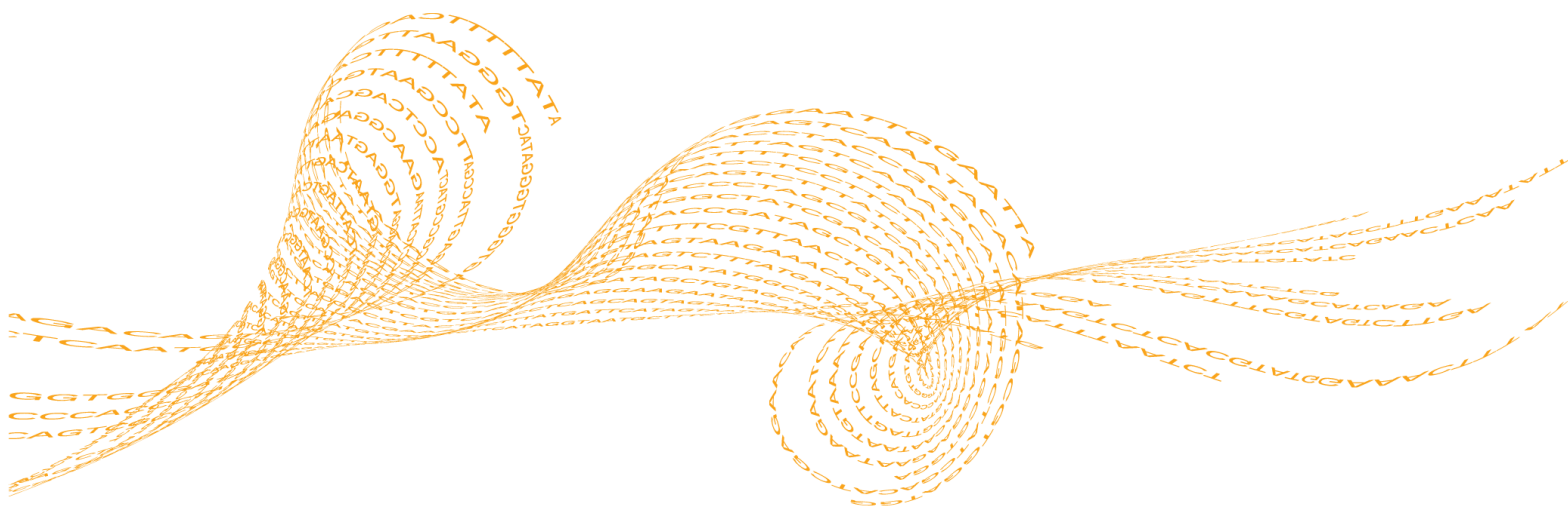


mtDNA Variant Processor v1.0

BaseSpace App Guide

For Research Use Only. Not for use in diagnostic procedures.

Introduction	3
Workflow Diagram	4
Workflow	5
Log In to BaseSpace	6
Set Analysis Parameters	7
Analysis Methods	8
Analysis Output	9
Technical Assistance	



ILLUMINA PROPRIETARY

Doc # 1000000007931 v00
February 2016

illumina®

This document and its contents are proprietary to Illumina, Inc. and its affiliates ("Illumina"), and are intended solely for the contractual use of its customer in connection with the use of the product(s) described herein and for no other purpose. This document and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina. Illumina does not convey any license under its patent, trademark, copyright, or common-law rights nor similar rights of any third parties by this document.

The instructions in this document must be strictly and explicitly followed by qualified and properly trained personnel in order to ensure the proper and safe use of the product(s) described herein. All of the contents of this document must be fully read and understood prior to using such product(s).

FAILURE TO COMPLETELY READ AND EXPLICITLY FOLLOW ALL OF THE INSTRUCTIONS CONTAINED HEREIN MAY RESULT IN DAMAGE TO THE PRODUCT(S), INJURY TO PERSONS, INCLUDING TO USERS OR OTHERS, AND DAMAGE TO OTHER PROPERTY.

