Gene expression

circtools—a one-stop software solution for circular RNA research

Tobias Jakobi1,2,*, Alexey Uvarovskii1,2 and Christoph Dieterich1,2,*

1Department of Cardiology, Angiology, and Pneumology, University Hospital Heidelberg and 2DZHK (German Centre for Cardiovascular Research), Partner Site Heidelberg/Mannheim, Heidelberg, Germany

*To whom correspondence should be addressed.

Abstract

Motivation: Circular RNAs (circRNAs) originate through back-splicing events from linear primary transcripts, are resistant to exonucleases, are not polyadenylated and have been shown to be highly specific for cell type and developmental stage. CircRNA detection starts from high-throughput sequencing data and is a multi-stage bioinformatics process yielding sets of potential circRNA candidates that require further analyses. While a number of tools for the prediction process already exist, publicly available analysis tools for further characterization are rare. Our work provides researchers with a harmonized workflow that covers different stages of in silico circRNA analyses, from prediction to first functional insights.

Results: Here, we present circtools, a modular, Python-based framework for computational circRNA analyses. The software includes modules for circRNA detection, internal sequence reconstruction, quality checking, statistical testing, screening for enrichment of RBP binding sites, differential exon RNase R resistance and circRNA-specific primer design. circtools supports researchers with visualization options and data export into commonly used formats.

Availability and implementation: circtools is available via https://github.com/dieterich-lab/circtools and http://circ.tools under GPLv3.0.

Contact: tobias.jakobi@uni-heidelberg.de or christoph.dieterich@uni-heidelberg.de

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Circular RNAs (circRNAs) were initially discovered in the 1990s, but the novel class of RNAs was first described as ‘scrambled exons’ (Nigro et al., 1991). Two decades later, new studies employing next generation sequencing discovered a large repertoire of circRNAs in different cell types and provided first hints of potential regulatory functions (Hansen et al., 2013). CircRNAs originate through back-splicing events from linear primary transcripts, are resistant to exonucleases, typically not polyadenylated and are highly specific for cell type and developmental stage. Detection of circRNAs is usually based on the existence of chimeric reads that cover the back splice junction (BSJ) of the circRNA, i.e. the position where the 3’ tail of the linear RNA molecule is fused with the 5’ head to form a circle. To increase the sensitivity for circRNAs in RNA-seq experiments, the CircleSeq protocol was developed based on the circRNAs resistance to exonuclease (RNase R) treatment (Jeck et al., 2013). While circRNA detection tools are available (Gao and Zhao, 2018), workflows for automated analyses including functionality, such as structural analyses or first functional predictions, are still rare.

2 Materials and methods

circtools currently offers seven modules shown in Figure 1A. Detection of circRNAs from RNA-seq reads mapped by STAR (Dobin et al., 2013) is based on the DCC software [detect, Cheng et al. (2016)] and generally the first step in the analysis workflow since the TSV-formatted data files generated here are required for subsequent analyses. Briefly, the detection step produces raw counts...
for circRNAs by exploiting reads that cover the BSJ and additionally generates count tables for the linear host genes. In a second step, these raw counts can be combined with log files of the STAR aligner within the quickcheck module. The diagnostic diagrams generated by the module also show the dramatic effect of the CircleSeq protocol on the number of detected circRNAs in HepG2 and K562 cells (Fig. 1B). Depending on the experimental setup and cell type, the initial detection step usually yields several hundreds to thousands of potential circRNA candidates. This set of candidates should be filtered for subsequent analysis steps. The circtools module allows for the identification of circRNA candidates that show a clear enrichment of selected sequence features (i.e. RBP sites), supports the reconstruction of circRNAs, tests for host gene-independent expression, and only requires a working R installation. Required Python and R packages are automatically installed. While much of circtools core functionality is implemented in Python, most plotting functions have been implemented in R. New modules can be easily added to the software by extending the circtools Python base class. We intend to add more functionality in the future in order to provide a comprehensive bioinformatics toolbox and we also encourage researchers to contribute modules to circtools.

3 Discussion
circtools provides a well-tested, harmonized workflow for state-of-the-art circRNA research. The software covers different aspects in this endeavor: It performs initial quality checks, detects and reconstructs circRNAs, tests for host gene-independent expression, screens for enriched sequence features (i.e. RBP sites), supports the design of primers for qRT-PCR verification and visualizes and exports all analyses results. A complete experimental workflow and detailed methods are described in the online documentation and the Supplementary Material.

Acknowledgements
The authors would like to thank Thiago Britto-Borges, Etienne Boileau and Brandon Malone for their great input and insightful discussions. They thank the Kurian Lab for cell culture experiments and total RNA isolation.

Fig. 1. Overview of the circtools software showing different output visualizations. (A) General workflow of circtools. (B) Initial quality check after read mapping and circRNA detection. (C) Relative enrichment of circN4BP2L2 in four samples. (D) Visualization of BED tracks produced by circtools; circRNA predictions (green), differentially spliced exons for K562/HepG2 cell lines (orange), reconstructed circRNAs for both cell lines (light green, purple). (E) Enrichment of different RBP binding sites within circN4BP2L2. (F) Visualization of a automatically designed primer pair (green) bracketing the back-splice junction (black line) that separates the two fused exons (orange, red).
Funding

The work was supported by the Klaus Tschira Stiftung gGmbH and the German Centre for Cardiovascular Research (DZHK) to T.J., A.U. and C.D. were supported by the Klaus Tschira Stiftung gGmbH and the German Centre for Cardiovascular Research (DZHK).

Conflict of Interest: none declared.

References


