Supplementary Information for
**AIRVF: an integrated filtering toolbox for so-matic variant calling in Ion Torrent sequencing**

Sunguk Shin1, Hanna Lee1, Hyeonju Son1,2, Soonmyung Paik1 and Sangwoo Kim1,\*

1Severance Biomedical Science Institute and 2Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul 03722, South Korea

Supplementary Methods

M1. Materials and sequencing data generation

The AcroMetrix Oncology Hotspot Control (AOHC; Life Technologies, Seoul, Korea) samples were sequenced in Ion Torrent 318 v2 chip with the Ion PGMTM Template OT2 Kit (Life Technologies, Seoul, Korea) to obtain data for developing AIRVF and find the optimal parameters. The same material has been sequenced one more time independently to obtain data for validation. The AOHC standard material contained 555 true variants in 53 cancer related genes. The variant allele frequencies of the 555 true variants are ranged between 5 to 15%. The reference sequences for the 53 genes of AOHC and human reference genome sequences were downloaded from NCBI database (ftp://ftp.ncbi.nlm.nih.gov/genomes/H\_sapiens/ARCHIVE/BUILD.37.3/Assembled\_chromosomes/seq/). Of the 555 variants in AOHC, 500 variants were successfully targeted by Oncomine Cancer Panel (OCP) (Hovelson, *et al*., 2015). The libraries were quantified by real-time PCR using Ion Library TaqManTM Quantitation Kit (Thermo Fisher Scientific, Waltham, MA, USA). The templates were prepared using the Ion PGM Template OT2 200 Kit and Ion OneTouch 2 System according to the manufacturer’s instructions. A single DNA template was assessed on a single Ion 318 chip v2.

Additional public microbial data sets (run: SRR1653476, SRR1653477, SRR2013384, and SRR2013385) were downloaded from the SRA database (http://www.ncbi.nlm.nih.gov/sra/) using SRAtoolkit (version 2.5.7) to develop the quality filters without heterozygosity effects. In most studies about sequencing errors, microbial data have been used to assess sequencing error rates due to the lack of genetic heterogeneity. Three reference sequences for the microbial data sets were also downloaded (gi: 121602919, 134290884, and 219666071) and used for analysis.

M2. Somatic variant calling

To test pipelines for somatic variant calling, we used three different short read aligners and five variant callers. The raw reads were mapped using Bowtie2 (version 2.2.8), BWA (MEM algorithm; version 0.7.12-r1039), and TMAP (version 3.4.1). HaplotypeCaller (GATK version 3.5-0), LoFreq (version 2.1.3), MuTect2 beta version (GATK nightly build 2016-04-25-g7a7b7cd), SNVer (version 0.5.3), Torrent Variant Caller (version 4.2-13), and VarScan2 (version 2.3.9) were used to call variants. However, HaplotypeCaller and SNVer were not used for further analyses due to low sensitivities (Supplementary Table S2)*.* The optimized parameters for MuTect2 was --artifact\_detection\_mode --initial\_tumor\_lod 8.1 (Supplementary Figure S3). To test the vendor’s pipeline, Ion Reporter (version 5.0; Life Technologies) was used to filter variant calls from the Torrent Variant Caller.

M3. Development of AIRVF filters

To develop AIRVF filters, we first defined eight filters that work on distinct characteristics of erroneous reads based on previous studies. For each filter, we set the optimal thresholds based on the observed training data (the public microbial data and the AOHC training data, see Supplementary Method M1). The detailed analysis data to determine the threshold are described in the Supplementary Figures and Tables).

1. Raw-read level filters

Base-call errors in a sequencing step are a major source of erroneous variant calling. Previous studies showed that a strict base-call filtration can reduce false-positive variant calls in Complete Genomic and Illumina platforms (Reumers, et al., 2012). We assume that a similar approach regarding the Ion Torrent specific errors would lead to a reduction of errors. In total, five different filters were defined and used in AIRVF. Default parameters have been set based on the observation of the public microbial data. To prevent an excessive filtering that may cause a lack of coverage to detect somatic variants, the quality filtering thresholds were designed to retain at least 50% of the raw reads (Supplementary Figure S1).

* **Contaminant filter**:

In this filter, NGS reads generated from potential contaminant are removed. Each raw short read is queried in the BLAST+ (version 2.3.0) to find the match in the human reference genome. BLAST+ was run with default parameters. Unmatched sequences are removed from the raw data.

* **Flowgram filter**:

A few previous studies attempted to use flow values to determine the exact homopolymer lengths (Zeng, et al., 2013). Based on the technical note from the vendor (Thermo Fisher Scientific), however, current Ion Torrent PGM system implements a proper base-calling algorithm by base recalibration. Thus, we used the flow value to filter ambiguous reads, instead of reassigning homopolymer length. To build the filter, we defined the *ambiguous zone* in the flow value (Supplementary Figure S4) and NGS reads that contain flow values within the zones were removed. By default, AIRVF removes the erroneous reads with frequent ambiguous flow values (0.40-0.50; 1.23-1.50). The criteria (0.085) for the ambiguous flow value ratio was designed to selectively remove only significantly erroneous reads (Supplementary Figure S1).

