



Applied Biosystems SOLiD™ 4 System SETS Software User Guide

SOLiD Experimental Tracking Software (SETS) v4.0.1

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.

This user guide is the proprietary material of Life Technologies or its subsidiaries and is protected by laws of copyright. The customer of the SOLiD™ 4 System is hereby granted limited, non-exclusive rights to use this user guide solely for the purpose of operating the SOLiD™ 4 System. Unauthorized copying, renting, modifying, or creating derivatives of this user guide is prohibited.

Information in this document is subject to change without notice.

APPLIED BIOSYSTEMS DISCLAIMS ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. TO THE FULLEST EXTENT ALLOWED BY LAW, IN NO EVENT SHALL APPLIED BIOSYSTEMS BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF, WHETHER OR NOT FORESEEABLE AND WHETHER OR NOT APPLIED BIOSYSTEMS IS ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Notice to Purchaser: License Disclaimer

Purchase of this software product alone does not imply any license under any process, instrument or other apparatus, system, composition, reagent or kit rights under patent claims owned or otherwise controlled by Applied Biosystems, either expressly, or by estoppel.

TRADEMARKS:

The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners. Cy3 and Cy5 are registered trademarks of GE Healthcare. NanoDrop is a registered trademark of NanoDrop Technologies.

© Copyright 2010, Life Technologies Corporation. All rights reserved.

Part Number 4448411 Rev. B
04/2010

	Preface	7
	How to use this guide	7
	How to obtain support	8
Chapter 1	SETS Software Overview	9
	SOLiD™ 4 System overview	10
	SETS Software overview	12
	SETS homepage	13
Chapter 2	Prepare Run Settings	17
	ICS, SETS, and BioScope™ Software applications	18
	SOLiD™ 4 System Software workflow	19
	Log in to SETS	21
	Set analysis parameters	22
	Load a Reference Sequence	37
Chapter 3	Monitor the Run in SETS	41
	Overview	42
	View Run Metrics	42
	View Cycle Scans	46
	View sample data	49
	Troubleshoot failed jobs	51
	View History	57
	Cancel analysis midstream	58
	Cluster Status	60
	System Logs	61
	RSS feed	62
Chapter 4	View Reports	63
	View overall run reports	64
	View sample reports	77
	View Analysis Reports	85
	Master Report Tool	93

Contents

Chapter 5	Perform Reanalysis	97
	Overview	98
	Primary reanalysis	98
	Secondary reanalysis	100
Chapter 6	Support Multiplexing	103
	Fragment and Paired-End Multiplexing	104
	View Multiplexing Series in SETS	105
	Set up Multiplexing run in ICS	107
	Multiplexing Assignment Report	108
	Reassign Run	111
Chapter 7	Manage Administrative Tasks	113
	Export Data with SOLiD™ 4 System Software	114
	Auto export in SETS	115
	Manual export in SETS	124
	Event notifications and email services	125
	Alerts	128
	Delete runs	129
	Manage Users	130
Appendix A	Web Services	133
	Overview	133
	WSDL	134
Appendix B	Advanced Topic: Data Analysis Overview	137
	Fundamentals of color-space analysis	137
	The 2-base color coding scheme	137
	Color-coding process	138
	Required properties for a 2-base color code scheme	141
	How the system satisfies 2-base coding system requirements ..	142
	Principles of ligation-based chemistry and 2-base encoding	143

	Color-space and base space as applied in the SOLiD™ 4 System	145
	Color-space data	145
	Relationship between cycle and base position	145
	Color-space formats	146
	Complementing color-space data	147
	2-base encoding and error recognition	148
	Processing	148
	Data analysis considerations	151
	Understand Ns and color-space	151
	Find Single Nucleotide Polymorphisms (SNPs)	151
	Color-space rules for SNP detection	151
	SNPs and errors	153
	High accuracy for SNP detection	153
	SNP error rates	154
	Sampling	155
	Allele ratio	155
	Find polymorphisms in color-space	155
Appendix C	Advanced Topic: Data Management	157
	Saving data	157
	Approximate storage space for all data	157
	Raw image data	158
	Primary analysis results data	158
	Data transfer	160
	Exporting data	160
Appendix D	Advanced Topic: Primary Analysis	161
	What is primary analysis?	161
	Primary analysis overview	162
	Primary analysis workflow	163
	Primary analysis inputs and outputs	165
	Storage requirements for primary analysis	169
	Perform primary re-analysis	170

Contents

Redo the read-filtering process	175
Check the completion of primary analysis	175
Image metrics	177
Understanding quality values in the SOLiD™ System	180
Usage	181
Validation and observed quality	181
Key files generated by primary analysis	183
xxx_sequence.csfasta	183
xxx_sequence.QV.qual	184
.spch files	185
Procedures to verify primary analysis status	186
Troubleshooting analysis failure	187
Procedure for re-analyzing the image (primary analysis)	190

Appendix E	Software License Agreement	191
------------	---	------------

APPLIED BIOSYSTEMS END USER SOFTWARE LICENSE AGREEMENT	191
THIRD PARTY PRODUCTS	192
TITLE	192
COPYRIGHT	192
LICENSE	193
LIMITED WARRANTY and LIMITATION OF REMEDIES	196

Glossary	201
--------------------	-----

Documentation	207
-------------------------	-----

Related documentation	207
Send us your comments	208

How to use this guide

- Text conventions** This guide uses the following conventions:
- **Bold** text indicates user action. For example:
Type **0**, then press **Enter** for each of the remaining fields.
 - *Italic* text indicates new or important words and is also used for emphasis. For example:
Before analyzing, *always* prepare fresh matrix.
 - A right arrow symbol (▶) separates successive commands you select from a drop-down or shortcut menu. For example:
Select **File ▶ Open ▶ Spot Set**.
Right-click the sample row, then select **View Filter ▶ View All Runs**.

User attention words Two user attention words appear in Life Technologies user documentation. Each word implies a particular level of observation or action as described below:

Note: – Provides information that may be of interest or help but is not critical to the use of the product.

IMPORTANT! – Provides information that is necessary for proper instrument operation, accurate chemistry kit use, or safe use of a chemical.

How to obtain support

For the latest services and support information for all locations, go to:

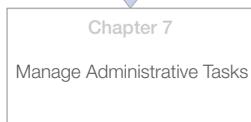
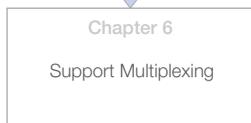
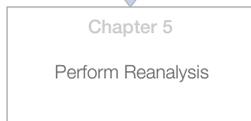
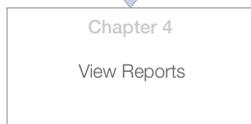
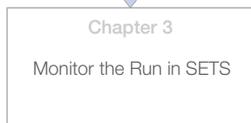
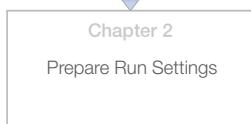
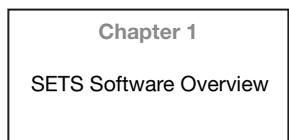
www.appliedbiosystems.com

At the Life Technologies web site, you can:

- Access worldwide telephone and fax numbers to contact Life Technologies Technical Support and Sales facilities
- Search through frequently asked questions (FAQs)
- Submit a question directly to Technical Support
- Order Life Technologies user documents, MSDSs, certificates of analysis, and other related documents
- Download PDF documents
- Obtain information about customer training
- Download software updates and patches

1

SETS Software Overview



This chapter covers:

- SOLiD™ 4 System overview 10
- SETS Software overview 12
- SETS homepage 13

SOLiD™ 4 System overview

IMPORTANT! The features presented in this document are related to the SOLiD™ 4 System SETS v4.0.1.

SOLiD™ system technology

The Applied Biosystems SOLiD™ 4 System provides parallel sequencing of clonally amplified DNA fragments linked to magnetic beads. The sequencing methodology is based on sequential ligation with dye-labeled oligonucleotides. All fluorescently labeled oligonucleotide probes are present simultaneously, competing for incorporation. After each ligation, fluorescence is measured before another round of ligation takes place.

This ligation-based chemistry eliminates dephasing. The use of a two-base encoding mechanism, which interrogates each base twice for errors during sequencing, discriminates between true polymorphisms and system noise. The SOLiD™ 4 System software allows customers to monitor runs in real-time, and provides basic data analysis tools.

The SOLiD™ 4 System consists of:

- SOLiD™ 4 Analyzer
- All ancillary equipment
- ICS software, which is described in the *Applied Biosystems SOLiD™ 4 System ICS Software Help*
- SOLiD™ Experiment Tracking System (SETS) software, which is described in this document

Naming convention for sequencing reads

The following shows the naming convention for the sequencing runs.

- Fragment
- Paired-end
- Mate-pair
- Multiplex fragment
- Multiplex paired-end
- WFA (not shown below)

Selecting a Fragment run and then selecting Multiplexing will result in creating a barcoded fragment library. The same concept applies to the Paired-End run.

Fragment sequencing



Paired-end sequencing



Mate-pair sequencing



Multiplex fragment sequencing



Multiplex paired-end sequencing



SETS Software overview

Overview The SOLiD™ Experiment Tracking System (SETS) software is a web-based application for viewing real-time data and completed run analysis reports from the SOLiD™ 4 System. The SOLiD™ 4 Analyzer uses ligation-based sequencing, then filters data to remove missing calls, nonligations, and/or redundancies.

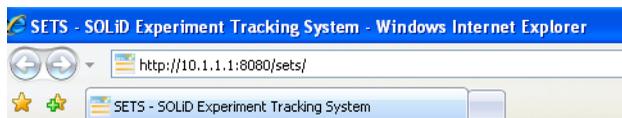
Data are analyzed at two levels:

- Primary analysis, which involves:
 - Image alignment
 - Color assignment
 - Color calls
 - Quality value assignments
- Secondary analysis, which compares color space calls to a reference sequence.

Note: Applied Biosystems recommends performing the secondary analysis off-instrument with BioScope™ Software. Refer to *BioScope™ Software for Scientists Guide: Data Analysis Methods and Interpretation* (PN 4448431).

IMPORTANT! Do not install BioScope™ Software on the SOLiD™ 4 Analyzer.

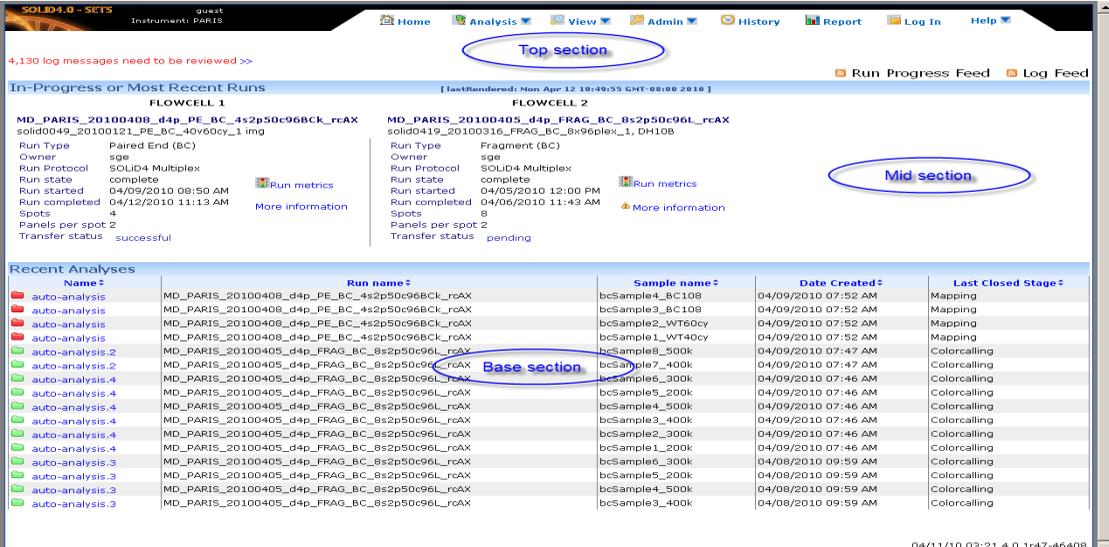
Opening SETS Access SETS by typing in the SETS software URL supplied by the Applied Biosystems Field Service Engineer:



This address opens the SETS software to the homepage and you are automatically logged-in as a “guest” user. As a guest, you can only view analyzed data and reports. Log in to the software to perform other functions. See “Log in to SETS” on page 21.

SETS homepage

The SETS software opens the homepage where you can view real-time and completed-run data generated by the SOLiD™ 4 Analyzer.



The screenshot displays the SETS software homepage. At the top, there is a navigation bar with links for Home, Analysis, View, Admin, History, Report, Log In, and Help. Below the navigation bar, a message indicates that 4,130 log messages need to be reviewed. The main content area is divided into three sections:

- Top section:** Contains navigation and error messages.
- Mid section:** Displays details for two flowcells (FLOWCELL 1 and FLOWCELL 2), including run type, owner, protocol, and completion status.
- Base section:** A table titled "Recent Analyses" listing various analysis runs with columns for Name, Run name, Sample name, Date Created, and Last Closed Stage.

Name	Run name	Sample name	Date Created	Last Closed Stage
auto-analysis	MD_PARIS_20100408_d4p_FE_BC_4s2p50c96Bck_rcAX	bcSample4_BC108	04/09/2010 07:52 AM	Mapping
auto-analysis	MD_PARIS_20100408_d4p_FE_BC_4s2p50c96Bck_rcAX	bcSample3_BC108	04/09/2010 07:52 AM	Mapping
auto-analysis	MD_PARIS_20100408_d4p_FE_BC_4s2p50c96Bck_rcAX	bcSample2_WT60cy	04/09/2010 07:52 AM	Mapping
auto-analysis	MD_PARIS_20100408_d4p_FE_BC_4s2p50c96Bck_rcAX	bcSample1_WT40cy	04/09/2010 07:52 AM	Mapping
auto-analysis.2	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample8_500k	04/09/2010 07:47 AM	Colorcalling
auto-analysis.2	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample7_400k	04/09/2010 07:47 AM	Colorcalling
auto-analysis.4	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample6_300k	04/09/2010 07:46 AM	Colorcalling
auto-analysis.4	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample5_200k	04/09/2010 07:46 AM	Colorcalling
auto-analysis.4	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample4_500k	04/09/2010 07:46 AM	Colorcalling
auto-analysis.4	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample2_400k	04/09/2010 07:46 AM	Colorcalling
auto-analysis.4	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample2_300k	04/09/2010 07:46 AM	Colorcalling
auto-analysis.4	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample1_200k	04/09/2010 07:46 AM	Colorcalling
auto-analysis.3	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample6_300k	04/08/2010 09:59 AM	Colorcalling
auto-analysis.3	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample5_200k	04/08/2010 09:59 AM	Colorcalling
auto-analysis.3	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample4_500k	04/08/2010 09:59 AM	Colorcalling
auto-analysis.3	MD_PARIS_20100405_d4p_FRAG_BC_8s2p50c96L_rcAX	bcSample3_400k	04/08/2010 09:59 AM	Colorcalling

The homepage is divided into three sections:

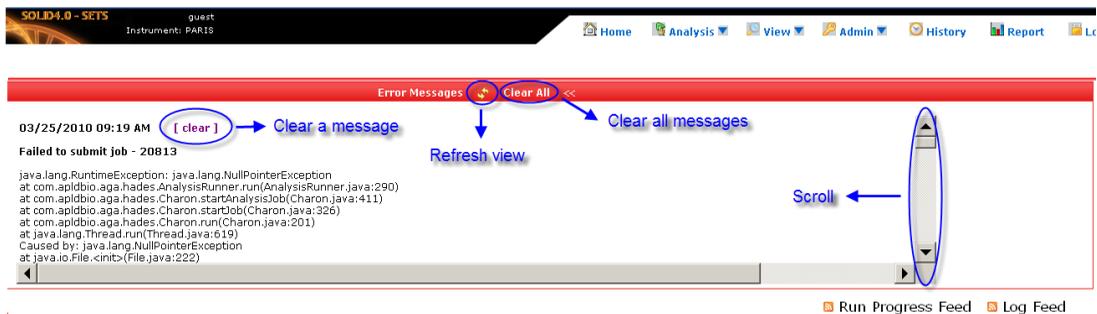
- Top section: Main menu and error messages (when present)
- Middle section: In-progress or the most recent runs (per flowcell)
- Base section: Recent analysis

Error messages

When there are error messages, the Error Messages section displays them at the top of the homepage after you click the **View Messages** arrow.



You can scroll through the messages, refresh the view for all messages received, or remove one message, or all messages from the field, as shown.



For more information on log messages see “System Logs” on page 61.

In progress or most recent run

This section displays the status of a current run or results for the most recently completed run. Flowcell 1 displays on the left-hand side of the window; Flowcell 2 displays on the right side.



To understand more about your run, click **Run metrics** or **More Information** within each flowcell section.

Run metrics is quality metrics in the form of traffic lights to help you assess the progress and measurable quality of the run. “More Information” provides a view of analyzed data for a particular run. For details on monitoring a run, see [Chapter 3, Monitor the Run in SETS](#).

Note: At any time you can return to the SETS homepage by clicking the  **Home** button on the main menu at the top of the screen.

Recent analyses This section displays the progress of data analysis for individual samples.

Every analysis is an active hyperlink.
 Click the analysis that you want to open its Run Details page.

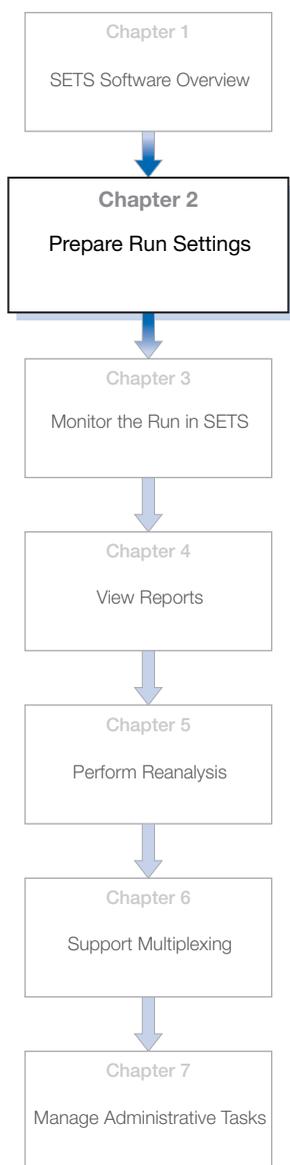
Name	Run name	Sample name	Date Created	Last Closed Stage
auto-analysis	MD_PARIS_20100212_r4500_MP_4s4p50c_AX	Sample4	02/12/2010 03:28 PM	Barcoding
auto-analysis	MD_PARIS_20100212_r4500_MP_4s4p50c_AX	Sample3	02/12/2010 03:28 PM	Barcoding
auto-analysis	MD_PARIS_20100212_r4500_MP_4s4p50c_AX	Sample2	02/12/2010 03:28 PM	Barcoding
auto-analysis	MD_PARIS_20100212_r4500_MP_4s4p50c_AX	Sample1	02/12/2010 03:28 PM	Barcoding
auto-analysis	MD_PARIS_20100212_r4500_PE_BC_4s4p2L50c_AX	bcSample2	02/12/2010 08:40 AM	Barcoding
auto-analysis	MD_PARIS_20100212_r4500_PE_BC_4s4p2L50c_AX	bcSample1	02/12/2010 08:40 AM	Barcoding
auto-analysis	MD_PARIS_20100211_d4500_PE_4s2p25c_AX	Sample4	02/11/2010 09:54 PM	Barcoding
auto-analysis	MD_PARIS_20100211_d4500_PE_4s2p25c_AX	Sample3	02/11/2010 09:54 PM	Barcoding
auto-analysis	MD_PARIS_20100211_d4500_PE_4s2p25c_AX	Sample2	02/11/2010 09:54 PM	Barcoding
auto-analysis	MD_PARIS_20100211_d4500_PE_4s2p25c_AX	Sample1	02/11/2010 09:54 PM	Barcoding
auto-analysis	MD_PARIS_20100211_d4500_PE_2s1p2L25c_AX	bcSample2	02/11/2010 04:01 PM	Barcoding
auto-analysis	MD_PARIS_20100211_d4500_PE_2s1p2L25c_AX	bcSample1	02/11/2010 04:01 PM	Barcoding
auto-analysis.4	MD_PARIS_20100211_d4500_PE_BC_2s1p2L25c_AX_iss1779_2292	bcSample2	02/11/2010 03:46 PM	Barcoding
auto-analysis.4	MD_PARIS_20100211_d4500_PE_BC_2s1p2L25c_AX_iss1779_2292	bcSample1	02/11/2010 03:46 PM	Barcoding
auto-analysis.3	MD_PARIS_20100211_d4500_PE_BC_2s1p2L25c_AX_iss1779_2292	bcSample2	02/11/2010 03:38 PM	Colorcalling
auto-analysis.3	MD_PARIS_20100211_d4500_PE_BC_2s1p2L25c_AX_iss1779_2292	bcSample1	02/11/2010 03:38 PM	Colorcalling

For every analysis, you can see:

- Run name
- Sample name
- Date created
- Last closed stage

2

Prepare Run Settings



This chapter covers:

- ICS, SETS, and BioScope™ Software applications 18
- SOLiD™ 4 System Software workflow. 19
- Log in to SETS 21
- Set analysis parameters. 22
- Load a Reference Sequence 37

ICS, SETS, and BioScope™ Software applications

To understand the relationship between the three software components in the SOLiD™ 4 System and how a user would interact between them in a typical workflow, see the [“SOLiD™ 4 System Software workflow”](#) on page 19.

Analysis settings

Before you start a run on the SOLiD™ 4 Analyzer using the Instrument Control Software (ICS), open the SETS software and set your primary analysis and the secondary analysis settings. See [“Set analysis parameters”](#) on page 22.

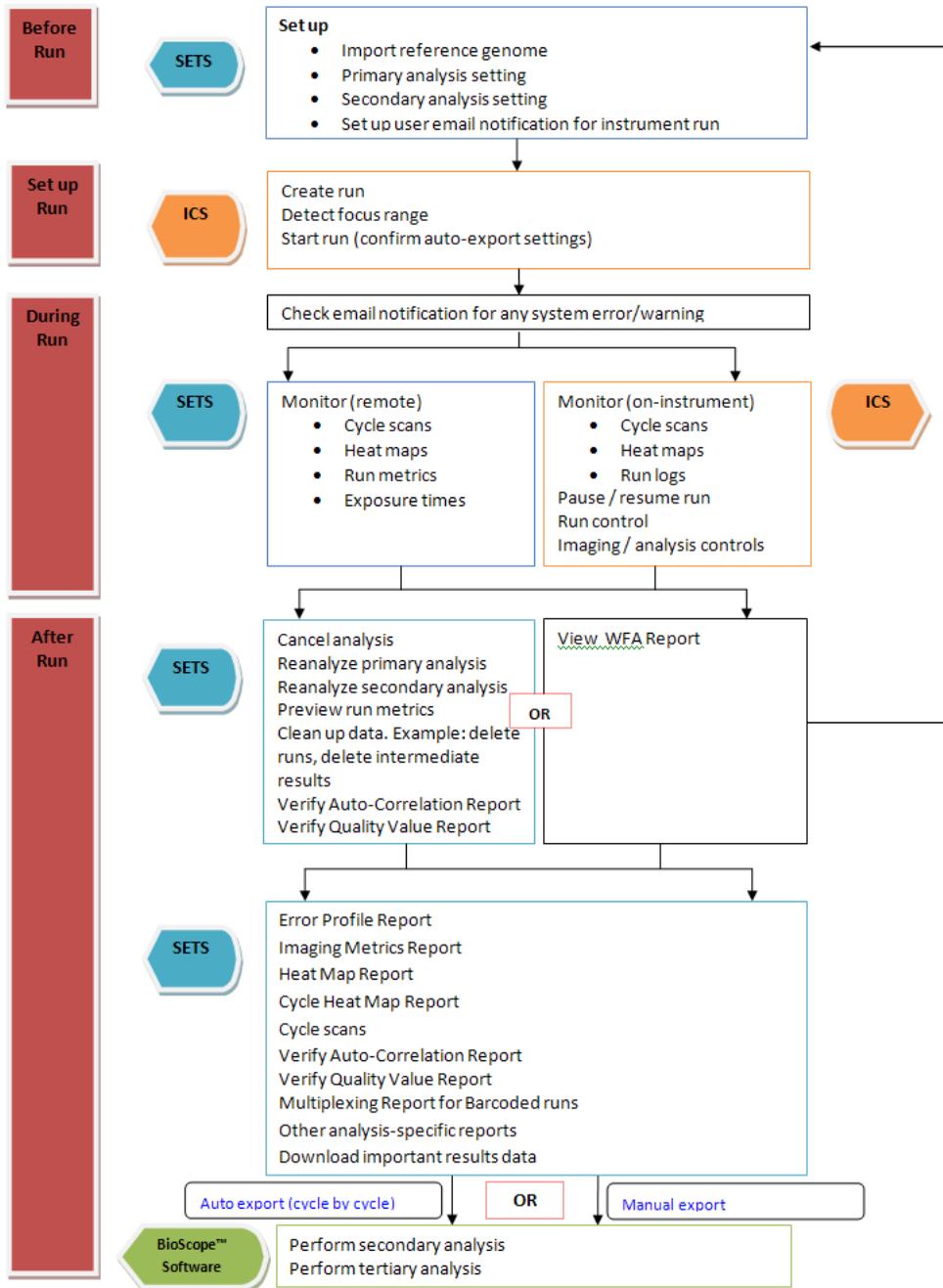
IMPORTANT! For more complex genomes (such as the human genome), barcoded samples, and larger reference (and reads) the secondary analysis should be performed off-instrument with the BioScope™ Software. Refer to *BioScope™ Software for Scientists Guide: Data Analysis Methods and Interpretation* (PN 4448431). If a sample contains more than the maximum allowable number of reads, the SETS software selects only up to the maximum allowed number of reads and the rest will not be included in the results of the secondary analysis.

IMPORTANT! Do not install BioScope™ Software on the SOLiD™ 4 Analyzer.

The Secondary analysis (sequence data alignments to a reference sequence) will not occur without a preloaded reference sequence. Perform secondary and tertiary analysis with the BioScope™ Software. See [“Load a Reference Sequence”](#) on page 37.

SOLiD™ 4 System Software workflow

The following figure shows the relationship workflow between ICS, SETS, and BioScope™ Software.



Log in to SETS

Logging in The default login status for the SOLiD™ SETS Software is a “guest” user. In the guest role, you can only view real-time run and completed-run data. To perform administrative functions (edit, reanalyze, or delete), log in as “admin” or “user.”

Note: The current instrument and user name always displays in the upper right corner of the SETS software homepage. For example:



To log in to SETS:

1. Click  **Log In** on the main menu to open the Login screen.

A screenshot of a 'Log in' dialog box. The title bar reads 'Log in'. It contains two input fields: 'User name:' with the placeholder text 'Your Name' and 'Password:' with a masked password of seven dots. Below the fields are two buttons: 'Submit' and 'Cancel'.

2. Enter the user name and assigned password in the Log In dialog.

Note: A Field Service Engineer (FSE) will assign the Administrator name and password at the time of installation. The “admin” can then assign new users. For more information, see [“Manage Users” on page 130](#).

3. Click **Submit**.

- Login permissions** The permissions for each login status are:
- Default login as “guest” allows you to view real-time and completed-run data
 - Log in as “user” allows you to:
 - View real-time and completed-run data
 - Edit analysis settings
 - Reanalyze data (primary and secondary reanalysis)
 - Export runs
 - Delete runs
 - Log in as “admin” allows you to perform all user actions and:
 - Load reference sequences
 - Manage users

Set analysis parameters

You can set parameters by selecting primary analysis settings and the application type (or secondary analysis settings and the application type), before starting a run in ICS.

You can perform filtering during the primary analysis.

IMPORTANT! For more complex genomes (such as the human genome), barcoded samples, and larger reference (and reads) the secondary analysis should be performed off-instrument with the BioScope™ Software. Refer to *BioScope™ Software for Scientists Guide: Data Analysis Methods and Interpretation* (PN 4448431).

If a sample contains more than the maximum allowable number of reads, the SETS software selects only up to the maximum allowed number of reads and the rest will not be included in the results of the secondary analysis.

IMPORTANT! Do not install BioScope™ Software on the SOLiD™ 4 Analyzer.

The Secondary analysis (sequence data alignments to a reference sequence) will not occur without a preloaded reference sequence. Perform secondary and tertiary analysis with the BioScope™ Software. See “Load a Reference Sequence” on page 37.

*How many samples will you need for your mask? 4

Sample Name	Library	Primary Analysis	Application	Secondary Analysis	Description	Remove
Sample1	defaultLibrary	default: primary	<Optional: Type or Se...	SINGLE_TAG_DH10B...		✗
Sample2	defaultLibrary	default: primary	<Optional: Type or Se...	SINGLE_TAG_DH10B...		✗
Sample3	defaultLibrary	default: primary	<Optional: Type or Se...	SINGLE_TAG_DH10B...		✗
Sample4	defaultLibrary	default: primary	<Optional: Type or Se...	SINGLE_TAG_DH10B...		✗

Define primary analysis settings

In SETS, you can define the traffic light thresholds as a part of your setup for primary analysis.

To modify the threshold settings:

1. Log into SETS.
2. Go to **Main Menu**
 - ▶ **Analysis** ▶ **Primary Analysis Settings**.
3. Click **[New...]** to begin modifying the threshold settings.



Primary Analysis Settings [New...]

Note: Click ✗ to delete an existing setting or ✎ to edit one.



4. Enter a new value for the Low or High percentage setting for any of the following parameters:
 - Stringency
 - Panel alignment
 - HQ bead
 - Bead retention

- Color balance
- Quality value
- Blur value
- Blur population
- Exposure value
- Exposure population
- Exposure time

Note: Refer to “Primary analysis parameters” on page 25 for details on these parameters.

Primary Analysis Setting

Owner Administrator

Name * Experiment April Experiment April Name your custom setting to easily identify it later

Description Control Beads ...

Minimum Calls 25

Stringency 20

Barcode Error Correction Level 1

Low and high thresholds for traffic light parameters:

	Low	High	
Panel Alignment	70 %	90 %	●
Bead Retention	80 %	95 %	●
HQ Bead	15 %	35 %	●
Color Balance	3 %	6 %	●
Quality Value	15.0	20.0	●
Blur Value	2.2	2.8	●
Blur Population	75 %	90 %	●
Exposure Value	0.05	0.15	●
Exposure Population	0 %	50 %	●
Exposure Time	200.0	800.0	●

Control the traffic lights by modifying threshold values for low and high, then click Save.

Advanced parameters

* = required fields

5. Click **Save** to keep the new settings.

Note: Click **Restore Defaults** at any time to reestablish the original (and recommended) default settings for primary analysis.

Primary analysis parameters

Primary analysis parameters control the traffic light “signals,” or thresholds that you will use in determining the quality of a run cycle, even to stop a run midstream if the traffic lights signal a poor quality run cycle.

The value for each setting is accessible in SETS and modifications can be made to increase or decrease each value for a custom run setting.

