workflow once the primer trim file names are updated to match the panel purchased. Please ensure these files are in the same directory from which the workflow is executed. The script is designed to only trim synthetic primer sequences at the ends of the reads to preserve the fidelity of variant calling. Please contact us with any questions.

Alignment

Once primers 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. Use Picard® “AddOrReplaceReadGroups” tool (Picard, Broad Institute) to Sort, Add ReadGroups & Index Bam during analysis.

On Target and Coverage Uniformity Analysis

For Accel-Amplicon 56G Oncology Panel v2 (Panels Containing Sample_ID Spiked-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:

- **On Target Percentage (% OT): Combined BED file:** After whole genome alignment, use the combined BED file containing 56G and Sample ID targets (available at www.swiftbiosci.com or by emailing TechSupport@swiftbiosci.com) to calculate overall % OT for the entire panel. Calculating % OT based on a BED file specific to 56G only or Sample_ID only will result in incorrect assessment of % OT, as the reads from the excluded panel will incorrectly contribute to off-target.
- **Coverage Uniformity Percentage (% CU): Per sub-panel BED file:** Use the BED file specific to 56G 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 56G 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 Panels Protocol.



Swift Biosciences, Inc.
58 Parkland Plaza, Suite 100 • Ann Arbor, MI 48103 • 734.330.2568 • www.swiftbiosci.com

© 2016, Swift Biosciences, Inc. The Swift logo and Accel-Amplicon 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. 16-1112, 09/16