

qPrimerDepot: a primer database for quantitative real time PCR

Wenwu Cui, Dennis D. Taub¹ and Kevin Gardner*

Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Bethesda, MD 20892-4605 and
¹Laboratory of Immunology, National Institute on Aging, Baltimore, MD 21224, USA

Received June 3, 2006; Revised September 25, 2006; Accepted September 29, 2006

ABSTRACT

Gene expression studies employing high throughput real time PCR methods require finding uniform conditions for optimal amplification of multiple targets, often a daunting task. We developed a primer database, qPrimerDepot, which provides optimized primers for all human and mouse RefSeq genes. These primers are designed to amplify desired templates under unified annealing temperature. For most intron-bearing genes, primers flank one of the largest introns thus minimizing background noise due to genomic DNA contamination. The qPrimerDepot database can be accessed at <http://primerdepot.nci.nih.gov/> and <http://mouseprimerdepot.nci.nih.gov/>.

INTRODUCTION

Real time PCR (RT-PCR) can rapidly, reproducibly and quantitatively determine changes in gene expression (1). Although microarray analysis can measure large scale gene expression levels simultaneously, its hybridization-related variation often demands validation by other methods. Routinely, RT-PCR is used to verify the observation from microarray studies. However, several artifacts can confound the analysis including: (i) amplification of undesired template secondary to mispriming or annealing at inappropriate temperatures; and (ii) susceptibility to RNA contamination with genomic DNA, especially when collecting samples from tumor tissues (2). Though the problem of genomic contamination is partially addressed by DNase treatment this method is often incomplete and its protracted use often diminishes the sensitivity of detection. This is a costly problem particularly when the detection of rare transcripts in precious tissue samples is desired (3).

Low cost methods for detecting fluorescent dyes which bind to double stranded DNA, such as SYBR Green, are most widely used and suitable for high throughput screening. Since these dyes are not sequence specific, careful consideration should be given to avoid generating extraneous

amplicons. One of the obstacles to high throughput RT-PCR gene expression studies in which multiple unique transcripts are simultaneously measured in 96 or 384 well formats, is the necessity to individually optimize each assay for each target (4). Currently, the criteria for successful determination by quantitative RT-PCR require that: (i) the optimal amplicon should be located in a non repetitive region without segments of low complexity, (ii) the optimal amplicon size should be ~100 bp to ensure the efficiency of *Taq* polymerase processivity, (iii) if possible, primers should be designed to flank intron–exon borders or primers anneal at a splice junction to distinguish genomic DNA from cDNA template, and (iv) primers have similar melting temperatures with 20–70% GC content (5,6). It is a laborious and error-prone chore to design RT-PCR primers that meet these requirements. Available resources for pre-designed primers are limited. RTPimerDB (<http://medgen.ugent.be/rtpimerdb/>), an online database, provides experimentally verified primer sets for 2699 human and 487 mouse genes (7,8). PrimerBank (<http://pga.mgh.harvard.edu/primerbank/>) is a well known resource that covers most known human (33 741) and mouse (27 681) genes (9). However, its primer algorithm is not designed to span introns and is therefore more prone to amplify contaminating genomic sequences. Here we describe qPrimerDepot, a primer database for RT-PCR analysis of >99% of human (23 400) and mouse (18 733) RefSeq genes. These primers sets are designed to be used under uniform annealing temperatures to facilitate their application in large scale high throughput assays. Moreover, to reduce the noise from contaminating genomic DNA (6), over 90% of the primer sets are designed to produce amplicons bridging exon:exon junctions of intron-bearing genes.

MATERIALS AND METHODS

Data processing

Sequences file (refMrna.zip) and intron/exon information tables (refGene.txt.gz) of 23 463 human and 18 737 mouse RefSeq genes (UCSC hg17 and mm6) were downloaded from UCSC genome browser (<http://hgdownload.cse.ucsc.edu/>).