Target values for parameters

Parameter	Target Value for Primary Analysis
Stringency	20
Panel Alignment	≥90%
HQ Beads	≥35%
Bead Retention	≥95%
Color Balance	±3%
QV Median	≥20
Blur Value	±2.2
Blur Population	±5%
Exposure Value	±0.05
Exposure Population	0
Exposure Time	±200

Parameters descriptions

Parameter	Description
Stringency	<p>Stringency is one criterion used to filter out junk reads in ReadBuilder stage. For filtered reads, csfasta and QV.qual are stored in the reject folder instead.</p> <p>The valid range of stringency is from 0 to 80. Any stringency less than 0 is treated as 0, where all reads will pass the stringency threshold.</p> <p>If stringency is set larger than 80, maximal possible stringency of 80 is used to avoid over filtering good beads.</p>
Panel Alignment	<p>Panel alignment refers to the ability of the primary analysis process to be able to take a channel ligation image and successfully match it to the corresponding focal map image.</p> <p>The stage does not reset to the exact same position in either the X, Y, or a combination of the two directions. If there is a Z plane difference, the focus can also change the characteristics of the image.</p> <p>A lack of signal from either a failure in chemistry or an error in reagents can also lead to insufficient bead information required to complete the analysis.</p> <p>There are alignment scores for each pairwise combination of ligation and focal map.</p> <p>A panel that fails to align has gone past a critical threshold.</p> <p>The scores, available in the underlying cycle *_PA.tab files, contain information about the panel specific information.</p> <p>The panel alignment metric is a measure of the panels that failed to align, resulting in the loss of a cycle's worth of information for a given panel count of beads.</p> <p>When seeing a decrease in panel alignment, explore other image quality metrics reported such as: blur value and population, exposure value, and times.</p>

Parameter	Description
HQ Beads	<p>When visualizing the beads on a satay plot, those that are on the axis and far from the origin are the beads that have a bright signal in only one channel. High quality beads are a summary of the “good” and “best” beads that are reported in the Cycle Scans Report (available in both ICS and SETS).</p> <p>There are two concept definitions necessary: intensity and angle.</p> <p>The concept of intensity is defined as the magnitude of the scaled intensity vector;</p> <p>The concept of angle is defined as $1 - \cos(\theta)$, where θ is the angle (in four-dimensional scaled intensity space) between the bead intensity vector and the axis of the called color.</p> <p>The intensity value can increase due to the total intensity across all channels independent of the intensity of the primary channel alone.</p> <p>Considering the angle and the intensity, a “good” bead is a usable bead that satisfies the following conditions:</p> <ul style="list-style-type: none"> • intensity >0.05 • $0.0125 < \text{angle} < 0.05$ <p>A “best” bead is a usable bead that satisfies these conditions:</p> <ul style="list-style-type: none"> • intensity >0.05 • angle < 0.0125 <p>A “best” bead is not a “good” bead because these are exclusive classes.</p> <p>The set of high quality beads, for this metric, is the merging of the two classes.</p> <p>These values are also highly correlated to the quality values (QV) for calls. A drop in the percentage of beads falling into the HQ category indicates that an accurate color calling is becoming more difficult.</p>
Bead Retention	<p>The number of usable beads in the analysis of a panel for any particular ligation cycle is nearly always smaller than the total number of beads found in the corresponding focal map image.</p> <p>Bead retention is simply the fraction of the total number of beads that are usable.</p> <p>Because the stage can drift slightly from one slide scan to the next, beads near the edges of a panel might not be imaged in all four colors.</p> <p>The software will call a color only for bead locations that are imaged in all four color-scans of a ligation cycle (i.e. any usable bead).</p> <p>Since beads are identified, or found, only during the focal map analysis stage, a call is made for such a bead location, even if the bead has physically moved from that position. In such cases, since the bead location is most likely now vacant, subsequent color calls will be based on data dominated by noise or, if present, a close neighbor bead. As such beads are still classified as usable, bead retention is not sensitive to such dislocations.</p>

Parameter	Description
Color Balance	<p>Because of the 2-base encoding utilized in the SOLiD™ 4 System, changes in underlying base balance should not significantly skew the color balance. In fact, the complete absence of a base would still leave all four colors present in approximately equal distribution.</p> <p>Color balance is the relative proportion of beads in a given cycle that are called as each of the four colors.</p> <p>Changes in the color balance can indicate that there are problems in the sequencing run.</p> <p>Signal drops, or absence in one channel, can be large enough that normalization does not compensate, resulting in a significant drop in beads called for a channel.</p> <p>It is important to note that library preparation methods can create either positive or negative selections that do change the color distribution at significant cycles (cycles corresponding to primer attachments or at the ends of a tag).</p> <p>If a shift in color balance is out of expectation, looking at imaging metrics—exposure in particular—for the channel that is either too high or low can explain the problem.</p> <p>If the images, alignment, exposure, quality values, and HQ beads all look within norms, then the color balance shift may be the result of the sample.</p> <p>For the color balance calculation, the colors are assumed to be in equal proportion.</p> <p>The real expected color balance would more accurately be determined by the dinucleotide frequency of the underlying DNA source. A target value of 6% would mean that all channels remain within +/- 6% of the 25% “ideal.”</p>
Quality Value	<p>Every bead has a color call and a quality value assigned to it after every cycle.</p> <p>The quality value is an empirically defined value based on a phred-like score equating to the confidence that the color called for that cycle is the correct one.</p> <p>In general, the brighter the bead, the greater the difference in signal between the primary and secondary colors and the higher the QV. This is similar to the HQ beads discussed previously, but where the HQ classification tracks the number of beads that fall into that threshold, the QV metric tracks the quality distribution of all the beads for a given cycle.</p> <p>A quality value is associated with beads that are in panels that do not align, or with beads that have no color call.</p> <p>The median QV of all calls is tracked and reported.</p> <p>In general, a higher QV is assigned when a bead is brighter, because the brighter the bead, the greater the differences between the primary and secondary color calls.</p>

Parameter	Description
Blur Value	<p>Given a 1 micron bead and the optics and detector characteristics of the signal acquisition system, a bead should have an apparent radius of approximately 2.2 microns.</p> <p>Deviations from this value are almost exclusively in the positive direction and can come from a number of sources.</p> <p>The most obvious is that of focus where the focal plane and the bead plane are not aligned, causing the bead to blur and grow larger.</p> <p>Another source of change can be in vibration of the sequencer or settling of the stage while the image is being acquired. These latter reasons are a motion blur while the first is a focal blur.</p> <p>For the purposes of determining the blur value for a given panel, a composite model of all beads in a region of a ligation image is created. The radius is just the center-to-edge measurement of that model. The deviation of this value from its nominal 2.2 is the metric being tracked. This is important because when the bead size becomes too large, the panel will not align, but when it aligns with a larger value, the region of the bead integrated to determine its intensity is too small.</p> <p>If only one channel of the three has a blur problem, the relative integrated signal for that channel is less accurate than the others, degrading the call quality.</p> <p>The blur value reported is the largest mean blur value of the four channels over all panels in one sample.</p>
Blur Population	<p>Using the blur metric discussed in Blur Value, rather than providing the average blur, this parameter reports the population of panels that have a blur value above a desired threshold.</p> <p>A larger than average size can be due to a small number of panels with large outlier values or a large number of panels with a larger size.</p> <p>The two values taken together can estimate whether there is an endemic problem over the whole sample (larger blur value but no outliers), a regional problem (an outlier population that spans channels), or a transient problem (outliers in a single channel).</p>

Parameter	Description
Exposure Value	<p>Typically the ideal exposure level of an image is one that maximizes the dynamic range of the content with some signal occupying the lowest (darkest) range and some information extending just to the saturation (highest) point. An overexposed image is one that compresses too much of the content at the top end of the range, creating a non-linearity in response and making it impossible to differentiate between the highest signals. An underexposed image is one that does not extend the content of the image across the possible range, thereby creating a low-end compression and failing to use the available range of the system.</p> <p>The calculated exposure value is a measure of how close the system is to capturing an image that maximizes dynamic range. Looking at the tail behavior of the signal distribution makes the determination. If the tail does not extend to the saturation point, the image is underexposed. In the case of overexposure, the tail has a peak at the far right side that is a measurement of the saturation. To get an estimate of the overexposure, the slope of the curve prior to the peak is extrapolated to a distance equal to the area of the saturated peak. The computed exposure quality metric is two-sided, extending from an ideal of 0 in the negative direction for underexposure and in the positive direction for overexposure and uses a log (2) scale. To fit this value into the traffic light model, the distance from ideal performance (not necessarily a value of 0) is used. Applied Biosystems has found that for purposes of our color calling algorithms, slightly overexposing the images yields better results.</p> <p>As with the blur value, the reported value is the largest exposure value of the four channels over all panels in the sample.</p>
Exposure Population	<p>As with the blur population, rather than providing the average exposure quality value, the exposure population value reports the number of panels that have an exposure quality beyond a desired threshold.</p> <p>A larger than average size can be due to a small number of panels with large outlier values or a large number of panels with a larger size.</p> <p>The two values, taken together, can estimate whether there is an endemic problem over the whole sample (larger deviation with no outliers), a regional problem (outlier population that spans channels), or a transient problem (outliers in a single channel).</p> <p>A difference from the blur metric, though, is that blur metric is invariant to the number of beads present in a panel where the exposure is more sensitive to the number of beads.</p> <p>A baseline exposure value per channel is determined based on a sampling strategy, but panels with either significantly more or fewer beads can be affected differently.</p>

Parameter	Description
Exposure Time	<p>In the SOLiD™ 4 System, every panel could technically have a different exposure value within a channel.</p> <p>This metric tracks the mean exposure time over all panels and reports the highest of the four channels. Increasing exposure times can indicate problems in the sequence.</p> <p>If the exposure time jumps too high too quickly, the signal on the beads is decreasing too quickly cycle over cycle, and there will likely be problems achieving longer read lengths.</p> <p>Vibration, or other issues, can also up the exposure time as the beads become larger due to blur and the exposures increase in order to get sufficient signal spread over a larger area.</p> <p>An increase in exposure time by itself is not a problem, but it is a warning sign that is valuable when taken into consideration with the other available run metrics to diagnose potential problems.</p> <p>Exposure times are ‘effective’ exposure times; they are a combination of both the exposure time and the gain.</p>

Application Descriptions

There are two types of applications available for the secondary analysis settings

- SingleTag
- IQ



SingleTag Pipeline Descriptions

These are the available SingleTag pipelines and their descriptions.

Pipeline	Description
Mapping	Used to find all hits between a set of reads and the reference sequence(s), based on the settings of anchor length, number of mismatches allowed, and anchor start.
MaToBAM	Writes a BAM file from a single-fragment match (.ma) file v4.0, and its associated quality value file, for SOLiD™ SETS Software.
Position Errors	Produces a file containing statistics on color mismatches vs read position. The output file has the format {primer.set}_positionErrors.txt.

IQ pipeline descriptions

These are the available IQ pipelines and their descriptions.

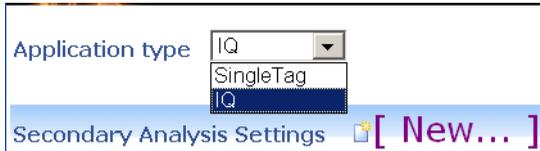
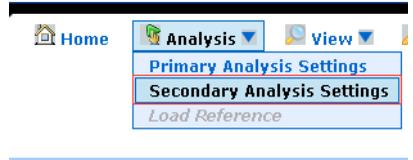
Pipeline	Description
IQ	Instrument qualification and troubleshooting. An independent pipeline, used by manufacturing and service, that is tied to a specific set of beads.

Define secondary analysis settings

In SETS, you can define the secondary analysis pipelines and parameters as a part of your setup for secondary analysis.

To modify any secondary analysis pipeline or parameter:

1. Log into SETS.
2. Go to **Main Menu**
 - ▶ **Analysis** ▶ **Secondary Analysis Settings**.
3. Select the Application type (**SingleTag** or **IQ**, as shown) to create a secondary settings.



4. Click **[New...]** to begin modifying the threshold settings for your application.



Note: Click  to delete an existing setting or  to edit one.



5. Provide a name for your new settings configuration, then select the reference genome to associate with this setting. See “[Load a Reference Sequence](#)” on page 37.

The screenshot shows the 'SingleTag Application Setting' dialog box. The 'Owner' is 'Administrator'. The 'Name *' field contains 'SingleTag_Administrator'. The 'Description' field is empty. The 'Reference genome' dropdown menu is set to 'DH10B_WithDup_FinalEdit_validated-sampled-1-sampled-1.fasta'. Below the fields is an 'Advanced parameters' section which is currently collapsed. At the bottom are 'Next' and 'Cancel' buttons. A note at the bottom left states '* = required fields'.

The screenshot shows the 'IQ Application Setting' dialog box. The 'Owner' is 'Administrator'. The 'Name *' field contains 'IQ_Administrator'. The 'Description' field is empty. The 'Reference genome' dropdown menu is set to 'DH10B_WithDup_FinalEdit_validated-sampled-1-sampled-1.fasta'. Below the fields is an 'Advanced parameters' section which is currently collapsed. At the bottom are 'Next' and 'Cancel' buttons. A note at the bottom left states '* = required fields'.

6. Click **Next** to view or select individual settings for each tag in a paired run.
Different settings can be applied to each tag for paired applications, as shown below.
7. Mouse over a pipeline’s name to see its description.

Secondary Analysis Pipelines for Setting 'SingleTag_Administrator'

Edit pipelines

On/Off	Name	Parameters
<input checked="" type="checkbox"/>	Mapping	View / Edit
<input checked="" type="checkbox"/>	MaToBAM	View / Edit
<input checked="" type="checkbox"/>	Position Errors	View / Edit

Restore Defaults

Save Back to Start Cancel

Mapping
The mapping pipeline in V3.5 is used to find all hits between a set of reads and the reference sequence(s), based on the settings of anchor length, number of mismatches allowed, and anchor start.

MaToBAM
Writes a BAM file from a single-fragment match (.ma) file, v3.5 and its associated quality value file for SOLiD3.5 software.

Position Errors
Makes a file containing statistics on color mismatches vs read position. The output file has the format {primer.set}_positionErrors.txt

Secondary Analysis Pipelines for Setting 'IQ_Administrator'

Edit pipelines

On/Off	Name	Parameters
<input checked="" type="checkbox"/>	IQ	View / Edit

Restore Defaults

Save Back to Start Cancel

IQ
Instrument qualification and trouble shooting. An independent pipeline used by manufacturing and service that is tied to a specific set of beads.

8. Click **View** to see the rollover summary of current settings.

Screenshot of the software interface showing the rollover summary for the Mapping pipeline. The 'View' link is circled, and a tooltip displays the following settings:

Clear zone	5
Maximum Hits	100
Minimum cutoff score for filtering	0
Mismatch Level	6
Mismatch penalty	-2.0
Run Classic Mapping	false

Note: Click to activate or deactivate a pipeline.

9. Click **Edit** to the right of a name to edit any active pipeline.

Position Errors for Setting SingleTag_Administrator

[Primer Set](#) any text

[Show first position as bases](#) a boolean (0 or 1)

[Finish](#) [Save and Edit Another Pipeline](#) [Restore Defaults](#) [Cancel](#)

[Primer Set](#)

The primer set label. For a mate pair based Gff3 file, should be 'F3,R3'.

[Show first position as bases](#)

For backward compatibility

10. Enter a new value or values, then click **Finish**. To edit more pipelines, click [Save and Edit Another Pipeline](#).

Note: Click to select show first position as bases.

IMPORTANT! Some pipelines are dependent upon others. When you select or deselect one, you may be affecting the activation state of the others.

Load a Reference Sequence

Specify a New Reference

If a specific reference sequence is not in the Reference Genome drop-down list (Secondary Analysis Settings view), you can load a new reference sequence before you start a run. Open the SETS software and perform steps 1 to 4 below to add a new reference sequence name to the list of choices you see in SETS.

Secondary analysis (sequence data alignments to a reference sequence) would not occur without a preloaded reference sequence.

IMPORTANT! For more complex genomes (such as the human genome), barcoded samples, and larger reference (and reads) the secondary analysis should be performed off-instrument with the BioScope™ Software. Refer to *BioScope™ Software for Scientists Guide: Data Analysis Methods and Interpretation* (PN 4448431).

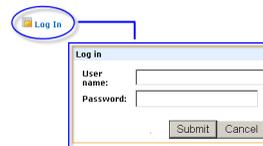
If a sample contains more than the maximum allowable number of reads, the SETS software selects only up to the maximum allowed number of reads and the rest will not be included in the results of the secondary analysis.

IMPORTANT! Do not install BioScope™ Software on the SOLiD™ 4 Analyzer.

The Secondary analysis (sequence data alignments to a reference sequence) will not occur without a preloaded reference sequence. Perform secondary and tertiary analysis with the BioScope™ Software. See [“Load a Reference Sequence” on page 37](#).

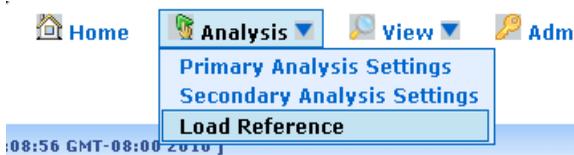
To upload a new reference sequence:

1. In the Home page log in as “Administrator.”



Note: The “Load reference” menu option under Analysis is not available unless you are logged in as an “Administrator.”

2. Click **Analysis** ▶ **Load Reference**.



The Load Reference Genome page opens.

Load Reference Genome?

The maximum file upload size is 2GB.?

Reference genome*

Name*

* = required fields

Note: For instructions on how to upload reference genomes larger than 2GB see [“To upload reference genome files larger than 2GB”](#) on page 39.

3. Click **Browse**, then select the appropriate fasta file for your reference sequence.

4. Enter a name for the selected file and click **Submit**.

The reference genome name that you enter in [step 4](#) appears in the Secondary Analysis Settings window in the drop-down list of reference genome.

IMPORTANT! After uploading the genome, the SETS Software validates the reference. The validation may take several minutes, during which the genome is not available. If validation fails, the new genome is not available in the drop-down list in the Secondary Analysis Settings and an error log is displayed in the homepage (under Error Messages).

To upload reference genome files larger than 2GB

Reference genome files larger than 2GB cannot be uploaded through the SETS Software due to the limitations set by browsers. To upload reference genomes larger than 2GB:

1. Use an FTP tool to first place the file inside the home directory, on the cluster.
2. Run the reference validation perl script to produce a validated reference.
3. Name the validated output and save it with the ".valid" extension.

Example

```
/share/apps/corona/bin/reference_validation.pl -r  
MY_REFERENCE.fasta -o MY_REFERENCE.fasta.valid
```

4. To ensure that the reference genome appears in the Reference genome box of the Secondary Analysis Settings list, move the reference genome file to the genomes folder, inside the reference folder.

Example

```
mv MY_REFERENCE.fasta.valid  
/share/reference/genomes/MY_REFERENCE.fasta
```

- From the Home page, navigate to the Secondary Analysis Settings and click **New**. Ensure that the reference genome file (MY_REFERENCE.fasta) appears in the Reference genome drop-down list, as shown in the example below.

The screenshot displays the "SingleTag Application Setting" form. The "Reference genome" field is set to "K12.fasta". A blue circle highlights the dropdown arrow next to "K12.fasta". A blue arrow points from this circle to a second screenshot below, which shows the dropdown menu expanded. The expanded menu lists "MY_REFERENCE.fasta" (circled in blue) and "K12.fasta". A blue circle also highlights the "Advanced parameters" field in the main form, with a blue arrow pointing to the expanded menu. A blue text annotation "To expand the 'Advanced Parameters'" is located to the right of the main form. The "Advanced parameters" field in the main form has a small "«" icon on its right side, which is also circled in blue. The "Advanced parameters" field in the expanded menu is empty. The "Reference genome" field in the expanded menu is also empty. The "Advanced parameters" field in the main form has a placeholder text "enter in this format: name1=value1;name2=value2, etc." at the bottom right. The "Next" and "Cancel" buttons are visible at the bottom of the main form. A note "* = required fields" is located at the bottom left of the main form.

SingleTag Application Setting

Owner Administrator

Name * SingleTag_Administrator

Description

Reference genome K12.fasta

Advanced parameters

enter in this format: name1=value1;name2=value2, etc.

Next Cancel

* = required fields

To expand the "Advanced Parameters"

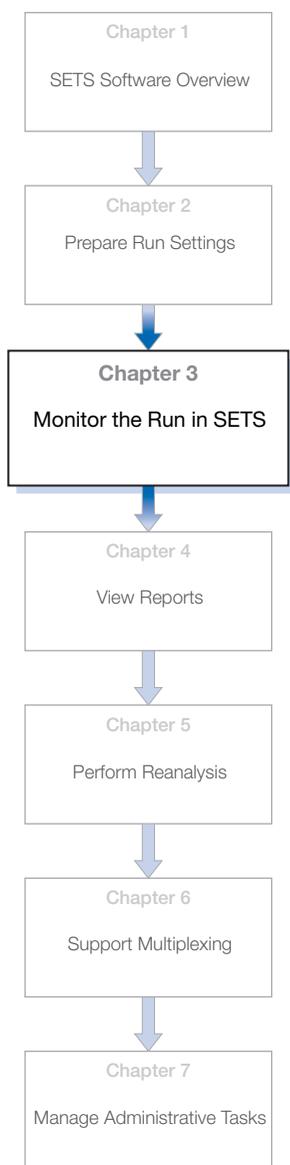
Reference genome MY_REFERENCE.fasta

Advanced parameters MY_REFERENCE.fasta

K12.fasta

3

Monitor the Run in SETS



This chapter covers:

- Overview 42
- View Run Metrics 42
- View Cycle Scans 46
- View sample data 49
- Troubleshoot failed jobs 51
- View History 57
- Cancel analysis midstream 58
- Cluster Status 60
- System Logs 61
- RSS feed 62

Overview

While executing a sequencing run, the instrument performs a Pre-scan. In a Pre-scan, all of the beads are labeled and their positions on the slide are recorded to derive a focal map followed by a single ligation cycle. The run can be remotely monitored in SETS by checking the Run Metrics, Cycle Scans, Heat Maps, and System Logs.

Note: To set up e-mail notification regarding instrument run and system information, see [Chapter 7, “Event notifications and email services”](#) on page 125.

View Run Metrics

View the Run Metrics

You can view a current run on the SOLiD™ 4 System and determine the quality of each cycle by looking at the traffic lights. The Run Metrics page in SETS displays all metrics per sample for the selected ligation cycle.

To view the quality of a run:

1. Open SETS and from the Run Details page, click **Run metrics**.

Run details: SR_20100118_FRAG [LastRendered: Wed Jan 20 19:44:41 GMT-08:00 2010]

SR_20100118_FRAG	Run metrics
Run Type	Fragment
Owner	SF
Run Protocol	SOLiD Opti
Run state	complete
Run started	01/18/2010 05:51 PM
Run completed	01/18/2010 06:00 PM
Spots	1
Panels per spot	1
Transfer status	

Actions

- Reanalyze Primary
- Reanalyze Secondary
- Delete
- Set Up Export...
- Reassign Run

Overall Reports

- Imaging Metrics Report
- Heat Map Report
- Cycle Heat Map Report
- Cycle Scans
- Exposure Time Report
- Flowcell Mask

Note: For in-progress runs, you can also access the Run metrics link from the homepage.

In-Progress or Most Recent Runs

FLOWCELL 1

r1ahead01_20100120_FRAG_2

Run Type	Fragment
Owner	sy
Run Protocol	SOLiD Opti
Run state	complete
Run started	01/20/2010 07:18 PM
Run completed	01/20/2010 07:29 PM
Spots	1
Panels per spot	1
Transfer status	

[Run metrics](#) [More information](#)

2. Select the primer set, primer, cycle, and version you want to monitor, then click **Display** to update the traffic lights.

Run Name: SR_20100118_FRAG

Sample1 - F3
SR_20100118_FRAG_F3_P5_04_V1

Status Item	Message	Threshold	Status
Panel Alignment	panel alignment of 100% >=	90%	● ● ●
Bead Retention	bead retention of 99% >=	95%	● ● ●
HQ Beads	HQ beads of 48% >=	35%	● ● ●
Color Balance	Color Deviation of fam of 9% >=	6%	● ● ●
Quality Value	median QV for sample of 22 >=	20.0	● ● ●
Blur Values	Blur Value of worst channel fam of 2.1626 <=	2.2	● ● ●
Blur Population	Blur Population of worst channel fam of 0% <=	75%	● ● ●
Exposure Values	Exposure Value of worst channel fam of 0.0 <=	0.05	● ● ●
Exposure Population	Exposure Population of worst channel fam of 0% <=	0%	● ● ●
Exposure Time	Exposure Time of worst channel cy3 of 402.0 <=	800.0	● ● ●

● - requires attention ● - might require attention ● - normal

Select Primer, Cycle, and Version, then click Display

3. Determine the quality of your run by looking at the traffic light indicators for in-progress runs.

Green - when a traffic light for a metric displays green, that parameter has a passing quality value and is hitting its target threshold.

Yellow - when a traffic light for a metric displays yellow, that parameter has a value that falls between the low and high threshold and is questionable in quality.

Red - when a traffic light for a metric displays red, that parameter has a value that is above the high threshold limit and the quality could be considered poor enough to fail; if a sample displays all red traffic lights during a run, you may want to abort that run immediately.

4. Click **Get summary** when you want a full summary of all the metrics for an entire run. **Get latest status** displays by default.

Run Name: solid0301_20081212_FC1_density_test_2x50

Run Name is displayed here

● - requires attention ● - might require attention ● - normal

120K - R3
Summary

Status Item	Message	Threshold	Status
Panel Alignment	panel alignment of 98% >	90%	● ● ●
Bead Retention	bead retention of 98% >	95%	● ● ●
HQ Beads	HQ beads of 23% >	15%	● ● ●
Color Balance	Color Deviation of tcr of 11% >	6%	● ● ●
Quality Value	median QV for sample of 8 <	15.0	● ● ●
Blur Values	Blur Value of worst channel cy5 of 2.1779 <	2.2	● ● ●
Blur Population	Blur Population of worst channel cy3 of 0.0 <	0.9	● ● ●
Exposure Values	Exposure Value of worst channel cy3 of 9.1167 >=	0.15	● ● ●
Exposure Population	Exposure Population of worst channel cy5 of 0.0 <	0.5	● ● ●
Exposure Time	Exposure Time of worst channel fam of 274.4790 <	800.0	● ● ●

Notice the traffic lights showing mixed quality results

140K - R3
Summary

Status Item	Message	Threshold	Status
Panel Alignment	panel alignment of 98% >	90%	● ● ●
Bead Retention	bead retention of 98% >	95%	● ● ●
HQ Beads	HQ beads of 24% >	15%	● ● ●
Color Balance	Color Deviation of tcr of 11% >	6%	● ● ●
Quality Value	median QV for sample of 8 <	15.0	● ● ●
Blur Values	Blur Value of worst channel cy5 of 2.2027 <	2.8	● ● ●
Exposure Values	Exposure Value of worst channel cy5 of 9.7198 >=	0.15	● ● ●
Exposure Population	Exposure Population of worst channel fam of 0.0 <	0.5	● ● ●
Exposure Time	Exposure Time of worst channel fam of 257.2840 <	800.0	● ● ●

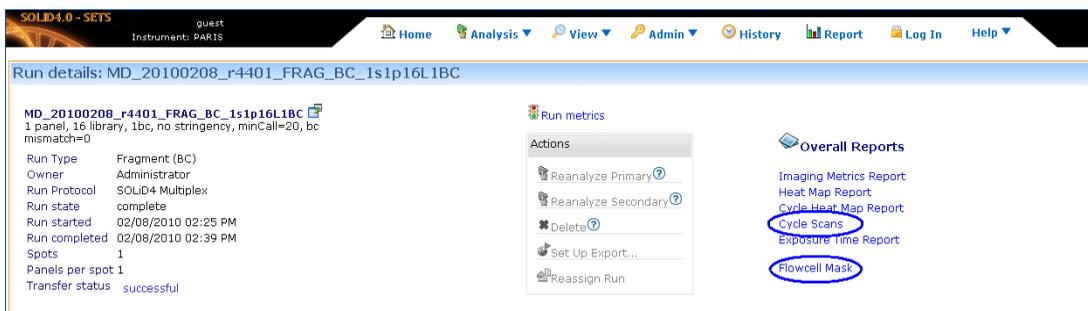
Note: Refer to the Imaging Metrics Report and the Exposure Time Report as additional resources when monitoring a run.

-  Overall Reports
- [Imaging Metrics Report](#)
 - [Heat Map Report](#)
 - [Cycle Heat Map Report](#)
 - [Cycle Scans](#)
 - [Exposure Time Report](#)
 - [Flowcell Mask](#)

View Cycle Scans

View Cycle Scans To view cycle scans of runs in progress or completed:

1. Go to **Run Details** ▶ **Overall Reports**.
2. Click **Cycle Scans** to see all the scans for the selected run.



The screenshot displays the SETS software interface for a specific run. The top navigation bar includes links for Home, Analysis, View, Admin, History, Report, Log In, and Help. The main content area is titled "Run details: MD_20100208_r4401_FRAG_BC_1s1p16L1BC".

On the left, the run details are listed:

- Run ID: MD_20100208_r4401_FRAG_BC_1s1p16L1BC
- 1 panel, 16 library, 1bc, no stringency, minCall=20, bc mismatch=0
- Run Type: Fragment (BC)
- Owner: Administrator
- Run Protocol: SOLiD4 Multiplex
- Run state: complete
- Run started: 02/08/2010 02:25 PM
- Run completed: 02/08/2010 02:39 PM
- Spots: 1
- Panels per spot: 1
- Transfer status: successful

In the center, the "Run metrics" section is visible, with an "Actions" menu containing the following options:

- Reanalyze Primary
- Reanalyze Secondary
- Delete
- Set Up Export...
- Reassign Run

On the right, the "Overall Reports" section lists several report types:

- Imaging Metrics Report
- Heat Map Report
- Cycle Heat Map Report
- Cycle Scans (circled in blue)
- Exposure Time Report
- Flowcell Mask (circled in blue)

SOLID4.0 - SETS guest
Instrument: PARIS

[Home](#) [Analysis](#) [View](#) [Admin](#) [History](#) [Report](#) [Log In](#) [Help](#)

[Back to Run page](#)

Cycle Scans

Run Name MD_20100208_r4401_FRAG_BC_1s1p16L1BC

Show used scans only [Bead Counts: Total | Average per panel] Exposure Times (ms)

Cycle	Scan		HM	Panels	Failed panels	Usable Beads (max=275,862)	Best Beads	Best+Good Beads	Bead Color Balance	Exposure Times (ms)			
	Ver.	Used								FAM	CY3	TXR	CY5
BC_PA_01	v.1	✓	🔴	1	0	-	38.8%	64.2%		-	-	-	-
BC_PB_01	v.1	✓	🔴	1	0	-	41.1%	65.1%		-	-	-	-
BC_PC_01	v.1	✓	🔴	1	0	-	46.3%	68%		-	-	-	-
BC_PD_01	v.1	✓	🔴	1	0	-	45.1%	67.3%		-	-	-	-
BC_PE_01	v.1	✓	🔴	1	0	-	48.5%	69.2%		-	-	-	-
F3_PA_01	v.1	✓	🔴	1	0	-	55.7%	69.4%		-	-	-	-
F3_PA_02	v.1	✓	🔴	1	0	-	45.4%	64.7%		-	-	-	-
F3_PA_03	v.1	✓	🔴	1	0	-	37.8%	59.9%		-	-	-	-
F3_PA_04	v.1	✓	🔴	1	0	-	31.3%	55.7%		-	-	-	-
F3_PA_05	v.1	✓	🔴	1	0	-	25.1%	49.7%		-	-	-	-
F3_PB_01	v.1	✓	🔴	1	0	-	53.9%	69.7%		-	-	-	-
F3_PB_02	v.1	✓	🔴	1	0	-	40.4%	62.9%		-	-	-	-
F3_PB_03	v.1	✓	🔴	1	0	-	33.1%	57.6%		-	-	-	-
F3_PB_04	v.1	✓	🔴	1	0	-	25.5%	50.5%		-	-	-	-
F3_PB_05	v.1	✓	🔴	1	0	-	21.8%	46.6%		-	-	-	-
F3_PC_01	v.1	✓	🔴	1	0	-	45.2%	64.3%		-	-	-	-
F3_PC_02	v.1	✓	🔴	1	0	-	34%	56.5%		-	-	-	-
F3_PC_03	v.1	✓	🔴	1	0	-	28.4%	52.4%		-	-	-	-
F3_PC_04	v.1	✓	🔴	1	0	-	22.9%	47.1%		-	-	-	-

A cycle can have multiple scan versions as indicated in the second column (Ver.), however, only one version is used in the latest analysis and is checked in the third column (Used).

To view cycle data:

1. Click on the version number hyperlink to display the cycle scan data for that particular version in a table at the bottom of the page.

SOLID4.0 - SETS guest
Instrument: PARIS

Cycle Scans

Run Name MD_20100208_r4401_FF

Cycle	Scan		Used	HM	Panels
	Ver.				
BC_PA_01	v.1		✓	🔥	1
BC_PA_01	v.1		✓	🔥	1
FS_PE_05	v.1		✓	🔥	1

BC_PA_01 v.1							Exposure Times (ms)				
Sample	Panels	Failed panels	Usable Beads	Best Beads	Best+Good Beads	Bead Color Balance	FAM	CY3	TXR	CYS	Plots
bcSample1	1	0	274,747	38.9%	64.5%		-	-	-	-	

Back to Run page

- Click **HM**  to open a Heat Map for any particular scan. Click **Close** when you have finished viewing the Heat Map.

SOLID4.0 - SETS guest
Instrument: PARIS

Cycle Scans

Run Name MD_20100208_r4401_FF

Cycle	Scan		Used	HM	Panels
	Ver.				
BC_PA_01	v.1		✓	🔥	1
BC_PA_01	v.1		✓	🔥	1

Run MD_20100208_r4401_FRAG_BC_1s1p16L1BC
Scan BC_PA_01_V1
Plot Type Best Beads (fixed)

- Click **Back to Run page** after you view cycle scans.

View sample data

Run Details Page The Run Details page displays all relevant Run Information that can be accessed by expanding individual samples in the sample tree in the left-hand panel of the page. Within each sample folder are analysis cycles that correspond to individual samples on a slide.

The screenshot displays the SETS software interface. On the left, a tree view shows a sample folder '1: bcSample1' expanded to 'auto-analysis', which contains a 'primary' folder. The right-hand panel is divided into sections: 'Sample: bcSample1' (Sample folder), 'Sample Reports' (Satay Report, Color Balance Report), 'Analysis: auto-analysis' (Displays the selected library), and 'Analysis Reports' (Quality Values Report, Auto-Correlation Report, Multiplexing Assignment Report). Below this, a 'Details for job: F3_PE_01_V1 (id=283)' panel shows a table of parameters:

Settings:	default primary(v.1)	cycle	1
Stage:	postScanSlidePrimary	export.dest	fozhdv10
Status:	finished	export.enabled	true
Created:	01/14/2010 04:58 PM	focal.map.dir	r1ahead01_20100114_FRAQ_BC/r1ahead01_20100114_FRAQ_BC_FOCALMAP_V1
Started:	01/14/2010 04:58 PM	focal.map.stg	r1ahead01_20100114_FRAQ_BC/r1ahead01_20100114_FRAQ_BC_FOCALMAP_V1
Completed:	01/14/2010 04:59 PM	focal.map.version	1
		images.dir.1	r1ahead01_20100114_FRAQ_BC/r1ahead01_20100114_FRAQ_BC_F3_P5_01_V1
		instrumentname	r1ahead01
		num.bases	5
		p1.channel	CV3
		p2.channel	2-2
		panels	2-2
		primer.id	5
		primer.set	F3
		probe.set	12
		run.name	r1ahead01_20100114_FRAQ_BC
		run.type	run
		sample.name	bcSample1
		spots	2
		version	1

Sample information

1. Click a sample in the sample tree to expand the folder to show general information about the sample, its analyses, and libraries. When you select an analysis or library, an Analysis/Library data panel displays in the lower half of the right-hand panel.
2. Click the sample folder drop-down arrow or the sample link [primary](#) to view individual analysis, or reanalysis cycles for each sample.

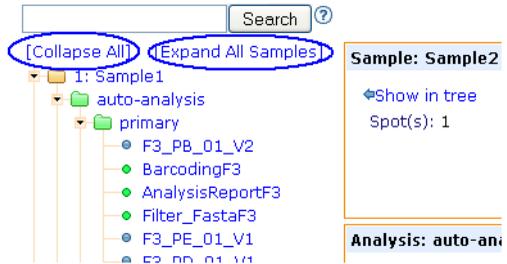
- Click one of the Overall Reports, or one of the Sample Reports (Satay or Color Balance Report) to view report data. For more information on reports, see [“View Reports” on page 63](#).

The sample tree shows the active pipelines as defined in your secondary settings.

Click **Collapse**

All to collapse all samples back to their original folder.

Click **Expand All Samples** to open all folders and view all samples at once.



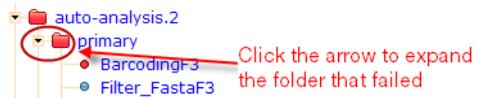
Note: If a pipeline has been inactivated from within the settings, it will not display.

Troubleshoot failed jobs

Sample analysis cycles

A blue folder indicates a completed analysis while a red folder indicates a failed analysis. To better understand why an analysis failed:

1. Click the arrow next to the analysis folder to view the analysis cycles for each sample (Run Details page).
2. Click the individual analysis job link to see details for that particular stage of analysis.



3. Select one of the job details for more information on each sample.



Job details **Beadfinding**

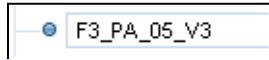
The beadfinding stage executes immediately after the Primer A cycle 1 scans have completed on-instrument.


Beadfinding

Details for job: F3_PA_05_V3 . (id=36661)		
Settings:	default primary(v.1)	cycle 1
Stage:	postScanSlidePrimary	focal.map.dir jm_1565/jm_1565_FOCALMAP_V1
Status:	finished	focal.map.stg jm_1565/jm_1565_FOCALMAP_V1/jm_1565_FOCALMAP_V1.STG
Created:	01/28/2010 03:11 PM	focal.map.version 1
Started:	01/28/2010 03:02 PM	images.dir.1 jm_1565/jm_1565_F3_PA_05_V3
Completed:	01/28/2010 03:02 PM	instrument.name R1AHEAD01
		num.bases 5
		p1.channel
		p2.channel CY3
		panels 2-2
		primer.id 2
		primer.set F3
		probe.set 12
		run.name jm_1565
		run.type run
		sample.name Sample1
		spots 2
		version 1

Colorcall Jobs

The job details (colorcall jobs) of the Post Scan Slide Primary stage are activated after each cycle is scanned. Incremental image analysis is called during this stage.



Details for job: F3_PA_05_V3 . (id=36661)

<p>Settings: default primary(v.1)</p> <p>Stage: postScanSlidePrimary</p> <p>Status: finished</p> <p>Created: 01/28/2010 03:11 PM</p> <p>Started: 01/28/2010 03:02 PM</p> <p>Completed: 01/28/2010 03:02 PM</p>	<p>cycle 1</p> <p>focal.map.dir jm_1565/jm_1565_FOCALMAP_V1</p> <p>focal.map.stg jm_1565/jm_1565_FOCALMAP_V1/jm_1565_FOCALMAP_V1.STG</p> <p>focal.map.version 1</p> <p>images.dir.1 jm_1565/jm_1565_F3_PA_05_V3</p> <p>instrument.name R1AHEAD01</p> <p>num.bases 5</p> <p>p1.channel</p> <p>p2.channel CY3</p> <p>panels 2-2</p> <p>primer.id 2</p> <p>primer.set F3</p> <p>probe.set 12</p> <p>run.name jm_1565</p> <p>run.type run</p> <p>sample.name Sample1</p> <p>spots 2</p> <p>version 1</p>
--	--

Filter Fasta

The Filter Fasta stage, Post Primer Set Primary, is performed after all cycles are imaged and the colorcalled data are ready for aggregation into final results.



