

CNV Calculator

Technical Guide

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



Notices

Manual Part Number

5993-0097

Edition A0, September 2018

Copyright

© Agilent Technologies, Inc. 2018

No part of this manual may be reproduced in any form or by any means (including electronic storage and retrieval or translation into a foreign language) without prior agreement and written consent from Agilent Technologies, Inc. as governed by United States and international copyright laws.

Manufacturer



Agilent Technologies Belgium S.A./N.V.
Galileilaan 18
B-2845 Niel
Belgium

Contact

Europe:

genomics_tech_europe@agilent.com
00800 345 34 600 (toll free) or
+49 69 8679 7730

US-Canada:

ngs.support@agilent.com
800-227-9770

Rest of World:

<https://www.agilent.com/en/contact-us/page>

Safety Notices

CAUTION

A **CAUTION** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in damage to the product or loss of important data. Do not proceed beyond a **CAUTION** notice until the indicated conditions are fully understood and met.

WARNING

A **WARNING** notice denotes a hazard. It calls attention to an operating procedure, practice, or the like that, if not correctly performed or adhered to, could result in personal injury or death. Do not proceed beyond a **WARNING** notice until the indicated conditions are fully understood and met.

Contents

1	Abbreviations	5
2	Product Description	6
3	Research Workflow	7
4	CNV Analysis Pipeline	9
	4.1 Input Information	9
	4.2 Automated Analysis Pipeline	11
	4.3 Output Generation	11
5	Version History	14

List of Figures

Figure 1: Automated Analysis pipeline: Terms of Use Pop Up	11
Figure 2: Example of Amplicon for Rejection	12
Figure 3: Example of Input Tab	12
Figure 4: Example of Dosage Quotient Tab	12
Figure 5: Example of CNV Results Tab	13
Figure 6: Download Option	13

1

Abbreviations

CNV	copy number variation
DQ	dosage quotient
IEM	Illumina Experiment Manager
IFU	instructions for use
MASTR	multiplex amplification of specific targets for resequencing
NGS	next-generation-sequencing
RC	read count
RUO	For Research Use Only. Not for use in diagnostic procedures.
SNV	single-nucleotide-variation

2 Product Description

The CNV Calculator is a client-side web application developed to identify or confirm the presence or absence of copy number variations (CNVs) in SureMASTR assay-derived next-generation sequencing (NGS) data. The Calculator is provided for Illumina MiSeq system users only. The automated analysis process facilitates assessment of the presence of CNVs in germline SureMASTR assays. Individual sample results are provided in an online CNV results output. This application replaces the existing CNV excel instructions for use (IFU) files that were available upon request.

3 Research Workflow

Web-based CNV analysis is available for a selected number of SureMASTR assays. These SureMASTR assays contain either (1) a separate set of control amplicons per plex (located on chromosomes different from the target genes) or (2) multiple amplicons on different chromosomes per plex whereby the target amplicons on different chromosomes within the plex are used as reference amplicons for CNV analysis. Importantly for CNV detection, all plexes in all samples should be amplified under exactly the same conditions (i.e. same master reaction mix and same PCR run) during the Multiplex and the Universal PCR reactions. CNV analysis should be performed on samples within the same sequencing run.

Read count comparison between target and control (or reference) amplicons is performed to calculate the dosage quotient (DQ) for amplicon *i* as described:

$$DQ_i = \frac{\text{Normalized RC target sample}}{\text{Average normalized RC all samples}}$$

Where Normalized RC:

$$\text{Normalized RC} = \frac{RC_i}{RC \text{ control amplicons}}$$

In case of separate control amplicons (1):

- The “normalized read count” per sample is calculated per plex as the read count of the amplicon of interest (*RC_i*) divided by the sum of read counts of all control amplicons in the respective plex.
- The average of the “normalized read counts” for all samples is calculated.
- The DQ is calculated as the “normalized read count” divided by the average of the normalized read counts.

In case of target amplicons on different chromosomes (2):

- The “normalized read count” per sample is calculated per plex as the read count of the amplicon of interest (*RC_i*) divided by the sum of read counts of the control (or reference) amplicons on different chromosomes of the respective plex.
- The average of the “normalized read counts” for all samples is calculated.
- The DQ is calculated as the “normalized read count” divided by the average of the normalized read counts.

A $DQ \geq 1.3$ indicates that the corresponding genomic fragment is considered present in 3 copies (duplication of one allele); a $DQ \leq 0.7$ indicates the genomic fragment is considered to be present in only 1 copy (deletion of one allele). When two or more subsequent amplicons are marked as a deletion or a duplication, this is a strong indication of a CNV and these will be listed as CNV results.