ILLUMINA DOES NOT ASSUME ANY LIABILITY ARISING OUT OF THE IMPROPER USE OF THE PRODUCT(S) DESCRIBED HEREIN (INCLUDING PARTS THEREOF OR SOFTWARE).

© 2016 Illumina, Inc. All rights reserved.

Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPRO, CytoChip, DesignStudio, Epicentre, ForenSeq, Genetic Energy, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, MiniSeq, MiSeq, MiSeqDx, MiSeq FGx, NeoPrep, NextBio, Nextera, NextSeq, Powered by Illumina, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

Introduction

The BaseSpace® mtDNA Variant Processor v1.0 App performs alignment and variant calling of samples against a reference mitochondrial DNA (mtDNA) genome. The app can analyze any part of the full circular genome using quality and coverage thresholds information. The app produces BAM and VCF output files. You can use the BAM files for alignment research with third party tools. You can view the analysis results of the VCF files in the mtDNA Variant Analyzer App.

Compatible Libraries

See the BaseSpace support page for a list of library types that are compatible with the mtDNA Variant Processor App.

Versions

The following components are used in the mtDNA Variant Processor App.

Software	Version
Isas (Analysis Software)	2.6.7.4
BWA-MEM (Aligner)	0.7.9

Workflow Requirements

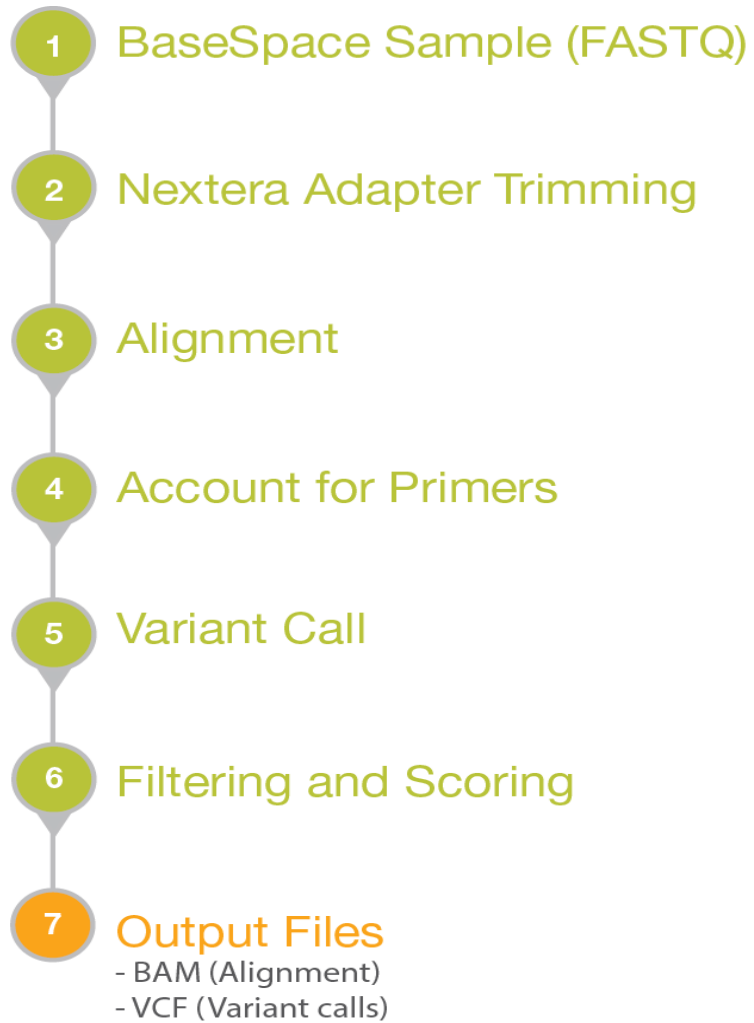
- ▶ Supports up to 96 samples per alignment.
- ▶ Requires alignment to the Revised Cambridge Reference Sequence (rCRS)
- ▶ Supports data from samples prepared with the Human mtDNA D-Loop Hypervariable Region Guide (document# 15034858) and Human mtDNA Genome Region Guide (document# 15037958).