Details for job: Filter_FastaF3 (id=36665)

Settings:	default primary(v.1)	focal.map.dir	jm_1565/jm_1565_FOCALMAP_V1
Stage:	postPrimerSetPrimary	focal.map.stg	jm_1565/jm_1565_FOCALMAP_V1/jm_1565_FOCALMAP_V1.STG
Status:	finished	focal.map.version	1
Created:	01/28/2009 03:13 PM	images.dir.1	jm_1565/jm_1565_F3_P2_01_V1
Started:	01/28/2009 03:04 PM	images.dir.2	jm_1565/jm_1565_F3_P1_01_V1
Completed:	01/28/2009 03:04 PM	images.dir.3	jm_1565/jm_1565_F3_P5_01_V1
		images.dir.4	jm_1565/jm_1565_F3_P4_01_V1
		images.dir.5	jm_1565/jm_1565_F3_P3_01_V1
		instrument.name	R1AHEAD01
		num.bases	5
		panels	2-2
		primer.set	F3
		probe.set	12
		read.length	5
		read.prefix	T
		run.name	jm_1565
		sample.name	Sample1
		spots	2

Filter Fasta fails

The Filter_Fasta job details will show a red icon indicating that the job has failed.

This problem can occur either at the Filter Fasta stage or at the Barcoding stage due to a bad hard-drive, a lack of disk space, or other system-related issues.

The software displays an error message and provides the Bead ID or Sequence number where the record is corrupted.

Example

Based on the original system-related issue, you will see an error message indicating that the “Fasta is empty” or the “Format is not right expecting sequence. Current line - >1_10_117_F3.”

Click **OK** to continue.

The error message will contain the problematic file’s location and record information.

To resolve, you need to fix any system-related issues first. Then, any of the following step may be applicable:

- Check the corrupted Fasta File manually; if possible fix it from the command line.
- Reanalyze primary analysis stage through SETS or manually by changing the database record/flag.
- If the SPCH files are also corrupted, report the error message to the Applied Biosystems Field Support Engineer.

Note: Report any critical error messages to the Applied Biosystems FSE.

Barcoding

IMPORTANT! Multiplexing reports can be created as soon as barcoding completes for a run.

Active only in multiplexing run; Barcoding job is performed after the FilterFasta job associated with each barcode tag. The barcoding job identifies and assigns each bead to a library. The output separates the other tag-based primary directories into library subdirectories within the Sample directory hierarchy.



Details for job: BarcodingF3 (id=36667)

Settings:	default primary(v.1)	focal.map.version	1
Stage:	postBarcodePrimary	instrument.name	R1AHEAD01
Status:	standby	is.barcode	0
Created:	01/28/2009 03:13 PM	panels	2-2
Started:		primer.set	F3
Completed:		run.name	jm_1565
		sample.name	Sample1
		spots	2

Analysis Report



[Show in tree](#)

Details for job: AnalysisReportF3 (id=596) Logs

Settings:	default primary(v.1)	focal.map.version	1
Stage:	analysisReportPrimary	instrument.name	r1ahead01
Status:	finished	is.barcode	0
Created:	01/26/2010 04:20 PM	panels	1-1
Started:	01/26/2010 04:22 PM	primer.set	F3
Completed:	01/26/2010 04:22 PM	read.length	5
		run.name	r1ahead01_20100126_FRAG_4
		sample.name	Sample1
		spots	1

View History

View Run History in SETS

1. Click **History** on the Main Menu to view the history of runs performed on the SOLiD™ 4 System.

Run name	Run Type	Samples	Libraries	Start date	Completion date	Results Dir	Images Dir
riahead01_20100120_FRAG_2	Fragment	1	1	01/20/2010 07:18 PM	01/20/2010 07:29 PM		
riahead01_20100120_FRAG	Fragment	1	1	01/20/2010 07:04 PM	01/20/2010 07:07 PM		
SR_20100120_WFA_WFA	WFA	1	1	01/20/2010 03:18 PM	01/20/2010 03:23 PM		
SR_20100120_WFA	WFA	1	1	01/20/2010 03:01 PM	01/20/2010 03:06 PM		
SR_20100118_FRAG	Fragment	1	1	01/18/2010 05:51 PM	01/18/2010 06:00 PM		
riahead01_20100114_FRAG_BC	Fragment (BC)	2	2	01/14/2010 04:50 PM	01/14/2010 04:58 PM		
riahead01_20100113_FRAG_8	Fragment	2	2	01/13/2010 05:27 PM	01/13/2010 05:44 PM		
riahead01_20100113_FRAG_6	Fragment	2	2	01/13/2010 03:59 PM	01/13/2010 04:02 PM		
riahead01_20100113_FRAG_4	Fragment	2	2	01/13/2010 03:26 PM	01/13/2010 03:29 PM		
riahead01_20100113_FRAG_BC_3	Fragment (BC)	2	2	01/13/2010 02:38 PM	01/13/2010 02:47 PM		
riahead01_20100113_FRAG_BC_2	Fragment (BC)	2	2	01/13/2010 01:55 PM	01/13/2010 02:04 PM		
riahead01_20100113_FRAG_3	Fragment	1	1	01/13/2010 01:44 PM	01/13/2010 01:47 PM		
riahead01_20100113_PE_BC	Paired End (BC)	2	2	01/13/2010 11:06 AM	01/13/2010 11:43 PM		
riahead01_20100113_FRAG_BC	Fragment (BC)	1	1	01/13/2010 10:54 AM	01/13/2010 11:03 AM		
riahead01_20100113_FRAG_2	Fragment	1	1	01/13/2010 10:16 AM	01/13/2010 10:19 AM		
riahead01_20100113_FRAG	Fragment	1	1	01/13/2010 09:18 AM	01/13/2010 09:29 AM		
riahead01_20100112_FRAG_5	Fragment	1	1	01/12/2010 05:44 PM	01/12/2010 05:47 PM		
riahead01_20100112_FRAG_4	Fragment	1	1	01/12/2010 05:42 PM	01/12/2010 05:44 PM		
riahead01_20100111_FRAG_stringency_test_01	Fragment	1	1	01/11/2010 04:09 PM	01/11/2010 04:12 PM	108.39 MB	88.01 MB
riahead01_20100112_FRAG_3	Fragment	2	2				
riahead01_20100113_FRAG_5	Fragment	1	1				
riahead01_20100112_FRAG	Fragment	2	2				
riahead01_20100112_FRAG_2	Fragment	2	2				
riahead01_20100113_FRAG_7	Fragment	2	2				
riahead01_20100114_WFA_WFA	WFA	1	1				
riahead01_20100113	Fragment	1	1				
riahead01_20100113_MP	Mate Pair	1	1				
riahead01_20100114_PE	Paired End	1	1				

A list of all runs displays by name and indicates the start and end date of each run.

2. Click **Export View** to see the Run Export History view.



3. Click **Recalculate Disk Space** when needed.

Note: Performing “Recalculate Disk Space” takes time to complete. Once started, disk space consumption (results data and image data) for all runs in Run History is calculated.

View Run Export History

The Run Export History view displays the runs that have been exported, as well as those runs currently being exported. This view shows the progress of the export and the specific location where each run was exported.

Run name	Start date	Completion date	Progress	Time Initiated	History
aaa_rithy	12/22/2008 11:36 AM	12/22/2008 11:42 AM	cancelled		
aaabbb_rithy	12/22/2008 02:49 PM	12/22/2008 02:52 PM	failed		
aabb_rithy	12/22/2008 03:00 PM	12/22/2008 03:15 PM	successful	12/22/2008 03:00 PM	Imported from 167.116.121.119 on 12/22/2008 03:04 PM
ACC14_CHAMPAGNE_20081021_ICsr21_BC_auto_export_clean_run	10/21/2008 02:09 PM		in progress	10/22/2008 02:16 PM	50% imported from 167.116.196.101
AUTOEXPORT_TEST_2	07/18/2008 11:27 AM	07/18/2008 11:35 AM	successful	07/23/2008 03:10 PM	Imported from 167.116.121.207 on 07/23/2008 03:04 PM
d_MP_20080707_1	07/07/2008 04:34 PM	07/07/2008 05:08 PM	successful	07/16/2008 01:18 PM	Imported from 167.116.122.181 on 07/16/2008 01:26 PM
e_MP_1_20080709_2	07/09/2008 02:53 PM	07/10/2008 11:34 AM	successful	12/18/2008 01:17 PM	Imported from 167.116.122.181 on 12/18/2008 01:24 PM

Cancel analysis midstream

Cancel in-progress analyses

You can cancel an analysis from the Run Details view in SETS. Only analyses in-progress are available for cancellation. The analysis will display as an active link if the analysis is in-progress and the run state is complete (no chemistry is running).

1. Click **Cancel analysis** for the selected run.

A complete list of all cancellable analyses displays.

Note: If **Cancel analysis** is disabled, it is possible that the run is no longer in-progress. Mouse over to see the explanation.

Sample: Sample1

Show in tree

Spot(s): 2

Sample Reports

Satay Report

Color Balance Report

Analysis: auto-analysis.3 - Library: defaultLibrary

Show in tree

Cancel analysis

Description:

Date created: 09/06/2009 03:05 PM

Last completed job: Barcoding

Data Files

Matching Result in Gif format

2. Select the analysis to cancel, then click **Submit**.
3. Cancel more than one analysis at a time from the list by clicking **Check all**, then **Submit**.
 - All the jobs for selected analyses will display gray (cancelled).
 - Any job still processing in the queue will be stopped and removed from the queue.

Cancel Analysis - r1ahead01_20090906_4

Sample1

<input type="checkbox"/> auto-analysis
<input checked="" type="checkbox"/> auto-analysis.2
<input type="checkbox"/> auto-analysis.3

[Check all](#) | [Uncheck all](#)

Cluster Status

Use the Cluster Status view to monitor the available disk space and progress of a current analysis job, as well as jobs that are scheduled to run. Specifically, the Cluster Status view shows:

- **Average Load (5 min)** – Average computing work load for each 5 minutes. If this is always 95%±, then the cluster may be overloaded or a node may be down.
- **Free Disk Space** – Available disk space for images (/data/images) and available disk space for data (/data/results). For example, 1.27 TB/ 8.86 TB (14.38%) (41,773 cycle-panels) indicates there are 1.27 TB of disk space out of 8.86 TB total space available to store images, and the disk is able to store up to 41,773 more cycle-panels.
- **Current Analysis Job Progress** – Displays which jobs are currently in progress and their completion percentage.
- **Job Queue Statistics** – Information about the job queues on the cluster. Search for blocked jobs.

Cluster Status [lastRendered: Mon Jan 26 14:51:59 PST 2009]

Cluster name: solid0308.sequencer
 CPUs total: 32
 Compute nodes up (ganglia): 3
 Avg load (5min): 14.28%
 Free disk space (images): 1.27 TB / 8.86 TB (14.38%) (41,773 cycle-panels)
 Free disk space (results): 1.63 TB / 3.93 TB (41.43%)

Current analysis jobs

Job Name	Job Id	Date Started	Progress	Analysis Name	Run Name	Sample Name
F3_P5_07_V1	743	01/26/2009 01:36 PM	<div style="width: 100%; height: 10px; background-color: green;"></div>	auto-analysis.2	solid0308_20090109_DH10B_2x50_FC1	Sample1
F3_P5_06_V1	742	01/26/2009 12:08 PM	<div style="width: 100%; height: 10px; background-color: green;"></div>	auto-analysis.2	solid0308_20090109_DH10B_2x50_FC2	Sample1

Queue statistics

2 of 3 Nodes Active (66.67%)
 Total Jobs: 4
 Active Jobs: 2
 Idle Jobs: 0
 Blocked Jobs: 2

Queue	# of running jobs	# of queued jobs
pipeline	0	0
tracking	0	2
feedback	2	0
batch	0	0
reanalyze	0	0
interactive	0	0
secondary	0	0

System Logs

The System Log View (**Admin ▶ System View**) allows you to view error logs on specific software components and to check the severity and the date the problem occurred.

Make your selections, then click **Search** or **Display Last 24 Hours**.

The screenshot shows the SETS System Log View interface. At the top, there is a navigation bar with links for Home, Analysis, View, Admin, History, Report, Log In, and Help. Below the navigation bar, there are search filters for Criteria Search and Text Search. The filters include:

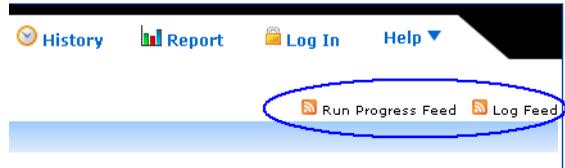
- Severity:** fatal, error, warn, info, debug (all checked). Buttons: Check all, Uncheck all.
- Component:** secondary, primary, instrument, daemon, sets (all checked). Buttons: Check all, Uncheck all.
- Clear Flag:** On, Off, Both (Both selected).
- Date:** All, Today, Last week, Specific dates: from 01/20/2010 to 01/20/2010. Format: (MM/DD/YYYY).

Below the filters is a Search button. The main area displays a table of log entries:

Date Created	Message	Severity	Component	Details	
01/20/2010 08:07 PM	Folder does not exist	info	sets (4136)	Cannot find folder/data/results/r1ahead01/r1ahead-01_20100120_FRAG/Sa More...	✖
01/20/2010 08:07 PM	Folder does not exist	info	sets (4135)	Cannot find folder/data/results/r1ahead01/r1ahead-01_20100120_FRAG/Sa More...	✖
01/20/2010 08:07 PM	DashboardBean cannot get samples for run 26	debug	sets (4134)	java.lang.NullPointerException at com.apldbio.aga.sets.view.DashboardB More...	✖
01/20/2010 08:07 PM	DashboardBean cannot get samples for run 26	debug	sets (4133)	java.lang.NullPointerException at com.apldbio.aga.sets.view.DashboardB More...	✖
01/20/2010 07:50 PM	Folder does not exist	info	sets (4132)	Cannot find folder/data/results/r1ahead01/SR_2010-0118_FRAG/Sample1/r More...	✖
01/20/2010 07:50 PM	Folder does not exist	info	sets (4131)	Cannot find folder/data/results/r1ahead01/SR_2010-0118_FRAG/Sample1/r More...	✖
01/20/2010 07:45 PM	Folder does not exist	info	sets (4130)	Cannot find folder/data/results/r1ahead01/SR_2010-0118_FRAG/Sample1/r More...	✖
01/20/2010 07:45 PM	Folder does not exist	info	sets (4129)	Cannot find folder/data/results/r1ahead01/SR_2010-0118_FRAG/Sample1/r More...	✖

RSS feed

You can observe real-time data from a current, in-progress run and/or from instrument logs by using any available RSS feed reader (Internet Explorer 7.0, Firefox, Sage, etc.). The feed reader links [Run Progress Feed](#) [Log Feed](#) are located on the right side of the page above the “In-progress or most recent runs” section on the homepage.

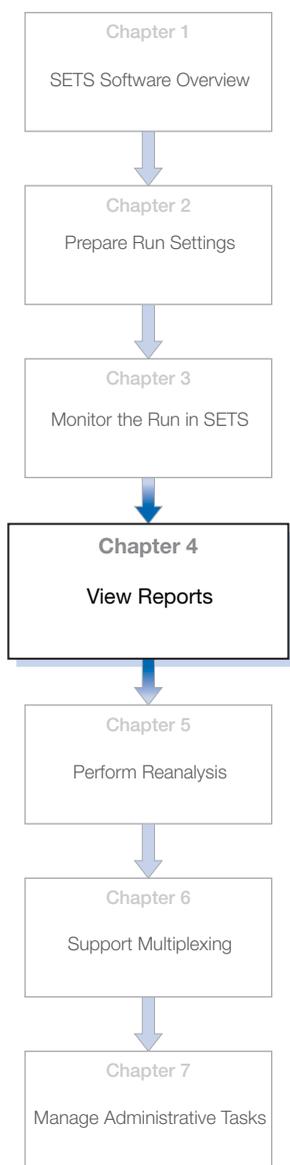


To set up the RSS feed reader:

1. Click the appropriate Feed button: [Run Progress Feed](#) [Log Feed](#).
2. Copy the web address for each feed, then paste it into the appropriate location in your browser.

4

View Reports



This chapter covers:

- View overall run reports 64
- View sample reports 77
- View Analysis Reports 85
- Master Report Tool 93

View overall run reports

You can view information about a completed run by opening one of the overall run reports:

- Imaging Metrics
- Heat Map Report
- Cycle Heat Map Report
- Cycle Scans (View)
- Exposure Time
- Flowcell Mask
- WFA Report (will display only when applicable)

 **Overall Reports**[WFA Report](#)[Imaging Metrics Report](#)[Heat Map Report](#)[Cycle Heat Map Report](#)[Cycle Scans](#)[Exposure Time Report](#)[Flowcell Mask](#)

Note: Most reports can be viewed in both real-time and completed run data.

Imaging Metrics Report

The Imaging Metrics report displays the cycle scan heat map view for all four channels (FAM™, CY3®, Texas Red®, and CY5® dyes) for the selected metric type, (for example, exposure).

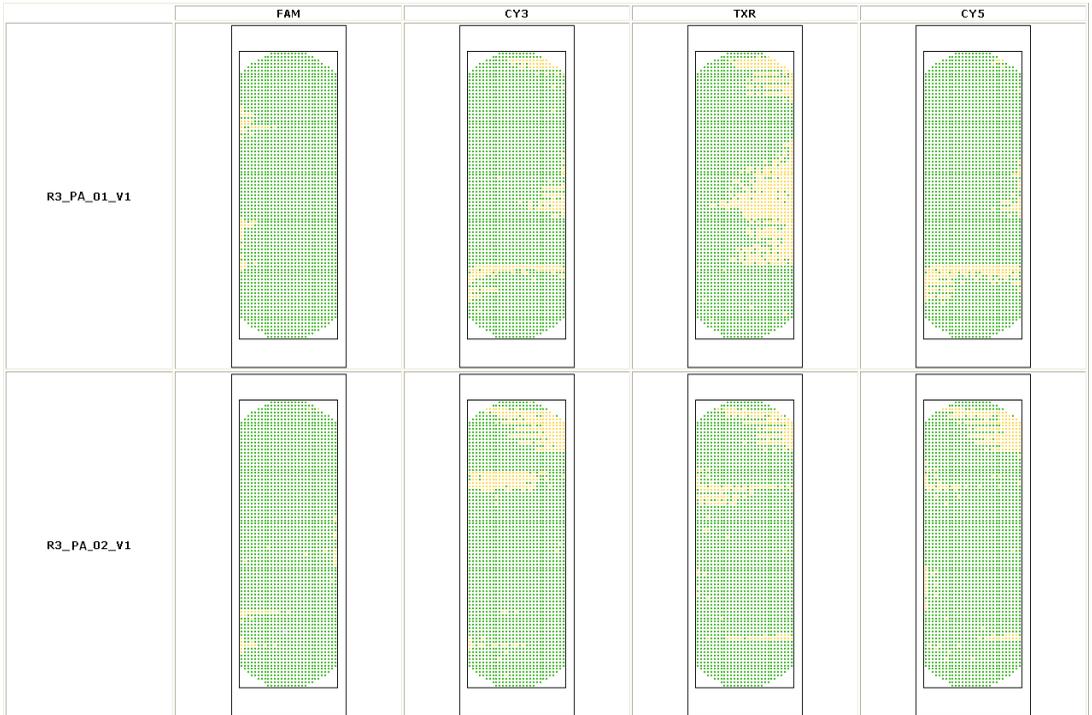
Definitions of all the metric types can be found in “Primary analysis parameters” on page 25.

Click **Imaging Metrics Report (Run Details ▶ Overall Reports)**.

Imaging Metrics Report

Run: solid0308_20090109_DH10B_2x50_FC1 Fixed Scale: Primer Set: R3
 Display Type: Exposure

Flowcell Mask



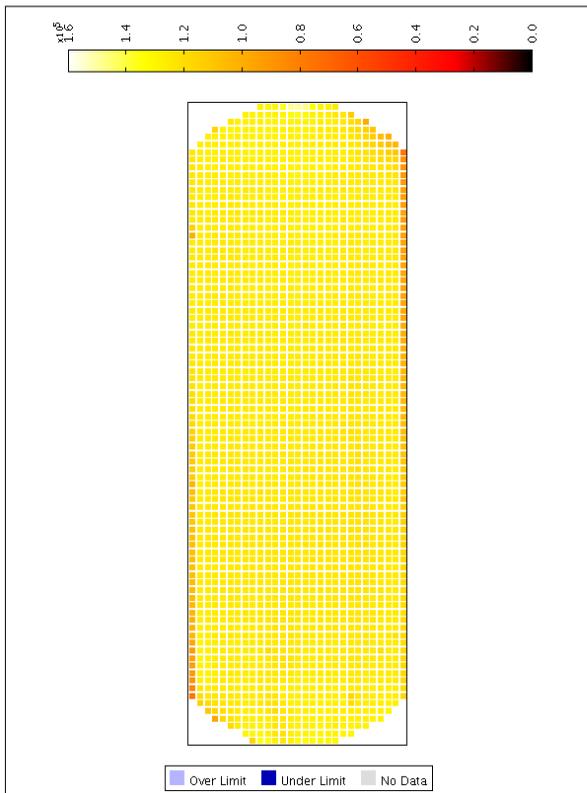
Heat Map Report Heat maps are generated from the results of the analysis of the focal map images. Heat maps give a colorized view of the selected metric across all panels for a run. Use this report to view experimental data mapped to the physical layout of the flowcell slide.

Click **Heat Map Report (Run Details ▶ Overall Reports)**.

Heat Map Report

Run solid0308_20090109_DH10B_2x50_FC1 Fixed Scale: Type: Bead Count Display

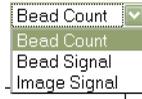
Flowcell Mask



Look for uniform deposition of beads on the slide. The actual average bead deposition density/panel value should be similar in value to the targeted average bead deposition density. A large number of missing panels could indicate a deposition problem.

Note: Heat Map Reports are available after beadfinding occurs.

Select a heat map type using the drop-down menu:

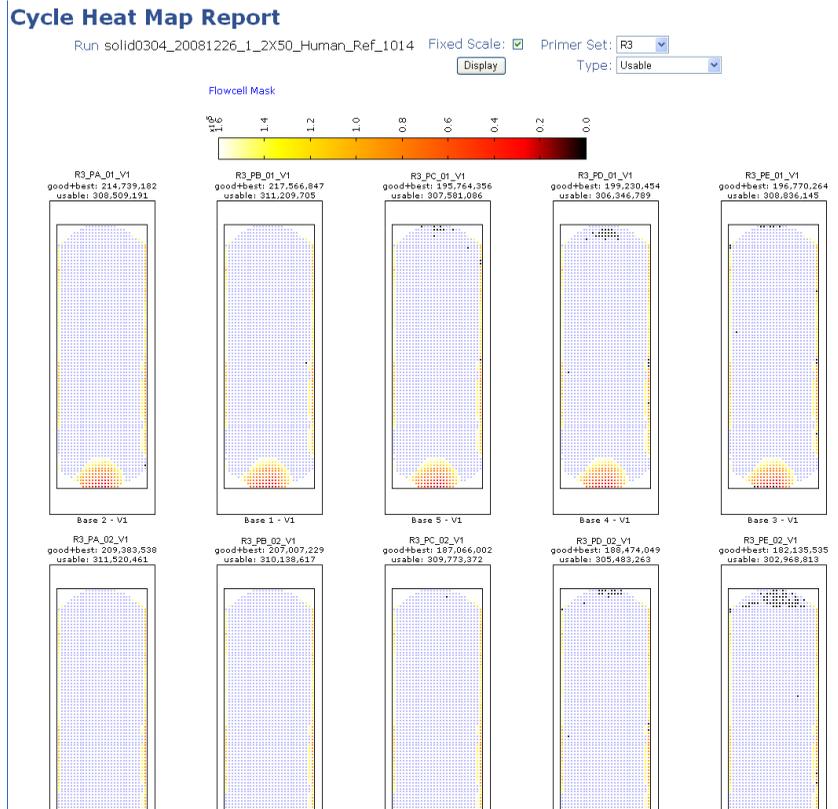


- **Bead Count** - number of beads per panel (Focal Map)
- **Bead Signal** - average signal intensity of the beads in a panel
- **Image Signal** - overall image signal intensity for a panel

Cycle Heat Map Report

This report displays the Cycle Scan Heat Map view for selected beads. See [“Bead definitions” on page 68](#).

Click **Cycle Heat Map Report (Overall Reports)**.



Bead definitions

Bead Type	Definition
Usable Beads	<p>Beads that, in a particular ligation cycle, are mapped in each of the four images. Typically, most beads in a panel are usable.</p> <p>The ones that are not lie near the edge of the panel, which for current purposes is defined as the area imaged by the corresponding reference or focal map image.</p> <p>The ligation fluorescence images must be shifted a few to tens of pixels to register them with the reference image. If, for example, a ligation image must be shifted 10 pixels to the left to register with the reference image, beads within 10 pixels of the right edge of the panel will not be mapped in that ligation image.</p> <p>Another common case is a panel-cycle where one of the ligation images fails to register. In this case, none of the beads are usable.</p>
Good Beads	<p>Usable beads that satisfy all of the following conditions:</p> <p>intensity >0.05</p> <p>$0.0125 < \text{angle} \leq 0.05$</p>
Best Beads	<p>Usable beads that satisfy all of the following conditions:</p> <p>intensity >0.05</p> <p>angle ≤ 0.0125</p>

Cycle Scans view Use the Cycle Scans window to distinguish normal runs from problematic runs.

Click **Cycle Scans (Run Details ▶ Overall Reports)**.

Cycle Scans
Run Name MD_20100208_r4401_FRAG_BC_1s1p16L1BC

Show used scans only [Bead Counts: Total | Average per panel]

Cycle	Scan			Panels	Failed panels	Usable Beads (max=275,862)	Best Beads	Best+Good Beads	Bead Color Balance	Exposure Times (ms)			
	Ver.	Used	HM							FAM	CY3	TXR	CYS
BC_PA_01	v.1	✓	🔒	1	0	-	38.8%	64.2%		-	-	-	-
BC_PB_01	v.1	✓	🔒	1	0	-	41.1%	65.1%		-	-	-	-
BC_PC_01	v.1	✓	🔒	1	0	-	46.3%	68%		-	-	-	-
BC_PD_01	v.1	✓	🔒	1	0	-	45.1%	67.3%		-	-	-	-
BC_PE_01	v.1	✓	🔒	1	0	-	48.5%	69.2%		-	-	-	-
F3_PA_01	v.1	✓	🔒	1	0	-	55.7%	69.4%		-	-	-	-
F3_PA_02	v.1	✓	🔒	1	0	-	45.4%	64.7%		-	-	-	-
F3_PA_03	v.1	✓	🔒	1	0	-	37.8%	59.9%		-	-	-	-
F3_PA_04	v.1	✓	🔒	1	0	-	31.3%	55.7%		-	-	-	-
F3_PA_05	v.1	✓	🔒	1	0	-	25.1%	49.7%		-	-	-	-
F3_PB_01	v.1	✓	🔒	1	0	-	53.9%	69.7%		-	-	-	-
F3_PB_02	v.1	✓	🔒	1	0	-	40.4%	62.9%		-	-	-	-
F3_PB_03	v.1	✓	🔒	1	0	-	33.1%	57.6%		-	-	-	-
F3_PB_04	v.1	✓	🔒	1	0	-	25.5%	50.5%		-	-	-	-
F3_PB_05	v.1	✓	🔒	1	0	-	21.8%	46.6%		-	-	-	-
F3_PC_01	v.1	✓	🔒	1	0	-	45.2%	64.3%		-	-	-	-
F3_PC_02	v.1	✓	🔒	1	0	-	34%	56.5%		-	-	-	-
F3_PC_03	v.1	✓	🔒	1	0	-	28.4%	52.4%		-	-	-	-
F3_PC_04	v.1	✓	🔒	1	0	-	22.9%	47.1%		-	-	-	-

The top section of the Cycle Scans window lists all the scans per ligation cycle for the slide, and the bottom section shows scan information sorted by sample for the scan selected in the top section. Use the parameters shown in the table on the next page to assess the progress of the sequencing run.

See “View Cycle Scans” on page 46 for more details.

Parameter	Normal Run	Problematic Run
Failed panels	In general, the number of failed panels should be relatively small and consistent.	Run begins with extremely high number of failed panels or dramatic increase in any subsequent ligation cycle for each sequencing primer.
Fraction of Best + Good beads/Usable beads	The fraction can vary depending on the quality of the library, the efficiency of the PCR, and the enrichment process. As a guideline, the fraction is around 0.5 to 0.6 in the first ligation cycle of each primer and drops to 0.2 to 0.3 in the last cycle.	A significant drop in the fraction of good beads in the initial ligation cycles is a reason to pause the run and troubleshoot the performance.
Effective exposure times	Gradual increase from ligation cycles 1 to 5 or higher for each sequence primer. Performance varies from slide to slide and as a function of the age of the SOLiD™ Light Source. As a guideline, the effective exposure time is typically 20 to 40 ms in the first ligation cycle and increases to 100 to 300 ms in the fifth cycle.	250 ms or greater in ligation cycle 1 or when instrument times out when the effective exposure time exceeds 500 ms. Long exposure times may indicate replacement of the SOLiD™ Light Source.
Satay plots	The first cycle of any primer should show a relatively “clean” Satay plot, with most points clustered on the four color axes and with minimal fraction of the points clustered around the origin. The quality of the Satay plot typically degrades gradually with each ligation cycle for a single primer cycle, becoming more “fuzzy” in the last cycle.	An abnormal “fuzzy” Satay plot in the first cycle is a reason to pause the run and troubleshoot the performance.

Exposure Time Report

The Exposure Time report displays the average exposure time of run cycle scans for each of the four channels (6-FAM[™], Cy3[®], Texas Red[®], and Cy5[®] dyes).

In this report, you can:

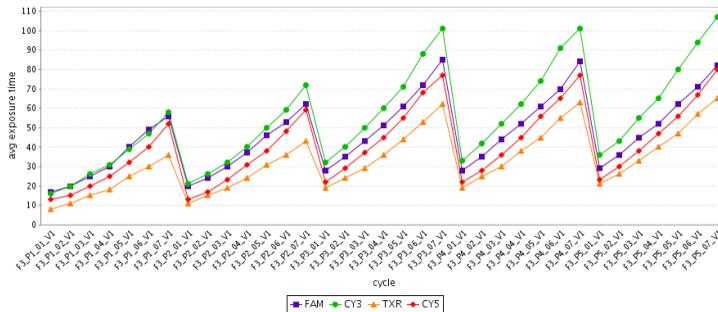
- Mouse over any data point to see the called value.
- Export the Average Exposure Time table in a CSV format.

Click **Exposure Time Report (Run Details ▶ Overall Reports)**.

Exposure Time Report

[Back to Run page](#)

Run solid0419_20090908_DH108_DH20_1x35_2

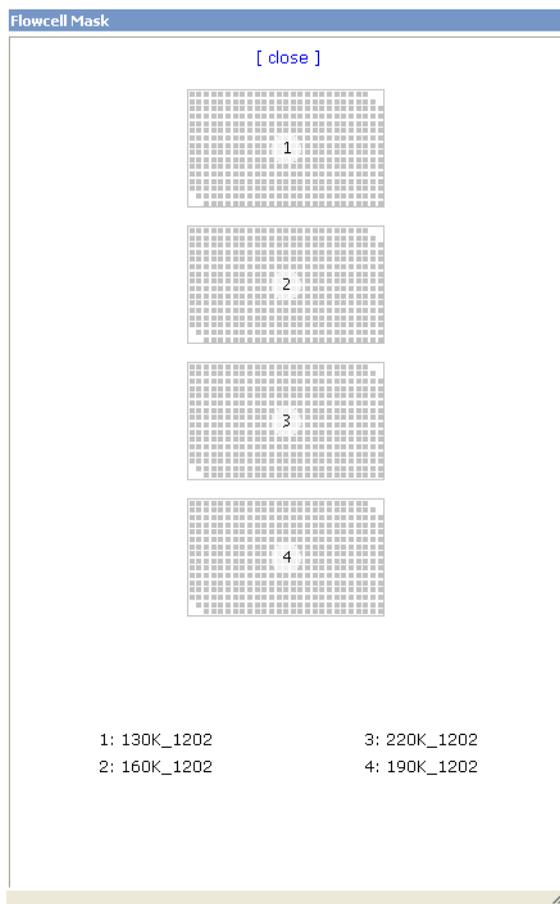


[Export to CSV file](#)

Average Exposure Time Table

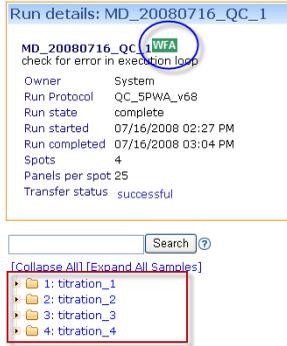
Scan	Scan Time	FAM Exposure	CY3 Exposure	TXR Exposure	CY5 Exposure
F3_P1_01_V1	09/09/2009 01:03 AM	17	16	8	13
F3_P1_02_V1	09/09/2009 03:45 AM	20	20	11	15
F3_P1_03_V1	09/09/2009 06:28 AM	25	26	15	20
F3_P1_04_V1	09/09/2009 09:10 AM	30	31	18	25
F3_P1_05_V1	09/09/2009 11:52 AM	40	39	25	32
F3_P1_06_V1	09/09/2009 02:35 PM	49	47	30	40
F3_P1_07_V1	09/09/2009 05:22 PM	56	58	36	52
F3_P2_01_V1	09/09/2009 08:33 PM	20	21	11	13
F3_P2_02_V1	09/09/2009 11:18 PM	24	26	15	17
F3_P2_03_V1	09/10/2009 02:10 AM	30	32	19	23
F3_P2_04_V1	09/10/2009 04:52 AM	37	40	24	31
F3_P2_05_V1	09/10/2009 07:34 AM	46	50	31	38
F3_P2_06_V1	09/10/2009 10:17 AM	53	59	36	48
F3_P2_07_V1	09/10/2009 12:53 PM	62	72	43	59
F3_P3_01_V1	09/10/2009 05:41 PM	28	32	19	22
F3_P3_02_V1	09/10/2009 08:24 PM	35	40	24	29
F3_P3_03_V1	09/10/2009 11:19 PM	43	50	29	37

Flowcell Mask view Use the Flowcell Mask report to view a representation of the sample slide. Each numbered area on the flowcell mask is a spot that corresponds to a specific sample location on the slide.



