Coffalyser digitalMLPA™ User Manual

For versions:
211020.1245
220221.1522
221020.1234
240719.0943
241105.1259
251008.1227

Instructions For Use

Coffalyser digitalMLPA™ is for Research Use Only (RUO).
Not for use in diagnostic procedures.

General information
Coffalyser digitalMLPA™ is Research Use Only (RUO) software designed for the analysis of digitalMLPA™ data generated using a NXtec™ probemix* as described in the digitalMLPA NXtec Protocol.
* To be used in combination with a NXtec Reagent Kit and one or multiple NXtec barcode plates.

Latest version of this document can be found on www.mrcholland.com.

Manufacturer: MRC Holland BV, Willem Schoutenstraat 1, 1057 DL Amsterdam, The Netherlands
Website: www.mrcholland.com; Phone: +31 888 657 200
E-mail: info@mrcholland.com (information & technical questions); order@mrcholland.com (for orders).

TABLE OF CONTENTS

IMPORTANT NOTIFICATION

Probe-specific information concerning certain genetic characteristics (e.g. whether a probe is located in a promoter region) and experimental sensitivity (e.g. sensitivity to salt) are not included in the Coffalyser digitalMLPA reports. When interpreting the data, it is therefore essential to consult the relevant product files provided by MRC Holland: the Product Description, and the Probe Information File (and other product-specific files, if available).

Coffalyser digitalMLPA has an expiration date. This date is displayed at the bottom of the initial Coffalyser digitalMLPA window when opened. A new version will be made available by MRC Holland before the expiration date has passed. Changes between versions are described in the Coffalyser digitalMLPA Release Notes, present in the "_Documentation" folder in the downloaded software package and via https://www.mrcholland.com/r/coffalyser-digitalmlpa-release-notes on the MRC Holland support portal.

In case of alerts, warnings and important information on updates that should be implemented immediately, MRC Holland will contact you via email. It is therefore essential to keep the contact information in your MRC Holland account up-to-date (www.mrcholland.com).

Software version is displayed in the Coffalyser digitalMLPA Release Notes present in the "_Documentation" folder of the downloaded software package. The latest version of this manual is available for download on www.mrcholland.com. The manual is also included in the "_Documentation" folder in the downloaded software package. Please ensure you always use the latest version of the manual corresponding with your software version.

Please note that all data used and generated by Coffalyser digitalMLPA, like FASTQ files, PDF and Excel reports, Coffalyser Definition Files and Coffa files, may contain sensitive information. It is the responsibility of the user to ensure data is stored in secure locations to prevent unauthorized persons access to these data.

1. GLOSSARY

Barcode	A barcode plate is a microtiter plate with 96 different barcode solutions. A barcode solution is a NXtec reagent that is added in the first step of digitalMLPA and that should be unique for each sample within a sequencing run. Each barcode solution contains an oligonucleotide (barcode oligo) that is incorporated in each digitalMLPA amplicon in order to distinguish digitalMLPA data from other sequencing data and to assign amplicons to the corresponding sample. The unique sequence part of the barcode oligo is called the barcode.
Coffalyser Definition File	A file (*.cdf) containing the information required for the analysis of an experiment with Coffalyser digitalMLPA. This file specifies analysis settings, file locations, Product Sheets and other data such as sample names, sample types and gender.
Coffa file	A file (*.coffa) containing the sequencing output and several analysis properties of a single digitalMLPA reaction. It is generated from a FASTQ file by Coffalyser digitalMLPA.
FASTQ	A file containing the sequencing data obtained on an Illumina® Next Generation Sequencer.
Intra ratio	Ratio obtained after the first step of normalisation (intra-normalisation) in which read counts are converted to relative values by normalising target probe read counts against the read counts of reference probes within a sample. See APPENDIX I – NORMALISATION for more information.
Inter ratio	Ratio obtained after the second step of normalisation (inter-normalisation) in which relative probe values of each sample are compared to those of the reference samples. See APPENDIX I – NORMALISATION for more information.
No-DNA sample	A sample from a no-DNA control reaction.
Pooled DNA source	A sample that consists of a mixture of DNA samples.
Positive sample	A sample with a known mutation and/or copy number variation.
Product Sheet	A file in the Coffalyser digitalMLPA package that contains information for the software on how to analyse the data obtained using a specific NXtec probemix lot.
Reference probe	A probe that detects a sequence that is expected to have a normal copy number in (almost) all samples. Reference probes are used in the first part of data normalisation to normalise other probes against.
Reference sample	A sample from a healthy individual that is expected to have a normal copy number in the regions targeted by all probes. Reference samples are used in the second part of data normalisation to normalise other samples against.
Sample Track Probe	NOT CURRENTLY AVAILABLE. An oligo that is added to a sample before or during DNA extraction. The resulting probe is recognised by Coffalyser digitalMLPA thereby aiding in sample identification and detection of potential sample swaps.
SD	Sample DNA (SD) is a DNA sample with a specific genotype provided by MRC Holland for a specific application, e.g. an artificial mutation-positive DNA sample.

SNP code	Most NXtec probemixes contain a set of SNP-specific probes that generate a SNP code, also referred to as a *sample identifier* since DNA samples from the same individual will have the same identifier. This SNP code consists of 39 characters each corresponding to one specific SNP that are always in the same order. Every position can have four different options: 0, 1, 2 or ?, where 0 and 2 are SNPs that are homozygous for allele A or B respectively, while 1 represents a heterozygous SNP. If one of these options cannot be determined, this results in a “?”. The presence of multiple "?" signs can indicate sample DNA contamination. This will be mentioned in the Quality table. Please note that a SNP code is not listed, and the associated contamination check is not executed when a sample is defined as "pooled DNA source”. Furthermore, for some probemixes the contamination check is not performed, as detection of sample DNA contamination is not always possible, e.g. in tumour derived samples. Please consult the probemix-specific Product Description for more information. The SNP code ends with a version number, e.g. v1 for version 1. More information on these SNP probes is described in the digitalMLPA NXtec Protocol.
Test sample	A sample being investigated.

2. BASIC CONCEPTS OF COFFALYSER digitalMLPA

digitalMLPA products are intended to be used with Illumina sequencers. These sequencers generate one or more large FASTQ files.

The digitalMLPA data analysis flow consists of the following steps:

- **Data conversion:** Reads from a FASTQ file that are recognised as digitalMLPA reads are counted and assigned to the digitalMLPA probes. These reads are saved per digitalMLPA reaction in Coffa files (*.coffa, one file per NXtec barcode). These files are typically less than 1 MB in size.
- **Analysis:** The obtained digitalMLPA probe read counts are normalised, see APPENDIX I – NORMALISATION for details, and evaluated for quality and the presence of aberrant ratios.
- **Reporting:** Reports are generated per experiment (Overview.pdf and Ratios.xlsx) and per sample (General Details.pdf, Sample Results 1.pdf and/or No Dna.pdf).

Coffalyser digitalMLPA requires a Coffalyser Definition File (*.cdf) to perform data analysis. This file specifies analysis settings, file locations, Product Sheets and other data such as sample names, sample types and gender. This file will be created when analysing your data using MRC.DataAnalysis.Client.exe. In addition, Coffalyser digitalMLPA contains a Definition File Editor (MRC.DataAnalysis.Definitions.Editor.exe) for easy creation and editing of Coffalyser Definition Files in case you want to prepare a file but not yet perform the analysis. A standard text editor such as Notepad can also be used to edit Coffalyser Definition Files.

It is possible to generate Coffa files as a separate step via a command line interface. An NGS service provider who pools Illumina sequencer experiments from several customers in a single Illumina run can perform the data conversion to Coffa files and provide each customer with the Coffa files of their experiment. Coffa files contain all relevant information for data analysis. Each customer can subsequently run the analysis and reporting steps on their own computer with Coffalyser digitalMLPA. For detailed information on command line interface, see APPENDIX IV – COMMAND LINE INTERFACE (CLI).

2.1 SYSTEM REQUIREMENTS

Coffalyser digitalMLPA does not require installation. For successful operation of the software, the following requirements should be met:

- A genuine copy of Windows 10 (32- or 64-bit) with the latest updates installed.
- .NET Framework version 4.8.
- An Intel i5 processor or equivalent.
- 4 GB of available RAM.
- Sufficient free disk space with a minimum of the size of the FASTQ files to be processed.

For optimal operation, we recommend:

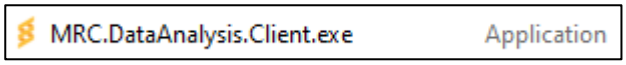
- An Intel i7 processor or equivalent.
- 8 GB of available RAM.
- A solid state drive (SSD) with sufficient free disk space.

Note: (1) The use of a USB stick to store files may greatly reduce the speed of the analysis. (2) An internet connection is not required to run Coffalyser digitalMLPA. (3) Analysis of very large FASTQ files may take a considerable amount of time. This time can be shortened by using a more powerful computer.

3. GETTING STARTED

3.1 OBTAIN A COPY OF COFFALYSER digitalMLPA

1. Log in to your MRC Holland account on www.mrcholland.com.
2. Navigate to software in the account menu and select Coffalyser digitalMLPA.
3. Accept the End User License Agreement (EULA) and download the Coffalyser digitalMLPA – software package.
4. Save the package on the desired location.
5. Unzip the file.
6. Optional: Create a shortcut for the MRC.DataAnalysis.Client.exe and place it on the desktop.



Notes:

- You may unpack and run Coffalyser digitalMLPA from a local hard-disk location or from a network location. A network location allows multiple users to use the same version, and central version management.
- Please note that network security settings or virus scanners may prevent Coffalyser digitalMLPA from running from network locations.

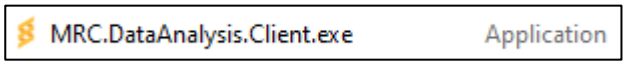
3.2 ANALYSE YOUR DATA

The time required for analysis of FASTQ files is highly dependent on the location of the FASTQ file and the location chosen for the generated Coffa files and the temporary files.

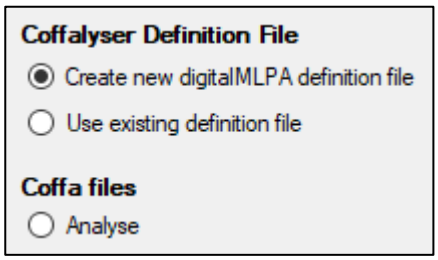
Coffalyser digitalMLPA can perform an analysis using zipped FASTQ files. Unzipping FASTQ files before analysis or copying a FASTQ file to a local disk is not necessary.

3.2.1 CREATE A NEW EXPERIMENT

1. Double click on MRC.DataAnalysis.Client.exe or the shortcut you created to start Coffalyser digitalMLPA.



2. Under Coffalyser Definition File select *Create new digitalMLPA definition file* and click **Next >**.



3. Navigate to the location where you want to save the Coffalyser Definition File (*.cdf).

4. Enter a name in the field *File name:* and click **Save**.

   The Coffalyser - Data Conversion - digitalMLPA - General Settings window will open.

5. Leave the analysis settings on *Default profile*.

6. Select the appropriate barcode plate lot collection from the drop-down menu.

   The selected barcode plate lot collection should encompass all barcode plate lots that were used in your experiment. You can find the lot number on the label of the barcode plate.

   LOT 03-012-250514.

   Demultiplexed barcode collections should only be selected when sequencing reads are generated using a custom read 1 primer resulting in demultiplexed FASTQ files. Email info@mrcholland.com for more information.

7. Click **Configure** to set up your experiment.[a]

   The Coffalyser - Samples window will open.

8. Select all barcodes and/or well positions on the barcode plate that were used in your experiment.

9. Right click on any of the selected barcodes and select *Create New Experiment (Manual)*.

   The Coffalyser - Experiment Overview window will open.

10. Enter the experiment name in the designated text field.

11. Select the experiment type from the drop-down menu.

12. Select the NXtec probemix used from the *product* drop-down menu. In case the used probemix is not available in the drop-down menu, please ensure you have downloaded the latest version of the Coffalyser digitalMLPA package from the website.

[a] In case an experiment is not (completely) configured, Coffalyser digitalMLPA will run using the default settings. Please be aware that these may not be correct for your experiment! MRC Holland recommends to always complete the experiment details to increase the likelihood of detecting mistakes.

13. Select the lot number of the NXtec probemix used from the *sheet* drop-down menu. In case the used probemix lot is not available in the drop-down menu, please ensure you have downloaded the latest version of the Coffalyser digitalMLPA package from the website.

14. Enter the name for each sample in the column *sample* or leave it as the barcode ID.

    You can copy a column of sample names from another file. Right click on the first row of the range of samples and select *Paste Sample Details From Clipboard*.
    Note: to avoid layout problems in the Sample Results 1.pdf report, please use short names for reference samples.

15. Select the sample type for each sample from the drop-down menu in the column *type*.

    You can also do this batch-wise: select the desired samples and type “n” for No DNA, “p” for Positive sample, “r” for Reference, “s” for SD, “t” for Test or “u” for Undefined.

    The various sample type options are explained in the GLOSSARY. If sample type is left as *Undefined*, it will be analysed as a Test sample. In case no samples are indicated as *Reference*, all samples with type *Test* and *Undefined* will be used for normalisation (for details see APPENDIX I – NORMALISATION). Only use the sample type *Reference* for dedicated reference samples. For correct experiment setup, please check the digitalMLPA NXtec Protocol and the probemix-specific Product Description.

16. For samples that consist of a mixture of DNA samples, select *Pooled DNA source* from the drop-down menu in the column *options*. For samples from a single source, leave this on *default*.

17. Optional: select the gender of each sample from the drop-down menu in the column *gender*.

    You can also do this batch-wise: select the desired samples and type “f” for Female or “m” for Male.

