

GeneMarker[®] HTS

Software

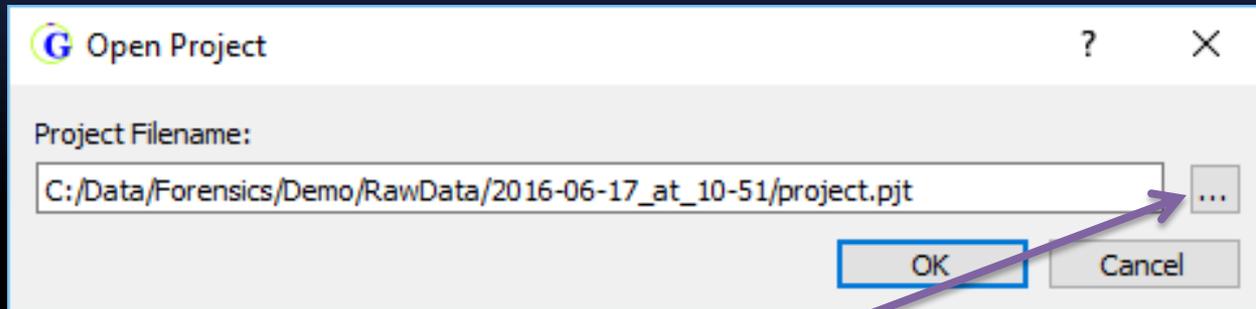
Quick Start Guide



Upon launching the software, the user will have the option to start a *New* project or *Open* a previously saved project.



Opening a Project



Using the ellipsis button, the user can select a previous project file (project.pjt). After clicking *OK*, the project will be available to open in a *Sample Viewer* window.

Setting up a New Project

Creating a new project

When *New* is selected, the *New Project* window will be displayed. In this window, the user can specify their data and choose alignment settings for their project.

Selecting the ? button will allow users with a touch screen device to touch a setting and display a tool tip.

Project Folder:

Location to Save the Project File

Reference Filename:

Fullpath of the Reference File

BuiltIn Motif Template
 Customized Motif

Fullpath of the Custom Motif File

Name	File 1	File 2
------	--------	--------

Alignment Options

Remove PCR Duplicates
 Motifs

Identity
90

Soft Clipping at 3bp Q ≤
25

Sequencer:
 Illumina
 Ion Torrent
 Other

Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel



Creating a new project

New Project

Project Folder:
Location to Save the Project File

Reference Filename:

Samples can be loaded using the *Add* button at the bottom of the *New Project* window. If paired reads are selected, they will be displayed together.

Compressed (**fastq.gz**) or uncompressed (**.fastq**) sequence files are the accepted input. Sample files can be removed individually or all at once using the *Remove* and *Remove All* buttons.

Identity
90
Soft Clipping at 3bp Q ≤
25
Sequencer:
 Illumina
 Ion Torrent
 Other

Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel



Creating a new project

The software will automatically group paired reads into the same sample, but this can be adjusted by right-clicking on rows in the table.

New Project

Project Folder:
Location to Save the Project File

Reference Filename:
Fullpath of the Reference File

BuiltIn Motif Template
 Customized Motif

Fullpath of the Custom Motif File

Use Default Reference

Create Motif Edit Motif

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz
SID004_R1.fastq	SID004_R1.fastq.gz	SID004_R2.fastq.gz
SID005_R1.fastq	SID005_R1.fastq.gz	SID005_R2.fastq.gz
SID006_R1.fastq	SID006_R1.fastq.gz	SID006_R2.fastq.gz

Alignment Options

Remove PCR Duplicates
 Motifs

Identity
90

Soft Clipping at 3bp Q ≤
25

Sequencer:
 Illumina
 Ion Torrent
 Other

Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel



Creating a new project

New Project

Project Folder: ...

Reference Filename: ...

BuiltIn Motif Template
 Customized Motif

Fullpath of the Custom Motif File ...

Use Default Reference

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz
SID004_R1.fastq	SID004_R1.fastq.gz	SID004_R2.fastq.gz
SID005_R1.fastq	SID005_R1.fastq.gz	SID005_R2.fastq.gz
SID006_R1.fastq	SID006_R1.fastq.gz	SID006_R2.fastq.gz

Alignment Options

Remove PCR Duplicates
 Motifs

Identity: 90

Soft Clipping at 3bp Q ≤ 25

Sequencer:
 Illumina
 Ion Torrent
 Other

Sample names are automatically generated from filenames, but they can be edited by double-clicking the name in the input table.



Creating a new project

The *Filter Settings* and *Table Settings* buttons will allow the user to adjust settings to meet their SOP or select *Default* to return them to their default values.

Selecting *OK* will save the selected settings, but they may be adjusted after alignment. These settings control the criteria for calling variants and for displaying information.

Fullpath of the Custom Motif File

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz
SID004_R1.fastq	SID004_R1.fastq.gz	SID004_R2.fastq.gz
SID005_R1.fastq	SID005_R1.fastq.gz	SID005_R2.fastq.gz
SID006_R1.fastq	SID006_R1.fastq.gz	SID006_R2.fastq.gz

Alignment Options

- Remove PCR Duplicates
- Motifs

Identity: 90

Soft Clipping at 3bp Q ≤ 25

Sequencer:

- Illumina
- Ion Torrent
- Other

Buttons: Add, Remove, Remove All, Filter Settings, Table Settings, Load Template, None, OK, Cancel



Filter Settings

Filter Settings

Table Filter Settings

Entire Reference Input Region

Start 16024 End 527

Variant Percentage \geq 2.0%

Variant Allele Coverage \geq 10

Total Coverage \geq 200

Allele Score Difference \leq 10

Allele Balance Ratio

SNP \leq 2.5

Indel \leq 5.0

Load Default OK Cancel

The *Table Filter Settings* dialog allows for variant calling settings to be adjusted.

The settings can be returned to their original state by using the *Default* button. A project settings file can be imported using the *Load* button.



Filter Settings

Table Filter Settings

Entire Reference Input Region

Start End

Variant Percentage \geq

Variant Allele Coverage \geq

Total Coverage \geq

Allele Score Difference \leq

Allele Balance Ratio

 SNP \leq

 Indel \leq

Load Default **OK** Cancel

Reference area where variants may be called



Filter Settings

Table Filter Settings

Entire Reference Input Region

Start 16024 End 527

Variant Percentage \geq 2.0%

Variant Allele Coverage \geq 10

Total Coverage \geq 200

Allele Score Difference \leq 10

Allele Balance Ratio

SNP \leq 2.5

Indel \leq 5.0

Load Default OK Cancel

← Percent of reads at a position with the variant



Filter Settings

G Table Filter Settings ? X

Entire Reference Input Region

Start 16024 End 527

Variant Percentage \geq 2.0%

Variant Allele Coverage \geq 10

Total Coverage \geq 200

Allele Score Difference \leq 10

Allele Balance Ratio

SNP \leq 2.5

Indel \leq 5.0

Load Default **OK** Cancel

Number of reads with the variant and Total number of reads



Allele Score Difference

Table Filter Settings

Entire Reference Input Region

Start 16024 End 527

Variant Percentage \geq 2.0%

Variant Allele Coverage \geq 10

Total Coverage \geq 200

Allele Score Difference \leq 10

Allele Balance Ratio

SNP \leq 2.5

Indel \leq 5.0

Load Default OK Cancel

Each allele (A, C, G, T, deletion, and any insertions) has quality scores calculated as the median of the base quality scores from the original data. Deletion base quality scores don't exist, so they are based on the quality scores of surrounding bases.

Each minor allele has its score subtracted from the major allele score and the difference must be less than or equal to the set value.

Both directions are checked individually. If the score difference is higher and the minor allele has at least 5 reads the variant is assumed to be a false positive and is filtered out.



Allele Balance Ratio

Table Filter Settings

Entire Reference
 Input Region

Start: 16024 End: 527

Variant Percentage: \geq 2.0%

Variant Allele Coverage: \geq 10

Total Coverage: \geq 200

Allele Score Difference: \leq 10

Allele Balance Ratio

SNP: \leq 2.5

Indel: \leq 5.0

The balance ratio is the maximum value of:

- Allele forward % / Total forward %
- Total forward % / Allele forward %
- Allele reverse % / Total reverse %
- Total reverse % / Allele reverse %

This is calculated for each allele. Indels and SNPs have different cutoffs for filtering.

- 2 forward and 79 reverse reads with the allele
- 1758 forward and 6055 reverse reads overall
- Balance is the maximum of:
 - $0.02469/0.22501 = 0.9113$
 - $0.22501/0.02469 = \mathbf{9.1134}$
 - $0.97531/0.77499 = 1.2585$
 - $0.77499/0.97531 = 0.7946$

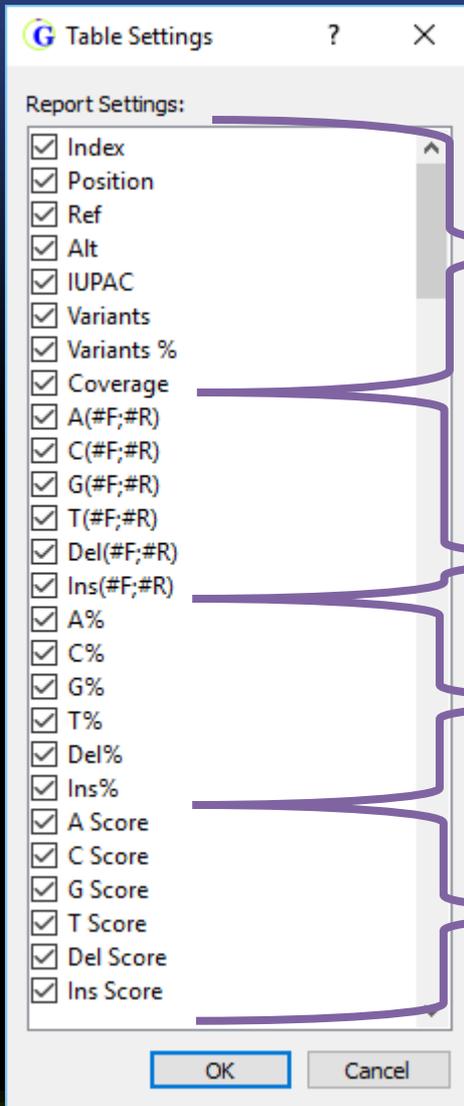
This allele would be filtered with the default settings because $9.1134 > 2.5$

Alleles with frequency $\geq 85\%$ are not filtered using this setting



Table Settings

Table Settings



- Index (number the variants)
- Position of Variant
- Reference Allele
- Alternative Allele
- IUPAC (ambiguity) formatted variant call
- Standard variant call
- Variant Frequency
- Total Coverage at the Position

Allele Counts (Forward and Reverse)

Allele Percentage

Median Basecall Quality (Forward and Reverse)



Variant Calls -IUPAC

When there are multiple alleles at a position, ambiguity codes are used when reporting in the “IUPAC” column.

Code	Alleles
M	A, C
R	A, G
W	A, T
S	C, G
Y	C, T
K	G, T
V	A, C, G
H	A, C, T
D	A, G, T
B	C, G, T
N	A, C, G, T

Mixtures of a deletion and other alleles use a lowercase letter.



Creating a new project

G New Project ? X

Project Folder:
Location to Save the Project File ...

Reference Filename:
Fullpath of the Reference File ...

BuiltIn Motif Template
 Customized Motif

Fullpath of the Custom Motif File ...

Use Default Reference
Create Motif Edit Motif

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq		gz
SID004_R1.fastq		gz
SID005_R1.fastq		gz
SID006_R1.fastq		gz

Alignment Options
 Remove PCR Duplicates
 Motifs
Identity
90
Soft Clipping at 3bp Q ≤
25
Sequencer:
 Illumina
 Ion Torrent
 Other

Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel

Selecting *Use Default Reference* will make the software use the revised Cambridge reference sequence (rCRS) for the alignment.



Creating a new project

New Project

Project Folder:
Location to Save the Project File

Reference Filename:
Fullpath of the Reference File

BuiltIn Motif Template
 Customized Motif

Use Default Reference

Create Motif Edit Motif

Fullpath of the Custom Motif File

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz

Alignment Options
 Remove PCR Duplicates
 Motifs
Identity: 90
Soft Clipping at 3bp Q ≤

Other

Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel

Alternatively, a *Reference file* may be loaded. Using the ellipsis icon to the right of the Reference field, navigate to the location of the desired Reference file. The software supports input of FASTA or GBK files for the reference sequence.