* **Read length filter**:

The read length filter aims to remove NGS reads of an abnormally short length based on the previous reports that short reads may represent a problem in the extension of DNA fragments or from multi-templated beads (Huse, et al., 2007). The cut-off value for abnormally short reads can either be defined by a hard threshold (50bp), or can be drawn from the designated target region. When the latter is used, AIRVF takes the genomic target regions (BED file) and use the shorted amplicon length.

* **Quality filter**:

In the quality filter, AIRVF takes the base-call quality scores to removes reads with a bad quality. The quality filter works in two ways. One is the average quality filter that considers the arithmetic average quality scores of the all based in a read. The other is the minimum quality filter, which removes reads that contain any number of bases with a base-call quality below a threshold. The rational for the minimum quality filter is that the Ion Torrent platform can be regionally unstable , particularly due to the gorges in homopolymeric and primer regions (Whiteley, et al., 2012). The regional instability could be also observed in the microbial dataset (Supplementary Figure S5). To determine the thresholds for the average and the minimum quality score, we tested the filters on the microbial data with different values (data not shown). Although not globally optimized, we concluded that 25 and 8 for the average and the minimum quality filters could efficiently remove reads with low quality without losing too much portion of the raw data (>50% of the original reads, see Supplementary Figure S1).

* **Quality trimming**:

Quality trimming does not filter out reads, but cuts both ends of reads to remove bases of a low quality. Quality trimming also targets the sequencing errors that are caused by damaged primers (Shin and Park, 2014). AIRVF checks and cuts bases from the both ends of reads until it finds a base of quality score >27.

2. Mapped-read level filters
To identify the features of mapped reads, we categorized the mapped reads, which contained both true and false positives, into 5 groups depending on the frequency of true and false positives (Supplementary Table S3). To characterize the mapped reads, we defined five different regions (i.e., inaccurate, false-positive-rich, undetected, all mapped, and accurate) in this study (Supplementary Figure S6). Inaccurate regions had 3 or more continuous undetected calls within 100 bp and two or more false positive calls near the continuous undetected calls (within 100 bp). False-positive-rich regions had 4-5 (4 for MuTect2 beta version and 5 for VarScan2) or more false positive calls within 100 bp and no undetected true variants near the false positive calls (within 100 bp). Undetected regions had 2-3 (2 for VarScan2 and 3 for MuTect2 beta version) or more continuous undetected variants within 100 bp without false positive calls near the undetected variants (within 100 bp). All mapped regions designate all the regions mapped by mapping tools. Accurate regions denote short regions that have 4 or more true variant calls within 100 bp and no false positive call/undetected variants near true variant calls (within 100 bp). Generally, mapping quality was not a critical factor for the removal of erroneously mapped reads (Supplementary Figure S7). However, false-positive-rich regions had more insertions and deletions than did accurate regions, and accurate regions had fewer insertions and deletions than did all mapped regions (Supplementary Figure S2). To maximize precision, we developed a per-read indel filter generally allowed 1 insertion and 1 deletion per read (Supplementary Table S4).

* **Per-read indel filter**:

Based on our findings, the SAM files were filtered to efficiently remove the erroneous reads containing too many insertions and deletions (Supplementary Figure S2). Generally, the highest precision was observed with 1 insertion and 1 deletion per read (Supplementary Table S5). The VarScan2 pipelines demonstrated a significantly increased precision and a slightly decreased sensitivity (see Fig. 1).

3. Variant level filters
The expected results from AOHC were regarded as true positives. Any variants in the putative synthetic DNA regions (the regions between true positives within 100 bp) were regarded as false positives. To develop the parameter based filter, the differences in almost fields in vcf files between true and false positives were analyzed. To distinguish between true and false positives, we referred the alleles of hotspot mutations from my cancer genome (https://www.mycancergenome.org/).

* **Parameter based filter**:

The differences between true and false positive calls were identified for the development of the variant filters (The difference from LoFreq was not shown due to the low number of false positives). The SNV/indel ratio, HCNT, tumor LOD value, AD2/AD1, AF, and QSS2 differed between the true and false positive calls from the MuTect2 beta version vcf file (Supplementary Figure S8). Surprisingly, the ratio of the triallelic\_site filter among the true positive calls was slightly greater than that among false positive calls. The SNV/indel ratio, read number supporting variant allele, VAF, Strand1/Strand2, average base quality of variant-supporting read bases, and chi-square value testing the association between the +/- strand and reference/variant allele (Supplementary Table S1) were different in the VarScan2 vcf file (Supplementary Figure S9). Interestingly, some false positive calls had extremely high chi-square values. In the results from Torrent Variant Caller and Torrent Variant Caller-Ion Reporter pipelines, the filter, AF, rowtype, call, ID.count, ALT.count, INFO.O.HS, QUAL, INFO.A.FSAF were significantly different between the true and false positive calls (Supplementary Figure S8). We also analyzed the differences between the true and false positive calls from Torrent Variant Caller (Supplementary Figure S10) and Torrent Variant Caller-Ion Reporter (Supplementary Figure S11). Significantly different fields were used to establish the combinatorial conditions for the removal of the false positive calls (Supplementary Table S6, S7, S8). Surprisingly, FPR was decreased up to 534 removing ~99% FPR, and the sensitivity was slightly decreased (4.4%) in the AIRVF and VarScan2 pipeline. The AIRVF and VarScan2 pipeline detected relatively more variants with a low VAF (Supplementary Figure S12; 5-15%, 15-35%, and 50%). However, our manual analysis revealed that VarScan2 cannot sensitively detect long indels (Supplementary Table S9). Almost all fields in the vcf files were analyzed to identify the differences between the true and false positive calls. The FFPE filters were developed using common values among the putative true and false positives in the cancer genomes. The putative true positives and false positives in cancer genomes were determined by the alleles and information regarding the hotspot mutations. We only considered the false positive calls within 100 bp of the true positive calls to calculate precision and the FPR to exclude genomic variants that were not shown in AOHC Representative Performance Data.