IF YOU ADDED SAMPLE TRACK PROBES[b] TO YOUR SAMPLES, CONTINUE WITH STEP 18. OTHERWISE PROCEED WITH STEP 21.

18. Click in the column *tracker(s)*.

    The Coffalyser - Select Sample Track Probes window will open.

19. Select the Sample Track Probe(s) used and click **OK**.

20. Repeat steps 18 and 19 for other samples, if applicable.

21. Click **OK** to save and exit the *Coffalyser – Experiment Overview* window.

TO SET UP ANOTHER EXPERIMENT BEFORE STARTING THE ANALYSIS, FOLLOW THE PROCEDURE FROM STEP 8 ONWARDS.

22. Click **OK** to save the experiment(s) and close the *Coffalyser – Samples* window.

23. Select the handling of the analysis of the samples not included in the definition of the experiment(s), i.e. samples not defined in the steps above, by choosing one of the options from the *undefined samples* drop-down menu. By selecting *Detect and include* also reads from samples (barcodes) that are not defined will be analysed and output will be collected in an Undefined Experiment folder, see step 30. By selecting *Ignore and exclude* only the defined samples will be analysed.

24. To select the output directory click the upper **Select** button.

25. Browse to the desired folder, select it and click **OK**.

    For each run, Coffalyser digitalMLPA will create an output sub-directory in your selected location, named using the date and time of the analysis. For example: “C:\Coffalyser digitalMLPA Analysis Run01\20200322_110143”.

[b] Sample Track Probes are currently not available.

Please note that the parent folder should not contain more than 100 subfolders.

26. Optional: click the lower **Select** button to change the temp directory.

    The temp directory is by default the same as the output directory selected in the previous step.

    Note: the speed of the analysis is influenced by the location of the temp directory. Preferably, the temp directory should be located on a local disk. A network drive is not recommended for this purpose.

27. To select the FASTQ file(s) to be analysed, click **Add File**.

28. Browse to the location of the FASTQ files, select it and click **Open**.

29. Click **OK** to start data analysis.

    A progress bar will appear in the Coffalyser - Run Wizard.

30. Click **Open Results Folder** to access the experiment subfolder in the output directory.

    In this folder a subfolder is created for each experiment that was defined. This contains the Coffa files and all generated reports.

    In addition, one or more subfolder(s) called Undefined Experiment might have been created when undefined samples are set as *Detect and include*. This folder contains files for barcodes detected in the FASTQ file that were not defined in your experiment(s) and this output is grouped based on detected product lot.

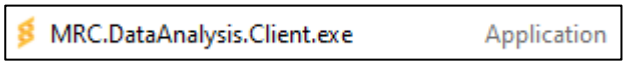
3.2.2 REANALYSE EXISTING EXPERIMENTS

Coffalyser digitalMLPA allows you to reanalyse existing experiments. This can be done by using an existing Coffalyser Definition File (*.cdf) and FASTQ files as input, or by using Coffa files (*.coffa). Reanalysis using Coffa files is much faster, but cannot be used in the following cases:

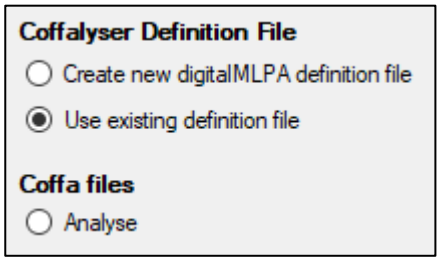
- When the wrong probemix or lot number was selected in the first analysis.
- When the wrong barcode plate lot was selected in the first analysis.
- When you want to change the output directory and/or temp directory.
- When Coffa files were generated using a different version of the Coffalyser digitalMLPA software that is not compatible with the version used for reanalysis.

3.2.2.1 REANALYSE AN EXPERIMENT USING EXISTING COFFALYSER DEFINITION FILE

1. Double click on MRC.DataAnalysis.Client.exe or the shortcut you created to start Coffalyser digitalMLPA.



2. Under Coffalyser Definition File select *Use existing definition file* and click **Next >**.



3. Navigate to the location of the Coffalyser Definition File, select the file and click **Open**.

The Coffalyser - Data Conversion - digitalMLPA - General Settings window will open.

IF YOU WISH TO EDIT YOUR EXPERIMENT, CONTINUE WITH STEP 4. OTHERWISE PROCEED WITH STEP 13.

4. Click **Configure** to edit your experiment(s).

   The Coffalyser - Samples window will open.

5. With the *Samples* window open begin editing. For example, to select barcodes that were missing in the initial experiment setup, right click on the selected barcodes and select *Add To Experiment > Your experiment*.

6. Right click and select *Edit Experiment*, if adjustments are needed. Alternatively, double-click on one of the three rightmost columns (experiment, type or sheet) of any sample included in that experiment.

   The Coffalyser - Experiment Overview window will open.

7. Make the desired adjustments to the experiment (e.g. experiment name, sample type, sample name).

8. Click **OK** to save and exit the *Coffalyser - Experiment Overview* window.

9. Click **OK** to save the experiment(s) and close the *Coffalyser - Samples* window.

10. When applicable, change the output directory by clicking the upper **Select** button. Browse to the desired folder, select it and click **OK**.

11. When applicable, change the Temp directory by clicking the lower **Select** button. Browse to the desired folder, select it and click **OK**.

12. When applicable, select another FASTQ file by clicking **Add File**. Browse to the location of the FASTQ file, select it and click **Open**. To remove the FASTQ file that was already imported, select it and click **Remove File**.

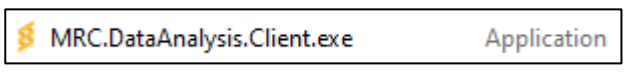
13. Click **OK** to start data analysis.

    A progress bar will appear in the Coffalyser - Run Wizard.

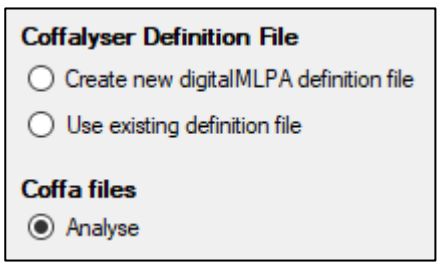
14. Click **Open Results Folder** to access the experiment subfolder in the output directory.

3.2.2.2 REANALYSE AN EXPERIMENT USING COFFA FILES

1. Double click on MRC.DataAnalysis.Client.exe or the shortcut you created to start Coffalyser digitalMLPA.



2. Under Coffa data files select *Analyse* and click **Next >**.



3. Browse to the folder that contains the Coffa files you wish to reanalyse, select it and click **OK**.

The Coffalyser - Data Analysis - General Settings window will open.

IF YOU WISH TO EDIT YOUR EXPERIMENT, CONTINUE WITH STEP 4. OTHERWISE PROCEED WITH STEP 9.

4. Click **Configure** to edit your experiment(s).

   The Coffalyser - Samples window will open.

5. Right click on any cell and select *Edit Experiment*.

   The Coffalyser - Experiment Overview window will open.

6. Make the desired adjustments to the experiment (e.g. experiment name, sample type, sample name).

7. Click **OK** to save and exit the *Coffalyser - Experiment Overview* window.

8. Click **OK** to save the experiment(s) and close the *Coffalyser - Samples* window.

9. Click **OK** to start data analysis.

   A progress bar will appear in the Coffalyser - Run Wizard.

10. Click **Open Results Folder** to access the experiment subfolder in the output directory.

4. COFFALYSER digitalMLPA REPORTS

Coffalyser digitalMLPA generates four different types of reports after data analysis.

Per experiment there are two types of reports:

- **Overview (PDF report):** This report gives a short overview of the general information of all samples included in an experiment. This includes sample name, barcode (well), sample type, errors encountered during analysis and quality of the sample (see chapter 4.1). We recommend first checking the overview report for a general impression of the quality of the experiment.
- **Ratios (Excel file):** This report provides an overview of the inter ratios, read counts and intra ratios for all relevant samples and probes in an experiment (see chapter 4.2).

Per sample, there are two types of reports:

- **Sample Results 1 (PDF report):** This report includes all information of a single sample including: the general details of a sample (e.g. sample name and gender), the experiment settings, quality details, results summary plot and aberrant region(s) that were found (see chapter 4.3).
- **General Details (PDF report):** This report includes the general details of a single sample, including the sample name, experiment settings, quality details and more information on errors if applicable (see chapter 4.4). It is useful for getting the first impression if the data from a sample is of sufficient quality.

Note: For a no-DNA sample only a single no-DNA PDF report will be generated (see chapter 4.5).

When printing reports from an experiment, select *Fit* in the print settings to fit all columns of the tables on the printed page(s) (see Figure 1).

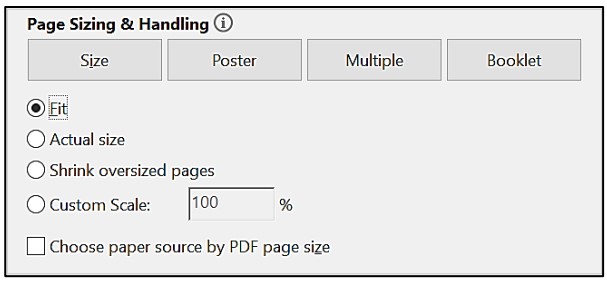
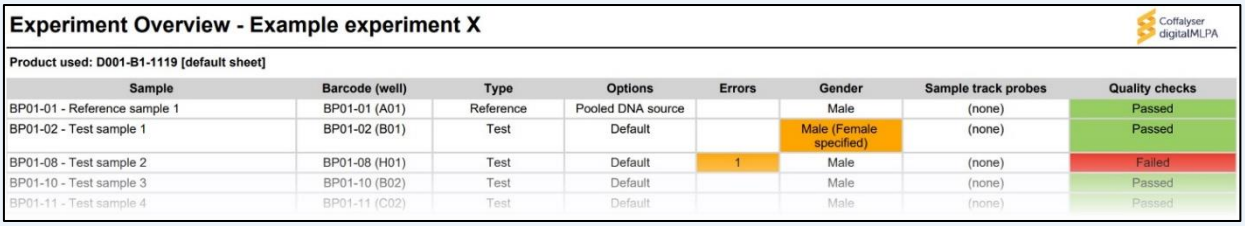


Figure 1: Print settings capture

4.1 [EXPERIMENT NAME] – OVERVIEW (PDF REPORT)

This report displays an overview of all samples within an experiment. It can be used to get a first impression of the quality of an experiment, such as if the quality of your samples was sufficient or if there were any errors during the analysis. The Sample Results 1.pdf report and/or General Details.pdf report will show more information per sample.



Experiment Overview - Example experiment X

Product used: D001-B1-1119 [default sheet]

Sample	Barcode (well)	Type	Options	Errors	Gender	Sample track probes	Quality checks
BP01-01 - Reference sample 1	BP01-01 (A01)	Reference	Pooled DNA source		Male	(none)	Passed
BP01-02 - Test sample 1	BP01-02 (B01)	Test	Default		Male (Female specified)	(none)	Passed
BP01-08 - Test sample 2	BP01-08 (H01)	Test	Default	1	Male	(none)	Failed
BP01-10 - Test sample 3	BP01-10 (B02)	Test	Default		Male	(none)	Passed
BP01-11 - Test sample 4	BP01-11 (C02)	Test	Default		Male	(none)	Passed

TOP

Product used displays the name of the Product Sheet used for the analysis.

COLUMNS

Sample
Displays the name that was specified for a sample in the Coffalyser Definition File. If no name was specified, the name of the barcode is displayed.

Barcode (well)
Displays the name of the barcode used for a sample. This includes the barcode plate reference (e.g. BP01 or BP02), the barcode number (between 01 and 96), and the well position of a given barcode on the barcode plate (listed between brackets). E.g. well A1 of BP01 that contains barcode 1 = BP01-01 (A01), and well H12 of BP01 that contains barcode 96 = BP01-96 (H12).

Type
Displays the type of sample (e.g. Reference, Test, No DNA) as specified in the Coffalyser Definition File.

Options
Displays the word “Default” unless a sample was indicated as being from a “pooled DNA source” in the Coffalyser Definition File.

Errors
Displays the number of errors, if any, that occurred during the analysis of a sample. If an error is displayed this often indicates that no reaction analysis and/or comparative analysis was performed due to the error (meaning no Sample Results 1 report was generated). A detailed explanation of the error can be found in the General Details report of the sample.

Gender
Displays the detected gender of a sample. When Coffalyser digitalMLPA cannot reliably predict the gender (e.g. only one Y-control probe is found while three Y-control probes are expected in male samples), it will display the term "Ambiguous". If you have specified the gender of a sample in the Coffalyser Definition File and there is a discrepancy between the findings of Coffalyser digitalMLPA, this will be indicated in this column with an orange colour as well as a reference to the gender specified (see the example above and in section 4.3.3).

Sample track probes
Displays the detected Sample Track Probe(s)[c]. A warning is given if a different Sample Track Probe is detected than the one(s) specified in the Coffalyser Definition File.

Quality checks
Displays if the reaction passed or failed the digitalMLPA quality criteria. More specific information on quality is listed in the Sample Results 1 report and the General Details report. If a sample is failed, the data should not be interpreted. If a sample is passed with warnings, data should be interpreted with caution.

ROWS

Data of samples is listed in rows.

[c] Sample Track Probes are not currently available.

4.2 [EXPERIMENT NAME] – RATIOS (EXCEL FILE)

Note: This file should be used with caution. Warnings on the quality of each individual sample should be carefully checked in the PDF reports.

