

## TECHNICAL NOTE

### The Ion AmpliSeq™ MH-74 Plex Microhaplotype Research Panel

#### Introduction

The analysis of alternative marker types can complement STR-CE results for challenging samples. Microhaplotypes have arisen as an alternative marker to traditional STR-CE typing due to the additive-enhanced mixture analysis and biogeographical ancestry prediction capabilities that favor its use for forensic samples. The Ion AmpliSeq MH-74 Plex Research Panel (MH-74 panel) is a 157 – 325 bp assay covering 74 microhaplotypes (230 SNPs) selected from a set of 130 microhaplotypes previously characterized by the Kidd Laboratory [1]. The MH-74 panel is available as a community panel on [www.ampliseq.com](http://www.ampliseq.com) and was described in a 2020 publication by Oldoni, *et al* [2].

This technical note describes an analysis workflow for the MH-74 panel using Torrent Suite™ Software and the 'HID-Microhaplotype-Research-PluginV1.5'.

#### Laboratory Processing

Processing samples for the MH-74 panel can be performed as per the instructions for custom Ion AmpliSeq SNP Panels in the *Precision ID SNP Panels with the HID IonS5™/HID Ion GeneStudio™ S5 System Application Guide* (Publication Number MAN0017767). Libraries may be prepared either manually or with the Ion Chef™ System, as per these instructions.

#### Data Analysis

Data from the MH-74 panel may be analyzed in Torrent Suite Software with the HID-Microhaplotype-Research-PluginV1.5. For the analysis, the following target and hotspot files are also required:

- mh74\_target\_v1.0.bed
- mh74\_hotspot\_v1.0.bed

These files and the HID-Microhaplotype-Research-PluginV1.5 are available from your local Thermo Fisher Scientific support representative.

Analysis of MH-74 data in this plugin has been tested with Torrent Suite Software v5.10.0. The analysis may work with other versions of Torrent Suite, but this has not been tested.

To perform analysis, first install the Microhaplotype-Research-PluginV1.5 in the 'Plugins' section of Torrent Suite Software and install the target and hotspot files named above in the 'References' section of Torrent Suite. On installation, associate the target and hotspot files with the hg19 human genome reference.

Next, open the run report with the MH-74 data to be analyzed. Select 'Plugins > 'Select Plugin to Run' from the dropdown menu at the top of the report, and in the resulting window, select 'HID-Microhaplotype-Research-PluginV1.5'.



The Configure Plugin screen will appear:

Select 'hg19' in the 'Reference' dropdown at the top of the screen. The Target and Hotspot dropdown boxes will then appear.

Select 'mh74\_target\_v1.0.bed' and 'mh74\_hotspot\_v1.0.bed' in each of these respectively as shown below. The bed files are specifically formatted to work with the microhaplotype plugin and should not be altered.

Configure Plugin

HID Microhaplotype Research PluginV1.5

Reference: hg19 (Homo sapiens) ▼

Target: mh74\_target\_v1.0 ▼

Hotspot: mh74\_hotspot\_v1.0 ▼

Parameters

Global Parameters

mh22KK-061 ▼

Min total read coverage per position	20	Min total read coverage per position	20
Min # of allele count to include in report	5	Min # of allele count to include in report	5
Min allele frequency (heterozygous)	10	Min allele frequency (heterozygous)	10
Min allele frequency (homozygous)	90	Min allele frequency (homozygous)	90

