Agilent SureCall 4.1.1

Product Name and Version Number

Agilent SureCall v4.1.1.15

Product Number

G4980AA – SureCall Client 6-month named license supports installation of one client and server (to host the SureCall database) on one machine. For additional client only installations that connect to the same database on the central server, additional copies of this license are needed. There is no limit on the number of free-of-charge licenses provided.

The software comes with 2 installers: (1) SureCall installer (2) GenAligners (contains BWA, BWA-MEM).

Overview

SureCall is a research desktop application combining both novel and widely accepted open-source algorithms for end-to-end NGS data analysis from alignment to categorization and annotation of mutations. SureCall addresses the critical need for an easy-to-use analysis tool that incorporates the most widely accepted open source libraries and algorithms, augments them with tools specific to Agilent assays and deploys them in a convenient and user-friendly manner. SureCall provides five different types of analysis: Single, Pair, Trio, SureSelect All-In-One (AIO), and OneSeq CNV and Mutation analysis. Analysis in SureCall begins with raw reads from Illumina HiSeq/MiSeq sequencing of genomic DNA enriched with HaloPlex or SureSelect target enrichment reagents. After removal of the adapter sequences and lower-quality bases from the end of each of each read, the reads are aligned to the reference genome using BWA-MEM or BWA. Subsequently, the appropriate variant caller is selected to detect variants in a sample. The SNPPET SNP caller is an Agilent algorithm, which is optimal for detecting low-frequency single nucleotide variants with high sensitivity and specificity. SNP filtering, mutation classification and annotations are applied to the called variant list as part of the analysis workflow. VCF export is also available for any further downstream data interpretation needs.

Single Sample analysis

Run a single sample analysis when you want to find mutations, insertions or deletions (indels), and translocations in individual samples. SNPPET, an in-house algorithm developed specifically for the detection of low allele frequency variants, is used to call mutations. For samples that were target-enriched using Agilent’s HaloPlexHS or SureSelectXT HS, duplicate reads will be flagged and merged, allowing for an even more accurate detection of alleles at low frequencies. Several tools are then used to provide input for the mutation classification. Each mutation is evaluated based on its location, amino acid change, and effect on protein function (SIFT). Further information regarding the mutation is then aggregated from
various public sources, including NCBI, COSMIC (Catalog of Somatic Mutations in Cancer), PubMed, and Locus-Specific Databases. In addition, SureCall also supports variant annotation with NCBI ClinVar files with a local database source. After collecting the various inputs for classification, the proprietary mutation classifier evaluates the significance of the mutation following default or customized guidelines. Each mutation is then categorized with the user triaging each mutation and reviewing supporting evidence in the built-in viewer, including raw data and confidence measures, as well as links to external databases such as OMIM, dbVar, dbSNP, etc.

**Pair analysis**
A pair analysis can have two different applications: 1) to determine copy number changes in a test sample relative to a reference that does not have a copy number change in your region of interest, or 2) to find somatic mutations in a tumor sample by comparing it to a normal sample.

**Trio analysis**
Select trio analysis to find mutations and indels in a trio of samples, typically mother, father and child. The analysis focuses on de novo mutations, i.e., mutations that are only found in the child and mutations that are homozygous in the child but not in either parent.

**SureSelect AIO analysis**
Select AIO analysis to identify copy number variants (CNVs), point mutations, indels, and translocations in samples that were target-enriched using one of Agilent's SureSelect AIO panels. You need to provide a reference sample (either matched or unmatched) to which SureCall compares the experimental sample.

**OneSeq CNV and Mutation analysis**
A OneSeq analysis simultaneously finds CNVs, copy-neutral LOH, point mutations, and indels in a single sample. The OneSeq workflow type is only suitable for samples that were target-enriched using Agilent's OneSeq kits. Copy number changes are detected by comparing an experimental sample to a known reference sample. The in-house developed SNP calling algorithm SNPPET is used to call point mutations and indels. The high-frequency, minor allele SNPs covered by the OneSeq backbone design are used to determine copy-neutral LOH.

