

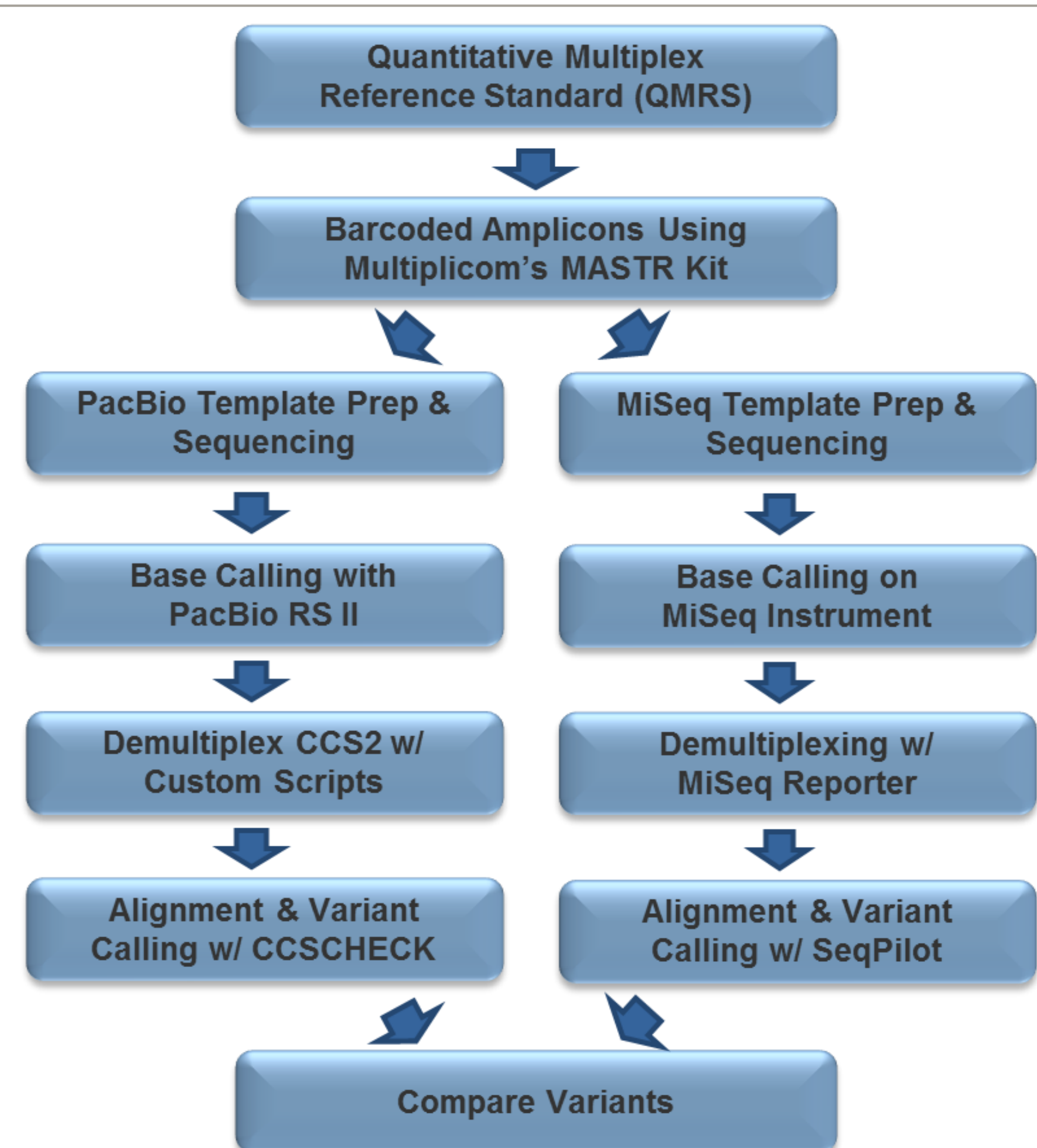


## Introduction

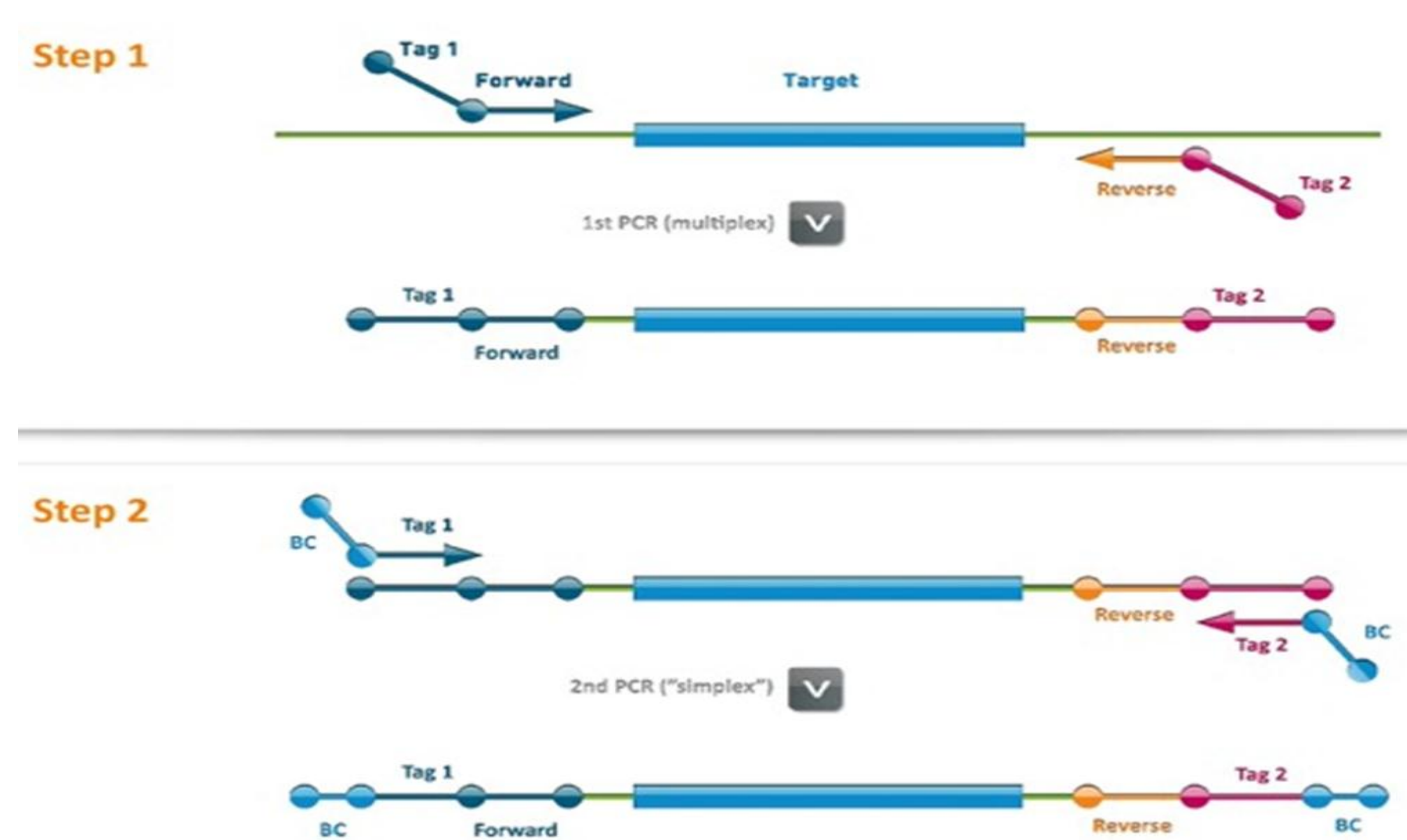
The sensitivity, speed, and reduced cost associated with Next-Generation Sequencing (NGS) technologies have made them indispensable for the molecular profiling of cancer samples. For effective use, it is critical that the NGS methods are not only robust, but can also accurately detect low-frequency somatic mutations. Single Molecule, Real-Time (SMRT) Sequencing offers several advantages, including the ability to sequence single molecules with very high accuracy using the circular consensus sequencing (CCS) approach. To develop and test any NGS-based molecular assay it is necessary to have genetically defined, human genomic reference standards.

Here we characterize SMRT Sequencing for the detection of low-frequency somatic variants using the Quantitative Multiplex Reference Standard (QMRS) from Horizon Discovery, combined with amplification of the variants using the Multiplicom Tumor Hotspot MASTR™ Plus assay. The Horizon Discovery reference sample contains precise allelic frequencies from 1% to 24.5% for major oncology targets verified using digital PCR. It recapitulates the complexity of tumor composition and serves as a well-characterized control. The control sample was amplified using Multiplicom's Tumor Hotspot MASTR™ Plus assay that targets 252 amplicons (121-254 bp) from 26 relevant cancer genes, including all 11 variants in the QMRS sample. The amplicons were sequenced and analyzed using SMRT Sequencing to identify the variants and determine the observed frequency. The random error profile and high accuracy CCS reads make it possible to accurately detect low-frequency somatic variants.

