Agilent Technologies

Release Notes

Agilent SureCall 4.0

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 four different types of analysis: Single, Pair, Trio, and OneSeq CNV and Mutation analysis. Analysis in SureCall begins with raw reads from Illumina HiSeq/MiSeq sequencing, or aligned reads from Ion Torrent 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 SNP caller is selected to detect variants in a sample. The SNPPET SNP caller is an Agilent algorithm, which is optimal for detecting low-frequency 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. SAMTools or SNPPET, an in-house algorithm developed specifically for the detection of low allele frequency variants, can be used to call mutations. For samples that were target-enriched using Agilent's HaloPlex^{HS} or SureSelect^{XTHS}, 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.

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 4.0

- **SureSelect$^{XT\ HS}$ analysis with the ability to detect low frequency variants with high confidence using Molecular Barcodes**
  Agilent's in-house developed algorithm called ***LocateIt*** has been updated in SureCall 4.0 to support SureSelect$^{XT\ HS}$ samples in addition to HaloPlex^{HS} samples which were supported in the previous version of SureCall. The algorithm updates allow for the identification of duplicate reads more efficiently using molecular barcodes, hence significantly improving base calling accuracy even at low allelic frequencies compared to conventional NGS methods.
- **Alleviation of the effects of index hopping**
  In SureCall 4.0, a brand new parameter has been introduced in the analysis method to alleviate and control false positive calls due to index hopping occurrences.
- **Expanded QC metrics to support SureSelect$^{XT\ HS}$ and HaloPlex^{HS} enriched samples**
  - New QC charts for high-sensitivity (HS) samples that plot amplification level per molecular barcode within covered regions and outside covered regions.
  - New QC metrics for HS analysis to report number of reads filtered from analysis due to low quality sequencing in the barcode region.
  - New QC metric for HS analysis to report number of reads filtered during deduplication. This can be due to threshold set in analysis method, e.g., Minimum number of read pairs per barcode, Barcode quality etc.

- **Translocation detection support**
  Data generated from SureSelect^{XT} or SureSelect$^{XT\ HS}$ enriched samples can now be analyzed in SureCall 4.0 for the presence of translocation events. The brand new in-house developed algorithm reliably detects translocation events based on the split reads method. In addition, SureCall 4.0 provides a brand new table in the software's Triage View and a split-screen format in the Genome Viewer so that users can effectively visualize breakpoints along with gene annotations of each reported translocation event in a sample.
- **Re-Annotation feature**
  SureCall 4.0 provides a feature to re-annotate variants for an analyzed sample. This allows you to submit an annotation-only job. The feature can be used to fetch the latest annotations available or, in the case of a network connectivity failure during a workflow job, to complete variant annotation that may not have been successful during the first attempt.
- **Software installation improvements**
  - A new smart install feature verifies if the system meets minimum specifications for OS and hardware prior to installation to help prevent potential installation errors.
  - New security measures ensure that only users with Admin privileges can uninstall the SureCall 4.0 server.

System Requirements

SureCall 4.0.1 is only supported on Windows operating systems (64-bit Windows 7 Enterprise, Windows 10 Enterprise, or Windows Server 2008). 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 (P/N G4890-90005) or the SureCall web site www.agilent.com/genomics/surecall for a complete list of minimum and recommended system requirements and installation instructions.

Workflow analysis memory settings: In SureCall, default memory allocated to workflow analysis is 8 GB. This memory settings may need to increase for larger data analysis (such as. Exome, HaloPlex HS and SureSelect XT HS) from Admin > Memory Management tab.

Installation Instructions

New installation
Refer to instructions in http://www.agilent.com/cs/library/usermanuals/public/G4890-90005.pdf.

Points to note for upgrade from older version of SureCall to SureCall 4.0:

- Upgrade is only supported from SureCall 3.0 or SureCall 3.5 released version (v3.0 or v3.5.1.46) to SureCall 4.0.
- SureCall versions prior to 3.0 (i.e. 2.1, 2.0 and 1.0) will not be upgraded to 4.0. The user must first upgrade from version 1.0/1.1/2.0 to version 3.0 using SureCall 3.0 installer and then to version 4.0 with the SureCall 4.0 installer.

- PostgreSQL server of earlier version will be upgraded (in case of SureCall 3.0 to SureCall 4.0). There will not be a new server installation. Hence, after upgrade, server and client installation folders will be saved to different locations on disk. If an uninstallation is required, the client and server will need to be uninstalled separately.

Upgrade instructions from earlier (v3.0 or later) SureCall version

1. Double-click the Agilent SureCall 4.0.1.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, v4.0 installation will be resumed.
7. Select Both Client and Server option and click Next. The installer will prompt that SureCall server already exists and it will be upgraded to latest version.
8. Click OK and proceed with the installation.
9. The installer will install SureCall 4.0 client application and upgrade existing SureCall server to 4.0.
10. ***Optional:*** Install GenAligners v3.0 if start with unaligned FASTQ files.