**New Key Features of SureCall 4.1**

- **Analysis of SureSelect All-in-One (AIO) Cancer Panels**
  With new analysis methods, expanded triage tools, and a new Gene Level Report, SureCall 4.1 supports analysis of sequencing results from samples enriched with a SureSelect AIO panel. The AIO analysis methods are capable of detecting SNPs/indels, translocations, and CNVs in a single experimental sample that is analyzed alongside a user-provided reference sample.

- **Sample Result Upload to Alissa Interpret**
  For Agilent target enrichment users who use both SureCall and Alissa Interpret for NGS analysis, SureCall 4.1 allows variant results (i.e., vcf and QC associated report files) to be uploaded from SureCall directly to Alissa Interpret for further data interpretation.

- **HG38 Support**
  SureCall now supports human genome build 38. Designs created in SureDesign using hg38 can be imported into SureCall 4.1 and used for NGS analysis.

- **MultiSample Analysis Improvements**
  Algorithms used in the Trio and Tumor-Normal Pair analysis types have been improved to better identify de novo and somatic mutations from proband and tumor samples, respectively. The CNV caller algorithms for Pair analysis methods have also been improved to effectively call CNVs down to the exon level.
- **Overall Software Functionality Improvements**
  - For HaloPlex\textsuperscript{HS} and SureSelect\textsuperscript{XT HS} data, SureCall 4.1 calculates the QC metric "Number of reads in covered regions" both before and after de-duplications.
  - SureCall 4.1 allows you to add sample notes while in Triage View.
  - When generating reports (e.g. mutation report, translocation report, etc.), SureCall 4.1 provides the option to include any Sample notes in the report.
  - For Translocation analyses, SureCall 4.1 allows you to suppress reported translocations in Triage View. The suppressed translocations are not included in the report and final vcf.

**System Requirements**

SureCall 4.1.1 is only supported on Windows operating systems (64-bit Windows 7 Enterprise, Windows 10 Enterprise and Professional, or Windows Server 2016). Additionally, only the English language versions of these operating systems are supported. If using a non-English version of Windows, switch the language to English before installing SureCall.

See the SureCall Installation Guide (publication G4890-90006) or the SureCall website (https://www.agilent.com/en/download-software-surecall) for a complete list of minimum and recommended system requirements and installation instructions.

**Workflow analysis memory settings**: In SureCall, the default memory allocated to workflow analysis is 8 GB. This memory settings may need to be increased for analysis of larger data sets (e.g., Exome, HaloPlex\textsuperscript{HS} and SureSelect\textsuperscript{XT HS}). If needed, first install additional RAM in your computer. Then, increase the memory allocation in SureCall from the Admin > Memory Management screen of the software.

**Installation Instructions**

**New installation**

**Points to note for upgrade from older version of SureCall to SureCall 4.1:**

- Upgrade to SureCall 4.1 is only supported from released versions of SureCall 3.5 or SureCall 4.0 (i.e., v3.5.1.46 or v4.0.1.46).
- SureCall versions prior to 3.5 (i.e., 2.1, 2.0, 3.0) cannot be directly upgraded to 4.1. You must first upgrade from version 1.0/1.1/2.0/3.0 to version 3.5 using the SureCall 3.5 installer. Then, you can upgrade to version 4.1 with the SureCall 4.1 installer.
- During upgrade to SureCall 4.1, the earlier version of the PostgreSQL server is upgraded, but there is not a new server installation. Hence, after upgrade, server and client installation folders are saved to different locations on disk. If you need to uninstall SureCall, you must uninstall the client and server separately.

