Applied Biosystems

Sequencing Analysis Software Version 5.1

The Applied Biosystems DNA Sequencing Analysis Software v5.1 is designed to analyze, display, edit, save, and print sample files generated from Applied Biosystems DNA analyzers and other ABI PRISM genetic analyzers. Features include:

- Novel basecaller algorithm that performs base calling for pure and mixed base calls
- Generation of quality values to provide basecall accuracy information for pure and mixed base calls
- Analysis report to help troubleshoot and provide easy assessment of data quality
- New Sample Manager interface
- Optional feature to generate an audit trail of base changes

Concepts Used in Sequencing Analysis Software v5.1

- Analysis Protocol – Contains all the settings necessary for analysis and used to perform basecalling and post processing.
- Analysis Report – Shows the status of the data analysis. The report can be used to help troubleshoot and provide easy assessment of data quality.
- Clear Range – Region of sequence that remains after excluding the low-quality or error-prone sequence at both the 5′ and 3′ ends.
- Length of Read (LOR) – Measurement of the length of quality bases. The LOR is user definable in the Display Settings dialog box and is displayed in the Analysis report.
- Quality Values (QVs) – Per-base estimate of the base calling accuracy.
- Sample Score – Average quality value of the bases in the clear range sequence for that sample.
- KB Basecaller – New algorithm that calculates mixed or pure bases and quality values.
- ABI basecaller – Algorithm used in Sequencing Analysis software (v3.7 and earlier), which identifies pure bases.

Input Sample Files

Sequencing Analysis software is compatible with sample files generated from the:

- Applied Biosystems 3730/3730*xl* DNA Analyzer
- ABI PRISM® 3700 DNA Analyzer
- ABI PRISM® 3100/3100-*Avant* Genetic Analyzer
- ABI PRISM® 377 DNA Sequencer
- ABI PRISM® 310 Genetic Analyzer

Output Files for Sequencing Analysis

- Analyzed ABI files (.ab1)
- Text file of the sequence in ABI or FASTA format (.seq)
- Phred (.phd.1) files
- Standard chromatogram format (.scf) files
- Analysis Report

Program Overview

Chapter 2

All analysis in Sequencing Analysis software occurs in the Sample Manager. Analysis and review of your samples involve six steps:

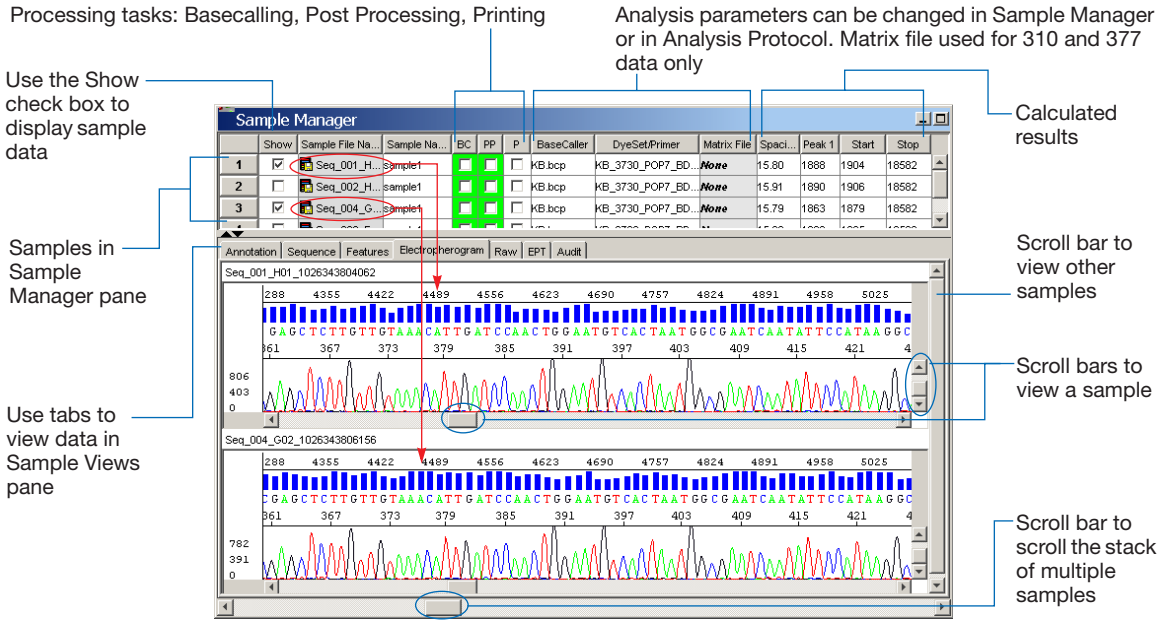
1. **Add sample(s) to the Sample Manager.**
2. **Show the sample data.**
3. **Edit and apply an analysis protocol (optional).**
4. **Analyze the data.**
5. **Review overall results and generate an analysis report.**
6. **Review samples and edit bases.**

