Highly sensitive and cost-effective detection of somatic cancer variants using single-molecule, real-time sequencing

Anand Sethuraman[1], Kevin Eng[1], Primo Baybayan[1], Lien Heyrman[2], Jurgen Del Favero[2], Catherine Huang[3], Bharathi Anakella[3] and Steve Kujawa[1]
[1] Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025
[2] Multiplicom N. V., Galileilaan 18, 2845 Niel, Belgium
[3] SeraCare Life Sciences, 37 Birch Street, Milford, MA 01757

Introduction

Next-Generation Sequencing (NGS) technologies allow for molecular profiling of cancer samples with high sensitivity and speed at reduced cost. For efficient profiling of cancer samples, it is important that the NGS methods used are not only robust, but capable of accurately detecting low-frequency somatic mutations. Single Molecule, Real-Time (SMRT®) Sequencing offers several advantages, including the ability to sequence single molecules with very high accuracy (>QV40) using the circular consensus sequencing (CCS) approach. The availability of genetically defined, human genomic reference standards provides an industry standard for the development and quality control of molecular assays for studying cancer variants. Here we characterize SMRT Sequencing for the detection of low-frequency somatic variants using the Quantitative Multiplex DNA Reference Standards from Horizon Discovery, combined with amplification of the variants using the Multiplicom Tumor Hotspot MASTR Plus assay.

First, we sequenced a reference standard containing precise allelic frequencies from 1% to 24.5% for major oncology targets verified using digital PCR. This reference material recapitulates the complexity of tumor composition and serves as a well-characterized control. The control sample was amplified using the Multiplicom Tumor Hotspot MASTR Plus assay that targets 252 amplicons (121-254 bp) from 26 relevant cancer genes, which includes all 11 variants in the control sample.

