miRkit: R Framework Analyzing miRNA PCR Array Data

Maria Tsagiopoulou
CERTH: Ethniko Kentro Ereunas & Technologikes Anaptyxes https://orcid.org/0000-0002-1653-0327
Anastasis Togkousidis
CERTH: Ethniko Kentro Ereunas & Technologikes Anaptyxes
Nikolaos Pechlivanis
CERTH: Ethniko Kentro Ereunas & Technologikes Anaptyxes
Maria Christina Maniou
CERTH: Ethniko Kentro Ereunas & Technologikes Anaptyxes
Aristea Batsali
University of Crete: Panepistemio Kretes
Angelos Matheakakis
University of Crete: Panepistemio Kretes
Charalampos Pontikoglou
University of Crete: Panepistemio Kretes
Fotis Psomopoulos (fpsom@certh.gr)
Institute of Applied Biosciences, Centre of Research and Technology Hellas, Thessaloniki, Greece.
https://orcid.org/0000-0002-0222-4273

Research note

Keywords: miRNA, qPCR, RT-PCR, PCR array, GO, KEGG

Posted Date: June 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-645992/v1

Abstract

Objective: The characterization of microRNAs (miRNA) in recent years is an important advance in the field of gene regulation. To this end, several approaches for miRNA expression analysis and various bioinformatics tools have been developed over the last few years. It is a common practice to analyze miRNA PCR Array data using the commercially available software, mostly due to its convenience and ease-of-use.

Results: In this work we present miRkit, an open source framework written in R, that allows for the comprehensive analysis of RT-PCR data, from the processing of raw data to a functional analysis of the produced results. The main goal of the proposed tool is to provide an assessment of the samples' quality, perform data normalization by endogenous and exogenous miRNAs, and facilitate differential and functional enrichment analysis. The tool offers fast execution times with low memory usage, and is freely available under a MIT license from https://github.com/BiodataAnalysisGroup/miRkitl. Overall, miRkit offers the full analysis from the raw RT-PCR data to functional analysis of targeted genes, and specifically designed to support the popular miScript miRNA PCR Array (Qiagen) technology.

Introduction

MicroRNAs are small non-coding RNA molecules with a critical role in gene expression regulation (1). They are implicated in mRNA post-transcriptional modulation in the cell as well as released into circulation and transferred to other target cells (2). For this reason, and beyond their key role in intracellular pathways (2), miRNAs have emerged as biomarkers in clinical medicine (3) and are thought to represent appealing novel therapeutic modalities (4). Also, the expression levels of miRNAs are known to be deregulated in diseases and malignancies (5, 6).

Various approaches have been used to profile the expression of miRNAs (4) such as RT-PCR arrays, microarrays, small RNA-seq (7). Quantitative real-time PCR (RT-PCR) assays are sensitive and specific in detecting and quantifying the expression of miRNAs in the human miRNA genome (miRNome) (4). Within this context, the commercially available human miRNome miScript miRNA PCR Array (Qiagen) can be used to profile the 1066 most abundantly expressed and best characterized miRNA sequences in the human miRNome, as annotated in miRBase Release 16 (www.miRBase.org).

Raw RT-PCR data are typically analyzed using the manufacturer's software. In fact, several open-source packages analyze RT-PCR data with the traditional Ct (threshold cycle) quantification approach (8), ignoring more sophisticated and publicly available methods to analyze the expression profiles. To assess the potential benefits of analyzing this data with an open source tool, we developed miRkit, a framework written in R, specific for miScript miRNA PCR Array (Qiagen) technology. The proposed toolset offers full analysis of the raw RT-PCR data, including quality control of the samples, normalization of endogenous and exogenous controls, differential expression analysis and functional analysis of targeted genes.