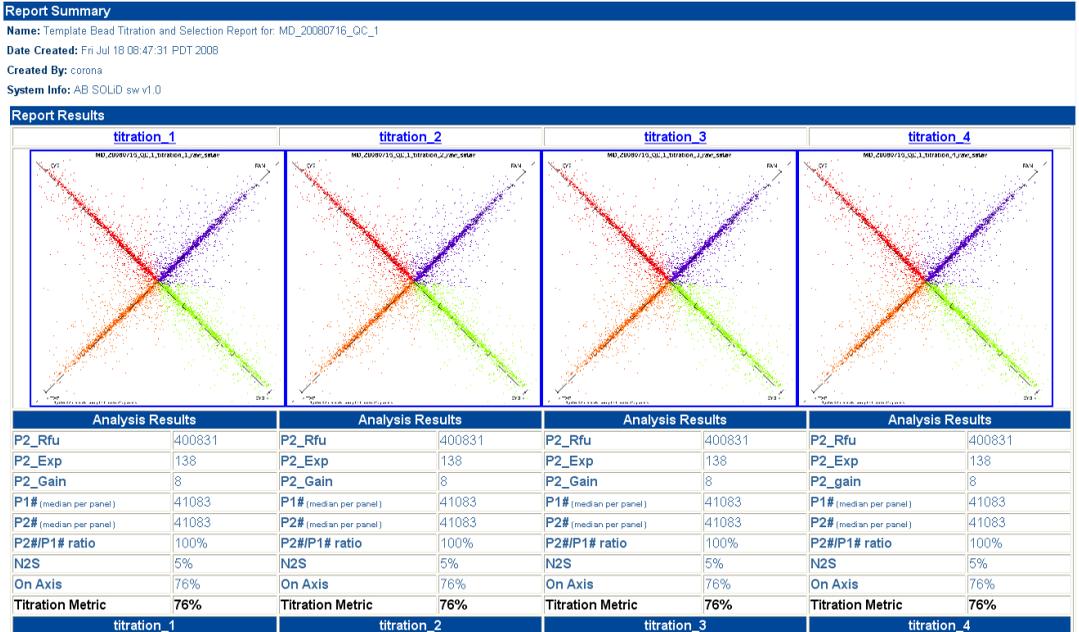
Note: The numbered areas on the Flowcell Mask also correspond to the numbered sample names in the sample tree, which appears at the bottom left of the Run Details page.

WFA report A WFA run on the SOLiD™ 4 System is uniquely labelled with a green WFA stamp and titration sample folders. When there is a valid WFA run, then the WFA report will be accessible in the Run Details page. Click **WFA Report (Overall Reports)**.

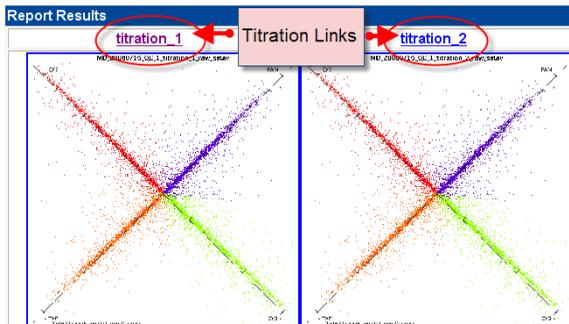


- Overall Reports**
- WFA Report
 - Imaging Metrics Report
 - Heat Map Report
 - Cycle Heat Map Report
 - Cycle Scans
 - Exposure Time Report
 - Flowcell Mask

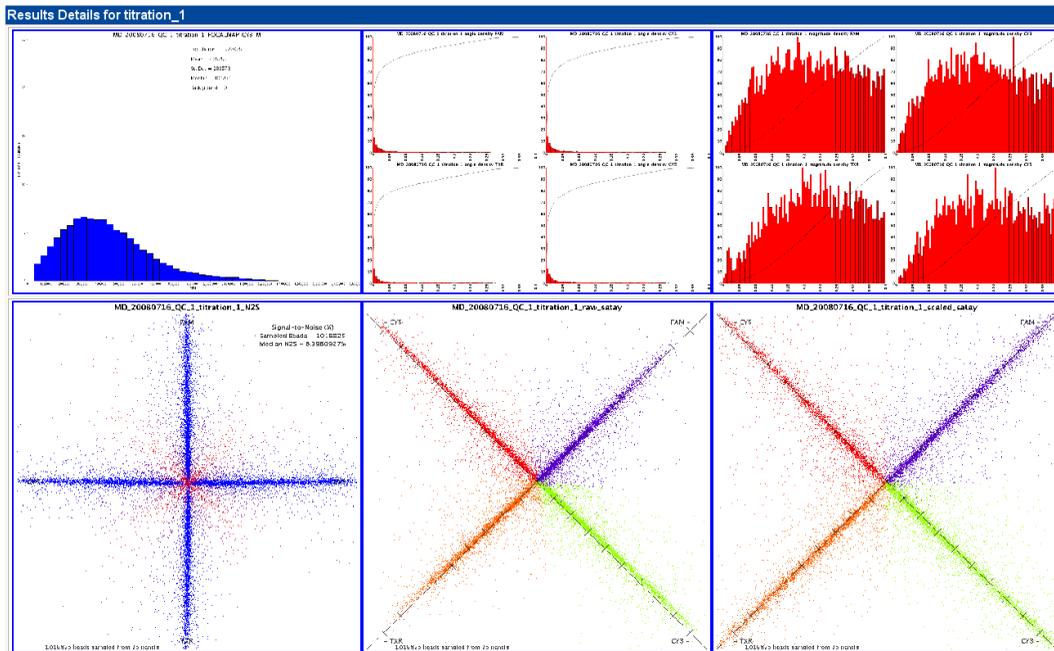
This example shows one WFA report for a single 4-well slide.



Click on individual titration links for more detail.



The Results Details for each titration opens to view.



Determine the optimal titration point

Determine which titration has the highest titration metric. This titration is the optimal titration point.

Determine the bead deposition density for a sequencing run

1. Calculate the concentration of P2-positive beads using the formula below, where X is the volume of templated beads used for the WFA sample (equivalent to 15 million beads) and P2# is given by the WFA report:

$$X \mu\text{L} = \frac{15 \times 10^6 \text{ beads}}{\# \text{ beads}/\mu\text{L} \text{ (according to NanoDrop®)}}$$

$$Y \text{ P2-positive beads}/\mu\text{L} = \frac{\text{P2\# beads/panel} \times 426 \text{ panels}}{X \mu\text{L}}$$

Example:

For a sample with a concentration of 500,000 beads/ μL measured by NanoDrop® ND-1000, where the WFA report indicates a P2# value of 20,000 beads/panel.

$$X \mu\text{L} = \frac{15 \times 10^6 \text{ beads}}{500,000 \text{ beads}/\mu\text{L}} = 30 \mu\text{L}$$

$$Y \text{ P2-positive beads}/\mu\text{L} = \frac{20,000 \text{ beads/panel} \times 426 \text{ panels}}{X \mu\text{L}}$$

$$= 284,000 \text{ P2-positive beads}/\mu\text{L}$$

2. Use the calculated concentration in place of the value determined by the NanoDrop® ND-1000 Spectrophotometer for more accurate deposition densities when preparing slides for sequencing. It is recommended to use this resulting calculated bead concentration to determine the volume of beads for deposition and multiply that volume by 120%:

$$Z \text{ } \mu\text{L bead solution} = \frac{\text{Desired \#P2-positive beads to be deposited}}{Y \text{ P2-positive beads}/\mu\text{L}}$$

Example:

For a sample with a concentration of 284,000 P2-positive beads/ μL to be deposited in one well of an 8-well SOLiD™ Deposition Chamber at a density of 300,000 beads/panel.

$$Z \text{ } \mu\text{L bead solution} = \frac{56 \text{ million P2-positive beads}}{284,000 \text{ P2-positive beads}/\mu\text{L}} \times 120\%$$

= 237 μL bead solution

View sample reports



Satay Reports

Use this report to view analysis cycle signals, which are converted into X,Y plots. For each cycle of sequencing, these plots show the relative intensities of the signals for each bead and how well separated the signal is across the four dye channels.

Each data point is an individual color call (i.e., an individual bead). Because total data points can exceed 20 million, only a subset of randomly selected points is displayed for each primer set (F3 and R3) and for various types of data (raw, scaled, angled, intensity, and N2S).

IMPORTANT! A satay report that collapses towards the origin or drifts away from the axes often indicates low signals or multiple DNA templates. If you observe such a satay report in the *early* cycles, consider aborting the run.

1. Click **Satay Report (Run Details ▶ Overall Reports)**.
2. Select the Primer Set and Data type of interest:
 - Raw – “Raw data” on page 80
 - Scaled – “Scaled data” on page 81
 - N2S – “N2S data” on page 82
 - Angle – “Angle data” on page 83
 - Intensity – “Intensity data” on page 83

3. Select the Dye Channel of interest:
 - 6-FAM™
 - Cy3®
 - Texas Red®
 - Cy5®
4. Click **Display**. The window opens to display a thumbnail plot for each cycle.

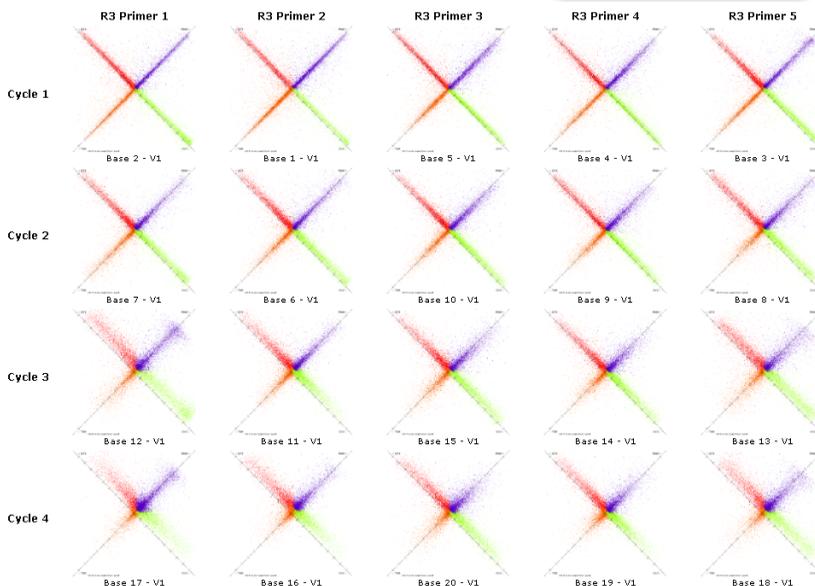
Satay Reports

Run MD_20081126_r23_MP
Sample Sample1

Primer Set: Channel:

Data:

Select the Primer Set and Data type to display



5. Click each thumbnail to see a larger, more detailed view.
6. In each data window, scroll down to view graphs for all cycles.

Resizing images

When you view the Satay images in SETS, your browser may automatically resize the large image to fit the screen. You may need to make an extra mouse click to view the full size image. The following instructions allow you to turn the browser's automatic image resizing function ON or OFF.

To resize image window in Internet Explorer:

1. On the Tools menu, click **Internet Options**.
2. Click the **Advanced** tab, then scroll down to the multimedia section.
3. Select or clear the **Enable Automatic Image Resizing** check box.

To resize image window in Firefox:

1. In the Firefox address bar, type *about:config*.
2. In the filter bar, type *browser.enable*.
3. Double-click *browser.enable_automatic_image_resizing*. The value changes to False.

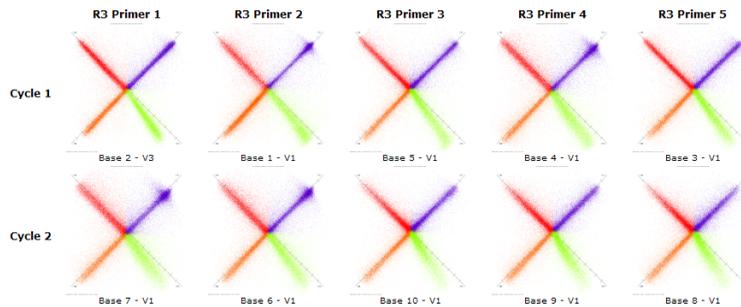
Raw data

Raw data images display the Satay plot for beads using raw signal intensities from the image analysis algorithm.

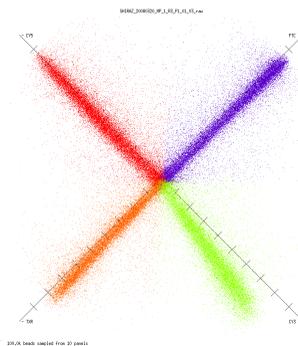
Satay Reports

Run SHIRAZ_20080320_MP_1
Sample Sample1

Primer Set: R3
Data: raw Channel: FTC



Note: The plot is calculated by subtracting the CY3 dye intensity from the CY5 dye intensity, then subtracting the FAM dye intensity.



Note: The CY3 dye is offset from the axis in this typical raw data view compared to the graph showing processed data.

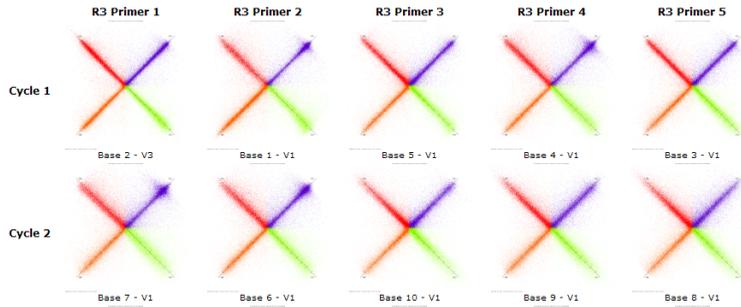
Scaled data

Scaled data images display the Satay plot for beads after signal intensities in each channel are normalized.

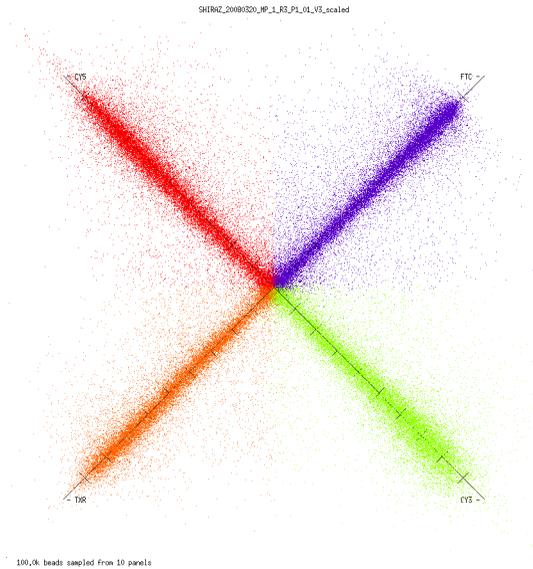
Satay Reports

Run SHIRAZ_20080320_MP_1
Sample Sample1

Primer Set: R3
Data: **scaled** Channel: FTC



Enlarged view:



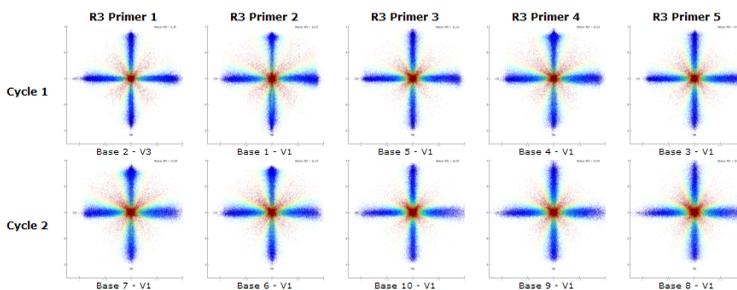
N2S data

N2S data plots show the distribution of high- and low-quality beads using a measure of the noise-to-signal (N2S) ratio.

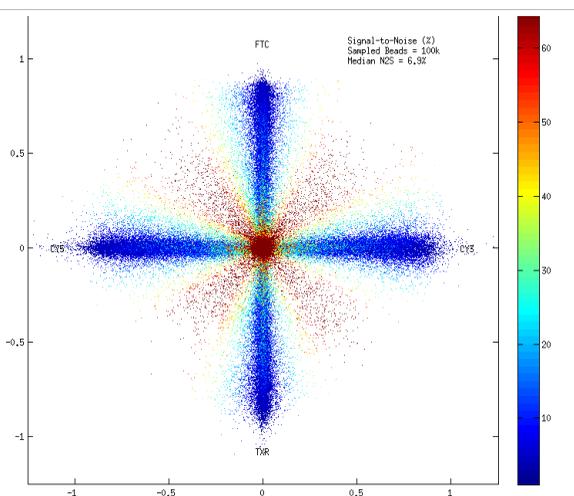
Satay Reports

Run SHIRAZ_20080320_MP_1
Sample Sample1

Primer Set: R3
Data: N2S Channel: FTC



Enlarged view:

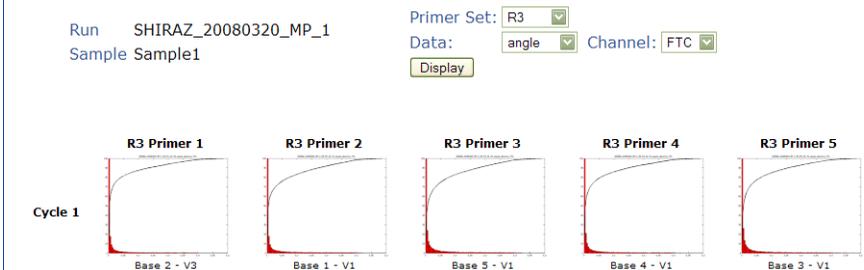


Angle data

Angle data histograms show the distribution of the angle between the 4-color signal vector and the canonical signal for each dye channel (6-FAM™, Cy3®, Texas Red®, and Cy5® dyes).

Note: Acceptable or good beads have angles close to zero.

Satay Reports

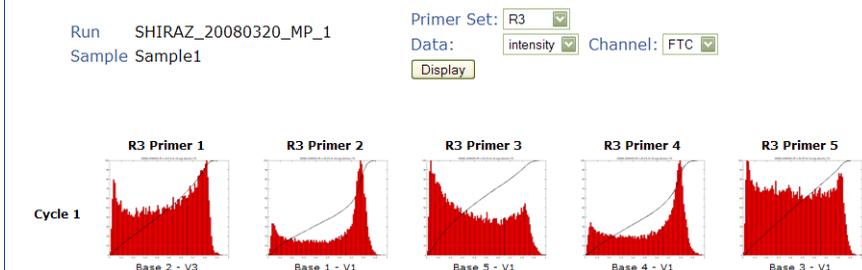


Intensity data

Intensity data histograms show the distribution of the intensities for each bead in each of the four dye channels (6-FAM™, Cy3®, Texas Red®, and Cy5® dyes).

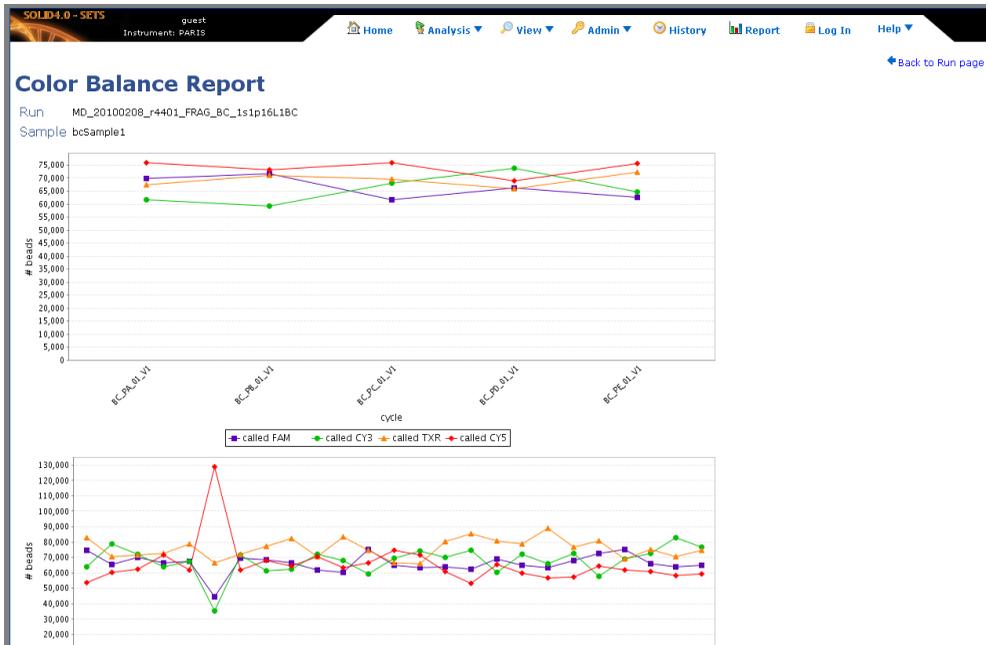
Note: Good beads have high intensities.

Satay Reports



Color Balance Report

Use this report to view the number of beads that have a particular color call for a specific ligation cycle. For most samples, the distribution of colors is relatively uniform across the four channels. However, a strong deviation from uniformity or an imbalance trend could indicate problems in the sequencing chemistry or imaging process, or a bias in the library preparation.



In this report, you can:

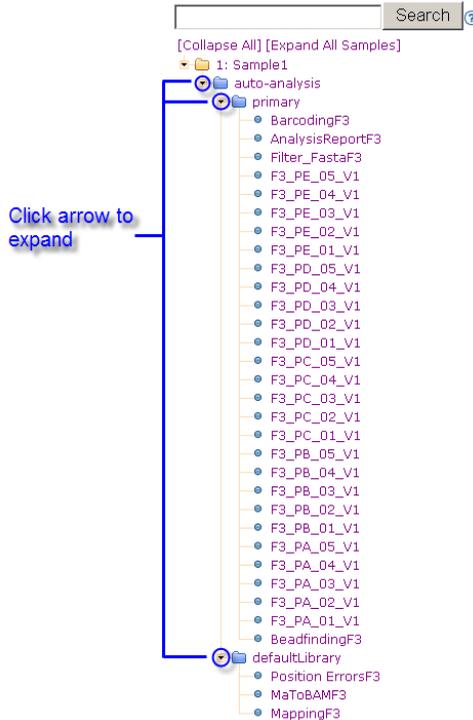
- Mouse over any data point to see the called value.
- Export the Scan Bead Summary table to a CSV file.

View Analysis Reports

View sample files Auto-analysis creates fasta files and various report graphs from the analyzed sample data. Primer sets F3 and R3 are represented by links to csfasta files for F3 and R3.

Note: When the analysis is successful for F3 and R3 primer sets, a csfasta file is generated. The “.csfasta” format is a text-based format that displays the base-pair sequences in color space (that is, bases display as 0, 1, 2, 3).

Click each auto-analysis folder’s expansion arrow to see its contents.



Select a library hyperlink to view the details.

Search

[Collapse All] [Expand All Samples]

- 1: Sample1
 - auto-analysis
 - primary
 - BarcodingF3
 - AnalysisReportF3
 - Filter_FastaF3
 - F3_PE_05_V1
 - F3_PE_04_V1
 - F3_PE_03_V1
 - F3_PE_02_V1
 - F3_PE_01_V1
 - F3_PD_05_V1
 - F3_PD_04_V1
 - F3_PD_03_V1
 - F3_PD_02_V1
 - F3_PD_01_V1
 - F3_PC_05_V1
 - F3_PC_04_V1
 - F3_PC_03_V1
 - F3_PC_02_V1
 - F3_PC_01_V1
 - F3_PB_05_V1
 - F3_PB_04_V1
 - F3_PB_03_V1
 - F3_PB_02_V1
 - F3_PB_01_V1
 - F3_PA_05_V1
 - F3_PA_04_V1
 - F3_PA_03_V1
 - F3_PA_02_V1
 - F3_PA_01_V1
 - BeadfindingF3
 - defaultLibrary
 - Position ErrorsF3
 - MaToBAMF3
 - MappingF3

Sample: Sample 1

Show in tree

Spot(s): 1

Sample Reports

Satay Report

Color Balance Report

Analysis: auto-analysis - Library: defaultLibrary

Show in tree

Cancel analysis

Description:

Date created: 03/30/2010 12:41 PM

Last completed job: Position Errors

Data Files

Mapping Summary F3 (1.26 KB)

csFasta file F3 (10.09 MB)

QV file F3 (20.85 MB)

Matching Result: in Gff format

Analysis Reports

Quality Values Report

Auto-Correlation Report

Error Profiles Report

Depth of Coverage Report

Mate Pair Report

Show in tree

Details for job: MappingF3 (id=1287)

Settings:	SingleTag_Test(v.1)	app.id	single.tag
Stage:	postPrimerSetSecondary	instrument.name	r1ahead01
Status:	finished	library.name	defaultLibrary
Created:	03/30/2010 12:55 PM	num.bases	25
Started:	03/30/2010 12:57 PM	panels	1-1
Completed:	03/30/2010 12:58 PM	plugin.id	mapping.run
		primer.set	F3
		probe.set	12
		read.length	25
		run.name	r1ahead01_20100330_FRAG_3
		sample.name	Sample1
		spots	1

View analysis Data Files In the sample folder, you can view and download the following data files for each sample.



- Mapping Summary F3
- Mapping Summary R3
- Mapping Summary F5-P2
- Mapping Summary F5-BC
- csfasta file F3
- csfasta file R3
- csfasta file F5-P2
- csfasta file F5-BC
- QV file F3
- QV file R3
- QV file F5-P2
- QV file F5-BC

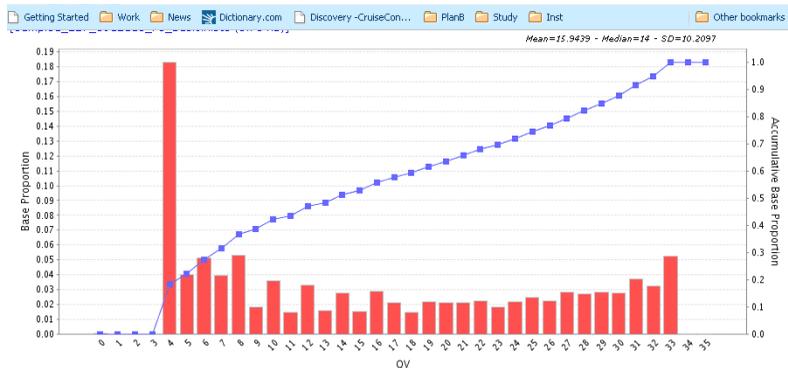
View Analysis Reports Sample analysis creates various reports for individual sample files. You can view and download the following reports:



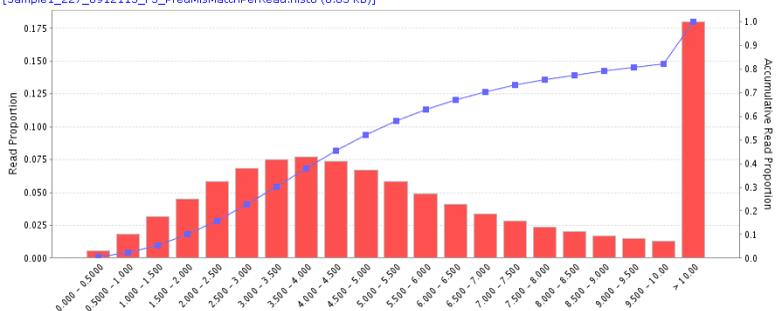
- Multiplexing Assignment Report – [page 92](#)
- Quality Values Report – [page 87](#)
- Auto-Correlation Report – [page 89](#)
- Error Profiles Report – [page 91](#)

Quality Values Report The Quality Values Report contains the quality values for each base sequenced.

Click **Quality Values Report (Run Details ▶ Analysis Reports)**.

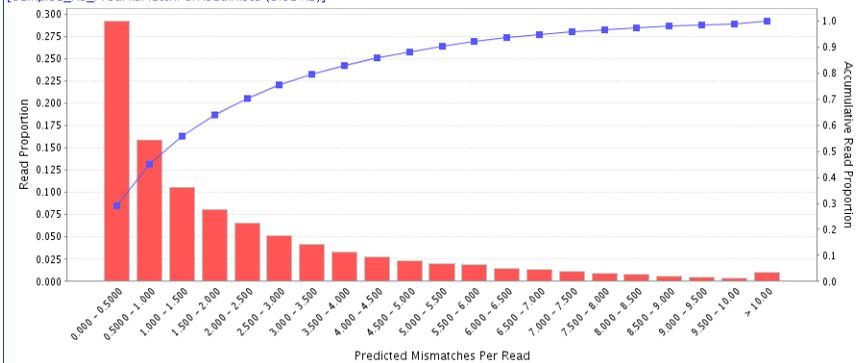


[Sample1_227_0912113_F3_PredMisMatchPerRead.histo (0.85 KB)]

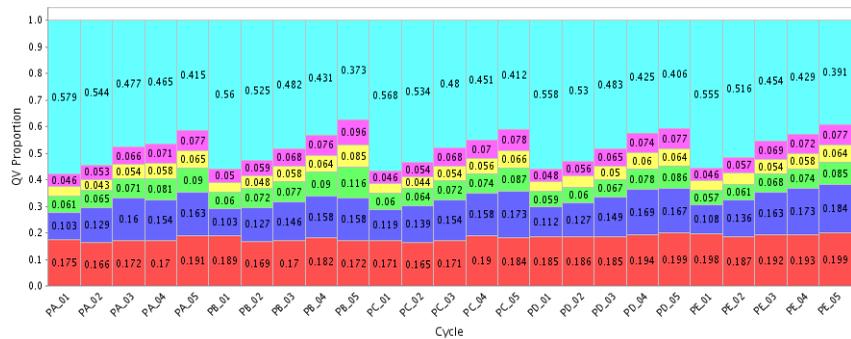


The bar chart shows the proportion of color calls with a particular quality value.

[Sample1_R3_PredMisMatchPerRead.histo (0.81 KB)]



With quality values, it is possible to predict the expected number of mismatches to the reference genome caused by measurement errors as depicted in the bar chart above. The blue line shows the accumulative read proportion with predicted mismatches. For this run example, about 28% of the color space calls are predicted to have 0 to 0.5 mismatches. About 80% of the bases are predicted to have approximately 0 to 3 mismatches.



The final graph included with the Quality Values Report shows the proportions of six groups of quality values versus the primers and cycles in the run. The proportion of the highest quality bases tend to decrease in later cycles and increase after the primer resets. (PAC1 represents Primer A Cycle 1, PBC5 represents Primer B Cycle 5).

Auto-Correlation Report

Use this report to view the correlation between pairs of color calls in the raw reads.

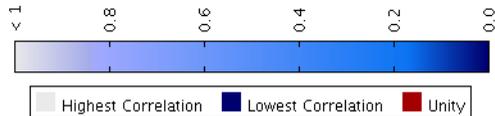
Click **Auto Correlation Report (Run Details ▶ Analysis Reports)**.

Auto-Correlation Report

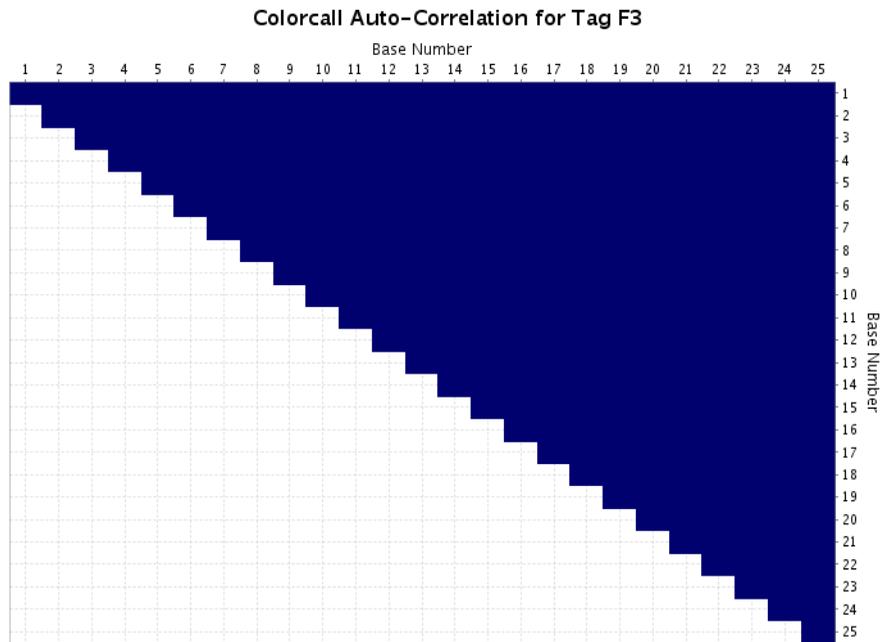
Run r1ahead01_20100113_FRAG Analysis primary_reanalysis
Sample Sample1 Library s1l02

Raw Data[F3_autoCorrelation.txt (2.82 KB)]

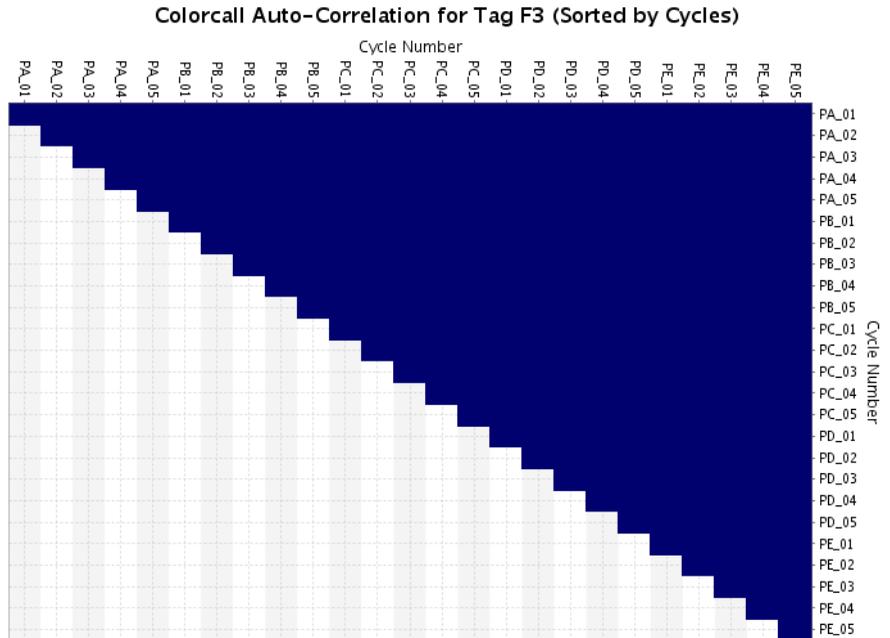
Correlation Scale



Tag F3



The Auto-correlation of 'base' position “i” with position “j” for Tag F3 displayed as a heat map, with high correlation as dark red/bright green, and low correlation as dark blue. The correlation is calculated as an “r” coefficient between each possible pair of positions.



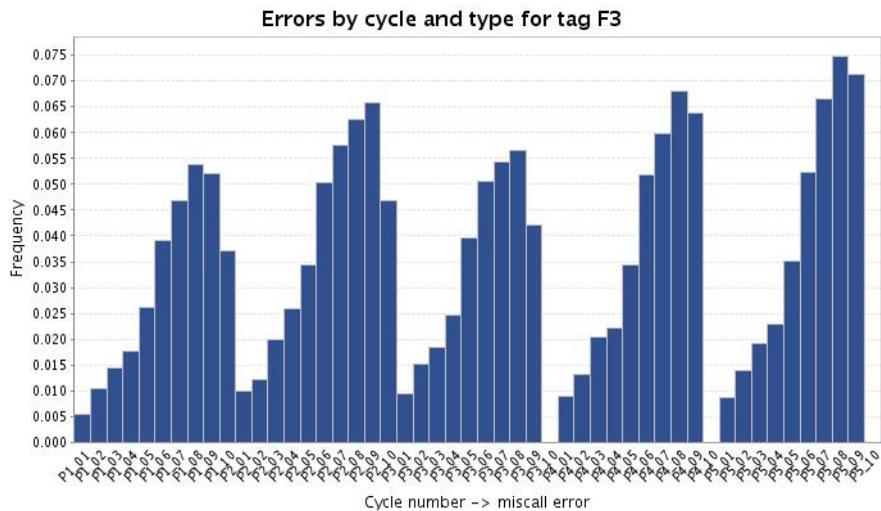
The Same data as the one on the previous page but sorted in the order of the chemistry cycles on the sequencing instrument.

Light blue shading or dark red shading shows stronger correlations between the two cycles. Such correlations indicate potential problems with swapped reagents, incomplete cleavage, and other sequencing chemistry problems.