Ignore an allele if a read does not cover the entire marker ☒

Do not attempt to find new positions in target (rare alleles) ☒

Write a .bam file with allele tags ☒

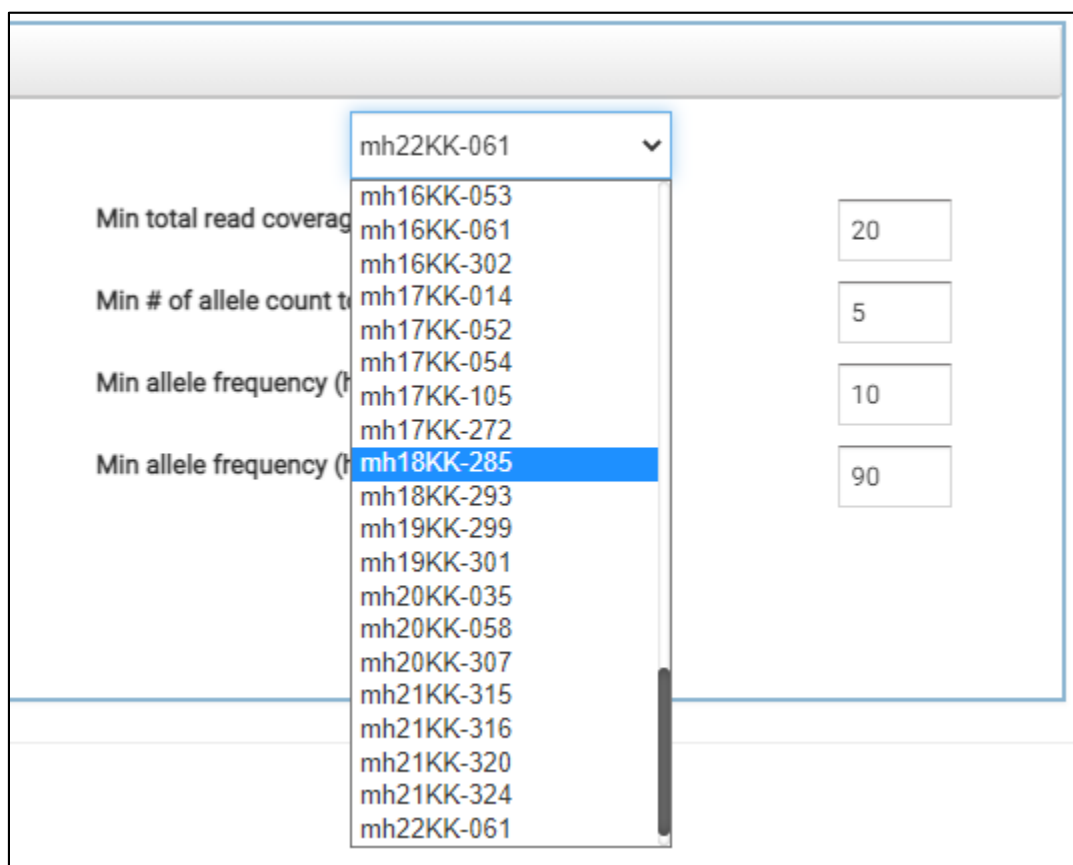
Do just a short testrun with only 2 barcodes ☐

Submit

The desired analysis parameters can be configured in the 'Parameters' section of the screen. By default, the options 'Ignore an allele if a read does not cover the entire marker', 'Do not attempt to find new positions in target (rare alleles)', and 'Write a .bam file with allele tags' are checked. The results of unchecking these options will be discussed later in this document.

'Global Parameters' are shown to the left of the screen. Settings can be chosen here that will be applied to every locus in the panel for the analysis.

If locus-specific settings are desired, these can be selected in the fields on the right. After the target file is selected, the markers in the MH-74 panel will populate the dropdown box as shown. Any changes made to the locus-specific will persist even if subsequent changes are made to thresholds in the Global Parameters section.



mh22KK-061 ▼

- mh16KK-053
- mh16KK-061
- mh16KK-302
- mh17KK-014
- mh17KK-052
- mh17KK-054
- mh17KK-105
- mh17KK-272
- mh18KK-285**
- mh18KK-293
- mh19KK-299
- mh19KK-301
- mh20KK-035
- mh20KK-058
- mh20KK-307
- mh21KK-315
- mh21KK-316
- mh21KK-320
- mh21KK-324
- mh22KK-061