## Workflow

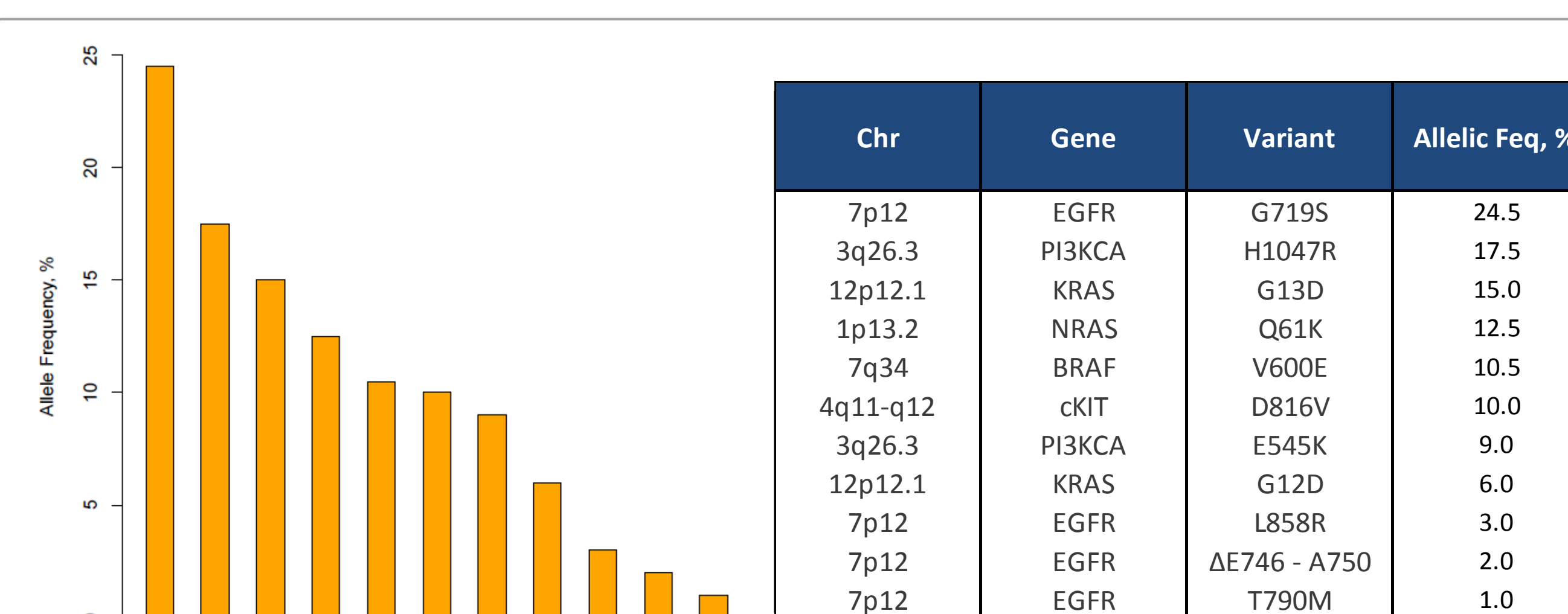


**Figure 1** The Quantitative Multiplex Reference Sample was amplified, barcoded and sequenced on either the PacBio RS II or MiSeq instrument. The 11 verified mutation frequencies were compared.