*To whom correspondence should be addressed at National Institutes of Health, Advanced Technology Center, Room 134C, 8717 Grovemont Circle, Bethesda, MD 20892-4605, USA. Tel: +1 301 496 1055; Fax: +1 301 435 7558; Email: gardnerk@mail.nih.gov

To assure amplicons free of repetitive elements and sequences of low complexity (10), we utilized Biowulf, a high-performance Linux cluster at the National Institutes of Health, to mask the repetitive elements using the Repeat-Masker application with built-in MaskerAid (11).

Primer3 (12) was used to design primers for each RefSeq entry, with the following parameters: for intronless genes (5.5% of human RefSeq genes and of 12.4% of mouse RefSeq genes), primers were set to be between 17 and 27 bp with 20 bp as optimum, and melting temperature was set to be between 57 and 63°C with 60°C as optimum, all other parameters, such as PRIMER_SELF_ANY and PRIMER_SELF_END, were set to default (8.0 and 3.0, respectively) to assure low self-complementarity. All cDNA amplicons were 90–150 bp in size to ensure *Taq* polymerase efficiency. For 99% of intron-bearing genes (94.5% human genes and 88.6% mouse genes bear at least one intron), primers were designed to flank or cross an exon-intron border in which the intron was one of the top three largest in the gene of interest. Thus, contamination by genomic DNA would generate either a longer product, which can be detected by melting curve analysis, or no product if the contamination template length (intron > 3 Kb) is too long for *Taq* polymerase to traverse during the extension period.

The BLAST algorithm was used via the NIH biowulf Linux cluster to evaluate all primers against corresponding RefSeq databases. The criteria for possible mis-priming requires that both primers have at least 15 matches in another RefSeq entry (i.e. expectation value, $e < 1$) (13,14). The BLAST result revealed that 891 of human and 420 of mouse primers may mis-prime to other RefSeq sequences. Sequence alignments of query and hit RefSeq using the BLAST2 algorithm (15) were performed and primer pairs which had <80% identities were filtered out of the database. Annotations are presented in the user interface for the individual primer sets that could 'mis-prime' another RefSeq gene with >80% identity. The sources of these mis-primed RefSeq will vary, but may include redundancy within the RefSeq database, transcript variants and paralogs of high sequence similarity. Each of these possibilities can be assessed by a direct link that is provided to *in silico* PCR (<http://genome.ucsc.edu/cgi-bin/hgPcr?command=start>) for all primer pair sets. This link allows the user to rapidly identify amplicon locations in the mouse and human genomes so that primer specificity can be visually assessed and validated.

qPrimerDepot can be accessed at <http://primerdepot.nci.nih.gov/> or <http://mouseprimerdepot.nci.nih.gov/> by querying the database with a RefSeq ID or a gene name. Batch query service is available upon request if user provides standard gene name or accession number. Flat files and MySQL dump file which have all primer information are also available upon request.

Experimental validation

Reverse transcription was applied with Omniscript RT Kit following manufacturer's protocol (Qiagen). A 20 µl RT reaction included 2 µg Universal reference RNA (Stratagen), 1 µM Oligo-dT primer, 2 µl of 10× RT buffer, 0.5 mM each dNTP, 10 U of RNase inhibitor, 4 U of Omniscript

Reverse Transcriptase, and DEPC-treated water. The reaction mix was incubated at 37°C for 60 min. After the reaction, the mix was diluted 1:5 with water for PCR analysis.

Primer sequences were extracted from our database and synthesized by Integrated DNA Technologies (Coralville, IA, USA). Quantitative RT-PCR was carried out in a DNA Engine Opticon-2 Real Time PCR Detection System (MJ Research). In brief, each 20 µl reaction mix comprises 0.3 µM primers (both 5' and 3' primers), 1 µl template from reverse transcription and 10 µl 2 × QuantiTect SYBR Green PCR Master Mix (Qiagen). Each reaction mix was incubated at 95°C for 15 min, 40 cycles of 95°C for 15 s and 60°C for 1 min. A melting curve analysis which read every 0.3°C was followed to assess the homogeneity of a PCR product. Real-time PCR results were analyzed using the software provided by the manufacturer.

