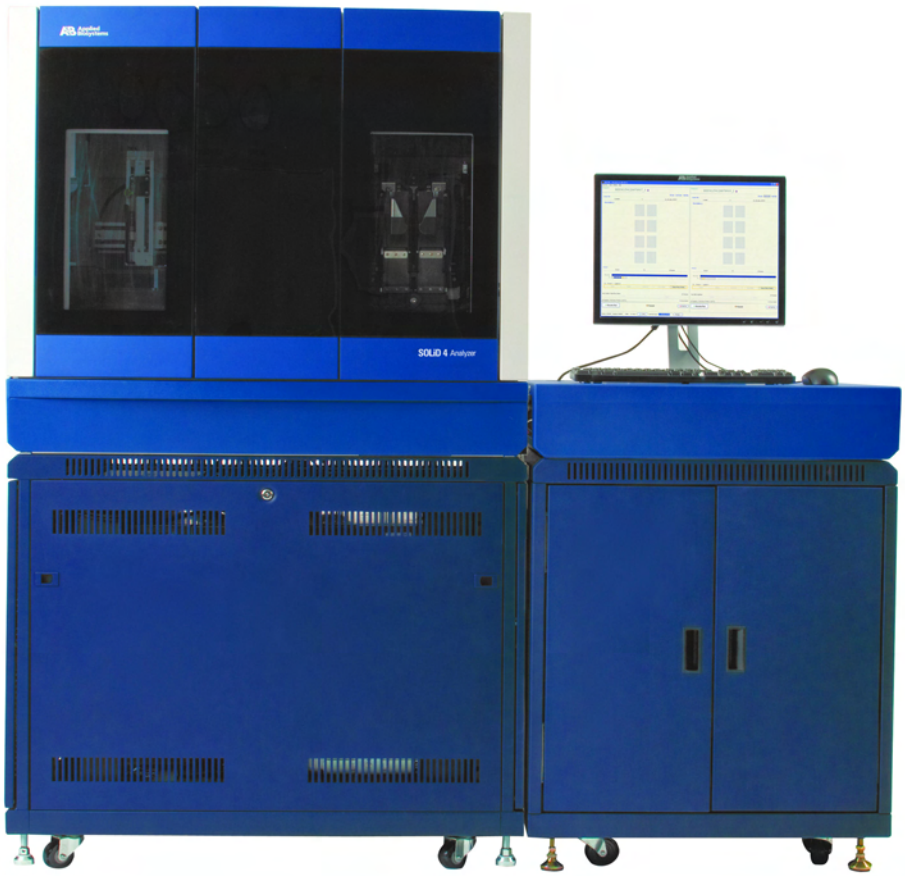
applied biosystems™
by life technologies™

Applied Biosystems SOLiD™ 4 System

SOLiD™ SAGE™ Analysis Software v1.10 Guide

SOLiD™ SAGE™ Tag Preparation | Templated Bead Preparation | Instrument Operation



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

Part no. MAN0001685

Rev. date 29 June 2010

Table of Contents

Description of the Software 4
Download and Installation 5
Using the Software 6
Getting Started 6
Mapping Sequence Files 7
Mapping Multiple Sequence Files with Different Parameters 11
Comparing Two Mapped Samples 12
Resetting and Exiting the Program 13
Appendix 14
Frequently Asked Questions 14
Examples of Files and Formats 15
Technical Support 19
End User License Agreement 20

Description of the Software

Overview

SOLiD™ SAGE™ Analysis Software v1.10 is a Linux-based program that takes the raw data files from SOLiD™ SAGE™ sequencing reads and matches them to known sequences in your reference database of choice. It is designed for use with the SOLiD™ SAGE™ Kit or the SOLiD™ SAGE™ Kit with Barcoding Adaptor Module, which generates libraries of 27-bp tags for all transcripts in a cell.

Using the simple software interface, you create a project name and location, and then identify the directories containing:

- Sequences from a SOLiD™ System read (csfasta formatted)
- A fasta-formatted database of reference sequences (*e.g.*, from RefSeq).

The software then automatically matches the experimental data with the reference data or compares two experimental data sets. The software identifies tags, tabulates tag abundances, and displays associated gene descriptions. It also sorts multiplexed libraries by their barcodes. It presents the data in a tabular format that is easy to read and export to spreadsheet or other programs for further analysis.

How Tags are Processed

The user selects the folders containing the SOLiD™ SAGE™ reads and reference sequences, and specifies the parameters of tag length and number of mismatches allowed. The software then maps the reads using only a portion of the reference sequences, to greatly reduce processing time. The algorithm extracts tag-length fragments flanking the *Nla* III sites (CATG) in the reference sequences, and concatenates unique fragments together with a spacer of NNN. Only the first tag-length bases of SOLiD™ reads are used in mapping.

Output Files

All output files are in tab-delimited text format. The output files for the individual barcoded libraries in a multiplex SOLiD™ sequencing run are stored in subdirectories of the main project directory. The output files are:

- A mapping output file that lists each tag, its frequency of occurrence, its GenBank Identifier (GI) number, and a description of the identified gene
- A mapping results file that provides more detailed information, including SOLiD™ sequencing read IDs and mismatches
- A file comparing the tags in two different samples
- A file calculating the abundances of repeat reads

To transfer the files to a Windows-based computer for analysis (e.g., by Microsoft Excel), you can use WinSCP or equivalent.

Download and Installation

System Requirements

SOLiD™ SAGE™ Analysis Software v1.10 has the following system requirements:

Linux operating system

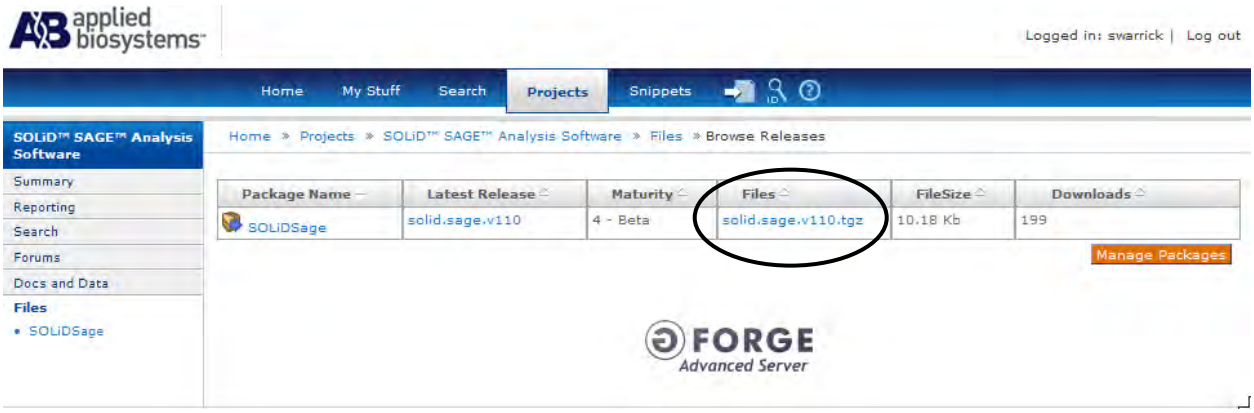
Perl Tk

Reference sequences, in NCBI fasta format (*e.g.*, RefSeq files)

≥8 gigabytes of memory—varies depending on the size of the reference genome, selected tag length, and selected number of allowed mismatches

Downloading and Installing SOLiD™ SAGE™ Software

1. To download the software, go to: http://solidsoftwaretools.com/gf/project/solid-sage
2. Click on the **Files** link in the left navigation bar, and then select the "scripts" tgz file (e.g., **solid.sage.v110.scripts.tgz**) to begin the software download.



Completing the Installation

1. Extract the downloaded **.tgz** file into the directory where you want to install the software.
2. Open **.bash_profile** in your home directory with a text editor program.
3. Add the following line, substituting the actual directory name from Step 1: **PATH=$PATH:/solid_sage_directory**
4. Add the line: **export PATH**
5. Save **.bash_profile** and exit the text editor.
6. Logout and re-login.

Downloading Sample Data

1. To download sample data, go to: http://solidsoftwaretools.com/gf/project/solid-sage
2. Click on the **Docs and Data** link in the left navigation bar, and then select **Sample Data**.
3. Select the "samples" tgz file (e.g., **solid.sage.v110.samples.tgz**) to begin the data file download.
4. The tgz file contains sample data for each mapping and comparison option, and results are provided in a corresponding directory. An example Description File for mapping multiple sequence files with different parameters is provided as well (see page 11).