For more information, refer to the chapter or page numbers in the *Applied Biosystems DNA Sequencing Analysis Software v5.1 User Guide* (P/N 4346366) that are indicated in the blue boxes.

The Sample Manager

Pages 3-5 & 5-2

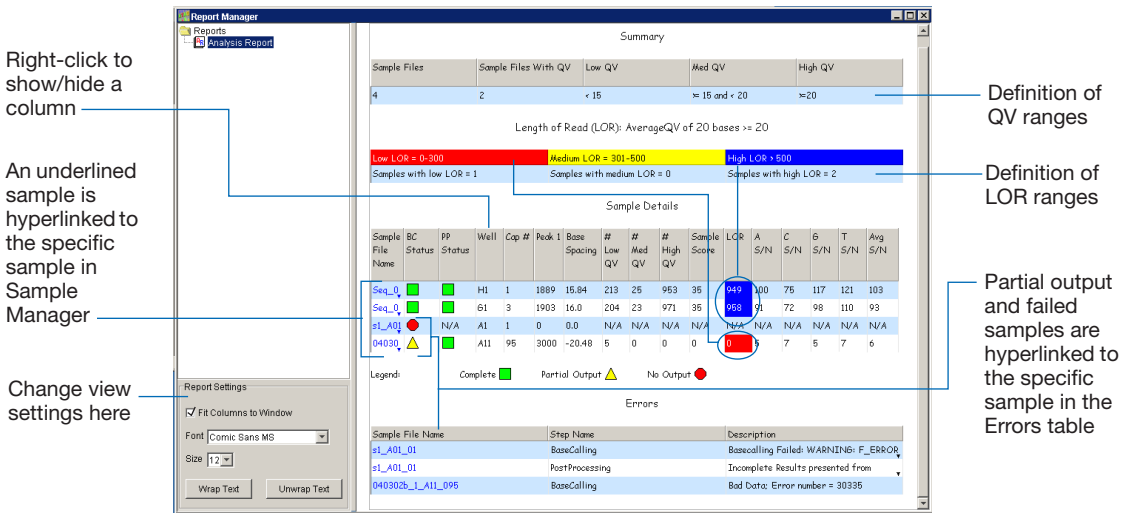
The Sample Manager is a window that displays sample files and the various analysis parameter values. All data is analyzed or reanalyzed from this window pane. Single or multiple sample views are displayed here.



The Analysis Report

Chapter 7

An Analysis Report shows the status of data analysis. An analysis report can be generated for any samples added to the Sample Manager. If the data is analyzed, the report displays a summary of QVs and LORs, as well as individual sample information and errors. If the data is unanalyzed, the report displays status information. It can be exported as a tab-delimited file and opened in Microsoft® Excel software for trend analysis.



Data Analysis and Review

All analysis in Sequencing Analysis software occurs in the Sample Manager. You can perform the analysis and review of your sequencing sample files by following the steps below.

To launch the analysis software, double-click the Sequencing Analysis v5.1 desktop icon.

Step 1: Add sample(s) to the Sample Manager.

Page 3-10

Use the Add Samples function to add samples to the Sample Manager for analysis, printing, viewing, or editing data.

1A. Click (Add Sample).

1B. Select the files that you want to add to the Samples To Add section of the dialog box.

To add ...	Do this ...
A single file to the list	Select the file, then click **Add Selected Samples**.
Multiple files	Shift-click to select continuous samples or Ctrl-click to select discontinuous samples, then click **Add Selected Samples**.
All samples in a single folder	Select the folder, then click **Add Selected Samples**.