Error Profiles Report

The Error Profiles Report gives the frequency of errors as a function of base position in the read and type of error produced for a particular tag.

Click the **Error Profiles Report** link (Run Details Analysis Reports) to view the graphs that show the number of mismatches to the reference sequence, shown by position on read (top) and by cycle (bottom). Mouse over to view the calculated values for each bar.



Multiplex Assignment Report

To see an example of multiplex assignment report, go to [“Multiplexing Assignment Report” on page 108](#).

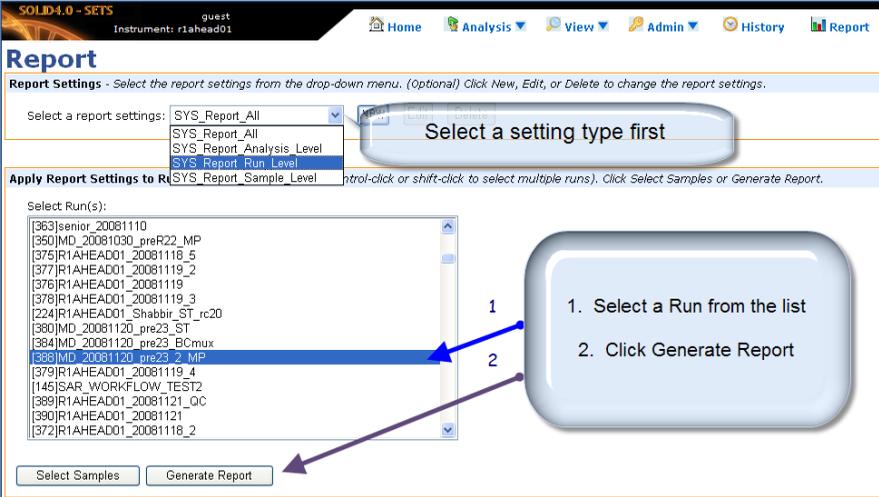
Master Report Tool

Generate a Master Report

SETS provides a powerful report building tool to assist you in generating a summary of several reports bundled into a comprehensive one, customized with only the data that you want to see.

1. From the main menu, click  **Report**.
2. Select a report setting from the drop-down menu.
3. Select at least one run from the full list of runs, then click **Generate Report**.

To select multiple runs, use the Control key and the scroll bar.



The screenshot shows the 'Report' tool interface. At the top, there is a navigation bar with 'Home', 'Analysis', 'View', 'Admin', 'History', and 'Report' menus. Below this, the 'Report Settings' section has a dropdown menu for 'Select a report settings:' with options: 'SYS_Report_All', 'SYS_Report_Analysis_Level', 'SYS_Report_Run_Level', and 'SYS_Report_Sample_Level'. A callout box points to this menu with the text 'Select a setting type first'. Below the settings, the 'Apply Report Settings to Run(s):' section contains a list of runs. One run, '[358]JMD_20081120_pre23_MP', is highlighted in blue. A callout box with two arrows points to this run and the 'Generate Report' button, with the text: '1. Select a Run from the list' and '2. Click Generate Report'. At the bottom, there are 'Select Samples' and 'Generate Report' buttons.

4. To specify which analyses are included in the Master Report, click **Select Samples** and check only the samples that you want included in the report.



The 'Select Samples' dialog box shows a table with two columns: 'Sample' and 'Run'. The first row is checked, and the second row is unchecked. Below the table are 'Select Analyses' and 'Generate Report' buttons.

Sample	Run
<input checked="" type="checkbox"/> Sample1_st	IL_CHAMPAGNE_20081022_ST_auto_export
<input type="checkbox"/> Sample1	JM_RPT_PRIMER_20080819

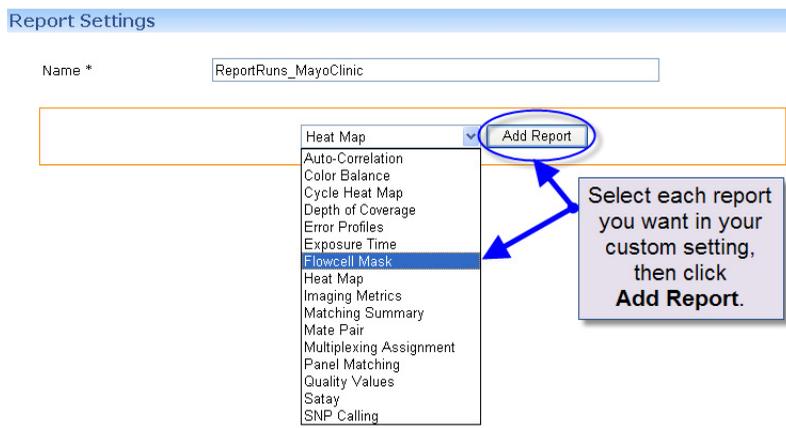
Create a custom Report Setting

To create a custom report setting:

1. Click **New** in the **Report Settings** view (next to the Name field).

IMPORTANT! You must be logged in before the software associates the new report setting to your user profile.

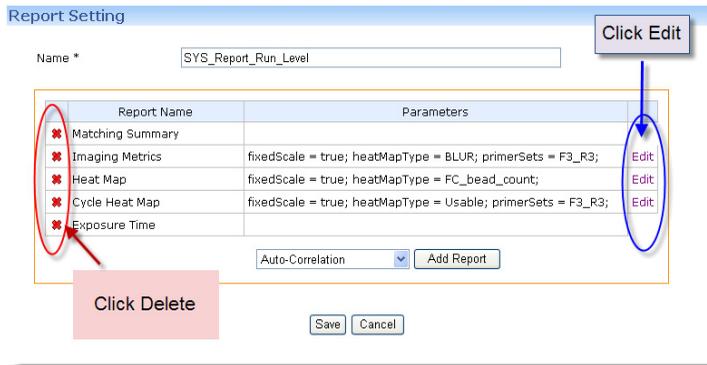
2. Select the reports that you want in your custom setting by selecting one at a time, then click **Add Report**.



3. Click **Save**.

Edit and save new Report Settings

To edit report parameters, click **Edit**. To delete a report from your custom report setting, click **Delete**.



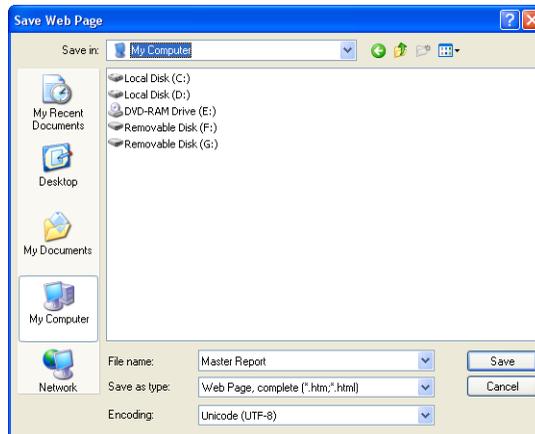
Click **Save** when you are finished making edits to the report setting.

IMPORTANT! You must be logged in to save a new report setting or edit an existing setting. If you are not logged in, any changes that you make are temporary and will apply only to the current session.

Save reports

After creating a Master Report with the comprehensive report-building tool (see “[Master Report Tool](#)” on page 93), use your browser menu to save the report.

Go **File ▶ Save Page As ▶ Save as type ▶ Web Page**.



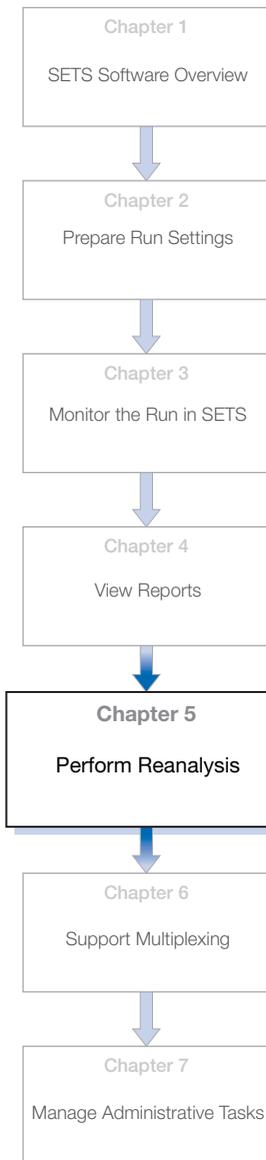
Click **Save** to save the entire page as one html file and create a folder by the same name containing the associated report graphics.

Print Reports Use the Print function on your browser to print reports.

Note: In some cases, you may need to adjust your browser to print the data reports.

5

Perform Reanalysis



This chapter covers:

- Overview 98
- Primary reanalysis 98
- Secondary reanalysis 100

Overview

You can perform two types of reanalysis on the SOLiD™ 4 Analyzer.

- Primary
- Secondary

When you perform primary reanalysis of a sample, SETS keeps the original focal map and color-calling data and recollects the filter FASTA and run analysis data. When you perform secondary reanalysis of a sample, only the run analysis data are collected again.

Note: You must be logged in with “user” or “admin” privileges in order to perform reanalysis.

Primary reanalysis

Overview Primary reanalysis starts at the Filter Fasta stage after color calling. After you select scans and reanalyze, a new analysis workflow is created in the database resulting in a new set of FASTA files to be generated from the pipeline. Reports that are based on FASTA file inputs reflect changes created by reanalysis.

Note: Make sure to use the correct version of the cycle scans that you want to reanalyze. Reanalysis is applied to all samples in a run.

1. Click **Reanalyze Primary** to reanalyze the primary analysis data.

Click the  icon for more information.



MD_CHAMPAGNE_20090108_r24_BCim_8s_8spt_16lib_2: Primary Reanalysis

Select the scan to reanalyze, then click Reanalyze.

Name: primary_reanalysis

Description:

[Reanalyze] [Cancel]

[Bead Counts: Total Average per panel]											Exposure Times (ms)		
Cycle	Scan			Panels	Failed panels	Usable Beads (max=561,385)	Best Beads	Best+Good Beads	Bead Color Balance	FAM	CY3	TXR	CYS
	Mask	Ver.	Used										
BC_PA_01	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
BC_PB_01	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
BC_PC_01	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
BC_PD_01	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
BC_PE_01	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
F3_PA_01	<input checked="" type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
F3_PA_02	<input checked="" type="checkbox"/>	Mask this cycle. Masked cycles will be represented by a '.' in the FASTA output file.								-	-	-	-
F3_PA_03	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
F3_PB_01	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
F3_PB_02	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-
F3_PB_03	<input type="checkbox"/>	v.1	<input checked="" type="radio"/>	8	0	-				-	-	-	-

Selecting the **Mask** column allows the analysis to mask, or ignore, this cycle. Masked cycles will be represented by a '.' in the FASTA output file.

If the cycle has more than one scan version, click the radio button next to the version that you want to select for reanalysis.

Scan version 3 is selected

	v.1	<input type="radio"/>		2,357	1036	(44%)
	v.2	<input type="radio"/>		2,357	11	-
	v.3	<input checked="" type="radio"/>		2,357	0	-
R3_PC_02	v.1	<input type="radio"/>		2,357	77	(3%)
	v.2	<input type="radio"/>		2,357	6	-
	v.3	<input checked="" type="radio"/>		2,357	0	-

Secondary reanalysis

Overview Secondary reanalysis starts at the Single Tag Analysis stage. This workflow stage generates application-specific results such as base space sequences, alignment, tag counting from the colorspace sequence reads, and ".csfasta" file. The external analysis pipelines will also be executed if they are plugged-in and turned-on in the selected analysis setting.

r1ahead01_20100410_PE_2: Secondary Reanalysis

Name: secondary_reanalysis

Description:

Step 1: Select one or more Sample(s) --> (Optional) Enter the library keyword filter --> Click on Add Libraries

Sample1

Library Keyword filter:

Step 2: Select one or more Libraries --> Select the Application Type --> Select the Secondary Analysis Settings --> Click on Add

None selected

Application Type: SingleTag

Secondary Analysis Settings: SingleTag_sy

None Selected

The secondary reanalysis allows you to selectively apply secondary analysis settings to individual samples. You are not forced to reanalyze all samples as you are in primary reanalysis. The downstream effect is that a new analysis is created. All Single Tag analyses are rerun.

To perform a secondary reanalysis:

1. Click **Reanalyze Secondary** to reanalyze the secondary analysis data.

Click the  icon for more information.



2. Check the sample, or samples to reanalyze, specify a library keyword (optional), then click **Add Libraries**.
3. Select the library or libraries from the left panel to apply the Secondary Analysis Settings to, then click **Add**.
4. Verify that libraries, samples, and settings are set up correctly, then click **Reanalyze**.

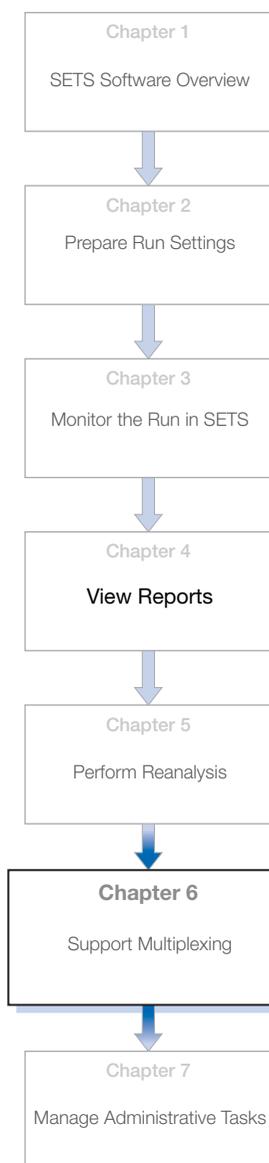
Note: Upon clicking Reanalyze, you will be warned that reanalysis might affect the performance of the instrument system. At this point, you can cancel reanalysis.

Effect on the current run After you initialize SETS to reanalyze your samples, you are returned to the Run Portal page where a notification of reanalysis is displayed.

Results If you continue with reanalysis, you can monitor progress by expanding samples in the sample tree. In the Run Details page, you will see a new analysis under each sample.

6

Support Multiplexing



This chapter covers:

- Fragment and Paired-End Multiplexing 104
- View Multiplexing Series in SETS 105
- Set up Multiplexing run in ICS 107
- Multiplexing Assignment Report 108
- Reassign Run 111

Fragment and Paired-End Multiplexing

Overview of Multiplexing and Barcodes

Multiplexing in general, and barcodes in particular, enable the SOLiD™ 4 System to target many applications, from deep sequencing of a single source to extreme parallel sequencing of a large number of samples over limited regions - sometimes on the same slide.

The additional stretches of DNA sequence, known as barcodes, can be used to segregate pooled library samples during sequencing back into their original (logical) single library groupings. Barcodes—in their simplest form—allow the pooling of samples on a slide without the loss of surface area creating geographic isolations. Multiplexing occurs at both the whole slide level as well as at the geographically isolated (mask) region level.

A barcode, by definition, is a tag with a known sequence that is attached to a sequencing primer on a bead during library preparation. When sequenced, it is the barcode that identifies and classifies the bead to a named sample. This barcoding, or indexing, approach allows multiplexing at the sequencing level.

Following sequencing of the target DNA, additional rounds of ligation-based sequencing are performed using primer sets complimentary to the barcode. The resulting reads can then be sorted by the barcode and aligned in groups to the reference sequence.

Barcodes are assigned to a library both conceptually and physically (as specific tags) as part of the library preparation. If more than one barcode is assigned to a single library during the library preparation step, the conceptual grouping is referred to as a barcode group. All barcodes in the pool will assign the underlying beads to the same original library for analysis.

The SOLiD™ 4 System supports multiplexing for both fragment sequencing and paired-end sequencing. See [“Naming convention for sequencing reads” on page 11](#).

Note: Multiplexing is not available for mate-paired libraries.

View Multiplexing Series in SETS

Access Multiplexing Series runs in SETS

The Multiplexing Series view in SETS shows all the barcode kits in the system to date. When a multiplexing series run is created in ICS, only the multiplexing series names are displayed.

To view multiplexing series details, click **View ▶ Multiplexing Series**.



Select the multiplexing series run that you want to view.

Every multiplexing series file includes:

- Name of the multiplexing kit
- Colorspace code
- Basespace sequence

SOLID4.0 - SETS
Instrument: riahead01

Home Analysis View Admin History Report Log In Help

Multiplexing Series

SOLID Multiplexing Early Access set -

Name	Colorspace	Basespace
Barcode 1	00032	GGGGCTT
Barcode 10	02103	GGACCGC
Barcode 11	32212	GCTCAGG
Barcode 12	22331	GAGCGTT
Barcode 13	33001	GCCGGTC
Barcode 14	33122	GCGTCTG
Barcode 15	13233	GTAGCGT
Barcode 16	03310	GGCGTTT
Barcode 2	00111	GGTGTTG
Barcode 3	20200	GAAAGGG
Barcode 4	30323	GCCGATG
Barcode 5	31013	GCAAAGA
Barcode 6	01130	GGTGCCC
Barcode 7	01221	GGTCTGG
Barcode 8	21302	GACGGAG
Barcode 9	02020	GGAAAGG

Note: Refer to the advanced topic appendices in this guide to learn more on colorspace and basespace.

Set up Multiplexing run in ICS

Create a Fragment and Paired-End Multiplexing run in ICS

There are two ways to set up a multiplexing run in the ICS:

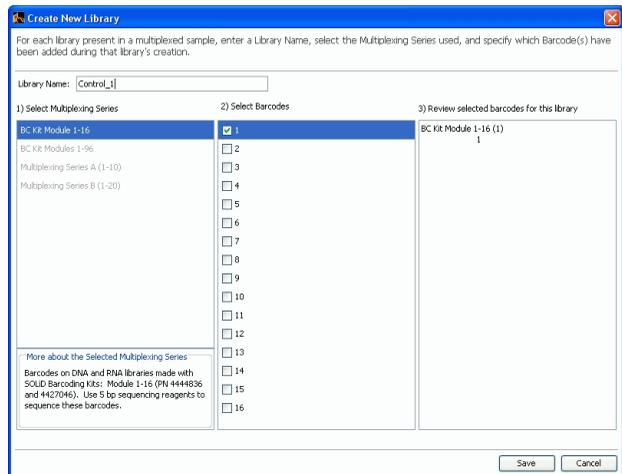
- Creating a multiplexing run using the Run Wizard
- Importing a multiplexing run

Note: Refer to the *Applied Biosystems SOLiD™ 4 System ICS Software Help* or *Applied Biosystems SOLiD™ 4 System Instrument Operation Guide* (PN 4448379) for detailed instructions on these procedures.

Assign Libraries in ICS

Continuing in the Run Wizard, you manually configure each barcode library to generate a new run definition. You need to save the run definition before you assign the run to one of the two flowcells.

Using the Run Wizard, assign libraries to the appropriate barcode from the appropriate multiplexing series by creating a run definition file.



Note: Refer to the *Applied Biosystems SOLiD™ 4 System ICS Software Help* or *Applied Biosystems SOLiD™ 4 System Library Preparation Guide* (4445673) for detailed instructions on these procedures.

Multiplexing Assignment Report

Reports for Multiplexing Series runs

To view multiplexing run analysis results, select the sample folder ▶ analysis folder ▶ library folder.

Note: Enter a library name in the Search field to automatically expand the sample tree to that library.

The screenshot displays the Multiplexing Assignment Report interface. On the left, a tree view shows the sample structure. The tree is expanded to show sample 1 (bcSample1) and sample 2 (bcSample2). Under sample 1, the 'auto-analysis' folder is selected, which contains a 'primary' folder. The 'primary' folder contains various analysis folders such as 'BarcodingF3', 'AnalysisReportF3', 'Filter_FastaF3', 'F3_PE_01_V1', 'F3_PD_01_V1', 'F3_PC_01_V1', 'F3_PB_01_V1', 'F3_PA_01_V1', 'BarcodingBC', 'AnalysisReportBC', 'Filter_FastaBC', 'BC_PE_01_V1', 'BC_PD_01_V1', 'BC_PC_01_V1', 'BC_PB_01_V1', 'BC_PA_01_V1', and 'BeadfindingBC'. The 'lib1' folder is also visible, containing an 'auto-analysis' folder. The 'lib2' folder is also visible.

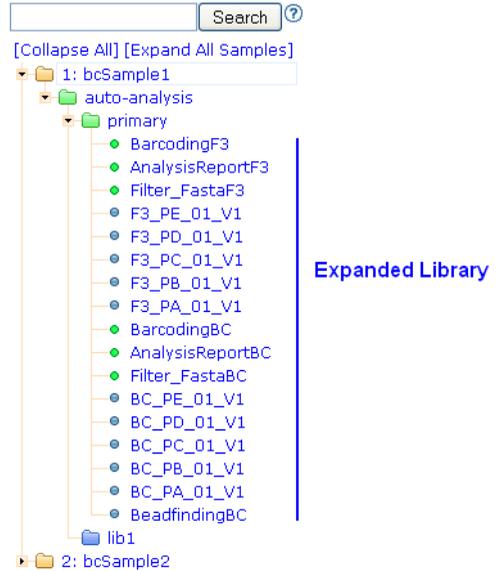
On the right, the 'Sample: bcSample1' details are shown. The 'Sample Reports' section includes 'Satay Report' and 'Color Balance Report'. The 'Analysis: auto-analysis' section shows 'Description: Colorcalling', 'Date created: 01/14/2010 04:51 PM', and 'Last completed job: Colorcalling'. The 'Data Files' section is empty. The 'Analysis Reports' section includes 'Quality Values Report', 'Auto-Correlation Report', and 'Multiplexing Assignment Report'. The 'Details for job: F3_PE_01_V1 (id=283)' section shows the following information:

Settings:	default primary(v.1)	cycle	1
Stage:	postScanslidePrimary	export.dest	foshtdvv10
Status:	finished	export.enabled	true
Created:	01/14/2010 04:58 PM	focal.map.dir	r1ahead01_20100114_FRAG_BC/r1ahead01_
Started:	01/14/2010 04:58 PM	focal.map.stg	r1ahead01_20100114_FRAG_BC/r1ahead01_
Completed:	01/14/2010 04:59 PM		r1ahead01_20100114_FRAG_BC_FOCALMAF
focal.map.version	1		
images.dir.1	r1ahead01_20100114_FRAG_BC/r1ahead01_		
instrument.name	r1ahead01		
num.bases	5		
p1.channel			
p2.channel	CV3		
panels	2-2		
primer.id	5		
primer.set	F3		
probe.set	12		
run.name	r1ahead01_20100114_FRAG_BC		
run.type	run		
sample.name	bcSample1		
spots	2		
version	1		

There is a wide range of reports that you can generate from the results data, one of which is the Multiplexing Assignment Report.

Multiplexing Assignment report

The Multiplexing Assignment report is a report specific to barcode runs. There is one report generated per analysis.



When you have selected your target analysis folder, click



Multiplexing Assignment Report

Run MD_20081126_r23_BCmux
Sample bcSample1

Total % Assigned : 53.77%

Raw Data

[BarcodeStatistics.20081126204223303.txt (0.47 KB)]

Export to CSV file

Barcode Statistics				
Library	Barcode	0 Mismatches	1 Mismatch	Total
L1	SOLID BC4	3276	694	3970
L1	SOLID BC3	2846	543	3389
L1	SOLID BC1	3868	1473	5341
L1	SOLID BC5	3157	1040	4197
L1	SOLID BC2	3416	718	4134
L1	Subtotals	16563	4468	21031
L2	SOLID BC6	3271	1142	4413
L2	SOLID BC8	3111	669	3780
L2	SOLID BC7	3485	1011	4496
L2	SOLID BC10	4142	1279	5421
L2	SOLID BC9	3256	1407	4663
L2	Subtotals	17265	5508	22773
unassigned	unresolved	NA	NA	37665
unassigned	Subtotals	0	0	37665
All Beads	Totals	33828	9976	81469

The Multiplexing Assignment report shows the number of barcode reads assigned to each library, along with the subtotals for all libraries. It also shows the number of reads that were not assigned to any specified library. These numbers include both passed and failed reads from the filter fasta stage.

The value in the “0 Mismatch” column is the number of barcode reads that match exactly to the barcode shown in the barcode column. The “1 Mismatch” column is the number that missed matching that barcode by one.

Reads that did not match any of the barcodes at the specified level are counted in the unassigned library with the name unresolved. It is possible that barcodes in the kit that were not assigned to a library are also matched. In that case, the unassigned library may show barcodes named with the color sequence of the matched barcode. Two additional libraries may appear:

- For beads that have barcode read but no F3 read (missing-f3).
- For reads that have an F3 read but no barcode read (missing-bc).

The “Total % Assigned” is the sum of the library subtotals divided by the “All Beads Total” expressed as a percent.

Import a Multiplexing Series

Follow the steps that use the Command line to import a multiplexing series found in the Applied Biosystems *SOLiD™ 4 System ICS Software Help*.

Multiplexing report

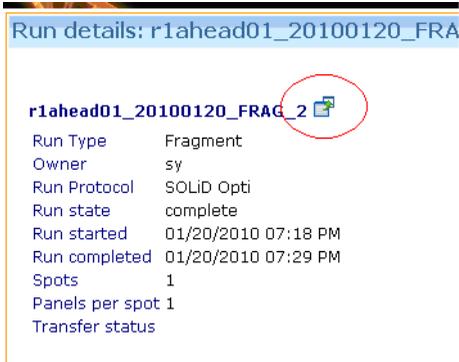
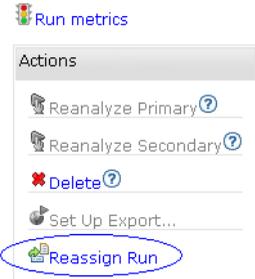
Multiplexing report can be attained as soon as barcoding is complete for a run.

Reassign Run

1. Download the sample sheet of the run that you want to reassign.
2. Open the sample sheet in excel and edit the barcoding reassignment or spot reassignment.

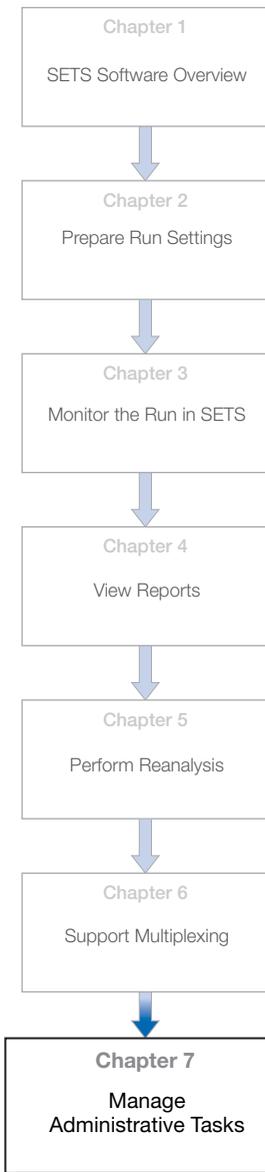
Note: When reassigning the spots, ensure that the new, or edited, sample sheet has the same number of spots.

3. Log in as Administrator.
4. From Home page, click **Reassign Run**.
5. Click **Log on** ► enter your name ► enter password.
6. From the Home page, click on the analyses result of your choice, then click **Reassign Run**. Importing a new run definition for an existing run will force regeneration of all results.



7

Manage Administrative Tasks



This chapter covers:

- Export Data with SOLiD™ 4 System Software 114
- Auto export in SETS. 115
- Manual export in SETS 124
- Event notifications and email services 125
- Alerts 128
- Delete runs 129
- Manage Users 130

Export Data with SOLiD™ 4 System Software

The SOLiD™ 4 System software supports:

- Auto export
- Manual export

Auto export You can set up SETS to auto-export data (.spch files) on a cycle-by-cycle basis for every run, or per run. Enable auto-export in SETS prior to a run.

IMPORTANT! Ensure that the off-instrument has the auto-export software installed.

Manual export When opting to perform a manual export, you can configure the export to be all data or specific data. See [“Manual export in SETS” on page 124](#) for details.

Auto export in SETS

To auto-export run data to the offline cluster, log into SETS, then go to **Admin ▶ Change Preferences**.



1. On the Edit User Preferences and Profile page, enter the host name or IP address of the server where you will export your runs to.

IMPORTANT! Prior to performing auto export, ensure that the off-instrument cluster has enough disk space.

2. Check the **Auto-export after primary analysis** box.



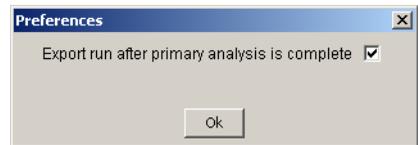
3. Click **Save**.

IMPORTANT! When a run is created, the “Export to Host” and “Auto-export after primary analysis” preferences are saved to *that* run. Any more changes to the preferences in the Edit User Preferences page will not affect the values of *previous* runs.

IMPORTANT! RSA keys must be set up between machines in order for the host name to be accepted. See [“Manually configure the RSA keys” on page 121](#).

Status	Time Initiated	History
successful	01/26/2009 02:57 PM	Exported to foshtdv02 on 01/26/2009 02:57 PM
failed	01/26/2009 01:19 PM	Could not execute: /usr/bin/rsync -vz pipeline@foshtdv02:/data/results. Host key verification failed.
failed	01/26/2009 01:18 PM	The run RV TEDESCHI_20090126_WFA already exists on the destination: foshtdv02
failed	01/26/2009 01:12 PM	Could not execute: /usr/bin/rsync -vz pipeline@foshtdv02:/data/results. Host key verification failed.

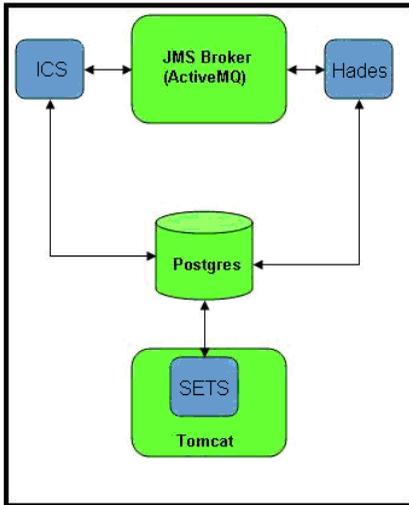
Note: Once auto export is configured in the SETS software, ensure that the auto export option is selected in the ICS, before starting the run.



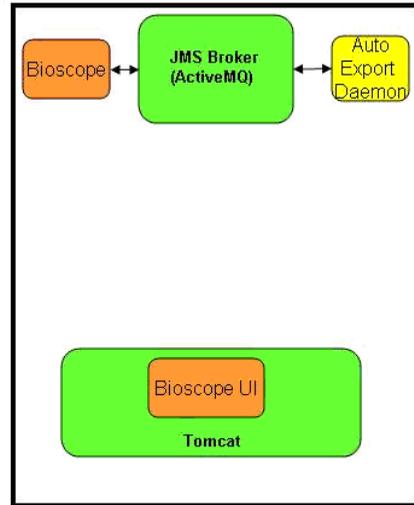
IMPORTANT! If run export fails displaying a similar message to the one above, it is usually due to a change in the host IP address or a loss in the RSA key files. Manually configuring the RSA keys should resolve the problem.

The following diagram shows the deployment diagram on remote machine (the offline cluster).

Instrument Cluster



Remote Cluster



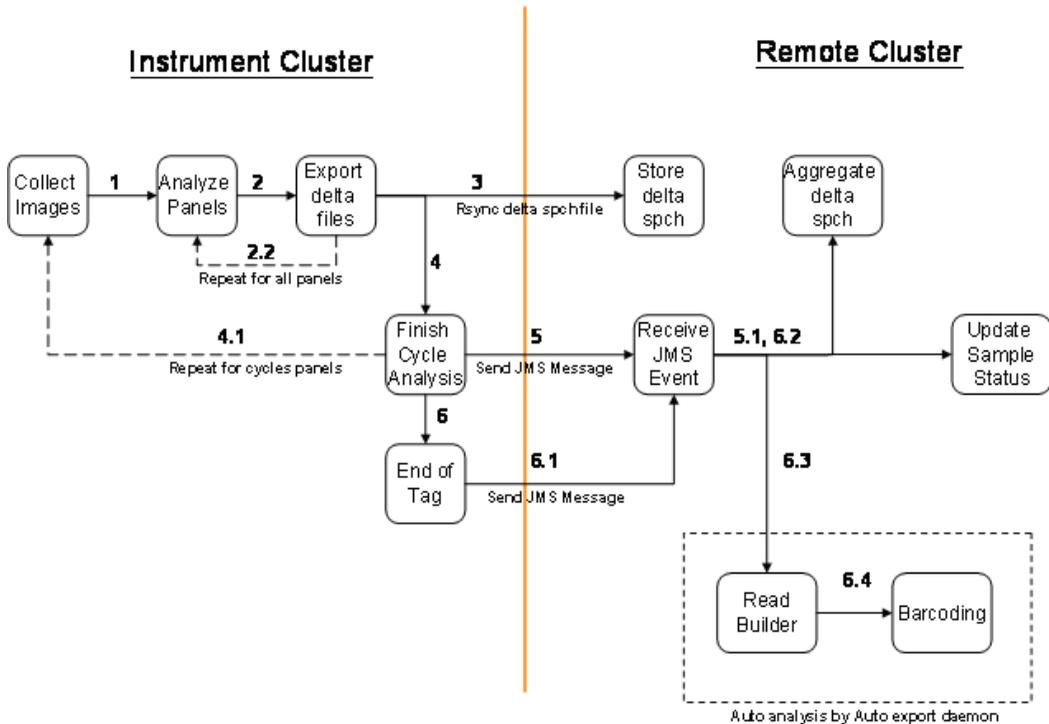
- Disco installer
- Bioscope installer
- System installer
- Auto-export installer

Auto export can work in two modes:

1. With (Java Messaging Service) JMS broker and Export Daemon.
2. Without JMS broker and Export Daemon.

Auto export with JMS broker and Export Daemon

In this mode, all the delta spch is automatically transferred to remote location. On the remote side, the auto export daemon automatically runs readbuilder and barcoding. See the diagram below.



Auto export without JMS broker and Export Daemon

In this mode, the user can manually run readbuilder and barcoding. When auto-export is performed on a run, all the jobs folders are also exported to the remote cluster.

Jobs folders are typically located at:

```
/data/results/<instrument name>/<run name>/<sample name>/jobs.
```

To run read builder:

1. Log in to the linux cluster as user which can submit jobs to pbs (pipeline).
2. Change directory (cd) to the postPrimerSetPrimary under jobs folder.