Creating a new project

Project Folder:
Location to Save the Project File

Reference Filename:
Fullpath of the Reference File

In the *Project Folder* field, a location can be selected for the data output by the program. A location can be set using the ellipsis button to the right of the field, or it can be typed manually. The folder will be created if it does not exist.

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz
SID004_R1.fastq	SID004_R1.fastq.gz	SID004_R2.fastq.gz
SID005_R1.fastq	SID005_R1.fastq.g	
SID006_R1.fastq	SID006_R1.fastq.g	

Alignment Options
 Remove PCR Duplicates
 Motifs
Identity
90
Soft Clipping at 3bp Q ≤
25

If there is a previous project saved in the selected folder, the program will ask the user if they would like to overwrite the existing project file. If *No* is selected, the program will revert to the *New Project* window and a new folder name can be entered.

Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel



Creating a new project

New Project

Project Folder:
Location to Save the Project File

Reference Filename:
Fullpath of the Reference File

BuiltIn Motif Template
 Customized Motif

Fullpath of the Custom Motif File

Name	Reference File	Custom Motif File
SID001_R1.fastq		
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz
SID004_R1.fastq		
SID005_R1.fastq		
SID006_R1.fastq		

Use Default Reference

Create Motif Edit Motif

Use Default Project Folder

The project folder was not set, would you like to use:
C:/Data/Forensics/Demo/RawData/2016-06-17_at_08-40/

Yes No

Alignment Options

Remove PCR Duplicates
 Motifs

Identity: 90

Soft Clipping at 3bp Q ≤: 25

Sequencer:
 Illumina
 Ion Torrent
 Other

Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel

If the *Project Folder* field is empty when the user selects OK, the program will suggest a name for a new folder based on the current date and time. The user will receive a pop-up message to confirm the folder name.



Setting up your project

The *Identity* and *Soft Clipping* settings will change depending on the *Sequencer* selected. If the user manually changes either setting, the program will change the selected *Sequencer* to *Other*.

Name	File 1	File 2
SID001_R1.fastq		
SID002_R1.fastq		
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz
SID004_R1.fastq		
SID005_R1.fastq		
SID006_R1.fastq	SID006_R1.fastq.gz	SID006_R2.fastq.gz

Identity: Reads that are less similar to the reference than this percentage are not aligned.

Soft Clipping: The 3' end of reads are trimmed when basecall quality is low.

Note:
Soft-clipping is also performed on the 3' ends of reads when mismatched bases are found near the end of the alignment.

Alignment Options

- Remove PCR Duplicates
- Motifs

Identity: 90

Soft Clipping at 3bp Q ≤ 25

Sequencer:

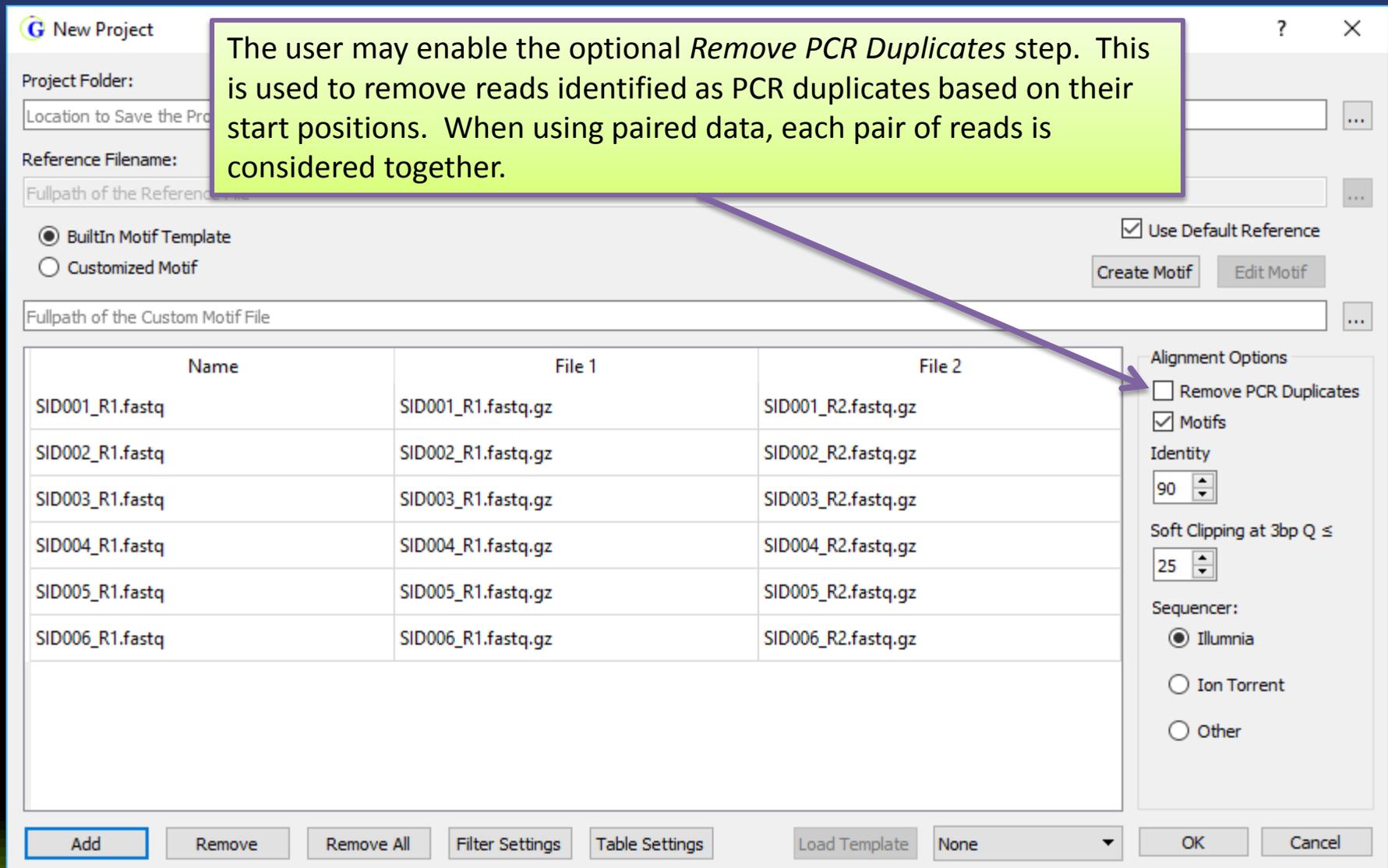
- Illumina
- Ion Torrent
- Other

Buttons: Add, OK, Cancel



Setting up your project

The user may enable the optional *Remove PCR Duplicates* step. This is used to remove reads identified as PCR duplicates based on their start positions. When using paired data, each pair of reads is considered together.



New Project

Project Folder:

Reference Filename:

BuiltIn Motif Template Customized Motif

Use Default Reference

Fullpath of the Custom Motif File

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq	SID002_R1.fastq.gz	SID002_R2.fastq.gz
SID003_R1.fastq	SID003_R1.fastq.gz	SID003_R2.fastq.gz
SID004_R1.fastq	SID004_R1.fastq.gz	SID004_R2.fastq.gz
SID005_R1.fastq	SID005_R1.fastq.gz	SID005_R2.fastq.gz
SID006_R1.fastq	SID006_R1.fastq.gz	SID006_R2.fastq.gz

Alignment Options

Remove PCR Duplicates

Motifs

Identity:

Soft Clipping at 3bp Q ≤

Sequencer:

Illumina Ion Torrent Other



Setting up your project

The user has the option to align their data with or without using a Motif file. Motifs help with the alignment of the data.

If the user chooses to align using the *Motif* option, they may use the built-in motif data by selecting *BuiltIn Motif Template*.

If the user would like to use a *Customized Motif* file, it may be imported using the ellipsis icon to the right of the *Custom Motif* field.

New Project

Project Folder:
Location to Save the Project File

Reference Filename:
Fullpath of the Reference File

BuiltIn Motif Template
 Customized Motif

Use Default Reference
Create Motif Edit Motif

Fullpath of the Custom Motif File

Name	File 1	File 2
SID001_R1.fa		
SID002_R1.fa		
SID003_R1.fa		
SID004_R1.fa		
SID005_R1.fa		
SID006_R1.fa		

Alignment Options

Remove PCR Duplicates
 Motifs

Identity: 90

Soft Clipping at 3bp Q ≤ 25

Sequencer:
 Illumina
 Ion Torrent
 Other

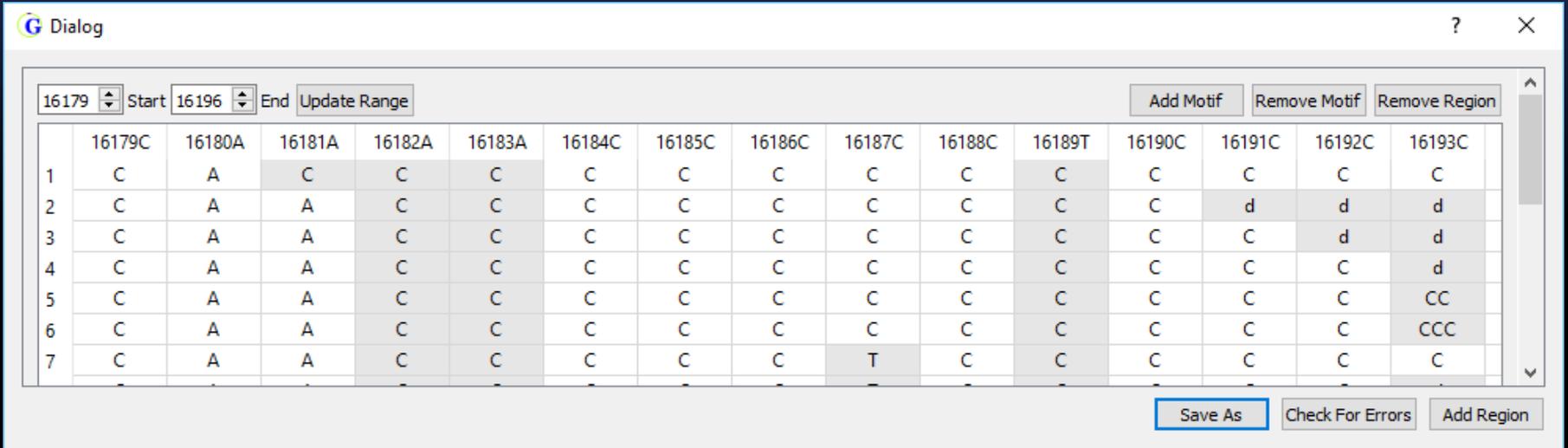
Add Remove Remove All Filter Settings Table Settings Load Template None OK Cancel



Motifs

Setting up your project

The *Edit Motif* icon opens the custom motif in the *Motif Editor* window.
The *Create Motif* icon will open the built-in motif file.



The screenshot shows the Motif Editor window with a grid of motifs. The grid has 7 rows and 16 columns. The columns are labeled with motif IDs: 16179C, 16180A, 16181A, 16182A, 16183A, 16184C, 16185C, 16186C, 16187C, 16188C, 16189T, 16190C, 16191C, 16192C, and 16193C. The rows are numbered 1 through 7. The motifs are represented by letters: C, A, T, d, and CC. The 'Save As' button is highlighted in blue.

	16179C	16180A	16181A	16182A	16183A	16184C	16185C	16186C	16187C	16188C	16189T	16190C	16191C	16192C	16193C
1	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C
2	C	A	A	C	C	C	C	C	C	C	C	C	d	d	d
3	C	A	A	C	C	C	C	C	C	C	C	C	C	d	d
4	C	A	A	C	C	C	C	C	C	C	C	C	C	C	d
5	C	A	A	C	C	C	C	C	C	C	C	C	C	C	CC
6	C	A	A	C	C	C	C	C	C	C	C	C	C	C	CCC
7	C	A	A	C	C	C	C	C	T	C	C	C	C	C	C

The *Motif Editor* window makes it easy to add new motifs or edit existing motifs. It is also possible to add new motif regions. The *Check for Errors* button will ensure that all motifs in the file are valid. After editing the user can save the motifs to a new file using the *Save As* button.



Setting up your project

New Project [?] [X]

Project Folder:
Location to Save the Project File [...]

Reference Filename:
Fullpath of the Reference File [...]

BuiltIn Motif Template Use Default Reference
 Customized Motif [Create Motif] [Edit Motif]

Fullpath of the Custom Motif File [...]

