ThermoFisher SCIENTIFIC
The world leader in serving science

TECHNICAL NOTE

Analysis of the Ion AmpliSeq™ Body Fluid Identification RNA Research Panel and Ion AmpliSeq™ Body Fluid Identification DNA Research Panel

Table of Contents

Introduction

The evidentiary value of DNA profiles varies depending on the context in which the DNA was found. Linking a DNA profile to a particular cellular phenotype may aid in assessing evidentiary relevance and value. The Ion AmpliSeq™ Body Fluid Identification RNA Research Panel (BFID-mRNA panel) is a 90 – 245 bp assay targeting 23 body-fluid specific genes with 36 amplicons. Within 20 genes are 46 coding region SNPs (cSNPs) for human identification. To serve as both a body fluid identification assay and body fluid association assay, a second adjunct panel was developed targeting the 46 cSNPs in genomic DNA. The Ion AmpliSeq™ Body Fluid Identification DNA Research Panel (BFID-gDNA panel) is a 210 – 275 bp assay targeting 36 amplicons across 13 chromosomes. The 36 amplicons contain the 46 SNPs corresponding to the cSNPs in the mRNA version of the panel.

The cSNPs can be evaluated for concordance between both panels. This paired panel approach enables the identification of the body fluid type and the donor of the fluid. For example, in a blood and semen mixture sample, the assay could allow the association of the blood with donor A and the semen with donor B. Using cSNPs in various body fluid genes in conjunction with gDNA SNPs for the Ion AmpliSeq panels was evaluated by the Ballantyne & Haas laboratories in 2019 (Hanson, *et al* [1]) and further characterized and enhanced in their 2023 publication (Hanson, *et al* [3]).

This technical note describes an analysis workflow for the BFID-mRNA and BFID-gDNA panels using Ion Torrent next-generation sequencing and supporting software.

Required Files

The following file package is required for the analysis of the BFID-mRNA/gDNA panels and can be obtained by contacting your local Thermo Fisher Scientific HID support representative:

- IonAmpliseq_BFID_Analysis_Files_v1.0.zip

Laboratory Processing

Laboratory processing of samples for the BFID-gDNA panel can be performed as per the instructions for custom Ion AmpliSeq SNP Panels in the *Precision ID SNP Panels with the HID Ion S5™/ HID Ion GeneStudio™ S5 System Application Guide* (Publication Number MAN0017767). For processing samples for the BFID-mRNA panel, the RNA sample must first be reverse transcribed to produce cDNA. Libraries may be prepared manually or with the Ion Chef™ System.

When setting up the run plan for the panels, use reference genome and target files as in Table 1.

Table 1. Reference and Target files for the BFID panels. The h19 reference is built-in to Torrent Suite Software (TSS). The analysis file package listed on page 2 includes the other required files. If necessary, consult Chapter 10 of the TSS User Guide for instructions on installing these files in TSS [4].

Panel	Reference Genome	Target File
BFID-gDNA panel	hg19	BFID_gDNA_target_v1.bed
BFID-mRNA panel	BFID_mRNA_Genes_v1.fa	BFID_mRNA_target_v1.bed

Note: *Hotspot files are not required for the run plan. These are unnecessary to generate the BAM files produced in primary analysis in Torrent Suite Software. Hotspot files are required later in the analysis after generating the BAM files.*

Data Analysis

Install Java

64-bit Java must be installed to analyze the BFID panels. To check the Java version, either open a Terminal (MacOS) or Command Prompt (Windows) and type:

```
java -version
```

If a display similar to Figure 1 is seen, with the text “64-Bit,” a compatible version of Java is installed.

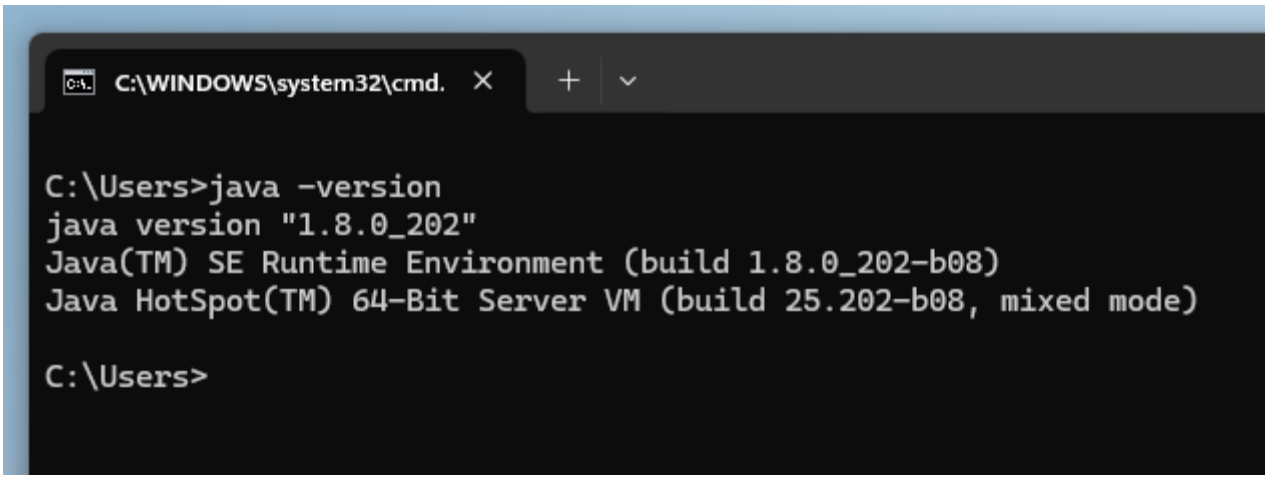

Figure 1. Checking the Java version via Command Prompt on Windows.

If Java is not installed or the version is not 64-bit, it must be installed or updated. This Java version can be obtained from https://www.java.com/en/download/manual.jsp. Select and install the correct Java version for your operating system. After installation, perform the check listed above again to confirm success.

Transfer BAM Files

Mapped BAM files must be transferred from the Torrent server to the local computer for analysis. To do this:

1. Connect to Torrent Suite Software in the browser.
2. Navigate to the run report for the sample(s) in question.
3. Click ‘Output Files’ on the left of the page (Figure 2).
4. Click the ‘BAM’ and ‘BAI’ buttons (Figure 2) for the samples you wish to download.
5. Save the downloaded BAM and BAI files into a folder for analysis.

