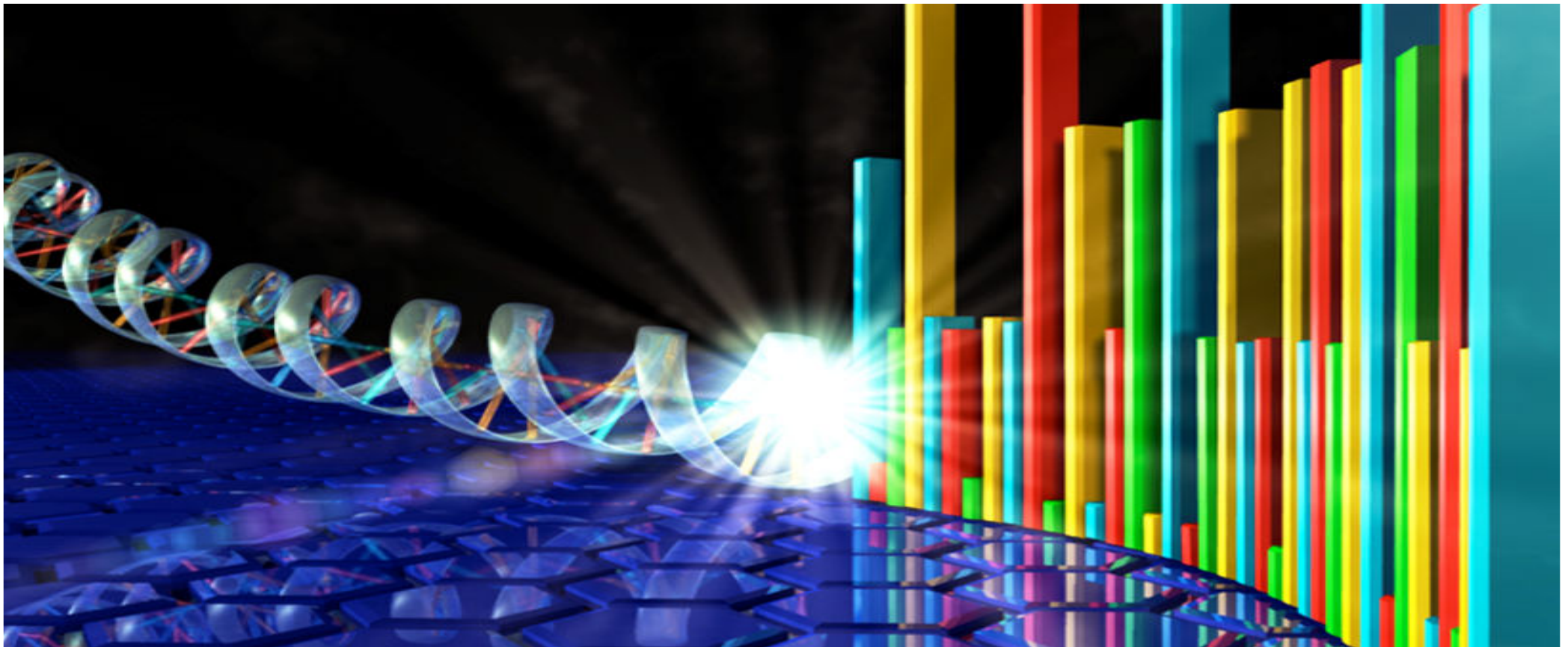


The Genome Sequencer FLX™ Software 2.0.00 Titanium

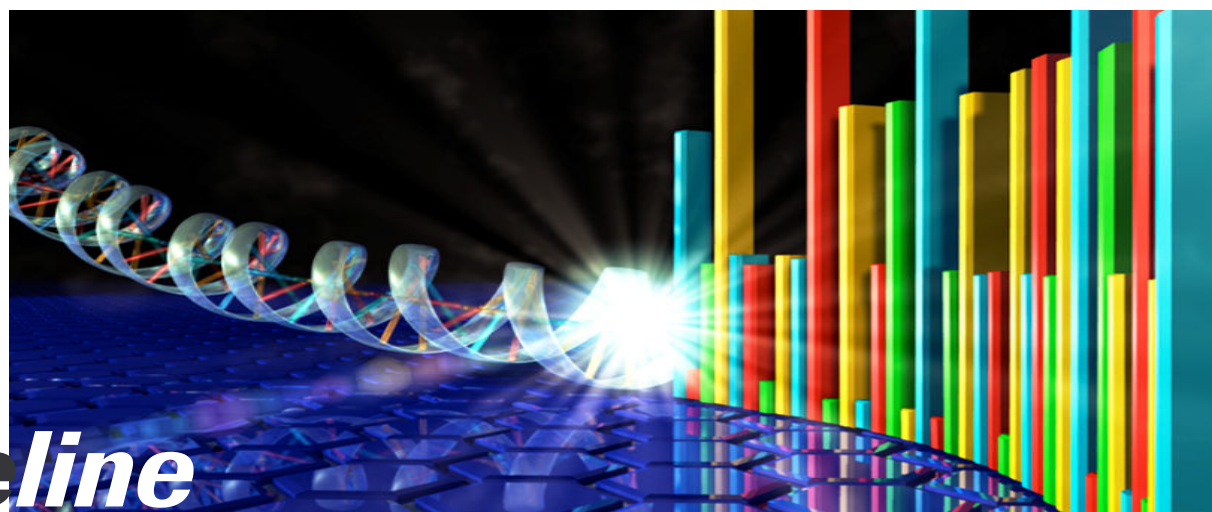
What is new ?



Software Package List

Package	On Instrument	Off Instrument	RPM Available
rigControl / Sequencer	Only	Never	No
gsRunProcessor	Required	Yes	OpenMPI-Yes Self Contained -Yes
gsReporter	Required	Yes	Yes
gsRunProcessorManager	Required	Optional*	Yes
gsRunBrowser	Required	Optional	No
gsAssembler	Never	Optional	No
gsMapper	Never	Optional	No
gsAmplicon	Never	Optional	No
gsSupportTool	Optional	Optional	Yes

* Required if users want to launch jobs with gsRunBrowser



Pipeline

Major Changes in the Pipeline

- Elimination of IPC in lieu of industry standard Message Passing Interface (MPI)
 - No more sMinProc.sh
 - No more /usr/local/rig/ required (off instrument)
 - No more pipeline configuration files
 - No more internal shell scripts
 - The pipeline can run on nearly any number of computers simultaneously (“cluster”)
- Retains the familiar runImagePipe, runAnalysisPipe, runAnalysisFilter commands
- Single executable called “gsRunProcessor” contains all the pipeline code
- Introduction of “processing scripts”

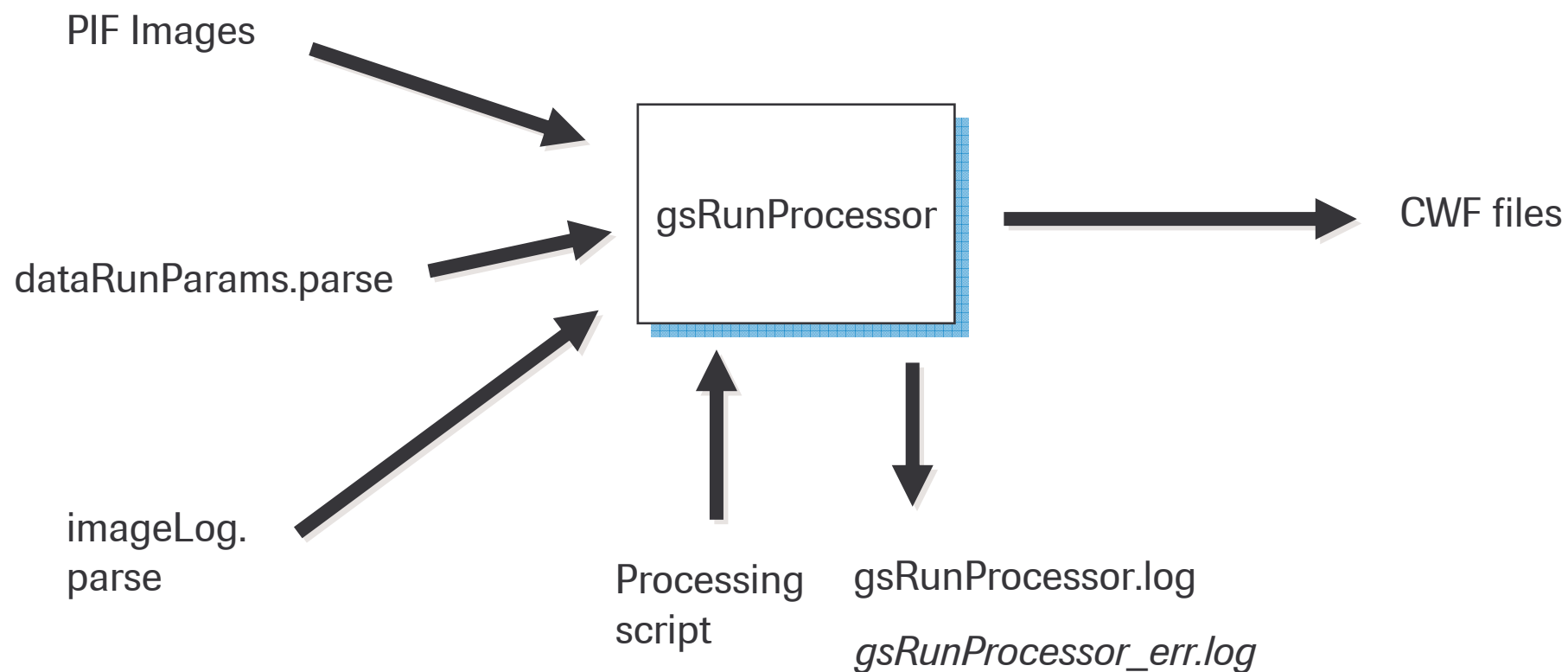
Data Set Layout for 2.0.00

- R_2008_06_04_10_11_24_rig2_jkeller_jktestrun } Image Capture
 - rawImages/
 - D_2008_06_04_10_11_32_rig2_imageProcessingOnly } Image Processing Pipeline
 - regions/
 - 1.cwf, 2.cwf
 - gsRunProcessor.log
 - D_2008_06_04_14_00_10_dataRig10_signalProcessing } Signal Processing Pipeline
 - regions/
 - 1.cwf, 2.cwf
 - Output from gsReporter (.txt, csv, etc.)
 - sff/
 - FD392FSF01.sff, FD392FSF02.sff
 - gsRunProcessor.log
 - 454DataProcessingDir.xml
 - Output from gsReporter (.txt, .fna., .qual, etc)