Name	File 1	File 2
SID001_R1.fastq	SID001_R1.fastq.gz	SID001_R2.fastq.gz
SID002_R1.fastq		
SID003_R1.fastq		
SID004_R1.fastq		
SID005_R1.fastq		
SID006_R1.fastq	SID006_R1.fastq.gz	SID006_R2.fastq.gz

Alignment Options
 Remove PCR Duplicates
 Motifs
Identity: 90
Soft Clipping at 3bp Q ≤ 25
Sequencer:
 Illumina
 Ion Torrent
 Other

[Add] [Remove] [Remove All] [Filter Settings] [Table Settings] [Load Template] [None] [OK] [Cancel]

If the user does not want to align using a motif file, the *Motif* option can be turned off. When the *Motif* option is turned off, all of the *Motif* options will be grayed out.



Using Motifs

Alignment traditionally is focused on minimizing the number of differences between the read and the reference. This isn't optimal for forensic analysis since there is an established convention for the positioning of many common variants.

	16180					16185					16190					16195		
Reference	A	A	A	A	C	C	C	C	C	T	C	C	C	C	A	T	G	
Optimal Alignment	A	A	-	-	C	C	C	C	C	T	C	C	C	CC	A	T	G	
Forensic Alignment	A	A	C	C	C	C	C	T	C	C	C	C	C	-	A	T	G	

```
#16179-16196
16181C, 16182C, 16183C, 16189C
16182C, 16183C, 16189C, 16191d, 16192d, 16193d
16182C, 16183C, 16189C, 16192d, 16193d
16182C, 16183C, 16189C, 16193d
16182C, 16183C, 16189C, 16193.1C
16182C, 16183C, 16189C, 16193.1C, 16193.2C
16182C, 16183C, 16187T, 16189C
16182C, 16183C, 16187T, 16189C, 16193d
16182C, 16183C, 16189C
```

Each motif is a list of variant calls. These variant calls are translated into an expected sequence.

When using “motif alignment”, the alignments of reads crossing motif regions (specified by the “#” line) with an expected sequence will be adjusted to match the list of variants.

Reads starting or ending in motif regions are soft-clipped.



Motif File Format

```
1 #16180-16198
2 16181C, 16182C, 16183C, 16189C
3 16182C, 16183C, 16189C
4 16182C, 16183C, 16189C, 16191d, 16192d, 16193d
5 16182C, 16183C, 16189C, 16192d, 16193d
6 16182C, 16183C, 16189C, 16193d
7 16182C, 16183C, 16189C, 16193.1C
8 16182C, 16183C, 16189C, 16193.1C, 16193.2C
9 16182C, 16183C, 16187T, 16189C
10 16182C, 16183T, 16188T, 16189C
11 16182C, 16183C, 16187T, 16189C, 16193d
12 16183C, 16184A, 16189C
13 16183C, 16188T, 16189C, 16193.1C
14 16183C, 16189C
15 16183C, 16189C, 16190T, 16193.1C
16 16183C, 16189C, 16190T, 16193.1C, 16193.2C
17 16183C, 16189C, 16191T, 16193.1C
18 16183C, 16189C, 16191T
19 16183C, 16189C, 16193.1C
20 16183C, 16193.1C
21 16184A, 16189C
22 16184A
23 16185T, 16189C
24 16185T
25 16186T, 16189C
26 16186T
27 16187T, 16189C
28 16187T
```

It's possible to edit motif files manually, but the built-in editor is recommended.

Lines starting with “#” define the region of motifs listed below it.

#start-end (inclusive)

Mutations are listed with standard forensic nomenclature:

- **Deletion:** 16191d
- **Insertion:** 16193.1C
- **SNP:** 16184A

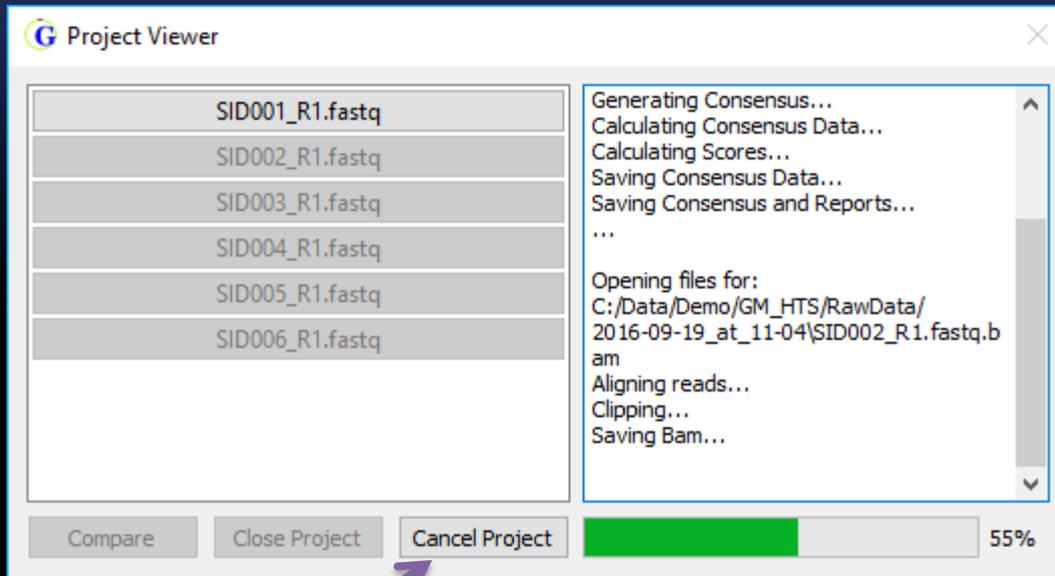
Important:

- Spacing of text is not important, but commas are.
- Regions cannot overlap



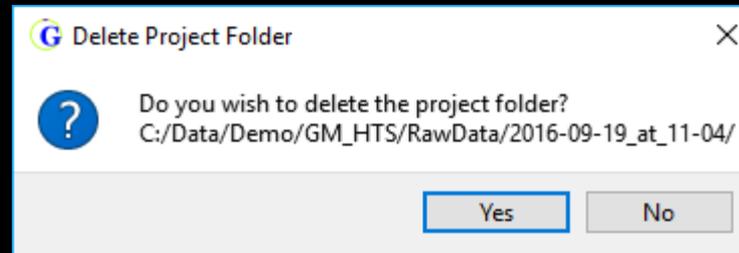
Sample Processing

After the all the desired settings are chosen, selecting *OK* will being alignment.



When a sample is finished it is possible to click on the button in the list of projects to open it- even before all samples finish processing

Projects can be canceled using the *Cancel Project* button. The *Project Viewer* will be closed after the next alignment finishes.



If the project is cancelled, the program will ask the user if they would like to delete the project folder that was created.



Viewing the Results

Viewing the Results

Clicking on a sample name will open it in a *Viewer* window.

The image shows two windows from the GeneMarker HTS software. The 'Project Viewer' window on the left lists six sample files: SID001_R1.fastq, SID002_R1.fastq, SID003_R1.fastq, SID004_R1.fastq, SID005_R1.fastq, and SID006_R1.fastq. A purple arrow points from the text above to the first sample name. The 'SID001_R1.fastq.bam Viewer' window on the right displays genomic data for the first sample. It includes a coverage plot at the top, a reference sequence, a consensus sequence, and a pile-up view. Below these is a table of variants.

Index	Position	Ref	Alt	IUPAC	Variants	Variants %	Coverage	A(#F;#R)	C(#F;#R)	G(#F;#R)	T(#F;#R)	Del(#F;#R)	Ins(#F;#R)	A%	C%	G%
1	73	A	G	73G	A73G	99.83	1818	2;1	0;0	888;927	0;0	0;0	0;0	0.16	0.00	99.83
2	195	T	C	195C	T195C	99.50	1830	1;0	915;906	0;0	2;1	0;5	0;0	0.05	99.50	0.00
3	263	A	G	263G	A263G	99.78	1822	1;2	1;0	903;915	0;0	0;0	0;0	0.16	0.05	99.78
4	489	T	C	489C	T489C	99.72	1814	0;0	911;898	1;0	0;4	0;0	0;0	0.00	99.72	0.05
5	16114	C	T	16114T	C16114T	99.94	1817	0;0	1;0	0;0	910;906	0;0	0;0	0.00	0.05	0.00



Viewing the Results

Multiple samples from the sample project can be opened at the same time. The *Viewer* windows will zoom and scroll horizontally in unison but they will not scroll vertically in unison. Any changes made to the *Filter Settings* will be applied to all *Viewers* of the same project.

If there are two separate projects open at the same time, the *Viewers* between projects are not linked.

Sample02.bam Viewer For Internal Use Only

Filter Settings Load Report Save Report Table Settings Toggle Table Coverage Report

Blocks 1899
 Variants 1266
 Zoom Global 633

Reference
Consensus
Pile-Up

Sample02.bam Major Minor Both

Index	Position	Ref	Alt	IUPAC	Variants	Variants %	Coverage	A(##;#R)	C(##;#R)	G(##;#R)	T(##;#R)	Del(##)
1	73	A	G	73R	A73G	83.90	1765	119;164	0;0	719;762	0;0	1;0
2	73	A	A	73R	A73A	16.03						
3	195	T	C	195Y	T195C	4.26	2134	0;0	51;40	0;0	1044;999	0;0
4	200	A	G	200R	A200G	12.85	2077	948;859	0;1	126;141	1;1	0;0
5	263	A	G	263G	A263G	99.83	1789	0;3	0;0	922;864	0;0	0;0
6	489	T	C	489Y	T489C	75.45	1968	0;0	749;736	0;0	226;257	0;0
7	489	T	T	489Y	T489T	24.54						

Sample01.bam Viewer For Internal Use Only

Filter Settings Load Report Save Report Table Settings Toggle Table Coverage Report

Blocks 1839
 Variants 1226
 Zoom Global 613

Reference
Consensus
Pile-Up

Sample01.bam Major Minor Both

Index	Position	Ref	Alt	IUPAC	Variants	Variants %	Coverage	A(##;#R)	C(##;#R)	G(##;#R)	T(##;#R)	Del(##)
1	73	A	G	73G	A73G	99.94	1674	0;0	0;0	796;877	0;1	0;0
2	153	A	G	153R	A153G	4.39	2141	1007;1039	0;0	42;52	0;0	0;1
3	263	A	G	263G	A263G	99.94	1757	0;0	0;0	869;887	0;1	0;0
4	489	T	C	489Y	T489C	4.52	2255	0;0	46;56	0;0	1100;1053	0;0
5	709	G	A	709R	G709A	12.86	2153	139;138	1;0	987;886	0;1	0;1
6	750	A	G	750G	A750G	99.71	1760	2;2	0;0	893;862	0;1	0;0
7	827	A	G	827R	A827G	74.86	1886	236;237	0;0	702;710	0;1	0;0