Next, we sequenced control samples prepared by SeraCare Life Sciences, which contained a defined mutation at allelic frequencies from 10% down to 0.1%. The wild type and mutant amplicons were serially diluted, 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
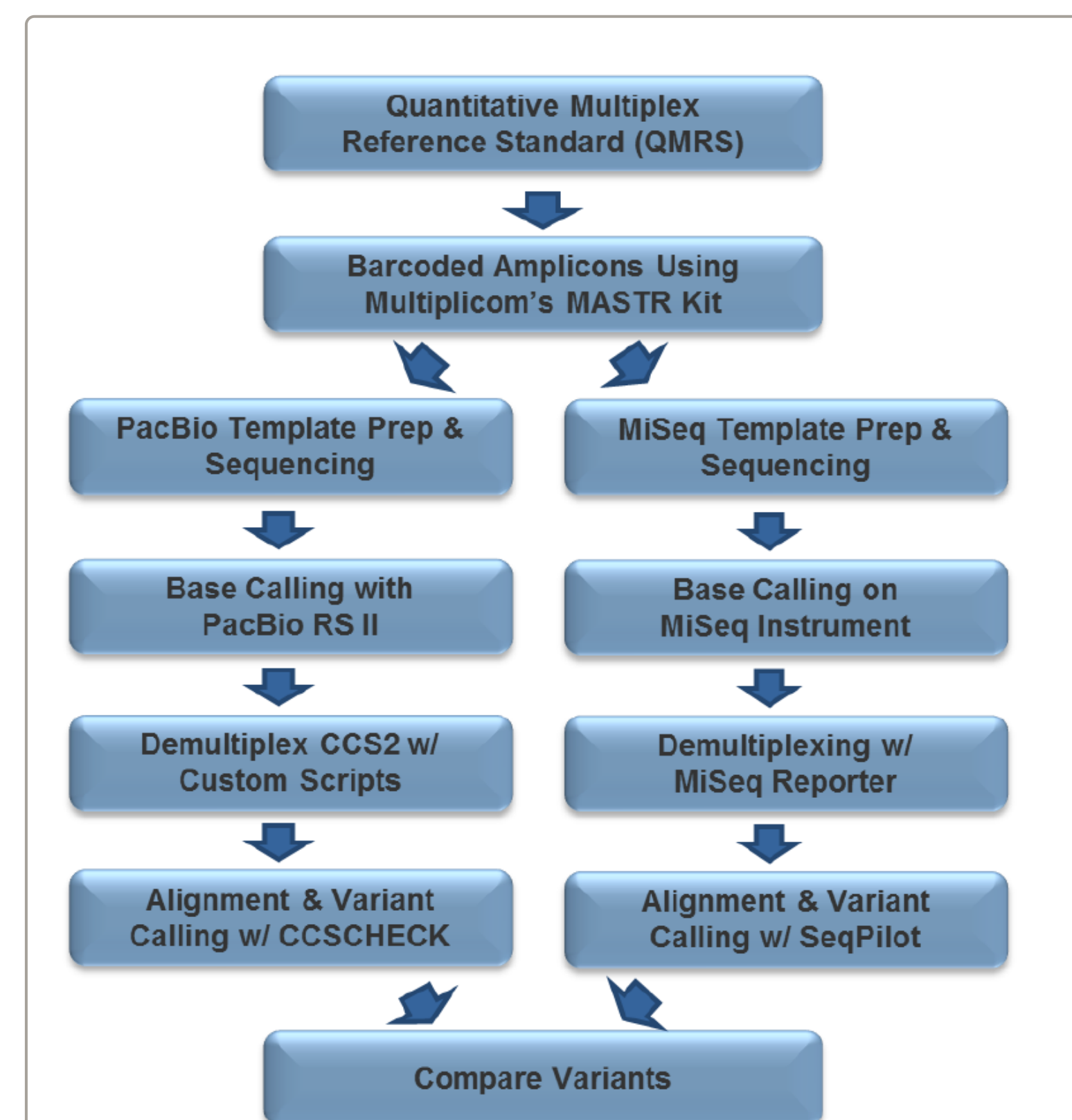