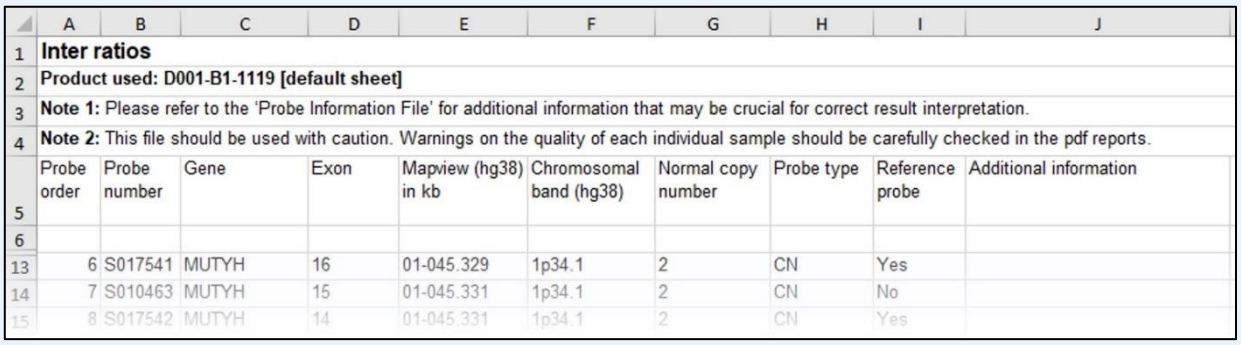
The Ratios Excel file can be used to investigate all samples in one experiment. It consists of five Excel sheets:

- **Inter ratios:** This sheet contains the inter ratios of all relevant reference, test and control probes in every sample. See chapter 5 and APPENDIX I – NORMALISATION for more information.
- **Reference sample quality:** This sheet contains information about the reference population used in the experiment for all relevant probes. This sheet is especially useful for finding the root cause of a warning or error for reference sample quality in an experiment.
- **Read counts:** This sheet contains the total read counts of all relevant reference, test and control probes in every sample.
- **Intra ratios:** This sheet contains the intra ratios, these are the ratios obtained after intra-normalisation calculations (reaction analysis). More information on intra-normalisation is described in APPENDIX I – NORMALISATION.
- **Software version:** This sheet contains the information about the version of each software module that was used for analysis of the experiment.

Please note that use of Conditional Formatting in Excel can aid in identifying deviating probe values. In addition, the Freeze Panes option in Excel allows for an area of a worksheet to remain visible while you scroll to another area of the worksheet, creating an easy overview of the most important information.

4.2.1 GENERAL AND PROBE INFORMATION FIELDS

Fields in the Ratios Excel file with general and probe information that are common between the first four sheets (Inter ratios, Reference sample quality, Read counts and Intra ratios) are described below.



	A	B	C	D	E	F	G	H	I	J
1	**Inter ratios**									
2	**Product used: D001-B1-1119 [default sheet]**									
3	**Note 1:** Please refer to the 'Probe Information File' for additional information that may be crucial for correct result interpretation.									
4	**Note 2:** This file should be used with caution. Warnings on the quality of each individual sample should be carefully checked in the pdf reports.									
5	Probe order	Probe number	Gene	Exon	Mapview (hg38) in kb	Chromosomal band (hg38)	Normal copy number	Probe type	Reference probe	Additional information
6										
13	6	S017541	MUTYH	16	01-045.329	1p34.1	2	CN	Yes	
14	7	S010463	MUTYH	15	01-045.331	1p34.1	2	CN	No	
15	8	S017542	MUTYH	14	01-045.331	1p34.1	2	CN	Yes	

TOP

Product used
Displays the name of the Product Sheet used for the analysis.

Notes
Displays essential information pertaining to the interpretation of results.

COLUMNS

Probe order
Displays the order number of the probes as specified in the Product Sheet.

Probe number
Displays the probe number of the digitalMLPA probe as designated by MRC Holland.

Gene
Displays the name of the gene that the digitalMLPA probe targets.

Exon
Displays the (nearest) exon number of the gene listed in the column "Gene" that the digitalMLPA probe targets.

Mapview (hg38) in kb
Displays the mapview location of the probe's target sequence, based on genome build hg38.

Chromosomal band (hg38)
Displays the chromosomal band location of the probe's target sequence, based on genome build hg38.

Normal copy number
Displays the copy number expected in samples from unaffected male individuals.

Probe type
Displays the probe type, this will be either: CN: copy number probe; CTRL: control probe (with the function specified in column "Gene"); MUT: mutation-specific probe; or WT: wild-type sequence specific probe at the location of a mutation.

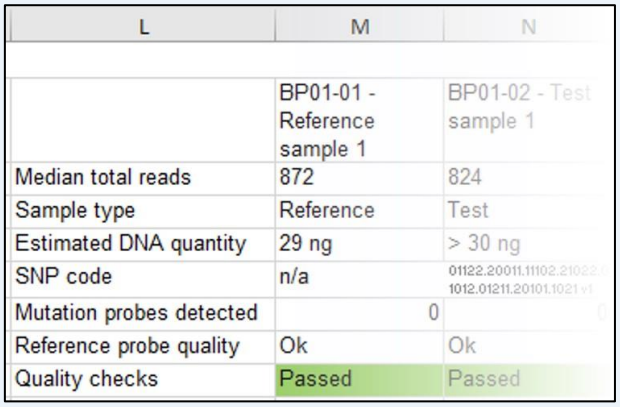
Reference probe
Displays if the probe is used as a reference probe during data normalisation. If a probe is used as a reference probe "Yes" will be listed in this column, otherwise "No" is listed.

Additional information
Displays important additional information (if available) that may help with data interpretation.

ROWS
Data of probes is listed in the rows.

4.2.2 SAMPLE INFORMATION FIELDS

Fields in the Ratios Excel file with sample information that are common between the Inter ratios, Read counts and Intra ratios sheets are described below.



L	M	N
	BP01-01 - Reference sample 1	BP01-02 - Test sample 1
Median total reads	872	824
Sample type	Reference	Test
Estimated DNA quantity	29 ng	> 30 ng
SNP code	n/a	01122.20011.11102.21022.0 1012.01211.20101.1021 v1
Mutation probes detected	0	0
Reference probe quality	Ok	Ok
Quality checks	Passed	Passed

COLUMNS

Each column represents a sample included in the experiment.

ROWS

Median total reads
Displays the median number of reads obtained for reference probes in each sample.

Sample type
Displays the sample type as defined in the Coffalyser Definition File.

Estimated DNA quantity
Displays the estimated DNA amount of the sample used in the digitalMLPA reaction. The DNA quantity estimation algorithm in CDM has been validated on MiSeq and MiniSeq NGS machines using a probemix with average number of probes. The accuracy of the estimation may be affected by the use of other NGS machines and the amount of probes present in the probemix.

SNP code
Displays the SNP code, generated based on a set of SNP-specific probes. See GLOSSARY for details.

Mutation probes detected
Displays the number of mutation-specific probes that gave reads above threshold in the sample. Not all probemixes contain mutation-specific probes.

Reference probe quality
Displays the estimated quality of the normalisation in a sample. This is based on the variation between intermediate ratios (see APPENDIX I – NORMALISATION). A warning highlighted in orange is displayed when the quality was found to be suboptimal (see APPENDIX II – QUALITY TABLE for more information).

Quality checks
Displays the overall outcome of the quality evaluation of a sample. The quality checks will either be “Passed” or “Failed”. The quality checks are further specified in the Sample Results 1 and General Details reports.

4.2.3 REFERENCE SAMPLE QUALITY SHEET

This sheet is especially useful for finding the root cause of a warning or error for reference sample quality in an experiment. Please note that due to a limitation in the software, the data displayed for the X- (non-pseudoautosomal region, non-PAR) and Y-chromosome probes cannot be directly used for troubleshooting. See APPENDIX III – TROUBLESHOOTING and our website for more information.

When defined reference samples are used in the experiment:

L	M	N	O	P	Q	R	S	T
Reference population statistics						*Reference sample inter ratios*	*Reference sample inter ratios*	*Reference sample inter ratios*
[REF] Average	[REF] Median	[REF] Minimum	[REF] Maximum	[REF] StDev		BP01-01 - Reference sample 1	BP01-03 - Reference sample 2	BP01-07 - Reference sample 3
1,00	0,99	0,96	1,05	0,04		0,99	1,05	0,96
1,00	1,01	0,97	1,03	0,03		0,97	1,01	1,03

When no reference samples are defined in the experiment (all test and undefined samples are used for normalisation):

L	M	N	O	P	Q
Reference population statistics					
[REF] Average	[REF] Median	[REF] Minimum	[REF] Maximum	[REF] MAD * 1.4826	*In this experiment no reference samples were defined and therefore population analysis was applied. For inter ratios of all samples see the Inter ratios tab.*
0,99	1,00	0,91	1,09	0,05	
1,00	1,00	0,92	1,13	0,04	

COLUMNS

[REF] Average
Displays the average inter ratio of the samples (reference samples if defined in the experiment) used during inter-normalisation.

[REF] Median
Displays the median inter ratio of the samples (reference samples if defined in the experiment) used during inter-normalisation.

[REF] Minimum
Displays the minimum inter ratio of the samples (reference samples if defined in the experiment) used during inter-normalisation.

[REF] Maximum
Displays the maximum inter ratio of the samples (reference samples if defined in the experiment) used during inter-normalisation.

When defined reference samples are used in the experiment (upper image):

[REF] StDev
Displays the standard deviation per probe in the reference samples.

Reference sample inter ratios
Displays the inter ratios per reference sample (identical to data present in Inter ratios sheet).

When no reference samples are defined in the experiment (lower image):

[REF] MAD * 1.4826
Displays the median absolute deviation (MAD) per probe in the samples used during inter-normalisation, corrected by the factor 1.4826 to make it comparable to standard deviation.

StDev and corrected MAD interpretation
If the StDev (defined reference samples used) or the corrected MAD (no reference samples defined) is above specific set thresholds, the value is highlighted in orange (warning) or red (error). This indicates that the StDev/corrected MAD of this probe within the population used for inter-normalisation exceeds the set limits of accepted variation, i.e. too much spreading present. Result interpretation of the affected highlighted probe(s) should be done with care in all samples. The estimation of reference sample quality displayed in the Quality table of the PDF reports (see APPENDIX II – QUALITY TABLE) is based on the above described StDev/corrected MAD.

ROWS

Data of relevant probes is listed in the rows. Note, not all probes in the probemix are included in the reference sample quality check.[d]

[d] For female samples Y target probes are listed, but not included in the reference sample quality check.

4.3 [BARCODE – SAMPLE NAME] – SAMPLE RESULTS 1 (PDF REPORT)

For every sample in an experiment that undergoes successful comparative analysis a Sample Results 1 report is generated. This report consists of several sections that contain information about the sample, experiment settings, analysis details, quality details and aberrant region(s) that were found.

4.3.1 SAMPLE QUALITY

These boxes indicate the overall outcome of the sample quality evaluation. It can be either "Failed" or "Passed". When the sample quality is failed, the results are not reliable. If a sample is passed with warnings, data should be interpreted with caution.

Detailed information on the sample quality is located on the first page of the report. On this page the quality is divided in different components. More information on this part of the report can be found in the section 4.3.8. QUALITY DETAILS – QUALITY TABLE.

4.3.2 GENERAL INFORMATION - SOFTWARE INFORMATION

This table indicates when each step of the data analysis was executed and which version of the Coffalyser digitalMLPA software module was used in this particular step. When a step has not been completed (e.g. due to an error), it is also shown here.

- **Configuration:** This is part of the software package and does not have to be executed.
- **Data conversion:** This is the conversion of digitalMLPA data from FASTQ to Coffa files.
- **Fragment analysis:** This is the first step of data analysis. In the fragment analysis step, each individual sample is analysed using reference probes for comparison (reaction analysis).
- **Comparative analysis:** This is the second step of data analysis. In the comparative analysis step, test samples are compared to the reference sample population.
- **Extended analysis:** This is only applicable for certain probemixes that require an extended analysis.
- **Reporting:** This is the step where reports are generated.

4.3.3 GENERAL INFORMATION - SAMPLE

The table *Sample* contains the following information:

- **Sample name (type):** This displays the name and type that was specified for the sample in the Coffalyser Definition File. If no name was specified, the name of the barcode will be displayed.
- **Barcode – well (lot):** This displays the name of the barcode used for the sample, followed by the well position of that barcode on the barcode plate and the detected barcode plate lot. The barcode name includes the barcode plate reference (e.g. BP01 or BP02) and the barcode number (between 01 and 96). E.g. well A1 of BP01 that contains barcode 01 will be displayed as BP01-01 – A01, and well H12 of BP01 that contains barcode 96 will be displayed as BP01-96 – H12. The barcode plate lot is detected automatically and should match the lot number on the barcode plate used. This is displayed as e.g. 03-007-yymmdd.
- **SNP code:** This displays the SNP code, or sample identifier, of a sample and the version number of the SNP probe collection. Additional information is described in the GLOSSARY and in the digitalMLPA NXtec Protocol.
- **Gender:** This displays the detected gender of the sample. When Coffalyser digitalMLPA cannot reliably determine the gender (e.g. only one Y-control probe is found while three Y-control probes are expected in male samples), it will display the term "Ambiguous". If you have specified the gender in the Coffalyser Definition File and there is a discrepancy between the findings of Coffalyser digitalMLPA, this will be indicated here. See Figure 2 for an example.

Figure 2: Image from a report where there was a discrepancy between the gender identified by Coffalyser digitalMLPA and the gender indicated in the Coffalyser Definition File.

- **Estimated DNA quantity:** This displays the estimated sample DNA amount used in the digitalMLPA reaction.
- **Sample track probes:** This displays the detected Sample Track Probe(s)[e]. A warning is given if a different Sample Track Probe is detected than the one(s) specified in the Coffalyser Definition File.