* **Sample filter (FFPE)**:

The FFPE filters only for the VarScan2 pipelines were developed as a part of AIRVF to remove putative false positive calls with a very low VAF from FFPE samples (Supplementary Figure S13) using values among putative the false and true positives (Supplementary Table S10) since the Torrent Variant Caller-Ion Reporter pipeline was already optimized for FFPE samples.

M4. Assessment of accuracy in somatic variant calling

The calls from the AOHC samples were assayed for the development of AIRVF (Figure 1). To compare the sensitivity of the mapping algorithms for Ion TorrentTM PGM, we chose Bowtie, BWA, and TMAP because these aligners are widely used and optimized for the alignment of long reads or Ion Torrent data. HaplotypeCaller, LoFreq, Samtools, MuTect2 beta version, SNVer, and VarScan2 were compared because these tools are rapid and widely used in medical research studies. The TMAP-HaplotypeCaller and BWA-VarScan2/TMAP-VarScan2 pipelines demonstrated the high precision and sensitivity, respectively. However, AIRVF was not developed for HaplotypeCaller, Samtools, and SNVer due to their low sensitivity (Supplementary Table S2). Additionally, these tools have a high sensitivity with low VAFs in HiSeq data. Additionally, the vendor’s default pipeline, Torrent SuiteTM Software (TS) and Torrent Variant Caller plugin, was compared. The expected results from AOHC were regarded as true positives. Any variants in the putative synthetic DNA regions (the regions between true positives within 100 bp) were regarded as false positives. Filters were stepwisely developed from the raw-read filters to variant filters with assessing steps. The FPR from the Torrent Variant Caller-Ion Reporter pipeline was low with up to 1,468 FP mutations / Mbp. Undetected variants and long homopolymers are often collocated (Supplementary Figure S14). We selected Torrent Variant Caller, MuTect2 beta version, and VarScan2 for the development of AIRVF because these tools outperformed the other variant calling tools. However, only the variant filter was developed for Torrent Variant Caller because it demonstrated relatively high precision and low sensitivity in addition to providing low-coverage data by filtering reads that decrease its sensitivity. The parameters of MuTect2 beta version were optimized to increase sensitivity (--artifact\_detection\_mode --initial\_tumor\_lod 8.1).

Supplementary Figures

Figure S1. Error rates and read numbers after applying filters. CR = Contaminant removal, FF = Flowgram filter, RLF = Read length filter, AQF = Average quality filter, MQF = Minimum quality filter, T = Trimming. Sequencing error rates from run SRR1653476, SRR1653477, SRR2013384, and SRR2013385 were designated as 318v2-400bp-1-E, 318v2-400-bp2-E, 318v2-HiQ-1-E, and 318v2-HiQ-2-E, respectively. Percentage changes of read numbers from run SRR1653476, SRR1653477, SRR2013384, and SRR2013385 were designated as 318v2-400bp-1-R, 318v2-400-bp2-R, 318v2-HiQ-1-R, and 318v2-HiQ-2-R, respectively.

a

b

Figure S2. The occurrence frequency of insertions and deletions according to the accuracy evaluation of variant calling. The occurrence frequency of (**a**) insertions and (**b**) deletions in mapped reads by TMAP (MT) and Bowtie2 (VS) after quality filtering. Inaccurate, false-positive-rich, undetected, all mapped, accurate regions were defined from the MuTect2 beta version (MT) and VarScan2 (VS) results (See Supplementary Table S1).

Figure S3. Precision, sensitivity, and FPR according to initial LOD threshold. Results from QF-TMAP-MRF-MT pipeline (QF = Quality filtering, MRF = Mapped read filtering, MT = MuTect2 beta version). Precision was high around initial LOD threshold 8.1.

Figure S4. Distribution of flow values from run SRR2013385. Ambiguous zones were set from 0.40 to 0.50 and from 1.23 to 1.50.



Figure S5. Positional quality scores in the Ion TorrentTM PGM data. Red, pink, blue, and azure lines designate the positional quality scores from run SRR1653476, SRR1653477, SRR2013384, and SRR2013385, respectively. Thick gray lines represent the primer regions.