Figure 1 The Quantitative Multiplex Reference Sample was amplified, barcoded and sequenced on either the PacBio RS II or MiSeq instrument. The 11 verified mutation frequencies were compared.
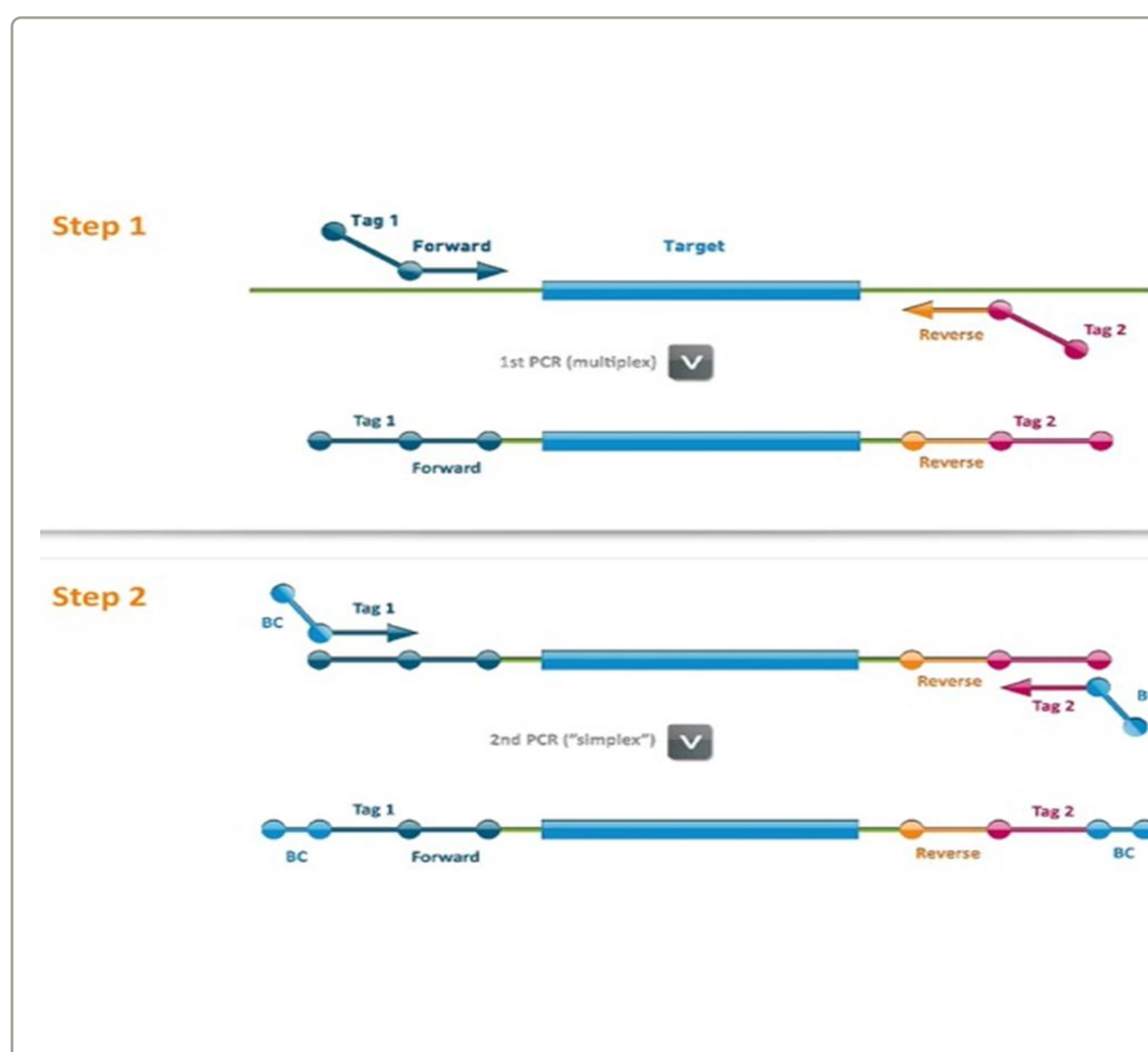
Figure 2 The MASTR™ assay employs a 2-step PCR protocol enabling specific amplification of the regions of interest (Step 1) followed by the incorporation