Getting Started

Launching the Program

To launch the software, type **solid.sage.v110.pl**. (The software should be in your working path, so you can launch it from anywhere.)

If you have a problem launching the software, try entering the full path of the directory where you extracted the package followed by the program name.

Example launch commands:

solid.sage.v110.pl

solid.sage.v110.pl &

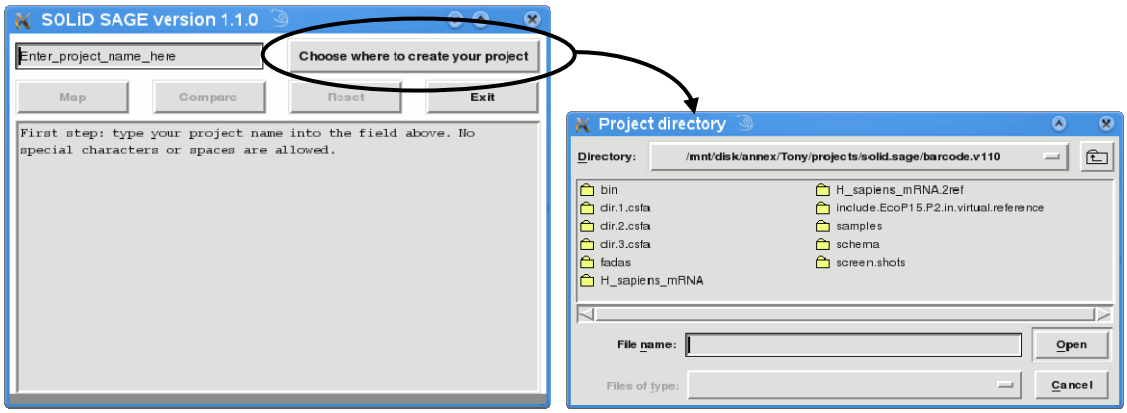
solid_sage_directory/solid.sage.v110.pl &

The program screen will open.

Creating a Project

1. To begin, enter a project name in the top left field, using only alphanumeric characters with no spaces.
2. Next, click on **Choose where to create your project** to specify the file path to the project directory. A directory with the project name you entered will be created containing temporary analysis files and final output files. (Make sure a directory with that project name does not already exist in that location.)

   Note: You must have "write" permission to the location of the project directory.
3. When you have made your selection, the full path will be displayed in the top left box, and the **Map** and **Compare** buttons will become available.



Next Steps

With the project name and location created, you can now proceed to:

Sequence Files and, starting on the next page

Comparing Two Mapped Samples, page 11

Generating Repeat Reads Statistics, page **Error! Bookmark not defined.**

Mapping Sequence Files

Sequence Files from the SOLiD™ System

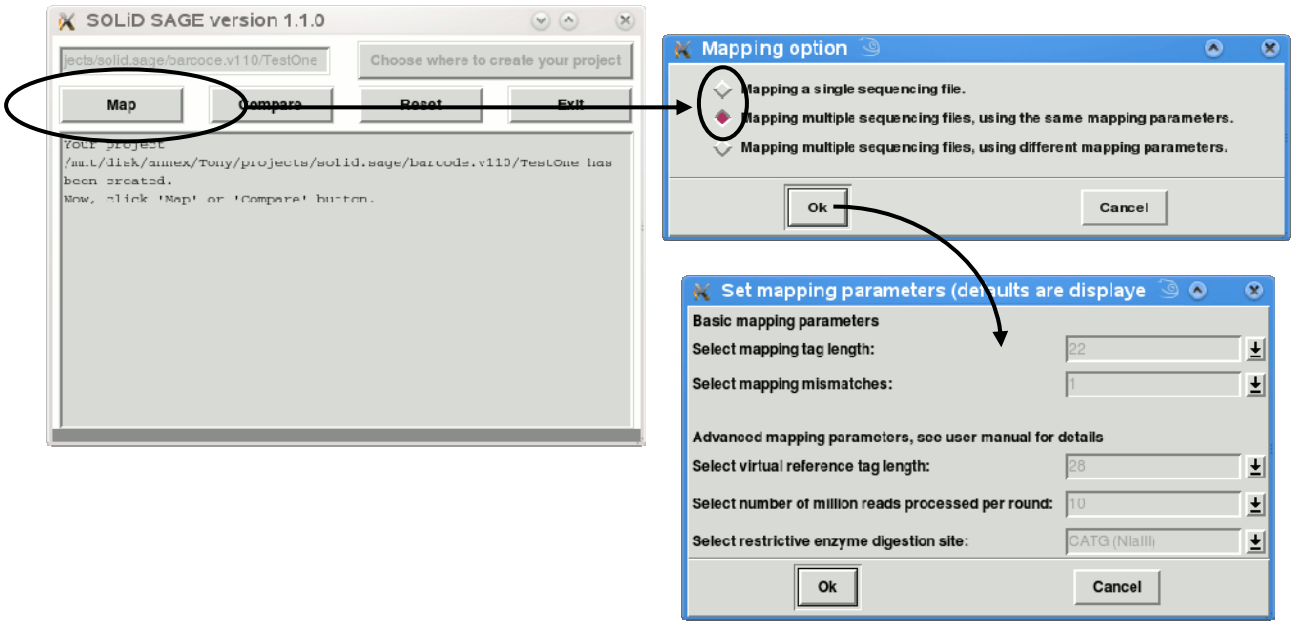
Sequences files generated by the SOLiD™ System are in the color-space fasta format (**.csfasta**). SOLiD™ SAGE™ Analysis Software v1.10 can analyze a single .csfasta results file or multiple files (e.g., from different barcoded libraries in a single run). This section describes mapping sequence files using a single set of mapping parameters.

Note: All csfasta files to be analyzed must be located in the same directory.

Besides read definition lines (beginning with a > sign) and the actual read sequence (beginning with T followed by the digit 0/1/2/3), there may be some commenting lines which begin with a # sign at the beginning of the file. The first read definition line should appear within the first 100 lines of the input reads. If it appears later than 100 lines, you will need to trim the preceding comment lines.

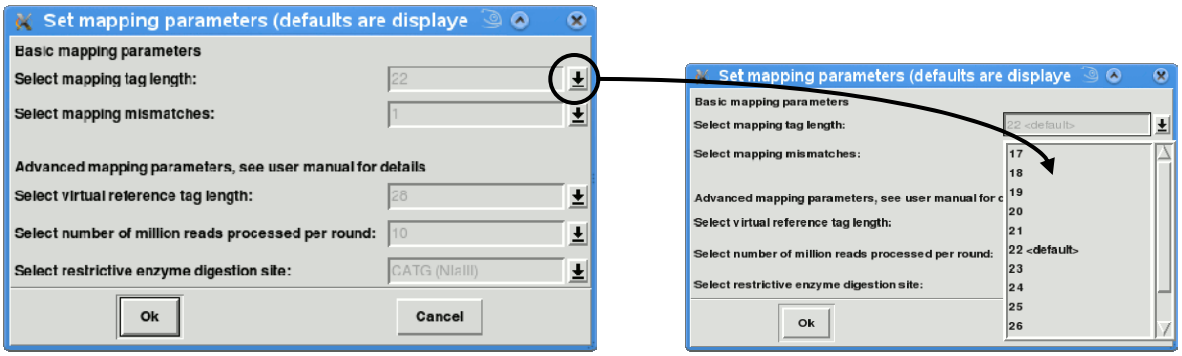
Mapping a Single File or Multiple Files with the Same Parameters

1. Click on the **Map** button.
2. Select either **Mapping a single sequencing file** or **Mapping multiple sequencing files using the same parameters**.
3. Click on **OK**. The **Set mapping parameters** dialog will open.



Selecting Mapping Parameters

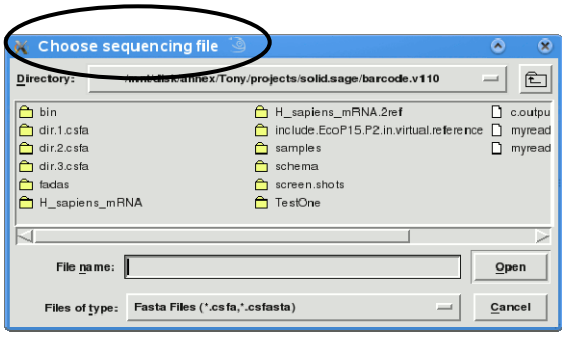
In the **Set mapping parameters** dialog, the default parameters are preselected. Click on the pulldown arrow to change each selection, then click on **OK**.