3. Run the following command:

```
jprimaryanalysis.sh -dir /data/results/<instrument  
name>/<run name>/<sample name>/jobs  
/postPrimerSetPrimary.<job id> --ini=parameters.ini
```

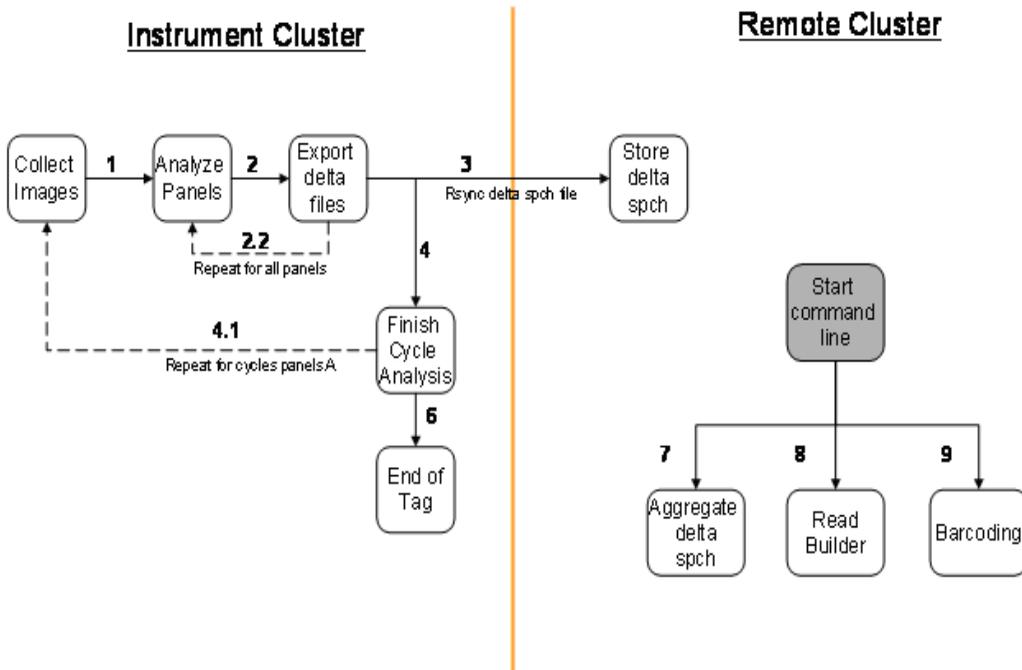
This will run readbuilder and generate csfasta output.

To run barcoding:

1. Log in to the linux cluster as user which can submit jobs to pbs (pipeline)
2. Change directory (cd) to the postBarcodePrimary under jobs folder.
3. Run the following command:

```
jprimaryanalysis.sh -dir /data/results/<instrument  
name>/<run name>/<sample name>/jobs /  
postBarcodePrimary.<job id> --ini=parameters.ini
```

This will run barcoding and split the csfasta into libraries as defined in the original run configuration. See the diagram below.



Example of saving preferences

1. User A logs in, and enters **Host 1** as the host name and checks **Auto-export after primary analysis** (Edit User Preferences page).
2. Then User A logs into ICS and creates a run “Run1_FC1.” User A’s preferences are applied to this run.
3. Next, User A goes back to SETS and changes the preferences to Export to Host 2.
4. User A returns to ICS and creates another run called “Run2_FC2.” This run copies the most recent preferences, including the new host name, and exports data once the run is complete.

With the current setup, User A has created Run1 on Flowcell 1 auto-exporting to one host (Host 1) and Run 2 on Flowcell 2 to a different host (Host 2).

Manually configure the RSA keys

When you export, your runs will be exported to the server called “hostName”. For example, assume you are exporting a run from server_src to server_dest. Enter **server_dest** in the hostName field and check the **Auto-export** box.

If you get an error message, follow the steps below:

1. From the command line, confirm that each machine can see the other by the hostname on the network:
 - Use the secure shell (ssh) to log into the server_src, then type **ping server_dest**.
 - Use the secure shell (ssh) to log into the server_dest, then type **ping server_src**.

In both cases, you should see bytes returned.

2. If you have access to remoteKey.sh, execute the following commands: (If you do *not* have access, go to [step 3](#)).
 - Use the secure shell (ssh) to log into the server_src as corona, then type **remoteKey.sh pipeline@server_dest**.
 - At the prompt for corona's password of server_dest, type in the password of corona user.
 - Use the secure shell (ssh) to log into the server_src as pipeline, then type **remoteKey.sh pipeline@server_dest**.
 - At the prompt for pipeline's password of server_dest, type the password of pipeline user.
 - Go to [step 4](#).
3. If you do NOT have access to **remoteKey.sh**, execute these commands:
 - Use the secure shell (ssh) to log into the server_dest as pipeline, then go to the ssh directory
 - ssh pipeline@server_dest.
 - cd ~/.ssh.
 - Execute the following commands which copy the appropriate public keys onto server_dest:
 - scp corona@server_src:~/.ssh/id_rsa.pub/remote.pub then enter corona's password.
 - cat remote.pub >> authorized_keys.

- scp
pipeline@server_src:~/.ssh/id_rsa.pub./remote.pub
then enter pipeline's password
- cat remote.pub >>authorized_keys.
- Exit.

4. Test

Execute the following commands:

- ssh corona@server_src
- ssh pipeline@server_dest
- Exit.
- Exit from server_src
- ssh pipeline@server_src
- ssh pipeline@server_dest
- Exit.

Change export for current run only

To modify the export preferences for the current run only and keep run data stored locally:

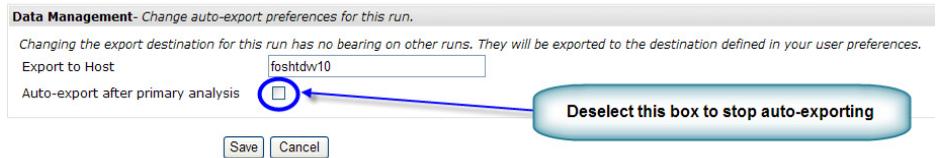
1. Go to **Run Details** ▶ **Action box** ▶ **Host Configuration**.



2. Uncheck **Auto-export after primary analysis**.

Any subsequent secondary analysis for the current run will now occur on the local cluster and run data will not be exported.

Edit r1ahead01_20090906_3's Export Preferences



3. Click **Save**.

Manual export in SETS

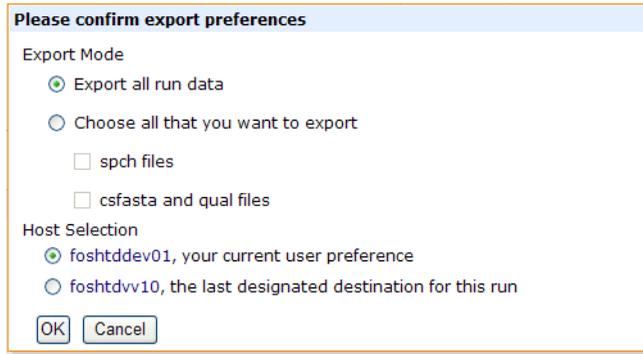
Set up manual export

1. Log into SETS, then go to **Run Details** ▶ **Action box**.
2. Click **Set Up Export...**

Note: If the Host has not yet been configured in the Preferences, this link will be disabled.



3. Select one of the two export options, then the host location:



4. Click **OK** to save your settings.
5. On the Run Details page, click **Export** (action box) to manually export the run.



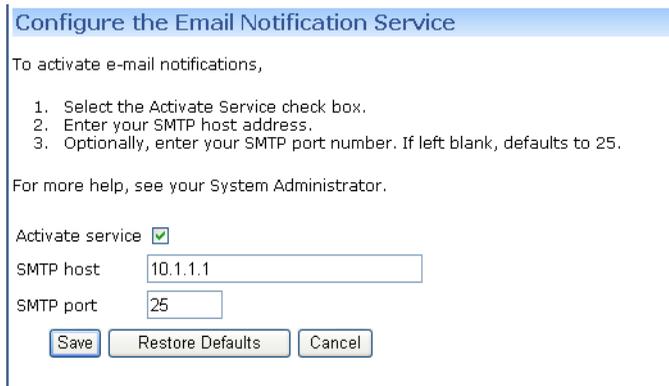
Event notifications and email services

Configuration setup There is a one-time setup requirement for notifications and emails to work properly. Work with your IT representative to configure the correct SMTP host and port information for your network.

1. Log in to the SETS software.
2. Click **Admin** ► **Email Service**.



3. Check the **Activate Service** checkbox.
4. Enter your SMTP host number.
5. Enter you SMTP port number, or leave the port number blank to allow the system to assign the port number 25 by default.



6. Click **Save**.

Event notification Event notification includes:

- Email notification
- Default actions when an event occurs

To set up event notifications:

1. Log in to the SETS software, if you are not already logged in.
2. Click **Admin** ▶ **Change Preferences**.



3. Enter the email address, or addresses, where you want to receive the event notifications.

Note: If you enter multiple email addresses, separate each with a comma.



4. Select the run-specific events you want to be notified about and the default actions to take under each category shown (System Errors, Fatal Run Errors, Informational Updates, and Completion Updates).

System Errors
Email: Node Failure Power Failure Instrument Errors Job Scheduler Down

Fatal Run Errors
Instrument Run Errors
Email: All Runs My Run(s) Only Ignore

Informational Updates
Run Metrics
Email: All Runs My Run(s) Only Ignore
Default Actions: Stop Export Pause Run

Low signal in spot(s)
Email: All Runs My Run(s) Only Ignore
Default Action: Pause Run

Completion Updates
Primer Complete
Email: All Runs My Run(s) Only Ignore

Primer Set Complete
Email: All Runs My Run(s) Only Ignore

Run Complete
Email: All Runs My Run(s) Only Ignore

Manual Export Complete
Email: All Runs My Run(s) Only Ignore

Note: When you select “My Runs Only” you will receive email notifications for the runs created only by you. When you select “All Runs” you will receive email notifications for all runs on the instrument.

5. Click **Save** to begin receiving email notifications.

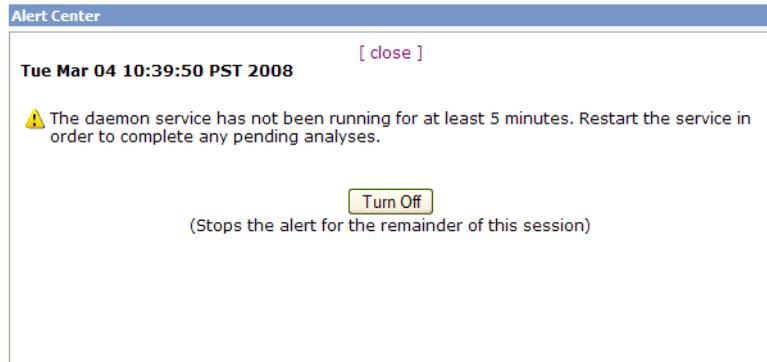
Default actions Default actions are available for these events:

- Run metrics (stop export)
- Run metrics (pause run)
- Low signal in spots (pause run)

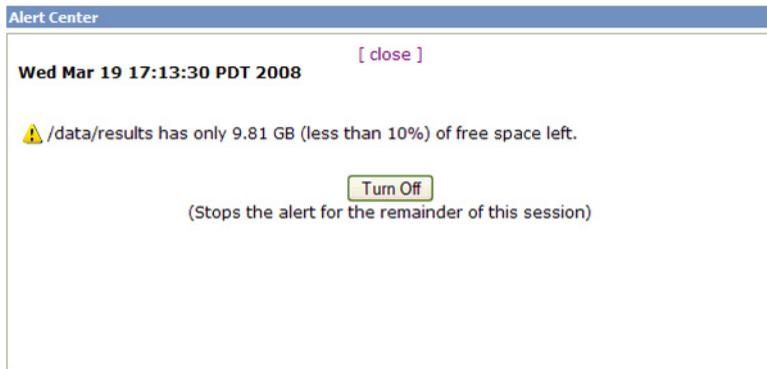
Run Metrics event notifications are triggered whenever one or more of the quality or other metrics falls into the red zone during a run.

Alerts

Job Manager If the Job Manager daemon service is not running, you may see this alert in SETS:



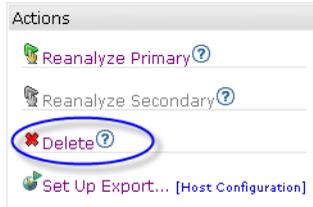
Results disk space alert If you are working in SETS, and you see this alert, you need to clean up disk space by either exporting or deleting data.



Note: To change the default setting of your minimum space threshold, change the property file (runDirectory.properties). After making changes, restart ICS and the SETS server.

Delete runs

When you need to clean up disk space and want to delete runs, go to a particular run, then click **Delete** (Run Details page).



Select a delete option when prompted.



- Selecting intermediate results files deletes:
 - RUN_FOLDERS/SAMPLE_NAME/"jobs" folder
 - RUN_FOLDERS/SAMPLE_NAME/"results.*"/"colorcalls"/ All folders except "CACHE"
 - RUN_FOLDERS/SAMPLE_NAME/"results.*"/"intermediate.*" folder
 - RUN_FOLDERS/SAMPLE_NAME/"results.*"/"primary.*"/"reject" folder
 - RUN_FOLDERS/SAMPLE_NAME/"results.*"/"libraries"/ LIBRARY_NAME/"intermediate.*" folder
 - RUN_FOLDERS/SAMPLE_NAME/"results.*"/"libraries"/ LIBRARY_NAME/"primary.*"/"reject" folder

- Selecting images deletes all imaging data associated with that run.
- Selecting all run data except database records deletes all results and imaging data associated with that run.
- Selecting all run data deletes all results, imaging data, and database records associated with that run.
- Selecting database records deletes the database records for that run, but nothing on the file system.

Note: After deletion, the run will not be viewable in SETS.

Manage Users

An administrator can add new users, create passwords, change user preferences, and inactivate current users under **Admin ▶ Manage Users**.

See “[Log in to SETS](#)” on page 21 for information about “admin”, “user”, and “guest” (default) log in permissions.

Create a New user

1. In the SETS Home page, go to **Admin ▶ Manage Users**.
The **Current Users** window opens.
2. Click **New** to open the Create a New User dialog box.



3. Enter a name in the User name field to add a new user.

Create a new user

User name *	<input type="text" value="New Name"/>
Role	<input type="text" value="user"/>
Password *	<input type="password"/>
Re-enter password *	<input type="password"/>

4. Select the Role – “guest”, “user”, or “admin” – (See [“Log in to SETS” on page 21](#) for information about each role.)
5. Enter a password for that user, then re-enter the same password.
6. Click **Submit**.

Inactivate a user

In the Current Users table, under the Activation column, click the check mark next to a current user name to inactivate that user.

Click **Inactivate** at the “Are you sure” prompt.

Deleting a current user deletes all analysis settings associated with that user.

The screenshot displays the 'Current Users' management interface in SOLiD4.0 - SETS. The interface includes a header with the user 'Administrator (admin)' and instrument 'riahead01'. A 'New...' button is circled in blue. Below is a table of current users with columns for Name, Role, Edit, and Activate/Deactivate. The 'sy' user is highlighted, and its 'Activate/Deactivate' checkbox is checked and circled in blue. A 'User Inactivation' dialog box is open, asking 'Are you sure you want to inactivate sy?' with 'Inactivate' and 'Cancel' buttons.

Name	Role	Edit	Activate/Deactivate
Administrator	admin	Password Preferences	<input type="checkbox"/>
lab_user	user	Password Preferences	<input checked="" type="checkbox"/>
sr	user	Password Preferences	<input checked="" type="checkbox"/>
sy	user	Password Preferences	<input checked="" type="checkbox"/>
System	guest	Password Preferences	<input type="checkbox"/>

User Inactivation
Are you sure you want to inactivate sy?



Web Services

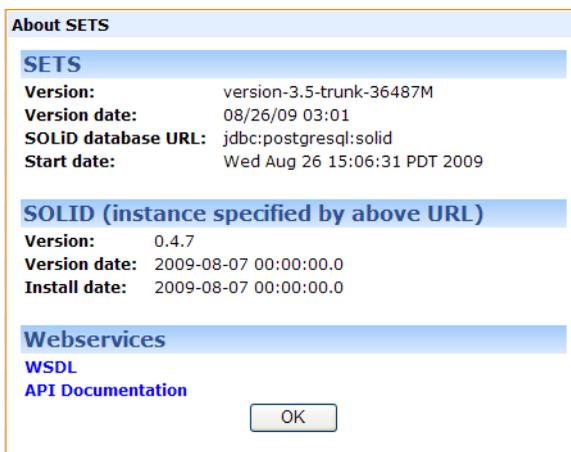
Overview

Within SETS, you can view the status of runs and control the analysis. SETS also exposes some of its functionalities through web services. Web services provides an open channel that allows you to integrate directly with the system.

Client code Web-based API is a method used to send information from one computer system to another. Write web service client code, then point the client (computer/analyzer system) to the SETS software WSDL.

WSDL

WSDL defines an API that can be used to programmatically retrieve information from the SOLiD™ 4 System. The WSDL definition (for SETS web services) is available through the SETS About menu.



The URL is:

```
[SETS service URL]/webservice/solid?wsdl
```

For example, if SETS is installed on a host named "foo," the [SETS service URL] might be:

```
http://foo:8080/sets.
```

The corresponding WSDL definition is available at:

```
http://foo:8080/sets/webservice/solid?wsdl
```

API documentation

On the iExternalService, go to the Method Summary table to see the services available.

Method Summary	
ClusterStatus	getClusterStatus () Get information about the status of the cluster.
java.util.List<Run>	getLastStartedRuns () Get the runs last started on flowcell1 and flowcell2.
java.util.List<LogMessage>	getLogMessages (java.lang.String severity, java.lang.String component, java.lang.String startDate, java.lang.String endDate, java.lang.String cleared) This method retrieves a list of log messages.
java.util.List<Run>	getRun (java.lang.String runName, java.lang.String instrumentName) Get a run by instrument and runName.
java.util.List<Run>	getRunsByStatus (java.lang.String status) Get runs that match one of the statuses listed below.

Binding A Java binding to the web services is available on request. Bindings for other languages can be readily generated based on the SETS web services WSDL definition.



B

Advanced Topic: Data Analysis Overview

The topics provided in this appendix are intended for advanced users of the SOLiD™ 4 System and do not apply to the typical user.

Fundamentals of color-space analysis

The Applied Biosystems SOLiD™ 4 System sequencing technology is based on sequential ligation of dye-labeled oligonucleotides. This technology makes possible massive parallel sequencing of clonally amplified DNA fragments. Features of this system, such as mate-paired analysis and 2-base encoding, enable studies of complex genomes by providing a greater degree of accuracy. This section describes the principles of 2-base encoding and the benefits of performing analysis in the di-base alphabet, known as *color-space analysis*.

The 2-base color coding scheme

Until recently, most DNA sequencing was performed using the chain termination method developed by Frederick Sanger. (Refer to the paper by Sanger F., Coulson A. R., 1975, A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 94(3): 441-448.) This type of sequencing is often referred to as *Sanger sequencing*. Sanger sequencing data is also encoded in color-space by the four fluorescent dyes used in the sequencing chemistry and displayed as peaks in an electropherogram. In Sanger sequencing, each color, representing only a single nucleotide, is automatically translated to A, C, G, or T. With the SOLiD™ 4 System, each color represents four potential 2-base combinations (see [Figure 1](#)). The conversion into nucleotide base space is usually done after the sequence is aligned to a reference genome transcribed in color-space. As an alternative, translation can occur following the generation of a consensus sequence.

[code]	0	1	2	3
[dye]	FAM	Cy3	TXR	Cy5
(XY) ₁	AA	AC	AG	AT
(XY) ₂	CC	CA	GA	TA
(XY) ₃	GG	GT	CT	CG
(XY) ₄	TT	TG	TC	GC

Figure 1 SOLiD™ 4 System's 2-base color-coding scheme

Color-coding process

The column under code i (0, 1, 2, or 3) lists the corresponding dye and the di-base probes (adjacent nucleotides) encoded by color i . For example, GT is labeled with Cy3® dye and coded as “1”.

The instrument takes the following steps to encode a DNA sequence. Consider the example ATCAAGCCTC:

1. Start at the 5' end.
2. Replace the di-base AT at this position with its corresponding code 3 from the table.
3. Advance by one base, which shows the TC di-base (code 2).
4. Continue to advance by one base, as shown below.

```
Base Sequence: A T C A A G C C T C
Color String:   3 2 1 0 2 3 0 2 2
```

This process encodes a k -mer of bases as a $(k-1)$ -mer of colors. Although this color string codes for four different k -mers, knowledge of the type and position of any of its k bases helps to encode the sequence. For SOLiD™ 4 System sequencing applications, the

instrument prepends the leading base A in the example above to result in a k -mer A321023022 to reconstruct the base sequence. The SOLiD™ 4 System generates its reads in this encoded form, as shown in [Figure 2](#).

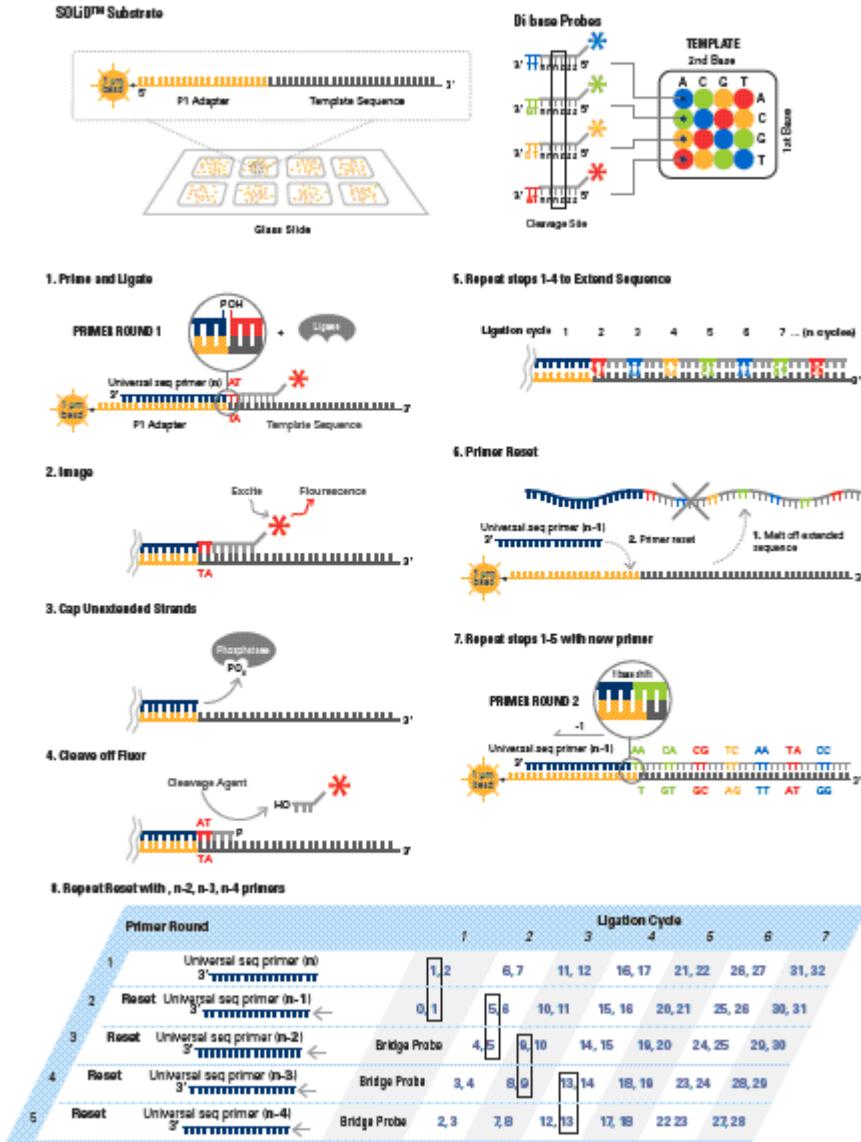


Figure 2 Sequence by ligation using di-base labeling probes on the SOLiD™ 4 System

Required properties for a 2-base color code scheme

The 2-base color-coding scheme satisfies the requirements in the following list. This conclusion can be observed by treating the properties as requirements and constructing the color code from them. This section addresses only bases, not other International Union of Biochemistry (IUB) codes. Let $B = \{A, C, G, T\}$.

The color code should satisfy the following requirements. For all bases b, d, e in B :

1. The available colors are 0, 1, 2, and 3.
 $\text{color}(bd) \in \{0, 1, 2, 3\}$.
2. Two different di-base probes that have the same first base result in different colors.
 $\text{color}(bd) \neq \text{color}(be)$ if $d \neq e$.
For example, $\text{color}(AC) \neq \text{color}(AG)$.
3. A di-base probe and its opposite result in the same color.
 $\text{color}(bd) = \text{color}(db)$.
For example, $\text{color}(AC) = \text{color}(CA)$.
4. Mono- and di-base probes result in the same color.
 $\text{color}(bb) = \text{color}(dd)$.
In other words, $\text{color}(AA) = \text{color}(CC) = \text{color}(GG) = \text{color}(TT)$.
The following are interesting properties that follow from the above four requirements. Property 5 follows from requirements 2 and 3 and makes constructing the color code easier.
5. Two different di-base probes that have the same second base result in different color codes: $\text{color}(bd) \neq \text{color}(cd)$, if $b \neq c$.
For example, $\text{color}(AC) \neq \text{color}(TC)$.
Property 6 also follows from requirements 1-4, but it is most easily verified against the completed code (see [Figure 3](#), Panel E).

- A di-base probe and its complement result in the same color.
 $\text{color}(\text{bcdc}) = \text{color}(\text{dcbc})$.
For example, $\text{color}(\text{AC}) = \text{color}(\text{TG})$.

How the system satisfies 2-base coding system requirements

[Figure 3](#) lists the colors for each di-base probe. For example, the value in row C and column T is the color 2 for di-base CT.

Requirements 1 and 2 require that all colors are present in the first row. Because the system can use any one-to-one mapping between the actual dyes and the labels 0, 1, 2, and 3 (provided that requirements 1 and 2 are satisfied), the first row can be labeled as shown in [Figure 3](#), Panel B.

Requirement 3, that $\text{color}(\text{bd}) = \text{color}(\text{db})$, gives a unique labeling for column A (see [Figure 3](#), Panel C).

Requirement 4, that $\text{color}(\text{bb}) = \text{color}(\text{AA})$, gives a unique labeling for the diagonal (see [Figure 3](#), Panel D).

Finally, requirements 1, 2, and 5 state that every color must appear in every row and every column exactly once (see [Figure 3](#), Panel E).

The table (see [Figure 3](#), Panel E) is easy to memorize and work with because, by virtue of Property 3, di-base probes can be thought of as two-element sets for assigning colors. The di-base probes that start with A result in colors 0, 1, 2, and 3 respectively.

Therefore:

- AA, CC, GG, TT all are assigned color 0.
- AC and CA are assigned color 1, and so must GT and TG.
- AG and GA are assigned color 2, and so must CT and TC.
- AT and TA are assigned color 3, and so must CG and GC.

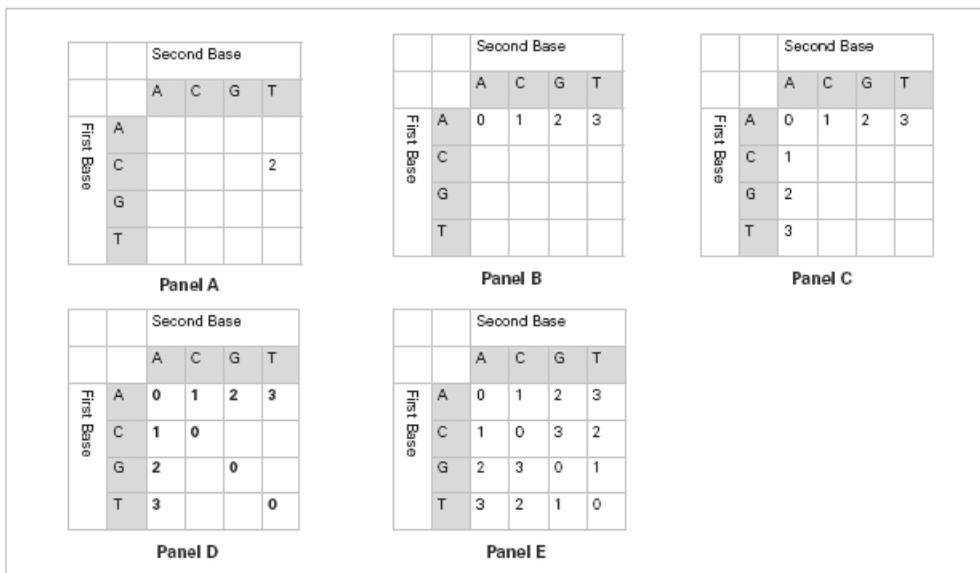


Figure 3 Requirements that assign color for a 2-base code

Principles of ligation-based chemistry and 2-base encoding

The SOLiD™ 4 System's sequencing technology is based on sequential ligation of dye-labeled oligonucleotide probes. Each probe assays two base positions at a time (Figure 2). The system uses four fluorescent dyes to encode for the sixteen possible 2-base combinations. Multiple ligation cycles of probe hybridization, ligation imaging, and analysis are performed to extend the strand from a primer hybridized to a ligated adaptor by the immobilized bead (P1 adaptor). The resulting product is then removed and the process repeated for 5 more rounds with primers hybridized to positions n-1, n-2, and so on, in the P1 adaptor.

There are several fundamental properties unique to ligation-based sequencing. These properties contribute to the high accuracy inherent in the SOLiD™ 4 System. The advantages of these properties and their contribution to data quality are:

- Two bases are interrogated in each ligation reaction, increasing specificity.

- The primer is periodically reset for five or more independent rounds of extension reactions, improving the signal-to-noise ratio of the system.
- Each base is interrogated twice in two independent primer rounds, increasing confidence in each call.
- Four dyes are used to encode for sixteen possible 2-base combinations. The design of the encoding matrix enables built-in error-checking capability.

Color-space and base space as applied in the SOLiD™ 4 System

The final files generated by the SOLiD™ 4 System are in base space. To maximize the built-in error correction of 2-base encoding, all file analysis is conducted in color-space prior to final results.

Color-space data

Rather than reading one base per cycle, the software measures information on two bases simultaneously. In each cycle, the software calls one of four colors (color-space call). Because each ligation measurement event measures two bases, all bases (except for the final base of a read) are interrogated twice, providing for an additional level of error correction.

Relationship between cycle and base position

Figure 4 shows the relationship between color-space, base position, and the sequencing chemistry. *Base number* refers to base position. Base 0 is the last base of the adapter and is not part of the target sequence. The five primer lines show the order in which the data was generated (0-indexed).

Note: In all files, only processed data refers to a color-space position.

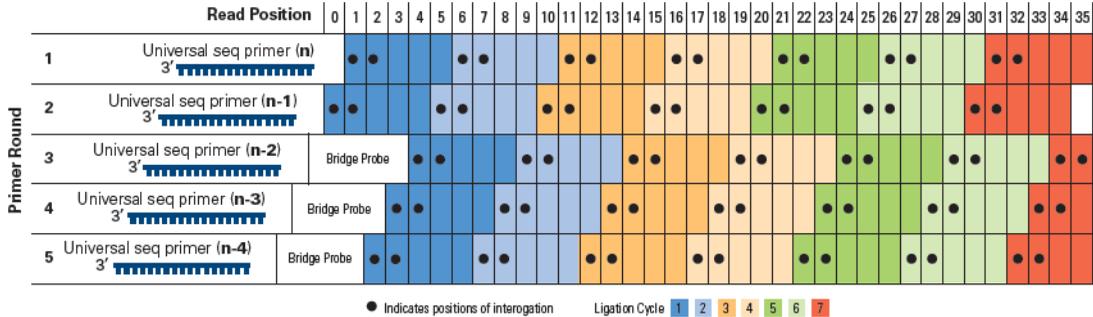


Figure 4 Relationship between color-space, base position, and sequencing chemistry

Color-space formats

Color-space data are presented in three slightly different formats. In two of the formats, a base (A, C, G, or T) is appended to the color-space calls.

Note: Color-space data are self-complementary: In some situations, when you might expect to see complemented data (for example, reverse), the data appear the same. For example, AC = 1, TG = 1.

The different types of color-space data are:

- **Processed color-space data:** Consists of a numeric string prefixed (suffixed if reversed) by a single base. The base that precedes the numeric (color code) data is the first base of the actual sequence (in base space, not color-space).
- **Unprocessed color-space data:** Consists of a numeric string prefixed by a single base. This base is the final base of the sequencing adapter and is not part of the target sequence. It is included to disambiguate the first color call.

Complementing color-space data

Color-space data are self-complementary as shown in [Figure 5](#).

		2 nd Nucleotide			
		A	C	G	T
1 st Nucleotide	A	0	1	2	3
	C	1	0	3	2
	G	2	3	0	1
	T	3	2	1	0

Sequence:														
Base	A	G	C	T	C	G	T	C	G	T	G	C	A	G
Color-space	2	3	2	2	3	1	2	3	1	1	3	1	2	
Complemented:														
Base	T	C	G	A	G	C	A	G	C	A	C	G	T	C
Color-space	2	3	2	2	3	1	2	3	1	1	3	1	2	

Figure 5 Color-space data

2-base encoding and error recognition

The error-checking abilities of the 2-base encoding schemes have *not* been used to correct any of the data provided in these files.

Example:

Reference = 2 3 2 2 **3** 1 2 3 1 1 3 1 2

Observed = 2 3 2 2 **0** 1 2 3 1 1 3 1 2

Note: Notice the single color-space error.

In this example, a single color-space error occurs. The most likely explanation for the observed 0 is that it is a measurement error. Because a single color-space change is not allowed, a change to one of the adjacent bases is needed for a real SNP. This correction requires multiple measurement errors, leaving the most likely explanation that the 0 is a 3. The fact that the two surrounding bases are the same as the reference is further evidence that correcting the 0 to a 3 is acceptable. Any single color error can likely be corrected, especially when using multiple aligned reads.

Processing

Typically, a panel is imaged four times per cycle (once per channel). The color-space calls are carried out on a per-cycle basis. Each data point represents an individual bead with four intensity values. Beads are assigned a color-space call using a clustering algorithm. As part of the clustering process, data are scaled and baselined, quality values are assigned, and color call is assigned.

Results of primary analysis for each panel are stored in files with the “.spch” extension (Solid Panel Cache HDF5). The “.spch” file is used as a cache; that is, primary analysis both writes results to the file and reads results from the file as analysis proceeds on a cycle-by-cycle basis. Data within the spch file are organized by cycle. When all cycles have been analyzed, a final processing step is run to convert the data from the cycle-based format to the “read-based” ASCII formats (“.csfasta” and “.qual”).

The spch file is in the HDF5 format. Details on the format are available at:

<http://www.hdfgroup.org/HDF5/>.

Tools to view the contents and extract results are available at:

<http://www.hdfgroup.org/hdf-java-html/hdfview/>.

Prior to secondary analysis, results are filtered by removing all tags with missing data. Filtering removes all incomplete beads and reads with missing calls. The filtered data are then processed to generate the *xxxx.csfasta* file. As part of this process, a known base (from the adapter sequence) is prepended to the color-space data, allowing disambiguation of the first color call. Each read is treated separately, even if it is part of a mate-paired tag. The file contains the color-space read for each tag, and the color-space data are used as the basis of all subsequent results. The data are then aligned in color-space to the color-space reference sequence.

During analysis and alignment:

- Color-space reference sequences are derived by converting the base-space reference sequence to color-space (this process occurs internally; you supply the reference sequence in base space).
- All alignment is done in color-space.
- All the alignments are with individual tags; the mate-paired information is not used in the initial alignments.
- After the single reads are analyzed, a further round of analysis is performed in which the mate-pair information is used.
- The paired tags are analyzed in a two-step process:
 - a. The tags where both the forward and reverse tags passed (both tags were matched to a reference sequence) are analyzed.

- b. Those mate-pairs where only one of the tags matched to the reference sequence (in color-space) are re-analyzed, and the non-matching tag is compared to the reference sequence.

Because analysis can be constrained by the knowledge of the allowed distances, this process allows alignment with more disagreements and the use of a more CPU-intensive alignment algorithm.

Note: Where possible, all data are presented in a fasta-compatible format to facilitate use in a variety of applications. Final output files are in base space and relative to the submitted reference (for example, consensus file, variation file).

Data analysis considerations

Understand Ns and color-space

In the SOLiD™ 4 System, unknown color-space calls are represented as ‘.’ or *N*. By default, the system filters out bases with a single ‘.’ in the sequence and leaves only reads with color-space calls present at all positions. You can change this behavior by masking a specific position of a read. For example, if position 21 of a 35 bp run is missing color-space calls, you can mask this position.

