

# Ion AmpliSeq™ Exome Hi-Q Effective Regions File

Publication Number MAN0018645 Revision B.0

## Product information

The Ion AmpliSeq™ Exome Hi-Q Effective Regions file is a BED file in which poorly performing flanking regions have been trimmed away from the affected amplicons. To complement this effort, Torrent Variant Calling has also been optimized for the area that is covered by the Effective Regions File. This file is used to analyse data from sequencing runs from Ion AmpliSeq™ Exome RDY panels.

We extensively evaluated Ion AmpliSeq™ Exome Hi-Q assay performance and identified a set of regions where the Ion AmpliSeq™ Exome Hi-Q assay performs poorly. These regions are typically at the edges of amplicons where sequence errors interact with alignment to cause false positives. Additionally, there are a few amplicons where the primers fail to successfully amplify and cause false negatives.

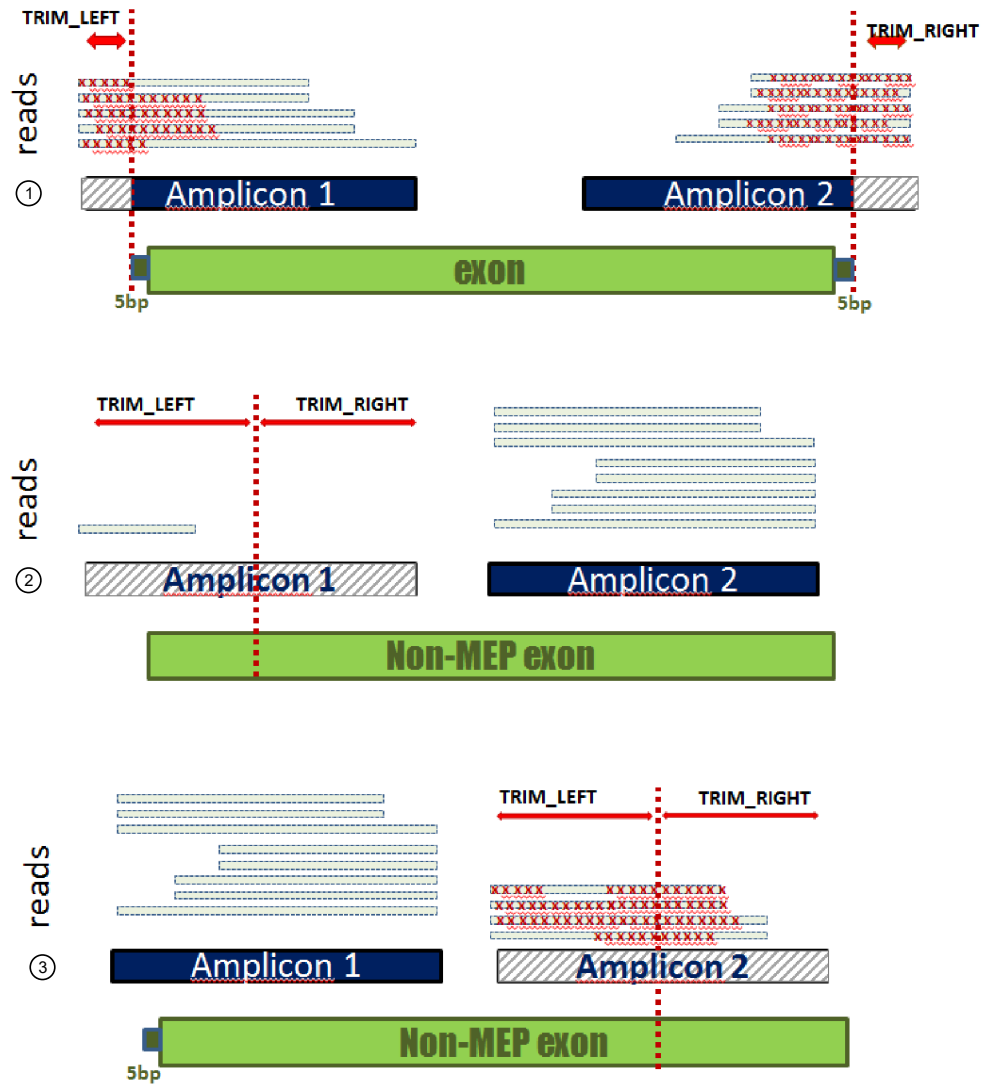
The Ion AmpliSeq™ Exome Hi-Q design targets exons and typically includes flanking regions since primer placement is often well into an intron. Most of the poorer performing regions are in intronic flanking regions. To correct for inaccurate calling, we used a method of selectively excluding (trimming) such regions during Variant Calling.