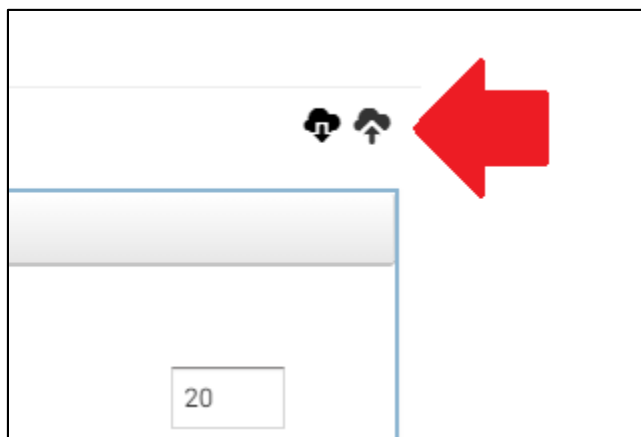
Min total read coverage: 20

Min # of allele count: 5

Min allele frequency (f): 10

Min allele frequency (f): 90

From this list, the desired marker(s) can be chosen, and analysis parameters entered. If desired, analysis parameters can be saved and loaded from a json file by use of these buttons:



Save Load

20

A description of the available analysis parameters is as follows:

Analysis Parameter	Description	Default Value
<b>Min total read coverage per position</b>	Minimum total read coverage to be considered for variant calling. If total number of reads for a marker are below this value, alleles will not be called at that marker.	20
<b>Min # of allele count to include in report</b>	Minimum read coverage for an allele. If the number of reads for an allele are below this value, the alleles will not be called.	5
<b>Min allele frequency (heterozygous)</b>	Minimum value for allele frequency for heterozygous markers. If two or more alleles are detected at a marker, any single allele must have coverage of at least this percentage of total reads at the marker to be called.	10
<b>Min allele frequency (homozygous)</b>	Minimum value for allele frequency for homozygous markers. A single allele at a marker must have coverage of at least this percentage of total reads at the marker to be called.	90
<b>Ignore an allele if a read does not cover the entire marker</b>	If true, only reads that cover all SNPs in the marker will be used. If false, the positions not entirely covered will be marked with *	True
<b>Do not attempt to find new positions in target ('rare' alleles)</b>	If true, only the SNPs in the hotspot will be considered for allele calling. If false, will attempt to find new variants in the target	True
<b>'Write a .bam file with allele tags'</b>	If true produces an output BAM file for the plugin run	True
<b>Do just a short test run with only 2 barcodes</b>	If true will do a short run for testing purposes.	False

When the desired parameters have been selected, press 'Submit'. The plugin will run. When complete, results appear in the Plugins section of the Run report:



To view the results, click on the 'report.html' button shown in Plugin result. The following screen will appear.

Results Overview					
Download result .zip files: <a href="#">Batch ZIP (no BAM)</a> <a href="#">Batch Result ZIP file with BAM files</a>					
Combined result files: <a href="#">SNP combined (no rare)</a>					
HTML File	EXCEL File	CSV File	SNP CSV File (no rare)	BAM File	BAI File
<a href="#">IonCode_0119.html</a>	<a href="#">IonCode_0119.xlsx</a>	<a href="#">IonCode_0119.csv</a>	<a href="#">IonCode_0119_SNP_norare.csv</a>	<a href="#">IonCode_0119_out.bam</a>	<a href="#">IonCode_0119_out.bam.bai</a>
<a href="#">IonCode_0117.html</a>	<a href="#">IonCode_0117.xlsx</a>	<a href="#">IonCode_0117.csv</a>	<a href="#">IonCode_0117_SNP_norare.csv</a>	<a href="#">IonCode_0117_out.bam</a>	<a href="#">IonCode_0117_out.bam.bai</a>
<a href="#">IonCode_0123.html</a>	<a href="#">IonCode_0123.xlsx</a>	<a href="#">IonCode_0123.csv</a>	<a href="#">IonCode_0123_SNP_norare.csv</a>	<a href="#">IonCode_0123_out.bam</a>	<a href="#">IonCode_0123_out.bam.bai</a>
<a href="#">IonCode_0122.html</a>	<a href="#">IonCode_0122.xlsx</a>	<a href="#">IonCode_0122.csv</a>	<a href="#">IonCode_0122_SNP_norare.csv</a>	<a href="#">IonCode_0122_out.bam</a>	<a href="#">IonCode_0122_out.bam.bai</a>
<a href="#">IonCode_0121.html</a>	<a href="#">IonCode_0121.xlsx</a>	<a href="#">IonCode_0121.csv</a>	<a href="#">IonCode_0121_SNP_norare.csv</a>	<a href="#">IonCode_0121_out.bam</a>	<a href="#">IonCode_0121_out.bam.bai</a>
<a href="#">IonCode_0120.html</a>	<a href="#">IonCode_0120.xlsx</a>	<a href="#">IonCode_0120.csv</a>	<a href="#">IonCode_0120_SNP_norare.csv</a>	<a href="#">IonCode_0120_out.bam</a>	<a href="#">IonCode_0120_out.bam.bai</a>
<a href="#">IonCode_0118.html</a>	<a href="#">IonCode_0118.xlsx</a>	<a href="#">IonCode_0118.csv</a>	<a href="#">IonCode_0118_SNP_norare.csv</a>	<a href="#">IonCode_0118_out.bam</a>	<a href="#">IonCode_0118_out.bam.bai</a>
<a href="#">IonCode_0124.html</a>	<a href="#">IonCode_0124.xlsx</a>	<a href="#">IonCode_0124.csv</a>	<a href="#">IonCode_0124_SNP_norare.csv</a>	<a href="#">IonCode_0124_out.bam</a>	<a href="#">IonCode_0124_out.bam.bai</a>