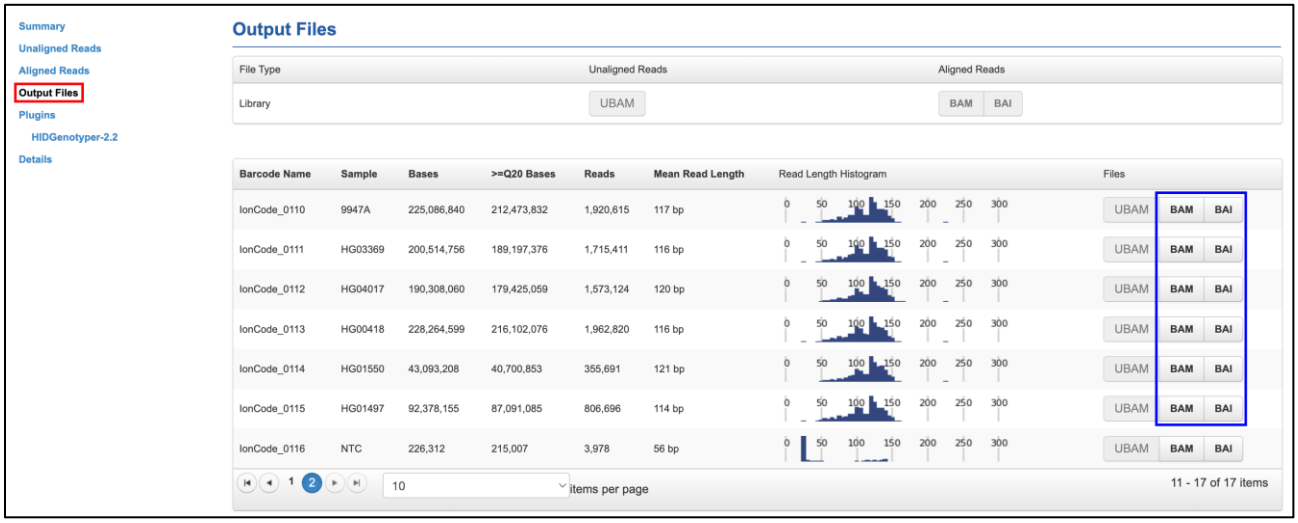

Figure 2. Accessing BAM and BAI Output files. Downloading BAM/BAI files from Torrent Suite Software. 'Output File' shown in red and 'BAM' and 'BAI' buttons for the required samples shown in blue.

Extract and Download Required Files

The mh.jar application file is required to process the BAM files. This file is contained within the IonAmpliseq_BFID_Analysis_Files_v1.0 file package described on Page 2. Extract and save this file to a folder on the local computer before beginning the analysis.

The files in Table 2 are required to analyze the BAM files in the mh.jar application. Except for hg19.fasta, these files are also in the IonAmpliseq_BFID_Analysis_Files_v1.0 file package described on Page 2. Extract and save the files to a folder on the local computer before beginning analysis.

Table 2. Files required for analysis of BFID panel BAM files. These must be saved locally. The h19.fasta file can be downloaded from Torrent Suite Software. The analysis file package listed on page 2 includes the other required files.

File	BFID-gDNA panel	BFID-mRNA panel
Reference Genome File	hg19.fasta	BFID_mRNA_Genes_v1.fa
Reference Index File	hg19.fasta.fai	BFID_mRNA_Genes_v1.fa.fai
Target File	BFID_gDNA_target_v1.bed	BFID_mRNA_target_v1.bed
Hotspot File	BFID_gDNA_hotspots_v1.bed	BFID_mRNA_hotspots_v1.bed

To obtain the hg19.fasta file, download a copy from the Torrent Suite Software. This file must also be saved to the local computer for analysis. To do this:

1. Connect to Torrent Suite Software in the browser.
2. Click the 'Cog' icon at the top right and choose 'Reference Sequences' in the menu that appears.
3. Click 'hg19' in the list of Reference Sequences (Figure 2).
4. Right-click' hg19.fasta' in the middle of the screen and choose 'Save link as…' (Figure 3).
5. Save the file to the desired folder.
6. Ensure that the hg19.fasta.fai index file from the IonAmpliseq_BFID_Analysis_Files_v1.0 file package is saved to the same folder.