Find Single Nucleotide Polymorphisms (SNPs)

SNPs are single base-pair mutations that usually consist of two alleles. Color-space rules allow the software to detect only valid SNP sites. According to the color-space rules, one isolated color change is always an error when you consider single nucleotide changes. Two adjacent color changes can be either an error or a valid SNP. To detect two adjacent SNPs, look for three adjacent color changes. There are very specific rules to determine which color changes are valid and which are not.

Color-space rules for SNP detection

In the simplest case, where you have a single base change, the following rules apply ([Figure 6](#)):

- For any given reference, there are only three valid double-color changes.
- For the other 12 possibilities, the 6 single-color and 6 two-color changes are invalid, since they would require multiple changes within that genomic region. For these complicated genomic changes, the software uses a more sophisticated algorithm.

Single SNP SitesIf 2 colors present: e.g. If 1 color present: e.g. 

- Reverse the colors



- Use other 2 colors, both combinations



- Can be the other 3 color pairs:

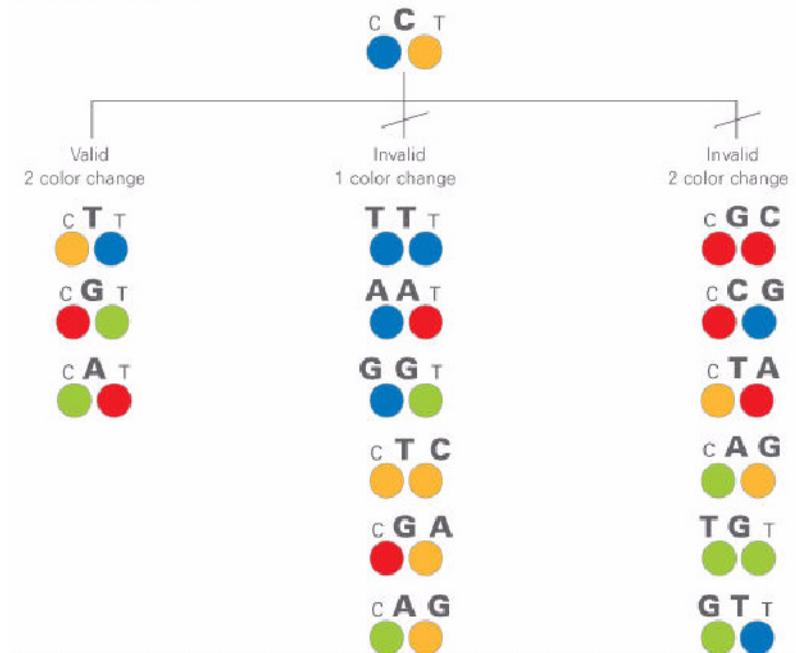
**Example using rules above**

Figure 6 Color-space rules for single nucleotide polymorphism (SNP) sites

Color-space rules are automatically applied during data analysis using the SOLiD™ 4 System's software analysis tools. You do not need to apply these filters manually.

SNPs and errors

Because a SNP generates two adjacent and valid color-space mismatches to the reference sequence, the software allows a minimum of two color-space mismatches during alignment to the reference sequence. Depending on the complexity of the alignment, the optimum number of mismatches in alignment is at least three. If only two errors are allowed, then whenever there is a SNP (two mismatches) and a measurement error, the third error causes omission of that read. You can override the omission by using an alignment program setting that counts two adjacent mismatches as a single mismatch. Therefore, two adjacent mismatches and a measurement error are classified as two errors, whereas three non-adjacent errors are classified as three errors.

High accuracy for SNP detection

2-base encoding provides high system accuracy and built-in error checking because it enables discrimination between measurement errors and true sequence polymorphisms. The SOLiD™ software interrogates each base twice in two independent reactions. Information about each base is included in two adjacent pieces of color-space data. Because the system uses four fluorescent dyes, there are 16 possible two-color combinations (Figure 7).

	Blue	Green	Yellow	Red
Blue	BB	BG	BG	BR
Green	GB	GG	GY	GR
Yellow	YB	YG	YY	YR
Red	RB	RG	RY	RR

Figure 7 Color-space transitions

A SNP must have two adjacent color changes. Single-color mismatches are not evidence of a SNP. This feature allows SOLiD™ software to distinguish errors in measurement from true SNPs.

SNP error rates

When you use 2-base encoding, any SNP in the original sequence is represented as two adjacent mismatches in color-space. Only three of nine possible adjacent mismatches can correspond to a real SNP. Suppose the raw sequencing error rate for a species is 1% per base, and for the same species that is sequenced, the SNP rate is about 0.1%. For a sequencing project of M total bases, there are about $0.001M$ real SNP occurrences in the data set. Among these SNPs, $0.001M \times 0.02$ total cases appear as single mismatches or two invalid mismatches because a sequencing error happens at one of the alleles of the SNP. Only 2% of the real SNPs fail to appear as two adjacent, valid mismatches. If only two adjacent, valid mismatches are treated as candidates for SNP detection, then there is a 2% false negative rate. For any data set, two adjacent and valid mismatches can be caused by sequencing errors. However, for a total of M bps sequenced, there are $0.00003M$ total occurrences of two adjacent mismatches. Note that there are about $0.001M$ adjacent valid mismatches from real SNPs, among all two adjacent, valid mismatches observed in a particular data set. Of these, 97% are from real SNPs, and only 3% may be from sequencing errors. This is a 97% true discovery rate for that particular data set. Because there are about $0.01M$ total sequencing errors in the data set caused by 2-base encoding and the software's ability to remove all single base mismatches, the error is reduced from $0.01M$ to $0.00003M$. This is a reduction of 300 times, making the effective error rate 0.003%. This calculation illustrates the power of 2-base encoding in resequencing and finding SNPs.

Sampling

If coverage at a genome position is low, the second allele might often not be sampled, even if it is present. Heterozygotes are expected to be underrepresented at low coverage. (They can be called as a homozygote for one of the two alleles.) It is important to have a low false positive rate for SNP detection, because SNPs are expected to exist at only 1 in 1,000 positions for humans and some other species. The false positive rate in an individual species should be at least an order of magnitude lower than this, to avoid a high false discovery rate. An error model that incorporates quality values (QV) can increase the overall accuracy of SNP detection. The SNP detection algorithms of the SOLiD™ 4 System use explicit error models generated from each run and biologically meaningful prior probabilities to evaluate evidence of a SNP. 2-base encoding provides a built-in way to distinguish errors from true alleles to manage the false positive rate.

Allele ratio

If the sample preparation method or some other cause produces results in allele ratios that are considerably different from the expected 50:50 ratio, it is more difficult to detect heterozygosity. Detecting heterozygosity then requires more sequence coverage.

Find polymorphisms in color-space

The SOLiD™ 4 System can detect complicated genomic variations such as adjacent SNPs, insertions, deletions, and structural rearrangements.

For other analysis tools, visit the AB SOLiD™ Software Community website:

http://www3.appliedbiosystems.com/AB_Home/applicationstechnologies/SOLiDSYSTEMSequencing/SoftwareCommunityDataAnalysisResourcesforScientistsDevelopers/index.htm

An example of various polymorphisms in color-space is shown in the figure below.

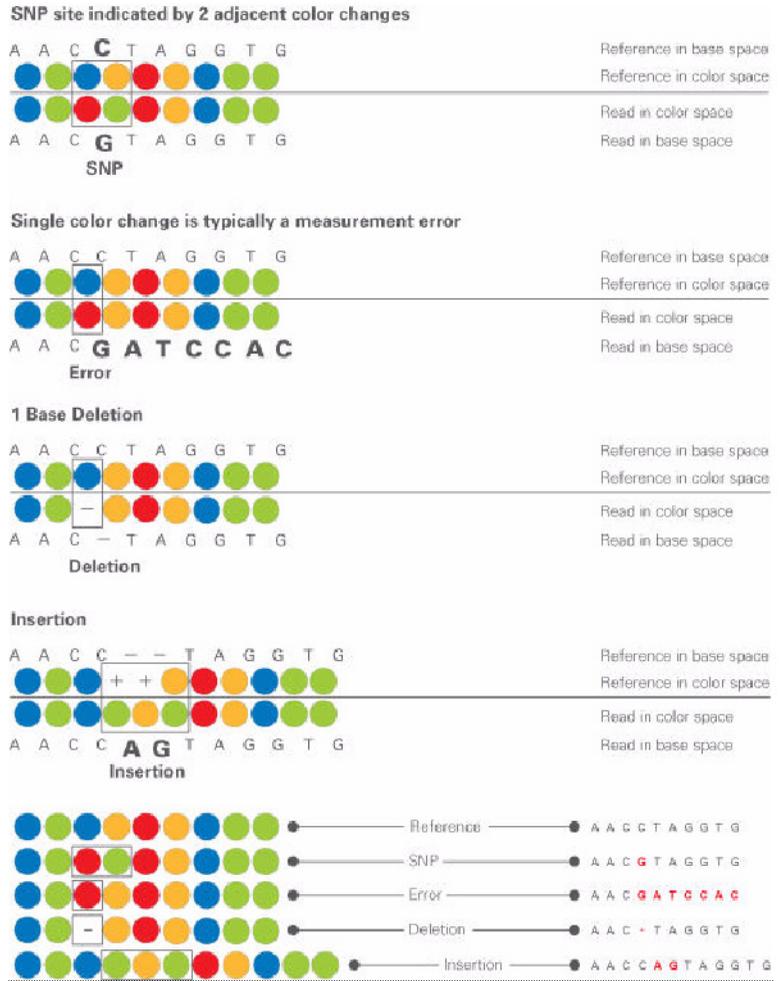


Figure 8 Examples of polymorphisms in color-space



Advanced Topic: Data Management

The topics provided in this appendix are intended for advanced users of the SOLiD™ 4 System and do not apply to the typical user.

Saving data

This appendix describes data storage for the following:

- All data
- Raw image data
- Primary analysis results data

Approximate storage space for all data

The following table shows the size of all data for archival and storage consideration. The following is estimated file sizes.

Table 1 Total Data size for archival and storage consideration

50 nt tag/ 300k/panel, 2357 panels	Image data size[‡]	Primary analysis results size in “.spch” format	Primary Analysis Data size (flatfile: “.csfasta”, “_QV.qual”, “.stats”)
1 slide -1 tag	1.84 TB	646 GB	170 GB
1 slide - 2 tags	3.6 TB	1.29 TB	340 GB
2 slides - 1 tag	3.6 TB	1.29TB	340 GB
2 slides - 2 tags	7.2 TB	2.58 TB	680 GB

‡ Image data is not needed after analysis is complete.



Raw image data

The data should be kept until the full completion of primary analysis is verified and primary analysis results are archived. After confirming that the image files are no longer needed, you can delete these image files to make room for additional sequencing runs.

Primary analysis results data

The image analysis results are stored in “.spch” files, which are located in:

`$sampleResultFolder/colorcalls/CACHE/.`

Table 2 Image analysis results files

File type	File size	Comments
.spch	646 GB	<ul style="list-style-type: none"> • Should be archived for the lowest-level non-image backup. An estimate for the disk space needed for one “.spch” file (per panel) is $n*(18*m+17)$ bytes, where n is the number of beads, and m is the number of cycles (F3 and R3). • For 50 cycles (1x50) and 300K beads, this yields 275 MB (per panel). • A full slide has 2357 panels, which means that 646 GB are required for a full slide (50 bp single tag run).

The color call results, including color-space raw reads and QV data, are located in:

`$sampleResultFolder/primary.[17-digit timestamp]/reads/.`

Table 3 Color call results files

File type	File size	Comments
.csfasta	56 - 80 GB	Should be archived.
_QV.qual	112 - 140 GB	Should be archived.
.stats	< 10 KB	Should be archived.
postPrimerSetP primary.[JobID]	364 - 420 GB	Temporary files. They are removed after completion. Located in <code>\$sampleJobFolder/.</code>



There are also color-space raw reads that were filtered in the sequence generation process. These raw reads include duplicate reads and reads with less than the expected read length. They can be found in:

`$sampleResultFolder/primary.[17-digit timestamp]/reject/`

File type	File size	Comments
.csfasta.reject	< 1 GB	Should be archived.
_QV.qual.reject	< 1 GB	Should be archived.

The following files can be ignored. They can be archived if needed.

File type	File size	Comments
Intensity files	4 x 336 - 420 = 1344- 1680 GB	Optional. Not generated by default. \$sampleResultFolder/primary.[17digit timestamp]/reads/ or \$sampleResultFolder/primary.[17digit timestamp]/reject/

Data transfer

Data transfer from the SOLiD™ 4 System is supported and tested on a 1-GB LAN. Applied Biosystems strongly recommends that you set up a dedicated 1-GB network between the SOLiD™ 4 System and the rest of your network. To ensure the appropriate transfer speed, you need to configure your network according to one of the suggested solutions:

Current network	Solution
1-GB LAN (shared)	On a shared network, the effective network speed is much lower than 1GB. Use or install a <i>dedicated</i> 1-GB network between the instrument and the rest of your network to ensure top transfer speeds.
No Gigabit LAN	Install a gigabit switch with the shortest path spanning tree. Connect all the SOLiD™ 4 Analyzer outgoing connections to the switch so that data transfer between the SOLiD™ 4 Analyzer and the rest of the network <i>only</i> occurs through the extra switch at 1 Gbps.

Exporting data

You export data from the SOLiD™ 4 Analyzer using SOLiD™ 4 Experiment Tracking System (SETS) software. You can set up SETS to auto-export data (csfasta and “.spch” files) for every run, or per run. You can also export files manually to a specified destination.

For detailed procedures, see [Chapter 7, “Manage Administrative Tasks”](#) on page 113.



Advanced Topic: Primary Analysis

The topics provided in this appendix are intended for advanced users of the SOLiD™ 4 System and do not apply to the typical user.

What is primary analysis?

Primary analysis is the process of extracting color calls or assignments (.csfasta file) from images captured from a sequencing run, calculating a quality value for each color call, and writing the color-call value to a QV file (.qual).

Primary analysis also provides real-time feedback about a sequencing run.

Primary analysis overview

Table 4 lists the stages in primary analysis, their inputs, and their outputs.

Table 4 Primary analysis stages, inputs, and outputs

Analysis stage	Inputs	Outputs	Steps
postBeadFinding Primary	Focal map and first ligation cycle ".tiff" images	<ul style="list-style-type: none"> • Bead location (in ".spch" files) • Image statistics and metrics (in *_FOCALMAP_summary.tab files) 	For each panel, analyze the focal map and first ligation cycle images to identify beads.
postScanSlide Primary	<ul style="list-style-type: none"> • .tiff images from each ligation cycle • Focal map ".tiff" images 	<ul style="list-style-type: none"> • Intensities, Color calls, Quality metrics, QV, in ".spch" files • Satay ".png" files • Summary data (tab) • Run status info (traffic lights) (tab). 	<ol style="list-style-type: none"> 1. Align images. 2. Extract intensities. 3. Classify colors (color calls). 4. Calculate quality value (QV). 5. Generate satay images. 6. Accumulate run status information.
postPrimerSet Primary	All ".spch" files	<ul style="list-style-type: none"> • Color-space raw reads (".csfasta") • QV (".qual") • Reads statistics file (".stats" file). 	<ol style="list-style-type: none"> 1. Assemble color-space reads per panel. 2. Concatenate per-panel reads into one single ".csfasta" file (optional). 3. Filter reads shorter than specified read length or duplicate reads (optional).

Primary analysis workflow

In the primary analysis workflow, each spot on the sample slide is broken down into a number of panels. Each panel is defined by the area that is imaged by the camera in one exposure. Primary analysis handles each panel independently of the others. For each panel, primary analysis comprises a sequence of steps. When the Instrument Control Software (ICS) has finished collecting the images for each step, it creates a job workflow in the SOLiD™ 4 System database. The Job Manager spots the job workflows and prepares and submits the needed jobs to the PBS resource manager. When a job completes, the Job Manager updates the SOLiD™ 4 System database with the appropriate analysis status.

Each primary analysis step falls into one of the following three classes:

- **Bead finding:** Focal map and first ligation cycle images are analyzed to identify and locate beads. Additionally, various image statistics and metrics are computed and recorded.
- **Color calling:** After a ligation cycle, the four color images are registered to the focal map image. The fluorescent intensity for each color is estimated for each bead. These color intensities are analyzed to determine a color call and quality value for each bead for that cycle.
- **Read building:** After a complete run, the cycle-wise results are assembled into color reads. After some optional filtering, these reads are then ready for subsequent secondary analysis (e.g., mapping to a reference sequence).

See [Figure 9](#) for the primary analysis workflow. Multiple primary analysis jobs can be run on the Linux computer cluster. The analysis results are written to various files on this cluster. Upon completion, the primary analysis output includes a summary statistics file that shows a summary of the overall color calls. This file can be used to evaluate whether primary analysis completed successfully.

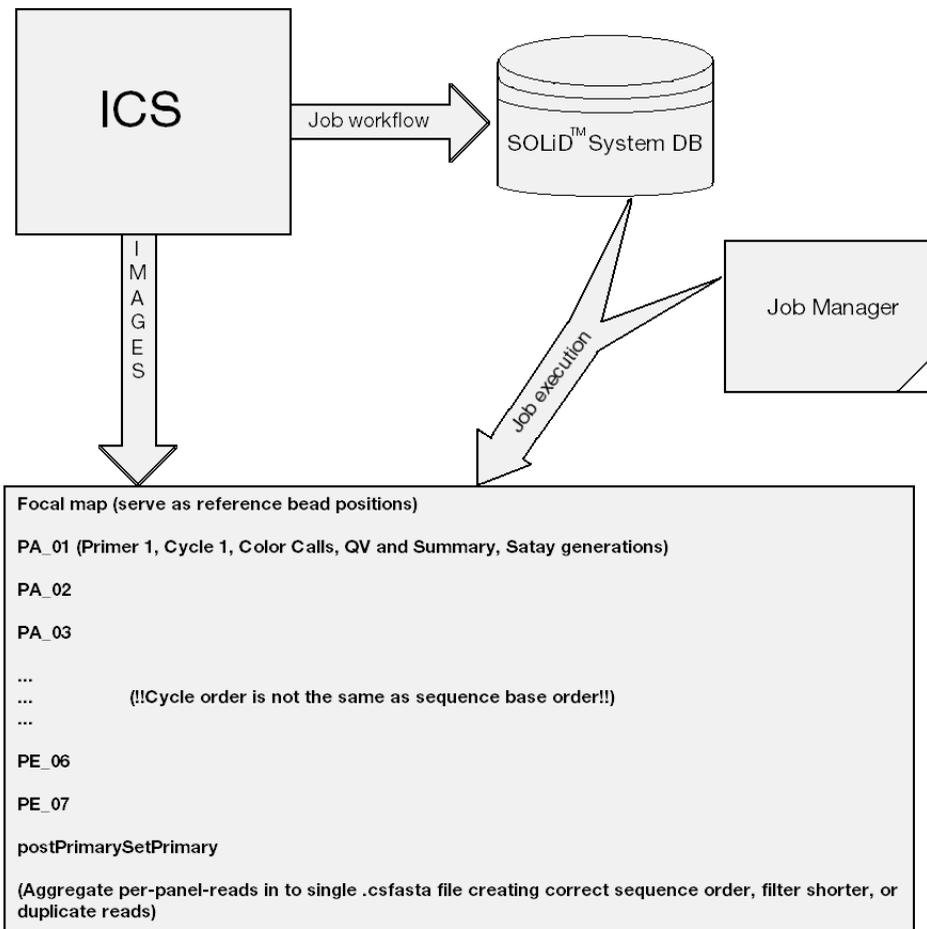


Figure 9 Primary analysis workflow

Primary analysis inputs and outputs

The input images are in the following files:

Focal Map

```
/data/images/{run.name}/{run.name}_FOCALMAP_V1
```

Ligation Cycle

```
/data/images/{run.name}/{run.name}_F3_P1_01_V1
/data/images/{run.name}/{run.name}_F3_P1_02_V1
...
...
/data/images/{run.name}/{run.name}_R3_P1_01_V1
...
...
$sampleJobFolder =
/data/results/{instrument.name}/{run.name}/{sample.name}
/jobs
$sampleResultFolder =
/data/results/{instrument.name}/{run.name}/{sample.name}
/results
```

Job analysis parameter and temporary files for the Focal Map

```
$sampleJobFolder/postBeadFindingPrimary.[JobID]
```

Job analysis parameters and temporary files for ligation cycle color calls

```
$sampleJobFolder/postScanSlidePrimary.[JobID]
$sampleJobFolder/postScanSlidePrimary.[JobID]
...
...
```

Job analysis parameters and temporary files for postPrimerSetPrimary

```
$sampleJobFolder/postPrimerSetPrimary.[JobID]
```

Image analysis results ".spch" files (SOLiD Panel Cache HDF5)

```
$sampleResultFolder/colorcalls/CACHE/{run.name}_{panelID}.spch (one full slide has 2357 panels)
```

Satay ".png" files

```
$sampleResultFolder/cycleplots
```

Summary data

```
$sampleResultFolder/colorcall_summary/{run.name}_FOCALMAP_V1.tab
```

```
$sampleResultFolder/colorcall_summary/{run.name}_F3_P1_01_V1.tab
```

```
..
```

```
$sampleResultFolder/colorcall_summary/{run.name}_R3_P1_01_V1.tab
```

```
..
```

```
$sampleResultFolder/colorcall_summary/{run.name}_BC_P1_01_V1.tab
```

Run status information (traffic light)

```
$sampleResultFolder/traffic_lights/
```

Color-space F3 raw reads (.csfasta)

```
$sampleResultFolder/primary.[17-digittimestamp]/reads/{run.name}_{sample.name}_F3.csfasta
```

Color-space BC raw reads (.csfasta)

```
$sampleResultFolder/primary.[17-digittimestamp]/reads/{run.name}_{sample.name}_BC.csfasta
```

Color-space F3 raw reads for library (.csfasta)

```
$sampleResultFolder/libraries/{library.name}/primary.[17-  
digittimestamp]/reads/{run.name}_{sample.name}_F3.csfas  
ta
```

Color-space BC raw reads for library (.csfasta)

```
$sampleResultFolder/libraries/{library.name}/primary.[17-  
digittimestamp]/reads/{run.name}_{sample.name}_BC.csfast  
a
```

F3 QV (.qual) file for library

```
$sampleResultFolder//libraries/{library.name}/primary.[1  
7-  
digittimestamp]/reads/{run.name}_{sample.name}_F3_QV.qua  
l
```

BC QV (.qual) file for library

```
$sampleResultFolder//libraries/{library.name}/primary.[1  
7-  
digittimestamp]/reads/{run.name}_{sample.name}_BC_QV.qua  
l
```

F3 Primary analysis statistics file (.stats)

```
$sampleResultFolder/primary.[17-  
digittimestamp]/reads/{run.name}_{sample.name}_F3.stats
```

BC Primary analysis statistics file (.stats)

```
$sampleResultFolder/primary.[17-  
digittimestamp]/reads/{run.name}_{sample.name}_BC.stats
```

F3 Primary analysis statistics file (.stats) for library

```
$sampleResultFolder/libraries/{library.name}/primary.[17-  
digittimestamp]/reads/{run.name}_{sample.name}_F3.stats
```

BC Primary analysis statistics file (.stats) for library

```
$sampleResultFolder/libraries/{library.name}/primary.[17-  
-  
digittimestamp]/reads/{run.name}_{sample.name}_BC.stats
```

Reads that have less than expected read length or duplicated

```
$sampleResultFolder/primary.[17-digit timestamp]/reject
```

Reads that have less than expected read length or duplicated for library

```
$sampleResultFolder/libraries/{library.name}/primary.[17-  
-digit timestamp]/reject
```

Reads without barcode reads

```
$sampleResultFolder/libraries/{library.name}/missing-bc  
$sampleResultFolder/libraries/{library.name}/missing-f3
```

Reads that could not be assigned to a library

```
$sampleResultFolder/libraries/{library.name}/unassigned
```

Barcode assignment statistics file/report

```
$sampleResultFolder/libraries/BarcodeStatistics.[17-  
digittimestamp].txt
```

Storage requirements for primary analysis

The storage requirements values given in the table below assume a 50-base tag run in one flowcell. If you use a tag size that is larger or smaller than 50 bases, the storage requirements differ from those given in [Table 5](#).

Table 5 Primary analysis storage requirements

Files	Storage required	Notes about archiving
.csfasta	56 - 80 GB	Should be archived.
_QV.qual	112 - 140 GB	Should be archived.
.spch	646 GB	Should be archived for the lowest-level non-image backup. An estimate for the disk space needed for one ".spch" file (per panel) is $n*(18*m+17)$ bytes, where n is the number of beads, and m is the number of cycles (F3 and R3). For 50 cycles (1x50) and 300K beads, this yields 275 MB (per panel). A full slide has 2357 panels, which means that 646 GB are required for a full slide (50 bp single tag run).
postPrimerSetPrimary.[JobID]	364 - 420 GB	Temporary files. Most of these are removed upon completion.
Intensity files	4 x 364 - 420 = 1344 - 1680 GB	Optional. Not generated by default. It is strongly recommended that intensity files be generated with offline ReadBuilder, which is part of the BioScope™ Software.

Perform primary re-analysis

To perform primary analysis manually, use the following command:

```
jprimaryanalysis.sh --ini=PrimaryParameters.ini
```

Note: Only buildReads tasks are executed for re-analysis. Neither beadFinding nor colorCalling are enabled for re-analysis in the pipeline.

The PrimaryParameters.ini file can be found inside the JobID folder with the file named **parameters.ini**.

```
$sampleJobFolder/postPrimerSetPrimary
```

To prepare for the re-analysis:

1. Create a new analysis directory, such as:
postPrimerSetPrimary.m1.
2. Copy the following file onto the new directory.
parameters.ini
3. Edit the following file to ensure that all the paths are correct.
PrimaryParameters.ini

4. Remove parameters marked for daemon property.

For example:

```
analysis.job.id
```

(Default behavior does not accept non-integer.)

```
update.solid.script
```

(Used by the daemon for updating the SOLiD™ 4 System database.)

```
logging.socket
```

(Used by the daemon for logging analysis progress.)

5. Enter

```
jprimaryanalysis.sh --ini=PrimaryParameters.ini
```



The following is an example of the PrimaryParameters.ini file:

```
#####
##- Primary Analysis Settings
#####

##- Instrument, sample, etc
instrument.name=Solid
run.name=Solid_20080310_1
sample.name=DH10B

panels=1-2357
primer.set=F3
read.length=35
read.prefix=T

#####
##- Analysis Job Parameters
focal.map.dir=Solid_20080310_1/Solid_20080310_1_FOCALMAP_V1
focal.map.stg=Solid_20080310_1/Solid_20080310_1_FOCALMAP_V1/Solid_20080310_1_FOCALMAP_V1.STG
focal.map.version=1

##- position map. Used for generating correct sequence order
##-
images.dir.1=Solid_20080310_1/Solid_20080310_1_F3_P5_01_V1
images.dir.2=Solid_20080310_1/Solid_20080310_1_F3_P4_01_V1
images.dir.3=Solid_20080310_1/Solid_20080310_1_F3_P3_01_V2
```

```
images.dir.4=Solid_20080310_1/Solid_20080310_1_F3_P2_01_
V1
images.dir.5=Solid_20080310_1/Solid_20080310_1_F3_P1_01_
V1
images.dir.6=Solid_20080310_1/Solid_20080310_1_F3_P5_02_
V1
images.dir.7=Solid_20080310_1/Solid_20080310_1_F3_P4_02_
V1
images.dir.8=Solid_20080310_1/Solid_20080310_1_F3_P3_02_
V1
images.dir.9=Solid_20080310_1/Solid_20080310_1_F3_P2_02_
V1
images.dir.10=Solid_20080310_1/Solid_20080310_1_F3_P1_02_
_V1
images.dir.11=Solid_20080310_1/Solid_20080310_1_F3_P5_03_
_V1
images.dir.12=Solid_20080310_1/Solid_20080310_1_F3_P4_03_
_V1
images.dir.13=Solid_20080310_1/Solid_20080310_1_F3_P3_03_
_V1
images.dir.14=Solid_20080310_1/Solid_20080310_1_F3_P2_03_
_V1
images.dir.15=Solid_20080310_1/Solid_20080310_1_F3_P1_03_
_V1
images.dir.16=Solid_20080310_1/Solid_20080310_1_F3_P5_04_
_V1
images.dir.17=Solid_20080310_1/Solid_20080310_1_F3_P4_04_
_V1
images.dir.18=Solid_20080310_1/Solid_20080310_1_F3_P3_04_
_V1
images.dir.19=Solid_20080310_1/Solid_20080310_1_F3_P2_04_
_V1
images.dir.20=Solid_20080310_1/Solid_20080310_1_F3_P1_04_
_V1
images.dir.21=Solid_20080310_1/Solid_20080310_1_F3_P5_05_
_V1
```



```
images.dir.22=Solid_20080310_1/Solid_20080310_1_F3_P4_05_V1
images.dir.23=Solid_20080310_1/Solid_20080310_1_F3_P3_05_V1
images.dir.24=Solid_20080310_1/Solid_20080310_1_F3_P2_05_V1
images.dir.25=Solid_20080310_1/Solid_20080310_1_F3_P1_05_V1
images.dir.26=Solid_20080310_1/Solid_20080310_1_F3_P5_06_V1
images.dir.27=Solid_20080310_1/Solid_20080310_1_F3_P4_06_V1
images.dir.28=Solid_20080310_1/Solid_20080310_1_F3_P3_06_V1
images.dir.29=Solid_20080310_1/Solid_20080310_1_F3_P2_06_V1
images.dir.30=Solid_20080310_1/Solid_20080310_1_F3_P1_06_V1
images.dir.31=Solid_20080310_1/Solid_20080310_1_F3_P5_07_V1
images.dir.32=Solid_20080310_1/Solid_20080310_1_F3_P4_07_V1
images.dir.33=Solid_20080310_1/Solid_20080310_1_F3_P3_07_V1
images.dir.34=Solid_20080310_1/Solid_20080310_1_F3_P2_07_V1
images.dir.35=Solid_20080310_1/Solid_20080310_1_F3_P1_07_V1

#####
##- Daemon Properties
##- These are optional
##-
analysis.ini=/data/results/Solid/Solid_20080310_1/DH10B/
jobs/postPrimerSetPrimary.m1/PrimaryParameters.ini
analysis.name=Filter_Fasta
```

```
analysis.pipeline=Class:com.apldbio.aga.analysis.primary
.core.PrimerSetPrimaryStage

analysis.stage.name=postPrimerSetPrimary

#####

##- Run Directories

##- The analysis output directories. Make sure that
these are correct.

##-

analysis.run.dir=/data/results/Solid/Solid_20080310_1/DH
10B

analysis.results.dir=/data/results/Solid/Solid_20080310_
1/DH10B/results.01

summary.dir=/data/results/Solid/Solid_20080310_1/DH10B/r
esults.01/basecall_summary

colorcallfolder=/data/results/Solid/Solid_20080310_1/DH1
0B/results.01/colorcalls

cycleplots.dir=/data/results/Solid/Solid_20080310_1/DH10
B/results.01/cycleplots

analysis.sample.dir=/data/results/Solid/Solid_20080310_1
/DH10B

analysis.work.dir=/data/results/Solid/Solid_20080310_1/D
H10B/jobs/postPrimerSetPrimary.m1

reads.result.dir.1=/data/results/Solid/Solid_20080310_1/
DH10B/results.01/primary.m1/reads

read.dir=/data/results/Solid/Solid_20080310_1/DH10B/resu
lts.01/primary.m1/reads
```



```
Average: 34762.8 / 38429.3 (90.6%): (Duplicates: 3.4  
[0.0%]) (TooManyErrors: 3663.1 [9.4%]) Usable (2251 p)  
(95.5%): 81935898 / 86395269 (94.8%): (Duplicates:  
8085.0 [0.0%]) (TooManyErrors: 4451286.0 [5.2%]) Usable  
Average: 36399.8 / 38380.8 (94.9%): (Duplicates: 3.6  
[0.0%]) (TooMnyErrors: 1977.5 [5.1%])
```

The last several lines show the summary of the primary analysis. If this file is empty or the last four lines are missing, the primary analysis has not finished, or it encountered a problem and terminated.

Image metrics

Image metrics are measures of the quality of the images produced by the SOLiD™ 4 Analyzer. The principal image metrics are the blur metric and the exposure metric.

Blur metric is a real number greater than zero that provides a measure of the average apparent radius of the beads in an image. The blur metric is computed for every image that is analyzed by primary analysis. The blur metric supports diagnosing instrument focusing and vibration issues. [Figure 10](#) gives two examples of the blur metric:

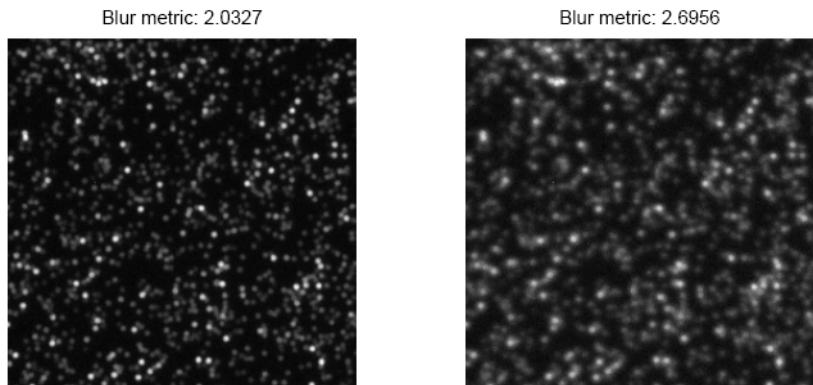


Figure 10 Two examples of the blur metric

The *blur population* reports the population of panels that have a blur value above a given threshold. A larger average size can be caused by a small number of panels with large outlier values or a large number of panels with a large size. The two values taken together can indicate whether there is a localized problem over the whole sample (larger blur value but no outliers), a regional problem (an outlier population that spans channels), or a transient problem (outliers in a single channel).

Exposure metric is a real number that measures whether and by how much an image is over- or under-exposed, relative to a particular definition of ideal exposure. A value of zero indicates ideal exposure. A value greater than zero indicates over-exposure, and one less than zero indicates under-exposure. The scale is logarithmic with base two; a value of 1 indicates that the image is exposed twice as long as ideal.

The typical ideal exposure level of an image is one that maximizes the dynamic range of the content, with some signal occupying the lowest (darkest) range and some information extending to the saturation (highest) point.

- An overexposed image compresses too much of the content at the top end of the range. Overexposure creates a non-linearity in response and makes it impossible to differentiate between the highest signals.
- An underexposed image does not extend the content of the image across the possible range. Underexposure creates a low-end compression and fails to use the available range of the system.

Note: For purposes of color-calling algorithms, slightly overexposing the images yields better results.

