An accessible database for mouse and human whole transcriptome qPCR primers

Amit Zeisel*, Assif Yitzhaky, Noa Bossel Ben-Moshe and Eytan Domany*

Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Real time quantitative polymerase chain reaction (qPCR) is an important tool in quantitative studies of DNA and RNA molecules; especially in transcriptome studies, where different primer combinations allow identification of specific transcripts such as splice variants or precursor messenger RNA. Several softwares that implement various rules for optimal primer design are available. Nevertheless, as designing qPCR primers needs to be done manually, the repeated task is tedious, time consuming and prone to errors.

Results: We used a set of rules to automatically design all possible exon–exon and intron–exon junctions in the human and mouse transcriptomes. The resulting database is included as a track in the UCSC genome browser, making it widely accessible and easy to use.

Availability: The database is available from the UCSC genome browser (http://genome.ucsc.edu/), track name 'Whole Transcriptome qPCR Primers' for the hg19 (Human) and mm10 (Mouse) genome versions. Batch query is available in the following: http://www.weizmann.ac.il/complex/compphys/software/Amit/primers/batch_query_qpcr_primers.htm

Contact: amit.zeisel@weizmann.ac.il or eytan.domany@weizmann.ac.il

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

1 INTRODUCTION

Quantitative polymerase chain reaction (qPCR) is an efficient tool to quantify changes in abundance of DNA/RNA molecules. Its main advantage is that it has a large dynamic range and requires only basic sample preparation steps (reverse transcription of RNA). Double-stranded DNA dye (such as SYBR) based qPCR quantifies the amount of double-stranded DNA in a sample. A pair of specific primers amplifies a short (usually 60–130 bp) DNA region, and the emitted light is measured after every cycle. The qPCR platforms allow measurement in parallel (on the same plate) of many different reactions, which saves time and work and reduces technical bias. Primer design for qPCR is a relatively simple but repetitive time-consuming task. The first step is to choose a target sequence, for which one then has to define constraints about the primers' product and reaction conditions. Then, available software (e.g. primer3, Rozen and Skaletsky, 2000; Untergasser *et al.*, 2012) is used to

get the set of best suggested primer pairs. When browsing RNA transcripts, the target sequence should be determined and selected carefully because of constraints such as product size, amplification of a single product and, in particular, the need to amplify a specific isoform of a given gene. Moreover, as qPCR platforms enable performing measurements for a large number of genes in parallel, the user has to repeat many times the task of target sequence selection and primer design manually, which can be tedious and prone to human errors. As an example, in the most frequent case, of measuring mRNA, one wishes to amplify an exon–exon junction, picking the sequences of both exons and imposing the constraint of having the forward primer on the 5′ exon and the reverse primer on the 3′ exon; see also Sandhu and Acharya (2005) discussed in Supplementary Information.

2 APPROACH

We carried out an automatic primer design procedure to target all possible transcripts and all 'possible' exon–exon and intron–exon junctions in the human and mouse transcriptomes. Not all consecutive exon–intron–exon triplets generate 'possible' primer pairs (i.e. satisfy the constraints). The resulting suggested primer pairs can be easily viewed as a track in the UCSC genome browser, making the application available to a broad audience. Our procedure uses the UCSC gene model as a reference of the gene structure and the primer3 software (release 2.3.2, http://frodo.wi.mit.edu/) as the design engine. The software goes over all possible consecutive exon–intron–exon triplets in the transcriptome and applies our design rules. We search for a primer pair to amplify the intron–exon junction and then use the reverse primer (interrogating the 3′ exon) to find a forward primer on the 5′ exon. The two primer pairs share the same reverse primer (see Fig. 1A). To provide a choice of primer pairs (i) the last step is done twice to yield another forward primer, and (ii) the entire process is done twice for each triplet.

3 METHODS

Detailed design rules:

(1) Do not address the first exon–exon junction to avoid problems related to abortive transcription.

(2) Intron length should be >800 bp to avoid problems of double products in amplification.

(3) We design only junction primers so that forward and reverse primers must flank the junction.

(4) Melting temperature should be between $60 < Tm < 63$ [optimal 60.5 according to Breslauer *et al.* (1986) tables].

*To whom correspondence should be addressed.
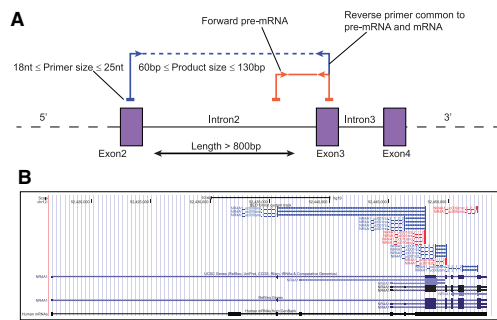
Fig. 1. (**A**) Basic design principles. In the isoform shown, the junction between exons 2 and 3 is amplified (intron >800 bp), whereas between 3 and 4, it is not. (**B**) View from the UCSC genome browser track

(5) Primer length should be 18–25 bp.