Results for each sample can be viewed in the browser with the links in the 'HTML File' column. Alternatively, the same data can be downloaded in xlsx or csv files. The result screen also contains a link to BAM / BAI files for the samples and a link to another version of the csv file with 'norare' appended to the file name. The 'norare' file contains a condensed version of the genotype of the sample in question but does not include any rare alleles that may have been found if the 'Do not attempt to find new positions in target' parameter described above was unchecked. Only SNPs present in the hotspot file are included in the 'norare' file.

Results in the .xlsx and .csv files are grouped by the microhaplotype as determined by the microhaplotype identifier. Data displayed for each microhaplotype are, chromosome, SNP positions, phased SNP allele calls, counts for microhap alleles detected, and percent coverage for the alleles.

Results in the first tab, 'MH Alleles', with default options, are shown below with annotations.

MH identifier	Chromosome	# SNP sites	Base position of SNPs		# Coverage	% Coverage
<b>mh01KK-106</b>	<b>chr1</b>	<b>4 locations</b>				
<b>Location</b>	<b>4227464</b>	<b>4227561</b>	<b>4227624</b>	<b>4227634</b>	<b>Count</b>	<b>Percent</b>
<b>Reference</b>	<b>C</b>	<b>A</b>	<b>G</b>	<b>G</b>	<b>46115</b>	
<b>Allele 1</b>	T	A	G	G	23314	50.556
<b>Allele 2</b>	C	A	G	A	22801	49.444

The second tab 'MH Alleles Condensed 1' contains the allele data in column format with the number of columns generated based on the number of alleles detected.

Marker	# Alleles	Allele	Count	Allele	Count
mh01KK-172 chr1	1	ACG	30051		
mh01KK-001 chr1	2	CG	15709	CA	13513
mh01KK-106 chr1	2	TAGG	23314	CAGA	22801
mh01KK-205 chr1	2	TTAG	17039	TTGG	16885
mh01KK-211 chr1	1	ATT	38745		
mh01KK-117 chr1	2	CACC	17543	AGCT	10804
mh01KK-002 chr1	2	GG	14269	AA	12959

If the option 'Ignore an allele if a read does not cover the entire marker' is unchecked the .xlsx file will display alleles where the SNP is absent from the allele with an asterisk '\*', as shown below. These partial alleles will generate new alleles with their own coverages in the result file, frequently triggering a 'Mixture?' detection warning.

mh01KK-117	chr1	4 locations					
Location	204633340	204633397	204633500	204633526	Count	Percent	Mixture?
Reference	A	A	G	C	30712		
	C	A	C	C	17543	57.121	
	A	G	C	T	10804	35.178	
	A	G	*	*	1284	4.181	
	*	*	C	T	1081	3.52	