Main Text

Implementation

Input data

The main R script of the workflow reads the input data from three distinct files stored in a folder:

(1) Count table: This is the main data file, containing the different samples on columns and the measurement of each well on the rows. The proposed tool is applicable on miScript miRNA PCR Array (Qiagen) which contains 384 wells and examines 372 miRNAs, 12 controls. Specifically, each well of 372/384 contains a miScript Primer Assay for a miRNome or pathway/disease/functionally-related mature RNA. Moreover, 2 wells contain replicate C. elegans miR-39 miScript Primer Assays and can be used as an alternative normalizer for array data (Ce), 6 wells contain an assay for a different snoRNA/snRNA that can be used as a normalization control for the array data. Finally, there are two wells which contain replicate miRTC Primer Assays (RTC) and two wells that contain positive PCR controls (PPC).

(2) Metadata: This file includes a list of sample IDs and the corresponding group e.g. normal/tumor

(3) Annotation of miRNAs well: A file that links the information of the well with the examined miRNA.

Workflow

The framework is implemented into three distinct phases; (1) QC and normalization, (2) differential analysis and (3) functional analysis. Specifically:

(1) QC and normalization.

The first phase takes as input the initial count table. The quality control process examines the maximum percentage of not detected or not available values (NA's) in each column, as defined by the user. Moreover, the ratio between reverse transcription control (RTC) assay, which detects an artificial RNA template, and positive PCR controls (PPC), which monitor for PCR inhibitors, is calculated and a standard threshold is used to validate the reverse transcription efficiency.

The data normalization module includes the option of endogenous and exogenous miRNA approach. The output of this step is the normalized data matrix that includes the samples which passed the NA's criterion. Additionally, a visualization option is available, which allows to generate figures that are automatically stored within the analysis folder, and include an upset plot for the NA's distribution and boxplots with counts before and after the normalization.

(2) Differential analysis

This module is performed using the limma package in R (9). The output includes the differentially expressed miRNAs using a user-defined adjusted p-value as a threshold. Moreover, a hierarchical cluster analysis is performed at this stage and a corresponding heatmap is constructed and stored in the analysis directory.

(3) Functional analysis

The downstream analysis links the differentially expressed miRNAs with the regulated genes using the multiMIR package, which includes several databases such as mirtarbase (10), tarbase (11), diana_microt (12) etc for both predicted and vali-dated targets. Moreover, the targeted genes of the differentially expressed miRNAs are used for KEGG and Gene ontology (GO) enrichment analysis, as facilitated by the *enrichR* package(13). Finally, barplots that present the results of the enrichment analysis are stored in the analysis folder.

At the end of the entire process, a report file is automatically exported. The report contains information of the particular execution process, including the user-defined criteria, the rationale for the excluded samples, the overall time required for execution and the total memory usage.

Case study

The miRkit was applied on artificial data provided by Qiagen. The selected NAs percentage threshold was set to 10% and all samples passed the control of NAs and RTC. The data normalization step was performed using the endogenous miRNAs. The total execution time was 18 minutes and 51.02 seconds (Table 1) and the memory usage was 274.7 MB. Detailed instructions are available in the repository of the tool (https://github.com/BiodataAnalysisGroup/miRkit) as well as the sample input and the output presented above.

Table 1
Table with the execution times in each phase

Phase	Execution time
Quality control	10.42sec
Differential analysis	2.32sec
Functional analysis	18min 38.25sec

Qualitative comparison to HTqPCR tool

We used HTqPCR (14), a software toolbox for dealing with RT-PCR data in order to compare its functionalities with miRkit. The main differences are listed in Table 2. A clear advantage of miRkit is the automatic process of all phases. Moreover, the linking of the statistically significant miRNAs with public databases gives the user the opportunity to complete the functional analysis within the frame of miRkit.

