

ChIPprimersDB: a public repository of verified qPCR primers for chromatin immunoprecipitation (ChIP)

Stefan Kurtenbach^{*,†}, Rohit Reddy[†] and J. William Harbour[†]

Bascom Palmer Eye Institute, Sylvester Comprehensive Cancer Center, and Interdisciplinary Stem Cell Institute, University of Miami Miller School of Medicine, Miami, FL 33136, USA

Received July 18, 2018; Revised August 27, 2018; Editorial Decision August 29, 2018; Accepted September 05, 2018

ABSTRACT

Chromatin immunoprecipitation (ChIP) has ushered in a new era of scientific discovery by allowing new insights into DNA-protein interactions. ChIP is used to quantify enriched genomic regions using qPCR, and more recently is combined with next generation sequencing (ChIP-seq) to obtain a genome wide profile of protein binding sites. Nevertheless, ChIP-qPCR remains an integral component of this technology for quality control purposes, before the library preparation and sequencing steps. In addition, ChIP-qPCR remains more time- and cost-effective for many focused projects in which the DNA regions of interest are already known. However, the DNA oligonucleotide primers needed for ChIP-qPCR are more challenging to design than for other qPCR projects. Here, we present the first public repository for ChIP oligonucleotides that have been verified to perform well in ChIP-qPCR experiments. ChIPprimersDB was developed by manual screening of publications to ensure primer quality and provide additional specific information on the ChIP experiments where the primers have been used. In addition to the primer sequences, the database includes information about the antibody, cells and tissues used in the experiment, information on the experimental design, and a direct link to the original publication. The database is linked at <https://umiamihealth.org/bascom-palmer-eye-institute/research/clinical-and-laboratory-research/ocular-oncology-laboratory/chip-primers> and hosted at <https://www.chipprimers.com/>.

INTRODUCTION

Polymerase chain reaction (PCR) is widely used to amplify specific target DNA sequences in various applications. This targeted amplification is achieved by oligonucleotide primers flanking the sequence of interest that initiate the polymerase reaction. Since the invention of PCR in 1983 (1–3), the method has been widely adopted and modified to suit various purposes (4), including qPCR (5) and qRT-PCR (5,6), allowing the real-time quantification of PCR amplicons. Today, qPCR still remains the most sensitive technique for measuring minute quantities of nucleic acids in research and diagnostics applications.

The quality of PCR amplification is highly dependent on the specificity and efficiency of the primers. Specificity can be assessed by analyzing melting curves and separation of the PCR products on an agarose gel to verify the correct amplicon size, alongside Sanger sequencing. Primer efficiency is calculated by performing PCR with progressive template dilutions and is a measure of the amount of amplification per cycle. Numerous primer databases have been established to guide the selection of high quality primers (7–18). Further, publicly available algorithms allow the design of high quality qPCR primers that meet specific characteristics (19).

More recently, qPCR has been adopted to quantitate the enrichment of DNA fragments in chromatin immunoprecipitation (ChIP) experiments. ChIP allows for the quantitation of protein binding enrichment at specific genomic regions, thereby providing a new window into chromatin organization and gene regulation. This method employs chemical crosslinking to crosslink (or ‘fix’) DNA-protein interactions, and the chromatin is subsequently sheared into small fragments using enzymatic or physical methods. The DNA fragments are subjected to immunoprecipitation with antibodies against proteins of interest that bind directly or indirectly to genomic DNA. After immunoprecipitation, the DNA is released from its interacting proteins and an-

*To whom correspondence should be addressed. Tel: +1 305 326 6020; Email: Stefan.Kurtenbach@med.miami.edu

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Present address: Stefan Kurtenbach, Department of Ophthalmology, University of Miami Miller School of Medicine Biomedical Research Building, Room 832, 1501 NW 10th Avenue, Miami, FL 33136, USA.

alyzed for enrichment by qPCR. Although ChIP-qPCR is widely used, it still remains challenging and time intensive, in part due to the process of developing optimal PCR primers. More recently, drawing a genome-wide picture of protein-DNA interactions has been made possible through the development of ChIP followed by next generation sequencing (ChIP-seq) (20–22). Although this method is very powerful, it has not supplanted the need for ChIP-qPCR, which plays a crucial role in quality control before sequencing and validation of ChIP-seq findings. In addition, ChIP-qPCR is still used for focused experiments and verification of ChIP-seq findings.

Designing primers for ChIP-qPCR is significantly more challenging than for other qPCR methods for several reasons: (i) ChIP primers must target very specific regions, thereby limiting the options for primer design. This is particularly relevant for proteins with narrow binding regions, like transcription factors, where primers verified to work for one antibody in a promoter region might not work efficiently in another ChIP experiment. (ii) The quality of the DNA is reduced by the mechanical shearing as well as the chemical crosslinking. (iii) The quantity of available DNA is typically low: frequently < 5 ng. (iv) Intron-spanning primers are used to enhance specificity in qRT-PCR reactions, but as the template in ChIP is genomic DNA, this is not possible.

Even though ChIP primer information is provided in the materials and methods section of publications, it remains very time consuming to find suitable primers in the literature. Many publications describe ChIP-qPCR results with mathematical significance, however, the actual fold-change or the controls used prevent a determination of the general suitability of these primers. Alternate or outdated gene names in some publications constitute another difficulty. Despite the challenges in designing ChIP-qPCR primers, to the best of our knowledge, there is no ChIP primer repository/database available thus far. The current lack of a ChIP primer database can be attributed to the complexity of extracting high quality information of ChIP experiments from publications—a process that cannot be automated in a meaningful way. To address this need, we describe herein a database for published and verified ChIP-qPCR primer sequences, curated by manual screening, providing researchers a user-friendly interface to access and compare ChIP-qPCR primers, and get specifics on the experimental conditions in which the primers were used. The database is linked at <https://umiamihealth.org/bascom-palmer-eye-institute/research/clinical-and-laboratory-research/ocular-oncology-laboratory/chip-primers> and hosted at <https://www.chipprimers.com/>.