## Trimming approach

The variantCaller plugin (TVC) is highly optimized and determines whether a variant is truly present by using filters that filter out possible variants that are likely false. The metric used to measure performance is the regional density of filtered variants instead of the known truth. This metric guards against fitting to truth in a particular sample.

We applied the following rules:

- Selectively trim 9.4 MB of intronic regions from the amplicons with the largest number of filtered candidates within introns (that is, we targeted false positives in introns),
- Trim 0.4 MB of non-Medical Exome Project (MEP) exonic regions from amplicons with consistently very low read coverage (that is, we targeted false negatives),
- Trim 1.4 MB of non-MEP exonic regions from amplicons with the largest number of filtered candidates within introns (that is, we targeted false positives in exons).



- ① Trimming intronic domain based on the number of filtered candidates, for example, targeting false positives in introns.
- ② Trimming non-MEP exonic and intronic domain based on median coverage, for example targeting false negatives.
- ③ Trimming non-MEP exonic domain based on the number of filtered candidates, for example targeting false positives in exons.

## Trim without removing amplicons

By adding TRIM INFO to the designed BED file, we changed effective Ion AmpliSeq™ coverage by specifying TRIM\_RIGHT/TRIM\_END in the INFO field of the design regions file and modifying TVC to merge amplicons within (higherquality) trimmed boundaries.

By modifying BED file, we maintained the integrity of existing Torrent Suite™ plugins as well as any internal Torrent Suite™ Software that requires amplicon start/end positions—as these remain unchanged when a BED file including the TRIM directives is applied.

## Effect on medically relevant genes

There are several lists of possibly medical relevant genes. We confirmed the result of trimming against a curated list that includes 4,631 "genes associated with disease" presented during a talk given by Santani et al. at the ACMG Annual Clinical Genetics Meeting, March 2014. This list was part of the Medical Exome Project, a 2013 initiative sponsored by Emory Genetics Laboratory, Children's Hospital of Pennsylvania and Harvard's laboratory of Molecular Medicine. We also checked against claimed coverage for several other assays. The results are presented in the following table.

	Estate	Exonic	Overlap with Medical Exome Project	Overlap with AmpliSeq Exome
Ion AmpliSeq™ Exome baseline (prior to trim)	57.9	36.7	10.0	57.9
Ion AmpliSeq™ Exome post-trim	46.5	34.8	10.0	46.5
SureSelect	50.6	34.0	9.2	40.7
TrueSight	11.9	11.3	8.8	11.3
RapidCapture	37.3	35.2	10.2	33.6

**Note:** The following table contains 28 genes in the MEP genes that are not present in Ion AmpliSeq™ Exome (trimmed or untrimmed). Genes that do not encode exons, such as those encoding miRNA, are excluded. Contact technical support for more details.

Genes						
C4A	H19	MIR4691	MSH5	RMST	UOX	C4B
MIAT	MIR486	NBEAP1	RNU4ATAC	WT1-AS	CEBPA-AS1	MIR17HG
MIR621	NCF1C	SNORD50A	XIST	DSCR8	MIR184	MIR96
OTOG	SULT1A3	DUX4	MIR425	MIRLET7E	RMRP	TERC

## Results

For a NA12878 sample using the NIST 2.18 "highly confident" truth set for this sample, we see an improvement in sensitivity and specificity using TVC with the trimmed design BED file versus TVC with the untrimmed design BED file. We also see a reduction in the number of variants that are called in the intersection of the NIST 2.18 truth set from ~37,000 to ~27,000. This lower number is primarily due to the removal of the poorly performing flanking intronic sequences and is accompanied by a marked reduction in INDEL FP and SNP FP.

The following table contains the results of using the NIST 2.18 truth set on an example NA12878 sample sequencing run in duplex with Ion PI™ Hi-Q™ on the Ion Proton™ System, with and without trimming.

