VIVA-NUS CenTRAL
Centre for Translational Research in Acute Leukaemia (Molecular)
Department of Paediatrics, Yong Loo Lin School of Medicine

Standard Operating Procedures (SOP)

Title: Droplet Digital PCR (ddPCR) MRD Analysis Pipeline

Document Ref No.:	BP-05
Version No.:	1.1
Issue Date:	26 Jul 2021
Review Date:	02 Sep 2021

Reviewed by:
Name: Evelyn Lim

Approved by:
Name: A/Prof Allen Yeoh Eng Juh

1. BACKGROUND

Droplet Digital Polymerase Chain Reaction (ddPCR) was develop to provide high precision, absolute quantitation of nucleic acid target sequences. ddPCR measures absolute quantities by counting nucleic acid molecules encapsulated in discrete, volumetrically defined water-in-oil droplet partition. This platform may be used for MRD quantitation.

The use of an analysis pipeline would streamline and standardise the ddPCR MRD analysis. It serves to improve efficiency and reduce the error rates in data input. From the raw data format, the MRD levels could be calculated automatically and a run report would be generated for the sample. MRD levels

2. PURPOSE

This Standard Operating Procedure (SOP) provides an overview of the main steps involved in setting up the

3. SCOPE

This SOP is applicable to all CenTRAL (Molecular) Bioinformatics personnel involved in the implementation, maintenance and/or upgrading of the bioinformatics pipeline.

4. DEFINITIONS

4.1 **Immunoglobulin (IG) gene:** the gene that encodes immunoglobulin/antibody.

4.2 **T-cell receptor (TCR) genes:** the genes that encode T-cell receptors.

4.3 **Minimal Residual Disease (MRD)**: Monitor sub-microscopic minimal or traces of residual disease or in ALL.

4.4 **Ig/TCR rearrangements**: Specific rearrangements in the B-cells and T-cells resulting from V(D)J recombination.

4.5 **Mononuclear Cells (MNCs)**: Pooled mononuclear cells obtained from at least 5 healthy subjects.

4.6 **Droplet Digital Polymerase Chain Reaction (ddPCR):** a methodology that is based on water-oil emulsion droplet technology. A sample is fractionated into

maximally 20,000 droplets, and PCR amplification of the template molecules occurs in each individual droplet.

4.7 **Clusters:** The droplets in each cluster are similar to each other. They are formed closely enough to be distinguished clearly from other clusters.

4.8 **Two K-means clustering:** It is an iterative clustering that aims to find local maxima in each iteration. When there is no further switching of data points for two successive repeats, it will mark the termination of the algorithm.

5. RESPONSIBILITIES

5.1 Medical Director

5.1.1 Ensure the Bioinformatics Team Leader is qualified, trained and competent to design and set up the bioinformatics pipeline for the purpose of this SOP.

5.1.2 Ensure adequate resources for Bioinformatics Team Leader to perform the work tasks in this SOP.

5.1.3 Ensure Bioinformatics team are compliant with good bioinformatics practices and any applicable regulatory requirements.

5.1.4 Ensure Bioinformatics Team Leader provides the relevant information and databases supported by clinical evidence from literature. This knowledge will be used to provide the necessary information to build the pipeline.

5.1.5 Provides clinical insights that are relevant and useful in setting up the bioinformatics pipeline to achieve the desired results.