**Upgrade instructions from earlier (v3.5 or later) SureCall version**

1. Double-click the Agilent SureCall 4.1.1.x.exe file to start the installation wizard. You will be prompted that a version of SureCall client already exists on the local machine.
2. Click OK to proceed with uninstalling the existing Agilent SureCall client. The Uninstaller of the existing SureCall installation is launched.
3. Click Next to proceed.
4. Select "Uninstall specific features", and click Next.
5. In the top panel, check the Client checkbox, and click Uninstall. **Note:** Do not remove Server as doing so will remove all previously analyzed samples from the database.
6. After client uninstallation of earlier version is complete, you are automatically returned to the v4.1 installation wizard.
7. Select "Both Client and Server" option and click Next. The installer notifies you that SureCall server already exists and will be upgraded to latest version.
8. Click OK and proceed with the installation.
9. The installer installs SureCall 4.1 client application and upgrades existing SureCall server to 4.1.
10. **Optional:** Install GenAligners v3.0 if you plan to use unaligned FASTQ files in your analysis workflows.

### Default Analysis Method changes

<table>
<thead>
<tr>
<th>Analysis type</th>
<th>Analysis method</th>
<th>SureCall v4.0</th>
<th>SureCall v4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single sample</td>
<td>Default HaloPlex method</td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td></td>
<td>Default SureSelect method</td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td></td>
<td>Default SureSelect method</td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td>Pair Sample</td>
<td>Default HaloPlex Tumor normal method</td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td></td>
<td>Default SureSelect Tumor normal method</td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td></td>
<td>Default HaloPlex Copy Number method</td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td></td>
<td>Default SureSelect Copy number method</td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td>Trio</td>
<td>Default HaloPlex and SureSelect method</td>
<td>Variants Score Threshold</td>
<td>Not Present</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimum Quality for base</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Report multiple allele at locus check box</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Variants quality Threshold</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimum Allele frequency</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimum no of read supporting Denovo Variant</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denovo variant quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denovo variant genotype quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not present</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denovo probability threshold</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fraction of variant allele in base pile up</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimum read depth</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimum no of read supporting Indel</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BAQ SNP caller (Legacy)</td>
<td>Not Present</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Version of Annotation</td>
<td>Not Present</td>
</tr>
<tr>
<td>OneSeq</td>
<td>Default OneSeq Backbone CN</td>
<td>Version of Annotation</td>
<td>Not Present</td>
</tr>
</tbody>
</table>
SureCall 4.1 result differences

Variant results in SureCall 4.1 are expected to be different as compared with earlier SureCall release build (v4.0) because of:

1. The analysis method option "Report multi-allelic variant" has been removed in SureCall 4.1. Reporting of multi-allelic variants is now performed in all analyses and cannot be removed from custom analysis methods. For analyses in previous versions of SureCall that did not include this option, results may now differ.
2. The deduplication algorithm, previously used only for HaloPlex, HaloPlexHS and SureSelectXT HS data, is now used in analyses of non-HS SureSelect data. Some minor differences with regards to QC metrics and results in non-HS SureSelect data may be observed due to the change.
3. The Translocation caller algorithm includes updates that may result in changes to translocation calls.
4. The algorithm for detecting indels includes updates that may result in changes to the reported Position and REF/ALT sequences of indel calls.
5. The Pair analysis algorithms for Tumor-Normal samples include updates that may result in different results for those analyses.
6. The Trio analysis algorithms include updates that may result in different results for those analyses.
7. In case of HaloPlex/HaloPlex HS unaligned workflow, insert size calculation is corrected. Due to this fix, aligner output will be changed and hence QC, variant results may be changing.
8. To obtain annotations for variants, SureCall 4.1 now uses an internally-maintained Knowledge Service instead of NCBI Variation Reporter. Due to this change, the application may provide different annotations for the same variant when compared to earlier versions.

Note: Starting with version 4.1, SureCall no longer supports analysis of data generated on an Ion Torrent sequencer. However, if you have aligned data from the Ion Torrent platform that was aligned in Torrent Server using the TMAP aligner tool, you can use a previous version of SureCall (version 4.0 or earlier) to analyze your aligned BAM files. Contact Agilent Technical Support (informatics_support@agilent.com) to obtain an installer for a previous version of SureCall.