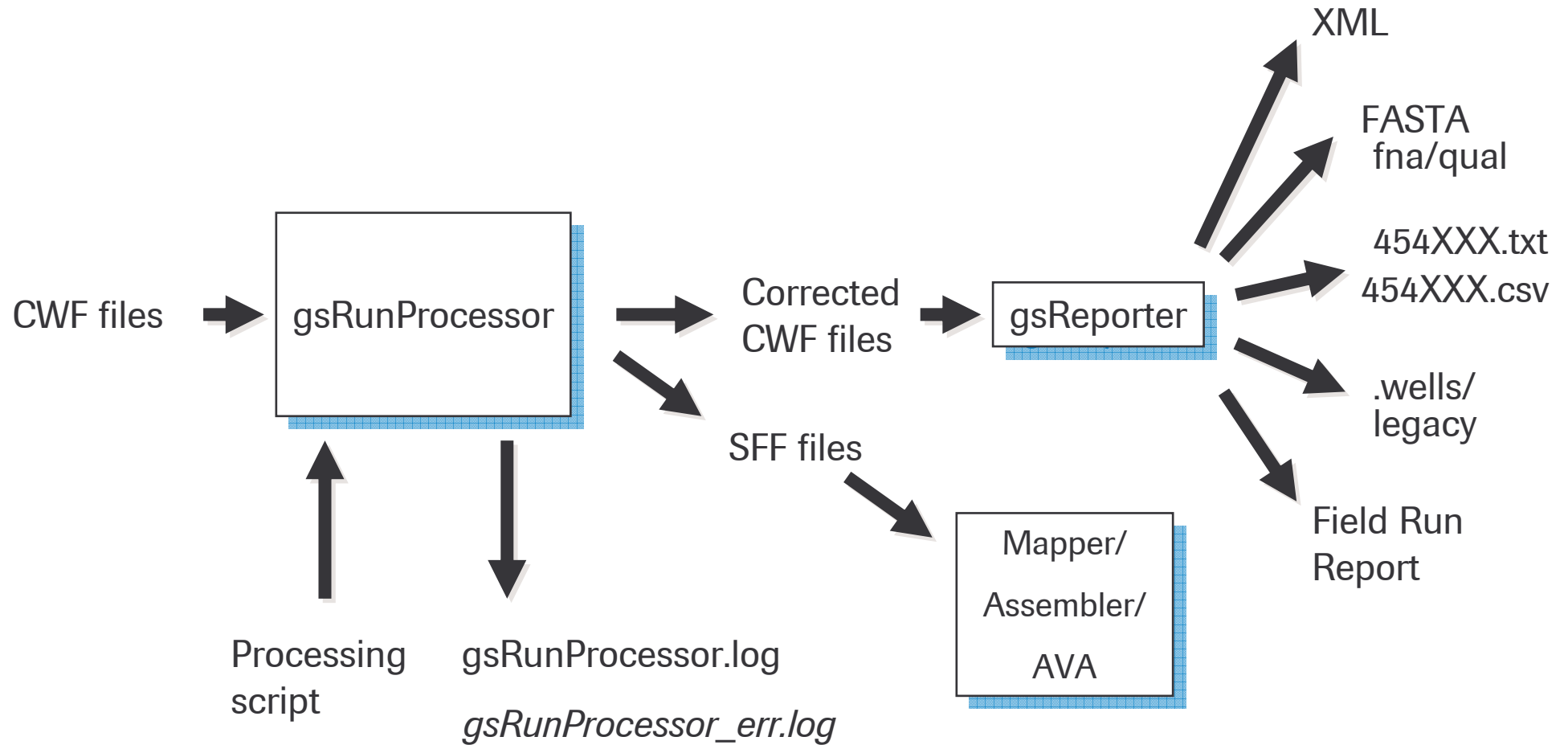
Composite Wells File (.cwf)

- “.wells” file replacement
- Self-contained
- “Package” format based on OpenOffice “.odt”
 - XML based meta-data
 - XML based metrics
 - Full processing history of the data
 - Called bases and quality scores
 - ... and the well flowgrams themselves
- Compressed
 - About half the size of a comparable .wells file for the same number of wells and flows

Data Flow For Image Processing



Data Flow For Signal Processing



Recommended Data Workflow

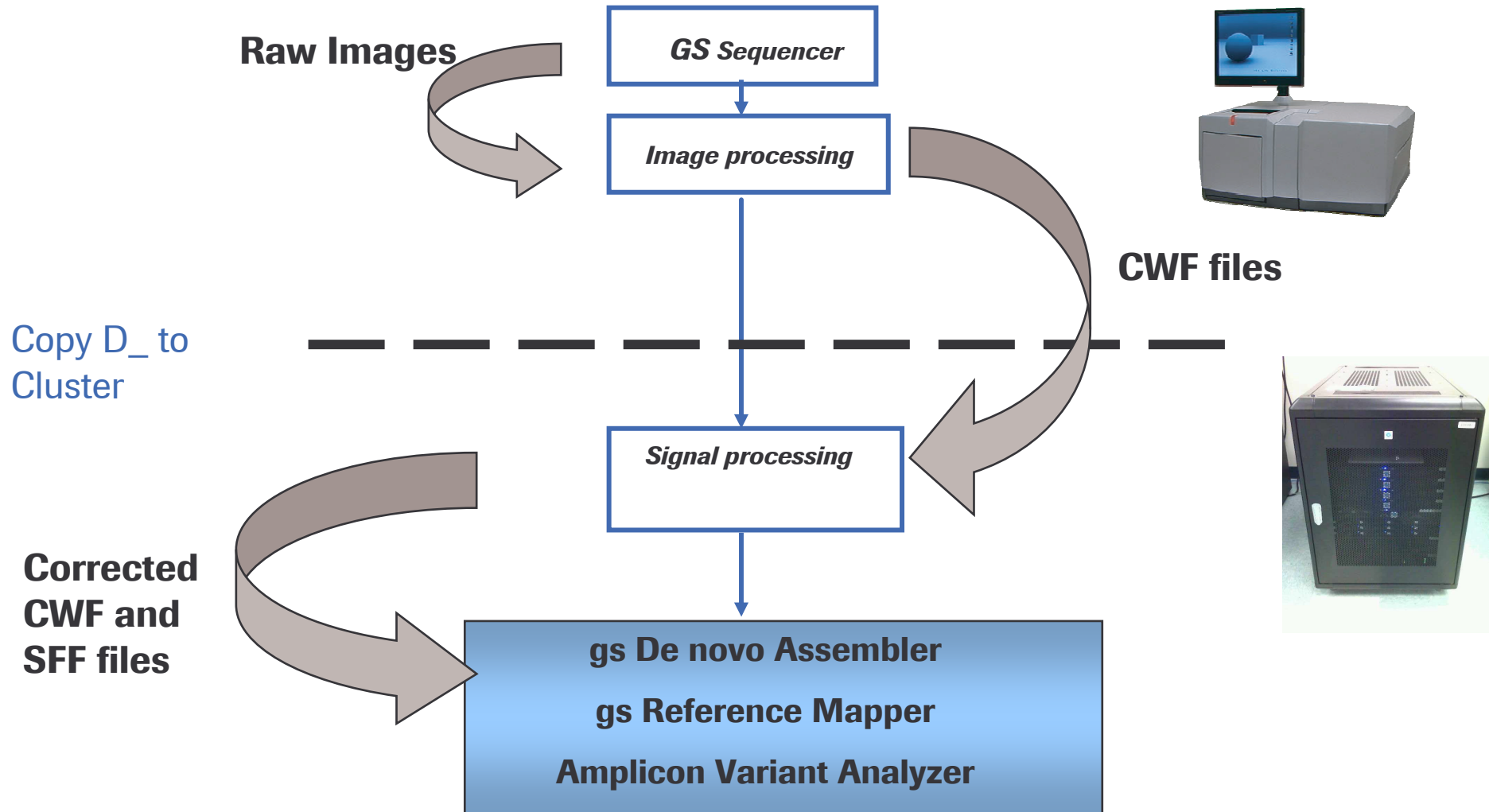
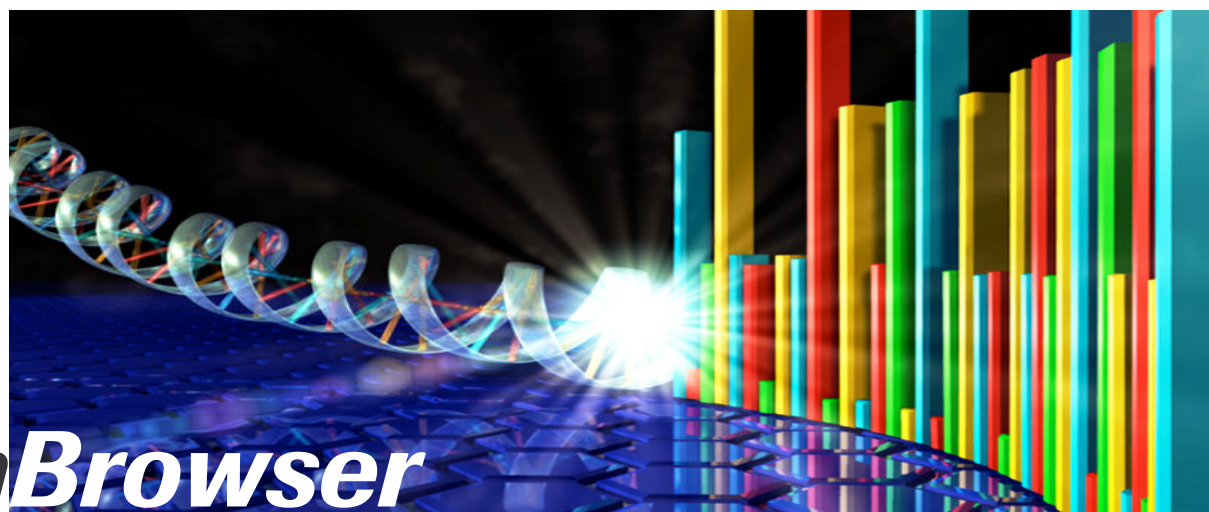


Image and Signal Processing Summary

GS Run Processor

Sequencing Kit	Number of Cycles	Raw Images Only	Raw Images Plus Image Processing	Raw Images Plus Full Processing
XLR70	200	28 Gb	30 Gb	37 Gb
XLR70	150	21 Gb	23 Gb	28 Gb
XLR70	100	14 Gb	16 Gb	18 Gb
LR70	100	14 Gb	15 Gb	16 Gb
LR70	42	5 Gb	6 Gb	7 Gb
LR25	100	6 Gb	7 Gb	8 Gb
LR25	42	3 Gb	4 Gb	4 Gb
SR70	42	5 Gb	6 Gb	7 Gb



