



Release Notes

Agilent AGeNT v3.0.6

Product Name and Version Number

Agilent AGeNT v3.0.6

Product Part Number

G5762AA

Product Description

The Agilent Genomics NextGen Toolkit (AGeNT) is a Java-based suite of command-line software tools for pre-processing NGS data. It includes tools for adaptor trimming and molecular barcode-aware PCR duplicate identification and removal for Agilent SureSelect and HaloPlex assays. AGeNT runs on Linux, Windows, and Mac command-lines with minimal configuration, and users can easily integrate it into their custom bioinformatics pipelines.

New Key Features of AGeNT v3.0

- CReaK (Consensus Read Kit), a new molecular barcode-aware PCR duplicate tool for SureSelect XT HS and XT HS2 data that replaces LocatIt for the SureSelect assays.
- CReaK specifically addresses the large memory resource requirements in the LocatIt application and enables users to process data with substantially less memory compared to LocatIt.
- CReaK also adds a robust algorithm for handling chimeric alignments that consist of read segments that are far from each other on the same chromosome or on different chromosomes.
- CReaK reports essential statistics with clear definitions that can be traced directly to reads in the output BAM file. Statistics include: number of processed reads, number of reads passing the filter steps, number of correctly paired reads, number of read pairs flagged as duplicates in the input file, number of read pairs with at least one unmapped mate, number of chimeric read pairs, number of read pairs called as single consensus, number of read pairs called as duplex consensus, number of read pairs called as chimeric consensus, total duplicates identified, number of read pairs that failed the consensus filter, and number of read pairs called as single consensus that failed the quality filter in duplex mode.

For Research Use Only. Not for use in diagnostic procedures.

- CReaK provides additional information about the consensus mode selected (SINGLE, HYBRID, DUPLEX).
- CReaK allows filtering of the inputs based on average base quality in the read and MBC sequences, mapping quality, and single/duplex consensus read depth.
- This release also introduces new features for the AGeNT Trimmer tool, including: improved quality trimming, polyG trimming on the 3' end to support dual color Illumina sequencing chemistry, and unaligned BAM output.

For a complete list of all available options in AGeNT v3.0, refer to the README file, which is available as a download on the [AGeNT product page](#) and included in the AGeNT zip package.

For users migrating from LocatIt to CReaK, several features and parameters differ. The table below provides a detailed comparison of the usage differences between the two tools.

LocatIt/CReaK parameter comparison:

Input description	LocatIt	CReaK	Changes from LocatIt
Input BAM file	input_bam_file_name	input_bam_file_name	No change
Output BAM file name	-o <output_file_name>	-o <output_file_name>	No change
MBC sequences.	MBC_file_1..N	N/A	In XT HS and XT HS2 single consensus mode, LocatIt requires a separate MBC file (either the index 2 read FASTQ or the MBC.txt file produced by Trimmer). CReaK requires the aligned BAM to contain the MBC tags already. See the Best Practices document for examples of how to insert the MBC tags into the aligned BAM file.
Covered regions file	-b <covered.bed>	-b <covered.bed> or --bed-file <covered.bed>	In LocatIt, this option is identical to the -b option for SureSelect assays.

For Research Use Only. Not for use in diagnostic procedures.

Input description	LocatIt	CReaK	Changes from LocatIt
Covered regions file.	-l <covered.bed>	N/A	LocatIt has two redundant arguments for this option for SureSelect data. CReaK eliminates the redundant option.
Consensus mode selection (single, hybrid, duplex)	-i or -v2Duplex or -v2Hybrid or -v2Only (deprecated, same as -v2Duplex)	-c <consensus_mode> where consensus mode is one of: SINGLE, HYBRID, DUPLEX	CReaK simplifies the confusing consensus mode selection to one parameter with predefined values. For a more in depth review of the consensus modes, please see the AGeNT/CReaK README.
Barcode distance.	-d <number>	-d <number> or --MBC-mismatch <number>	No change.
Input read filtering master option.	N/A	-f or --input-read-filtering	This option is required in CReaK, even if no extra input read filtering is enabled. With no additional filtering options specified, CReaK will filter unmapped (SAM flag 0x4), secondary (SAM flag 0x100), and supplementary (SAM flag 0x800) reads. LocatIt also filters these reads. This basic input read filtering cannot be disabled.