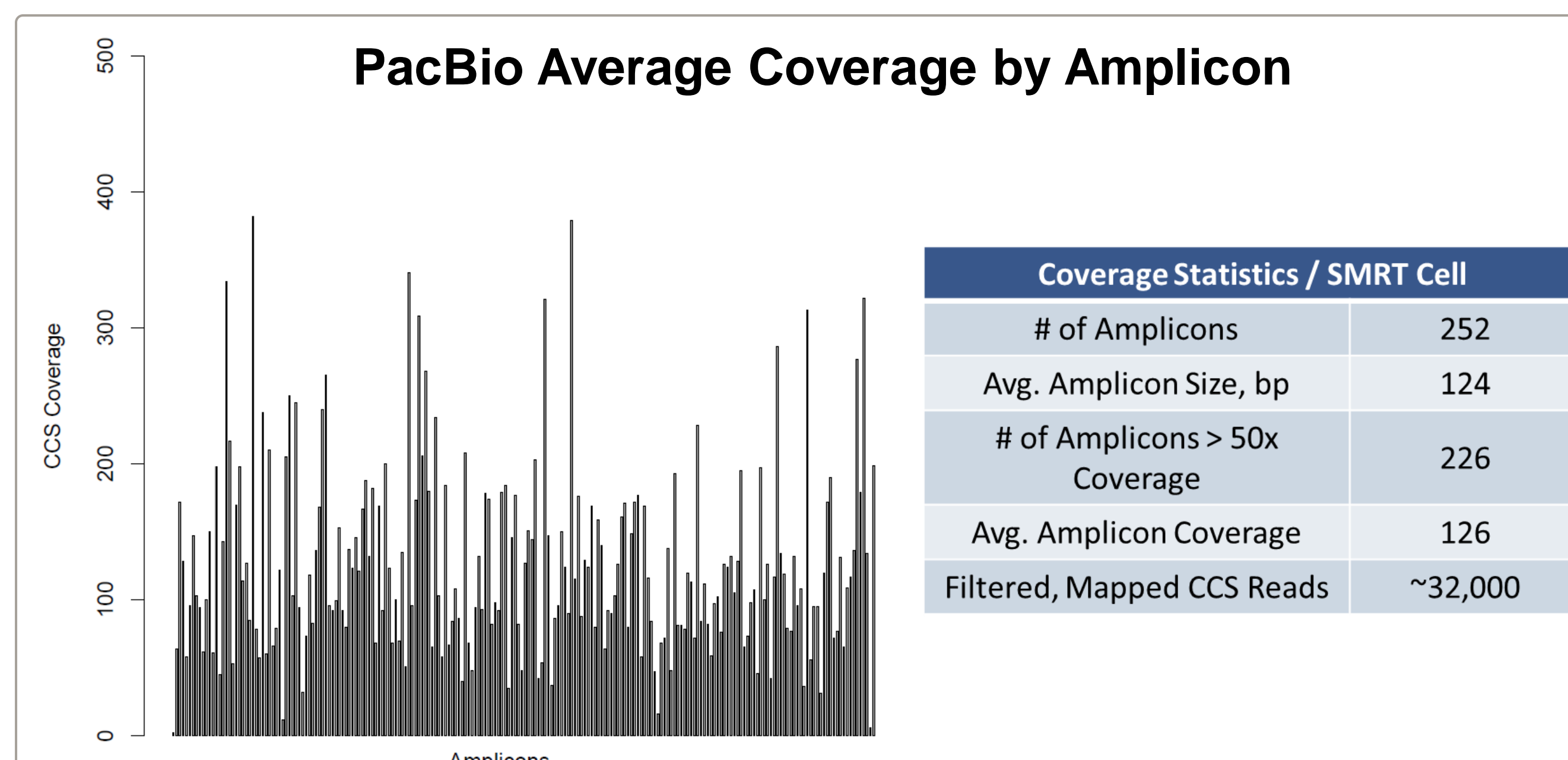
**Figure 2** The MASTR™ assay employs a 2-step PCR protocol enabling specific amplification of the regions of interest (Step 1) followed by the incorporation of molecular barcodes (Step 2) to unambiguously link each read to the originating sample.