5.1.6 Ensure the bioinformatics pipeline generates robust and accurate results.

5.1.7 Approve the bioinformatics pipeline.

5.1.8 Shall maintain the ultimate responsibility for the result analysis generated.

5.1.9 Review and approve Quality Improvement Plan when required.

5.2 Bioinformatics Team Leader

5.2.1 To conduct adequate training and ensure the competency of the Bioinformatics personnel involved in using this SOP.

5.2.2 Ensure Bioinformatics personnel is compliant with good bioinformatics practices and any applicable regulatory requirements.
5.2.3 Monitor, document and implement patch-releases, upgrades and other updates to the bioinformatics pipeline where appropriate.
5.2.4 Address any bioinformatics procedure or any arising issue raised by Staff.
5.2.5 Review and monitor the progress of the corrective and preventive action taken where applicable.
5.2.6 Review workflow processes and implement Quality Improvement Plan if required.
5.2.7 Verify the quality control checks of the bioinformatics pipeline and result analysis.
5.2.8 Ensure bioinformatics pipeline generate data analysis results within the stipulated turnaround time (+ 1-2 days deviation is acceptable).
5.2.9 May designate selected duties and/or responsibilities to Laboratory Staff to assist in recording and documenting all relevant information, including version traceability of bioinformatics algorithms and software tools.

5.3 Designated Laboratory Staff

5.3.1 Ensure adequate training is given; and have passed work competency assessment to perform the assigned tasks.
5.3.2 Inform laboratory management on the resources required in order to perform the task where applicable.
5.3.3 Understand and follow the instructions in this SOP including fulfilling the Quality Control and any applicable regulatory requirements.
5.3.4 Ensure bioinformatics pipeline is robust and generates accurate results.
5.3.5 Raise any bioinformatics procedure, quality control measure, or any arising issue with the Bioinformatics Team Leader or Laboratory Management; resolve together where applicable.
5.3.6 Submit timely analysed results for verification by the Bioinformatics Team Leader and Medical Director.
5.3.7 Ensure generated result analysis report meet the expected turnaround time.
5.3.8 Proper recording and documenting of all relevant bioinformatics information and results including version no. traceability records etc.

5.3.9 Maintain good documentation practices.

6. EQUIPMENT, SOFTWARE AND RESOURCES

6.1 Equipment

Item	Brand
Workstation with Windows OS	Any

6.2 Softwares

Tools	Purpose & Availability
R	*R* is a free *software* environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.
MiKTex	MiKTeX is a free distribution of the TeX/LaTeX typesetting system and contains related programmes. It has tools which prepares documents using the TeX/LaTeX markup language and a TeX editor.

7. PROCEDURES

7.1 Installation and Set-up

7.1.1 Download and install R software.
7.1.2 Download and install MiKTeX.
7.1.3 Run the R software as administrator.
7.1.4 Run the following commands on R:
Install.packages(“rmarkdown”)
if(!require(installr)) { install.packages("installr"); require(installr)}
install.pandoc()
7.1.5 Restart the computer as instructed by pandoc

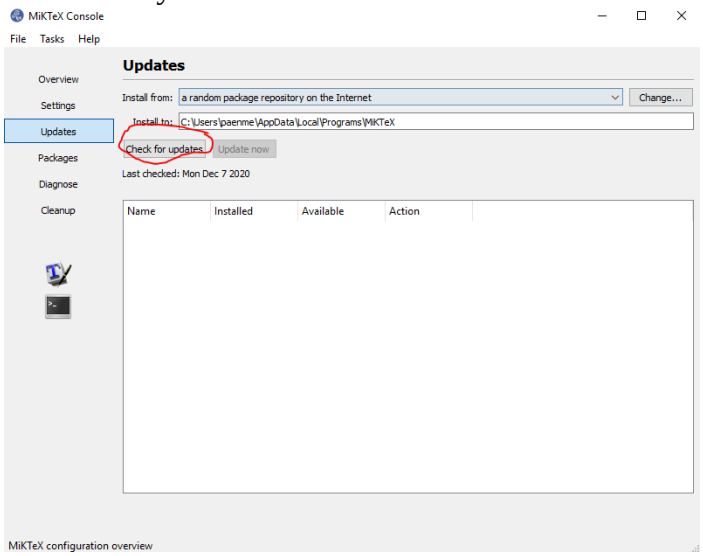
7.1.6 Download the pipeline files, and configure

7.1.7 When running for the first time, the program may ask to install some .sty file needed by TeX. Install the needed files.