Viewing the Results

Global view

Reference and Consensus Sequences

Pileup

Toggle between different reports

Sample Name

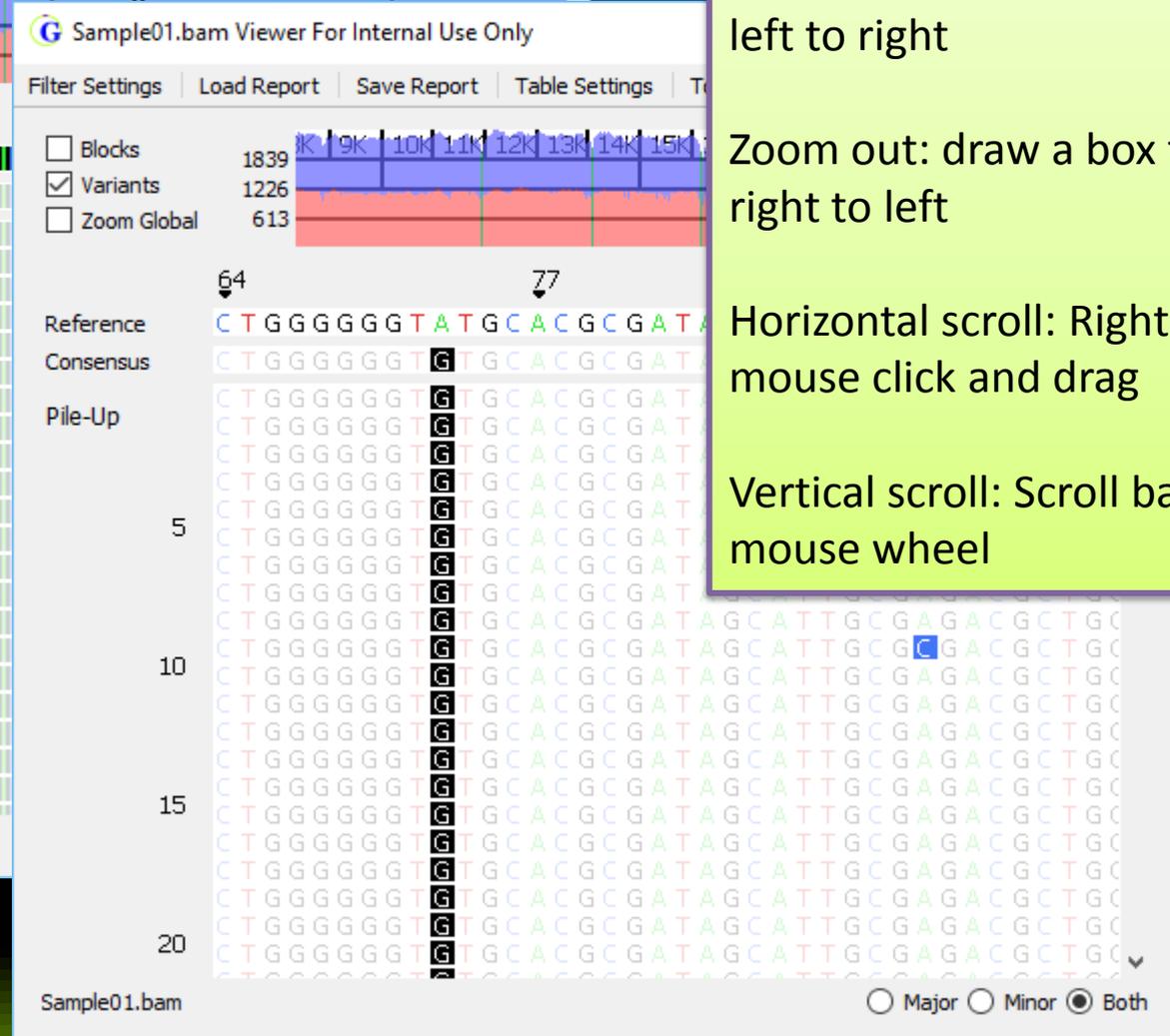
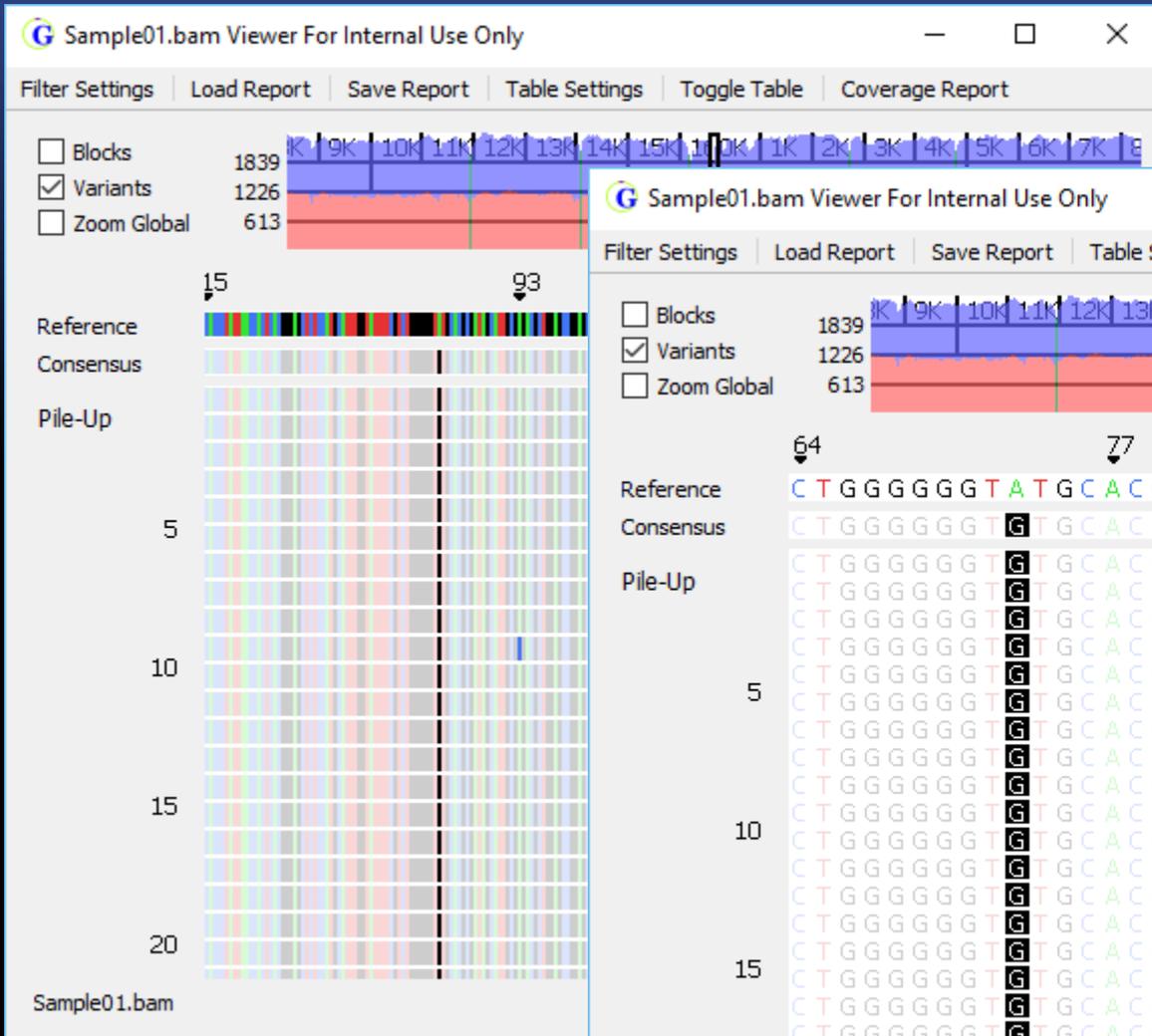
Result Table

The *Global View* shows the depth of coverage with forward read coverage in **blue** and reverse read coverage in **red**.

Index	Position	Ref	Alt	IUPAC	variants	variants %	Coverage	A(#F;#R)	C(#F;#R)	G(#F;#R)	T(#F;#R)	Del(#F;#R)	Ins(#F;#R)	A%	C%	G%
1	73	A	G	73G	A73G	99.83	1818	2;1	0;0	0;0	0;0	0;0	0;0	0.16	0.00	99.83
2	195	T	C	195C	T195C	99.50	1830	1;0	0;0	0;0	2;1	0;5	0;0	0.05	99.50	0.00
3	263	A	G	263G	A263G	99.78	1822	1;2	1;0	903;915	0;0	0;0	0;0	0.16	0.05	99.78



Viewing the Results



Navigation tools:

Zoom in: draw a box from left to right

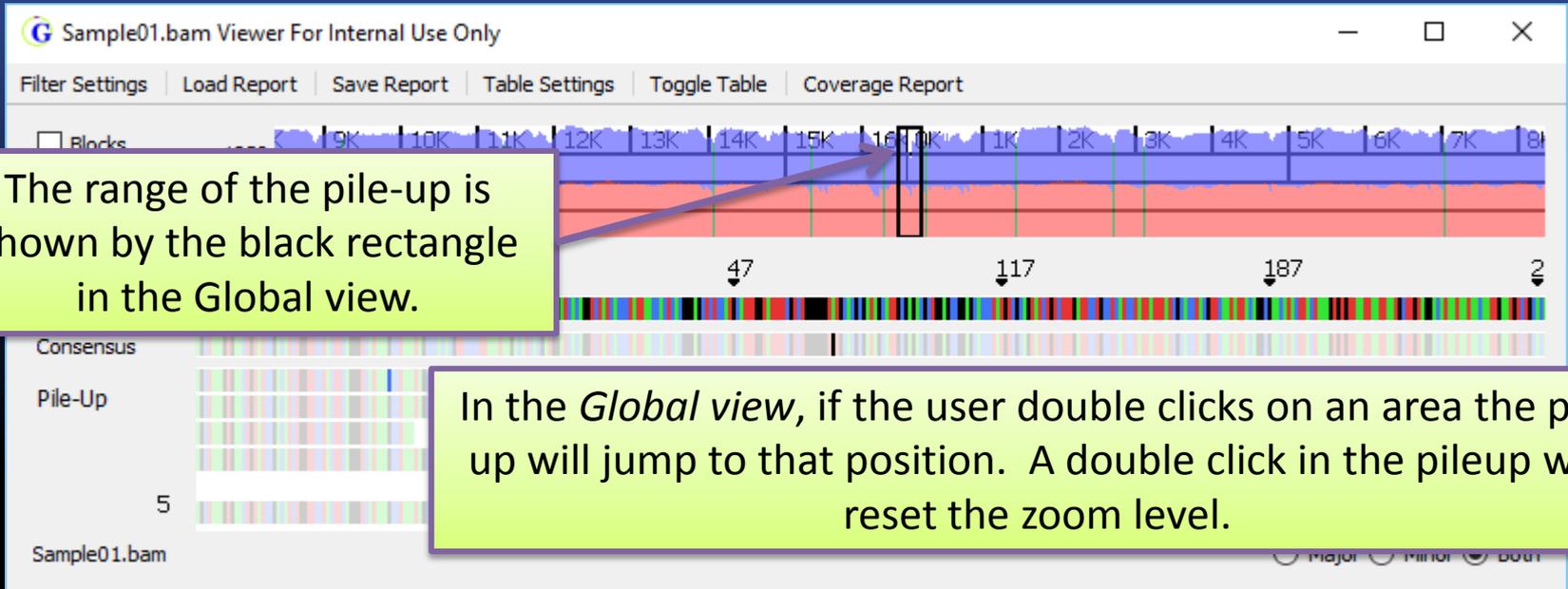
Zoom out: draw a box from right to left

Horizontal scroll: Right mouse click and drag

Vertical scroll: Scroll bar or mouse wheel



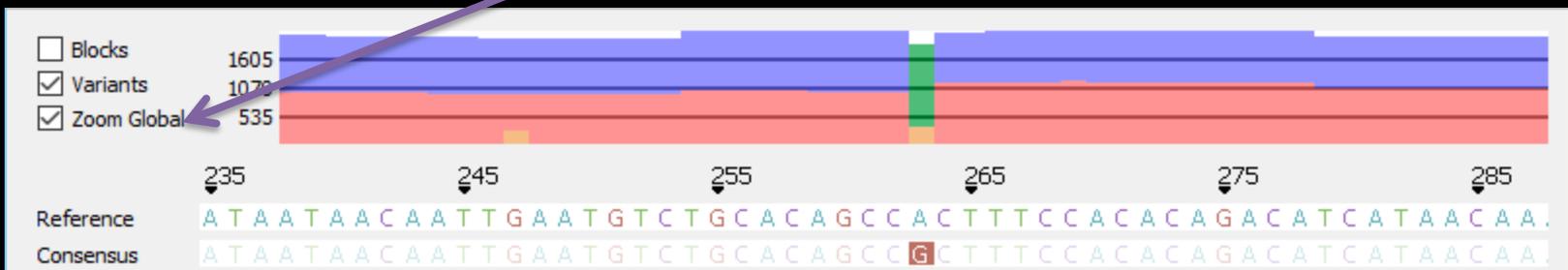
Viewing the Results



The range of the pile-up is shown by the black rectangle in the Global view.

In the *Global view*, if the user double clicks on an area the pile up will jump to that position. A double click in the pileup will reset the zoom level.

Zoom Global will zoom the global view to only the pileup region



The **teal lines** in the *Global View* represent the location of variants. The **orange lines** represent the percent of reads that didn't match the consensus sequence at that position.