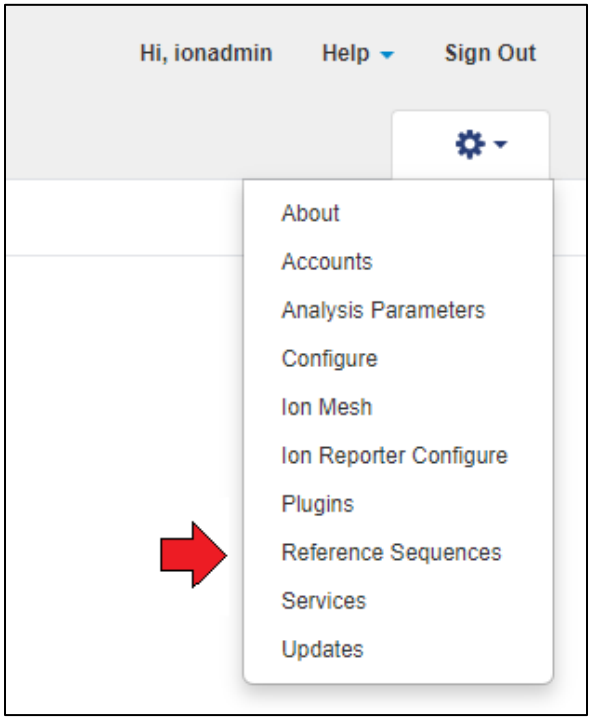

Figure 3. Accessing the 'Reference Sequences' section of Torrent Suite Software

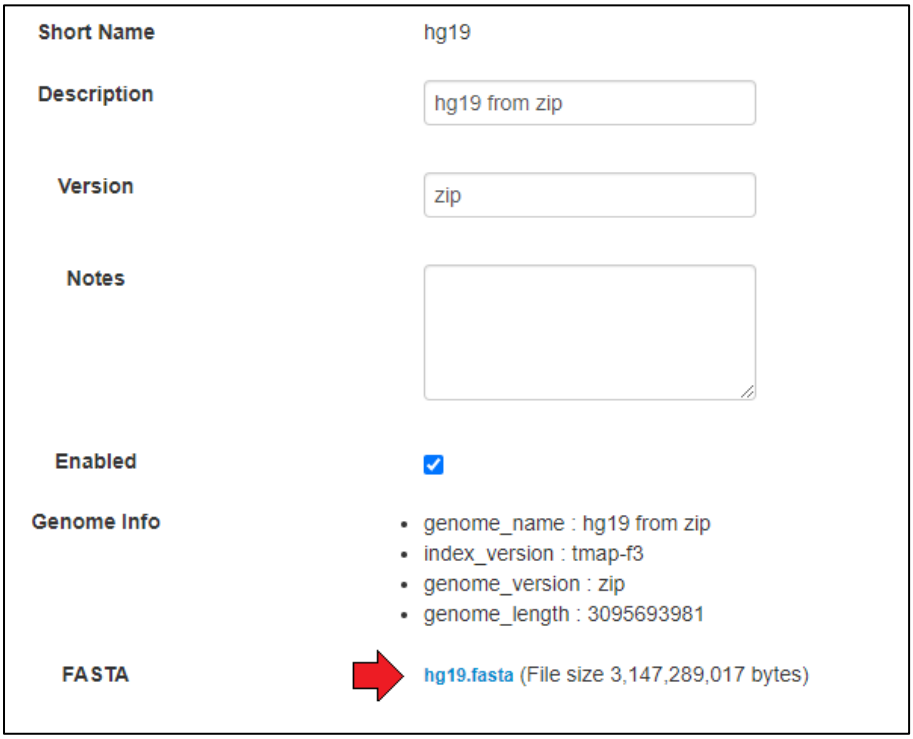

Figure 4. Accessing the hg19 reference file in the 'Reference Sequences' section of Torrent Suite Software.

Note: *The hg19.fasta file is approximately 3 GB in size and may take some time to download, depending on network speed.*

Analysis of the BFID-gDNA Panel

To analyze the BFID-gDNA panel BAM files:

1. Launch the mh.jar application by double-clicking its icon.
2. Select the folder with gDNA BAM files to analyze in the window that appears.
    a. If more than one BAM is found in the selected folder, a pop-up window will appear, allowing the user to select which files to analyze (Figure 5).
    b. Highlight the desired files and click 'Yes'.
3. Select the hg19.fasta file as the reference genome in the next window.
4. Select the BFID_gDNA_target_v1.bed as the target file in the next window.
5. Select the BFID_gDNA_hotspots_v1.bed as the target file in the next window.

Note: *If the target and hotspot files are saved in the same folder as the BAM files, these will be selected automatically.*

6. The 'Program Parameters' window appears. Leave the parameters at the default and click 'Ok' to start the analysis (Figure 6)

Note: *The 'Target application' field in the 'Program Parameters' can be left with blank default values. The software defaults to processing gDNA SNP data.*

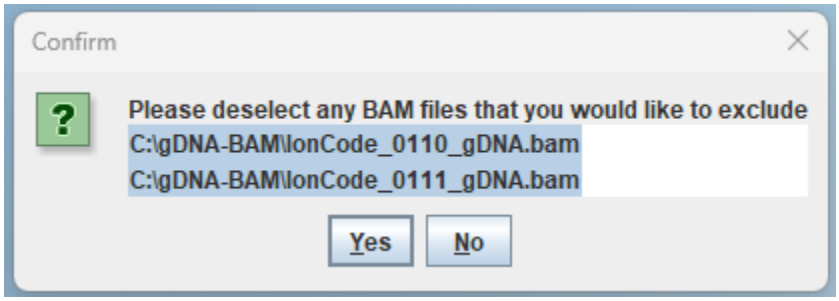

Figure 5. The BAM selection window in mh.jar.

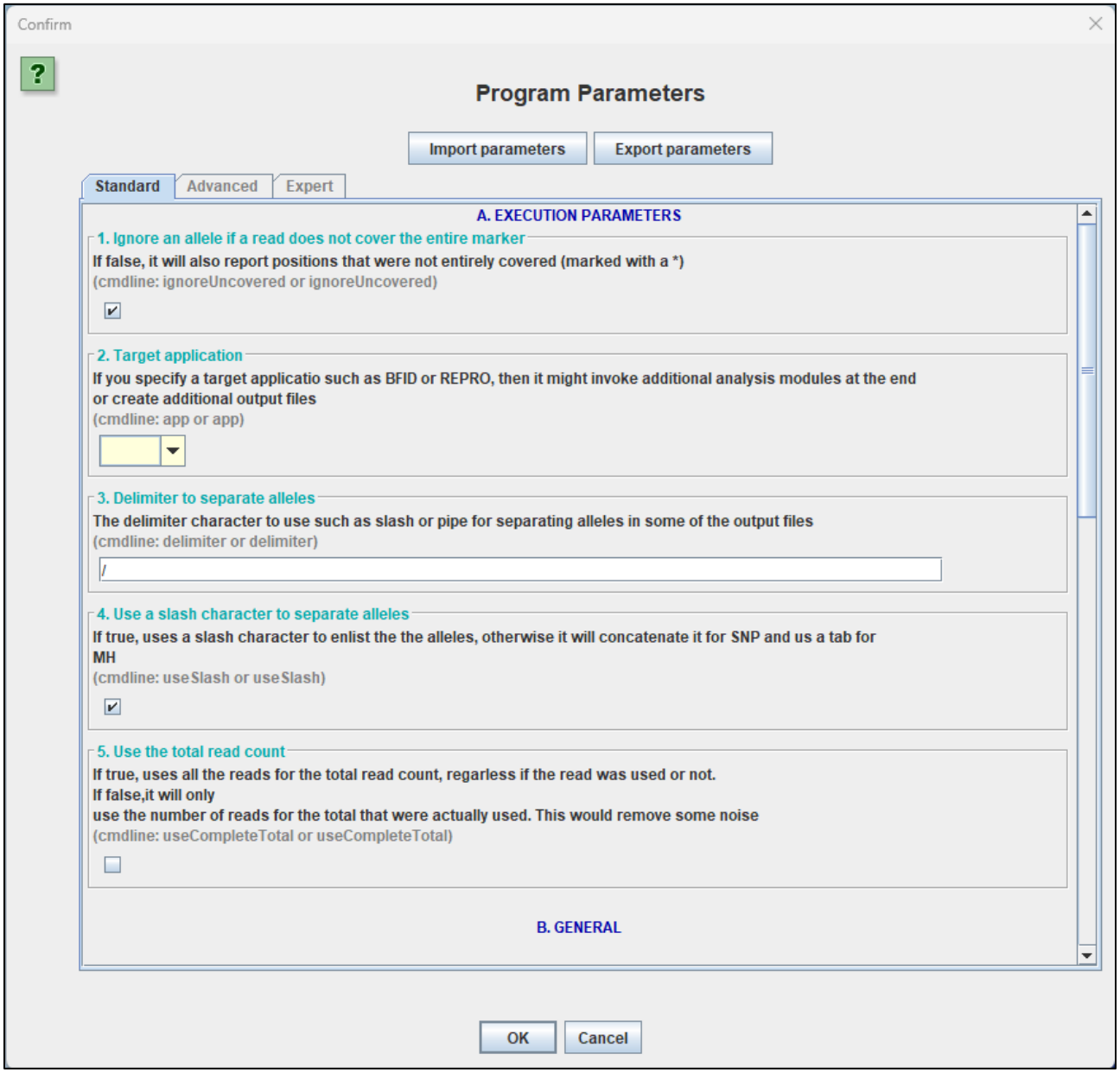

Figure 6. 'Program Parameters' screen when running the mh.jar application for BFID-gDNA samples.