NOTE

CNV analysis calculations are always made “per plex”. When amplifying many samples or for SureMASTR assays with many plexes, try to combine as many samples per plex within one PCRrun.

NOTE

For the correct calculation of the “average normalized read count” (in the calculation of the DQ as described above), the set of samples should meet the following requirements:

- Minimum 10 samples should be included per CNV analysis to guarantee sufficient reference samples.
- When using a set of reference samples known to have no CNV, the libraries of these samples should be constructed together with the samples with unknown CNV status.
- Identical CNVs cannot be present in more than 15 % of the samples of a sequencing run.

NOTE

Since polymorphisms in primer sites may lead to amplification of only one of the alleles, resulting in a false positive $DQ \leq 0.5$, a detected CNV is only considered to be valid when 2 adjacent amplicons show a significantly altered DQ and/or when confirmed by an independent method.

NOTE

Compared to single nucleotide variation (SNV) analysis, deeper sequencing is required for CNV analysis. The minimal coverage recommended for Agilent’s SureMASTR assays for CNV analysis in combination with an Illumina MiSeq system is 200 read pairs per amplicon. Amplicons that do not reach this minimal coverage will be marked in the CNV Calculator. For the SureMASTR PED, CNV analysis can only be performed in combination with Illumina MiSeq Reagent Kit v3.

4 CNV Analysis Pipeline

The CNV analysis pipeline consists of the following steps:

1. Input information
 - Generate read count information for all samples
 - Assemble all other necessary information
2. Automated analysis pipeline
 - User upload and consistency check of the data
 - Filtering step 1 based on initial coverage data
 - Normalization of coverage data
 - Filtering step 2 based on normalized coverages and preliminary CNVs
 - CNV detection based on normalized coverage data
3. Output generation
 - Result tables can be downloaded as tab-delimited files

4.1 Input Information

The user input consists of coverage data for a single sequencing run and selection of a single SureMASTR assay. This includes an AmpliconCoverage.tsv input file derived from the MiSeq Reporter v2.5 or higher versions and the name of the SureMASTR assay.

To generate the input files using the MiSeq Reporter v2.5 or higher versions software, an adapted CustomAmplicon Manifest file is necessary. The Amplicon Manifest files for these SureMASTR assays are available on request by contact customer service.

Create a Sample Sheet with the Illumina Experiment Manager (IEM) software by following the guidelines below:

NOTE Use the latest version of the IEM software

- On a Windows PC extract the files from '5993-0006_A1_SureMASTR SamplePrepKit for IEM or MiSeq'.
- Save the files at the following directories:
 - 'GenerateFASTQ.txt' and 'TruSeqAmplicon.txt' at ...\\Illumina\\Illumina Experiment Manager\\Applications; overwrite the existing files.
 - SureMASTR MID 1-192.txt' at ...\\Illumina\\Illumina Experiment Manager\\SamplePrepKits.
- Open the IEM software.
- Create the sample sheet.
- For the Instrument Selection, choose '**MiSeq**', click '**Next**'.
- For the MiSeq Application Selection, choose '**Targeted Resequencing**' (Category) and '**TruSeq Amplicon**' (Application). Click '**Next**'.
- Fill out the Workflow Parameters, with these specific settings:

- Reagent Cartridge Barcode
- Library Prep Kit: select '**SureMASTR MID 1-192**'
- Index Reads: select '**2**'
- Read Type: select '**Paired End**'
- Cycles Read 1: specified in the Assay Specification sheet, Section 6
- Cycles Read 2: specified in the Assay Specification sheet, Section 6
- Select the 3 '**Custom Primers**' (Read 1, Index Read and Read 2).
- Select both '**Use Adapter Trimming**' boxes. Click '**Next**'.
- Add rows to the Sample Sheet. Select Indexes from the list. Select the correct Amplicon Manifest file. Click '**Finish**' and save the Sample Sheet.

When the MiSeq run is finished, an AmpliconCoverage.tsv file per Amplicon Manifest file is generated in the \Data\Intensities\BaseCalls\Alignment folder of the run, i.e. all samples attributed to the same Amplicon Manifest file will be grouped in the same AmpliconCoverage.tsv file. When combining different SureMASTR assays into one sequencing run, select multiple Amplicon Manifest files and indicate the correct Amplicon Manifest file per sample to allow different CNV analyses. All different SureMASTR assays within one run require the use of an adapted Amplicon Manifest file, even when CNV analysis will be restricted to part of the samples. Note that in case the Amplicon Manifest file of the wrong assay was attributed to samples, this will result in no read counts for the amplicons and will lead to warnings (shown by the exclamation marks) in the CNV Calculator analysis. CNV analyses in the CNV Calculator are restricted per SureMASTR assay.