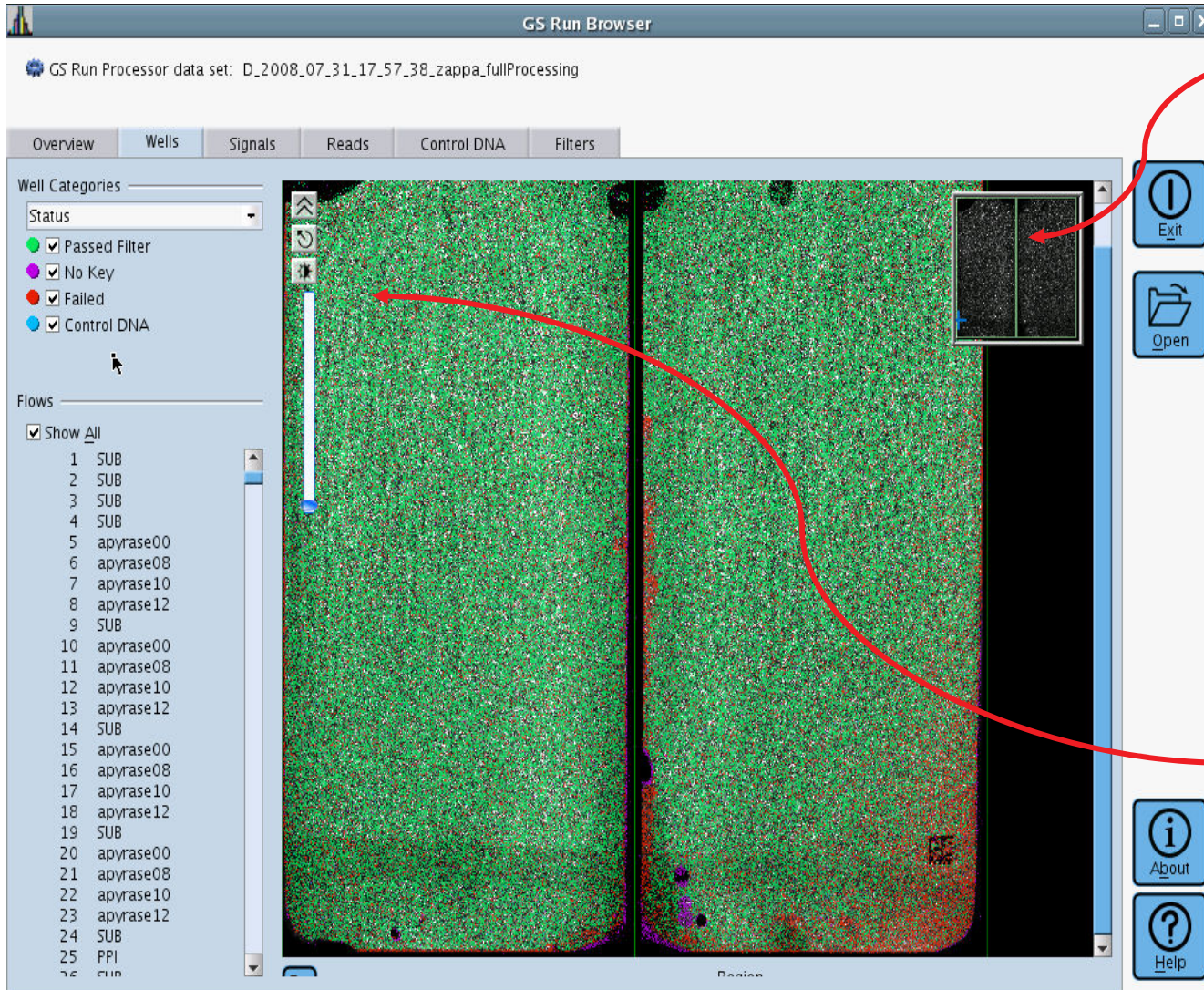
*gsRun*Browser



gsRunBrowser - Highlights

- Manage higher well counts
- Ability to browse new cwf data format
- Image and Wells Tab have been combined
- Mapping Tab has been removed
- Improve image and wells rendering and user interaction

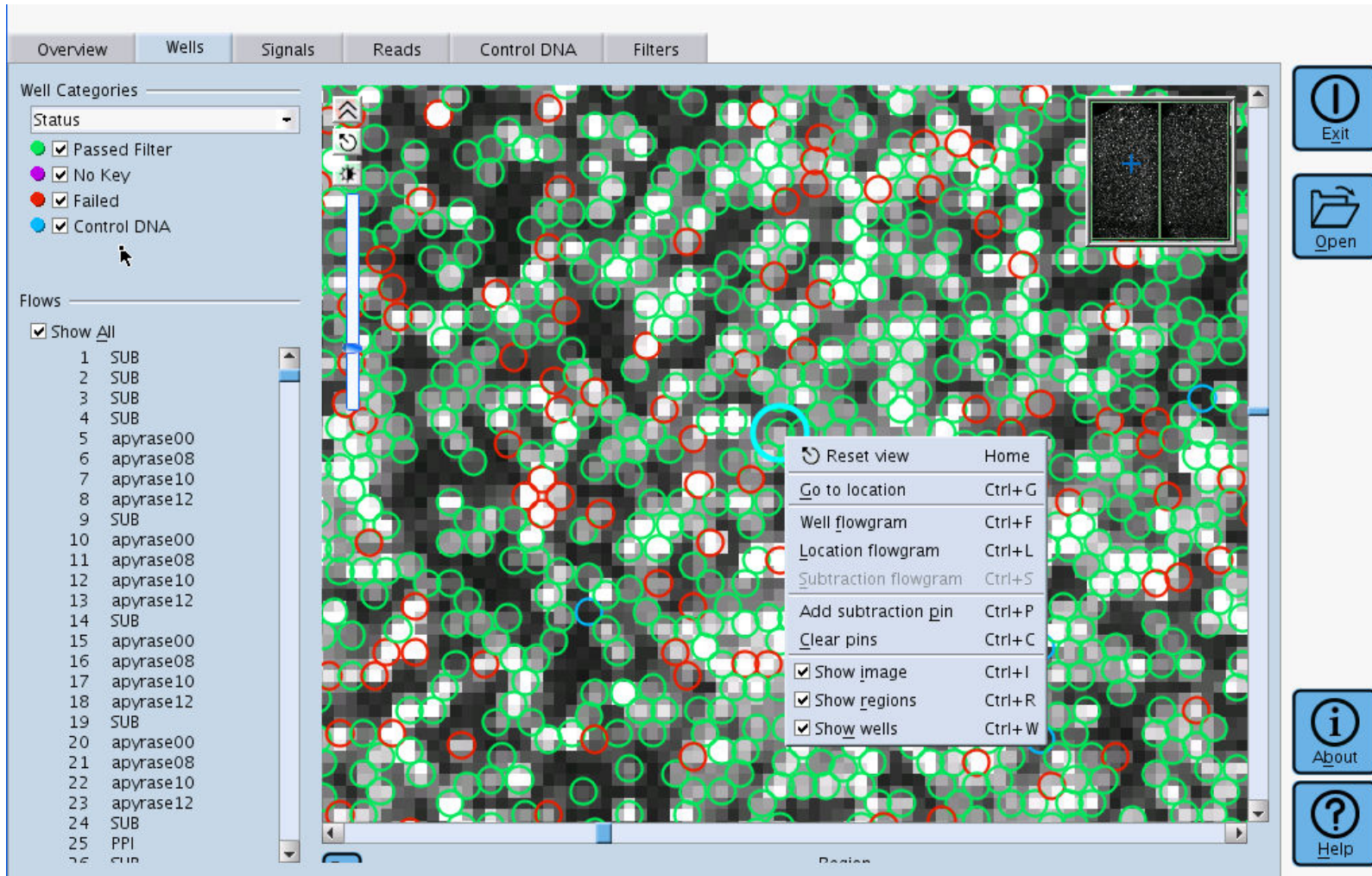
gsRunBrowser – Wells Tab



- Scaled out overview

- Collapse button
- Reset button
- Set contrast button
- Scaling slider

gsRunBrowser – Wells Tab



The screenshot displays the 'Wells' tab in the gsRunBrowser application. The main area is a grid of wells, each represented by a small square. Many wells are highlighted with colored circles: green for 'Passed Filter', red for 'Failed', and blue for 'Control DNA'. A context menu is open over a well, showing the following options and shortcuts:

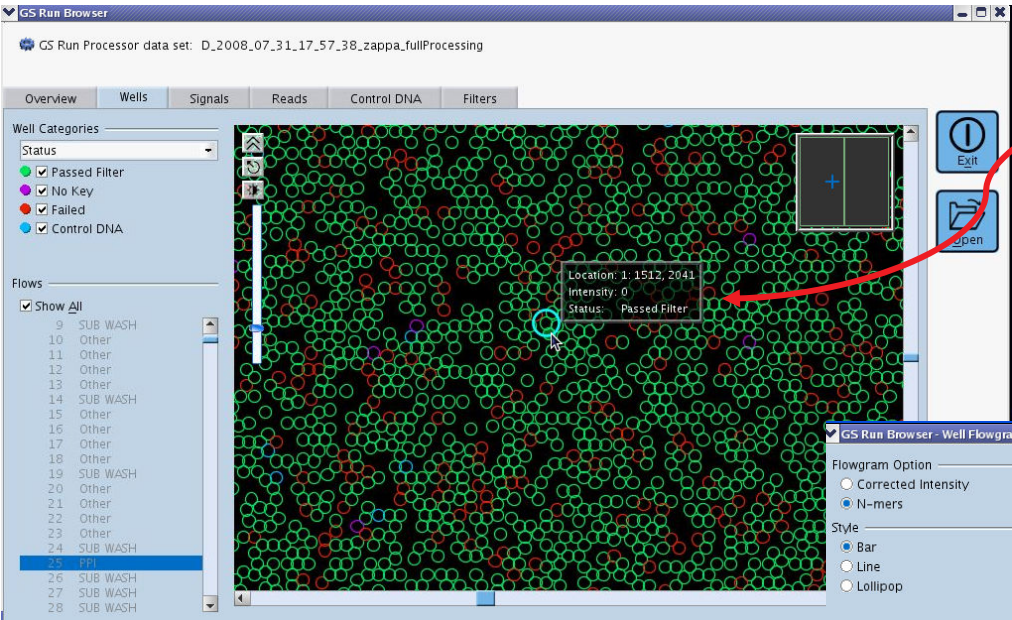
- Reset view (Home)
- Go to location (Ctrl+G)
- Well flowgram (Ctrl+F)
- Location flowgram (Ctrl+L)
- Subtraction flowgram (Ctrl+S)
- Add subtraction pin (Ctrl+P)
- Clear pins (Ctrl+C)
- Show image (Ctrl+I)
- Show regions (Ctrl+R)
- Show wells (Ctrl+W)

The left sidebar contains two sections: 'Well Categories' with a 'Status' dropdown and four checked items (Passed Filter, No Key, Failed, Control DNA), and 'Flows' with a 'Show All' checkbox and a list of 26 flow entries (1-26) with their respective flow names (SUB, apyrase00, apyrase08, apyrase10, apyrase12, PPI, SUB).

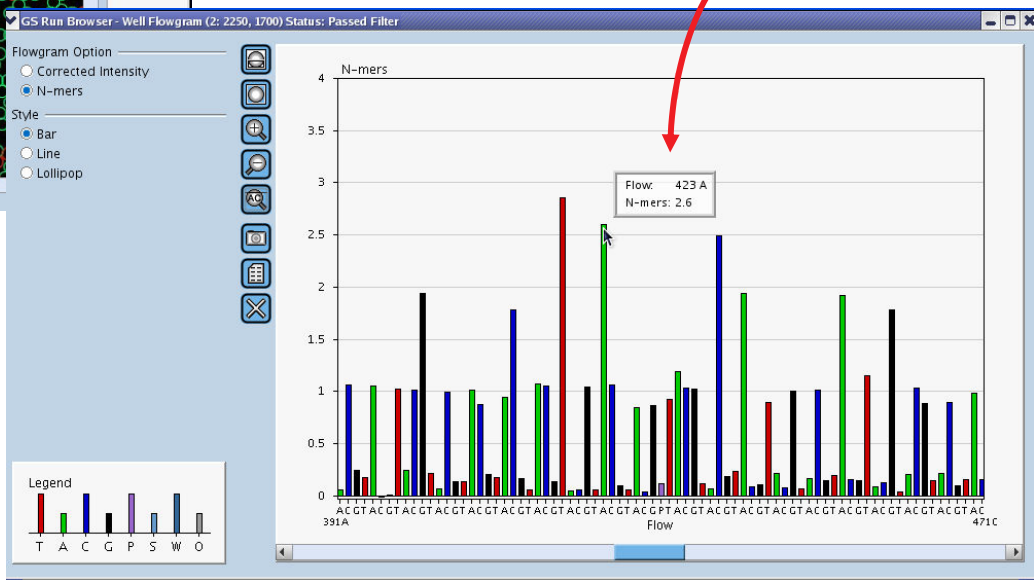
The top navigation bar includes 'Overview', 'Wells', 'Signals', 'Reads', 'Control DNA', and 'Filters'. On the right side, there are four buttons: 'Exit', 'Open', 'About', and 'Help'.