Figure 11 gives two examples of the exposure metric:

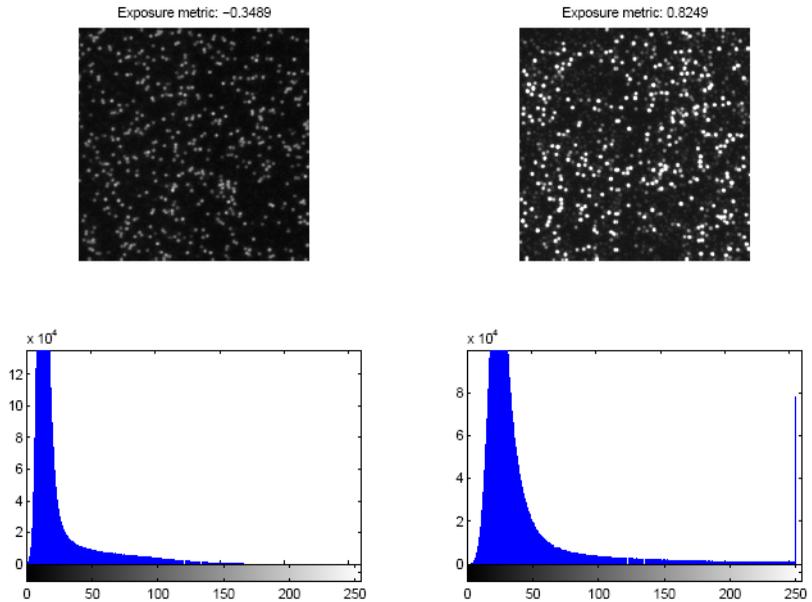


Figure 11 Two examples of the exposure metric

The exposure value calculated is a measure of how close the system is to capturing an image that maximizes the use of the dynamic range. Looking at the tail behavior of the pixel distribution makes the determination.

- If the tail does not extend to the saturation point, the image is underexposed.
- If the tail has a peak at the far right side that is a measurement of the saturation, the image is overexposed.

The *exposure population* reports the population of panels that have an exposure quality beyond a given threshold. A large average size can be caused by a small number of panels with large outlier values or a large number of panels with a large size. The two values taken

together can estimate whether there is a localized problem over the whole sample (larger deviation with no outliers), a regional problem (outlier population that spans channels), or a transient problem (outliers in a single channel).

A difference with the blur metric, however, is that whereas the blur metric is constant over the number of beads present in a panel, the exposure is more sensitive to the number of beads. A baseline exposure value per channel is determined based on a sampling strategy, but panels with either significantly more or fewer beads can be affected differently. As with the blur value, the reported value is the largest mean exposure value of the four channels over all panels in the sample.

Understanding quality values in the SOLiD™ System

The SOLiD™ primary analysis software analyzes the raw image data collected by the system and generates a color-space read for each tag. The software generates a quality value (QV) for each color call, which is an estimate of confidence for that color call.

The SOLiD™ 4 System QVs are similar to those generated by Phred and the KB™ Basecaller software for capillary electrophoresis sequencing. The quality value q for a particular call is mathematically related to its probability of error p .

$$q = -10 \log_{10} p$$

The algorithm used to predict QVs for the SOLiD™ 4 System is similar to Phred and the KB™ Basecaller software. (Refer to the paper, *Ewing B., Green P., 1998, Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Research 8:186-194.*)

The algorithm relies on training or calibration to a large set of control data and color calls for which the correct call is known. In the SOLiD™ 4 System, the correct call is determined by mapping the read to a known reference sequence.

Usage

Given a read with QVs, the expected number of errors in the read can be estimated. If the QV of the i -th call is q_i , then the expected number of errors in the read is:

$$m = \sum_{i=1}^n p_i$$

where

$$p_i = 1.00 \cdot 10^{-q_i/10}$$

$p_i = 1.00 \cdot 10^{-q_i/10}$ is the predicted probability of error for the i -th call. This probability includes the fact that QVs are conventionally rounded to the nearest integer. The average error rate for this read is m/n , which can be converted into an overall quality score for the whole read.

$$Q = -10 \log_{10} \frac{m}{n}$$

Validation and observed quality

The conventional method of validating QVs involves computing a so-called observed quality for all the calls with a particular predicted quality in a set of control data. The following figure shows such a validation.

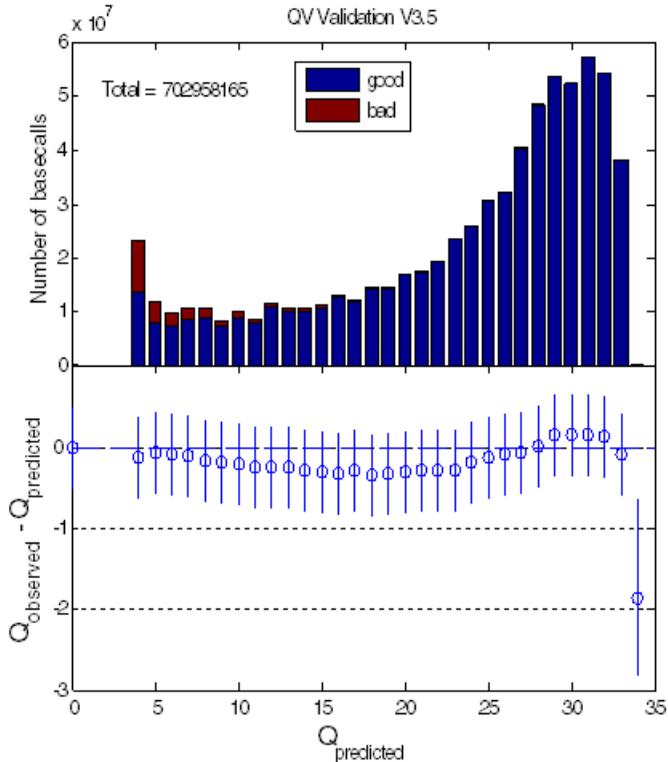


Figure 12 Statistical characteristics of quality values (QV) for the SOLiD™ 4 System

The top panel shows distribution of QVs. The bottom panel shows deviation of observed from predicted QVs.



Key files generated by primary analysis

Table 6 Key files generated by primary analysis

File name	Description
xxxx_sequence.csfasta	The raw data from all beads
xxxx_sequence.QV.qual	The quality values file
.spch files	Cache files for primary analysis

xxx_sequence.csfasta

The “.csfasta” file is a color-space fasta file that contains the color calls generated for each tag, with the last base of the primer prepended.

The format is:

```
>TAG_ID
Color_space
```

For example:

```
>1_88_1830_R3
G321113123201300232320
>1_89_1562_R3
G23133131233333101320
```

This file contains all the data that passed filtering, for complete reads (for example, 35 base read has 35 bases). This file is the input to the alignment tool.

xxx_sequence.QV.qual

The QV.qual files are fasta-like files that list the quality values in sequence order (not cycle order).

The format is:

```
TAG_ID
```

```
Quality Values:
```

For example:

```
>97_2040_1850_F3
```

```
38 36 26 33 41 26 24 33 28 31 27 23 5 35 32 31 11 10 24  
38 22 24 7 12 15 21 12 18 34 31 27 11 15 26 13 14 17 17  
13 12 8 5 17 5 12
```

```
>97_2040_1898_F3
```

```
41 41 41 38 32 29 39 24 23 36 32 38 25 30 28 21 27 33 34  
33 24 27 9 35 34 14 30 18 33 8 13 32 10 31 24 7 22 5 27  
30 21 5 0 27 9
```

All primary data (with exception of images) is stored in the file xxxx.spch. This file is in the HDF5 format. Tools to view and extract this data are available at:

<http://www.hdfgroup.org/HDF5/>.

The data within the ".spch" file is organized by cycle. You can extract data in any format for your own analysis pipelines.

.spch files

Results of primary analysis for each panel are stored in files with the ".spch" extension (Solid Panel Cache HDF5). The ".spch" file is used as a cache; that is, primary analysis both writes results to the file and reads results from the file as analysis proceeds on a cycle-by-cycle basis. Data within the ".spch" file is organized by cycle. When all cycles have been analyzed, a final processing step converts the data from the cycle-based format to the "read-based" ASCII formats (".csfasta" and ".qual").

".spch" files use the hierarchical binary format HDF5. Details on this format are available at:

<http://www.hdfgroup.org/HDF5/>.

Tools to view the contents and extract results are available at:

<http://www.hdfgroup.org/hdf-java-html/hdfview/>

Procedures to verify primary analysis status

To verify primary analysis status:

1. From SETS, click the **History** menu to access Run History view, then click the run name of interest.

The numbered items in the data tree on the left side of the page relate to the individual samples on the slide.

2. Click one of the data tree sample numbers on the left of the page to see the Library file name, Spot number, Sample name, and Reports information for that particular sample.

3. Click the drop-down arrow to the left of the sample folder to view individual analysis cycles for each sample.

4. Check that all the analysis completed successfully. A **blue dot** before the analysis name means that the analysis completed successfully.

Click the name of the analysis library to see additional information about the analysis.

The screenshot shows the SETS software interface. The top navigation bar includes Home, Analysis, View, Admin, History, Report, Log In, and Help. The main window title is 'Run details: MD_PARIS_20100204_F401_FRAC_1529_IQ'. Below the title, there is a summary of the run: Run name (MD_PARIS_20100204_F401_FRAC_1529_IQ), Owner (Administrator), Run method (SOLiD), Run status (complete), Run started (02/04/2010 01:18 PM), Run completed (02/04/2010 01:18 PM), Spots (2), and Transfer status (Paras per spot: 1). On the left, a data tree shows a folder for 'Sample: Sample1' containing a sub-folder for 'Analysis cycles' with a blue dot next to 'F3_PE_05_V1'. The main panel displays details for 'Sample: Sample1' and 'Analysis cycles' for 'F3_PE_05_V1', including the date created (02/04/2010 01:24 PM) and a table of analysis results.

Library	Method	Run Name	Run Date	Run Time	Run Status	Run Method	Run Owner	Run Started	Run Completed
Library	Method	Run Name	Run Date	Run Time	Run Status	Run Method	Run Owner	Run Started	Run Completed
Library	Method	Run Name	Run Date	Run Time	Run Status	Run Method	Run Owner	Run Started	Run Completed



Troubleshooting analysis failure

If there is a **red dot** before the analysis job name, the analysis job failed. To see details:

1. Click the name of the failed analysis job. The following example shows 2 failed analysis jobs: F3_PC_01_V1 and F3_PB_10_V1. The right portion of the figure shows the details of analysis F3_PC_01_V1.

The screenshot shows the software interface for sample MS0041. On the left, a tree view lists analysis jobs under 'auto-analysis'. Two jobs, 'F3_PC_01_V1' and 'F3_PB_10_V1', are marked with red dots, indicating they failed. On the right, the details for the failed job 'F3_PC_01_V1' are displayed. The 'Status' is 'failed', and the 'Description' is 'Colorcalling'. The 'Data Files' section shows 'Matching Result in Gff format'. The 'Analysis Reports' section includes 'Quality Values Report' and 'Auto-Correlation Report'. The 'Details for job: F3_PC_01_V1 (id=420)' section shows various parameters such as 'Settings: default primary(v.1)', 'Cycle: 1', 'Status: failed', 'Created: 02/11/2010 06:49 PM', 'Started: 02/11/2010 06:50 PM', and 'Completed: 02/11/2010 08:12 PM'.

2. Check Filter_FastaF3 and Filter_FastaR3 (if paired-end library) to see if the analysis job was successful. Check the details for the analysis job and each cycle to ensure that they are correct. See this example:

The screenshot displays the primary analysis interface. On the left, a tree view shows the analysis structure for 'Sample1_227_0912113'. The 'Filter_FastaF3' job is highlighted with a blue circle. The main panel shows the details for this job, including a description, date created, and last completed job. Below this, the 'Data Files' section lists various analysis reports. The 'Details for job Filter_FastaF3 (Id=2920)' section is expanded, showing a table of image directories and their corresponding file names.

Sample: Sample1_227_0912113

Show in tree
Spot(s): 1

Sample Reports
Satay Report
Color Balance Report

Analysis: auto-analysis

Show in tree
Cancel analysis

Description:
Date created: 01/22/2010 12:36 PM
Last completed job: Barcoding

Data Files

Analysis Reports
Quality Values Report
Auto-Correlation Report

Show in tree
Details for job Filter_FastaF3 (Id=2920) Logs

Settings:	default_primary(v.1)	bead.libraryType	fragment
Stage:	postPrimerSetPrimary	focal.map.dir	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
Status:	finished	focal.map.stg	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
Created:	01/28/2010 08:13 PM	focal.map.version	2
Started:	01/28/2010 08:50 PM	images.dir.1	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
Completed:	01/28/2010 09:33 PM	images.dir.10	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.11	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.12	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.13	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.14	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.15	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.16	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.17	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.18	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.19	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.2	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.20	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.21	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.22	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.23	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.24	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli
		images.dir.25	solid0049_20100121_FRAG_P1beadOligo_2/solid0049_20100121_FRAG_P1beadOli

The above example shows detailed information for one Filter_FastaF3 run. This run uses 5 primers (PA-PE) and for each primer, 5 cycles. This analysis used all 5 primers (PA-P5); for each primer, 5 cycles were included in the analysis job, indicating no missing images in the analysis.

Sometimes, an analysis job might be reported as successful (no red x or warning dot next to the analysis name), but some of the primer/cycle combinations were not included in the analysis. Even if the analysis job is successful, if it is incomplete, it should be treated as a failed analysis job.

3. Log in to the instrument and look in the following directory:

`{SampleResultsFolder}/primary.{17-digit timestamp}/reads`

There should be a file named **{run.name}_F3.stats** for F3 tag, or **{run.name}_R3.stats** for R3 tag. This file contains information about the color-space reads generation for each panel. It contains the number of reads and an indication of whether they have too many errors, are too short, or are duplicates.

The Totals (### p) line summarizes the number of panels found and the percentage of error, short, and duplicate reads.

After this line is a line beginning with Usable (### p) (##%). This line summarizes how many panels were usable and the percentage of erroneous, short, or duplicate reads among the usable panels.

If these two lines are missing, the primary analysis did not complete.

Procedure for re-analyzing the image (primary analysis)

If there was failure in the automatically-generated primary analysis pipeline, there are two ways to do primary re-analysis of the image. Before doing so, look at the primary log files to find out reason of failure. The log files are in the folder:

```
/data/result/{instrument.name}/{run.name}/{sample.name}/  
jobs/{stage.name}. [JobId]/analysis/
```

Also, ensure that all images are present at /data/images/{run.name}.

After finding out the reason, change primary analysis parameters.

To re-analyze through SETS:

1. Log in to SETS using an administrative account.
2. Modify the primary analysis settings. Either modify added parameters or add your new values in Additional parameter field.
3. Open the Run Details page for that run and do primary re-analysis.

To perform primary analysis manually:

1. Access the job folder at:

```
/data/result/{instrument.name}/{run.name}/{sample.n  
ame}/jobs/{stage.name}. [JobId]/
```

2. Modify the **parameters.ini** file with the desired value.
3. Enter the following command:

```
jprimaryanalysis.sh --dir=<parameter file directory>  
--ini=<parameter file name>
```

For example:

```
jprimaryanalysis.sh  
--  
dir=/data/results/DAEMONAC2/LAST_MP_1245_041008/Sam  
ple1/jobs/ postPrimerSetPrimary.3793 --  
ini=parameters.ini
```



Software License Agreement

APPLIED BIOSYSTEMS END USER SOFTWARE LICENSE AGREEMENT

FOR INSTRUMENT OPERATING AND ASSOCIATED
BUNDLED SOFTWARE
AND LIMITED PRODUCT WARRANTY

Applied Biosystems SOLiD™ 4 System SETS Software

NOTICE TO USER: PLEASE READ THIS DOCUMENT CAREFULLY. THIS IS THE CONTRACT BETWEEN YOU AND LIFE TECHNOLOGIES REGARDING THE OPERATING SOFTWARE FOR YOUR APPLIED BIOSYSTEMS WORKSTATION OR OTHER INSTRUMENT AND BUNDLED SOFTWARE INSTALLED WITH YOUR OPERATING SOFTWARE. THIS AGREEMENT CONTAINS WARRANTY AND LIABILITY DISCLAIMERS AND LIMITATIONS. YOUR INSTALLATION AND USE OF THE APPLIED BIOSYSTEMS SOFTWARE IS SUBJECT TO THE TERMS AND CONDITIONS CONTAINED IN THIS END USER SOFTWARE LICENSE AGREEMENT.

IF YOU DO NOT AGREE TO THE TERMS AND CONDITIONS OF THIS LICENSE, YOU SHOULD PROMPTLY RETURN THIS SOFTWARE, TOGETHER WITH ALL PACKAGING, TO APPLIED BIOSYSTEMS AND YOUR PURCHASE PRICE WILL BE REFUNDED.

This Applied Biosystems End User License Agreement accompanies an Applied Biosystems software product (“Software”) and related explanatory materials (“Documentation”). The term “Software” also includes any upgrades, modified versions, updates, additions and copies of the Software licensed to you by Applied Biosystems. The

term “Applied Biosystems,” as used in this License, means Applied Biosystems, LLC. The term “License” or “Agreement” means this End User Software License Agreement. The term “you” or “Licensee” means the purchaser of this license to use the Software.

THIRD PARTY PRODUCTS

This Software uses third-party software components from several sources. Portions of these software components are copyrighted and licensed by their respective owners. Various components require distribution of source code or if a URL is used to point the end-user to a source-code repository, and the source code is not available at such site, the distributor must, for a time determined by the license, offer to provide the source code. In such cases, please contact your Life Technologies representative. As well, various licenses require that the end-user receive a copy of the license. Such licenses may be found on the distribution media in a folder called “Licenses.” In order to use this Software, the end-user must abide by the terms and conditions of these third-party licenses. After installation, the licenses may also be found in a folder named “Licenses” located in the Software installation's root directory.

TITLE

Title, ownership rights and intellectual property rights in and to the Software and Documentation shall at all times remain with Applied Biosystems, LLC and its subsidiaries, and their suppliers. All rights not specifically granted by this License, including Federal and international copyrights, are reserved by Life Technologies or their respective owners.

COPYRIGHT

The Software, including its structure, organization, code, user interface and associated Documentation, is a proprietary product of Life Technologies or its suppliers, and is protected by international laws of copyright. The law provides for civil and criminal penalties for anyone in violation of the laws of copyright.



LICENSE

Use of the Software

1. Subject to the terms and conditions of this Agreement, Applied Biosystems, LLC grants the purchaser of this product a non-exclusive license only to install and use the Software to operate the single product in connection with which this License was purchased and to display, analyze and otherwise manipulate data generated by the use of such product. There is no limit to the number of computers on which you may install and use the Software to display, analyze and otherwise manipulate such data.
2. If the Software uses registration codes, access to the number of licensed copies of Software is controlled by a registration code. For example, if you have a registration code that enables you to use five copies of Software simultaneously, you cannot install the Software on more than five separate computers.
3. You may make one copy of the Software in machine-readable form solely for backup or archival purposes. You must reproduce on any such copy all copyright notices and any other proprietary legends found on the original. You may not make any other copies of the Software except as permitted under Section 1 above.

Restrictions

1. You agree that you will not copy, transfer, rent, modify, use or merge the Software, or the associated documentation, in whole or in part, except as expressly permitted in this Agreement.
2. You agree that you will not reverse assemble, decompile, or otherwise reverse engineer the Software.
3. You agree that you will not remove any proprietary, copyright, trade secret or warning legend from the Software or any Documentation.

4. You agree to fully comply with all export laws and restrictions and regulations of the United States or applicable foreign agencies or authorities. You agree that you will not export or reexport, directly or indirectly, the Software into any country prohibited by the United States Export Administration Act and the regulations thereunder or other applicable United States law.

5. You agree that you will not modify, sell, rent, transfer (except temporarily in the event of a computer malfunction), resell for profit, or distribute this license or the Software, or create derivative works based on the Software, or any part thereof or any interest therein. Notwithstanding the foregoing, if this Software is instrument operating software, you may transfer this Software to a purchaser of the specific instrument in or for which this Software is installed in connection with any sale of such instrument, provided that the transferee agrees to be bound by and to comply with the provisions of this Agreement.

Trial

If this license is granted on a trial basis, you are hereby notified that license management software may be included to automatically cause the Software to cease functioning at the end of the trial period.

Termination

You may terminate this Agreement by discontinuing use of the Software, removing all copies from your computers and storage media, and returning the Software and Documentation, and all copies thereof, to Life Technologies. Life Technologies may terminate this Agreement if you fail to comply with all of its terms, in which case you agree to discontinue using the Software, remove all copies from your computers and storage media, and return the Software and Documentation, and all copies thereof, to Life Technologies.

U.S. Government End Users

The Software is a “commercial item,” as that term is defined in 48 C.F.R. 2.101 (Oct. 1995), consisting of “commercial computer software” and “commercial computer software documentation,” as such terms are used in 48 C.F.R. 12.212 (Sept. 1995). Consistent with 48 C.F.R. 12.212 and 48 C.F.R. 227.7202-1 through 227.7202-4 (June 1995), all U.S. Government End Users acquire the Software with only those rights set forth herein.

European Community End Users

If this Software is used within a country of the European Community, nothing in this Agreement shall be construed as restricting any rights available under the European Community Software Directive, O.J. Eur. Comm. (No. L. 122) 42 (1991).

Regulated Uses

You acknowledges that the Software has not been cleared, approved, registered or otherwise qualified (collectively, “Approval”) by Applied Biosystems, LLC with any regulatory agency for use in diagnostic or therapeutic procedures, or for any other use requiring compliance with any federal or state law regulating diagnostic or therapeutic products, blood products, medical devices or any similar product (hereafter collectively referred to as “federal or state drug laws”). The Software may not be used for any purpose that would require any such Approval unless proper Approval is obtained. You agree that if you elect to use the Software for a purpose that would subject you or the Software to the jurisdiction of any federal or state drug laws, you will be solely responsible for obtaining any required Approvals and otherwise ensuring that your use of the Software complies with such laws.

LIMITED WARRANTY and LIMITATION OF REMEDIES

Limited Warranty. Applied Biosystems warrants that, during the same period as of the SOLiD™ 4 Analyzer for which this Software is an instrument operating software, the Software will function substantially in accordance with the functions and features described in the Documentation delivered with the Software when properly installed, and that for a period of ninety days from the beginning of the applicable warranty period (as described below) the tapes, CDs, diskettes or other media bearing the Software will be free of defects in materials and workmanship under normal use.

The above warranties do not apply to defects resulting from misuse, neglect, or accident, including without limitation: operation outside of the environmental or use specifications, or not in conformance with the instructions for any instrument system, software, or accessories; improper or inadequate maintenance by the user; installation of software or interfacing, or use in combination with software or products not supplied or authorized by Applied Biosystems; intrusive activity, including without limitation computer viruses, hackers or other unauthorized interactions with instrument or software that detrimentally affects normal operations; and modification or repair of the products not authorized by Applied Biosystems.

Warranty Period Commencement Date. The applicable warranty period for software begins on the earlier of the date of installation or three (3) months from the date of shipment for software installed by Applied Biosystems' personnel. For software installed by the purchaser or anyone other than Applied Biosystems, the warranty period begins on the date the software is delivered to you. The applicable warranty period for media begins on the date the media is delivered to the purchaser.

APPLIED BIOSYSTEMS MAKES NO OTHER WARRANTIES OF ANY KIND WHATSOEVER, EXPRESS OR IMPLIED, WITH RESPECT TO THE SOFTWARE OR DOCUMENTATION, INCLUDING BUT NOT LIMITED TO WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE OR MERCHANTABILITY OR THAT THE SOFTWARE OR DOCUMENTATION IS NON-INFRINGEMENT. ALL OTHER WARRANTIES ARE EXPRESSLY DISCLAIMED. WITHOUT LIMITING THE GENERALITY OF THE FOREGOING, APPLIED



BIOSYSTEMS MAKES NO WARRANTIES THAT THE SOFTWARE WILL MEET YOUR REQUIREMENTS, THAT OPERATION OF THE LICENSED SOFTWARE WILL BE UNINTERRUPTED OR ERROR FREE OR WILL CONFORM EXACTLY TO THE DOCUMENTATION, OR THAT APPLIED BIOSYSTEMS WILL CORRECT ALL PROGRAM ERRORS. APPLIED BIOSYSTEMS' SOLE LIABILITY AND RESPONSIBILITY FOR BREACH OF WARRANTY RELATING TO THE SOFTWARE OR DOCUMENTATION SHALL BE LIMITED, AT APPLIED BIOSYSTEMS' SOLE OPTION, TO (1) CORRECTION OF ANY ERROR IDENTIFIED TO APPLIED BIOSYSTEMS IN A WRITING FROM YOU IN A SUBSEQUENT RELEASE OF THE SOFTWARE, WHICH SHALL BE SUPPLIED TO YOU FREE OF CHARGE, (2) ACCEPTING A RETURN OF THE PRODUCT, AND REFUNDING THE PURCHASE PRICE UPON RETURN OF THE PRODUCT AND REMOVAL OF ALL COPIES OF THE SOFTWARE FROM YOUR COMPUTERS AND STORAGE DEVICES, (3) REPLACEMENT OF THE DEFECTIVE SOFTWARE WITH A FUNCTIONALLY EQUIVALENT PROGRAM AT NO CHARGE TO YOU, OR (4) PROVIDING A REASONABLE WORK AROUND WITHIN A REASONABLE TIME. APPLIED BIOSYSTEMS SOLE LIABILITY AND RESPONSIBILITY UNDER THIS AGREEMENT FOR BREACH OF WARRANTY RELATING TO MEDIA IS THE REPLACEMENT OF DEFECTIVE MEDIA RETURNED WITHIN 90 DAYS OF THE DELIVERY DATE. THESE ARE YOUR SOLE AND EXCLUSIVE REMEDIES FOR ANY BREACH OF WARRANTY. WARRANTY CLAIMS MUST BE MADE WITHIN THE APPLICABLE WARRANTY PERIOD.

LIMITATION OF LIABILITY

IN NO EVENT SHALL APPLIED BIOSYSTEMS OR ITS SUPPLIERS BE RESPONSIBLE OR LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY OR UNDER ANY STATUTE (INCLUDING WITHOUT LIMITATION ANY TRADE PRACTICE, UNFAIR COMPETITION OR OTHER STATUTE OF SIMILAR IMPORT) OR ON ANY OTHER BASIS FOR SPECIAL, INDIRECT, INCIDENTAL, MULTIPLE, PUNITIVE, OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE POSSESSION OR USE OF, OR THE INABILITY TO USE, THE SOFTWARE OR DOCUMENTATION, EVEN IF APPLIED BIOSYSTEMS IS ADVISED IN ADVANCE OF THE POSSIBILITY OF SUCH DAMAGES, INCLUDING WITHOUT LIMITATION DAMAGES ARISING FROM OR RELATED TO LOSS OF USE, LOSS OF DATA, DOWNTIME, OR

FOR LOSS OF REVENUE, PROFITS, GOODWILL OR BUSINESS OR OTHER FINANCIAL LOSS. IN ANY CASE, THE ENTIRE LIABILITY OF APPLIED BIOSYSTEMS' AND ITS SUPPLIERS UNDER THIS LICENSE, OR ARISING OUT OF THE USE OF THE SOFTWARE, SHALL NOT EXCEED IN THE AGGREGATE THE PURCHASE PRICE OF THE PRODUCT.

SOME STATES, COUNTRIES OR JURISDICTIONS LIMIT THE SCOPE OF OR PRECLUDE LIMITATIONS OR EXCLUSION OF REMEDIES OR DAMAGES, OR OF LIABILITY, SUCH AS LIABILITY FOR GROSS NEGLIGENCE OR WILLFUL MISCONDUCT, AS OR TO THE EXTENT SET FORTH ABOVE, OR DO NOT ALLOW IMPLIED WARRANTIES TO BE EXCLUDED. IN SUCH STATES, COUNTRIES OR JURISDICTIONS, THE LIMITATION OR EXCLUSION OF WARRANTIES, REMEDIES, DAMAGES OR LIABILITY SET FORTH ABOVE MAY NOT APPLY TO YOU. HOWEVER, ALTHOUGH THEY SHALL NOT APPLY TO THE EXTENT PROHIBITED BY LAW, THEY SHALL APPLY TO THE FULLEST EXTENT PERMITTED BY LAW. YOU MAY ALSO HAVE OTHER RIGHTS THAT VARY BY STATE, COUNTRY OR OTHER JURISDICTION.

GENERAL

This Agreement shall be governed by laws of the State of California, exclusive of its conflict of laws provisions. This Agreement shall not be governed by the United Nations Convention on Contracts for the International Sale of Goods. This Agreement contains the complete agreement between the parties with respect to the subject matter hereof, and supersedes all prior or contemporaneous agreements or understandings, whether oral or written. If any provision of this Agreement is held by a court of competent jurisdiction to be contrary to law, that provision will be enforced to the maximum extent permissible, and the remaining provisions of this Agreement will remain in full force and effect. The controlling language of this Agreement, and any proceedings relating to this Agreement, shall be English. You agree to bear any and all costs of translation, if necessary. The headings to the sections of this Agreement are used for convenience only and shall have no substantive meaning. All questions concerning this Agreement shall be directed to: Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404-1128, Attention: Legal Department.

Unpublished rights reserved under the copyright laws of the United States.

Applied Biosystems, LLC, 850 Lincoln Centre Drive, Foster City, CA 94404.



Glossary

Alignment Matrix	The alignment matrix is displayed as: AA is encoded as 0 CG is encoded as 3 AACG is encoded as 0 1 3
Color Codes	Four dye colors, encoded as follows: 6-FAM [™] = 0 CY3 [®] = 1 Texas Red [®] = 2 CY5 [®] = 3
Cycle Order	The order in which the data are generated.
Location	For processed files, describes the location of the prepended base on the base-space reference sequence.

Mate-Pair

Fragments from the genome paired during sequencing and separated by the same number of bases. For a known reference sequence of size 5 Mbp and a read length of 25 bp per tag, the system allows a total of six mismatches when placing the F3 and R3 tags. These mismatches can be distributed in any allocation across the F3 and R3 tags as long as the total number of mismatches does not exceed six. Any beads with F3 and R3 tags, which map under this mismatch allowance and which satisfy the mate-pair constraints, are considered to be a successfully placed mate-paired read. The mate-pair constraints are as follows:

The distance between the F3 and R3 tags must be within a prescribed range of distances determined by the insert size and library preparation.

The tags must map such that R3 comes before F3 reading $5' > 3'$.

The tags must map such that F3 and R3 map to the same strand on the reference sequence.

Multiplex sequencing

Sequencing of multiple barcoded fragment libraries on a single spot on a slide.

Panel

The region of the array that the camera can image at one time. Successive cycles result in four images for each panel (one per color). Beads are identified by position within a panel, that is, x,y coordinates are not unique values across panels. The combination of panel and x,y coordinates uniquely identifies a bead.

Paired-End	<p>Paired-end libraries are standard fragment libraries. Sequence reads are generated from both ends of the fragment (5'>3' and 3'>5'). These reads are paired during sequencing and separated by the same number of bases (typically 50-200 bases).</p> <p>The paired-end constraints are as follows:</p> <ul style="list-style-type: none">• The distance between the F3 and F5 tags must be within a prescribed range of distances determined by the insert size and library preparation.• The tags must map such that F3 comes before F5 reading 5' > 3'.• The tags must map such that F3 and F5 map to opposite strands on the reference sequence.
Quality Value	<p>Accuracy estimation of colorcalling using metrics similar to phred or KBTM Basecaller.</p>
Sequencing Run	<p>During a SOLiDTM 4 System sequencing run, two probe sets are used to maximize the fraction of mappable beads, read length, and sequencing throughput. (Mappable beads are beads, amplified with template, that map to the reference genome.) This protocol must be used for sequencing 50-bp reads of both mate-paired and fragment libraries. Unlike terminator-based sequencing, the SOLiDTM 4 System does not collect base-sequencing information. Instead, five rounds of primers (Primers A, B, C, D, and E) are used to sequence template by ligation of di-base labeled probes. For sequencing of fragment libraries, the set of primers used is specific to the P1 adaptor.</p>
SOLiDTM 4 Analyzer	<p>The sequencing instrument and its supporting computer cluster.</p>
SOLiDTM 4 System	<p>The SOLiDTM 4 Analyzer, its instrument control software (ICS), and ancillary equipment.</p>
Spot	<p>A series of panels, each panel with unique numbers for ordering across spots. When multiple samples are run on a single array, each array has its own spot. This is shown in the file name format S#. For example, S1 are files for Spot 1.</p>

Stringency

Stringency is one criterion used to filter out junk reads in ReadBuilder stage. For filtered reads, csfasta and QV.qual are stored in the reject folder instead. The valid range of stringency is from 0 to 80. Any stringency less than 0 is treated as 0, where all reads will pass the stringency threshold. If stringency is set larger than 80, maximal possible stringency of 80 is used to avoid over filtering good beads.

TAG-ID

A unique identifier for every tag, which consists of four components: *panel_xpixel_ypixel_tagtype*. Each tag has a TAG-ID. For example, *1_567_321_F3* describes a bead in panel 1 at coordinates 56, 321 (X,Y) with the R3 tag (second tag in a mate-pair).

Note: A Tag ID describes the bead and its data within files.

Values

A dot (.) is used in color space to show that there is no call for this position. Although low values may have been measured, during clustering it was determined that no call could be made.

A minus one (-1) in a set of calculated values (for example, N2S) indicates that there is no data for this point; that is, a dot (.) is in the color space for this position.

A zero (0) in a set of calculated values (for example, N2S) means zero (0). The calculated result after rounding is zero (0).

WFA

A Workflow Analysis (WFA) run analyzes a quadrant of a slide that undergoes a single ligation cycle. The quadrant contains beads deposited at a lower density than the density of beads deposited for a sequencing run. You can optimize sequencing results by performing workflow analysis (WFA) runs.

A WFA run determines the:

- **Optimal library concentration:** The library concentration for optimal preparation of templated beads using the library. You use this library concentration for any preparation of templated beads for that library when the scale of templated bead preparation is the same.
- **Bead enrichment efficiency:** The proportion of beads that have been successfully amplified using emulsion PCR (ePCR) as a fraction of the total number of beads prepared. You use this value to more accurately deposit successfully amplified beads for a sequencing run.

WFA runs require the same materials as those materials needed for sequencing runs. If you perform multiple WFA runs routinely, order additional SOLiD™ Instrument Buffer Kits.

Zero and One Indexing

For preprocessed data, the first color space call is position 1, which refers to the transition between the last base of the adapter and the first base of the read. For processed data, the positions are 0-based so that the first position (the prepended base) is 0 on both strands. The n^{th} position of the tag is numbered $(n - 1)$ in the forward direction and $-(n - 1)$ in the reverse direction.

Related documentation

The following related documents are shipped with the system:

Document	Part number	Description
<i>BioScope™ Software for Scientists Guide: Data Analysis Methods and Interpretation</i>	4448431	Provides a bioinformatics analysis framework for flexible application analysis (data-generated mapping, SNPs, and count reads) from sequencing runs.
<i>Applied Biosystems SOLiD™ 4 System ICS Software Help</i>	Not applicable	Describes the software and provides procedures for common tasks (see the Instrument Control Software).
<i>Applied Biosystems SOLiD™ 4 System Instrument Operation Guide</i>	4448379	Describes how to load and run the SOLiD™ 4 Analyzer for sequencing.

A portable document format (PDF) version of this guide is also available on the Applied Biosystems SOLiD™ 4 System SETS v4.0.1 Software Help menu.

Note: To open any user documentation included on the software CD, download Adobe Reader from www.adobe.com (Adobe ▶ Acrobat ▶ Reader).

Send us your comments

Life Technologies welcomes your comments and suggestions for improving its user documents. You can e-mail your comments to:

techpubs@appliedbiosystems.com

IMPORTANT! The e-mail address above is for submitting comments and suggestions relating *only* to documentation. To order documents, download PDF files, or for help with a technical question, see “[How to obtain support](#)” on page 8.

Part Number 4448411 Rev. B 04/2010



Applied Biosystems
850 Lincoln Centre Drive
Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.345.5224
www.appliedbiosystems.com

Technical Resources and Support
For the latest technical resources and
support information for all locations,
please refer to our Web site at
www.appliedbiosystems.com/support