Input description	LocatIt	CReaK	Changes from LocatIt
Interval filter option.	N/A	-fi or --interval-filter	CReaK adds an option for input read filtering based on the input BED file. Note that in the case of an input BAM with many alignments that partially overlap the target region, this may result in loss of read pairs (especially chimeric read pairs) prior to consensus calling.
MBC base quality filtering.	-q <quality_value>	-mm <quality_value> or --min-avg-MBC-qual <quality_value>	CReaK filters reads with minimum average MBC quality less than the specified value. LocatIt filters reads with any MBC base having quality less than the specified value.
Mapping quality (MAPQ filter) base quality filtering.	N/A	--mq <MAPQ_value> or --min-MAPQ <MAPQ_value>	CReaK adds filtering of input reads based on MAPQ.
Reads having any base that is lower than specified threshold will be filtered. Range is 0 to 45. Default value is 0.	-Q <quality value>	N/A--min-avg-read-qual -mr <quality value> or --min-avg-read-qual <quality value>	CReaK filters reads with minimum average read quality less than the specified value. LocatIt filters reads with any base having quality less than the specified value.
Consensus read filtering master option. Filtered reads will be flagged with SAM flag 0x200.	N/A	-F, --consensus-read-filtering	CReaK specifies that consensus reads will be filtered based on the consensus read filtering options. Consensus read filtering cannot be disabled.

Input description	LocatIt	CReaK	Changes from LocatIt
Minimum number of read pairs associated with an MBC/single consensus read pair (amplification level).	-m <number>	-MS <number> or --min-multiplicity-in-single <number>	CReaK renames the parameter, but does not change the functionality.
Minimum number of read pairs associated with duplex MBC/duplex consensus read pairs.	N/A	-MD <number> or --min-multiplicity-in-duplex <number>	CReaK adds a parameter for specifying the minimum number of read pairs that need to be present to create a duplex consensus read. The number indicates the total number of read pairs associated with the two single consensus read pairs that form the duplex consensus read pair.
Enable optical duplicate detection.	-c <number>	N/A	Specify the pixel distance between two PCR duplicates for them to be considered "sequencer" duplicates as opposed to library amplification duplicates. This value is not used in any of the LocatIt output.
Unsorted BAM/SAM output	-U	N/A	This option is not applicable to CReaK.
Sorted BAM/SAM output	-S	N/A	CReaK always produces sorted BAM output.

For Research Use Only. Not for use in diagnostic procedures.

Input description	LocatIt	CReaK	Changes from LocatIt
Allow chimeric read pairs.	-C	N/A	LocatIt filters out chimeric read pairs in non-SureSelect data types. This flag is automatically set when using "-i" mode. It is unnecessary in CReaK, which does not have to support non-SureSelect modes.
To remove/mask read1 read2 common overlap, half on each side.	-r	N/A	This option is not supported in CReaK.
Input file is BAM (default is SAM)	-IB	N/A	This option is not supported in CReaK.
Input file is SAM (default is SAM)	-IS	N/A	This option is not supported in CReaK.
Output file is BAM (default is BAM)	-OB	N/A	This option is not supported in CReaK. CReaK always outputs BAM format.
Output file is SAM (default is BAM)	-OS	N/A	This option is not supported in CReaK.
Renames the SAM tags so that downstream pipelines expecting different conventions can be used.	-P	N/A	This option is not supported in CReaK.
Keep the intermediate bam file. By default this is not applied and is only useful for debugging.	-K	N/A	This option is not supported in CReaK.

For Research Use Only. Not for use in diagnostic procedures.

Input description	LocatIt	CReaK	Changes from LocatIt
Sets the pairing cache size.	N/A	<code>-s <number></code> or <code>--cache-size <cache_size></code>	CReaK allows greater parameterization with parameters that impact memory consumption. The default value should cover most cases but may be increased if the output .stat file reveals an unreasonably large gap between the number of sam records passing input read filtering and the number of correctly-paired read pairs for MBC Consensus calling. A larger cache size will require more memory.
Remove duplicates and filtered consensus reads.	-R	<code>-r</code> or <code>--remove-dup-mode</code>	No change.
Location of temporary bam files used to store overflow of matches.	<code>-X <temp_directory></code> (to delete temp directory on exit) or <code>-t <temp_directory></code> (to keep temp directory on exit)	N/A	CReaK does not create these types of temporary files.