of molecular barcodes (Step 2) to unambiguously link each read to the originating sample.
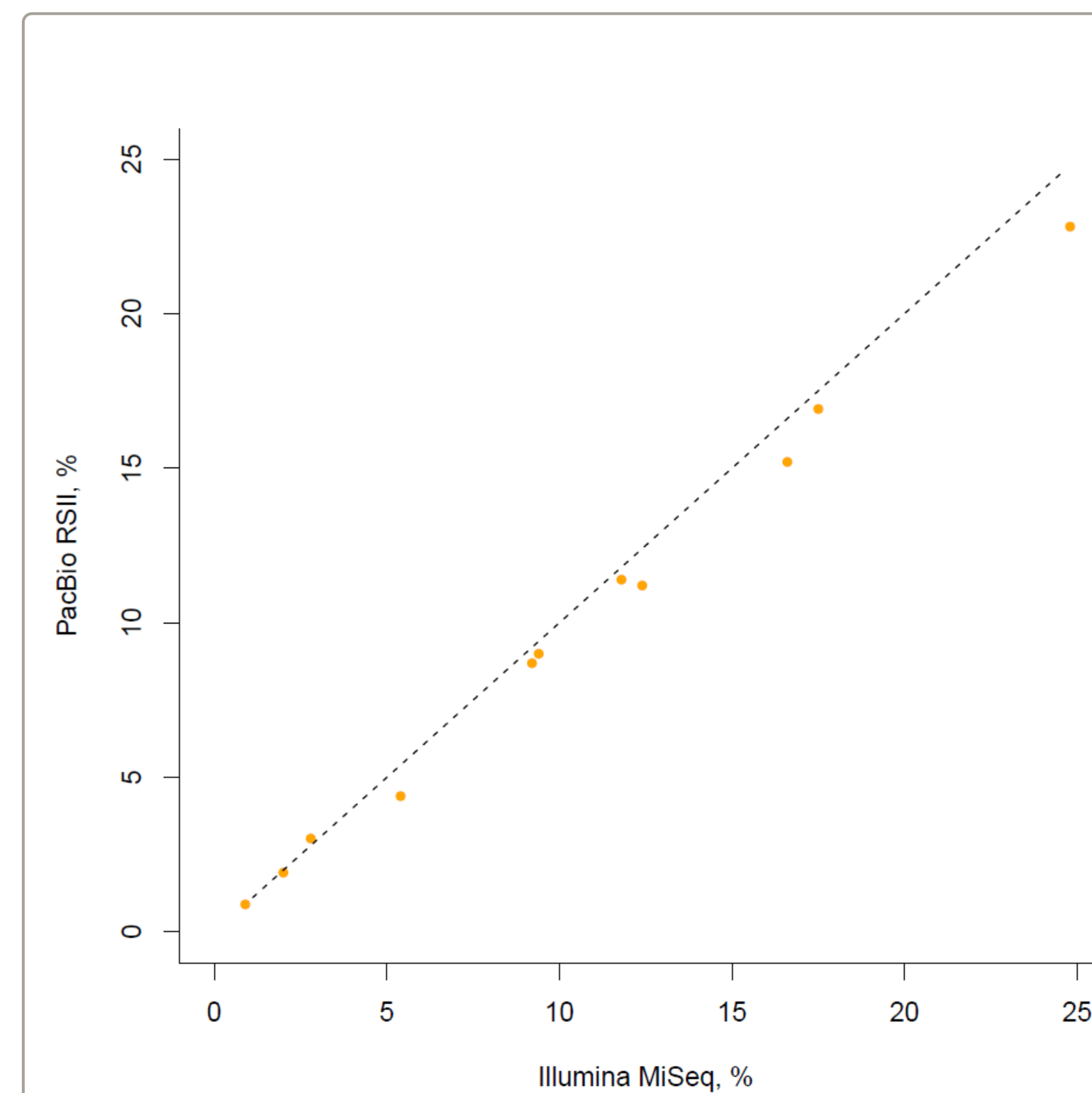
Verified Variants in the Control Sample
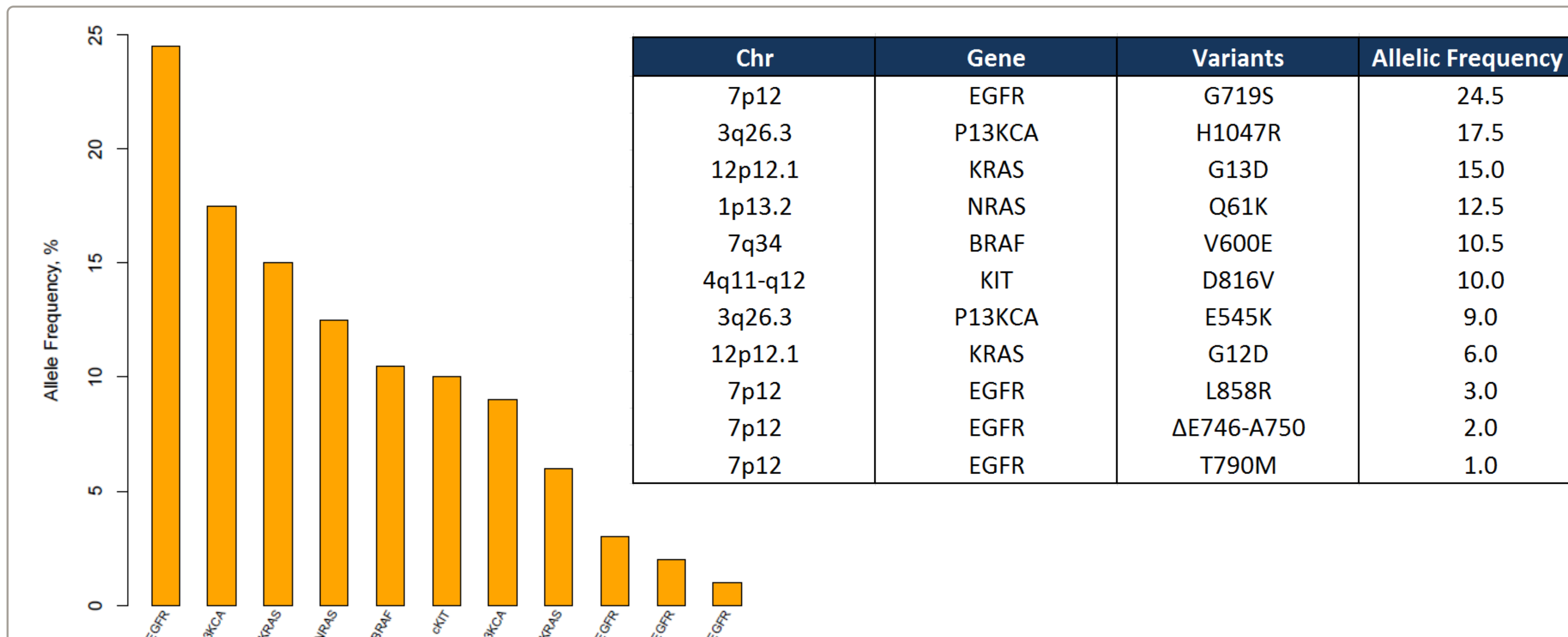
Figure 3 Allelic frequency of the 11 verified variants in the Horizon Discovery Quantitative Multiplex Reference Standard.