Figure S6. Conceptual illustration of inaccurate, false-positive-rich, undetected, all mapped, and accurate regions. Inaccurate regions mean short regions which have continuous undetected calls with false positive calls near them. False-positive-rich regions mean short regions which have false positive calls without undetected true variant near them. Undetected regions mean short regions which have continuous undetected variants without false positive calls near them. All mapped regions mean all the regions mapped by mapping tools. Accurate regions mean short regions which have true variant calls without false positive call and undetected variant near them.

Figure S7. Average mapping quality according to the accuracy evaluation of variant calling. QF = Quality filtering, Bowtie = Bowtie2, MT = MuTect2, VS = VarScan2. Inaccurate regions designate short regions which have 3 or more continuous undetected calls within 100 bp, and two or more false positive calls within 100 bp around undetected variants. False-positive-rich regions indicate short regions which have 4-5 (4 for MT and 5 for VS) or more false positive calls within 100 bp, and no undetected true variant within 100 bp around false positive calls. Undetected regions mean short regions which have 2-3 (2 for VS and 3 for MT) or more continuous undetected variants within 100 bp, and no false positive call within 100 bp around undetected variants. Accurate regions denote short regions which have 4 or more true variant calls within 100 bp, and no false positive call/undetected variant within 100 bp around true variant calls. See Supplementary Table S1.

a b

c d

e f

g

Figure S8. Different and similar results between true and false positive calls in MT data.Differences between true positive false positive calls from QF-TMAP-MRF-MT-PO pipeline (QF = Quality filtering, BWA = BWA-MEM, MRF = Mapped read filtering, MT = MuTect2 beta version, PO = Parameter optimization). (**a**) SNV, insertion, and deletion ratio. (**b**) MT filter result. (**c**) HCNT. (**d**) TLOD. (**e**) AD2 / AD1. Values were removed when AD1 = 0. (**f**) AF. (**g**) QSS2.

a b

c d

e f

Figure S9. Different results between true and false positive calls in VS data.Differences between true positive false positive calls from QF-Bowtie-MRF-VS pipeline (QF = Quality filtering, Bowtie = Bowtie2, MRF = Mapped read filtering, VS = VarScan2). (**a**) SNV, insertion, and deletion ratio. (**b**) A number of reads supporting variant allele. (**c**) The frequency of the variant allele. (**d**) Strands1 / Strands2 pair. (**e**) The average base quality of variant-supporting read base. (**f**) Chi-square value to test the association between +/- strand and reference/variant allele. Values were removed when expected values were 0.

**a b**Figure S10. Different results between true and false positive calls in TS (TVC) data.Differences between true positive false positive calls from TS (TVC) (TS = Torrent Suite, TVC = Torrent Variant Caller). (**a**) filter. (**b**) AF.

a b

c d

e f

g h

Figure S11. Different results between true and false positive calls in TVC-IR data. Differences between true positive false positive calls from TVC-IR pipeline (IR = Ion Reporter, TVC = Torrent Variant Caller). (a) Rowtype. (b) Call. (c) ID.count. (d) ALT.count. (e) INFO.O.HS. (f) QUAL. (g) INFO.A.FSAF. (h) INFO.A.AF.



Figure S12. Sensitivity of variants with different VAF. (IR = Ion Reporter, MT = MuTect2 beta version, TVC = Torrent Variant Caller, VS = VarScan2)

Figure S13. A number of detected cancer mutations (SNVs and indels) before and after applying FFPE filters according to VAF from the AIRVF and VS pipeline in cancer genomes. BWA-MEM algorithm was used for the AIRVF and VS pipeline (VAF = variant allele frequency, VS = VarScan2).



Figure S14. Detected and undetected variants in chromosome 4 from TMAP-VS pipeline. Blue and red circles designate detected and undetected variants, respectively. The bigger the circles, the more the variants.

Figure S15. Overview of the development and assessment of AIRVF. DNA samples from AOHC were sequenced and assayed for the development of AIRVF. Microbial data were used to identify conditions that selectively remove erroneous reads. We defined the accurate and inaccurate regions and analyzed the differences in the reads between these regions. Fields in the vcf files were analyzed to compare the true and false positives. To remove the frequent false positives with a low VAF from FFPE tissue sample, FFPE filters were developed.