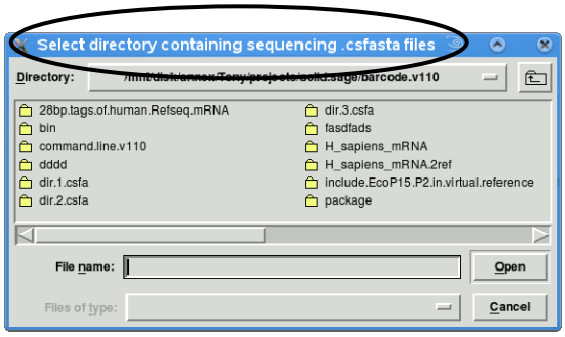
Parameter	Default	Explanation
Mapping tag length	22	The number of bases after the restriction enzyme site (CATG) in each tag used for mapping.
Mapping mismatches	1	The number of base mismatches allowed during mapping. Empirical data have shown that 1 mismatch provides the optimal balance of mapping accuracy and robustness. Increasing the number of mismatches allowed will decrease accuracy and increase processing time.
Virtual reference tag length	28	The length of each reference sequence following a restriction anchor site that is analyzed for mapping purposes.
Number of million reads processed per round	10	The number of tags (in millions) processed at one time by the software. If this number is less than the total number of tags in a run, the software will process the tags in multiple "rounds" of processing and then combine the data into a single result. This enables slower computers to process large numbers of tags without crashing, but also increases processing time.
Restriction enzyme digestion site	CATG (*Nla* III)	The default restriction enzyme used by the SOLiD™ SAGE™ System is *Nla* III. Alternative workflows may use a different restriction enzyme.

Selecting the Sequencing File(s)

After you select the mapping parameters, you will be prompted to select either a single .csfasta file or the directory containing multiple files, depending on your selection in the **Mapping option** dialog box. Multiple files must be located in the same directory, and all the files in that directory will be analyzed.



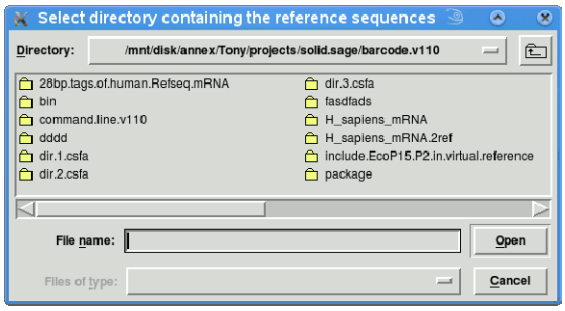
or



Selecting the Reference Directory

After you have selected a sequence file or directory and specified the mapping parameters, you will be prompted to select the directory containing the reference sequences.

Note: You do not select the reference files themselves—only the directory. The software will use every fasta file in that directory for its analysis. See **Reference File Format** below.



Reference File Format

In general, the reference files should have the same format as the fasta files from an NCBI RefSeq database. You can simply download the appropriate gDNA/mRNA database from NCBI RefSeq and use it as your reference.

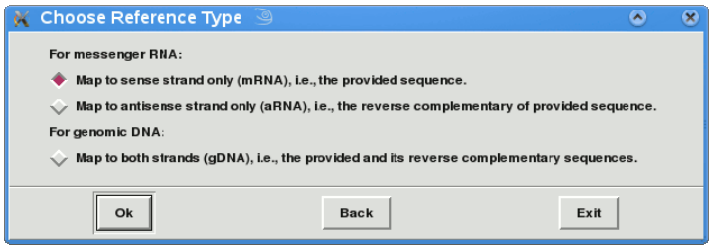
The reference files must be in the fasta format (**.fasta**). The definition lines of the reference sequences should have an NCBI RefSeq sequence format or have the following structure: >gi|xxx|ref|yyy

… where xxx is the GI number and yyy is the gene name. There is no length restriction on the GI number or gene name.

The fasta file can have multiple sequences, but each definition line should follow the structure defined above. No comment lines are allowed.

Selecting the Reference Type

After selecting the reference library, choose whether to map tags to the **sense strand** for mRNA, **antisense strand** for aRNA, or **both strands** for genomic DNA.



Note on Virtual Reference Sequences

The program generates a virtual reference sequence that is much shorter than the actual reference sequence so that the analysis will run faster. The virtual reference sequence consists of fragments of a pre-set length—either the default 28 base pairs that follow the *Nla* III anchor site CATG or a user-selected tag length (as selected under **Select mapping parameters**).

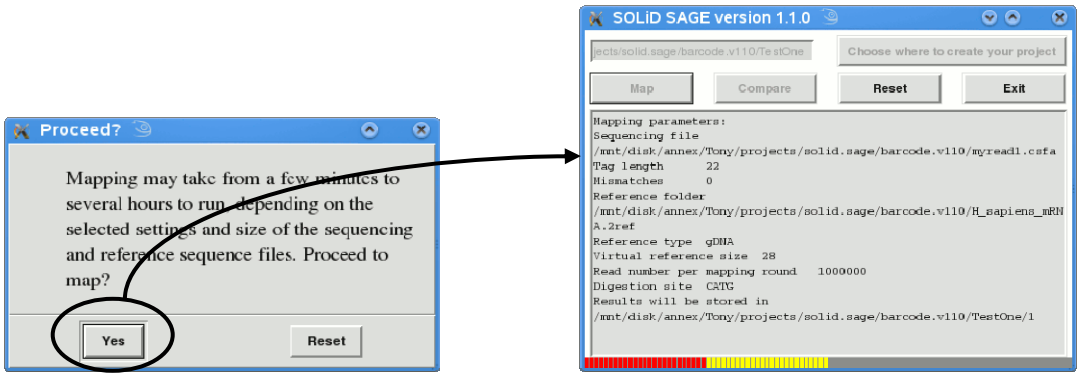
For mRNA or aRNA mapping, only the anchor site CATG and the following bases on the sense or antisense strand will be used as the virtual reference.

If genomic DNA mapping, CATG and the following bases on both strands—the provided sequence and its reverse complementary sequence—will be used as the virtual reference.

Running the Analysis

After you select the reference type and click on **OK**, you will be prompted to proceed to mapping. Click on **Yes**. After a short processing delay, a progress bar will appear at the bottom of the main software window.

The length of time required for mapping will vary depending on the size of the reference database, the selected tag length, and the number of mismatches allowed. Complex analyses may take several hours.



Output and Results Files

For each fasta file in the reference directory, the SOLiD™ SAGE™ Analysis Software generates a corresponding subdirectory in the selected project directory.

In each subdirectory, two files are generated by the mapping program:

A mapping Output file (**output.tab**) that lists each tag, its frequency of occurrence, its associated GenBank Identifier (GI) number, and a brief description of the identified gene

A mapping Results file (**results.tab**) that provides additional information, including SOLiD™ sequencing read IDs and mismatches

Output file example:

	A	B	C	D
1	Tag	Count	GI	Description
2	CATGAAAAAACTCCAAATAAGAGAATC	1	GI33413399	>gi|33413399|ref|NM_001984.1| Homo sapiens esterase D/formylglutathione hydrolase (ESD), mRNA
3	CATGAAAGGGTCACTTCTGTAATAGTG	1	GI81158221::GI81158223::GI81158225	>gi|81158221|ref|NM_001037133.1| Homo sapiens neuronal cell adhesion molecule (NRCAM), transcript variant 3, mRNA::>gi|81158223|ref|NM_005010.3| Homo sapiens neuronal cell adhesion molecule (NRCAM), transcript variant 2, mRNA::>gi|81158225|ref|NM_001037132.1| Homo sapiens neuronal cell adhesion molecule (NRCAM), transcript variant 1, mRNA
4	CATGAACAACCGGCTGGCCGAGACCAG	1	GI11545760	>gi|11545760|ref|NM_022055.1| Homo sapiens potassium channel, subfamily K, member 12 (KCNK12), mRNA
5	CATGAACTTGATACGTCCGTGTGTCCC	1	GI53759150	>gi|53759150|ref|NM_005063.4| Homo sapiens stearoyl-CoA desaturase (delta-9-desaturase) (SCD), mRNA
6	CATGAAGATGATATGAGGCCGGGGCGG	1	GI52145308	>gi|52145308|ref|NM_032808.5| Homo sapiens leucine rich repeat and Ig domain containing 1 (LINGO1), mRNA
7	CATGAAGGAAGATCCCACAGTCTCAGC	1	GI4758483	>gi|4758483|ref|NM_004832.1| Homo sapiens glutathione S-transferase omega 1 (GSTO1), mRNA
8	CATGACAGCCCTCTGCTCTTGAGTACC	1	GI39812204	>gi|39812204|ref|NM_025164.3| Homo sapiens KIAA0999 protein (KIAA0999), mRNA
9	CATGACGGAACAATAGGACTCCCCAGG	2	GI38505192	>gi|38505192|ref|NM_000954.5| Homo sapiens prostaglandin D2 synthase 21kDa (brain) (PTGDS), mRNA
	CATGACGTGTCTATGTCAAAAGTTCTT	1	GI115583669	>gi|115583669|ref|NM_003253.2| Homo sapiens T-cell lymphoma invasion and