## Verified Variants in the Control Sample



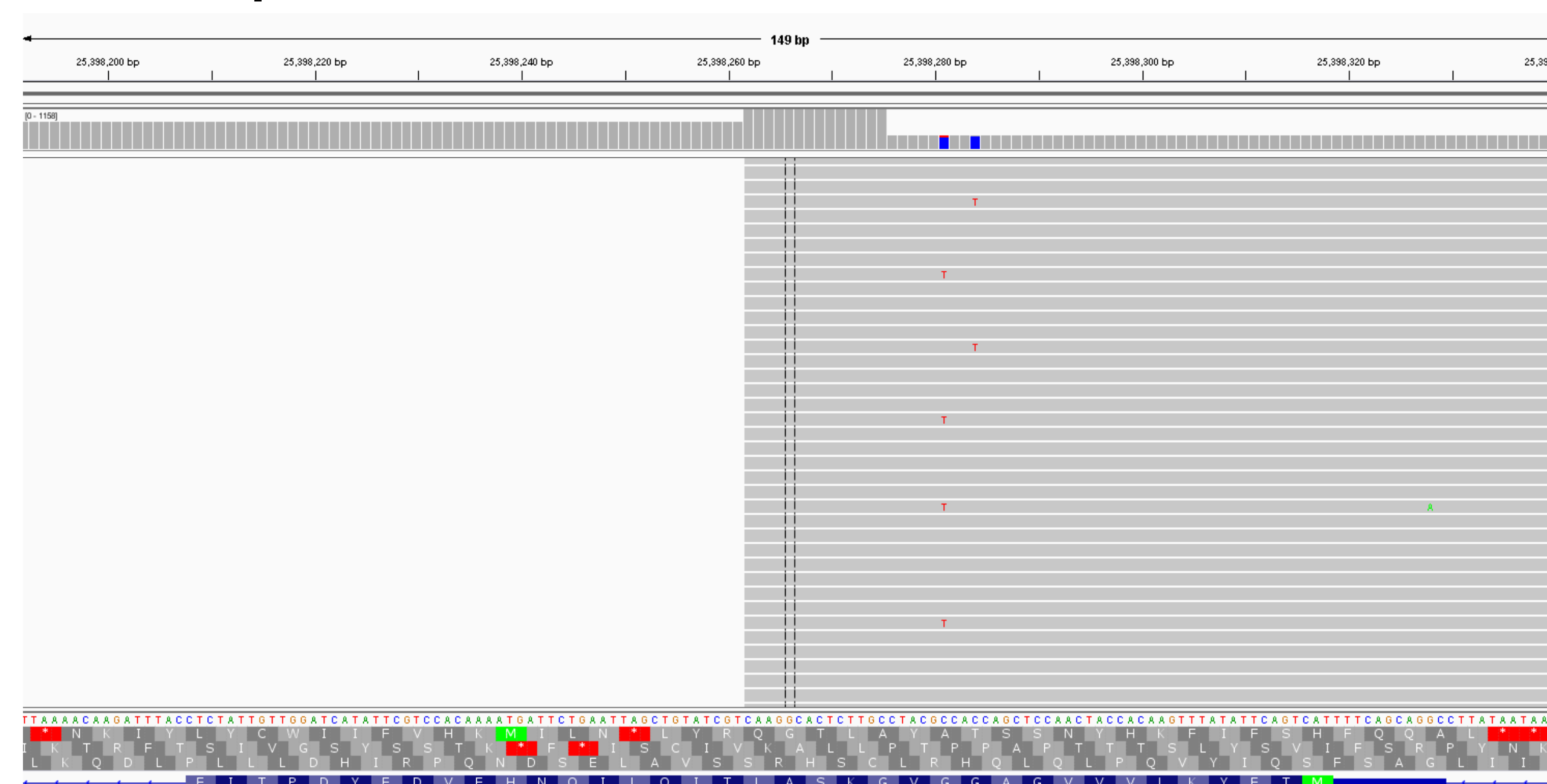
**Figure 3** Allelic frequency of the 11 verified variants in the Horizon Discovery Quantitative Multiplex Reference Standard.