Supplementary Tables

Table S1: Table to calculate chi-square.

|  |  |
| --- | --- |
| Read1Plus | Read1Minus |
| Read2Plus | Read2Minus |

Table S2: Comparison of low-frequency variant callers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Aligner | Caller | Precision\* (%) | Sensitivity (%) | FPR |
| Bowtie2 | HaplotypeCaller | 90.43 | 41.60 | 2,937 |
| SNVer | 65.03 | 18.60 | 6,675 |
| BWA-MEM | HaplotypeCaller | 90.49 | 51.40 | 3,604 |
| SNVer | 61.85 | 21.40 | 8,811 |
| TMAP | HaplotypeCaller | 90.60 | 48.20 | 3,337 |
| SNVer | 67.72 | 21.40 | 6,808 |
| Abbreviations |
| **FPR:** False positive mutations / Mbp |

Table S3: Chromosomes and positions of regions according to the accuracy evaluation of variant calling.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pipeline | Region | Chromosome | Start | End |
| QF-TMAP-MT | Inaccurate | 4 | 153,243,998 | 153,244,196 |
| 7 | 55,259,420 | 55,259,538 |
| 17 | 7,578,449 | 7,578,679 |
| 17 | 7,579,495 | 7,579,595 |
| False-positive-rich | 5 | 112,175,011 | 112,175,130 |
| 5 | 112,175,567 | 112,175,728 |
| 8 | 38,285,851 | 38,285,975 |
| 16 | 68,856,011 | 68,856,133 |
| 17 | 7,578,042 | 7,578,267 |
| 18 | 48,575,099 | 48,575,213 |
| 22 | 24,145,432 | 24,145,598 |
| Undetected | 3 | 178,936,023 | 178,936,179 |
| 12 | 112,888,118 | 112,888,228 |
| 12 | 112,926,835 | 112,926,961 |
| 13 | 49,039,149 | 49,039,232 |
| Accurate | 3 | 10,183,732 | 10,183,854 |
| 3 | 10,191,418 | 10,191,527 |
| 3 | 41,266,029 | 41,266,210 |
| 4 | 55,592,157 | 55,592,246 |
| 7 | 55,242,411 | 55,242,540 |
| 10 | 89,690,782 | 89,690,877 |
| 10 | 89,711,914 | 89,712,021 |
| 12 | 25,398,186 | 25,398,304 |
| 13 | 49,037,846 | 49,037,966 |
| 17 | 7,573,976 | 7,574,068 |
| QF-Bowtie-VS | Inaccurate | 4 | 153,243,998 | 153,244,196 |
| 17 | 7,578,449 | 7,578,679 |
| 17 | 7,579,292 | 7,579,503 |
| False-positive-rich | 7 | 55,241,635 | 55,241,729 |
| 17 | 7,578,370 | 7,578,563 |
| Undetected | 4 | 153,245,410 | 153,247,238 |
| 16 | 68,855,845 | 68,856,022 |
| Accurate | 3 | 10,183,732 | 10,183,854 |
| 3 | 10,191,418 | 10,191,527 |
| 3 | 178,936,023 | 178,936,179 |
| 4 | 55,592,157 | 55,592,246 |
| 5 | 112,175,740 | 112,176,035 |
| 7 | 116,423,407 | 116,423,492 |
| 9 | 139,399,337 | 139,399,447 |
| 10 | 89,690,782 | 89,690,877 |
| 10 | 89,692,813 | 89,693,032 |
| 10 | 89,711,914 | 89,712,021 |
| 12 | 25,398,186 | 25,398,304 |
| 12 | 112,926,835 | 112,926,961 |
| 13 | 49,037,846 | 49,037,966 |
| 17 | 7,573,976 | 7,574,068 |
| 17 | 7,577,432 | 7,577,645 |
| 18 | 48,604,658 | 48,604,774 |
| QF = Quality filtering, Bowtie = Bowtie2, MT = MuTect2 beta version, VS = VarScan2 |

Table S4: Conditions for the removal of mapped reads.

|  |  |  |
| --- | --- | --- |
| Pipeline | Insertion per read | Deletion per read |
| QC-Bowtie-MRF-MT (PO)-VF, QC-Bowtie-MRF-VS-VF, QC-BWA-MRF-VS-VF, QC-TMAP-MRF-VS-VF | >1 | >1 |
| QC-BWA-MRF-MT (PO)-VF, QC-TMAP-MRF-MT (PO)-VF | >4 | >1 |
| QF = quality filtering, Bowtie = Bowtie2, BWA = BWA-MEM, MRF = mapped read filtering, MT = MuTect2 beta version, VS = VarScan2, PO = parameter optimization, Conditions for MuTect2 stable version will be different from MuTect2 beta version. |

Table S5: Precision and sensitivity according to the allowed maximum number of insertions and deletions per read.

**a**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Precision | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletion | 74.28% | 73.18% | 72.81% | 72.83% | 72.31% |
| 2 deletions | 72.94% | 71.56% | 71.13% | 71.30% | 71.27% |
| 3 deletions | 72.88% | 71.15% | 71.14% | 72.27% | 71.19% |
| 4 deletions | 72.93% | 71.81% | 72.13% | 71.65% | 72.33% |
| 5 deletions | 72.31% | 71.56% | 72.35% | 71.86% | 71.50% |
| Sensitivity | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletions | 82.00% | 78.60% | 79.80% | 80.40% | 79.40% |
| 2 deletions | 81.40% | 78.00% | 81.80% | 80.00% | 78.40% |
| 3 deletions | 79.00% | 79.40% | 78.40% | 80.80% | 77.60% |
| 4 deletions | 79.20% | 81.00% | 79.20% | 81.40% | 80.00% |
| 5 deletions | 79.40% | 80.00% | 80.60% | 81.20% | 80.80% |