RESULTS AND DISCUSSION

Our database comprises pre-designed primers for 42 133 mouse and human RefSeq genes (Table 1). For most genes three unique sets of primers are provided (96.3% of total). The database provides a simple user interface where the user may enter either the HUGO approved gene symbol or the RefSeq gene identifier (Figure 1). The database graphic output provides information on the primary transcript location, number of introns, primer sequence, primer length, GC%, amplicon size, and genomic amplicon size. Also a direct link is provided for location of the genomic amplicon by *in silico* PCR (Figure 2).

To experimentally evaluate the primer quality, 288 genes were arbitrarily selected from a list of genes known to function in the immune response. 288 primer sets were retrieved from the database and synthesized in 96-well plates. Universal human reference RNA was reverse transcribed

Table 1. Summary statistics for qPrimerdepot primerDb

		Intron-bearing (%)	Intronless (%)	All entries (%)
<i>Homo sapiens</i>				
Intron-bearing genes	22 176	100.00		94.51
Primer overlap an intron	21 910	98.80		93.38
No intron considered	212	0.96		0.90
Intronless genes	1287		100.00	5.49
Primer provided for intronless genes	1278		99.30	5.45
RefSeq genes	23 463			100.00
Total genes with primer design	23 400			99.73
<i>Mus musculus</i>				
Intron-bearing genes	16 411	100.00		87.59
Primer overlap an intron	16 293	99.28		86.96
No intron considered	115	0.70		0.61
Intronless genes	2326		100.00	12.41
Primer provided for intronless genes	2325		99.96	12.41
RefSeq genes	18 737			100.00
Total genes with primer design	18 733			99.98

qPrimerDepot --- A quantitative real time PCR primer database

This database provides qRT PCR primers for 99.96% human RefSeq sequences.

For 99% of intron-bearing genes, the PCR product will cross an exon-exon border which overlaps one of the largest introns.

All primers have annealing temperatures approximately 60C.

Please input RefSeq ID (such as NM_XXXXXX) or [HUGO](#) gene name (such as JUN) here:

VEGF	
Submit Query	Reset

Figure 1. Web input interface for qPrimerDepot (see text).

Primer Information

Gene Name: VEGF	RefSeq ID: NM_003376	Chromosome: chr6	Strand: +
Transcription: 43845923-43860314	Number of introns: 7		
Right primer sequence: AGCTGCGCTGATAGACATCC 55 Tm: 60.535	Position: 182	Length: 20	GC%:
Left primer sequence: CTACCTCACCATGCCAAGT 55 Tm: 59.989	Position: 79	Length: 20	GC%:
cDNA amplicon size:104	Estimated genomic amplicon size:3180	in silico PCR	
Right primer sequence: GCAGTAGCTGCGCTGATAGA 55 Tm: 59.498	Position: 187	Length: 20	GC%:
Left primer sequence: CCTTGCTGCTCTACCTCCAC 60 Tm: 60.012	Position: 69	Length: 20	GC%:
cDNA amplicon size:119	Estimated genomic amplicon size:3195	in silico PCR	
Right primer sequence: CACACAGGATGGCTTGAAGA 50 Tm: 59.831	Position: 265	Length: 20	GC%:
Left primer sequence: AGGGCAGAATCATCACGAAG 50 Tm: 60.218	Position: 130	Length: 20	GC%:
cDNA amplicon size:136	Estimated genomic amplicon size:3212	in silico PCR	

Figure 2. Web output interface for qPrimerDepot (see text).

and used as template in PCR to examine the quality of the primers.

Given the variation of transcript abundance, melting curve analysis followed by gel electrophoresis has been suggested

to verify RT-PCR products (6). Melting curve analysis revealed that 94.1% generate unique product. Visualization by the less sensitive ethidium bromide stain shows that >70% of the primer sets produce an amplicon of the correct

