

LifeScope™ Genomic Analysis Software v 2.5

Release Notes

LifeScope™ Genomic Analysis Software is the optimal application for analysis and management of 5500 Series Genetic Analyzer data. You can map reads to a reference sequence, then detect many forms of genomic variation using LifeScope™ Software’s state-of-the-art algorithms and workflows.

These release notes contain features of the software, changes in the current release, installation and configuration notes, resolved issues, known issues and limitations, tips for using the system, and workarounds. This release note is updated on an as-needed basis.

LifeScope™ Software v 2.5

November 2011

Features

Low frequency variant detection workflow

The low frequency variant detection workflow uses the diBayes algorithm (with highly optimized parameters) to reveal SNPs present in a targeted resequencing experiment. The detection range can range from 1%-5% based on experimental conditions, acceptable sensitivity and specificity, and sequencing coverage. This workflow is ideal for detecting low frequency variants in cancer-related research projects.

Processing time improvements

Release 2.5 has general speed improvements. In particular, the processing time required for small indel runs is improved, due to parallelization within the module. Processing time comparisons with the 2.0 release are shown in the following table:

Application type	Data description	LS 2.0 (hours)	LS 2.5 (hours)	Difference (hours)
Whole genome mate-pair resequencing	600M reads	27:30	21:54	~5.5
Targeted resequencing paired-end	142M reads, 50x25	09:54	09:27	~0.5
Whole transcriptome paired-end	650M, indexed	49:46	40:44	~9
Small RNA	125M reads, indexed	03:18	02:23	~1

Changes in the 2.5 release

Mark duplicates

For multiple BAM files of paired data, the default behavior of the mark duplicates module is changed in 2.5. The previous default was to scan all BAM files first, and in a second phase mark duplicates for all BAM files concurrently. With the new default, the mark duplicates phase considers one BAM file at a time. The new default typically requires more time and less memory. The mark duplicates parameter `process.one.bamfile.only` controls this behavior.

Documentation

The LifeScope™ Software user guides include instructions for installation, licensing, and administration. Visit this site to download the LifeScope™ Software user guides (registration and login are required):

http://solidsoftwaretools.com/gf/project/lscope_release

The following are the LifeScope™ Software user guides:

- *LifeScope™ Genomic Analysis Software User Guide* (Part No. 4471877)
Describes the bioinformatics analysis framework for flexible application analysis (data-generated mapping, SNPs, count reads, etc.) from sequencing data.
- *LifeScope™ Genomic Analysis Software Command Shell User Guide* (Part No. 4471875)
Describes how to run secondary and tertiary data analyses with the software's the command shell interface.

Note: LifeScope™ Software must **NOT** be installed as `root`.

Note: The “Set up a workstation with remote submission” procedure described in the user guides installation instructions is **not recommended** and is **not supported**.

Installation instructions are described in Part II, “Installation”, in each user guide.

Index IDs and index names, for barcoded runs

XSQ files created with SOLiD Instrument Control Software (ICS) versions earlier than 1.1.1 allow the index ID and index name to be different. Mistakenly using the index name in place of the index ID gives erroneous results. The LifeScope™ Software GUI correctly uses the index ID, but the command shell allows you to enter the index name instead of the index ID. If the index ID and the index name are different, your analysis fails or produces erroneous results.

This issue does not apply to GUI users, or to XSQ files created with ICS version 1.1.1 or later.

Command shell users can determine the index ID from the `lscope.sh` shell command `cat xsqfile` in the reads repository. In `/reads`, the `cat` command with an XSQ file argument displays the libraries and barcodes contained in the XSQ file. For example, use the following command output to determine the index ID

of your data, and use the index ID given in the [idx:] fields with `lscope.sh add xsq` commands and groupfiles.

```
server:/reads> cat Indexing3panels.xsq
[lib:Library1] [idx:1] [MatePair] [50,50] [mixed]
[lib:Library1] [idx:13] [MatePair] [50,50] [mixed]
[lib:Library1] [idx:5] [MatePair] [50,50] [mixed]
[lib:Library1] [idx:9] [MatePair] [50,50] [mixed]
[lib:Library2] [idx:10] [MatePair] [50,50] [mixed]
```

You can also use the `print-library.sh` command to find the Index ID:

```
print-library.sh Indexing3panels.xsq
Library Name      Index Name      Index ID
Unclassified     Unclassified    0
Library1         BC12            1
Library1         BC15            5
Library1         BC3             9
Library1         BC4             10
Library1         BC7             13
```

Resolved issues

The following issues are resolved in the 2.5 release:

- The GUI now verifies the `analysis.regions.file` parameter for targeted resequencing and low frequency variant workflow runs. If the `analysis.regions.file` parameter is not set, these workflows fail during tertiary processing.

To run a targeted resequencing or low frequency variant analysis, you must set the `analysis.regions.file` parameter to your regions of interest BED file:

- When using the GUI, set the `analysis.regions.file` parameter in the Set General Parameters view.
- When using the command shell, set the parameter in your `global.ini` files, with this command: `set analysis.regions.file /path/to/BED_file`

Known issues and limitations

This section describes known issues and limitations with the 2.5 release.

- Internet access is required to do the following with LifeScope™ Software:
 - During installation
 - To use the graphical user interface (GUI)
 - 03018 To open LifeScope™ Software results files in the Interactive Genomics Viewer (IGV)
- If you are using a proxy server, please check with your system administrator for the settings to enable JNLP downloads. This is required for LifeScope™ Software client start-up.

- If LifeScope™ Software client fails to open in your browser, follow these steps as a workaround:
 1. Right-click on the Analysis Portal or Admin Portal button, and select **Save Target As ...** or similar menu item
 2. Save the file `LifeScopeClient.jnlp` to your local machine.
 3. Run this command to start LifeScope™ Software:

```
javaws LifeScopeClient.jnlp
```
- LifeScope™ Software requires Perl v.5.8.5 (or later). However, the inversion module requires a Perl version up through 5.8.8. The most recent version of Perl has a backwards compatibility issue which causes the inversion module to fail.
- With LifeScope™ Software and IGV 2.0, the IGV browser does not add a track to an existing IGV window. Instead, another instance of IGV is opened. To avoid opening the second instance of the IGV browser, download the data locally, then go to IGV, and open the file from IGV.
- 02298, 03060: The results of the SNPs and small indels modules are not available in VCF format.
- 02684 LifeScope™ Software does not support mapping multiple organisms to separate references in one analysis run.
- 02688 The tertiary modules do not support merged input BAM files with data from multiple runs.
- 02701 While importing a CSFASTA or QUAL file, status is not properly displayed if the analysis fails before it is started.
- 02771 The import of an XSQ file to the LifeScope™ Software reads repository does not generate a log file. If an error occurs during import, the error is not captured.
- 02865 LifeScope™ Software reports significantly more duplicates than does Picard. There are a number of underlying reasons why our mark duplicates module is more aggressive in marking duplicates than Picard:
 - By design, LifeScope™ Software marks reads with the same read start position, even if they have soft or hard clipping at the start of the read, as duplicates. We consider this to be desirable behavior.
 - There are also differences in the handling of reads mapped to different chromosomes, and of reads that are not “proper pairs” (because of orientation, strand, or distance).
 - In the case of reads from multiple BAM files and with non-unique bead IDs across multiple files, LifeScope™ Software marks reads with the same bead ID as duplicates.

LifeScope™ Software 2.5 by default marks duplicates only within a single BAM file, with an option to mark duplicates across multiple BAM files. The multiple BAM option is appropriate for multiple barcodes from the same sample run on the same lane. For such runs, our mark duplicates module behaves similarly to Picard and marks around the same number of duplicates as Picard. The duplicates detected are not exactly the same, because the algorithms differ slightly.

- 02902 Modifying files under the `lifetech` directory of the reference repository is not supported. Header and other information in the reference file is case sensitive. For example, changing the case of a contig name in a reference file causes program failures. Adding files under the `lifetech` directory also is not supported. Changing the information in any respect is not supported.
- 02925 LifeScope™ Software does not support run of mixed library types, such as mixing mate-pair, paired-end, and fragment libraries in one run.
- 02928 Secondary processing creates one BAM file per lane, and typically generates multiple BAM files, which are input to tertiary processing. LifeScope™ Software does not support generating a single BAM file per sample, if the sample is analyzed on multiple lanes.
- 02948 For a job with a large (more than a few thousand) number of contigs, change the system property `max.head.node.java.heap.space` from the default of 64 to 128 or 256. Edit the file `<installdir>/etc/analysis/system.properties` to modify the `max.head.node.java.heap.space` property.
For example, a run with 40,000 contigs required 128 for BAMStats and 256 for SNPs.