**b**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Precision | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletion | 68.64% | 67.19% | 67.79% | 68.80% | 68.50% |
| 2 deletions | 68.15% | 66.21% | 67.37% | 67.23% | 65.38% |
| 3 deletions | 68.19% | 66.82% | 65.36% | 66.97% | 65.53% |
| 4 deletions | 68.41% | 67.57% | 65.77% | 67.23% | 66.37% |
| 5 deletions | 68.50% | 68.48% | 66.77% | 67.32% | 66.67% |
| Sensitivity | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletions | 88.00% | 86.40% | 88.80% | 88.20% | 87.00% |
| 2 deletions | 88.60% | 87.00% | 89.20% | 87.80% | 88.00% |
| 3 deletions | 89.60% | 88.20% | 86.80% | 89.20% | 88.60% |
| 4 deletions | 88.80% | 90.00% | 88.00% | 88.20% | 90.00% |
| 5 deletions | 87.00% | 88.20% | 86.40% | 89.40% | 87.60% |

**c**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Precision | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletion | 74.29% | 73.50% | 74.69% | 74.82% | 74.34% |
| 2 deletions | 74.29% | 72.65% | 72.51% | 73.01% | 74.02% |
| 3 deletions | 74.39% | 73.23% | 73.06% | 74.01% | 74.39% |
| 4 deletions | 74.02% | 73.93% | 73.45% | 73.60% | 73.09% |
| 5 deletions | 73.01% | 73.94% | 73.79% | 73.15% | 72.77% |
| Sensitivity | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletions | 83.80% | 83.20% | 85.60% | 83.80% | 84.00% |
| 2 deletions | 83.80% | 83.40% | 84.40% | 84.40% | 87.20% |
| 3 deletions | 84.80% | 84.80% | 83.00% | 86.00% | 86.00% |
| 4 deletions | 83.20% | 86.20% | 85.20% | 84.20% | 84.20% |
| 5 deletions | 84.40% | 84.00% | 85.60% | 85.00% | 85.00% |

**d**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Precision | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletion | 81.24% | 78.54% | 77.89% | 77.50% | 77.50% |
| 2 deletions | 79.00% | 76.61% | 61.85% | 75.76% | 75.76% |
| 3 deletions | 78.61% | 76.24% | 75.64% | 75.40% | 75.40% |
| 4 deletions | 78.61% | 76.24% | 75.64% | 75.40% | 75.40% |
| 5 deletions | 78.61% | 76.24% | 75.64% | 75.40% | 75.40% |
| Sensitivity | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletions | 94.40% | 94.40% | 94.40% | 94.40% | 94.40% |
| 2 deletions | 94.80% | 95.00% | 95.00% | 95.00% | 95.00% |
| 3 deletions | 94.80% | 95.00% | 95.00% | 95.00% | 95.00% |
| 4 deletions | 94.80% | 95.00% | 95.00% | 95.00% | 95.00% |
| 5 deletions | 94.80% | 95.00% | 95.00% | 95.00% | 95.00% |

**e**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Precision | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletion | 77.60% | 75.87% | 75.28% | 75.04% | 74.92% |
| 2 deletions | 75.55% | 73.13% | 72.91% | 72.80% | 72.47% |
| 3 deletions | 75.08% | 72.69% | 72.47% | 72.36% | 56.62% |
| 4 deletions | 75.08% | 72.69% | 72.47% | 72.36% | 72.03% |
| 5 deletions | 75.08% | 72.69% | 72.47% | 72.36% | 72.03% |
| Sensitivity | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletions | 95.60% | 95.60% | 95.60% | 95.60% | 95.60% |
| 2 deletions | 95.80% | 95.80% | 95.80% | 95.80% | 95.80% |
| 3 deletions | 95.80% | 95.80% | 95.80% | 95.80% | 95.80% |
| 4 deletions | 95.80% | 95.80% | 95.80% | 95.80% | 95.80% |
| 5 deletions | 95.80% | 95.80% | 95.80% | 95.80% | 95.80% |

**f**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Precision | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletion | 80.75% | 78.74% | 77.96% | 77.83% | 77.83% |
| 2 deletions | 77.94% | 75.63% | 75.16% | 75.04% | 75.04% |
| 3 deletions | 76.73% | 74.57% | 73.99% | 73.88% | 73.88% |
| 4 deletions | 76.73% | 74.57% | 73.99% | 73.88% | 73.88% |
| 5 deletions | 76.73% | 74.57% | 73.99% | 73.88% | 73.88% |
| Sensitivity | 1 insertion | 2 insertions | 3 insertions | 4 insertions | 5 insertions |
| 1 deletions | 94.80% | 94.80% | 94.80% | 94.80% | 94.80% |
| 2 deletions | 95.40% | 95.60% | 95.60% | 95.60% | 95.60% |
| 3 deletions | 95.60% | 95.60% | 95.60% | 95.60% | 95.60% |
| 4 deletions | 95.60% | 95.60% | 95.60% | 95.60% | 95.60% |
| 5 deletions | 95.60% | 95.60% | 95.60% | 95.60% | 95.60% |

Precision and sensitivity form (**a**) Bowtie-MT, (**b**) BWA-MT, (**c**) TMAP-MT, (**d**) Bowtie-VS, (**e**) BWA-VS, (**f**) TMAP-VS pipelines (Bowtie = Bowtie2, BWA = BWA-MEM, MT = MuTect2 beta version, VS = VarScan2).