7. The application will then process the BAM files. It can take multiple minutes to process each sample, depending on the file size and the computer's processing power. A pop-up will approximate the completion time if multiple files are being processed. Click 'Ok'. A progress window is displayed during the analysis.

Note: *The estimated time to complete the analysis will often be overestimated– the analysis can be completed much faster than the estimate.*

8. When the analysis is finished, the result for the last sample will be displayed along with the comment "No analysis running" in the bottom left corner (Figure 7). The samples processed are listed at the top with a green check next to each. Close this window.

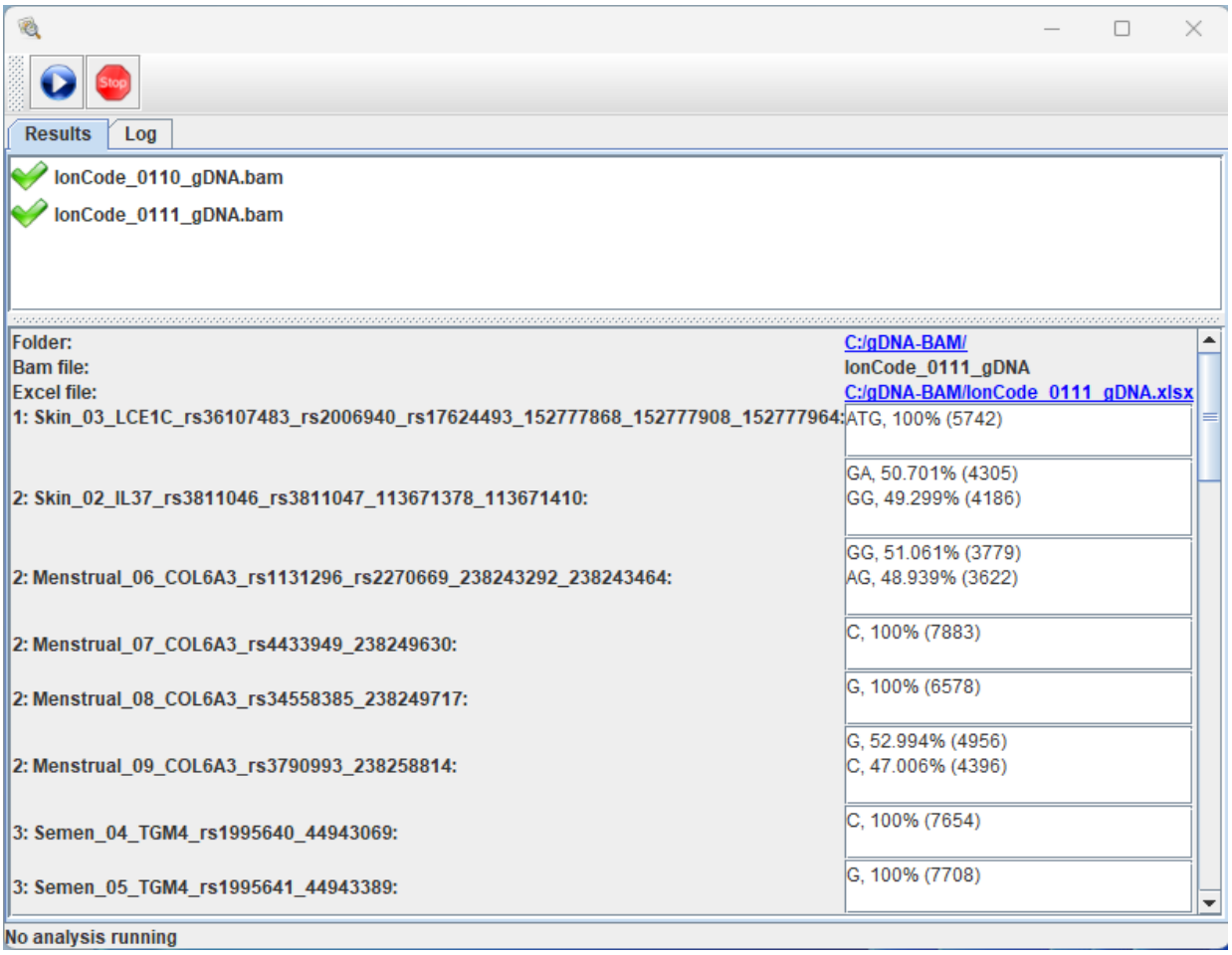

Figure 7. The Progress Window when running the mh.jar application for BFID-gDNA samples is shown when analysis has been completed.

9. Multiple results files are produced for each BAM file. The files are saved in the same folder as the BAM file. An example output is shown in Figure 8.

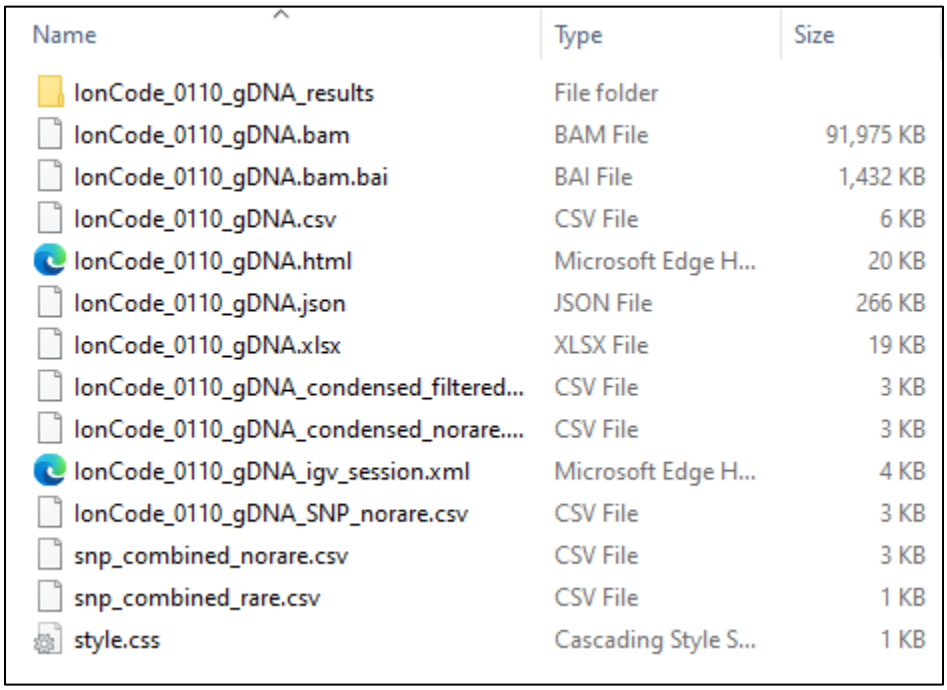

Figure 8. The result files produced by mh.jar during the analysis of a BFID-gDNA BAM file. These are automatically saved to the same folder as the input BAM/BAI files.