Issues Fixed in SureCall 4.1

1. When SureCall attempts to contact the SureDesign website while the SureDesign website is down, SureCall needs to display an error message explaining that SureDesign is currently down. (TT#257861)
2. For multi-sample analysis types (e.g, Pair and Trio), the variants are not color-coded in the Triage View Mutation table according to zygosity (HOM/HET), as they are in a single sample analysis. (TT#259320)
3. For HaloPlexHS analysis, the analysis method parameter "Minimum number of reads supporting variant allele" does not work as expected. (TT#259971)
4. When viewing results in Triage View opened from a variant search, the option to generate a QC report of those results does not work. (TT#268764)
5. SNP calling in padded regions is inconsistent across samples in a Trio analysis. (TT#270208)
6. CNV calls in a Pair analyses are not included in the VCF. (TT#274323)
7. In analyses with HaloPlexHS or SureSelectXT HS data, the workflow fails if the Remove Duplicates parameter is turned off. (TT#274513)
8. In some cases, when setting up a workflow, the sample files selected for analysis fail to appear on all workflow setup screens. (TT#275841)
9. Analysis workflows fail if the design file denotes the chromosome X name as chrx instead for chrX. (TT#275952)
10. For some genes with multiple transcripts, the tooltips in IGV do not always display the correct transcript name. (TT#276263)
11. In some cases, the software fails to generate a Known Variant report when allele frequency data is selected for inclusion in the report. (TT#276410)

12. In some HaloPlex workflows, the software fails to sort the amplicon BED file, resulting in an error when opening Triage View. (TT#277020)

13. If the track used for annotation in an analysis contains a space in its name, then track annotation fails. (TT#277238)

14. For workflows with aligned BAM files, if the chromosome order used in the BAM file is different from the standard order, then the QC metrics are reported incorrectly. (TT#277652)

15. In some cases, analysis jobs with a large size data set fail to complete. (TT#278374)

16. If the covered.bed file of a design contains blank cells in the annotation column (column 4), then analysis jobs using that design fail. (TT#280105)

17. If a design includes regions from non-standard chromosomes, then analysis jobs using that design fail. (TT#280858)

18. In Pair analyses with HaloPlexHS data, the software may identify variants in off-target regions. (TT#282807)

19. For some genes, variants are not properly called when a track filter is used in the analysis method. (TT#283084)

20. For some detected variants, HGVS genomic annotations are missing in the VCF. (TT#283289)

Known/Open issues

1. When a FASTQ sample file is submitted for re-analysis, the resulting file name length may exceed the 255-character limit, which results in the report links in job summary not working. (TT# 283167)

2. In a multi-sample analysis (e.g., Pair or Trio) with large sample files and a large design, the workflow job may fail. (TT# 283621)

3. In a SureSelect All-In-One analysis, the workflow job fails when the same input file is used for the sample and reference. (TT# 283680)

4. When analyzing some especially large data sets, the job may fail due to memory limitations. (TT# 284436)

5. When importing a SureSelect AIO Reference file, if the file name includes any special characters, then the user is forced to rename the sample. (TT# 284485)

6. When there is no internet connection during an analysis job, the job summary report fails to include links to output files. (TT# 284724)

7. For OneSeq analysis, 'Number of reads' metric value is not getting displayed in QC metrics PDF although it is displayed for reference. (TT# 286256)

8. For translocation records with multiple mates, the VCF does not follow 4.2 specifications hence mates are shown correctly in Interpret. (TT# 286877)

9. When running an analysis with a custom analysis method in which the SNP Filter parameter is disabled, SureCall may identify fewer SNPs than it would if the SNP Filter parameter were enabled. (TT#259762)
   
   NOTE: To maximize SNP detection sensitivity, Agilent recommends only using analysis methods in which SNP filtering is enabled (i.e., the SNP Filter check box is marked in the analysis method configuration). SNP filtering is enabled in all default analysis methods that come preloaded with SureCall.