Reference Genomes

The genome for alignment is the Revised Cambridge Reference Sequence (rCRS), a mtDNA genome used by the Forensic Genomics community.

Workflow Diagram

Figure 1 mtDNA Variant Processor Workflow



The mtDNA Variant Processor App supports both mtDNA D-Loop, custom regions, and whole genome samples. See *Analysis Methods* on page 8.

Nextera Adapter Trimming

- ▶ Nextera adapters are removed from the forward and reverse reads.
- ▶ Trimming is repeated until there are no more than 3 adapter bases on the end of the read.
- ▶ Reads are discarded when they are excessively trimmed or have short amplicons.

Alignment

- ▶ BWA-MEM performs the alignment.
- ▶ Circular alignment is handled across the origin by identifying the true start and end of reads.
- ▶ Realigns regions with indels to improve alignment and shifts indels to a 3' alignment.

Account for Primers

- ▶ Manifest identifies the primers and amplicons.
- ▶ Removes primer contributions from the reads for accurate variant calling.

Variant Calling

- ▶ Identifies nucleotides that are greater than the interpretation threshold for the position.
- ▶ Generates percentage of reads for a nucleotide that exceeds the analysis threshold (at).
- ▶ Generates a percentage of reads for a nucleotide that exceeds the interpretation threshold (it).
- ▶ Generates the interpretation threshold indicator for positions that have a nucleotide between the analysis threshold and the interpretation threshold.
- ▶ Reports base call codes. See *Base Call Codes* on page 11.

Filtering and Scoring

- ▶ Removes reads that do not meet the minimum quality score for a base call.
- ▶ Calculates a score from the minimum quality score for base call, alignment score, and presence of low abundance reads at the position.

Output Files

- ▶ VCF files and BAM files.
- ▶ Use the mtDNA Variant Analyzer App to view the analysis results.

Log In to BaseSpace

- 1 Navigate to BaseSpace at basespace.illumina.com, enter your user name and password, and click **Login**.
- 2 Click the **Apps** tab.
- 3 In Categories, click **Targeted Sequencing**, and then click **mtDNA Variant Processor**.
- 4 From the drop-down list, select **version 1.0.0**, and then click **Launch** to open the app. The mtDNA Variant Processor parameters window displays.
- 5 Set the parameters for analysis.
For additional training on BaseSpace, see the BaseSpace support page at support.illumina.com/sequencing/sequencing_software/basespace/training.html

Set Analysis Parameters

- 1 In the App Session Name field, enter the analysis name.
By default, the analysis name includes the app name, followed by the date and time that the analysis session starts.
- 2 From the Project field, select the project that stores the app results.
- 3 From the Samples field, browse to the sample you want to analyze, and select the checkbox.
- 4 From the Minimum Base Call Quality Score for a Call (Q) field, enter the quality score.
The default value is 30. At each coordinate, reads at or above the Q value can contribute to the variant call.
- 5 From the Analysis Threshold % field, enter the percentage.
The default value is 10. If a nucleotide call is at less than the value of the Analysis Threshold (%), then the nucleotide call is not visible or reported as part of the call.
- 6 From the Interpretation Threshold % field, enter the percentage.
The default is 25. A nucleotide with a total number of reads greater than the interpretation threshold percentage contributes to the call for the coordinate. When more than one base exceeds this percentage for a coordinate, the app reports the base call code for the bases. See *Base Call Codes* on page 11.
- 7 From the Minimum Read Count field, enter the number of reads required at the coordinate to produce a base call.
The default value 10.
- 8 From the Genome field, select the genome to be analyzed.
- 9 From the PCR Primer Description field, select from the following options:
 - ▶ Whole Mito Genome Amplicon Manifest
 - ▶ DLoop Amplicon Manifest
 - ▶ Custom Manifest.
- 10 [Optional] If you select **Custom Manifest**, complete the following steps:
 - a Enter the start coordinate (5', 0-based) for the forward PCR primer.
 - b Enter the start coordinate (5', 0-based) for the reverse PCR primer.
 - c Enter the base length of the forward PCR primer.
 - d Enter the base length of the reverse PCR primer.