Results file example:

	A	B	C	D	E
1	Tag_Seq	GI_num	GI_Pos	Read_ID	Mismatch
2	CATGTGCAAATAAATGTGGCTTAGACT	133908618	3298	>665_1155_42_F3	0
3	CATGGTAATAAAATATGAATGATAAAA	189083841	2125	>665_817_382_F3	0
4	CATGGTAATAAAATATGAATGATAAAA	189083835	2280	>665_817_382_F3	0
5	CATGGTAATAAAATATGAATGATAAAA	189083837	2146	>665_817_382_F3	0
6	CATGGTAATAAAATATGAATGATAAAA	189083839	2128	>665_817_382_F3	0
7	CATGGTAATAAAATATGAATGATAAAA	51593094	2283	>665_817_382_F3	0
8	CATGGTAATAAAATATGAATGATAAAA	132814488	2268	>665_817_382_F3	0
9	CATGTGATGGGCATTGAGCCACACCTC	34147410	1140	>665_1807_1132_F3	0
10	CATGAGGAGCTCGGCTTAAAATGTCTT	40254847	2324	>665_1994_720_F3	0
11	CATGAAGATGATATGAGGCCGGGGCGG	52145308	1899	>668_309_447_F3	0

Transferring Files

To transfer the output files to a Windows-based computer for analysis using a spreadsheet program such as Microsoft Excel, you can use WinSCP or equivalent.

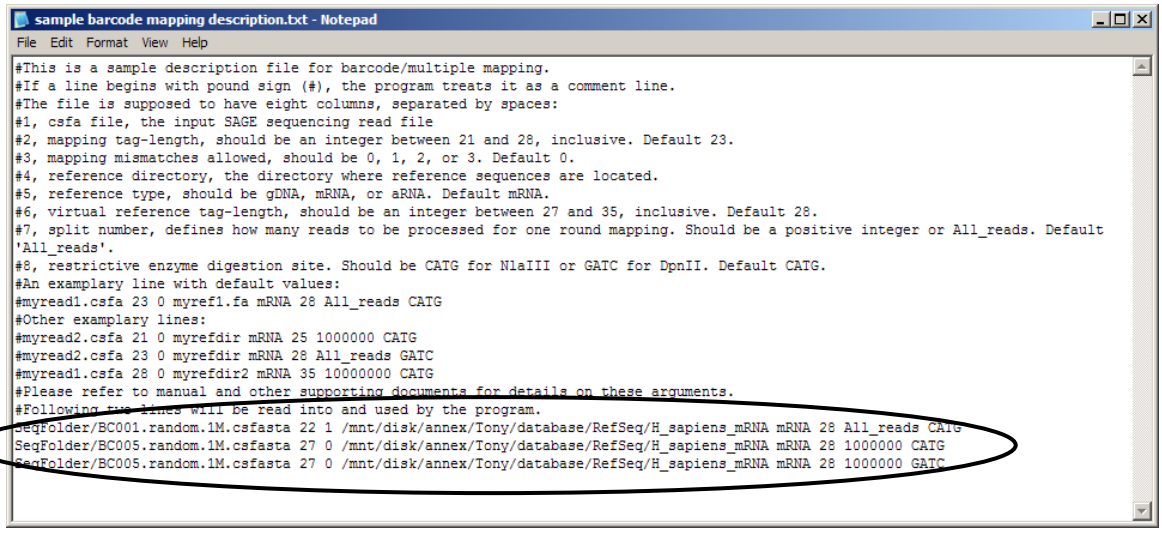
Mapping Multiple Sequence Files with Different Parameters

Description File

You can map multiple sequence files (in .csfasta format) with different mapping parameters using a **Description File** that contains all the details about each sequence file, the parameters to apply, and the reference directories to use.

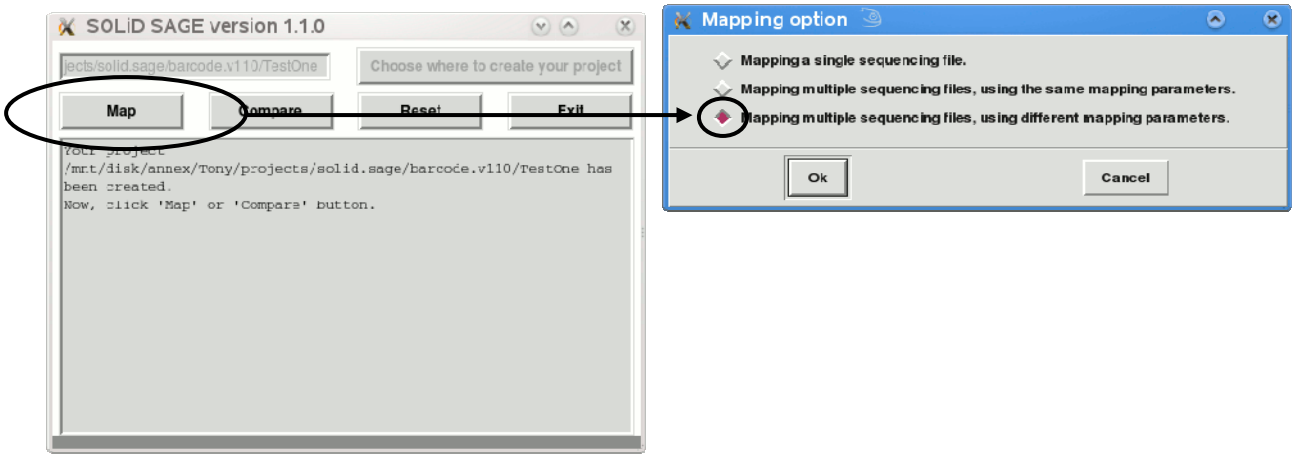
An example Description File is available as part of the sample data download. See **Downloading Sample Data** on page 5.

The example Description File is shown below. The comment lines at the top describe the different elements, and an example format is shown in the circled area, with a single space between each element.



Creating and Selecting the Description File

1. Create the Description File using the example provided with the software and shown above.
2. Launch the program and create a project, then click on the **Map** button.
3. Select **Mapping multiple sequencing files using different mapping parameters**.
4. Click on **OK**. The **Open mapping description file** dialog will open.
5. Select the Description File you created and click on **Open**.
6. The software will proceed to mapping, as described in **Running the Analysis**.



Comparing Two Mapped Samples

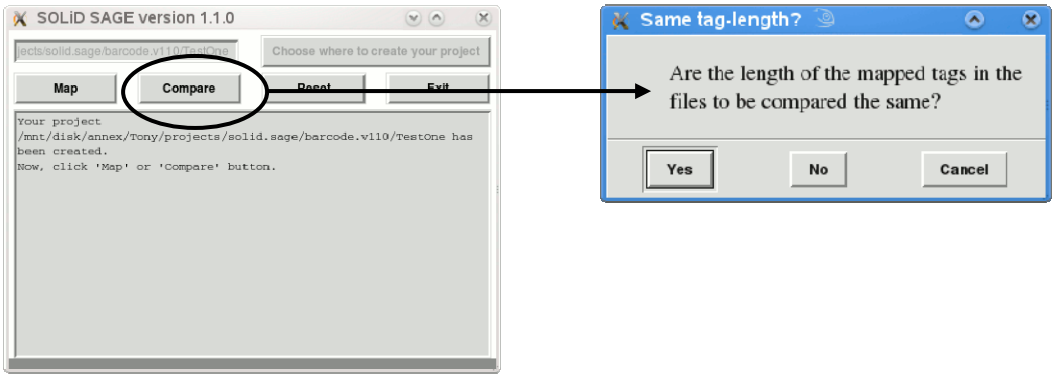
Introduction

You can compare two previously mapped libraries, as described in this section.