Viewing the Results

Default Color Key:

White area/space: in
between reads

Deletion: **Burnt Orange**

Insertion: **Purple**

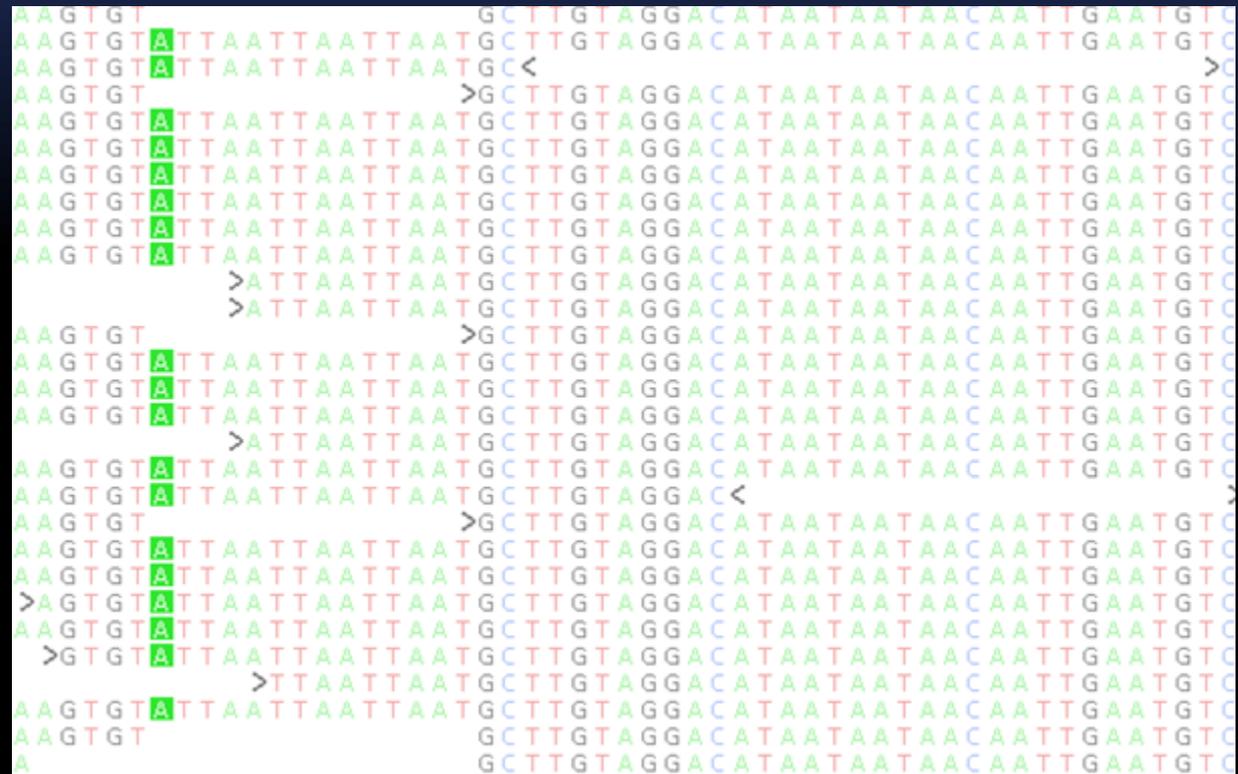
A: **Green**

C: **Blue**

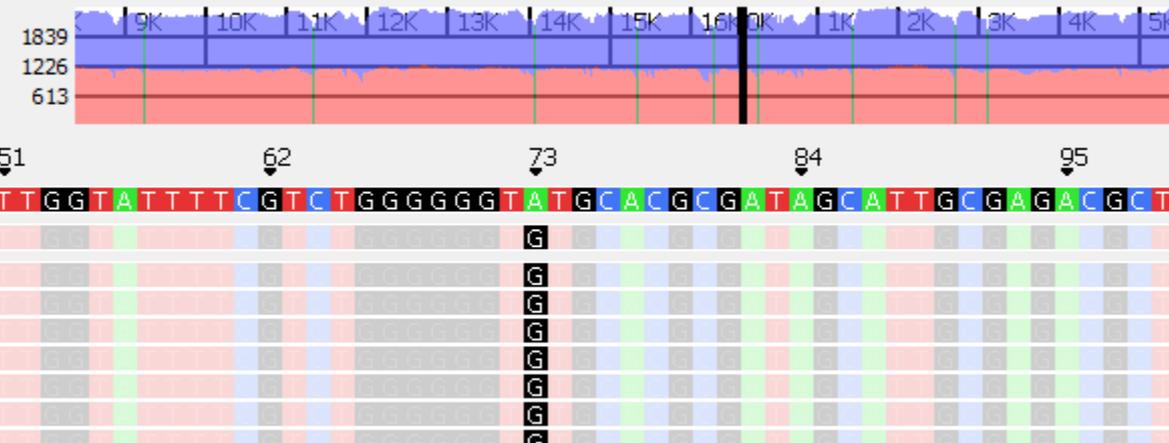
G: **Black**

T: **Red**

N: **Gray**



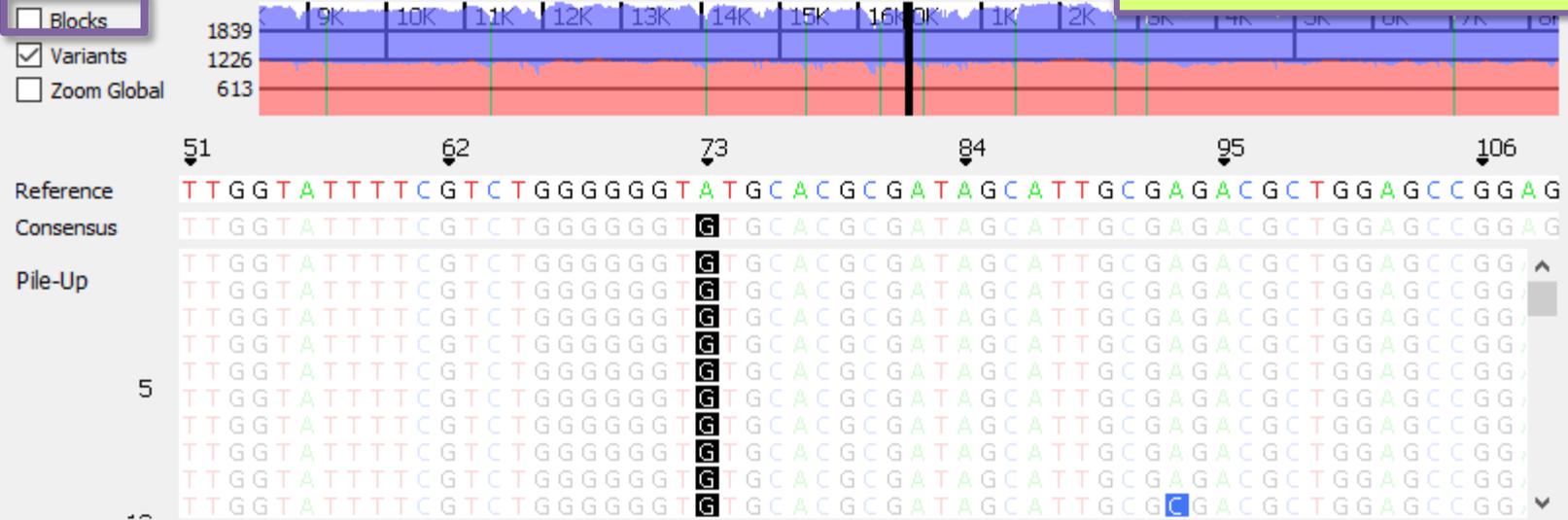
- Blocks
- Variants
- Zoom Global



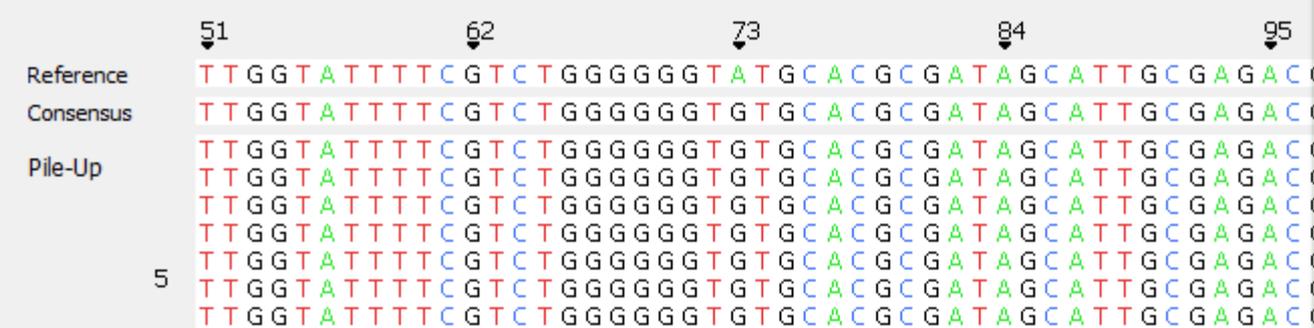
When *Blocks* is selected, the background becomes colored and the font is white.

When *Blocks* is not selected, the text will be colored and displayed on a white background. *Blocks* is turned off by default.

- Blocks
- Variants
- Zoom Global



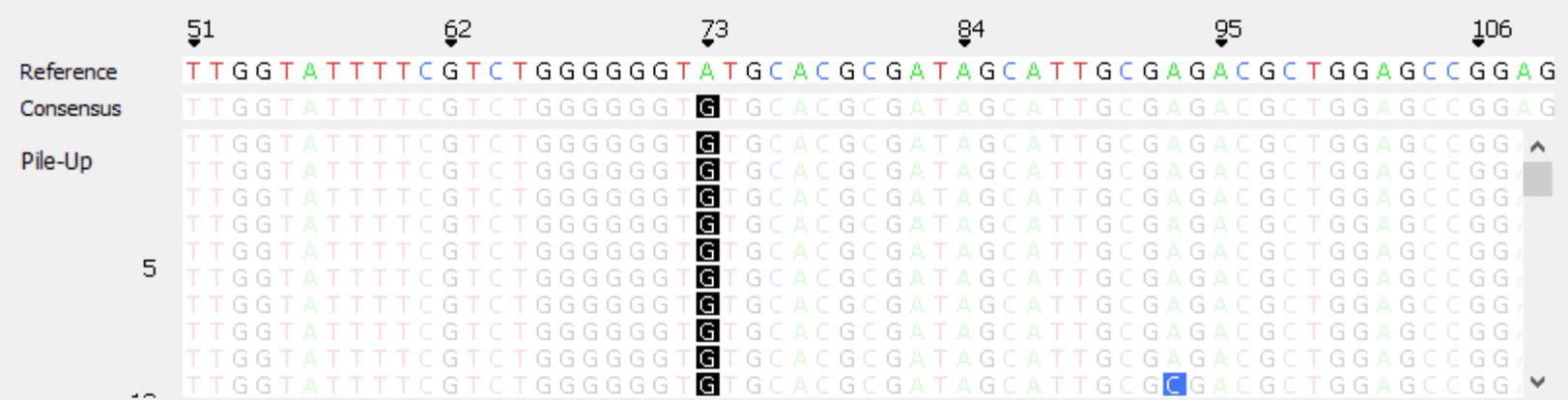
- Blocks
- Variants
- Zoom Global



When *Variants* is selected, the variants will be shown in block format and non-variants are dimmed. *Variants* is selected by default.

When *Variants* is not selected, the variants will not be highlighted from the other nucleotides.

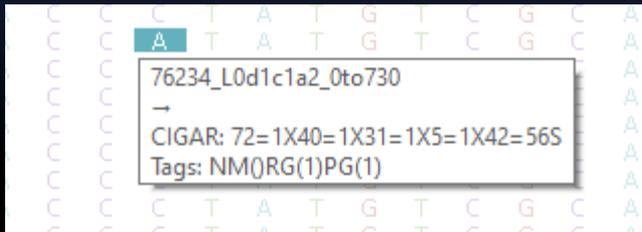
- Blocks
- Variants
- Zoom Global



Viewing the Results

To show information for a specific read:

- Hold SHIFT and center mouse wheel click
- Hold CTRL + SHIFT and left-click with mouse



Keyboard Shortcuts:

- Ctrl + f: The pile-up will jump to center on the next entry in the table.
- Ctrl + b: The pile-up will jump to center on the previous entry in the table.

To show information for a specific variant:

- Center mouse wheel click
- Hold CTRL and left-click with mouse

A screenshot showing a variant information popup. The popup contains the following text:

Position	73
Nucleotide	G
Direction	←
A	3(0%, +2, -1)
C	0(0%)
G	1815(99%, +888, -927)
T	0(0%)
Inserts	0(0%)
Deletes	0(0%)
Coverage	1818
Balance Ratio(G)	1.001172



Viewing the Results

The screenshot displays a Pile-Up view of sequencing data. At the top, three positions are marked: 695, 704, and 713. The alignment consists of multiple rows of colored letters representing nucleotides (A, T, G, C). A context menu is open over the alignment, listing the following options:

- Add A for position 699
- Add C for position 699
- Add G for position 699
- Add T for position 699
- Add del for position 699
- Go to closest position in table
- Copy Read's Sequence

Two purple arrows point from the text boxes to the 'Go to closest position in table' and 'Copy Read's Sequence' options in the menu.