If the option 'Do not attempt to find new positions in target ('rare' alleles)' is unchecked the .xlsx file will contain additional tabs where new SNPs (i.e. SNPs not present in the hotspot file) are detected meeting the global thresholds. The tab 'MH Alleles (with rare)' will contain the data including these SNPs. They will be highlighted in pink with an asterisk '\*' following the position as shown below. These rare SNPs may represent a new allele (as in the sample below) or may also represent noise. Rare SNP sites should be evaluated by looking at the IGV alignment to confirm their presence.

mh20KK-058	chr20	7 locations							
Location	48844260	48844293	48844348*	48844365	48844368*	48844389*	48844402*	Count	Percent
Reference	T	A	G	T	T	T	T	51063	
	C	A	T	C	T	G	C	25745	50.418
	C	A	T	C	A	G	C	25318	49.582

If the option 'Do not attempt to find new positions in target ('rare' alleles)' is unchecked the 'report.html' button link will also produce a page with additional links to the rare SNP data in .csv format.

Results Overview						
Download result .zip files: <a href="#">Batch ZIP (no BAM)</a> <a href="#">Batch Result ZIP file with BAM files</a> Combined result files: <a href="#">SNP combined (no rare)</a> <a href="#">SNP combined (rare)</a>						
HTML File	EXCEL File	CSV File	SNP CSV File (no rare)	SNP CSV File (rare)	BAM File	BAI File
<a href="#">IonCode_0119.html</a>	<a href="#">IonCode_0119.xlsx</a>	<a href="#">IonCode_0119.csv</a>	<a href="#">IonCode_0119_SNP_norare.csv</a>	<a href="#">IonCode_0119_SNP_rare.csv</a>	<a href="#">IonCode_0119_out.bam</a>	<a href="#">IonCode_0119_out.bam.bai</a>
<a href="#">IonCode_0117.html</a>	<a href="#">IonCode_0117.xlsx</a>	<a href="#">IonCode_0117.csv</a>	<a href="#">IonCode_0117_SNP_norare.csv</a>	<a href="#">IonCode_0117_SNP_rare.csv</a>	<a href="#">IonCode_0117_out.bam</a>	<a href="#">IonCode_0117_out.bam.bai</a>
<a href="#">IonCode_0123.html</a>	<a href="#">IonCode_0123.xlsx</a>	<a href="#">IonCode_0123.csv</a>	<a href="#">IonCode_0123_SNP_norare.csv</a>	<a href="#">IonCode_0123_SNP_rare.csv</a>	<a href="#">IonCode_0123_out.bam</a>	<a href="#">IonCode_0123_out.bam.bai</a>
<a href="#">IonCode_0122.html</a>	<a href="#">IonCode_0122.xlsx</a>	<a href="#">IonCode_0122.csv</a>	<a href="#">IonCode_0122_SNP_norare.csv</a>	<a href="#">IonCode_0122_SNP_rare.csv</a>	<a href="#">IonCode_0122_out.bam</a>	<a href="#">IonCode_0122_out.bam.bai</a>
<a href="#">IonCode_0121.html</a>	<a href="#">IonCode_0121.xlsx</a>	<a href="#">IonCode_0121.csv</a>	<a href="#">IonCode_0121_SNP_norare.csv</a>	<a href="#">IonCode_0121_SNP_rare.csv</a>	<a href="#">IonCode_0121_out.bam</a>	<a href="#">IonCode_0121_out.bam.bai</a>
<a href="#">IonCode_0120.html</a>	<a href="#">IonCode_0120.xlsx</a>	<a href="#">IonCode_0120.csv</a>	<a href="#">IonCode_0120_SNP_norare.csv</a>	<a href="#">IonCode_0120_SNP_rare.csv</a>	<a href="#">IonCode_0120_out.bam</a>	<a href="#">IonCode_0120_out.bam.bai</a>
<a href="#">IonCode_0118.html</a>	<a href="#">IonCode_0118.xlsx</a>	<a href="#">IonCode_0118.csv</a>	<a href="#">IonCode_0118_SNP_norare.csv</a>	<a href="#">IonCode_0118_SNP_rare.csv</a>	<a href="#">IonCode_0118_out.bam</a>	<a href="#">IonCode_0118_out.bam.bai</a>
<a href="#">IonCode_0124.html</a>	<a href="#">IonCode_0124.xlsx</a>	<a href="#">IonCode_0124.csv</a>	<a href="#">IonCode_0124_SNP_norare.csv</a>	<a href="#">IonCode_0124_SNP_rare.csv</a>	<a href="#">IonCode_0124_out.bam</a>	<a href="#">IonCode_0124_out.bam.bai</a>