**Table 1 SNP**

TVC Version	TP	FP	FN	Sensitivity	PPV	FP/MB
4.4	26732	228	319	98.8	99.2	6.4
4.4 (no trim)	36086	483	627	98.3	98.7	11.1

**Table 2 INDELS**

TVC Version	TP	FP	FN	Sensitivity	PPV	FP/MB
4.4	438	63	173	71.7	87.4	1.8
4.4 (no trim)	748	135	504	59.7	84.7	3.1

**Table 3 Combined**

TVC Version	Sensitivity	PPV	FP/MB
4.4	98.2	98.9	8.2
4.4 (no trim)	97.0	98.3	14.1

## Conclusion

We have excluded poorer performing regions of the Ion AmpliSeq™ Exome that are enriched for false positives or have coverage failures. We recommend that you use our Effective Regions BED file to improve your sensitivity and specificity.

**Note:** Ion Reporter™ Software workflows incorporate the trimmed BED file.

## Download effective regions BED file

After a variantCaller plugin run completes, you can access variantCaller run results from the run report screen in Torrent Suite™ Software. On the run report screen, in the variantCaller section you can:

- Review the summary of the variantCaller plugin run for each barcode used.
- Review the library type, reference genome, targeted regions, hotspots file, and parameter settings that were used in the run.
- Download data files for all barcodes and each individual barcode or sample.

You can also access the detailed variantCaller plugin summary report for each barcode or sample from the variantCaller section on the run report screen. In the detailed variantCaller plugin summary report browser, you can:

- View variant call information by allele for the specific barcode, such as allele location on the chromosome, allele annotations, coverage metrics, and quality metrics.
- Review the library type, reference genome, target regions, hotspots, and parameter settings that were used in the run.
- Download BED files and the parameters file that are used for the specific barcode.
- Download BAM and BAI files for the mapped and TVC-processed reads.
- Download data files for variant calls and coverage for the specific barcode.
- View variant calls in IGV.

The variantCaller plugin supports SNPs, MNPs, INDELS, and complex alleles as input candidates at genomic positions with the target regions file. If the variant is outside of the target regions, then the variant is not generated as a candidate and is not further evaluated, even if the variant is specified in the hotspots file.

1. In the **Data** tab, click **Completed Runs & Reports**.
2. In the **Completed Runs & Reports** list, find the run of interest, then click the report link in the **Report Name** column in the row of the run.
3. In the left navigation menu, click **variantCaller** to navigate to the variantCaller results section.
  - If the sequencing run contains barcodes, the plugin report includes a list of the barcodes that were used and file download options for all barcodes and each individual barcode.

variantCaller (v5.4.0.46) variantCaller.html Completed

Library type: AmpliSeq  
Reference genome: hg19  
Targeted regions: AmpliSeqExome.20141113 designed  
Hotspot regions: none  
Configuration: Generic - Proton P1 or S5/S5XL (540) - Germ Line - Low Stringency  
Output directory: variantCaller\_out.2410835

Download all barcodes: [VCF.ZIP](#) [XLS.ZIP](#) [XLS](#) [COV](#)

Please note: Variant calling was carried out for all barcodes with reference genome as specified above

Barcode Name	Sample	Variants	Downloads
<a href="#">IonXpress_001</a>	1	37,426	<a href="#">VCF.GZ</a> <a href="#">VCF.GZ.TBI</a> <a href="#">gVCF.GZ</a> <a href="#">gVCF.GZ.TBI</a> <a href="#">XLS</a>
<a href="#">IonXpress_002</a>	2	37,237	<a href="#">VCF.GZ</a> <a href="#">VCF.GZ.TBI</a> <a href="#">gVCF.GZ</a> <a href="#">gVCF.GZ.TBI</a> <a href="#">XLS</a>

20 items per page 1 - 2 of 2 items

**Table 4 Downloads options for all barcodes that were used in the plugin run.**

Download option	Description
VCF.ZIP	A compressed directory that contains separate Variant Call Format (VCF) files for each barcode.
XLS.ZIP	A compressed directory that contains separate XLS files for each barcode.
XLS	A file that contains a list of alleles for all barcodes in a tab-separated file that can be opened by Microsoft™ Excel™.
COV	A file that contains the coverage of the variant call results for all barcodes in tab-separated file format, which can be opened in Microsoft™ Excel™.

- If the sequencing run does not contain barcodes, the plugin report contains information for the sample that is used in the run and the file download options for the sample.