The .xlsx (Microsoft Excel) file is a spreadsheet containing the results for the SNPs in each amplicon. The amplicons are named after the tissue their target gene represents.

The gDNA result serves as a genotype for the SNPs sample in question to compare to the result produced by the mRNA panel. The mRNA result indicates the body fluid from which the sample originated.

An example result is shown in Figure 9.

Note: *Refer to The Ion AmpliSeq™ MH-74 Plex Microhaplotype Research Panel Technical Note (Thermo Fisher Scientific, 2021) for a detailed technical explanation of the file contents [5].*

Bam file:	C:\gDNA-BAM\IonCode_0110_gDNA.bam				
Date:	Wed Jul 31 14:43:49 BST 2024				
Content:	This sheet contains no rare alleles				
Skin_03_LCE1C_rs36107483_rs2006940_rs17624493_152777868_152777908_152777964	**chr1**	**3 locations**			
Location	**152777868**	**152777908**	**152777964**	**Count**	**Percent**
Reference	**C**	**T**	**A**	**2995**	
	A	T	G	1669	55.726
	G	C	G	1326	44.274
Skin_02_IL37_rs3811046_rs3811047_113671378_113671410	**chr2**	**2 locations**			
Location	**113671378**	**113671410**	**Count**	**Percent**	
Reference	**G**	**G**	**3696**		
	G	A	3696	100	
Menstrual_06_COL6A3_rs1131296_rs2270669_238243292_238243464	**chr2**	**2 locations**			
Location	**238243292**	**238243464**	**Count**	**Percent**	
Reference	**T**	**G**	**3290**		
	A	G	1791	54.438	
	G	G	1499	45.562	
Menstrual_07_COL6A3_rs4433949_238249630	**chr2**	**1 locations**			
Location	**238249630**	**Count**	**Percent**		
Reference	**C**	**3888**			
	C	3888	100		

Figure 9. An example result of analysis of a BFID-gDNA BAM file.

Analysis of the BFID-mRNA Panel

Analysis of BAM files from the BFID-mRNA Panel is similar to the gDNA panel. The mh.jar file is also used.

To analyze the BFID-mRNA panel BAM files:

1. Launch the mh.jar application by double-clicking its icon.
2. Select the folder with mRNA BAM files to analyze in the window that appears.
    a. If more than one BAM is found in the selected folder, a pop-up window will appear, allowing the user to select which files to analyze. Highlight the desired files and click 'Yes'.
3. Select the BFID_mRNA_Genes_v1.fa file as the reference genome in the next window.
4. Select the BFID_mRNA_target_v1.bed as the target file in the next window.
5. Select the BFID_mRNA_hotspots_v1.bed as the target file in the next window.

Note: *If the target and hotspot files are saved in the same folder as the BAM files, these will be selected automatically.*

6. The 'Program Parameters' window appears. Choose 'Target Application' as 'BFID'.
7. Leave the other parameters at the default.
8. Click 'Ok' to start the analysis (Figure 10).

Note: *Be sure to select 'BFID' as the Target Application for the mRNA analysis. This step differs from the gDNA analysis in the previous section, where that field is left blank.*

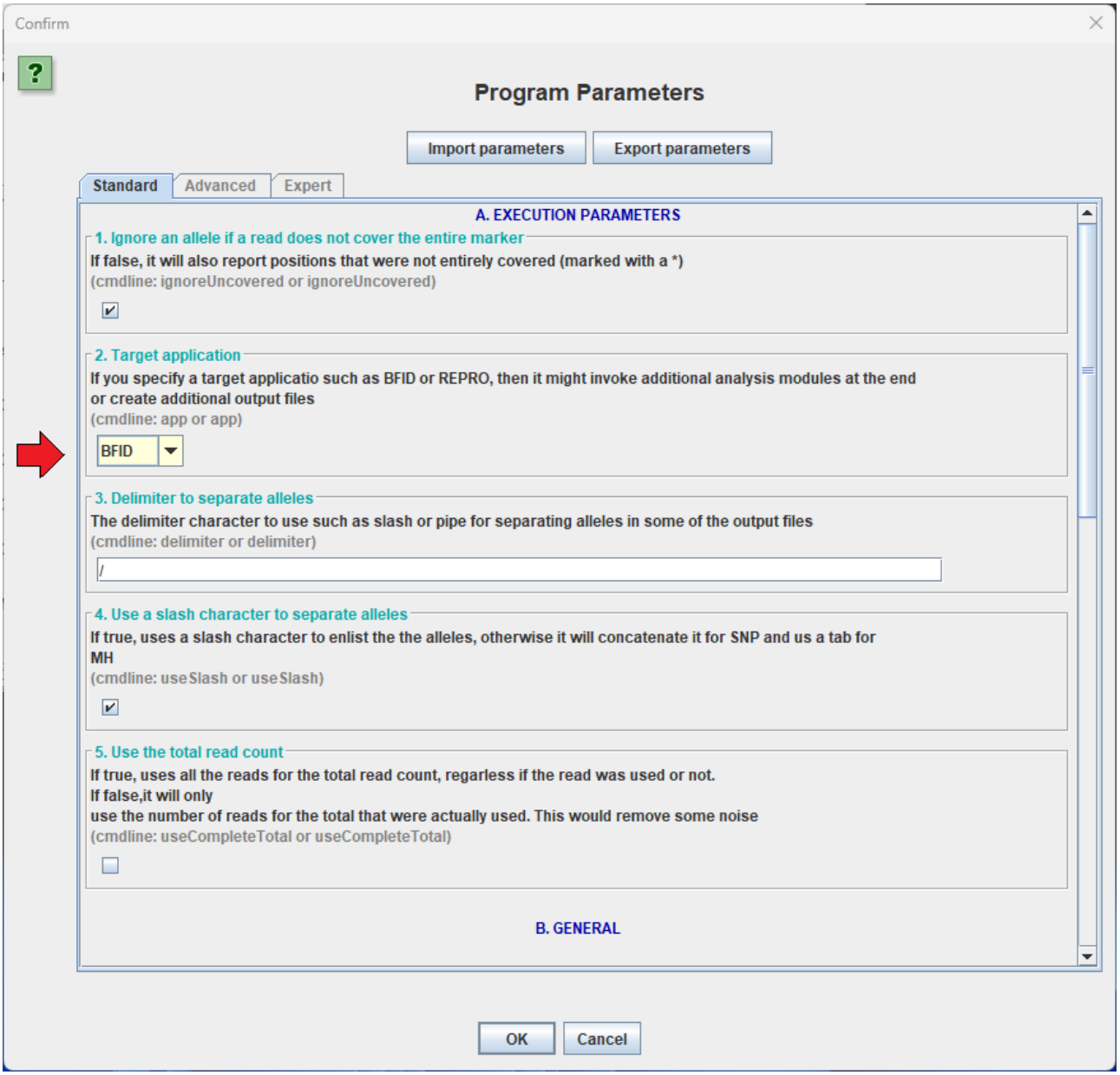

Figure 10. 'Program Parameters' screen when running the mh.jar application for BFID-mRNA samples. Select 'BFID' under 'Target Application'.