The html link opens a new browser tab with the data for the sample selected. If the plugin was launched with the 'Do not attempt to find new positions in target ('rare' alleles)' box checked the data will be presented in four columns: Location, Allele, Coverage (count), and Coverage % for each of the microhaplotypes in the panel. Links are also provided for downloading the result .xlsx, norare.csv, and out.bam files.

IonCode_0119_rawlib			
Excel report: <a href="#">IonCode_0119_rawlib.xlsx</a> SNP csv: <a href="#">IonCode_0119_rawlib_SNP_norare.csv</a> BAM: <a href="#">IonCode_0119_rawlib_out.bam</a>			
mh01KK-172 1:1486834-1487059			
Location	Allele	Coverage (count)	Coverage %
chr1:1486834	ACG	30051	100
mh01KK-001 1:3743132-3743319			
Location	Allele	Coverage (count)	Coverage %
chr1:3743132	CG	15709	53.76
chr1:3743132	CA	13513	46.24

If the 'Do not attempt to find new positions in target ('rare' alleles)' box is unchecked the data will be presented in five columns: Location, Allele, Coverage (count), Coverage %, and Allele with positions for rare SNPs for each of the microhaplotypes in the panel. When a rare SNP is present the Allele column display introduces a dash '-' indicating the next base displayed is a SNP not present in the hotspot file. The chromosome position of the rare SNP is displayed immediately preceding the SNP base in the 'Allele with positions for rare SNPs' column. In the example below for mh01KK-172 the 'A' SNP at position 1486846, 'G' SNP at position 1486938, and 'T' SNP at position 1486939 are all true variants or reference calls, however, they are most likely linked with the SNPs in the hotspot file and do not distinguish a new allele. In contrast the 'T' at position 1486846 and



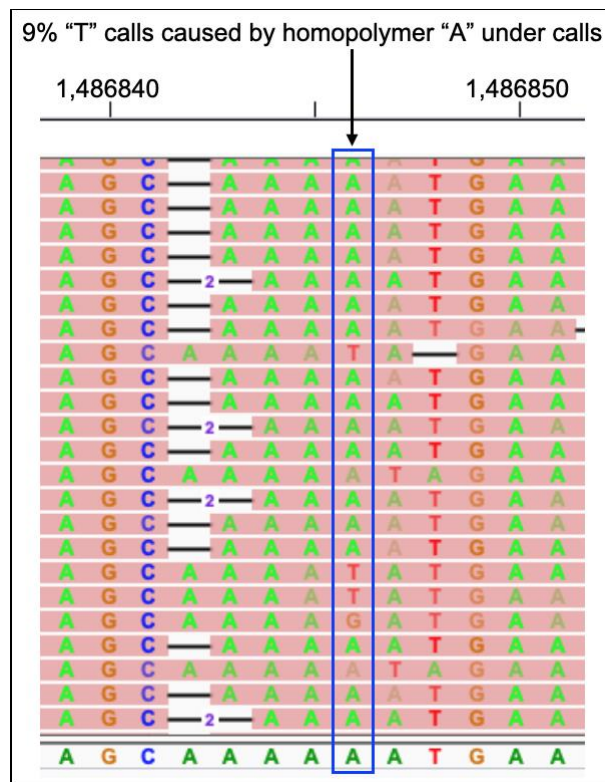
'G' at position 1486939 are the result of sequence noise and have generated low coverage false alleles. The 'A' SNP at position 1486846 is not a variant, it matches the reference hg19 call, however, the 'T' at position 1486846 is a variant when compared to the hg19 reference sequence and therefore all alleles of this microhaplotype will have a call out for position 1486846.