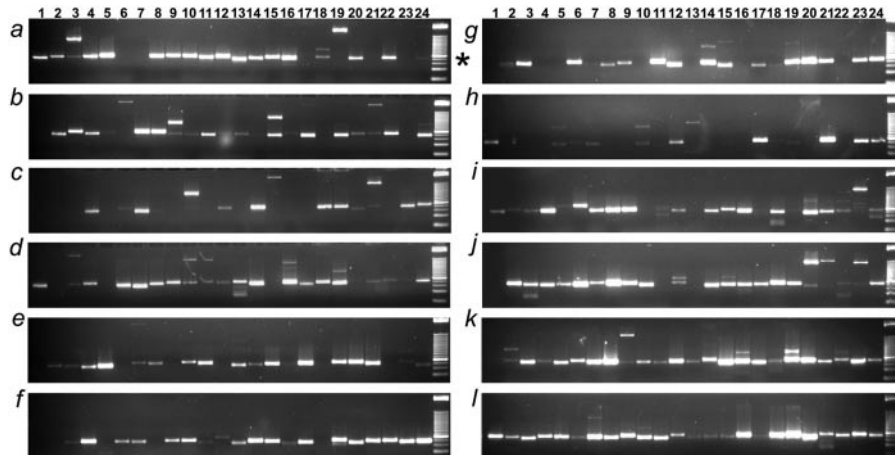


Figure 3. Validation of pPrimerDepot primer sets on 3% ethidium bromide stained agarose/acrylamide (3:1) gels after 38 cycles of amplification. Asterisk indicates 125 bp size marker.

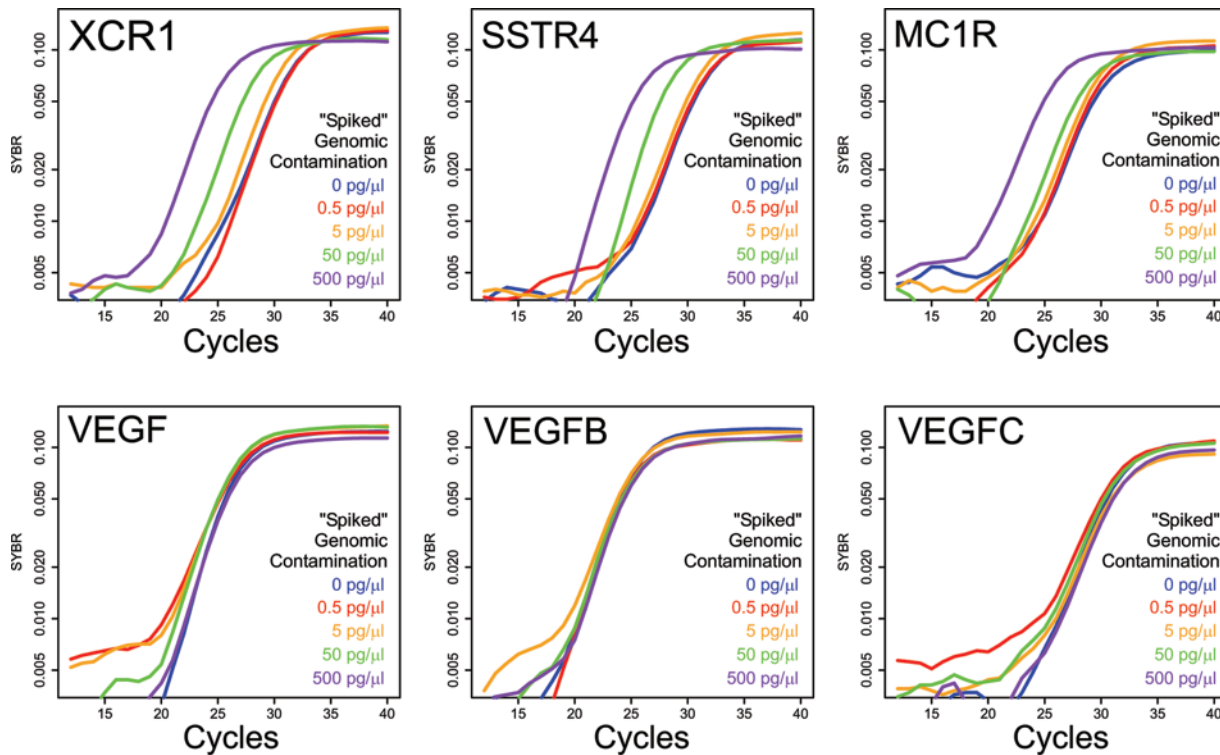


Figure 4. Real-time PCR amplification profiles generated using qPrimerDepot primer sets for non intron-bearing (XCR1, SSTR4, MC1R) and intron-bearing (VEGF, VEGFB, VEGFC) genes compared after the addition of increasing amounts of contaminating genomic DNA.