1C. Click **OK** in the Add Samples dialog box.

The dialog box closes and the selected files are added to the Sample Manager window.

Step 2: Show the sample data.

Page 3-13

2A. Use the Sample Manager or Sample Navigator view to show data for one or more sample files.

To show the data for ...	Do this ...
A single sample	Double-click the sample file name or select the corresponding **Show** check box.
Multiple samples	Shift-click, Shift-drag or Ctrl-click the sample row numbers to select the sample files, then click Show.
All samples	Click the empty box above row number 1 or Shift-drag the sample row numbers to select all samples, then click Show.

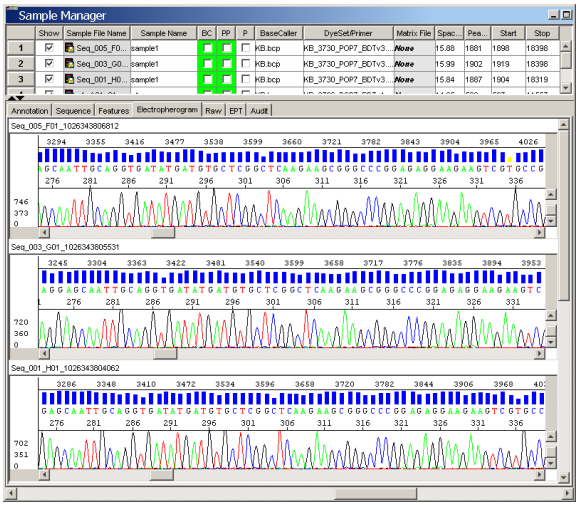


Figure 1 Samples in the Sample Manager view

2B. To show the data in the Sample Navigator, select **View > Sample Navigator** or click .

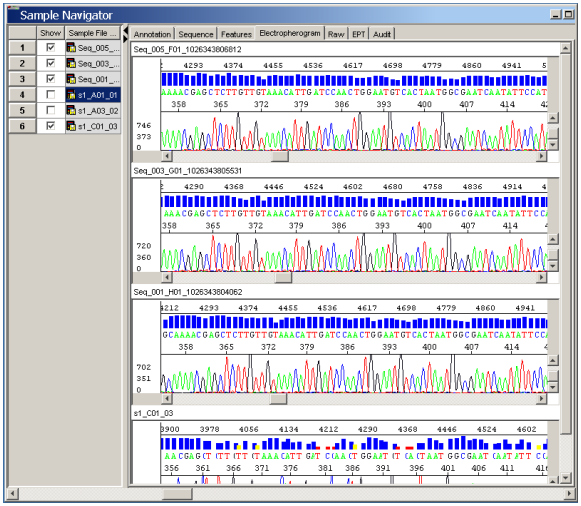


Figure 2 Samples in the Sample Navigator view

Step 3: Edit and apply an analysis protocol (optional).

Page 8-13

Use the Analysis Protocol Manager function to change the analysis protocol settings for samples.

3A. Select the sample rows of interest in the Sample Manager.
- Use Shift-click to select continuous samples.
- Use Ctrl-click to select discontinuous samples.

3B. Select **Analysis > Analysis Protocol Manager**.

3C. In the Analysis Protocol column, select the protocol you want to edit, then double-click the protocol name.

3D. Make changes in the General, Basecalling, Mixed Bases, and Clear Range tabs, as appropriate.

3E. Click **OK** to save the protocol and close the Sequence Analysis Protocol Editor.

3F. Click:
- **Apply to Selected Samples** to apply the protocol to the samples selected in step 3A, or
- **Apply to All Samples** to apply the protocol to all the samples in the Sample Manager.

IMPORTANT! Do not skip this step.

3G. Click **Done** to close the Analysis Protocol Manager.

Step 4: Analyze the data.

Chapter 5

Use to start the basecalling, post processing, and printing tasks that you selected.

Click  (Start Analysis) and Sequencing Analysis software will automatically perform the following tasks:

Basecalling

When the BC (basecalling) parameter is selected, the basecaller performs the either of following tasks:

- Calls the bases with the KB basecaller
  - If the mixed base option is selected in the analysis protocol, then mixed bases are called.
  - If the mixed base option is not selected, then pure bases are called (A, C, G, and T only).
  - Calculates quality values (QVs) for pure and mixed bases, if the mixed base option is selected.
- Calls the bases with the ABI basecaller
  Assigns A, C, G, T, or N to every base (no mixed-base calling and QV options)

The Sequencing Analysis software then generates optional file formats (.seq, .phd.1 and/or .scf).