Table S6: Conditions to filter variant calls from LF.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Order\* | Variant | Quality | DP | AF | SB | Chi | Result |
| 1 | SNV | <100 |  | <0.015 |  |  | FP |
| 2 | SNV | <300 | <2,000 | <0.035 |  | >0.8 | FP |
| 2 | SNV | <700 | <3000 | <0.05 | >21 | >8 | FP |
| FP = false positive, TP = true positive\*Order of determination. Results will not be changed by the latter.If results are not determined by above conditions, they will be true positives. |

Table S7: Conditions to filter variant calls from MT.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Order\* | Variant | Filter | TLOD | HCNT | AD2/AD1 | AF | QSS2 | Result |
| 1 |  |  |  |  | AD1=0 |  |  | TP |
| 2 | SNV | PASS | <12.8 |  |  |  |  | FP |
| 2 | SNV | PASS | <105.0 |  | <0.03 |  |  | FP |
| 3 | SNV | PASS | <105.0 |  |  | <0.03 |  | FP |
| 5 | SNV | triallelic\_site |  |  |  |  |  | TP |
| 6 | indel | PASS | <22.0 |  |  | <0.03 |  | FP |
| 7 | indel | PASS | <105.0 | <9 |  | <0.05 |  | FP |
| 8 | indel | PASS |  |  | <0.03 |  |  | FP |
| 9 | indel | PASS |  |  |  | <0.03 |  | FP |
| 10 | indel | PASS |  |  |  |  | <1,000 | FP |
| 11 | indel | triallelic\_site | <22.0 |  |  |  |  | FP |
| 12 | indel | triallelic\_site |  | <9 |  | <0.09 |  | FP |
| FP = false positive, TP = true positive\*Order of determination. Results will not be changed by the latter.If results are not determined by above conditions, they will be true positives. |

Table S8: Conditions to filter variant calls from VS.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Order\* | Variant | Strands1 or Strands2 | Read1 or Read2 | Read1Plus, Read1Minus, Read2Plus, or Read2Minus | VarFreq | MapQual2 | Chi-square | Result |
| 1 |  | Strands1=0 or 1 |  |  |  |  |  | TP |
| 2 | indel | Strands2=2 | $$Read2<The highest Read2×0.005$$ |  |  |  |  | FP |
| 3 | indel | Strands2=2 |  | Read2Plus/Read1Plus<0.030 & Read2Minus/Read1Minus<0.030 |  |  |  | FP |
| 4 | indel | Strands2=2 |  | Read2Plus/Read1Plus<0.065 & Read2Minus/Read1Minus<0.065 | <3.5% |  |  | FP |
| 5 | indel | Strands2=2 |  |  | <1.3% |  |  | FP |
| 6 | indel | Strands2=2 |  |  | <3.2% | <24 |  | FP |
| 7 | indel | Strands2=2 |  |  |  |  | >47 | FP |
| 8 | indel | Strands2=2 |  | Read2Plus<9 & Read2Minus<9 |  |  |  | FP |
| 9 | indel | Strands2=2 |  | Read2Plus/Read1Plus<0.100 & Read2Minus/Read1Minus<0.100 | <7.3% | <24 | >10 | FP |
| 10 | indel | Strands2=1 |  |  | <12.0% | <28 |  | FP |
| 11 | indel | Strands2=1 |  |  |  |  | >50 | FP |
| 12 | SNV |  | Read1<20 |  |  |  |  | FP |
| 13 | SNV | Strand2=2 |  |  | >50.0% |  |  | TP |
| 14 | SNV | Strand2=2 |  |  |  |  | >100 | FP |
| 15 | SNV | Strand2=2 |  |  | <20.0% | <25 |  | FP |
| 16 | SNV | Strand2=1 |  |  | <1.3% |  |  | FP |
| 17 | SNV | Strand2=1 |  | Read2Plus=0 & Read2Minus/Read1Minus<0.033 |  |  |  | FP |
| 18 | SNV | Strand2=1 |  |  |  | <25 |  | FP |
| FP = false positive, TP = true positive\*Order of determination. Results will not be changed by the latter.If results are not determined by above conditions, they will be true positives.These conditions require sequencing depth of over 700x. |

