CIRCprimerXL: Convenient and High-Throughput PCR Primer Design for Circular RNA Quantification

Marieke Vromman1,2, Jasper Anckaert1,2, Jo Vandesompele1,2* and Pieter-Jan Volders1,2

1OncoRNALab, Cancer Research Institute Ghent (CRIG), Ghent, Belgium, 2Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Circular RNA (circRNA) is a class of endogenous non-coding RNA characterized by a back-splice junction (BSJ). In general, large-scale circRNA BSJ detection is performed based on RNA sequencing data, followed by the selection and validation of circRNAs of interest using RT-qPCR with circRNA-specific PCR primers. Such a primer pair is convergent and functional on the circRNA template but divergent and non-functional on the linear host gene. Although a few circRNA primer design pipelines have been published, none of them offer large-scale, easy-to-use circRNA primer design. Other limitations are that these tools generally do not take into account assay specificity, secondary structures, and SNPs in the primer annealing regions. Furthermore, these tools are limited to circRNA primer design for humans (no other organisms possible), and no wet-lab validation is demonstrated. Here, we present CIRCprimerXL, a circRNA RT-qPCR assay design pipeline based on the primer design framework primerXL. CIRCprimerXL takes a circRNA BSJ position as input, and designs BSJ-spanning primers using Primer3. The user can choose to use the unspliced or spliced circRNA sequence as template. Prior to primer design, sequence regions with secondary structures and common SNPs are flagged. Next, the primers are filtered based on predicted specificity and the absence of secondary structures of the amplicon to select a suitable primer pair. Our tool is both available as a user-friendly web tool and as a stand-alone pipeline based on Docker and Nextflow, allowing users to run the pipeline on a wide range of computer infrastructures. The CIRCprimerXL Nextflow pipeline can be used to design circRNA primers for any species by providing the appropriate reference genome. The CIRCprimerXL web tool supports circRNA primer design for human, mouse, rat, zebrafish, Xenopus tropicalis, and C. elegans. The design process can easily be scaled up for the qPCR assay design of tens of thousands of circRNAs within a couple of hours. We show how CIRCprimerXL has been successfully used to design qPCR assays for over 15,000 human circRNAs of which 20 were empirically validated. CIRCprimerXL software, documentation, and test data can be found at: https://github.com/OncoRNALab/CIRCprimerXL. CIRCprimerXL is also implemented as a webtool at: https://circuitxltl.cmgg.be.

Keywords: RT-qPCR – real-time quantitative polymerase chain reaction, circular RNA (circRNA), primer design tool, back-splice junction, circRNA validation
1 INTRODUCTION

Circular RNA (circRNA) is a class of endogenous non-coding RNA that is widespread and abundant in many organisms. Although the exact function and working mechanism are lacking for the majority, some circRNAs are aberrantly expressed in pathological conditions such as cancer (Meng et al., 2017; Lux and Bullinger 2018). CircRNAs are formed through back-splicing, a non-canonical form of alternative splicing. Back-splicing results in a covalently closed loop characterized by a non-linear back-splice junction (BSJ) between a splice donor and an upstream splice acceptor (Figure 1A) (Salzman et al., 2012). The detection and quantification of circRNAs is based on the presence of their characteristic BSJ.

In general, large-scale circRNA detection is performed based on RNA sequencing data, using specific circRNA detection algorithms (Gao and Zhao 2018; Hansen et al., 2015; Zeng et al., 2017; Jakobi and Dieterich 2019). The circRNAs of interest are then confirmed by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) after ribonuclease R (RNase R) treatment (Vromman et al., 2021). RNase R is an enzyme that degrades linear RNA and thus reveals false-positive circRNA signals originating from endogenous trans-splicing (Vincent and Deutscher 2006; Pfafenrot and Preußer 2019). For RT-qPCR validation, circRNA-specific PCR primers are required (Figure 1B). Since these primers have a divergent orientation on the linear template, the linear host-gene will not be detected. Although a few circRNA primer design pipelines have been published, none of them offer large-scale, easy-to-use circRNA primer design (Dudekula et al., 2016; Jakobi, Uvarovskii, and Dieterich 2018; Panda and Gorospe 2018; Zhong et al., 2018). Other limitations are that these tools generally do not take into account assay specificity, spliced/unspliced nature of template, secondary structures, and SNPs in the primer annealing regions, and that no wet-lab validation is demonstrated. Here, we present CIRCprimerXL, a high-throughput and user-friendly circRNA qPCR assay design pipeline and web interface based on the primer design framework primerXL (Lefever et al., 2017).