Table 2
Comparison of functionalities offered by miRkit and HTqPCR

	miRkit	HTqPCR
Language	R	R
Input	Standard output from miScript miRNA PCR Array (Qiagen) technology/ Handles data from multiple plates	Data preprocessing is required/ Only single-plate data, consisting of either 96 or 384 wells
Usage	Automatic	Manual (users need to write their one code)
Quality Control	YES	YES
Data Filtering	YES Automatically excluding samples based on NAs and on RTC	NO No standard way implemented
Normalization	YES endogenous/ exogenous genes	YES scaling up the values or changing the total distribution of values.
Clustering	YES	YES
Differential Analysis	YES	YES
Link miRNAs with databases	YES (mirtarbase, tarbase, diana_microt, etc)	NO
Enrichment analysis of deregulated genes	YES (KEGG and GO)	NO

Overall, miRkit implementation supports a fast execution with low memory usage in order to (i) perform quality control of the samples and data normalization, (ii) identify significant differ-ences on the expression profiles of miRNAs, and (iii) link the significant miRNAs with the targeted genes and biological processes. In each step of the process, the tool produces also the relevant visual representations of the results.

Compared to the traditional commercial software for analyzing RT-PCR data, miRkit aims to become fully aligned to the FAIR principles (Findable, Accessible, Interoperable, Reusable) for Research Software (15). It is freely available on GitHub and is accompanied by detailed documentation and examples, in order to facilitate the reproducibility of the presented results. Our method provides a new perspective towards analyzing RT-PCR data. Also, it supports efficient data discovery using the gold standard approach of

limma analysis and linking the information with publicly available databases to extract the biological meaning.

Abbreviations

miRNA
microRNAs
RT-PCR
Reverse transcription polymerase chain reaction
miRNome
human miRNA genome
Ct
threshold cycle
RTC
reverse transcription control
PPC
positive PCR controls
NA
not available value
GO
Gene ontology
FAIR
Findable, Accessible, Interoperable, Reusable

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the GitHub repository, https://github.com/BiodataAnalysisGroup/miRkit

Competing interests

The authors declare that they have no competing interests.

Funding

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning 2014-2020" in the context of the project "microRNA from Bone Marrow Mesenchymal Stem Cell-derived exosomes and defective hematopoiesis in Myelodysplastic Syndromes" (MIS 5048511). This research has also been co–financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (GenOptics, project code: T2E1 K-00407).

Authors' contributions

MT designed the study and wrote the manuscript. AT, NP and MCM developed the tool, analyzed the data. AB and AM provided the data and reviewed the submitted version. CP supervised the study and reviewed the submitted version. FP designed, supervised the study, and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgements

None

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and material

The tool offers fast execution times with low memory usage, and is freely available under a MIT license from https://github.com/BiodataAnalysisGroup/miRkit. The miRkit was applied on artificial data provided by Qiagen.

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281–97.
2. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol (Lausanne). 2018;9:402.

3. Trang P, Weidhaas JB, Slack FJ. MicroRNAs as potential cancer therapeutics. Oncogene. 2008;27(Suppl 2):52-7.
4. Hydbring P, Badalian-Very G. Clinical applications of microRNAs. F1000Res. 2013;2:136.
5. Ha TY. MicroRNAs in Human Diseases: From Cancer to Cardiovascular Disease. Immune Netw. 2011;11(3):135–54.
6. Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal Transduct Target Ther. 2016;1:15004.
7. Git A, Dvinge H, Salmon-Divon M, Osborne M, Kutter C, Hadfield J, et al. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA. 2010;16(5):991–1006.
8. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2^(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostat Bioinforma Biomath. 2013;3(3):71–85.
9. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
10. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 2011;39(Database issue):D163-9.
11. Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: A comprehensive database of experimentally supported animal microRNA targets. RNA. 2006;12(2):192–7.
12. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res. 2013;41(Web Server issue):W169-73.
13. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.
14. Dvinge H, Bertone P. HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. Bioinformatics. 2009;25(24):3325–6.
15. Lamprecht A-L, Garcia L, Kuzak M, Martinez C, Arcila R, Del Pico M, Angel EDominguezD, van de Sandt V, Ison S, Martinez J, McQuilton PA, Valencia P, Harrow A, Psomopoulos J, Gelpi F, Hong JL,C. N., Goble, C. Towards FAIR principles for research software. Data Science. 2020;3:37–59.