4.3.4 GENERAL INFORMATION – EXPERIMENT SETTINGS

The table *Experiment settings* contains the following information:

- **Experiment name (type):** This displays the name of the experiment. This can be either the name that was indicated in the Coffalyser Definition File, or if this was not defined, it will display Undefined Experiment 01 etc. Type refers to the experiment type as indicated in the Coffalyser Definition File (e.g. digitalMLPA).
- **Product name:** This displays the name of the Product Sheet used for the analysis.
- **Product detection:** This displays whether product identity detection was successful. The software will detect the product lot that was used based on a set of identifiers of which the combination is unique for a certain probemix lot. In Table 1 the different possibilities for this cell are given, what is displayed in the Quality table and an explanation.

Table 1: Overview of product detection related information in the report.

Product detection in Experiment settings	Product detection in Quality Table	Explanation
Identity detected	Ok, not displayed	Product Sheet detection by the software was successful and corresponds to the Product Sheet selected in the Coffalyser Definition File (if one was specified).
Identity NOT detected	Error [f]	The detected identifiers do not correspond to a Product Sheet (product lot) that is known by the software.[g] Ensure the experiment is preformed according to protocol, probemix lots are not mixed and you are using the latest software version.
CONFLICTING identity detected	Error	The Product Sheet (product lot) selected in the Coffalyser Definition File does not correspond to the product lot that is detected by the software.[f] Ensure the correct Product Sheet is selected in the Coffalyser Definition File.
Identity detected, x disqualified identifier(s)	Error	Product Sheet detection by the software was successful, but additional identifier(s) are observed in small amounts. Ensure there is no contamination during your experiment, e.g. from another probemix lot.
Identity NOT detected, x disqualified identifier(s)	Warning [g]	The detected identifiers do not correspond to a Product Sheet (product lot) that is known by the software and additional identifier(s) are observed in small amounts.[f] In rare cases identifier probes may give lower read counts than expected. Ensure the experiment is performed according to protocol and contact MRC Holland at info@mrcholland.com.

[e] Sample Track Probes are not currently available.

[f] This error is only displayed in the Quality Table in case Fragment Analysis (reaction analysis) is successfully performed.

[g] Please note that the Product Sheet selected in the Coffalyser Definition File will be used for analysis.

Product detection in Experiment settings	Product detection in Quality Table	Explanation
CONFLICTING identity detected, x disqualified identifier(s)	Error	The Product Sheet (product lot) selected in the Coffalyser Definition File does not correspond to the product lot that is detected by the software.[f] Also additional identifier(s) are detected in small amounts. Ensure the correct Product Sheet is selected in the Coffalyser Definition File and that there is no contamination during your experiment, e.g. from another probemix lot.

- **Digital signatures:** This displays whether all Coffalyser digitalMLPA files in the package on your local computer used for the data analysis are identical to those provided by MRC Holland. If it states "All signatures valid", this means these files are in their original form.
- **Probe catalogue:** This displays the Probe Catalogue version that was used for the analysis, including the release date and time. The Probe Catalogue contains (a reference to) information on all digitalMLPA probes.
- **Source files:** This displays the location and name(s) of the FASTQ file(s) used.

4.3.5 GENERAL INFORMATION – ANALYSIS DETAILS

The table *Analysis details* contains the following information:

- **Normalisation method:** This displays the normalisation method used.
- **Coffa file location:** This displays the location where the Coffa file was saved when the report was generated.
- **Coffa file id:** This displays the unique Coffa file identification code.
- **Reference samples:** This displays the reference samples used in the analysis of the sample.
  Note: If many reference samples are included in the experiment or if the reference samples have long names, sections Reads (4.3.6) and Probes (4.3.7) may overlap with the Chart (4.3.9). To prevent this, we strongly recommend the use of short names for the reference samples.

4.3.6 GENERAL INFORMATION – READS

The table *Reads* provides more information on the reads associated with a sample.

- **Recognised:** The column *Items* displays the number of unique sequences (items) that were recognised in this sample as belonging to a specific digitalMLPA probe present in the probemix used. This includes all target probes, reference probes and control probes. Note: this number does not have to be equal to the number of probes present in the probemix (e.g. reads for certain SNP-specific probes may not have been found).
  The column *Reads* displays the total number of reads that have the sample barcode and are recognised as digitalMLPA probes present in the probemix used.
- **Unrecognised:** This displays the number of reads that contain the barcode sequence of the sample, but could not be assigned to a digitalMLPA probe present in the probemix used, e.g. by too many errors in the read.
- **Median total reads:** This displays the median number of reads assigned to the reference probes in a sample. This is a measure for the read depth in a sample.
- **Median distinct reads:** This displays the median number of distinct reads assigned to the reference probes in a sample. Distinct reads are based on the number of independent ligation events (recognised by unique molecular identifiers; UMIs). The unique reads check in the Quality table (see APPENDIX II – QUALITY TABLE) is based on this metric.

4.3.7 GENERAL INFORMATION – PROBES

The table *Probes* contains information on different types of probes in the probemix. The number of *Expected* (as defined in the Product Sheet) and *Detected* probes are displayed. If there are more probes detected than expected, this number will be displayed in the column *Unexpected*.

- **Test probes:** This displays the number of probes that were inter-normalised. This includes both target probes and reference probes in a probemix. In reference samples, the number of expected and detected test probes must be the same. In test samples, the number of detected test probes may be lower than the number of expected test probes, for instance due to a homozygous deletion.
- **Reference probes:** This displays the number of probes used as reference probes during data normalisation. Reference probes are expected to have a normal copy number in all samples, and therefore should show the same number of expected and detected probes. However, in some samples, differences may occur in the number of expected and detected reference probes if copy number changes exist in the region where the reference probe is located. In case too many reference probes are missing a warning or error will appear in the Quality table (see section 4.3.8) as inter-normalisation may not be reliable anymore.
- **Y-control probes:** This displays the number of control probes that target the Y chromosome and are used to determine the gender of a sample. For female samples both the expected and detected number of Y probes should be 0.
- **Mutation probes:** This displays the number of mutation-specific probes in the probemix. The expected number of mutations is written as 0-n, with n being the number of mutation-specific probes in the probemix. If a mutation-specific probe generated sufficient reads (because the mutation is present in the sample), it is displayed in the column *Detected*. More information on the mutation-specific probes (if present in a probemix) can be found in the table *Mutation specific probes* further down in the Sample Results 1 report (see section 4.3.10).

4.3.8 QUALITY DETAILS – QUALITY TABLE

Quality assessment is an essential part of digitalMLPA data analysis: based on this evaluation it is determined if the data is of sufficient quality for reliable result interpretation. It may also help to identify possible problems with the sample and/or experiment.

Coffalyser digitalMLPA assesses the quality of each individual sample (*Reaction analysis*), and the quality of each sample compared to the reference population (*Comparative analysis*). It displays the results of these evaluations in the Quality table.

NXtec probemixes contain a large number of control probes for a very thorough quality evaluation.

For each category of control probes (e.g. denaturation control) multiple probes are included in every NXtec probemix. The software checks if the control probes meet the set criteria. In the *Reaction analysis*, the intra ratios of the control probes are compared to what is expected based on internal testing data generated at MRC Holland. During *Comparative analysis*, the inter ratios of the control probes are compared between the test sample and reference population. Control probe ratios should be similar in the test sample and reference population, which is necessary for reliable data interpretation.

When there are differences (outside of the defined normal range) in the evaluation of results of the control probes in the *Reaction analysis* and/or *Comparative analysis*, a warning or error is displayed in the reports. If warnings or errors are present, analysis results should be examined and interpreted cautiously.

For more information on the functions of the different quality checks present in the table see APPENDIX II – QUALITY TABLE.

4.3.9 RESULT SUMMARY – CHART(S)

In the *Inter ratios* plot, every dot represents a test or reference probe. They are sorted by gene or chromosomal location (X-axis), depending on the probemix. On the Y-axis the inter ratio of the probes is shown. This plot displays an overview of the variation of a sample.

The *Inter ratios aberrant region(s)* plot shows regions containing aberrant probes in more detail.

Please note, for detailed analysis refer to the *Aberrant region(s)* table (see section 4.3.11) as this graph does not display the results of the statistical analyses.

4.3.10 RESULT SUMMARY – MUTATION-SPECIFIC PROBE(S)

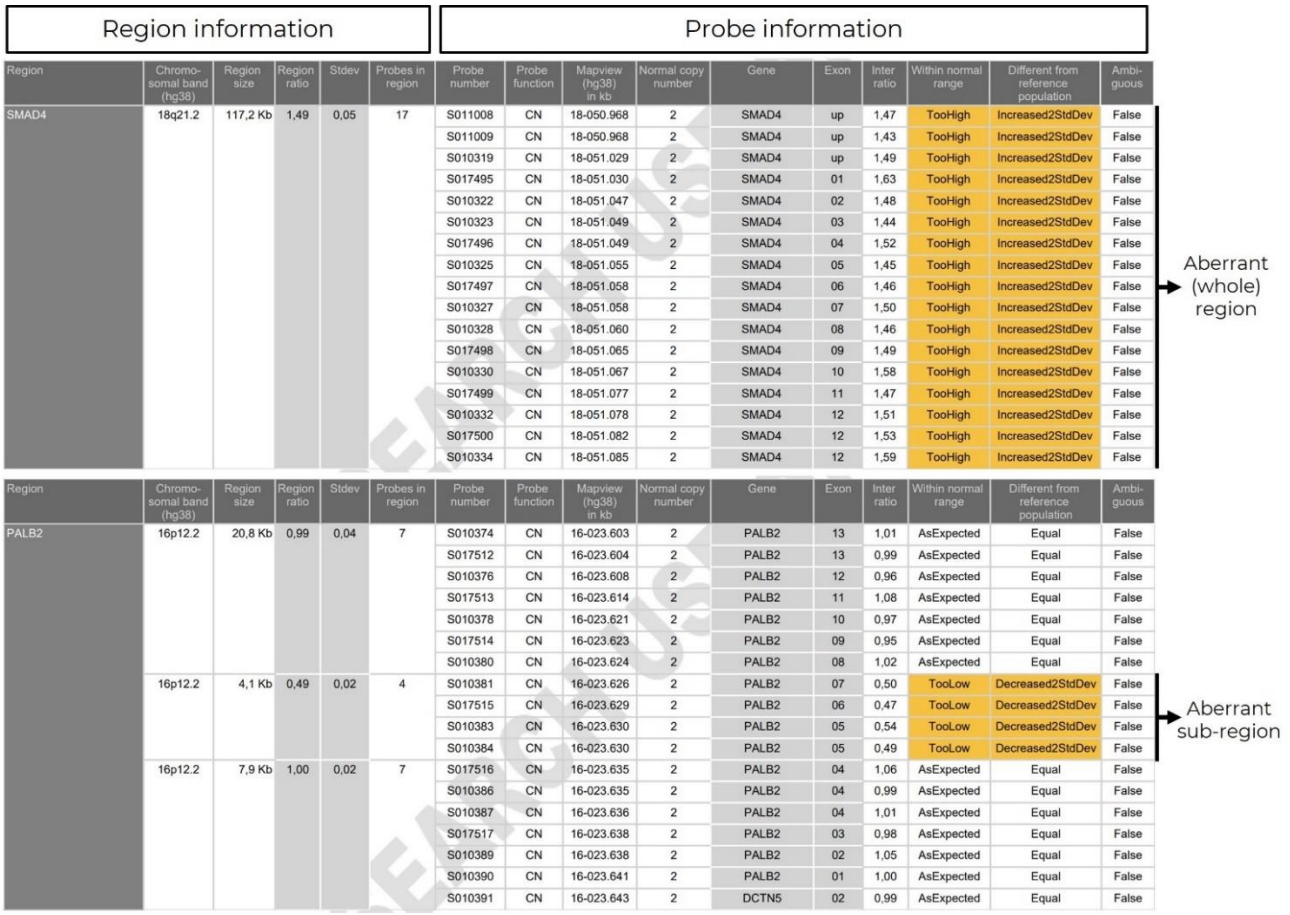
The table *Mutation specific probes* lists all digitalMLPA mutation-specific probes in the probemix.

- **Gene:** This displays the gene name that the probe targets.
- **Mutation:** This displays the mutation that the probe targets.
- **Probe:** This displays the probe number as designated by MRC Holland.
- **Mutation Status:** This displays if the mutation is present or absent in a sample. The mutation is indicated as present when the number of assigned reads for the mutation-specific probe is above the defined background threshold determined by MRC Holland.
- **Additional information:** This displays any additional information about the probe or mutation.

4.3.11 RESULT SUMMARY – ABERRANT REGION(S) TABLE

The *Aberrant region(s) table* displays regions with aberrant probe inter ratios. Aberrant is defined as deviating from what is expected (*Within normal range* column, see Figure 3); it does not necessarily indicate that the sample has a true deletion or duplication! Deviation of a probe may be due to a SNP or experimental factor causing variation. Results must be confirmed by another probemix or technique, especially if only one probe is deviating! If all probes in the probemix are within the normal range, no table is displayed.

Note: If (many) reference probes are listed in the *Aberrant region(s) table* of the Sample Results 1 report, indicating deviating inter ratios, this may be an indication that the reference sample collection was not optimal. This will also result in a warning for reference sample quality on the last page of the Sample Results 1 report. Which probes are used as reference probes can be found in the product-specific Product Information File and in the Ratios Excel file.