2 METHODS

2.1 Implementation

The CIRCprimerXL scripts are written in Python (version 3.8.10). The pipeline is written using Nextflow (version 20.10.0.5430) and runs in a Docker container (version 20.10.11).

CIRCprimerXL is also implemented as a web application. The backend of the web application is built with Python (version 3.8-slim-buster) and the Flask framework (version 2.0.2), while the frontend is based on Bootstrap (version 5.1.). CIRCprimerXL is therefore compatible with the following web browsers: Chrome, Opera, Microsoft Edge, Firefox, and Safari.

2.2 CIRCprimerXL Workflow

A general overview of the pipeline can be found in Figure 1C.

STEP 1: The sequence around the BSJ is retrieved from a reference genome (default: homo sapiens) using fastahack (https://github.com/ekg/fastahack, RRID: SCR_016090, version 1.0). If both the start and end position of the circRNA are known splice sites or overlap with a known exon, the introns are removed from the template sequence based on an exon annotation file (default: Ensembl GTF, version GRCh38.104). If a list of canonical transcripts and/or transcripts of interest is provided (format: ENST transcript ids), it is used for transcript selection (default: RefSeq canonical transcripts, version GRCh38. v0.95). Alternatively, if the start and/or end position of the circRNA are intronic, exon-intron boundaries and intronic regions are included in the template sequence. By default, 150 nucleotides on both sides of the BSJ are concatenated to form the 300-nucleotide template sequence. If a circRNA is smaller than the requested template size, the template size is reduced to the circRNA size.

STEP 2: All common SNPs present in the template sequence are retrieved and flagged for exclusion (default SNP database: http://hgdownload.soe.ucsc.edu/gbdb/bg38/snp/dbSnp153Common.bb). The template sequence is screened for secondary structures using NUPACK (version 4.0.0.23) (Fornace, Porubsky, and Pierce 2020). All nucleotides involved in a secondary structure are flagged.

STEP 3: Twenty (by default) primer pairs are designed for each circRNA using Primer3 RRID:SCR_003139, version 2.5.0) (Untergasser et al., 2012). For this design, the Primer3 settings previously optimized in primerXL are used (Lefever et al., 2017). The most important parameters are primer length (nucleotides) = 16-20-30 (min-opt-max); primer Tm: 58-59-60 (min-opt-max); max Tm difference: 2°C; primer GC %: 30-50-80 (min-opt-max); concentration of monovalent cations: 50 mM; concentration of divalent cations: 3 mM; amplicon product size: 50-250 nucleotides. Furthermore, the nucleotides flagged for SNPs and secondary structures from STEP 2 are avoided.

STEP 4: Each primer pair is then screened for off-target activity by mapping the primer pair to a reference transcriptome (default: human cDNA and ncRNA) using bowtie (RRID:SCR_005476, version 1.3.0) (Langmead et al., 2009). When a match is found in the reference transcriptome, the primer pair is discarded, unless there are multiple mismatches present in both primers (Table 1) (Lefever et al., 2013). Simultaneously, the amplicon is screened for possible secondary structure formation using NUPACK (Fornace, Porubsky, and Pierce 2020), as this can prevent efficient amplification (Hoebeeck et al., 2005). If the amplicon has a secondary structure with a deltaG (Gibbs free energy) lower than −15 kcal/mol, the primer pair is rejected. If there is a secondary structure present with a deltaG between −15 kcal/mol and −5 kcal/mol, but the
secondary structure is not in the primer sequence, or the delta G is higher than $-5$ kcal/mol, the primer pair is not rejected. The list of filtered primers is then sorted by increasing amplicon size. The selected primer pair is the one with the smallest amplicon size, as it has a higher chance of amplifying the BSJ sequence (smaller chance of priming to introns or neighboring exons).