Because changes to system properties affect all runs of all users, you could consider changing the property back to its original value after the large contig run completes. Setting `max.head.node.java.heap.space` to 128 or 256 could present a problem if your head node is extremely loaded.

- 02973 In LifeScope™ Software v2.1 and v2.5, BAMStats output incorrectly reports the `IsECC` field as “N” for ECC data. The program checks for data from the ECC 6th primer, and this check fails even on data from ECC runs. The “N” can be ignored.
- 03048 When the cluster’s compute nodes are not visible from the head node, the error message displayed does not well describe the actual problem. The error message is an `RRSValidatorException`: “All Records in the `rrs` file are invalid.” The full stack trace is:

```

Error:
05 Oct 2011 00:21:12,057 INFO [main] PluginSubJobLauncherImp:43 -
05 Oct 2011 00:21:12,059 FATAL [main] PluginRunner:87 -
com.apldbio.aga.common.model.plugins.ModuleException:
com.lifetechnologies.bioscope.readset.RRSValidatorException: All Records in the rrs
file are invalid.
at com.lifetechnologies.secondary.Splitter.getValidTable(Splitter.java:163)
com.lifetechnologies.secondary.FragmentMapper.setGroupMap(FragmentMapper.java:214)
com.lifetechnologies.secondary.FragmentMapper.validateCommonParams(FragmentMapper.java
:148)
com.lifetechnologies.secondary.PairMapper.validateParams(PairMapper.java:119)
com.apldbio.aga.analysis.jobms.PluginRunner.preparePipeline(PluginRunner.java:402)
com.apldbio.aga.analysis.jobms.PluginRunner.doMain(PluginRunner.java:450)
com.apldbio.aga.analysis.jobms.PluginRunner.safeMain(PluginRunner.java:721)
com.apldbio.aga.analysis.jobms.PluginRunner.main(PluginRunner.java:699)

```

- 02893 No error message shown if you import a BAM or XSQ file which is actually a broken symbolic link in the file system. The Import Tab shows the import status as failed, but no error is given during the import.

- 03054 The parameter `refcor.base.filter.qv` is provided to filter out reads with low quality values, and bases with a quality value below the value of this parameter are replaced with N in the BAM output. These bases are usually not mismatches, but the SAMTools `fillmd` utility counts N positions as mismatches, and the presence of the Ns causes a discrepancy in the number of mismatches for SAMTools. Then SAMTools, when using `fillmd`, reports an error such as the following: `fillmd [bam_fillmd1] different NM for read '813_939_878': 0 -> 3`
- 03055 When running a targeted resequencing workflow in the command shell, if you turn off the SAET error correction module, the following parameter changes must be made:
 - Set `saet.run=0` in the `secondary/saet.ini` file.
 - Set the `analysis.input.readset.file` parameter to `secondary/<reference>/readsets.rrs` in your mapping INI file (`secondary/fragment.mapping.ini` or `secondary/pair.mapping.ini`).

Make sure you **do not** set the `analysis.input.readset.file` parameter in the `global.ini` file.

If the `analysis.input.readset.file` parameter is set to the `.rrs` file in both `global.ini` file and `fragment.mapping.ini` file, the framework writes the enrichment results to the incorrect directory. This error happens because `analysis.sample.name` is not appended to the `enrichment.output.dir` parameter.

- 03058 LifeScope™ Software supports the following characters in contig names:

Alphanumeric	A-Z, a-z, 0-9
Underscore	_
Period	.
Pipe symbol	

- 03075 This issue involves an intermittent job failure that occurs during analysis runs. The failures are either machine-specific or configuration-specific: on most installations these failures do not occur. The following is an example of the error message:
`/share/lifescopes/lifescopes/bin/java_app.sh: line 128: 82957 Killed`

The error message is found in the module logs in the `output/log/TORQUE` directory. Because the failures are not associated with a specific module, the error message typically does not appear in the same log file. The example above is from a small RNA TORQUE log file: `outputs/log/TORQUE/smallRNAmapping/secondary-hg18-smallrnmapping.3.run.20111110232334079.log`.

We are currently investigating this issue. This failure happens in only a small fraction of total runs. Most of the failures, about 80%, occur during secondary analysis, after the `refcor` module. About 10% of the failures are seen in the CNV module, and about 10% occur in other modules.

The workaround is to restart the job. Consider using the `resume` command to avoid repeating completed modules.

LifeScope™ Software v 2.1

August 2011

Features

The main LifeScope™ Software v 2.1 updates are listed below, and are described in more detail in the [Changes in the 2.1 release](#) section.

Implemented ECC-corrected color space

Exact Call Chemistry (ECC) is an optional primer round on v1.1 of the 5500 Genetic Analyzer. ECC-corrected color provides an increase in call accuracy and quality, in mapping throughput, and in variant detection.

Improved large indel detection

The insert size distribution is computed within the large indel module. This modification results in a greater number of large indels being detected.

Modified SNP analysis for targeted resequencing workflow

In LifeScope™ Software v2.1, the methodology for counting SNPs associated with targeted regions is modified to follow the same approach as seen in BioScope™ Software 1.3.x releases. This modification results in more SNPs being reported.

Changes in the 2.1 release

ECC-corrected color space

With ECC-corrected color calls, after mapping we observe increased accuracy and increased quality of the calls, compared to non-ECC color calls. These increases enable analysis with an accuracy of 99.99%. The following table shows accuracy for base space calls translated from ECC-corrected color calls (data on file with Life Technologies):

Type of call	QV threshold	Percentage of QVs >= threshold	Error rate, accuracy
Synthetic beads	25	97%	1 in 10 ⁻⁴ , 99.99%
E. coli data	40	90%	1 in 10 ⁻⁴ , 99.99%

ECC-corrected color also enables the following increases in throughput and in the number of variant calls:

Type	Improvement
Mapping throughput	2%
SNPs detection	2-4%
Small indel detection	13%

To benefit from ECC-corrected color space, use the following:

- Sequencing data produced by the 5500 Series Genetic Analyzer, *with* the optional ECC primer run and ICS (Instrument Control Software for 5500 systems) v1.1
- The new analysis space preference parameter (see [Analysis space preference](#))
- SNPs color QV parameter settings optimized for ECC (see [SNPs color QV filter settings](#))

Note: ECC-corrected color is not available for paired-end libraries.

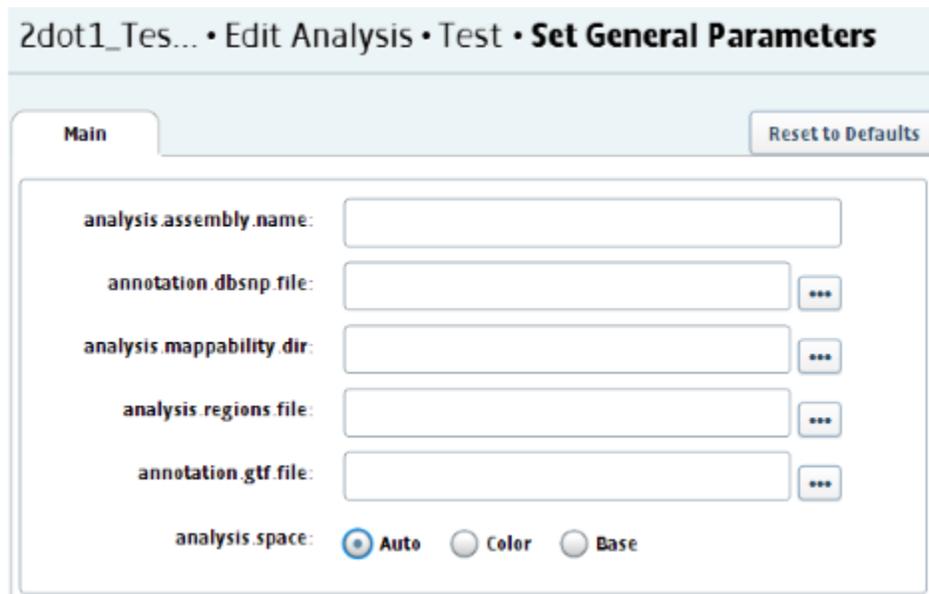
Analysis space preference

LifeScope™ Software v2.1 introduces a new global workflow parameter `analysis.space`, which allows you to enter your preference for base space or color space analysis. Three tertiary modules – SNPs, small indel, and splice finder – use your preference, in conjunction with information about whether color space is present in the BAM file and whether mapping is done in base space or color space, to determine their analysis space.