The user may right-click at any position in the *Pile-Up* and choose *Go to closest position in table* and the *Results Table* will highlight the closest position.

There is also the option to copy the sequence from any read.



Viewing the Results

Sample01.bam Viewer For Internal Use Only

Filter Settings | Load Report | Save Report | Table Settings | Toggle Table | Coverage Report

Blocks 1839 | 9K | 10K | 11K | 12K | 13K | 14K | 15K | 16K | 17K | 18K | 19K | 20K | 21K | 22K | 23K | 24K | 25K | 26K | 27K | 28K | 29K | 30K | 31K | 32K | 33K | 34K | 35K | 36K | 37K | 38K | 39K | 40K | 41K | 42K | 43K | 44K | 45K | 46K | 47K | 48K | 49K | 50K

The *Results Table* is displayed at the bottom of the *Sample Viewer* window.

Reference: G A T C A C A G G T C T A T C A C C C T A
Consensus: G A T C A C A G G T C T A T C A C C C T A
Pile-Up: G A T C A C A G G T C T A T C A C C C T A
5: G A T C A C A G G T C T A T C A C C C T A
10: G A T C A C A G G T C T A T C A C C C T A

Sample01.bam

Major Minor Both

Index	Position	Ref	Alt	IUPAC	Variants	Variants %	Coverage	F(#;#R)	C(#F;#R)	G(#F;#R)	T(#F;#R)	Del(#F;#R)	Ins(#F;#R)	
1	73	A	G	73G	A73G	99.94	1674	0;0	0;0	796;877	0;1	0;0	0;0	
2	153	A	G	153R	A153G	4.39	2141	1007;1039	0;0	42;52	0;0	0;1	0;0	
3	263	A	G	263G	A263G	99.94	1757	0;0	0;0	869;887	0;1	0;0	0;1	
4	489	T	C	489Y	T489C	4.52	2255	0;0	46;56	0;0	1100;1053	0;0	0;0	
5	709	G	A	709R	G709A	12.86	2153	139;138	1;0	987;886	0;1	0;1	0;1	

The *Results Table* can be made larger or smaller by left-clicking and dragging above the table.

The user may choose to display the major, minor, or all variants.

When a position in the *Results Table* is clicked on, the pile-up will jump to that position.



Allele Calls

Major alleles are the highest frequency allele at that position

Minor alleles are any other alleles that have frequency > the minimum threshold

AUT0500037(0.8)_AUT0500009(0.15)_AUT0500182(0.05)_100k_250bp.bam Major Minor Both

Index	Position	Ref	Alt	Variants	Variants %	A%	C%	G%	T%	Del%	Ins%
1	73	A	G	A73G	83.71	16.17	0.04	83.71	0.01	0.04	0.00
2	73	A	A	A73A	16.17						
3	94	G	A	G94A	4.70	4.70	0.02	95.24	0.01	0.00	0.00
4	189	A	G	A189G	14.50	85.44	0.01	14.50	0.02	0.00	0.00
5	195	T	C	T195C	79.13	0.00	79.13	0.02	20.81	0.01	0.01

AUT0500037(0.8)_AUT0500009(0.15)_AUT0500182(0.05)_100k_250bp.bam Major Minor Both

Index	Position	Ref	Alt	Variants	Variants %	A%	C%	G%	T%	Del%	Ins%
1	73	A	G	A73G	83.71	16.17	0.04	83.71	0.01	0.04	0.00
2	195	T	C	T195C	79.13	0.00	79.13	0.02	20.81	0.01	0.01
3	263	A	G	A263G	99.87	0.02	0.04	99.87	0.02	0.02	0.00
4	315	C	insC	C315insC	99.80	0.01	99.95	0.01	0.00	0.00	99.80
5	16126	T	C	T16126C	78.44	0.01	78.44	0.01	21.50	0.00	0.01

AUT0500037(0.8)_AUT0500009(0.15)_AUT0500182(0.05)_100k_250bp.bam Major Minor Both

Index	Position	Ref	Alt	Variants	Variants %	A%	C%	G%	T%	Del%	Ins%
1	73	A	A	A73A	16.17	16.17	0.04	83.71	0.01	0.04	0.00
2	94	G	A	G94A	4.70	4.70	0.02	95.24	0.01	0.00	0.00
3	189	A	G	A189G	14.50	85.44	0.01	14.50	0.02	0.00	0.00
4	195	T	T	T195T	20.81	0.00	79.13	0.02	20.81	0.01	0.01
5	309	C	insC	C309insC	4.75	0.00	99.89	0.01	0.02	0.05	4.75

Reference alleles may be reported as minor alleles when their frequency fits the criteria because they may represent a secondary mutation back from the major allele.

Major alleles matching the reference are not reported

Insertions are reported as *Major* alleles if they are above 50% and *Minor* if they are below

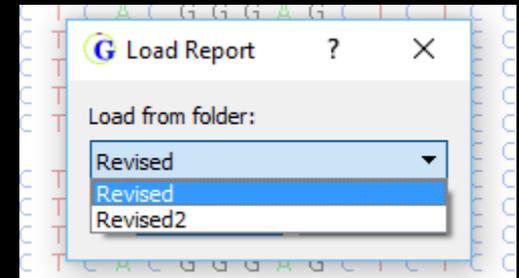
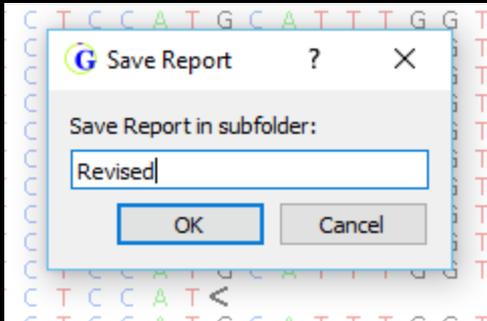


Adding and Removing Variants

Variants may be manually added (right-click in the pileup) or removed (right-click in the table). Added variants are **shaded green** and removed variants are **shaded red**.
Either kind of edit can include comments

Index	Position	Ref	Alt	Variants	Variants %	Coverage	A%	C%	G%	T%	Del%	Ins%
1	73	A	G	A73G	99.83	1805	0.11	0.00	99.83	0.05	0.00	0.00
2	114	C	A	C114A	12.19	2001	12.19	87.75	0.00	0.04	0.00	0.04
3	143	G	C	G143C	0.04	2240	0.00	0.04	99.95	0.00	0.00	0.00
4	146	T	C	T146C	12.01	2206	0.00	12.01	0.00	87.98	0.00	0.00
5	152	T	C	T152C	16.98	2220	0.00	16.98	0.00	83.01	0.00	0.00
6	195	T	C	T195C	16.75	2172	0.00	16.75	0.00	83.24	0.00	0.00

The report can be saved as a new report after changes to settings or manual edits. The original project is opened by default, but all previously saved reports are maintained in subfolders with changelogs.



Viewing the Results

 Sample01.bam Viewer For Internal Use Only

Filter Settings | Load Report | Save Report | Table Settings | Toggle Table | Coverage Report

The **Filter Settings** button will open the *Table Filter Settings* window that was available in the *New Project* window.

The **Load Report** button will allow the user to change to a different saved report.

The **Save Report** button will allow the user to save the report using the current filter settings and edited variants.

The **Table Settings** button will open the *Table Settings* window. Here the user can choose what information is displayed in the table.

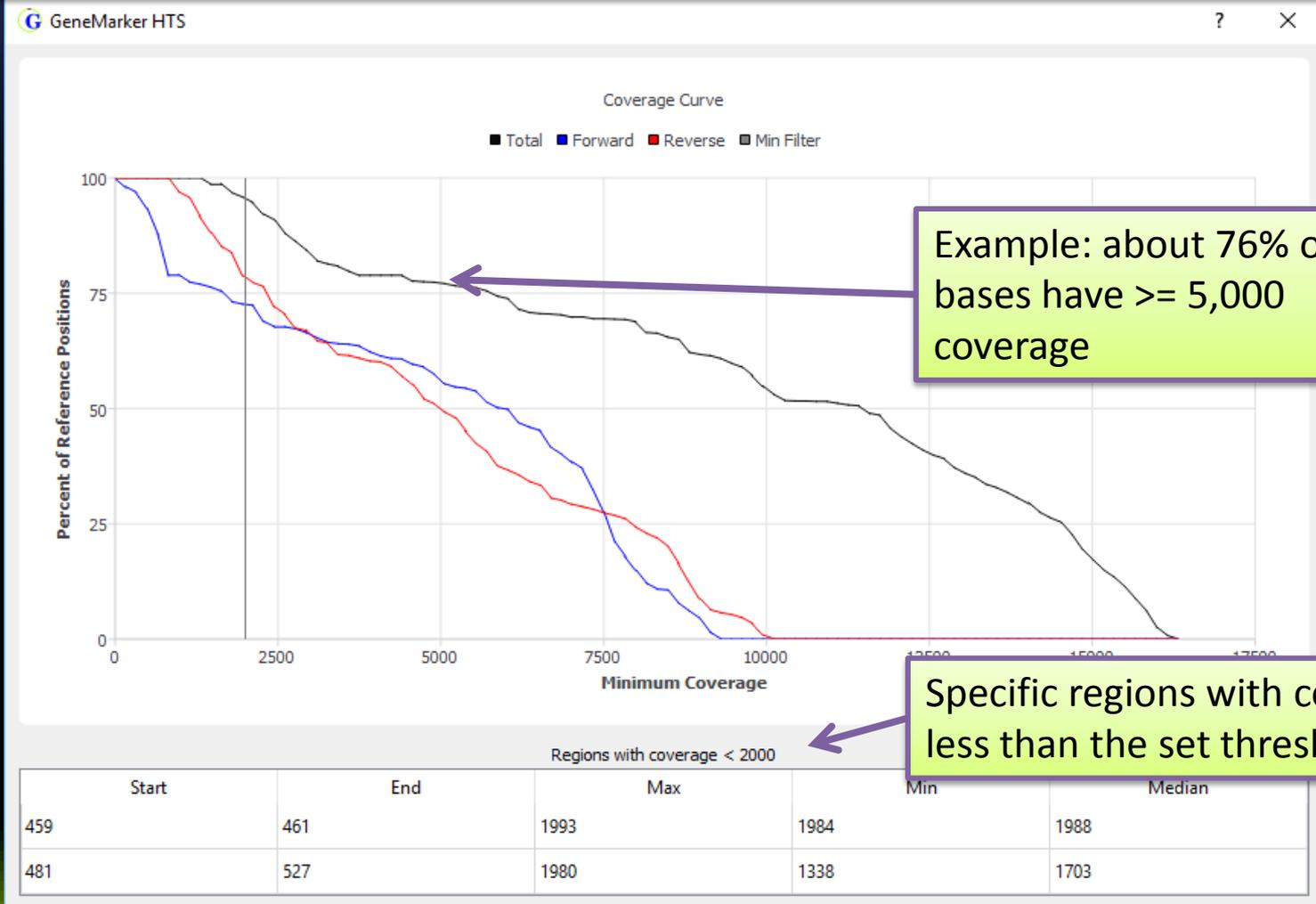
The **Toggle Table** button will allow the user to display or hide the table below the pile-up.

The **Coverage Report** button will open the coverage report.