Region	Chromosomal band (hg38)	Region size	Region ratio	Stdev	Probes in region	Probe number	Probe function	Mapview (hg38) in kb	Normal copy number	Gene	Exon	Inter ratio	Within normal range	Different from reference population	Ambiguous
SMAD4	18q21.2	117,2 Kb	1,49	0,05	17	S011008	CN	18-050.968	2	SMAD4	up	1,47	TooHigh	Increased2StdDev	False
						S011009	CN	18-050.968	2	SMAD4	up	1,43	TooHigh	Increased2StdDev	False
						S010319	CN	18-051.029	2	SMAD4	up	1,49	TooHigh	Increased2StdDev	False
						S017495	CN	18-051.030	2	SMAD4	01	1,63	TooHigh	Increased2StdDev	False
						S010322	CN	18-051.047	2	SMAD4	02	1,48	TooHigh	Increased2StdDev	False
						S010323	CN	18-051.049	2	SMAD4	03	1,44	TooHigh	Increased2StdDev	False
						S017496	CN	18-051.049	2	SMAD4	04	1,52	TooHigh	Increased2StdDev	False
						S010325	CN	18-051.055	2	SMAD4	05	1,45	TooHigh	Increased2StdDev	False
						S017497	CN	18-051.058	2	SMAD4	06	1,46	TooHigh	Increased2StdDev	False
						S010327	CN	18-051.058	2	SMAD4	07	1,50	TooHigh	Increased2StdDev	False
						S010328	CN	18-051.060	2	SMAD4	08	1,46	TooHigh	Increased2StdDev	False
						S017498	CN	18-051.065	2	SMAD4	09	1,49	TooHigh	Increased2StdDev	False
						S010330	CN	18-051.067	2	SMAD4	10	1,58	TooHigh	Increased2StdDev	False
						S017499	CN	18-051.077	2	SMAD4	11	1,47	TooHigh	Increased2StdDev	False
						S010332	CN	18-051.078	2	SMAD4	12	1,51	TooHigh	Increased2StdDev	False
						S017500	CN	18-051.082	2	SMAD4	12	1,53	TooHigh	Increased2StdDev	False
						S010334	CN	18-051.085	2	SMAD4	12	1,59	TooHigh	Increased2StdDev	False

Region	Chromosomal band (hg38)	Region size	Region ratio	Stdev	Probes in region	Probe number	Probe function	Mapview (hg38) in kb	Normal copy number	Gene	Exon	Inter ratio	Within normal range	Different from reference population	Ambiguous
PALB2	16p12.2	20,8 Kb	0,99	0,04	7	S010374	CN	16-023.603	2	PALB2	13	1,01	AsExpected	Equal	False
						S017512	CN	16-023.604	2	PALB2	13	0,99	AsExpected	Equal	False
						S010376	CN	16-023.608	2	PALB2	12	0,96	AsExpected	Equal	False
						S017513	CN	16-023.614	2	PALB2	11	1,08	AsExpected	Equal	False
						S010378	CN	16-023.621	2	PALB2	10	0,97	AsExpected	Equal	False
						S017514	CN	16-023.623	2	PALB2	09	0,95	AsExpected	Equal	False
						S010380	CN	16-023.624	2	PALB2	08	1,02	AsExpected	Equal	False
	16p12.2	4,1 Kb	0,49	0,02	4	S010381	CN	16-023.626	2	PALB2	07	0,50	TooLow	Decreased2StdDev	False
						S017515	CN	16-023.629	2	PALB2	06	0,47	TooLow	Decreased2StdDev	False
						S010383	CN	16-023.630	2	PALB2	05	0,54	TooLow	Decreased2StdDev	False
						S010384	CN	16-023.630	2	PALB2	05	0,49	TooLow	Decreased2StdDev	False
	16p12.2	7,9 Kb	1,00	0,02	7	S017516	CN	16-023.635	2	PALB2	04	1,06	AsExpected	Equal	False
						S010386	CN	16-023.635	2	PALB2	04	0,99	AsExpected	Equal	False
						S010387	CN	16-023.636	2	PALB2	04	1,01	AsExpected	Equal	False
						S017517	CN	16-023.638	2	PALB2	03	0,98	AsExpected	Equal	False
						S010389	CN	16-023.638	2	PALB2	02	1,05	AsExpected	Equal	False
						S010390	CN	16-023.641	2	PALB2	01	1,00	AsExpected	Equal	False
						S010391	CN	16-023.643	2	DCTN5	02	0,99	AsExpected	Equal	False

Figure 3: Examples of *Aberrant region(s) tables* depicting when either an entire region is called as aberrant (top) or only a part to a region is called as aberrant (bottom).

4.3.11.1 REGION INFORMATION

Probes are categorised by region; this can be a chromosomal region or a gene region, depending on the probemix. The first six columns give information on the region.

- ***Region*:** This displays the name of the region (e.g. (part of) a chromosome or gene).

If only a part of a region is aberrant (=aberrant sub-region), the region is divided. This will be visible from column 2 – *Chromosomal band (hg38)* to column 6 – *Probes in region*.

- **Chromosomal band (hg38):** This displays the band on the chromosome where the aberrant (sub-) region is located, based on genome build hg38.
- **Region size:** This displays the distance between the first and last probe in this region.
- **Region ratio:** This displays the average over the inter ratios of the probes in the aberrant (sub-) region.
- **Stdev:** This displays the standard deviation calculated over the inter ratios of the probes in the aberrant (sub-) region.
- **Probes in region:** This displays the number of probes present in the aberrant (sub-) region.

4.3.11.2 PROBE INFORMATION

The last ten columns of the *Aberrant region(s) table* display information per probe in the region.

- **Probe number:** This displays the probe number of the digitalMLPA probe as designated by MRC Holland.
- **Probe function:** This displays the probe type. CN: copy number probe; WT: wild-type sequence specific probe.
- **Mapview (hg38) in kb:** This displays the mapview location of the probe's target sequence, based on genome build hg38.
- **Normal copy number:** This displays the expected copy number of the target sequence of a probe in samples from healthy male individuals.
- **Gene:** This displays the name of the gene that the digitalMLPA probe targets.
- **Exon:** This displays the (nearest) exon number of the gene listed in "Gene" that the digitalMLPA probe targets.
- **Inter ratio:** This displays the inter ratio of the probe.
- **Within normal range:** This displays the category that the inter ratio of a probe falls into (see chapter 5 for more information).

Note: Probes with inter ratios in the normal range are only displayed in the *Aberrant region(s) table* when another probe in the same region has an aberrant inter ratio.

- **Column *Different from reference population*:** This displays if the inter ratio of the probe in a sample is different compared to the inter ratios of the same probe in the reference population. This can state equal or be listed as increased or decreased by one or two standard deviations. See chapter 5 for more information.
- **Column *Ambiguous*:** This displays if the probe result is found to be ambiguous or not (indicated by 'True' or 'False', respectively). See chapter 5 for more information.

4.4 [BARCODE – SAMPLE NAME] – GENERAL DETAILS (PDF REPORT)

The General Details report displays an overview of the general details and the quality of a sample. In addition, errors mentioned in the Overview report are further explained in this report. This report can be especially helpful for troubleshooting in case no Sample Result 1 report is generated. Most tables in this report are identical to the tables presented on the first page of the Sample Results 1 report. Only the sections Analysis details and Errors are different, as discussed below. Please refer to section 4.3 for information on the other tables.

4.4.1 GENERAL INFORMATION - ANALYSIS DETAILS

The table *Analysis details* contains the following information:

- **Analysis settings:** This displays the analysis mode used.
- **Normalisation method:** This displays the normalisation method used.
- **Coffa file location:** This displays the location where the Coffa file was saved when the report was generated.
- **Coffa file id:** This displays the unique Coffa file identification code.
- **Extended module:** This displays whether an extended analysis was performed.
- **Read count type:** This displays the type of reads used for calculations. In most probemixes this is the total read count (indicated by *Total*).
- **Reference samples:** This displays the reference samples used in the analysis of the sample.

4.4.2 QUALITY DETAILS - ERRORS

If applicable, errors are shown in the digitalMLPA Run Wizard directly after data analysis and are described in more detail in the General Details report. More information about interpretation of these error messages can be found in APPENDIX III – TROUBLESHOOTING.

4.5 [BARCODE – SAMPLE NAME] – NO DNA (PDF REPORT)

For a no-DNA sample that is defined as such in the Coffalyser Definition File, a no-DNA report is generated instead of the Sample Results 1 and General Details reports. In the no-DNA report, a general overview of the sample is given. The tables in this report are highly similar to those of the other two PDF reports, the differences are described below.

4.5.1 GENERAL INFORMATION – SOFTWARE INFORMATION

This table is identical to what is presented in the Sample Results 1 and General Details reports, except that Comparative analysis is not shown as the sample is not subjected to inter-normalisation.

4.5.2 GENERAL INFORMATION – SAMPLE

- **SNP code** and **Gender** are not displayed, as these are not applicable to a no-DNA A sample.
- **Estimated DNA quantity**: When a no-DNA sample is clean, there should be an insufficient number of reads available for DNA quantity estimation. When a no-DNA sample contains many reads that can be assigned to probes, e.g. due to contamination, a value is printed and highlighted in orange.
- **Median distinct reads**: This displays the median number of distinct reads of the reference probes in a sample. Distinct reads are based on the number of independent ligation events (recognised by unique molecular identifiers; UMIs). In a clean no-DNA sample a very low number of median distinct reads is expected.

4.5.3 GENERAL INFORMATION – ANALYSIS DETAILS

This table is identical to what is presented in the General Details report, except that Reference samples are not shown as the sample is not subjected to inter-normalisation.

5. ANALYSIS OF RESULTS

5.1 DISPLAY OF RESULTS

5.1.1 COMPARISON TO COPY NUMBER THRESHOLDS

Inter ratios are categorized based on the copy number threshold set for each probe in a specific probemix. Based on the category, the software applies a colour highlight to the probe result in the column *Within normal range* of the *Aberrant region(s) table* in the Sample Results 1 report and in the sheet *Inter ratios* of the Ratios Excel file.

Table 2 shows an example of the possible categories, the corresponding information, and the highlight colours of the column *Within normal range* and the sheet *Inter ratios*. Refer to the NXtec probemix-specific Product Description for interpretation of results.

Table 2: Categories of inter ratios and the corresponding information and background colour.

Category	**Copy number threshold** (example)	**Within normal range column** (Sample Results 1 PDF report)	**Inter ratios sheet** (Ratios Excel file)
1.	>1.3	Too high; Orange	Dark blue
2.	1.2 – 1.3	High; Yellow	Light blue
3.	0.8 – 1.2	As expected; None [h]	None
4.	0.7 – 0.8	Low; Yellow	Light orange
5.	<0.7	Too low; Orange	Dark orange

5.1.2 COMPARISON TO REFERENCE POPULATION

In addition to calculating inter ratios and comparing those to copy number thresholds, Coffalyser digitalMLPA also estimates the difference in the inter ratio of a sample compared to that of the reference population per probe. This deviation between the inter ratios of a probe can be either: Equal (negligible difference between the ratios), differ by 1 standard deviation decrease/increase (non-significant difference between the ratios), or differ by 2 standard deviations decrease/increase (significant difference between the ratios).

The results of the estimate of deviation are displayed in the column *Different from reference population* of the *Aberrant region(s) table* in the Sample Results 1 report and in the sheet *Inter ratios* of the Ratios Excel file. Table 3 shows the possibilities, the highlight colours and formatting of the column *Different from reference population* and the sheet *inter ratios*.

Table 3: Background colour and formatting based on deviation from reference population.

Deviation from the reference population	**Different from reference population column** (Sample Results 1 PDF report)	**Inter ratios sheet** (Ratios Excel file)
Equal	Equal ; None	None
Decreased or Increased by 1 Standard Deviation	Decreased 1 StdDev / Increased 1 StdDev; Yellow	***Bold + italic***
Decreased or Increased by 2 Standard Deviation	Decreased 2 StdDev / Increased 2 StdDev; Orange	**Bold**

5.1.3 INCOMPLETE DATA

When data is not complete, for example when not enough reference samples have been included in the digitalMLPA experiment, formatting of results in the Sample Results 1 report and the Ratios Excel file is different than described above. This is shown in Table 4. In such cases, also a warning on *Detected probes that did not meet the minimum requirements to derive a classification* will appear in the Quality table.

[h] Probes with inter ratios in the normal range are only displayed in the *Aberrant region(s) table* of the Sample Results 1 report when another probe in the same region has an aberrant probe ratio.

Table 4: Visualisation of the Sample Results 1 report and the Ratios Excel file for probes that did not meet the minimum requirements to derive a classification. Information displayed and background colour.

Cause	**Sample Results 1 PDF report**			**Inter ratios sheet** (Ratios Excel file)
	Inter ratio	Within normal range	Different from reference population	
Probe reads are not normalised to a sufficient number of reference samples. E.g. only two reference samples were included in the experiment or analysis of Y-probe reads in a male test sample when only one male reference sample was used.	Ratio; None	Unknown; Red	Unknown; Red	Grey
Probe reads are not normalised to reference samples. E.g. analysis of Y-probe reads in a male test sample when no male reference samples were used. Please note, Y probes in females are also displayed as depicted on the right, but this can be regarded as expected. For this situation the *Detected probes that did not meet the minimum requirements to derive a classification* warning is not displayed in the Quality table.	?; Red	Unknown; Red	Unknown; Red	Grey
Probe has zero assigned reads. E.g. complete deletion of DNA sequence targeted by probe.	0.00; Purple	Unknown; Purple	Unknown; Purple	Grey

5.2 INTERPRETATION

In addition to this manual please consult the probemix-specific Product Description, Probe Information File and, if available, other product-specific files provided by MRC Holland for interpretation of results.

For result interpretation consult all three rightmost columns (*Within normal range*, *Different from reference population* and *Ambiguous*) of the *Aberrant region(s) table* in the Sample Results 1 report. Table 5 provides an overview of the combination of calls in the columns *Within normal range*, *Different from reference population* and *Ambiguous* that may aid in result interpretation.