The recommended usage scenarios for the `analysis.space` parameter are:

- For color space analysis, map in color space and set `analysis.space` to `color` (or use the default value of `auto`). The following mapping parameter settings are required:
 - `mapping.in.base=0` (default)
 - `bamgen.refcor.addcs=1` (default)
- For base space analysis, map in base space and set `analysis.space` to `base` (or use the default value of `auto`). The following mapping parameter setting is required:
 - `mapping.in.base=1`

The following figure shows the Set General Parameters screen with the new `analysis.space` parameter and its default setting.



For the command shell, an example command is:

```
set analysis.space color tertiary/global.ini
```

If you create your own INI files, set the `analysis.space` parameter in your `tertiary/global.ini` file. An example setting is:
`set analysis.space=base`

The possible conditions determining the analysis space are given in the following table. **The analysis space used ...** column refers to the SNPs, small indel, and splice finder tertiary modules. (This table is provided for reference only. Recommended scenarios are given above.)

With these conditions ...			The analysis space used ...
Mapping space	Color present in BAM	analysis.space value	
Base	No	Base	Base
Base	No	Color	Base
Base	No	Auto	Base
Base	Yes	Base	Base
Base	Yes	Color	Color
Base	Yes	Auto	Base
Color	No	Base	Base
Color	No	Color	Base
Color	No	Auto	Base
Color	Yes	Base	Base
Color	Yes	Color	Color
Color	Yes	Auto	Color

SNPs color QV filter settings

In order to take advantage of ECC improvements, specific SNPs color QV parameter settings are recommended in the fragment and mate-pair workflows, for ECC data that is analyzed in color space. The setting changes are recommended for these workflows:

- Fragment targeted resequencing
- Fragment whole genome
- Mate-pair whole genome

For non-ECC runs, including paired-end runs, the 2.0 defaults are recommended. The paired-end workflows retain the 2.0 defaults. The following list summarizes the recommended values and default values:

- New 2.1 default values:
`dibayes.het.min.nonref.color.qv=14`
`dibayes.hom.min.nonref.color.qv=8`
- Recommended values for 2.1 ECC data:
`dibayes.het.min.nonref.color.qv=20`
`dibayes.hom.min.nonref.color.qv=8`
- Recommended values for non-ECC data (also the 2.0 default values):
`dibayes.het.min.nonref.color.qv=7`
`dibayes.hom.min.nonref.color.qv=7`

The 2.1 defaults are compromise settings to handle both ECC and non-ECC data. If you know whether your run contains ECC data, use the specific ECC or non-ECC recommended values given above.

Small indel alignment compatibility filter allowed values

The `small.indel.alignment.compatibility.filter` parameter is changed in 2.1. This parameter toggles an optional check for color space compatibility around the gap, and now has two allowed values:

- 0: Disable the check.
- 1: Enable the check (default).

Installation

For new users

For new users (who do not have a 2.0 installation), the software acquisition and installation process involves these steps:

1. Purchase the software.
2. Visit the LifeScope™ Software project page:
<http://solidsoftwaretools.com/gf/project/lifescopy>
3. On the project page, you must:
 - a. Read and accept the LifeScope™ Software End User License Agreement (EULA).
 - b. Create an email with the following information:
 - Your acceptance of the LifeScope™ Software End User License Agreement
 - Your LifeScope™ Software license key
 - Your solidsoftwaretools account name
 - c. Send the email to:
lifescopy@lifetech.com
Instructions to download LifeScope™ Software are emailed to you within one business day.
4. The LifeScope™ Data Drive is shipped to you. This drive contains the initial reference repository and other files used with the hg18 and hg19 genomes. These files are required for installation (as described in the installation instructions).
Note: When you purchase a LifeScope™ Workstation or Cluster, these files are pre-installed, and you should retain the Data Drive as a backup copy.
5. Go to the download page, as directed by the emailed instructions.
6. Download the following:
 - a. The LifeScope™ Software `.tar.gz` file (for example, `LifeScope-2.1r8380-2011072833223344.tar.gz`)
 - b. The LifeScope™ Software installation script (`install.sh`)
 - c. The LifeScope™ Software performance verification scripts
 - d. (*Optional*) LifeScope™ Software example workflows
 - e. The LifeScope™ Software Advanced user guide, which contains installation instructions
7. Follow the instructions in the user guide to install LifeScope™ Software. Please also read the [Installation and Configuration](#) section for LifeScope™ Software 2.0, in these release notes.

8. Verify the installation (as described in the installation instructions).

Installing 2.1 on the same machine as your 2.0 installation

For existing LifeScope™ Software 2.0 users, 2.1 is installed as a separate tree – it does not upgrade or modify your 2.0 installation tree. The 2.1 installer reads your preferences from your 2.0 installation, copies your 2.0 licenses and user database to the 2.1 installation, and creates links to your 2.0 repositories. The 2.1 installation uses the same reference, projects, reads, and bams repositories as your 2.0 installation. (You do not use the 2.0 Data Drive with the 2.1 upgrade.)

The upgrade process requires you to manually set users' PATH variables according to the new v2.1 location. The installer provides instructions on completion of the upgrade.

After installing 2.1, you are still able to use your 2.0 installation (provided you manage the PATH variable entries). However, the 2.0 and 2.1 installations *cannot* run simultaneously.

Note: The restriction that 2.0 and 2.1 installations cannot run simultaneously is based on the license server. Two installations that use the same license server cannot run simultaneously, even if the installations are on different machines.

During this process you supply two directories that are not in the v2.0 installation directories:

- A location to download the v2.1 software and the installation script
- A location to install the v2.1 upgrade

Follow these steps to install 2.1:

1. Choose a download directory that is not your LifeScope™ Software installation directory.
2. Go to the LifeScope™ Software download page.
If you do not know the download page, contact your local Field Bioinformatics Specialist or send an email asking for the location to:
lifescopel@lifetech.com
Include your LifeScope™ Software license key and your solidsoftwaretools account name in the email.
3. Download (into the same directory):
 - The LifeScope™ Software `.tar.gz` file (for example, `LifeScope-2.1r8380-2011072833223344.tar.gz`)
 - The LifeScope™ Software installation script (`install.sh`)
4. Run this UNIX® command:

```
chmod +x install.sh
```


to make the downloaded `install.sh` file executable.
5. Run the installer:

```
./install.sh
```
6. Installation options include:
 - 1) Install the LifeScope™ Genomic Analysis Software.
 - 2) Upgrade your current installation of LifeScope™ Genomic Analysis Software.
 - 3) Re-configure the LifeScope™ Software.
 - 4) Request a LifeScope™ Software license.

- 5) Register a LifeScope™ Software license.
- 6) Exit this application.

To upgrade your existing LifeScope™ Software v2.0, type 2.

7. At the following installer prompt, enter your current v2.0 installation directory:
Please enter the existing INSTALL location for LifeScope:
8. At the following installer prompt, enter the directory where v2.1 is to be installed:
Please enter the existing UPGRADE location for LifeScope:

Either specify an empty directory or let the installer create the upgrade directory for you. V2.1 is installed, and the installer applies your current v2.0 configuration and preferences to v2.1.

The installer displays configuration information retrieved from the v2.0 installation and displays progress information as it proceeds with the upgrade. At the completion of the upgrade, the installer gives instructions on updating the PATH variable for LifeScope™ Software users.

9. Follow the installer instructions to update the PATH environment variable. Remove the v2.0 entry in the PATH.

Installing 2.x on the same machine as a working BioScope™ installation

If you currently run BioScope™ Software on your cluster, and need to support both BioScope™ and LifeScope™ users on the same cluster, follow these steps:

1. Remove the file `bioscope_profile.sh` from the `/etc/profile.d` directory (the script cannot be in any `/etc` init directory). Having this script in a system init directory prevents LifeScope™ Software from running properly. (Save the script to your own private location, in order to configure the environment for BioScope™ Software users, as described in step 3.)
2. Users who run LifeScope™ Software must not have the environment variable `BIOSCOPE_ROOT` set and must not have the BioScope™ Software `/bin` directory in their `PATH`.
3. Users who run BioScope™ Software should configure their environment by copying the content of the file `/etc/profile.d/bioscope_profile.sh` into their own shell setup file (such as `.bash_profile`).