- New right click mouse options
- New Keyboard shortcuts

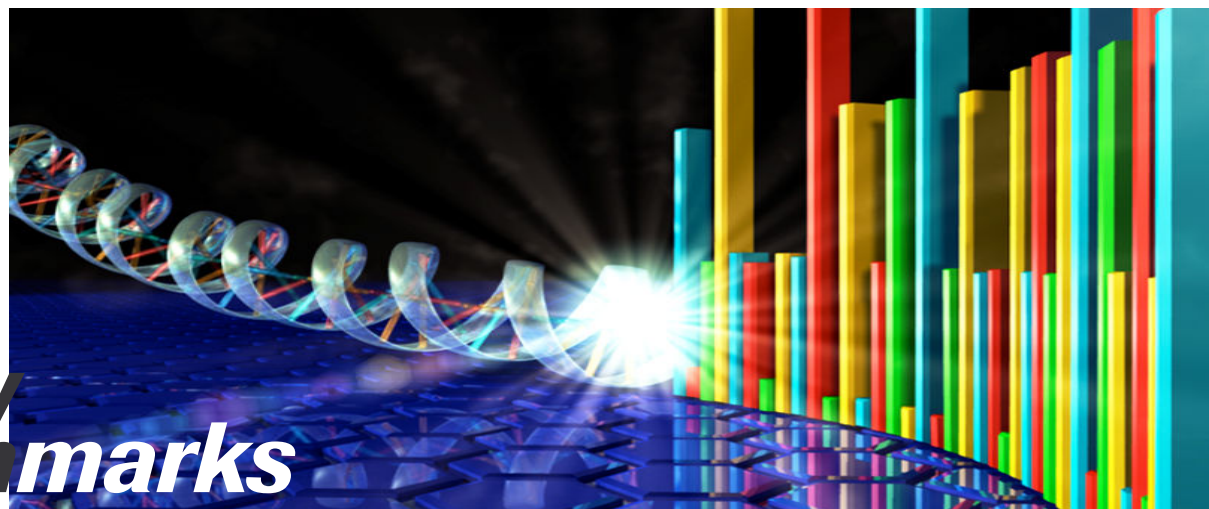
gsRunBrowser - Tooltips



- New Mouse over tooltips



New Benchmarks

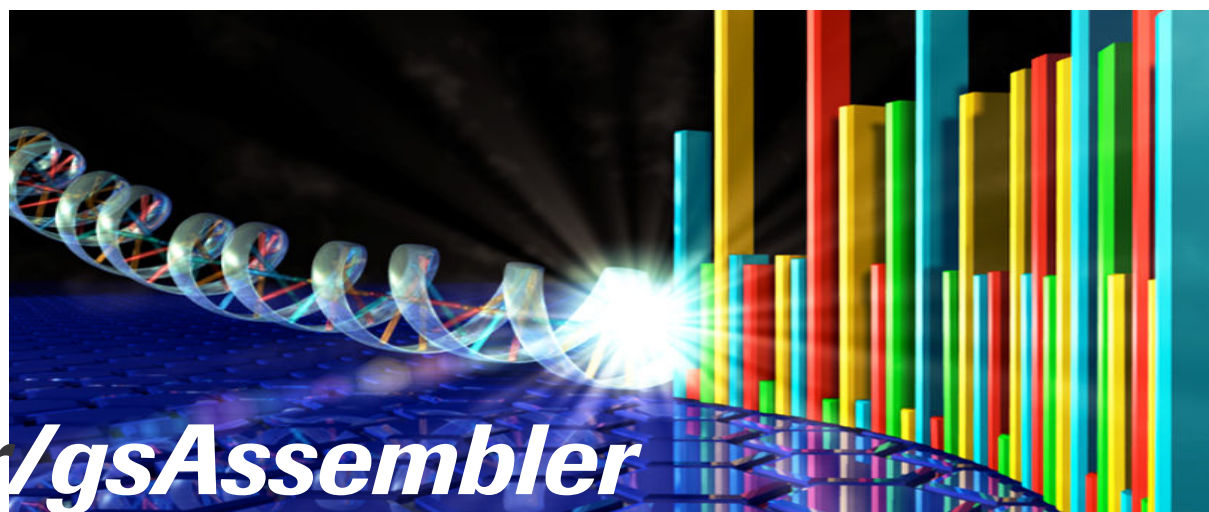


New Benchmarks

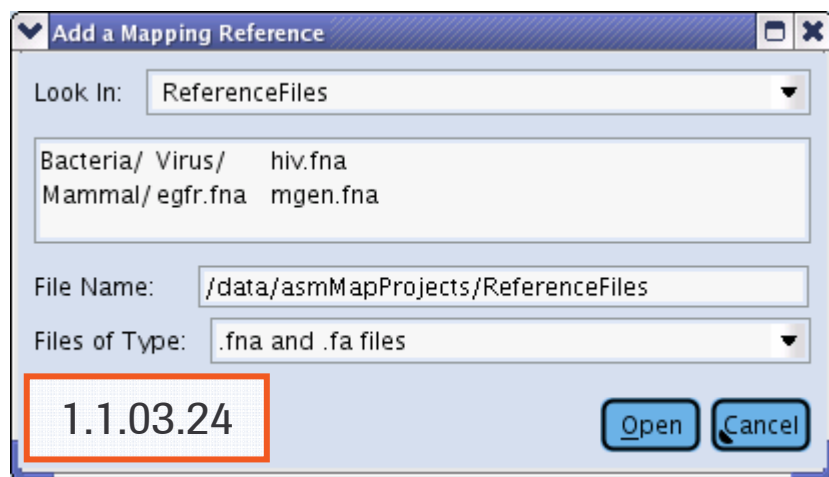
GS Run Processor and Metrics Files

Metrics File	Benchmark	Value
454QualityFilterMetrics.txt	totalKeyPass/totalRawWell	>90%
454QualityFilterMetrics.txt	totalPassFiltering/TotalKeyPass	>65%
454BaseCallerMetrics.txt	Average Read Length	>300bp
454BaseCallerMetrics.txt	Total Bases for 2 region PTP	400 million bases

gsMapper/gsAssembler

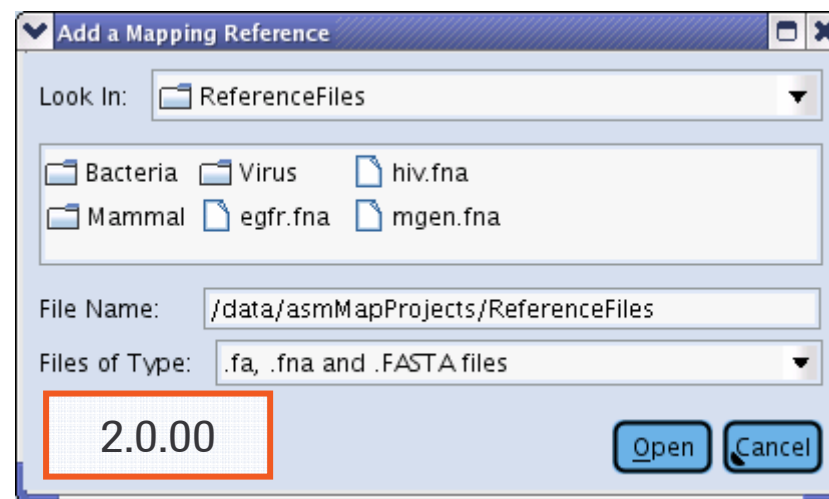


Icons Now Present in all File Dialogs



- All FileChooser dialogs have been augmented in 2.0.00 with folder and file icons to make them easier to read and use

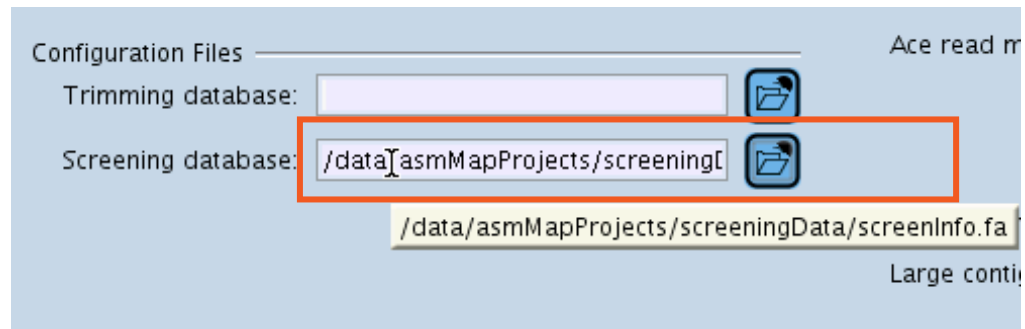
- 2.0.00 FileChooser consult historic, user-specific values for both gsMapper and gsAssembler to jump into the last opened directory



gsAssembler/gsMapper Parameter File Tooltips



- Tooltips have been added to display the full path if you hover over the text area.



gsAssembler/gsMapper Result File Tab



The screenshot shows a file browser on the left with the following files listed: 454AlignmentInfo.tsv, 454AllContigs.fna, 454AllContigs.qual, 454Contigs.ace (selected), 454LargeContigs.fna, 454LargeContigs.qual, 454NewblerMetrics.tx, 454NewblerProgress., 454ReadStatus.txt, and 454TrimStatus.txt.

The main pane displays the following information for 'File:/data/asmMapProjects/HIVAsmTest/assembly/454Contigs.ace':

- Lines: more than 50000
- Modified: Fri Apr 25 19:20:12 EDT 2008

The content of the file is displayed in a monospaced font, showing sequence alignment data for contigs AS, BQ, AF, BS, RD, QA, and DS. A callout box with a red border contains the text: "Files whose content is longer than 50,000 lines are so flagged and truncated. In 1.1.03.24 this cutoff was 10,000 lines."