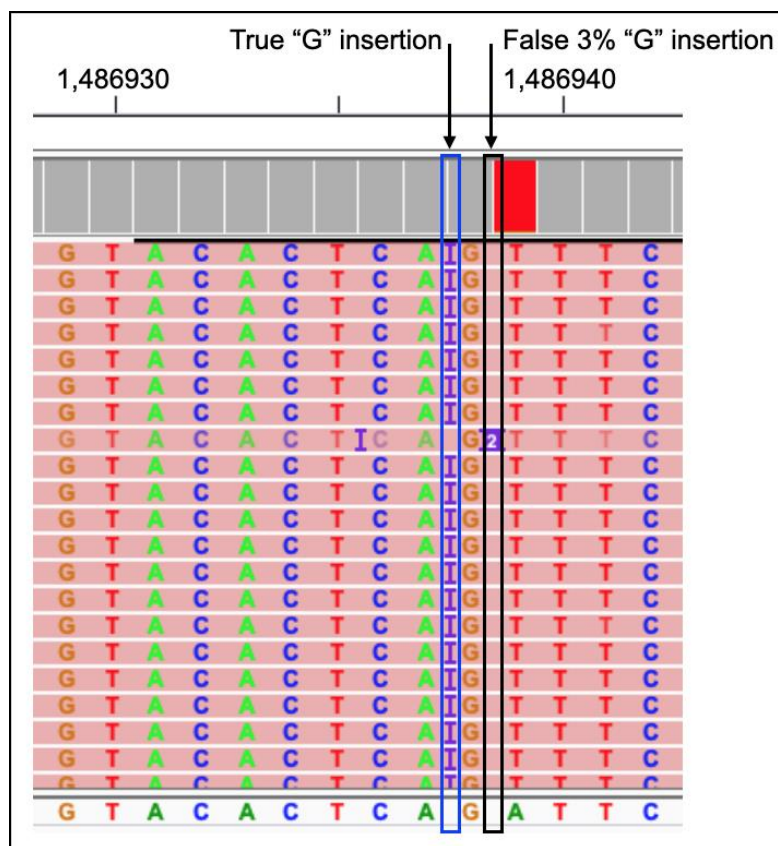
IonCode_0119_rawlib				
Excel report: <a href="#">IonCode_0119_rawlib.xlsx</a> SNP csv: <a href="#">IonCode_0119_rawlib_SNP_norare.csv</a> BAM: <a href="#">IonCode_0119_rawlib_out.bam</a>				
mh01KK-172 1:1486834-1487059				
Location	Allele	Coverage (count)	Coverage %	Allele with positions for rare SNPs
chr1:1486834	A-AC-G-TG	26394	88.15	A-1486846AC-1486938G-1486939TG
chr1:1486834	A-TC-G-TG	2692	8.99	A-1486846TC-1486938G-1486939TG
chr1:1486834	A-AC-G-GG	856	2.86	A-1486846AC-1486938G-1486939GG
mh01KK-001 1:3743132-3743319				
Location	Allele	Coverage (count)	Coverage %	Allele with positions for rare SNPs
chr1:3743109	-GC-GG	15504	53.86	-3743109GC-3743151GG
chr1:3743109	-GC-AA	13283	46.14	-3743109GC-3743151AA
mh01KK-106 1:4227464-4227634				
Location	Allele	Coverage (count)	Coverage %	Allele with positions for rare SNPs
chr1:4227464	TAGG	23314	50.56	TAGG
chr1:4227464	CAGA	22801	49.44	CAGA

As mentioned previously rare SNPs can indicate a new allele and add a distinction to the genotype obtained from a sample. Their presence should be confirmed by investigating their alignment to the genome in a viewer like IGV.

Looking at the alignments for the rare SNPs indicated above it becomes clear the 'T' at position 1486846 is caused by an under calling of the neighboring 'A' homopolymer stretch as shown in the figure to the right.