Resolved issues

The following issues are resolved in the 2.1 release:

02749 Large indel detection

In release 2.0, to identify candidate read pairs that support an indel, the large indel module uses the insert size distribution (mean and standard deviation) specified in the BAM header tags `IA` and `IS`. In 2.1, the large indel module no longer relies on the BAM header insert size tags. Instead the insert size distribution is computed within the module, using an approach that removes outliers, which are primarily due to artifacts of mapping, from the insert size distribution.

The Inter Quartile Range IQR ($Q3 - Q1$) is first computed from the entire insert size distribution (sampling 10 million BAM records). A new distribution of insert sizes is developed using only the values from $\text{Mean} - (3 * \text{IQR})$ to $\text{Mean} + (3 * \text{IQR})$, and the standard deviation is computed from the resulting distribution.

02773 Too few SNPs called in targeted resequencing workflows

In the targeted resequencing workflows, the `dibayes.ini` file is changed to set the `analysis.regions.file` parameter default to an empty string (empty single quotes).

02794 The SAET suppvotes parameter is corrected in targeted resequencing workflows

In the targeted resequencing workflows, the `saet.ini` file is corrected to set the parameter `saet.suppvotes` default to 3. 3 is the recommended value for targeted resequencing.

02811 Results from the whole transcriptome counts module differ between 2.0 and 2.1

Results from the whole transcriptome counts module are incorrect in the 2.0 release. This issue is addressed in the 2.1 release.

02799 Parameter analysis.mirbase.mature.file incorrectly set to readme file

In release 2.0, the shell command `set workflow hg18` (or `hg19`) incorrectly sets the parameter `analysis.mirbase.mature.file` to `README.txt`, and you must manually set the parameter correctly *after* the `set workflow` command. This issue is addressed in the 2.1 release.

Known issues and limitations

This section describes known issues and limitations with the 2.1 release. Please also read the [Known issues and limitations](#) section for LifeScope™ Software 2.0, in these release notes. Except for those issues noted as resolved in the 2.1 [Resolved issues](#) section, 2.0 [Known issues and limitations](#) still apply.

- When entering content into the instrument fields of the ICS user interface, the acceptable characters are letters, numbers, white spaces, underscores, and dashes.
- When running a targeted resequencing workflow, always specify a valid regions of interest file. The targeted resequencing workflows require a regions of interest file, but the user interfaces do not enforce the requirement. The interfaces also do not verify whether the regions of interest file specified is an empty file.
- Whenever possible, run mapping with 24 GB. When a mapping analysis is run with less than 19 GB of memory, the mapping module splits the reference. During gapped alignment of the second half of the genome, some alignments are reported at incorrect positions. The incorrect gapped alignments cause incorrect results (missing indels) during small indel analysis.
- The SNPs module does not accept with a mixture of ECC and non-ECC input data. If the BX and BY tags across *all* input BAMs do not have the same value, the SNPs module fails with the following error message:

The BAM files are not all consistent with each other. Only BAM files all originating from base-space instrument runs or all originating from color-space instrument runs can be analyzed together from SNP detection.

- LifeScope™ Software places certain character restrictions on file names and other names (in addition to the Linux filename conventions), as shown in the following table:

Use	Character	
		Hyphen ('-')

Reference file names	Not allowed	Not allowed
Assembly names	Not allowed	Not allowed
Sample group names	Not allowed	Not allowed
Names of INI and PLN files	Not allowed	Not allowed
Names of input CSFASTA and QUAL files	Not allowed	—
Names of input XSQ files	Not allowed	—
Project names	—	Not allowed
Analysis names	—	Not allowed

For files listed in the table, the restrictions apply to the base file name only, not to paths up to the filename.

In addition, project names and analysis names within LifeScope™ Software must contain only alphanumeric and underscore characters. Project names must not begin with an underscore.

When converting input CSFASTA files to XSQ files, the GUI replaces hyphens in the original file names with underscores in the XSQ file names.

- Licenses are released during normal completion of the software. If the software exits abnormally, it may not be able to release the license. The license is then only released when the license server is rebooted.
- Releases before 2.0 allowed users to supply the pairing insert sizes. Beginning in the 2.0 release, LifeScope™ Software always calculates these sizes internally. The insert size parameters are no longer available.
- The *LifeScope™ Genomic Analysis Software Command Shell User Guide* (Part no. 4465697) internal mapping parameter table incorrectly lists the following non-existent parameters:
`first.map.gap.min.non.matched`
`second.map.gap.min.non.matched.length`
 The correct parameter name is `mapping.gap.min.non.matched.length`.
- The *LifeScope™ Genomic Analysis Software Command Shell User Guide* (Part no. 4465697) pairing parameters table is missing the following two parameters:

<code>minimum.points.for.size.stat</code>	1000000	Minimum points in insert size distribution points to do compute statistics.
<code>memory.per.bam.merge</code>	3	Gigabytes of memory for one BAM merge thread.

- The *LifeScope™ Genomic Analysis Software Command Shell User Guide* (Part no. 4465697) XSQ Tools appendix does not include the `--laneNumber <arg>` option in the required arguments table or in the `sh convertToXSQ.sh` examples. You must use this option to specify the lane number in each use of the `convertToXSQ.sh` script.

LifeScope™ Software v 2.0

May 2011

Features

Supported analyses

LifeScope™ Software offers the following types of data analyses:

- Whole genome sequencing
- Targeted resequencing
- Whole exome sequencing
- Whole transcriptome RNA sequencing
- Small RNA sequencing
- MethylMiner™ mapping
- ChIP-Seq mapping
- Detection of wide range of genomic and RNA variation, including:
 - SNP detection
 - Large and small indel detection
 - Copy number variation detection
 - Inversion detection
 - Fusion transcript detection
 - Exon counting
 - Splice finding

Main features

- Seamless integration with the 5500 Series Genetic Analyzers
- Performance-tuned algorithms for the 5500 Series Genetic Analyzers and ECC Module
- Significant performance improvement over previous releases
- Push-button workflows, intuitive user interface, and secure project management
- Optimized mapping and smaller file formats
- Annotated variant reports, numerous charts, and select visualization tools for simple data interpretation
- Graphically-driven configuration of multistep analysis workflows
- Ability to save and reuse workflows
- Ability to resume a workflow without repeating completed analyses
- Secure project-based data management specific to your data analysis
- Projects that can be stored and reanalyzed

Tips

- Analyses that you run through the command shell interface are integrated with the GUI. Results from command shell analyses can be visualized in the GUI.
- Do not use 1.3 INI and PLN files for 2.0 analyses. The PLN file syntax has been enhanced, and some analysis parameters have changed.
- Global mapping is available in the LifeScope™ Software command shell (not in the GUI).
- For input that contains both base-space and color-space data, in order to perform secondary analysis in base space mode, set the `mapping.in.base` parameter to `true`. The default setting for

the `mapping.in.base` parameter is `false`, which causes analysis to be performed in color-space mode.

- This tip describes using the shell `resume` command to finish an analysis that fails on an intermediate step. After you log into the command shell, open your analysis and give the `resume` command:

1. Open the projects repository: `cd /projects`
2. Open your project: `cd myproj`
3. Open your analysis: `cd run1`
4. Give the `resume` command: `resume`

The analysis is restarted in the module that failed. The `resume` command does not rerun modules that completed successfully before the failure.

- You can add multiple input XSQ files to the reads repository by copying (or linking) the files directly into the repository directory on the local file system. After adding files by this method, run the `rebuild` command in the command shell. The `rebuild` command is required for the newly-added files to appear in the GUI Find Data view and in the command shell `ls` output. (The local file system directory for the reads repository is set during installation.)
- The new `cat` shell command provides summary information of XSQ and BAM file content. The GUI also provides this information in the Find Data view. The following is example output of the `cat` command:


```
[lib:Library1] [idx:10] [MatePair] [50,50] [mixed]
[lib:Library1] [idx:13] [MatePair] [50,50] [mixed]
```
- When viewing results for a completed analysis, use the **Ctrl** key to download multiple files (both log files and the Additional Files table).
- Refer to the LifeScope™ Software user guides for a description of the new project-based data management as well as the graphical and command shell interfaces:
 - *LifeScope™ Genomic Analysis Software User Guide* (Part no. 4465696)
 - *LifeScope™ Genomic Analysis Software Command Shell User Guide* (Part no. 4465697)

Installation and configuration

This section lists notes about LifeScope™ Software 2.0 requirements, installation, configuration, and administration, especially where these are different from previous releases.