Selecting the Sequencing Reads to Compare

1. Launch the program and create a project, as described in **Getting Started**.
2. Click on the **Compare** button in the program screen and follow the on-screen steps to select the files to compare. These will be Output files from previous mapping runs (**output.tab** format; see page 10). Do *not* compare results files.
3. The software will find common tags between the libraries and identify the corresponding tag counts and genes. When the comparison is complete, you will receive an alert.

Note: If the lengths of the tags in the two libraries are different, all tags will be truncated to the shortest length before comparison. Note that the gene description may not match the truncated tags as well as the non-truncated ones. Be careful when comparing libraries with different tag lengths.



Comparison File

The Comparison file generated by the analysis (**comparison.tab**) lists each tag, its frequency of occurrence in each sample, and its GenBank identification.

	A	B	C	D
1	Tag_Sequence	File1	File2	
2	CATGAAGACAGTGGCTGGCGGTGCCTG	4715	5430	(Based on non-truncated tag) >gi|78214519|ref|NM_000998.4| Homo sapiens ribosomal protein L37a (RPL37A), mRNA
3	CATGCACAAACGGTAGTTTTGTGTGTT	2361	2604	(Based on non-truncated tag) >gi|169202035|ref|XM_001726056.1| PREDICTED: Homo sapiens similar to hCG2027326 (LOC100129905), mRNA::>gi|169203086|ref|XM_001725509.1| PREDICTED: Homo sapiens similar to hCG2027326 (LOC100129905), mRNA::>gi|169203606|ref|XM_001726007.1| PREDICTED: Homo sapiens similar to hCG2027326 (LOC100129905), mRNA::>gi|68160923|ref|NM_001030.3| Homo sapiens ribosomal protein S27 (metallopanstimulin 1) (RPS27), mRNA
4	CATGACAACAAAGAAAAAGACCTTGTA	1736	1606	(Based on non-truncated tag) >gi|56550064|ref|NM_001008220.1| Homo sapiens complexin 2 (CPLX2), transcript variant 2, mRNA::>gi|56550103|ref|NM_006650.3| Homo sapiens complexin 2 (CPLX2), transcript variant 1, mRNA
5	CATGAATATGTGGGCTAAGAAATAGTT	1659	1980	(Based on non-truncated tag) >gi|17999531|ref|NM_004374.2| Homo sapiens cytochrome c oxidase subunit VIc (COX6C), mRNA
6	CATGAAGCTGAGGTCTTGAAGCAGCTG	1558	715	(Based on non-truncated tag) >gi|44889961|ref|NM_005563.3| Homo sapiens stathmin 1/oncoprotein 18 (STMN1), transcript variant 3, mRNA::>gi|44890049|ref|NM_203399.1| Homo sapiens stathmin 1/oncoprotein 18 (STMN1), transcript variant 2, mRNA::>gi|44890051|ref|NM_203401.1| Homo sapiens stathmin 1/oncoprotein 18 (STMN1), transcript variant 1, mRNA
7	CATGAACTAATACTACAATAAAGGATG	1533	2163	(Based on non-truncated tag) >gi|142358075|ref|NM_152350.2| Homo sapiens chromosome 17 open reading frame 45 (C17orf45), mRNA
	CATGATCGCTTTCTACACTGTATTACA	1442	1496	(Based on non-truncated tag) >gi|41406053|ref|NM_000484.2| Homo sapiens amyloid beta (A4) precursor protein (APP), transcript variant 1, mRNA::>gi|41406054|ref|NM_201413.1| Homo sapiens amyloid beta

Resetting and Exiting the Program

Reset

Clicking on **Reset** will reset all the options in the program screen, including the project name selection. You will be prompted to complete this action.

Exiting the Program

Click on **Exit** to exit the program. Depending on the state the program is in, there may be a small or lengthy delay before exiting.

You can also end the program by using standard Unix commands (*e.g.*, CTRL-C) from the terminal where you launch the program.

Appendix

Frequently Asked Questions

Is the Program Frozen?

The program will sometimes appear frozen during processing. This often happens when you first load the SOLiD™ reads or the reference sequences, or when you attempt to switch application windows while the SOLiD™ SAGE™ program is actively running. Allow the process to continue for a few minutes before force-quitting.

Why Am I Getting An Error Message?

The program includes various error messages, most of which are self-explanatory. If the error says something about "balloon" or "cancellation", you can ignore it. These errors will not affect your results.

How Long Does Mapping Take?

The running time of the mapping analysis depends on a few factors:

- The number of reads in the input SOLiD™ SAGE™ file
- The size of the reference genome
- Input tag length
- Number of allowed mismatches (more mismatches allowed = longer processing time)
- Number of million reads per round of processing (more rounds required = longer processing time)

A typical read file with octet SOLiD™ data mapped to the human Refseq mRNA dataset with a tag length = 27 and mismatch = 2 takes ~10 minutes on a computer with a 2.33 GHz Duo CPU and 8 GB of memory (note that only one CPU and ~2 GB memory are actually used).

If you are mapping to the human genome with mismatch = 0, it may take 2–3 hours and up to 15 hours with mismatch = 2.

Examples of Files and Formats

Input File Formats

SOLiD™ reads file:

This is a standard .csfasta file, which looks like:

Tue Dec 23 01:04:30 2008 comments follow

Cwd: /home/pipeline

Title: Solid0110_20081218_JMK26_Ribominus_JMK26_C_amp_1_

>443_12_55_F3

T2112001212101120200203200000000

>443_12_118_F3

T0000000000000000000000000000000

>443_12_170_F3

T2130011300121020330330321313 01

>443_12_201_F3

T0030000323000000000000000000000

>443_12_278_F3

T0000000000000000000000000000000

>443_12_294_F3

T0000000000000000000000000000000

>443_12_336_F3

T0000000000000000000000000000000

Reference files:

This is a standard NCBI fasta file, part of a database file set in a single directory.

Files look like:

>gi|155369268|ref|NM_001100917.1| Homo sapiens tetraspanin 19 (TSPAN19), mRNA

AAACAATCTCGATTCTAAATTG… … (bases follow) … … ACTGGTG

>gi|169212695|ref|XM_001716884.1| PREDICTED: Homo sapiens hypothetical protein LOC100132679 (LOC100132679), mRNA

ATGTGTGTATATATATATACACATATATATG… … (bases follow) … … ATGGATGTAT

Comparison input file:

Same as standard output file (see next page), with a file name of the format solidsageread.csfasta.taglength.mismatch.output.tab (sample file: solidsageSampleRead.csfasta.27.0.output.tab)

Output File Names

Output file: File name format: solidsageread.csfasta.taglength.mismatch.output.tab (sample file: solidsageSampleRead.csfasta.27.0.output.tab)

Results file: File name format: solidsageread.csfasta.taglength.mismatch.results.tab (sample file: solidsageSampleRead.csfasta.27.0.results.tab)

Comparison file: File name: comparison.tab

Mapping Output File Format

Tag Count GI Description

CATGAAAAAACTCCAAATAAGAGAATC 1 GI33413399 >gi|33413399|ref|NM_001984.1| Homo sapiens esterase D/formylglutathione hydrolase (ESD), mRNA

CATGAAAGGGTCACTTCTGTAATAGTG 1 GI81158221::GI81158223::GI81158225 >gi|81158221|ref|NM_001037133.1| Homo sapiens neuronal cell adhesion molecule (NRCAM), transcript variant 3, mRNA::>gi|81158223|ref|NM_005010.3| Homo sapiens neuronal cell adhesion molecule (NRCAM), transcript variant 2, mRNA::>gi|81158225|ref|NM_001037132.1| Homo sapiens neuronal cell adhesion molecule (NRCAM), transcript variant 1, mRNA

CATGAACAACCGGCTGGCCGAGACCAG 1 GI11545760 >gi|11545760|ref|NM_022055.1| Homo sapiens potassium channel, subfamily K, member 12 (KCNK12), mRNA

CATGAACTTGATACGTCCGTGTGTCCC 1 GI53759150 >gi|53759150|ref|NM_005063.4| Homo sapiens stearoyl-CoA desaturase (delta-9-desaturase) (SCD), mRNA

CATGAAGATGATATGAGGCCGGGGCGG 1 GI52145308 >gi|52145308|ref|NM_032808.5| Homo sapiens leucine rich repeat and Ig domain containing 1 (LINGO1), mRNA