The 'G' insertion seen at 1486938 (figure next page) is correct but the 3% 'G' call at 1486939 is caused by a 'C' over call pushing the 'G' base out of position.

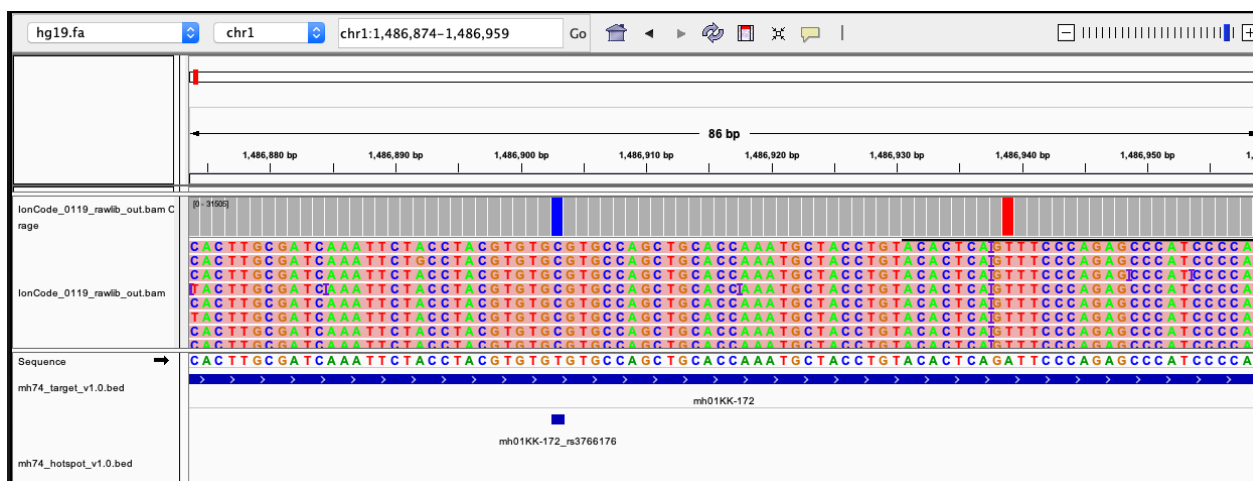




If the 'Write a .bam file with allele tags' is checked the plugin will try and produce a out.bam file for each sample. If the 'Write a .bam file with allele tags' is unchecked the links provided for the out.bam files will generate a 'Not Found' error in the browser.

To view the alignment of a sample that has been processed by the HID-Microhaplotype-Research-PluginV1.5 click the click the 'BAM File' and BAI File' links for the appropriate sample displayed on the 'Results Overview' page or alternatively download all the BAM files by clicking the 'Batch Result ZIP file with BAM files' link on the 'Results Overview' page. Launch the IGV viewer application and make sure the genome selected is hg19. Load the out.bam file for the sample by using the 'File' > 'Load from File...' option in the IGV menu. Select the out.bam file and click 'Open'. The hotspot and target bed files can also be uploaded and used as visual aids when viewing the alignment in IGV

Once the data has been loaded it is easiest to jump to the microhaplotype of interest by using the coordinates in the target bed file to jump to that location in the bam file alignment. On the next page is the result of uploading the bam and bed files, using the coordinates for mh01KK-172 to display the alignment, and then moving the sequence to the SNP and hotspot sites of interest.



## References

- [1] K.K. Kidd, W.C. Speed, A.J. Pakstis, D.S. Podini, R. Lagace, J. Chang, S. Wootton, E. Haigh, U. Soundararajan, Evaluating 130 microhaplotypes across a global set of 83 populations, Forensic Science International: Genetics. 29 (2017) 29–37.
- [2] F. Oldoni, D. Bader, C. Fantinato, S.C. Wootton, R. Lagacé, K.K. Kidd, D. Podini. A sequence-based 74plex microhaplotype assay for analysis of forensic DNA mixtures, Forensic Science International: Genetics 49 (2020) 102367.

## Revision History

Revision	Date	Description
A	20 October 2021	Initial publication