Post Processing

When the PP (post processing) parameter is selected in the Sample Manager, the software calculates the clear range.

Printing

When the P (printing) parameter is selected, the software prints the sample views after analysis and post processing.

Note: The views that are printed are defined in the Options dialog box. To change the defaults, select **Tools > Options**, then select the **Printing** tab.

Step 5: Review overall results and generate an analysis report.

Use the Sample Manager and the analysis report to review your data.

5A. Review your results in the Sample Manager.

a. Look for green, yellow, or red boxes for the BC parameter. Green indicates the process was successful, yellow indicates poor quality data, and red indicates failure. Page 5-8

Note: The yellow result applies only to samples analyzed with the KB basecaller.

b. Look for green or red boxes for the PP and/or P parameters. Green indicates the process was successful, and red indicates failure.

c. Review the base spacing, peak 1 location, and start and stop points. A red value in the Base Spacing column indicates that the spacing could not be calculated and the default value was used for analysis. Page 5-19

5B. Review the analysis report. Chapter 7

a. Click  (Analysis Report) to generate and display the report.

b. Review the data in the report.

c. To export the report, select **File > Export Report**. The file is exported in a tab-delimited format.

Step 6: Review samples and edit bases. Page 4-2

6A. Select a sample file.

6B. Review your results in the sample file:

a. Review the raw, analyzed, and EPT data.

b. Review low-quality basecalls and check for errors.

6C. Edit the bases, as needed.

- To navigate through the data, use the keyboard shortcuts in the table in the next column.
- The QVs change depending on the edits you make. If you:
  - Insert a base – No QV is added
  - Delete a base – QV is deleted
  - Change a base – QV has the same value but is displayed as a gray bar

6D. Save the sample file. The .seq file created when the data was analyzed is updated when you save the sample file.

Quality Values Page 6-4

The QV is a per-base estimate of the basecaller accuracy. The QVs are calibrated on a scale corresponding to:

$$QV = -10\log_{10}(Pe)$$

where *Pe* is the probability of error.

The KB basecaller generates QVs from 1 to 99.

Quality Value	Probability the Basecall Is Incorrect
10	10%
20	1%
30	0.1%
40	0.01%
50	0.001%

- Typical high quality pure bases have QVs 20 to 50
- Typical high quality mixed bases have QVs 10 to 50
- Size and color of QVs bars are identical for QVs 50 to 99

The QVs are optionally displayed as bars above each base in the Sequence and Electropherogram views. The QVs are also displayed in the analysis report.

You can change the colors and values of the QV bars in the Display Settings: Chapter 9

1. Select (Display Settings).
2. In the Sample File Display section, change the range for the QVs. Use the two sliders to define the low, medium, and high ranges.

QV Bar	Default Color and Range	Set the range to identify data that is …
Low	Red 0 to 14	Not acceptable
Medium	Yellow 15 to 19	Needs manual review
High	Blue 20 or higher	Acceptable