gsAssembler/gSMapper Result File Tab



454AlignmentInfo.tsv
454AllContigs.fna
454AllContigs.qual
454Contigs.ace
454LargeContigs.fna
454LargeContigs.qual
454NewblerMetrics.tx
454NewblerProgress.
454ReadStatus.txt
454TrimStatus.txt

```
ACTCAAGACTTCTGGAAAGTTCAATTAGGAATACACATCCCGCAGGGTTA
AAAAAGAAAAAATCAGTAACAG

QA 1 48 1 48
DS CHROMAT_FILE: EES65NE01ARQOE.1-26 PHD_FILE: EES65NE01ARQOE.1-26.phd.1 TIME: Thu Jul 27 12:33:48

RD EES65NE01AL6RP.1-26 108 0 0
G*T*A*CC**A*G*TA*AAA**TT**AAA*G****CC*A*GG*AA**TGG
ATGGCCCAAAGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAGCAT
TAATGGAA

QA 1 48 1 48
DS CHROMAT_FILE: EES65NE01AL6RP.1-26 PHD_FILE: EES65NE01AL6RP.1-26.phd.1 TIME: Thu Jul 27 12:33:48

RD EES65NE01A01X6.1-26 319 0 0
G*T*A*CC**A*G*TA*AAA**TT**AAA*G****CC*A*GG*AA**TGG
ATGGCCCAAAGTGAACAATGGCCATTGACAGAAGAAAAAATAAAGCA
TTAATGGAAATTTGTACAGAAATGGAAGGAAAGGGAAATTTCAAAAAAT
TGGGCCTGAAAATCCATACAACACTCCAGTATTTGCCATAAAGAAGAAAA
ACAGTGATAGATGGAGAAAATTAGTAGATTTAGAGAACTTAATAAGAGA
ACTCAAGACTTTTGGGAAGTTCAATTAGGGATACACATCCTGCAGGGTT
AAAAAGAGAAATCAGTAA

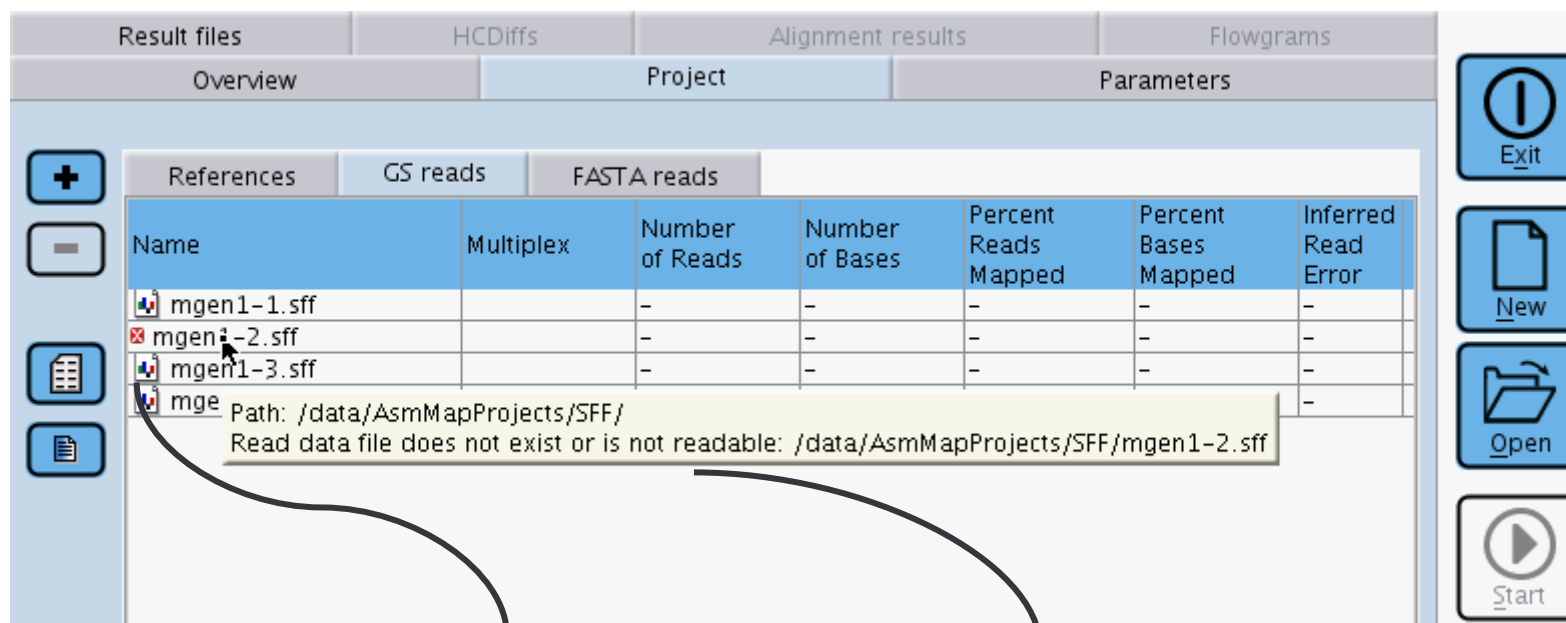
QA 1 48 1 48
DS CHROMAT_FILE: EES65NE01A01X6.1-26 PHD_FILE: EES65NE01A01X6.1-26.phd.1 TIME: Thu Jul 27 12:33:48


RD EES65NE01AJZEH.103-232 338 0 0
TGCCAGGAAGACGGAAACCAAAAAATAATAGGGGAATTGGAGTTTTGTC
AAAGTAAGAGAATATGATCAAATACCCATAGAAATCTGTGGACGTAAAGT
TGTA*CTACAG*TA*TTA*GT*A*G**GA**CC*T*AC*A*CC*TG*T*
CAA*CG*T*AA**TTGG**AA**G*AA**AT**C*T**AA**T***GACT
C*A*G*A**TT*GG**TT**GC*AC*TTT*AAA**TT**TT**CC**C*
ATT*A**GTCC**T*A*TT*GAA**AC*TG*T*A*CC**A*G*TA*AAA*
*TT**AAA*G****CC*A*GG*AA**TGGATGGCCCAA
```

File output view truncated at 50000 lines.
The entire file may be read from /data/asmMapProjects/HIVAsmTest/assembly/454Contigs.ace

In the case of truncated output the 2.0.00 software additionally reminds the user of the truncation at the bottom of the file view, and recapitulates the location on-disk where the full contents of the file can be found (this is a new feature, not present in 1.1.03.24)

Visible Validation of Project Input Files



Name	Multiplex	Number of Reads	Number of Bases	Percent Reads Mapped	Percent Bases Mapped	Inferred Read Error
mgen1-1.sff		-	-	-	-	-
 mgen1-2.sff		-	-	-	-	-
mgen1-3.sff		-	-	-	-	-
mge		-	-	-	-	-

Path: /data/AsmMapProjects/SFF/
Read data file does not exist or is not readable: /data/AsmMapProjects/SFF/mgen1-2.sff

- Red “x” appears adjacent to Reference (Mapper), GS read or FASTA read (Mapper + Assembler) files that fail validation

- Mousing over the red “x” or anywhere in the cell with the red “x” provides a tooltip explaining the problem.

Vector Trimming/Screening Database



Project _____

Incremental reference mapper analysis
 Nimblegen sequence capture

Automate trimming

Define these parameters, add the read data files and add the reference files. Once complete, click Start to begin the process.

```
runMapping -vt TrimmingDB.fasta -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

Overlap Detection

Seed step: include consensus

Seed length: Pairwise alignment: None

Seed count: Simple format

```
runAssembly -vt TrimmingDB.fasta -read YourReads.sff
```

HT-per-seed limit: Ace/Consed: No files

HT location limit: Single ACE file for small genomes

Minimum overlap length:* Single ACE file

Minimum overlap identity:* ACE file per reference

```
runMapping -vs ScreeningDB.fasta -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

Alignment difference score:* Ace/Consed: Deleted

Repeat score threshold:* Raw

Trimmed

Single read variant

Configuration Files

Trimming database:

```
runAssembly -vs ScreeningDB.fasta YourReads.sff
```

Screening database:

Targeted Regions:

Genome annotation:

Known SNP:

Large contig threshold: Minimum contigs depth:

- 454 reads are trimmed at the ends by the trimming
- 454 reads are removed by the screening database

Expected Depth



Project _____

Incremental reference mapper analysis
 Nimblegen sequence capture
 Automatic trimming

Expected depth:*

Overlap Detection _____

Seed step:*
Seed length:*
Seed count:
Min location limit:
Minimum overlap length:*
Minimum overlap identity:*
Alignment identity score:*
Alignment difference score:*
Repeat score threshold:*

Output _____

Include consensus

Pairwise alignment: None
 Simple format
 Tabular
 No files
 Single ACE file for small genomes
 Single ACE file
 ACE file per reference
 Complete consensus folder
 Ace file mode
 Default
 Raw
 Trimmed

Configuration Files _____

Trimming database:

Screening database:

Targeted regions:

Genome annotation:

Known SNP:

All contig threshold:*
Large contig threshold:*
Minimum contigs depth:*
 Single read variant

Define these parameters, add the read data files and add the references files. Once complete, click 'Start' to begin the Reference Mapper analysis.

- Algorithm adjustment to filter out random-chance events at higher coverage (for example 150 fold coverage)

```
runMapping -e 150 -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

```
runAssembly -e 150 YourReads.sff
```

Seed Count



- Number of seeds required in a window before an extension is made

Project

Incremental reference mapper analysis
 Nimblegen sequence capture
 Automatic trimming

Expected depth:* 150

Overlap Detection

Seed step:* 12
Seed length:* 16
Seed count: 1

Minimum overlap length:* 40
Minimum overlap identity:* 90 %
Alignment identity score:* 2
Alignment difference score:* -3
Repeat score threshold:* 12

Configuration Files

Trimming database:
Screening database:
Targeted regions:
Genome annotation:
Known SNP:

Output

Include consensus

Pairwise alignment:

None
 Simple format
 Tabular format
 No files
 Single ACE file for small genomes
 Single ACE file
 ACE file per reference
 Complete contig folder
 Default
 Raw
 Trimmed

Single read variant

All contig threshold:* 100
Large contig threshold:* 500
Minimum contigs depth:* 1

Define these parameters, add the read data files and add the references files. Once complete, click 'Start' to begin the Reference Mapper analysis.

```
runMapping -sc 1 -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

```
runAssembly -sc 1 YourReads.sff
```

Ace File Read Mode



The screenshot displays the Roche software interface with various settings and command-line examples. The interface is divided into several sections:

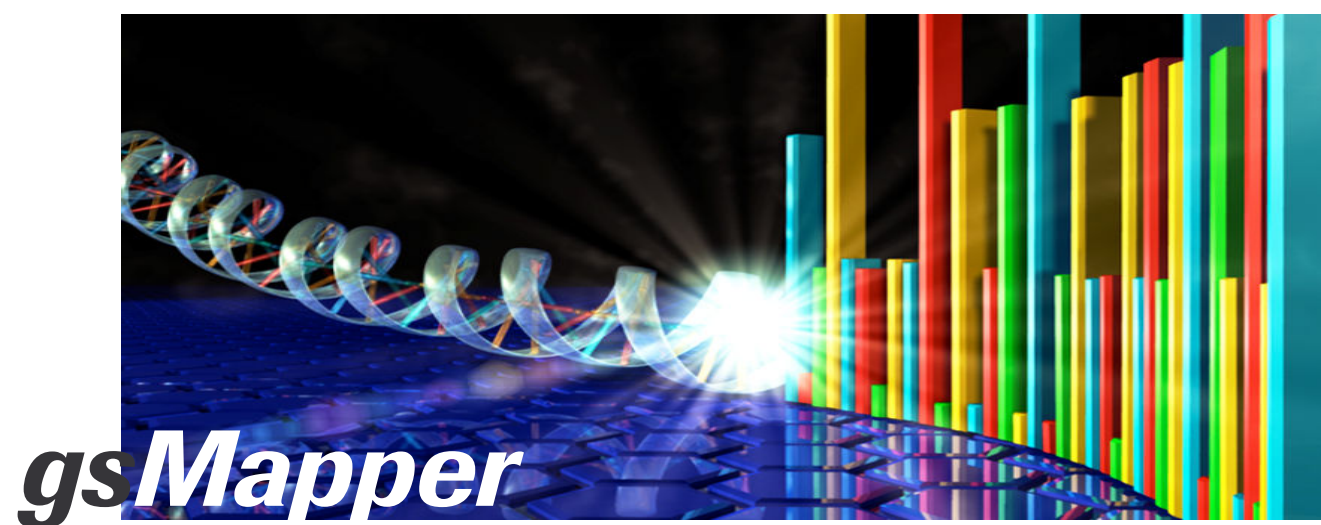
- Project:** Includes checkboxes for "Incremental reference mapping analysis" and "Automatic trimming". A text box contains the command: `runMapping -ad -ref MultipleReferenceFiles.fasta -read YourReads.sff`.
- Overlap Detection:** Includes fields for "Seed step" (12), "Seed length" (16), and "Seed count" (1). A text box contains the command: `runAssembly -ad YourReads.sff`.
- Output:** Includes a "Pairwise alignment" dropdown set to "None" and radio buttons for "Simple format", "Tabbed format", "Single ACE file for small genomes", "Single ACE file", and "ACE file per reference". A text box contains the command: `runMapping -ar -ref MultipleReferenceFiles.fasta -read YourReads.sff`.
- Configuration Files:** Includes fields for "Minimum overlap length" (40), "Minimum overlap identity" (90), "Alignment identity score" (2), "Alignment difference score" (-3), and "Repeat score threshold" (12). A text box contains the command: `runAssembly -ar YourReads.sff`.
- Configuration Files (continued):** Includes a "Trimming database" dropdown, "Screening database" dropdown, "Targeted regions" dropdown, "Genome annotation" dropdown, and "Known SNP" dropdown. A text box contains the command: `runMapping -at -ref MultipleReferenceFiles.fasta -read YourReads.sff`.
- Configuration Files (continued):** Includes fields for "Large contig threshold" (500) and "Minimum contigs depth" (4). A text box contains the command: `runAssembly -at YourReads.sff`.