Hardware requirements

- **RAM** – The recommendation for 2.0 is 24 GB minimum per compute node. With 8 cores per compute node, the recommendation is an average of 3 GB per core at the minimum.
- **Cluster** – LifeScope™ Software 2.0 runs on a BioScope™ Software 1.3 cluster, but does not achieve optimal execution performance.

Software requirements

- The following software products are required in previous releases but are not required in 2.0:
 - Java v1.6
 - Java Message Service (ActiveMQ) v5.3+
 - Rsync v.3.0+
 - Tomcat v
 - Installation of compatible versions of these software packages on all compute nodes, prior to the LifeScope™ Software installation:
 - Java

- Perl
- Python
- The following have changed requirements, versions, or options in 2.0:

Software	1.3 requirement	2.0 requirement
OS	Redhat 4.2 or CentOS 4.2 (or later)	Redhat 4.7 or CentOS 4.7 (or later)
Resource managers	<ul style="list-style-type: none"> • TORQUE v2.3+ or • Sun Grid Engine (SGE) v6.2+ 	<ul style="list-style-type: none"> • TORQUE v2.3+ • Sun Grid Engine (SGE) v6.2+ or • Platform Load Sharing Facility (LSF) 7 Update 6
TZ environment variable	TZ set to a valid time zone and exported	—
Queues	The <code>bs_primary</code> and <code>bs_secondary</code> queues must be available.	—

(Cells with a dash ‘—’ mean these is no requirement for that software.)

GUI requirements

The following lists the minimum requirements for client access (browser access to the graphical user interface):

- Windows® XP SP3, Linux® CentOS 4, or Mac OS® X v10.5
- 2 GB RAM
- 1024 × 768 display monitor
- Internet Explorer® 6 or Firefox® 3.6

Licensing

- The licensing server is new to LifeScope™ Software 2.0.
- The 2.0 release offers two licensing options: named users and concurrent users.
- If the following DNS requirements are not met, the licensing server software denies users access to LifeScope™ Software:
 - The DNS fully-qualified domain name and the local Linux hostname must match.
 - The forward and reverse lookup on the DNS against the hostname and its public-facing IP address must be consistent.

Pre-Install setup

- **Data drive** – The data drive distributed to LifeScope™ Software customers contains files required for the correct installation and operation of the software:
 - Files for the initial reference repository, with support for the hg18 and hg19 assemblies.
 - Data files for installation verification.
- **Examples** – The examples are a separate download (not included in the installer), and are used for installation verification. New users are recommended to use the standard workflows.

Installation options

- The previous release offers these installation options: a full install, a command-line install, and a command-line plus GUI install. In 2.0 these options are combined into one install type.
- The 2.0 installer offers different options based on hardware configuration:
 - **Standalone Workstation** – LifeScope™ Software is installed completely on the local workstation. There is no access to a cluster. Resource management software is required on the workstation.
 - **Cluster** – LifeScope™ Software is installed on a cluster with a headnode, a set of compute nodes, shared storage, node local storage, and resource management software.
 - **Workstation with Remote Submission** – The local workstation, instead of the cluster headnode, is used for running the LifeScope™ Software server. Additional configuration is required to support submitting analyses to the remote cluster.

User authentication

User authentication realm options are new in 2.0. The following authentication realms are supported:

- **LifeScope** – Users are created in the LifeScope™ Software user repository. User Account Management (including account creation) is performed with the LifeScope™ Software Admin Portal (included in this installation). There is no relationship between the LifeScope™ Software users accounts and Linux® host system users or LDAP users.
- **LDAP** – User accounts on LDAP-compliant authentication server (OpenLDAP, Active Directory, etc.) are eligible to be LifeScope™ Software users. User Account Management is performed with the LifeScope™ Software Admin Portal (included in this installation).
- **Host** – Local (on the installation machine or cluster) Linux users are all LifeScope™ Software users.

Port numbers

The following table lists the port numbers used by 2.0 and the previous release.

Required by...	1.3 default port	2.0 default port
Tomcat	8080	n/a
LifeScope™ Software server	—	9998
LifeScope™ Software licensing server	—	27000

(Cells with a dash ‘—’ mean these is no requirement for that software.)

URLs

The following are the default URLs for 2.0 client access and cloud portal:

- Client access – <http://IPaddress:9998/LifeScope.html>
- Cloud portal – <http://www.LifeScopeCloud.com>

User administration

The GUI Admin Portal, which supports user administration (among other functionality), is new in 2.0. User roles (admin and user) are also new in 2.0.

Known issues and limitations

This section describes known issues and limitations with the current release. Known issues from previous releases have been resolved in 2.0, with the exception of issues noted here.

Installation

- Depending on your selection of user authentication realm during installation, you must create the default user account `lifescopes` as a user in the realm:
 - **Host realm** – Create the user `lifescopes` as a system user on the machine where LifeScope™ Software is installed.
 - **Lightweight Directory Access Protocol (LDAP) realm** – Create the user `lifescopes` as an LDAP user.

Project and analysis names

- When creating names for your projects and analyses, you must create names without spaces or special symbols (except underscore).

Reference files

- Make sure that there are only filter reference files in the `referenceData/*/assembly_name` folders. Genome reference files *must* be placed in the designated `reference` subfolders. Placing a reference in the incorrect location causes invalid analysis results. For a description of the reference repository and its structure, refer to the *LifeScope™ Genomic Analysis Software Command Shell User Guide* (Part no. 4469657).

Using `referenceData/lifetech/hg18` as an example, here are the correct locations for the filter and genome reference files:

- Filter: `referenceData/lifetech/hg18`
- Genome: `referenceData/lifetech/hg18/reference`
- When multiple files exist in the designated reference repository directory, the reference for the current analysis is selected according to the following sort order of the filenames in that directory:
 - a. Numeric
 - b. Capital letters
 - c. Lower-case letters

Input handling

- The whole transcriptome analysis (WTA) module currently does not validate the input GTF genome annotation file. To get accurate analysis results, it is imperative to ensure that the GTF file matches the genome reference. Otherwise, the module may produce erroneous output without any warning messages.
- If an empty GTF file or a GTF file with no exons is passed to the WTA module, then the analysis fails with java runtime exception:

```
28 Apr 2011 04:37:29,389 - LifeScope version: LifeScope-v2.0-r0_87695_2011042716
5203
28 Apr 2011 04:37:29,750 - process-run runCommandHelper(qsub)
28 Apr 2011 04:37:29,752 - process-run success.
28 Apr 2011 04:37:29,753 - process-run startOutputAndReadThreads
28 Apr 2011 04:37:29,754 - process-run success.
28 Apr 2011 04:37:29,754 - process-run getOutputStream
```

```

28 Apr 2011 04:37:29,754 - process-run success.
28 Apr 2011 04:37:29,755 - process-run waitfor
28 Apr 2011 04:37:29,770 - process-run success. exit code = 0
28 Apr 2011 04:37:29,773 - Task secWT_ExonSequenceExtractor running module
wtexonsequenceextractor submitted.
28 Apr 2011 04:37:30,946 - Task secWT_ExonSequenceExtractor running module
wtexonsequenceextractor started.
28 Apr 2011 04:37:31,456 - Task secWT_ExonSequenceExtractor running module
wtexonsequenceextractor failed. Reason: java.lang.RuntimeException
28 Apr 2011 04:37:31,463 - Analysis failed.

```

- The small RNA analysis module currently does not validate the input precursor sequences file. If an empty precursor sequences file is provided, the module gives an inaccurate error message:

```

Task srCount running module counts failed. Reason: Module Error: Parameter
'SmallRNA.precursor.gff.file' is required, if Mature Form file is specified.