**variantCaller** v5.4.0.46 (2083138) [View Log](#) [Delete](#)  
 Completed 2.19 GB

[variantCaller.html](#)

Library type: AmpliSeq  
 Reference genome: hg19  
 Targeted regions: Oncomine\_BRCA\_Research\_Assay.20170303.designed  
 Hotspot regions: Oncomine\_BRCA\_Research\_Assay.20170316.hotspots.blist.318  
 Configuration: Custom  
 Output directory: variantCaller\_out.2083138  
 Please note: Variant calling was carried out for all barcodes with reference genome as specified above

Sample	Variants	Hotspot Variants	Downloads
<a href="#">Sample_1</a>	26	0	<a href="#">VCF.GZ</a> <a href="#">VCF.GZ.TBI</a> <a href="#">gVCF.GZ</a> <a href="#">gVCF.GZ.TBI</a> <a href="#">XLS</a>

◀ 1 ▶
 Items per page
 1 - 1 of 1 items

- To open the detailed variantCaller plugin report for a specific barcode or sample, do one of the following in the variantCaller run summary table:
  - In the **Barcode name** column in a barcoded sequencing run report, click the barcode name link.
  - In the **Sample** column in a report from a sequencing run that does not contain barcodes, click the sample name link.

## Modify a BED file

You can modify an existing target regions or hotspots BED file, if necessary, to remove regions from the file for which you do not want variants called.

If you modify a BED file, you must modify it *before* it is uploaded to Torrent Suite™ Software. A target regions or hotspots BED file cannot be modified after the file is uploaded.

To modify a BED file:

1. Copy the BED file, then rename the new file in a way that reflects changes you make to the regions being analyzed.
2. Open the BED file with a text editor.
3. Delete the lines for regions you do not want.
4. Save the file.
5. Upload the modified file as described in the help system for your version of Torrent Suite™ Software.

**Note:**

- If the region (or regions) appears in both your target regions BED file and in your hotspots BED file, you must delete the line for those regions from both types of BED file.
- Do not modify a VCF hotspots file. We recommend that you upload it first to convert it to a BED file before modifying it.

## Customer and technical support

Visit [thermofisher.com/support](http://thermofisher.com/support) for the latest service and support information.

- Worldwide contact telephone numbers
- Product support information
  - Product FAQs
  - Software, patches, and updates
  - Training for many applications and instruments
- Order and web support
- Product documentation
  - User guides, manuals, and protocols
  - Certificates of Analysis
  - Safety Data Sheets (SDSs; also known as MSDSs)

**Note:** For SDSs for reagents and chemicals from other manufacturers, contact the manufacturer.

## Limited product warranty

Life Technologies Corporation and/or its affiliate(s) warrant their products as set forth in the Life Technologies' General Terms and Conditions of Sale at [www.thermofisher.com/us/en/home/global/terms-and-conditions.html](http://www.thermofisher.com/us/en/home/global/terms-and-conditions.html). If you have any questions, please contact Life Technologies at [www.thermofisher.com/support](http://www.thermofisher.com/support).





Life Technologies Corporation | 200 Oyster Point Blvd | South San Francisco, CA 94080 | USA

For descriptions of symbols on product labels or product documents, go to [thermofisher.com/symbols-definition](https://www.thermofisher.com/symbols-definition).

The information in this guide is subject to change without notice.

**DISCLAIMER:** TO THE EXTENT ALLOWED BY LAW, THERMO FISHER SCIENTIFIC INC. AND/OR ITS AFFILIATE(S) WILL NOT BE LIABLE FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING YOUR USE OF IT.

**Revision history:** Pub. No. MAN0018645

Revision	Date	Description
B.0	14 August 2020	Revision of <i>Ion AmpliSeq™ Exome Hi-Q™ Effective Regions File Technical Note</i> (Released: 27 February 2015).

Important Licensing Information: These products may be covered by one or more Limited Use Label Licenses. By use of these products, you accept the terms and conditions of all applicable Limited Use Label Licenses.

Trademarks: All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. Microsoft™ Excel™ is a trademark of Microsoft Corp.

©2020 Thermo Fisher Scientific Inc. All rights reserved.

[thermofisher.com/support](https://www.thermofisher.com/support) | [thermofisher.com/askaquestion](https://www.thermofisher.com/askaquestion)

[thermofisher.com](https://www.thermofisher.com)

14 August 2020

**ThermoFisher**  
SCIENTIFIC