## PacBio Coverage



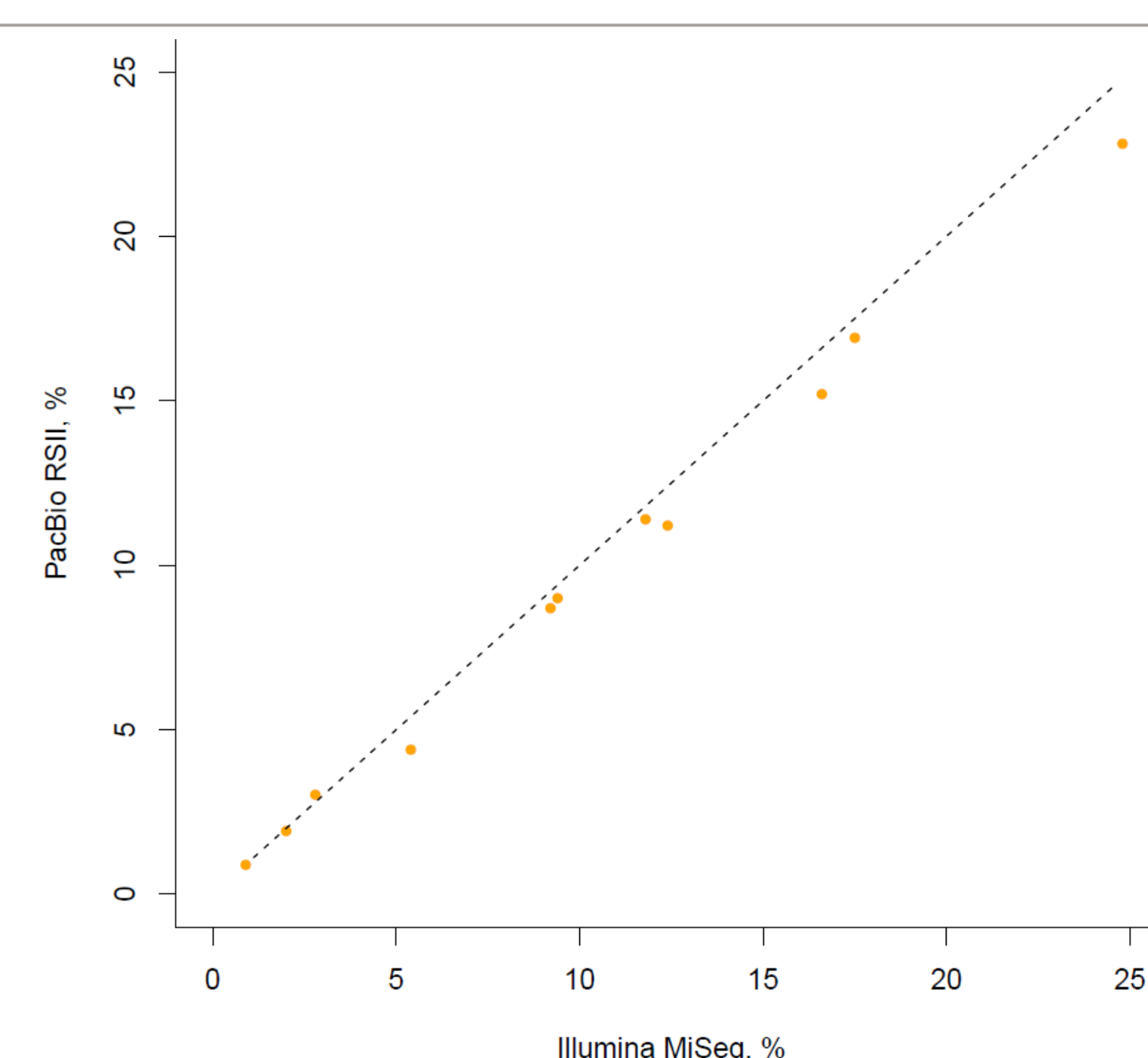
**Figure 4** The Multiplicom Tumor Hotspot MASTR™ Plus assay generates 252 amplicons. All amplicons were sequenced (shortest to longest, left to right), but analysis focused only on amplicons containing the 11 verified variants.

## Representative PacBio Reads from KRAS Exon 2



**Figure 5** Representative PacBio reads from exon 2 of KRAS. The two verified KRAS mutations (G12D and G13D) are clearly observed.

