BITS :: Call for Abstracts 2019 - Oral communication

Type	Oral communication
Session	(Multi-)Omics Data Integration and Analysis
Title	PIPE-T: A new Galaxy tool for analyzing RT-qPCR expression data
All Authors	Nicolò Zanardi(1), Martina Morini (1), Maria Carla Bosco (4), Varesio L(1), Eva A(1), Cangelosi D(1)
Affiliation	
(1) Laboratory of Molecular Biology, Istituto G. Gaslini, Genova, Italy	

Motivation

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is an accurate, sensitive, and fast method to measure gene expression. HTqPCR is a well-known R\Bioconductor package for the high-throughput analysis of RT-qPCR data. HTqPCR provides several functions and parameter options for quality assessment, data filtering, data normalization, differential expression, and visualization of RT-qPCR data. However, R-based analysis suffers of some known limitations. Users with biological background who want to exploit the functionalities of R based packages need non-trivial coding skills. Furthermore, the lack of a simple framework on reusing, sharing, and communicating experimental procedures and results limits the reproducibility, transparency, and accessibility of the R-based analysis. Galaxy is an open, collaborative, web-based, genomic workbench for a reproducible, transparent, and accessible science. Galaxy provides a very active developer community. More than 6000 public tools and workflows are freely installable from the Galaxy ToolShed repositories. Anybody can contribute with new tools and workflows. To this end, Galaxy provides the possibility of using fresh installations of R and Python, a Conda dependency resolver, a step-by-step documentation, a simple graphical interface, and GitHub integration. In the present work, we developed PIPE-T, a new tool for analyzing RT-qPCR expression data integrating the functionalities implemented in various R packages into one reusable, transparent, accessible, and easy to use Galaxy wrapper.

Methods

To develop PIPE-T we used R language and programming environment version 3.5.0 and Galaxy software system version 18.09. We deployed PIPE-T into the Galaxy main ToolShed repository allowing researchers to download and install our wrapper. PIPE-T requires the following R package dependencies to run: r-base, libgcc, bioconductor-htqpcr, bioconductor-rankprod, bioconductor-impute, r-bbmisc, r-psych, r-zoo, bioconductor-nondetects, and r-hmisc. To automatically resolve PIPE-T dependencies we used Conda package manager. For allowing users uploading their own input files, we exploit the functionalities of the “Upload file from your computer” Galaxy tool. PIPE-T code is freely available into GitHub.

Results

PIPE-T requires two input files to run and produces seven output files. Input files are: a List collection of tab-separated text files, for summarizing the resulting data of the RT-qPCR experiment, and a tab-separated text file, for associating each filename with a treatment group. For analyzing RT-qPCR data, we implemented the following computational procedures: File uploading and parsing, Ct filtering and categorization, Normalization, Transcript filtering and imputation, and Differential Expression Analysis.
File uploading and parsing procedure allows users to upload files into Galaxy and to check file format. PIPE-T accepts seven distinct file formats among the most used in the field including EDS, SDS, LightCycler, CFX, OpenArray, and BioMark file formats. Not all expression values are reliably measured in a RT-qPCR experiment. Ct filtering and categorization procedure assigns a category to all features among “Ok”, “Undetermined”, and “Unreliable” according to a user-specified range of values and/or on the basis of one of the failure experimental flags. Normalization procedure minimizes unwanted systematic technical and experimental variation in the data. PIPE-T offers six different normalization alternatives among the most used methods in the literature including Global mean, DeltaCt, Modified global mean, Quantile, Rank invariant, and Scale rank invariant.
To handle any missing value in the data, we developed the Transcript filtering and imputation procedure that filters out any transcript who exceeds a user-specified maximum number of missing values and imputes the remaining missing values with one of the most known imputation methods such as KNN, Cubic spline, Nondetects, Mean, and Median. Differential Expression Analysis procedure identifies candidate transcripts the expression of which discriminates between two predefined conditions. PIPE-T offers the possibility of selecting between two alternative methods: t-test and fold change, and Rank Product.
We tested PIPE-T on two example datasets the tab-separated text file of which were available in Gene Expression Omnibus repository with accession identifiers: GSE52847 and GSE43000. In both cases, our

tool successfully completed execution returning correct results.
In conclusion, we developed PIPE-T, a new Galaxy tool that offers several options for parsing, filtering, normalizing, imputing and analyzing RT-qPCR expression data. Integration of PIPE-T into Galaxy allows experimenters with strong bioinformatic background, as well as those without any programming expertise, to perform complex analysis in a simple to use, transparent, accessible, reproducible, and user-friendly environment.

Info
-
Figure
-

Availability	-

Corresponding Author	
Name, Surname	Davide, Cangelosi
Email	davidecangelosi@gaslini.org

Submitted on	29.04.2019