Bioinformatics, 2017, 1–3
doi: 10.1093/bioinformatics/btw839
Advance Access Publication Date: 5 January 2017
Applications Note

OXFORD

Gene expression

Chainy: an universal tool for standardized relative quantification in real-time PCR

Izaskun Mallona*, Anna Díez-Villanueva, Berta Martín and Miguel A. Peinado

Germans Trias i Pujol Health Science Research Institute (IGTP), Program for Personalized Medicine of Cancer, Badalona 08916, Catalonia, Spain

*To whom correspondence should be addressed.

Associate Editor: Ziv Bar-Joseph

Received on August 5, 2016; revised on December 15, 2016; editorial decision on December 29, 2016; accepted on December 31, 2016

Abstract

Summary: Chainy is a cross-platform web tool providing systematic pipelines and steady criteria to process real-time PCR data, including the calculation of efficiencies from raw data by kinetic methods, evaluation of the suitability of multiple references, standardized normalization using one or more references, and group-wise relative quantification statistical testing. We illustrate the utility of Chainy for differential expression and chromatin immunoprecipitation enrichment (ChIP-QPCR) analysis.

Availability and Implementation: Chainy is open source and freely available at http://maplab.cat/chainy

Contact: imallona@igtp.cat

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Real-time PCR has become the technique of choice in a wide range of quantitative applications including the analysis of gene expression and binding of proteins to chromatin, among others. In most settings the quantification of the target is referred to one or more references, the so-called normalization factor. The selection of a stable normalization factor is the key point to deliver reliable and accurate results (Bustin *et al.*, 2013). Surprisingly, it is not uncommon the arbitrary choice of one or more references without an appropriate appraisal of the normalization process, which has important drawbacks in the final results.

Nonetheless, references validation has become a publication trend. Validation papers evaluate a shortlist of reference genes and provide a ranked final panel. However, even though reference validation studies clearly provide good normalization factors for the assayed experimental conditions, it is impossible to say whether these results can be extrapolated to a different setting, specifically when conducted in a different lab. Despite the obvious advantages of evaluating the references stability as a part of the quantification workflow its inclusion is still marginal, partly because of the lack of open-source easy-to-use software.

To overcome these problems we have developed Chainy, an user-friendly tool in which the references validation is coupled to the relative quantification. Accepted inputs include raw fluorescence in multiple formats including the Real-time PCR Data Markup Language (RDML) standard and vendor-specific files for most popular thermocyclers, thus calculating efficiencies by kinetic methods (Guescini *et al.*, 2008; Liu and Saint, 2002; Schlereth *et al.*, 1998; Tichopad *et al.*, 2002); as well as semiprocessed data with already called Cq and efficiency values (Fig. 1).

Reference suitability analysis is conducted by the standard geNorm method (Vandesompele *et al.*, 2002) providing a stability ranking. The normalization factor with the minimum number of genes and the highest stability is also reported.

For differential quantification, Chainy computes fold changes between experimental groups while propagating errors during the whole procedure (Nordgård *et al.*, 2006). To statistically test whether the fold change are significant, the analyte quantities are taken from both groups (i.e. treatment and control) and randomly reallocated 100 times. For each randomization, a treatment versus control ratio is calculated and the number of shuffled ratios below

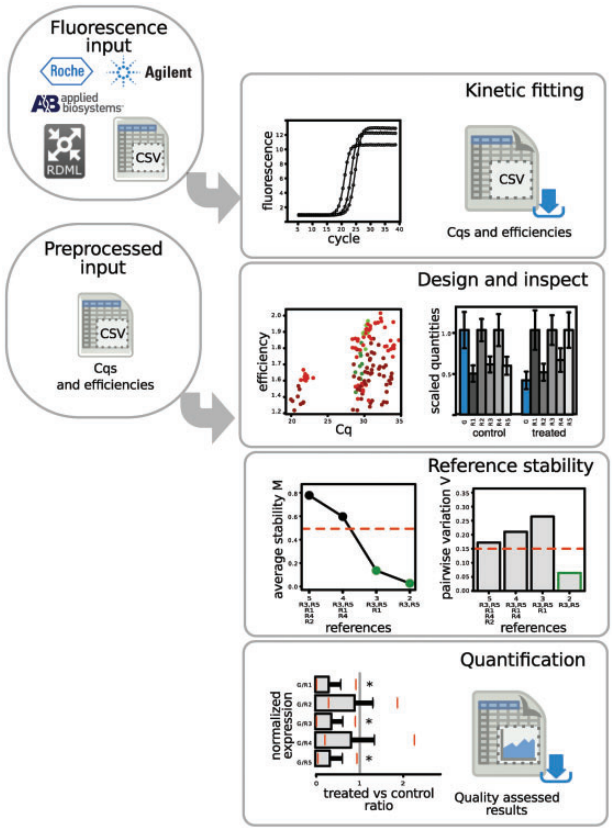

Fig. 1. Data flow. Chainy offers an integrative workflow with a graphical user interface. Chainy capabilities include efficiency and Cq calling from raw fluorescence, quality check, reference stability evaluation and statistically assessed relative quantification

or above the original ratio are reported as a p-value against the null hypothesis that the ratios calculated by permutation are just by chance (Pfaffl *et al.*, 2002). Following the MIQE guidelines (Bustin *et al.*, 2009), efficiencies are taken into account during the whole procedure and normalization against multiple reference genes is allowed.