**FIGURE 1 |** CircRNAs formation and primer design. (A) CircRNAs are formed through a process called back-splicing, which results in a covalently closed loop characterized by a back-splice junction (BSJ). (B) CircRNA-specific primers amplify the region surrounding the BSJ when annealing to the circRNA (convergent orientation). These primers will not amplify linear RNA (divergent orientation). (C) First, CIRCprimerXL retrieves the sequence surrounding the circRNA BSJ (step 1). Next, both common SNPs and regions with secondary structures are flagged (step 2) and avoided during primers design (step 3). Finally, the primers are filtered based on their specificity and secondary structures in the amplicon (step 4).
Finally, a list containing a final primer pair for each circRNA is compiled and returned to the user, along with the complete list of evaluated primers (more than one primer pair can pass all filters) and a report showing primer design rates and reasons for failure.

### 2.3 Performance Tests

To evaluate the performance of CIRCprimerXL, a circRNA dataset was generated by detecting circRNAs with two different circRNA detection tools, find_circ (Memczak et al., 2013) and CIRCexplorer2 (RRID:SCR_021664) (Zhang et al., 2016) in a deeply sequenced total RNA from SW480 colon carcinoma cells (SRA: SRS11316475). From this circRNA dataset, a random subset of 2000 circRNAs was used to investigate the primer design success rate of CIRCprimerXL. To evaluate the scalability and run time of CIRCprimerXL, the pipeline was used to design primers for a random subset of 15,000 circRNAs.

| Number of mismatches in primer 1 | Number of mismatches in primer 2 | Sum of the number of mismatches | Specificity filter setting |
|----------------------------------|----------------------------------|---------------------------------|--------------------------|
| 0                                | 0                                | 0                               | off-target               |
| 0                                | 1                                | 1                               | off-target               |
| 1                                | 0                                | 1                               | off-target               |
| 2                                | 2                                | 2                               | off-target               |
| 3                                | 1                                | 3                               | off-target               |
| 3                                | 1                                | 4                               | off-target               |
| 2                                | 1                                | 3                               | no off-target            |
| 0                                | 3                                | 4                               | no off-target            |
| 3                                | 2                                | 5                               | no off-target            |
| 2                                | 3                                | 6                               | no off-target            |

**TABLE 1** | Tolerated off-targets for circRNA primers (based on (Lefever et al., 2013)). When the specificity filter of CIRCprimerXL is set to “strict” (default), primers with predicted off-targets are discarded, unless there are at least 4 mismatches for a single primer or a total of at least 5 mismatches between the primers and the potential off-target sequence. Alternatively, the specificity setting can be set to “loose”, where at least 5 mismatches for a single primer or a total of at least 4 mismatches between the primers and the potential off-targets are also allowed.

**FIGURE 2** | CIRCprimerXL web interface. CIRCprimerXL is implemented as a user-friendly web-interface, where multiple parameters (Table 2) can be adjusted.
2.4 Computer System’s Specifications
For smaller runs (up to 200 circRNAs), CIRCprimerXL was run on a standard 64-bit laptop computer running macOS (Monterey 12.0.1), equipped with 16 GB of memory, and 2.5 GHz Dual-Core Intel Core i7 processor. For larger runs (up to 15,000 circRNAs), a high-performance cluster (HPC) (Linux) was used (requesting 40 GB memory and 16 CPUs).