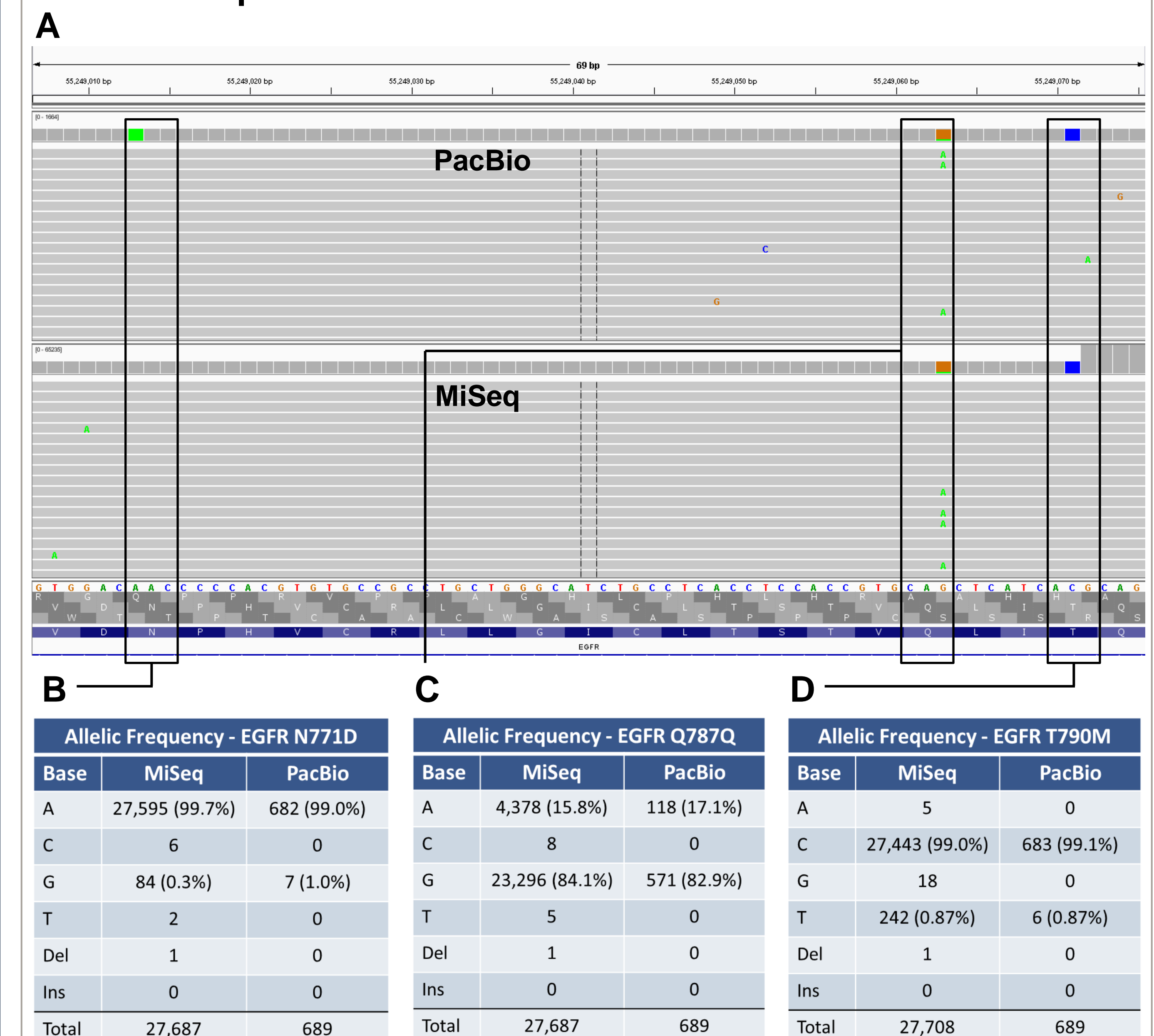
## Comparison of the Verified Variants



**Figure 6** A comparison of the observed allelic frequencies of the 11 verified variants in the QMRS sample.

## Comparison of Variants

### Comparison of Observed Variants in EGFR Exon 20



**Figure 7** A comparison of three observed variants with frequencies greater than 0.8% in exon 20 of EGFR (Panel A). Variants B and C have not been verified in the QMRS sample. Variant D is the T790M mutation with a verified allelic frequency of 1%. Even though the PacBio data has 40x less coverage, the observed allelic frequencies from each platform have strong concordance (Panels B-D). The high accuracy of the PacBio CCS reads only generates base calls for the reference or the observed variant.

## Summary and Resources

### Summary:

- This work describes a complete PacBio workflow for amplifying, barcoding, sequencing and calling variants from a DNA reference sample
- SMRT Sequencing identifies all 11 verified variants, ranging from 1% - 24.5%, in the reference sample
- The PacBio allelic frequencies of the 11 variants are concordant with the MiSeq data, using significantly fewer PacBio reads to produce the same result
- PacBio CCS reads are highly accurate and can detect somatic variants down to 1%

### Multiplicom Info Available Here:

<http://www.multiplicom.com/product/tumor-hotspot-mastr-plus>

### PacBio Info Available Here:

<http://www.pacb.com/applications/targeted-sequencing/>

### Data & Analysis Info Available Here:

<https://github.com/PacificBiosciences/pbccs>

<https://github.com/PacificBiosciences/ccscheck>

<https://github.com/PacificBiosciences/PacBioFileFormats/wiki/BAM-recipes#splitting>

### Acknowledgements

The authors would like to thank Joachim De Schrijver and Nigel Delaney for their assistance with the data analysis and scripts.