Table 5: Overview of combination of calls and highlights in the Sample Results 1 report.

Within normal range	**Different from reference population**	**Ambiguous**	**Interpretation**
As expected	Equal	False	Sample probe result is within the set boundaries for this probe and is similar to results of the reference population.
As expected	Decreased 1 StdDev or Increased 1 StdDev	False	Sample probe result is within the set boundaries for this probe and is not significantly different from the reference population.[i]
As expected	Decreased 2 StdDev or Increased 2 StdDev	False	Sample probe result is within the set boundaries for this probe and is significantly different from the reference population.[i]
Low or High	Equal	False	Sample probe result is outside the set inner boundaries for this probe and is similar to results of the reference population.[i]
Low or High	Decreased 1 StdDev or Increased 1 StdDev	False	Sample probe result is outside the set inner boundaries for this probe and is not significantly different from the reference population.[i]
Low or High	Decreased 2 StdDev or Increased 2 StdDev	False	Sample probe result is outside the set inner boundaries for this probe and is significantly different from the reference population.[i]
Too low or Too high	Equal	True	Sample probe result is outside the set boundaries for this probe and is similar to results of the reference population.[i]

[i] Proceed with caution when interpreting this data.

Within normal range	Different from reference population	Ambiguous	Interpretation
Too low or Too high	Decreased 1 StdDev or Increased 1 StdDev	False	Sample probe result is outside the set boundaries for this probe and is not significantly different from the reference population.[i]
Too low or Too high	Decreased 2 StdDev or Increased 2 StdDev	False	Sample probe result is outside the set boundaries for this probe and is significantly different from the reference population.
Unknown	Unknown		Sample probe result is not classified as insufficient reference samples were available during inter-normalisation. No conclusion can be drawn for this probe.
Unknown	Unknown		Sample probe result is not classified as zero reads are assigned to the probe.[i]

If the inter ratio of a probe falls into the highest or lowest category (resulting in 'Too high' or 'Too low' in the column *Within normal range*), it does not necessarily mean that the probe's target sequence is indeed duplicated or deleted. It is possible that the inter ratio of a probe in a sample is not significantly different from the reference population used in the analysis. For instance, if a probe is highly variable in the reference samples used, the column *Ambiguous* will state 'True' in this case. Results for probes showing this warning should be interpreted with caution!

When a probe deviates by 2 standard deviations from the reference population AND multiple adjacent probes show the same result the likelihood that the observed result is due to a copy number change is increased.

APPENDIX I – NORMALISATION

Coffalyser digitalMLPA uses a series of normalisation steps and calculations to compute the inter ratios of digitalMLPA probes. In this process and the associated population statistics, it is assumed that digitalMLPA data follows a normal distribution.

In a process called intra-normalisation, Coffalyser digitalMLPA converts absolute read counts of a sample into relative values (intra ratios) by normalising target probe read counts against the read counts of every reference probe in the sample[j]. This is done for every probe and every sample. During inter-normalisation, Coffalyser digitalMLPA compares the relative probe values of each sample to those of the reference population, resulting in inter ratios.

A simplified version of the complete normalisation process is as follows:

STEP 1

The read count of target probe 1 (Tp1) is divided by the read count of reference probe 1 (Rp1) in sample 1 to give a relative value for the target probe. The same is done for reference sample 1. The relative target probe value in the sample is then divided by the relative target probe value in the reference sample resulting in an intermediate ratio.

This calculation is then repeated using every **reference probe** included in the probemix for the target probe. This results in the same number of intermediate ratios for target probe 1 as there are reference probes in the probemix. Next, the median value of these intermediate ratios is determined. See the equation below.

$$Median\left(\frac{(Tp1\ in\ sample\ 1/Rp1\ in\ sample\ 1)}{(Tp1\ in\ reference\ sample\ 1/Rp1 in\ reference\ sample\ 1)},\ldots,\frac{(Tp1\ in\ sample\ 1/Rp_n\ \ in\ sample\ 1)}{(Tp1\ in\ reference\ sample\ 1/Rp_n\ in\ reference\ sample\ 1)}\right)$$

STEP 2

Step 1 is repeated using every **reference sample** included in the analysis. This results in as many median values for target probe 1 in sample 1 as there are reference samples in the analysis.

Coffalyser digitalMLPA then calculates the average value over these median values. This results in the inter ratio of target probe 1 in sample 1.

These calculations are repeated for all probes in the probemix and all samples in the experiment.

NO REFERENCE SAMPLES DEFINED IN ANALYSIS

If no reference samples are defined in an analysis, test and undefined samples are used for normalisation. In step 2, instead of taking the average over the previously calculated median values, Coffalyser digitalMLPA calculates the inter ratio for a probe by taking the median of these medians.

[j] The intra ratios available in the Ratios Excel file are based on the read counts of a test probe divided by the median read counts (rounded up to an integer) of the reference probes in the same sample.

APPENDIX II – QUALITY TABLE

In two Coffalyser digitalMLPA reports, General Details.pdf and Sample Results 1.pdf, a table is present outlining the detailed quality checks for the sample. This table is divided into two categories: (1) Reaction analysis, this focuses on the quality checks for each individual sample, and (2) Comparative analysis, this focuses on the quality checks from the comparison of a sample to the reference population. The various quality checks are explained in Table 6 below.

Quality checks can have three different classifications based on the presence and/or severity of the issue: OK, Warning and Error. Note: some quality checks are only displayed when a warning or error is identified (see Table 6).

The overall quality of the sample, as depicted by the Quality checks "Passed" or "Failed" boxes in green or red respectively (see 4.3.1), is based on the cumulative observations from all quality checks described in Table 6. Please note that the overall quality of a sample is based on a weighted scoring system that takes the number of control probes affected per category into account. This quality scoring is not visible in reports or completely reflected in the warning and error messages. Therefore, it is possible that samples with identical warning and error messages displayed on a report will have a different overall quality assessment. For more information on the control probes affected, check the intra and inter ratios of the control probes in the Ratios Excel file.

Please refer to the digitalMLPA NXtec Protocol for more information on the various quality checks and troubleshooting.

Table 6: Overview of quality checks which may appear in the Quality table.

Description	Category	Explanation
Product detection [g,k]	Reaction analysis	This quality classification is based on the success of product identity detection. When applicable, it also assesses whether product detection corresponds to the Product Sheet selected in the Coffalyser Definition File. See section 4.3.4 for more details.
Missing reference probes [k]	Reaction analysis	This quality classification is based on the assigned read counts for reference probes. Data analysis may not be reliable when too many reference probes have insufficient read numbers. In case an error is displayed for this quality check, comparative analysis will not be performed and the Sample Results 1 report will not be available.
Read depth	Reaction analysis	This quality classification is based on the median number of total reads of the reference probes for a sample, displayed in the Reads table as "Median total reads". When read depth is too low, an error is displayed as data analysis may not be reliable.
Unique reads [k]	Reaction analysis	This quality classification is based on the median number of distinct reads of the reference probes for a sample, displayed in the Reads table as "Median distinct reads". When the median number of distinct reads is too low, an error is displayed as data analysis may not be reliable.
Sequence quality	Reaction analysis	This quality classification is based on the quality of the sequencing reads. When many reads contain errors, a warning or error is displayed. To check read quality we recommend using Illumina's BaseSpace® Analysis Environment or Illumina's Sequence Analysis Viewer Software.
Unrecognised reads	Reaction analysis	This quality classification is based on the amount of reads that cannot be reliably assigned to a probe. When many digitalMLPA reads are not reliably assigned to one of the probes present in the detected or selected Product Sheet, a warning or error is displayed.

[k] These quality checks will only be displayed in the *reaction analysis* part of the Quality table in the event of a warning or error.

Description	Category	Explanation
DNA contamination [k]	Reaction analysis	This quality classification is based on the SNP code generated by the set of 39 SNP-specific probes. When there are three or more "?" signs, a warning or error is displayed. Please note that this contamination check is not performed when a sample is defined as "pooled DNA source". Furthermore, for some probemixes the contamination check is not performed, as detection of sample DNA contamination is not always possible, e.g. in tumour derived samples. Please consult the probemix-specific Product Description for more information.
Sample uniformity [k]	Reaction analysis	This quality classification is based on reads being evenly spread over the samples within one experiment. If a significantly different number of reads is assigned to a barcode as compared to the median of the other samples in the experiment, a warning or error is displayed.
X-presence control [k]	Reaction analysis	This quality classification is based on the set of control probes in a probemix that indicate the presence of X-chromosome sequences. If insufficient reads are detected for one or more of the X-chromosome control probes, a warning or error is displayed.
Y-presence control [k]	Reaction analysis	This quality classification is based on the set of control probes in a probemix that indicate the presence of Y-chromosome sequences. If one or more but not all Y-chromosome control probes are detected (visible in the *Probes* table, see section 4.3.7), a warning or error is displayed.
Sample DNA denaturation	Reaction analysis & comparative analysis	This quality classification is based on the set of control probes in a probemix that indicate if sample DNA denaturation was complete. A warning or error in the Reaction analysis category indicates that sample DNA denaturation was incomplete. A warning or error in the Comparative analysis category indicates that the test sample is not denatured to a similar extent as the reference population.
Sample DNA depurination [k]	Reaction analysis & comparative analysis	This quality classification is based on the set of control probes in a probemix that indicate the extent of sample DNA depurination. A warning or error in the Reaction analysis category indicates that the sample DNA is depurinated. A warning or error in the Comparative analysis category indicates that the test and reference population do not have the same level of depurination.
Sample DNA fragment length [k]	Reaction analysis & comparative analysis	This quality classification is based on the set of control probes in a probemix that indicate the sample DNA fragment length. A warning or error in the Reaction analysis category indicates that the sample DNA is heavily fragmented. A warning or error in the Comparative analysis category indicates that the test and reference population do not have the same level of DNA fragmentation.
Hybridisation Tm	Reaction analysis & comparative analysis	This quality classification is based on the set of control probes in a probemix that are dependent on the overnight hybridisation. A warning or error in the Reaction analysis category indicates that the hybridisation conditions were outside of the designated margins. A warning or error in the Comparative analysis category indicates that the hybridisation conditions differ between the test and the reference population.
Hybridisation completeness	Reaction analysis & comparative analysis	This quality classification is based on the set of control probes in a probemix that check the extent of the hybridisation reaction. A warning or error in the Reaction analysis category indicates that hybridisation was incomplete. A warning or error in the Comparative analysis category indicates that the level of hybridisation differs between the test and the reference population.

Description	Category	Explanation
Ligation start temperature	Reaction analysis	This quality classification is based on the set of control probes in a probemix that indicate if the ligation reaction was started at room temperature instead of 48°C. When these probes are outside specified borders, a warning or error is displayed.
Ligase activity	Reaction analysis & comparative analysis	This quality classification is based on the set of control probes in a probemix that check the ligase activity. A warning or error in the Reaction analysis category indicates that ligase activity is outside of the designated margins. A warning or error in the Comparative analysis category indicates that the ligase activity differs between the test and the reference population.
Ligase inactivation	Reaction analysis	This quality classification is based on the set of control probes in a probemix that detect when there is simultaneous ligase and polymerase activity, for example when there is incomplete ligase inactivation. When there is simultaneous activity of these enzymes a warning or error is displayed.
Polymerase activity	Reaction analysis & comparative analysis	This quality classification is based on the set of control probes in a probemix that check the polymerase activity. A warning or error in the Reaction analysis category indicates that polymerase activity is outside of the designated margins. A warning or error in the Comparative analysis category indicates that the polymerase activity differs between the test and the reference population.
Reference probe quality	Comparative analysis	This quality classification is based on the estimated quality of the data normalisation within a sample based on the variation between intermediate ratios (see APPENDIX I – NORMALISATION). If a large amount of variation is present between intermediate ratios a warning or error is displayed. This typically indicates that reference probes are not stable in the sample. Please note that in case of high amplifications (e.g. in tumour data), an RPQ warning or error can be triggered although reference probes are stable.
Reference sample quality	Comparative analysis	This quality classification is based on the estimated quality of the data normalisation in the experiment as indicated by the variation between inter ratios (standard deviation with defined reference samples or corrected median absolute deviation (MAD) when no reference samples are defined) in samples used for inter-normalisation. If a large amount of variation is present, a warning or error is displayed. This may indicate a sub-optimal choice of reference samples or, in case no references are defined, a suboptimal test population (e.g. too small, too many samples with the same aberration). More information can be found in the Reference sample quality sheet in the Ratios Excel file, see section 4.2.3 and on the MRC Holland support portal at https://support.mrcholland.com.
Detected probes that did not meet the minimum requirements to derive a classification [I]	Comparative analysis	This quality classification is based on whether there are probes present in the sample that could not be adequately inter-normalised, e.g. by an insufficient number of reference samples. As a result, these probes are not classified (see section 5.1.3), and a warning is displayed.
Number of reference samples [I]	Comparative analysis	This quality classification is based on whether enough reference samples are included in the experiment. An error is displayed when this is not the case.

[I] These quality checks will only be displayed in the *comparative analysis* part of the Quality table in the event of a warning or error.

APPENDIX III – TROUBLESHOOTING

The tables below list common errors encountered during data analysis with Coffalyser digitalMLPA. If you encounter problems or software errors not described below, please contact info@mrcholland.com and include a screenshot of the error if applicable.