CATGAAGGAAGATCCCACAGTCTCAGC 1 GI4758483 >gi|4758483|ref|NM_004832.1| Homo sapiens glutathione S-transferase omega 1 (GSTO1), mRNA

CATGACAGCCCTCTGCTCTTGAGTACC 1 GI39812204 >gi|39812204|ref|NM_025164.3| Homo sapiens KIAA0999 protein (KIAA0999), mRNA

CATGACGGAACAATAGGACTCCCCAGG 2 GI38505192 >gi|38505192|ref|NM_000954.5| Homo sapiens prostaglandin D2 synthase 21kDa (brain) (PTGDS), mRNA

CATGACGTGTCTATGTCAAAAGTTCTT 1 GI115583669 >gi|115583669|ref|NM_003253.2| Homo sapiens T-cell lymphoma invasion and metastasis 1 (TIAM1), mRNA

Mapping Results File Format

Tag_Seq GI_num GI_Pos Read_ID Mismatch

CATGTGCAAATAAATGTGGCTTAGACT 133908618 3298 >665_1155_42_F3 0

CATGGTAATAAAATATGAATGATAAAA 189083841 2125 >665_817_382_F3 0

CATGGTAATAAAATATGAATGATAAAA 189083835 2280 >665_817_382_F3 0

CATGGTAATAAAATATGAATGATAAAA 189083837 2146 >665_817_382_F3 0

CATGGTAATAAAATATGAATGATAAAA 189083839 2128 >665_817_382_F3 0

CATGGTAATAAAATATGAATGATAAAA 51593094 2283 >665_817_382_F3 0

CATGGTAATAAAATATGAATGATAAAA 132814488 2268 >665_817_382_F3 0

CATGTGATGGGCATTGAGCCACACCTC 34147410 1140 >665_1807_1132_F3 0

CATGAGGAGCTCGGCTTAAAATGTCTT 40254847 2324 >665_1994_720_F3 0

Comparison File Format

Tag_Sequence File1 File2

CATGAAGACAGTGGCTGGCGGTGCCTG 4715 5430 (Based on non-truncated tag) >gi|78214519|ref|NM_000998.4| Homo sapiens ribosomal protein L37a (RPL37A), mRNA

CATGCACAAACGGTAGTTTTGTGTGTT 2361 2604 (Based on non-truncated tag) >gi|169202035|ref|XM_001726056.1| PREDICTED: Homo sapiens similar to hCG2027326 (LOC100129905), mRNA::>gi|169203086|ref|XM_001725509.1| PREDICTED: Homo sapiens similar to hCG2027326 (LOC100129905), mRNA::>gi|169203606|ref|XM_001726007.1| PREDICTED: Homo sapiens similar to hCG2027326 (LOC100129905), mRNA::>gi|68160923|ref|NM_001030.3| Homo sapiens ribosomal protein S27 (metallopanstimulin 1) (RPS27), mRNA

CATGACAACAAAGAAAAAGACCTTGTA 1736 1606 (Based on non-truncated tag) >gi|56550064|ref|NM_001008220.1| Homo sapiens complexin 2 (CPLX2), transcript variant 2, mRNA::>gi|56550103|ref|NM_006650.3| Homo sapiens complexin 2 (CPLX2), transcript variant 1, mRNA

CATGAATATGTGGGCTAAGAAATAGTT 1659 1980 (Based on non-truncated tag) >gi|17999531|ref|NM_004374.2| Homo sapiens cytochrome c oxidase subunit VIc (COX6C), mRNA

Repeat Reads File Format

Total reads for SOLiD™SAGE™SampleRead.csfasta: 49999
Total unique reads for SOLiD™SAGE™SampleRead.csfasta: 48630
Top 100 abundant reads for SOLiD™SAGE™SampleRead.csfasta:
T1111111111111111111111111111111..1 148
T1111111110111111111111111111111..1 135
T1111011111011111111111111111111..1 127
T3111111111011111111111111111111..1 70
T1111011111111111111111111111111..1 46
T0000000000000000000000000000010..0 39
… …
Total unique reads for SOLiD™SAGE™SampleRead.csfasta at tag 26: 48102
Top 100 abundant reads for SOLiD™SAGE™SampleRead.csfasta at tag 26:
T11110111110111111111111111 172
T11111111110111111111111111 157
T11111111111111111111111111 156
T31111111110111111111111111 90
T00000000000000000000000000 57
T31111110110111111111111111 55
T11110111111111111111111111 55
… …
Total unique reads for SOLiD™SAGE™SampleRead.csfasta at tag 27: 48363
Top 100 abundant reads for SOLiD™SAGE™SampleRead.csfasta at tag 27:
T111111111111111111111111111 150
T111111111101111111111111111 140
T111101111101111111111111111 139
T311111111101111111111111111 73
T000000000000000000000000000 57
T111101111111111111111111111 50
T111101110101111111111111111 37
T311111101101111111111111111 36
… …
Total unique reads for SOLiD™SAGE™SampleRead.csfasta at tag 28: 48412
Top 100 abundant reads for SOLiD™SAGE™SampleRead.csfasta at tag 28:
T111111111111111111111111111 150
T111111111101111111111111111 140
T111101111101111111111111111 138
T311111111101111111111111111 73
T000000000000000000000000000 57
T111101111111111111111111111 49
T311111101101111111111111111 36

Technical Support

Web Resources

Visit the Invitrogen website at **www.invitrogen.com** for:

- Technical resources, including manuals, vector maps and sequences, application notes, MSDSs, FAQs, formulations, citations, handbooks, etc.
- Complete technical support contact information
- Access to the Invitrogen Online Catalog
- Additional product information and special offers

Contact Us

For more information or technical assistance, call, write, fax, or email. Additional international offices are listed on our website (www.invitrogen.com).

Corporate Headquarters:
5791 Van Allen Way
Carlsbad, CA 92008 USA
Tel: 1 760 603 7200
Tel (Toll Free): 1 800 955 6288
Fax: 1 760 602 6500
E-mail: tech_support@invitrogen.com

Japanese Headquarters:
LOOP-X Bldg. 6F
3-9-15, Kaigan
Minato-ku, Tokyo 108-0022
Tel: 81 3 5730 6509
Fax: 81 3 5730 6519
E-mail: jpinfo@invitrogen.com

European Headquarters:
Inchinnan Business Park
3 Fountain Drive
Paisley PA4 9RF, UK
Tel: +44 (0) 141 814 6100
Tech Fax: +44 (0) 141 814 6117
E-mail: eurotech@invitrogen.com

End User License Agreement

This is a legal agreement between you, the person or entity receiving software products, and Life Technologies Corporation, through its Invitrogen business unit, having offices at 5791 Van Allen Way, Carlsbad California 92008 USA ("Licensor"). This agreement is part of a package that includes one or more software products and certain electronic and/or written materials. This agreement covers your licensing of such software.

BY CLICKING YOUR ACCEPTANCE OF THIS END USER LICENSE AGREEMENT ("Agreement"), OR BY INSTALLING OR USING THE SOFTWARE (defined below) OR ANY OTHER COMPONENT OF THE PACKAGE, YOU ACKNOWLEDGE THAT YOU HAVE READ ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT, UNDERSTAND THEM, AND AGREE TO BE LEGALLY BOUND BY THEM. If you do not agree to the terms of this Agreement, you may not install or use the Software, and may return it to Licensor. In addition to the restrictions imposed under this Agreement, the Software installation instructions or release notes shall apply to your use of the Software.

As used in this Agreement: "Authorized Users" means, collectively, the personnel authorized by you to use the Software for your benefit. Unless otherwise expressly allowed by this Agreement, Authorized Users may include only your employees and agents having a need to know, and Authorized Users may not be entities or persons in the business of licensing or otherwise providing products or services competitive with the Software. "Array" means any Licensor product which is a microarray. "Designated Site" means your facilities or offices located at the postal address provided to Licensor for your billing and invoicing purposes, unless otherwise indicated in a license key provided to you. "Software" means the software product(s) accompanying this Agreement and the content therein; including the associated user manuals, user documentation and application program interfaces, and license key(s) provided, and any patches, updates, upgrades, improvements, enhancements, fixes and revised versions of any of the foregoing that may be provided to you from time to time, and any combination of the foregoing. "affiliate" means any entity Controlling, Controlled by, or under common Control with the referenced entity, where the term "Control" means the possession, direct or indirect, of the power to direct or cause the direction of the management and policies of an entity, whether through the ownership of voting securities, by contract, or otherwise.