Annotations on the right side of the screenshot:

- A red box highlights the "Ace read mode" dropdown menu, which is currently set to "Default".
- A bracket on the right side of the "Ace read mode" dropdown points to the text: "• Ace file output set to default (raw)".
- A bracket on the right side of the "Screening database" dropdown points to the text: "• Ace file output set to raw".
- A bracket on the right side of the "Targeted regions" dropdown points to the text: "• Ace file output set to trimmed".

Additional Output Highlights

- Automatic quality score based trimming of 454 reads (`-trim -notrim`)
- `454TrimStatus.txt` reports the trim status of each read
 - “TooShort” is added as an additional read status (trimmed shorter than 50 bases)
- `454PairStatus.txt` is created whenever paired-end reads are part of the dataset
 - One line summary for each pair
- Each paired-end file is now treated as a distinct library in order to compute mean distances separately



Mapper – Nimblegen sequence capture



Project _____

Incremental reference mapper analysis
 Nimblegen sequence capture
 Automatic trimming

Expected depth:*

Overlap Detection _____

Seed step:*
Seed length:*
Seed count:*

Output _____

Include consensus

Pairwise alignment: None
 Simple format

```
runMapping -n -reg targetedRegionsFile -annot GenomeAnnotationFile -snp  
KnownSNPFile -ref MultipleReferenceFiles.fasta -read YourReads.sff
```

Minimum overlap identity:* %
Alignment identity score:*
Alignment difference score:*
Repeat score threshold:*

Ace read mode: Default
 Raw
 Trimmed
 Single read variant

Configuration Files _____

Trimming database:

Screening database:

Targeted regions:

Genome annotation:

Known SNP:

All contig threshold:*
Large contig threshold:*
Minimum contigs depth:*

- Nimblegen sequence capture primer is trimmed

- The targeted regions file from Nimblegen
- The Genome annotation file (UCSC GoldenPath)
- The dbSNP file (UCSC GoldenPath)

gsMapper - New HCDiffs Tab



GS Reference Mapper

Project: aamap Ready for analysis

Overview	Project	Parameters	Result files	HCDiffs	Alignment results	Flowgrams											
Reference Accession Number	Start Position in Ref	End Position in Ref	Reference Bases	Variation Bases	Total Percent	Total Depth	Reference Amino Acids	Variation Amino Acids	Coding Frame	Region Name	Known SNP Info	Percent Forward	Percent Reverse	Num Forward Reads	Num Reverse Reads With Variation	Total Num Forward Reads	Total Num Reverse Reads
gii16271976 ref ...	1633791	1633792	AG	-	15.0	20						18.2	11.1	2	1	11	9
gii16271976 ref ...	1633794	1633796	CAA	-	15.0	20						9.1	22.2	1	2	11	9
gii16271976 ref ...	327550	327550	A	G	15.4	13						0.0	25.0	0	2	5	8
gii16271976 ref ...	68949	68950	CA	-	16.7	12						28.6	0.0	2	0	7	5
gii16271976 ref ...	706739	706739	-	A	16.7	12						33.3	0.0	2	0	6	6
gii16271976 ref ...	1475520	1475520	-	A	16.7	12						0.0	28.6	0	2	5	7
gii16271976 ref ...	1089482	1089482	-	T	18.2	11						0.0	25.0	0	2	3	8
gii16271976 ref ...	1633910	1633910	-	CA	18.2	11						0.0	22.2	0	2	2	9
gii16271976 ref ...	384197	384197	A	-	18.8	16						0.0	33.3	0	3	7	9
gii16271976 ref ...	420753	420753	-	A	20.0	10						0.0	33.3	0	2	4	6
gii16271976 ref ...	836345	836345	-	T	20.0	10						40.0	0.0	2	0	5	5
gii16271976 ref ...	207408	207408	-	T	21.4	14						0.0	25.0	0	3	2	12
gii16271976 ref ...	177845	177845	-	T	22.2	9						28.6	0.0	2	0	7	2
gii16271976 ref ...	1276329	1276329	G	AGT	22.2	9						33.3	0.0	2	0	6	3
gii16271976 ref ...	1358399	1358400	CG	-	22.2	9						0.0	100.0	0	2	7	2
gii16271976 ref ...	1633712	1633712	C	T	23.1	13						25.0	20.0	2	1	8	5
gii16271976 ref ...	1522093	1522093	A	-	27.3	11						50.0	0.0	3	0	6	5
gii16271976 ref ...	1608609	1608618	AAATGC...	TCAAT...	27.3	11						28.6	25.0	2	1	7	4
gii16271976 ref ...	173478	173478	A	G	28.6	7						0.0	66.7	0	2	4	3
gii16271976 ref ...	219116	219116	-	A	28.6	7						50.0	0.0	2	0	4	3
gii16271976 ref ...	1218513	1218513	-	T	30.0	10						42.9	0.0	3	0	7	3
gii16271976 ref ...	1715765	1715765	-	T	30.0	10						37.5	0.0	3	0	8	2
gii16271976 ref ...	1633794	1633794	-	TCAAC...	30.0	20						27.3	33.3	3	3	11	9
gii16271976 ref ...	508004	508004	-	T	33.3	6						0.0	66.7	0	2	3	3
gii16271976 ref ...	375636	375636	-	A	33.3	9						16.7	66.7	1	2	6	3
gii16271976 ref ...	517873	517878	CTCCGA	TC	33.3	9						25.0	40.0	1	2	4	5
gii16271976 ref ...	1071476	1071476	G	-	33.3	9						66.7	16.7	2	1	3	6
gii16271976 ref ...	1392842	1392842	-	T	33.3	9						33.3	33.3	2	1	6	3
gii16271976 ref ...	1633901	1633901	A	T	33.3	12						50.0	30.0	1	3	2	10
gii16271976 ref ...	1633789	1633790	TG	-	35.0	20						45.5	22.2	5	2	11	9
gii16271976 ref ...	327831	327831	G	A	35.7	28						27.3	41.2	3	7	11	17
gii16271976 ref ...	917015	917015	T	-	36.4	11						80.0	0.0	4	0	5	6
gii16271976 ref ...	1369478	1369482	TCGTC	-	37.5	8						33.3	50.0	2	1	6	2
gii16271976 ref ...	1369496	1369496	T	-	37.5	8						33.3	50.0	2	1	6	2
gii16271976 ref ...	1369499	1369501	CGC	-	37.5	8						33.3	50.0	2	1	6	2
gii16271976 ref ...	838375	838375	-	A	40.0	5						0.0	100.0	0	2	3	2
gii16271976 ref ...	964470	964471	AC	TA	40.0	5						0.0	100.0	0	2	3	2
gii16271976 ref ...	1497105	1497105	-	T	40.0	5						0.0	50.0	0	2	1	4
gii16271976 ref ...	754285	754285	-	A	40.0	10						50.0	33.3	2	2	4	6
gii16271976 ref ...	1633775	1633775	A	G	41.2	17						55.6	25.0	5	2	9	8
gii16271976 ref ...	421042	421042	T	ATTA	41.7	12						62.5	0.0	5	0	8	4
gii16271976 ref ...	853546	853547	AT	-	50.0	4						0.0	66.7	0	2	1	2

Exit

New

Open

Start

Stop

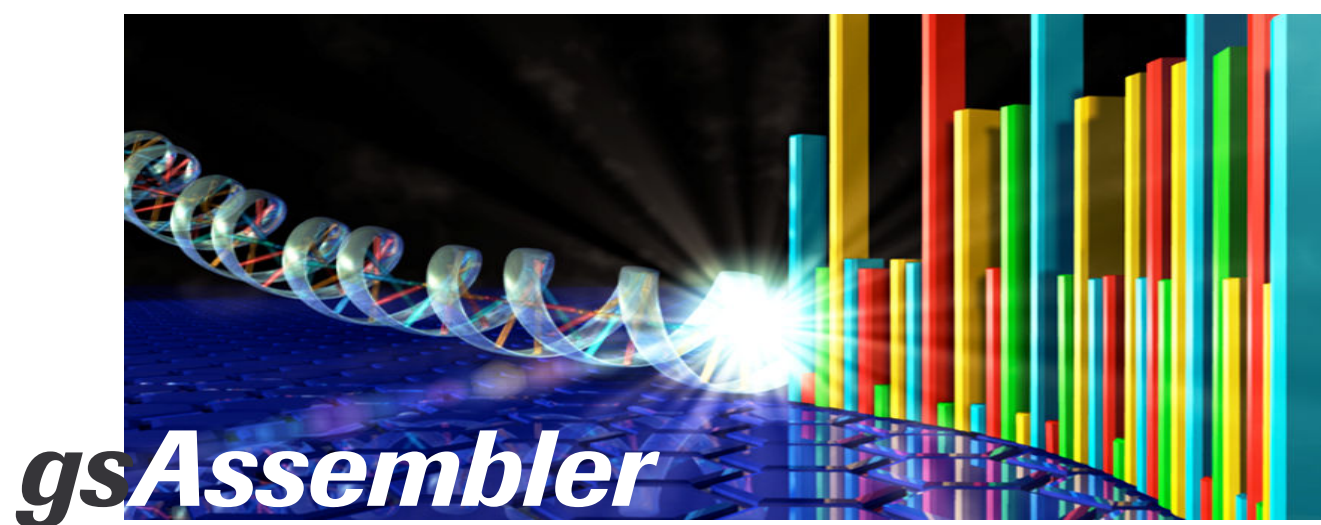
About

Help

gsMapper - HCDiffs Tab Capabilities



- Parsed Contents of HCDiffs file with live link to Alignment Results
 - Right-click on any row and select the “Show Alignments” menu item to go directly to the given HCDiff variation in the “Alignment results” Tab view
 - The first base of the variation will appear leftmost in the view.
- Sortable Table
 - Click on header to sort by contents of column.
 - Click again on same header to reverse the sort (triangular indicator specifies which column was last used for sort operation and the ascending/descending direction of the sort)
 - Sort is “stable”: Subsequent sorts that have tied values will not re-order previously sorted rows, allowing the grouping of data via nested sorting operations by sorting different columns in succession.



Assembler – Large option



Project _____

Incremental de novo assembler analysis

Large or complex genome

Expected depth:*

Overlap Detection _____

Seed step:*

Seed length:*

Seed contig*

Minimum overlap length*

Minimum overlap identity*

Alignment identity score:*

Alignment difference score:*

Configuration Files _____

Trimming database:

Screening database:

Output _____

Include consensus

Pairwise alignment: None

Simple format

Ace/Consed

No files

Single ACE file for small genomes

Single ACE file

ACE file per contig

Complete consed folder

Ace read mode: Default

Raw

Trimmed

All contig threshold:*

Large contig threshold:*

Define these parameters and add the read data files. Once complete, click 'Start' to begin the De novo Assembly analysis.