2.5 RT-qPCR Validation
2.5.1 Primer Efficiency
Primer efficiency was tested using a 6-point 10-fold dilution series of synthetic DNA template positive controls, previously described in (Vromman et al., 2021).

2.5.2 Cell Culture
SW480 cells were cultured at 37°C, 0% CO₂ in Leibovitz’s L-15 medium (31415-029, ThermoFisher). 10% foetal bovine serum (FBS) (F7524, Sigma) and 1% Penicillin-Streptomycin (10,000 U/mL) (15140122, ThermoFisher) were added to the medium.

2.5.3 RNA Isolation
RNA was isolated from the cells using the miRNeasy Mini kit (217004, Qiagen) according to the manufacturer’s instructions, including an on-column DNase treatment (79254, Qiagen).

| Parameter       | Default Options                                                                 |
|-----------------|---------------------------------------------------------------------------------|
| species         | human, mouse, rat, zebrafish, Xenopus tropicalis, or Caenorhabditis elegans     |
| splicing        | yes                                                                             |
| template length | 150 (which results in a total template length of 300 nucleotides)               |
| ampiclon length | calculated based on template length                                             |
| primer GC settings | 30-50-80 (minimum—optimum—maximum)                                       |
| primer Tm settings | 58-59-60 (minimum—optimum—maximum)                                      |
| SNP filter      | strict, loose (see Table 1)                                                     |

**TABLE 2** | CIRCprimerXL web tool parameters. CIRCprimerXL takes multiple parameters to adjust the species, input template and amplicon size, GC and Tm primer settings, SNP, and specificity evaluation.

**FIGURE 3** | CIRCprimerXL has a high primer design success rate. When using default settings, CIRCprimerXL has a primer design success rate of almost 90% (here tested for 2000 circRNAs). This can be increased by optimizing the CIRCprimerXL parameters. The most improvement can be achieved by allowing a larger design space (template length of 500 instead of 300).
2.5.4 RNase R Treatment and RT-qPCR
RNase R treatment and RT-qPCR were performed following a previously described protocol (Vromman et al., 2021).

3 RESULTS
Here, we present CIRCprimerXL, a high-throughput and user-friendly circRNA primer design tool and web interface. CIRCprimerXL can be used to design circRNA primers for any species (using the command line version) and provides numerous user settings, including input template and amplicon size, SNP and secondary structure filtering, specificity evaluation, and multiple primers specifications.

3.1 Availability and Usage
CIRCprimerXL is implemented as a web application making it accessible through a web browser without the need for package or software installation.

The web interface of CIRCprimerXL (https://circprimerxl.cmgg.be) is an easy-to-use application to design primers for human, mouse, rat, zebrafish, *Xenopus tropicalis*, and *C. elegans* circRNAs (Figure 2). The parameters are described in detail in Table 2. The results are made available as multiple files, including a list containing the optimal primer pair for
TABLE 3 | CIRCprimerXL is efficient and can be used for large circRNA datasets. CIRCprimerXL duration times presented in this table are examples from single runs. The duration time can vary substantially based on the chosen parameters.

| Number of input circRNAs | Standard laptop | HPC (40 GB memory, 16 CPUs) |
|--------------------------|-----------------|-----------------------------|
| 2                        | 2 min           | 33 s                        |
| 20                       | 7 min 43 s      | 1 min 12 s                  |
| 200                      | 1 h 2 min 58 s  | 9 min 37 s                  |
| 2000                     | NA              | 2 h 47 min 03 s             |
| 15,000                   | NA              | 25 h 58 min 52 s            |

For empirical validation of CIRCprimerXL, 20 circRNAs were randomly selected and examined in detail. The PCR efficiency was tested on a dilution series of a synthetic DNA template positive control for each primer pair (Figure 4A). All primer pairs displayed an efficiency between 90 and 110%.