```

- For the small RNA analysis module, if the input precursor sequences file does not have the same format as GFF files downloaded from <http://www.mirbase.org/ftp>, then the following error message may be given:

```

24 Mar 2011 08:15:45,492 FATAL [main] PipelineCommandsGenerator:210 - Error in
generating miRBase invocation command. java.lang.NumberFormatException: For input
string: "dja2534"
24 Mar 2011 08:15:45,496 FATAL [main] PluginRunner:87 -
com.apldbio.aga.common.model.plugins.ModuleException: SmallRNA Mapper Failed. Reason:
Pipeline Exception :Error in generating miRBase invocation command.
java.lang.NumberFormatException: For input string: "dja2534"
at
com.lifetechnologies.smallrna.pipeline.mappingPlugin.SmallRNAFragmentMapper.run(SmallR
NAFragmentMapper.java:98)
com.apldbio.aga.analysis.jobms.PluginRunner.doMain(PluginRunner.java:451)
com.apldbio.aga.analysis.jobms.PluginRunner.safeMain(PluginRunner.java:696)
com.apldbio.aga.analysis.jobms.PluginRunner.main(PluginRunner.java:674)

```

- The small indel module does not properly handle input BAM files that lack the RG field in the header. If there is only a single input BAM file, the module runs without error messages and generates empty results. If multiple BAM files are used in the input and one of them lacks the RG field, then the module fails with a non-descriptive error message:

```

07 Mar 2011 01:00:43,414 INFO [0x2a9659dca0] sam:293 - loading index :
/data/results/RegressionDriver/CaseManager/knownData/testInput_era3/Resequencing/Human
/vv07_data0005_BAM_simulated_testDataFromValidation_HuRefLMP/smallIndel/HuRefLMP/Reseq
MP/output/pairing/F3-R3-Paired.bam
07 Mar 2011 01:00:43,439 INFO [0x2a9659dca0] sam:293 - loading index :
/data/results/RegressionDriver/CaseManager/knownData/testInput_era2/cnvData/bam_input/
Simulated_Reads_abl_bcr.bam
ERROR in 'splitread-pileup
/data/results/RegressionDriver/CaseManager/knownData/testInput_era3/Resequencing/Human
/vv07_data0005_BAM_simulated_testDataFromValidation_HuRefLMP/smallIndel/HuRefLMP/Reseq
MP/output/pairing/F3-R3-
Paired.bam, /data/results/RegressionDriver/CaseManager/knownData/testInput_era2/cnvData
/bam_input/Simulated_Reads_abl_bcr.bam
/data/results/RegressionDriver/CaseManager/knownData/validatedReference//genomes/hg18_
validated.fasta
/data/results/projects/joshish/smallIndel/smallIndel_SI_TC1073b/outputs/small.indel/sm
allIndel_SI_TC1073b.pas.sum
/data/results/projects/joshish/smallIndel/smallIndel_SI_TC1073b/outputs/small.indel/sm
allIndel_SI_TC1073b.pas
/data/results/projects/joshish/smallIndel/smallIndel_SI_TC1073b/outputs/small.indel/sm

```

```
allIndel_SI_TC1073b.ungapped -1 -1 8 1 2 -1 1000 94404 '' 3 Primary
ProperPair,Primary '' 1' at /share/apps/LifeScope-2.0.r0-
79785_201110306171255/bin/small-indel-tool.pl line 884.'
```

- If an empty GTF file or a GTF file with no exons is passed to the whole transcriptome analysis module, then the analysis fails with a Java runtime exception:

```
28 Apr 2011 04:37:29,389 - LifeScope version: LifeScope-v2.0-r0_87695_20110427165203
28 Apr 2011 04:37:29,750 - process-run runCommandHelper(qsub)
28 Apr 2011 04:37:29,752 - process-run success.
28 Apr 2011 04:37:29,753 - process-run startOutputAndReadThreads
28 Apr 2011 04:37:29,754 - process-run success.
28 Apr 2011 04:37:29,754 - process-run getOutputStream
28 Apr 2011 04:37:29,754 - process-run success.
28 Apr 2011 04:37:29,755 - process-run waitFor
28 Apr 2011 04:37:29,770 - process-run success. exit code = 0
28 Apr 2011 04:37:29,773 - Task secWT_ExonSequenceExtractor running module
wtexonsequenceextractor submitted.
28 Apr 2011 04:37:30,946 - Task secWT_ExonSequenceExtractor running module
wtexonsequenceextractor started.
28 Apr 2011 04:37:31,456 - Task secWT_ExonSequenceExtractor running module
wtexonsequenceextractor failed. Reason: java.lang.RuntimeException
28 Apr 2011 04:37:31,463 - Analysis failed.
```

- Read lengths 25 – 100 are supported. For data with a read length less than 25, mapping is turned off by setting the mapping scheme to 0.0.0.

XSQ and BAM conversion

- XSQConverter does not properly update the NumFragmentsPassed attribute. The correct behavior should be that the attribute is set to the number of fragments passing all filtering criteria.
- The 1.3 to 2.0 BAM converter generates the 2.0 converted BAM file at the output location specified in the INI file and not at the framework-specified output location.
- The 1.3 to 2.0 BAM converter module can convert only one 1.3 BAM at a time.

LifeScope™ Software server

- If the error message `read-only db connection` is seen and users cannot access LifeScope™ Software, enter the following commands:
 - `lscope-server.sh stop`
 - `/bin/rm <installdir>/lifescopeserver/UserDB/db*.lck`
 - `lscope-server.sh start`

If the problem persists, or if the error message is different (indicating a corrupt db), enter the following commands to delete the users repository:

- `lscope-server.sh stop`
- `rm -rf <installdir>/lifescopeserver/UserDB`
- `lscope-server.sh start`

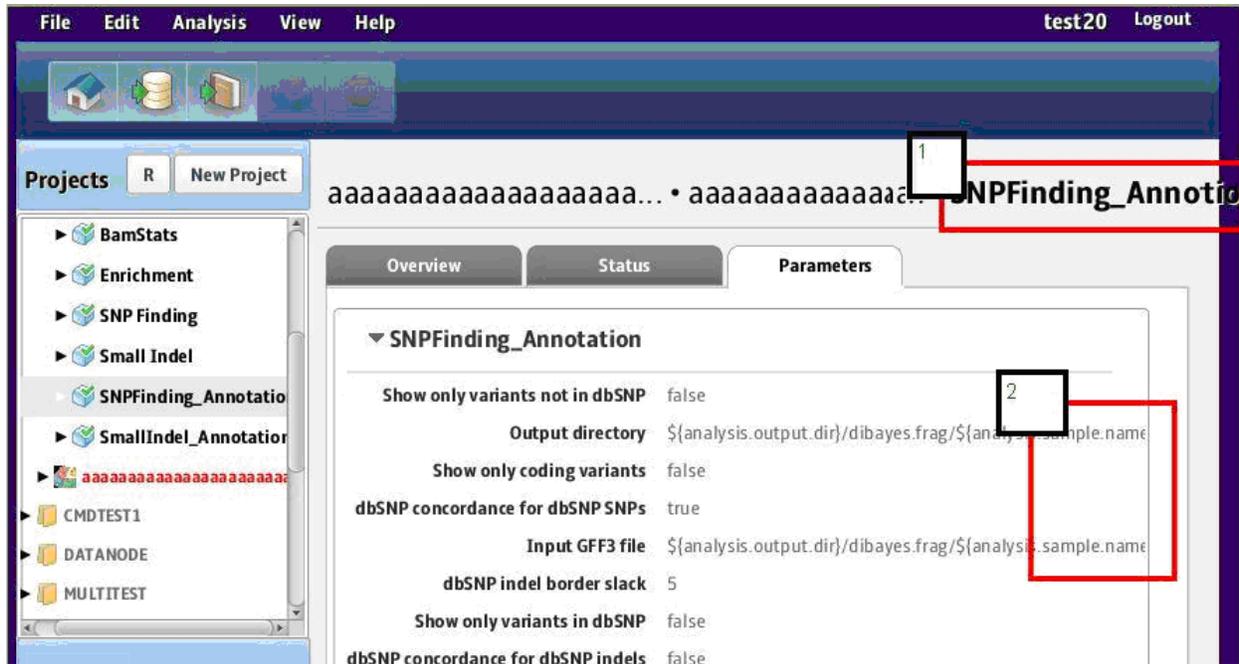
These commands remove all current LifeScope™ Software user accounts. The admin must recreate the user accounts.