- Algorithm adjustment to assemble large and/or complex genomes

```
runAssembly -large YourReads.sff
```

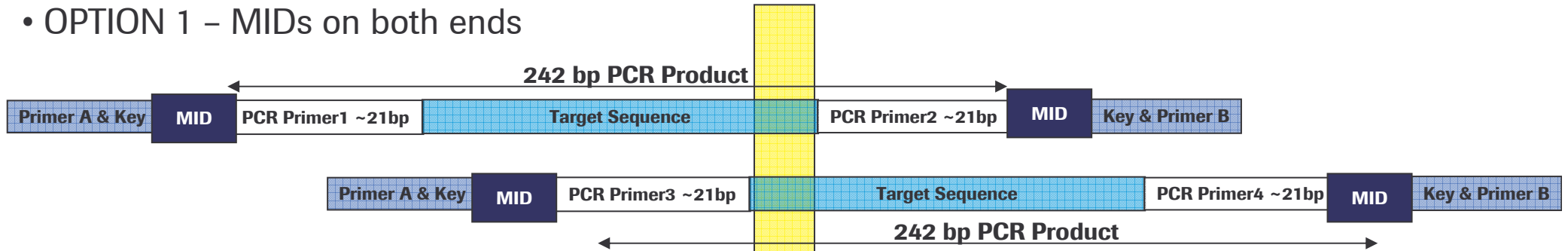


Amplicon Variant Analyzer

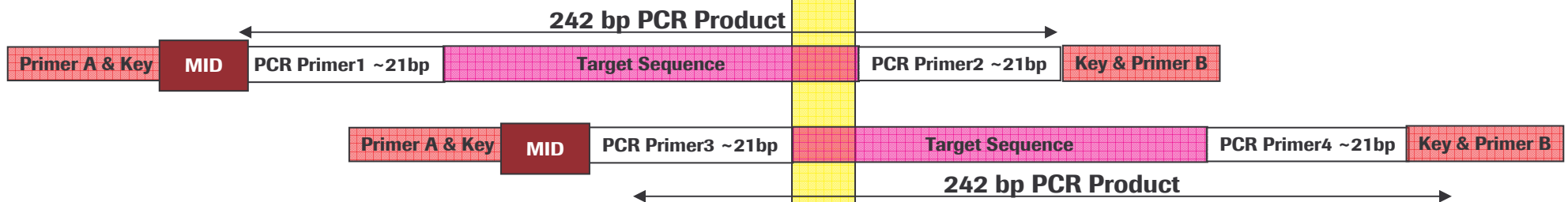
PCR product design options using MIDs



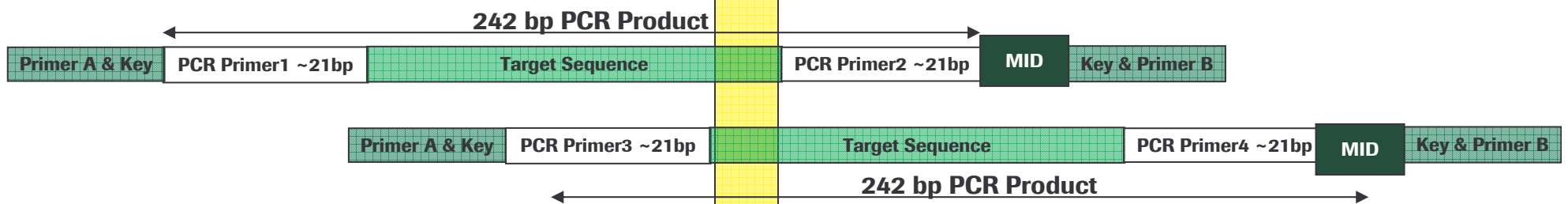
- OPTION 1 – MIDs on both ends



- OPTION 2 – MID on 5' end



- OPTION 3 – MID on 3' end



20 bp overlap
between products

Project Tab: Dragging Amplicons To Read Data Tree

GS Amplicon Variant Analyzer <3>

Project Name: EGFR_MSB
Location: /data/ampProjects/EGFR_MSB

Overview **Project** Computations Variants Global

References **Read Data**

EGFR_MSB

- ReadGrp_1
 - DGV590J01
 - Sample1
 - EGFR_20_1
 - EGFR_20_2
 - EGFR_20_3
 - DGV590J02
 - DGV590J03
 - Sample3
 - Sample4
 - EGFR_19_1
 - EGFR_19_2
 - Sample5
 - Sample6
 - Sample7
 - DGV590J04

Amplicons (12)												Read Data (<>)	
Name	...	A...	Pri...	Pri...	St...	End							
EGFR_18_1	EGF...	A...	GA...	CC...	23	66							
EGFR_18_2	EGF...	A...	AG...	CC...	60	1...							
EGFR_18_3	EGF...	A...	TG...	CC...	1...	1...							
EGFR_19_1	EGF...	A...	TC...	GA...	23	1...							
EGFR_19_2	EGF...	A...	TC...	GA...	67	1...							
EGFR_19_...	EGF...	A...	TC...	GA...	67	1...							
EGFR_20_1	EGF...	A...	CC...	GC...	1	1...							
EGFR_20_2	EGF...	A...	GC...	GC...	1...	1...							
EGFR_20_3	EGF...	A...	GG...	GA...	1...	2...							
EGFR_21_1	EGF...	A...	TC...	GA...	23	1...							
EGFR_21_2	EGF...	A...	GG...	AT...	1...	2...							
EGFR_22_1	EGF...	A...	CA...	CC...	21	1...							

Exit
New
Open
Save
About
Help

- Amplicons may now be dragged to the Project (root) node, any Read Data Group, or any Read Data node

- The drop will only be accepted if there is at least one Read Data below the level of the drag that has at most one Sample or Multiplexer currently associated with it

Multiplexing Features – New Tree & Tables

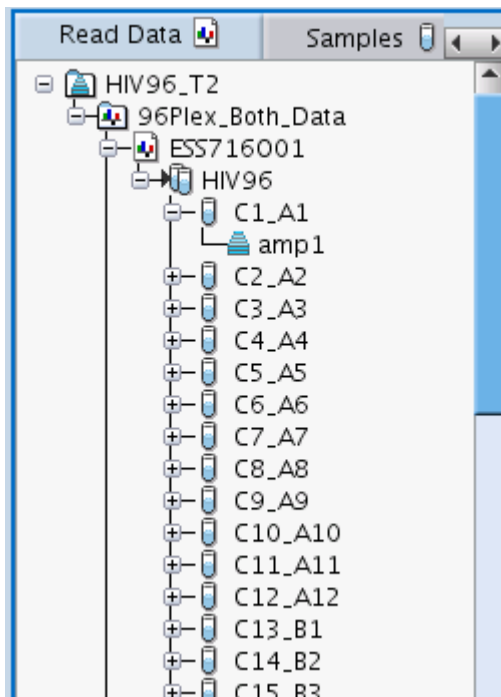


The screenshot shows the 'Project' tab in the GS Amplicon Variant Analyzer. On the left, a tree view shows the project structure: 'HIV96_T2' containing a folder '45 4Standard' which lists 'Mid1' through 'Mid14'. The 'MIDs' folder is highlighted with a red box. On the right, a table titled 'MIDs (14)' is displayed, listing details for each MID. The table has columns for Name, Annotation, Sequence, and Group. The 'Mid3' row is highlighted in blue. A red box highlights the table header and the 'MIDs (14)' title. A red box also highlights the 'Multiplexers (1)' tab, which is currently empty. Arrows point from the text on the right to these red boxes.

Name	Annotation	Sequence	Group
Mid1		ACGAGTGCGT	45 4Standard
Mid2		ACGCTCGACA	45 4Standard
Mid3		AGACGCACTC	45 4Standard
Mid4		AGCACTGTAG	45 4Standard
Mid5		ATCAGACACG	45 4Standard
Mid6		ATATCGCGAG	45 4Standard
Mid7		CGTGTCTCTA	45 4Standard
Mid8		CTCGCGTGTC	45 4Standard
Mid9		TAGTATCAGC	45 4Standard
Mid10		TCTCTATGCG	45 4Standard
Mid11		TGATACGTCT	45 4Standard
Mid12		TACTGAGCTA	45 4Standard
Mid13		CATAGTAGTG	45 4Standard
Mid14		CGAGAGATAC	45 4Standard

- “MIDs” Tree enumerating the MIDs of the project in relationship to their (optional) MID Group
- The Table section contains two new tables for defining the MIDs and Multiplexers of the Project

Multiplexing Features – Tree Nodes



- For Read Data sets that contain MID-based multiplexed data, a new node representing the Multiplexer will appear in the Read Data Tree of the Project Tab:

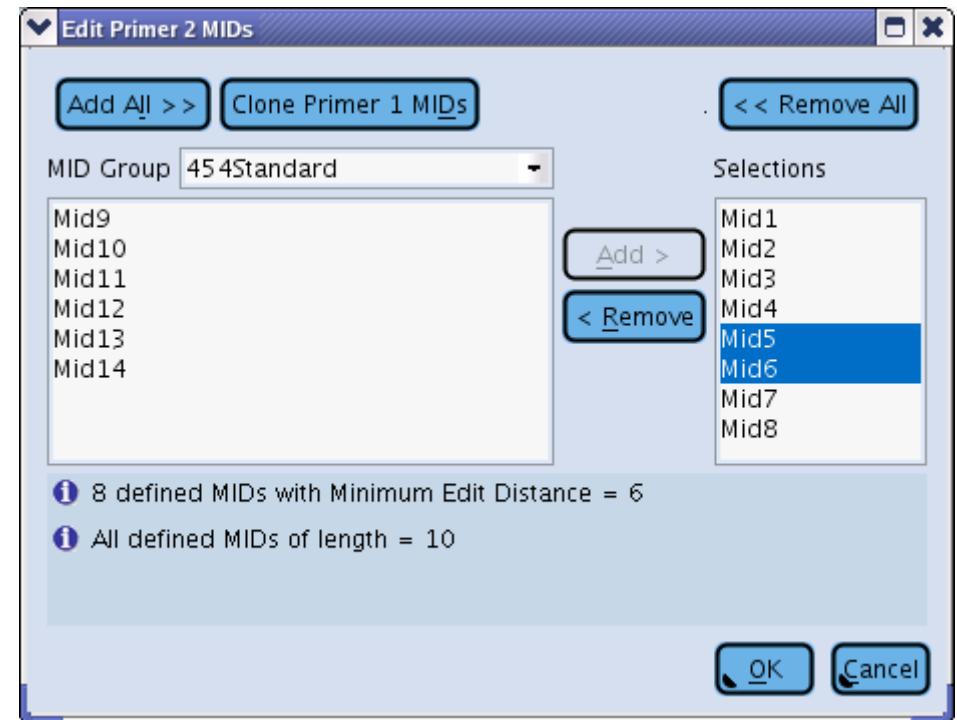
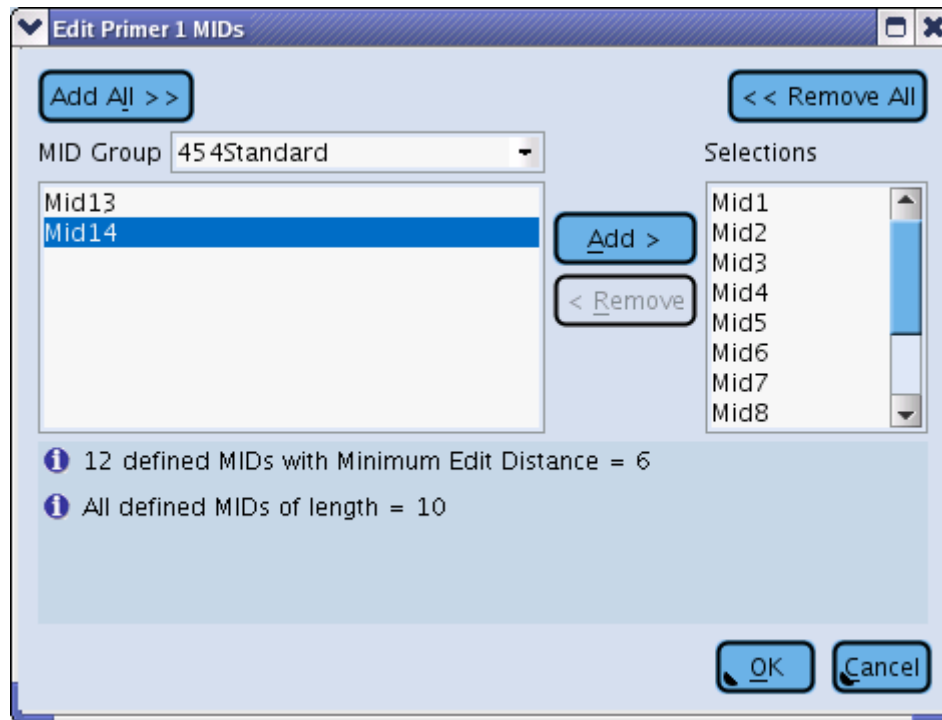


- Under this Multiplexer node will be the Samples encoded by MIDs of that Multiplexer, with the associated Amplicons appearing below those Samples.