To requeue sequencing experiments previously run with the Manifest workflow with the Amplicon Manifest workflow for CNV analysis in the MiSeq Reporter v2.5 software (or higher), follow the guidelines below:

- Copy the Manifest file into the correct run folder.
- Replace the sample sheet in the run folder by the newly generated sample sheet (steps 1-7, as described above).
- For the MiSeq Application Selection, either select the "requeue" checkbox of the run in the analyses view and select "requeue" or go to the "samplesheet" tab of the run and select "save and requeue".
- Upon completion of the run, the readcount files can be found in the Alignment folder of the run. Per Amplicon Manifest file, one AmpliconCoverage_M1.tsv file is made.

Note upon requeueing experiments, FASTQ files are overwritten. When requeueing is performed on a selected set of samples, FASTQ files of other samples disappear. It is therefore recommended to requeue complete sequencing experiments.

The web-based CNV Calculator is available for a set of assays listed under the Assay drop down menu. Contact customer service if you are interested in performing CNV analysis for other Germline SureMASTR assays.

4.2 Automated Analysis Pipeline

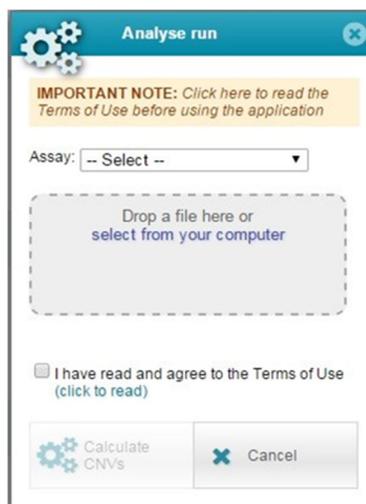


Figure 1: Automated Analysis pipeline: Terms of Use Pop Up

First select the assay (Click “**Select**”), import the “.tsv” files (use the button “**select from your computer**”), and read the “**Terms of Use**”, that specify:

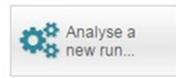
- Agilent does not retain any data submitted through the CNV Calculator.
- Customer shall use CNV Calculator with data obtained by following the procedure as described in this Technical Guide and relevant assays protocols or instructions.
- The CNV Calculator is provided a courtesy to Customer. *“To the extent permitted by applicable law, the CNV calculator is provided “as is” without warranties or conditions of any kind, whether oral or written, expressed or implied. Agilent disclaims any implied warranties or conditions of merchantability, satisfactory quality, non-infringement and fitness for a particular purpose.”*

Analyses can then be automatically started by clicking “**Calculate CNVs**”. A first analysis includes filter steps on three different levels (sample, amplicon and sample data within a single plex). The set of non-rejected amplicons and samples is subsequently used in the read count normalization. After normalization, a second round of filtering is applied. Finally, DQ values and CNVs are determined using the non-rejected amplicons.

4.3 Output Generation

The CNV Calculator application presents three tabs with information on the input and automated analyses. Each tab can be downloaded to a separate text file.

The first tab (“**Input**”, Figure 2) contains a copy of the MiSeq reporter data (AmpliconCoverage.tsv file) whereby amplicons are sorted according to their genomic position. Rejected amplicons in the calculation are displayed in grey. Rejection reasons are marked with a red exclamation mark and visible upon hovering (Figure 3).



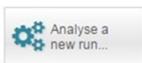
Input	Dosage Quotient																	CNV results																
BARD1_ex06_01	240	291	411	346	257	319	434	246	216	294	328	366	250	394	372	341	93	728																
BARD1_ex05_01	292	347	397	319	231	380	431	235	255	254	319	388	305	327	387	397	111	763																
BARD1_ex04_01	352	767	705	802	365	791	782	824	491	642	1039	646	634	974	776	837	572	775																
BARD1_ex03_05	857	939	1265	1249	872	1015	1330	1297	779	908	1248	1462	876	1257	1274	1454	811	639																
BARD1_ex03_04	828	1069	1171	983	783	1082	1355	833	811	1030	1110	1261	899	1257	1221	1114	372	1739																
BARD1_ex03_03	365	524	728	673	306	663	733	652	402	578	711	909	506	717	696	761	385	380																
BARD1_ex03_02	1173	1562	1755	1476	1119	1552	1987	1181	1047	1441	1593	1755	1266	1626	1722	1671	528	2398																
BARD1_ex03_01	1392	1846	2118	1789	1464	1838	2466	1583	1408	1427	1993	2253	1646	1935	2143	2093	898	2763																
BARD1_ex02_01	297	518	447	459	311	461	409	599	335	375	533	350	411	480	416	551	385	431																
BARD1_ex01_01	526	1226	747	982	699	958	1149	1226	567	660	1246	871	803	1155	1128	1400	827	1076																