Figures

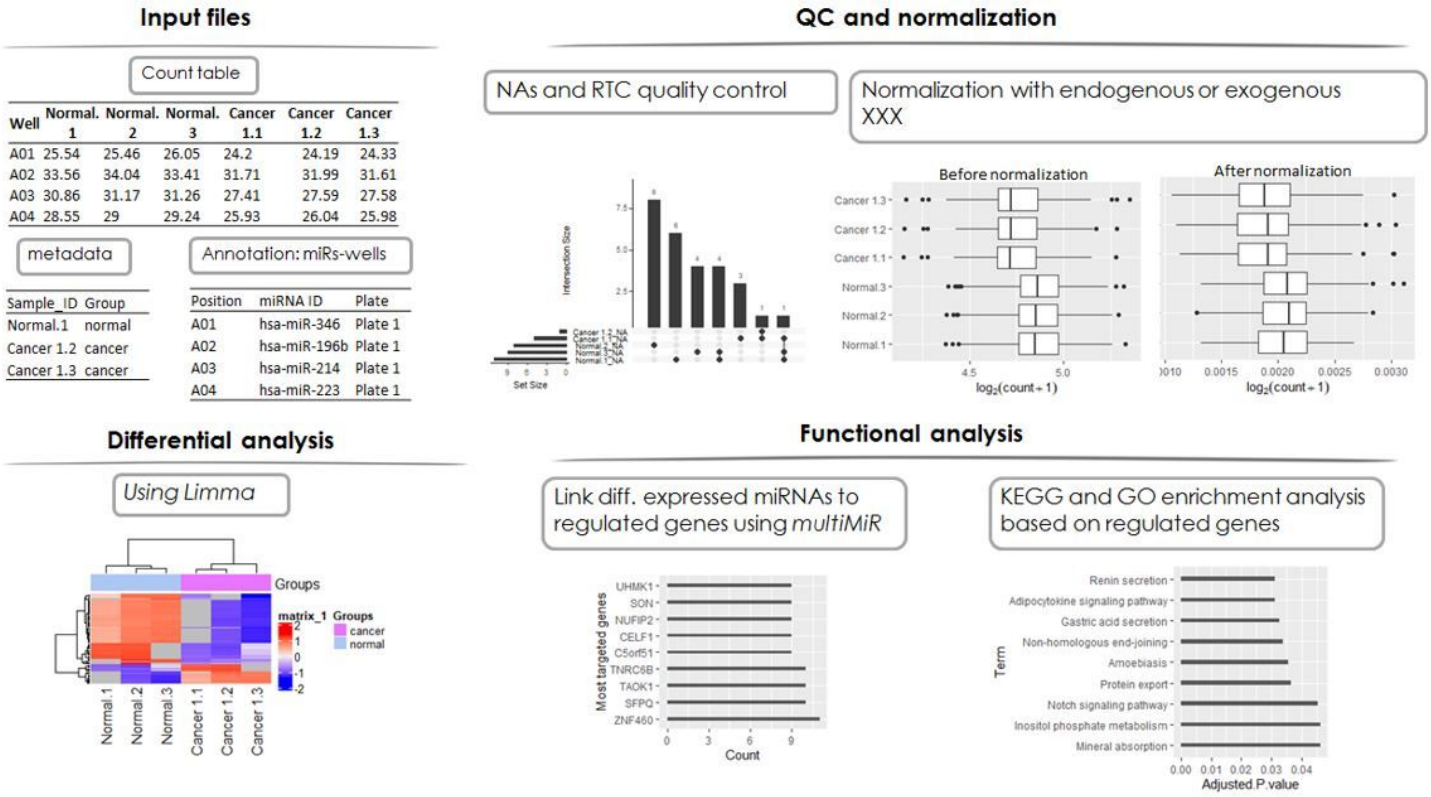

Figure 1

Graphical representation of working cases of the miRkit tool.