Multiplexing Features – MID Association Editors



- By double clicking on the Primer1 MID and Primer2 MID columns one can bring up editors for associating previously defined MIDs with the Primer1 and Primer2 positions of the Multiplexer:



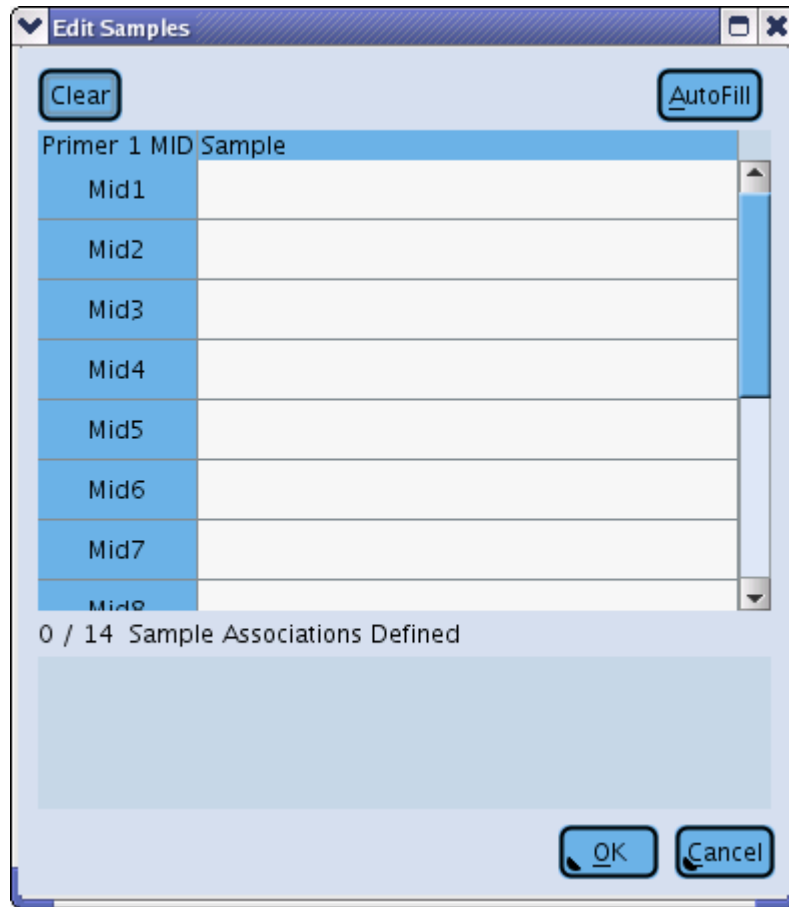
- These editors provide informative status messages regarding the mutual compatibility of the selected MIDs for use during multiplexing.

Multiplexing Features – Sample Editors

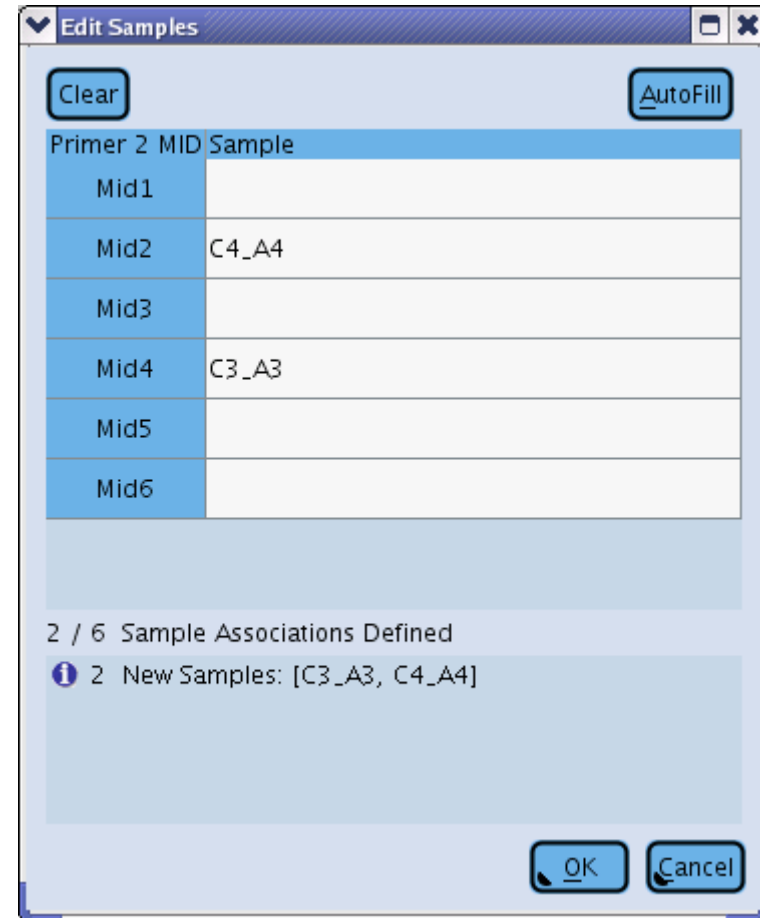


- By double clicking on the Samples column of the Multiplexer table one can define the relationship between combinations of Primer1 and Primer2 MIDs and Samples.
- Different editors are available depending on the type of Multiplexer Encoding, which is assignable via a dropdown editor on the Encoding column of the Multiplexer table, and will be one of the values:
 - **Both**: both MIDs are used to determine the Sample for each Read.
 - **Either**: there are MIDs at both ends of the sequence, but *either* one is sufficient to determine the Sample, with the MID in at the 5' end, in the read's orientation, being used for a given read
 - **Primer1**: there is only a MID sequence at the Primer1 end of the Read
 - **Primer2**: there is only a MID sequence at the Primer2 end of the Read.
- All editors provide informative status messages regarding any possible incompatibilities of the Multiplexer's associated MIDs, any incompletely specified or duplicate Sample mappings, or Samples that will be created new because their names don't match any Samples currently in the project.

Multiplexing Features – Sample Editors



Primer1 MID Encoding Sample Editor



Primer2 MID Encoding Sample Editor

Multiplexing Features – Sample Editors



▼ Edit Samples

Clear Flip Table AutoFill

▼ 1 / 2 ▶	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MID1	A1	B1	C1	D1	E1	F1	G1	H1
MID2	A2	B2	C2	D2	E2	F2	G2	H2
MID3	A3	B3	C3	D3	E3	F3	G3	H3
MID4	A4	B4	C4	D4	E4	F4	G4	H4
MID5	A5	B5	C5	D5	E5	F5	G5	H5
MID6	A6	B6	C6	D6	E6	F6	G6	H6
MID7	A7	B7	C7	D7	E8	F7	G7	H7
MID8	A8	B8	C8	D8	E8	F8	G8	H8
MID9	A9	B9	C9	D9	E9	F9	G9	H9
MID10	A10	B10	C10	D10	E10	F10	G10	H10
MID11	A11	B11	C11	D11	E11	F11	G11	H11
MID12	A12	B12	C12	D12	E12	F12	G12	H12

96 / 96 Sample Associations Defined

⚠ Sample E8 is used 2 times: [(MID5,MID7) (MID5,MID8)]

OK Cancel

“Both” Encoding Sample Editor

www.roche-applied-science.com

▼ Edit Samples

Clear Flip Table AutoFill

▼ 1 / 2 ▶	MID1	MID6	MID7	MID8	MID9	MID10
MID1						
MID2		B5				
MID3						
MID4						
MID8			B5			
MID9						
MID10						
MID11						

2 Samples / 8 Association Pairs

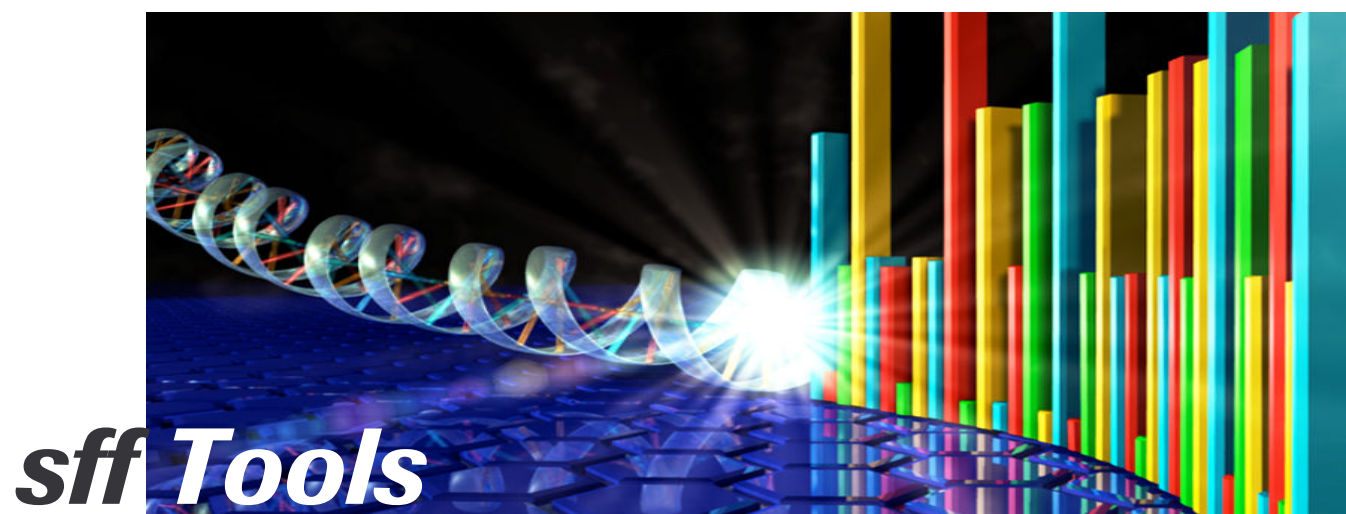
⚠ Asymmetric Design – Primer 1 and Primer 2 MID's differ

⚠ Sample B5 is used 2 times: [(MID6,MID2) (MID7,MID8)]

OK Cancel

“Either” Encoding Sample Editor





***sff* Tools**

SFF Tools

- `sffcalle` has been removed as it was only applicable to runs processed with 1.0.51 earlier
- `sffvolume` has been removed since the NCBI Trace Archive has shifted to the new Short Read Archive
- `sffrescore` has been added to rewrite SFF files with the new quality scores
- `sfffile` command now contains options to randomly pick a subset of reads (`-pick #`) or bases (`-pick #`)



Please contact Roche Applied Science Technical Support via:

phone: 1-800-262-4911

e-mail: us.gssupport@roche.com

gsMapper Project Tab / References / Accession Index



1.1.03.24 The Accession Index Column is *NEVER* Populated, and has been eliminated in 2.0.00

Name	Path	Number of Reads	Number of Bases	Accession Index
NC_000907.fna	/data/asmMapProjects/	1	1830138	