GENERAL

Problem	Cause
Program "MRC.DataAnalysis.Client.exe" is not able to start.	Virus scanners can prevent the program from running from a network drive. If you unpacked your files on a network drive, try to unpack the files on a local drive. If the problem persists, please contact MRC Holland at info@mrcholland.com.
.NET Framework is not up-to-date (older than version 4.8).	The latest .NET framework can be downloaded from https://dotnet.microsoft.com/download/dotnet-framework.
The following error message is shown when starting Coffalyser digitalMLPA: *Failed to load the configuration: 'failed to read the digitalMLPA library from the specified folder (xxx), error message = 'file not found (xxx)'*	Check the path name of the folder where Coffalyser digitalMLPA is stored, very long paths need to be shortened. Move the software to a higher folder or shorten the name of parent folders.
The following error message is shown when starting Coffalyser digitalMLPA: *No valid Coffalyser digitalMLPA license found in the configuration.*	It is necessary to renew the free license for Coffalyser digitalMLPA. To renew the license, download a new version of Coffalyser digitalMLPA from your account on www.mrcholland.com/. Remove or archive the old version, and extract the new package. No further installation is required.
The following message is shown when starting Coffalyser digitalMLPA: *The current license will expire in xx days!* *Update your configuration and/or software to continue using this product after the expiration.*	The license expiration date is approaching. To renew the license, download a new version of Coffalyser digitalMLPA from your account on www.mrcholland.com/. Remove or archive the old version, and extract the new package. No further installation is required.
Results reports suggest problems with Illumina Next Generation Sequencing data quality, e.g. overclustering.	To optimize your Illumina sequencing run for digitalMLPA, check the guidelines provided in the digitalMLPA NXtec Protocol. In addition, on the Illumina Support Center (support.illumina.com), various documents are available to help you achieve better results, for example optimizing cluster density. We also recommend checking the quality of your sequencing run in the Illumina BaseSpace® Analysis Environment (basespace.illumina.com) or Illumina Sequence Analysis Viewer Software.

EXPERIMENT DEFINITION

Problem	Cause
The following warning is shown when reusing an old Coffalyser Definition File: *The specified analysis settings xxx cannot be found. Do you want to continue?*	This message indicates that the configuration of Coffalyser digitalMLPA has changed compared to when the Coffalyser Definition File was made. The Coffalyser digitalMLPA Release Notes, available in the "_Documentation" folder of the software package, contain information about the changes between software versions. When you choose to continue, the Coffalyser Definition File will be updated using the new settings.
The following error message is shown when pasting sample details from the clipboard into the Coffalyser Definition File Editor: *Failed to paste the sample details:* *the samples on your clipboard does not match the number of selected rows*	This message indicates that the number of copied sample names being pasted into Coffalyser digitalMLPA is not equal to the number of selected samples in the Coffalyser Definition File Editor. Please ensure the correct number of samples is selected in Coffalyser digitalMLPA before pasting the sample names.
The following error message is shown when pasting sample details from the clipboard into the Coffalyser Definition File Editor: *Failed to paste the sample details:* *the samples on your clipboard do not fit in the remaining number of rows given your current start cell.*	This message indicates that the number of copied sample names being pasted into Coffalyser digitalMLPA exceeds the number of remaining samples in the range. Please ensure that the number of sample names you want to copy is smaller than or equal to the range of samples in the experiment in Coffalyser digitalMLPA.

DATA ANALYSIS (COFFALYSER digitalMLPA – RUN WIZARD)

Problem	Cause
The run is completed but error messages are shown in the summary window.	Click "Open Results Folder" and review the Overview.pdf and General Details.pdf reports for more information.
Output files of some of the samples/barcodes ended up in the "Undefined Experiment {nn}" folder.	Ensure that all the barcodes that were part of your experiment were included in the Coffalyser Definition File. Reanalysis can be done using all Coffa files belonging to the experiment (see section 3.2.2.2) or by using an updated Coffalyser Definition File (see section 3.2.2.1).
I received the following error message: *Detected a sample (xxx) without a sample barcode details object.*	This can happen if no reads are detected for a barcode that was selected and included in the Coffalyser Definition File. When defining experiments in the Coffalyser Definition File, ensure you only select the barcodes that are present in the data. Also ensure the experiment has been performed according to the instruction in the digitalMLPA NXtec Protocol, and both barcode solution and probemix are added to a sample.

Problem	Cause
I received the following error message: *The experiment (xxx) does not have a sheet.*	This can happen when the Product Sheet required for analysis of the digitalMLPA results is not available in the version of Coffalyser digitalMLPA used. Ensure Coffalyser digitalMLPA is up-to-date. The latest version can be retrieved through your account on www.mrcholland.com/.
I received the following error message: *Sample (xxx) missing too many reference probes.*	This error is given when not enough reads have been detected for a large number of reference probes. Ensure the experiment has been performed according to the instructions in the digitalMLPA NXtec Protocol in order to obtain sufficient reads for all probes (e.g. DNA input amount should be sufficient, ligation should be successful and read depth adequate). Also ensure the correct Product Sheet has been selected while making the Coffalyser Definition File and that no-DNA samples are selected as type "No DNA".
I received the following error message: *The experiment does not have any analysable samples.*	This can happen when none of the samples in an experiment can be analysed e.g. only no-DNA reactions are present in an (undefined) experiment. Ensure the experiment has been performed according to the instructions in the digitalMLPA NXtec Protocol and check whether all your samples are correctly defined in the Coffalyser Definition File.
I received the following error message: *Not a single experiment defined in the experiment definition file qualified for analysis.*	This error message is often accompanied by other error messages that explain what has gone wrong that prevented analysis of the digitalMLPA experiment(s). Ensure the experiment has been performed according to the instruction in the digitalMLPA NXtec Protocol and check whether all your samples are correctly defined in the Coffalyser Definition File.
I received the following error message: *The process cannot access the file xxx because it is being used by another process.*	This can happen when Coffa files are reanalysed, while reports of these Coffa files are still open. Close the reports and start the reanalysis again.
I received the following error message: *The parent folder contains more than 100 subfolders, which will prevent Microsoft Windows from showing the full list in the following form.*	Restructure the parent folder (e.g. combine output folders per week or month) so less than 100 subfolders are present in the output folder specified (see step 25 of the procedure in section 3.2.1).
I received the following error message: *The average read length (xx) for the first 500000 reads is too short (minimum required xx).*	The read length in the FASTQ file is too short for reliable analysis. Please check the instructions in the digitalMLPA NXtec Protocol and the probemix-specific Product Description for the minimum required read length and rerun the digitalMLPA samples on the sequencer (ensuring read length is sufficient). Also check for sequencing issues that could have affected the read length.

Problem	Cause
I received the following error message: *The following errors have been detected while initialising the definition:* *- failed to retrieve the search graph blob (xxx) specified in the first encountered value (xxx).*	This can happen during reanalysis of Coffa files that were initially generated using another version of Coffalyser digitalMLPA. To solve this, the digitalMLPA data can be reanalysed by starting with the FASTQ file as input, see sections 3.2.1 and 3.2.2.1.
I received the following error message: *Detected a sample (xxx) with multiple qualified barcode batch identifiers (xxx).*	This can happen if two different lots of the same barcode plate are combined in one sample (e.g. Barcode Plate 1 lot 03-007-xxxxxx and Barcode Plate 1 lot 03-009-xxxxxx). The three-digit number between dashes (e.g. -008-) indicates the barcode plate lot. Ensure you are using one barcode plate lot per sample.

RATIOS EXCEL FILE

Problem	Cause
The *Reference population statistics* in the *Reference sample quality* sheet is not completed in the Ratios Excel file.	This can happen when the first sample (lowest barcode number) that is analysed did not undergo comparative analysis because of an error, for example when a no-DNA sample is not indicated as such in the Coffalyser Definition File. To obtain the completed *Reference population statistics* in the *Reference sample quality* sheet, repeat the analysis of the experiment and ensure sample types are set correctly and/or omit the first sample that caused the error.
There is a *Reference sample quality* warning or error in the *Quality table*, but I cannot find the cause using the *Reference sample quality* sheet in the Ratios Excel file.	The data displayed in the *Reference population statistics* in the *Reference sample quality* sheet is not directly suitable for the troubleshooting of X- (non-PAR) and Y-chromosome probes as it only shows the statistics for male OR female samples rather than displaying separately the statistics for both male AND female samples by default. The values printed in the *Reference population statistics* for these probes are based on the gender of the first DNA sample (lowest barcode number) that is analysed. For more information, please refer to our website.

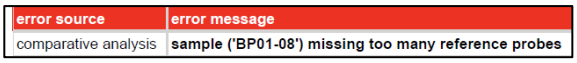
SAMPLE RESULTS 1 REPORT

Problem	Cause
Sample Results 1 report is not generated for a specific sample.	The cause of the error is described in the General Details Report of this particular sample.
A different gender than specified in the Coffalyser Definition File was detected by Coffalyser digitalMLPA (as displayed in the Overview file or the Sample Results 1 report).	This can occur in case of a sample swap or if a sample was incorrectly defined in the Coffalyser Definition File. It may also have a biological cause. For example, in older men, cells may lose the Y chromosome causing the sample to be detected as female.
Sequences for only two out of three control probes targeting the Y chromosome are detected in a male sample, as displayed in the *Probes* table under "Y-control probes".	This can be caused by a SNP at or around the ligation site of one of the control probes targeting the Y chromosome.

Problem	Cause
Sequences for one of the control probes targeting the Y chromosome are detected in a female sample, as displayed in the *Probes* table under "Y-control probes".	This could be due to a translocation of part of the Y chromosome onto an autosomal or X chromosome. Another possible explanation is sample contamination.
Sequences for only one of the control probes targeting the Y chromosome are detected in an SD sample provided by MRC Holland.	Some SD samples contain only a part of the Y chromosome as additional DNA is added to female DNA. Details of the SD can be found in the product-specific Product Description.
A warning is given for one control probe type (e.g. denaturation probes) in the comparative analysis, but not in the reaction analysis.	There is a difference (in quality, in the amount of contaminants present etc.) between this sample and the reference population. This can occur for example if another DNA extraction method was used for test samples compared to reference samples. For more information on the control probe warnings, please see APPENDIX II – QUALITY TABLE and the digitalMLPA NXtec Protocol.
A warning is given for one control probe type (e.g. denaturation probes) in the reaction analysis, but not in the comparative analysis.	This indicates that the quality of both the test sample and the reference population may be compromised. However, there is no difference (in quality, in the amount of contaminants present etc.) between this sample and the reference samples. For more information on the control probe warnings, please see APPENDIX II – QUALITY TABLE and the digitalMLPA NXtec Protocol.
A warning is given for reference sample quality.	When reference samples have been designated: 1. This can be caused by copy number variations in one or more of the reference samples. 2. This can be caused by a high standard deviation of one or multiple probes over the reference population. When no samples have been designated as reference samples: 1. This can be caused by the test sample population being too small, resulting in higher probe standard deviations. 2. This can be caused by inclusion of many samples with the same aberration. More information about this warning can be found in the Reference sample quality sheet in the Ratios Excel file, see section 4.2.3. Please note that due to a limitation in the software, the data displayed for the X- (non-PAR) and Y-chromosome probes cannot be directly used for troubleshooting. See APPENDIX III – TROUBLESHOOTING section RATIOS EXCEL FILE. It may be that one single reference sample is aberrant; in that case you might consider re-analysing the experiment without that specific sample being designated as a reference sample (provided that three or more reference samples remain). For more information, please also refer to APPENDIX II – QUALITY TABLE and the digitalMLPA NXtec Protocol.

Problem	Cause
The *Reads* and *Probes* sections overlap with the *Chart* in the Result Summary section of the Sample Results 1.pdf.	There is a problem with the report layout. To prevent this, we strongly recommend the use of short names for the reference samples.

GENERAL DETAILS REPORT

Problem	Cause and solution
The following error was reported: *Error source: Reaction analysis. Error message: sample's read count is too low.*  Error source: Reaction analysis Error message: sample's read count is too low	This error is given if no reads are detected for a barcode that was selected and included in the Coffalyser Definition File. When defining experiments in the Coffalyser Definition File, ensure you only select the barcodes that are present in the data. Also ensure the experiment has been performed according to the instruction in the digitalMLPA NXtec Protocol, and both barcode solution and probemix are added to a sample.
The following error was reported: *Error source: comparative analysis. Error message: sample x missing too many reference probes.*  error source: comparative analysis error message: sample ('BP01-08') missing too many reference probes	This error is given when not enough reads have been detected for a large number of reference probes, as this does not allow for reliable comparative analysis. This may happen when for example a no-DNA sample was not specified as such in the Coffalyser Definition File. Ensure the experiment has been performed according to the instructions in the digitalMLPA NXtec Protocol in order to obtain sufficient reads for all probes (e.g. DNA input amount should be sufficient, ligation should be successful and read depth adequate). Also ensure the correct Product Sheet is selected while making the Coffalyser Definition File and no-DNA samples are selected as type "No DNA".
The following error was reported: *Error source: comparative analysis. Error message: the experiment does not have any analysable samples.*	This can happen when none of the samples in an experiment can be analysed e.g. only no-DNA reactions are present in an (undefined) experiment. Ensure the experiment has been performed according to the instructions in the digitalMLPA NXtec Protocol and check whether all your samples are correctly defined in the Coffalyser Definition File.