7.1.8 Go to Windows, search for MiKTex console.

7.1.8.1 In settings, click check for updates. When the packages are loaded, click “update now”

7.1.8.2 In settings, check the box “always install missing packages on-the-fly”.



7.1.9 Go to Windows, search for MiKTex console.

7.1.10 Copy the pipeline files from ddPCR.

7.1.11 Create a “Export” folder to exporting the raw data files and report generation.

7.1.12 Change the directory to the newly created folder.

7.2 Sample Sheet

7.2.1 Create a excel sample sheet in the “Export” folder.

7.2.2 Label the reaction wells in 96-well format, as according to the following:

Reaction well	Run Type	Labeling
Diagnostic Sample	MRD Marker	Dx_Mk_<Patient ID>_<Marker>_<Input Amount>
Follow-up Sample		FU_Mk_<Patient ID>_<Marker>_<Input Amount>
MNC		MNC_Mk_<Patient ID>_<Marker>_<Input Amount>
NTC		H2O_Mk_<Patient ID>_<Marker Name>_0
Diagnostic Sample	Albumin	Dx_ALB_<Patient ID>_<Input Amount>
Follow-up Sample		FU_ALB_<Patient ID>_200x
MNC		MNC_ALB _200x_MNC<Batch Number>
NTC		H2O_ALB _0
Empty	Both	EMPTY

7.2.3 Save as .txt file.
7.2.4 Export the raw run files to the “export” folder.

7.3 Running the pipeline

7.3.1 Double click the silent mode.
7.3.2 Select sample sheet.
7.3.3 Enter the MNC dilution factor.
7.3.4 Enter the albumin concentration.
7.3.5 Enter the tumorload value.
7.3.6 The ddPCR run report will be generated in the “Export folder” created.
7.3.7 Check back on the generated report that the excluded outliers tallies correctly with the outliers on the 2D plot on Quantasoft software.
7.3.8 If the algorithm did not exclude the outliers (if any) correctly, rerun the algorithm interactive mode. In the interactive mode, preview the plot for each reaction well. Click to remove any outlying points (if any) or include points which were wrongly excluded (if any). Click the middle key when done to continue to the next plot. Do this for all the reaction wells. This step is not available in the silent mode.

7.4 Defining Positive Droplet

7.4.1 K-Means Clustering Algorithm was applied to the amplitude data of diagnostic DNA in order to classify each droplet as positive or negative.

7.4.2 The algorithm searches for partition of data space with locally optimal within-cluster sum of squared errors (SSE). Once the positive and negative clusters are defined, the cut-off was calculated as the average if the minimum amplitude of positive droplets and the maximum amplitude of negative ones (Hartigan and Wong, 1979).

7.4.3 The cut-off value was applied to the amplitude data of the other controls and samples for subsequent MRD calculation.

7.4.4 Outliers are excluded from analysis.

7.5 Merged Concentration

7.5.1 The merged concentration is calculated based on the concentration of individual wells using the following formula:
-ln [1-(Total positive events/Total events)]/0.00085

7.6 MRD Calculation

7.6.1 The absolute MRD of the follow-up sample is defined as the ratio of the concentration of the target gene against that of the Albumin gene:

$$MRD_{FU} = \frac{C_M}{D \times (C_{ALB}/2)} = \frac{C_M}{100 \times C_{ALB}}$$

where D is the dilution factor of the DNA for Albumin concentration (default: 200x).

7.6.2 The MRD level of the sample was then normalized as the MRD level of the diagnosis sample as follow:

$$MRD = \frac{MRD_{FU}}{MRD_{Dx}}$$

7.7 Result Interpretation

7.7.1 Each MRD result is interpreted by the algorithm as Positive, Negative, and Limit of quantitation. A sample classified into a result class based on binomial distribution.

