# Approaches for trimming primer sequences from amplicon NGS reads

**Aled Jones[1], Mikel Valganon[1], Joo Wook Ahn[2]**

1 - Genetics Laboratories, Viapath, London 2 - Genetics Laboratories, Guy's and St Thomas' NHS Foundation Trust, London

## Introduction

- Amplicon NGS assays use PCR to target select regions of interest.
- Sequencing reads from amplicon NGS assays include both the regions of interest and primer sequences.
- Where amplicons overlap, sequenced bases will be derived from both sample DNA and primer DNA.
- Overlapping amplicons can produce misleading read depth values.
- Variants within primer sequences may prevent the affected allele being amplified, leading to a false negative result.
- Variant callers struggle with variants (particularly insertions/deletions) at the extreme ends of reads.
- Primer trimming can resolve some of these issues.

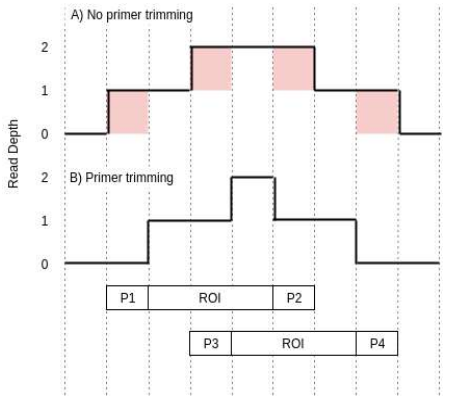


**Figure 1** – Without primer trimming (A), primers from overlapping amplicons inflate the reported read depth (shaded red areas) with unsequenced bases.

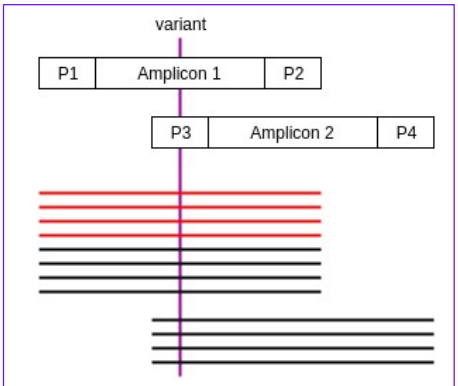


**Figure 2** – If a variant (which prevents amplification of that allele) falls within the primer of an overlapping amplicon, the allele frequency can be underestimated. In this example a variant is seen in amplicon 1 at 50% (4/8 reads), but without trimming primer 3 (P3) the allele frequency is underestimated at 33% (4/12 reads).

**Table 1 – Comparison of pre and post-alignment primer trimming**

| | Pre-alignment | Post-alignment |
|---|---|---|
| Trim based on | Number of bases, quality | Genomic coordinates |
| Primer sequences used in read mapping | No | Yes |
| Number of tools | More | 2 |
| Example tools | Trimmomatic, cutadapt, Flexbar | BAMclipper, Primerclip |

## Primer trimming

Primer trimming is necessary for accurate read depth calculations (Figures 1 and 2) and can be performed pre or post-alignment (Table1).

## Pre or post-alignment

Most tools perform pre-alignment trimming by removing a given number of bases from the end of reads (so incompatible with using different length primers) and then aligning them. Trimming primers at this stage prevents their use as "anchors" to a reference genome, which can be problematic for variants occurring near primer sequences.

Post-alignment trimming allows primers to first anchor reads before genomic coordinates are used to accurately trim primers.

## Our solution

We have implemented a post-alignment primer trimming using BAMclipper (Figure 3):

- This detects indels at the primer-amplicon boundary (Figure 4)
- Doesn't trim primers mid-read or include primer sequences in read depth (Figure 5)

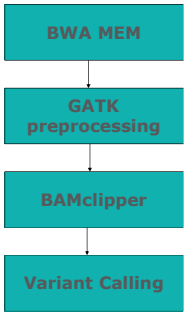


**Figure 3** – Post-alignment primer trimming is performed by BAMclipper.

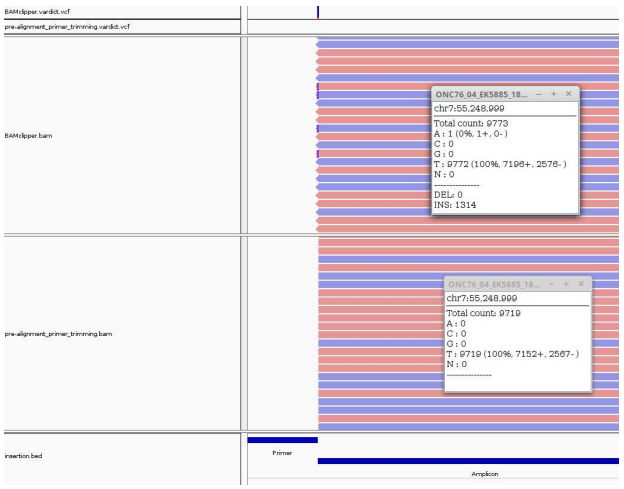


**Figure 4** – A 9bp insertion at the first base of the region of interest is detected when post-alignment primer trimming is performed (top BAM track) but not with pre-alignment primer trimming (bottom BAM track).

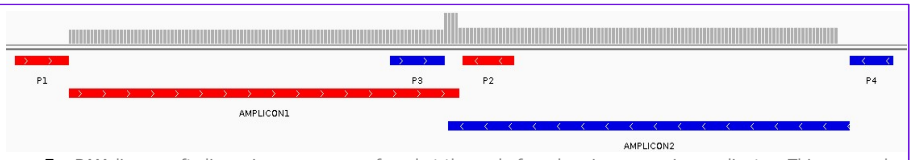


**Figure 5** – BAMclipper soft clips primer sequences found at the end of reads using genomic coordinates. This approach doesn't include primers in the coverage (Figure 1).

## Recommendations

- Trim primers post-alignment from amplicon NGS data
- BAMclipper works well as it uses genomic coordinates to accurately remove (soft clip) primer sequences from the ends of reads.

https://github.com/tommyau/bamclipper http://doi.org/f98hpw