3.3 Run Time
Due to the use of Nextflow and Docker, CIRCprimerXL runs fast and is highly scalable (Table 3).

3.4 Assessment of PCR Amplification Efficiency of 20 circRNA Primer Pairs
For empirical validation of CIRCprimerXL, 20 circRNAs were randomly selected and examined in detail. The PCR efficiency was tested on a dilution series of a synthetic DNA sample treated SW480 RNA, compared to a buffer-control SW480 RNA sample (Figure 4B). All circRNAs remained stable under RNase R treatment, in contrast to four linear control genes, that all (expectedly) show degradation (increase in Cq value).

3.5 Measurement of 20 circRNAs in With RT-qPCR
To validate the circRNAs, RNA isolated from the SW480 cells was used. All 20 primers were measured by RT-qPCR on RNase R treated SW480 RNA, compared to a buffer-control SW480 RNA sample (Figure 4B). All circRNAs remained stable under RNase R treatment, in contrast to four linear control genes, that all (expectedly) show degradation (increase in Cq value).

4 DISCUSSION
RT-qPCR combined with RNase R treatment is currently the most straightforward method to validate circRNAs that have been detected in sequencing data. For this, circRNA-specific primer design is essential. Here, we present CIRCprimerXL, a high-throughput and user-friendly circRNA primer design tool and web interface. CIRCprimerXL can be used to design circRNA primers for any species (using the command line version) and allows the user numerous other settings, including input template and amplicon size, SNP, and secondary structure filtering, specificity evaluation, and multiple primers specifications. CIRCprimerXL has a primer design success rate of almost 90% with default parameters, which can be increased by adjusting the settings. CIRCprimerXL is efficient and can be used for large circRNA datasets. Finally, we validated the PCR efficiency of 20 circRNA primer pairs designed by CIRCprimerXL and showed how these primer pairs can quantify circRNAs in total RNA from SW480 cells.

DATA AVAILABILITY STATEMENT
The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS
MV: software (pipeline), formal analysis, investigation, visualization, writing—original draft, writing—review and editing; JA: software (web interface); JV conceptualization, funding acquisition, supervision,
writing-review and editing; P-JV: software (pipeline and web interface), funding acquisition, supervision, writing-review and editing.

FUNDING

This work was supported by a Foundation Against Cancer grant (STK F/2018/1,267), Standup Against Cancer grant (STIVLK2018000601), and Concerted Research Action of Ghent University (BOF16/GOA/023) and the Research Foundation - Flanders (FWO 1253321N).

ACKNOWLEDGMENTS

We would like to thank Steve Lefever for helpful discussions.

REFERENCES

Dudekula, D. B., Panda, A. C., Grammatikakis, I., De, S., Abdelmohsen, K., and Gorospe, M. (2016). Circinteractome: A Web Tool for Exploring Circular RNAs and Their Interacting Proteins and MicroRNAs. RNA Biol. 13 (1), 34–42. doi:10.1080/15476286.2015.1128065

Fornace, M. E., Porubsky, N. J., and Pierce, N. A. (2020). A Unified Dynamic Programming Framework for the Analysis of Interacting Nucleic Acid Strands: Enhanced Models, Scalability, and Speed. ACS Synth. Biol. 9 (10), 2665–2679. doi:10.1021/acssynbio.9b00523

Gao, Y., and Zhao, F. (2018). Computational Strategies for Exploring Circular RNAs. *Trends Genet.* 34 (5), 389–400. doi:10.1016/j.tig.2017.12.016

Hansen, T. B., Venø, M. T., Damgaard, C. K., and Kjems, J. (2016). Comparison of Circular RNA Prediction Tools. *Nucleic Acids Res.* 44 (6), e58. doi:10.1093/nar/gkv1458

Hoebeek, J., Van Der Luijt, R., Poppe, B., De Smet, E., Yigit, N., Claes, K., et al. (2005). Rapid Detection of VHL Exon Deletions Using Real-Time Quantitative PCR. *Lab. Invest.* 85 (1), 24–33. doi:10.1038/lab.2004.240