Click the plus (+) sign to add additional rows.
- 11 Click **Continue**.
The mtDNA Variant Processor App begins analysis.
When analysis is complete, the app updates the status of the session and sends an email to notify you.

Analysis Methods

The mtDNA Variant Processor App uses the following methods to analyze the sequencing data. See *Workflow* on page 5.

- ▶ Illumina Secondary Analysis Software (Isas)
- ▶ BAM-MEM¹

¹ Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997v1[q-bio.GN].

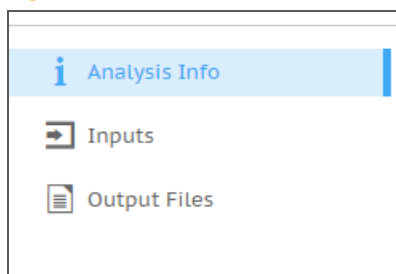
BWA-MEM

BWA-MEM is the most recent version of the Burrows-Wheeler Alignment algorithm. Optimized for longer read lengths of ≥ 70 bp, BWA-MEM has a significant positive impact on detection of variants, especially insertions and deletions.

Analysis Output

To view the results, click the **Projects** tab, then the project name, and then the analysis.

Figure 2 mtDNA Variant Processor Output Navigation Bar



Use the left navigation bar to access the following analysis output:

- ▶ **Analysis Info**—Information about the analysis session, including log files.
- ▶ **Inputs**—Overview of input settings.
- ▶ **Output Files**—Output files for the samples.

Analysis Info

The Analysis Info page displays the analysis settings and execution details.

Row Heading	Definition
Name	Name of the analysis session.
Application	App that generated this analysis.
Date Started	Date and time the analysis session started.
Date Completed	Date and time the analysis session completed.
Duration	Duration of the analysis.
Session Type	Multi-Node.
Status	Status of the analysis session. The status shows either Running or Complete and the number of nodes used.

Log Files

File Name	Description
SampleSheetUsed.csv	A copy of the sample sheet.
[Manifest name].txt	Manifest files used in the analysis.
WorkflowError.txt	Contains error messages created when running the workflow.
WorkflowLog.txt	Contains details about workflow steps, command line calls with parameters, timing, and progress.

Output Files

The Output Files page provides access to the output files for each sample analysis.

- ▶ BAM Files
- ▶ VCF Files

BAM File Format

A BAM file (*.bam) is the compressed binary version of a SAM file that is used to represent aligned sequences up to 128 Mb. SAM and BAM formats are described in detail at <https://samtools.github.io/hts-specs/SAMv1.pdf>.

BAM files use the file naming format of SampleName_S#.bam, where # is the sample number determined by the order that samples are listed for the run. In multi-node mode, the S# is set to S1, regardless the order of the sample.

BAM files contain a header section and an alignment section:

- ▶ **Header**—Contains information about the entire file, such as sample name, sample length, and alignment method. Alignments in the alignments section are associated with specific information in the header section.
- ▶ **Alignments**—Contains read name, read sequence, read quality, alignment information, and custom tags. The read name includes the chromosome, start coordinate, alignment quality, and the match descriptor string.