Chainy produces reliable quantifications when validated with known amounts of template (supplemental results). A full description of the data flow is available as supplemental methods, including the analysis of three experiments. The online user manual covers results interpretation and troubleshooting. Chainy is written in R/shiny and its source can be accessed at https://bitbucket.org/imallona/chainy/ (including a script to launch the app locally).

The graphical user interface is built to maximize usage versatility (Fig. 1). The simplest application estimates PCR efficiency from raw fluorescence inputs. However, full capabilities are better depicted when used as a comprehensive workflow from raw data to differential quantification. To exemplify the whole process we have applied Chainy to two classic experimental settings: differential gene expression in response to a drug treatment and quantification of a histone modification by chromatin immunoprecipitation (ChIP).

Regarding gene expression relative quantification (supplemental results), Chainy evaluates the adequacy of the reference genes selected by the user, but it can also perform an agnostic analysis of both candidates and non-candidates. The assessed reference genes are ranked according to their stability, facilitating the decision making. As for ChIP enrichment analysis, we propose a quantification strategy (supplemental results) aware of its intrinsic technical challenges, namely that most ChIP-qPCR settings present several positive and negative controls which can be used as references. Commonly used controls include a negative or background control (IgG immunoprecipitate obtained without specific antibody) and the so called input representing the total load of chromatin. Additional positive and negative controls include the analysis of loci with known chromatin states, for instance highly transcribed regions enriched in active chromatin marks (i.e. the GAPDH gene) or non-transcribed regions enriched in repressive marks (i.e. pericentromeric regions).

The normalization strategy with such a complex setting is often underestimated in ChIP-QPCR. However, the integrative annotation of the ChIP controls taking into account both the immunoprecipitation and the target locus allows to evaluate the two possible normalization scenarios: cross-loci and intra-locus. The cross-loci normalization, much alike as gene expression analysis, compares the target quantities to an external control locus which is expected to present an invariant level of the QPCR readout. On the other hand, the intra-locus method compares the target binding to both the chromatin quantity (using the input as reference) or to the nucleosome abundance (using an antibody specific to an invariant part of histone H3) of the same locus. These two normalization methods may be not comparable for two reasons: first, technically, they are subjected to different experimental procedures; and second, biologically, the results interpretation heavily relies on the yardstick used (i.e. comparing to the H3 abundance of the same locus takes into account nucleosome density). As for gene expression analysis, however, Chainy evaluates the set of references and the appraises both scenarios and their effect on the final result, that may be essential to correctly interpret the data.

In summary, Chainy is an user-friendly tool for QPCR references validation and relative quantification, accepts either raw fluorescence (thus calculating efficiencies) or semiprocessed data, and does not require a deep knowledge of the quantification methods to be correctly applied. Specially suited to treated versus control experimental designs, implements state-of-the-art options allowing reliable, traceable and reproducible real-time PCR quantifications.

Acknowledgements

We thank the developers of the versatile qpcR, NormqPCR and RDML packages and Iñaki Martinez de Ilarduya for his excellent technical support.

Funding

This work was supported by the Spanish Ministry of Economy and Competitiveness [FEDER, SAF2011/23638 and SAF2015-64521-R to M.A.P.]. CERCA Programme/Generalitat de Catalunya.

Conflict of Interest: none declared.

References

Bustin,S. *et al.* (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.

Bustin,S. *et al.* (2013) The need for transparency and good practices in the qPCR literature. *Nat. Methods*, **10**, 1063–1067. Commentary.

Guescini,M. *et al.* (2008) A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. *BMC Bioinformatics*, **9**, 1.

Liu,W. and Saint,D.A. (2002) Validation of a quantitative method for real time PCR kinetics. *Biochem. Biophys. Res. Commun.*, **294**, 347–353.

Nordgård,O. *et al.* (2006) Error propagation in relative real-time reverse transcription polymerase chain reaction quantification models: the balance between accuracy and precision. *Anal. Biochem.*, **356**, 182–193.

Pfaffl,M. *et al.* (2002) Relative expression software tool (rest[copyright]) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.*, **30**, e36–e36.

Schlereth,W. *et al.* (1998) Use of the recursion formula of the gompertz function for the quantitation of PCR-amplified templates. *Int. J. Mol. Med.*, **1**, 463–470.

Tichopad,A. *et al.* (2002) Improving quantitative real-time rt-PCR reproducibility by boosting primer-linked amplification efficiency. *Biotechnol. Lett.*, **24**, 2053–2056.

Vandesompele,J. *et al.* (2002) Accurate normalization of real-time quantitative rt-pcr data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, 1–12.