Coverage Report

The *Coverage Report* gives a quick overview of the coverage depth, specifically the fraction of all bases (entire reference, or range specified in filter settings) that have \geq some level of coverage (specified in filter settings)



Example: about 76% of bases have \geq 5,000 coverage

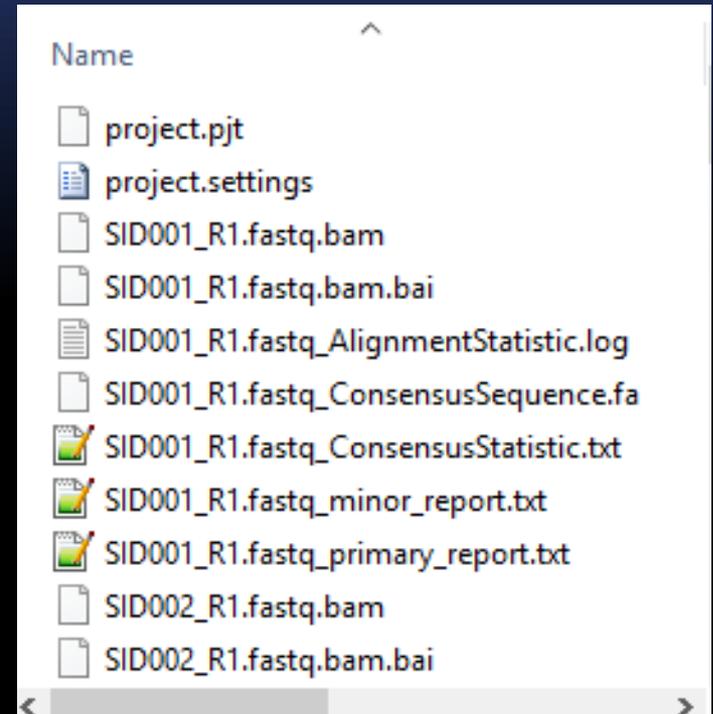
Specific regions with coverage less than the set threshold



Output Files

The program will output the following pieces of information for each sample in the project:

- **BAM/BAI files:** Alignment results
- **Alignment Statistic Log:** Summary of alignment results
- **Consensus Sequence:** Consensus sequence with primary alleles in FASTA format
- **Consensus Statistics:** Report listing the number of copies of each allele at each position in the sample.
- **Primary Allele Report:** Variant report for primary alleles
- **Minor Allele Report:** Variant report for minor alleles

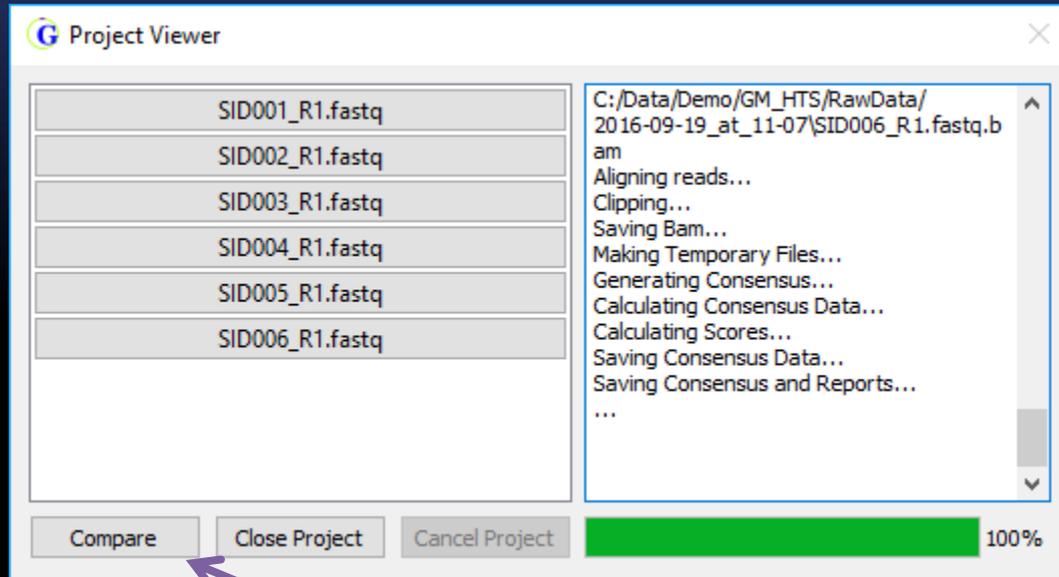


Project file and Project Settings file: Used by software to track settings and data



Comparison Viewer

Comparison Tool



All of the samples within a project can be opened in a *Comparison Viewer* using the "Compare" button

Comparison Tool

Sample to Sample

Comparison Viewer

Filter Settings | Major to Major | Major to Minor | Minor to Minor

Sample to Sample Comparison Proportion of Shared Variants: None All

	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa
SID001_R1.fa	100%	100%	41%	41%	20%	17%
SID002_R1.fa	100%	100%	41%	41%	20%	17%
SID003_R1.fa	41%	41%	100%	100%	10%	18%
SID004_R1.fa	41%	41%	100%	100%	10%	18%
SID005_R1.fa	20%	20%	10%	10%	100%	7%
SID006_R1.fa	17%	17%	18%	18%	7%	100%

Row: Both Major Minor Column: Both Major Minor

Showing (Row's Major ∩ Column's Major) / (Row's Major ∪ Column's Major) Show Percentage Use Jaccard Index

Variant Comparison

Variant Comparison Major Allele Minor Allele Low Coverage

	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa
A73A	0%	0%	0%	6%	99%	0%
A73G	99%	99%	99%	93%	0%	99%
A93G	0%	4%	0%	0%	0%	0%
T146C	0%	4%	0%	0%	0%	0%
C150T	0%	15%	0%	0%	0%	0%
T152C	0%	15%	0%	0%	0%	0%

Comparison Tool

The *Sample to Sample Comparison* table (top half) shows a similarity table for all samples in the project.

Comparison Viewer

Filter Settings | Major to Major | Major to Minor | Minor to Minor

Sample to Sample Comparison Proportion of Shared Variants: None All

	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa
SID001_R1.fa	100%	100%	41%	41%	20%	17%
SID002_R1.fa	100%	100%	41%	41%	20%	17%
SID003_R1.fa	41%	41%	100%	100%	10%	18%
SID004_R1.fa	41%	41%	100%	100%	10%	18%
SID005_R1.fa	20%	20%	10%	10%	100%	7%
SID006_R1.fa	17%	17%	18%	18%	7%	100%

Row: Both Major Minor

Column: Both Major Minor

Showing (Row's Major \cap Column's Major) / (Row's Major \cup Column's Major)

Show Percentage Use Jaccard Index

Comparisons can be made between major alleles, minor alleles, or both on each axis (row and column)

It is possible to show the absolute number of variants in common, or the percentage.

Comparison Tool

Comparison Viewer

Filter Settings | Major to Major | Major to Minor | Minor to Minor

Sample to Sample Comparison

Proportion of Shared Variants: None All

	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa
SID001_R1.fa	9/9	9/9	5/12	5/12	2/10	3/17
SID002_R1.fa	9/9	9/9	5/12	5/12	2/10	3/17
SID003_R1.fa	5/12	5/12	8/8	8/8	1/10	3/16
SID004_R1.fa	5/12	5/12	8/8	8/8	1/10	3/16
SID005_R1.fa	2/10	2/10	1/10	1/10	3/3	1/13
SID006_R1.fa	3/17	3/17	3/16	3/16	1/13	11/11

Row: Both Major Minor

Column: Both Major Minor

Showing (Row's Major \cap Column's Major) / (Row's Major \cup Column's Major)

Show Percentage Use Jaccard Index

The formula used to calculate similarity is shown. It is possible to use:

- A Jaccard Index measurement (Row AND Column / Row OR Column)
- Simple measurement (Row AND Column / Column)

Comparison Viewer

Filter Settings | Major to Major | Major to Minor | Minor to Minor

Sample to Sample Comparison

Proportion of Shared Variants: None All

	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa
SID001_R1.fa	9/9	9/9	5/8	5/8	2/3	3/11
SID002_R1.fa	9/9	9/9	5/8	5/8	2/3	3/11
SID003_R1.fa	5/9	5/9	8/8	8/8	1/3	3/11
SID004_R1.fa	5/9	5/9	8/8	8/8	1/3	3/11
SID005_R1.fa	2/9	2/9	1/8	1/8	3/3	1/11
SID006_R1.fa	3/9	3/9	3/8	3/8	1/3	11/11

Row: Both Major Minor

Column: Both Major Minor

Showing (Row's Major \cap Column's Major) / (Column's Major)

Show Percentage Use Jaccard Index

Diagonal cells (comparing a sample to itself) are colored black, while others are based on a color gradient from 0 to 100%

Comparison Tool

The *Variant Comparison* table (bottom half) simply shows the allele frequency of all variants called in at least one sample. Cells in the table are colored according to whether or not the variant was a major allele or minor allele in that sample, or if the total coverage was below the set threshold.

Variant Comparison							Major Allele	Minor Allele	Low Coverage
	SID001_R1.fa	SID002_R1.fa	SID003_R1.fa	SID004_R1.fa	SID005_R1.fa	SID006_R1.fa			
A73A	0%	0%	0%	6%	99%	0%			
A73G	99%	99%	99%	93%	0%	99%			
A93G	0%	4%	0%	0%	0%	0%			
T146C	0%	4%	0%	0%	0%	0%			
C150T	0%	15%	0%	0%	0%	0%			
T152C	0%	15%	0%	0%	0%	0%			

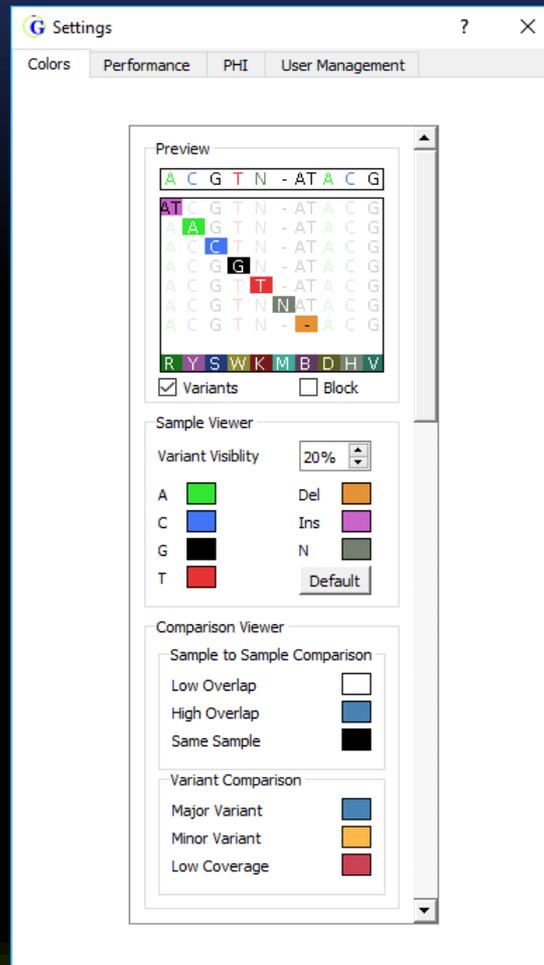
Program Settings

A settings button in the upper-left of the window allows for some customization and control of user management

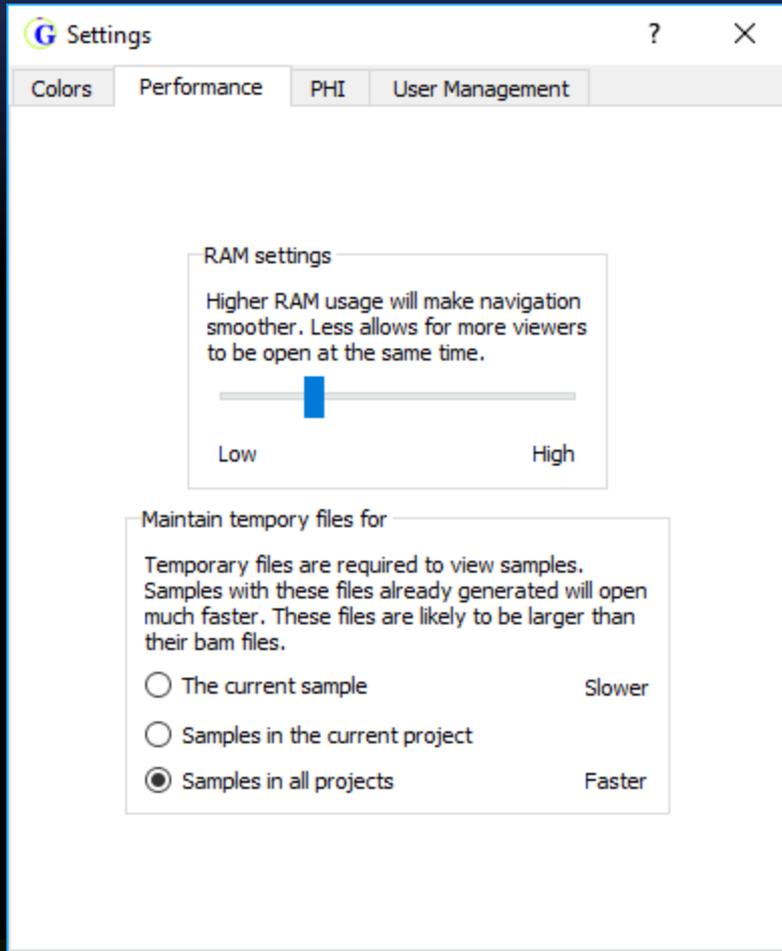


Program Settings - Colors

Colors used for highlighting in the pileup and in the comparison tool can be edited.