Chr	Gene	Variants	Allelic Frequency %
7p12	EGFR	G719S	24.5
3q26.3	P13KCA	H1047R	17.5
12p12.1	KRAS	G13D	15.0
1p13.2	NRAS	Q61K	12.5
7q34	BRAF	V600E	10.5
4q11-q12	KIT	D816V	10.0
3q26.3	P13KCA	E545K	9.0
12p12.1	KRAS	G12D	6.0
7p12	EGFR	L858R	3.0
7p12	EGFR	ΔE746-A750	2.0
7p12	EGFR	T790M	1.0

Table 1 Gene names and their variants in the Horizon Discovery Quantitative Multiplex Reference Standard.

Comparison of the Verified Variants

Figure 4 Representative PacBio reads from exon 2 of KRAS. The two verified KRAS mutations (G12D and G13D) are clearly observed.

Figure 5 A comparison of the observed allelic frequencies of the 11 verified variants in the QMRS sample.

Minor Variants

SeraCare provided constructs containing a single mutation in a partial coding region of AKT1. PCR-amplified wild-type (WT) and mutant (MUT) constructs were mixed and serial dilutions made down to 0.1% allelic frequency. Independent libraries were made for each dilution point and sequenced using the PacBio RS II instrument.
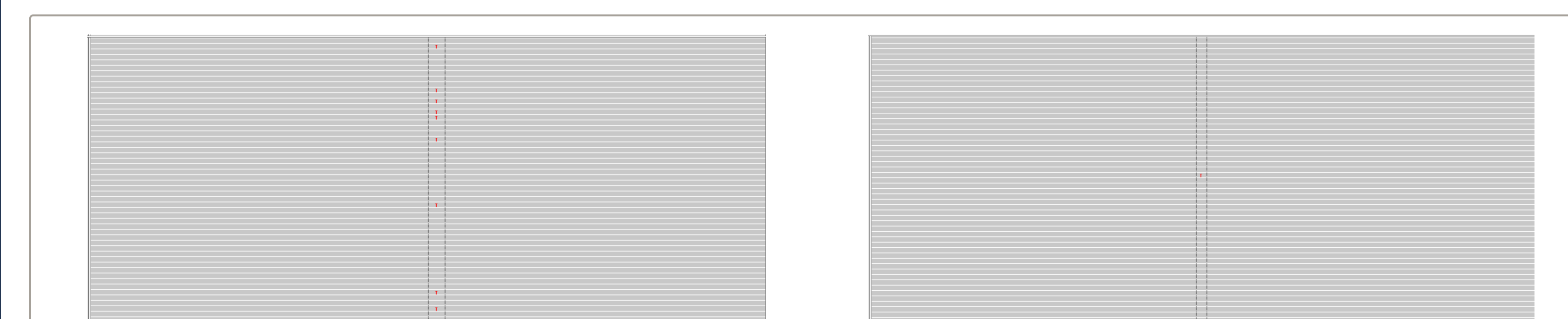