9. The application will then process the BAM files. It can take multiple minutes to process each sample, depending on the file size and the computer's processing power. A pop-up will approximate the completion time if multiple files are being processed. Click 'Ok'. A progress window is displayed during the analysis.

Note: *The estimated time to complete the analysis will often be overestimated– the analysis can be completed much faster than the estimate.*

10. When the analysis is finished, the result for the last sample will display along with the comment "No analysis running" in the bottom left corner. The samples processed are listed at the top with a green check next to each. Close this window.

11. Multiple results files are produced for each BAM file. The files are saved in the same folder as the BAM file. An example output is shown in Figure 11.

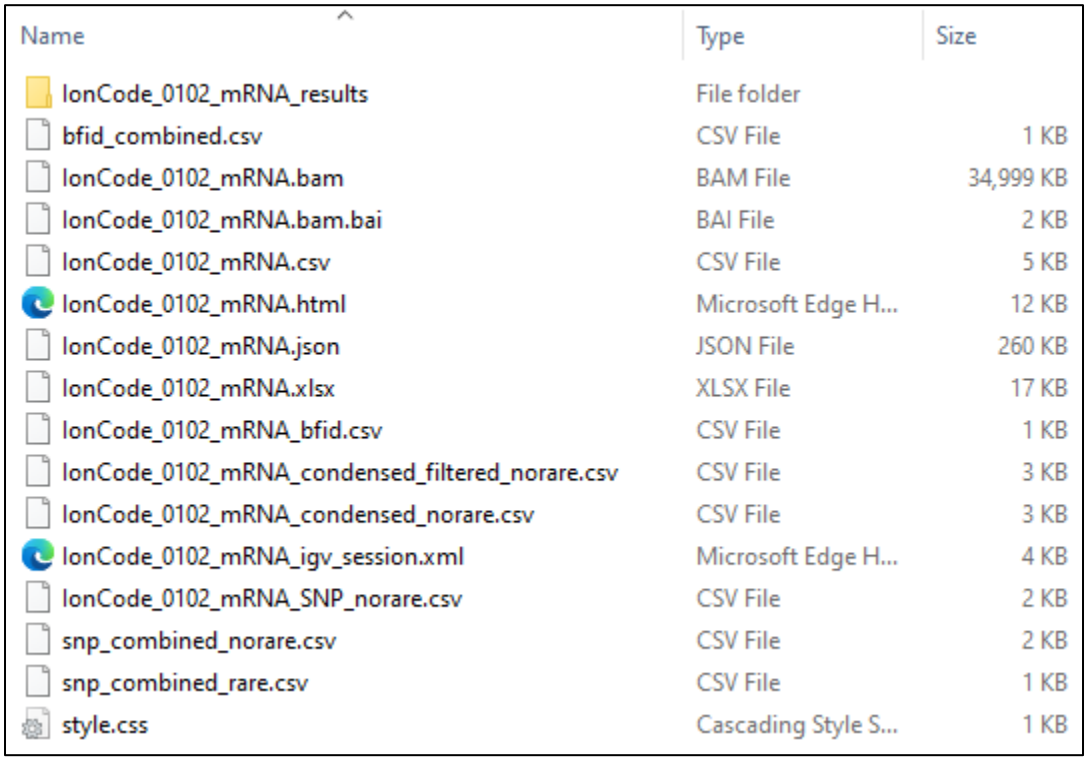

Figure 11. The result files produced by mh.jar during the analysis of a BFID-mRNA BAM file. These are automatically saved to the same folder as the input BAM/BAI files.

The .xlsx (Microsoft Excel) file is a spreadsheet containing the results for the SNPs in each amplicon. The amplicons are named after the tissue their target gene represents.

Note: *Refer to The Ion AmpliSeq™ MH-74 Plex Microhaplotype Research Panel Technical Note (Thermo Fisher Scientific, 2021) for a detailed technical explanation of the file contents [5].*

The mRNA result indicates the body fluid from which the sample originated. Only the tissues present in the sample should be represented by counts higher than background.

A gDNA unique portion of the PRM1 gene is used to estimate gDNA background amplification. Counts for PRM1_gDNA indicate the amount of background gDNA in the mRNA sample.

An example result from a mixed saliva and blood is shown in Figure 12.

gDNAPRM1_01_gDNA_prm1novel_30_160	PRM1_gDNA	2 locations			
Location	30	160	Count	Percent	
Reference			0		
Semen_01_KLK3_rs11573_rs1135766_89_95	NM_001648_2	3 locations			
Location	32	89	95	Count	Percent
Reference				0	
Saliva_01_HTN3_rs1849937_rs1136515_rs75067954_243_289_344	NM_000200_2	3 locations			
Location	243	289	344	Count	Percent
Reference	C	C	C	3457	
	C	T	C	1779	51.461
	C	C	C	1678	48.539
Saliva_02_MUC7_rs2306948_180	NM_152291_2	2 locations			
Location	169	180	Count	Percent	
Reference	G	C	714		
	G	C	714	100	
Blood_01_ANK1_rs504574_5169	NM_001142446_1RC	2 locations			
Location	5131	5169	Count	Percent	
Reference	T	C	8919		
	T	C	8919	100	
Blood_02_ANK1_rs7816734_3981	NM_001142446_1RC	2 locations			
Location	3981	4107	Count	Percent	
Reference	G	C	9588		
	A	C	5068	52.858	
	G	C	4520	47.142	
Skin_01_COL17A1_rs805701_4380	NM_000494_3RC	2 locations			
Location	4380	4425	Count	Percent	
Reference			0		

Figure 12. An example mRNA result for a mixed blood/saliva sample. The 'gDNAPRM1' marker at the top checks for potential gDNA contamination. The count for this marker is zero, which indicates no gDNA contamination. The above background read counts for the 'Saliva' and 'Blood' markers indicate the source of the sample. Other markers such as 'Skin' (at the bottom) and 'Semen', 'Vaginal', and 'Menstrual' (not shown) are not above background, indicating that the sample is not from these sample types.