These settings can adjust the RAM and disk space usage of the software to optimize resource usage



Ram Settings

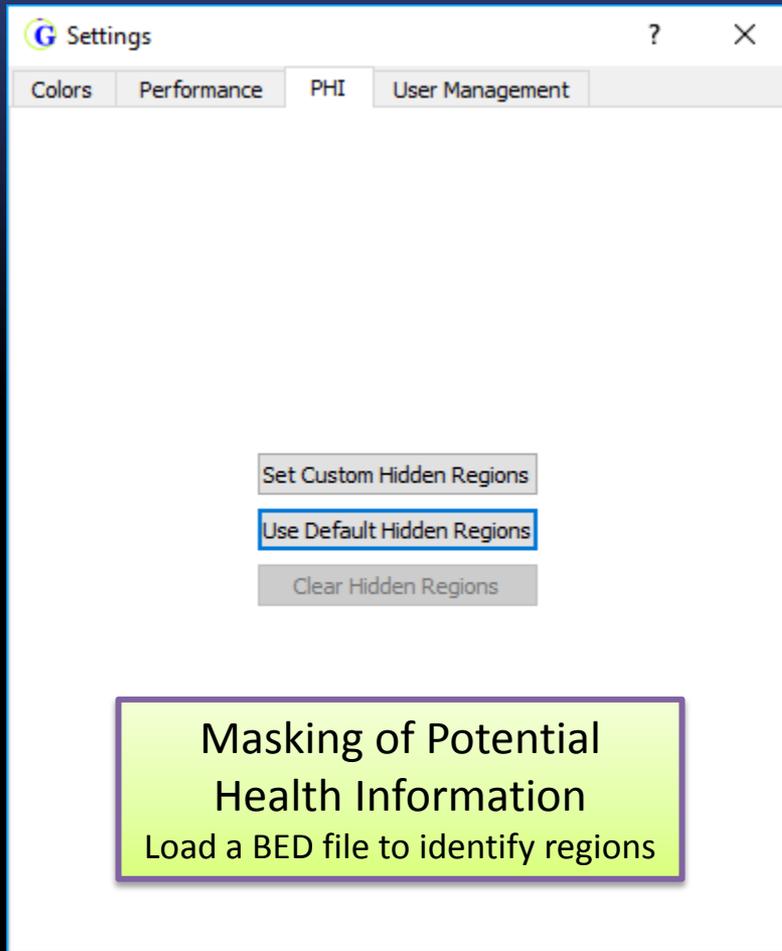
Setting this value higher will allow the program to use more RAM while viewing projects. This will enhance the speed of moving around in the pileup, but setting it lower will allow more projects to be opened at once.

Temporary Files

The temp files must be generated before opening a project. The files can be kept in order to save time whenever a sample is opened later on. This setting can allow them to be kept on-disk always, only for the currently opened project, or only for the currently opened sample. These files will take up extra disk space when they are kept.



Program Settings – Personal Health Information



The screenshot shows a 'Settings' window with tabs for 'Colors', 'Performance', 'PHI', and 'User Management'. The 'PHI' tab is active, displaying three buttons: 'Set Custom Hidden Regions', 'Use Default Hidden Regions' (highlighted with a blue border), and 'Clear Hidden Regions'. Below the buttons is a text box containing the text: 'Masking of Potential Health Information' and 'Load a BED file to identify regions'.



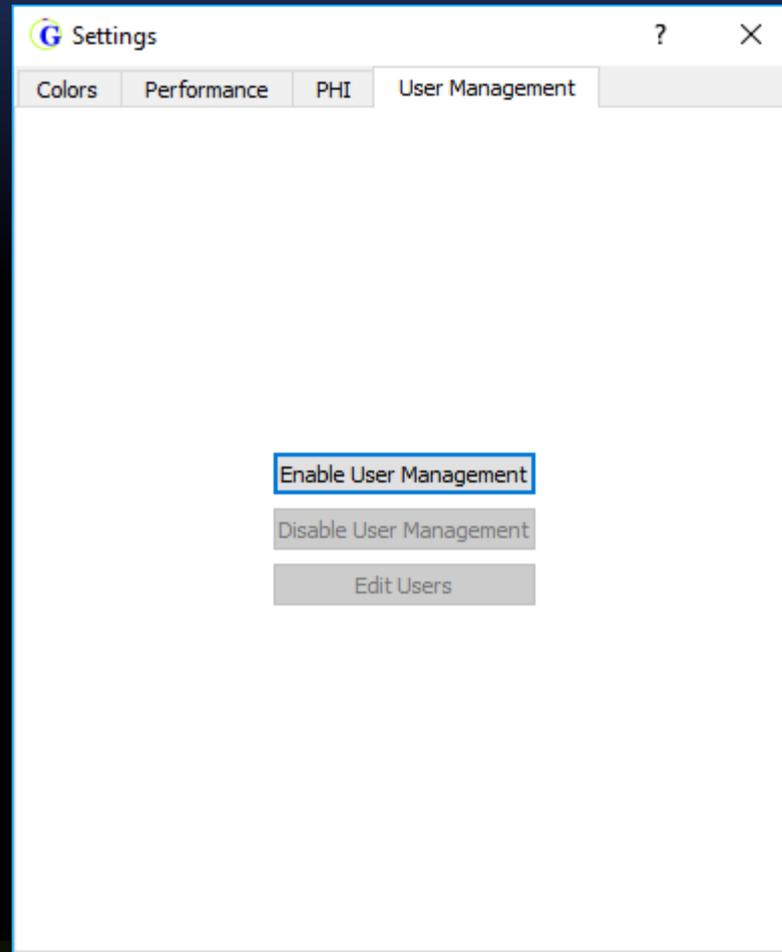
- Bases in the pileup are hidden
- Sequence cannot be copied for reads crossing these regions
- The CIGAR string is hidden from ctrl-shift-click
- Allele counts are hidden from all reports
- No variants are called

Default regions are available as marked in MitoMap



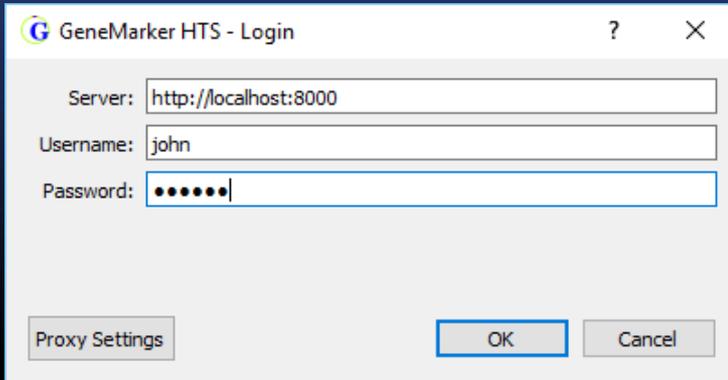
Program Settings – User Management

The software has User Management functionality which can be enabled or disabled by the administrator account of the current computer



User Management

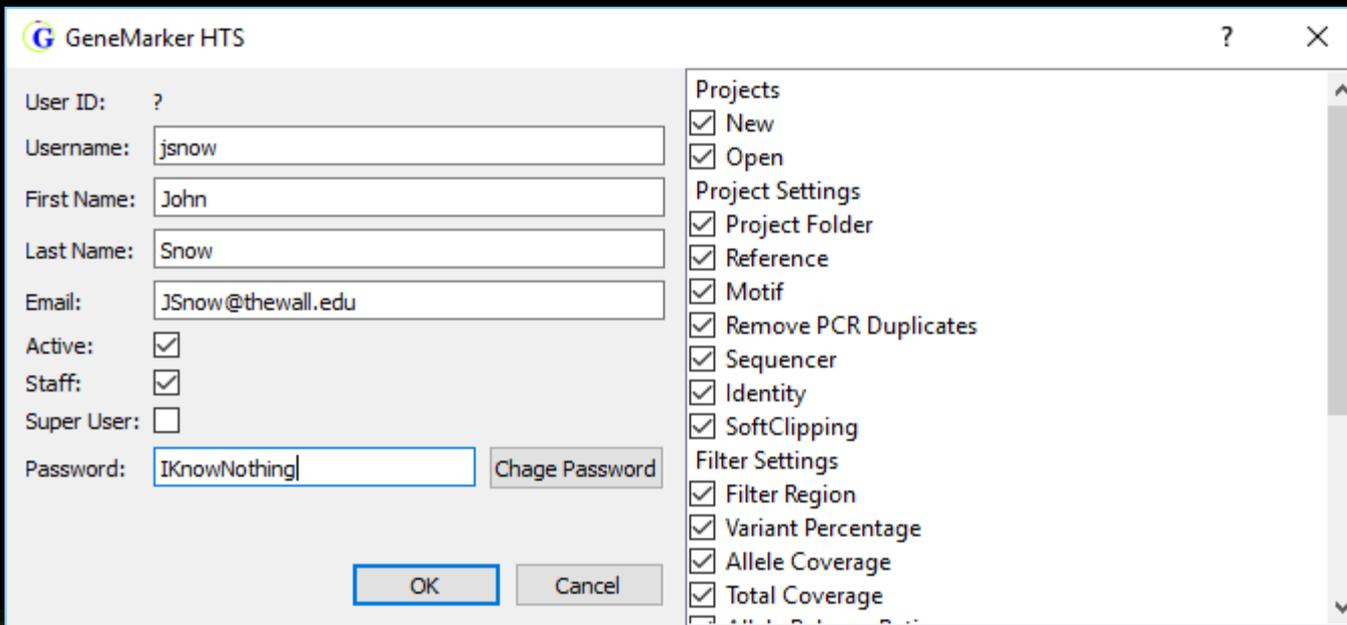
A windows administrator can enable or disable user management.



The screenshot shows the 'GeneMarker HTS - Login' dialog box. It contains the following fields and controls:

- Server:
- Username:
- Password:
- Buttons: Proxy Settings, OK, Cancel

User management is made possible by the SoftGenetics server, which may run on a different computer.



The screenshot shows the 'GeneMarker HTS' user management dialog box. It contains the following fields and controls:

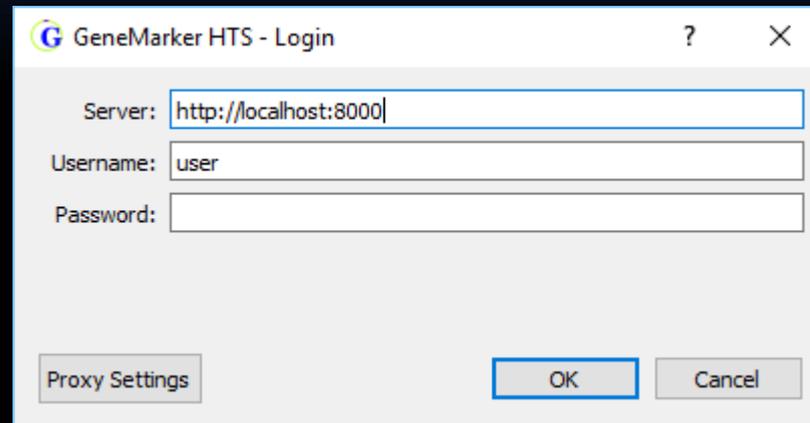
- User ID: ?
- Username:
- First Name:
- Last Name:
- Email:
- Active:
- Staff:
- Super User:
- Password:
- Buttons: OK, Cancel
- Permissions List (all checked):
 - Projects
 - New
 - Open
 - Project Settings
 - Project Folder
 - Reference
 - Motif
 - Remove PCR Duplicates
 - Sequencer
 - Identity
 - SoftClipping
 - Filter Settings
 - Filter Region
 - Variant Percentage
 - Allele Coverage
 - Total Coverage

Super users can create new users with limited permissions for almost any feature in the software



User Management

When user management is enabled, the software prompts the user for a login at startup.



The image shows a Windows-style dialog box titled "GeneMarker HTS - Login". The dialog has a title bar with a question mark and a close button. It contains three input fields: "Server:" with the text "http://localhost:8000", "Username:" with the text "user", and "Password:" which is empty. At the bottom, there are three buttons: "Proxy Settings", "OK", and "Cancel".

GeneMarker HTS - Login

Server:

Username:

Password:

Proxy Settings OK Cancel



Please contact tech_support@softgenetics.com
if further assistance is needed.

Visit our website for more information:
www.softgenetics.com

Thank you for using GeneMarker[®] HTS Software!