molecular weight that will amplify and be detected as a single species by quantitative PCR (Figure 3). Several of the failures detected by gel electrophoresis are likely due to very low abundance transcripts in the universal RNA, imperfect primer design, unanticipated high secondary mRNA structure, or erroneous exon annotation in UCSC Genome Browser. Approximately 88.5% of primer sets produced no product in the absence of reverse transcriptase and the remaining sets produced detectable product only beyond 34 cycles of amplification possibly, due to primer dimers.

The resistance of most qPrimerDepot primer sets to contaminating input genomic DNA is illustrated in Figure 4. Here the real-time amplification profiles of three intron-bearing genes (VEGF, VEGFB and VEGFC) was compared to that of three non intron-bearing genes (XCR1, SSTR4 and MC1R) after challenge with increasing concentrations of contaminating genomic DNA (0.5–500 pg/μl). As demonstrated in Figure 4, all three intron-bearing genes show robust resistance to >500 pg/μl of input genomic DNA. This is in stark contrast to the three non intron-bearing

genes where as little as 5 pg/ μ l produces a significant false signal.

CONCLUSION

Taking advantage of the intron/exon inventory of RefSeq genes and Primer3, a paradigm primer design tool, we designed primers which are contamination resistant for 99% of human and mouse RefSeq genes (Table 1). Since the majority of the primer sets will amplify desired templates under unified annealing temperatures, high throughput multiplex analysis is achievable at a reasonable cost. Empirical screening and validation of primer set performance conservatively suggests that 70–90% of the primer set designs are likely to perform effectively ‘right out of the box’ with no need to adjust the conditions of amplification. Therefore, qPrimerDepot is a valuable resource for qRT applications, especially in those circumstances requiring high throughput detection of rare transcripts in curated and/or patient-derived samples that often contain unavoidable contamination with genomic DNA.

ACKNOWLEDGEMENTS

This project has been supported by funds from the Intramural Research Program of the NIH, National Cancer Institute, Centers for Cancer Research and the National Institute on Aging. Funding to pay the Open Access publication charges for this article was provided by the Intramural Research Program of the NIH, National Cancer Institute, Centers for Cancer Research.

Conflict of interest statement. None declared.

REFERENCES

1. Walker,N.J. (2002) A technique whose time has come. *Science*, **296**, 557–559.
2. Kitlinska,J. and Wojciorowski,J. (1995) RNA isolation from solid tumor tissue. *Anal. Biochem.*, **228**, 170.
3. Bustin,S.A. (2002) Quantification of mRNA using real-time reverse transcription PCR (RT–PCR): trends and problems. *J. Mol. Endocrinol.*, **29**, 23–39.
4. Shaffer,C. (2005) PCR gains momentum with new applications. *Genetic Engineering News*, **25**, 24–29.
5. Bustin,S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.*, **25**, 169–193.
6. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (2005) Current protocols in molecular biology. John Wiley & Sons, New York.
7. Pattyn,F., Speleman,F., De Paepe,A. and Vandesompele,J. (2003) RTPrimerDB: the Real-Time PCR primer and probe database. *Nucleic Acids Res.*, **31**, 122–123.
8. Pattyn,F., Robbrecht,P., De Paepe,A., Speleman,F. and Vandesompele,J. (2006) RTPrimerDB: the real-time PCR primer and probe database, major update. *Nucleic Acids Res.*, **34**, D684–D688.
9. Wang,X. and Seed,B. (2003) A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res.*, **31**, e154.
10. Jurka,J. (2000) Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.*, **16**, 418.
11. Bedell,J.A., Korf,I. and Gish,W. (2000) MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics*, **16**, 1040–1041.
12. Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.
13. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
14. Wang,X. and Seed,B. (2003) Selection of oligonucleotide probes for protein coding sequences. *Bioinformatics*, **19**, 796–802.
15. Tatusova,T.A. and Madden,T.L. (1999) BLAST 2 S, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiology Lett.*, **174**, 247–250.