Input description	LocatIt	CReaK	Changes from LocatIt
Discard non-matching MBC reads as the input BAM is processed.	-I	N/A	Without this option, LocatIt stores unmatched MBC reads in memory until it finds them in the aligned BAM or exits. This option discards them instead to reduce memory load (can be used if both files are sorted in the same read-name sort order). It is not necessary in CReaK.
Enable memory efficient mode.	N/A	-e or --memory-efficient-mode	Forces single-threaded mode, which saves memory at the cost of computational time.
Displays version info	N/A	-v or --version	LocatIt will print version and help text when it is invoked with no parameters or unknown parameters
Displays help message	N/A	-h or --help	LocatIt will print version and help message if it is run with no parameters or unknown parameters. CReaK will also print the help message if it is run with no parameters or unknown parameters.

For Research Use Only. Not for use in diagnostic procedures.

Download Instructions

AGeNT is available for free download here <https://www.agilent.com/en/product/next-generation-sequencing/hybridization-based-next-generation-sequencing-ngs/ngs-software/agent-232879>.

Additional useful documents

AGeNT FAQ document: <https://www.agilent.com/cs/library/software/public/AGeNTFAQ.pdf>

AGeNT Best Practice: <https://www.agilent.com/cs/library/software/public/AGeNTBestPractices.pdf>

System Requirements for the Software

AGeNT 3.0 is compatible with following Operating Systems:

- RedHat Enterprise Linux 8.0
- Windows 11 Enterprise
- MacOS Monterey 12.5

AGeNT is compatible with Java version 11.

List of Issues Fixed in Version 3.0.4

LocatIt issues fixed by switching to CReaK:

- Invalid CIGAR strings are formed for consensus reads in some SureSelect datasets (TT#294529).
- MBC cache limit causes out of memory error (TT#294886).
- "-m" parameter in duplex mode is confusing (TT#295164). The consensus mode filtering options have been refactored and renamed.
- Partially fixed:
 - Properties file does not report percentages of reads (TT#291907). While CReaK does not report percentages, it does report values that can be used to calculate percentages.
 - Properties file does not report percentage of duplex consensus and single-strand consensus reads (TT#291908). While CReaK does not report percentages, it does report values that can be used to calculate percentages.

Trimmer issues:

- Quality trimming does not work (TT#291781).
- FASTQ files with read-names formatted in CASAVA 1.4 format (for example, "@HWUSI-EAS100R:6:73:941:1973#0/1") require "-IDEE_FIXE" parameter, which also causes a base quality conversion from Phred+64 to Phred+33. In files with this style of read-name but Phred+33 quality encoding, this behavior results in invalid quality string characters (TT#295775).
- Partially fixed:
 - Trimmer does not provide correct error if previously trimmed files are provided (TT#291909). This issue has been resolved for SureSelect XT HS and XT HS2 data, for which Trimmer is able to detect that trimming has already based on read name format.

For Research Use Only. Not for use in diagnostic procedures.

List of Issues Fixed in Version 3.0.5/3.0.6

Trimmer issues:

- The adaptor sequence for the "secondMateOppositeAdaptor" for Illumina is incorrect for dual barcode systems (XT LI, XT HS, XT HS2) (#297067).
- The mateOverlap parameter is hard-coded to "50", which does not work for insert sizes shorter than 50bp (#297068).

Documentation issue:

- Agent Documentation - Customers should be able to copy paste commands from manuals such as FAQ and software release notes (#297069).

List of Known Issues

LocatIt issues:

- BAM files containing empty reads (all N's) and/or empty MBC's fail with LocatIt (TT#288363).
- Properties file does not report percentages of reads (TT#291907). Partially fixed by switching from LocatIt to CReaK; see above.
- Properties file does not report percentage of duplex consensus and single-strand consensus reads (TT#291908). Partially fixed by switching from LocatIt to CReaK; see above.
- "-m" parameter in duplex mode is confusing (TT#295164). Fixed by switching from LocatIt to CReaK; see above.
- Mark duplicate mode changes secondary alignment locations in the output BAM (TT#295969).
- Invalid CIGAR strings are formed for consensus reads in some SureSelect datasets (TT#294529). Fixed by switching from LocatIt to CReaK; see above.
- MBC cache limit causes out of memory error (TT#294886). Fixed by switching from LocatIt to CReaK; see above.

AGeNT issues:

- AGeNT "-s" option does not work on RedHat Linux (TT#291765).

Trimmer issues:

- Invalid parameters do not show correct error message (TT#291905).
- Trimmer does not provide correct error if previously trimmed files are provided (TT#291909). Partially fixed in AGeNT 3.0; see above.
- Providing incorrect FASTQ file to AGeNT doesn't result in correct error message (TT#291910).
- Trimmer fails when input files contain spaces (TT#291914).

PR7000-3529