1. TAX OBLIGATIONS; SHIPMENT. If any authority imposes a duty, tax or similar amount (other than taxes based on Licensor's income), you agree to pay, or to promptly reimburse Licensor for, all such amounts. You are responsible for paying all shipping charges. All shipments by Licensor or its designee are FCA point of shipment (Incoterms 2000).

2. ACCEPTANCE. You will be deemed to have accepted the Software unless you provide written notice of rejection within ten (10) days after receipt of the Software or the applicable license key (whichever event occurs first). Any such notice must state the reason for rejection, and you may only reject the Software if it materially fails to comply with its accompanying documentation and user manual. If you reject the Software, Licensor's sole obligation and liability, and your sole and exclusive remedy, shall be for Licensor to use commercially reasonable efforts to deliver to you a replacement for the nonconforming Software, and if Licensor is not able to deliver a replacement for the Software, then the Agreement shall terminate.

3. GRANT OF SOFTWARE LICENSE. Subject to the terms and conditions of this Agreement, Licensor grants to you a non-exclusive, non-transferable license for Authorized Users to use the Software only to acquire and analyze data generated from an Array, and only for your internal operations and internal data processing purposes ("License"). You may install and operate the Software on a single designated computer located at the Designated Site. You may not network the Software for use at any other location at the Designated Site or any other site.

4. RESTRICTIONS ON USE. You acknowledge that you are receiving LICENSED RIGHTS only. The Software may only be used internally, by your Authorized Users, with the license key provided, for your copy(ies) of the Software. If any Software is provided on separate media (e.g., a CD-ROM), you may make a single copy solely for your internal backup purposes. You shall not directly or indirectly: (i) sell, rent, lease, distribute, redistribute or transfer any of the Software or any rights in any of the Software, including without limitation through "charge back" or any other selling, reselling, distributing or redistribution within your organization of any usage capacity you have licensed, without the prior express written approval of Licensor, (ii) modify, translate, reverse engineer (except to the limited extent permitted by law), decompile, disassemble, attempt to discover the source code for, create derivative works based on, or sublicense any of the Software, (iii) use any Software for the benefit of any third parties (e.g., in an ASP, outsourcing or service bureau relationship), or in any way other than in its intended manner, (iv) remove any proprietary notice, labels, or marks on or in Software, or (v) disable or circumvent any access control or related device, process or procedure established with respect to the Software, including the license key or any other part thereof. If the Software design permits modification, then you may only use such modifications or new software programs for your internal purposes and otherwise consistent with the License. You are responsible for all use of the Software and for compliance with this Agreement; any breach by you or any user of the Software shall be deemed to have been a breach by you. Licensor reserves all rights not expressly granted; no right or license is granted hereunder, express or implied or by way of estoppel, to any intellectual property rights other than as expressly set forth herein; and your receipt of a license to the Software does not by itself convey or imply the right to use the Software in combination with any other product(s). You agree and acknowledge that you have been provided sufficient information such that you do not need to reverse engineer the Software in any way to permit other products to interoperate with the Software.

5. NO SEPARATION OF COMPONENTS. The Software is licensed as a single product. Some Licensor software products combine separately available components into a single product (e.g., a software suite product may be comprised of multiple component products). When licensed as a combination product, the component parts may not be separated for use independently of the combination product. You must first purchase a license to each component of the combination product before you may use it independently of the combination product.

6. TERM AND TERMINATION. Unless otherwise agreed, the term of this Agreement shall continue until it is terminated. Licensor may terminate this Agreement (or any License) upon 15 days written notice if you breach this Agreement and do not cure the breach within 15 days following receipt of written notice of breach. Immediately upon any termination or expiration of this Agreement, you agree to: (a) pay all amounts owed to Licensor; (b) un-install and cease use of the Software for which your rights have been terminated; (c) upon request, return to Licensor (or destroy) all copies of the Software and any other Confidential Information or proprietary materials in your possession for which your rights have been terminated; and (d) upon request, certify in writing your compliance with (b) and (c), above.

7. CONFIDENTIALITY. You agree to protect Licensor's Confidential Information with the same degree of care used to protect your own confidential information (but in no event less than a reasonable standard of care), and not to use or disclose any portion of such Confidential Information to third parties, except as expressly authorized in this Agreement. You acknowledge that the Software, including its content, structure, organization and design constitute proprietary and valuable trade secrets (and other intellectual property rights) of Licensor and/or its licensors. The term "Confidential Information" means, collectively, non-public information that Licensor (and its licensors) provide and reasonably consider to be of a confidential, proprietary or trade secret nature, including but not limited to Software prices, and confidential elements of the Software and Licensor's (and its licensors') technology and know-how, whether in tangible or intangible form, whether designated as confidential or not, and whether or not stored, compiled or memorialized physically, electronically, graphically, photographically, or in writing. Confidential Information does not include any information which you can demonstrate by credible evidence: (i) is, as of the time of its disclosure or thereafter becomes part of the public domain through no fault of yours; (ii) was rightfully known to you prior to the time of its disclosure, or to have been independently developed by you without use of Confidential Information; and/or (iii) is subsequently learned from a third party not under a confidentiality obligation with respect to such Confidential Information. Confidential Information that is required to be disclosed by you pursuant to a duly authorized subpoena, court order, or government authority shall continue to be Confidential Information for all other purposes and you agree, prior to disclosing pursuant to a subpoena, court order, or government authority, to provide prompt written notice and assistance to Licensor prior to such disclosure, so that Licensor may seek a protective order or other appropriate remedy to protect against disclosure.

8. WARRANTY AND DISCLAIMER. Your use of the Software and its results is entirely at your own risk. Should the Software prove defective, you assume the entire cost of all service, repair or correction. The Software is made available to you "AS IS", AND LICENSOR MAKES NO REPRESENTATION OR WARRANTY REGARDING THE SOFTWARE, AND HEREBY DISCLAIMS ALL WARRANTIES, WHETHER EXPRESS, IMPLIED, OR STATUTORY, INCLUDING BUT NOT LIMITED TO ANY IMPLIED OR OTHER WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, NON-INFRINGEMENT OR NON-MISAPPROPRIATION OF INTELLECTUAL PROPERTY RIGHTS OF A THIRD PARTY, CUSTOM, TRADE, QUIET ENJOYMENT, ACCURACY OF INFORMATIONAL CONTENT, OR SYSTEM INTEGRATION. NO WARRANTY IS MADE THAT ANY SOFTWARE WILL OPERATE IN AN ERROR FREE, UNINTERRUPTED OR COMPLETELY SECURE MANNER, OR IN COMBINATION WITH THIRD PARTY HARDWARE OR SOFTWARE PRODUCTS. YOU ACKNOWLEDGE THAT LICENSOR HAS NO CONTROL OVER THE SPECIFIC CONDITIONS UNDER WHICH YOU USE THE SOFTWARE. ACCORDINGLY, EXCEPT FOR THE FOREGOING EXPRESS WARRANTY, LICENSOR CANNOT AND DOES NOT WARRANT THE PERFORMANCE OF THE SOFTWARE OR ANY PARTICULAR RESULTS THAT MAY BE OBTAINED BY THE USE OF THE SOFTWARE. THE SOFTWARE DOES NOT REPLACE YOUR OBLIGATION TO EXERCISE YOUR INDEPENDENT JUDGMENT IN USING THE SOFTWARE. The warranty made by Licensor may be voided by abuse or misuse. Licensor has no obligation to provide any support, error corrections, updates, or any other services related to the Software.

9. THIRD PARTY SOFTWARE AND DATABASES. You understand that third party products integrated into the Software or provided for use with the Software may be subject to additional terms and conditions and/or license agreements from the applicable third party vendor, which shall govern over conflicting terms of this Agreement for purposes of your relationship with the third party vendor. You agree not to use any such third party product on a stand-alone basis independent of the Software, unless you have purchased the appropriate license from the third party vendor for use of such products. NOTWITHSTANDING ANYTHING TO THE CONTRARY IN THIS AGREEMENT, ALL THIRD PARTY SOFTWARE, DATABASES AND OTHER PROGRAMS AND SOFTWARE COMPONENTS ARE PROVIDED "AS IS" WITHOUT ANY WARRANTY WHATSOEVER FROM LICENSOR. ANY DATABASES OR OTHER INFORMATION PROVIDED BY LICENSOR ARE DESIGNED TO SUPPLEMENT OTHER SOURCES OF INFORMATION, ARE NOT INTENDED TO REPLACE YOUR PROFESSIONAL DISCRETION AND JUDGMENT AND LICENSOR MAKES NO REPRESENTATIONS OR WARRANTIES REGARDING SUCH DATABASES OR INFORMATION, THEIR ACCURACY, COMPLETENESS OR OTHERWISE. Licensor agrees, upon request and as Licensor's sole liability and obligation, to have passed through to you (to the extent it may reasonably do so) any warranties and indemnifications provided by the applicable third party vendor of any third party products provided to you. To the extent any problem or liability arises from a third party product, you agree to seek recourse solely from the applicable third party vendor and not Licensor.