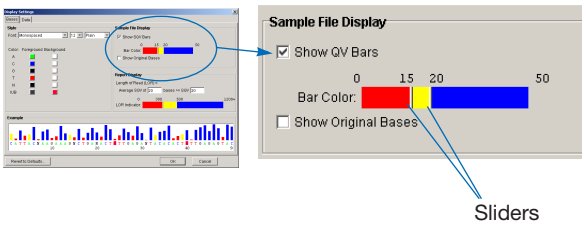


Figure 3 Display Settings: Bases Tab

3. Click **OK** to save the new settings and close the dialog box.

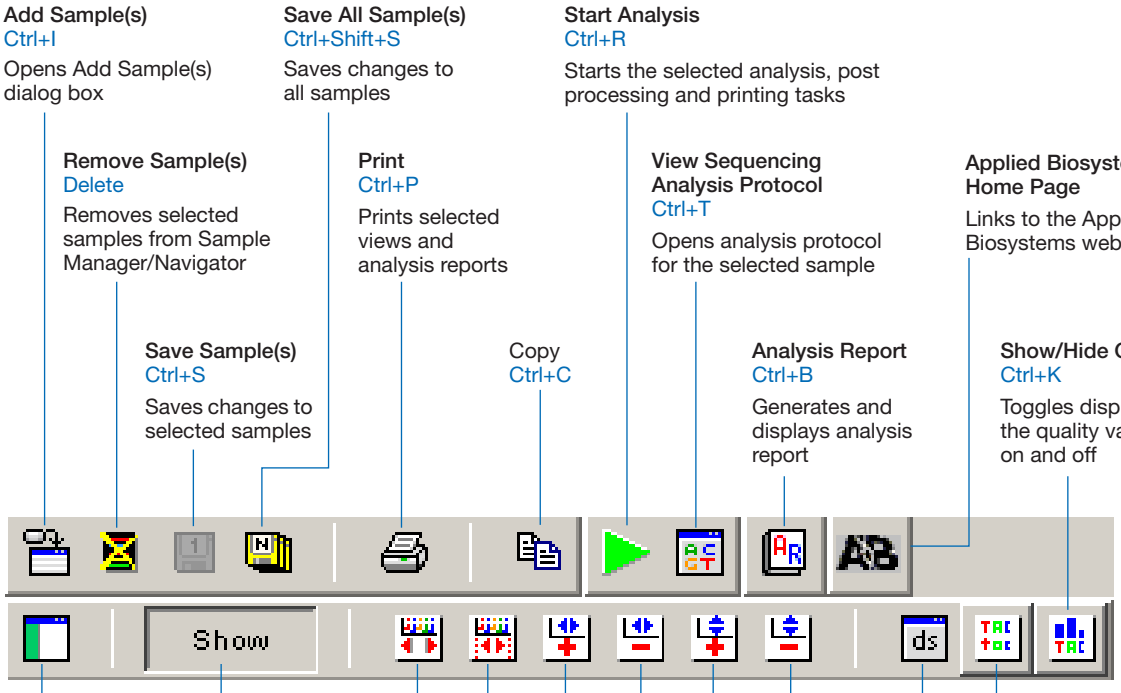
Shortcuts for Working with Data

Use the following shortcuts when working with data in the Sequence and Electropherogram views.

To move to the …	Press …
Next base	Right arrow
Previous base	Left arrow
Next/previous N	Tab key/Shift+Tab
Next/previous 10 bases	F5 key/Shift+F5
Next/previous low QV	F6 key/Shift+F6
Next/previous medium QV	F7 key/Shift+F7
Next/previous high QV	F8 key/Shift+F8

Toolbar

The toolbar displays buttons for the most commonly used software functions. See the graphic below for the name, keyboard shortcut, and description of each button.



Add Sample(s)
Ctrl+I
Opens Add Sample(s) dialog box

Remove Sample(s)
Delete
Removes selected samples from Sample Manager/Navigator

Save Sample(s)
Ctrl+S
Saves changes to selected samples

Save All Sample(s)
Ctrl+Shift+S
Saves changes to all samples

Print
Ctrl+P
Prints selected views and analysis reports

Copy
Ctrl+C

Start Analysis
Ctrl+R
Starts the selected analysis, post processing and printing tasks

View Sequencing Analysis Protocol
Ctrl+T
Opens analysis protocol for the selected sample

Analysis Report
Ctrl+B
Generates and displays analysis report

Applied Biosystems Home Page
Links to the Applied Biosystems web page

Show/Hide QV
Ctrl+K
Toggles display of the quality values on and off

Toggle
Ctrl+N
Toggles between Sample Navigator and Sample Manager views

Show/Hide
Ctrl+U
Displays selected sequence file data

Full View
Ctrl+[
Displays all data in a standard size window

Actual Size
Ctrl+]
Restores display to initial default zoom factor

Zoom In Horizontal
Ctrl+=
Enlarges view horizontally

Zoom Out Horizontal
Ctrl+Minus
Reduces view horizontally

Zoom In Vertical
Ctrl+Shift+=
Enlarges view vertically

Zoom Out Vertical
Ctrl+Shift+Minus
Reduces view vertically

Display Settings
Ctrl+Y
Opens Displays Settings dialog box

Show/Hide original sequence
Ctrl+J
Toggles display of original sequence on and off