GUI

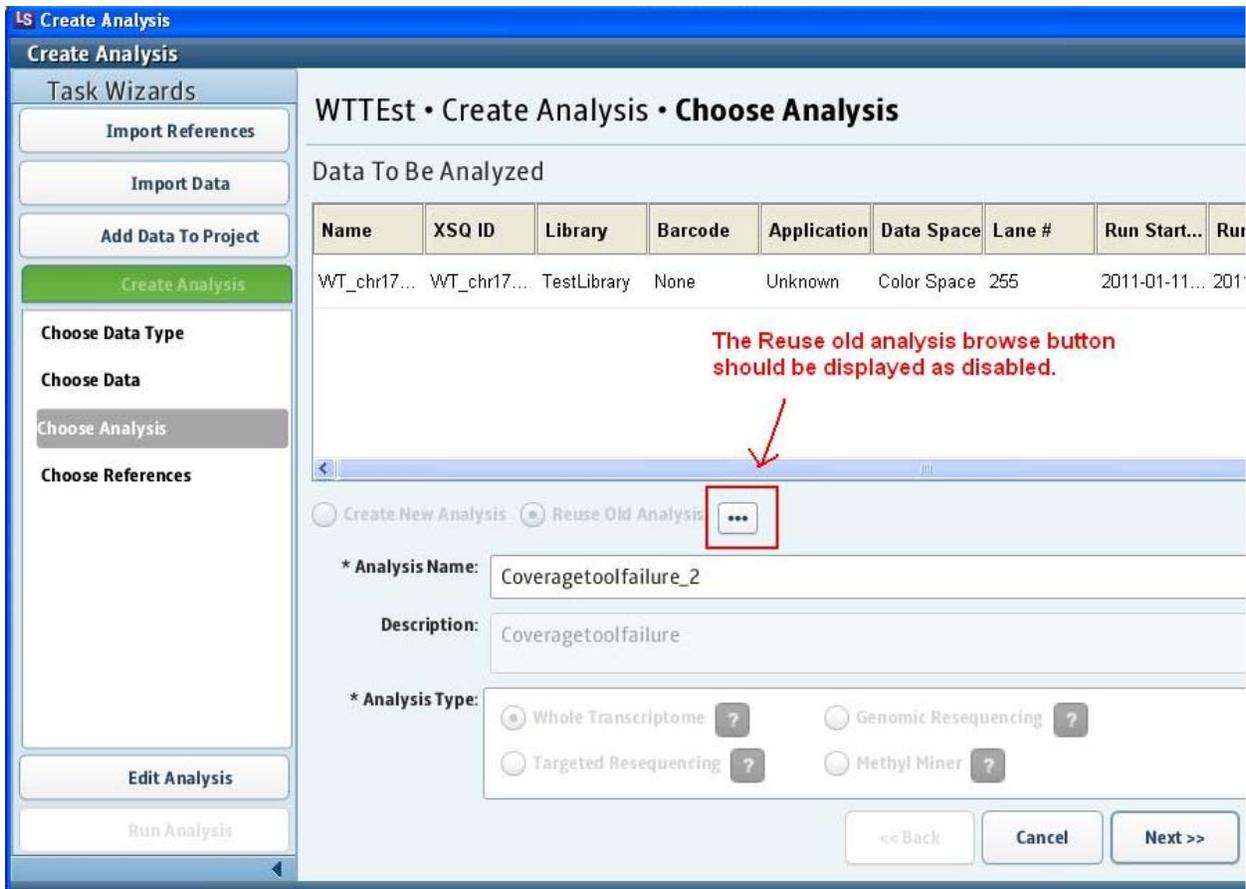
- If the GUI user guide is missing from the Help menu in the UI, request your system administrator to follow these instructions in order to setup access to the document:
 - a. Download the latest version from <http://solidsoftwaretools.com/gf/project/lifescopes>
 - b. Once the file is downloaded, rename the file as `GUI_UserGuide.pdf`.
 - c. Open a terminal and follow these steps:
 1. `$chmod -R 777 <lifescopes_install_dir>/ui`
 2. `$cp GUI_UserGuide.pdf <lifescopes_install_dir>/ui/`
 3. `$chmod -R 775 <lifescopes_install_dir>/ui`

Now the `UserGuide` menu item in the Help menu opens the user guide.

- When you submit an analysis, it may take several seconds before the Project Status view is updated in the Home View.
- If you provide a CSFASTA file as input, the folder containing the CSFASTA file ideally contains only one QUAL file. If the folder contains multiple QUAL files, both the first and last QUAL files are auto-selected. You must manually select the correct QUAL file and deselect the others.
- For analyses that are launched from the command shell interface (`lscope.sh`), the GUI does not properly display the parameters of secondary analysis (mapping and pairing) in the Parameters tab. Only parameters of tertiary analyses are currently displayed.
- For analyses that are launched with custom workflows, clicking on the Parameters tab leads to a pop-up message `Parameters cannot be displayed for a custom workflow as intended`. However, even though the message comes up, parameters from the previously selected analysis can still be seen in GUI.
- In the case of multiple read-sets of different library types added to an analysis, in order to use the **Select All** checkbox in the Choose data Type view, first select one read-set of the desired library type. Then the **Select All** checkbox selects all read-sets of that library type.
- On low resolution (1024 X 768 or under) displays, lengthy module names, analysis names, and path names may run off the screen. Please see the screenshot below with highlighted numbers 1 and 2:



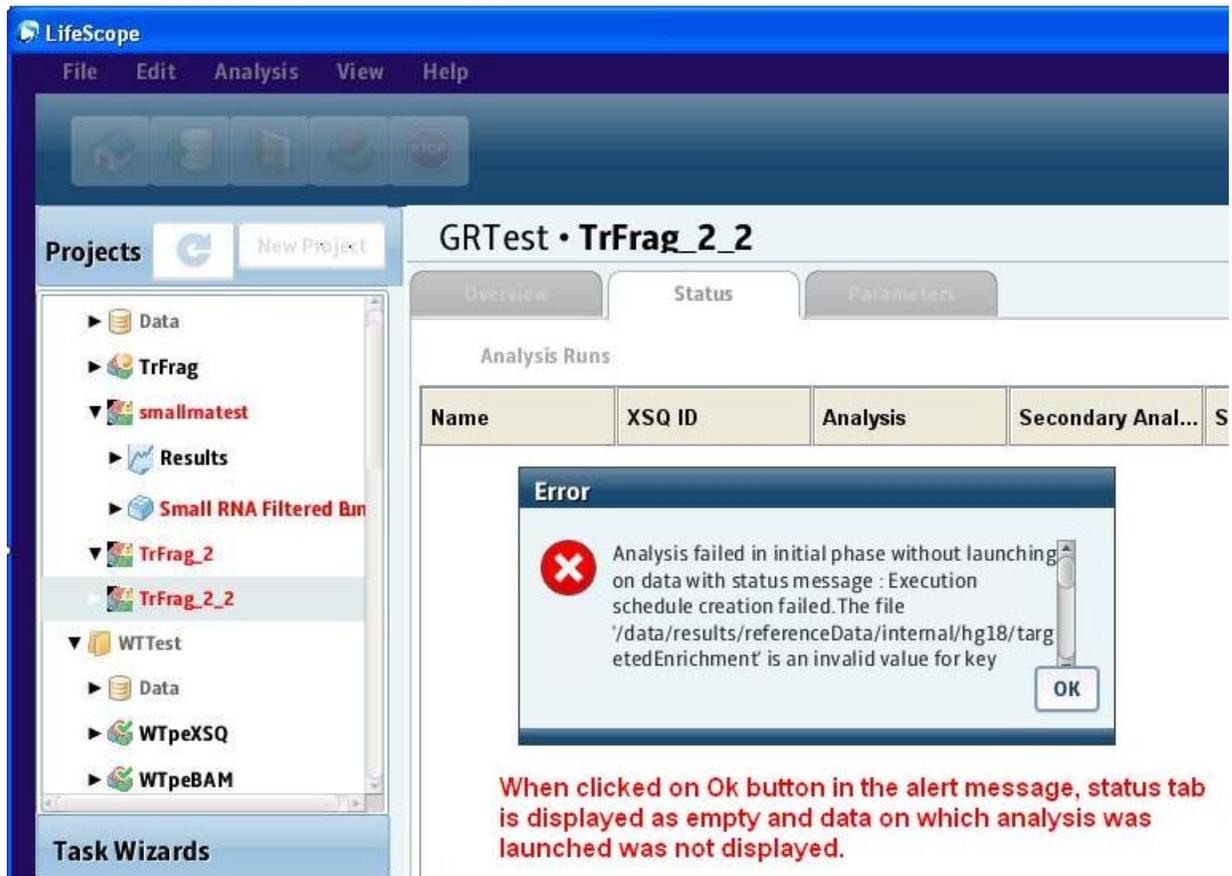
- In order to open an Edit menu for a text box, right-click on the text box.
- In the LifeScope™ Software Admin Portal, if one logs in as the master user (`lifescope`), it is currently possible to accidentally deactivate the master user, or to change its role from admin user to normal user. Care should be taken to avoid either of these actions, otherwise a system lockout may occur. In case of a system lockout, use the `resetpwd.sh` script to reset the master password.
- Selections made in the Set General Parameters view are not saved in the following situations:
 - When the analysis is reused.
 - When the analysis is edited, saved, and then edited again.
- If you deselect a module from a standard workflow, the deselected module is not available (cannot be selected in the GUI) later on under these situations:
 - You save the analysis without launching it, and then edit the analysis.
 - You reuse the analysis.
- Two methods to reuse an analysis are described in the “Reuse an analysis” section in Chapter 7 of the *LifeScope™ Genomic Analysis Software User Guide* (Part no. 4465696). These two methods use different data:
 - Method 1 uses the same data as the last time the analysis was performed.
 - With Method 2, you visit the Choose Data Type view and have the opportunity to select different input data.
- When the user selects a previously run analysis and clicks on “Reuse...” under the Analysis main menu, LifeScope™ Software navigates to the Choose Analysis screen in reuse mode. However, even in reuse mode, the **Reuse Old Analysis** button is enabled incorrectly, as shown in this screenshot:



Do not click the **Reuse Old Analysis** button when already in reuse mode.