(6) Product size should be 60–125 bp.

(7) We first design primers for the intron–exon junction [precursor messenger RNA (pre-mRNA)], we pick the two best primer pairs and then for both reverse primers, we design two options for the corresponding forward primer of the exon–exon junction (mRNA).

(8) We use the 'PRIMER MISPRIMING LIBRARY' of the *primer3* software: 'human' for human and 'rodent' for mouse transcriptome, respectively.

Manual:

To use the database via the UCSC genome browser, follow these steps:

(1) In the browser, select organism: 'human' or 'mouse'.

(2) Browse to your gene of interest.

(3) To view the track 'Whole transcriptome qPCR primers' scroll down the page to the panel of suggested tracks named 'Expression'. Turn the 'qPCR primers' to 'pack' (see Fig. 2A) and press refresh.

(4) Use the UCSC browser tools to zoom in to specific regions of interest (see example in Fig. 2B).

(5) View the primer pairs we suggest. In red are the primers designed to amplify pre-mRNA, and in blue are the primers designed to amplify mRNA. For each pre-mRNA primer pair, there are two corresponding mRNA primers (that use the same reverse primer, if possible). Each pair has a unique code, which stands for the gene name and the junction name. For example, the pair 'NR4A1 _uc009zmc.3_2_1' amplifies pre-mRNA, and the corresponding mRNA primers are 'NR4A1_uc009zmc.3_2_1_1' and 'NR4A1 _uc009zmc.3_2_1_2'. The naming scheme is as follows: 'NR4A1' is the gene symbol, 'uc009zmc.3' is the UCSC Genes identifier of the isoform, '2' identifies the exon–intron–exon triplet, '1' is the number (1 or 2) of the suggested pair for intron–exon junction. The names of primer pairs that cover intron–exon junctions end here. For the exon–exon junctions that use the same reverse primer, there is one additional number, which can be 1 or 2.

(6) Click on the primer pair you select; this will take you to a new page with details, Figure 2C.

(7) Clicking on the number next to 'Click here for primer details' opens a new page with the primer properties (see Fig. 2D).

(8) For design of a large number of primers, we recommend using our batch query, which converts a list of Primer Names to a table of primer details. Batch query is possible from our website, accessible either directly: Batch Query (http://www.weizmann.ac.il/complex/compphys/software/Amit/primers/batch_query_qpcr_primers.htm) or from the UCSC browser (under 'Description').
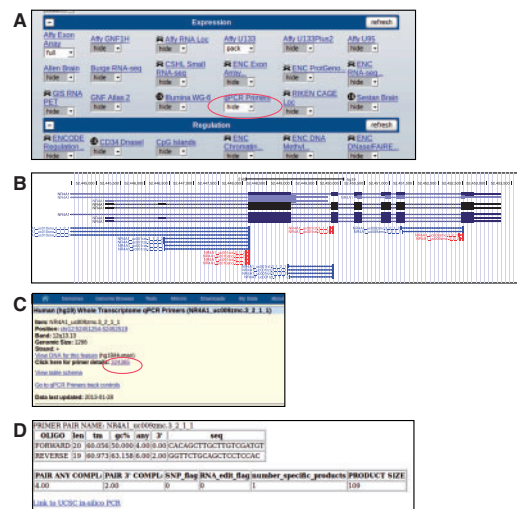
Fig. 2. Step-by-step guide to use the UCSC qPCR database track

(9) Outputs (both 7 and 8 aforementioned) contain warnings when SNPs or RNA editing sites appear in the primer-binding regions. Information about primer specificity is also provided (with link to *in silico* PCR)

4 DISCUSSION AND CONCLUSION

We created databases for primer pairs to be used for amplification of intron–exon and exon–exon junctions by qPCR, genome-wide, for human and mouse; in total, 534 301 pairs for human (and 518 230 for mouse) targeting 17 929 (and 18 039) different genes in human (and mouse). Many of the primer pairs in the database were tested successfully in various experiments [success rate >90%, see Supplementary File 1 and Supplementary Information and Zeisel *et al.* (2011)]. The databases are accessible as a track in the UCSC browser. Optionally users can access our web server to obtain, in batch mode, primer pairs for a large number of junctions of interest.

REFERENCES

Breslauer,K.J. *et al.* (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.

Rozen,S. and Skaletsky,H.J. (2000) *Primer3 on the WWW for general users and for biologist programmers.* Humana Press, Totowa NJ.

Sandhu,K.S. and Acharya,K.K. (2005) Exprimer: to design primers from exon–exon junctions. *Bioinformatics*, **21**, 2091–2092.

Untergasser,A. *et al.* (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.

Zeisel,A. *et al.* (2011) Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol. Syst. Biol.*, **7**, 529.