10. INDEMNIFICATION BY YOU. You agree to indemnify and defend Licensor, its licensors, and its affiliates, against any third party claims arising from or related to your use or misuse of the Software or any breach of the terms and conditions of this Agreement, and you agree to pay all costs, losses, damages, and attorneys' fees that a court finally awards, and all associated settlements.

11. LIMITATION OF LIABILITY. EXCEPT TO THE EXTENT PROHIBITED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR'S OR ITS AFFILIATES' TOTAL, AGGREGATE LIABILITY ARISING FROM OR RELATED TO THIS AGREEMENT OR THE SOFTWARE (INCLUDING FOR NEGLIGENCE, STRICT LIABILITY, BREACH OF CONTRACT, MISREPRESENTATION, AND OTHER CONTRACT OR TORT CLAIMS), EXCEED THE AMOUNT OF YOUR DIRECT DAMAGES ACTUALLY INCURRED, UP TO THE AMOUNT OF FEES PAID TO LICENSOR UNDER THIS AGREEMENT FOR THE SOFTWARE PRODUCT THAT IS THE SUBJECT OF THE CLAIM UNDERLYING THE DAMAGES. Some jurisdictions do not allow the exclusion or limitation of liability for certain damages, so the above limitation may not apply to you to the extent prohibited by such local laws; if so, then Licensor's liability for damages hereunder shall not exceed an amount equal to the amounts paid by you hereunder, or one hundred dollars ($100.00), whichever is less.

12. EXCLUSION OF DAMAGES. EXCEPT TO THE EXTENT PROHIBITED BY APPLICABLE LAW, UNDER NO CIRCUMSTANCES SHALL LICENSOR, ITS AFFILIATES, OR ANY OF THEIR SUPPLIERS OR LICENSORS BE LIABLE HEREUNDER FOR ANY OF THE FOLLOWING: (I) THIRD PARTY CLAIMS, (II) LOSS OR DAMAGE TO ANY SYSTEMS, RECORDS OR DATA, AND/OR (III) INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL, PUNITIVE, RELIANCE, OR COVER DAMAGES (INCLUDING WITHOUT LIMITATION FOR LOST PROFITS, LOST SAVINGS AND DAMAGE TO ANY DATA OR SYSTEMS); EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES AND EVEN IF A LIMITED REMEDY SET FORTH HEREIN FAILS OF ITS ESSENTIAL PURPOSE. YOU ARE SOLELY RESPONSIBLE AND LIABLE FOR VERIFYING THE ACCURACY AND ADEQUACY OF ANY OUTPUT FROM THE SOFTWARE, AND FOR ANY RELIANCE THEREON.

THE FEE FOR THE SOFTWARE, THE REMEDIES SET FORTH IN THIS AGREEMENT, THE LIMITS ON LIABILITY SET FORTH IN PARAGRAPHS 11 AND 12 AND THE OTHER PROVISIONS IN THIS AGREEMENT REFLECT THE ALLOCATION OF RISKS BETWEEN THE PARTIES. THIS PARAGRAPH IS AN ESSENTIAL ELEMENT OF THE BASIS OF THE BARGAIN BETWEEN THE PARTIES.

13. VERIFICATION. Licensor shall have the right to have on-site audits periodically conducted of your use of the Software. These audits will generally be conducted during regular business hours, and Licensor will use its reasonable efforts not to interfere unduly with your regular business activities. Licensor may also require you to accurately complete a self-audit questionnaire in a form Licensor may have

provided. If an audit reveals unauthorized use, you must promptly order sufficient licenses to permit all usage disclosed. If material unlicensed use is found, you also shall reimburse Licensor for all costs incurred in connection with the verification, including without limitation reasonable attorneys' fees.

14. LEGAL COMPLIANCE; RESTRICTED RIGHTS. Software is provided solely for lawful purposes and use. You shall be solely responsible for, and agree to comply with, all applicable laws, statutes, ordinances, and other governmental authority, however designated. Without limiting the foregoing, this Agreement is expressly made subject to any United States government laws, regulations, orders or other restrictions regarding export from the United States and re-export from other jurisdictions of equipment, computer hardware, software, technical data and information or derivatives of such equipment, hardware, software or technical data and information. You agree to comply with all applicable export and reexport control laws and regulations in regard to products (including computer hardware, software, deliverables, technical data, source code, or any other technology, equipment, and/or derivatives of such hardware, software, deliverables, technical data, source code, equipment, or any other technology) received from Licensor. You further certify that you will not, directly or indirectly, without obtaining prior authorization from the competent government authorities as required by those laws and regulations: (1) sell, export, reexport, transfer, divert, or disclose technical data or dispose of any product or technology received from Licensor to any prohibited person, entity, or destination; or (2) use the product or technology for any use prohibited by the laws or regulations of the United States. You will reasonably cooperate with Licensor and will provide to Licensor promptly upon request any certificates or documents as are reasonably requested to obtain approvals, consents, licenses and/or permits required for any payment or any export or import of products or services under this Agreement, at Licensor's expense. Your breach of this provision shall constitute cause for immediate termination of this Agreement. You agree to indemnify and hold harmless Licensor, its affiliates, and their respective officers, directors, employees and agents for your noncompliance with this Paragraph. Products acquired for use within or for any United States federal agency are provided with "LIMITED RIGHTS" and "RESTRICTED RIGHTS" as defined in DFARS 252.227-7013 and FAR 52.227-19.

15. GOVERNING LAW; SEVERABILITY. This Agreement shall be governed in all respects by the laws of the State of California, USA, without regard to its conflicts of law rules or principles. Any dispute arising out of or related to this Agreement shall be resolved only in the state or federal courts having subject matter jurisdiction in California. Each party hereby consents to the exclusive jurisdiction and venue of such courts. If any provision of this Agreement is held to be illegal or unenforceable for any reason, then such provision shall be deemed to be restated so as to be enforceable to the maximum extent permissible under law; the remainder of this Agreement shall remain in full force and effect.

GENERAL. This Agreement is the parties' complete agreement regarding its subject matter, superseding any prior oral or written communications, representations or agreements. You understand and agree that, to the extent Licensor permits you to use a non-Licensor purchase order or other form to order Software, Licensor does so solely for your convenience. Any terms in any such forms that purport to vary or are in addition to or inconsistent with any terms in this Agreement shall be deemed to be void and of no effect. Amendments or changes to this Agreement must be in mutually executed writings to be effective. Paragraphs 4, and 6 through 16, inclusive, shall survive the termination or expiration of this Agreement. The parties are independent contractors for all purposes under this Agreement. Neither party shall be liable for any delay or failure due to force majeure and other causes beyond its reasonable control; provided that the foregoing shall not apply to any of your payment obligations. Any notices under this Agreement to Licensor must be personally delivered or sent by certified or registered mail, return receipt requested, or by nationally recognized overnight express courier, to the address specified herein or such other address as Licensor may specify in writing. Such notices will be effective upon receipt, which may be shown by confirmation of delivery. All such notices shall be sent to the attention of General Counsel of Life Technologies Corporation (unless otherwise specified by Licensor). You may not assign or otherwise transfer this Agreement or any License without Licensor's prior written consent. This Agreement shall be binding upon and inure to the benefit of the parties' successors and permitted assigns. You agree, at Licensor's request and reasonable expense, to provide reasonable assistance and cooperation to Licensor and its designees, and to give testimony and execute documents and to take such further acts reasonably requested by the other to acquire, transfer, maintain, perfect, and enforce Licensor's intellectual property rights as described in this Agreement. To the extent you fail to do so, you appoint Licensor's or its affiliates' officers as your attorney in fact to execute documents on your (and your personnel's), successors' and assigns' behalf for this limited purpose.

Part Number MAN0001685 Rev. date 29 June 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