Figure 2: Example of Input Tab

	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6	Sample7	Sample8	Sample9	Sample10	Sample11	Sample12	Sample13	Sample14	Sample15	Sample16	Sample17	Sample18	
MUTYH_ex16_01	1935	2793	2981	3249	2064	2816	3116	3129	2237	2738	3476	4197	2845	4305	2832	3495	1837	1690	
MUTYH_ex15_01	2099	4114	3515	3758	1999	3353	4184	4046	2308	2747	4075	2780	2832	3589	3689	4242	2218	3372	
MUTYH_ex13-14_02	385	501	481	636	435	540	718	695	554	483	1071	596	433	851	747	616	385	459	
MUTYH_ex13-14_01	Unstable amplicon (part of many single CNV events)									230	166	408	220	165	333	340	256	155	156

Figure 3: Example of Amplicon for Rejection

The second tab (“Dosage Quotient”, 4) contains the DQ values and shows the chromosome and gene names for all amplicons. The background color of the normalized coverage depends on the DQ value. All DQ values ≥ 1.3 are highlighted in blue while all DQ values ≤ 0.7 are highlighted in red. A CNV is reported when two or more subsequent amplicons are marked as a deletion or a duplication. All CNVs are highlighted (red bold font and border, Figure 4).



Input	Dosage Quotient																	CNV results																
BARD1_ex06_01	chr2	BARD1	0.92	0.84	1.09	1.02	1.02	0.96	1.02	0.88	0.86	1.08	1.02	1.09	1.00	1.15	0.99	0.88	0.76	1.41														
BARD1_ex05_01	chr2	BARD1	1.07	0.96	1.01	0.90	0.88	1.10	0.97	0.81	0.98	0.90	0.95	1.11	1.17	0.92	0.99	0.98	0.87	1.42														
BARD1_ex04_01	chr2	BARD1	0.79	0.95	1.03	1.02	0.78	1.07	0.91	0.93	0.96	1.09	1.06	1.02	1.07	1.19	1.00	0.98	1.05	1.10														
BARD1_ex03_05	chr2	BARD1	1.15	0.98	0.98	0.99	1.09	0.90	1.02	0.98	0.97	0.96	1.04	0.96	0.92	0.94	1.02	1.00	1.07	1.03														
BARD1_ex03_04	chr2	BARD1	0.97	0.94	0.95	0.89	0.95	1.00	0.97	0.91	0.99	1.16	1.05	1.15	1.10	1.12	0.99	0.88	0.93	1.03														
BARD1_ex03_03	chr2	BARD1	0.91	1.01	1.05	0.98	0.70	1.09	1.04	0.91	0.93	1.13	1.10	1.10	0.98	0.99	1.03	0.96	0.94	1.13														
BARD1_ex03_02	chr2	BARD1	0.97	0.97	1.01	0.94	0.96	1.01	1.01	0.91	0.91	1.14	1.07	1.13	1.09	1.02	0.99	0.93	0.93	1.01														
BARD1_ex03_01	chr2	BARD1	0.92	0.92	0.97	0.91	1.01	0.96	1.00	0.98	0.97	0.91	1.07	1.16	1.13	0.97	0.98	0.93	1.27	0.93														

Figure 4: Example of Dosage Quotient Tab

The third tab (“CNV results”, 5) contains a list of all CNVs found by the CNV Calculator, ordered per sample identifier. CNVs are defined when two or more subsequent amplicons are marked as a deletion or duplication. DQ values are given for the deletions or duplications.



Figure 5: Example of CNV Results Tab



Figure 6: Download Option

The CNV Calculator is available on: https://suremastrapps.agilent.com/cnv_calc/

5 Version History

Version	Description
A0	Initial version. Content is based on previous IFU515 v170867.

www.agilent.com

© Agilent Technologies, Inc. 2018

Edition A0

5993-0097