7.7.2 Assuming the probability of being a blast was f, the cumulative probability (p) of presenting more than n blasts in a total of N cells could be expressed in the quantile function qbinom() of binomial distribution in R program:

$$n = qbinom((1-p), N, f)$$

7.7.3 The number of total cells N in a sample could be estimated from the quantity of ALB using the formula:

$$= D \times (C_{ALB}/2) \times V \times X = 2{,}000 \times C_{ALB} \times X$$

7.7.4 Estimation of the probability f from the MNC negative control was similar to MRD_{FU} calculation:

$$f = \frac{C_{M,MNC}}{D \times (C_{ALB,MNC}/2)} = \frac{C_{M,MNC}}{100 \times C_{ALB,MNC}}$$

7.7.5 Taken together, once N and f were known, n could be expressed as a function of p, where p=0.05 and p=0.0001 were used as the cut off for the negative and positive threshold values respectively:

$$MRD_p = \frac{n}{N} = \frac{qbinom((1-p), N, f)}{N}$$

8. QUALITY CONTROL

8.1 The algorithm will perform quality control checks as shown in the figure below. Some of the checks include the Minimum number of events, minimum amount of Follow-up concentrations, Assay controls for marker, Assay

controls for Albumin. When the criteria for all these are satisfactory, the overall QC result will be a "Pass".

Quality control

Table 5: Quality control summary

	QC.values	QC.status
Min events per well (>=8000)	9516	PASS
Min FU Alb conc. (copies/uL; >10000, GOOD; 2000-10000; ACCEPTABLE; <2000 POOR)	18379.458	GOOD
Assay controls for marker		
Positive (Dx marker conc., >=10 copies/uL)	64.557	PASS
NTC (H2O marker conc., <=0.1 copies/uL)	-0.000	PASS
Assay Controls for Alb		
Positive (MNC Alb conc., >=50 copies/uL)	69.785	PASS
NTC (H2O Alb conc., <=0.1)	-0.000	PASS
Overall QC result		**PASS**

9. RECORD MANAGEMENT

9.1 Analysis Pipeline

9.1.1 The ddPCR analysis shall be versioned and documented as a whole in general.

9.1.2 Change in version of any components in the bioinformatics pipeline is required depending on the relevance and/or when new update is needed to perform the work tasks as stated in this SOP.

9.1.3 Record any bioinformatics pipeline deviation from SOP and/or failures during bioinformatics process in Non-Conformity Logsheet (Appendix 10.2).

9.2 Data Storage

9.2.1 Raw output files from the ddPCR reader are stored on secure nus ndrive folder.

9.2.2 The generated run report shall be retained for at least 10 years.

10. REFERENCES

10.1 Della Starza, I., Nunes, V., Cavalli, M., De Novi, L.A., Ilari, C., Apicella, V., Vitale, A., Testi, A.M., Del Giudice, I., Chiaretti, S., Foà, R. and Guarini, A. (2016), Comparative analysis between RQ-PCR and digital-droplet-PCR of

immunoglobulin/T-cell receptor gene rearrangements to monitor minimal residual disease in acute lymphoblastic leukaemia. Br J Haematol, 174: 541-549. https://doi.org/10.1111/bjh.14082

10.2 Irene Della Starza, Vittorio Nunes, Federica Lovisa, Daniela Silvestri, Marzia Cavalli, Andrea Garofalo, Mimma Campeggio, Lucia Anna De Novi, Carlotta Oggioni, Andrea Biondi, Anna Guarini, Maria Grazia Valsecchi, Valentino Conter, Giuseppe Basso, Robin Foà, Giovanni Cazzaniga; Digital-Droplet PCR, an Accurate Method for IG/TR PCR-MRD Stratification in Childhood Acute Lymphoblastic Leukemia. *Blood* 2018; 132 (Supplement 1): 1544. doi: https://doi.org/10.1182/blood-2018-99-117954

10.3 Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. J R Stat Soc 1979, Series C (Applied Statistics); 28(1): 100-108

11. APPENDICES

11.1 ddPCR MRD Analysis Pipeline Version Logsheet
11.2 Non conformity logsheet
11.3 R-Script

12. AMENDMENT and REVIEW HISTORY

Version	Date	Amendment/Review	Recorded by
V1.1	2 Sep 2021	7.3 Edit procedure on running the pipeline. Indicated access route of generated algorithm to be the "Export" folder. Added section 8 on quality control	Evelyn