The mRNA result folder also contains .csv files that summarize the mRNA result.

The file named <sample_name>_bfid.csv contains the result for the sample in question.

The file named bfid_combined.csv contains the combined result for all samples in the folder. An example is shown in Figure 13.

BFID Summary per tissue															
Sample	Total count	Gdnaprm1 count	Semen count	Vaginal count	Menstrual count	Skin count	Saliva count	Blood count	Gdnaprm1 %	Semen %	Vaginal %	Menstrual %	Skin %	Saliva %	Blood %
VS29	92320	0	10	92302	0	0	8	0	0%	0.01%	99.98%	0%	0%	0.01%	0%
B7489	11280	0	0	0	14	0	0	11266	0%	0%	0%	0.12%	0%	0%	99.88%
SE2	343477	0	343449	6	0	0	8	14	0%	99.99%	0%	0%	0%	0%	0%
SA1	65891	0	0	143	0	0	65732	16	0%	0%	0.22%	0%	0%	99.76%	0.02%
VS29-SE463	176626	0	13690	161212	0	1613	25	86	0%	7.75%	91.27%	0%	0.91%	0.01%	0.05%
MB1-SE4vs	212526	0	168138	16222	27785	53	0	328	0%	79.11%	7.63%	13.07%	0.02%	0%	0.15%
VS8-SA2	37144	0	0	18251	41	150	18702	0	0%	0%	49.14%	0.11%	0.40%	50.35%	0%
B7479-SA75	61131	0	11	0	2331	0	5623	53166	0%	0.02%	0%	3.81%	0%	9.20%	86.97%

Figure 13. An example 'bfid_combined.csv' file containing results from single-source samples (shown in blue) and mixed source samples (shown in red) (Note: coloring is for illustrative purposes and does not appear in the original file).

mh.jar Parameter Settings

The following are the default mh.jar parameter settings used for the Body Fluid panels.

Table 3. Default recommended mh.jar parameters for the BFID gDNA and mRNA panels. Parameters are the same between the two panels except for 'Target Application'.

Parameter Name	gDNA Panel	mRNA Panel
Ignore an allele if a read does not cover the entire marker	true	true
Target application	<Blank> (MH)	BFID
Min allele frequency (heterozygous)	10	10
Min allele frequency (homozygous)	90	90
Min # targets with noise	1	1
Score threshold for flagging mixtures	30	30
Delimiter to separate alleles	/	/
Use a slash character to separate alleles	true	true
Ignore the rare alleles (that are not in the hotspot file)	true	true
Use the total read count	false	false
Debug position to show more info	0	0
Test marker name		
Min total read coverage per position	5	5
Min mutation percent to include	2.0	2.0
Min # of allele count to include in report	5	5
Min deletion percent to include	40	40
Min insertion percent to include	30	30
Min total read coverage per position to call rare allele	50	50
Min mutation percent to create new allele	5	5
Min deletion percent to include for new allele	60	60
Min insertion percent to include for new allele	50	50
Create tagged BAM	false	false
Find marker by position	false	false
Reuse existing result	false	false
Minimum nr of threads to use	1	1
Maximum nr of threads to use	1	1
If we are in a server environment (TS)	false	false
Show more debug info	false	false
Recursively process folders	true	true
Run name to use		
The test region start position	0	0
The test region end position	0	0

Results Analysis – Running the Script

Further analysis of the result files gained from processing each BAM file can be performed to associate the mRNA cSNP results to the gDNA SNPs results. Familiarity with UNIX and Perl is required.

Perl must be installed to perform the analysis [6]. The following procedure can be performed on a Windows computer with 'Cygwin', a free and open-source Unix-like environment [7].

To perform the analysis:

1. If necessary, copy the result files to the working directory for Cygwin.
2. In the output folders, after processing the BAM files with mh.jar, there is a file named '<sample_name>_condensed_norare.csv' for every sample. Use the following command to concatenate the '<sample_name>_condensed_norare. csv' files for the mRNA and gDNA results into two combined files, one for all of the mRNA results and one for all of the gDNA results.

```
cat *_condensed_norare.csv > <output_file_name>.txt
```

For example:

```
cat *_condensed_norare.csv > my_gDNA_batch.txt
```

3. Copy the Perl script called 'BFID_Analysis_v1.pl' from the IonAmpliseq_BFID_Analysis_Files_v1.0.zip file package to the same folder as the files created in Step 2.
4. Execute the 'BFID_Analysis_v1.pl' script:

```
perl BFID_Analysis_v1.pl -i <output_file_name>.txt -s X
```

Where '<output_file_name>.txt' is the file's name created in Step 3 and where 'X' equals either 'r' for mRNA file or 'd' for a gDNA file.

For example:

```
perl BFID_Analysis_v1.pl -i my_gDNA_batch.txt -s d
```

Or:

```
perl BFID_Analysis_v1.pl -i my_mRNA_batch.txt -s r
```

Note: *Adjust the path to the txt file in the above commands as needed if the script is in a different folder to the txt file.*

Note: *Run the script with no input variables to see the usage details: i.e, run: 'perl BFID_Analysis_v1.pl' to see a description of the '-i' and '-s' variables as described above.*

5. A successful run of the script will result in output text as in Figure 14.

```
$ perl BFID_Analysis_v1.pl -i my_gDNA_batch.txt -s d
Processing gDNA file my_gDNA_batch.txt
 Number of samples to process: 4
 Processing sample number: 1
  Printing sample 1
 Processing sample number: 2
  Printing sample 2
 Processing sample number: 3
  Printing sample 3
 Processing sample number: 4
  Printing sample 4
Processing gDNA input file my_gDNA_batch.txt complete
Creating summary file my_gDNA_batch_gDNA_summary.txt
done!
```

Figure 14. Command prompt output of a successful run with the BFID_Analysis_v1 perl script.

6. The script will produce two .txt files, a result file and a summary file, which can be opened in a text editor or spreadsheet program. The output file formats are the same for both the mRNA and gDNA processing.

Results Analysis – 'Result' Files

The output files produced in the previous section are described in this section.