- When the annotations module (to add genomic annotations) fails, its results files are not available in the View Results window.
- Information entered in Set General parameters is lost if the user clicks on **Save & Finish** instead of **Save & Review**. In that case, the user needs to click on **Edit Analysis** and reset the parameters.
- In the Find Data view, any read-sets already in the project are displayed as selected and are disabled from selection.
The following steps describe how to delete read-sets from a project:
 - a. Select the project in the left navigation panel.
 - b. Click **Add Data to Project**.
 - c. Go to the Group Data view.
 - d. Select the checkboxes of the read-sets or read-set group to be deleted from the project.
 - e. Click the **Delete** button just below the Groups in Project table.
 - f. Click **Cancel** on the Group data wizard.
- The LifeScope™ Software GUI does not automatically close a child window even though its parent window is closed. For example, while adding data to a project, if the user chooses to **Edit AND filter** in the Find Data page, a new dialog box would pop up. At this point, if the user aborts the process and closes the Add Data to Project window, the Edit AND filter dialog box would remain open even though the parent window is closed.

- If user clicks on a failed analysis or on its status tab, an error message is displayed as expected. However, the data on which analysis was run is not displayed, which makes it awkward to restart the analysis. An example is shown in this screenshot:



- The GUI does not support grouping BAM files together for tertiary analysis. In the GUI each BAM file is analyzed separately. The command shell does support combining BAM files into a group for a combined analysis.
- When you click the **Refresh** button, wait for the message All projects have been refreshed before making other selections in the GUI.
- The GUI does not allow you to change the grouping of input reads files. For a GUI run with a different grouping of input files, you must instead create a different analysis. The command shell does support changing the grouping of input reads files.

Command shell interface

- Avoid declaring the same parameter multiple times in either one INI file or multiple INI files (such as in a `global.ini` file and also in a mapping or `cnv` INI file). Only the last occurrence of the parameter is used in the analysis.

Mapping

- The third-party BFAST mapping module is not supported in 2.0.

Mapping statistics

- The BAMStats module counts N's in reference as missing coverage, when those positions really should not be considered while counting coverage. Consequently, the frequency distribution may claim that ~7.5% of the genome is uncovered, when in reality the figure is closer to ~0.4%.
- In order to generate mapping statistics on 1.3 BAM files, the 1.3 BAM files must first be converted to 2.0 BAM files. (This conversion can be done in both the GUI and the CLI.)
- Mapping statistics are not generated for the contents of unmapped or filtered BAM files that are created during 2.0 standard workflows.
- For charts that are titled *Distribution for Unique Alignments*, the term “unique alignments” is interpreted in this context as meaning primary alignments.

BAM files

- For your completed analyses that are run on XSQ data, intermediate BAM files are available in the project repository and may be used as input to tertiary analyses. The GUI displays these files as selectable for input. Command shell users must specify these files by absolute path on the local file system. The location of the project repository is set during installation. The path within the project repository to the intermediate BAM files on the file system is:

```
projects/user/project/analysis/outputs/bams
```
- When you use an intermediate BAM file as input in an analysis in the GUI, the reference is shown as already specified. Because in most cases the reference used in tertiary analyses must match the reference used during mapping, do not change the pre-selected reference.

Small indel module

- It is possible for an indel is reported twice in the output file. This situation occurs when an indel is detected separately in two different locations, from different sets of reads. After further processing, the two locations are resolved to the same location. This issue results in a second indel call with the same alleles and position, but different reads are being used to call the second one. These two entries represent a single insertion or deletion. (00057, Deferred from release 1.3)

Large indel module

- The large indel module might give a number of false positive calls when run on very high coverage (> several 100X) data. As a workaround reduce the number of reads per run to reduce coverage to 50X or less, and also set the parameter `large.indel.high.coverage` to 1. (Deferred from release 1.3)

Small RNA module

- The shell command `set workflow hg18` (or `hg19`) incorrectly sets the parameter `analysis.mirbase.mature.file` to `README.txt`. You must manually set the parameter correctly *after* the `set workflow` command.

Log files

- The `summary.log` file is originally intended to be a condensed version of the `analysis.log` file, containing only the most important INFO/ERROR messages from the latter. In practice, however, there is a fair amount of redundancy and `summary.log` still includes the majority of messages in `analysis.log` file.
- When LifeScope™ Software is run in single server mode, the last 600 characters of the module-specific logs may be repeated. This issue does not seem to affect cluster installations, nor does it show up in the main logs. Here is an example of the repetition from the end of the fragment mapping log:

```
19 Feb 2011 11:53:46,087 INFO [main] FragmentMapper:333 - sorting
/scratch/solid/corona/a1_fragmentMapping.main.20110219195232175/fragmentMapping.1.run.
20110219195232503/unsortedBam/fragmentMapping.1.run-mergedMa_F3-1-1-1.bam
19 Feb 2011 11:53:49,600 INFO [main] FragmentMapper:333 - sorting
/scratch/solid/corona/a1_fragmentMapping.main.20110219195232175/fragmentMapping.1.run.
20110219195232503/unsortedBam/fragmentMapping.1.run-mergedMa_F3-2-1-1.bam
19 Feb 2011 11:53:53,048 INFO [main] PluginRunner:715 - >>>> END of PluginRunner >>>>
date=2011-02-19 11:53:53.048 PST
19 Feb 2011 11:53:53,048 INFO [main] PluginRunner:716 - >>>> END of PluginRunner >>>>
date DURATION=1 minutes 20 secs
_F3-0-1-1.bam
19 Feb 2011 11:53:46,087 INFO [main] FragmentMapper:333 - sorting
/scratch/solid/corona/a1_fragmentMapping.main.20110219195232175/fragmentMapping.1.run.
20110219195232503/unsortedBam/fragmentMapping.1.run-mergedMa_F3-1-1-1.bam
19 Feb 2011 11:53:49,600 INFO [main] FragmentMapper:333 - sorting
/scratch/solid/corona/a1_fragmentMapping.main.20110219195232175/fragmentMapping.1.run.
20110219195232503/unsortedBam/fragmentMapping.1.run-mergedMa_F3-2-1-1.bam
19 Feb 2011 11:53:53,048 INFO [main] PluginRunner:715 - >>>> END of PluginRunner >>>>
date=2011-02-19 11:53:53.048 PST
19 Feb 2011 11:53:53,048 INFO [main] PluginRunner:716 - >>>> END of PluginRunner >>>>
date DURATION=1 minutes 20 secs
```

Resume command

- The `resume` command does not repeat the analysis of completed modules. However, a partially completed module is rerun from the beginning. For example, consider a workflow run that completes mapping and then fails halfway through the BAMStats module. When you resume this analysis, mapping is not rerun but the BAMStats processing is completely redone.

Targeted resequencing

- To run a targeted resequencing analysis, you must set the `analysis.regions.file` parameter to your regions of interest BED file:
 - When using the GUI, set the `analysis.regions.file` parameter in the Set General Parameters view.
 - When using the command shell, set the parameter in your `global.ini` files, with this command: `set analysis.regions.file /path/to/BED_file`

If the `analysis.regions.file` parameter is not set, the targeted resequencing run fails during tertiary processing.

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

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.

LIMITED USE LABEL LICENSE

No right to resell this product or any of its components is conveyed expressly, by implication, or by estoppel. For information on obtaining additional rights, please contact outlicensing@lifetech.com or Out Licensing, Life Technologies, 5791 Van Allen Way, Carlsbad, California 92008.

NOTICE TO PURCHASER: DISCLAIMER OF LICENSE

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. Internet Explorer is a registered trademark of Microsoft Corporation in the United States and other countries. Mozilla is a trademark of Mozilla Foundation Corporation.

© 2011 Life Technologies Corporation. All rights reserved.

Nov 2011