Jacobi, T., and Dieterich, C. (2018). Circertools-A One-Stop Software Solution for Circular RNA Research. *Bioinformatics* 35, 2326–2328. November: 1–3. doi:10.1093/bioinformatics/bty948

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L., and Salzberg (2009). Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biol.* 10 (3), R25. doi:10.1186/gb-2009-10-3-r25

Lefever, S., Pattyn, F., De Wilde, B., Coppiers, F., De Keulenaer, S., Hellemans, J., et al. (2017). High-Throughput PCR Assay Design for Targeted Resequencing Using Primer3XL. *BMC Bioinformatics* 18 (1), 400–409. doi:10.1186/s12859-017-1809-3

Lefever, S., Pattyn, F., Hellemans, J., and Vandesompele, J. (2013). Single-Nucleotide Polymorphisms and Other Mismatches Reduce Performance of Quantitative PCR Assays. *Clin. Chem.* 59 (10), 1470–1480. doi:10.1373/clinchem.2013.203653

Lux, S., and Bullinger, L. (2018). Circular RNAs in Cancer. *Adv. Exp. Med. Biol.* 1087 (1), 215–230. doi:10.1007/978-981-13-1426-1_17

Menzczak, S., Jens, M., Elefantnioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs Are a Large Class of Animal RNAs with Regulatory Potency. *Nature* 495 (7441), 333–338. doi:10.1038/nature11928

Meng, X., Li, X., Zhang, P., Wang, J., Zhou, Y., and Chen, M. (2017). Circular RNA: An Emerging Key Player in RNA World. *Brief Bioinform* 18 (4), 547–557. doi:10.1093/bib/bbw045

Panda, A. C., and Gorospe, M. (2018). Detection and Analysis of Circular RNAs by RT-PCR. *Bio Protoc.* 8 (6). doi:10.21769/bioprotoc.2775

Pfafenrot, C., and Pfeuffer, C. (2019). Establishing Essential Quality Criteria for the Validation of Circular RNAs as Biomarkers. *Biomol. Detect. Quantif.* 17 (November 2018), 100085. doi:10.1016/j.bdq.2019.100085

Salzmann, J., Gawad, C., Wang, P. L., Lacayo, N., and Brown, P. O. (2012). Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. *PLoS ONE* 7 (2), e30733. doi:10.1371/journal.pone.0030733

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3–new Capabilities and Interfaces. *Nucleic Acids Res.* 40 (15), e115–12. doi:10.1093/nar/gks596

Vincent, H. A., and Deutscher, M. P. (2006). Substrate Recognition and Catalysis by the Exoribonuclease RNase R. *J. Biol. Chem.* 281 (40), 29769–29775. doi:10.1074/jbc.M506744200

Vromman, M., Yigit, N., Verniers, K., Lefever, S., Vandesompele, J., and Volders, P. J. (2021). Validation of Circular RNAs Using RT-QPCR after Effective Removal of Linear RNAs by Ribonuclease R. *Curr. Protoc.* 1, e181–16. doi:10.1002/cp1.181

Zeng, X., Lin, W., Guo, M., and Zou, Q. (2017). A Comprehensive Overview and Evaluation of Circular RNA Detection Tools. *Plos Comput. Biol.* 13 (6), e1005420. doi:10.1371/journal.pcbi.1005420

Zhang, X. O., Dong, R., Zhang, Y., Zhang, J. L., Luo, Z., Zhang, J., et al. (2016). Diverse Alternative Back-Splicing and Alternative Splicing Landscape of Circular RNAs. *Genome Res.* 26 (9), 1277–1287. doi:10.1101/gr.20895.115

Zhong, S., Wang, J., Zhang, Q., Xu, H., and Feng, J. (2018). CircPrimer: A Software for Annotating CircRNAs and Determining the Specificity of CircRNA Primers. *BMC Bioinformatics* 19 (1), 292–295. doi:10.1186/s12859-018-2304-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.