The structure of the files labeled 'Result' (for example, 'gDNA_batch_gDNA_result.txt' or 'mRNA_batch_mRNA_result.txt') are as follows:

Line 1 in the file indicates which sample type was used to process the batch file, mRNA or gDNA. Example:

```
## Data processed as mRNA samples
```

Lines 3 - 7 display the processing date and the data and support file sources used by the mh.jar application for processing. Example:

```
Date, Mon Jun 10 13:31:09 PDT 2024
Bam file,
/Users/Analyst/Documents/mRNA_Data/bam_files/IonCode_0101_mRNA_Blood.bam
Hotspot file,
/Users/Analyst/Documents/mRNA_Data/bam_files/BFID_mRNA_hotspots_v1.bed
Target file,
/Users/Analyst/Documents/mRNA_Data/bam_files/BFID_mRNA_target_v1.bed
Fasta file, /usr/local/databases/BFID_mRNA_Genes_v1.fa
```

Lines 9 - 16 show a summary for the coverage of each body fluid and its percentage. Example:

Target	Percentage	Coverage
Blood	99.42	22037
Menstrual	0.58	128
Saliva	0	0
Semen	0	0

```
Skin                                    0           0
Vaginal                                 0           0
Total Coverage                                  22165
```

Line 17 (mRNA Only) – displays the percentage and coverage of the PRM1 gene amplicon which includes the intron sequence which indicates the presence of gDNA in the mRNA sample (column headers are same as for Lines 9 – 16). Example:

```
Background gDNA (PRM1gDNA)    0.00        0
```

Lines 19 - 55 (mRNA) 18 - 53 (gDNA) – show the detailed results for each target gene and their associated SNP. A summary of each column in the file is shown in Table 4 below, with a description of the contents and an example value.

Table 4. A description of Lines 19 – 55 (mRNA) / Lines 18 -53 (gDNA) of 'Result' analysis files produced with the BFID_Analysis_v1 perl script.

Column No.	Column Name	Description of Value	Example Value
1	MH Target	Target tissue / order number / gene / SNP(s) / SNP position(s)	Blood_04_SPTB_rs1741488 _rs1741487_5539_5545
2	Hit	Sequence used in alignment	NM_001024858_2RC
3	%total/all	Percent of all tissue targets	13.4
4	%target/gene	Percent in same tissue	13.48
5	mh counts	Total number of aligned sequences	2971
6	target	Target tissue	Blood
7	gene	Target gene	SPTB
8	mh gt	Genotype for target microhaplotype	ACA\ATG
9	mh counts	Allele counts for target microhaplotype	1691\1280
10	rs_id1	rs_id for first cSNP target	rs1741488
11	position	Position in alignment for first cSNP target	5539
12	allele	Genotype for first cSNP target	C\T
13	coverage	Allele coverage for first cSNP target	1691\1280
14	rs_id2	rs_id for second cSNP target	rs1741487
15	position	Position in alignment for second cSNP target	5545
16	allele	Genotype for second cSNP target	A\G
17	coverage	Allele coverage for second cSNP target	1691\1280

As in the example in Table 4, the 'mh gt' (the genotype of the target microhaplotype) may include SNP(s) not present in the reported cSNP(s). Additional SNPs or gene target bases are included in the microhaplotype to ensure gene coverage is mRNA specific by forcing the alignment to include exon sequence with no intronic sequence using the mh.jar parameter 'Ignore an allele if a read does not cover the entire marker' (See Table 3).

Results Analysis – 'Summary' Files

The structure of the files labeled 'Summary' (for example, 'mRNA_batch_mRNA_summary.txt' or 'gDNA_batch_gDNA_summary.txt') are as follows:

Line 1 in the file indicates which sample type was used to process the batch file, mRNA or gDNA. Example:

```
## Data processed as mRNA samples
```

Lines 3 - 7 display the microhaplotype and cSNP genotype individually for each reported polymorphism. A summary of each column in the file is shown in Table 5 below, with a description of the contents and an example value.

Table 5. A description of Lines 3 to 7 of 'Summary' analysis files produced with the BFID_Analysis_v1 perl script.

Column No.	Column Name	Description of Value	Example Value
1	Sample ID	Name of BAM file	IonCode_0101_mRNA_Blood
2	Tissue	Target gene's association	Blood
3	Gene	Gene designation	SPTB
4	rs_id(s)	NCBI assigned identifier for polymorphism	rs1741488
5	Genotype	Base(s) reported for rs_id(s)	C\T
6	Coverage	Coverage reported for genotypes in their respective order	1691\1280

For microhaplotype targets where a non-cSNP is included, the data will display the genotype for the microhaplotype but only reference the cSNP components. In the example for the SPTB gene aligning to NM_001024858_2RC, in the example in Table 4, the summary data file will have the following structure:

```
IonCode_0101_mRNA_Blood  Blood  SPTB  rs1741488_rs1741487  ACA\ATG  1691\1280
IonCode_0101_mRNA_Blood  Blood  SPTB  rs1741488            C\T      1691\1280
IonCode_0101_mRNA_Blood  Blood  SPTB  rs1741487            A\G      1691\1280
```

The first line shows the two cSNPs in the microhaplotype, while the genotype ('ACA\ATG') has three values in each allele. This data indicates that a non-cSNP was used to define the microhaplotype. Lines two and three show the genotype and coverage for the targeted cSNPs. This format allows for quick matching between mRNA and gDNA runs when looking for matching genotypes.

Additional Information

Panels can be found on the Ion AmpliSeq Designer website https://ampliseq.com/designer/home. Select "Fixed Panels – Community Panels", and check "Human Identification Research" on the left side of the page.

The gene targets: LEFTY2 (Menstrual), STATH (Saliva), and CYP2B7P1 (Vaginal), do not contain any cSNP targets. They only serve as tissue specificity markers.

It is recommended that laboratories establish their own background and noise thresholds along with minimum biomarker read (MBR) count thresholds. References 1 to 3 below are recommended for more details on the implementation of the panels, workflow, and data analysis.

References

[1] E. Hanson, S. Ingold, G. Dørum, C. Haas, R. Lagace, J. Ballantyne. Assigning forensic body fluids to DNA donors in mixed samples by targeted RNA/DNA deep sequencing of coding region SNPs using ion torrent technology, Forensic Sci Int Genet Supp Ser: 7(7):23–24 (2019).

[2] S. Ingold, G. Dørum, E. Hanson, J. Ballantyne, C. Haas. Assigning forensic body fluids to donors in mixed body fluids by targeted RNA/DNA deep sequencing of coding region SNPs, International Journal of Legal Medicine: 134, 473–485 (2020).

[3] E. Hanson, G. Dørum, M. Zamborlin, S. Wang, M. Gysi, S. Ingold, R. Lagace, C. Roth, C. Haas, J. Ballantyne. Targeted S5 RNA sequencing assay for the identification and direct association of common body fluids with DNA donors in mixtures, International Journal of Legal Medicine: 137, 13–32 (2023).

[4] Thermo Fisher Scientific. Torrent Suite Software v5.18 User Guide. Available here.

[5] Thermo Fisher Scientific. The Ion AmpliSeq™ MH-74 Plex Microhaplotype Research Panel Technical Note. Available here

[6] www.perl.org

[7] www.cygwin.com

Revision History

Revision	Date	Description
A00	August 15, 2024	Initial publication.