Table S9: Reference and alternate sequences of clinically identified variants.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Reference | Alternate | VAF (%) | AIRVF and TVC-IR | AIRVF and MT | AIRVF and VS |
| T | G | 19.37 | detected | detected | detected |
| TTAAGAGAAGCAACATCTCC | TC | 46.69 | detected | detected | undetected |
| TTAAGAGAAGCAACATCTCC | TC | 87.99 | detected | detected | undetected |
| AGGAATTAAGAGAAGCAAC | AAAC | 63.98 | detected | detected | undetected |
| AGGAATTAAGAGAAGCAAC | AAAC | 57.84 | detected | detected | undetected |
| AGGAATTAAGAGAAGCAAC | AAAC | 20.07 | detected | detected | undetected |
| GGAATTAAGAGAAGCAACATCT | GCAACCAACATCT | 29.61 | detected | detected | undetected |
| AGGAATTAAGAGAAGCAAC | AAAC | 88.26 | detected | detected | undetected |
| ACT | TCT | 37.79 | detected | detected | detected |
| ACT | TCT | 36.90 | detected | detected | detected |
| ACT | TCT | 31.99 | detected | detected | detected |
| ACT | TCT | 30.23 | detected | detected | detected |
| ACT | TCT | 48.16 | detected | detected | detected |
| ACT | TCT | 33.52 | detected | detected | detected |
| CC | AC | 59.97 | detected | detected | detected |
| C | T | 37.80 | detected | detected | detected |
| C | T | 19.47 | detected | detected | detected |
| CC | TC | 40.00 | detected | detected | detected |
| CC | TC | 38.79 | detected | detected | detected |
| C | T | 11.28 | detected | detected | detected |
| CC | TC | 50.14 | detected | detected | detected |
| CC | AC | 36.31 | detected | detected | detected |
| CC | TC | 41.89 | detected | detected | detected |
| C | A | 33.33 | detected | detected | detected |
| AGGAATTAAGAGAAGCAAC | AAAC | 57.08 | detected | detected | undetected |
| TTAAGAGAAGCAACATCTCC | TC | 67.14 | detected | detected | undetected |
| TTAAGAGAAGCAACATCTCCGAAAGC | TCTCCACA | 27.48 | detected | undetected | undetected |
| AGGAATTAAGAGAAGCAAC | AAAC | 46.12 | detected | detected | detected |
| CC | TC | 29.21 | detected | detected | detected |
| CC | AC | 29.72 | detected | detected | detected |
| Reference sequences, alternate sequences, and VAF were reported in TS (TVC)-IR tsv files. BWA-MEM algorithm was used for MT and VS. IR = Ion Reporter, MT = MuTect2 beta version, TVC = Torrent Variant Caller, VAF = variant allele frequency, VS = VarScan2. |

Table S10: Conditions to filter variant calls with low allele frequency from FFPE samples.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Order\* | Variant | Strands2 | VarFreq | Read2 | MapQual2 | Chi-square | Result |
| 1 | Indel | 1 |  |  |  |  | FP |
| 2 | Indel | 2 | <7.00% |  |  | >20 | FP |
| 3 | Indel | 2 | <7.00% |  | <29 |  | FP |
| 4 | Indel | 2 | <15.00% |  | <25 |  | FP |
| 5 | Indel | 2 |  | <200 | <18 | <36 | FP |
| 6 | SNV | 1 | <4.00% |  | >34 | <0.3 | TP |
| 7 | SNV | 1 |  |  |  |  | FP |
| 8 | SNV | 2 | <1.50% |  |  |  | FP |
| 9 | SNV | 2 | <2.00% | >10 | >32 | <0.03 | TP |
| 10 | SNV | 2 | <2.00% |  |  |  | FP |
| 11 | SNV | 2 | <3.00% | >15 | >31 | <0.20 | TP |
| 12 | SNV | 2 | <3.00% |  |  |  | FP |
| 13 | SNV | 2 | <4.00% | >15 | >29 | <0.20 | TP |
| 14 | SNV | 2 | <4.00% |  |  |  | FP |
| 15 | SNV | 2 | <5.00% | >18 | >29 | <0.20 | TP |
| 16 | SNV | 2 | <5.00% |  |  |  | FP |
| 17 | SNV | 2 | <6.00% | >18 | >29 | <0.20 | TP |
| 18 | SNV | 2 | <6.00% |  |  |  | FP |
| 19 | SNV | 2 | <8.00% | >2 | >29 | <0.30 | TP |
| 20 | SNV | 2 | <8.00% |  |  |  | FP |
| 21 | SNV | 2 |  |  | <29 |  | FP |
| FP = false positive, TP = true positive\*Order of determination. Results will not be changed by the latter.If results are not determined by above conditions, they will be true positives.These conditions require sequencing depth of over 700x. |

References

Hovelson, D.H.*, et al.* Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. *Neoplasia* 2015;17(4):385-399.

Huse, S.M.*, et al.* Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 2007;8(7):R143.

Reumers, J.*, et al.* Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nat Biotech* 2012;30(1):61-68.

Shin, S. and Park, J. Correction of sequence-dependent ambiguous bases (Ns) from the 454 pyrosequencing system. *Nucleic Acids Research* 2014;42(7):e51-e51.

Whiteley, A.S.*, et al.* Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *Journal of Microbiological Methods* 2012;91(1):80-88.

Zeng, F., Jiang, R. and Chen, T. PyroHMMsnp: an SNP caller for Ion Torrent and 454 sequencing data. *Nucleic Acids Res* 2013;41.