Figure 6. Pairwise alignment of the WT and MUT AKT1 sequence showing the expected SNP
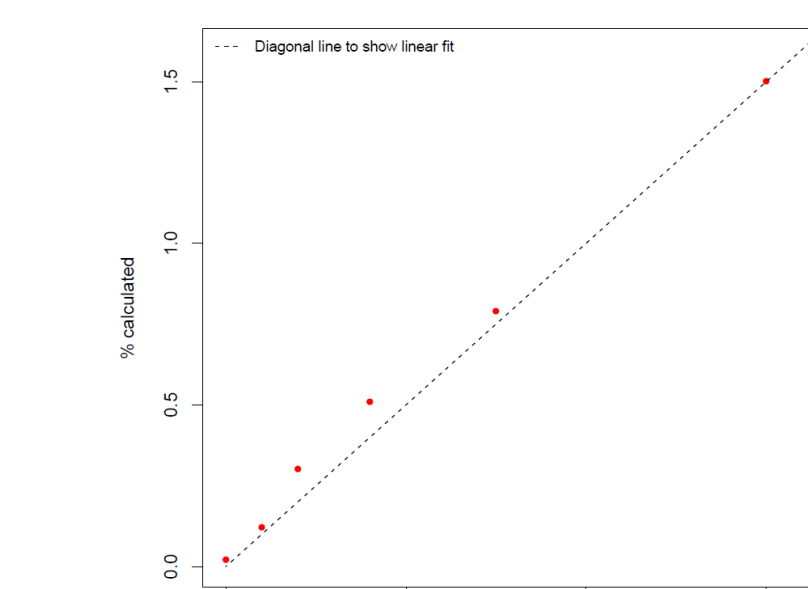
Figure 7. Serial dilution experiment results showing sensitivity of mutations detected down to 0.1%
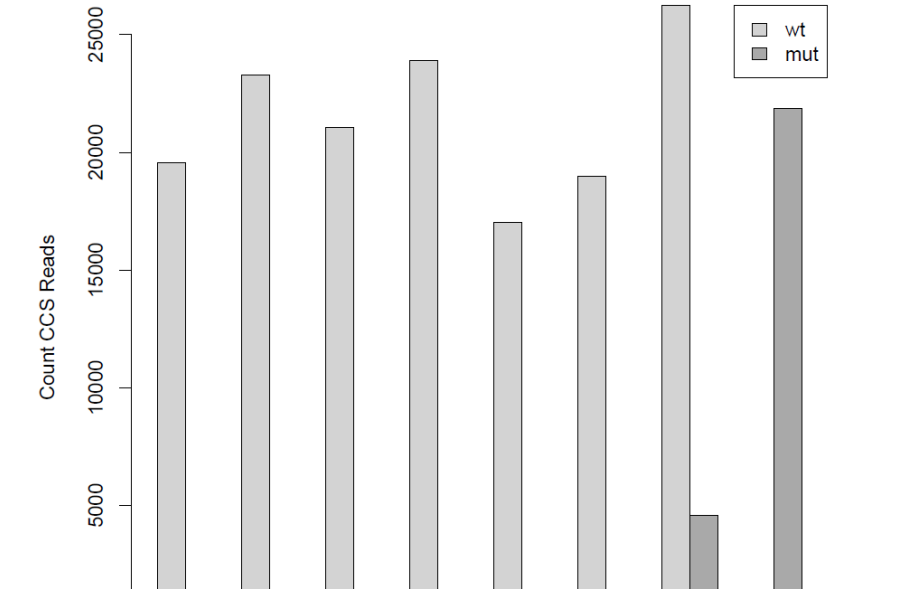
Figure 8. CCS Reads Histogram

Figure 9. IGV Alignments showing the AKT1 SNP in the 15% (left) and 1.5% (right) mix

Summary and Resources

Summary:
- This work describes a complete PacBio workflow for amplifying, barcoding, sequencing and calling variants from DNA reference samples
- SMRT Sequencing successfully identified 11 verified variants (1% - 24.5%) and the data are concordant with the MiSeq data
- PacBio CCS reads are highly accurate and can detect minor variants down to 0.1%

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 Nigel Delaney for his assistance with the data analysis and scripts.