Default Analysis Method changes

- New default method called 'Default SureSelect Method_TL' has translocation analysis option. Translocation analysis option is only available in SureSelect type Single Sample analysis methods. This option has a 'Minimum Number of reads' parameter with a default value of 5.
- 'Amplification level per barcode' parameter is now replaced by a new parameter called 'Minimum number of read pairs per barcode' with a default value of 2.
- For HaloPlex^{HS} and SureSelect$^{XT\ HS}$ analyses, two new parameters are part of 'Remove Duplicate' step of analysis method.
- Number of base errors allowed for merging molecular barcode, which has a default value of 1.
- Barcode base quality, which has a default value 25 for Haloplex^{HS} and 0 for SureSelect$^{XT\ HS}$
- SureSelect method now has 'Minimum number of reads supporting variant allele' with a default value of 3, which is applicable to XT HS analyses.
- New parameter named 'Remove overlap' is available in both HaloPlex & SureSelect methods and is enabled by default in SureSelect method. SureCall masks overlap when mate1 and mate2 read sequences overlap with each other. This helps to ignore redundancy for read number calculation in SureCall. Remove overlap option is applicable only for HaloPlex HS and SureSelect HS analysis.
- OneSeq analysis method has new parameter named 'Report SNPs from targeted regions only' in 'SNP Filter' section. This option is applied to filter SNPs to visualize during triage. The algorithm uses all SNPs in LOH calling.
- For SureSelect method, default value of region padding value is changed to 0 bp.

SureCall 4.0 result differences

Variant results in SureCall 4.0 are expected to be different as compared with earlier SureCall release build (v3.5) because of:

1. Changes to remove overlap when mate1 and mate2 read sequences have overlap with each other (as dictated by 'Remove Overlap' analysis option). This helps to ignore redundancy for read number calculations in SureCall. Variant calls may be changed due to this.
2. Changes in pre-processing and molecular barcode handling step, which results in changes to QC metrics.
3. Due to updates to the algorithm that identifies and merges duplicate reads, the percentage of duplicate molecules reported for a given sample is different in software versions 3.5 and 4.0.
4. For SureSelect methods, changed default value of region padding from 100 bp to 0 bp, which may result in fewer calls in SureCall 4.0.

Issues Fixed in SureCall 4.0

1. Workflow fails when non-standard chromosomes are present in BAM. (TT#263130)
2. For some variants, category calculated in SureCall 3.5 is not same as in v3.0. (TT#262565)
3. In some instances, SureCall fails to successfully generate the complete coverageInfo file. (TT#263364)
4. In the Triage View Mutation table, the Conservation Score column shows the GERP_RS score instead of the GERP_NR score from the VCF. (TT#268494)
5. Workflow jobs with single end split fastq files fail. (TT#264072)
6. In some cases, duplicate intervals of chrX are seen for SureSelect design downloaded from SureDesign. (TT#268954)
7. In some cases, HS analysis workflow completes and creates incomplete BAM file. (TT#273000)

Known/Open issues

1. For samples with a large number of variants, searching the Mutation table in Triage View is slow. (TT#218779)
2. In OneSeq analyses, sometimes the call is extended by 1 probe beyond what was visually selected as the boundaries of an aberration. (TT#244320 and 244322)
3. In Triage View, the regions below threshold do not include the last base of analyzable target region. (TT #248406)
4. Ensembl IDs have been updated in Ensembl database for new genome builds, etc., and some of the older IDs have been retired. For these retired IDs, the hyperlinks in the Mutation table point to a blank page. (TT#252572)
5. In a custom analysis method, setting the quality value cutoff to zero in the trimming parameters turns trimming off. With this setting, no trimming is performed and QC metrics corresponding to trimming are not available. However, Agilent recommends that users do perform quality trimming for low quality bases at the ends of reads. (TT#257037)
6. In certain OneSeq analyses, the CNV table in the Triage View is reporting data points not displayed in the genome viewer. (TT#257822)

7. In analysis method for Trio analysis, the parameter "Min number of reads supporting DeNovo variant" is applied to total read depth and not to reads with variant allele. (TT#258824)
8. For some Ion Torrent samples, the same deletion is reported multiple times in the Mutation table. (TT#259319)
9. The summary information for various effects is provided as a guideline in the job summary and includes count of variants with different annotations. However, the count for variants of one particular functional class is less than the actual number of variants for that class. (TT#258994)
10. In OneSeq Triage View, some of the SNPs do not have Design Type annotation, i.e., Backbone, Targeted, or Overlap. (TT#260476)
11. Translocation analysis does not call variants when BWA aligner is selected. (TT#273005)
12. Number of variant allele information is not calculated and functional for multi-sample analysis. (TT#272679, TT#266952)
13. Trio Analyses may miss calls in child sample when that call is present in parent samples. (TT#271253)
14. For translocation genome viewer, some viewing controls, e.g., zoom, do not work. (TT#271224)
15. Sometimes the link in the Triage View Mutation table to the NCBI Variation Reporter website becomes invalid and, consequently, annotation is not done with Variation Reporter. (TT#270299)
16. Program uses inconsistent formulas for calculating filtered read depth values in pair and trio analyses. (TT#262272)
17. For some variants, the zygosity assigned by the program (HOM/HET column in Mutation table) is incorrect. (TT#264022)
18. For some samples, the analysis fails to generate a complete BAM index file but the program does not fail the workflow as it should in such cases. (TT#265073)
19. Program does not call all of the known somatic variants in a pair analysis workflow that uses published sample data from a particular study. (TT#267872)
20. In certain pair analysis workflows, issues with read depth filtering in the reference sample may cause the program to miss some variant calls. (TT#272676)
21. In some analyses that include a known variant file, the inspected state assigned by the program is incorrect. (TT#264113,TT#264114,TT#264313, TT#260177)
22. Newly added analysis parameters, such as barcode quality, base errors for barcode merging, and minimum read supporting variant are not supported with Ion Torrent HaloPlex^{HS} data. (TT#273235)