The alignments section includes the following information for each or read pair:

- ▶ **RG:** Read group, which indicates the number of reads for a specific sample.
- ▶ **BC:** Barcode tag, which indicates the demultiplexed sample ID associated with the read.
- ▶ **SM:** Single-end alignment quality.
- ▶ **AS:** Paired-end alignment quality.
- ▶ **NM:** Edit distance tag, which records the Levenshtein distance between the read and the reference.
- ▶ **XN:** Amplicon name tag, which records the amplicon tile ID associated with the read.

BAM index files (*.bam.bai) provide an index of the corresponding BAM file.

VCF File Format

Variant Call Format (VCF) is a widely used file format developed by the genomics scientific community that contains information about variants found at specific positions in a reference genome.

Table 1 VCF Headings Table

Heading	Description
CHROM	The chromosome of the locus.
POS	The positional coordinate of the locus.
REF	The reference base for the POS and can only be A, C, T, G, or N.

Heading	Description
ALT	The called base for the POS and can only be A, C, T, G, N, period (.), or asterisk (*). The asterisk (*) is used when there is a deletion at the locus of an upstream POS.
QUAL	The field is not used.
FILTER	PASS—For a called position NoData—For a position without coverage.
INFO	The field is not used.
FORMAT	The field contains the key name for the information on each reported attribute in the SAMPLE column. The keys are colon-separated and reference the #FORMAT descriptors in the header of the file.
SAMPLE	The field contains the values for each field from the FORMAT column.

Base Call Codes

The mtDNA Processor App supports variant calling for more than one nucleotide at a position. When more than one nucleotide exceeds the interpretation threshold, the app reports the IUPAC code as the consensus for the position.

Table 2 Variant Base Call Codes¹ Table

Code	Base
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	Any base
-	Deletion
*	No call

When data for the indel and a base call at a position exceed the interpretation threshold, the app reports the ambiguous call code for an indel as the consensus for the position.

Table 3 Ambiguous Variant Base Call Codes Table

Code	Base
a	Adenine
c	Cytosine

Code	Base
g	Guanine
g	Thymine
r	a or g
y	c or t
s	g or c
w	a or t
k	g or t
m	a or c
b	c or g or t
d	a or g or t
h	a or c or t
v	a or c or g
n	Any base

¹Codes are from the International Union of Pure and Applied Chemistry (IUPAC).

For example, Figure 3 shows the presence of an ambiguous deletion at position 501. Because there are an equal number of reads that support a deletion call and a reference call of G, the app reports an ambiguous call of g as the consensus.

Figure 3 Example of Ambiguous Variant Base Calls

Coordinate	498	499	500	501	502	503
Call	A	C	T	g	A	C
Reads	A	C	T		A	C
	A	C	T		A	C
	A	C	T		A	C
	A	C	T		A	C
	A	C	T		A	C
	A	C	T		A	C
	A	C	T		A	C
	A	C	T	G	A	C
	A	C	T	G	A	C
	A	C	T	G	A	C
	A	C	T	G	A	C
	A	C	T	G	A	C
	A	C	T	G	A	C

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 4 Illumina General Contact Information

Website	www.illumina.com
Email	techsupport@illumina.com

Table 5 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Japan	0800.111.5011
Australia	1.800.775.688	Netherlands	0800.0223859
Austria	0800.296575	New Zealand	0800.451.650
Belgium	0800.81102	Norway	800.16836
China	400.635.9898	Singapore	1.800.579.2745
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	Taiwan	00806651752
Hong Kong	800960230	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000
Italy	800.874909		

Safety data sheets (SDSs)—Available on the Illumina website at support.illumina.com/sds.html.

Product documentation—Available for download in PDF from the Illumina website. Go to support.illumina.com, select a product, then select **Documentation & Literature**.



Illumina

5200 Illumina Way

San Diego, California 92122 U.S.A.

+1.800.809.ILMN (4566)

+1.858.202.4566 (outside North America)

techsupport@illumina.com

www.illumina.com