APPENDIX IV – COMMAND LINE INTERFACE (CLI)

For easy integration in your sequencing pipeline, Coffalyser digitalMLPA modules can be accessed through the Command Line Interface (CLI): MRC.DataAnalysis.{module name}.Client.exe using the Windows Command Prompt. Executing any of these CLI console applications without any parameters invokes the help function and displays an example call and a brief overview of the parameters (see screen capture below for an example).

```
*** EXAMPLE USAGE ***


       MRC.DataAnalysis.Analyses.Client.exe [-Definition] "C:\Path\" [-AnalysisType] (FragmentAnalysis|ComparativeAnalysis|ExtendedAnalysis|All)

MANDATORY PARAMETERS:

       -Definition   : path to the definition directory or file;
       -AnalysisType : the analysis type (FragmentAnalysis and/or ComparativeAnalysis and/or ExtendedAnalysis ... or just All);

c:\Digital Coffalyser\Program\sema4 [20210921-162643]>
```

CLI ANALYSIS WITH COFFALYSER DEFINITION FILE

Coffalyser digitalMLPA may be called with a Coffalyser Definition File (*.cdf) where parameters have been specified. Coffalyser digitalMLPA comes with a Definition File Editor for convenient creation of *.cdf files: MRC.DataAnalysis.Definitions.Editor.exe. You can either generate a Coffalyser Definition File for each analysis with the Definition File Editor, or use the Definition File Editor generated Coffalyser Definition Files as examples/templates and generate Coffalyser Definition Files from your experiment input files in your own pipeline.

COFFALYSER DEFINITION FILE HEADER

The Coffalyser Definition File header layout is as follows:

```
DefinitionVersion⟶⟶{version number (int)|CRLF
Technology⟶⟶⟶⟶DigitalMlpaCRLF
AnalysisSettings⟶⟶⟶{file id (guid)}CRLF
SampleBarcodes⟶⟶⟶{file id (guid)}CRLF
PathData⟶⟶⟶⟶{path to output directory, e.g.: "N:\NGS output\CoffalyserNet\"}CRLF
PathTemp⟶⟶⟶⟶{path to directory for temp files, e.g.: "C:\temp"}CRLF
PathsInputConversion⟶{path to input GASTQ(.gz) file(s)}CRLF
UndefinedSampleAction⟶{setting regarding undefined samples, e.g.: "Exclude"}ExcludeCRLF
```

Tabs are represented as yellow arrows, spaces are represented as yellow periods, CRLF represents “enter” (carriage return, line feed), and text in {curly brackets} indicate an explanation of the field.

An example of a *.cdf file header:

```
DefinitionVersion⟶⟶2CRLF
Technology⟶⟶⟶⟶DigitalMlpaCRLF
AnalysisSettings⟶⟶⟶518d1a77-1d51-4c59-86b6-f95a6c1be30aCRLF
SampleBarcodes⟶⟶⟶c467ebc7-99bc-80ed-25f5-109097d77b97CRLF
PathData⟶⟶⟶⟶N:\AnalysisOutput\Coffalyser digitalMLPACRLF
PathTemp⟶⟶⟶⟶C:\Coffalyser digitalMLPA\TmpFilesCRLF
PathsInputConversion⟶\\networklocation\FASTQs\MiniSeq-RUN93.fastq.gzCRLF
UndefinedSampleAction⟶ExcludeCRLF
```

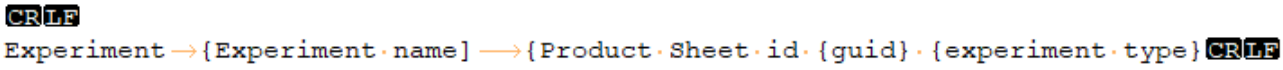
The file IDs of the AnalysisSettings and the SampleBarcodes may be found in the header of the AnalysisSettings and the SampleBarcodes files, which are located in the “_Configuration” sub-directory of your Coffalyser package. These Global Unique Identifiers (GUID) are filled in automatically when you create a Coffalyser Definition File in the editor. The PathData parameter specifies the location where Coffalyser digitalMLPA will create the output directory. This header line is optional; if the PathData is not specified, Coffalyser digitalMLPA will use the location of this Coffalyser Definition File as the output directory. The PathTemp parameter specifies a location for temporary working files (preferably on a local fast drive). The PathTemp line is optional; if not specified, Coffalyser digitalMLPA will use the PathData directory as the

location for the temporary working files. The UndefinedSampleAction line specifies how to treat any barcodes (samples) that were not defined in the *cdf file. The options are “Include” and “Exclude”, to respectively detect and include undefined samples in the analysis or ignore and exclude undefined samples from the analysis.

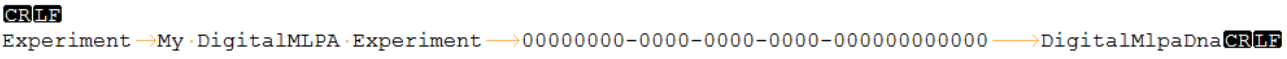
An analysis may be run with no further details in the Coffalyser Definition File. In that case, Coffalyser digitalMLPA will autodetect the NXtec probemix used in each sample and will create experiments by grouping all samples tested with the same NXtec probemix, provided that UndefinedSampleAction is set to Include. All samples will be treated as “Test samples”. In case your experiment contains samples of any other type (e.g. “No DNA”, “Positive Sample”, “Reference sample”, “SD”, “Undefined”) you must define these in the Coffalyser Definition File. MRC Holland recommends always explicitly defining experiment type, sample type, sample names, gender and if the sample is from a pooled DNA source to raise the likelihood of detecting mistakes. Not defining a sample from a pooled DNA source as such will lead to a DNA contamination error, failing the sample.

COFFALYSER DEFINITION FILE EXPERIMENT DEFINITION

Underneath the header in the Coffalyser Definition File the experiment definition is located. Each experiment section starts with a blank line, followed by an experiment header:

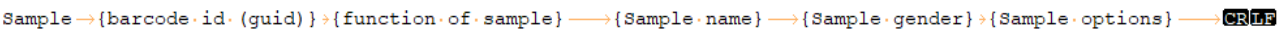


For example:

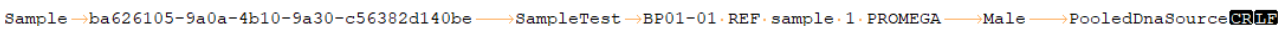


The experiment header line is single tab separated. The Product Sheet ID may be found in the header of the Product Sheet in the “_Configuration” sub-directory of your Coffalyser digitalMLPA package. If the Product Sheet ID is left to “00000000-0000-0000-0000-000000000000”, Coffalyser digitalMLPA will attempt to autodetect which Product Sheet to use. The Product Sheet id {GUID} can be found in the header of the Product Sheet .txt files which are located in the “_Configuration\Product Sheet” sub-directory of your Coffalyser digitalMLPA package.

The experiment header is followed by a line for each barcode in the experiment:



For example:



You may find the barcode IDs (GUID) in the SampleBarcodes file (for example "C:\Coffalyser\Coffalyser Package [20210922-145849]_Configuration\Sample Barcodes\Collection 3.txt").

The options for {function of sample} are: “NoDna”, “SampleReference”, “SampleTest”, “SampleReferencePositive”, “SampleDna” and “Undefined”.

The options for {Sample gender} are: “Ambiguous”, “Female” and “Male”.

The options for {Sample options} are: “PooledDnaSource” and “Default”.

RUNNING COFFALYSER DEFINITION FILE IN CLI

Once you have created your *.cdf file, you may call the DataConversion in the CLI using the following lines:

```
MRC.DataAnalysis.DataConversion.Client.exe -Definition
"C:\Path\MyDefinitionFile.cdf"
```

To call the Sample Analysis (including FragmentAnalysis for quality analysis of each sample and ComparativeAnalysis for comparison with the references) use the following command in the CLI:

```
MRC.DataAnalysis.Analyses.Client.exe -Definition "C:\Path\MyDefinitionFile.cdf"
-AnalysisType FragmentAnalysis,ComparativeAnalysis
```

Alternatively, instead of writing “FragmentAnalysis,ComparativeAnalysis” one can also write “all” to run all analyses modules.

And finally call Reporting (to generate the PDF reports as well as the Excel file) using the following command in the CLI:

```
MRC.DataAnalysis.Reporting.Client.exe -Definition "C:\Path\MyDefinitionFile.cdf"
```

CLI ANALYSIS WITHOUT A COFFALYSER DEFINITION FILE

The DataConversion module may be called without first making a Coffalyser Definition File. To do that, type into the CLI the following:

```
MRC.DataAnalysis.DataConversion.Client.exe -Technology digitalMLPA -
PathsInputConversion "C:\Path1\File1.fastq.gz" -PathData "C:\Path" -PathTemp
"C:\Path" -AnalysisSettings "Profile Name" -SampleBarcodes "Barcode Lot Name"
```

The –Technology parameter must be specified as “digitalMLPA” for digitalMLPA experiments.

The –PathsInputConversion parameter points to the FASTQ file(s) (multiple FastQ files can be selected simultaneously by listing them space separated).

The –PathData parameter points to the output directory.

The –PathTemp parameter points to a directory for the temporary files (optional, if not specified Coffalyser will use the PathData output directory).

The –AnalysisSettings parameter specifies which AnalysisSetting file to use (optional; if not specified, Coffalyser digitalMLPA will use the default from the configuration). The name of the AnalysisSettings can be found in the header of the files in the “Configuration\Analysis Settings” sub-directory of your Coffalyser digitalMLPA package.

The –SampleBarcodes parameter specifies which barcode collection file to use (optional; if not specified Coffalyser digitalMLPA will use the default from the configuration). The name of the SampleBarcodes can be found in the header of the files in the “_Configuration\Sample Barcodes” sub-directory of your Coffalyser digitalMLPA package.

The output of this conversion are Coffa files (*.coffa). The Coffa files are written to the output directory defined by the –PathData parameter. From here the Coffa files may be split up per experiment and distributed to the experiment owners for analysis.

The Coffa files may be analysed as they are. In this case, Coffalyser digitalMLPA will automatically create experiments by grouping all samples with the same NXtec probemix together. All samples will be treated as test samples. In case your experiment contains samples of any other type (e.g. “No Dna”, “Positive Sample”, “Reference sample”, “SD”, “Undefined”) you must define these in a Coffalyser Definition File. MRC Holland recommends always explicitly defining experiment type, sample type, sample names, gender and if the sample is from a pooled DNA source to raise the likelihood of detecting mistakes. You can set the sample type in the Coffalyser digitalMLPA configuration editor ("MRC.DataAnalysis.Definitions.Editor.exe") and by editing the *.cdf files.

If manual grouping of Coffa files is preferred, group Coffa files together in individual subfolders per experiment. Subsequently, the Sample Analysis may be called using the following command in the CLI:

```
MRC.DataAnalysis.Analyses.Client.exe -Definition "C:\Path1\" -AnalysisType
FragmentAnalysis,ComparativeAnalysis
```

The –Definition parameter should point to an experiment directory with Coffa files.

After finishing the analysis, reports can be generated. The Reporting module may be called without a Coffalyser Definition File using the following command in the CLI:

```
MRC.DataAnalysis.Reporting.Client.exe -Definition "C:\Path1\"
```

The –Definition parameter should point to a directory with Coffa files that have already undergone Sample Analysis.

The PDF reports and the Ratios Excel file are automatically generated in the directory containing the Coffa files.

Coffalyser digitalMLPA User Manual – Document History

Version 006 (10 October 2024)

- Configuration update: software version 251008.1227 added on the front page.
- Product names were updated throughout the document from SALSA® digitalMLPA™ to NXtec.
- The technique name was updated throughout the document from SALSA® digitalMLPA™ to digitalMLPA.
- The trademark sentence on page 4 was updated with NXtec.
- In section 3.2.1, step 6, instructions on how to select the barcode plate lot collection were updated.
- In section 3.2.1, a sentence was added to step 12 and 13 on what to do in case the used probemix (lot) is not available in the drop-down menu.
- In section 4.3.4, Table 1, the explanation for *Identity NOT detected – Error* was extended.

Version 005 (12 November 2024)

- Configuration update: software version 241105.1259 added on the front page.

Version 004 (29 July 2024)

- Configuration update: software version 240719.0943 added on the front page.
- Added a note on analysis of very large FASTQ files in chapter 2.
- Added a note on how to open the Experiment Overview window to section 3.2.2.1.
- Added information about interpretation of data in samples with passed or failed quality checks to section 4.1 and section 4.3.1.
- Added information about the validation of the DNA quantity estimation algorithm to section 4.2.2.
- Added a sentence stating that digitalMLPA data is assumed to be normally distributed to Appendix I Normalisation.
- Adjusted the footnote about intra ratios available in the Ratios Excel file in Appendix I Normalisation.
- Adjusted information on analysis without defined reference samples in Appendix I Normalisation.
- Added high amplifications as a possible trigger for RPQ warnings or errors in Appendix II Quality table.
- Removed the problem "*Failed to retrieve a sample's analyses fragment results collection (sample xxx).*" from Appendix III Troubleshooting, as it is no longer present in the software.
- Added the problem "*Error source: reaction analysis. Error message: sample's read count is too low."* to Appendix III Troubleshooting.
- Added a note on not defining a sample from a pooled DNA source as such in Appendix IV Command Line Interface (CLI).
- Minor textual and lay-out changes.

Version 003 (1 November 2022)

- Configuration update: software version 221020.1234 added on the front page.
- Added information to section 1 Glossary and Appendix II Quality table about the DNA contamination check not being performed for all probemixes.
- Notes added to section 3.2.1 Create new experiment, section 4.3.5 General information – analysis details and Appendix III Troubleshooting about avoiding layout problems in Sample Results 1.pdf.
- Added information about data in Reference sample quality sheet not being correct for X- (non-PAR) and Y-chromosome probes to section 4.2.3 Reference sample quality sheet and Appendix III Troubleshooting.

- Added incomplete Reference sample quality sheet problem to Appendix III Troubleshooting.
- Added Y chromosome loss as possible cause for detecting different gender in Appendix III Troubleshooting.

Version 002 (10 March 2022)

- Configuration update: software version 220221.1522 added on the front page.
- Paragraph added to Important notification section on the use of (registered) trademarks in this manual.

Version 001 (8 February 2022)

- First manual for software version 211020.1245.