MATERIALS AND METHODS

PubMed searches were performed using the search terms ‘ChIP-seq’, ‘ChIP-qPCR’ and ‘Chromatin Immunoprecipitation’ and publications were manually screened for ChIP-qPCR experiments. ChIP-qPCR primers were included when the experimental protocol and primer description were not ambiguous and study showed qPCR enrichment over a valid control of at least 5-fold. Gene names were verified with the NCBI database (<https://www.ncbi.nlm.nih.gov/gene/>) for correct nomenclature. In cases where old gene names or aliases had been used, the gene name was replaced by the correct name and a note was made in the database entry stating the original gene name used by the authors. Further, the complete gene name (‘Description’) was added, alongside the antibody used, the species, and the cell line or tissue which was used in the experiment. If the authors reported enrichment only when cells or animals that were not wild type (e.g. treated, stimulated, gene knockdown), we stated the conditions in the notes of the respective database entry. The PubMed ID was included and is displayed as a direct clickable link in the user interface opening a new browser window leading to the original publication on the NCBI website. The melting temperature (T_m) was calculated with OligoCalc (23) assuming 50 nM primers and 50 mM salt (Na^+). All entries were double-checked by two independent researchers.

Further, the complete gene name (‘Description’) was added, alongside the antibody used, the species, and the cell line or tissue which was used in the experiment. If the authors reported enrichment only when cells or animals that were not wild type (e.g. treated, stimulated, gene knockdown), we stated the conditions in the notes of the respective database entry. The PubMed ID was included and is displayed as a direct clickable link in the user interface opening a new browser window leading to the original publication on the NCBI website. The melting temperature (T_m) was calculated with OligoCalc (23) assuming 50 nM primers and 50 mM salt (Na^+). All entries were double-checked by two independent researchers.

RESULTS

Curating database entries of high quality primer sequences

A crucial part of determining ChIP enrichment at a genomic locus by qPCR is the use of valid controls. It is standard to perform PCR with control DNA in parallel, which can either be derived from an aliquot of the sonicated chromatin before the IP (‘input control’), or DNA from a mock ChIP performed with an IgG control antibody or an antibody specific for a non-nuclear antigen. Either of the two latter methods are accepted as valid controls as described in the *ENCODE ChIP-seq Guidelines* (24), and data is usually presented as ‘% of input’ after analysis. In addition, a primer set targeting a non-enriched region should be used. While some studies use a big change in the ‘% of input’ value to claim enrichment, it must be noted that without primers targeting non-enriched regions, a high ‘% of input’ value is technically not sufficient, even if other antibodies in the same experiment had lower ‘% of input’ values. The reason being that the ‘% of input’ value highly depends on the amount of DNA that is used in the PCR reactions, while antibodies, including control IgG antibodies, vary widely in their unspecific DNA pulldown. Hence, to curate a repository of high quality ChIP-qPCR primers, we manually screened publications for ChIP primers and only included primers that showed ≥ 5 -fold enrichment over a valid control and a non-enriched region. If no non-enriched regions were used as controls, we only included primers in cases where a ≥ 5 -fold enrichment was shown the same regions following different experimental conditions such as gene knockdown or drug treatment. While lower fold-changes may be biologically significant and are frequently reported in literature, a 5-fold cutoff increases the confidence and quality of the entries.

Database structure

Each primary database entry consists of a gene name/species pair. All gene names were verified with the NCBI database for up-to-date names and abbreviations and corrected if incorrect or out-of-date. Each primary gene name/species entry may have multiple primer pair entries (primer sequences), which in turn can have multiple

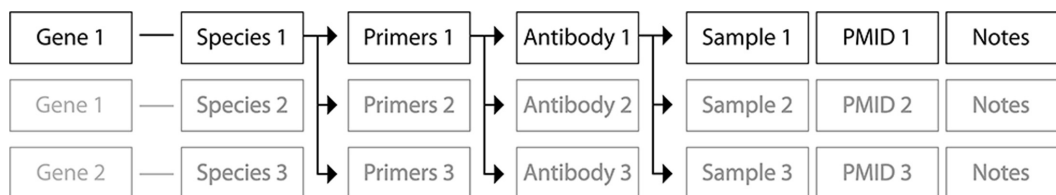


Figure 1. ChIPprimersDB Database structure. Each Gene/Species combination may have multiple primer entries, in turn allowing for multiple antibody entries. Each antibody may include multiple ChIP experiments, comprised of the sample used, the PMID as a clickable link, and Notes.

ChIPprimersDB
A public repository for verified ChIP qPCR primers

sox2 Advanced Search ▾

Primers: 8, Entries: 4, Genes: 3, Species: 2

Gene ID	Description	Species	
CASP3	caspase 3	Human	Show primers (1) Download
GLI2	GLI family zinc finger 2	Human	Show primers (1) Download
Sox2	SRY (sex determining region Y)-box 2	Mouse	Show primers (2) Download

← Previous 1 Next →

Figure 2. Example of a search performed on ChIPprimersDB. By default, the search function will search all database fields. A search for Sox2 results in three results, where one is the Sox2 gene entry containing primers for the Sox2 gene. The ‘Show primers’ button indicates that two different verified primer pairs are available for this gene. The genes CASP3 and GLI2 are listed in the search results, as those genes were found enriched in ChIP experiments for Sox2. Using the advanced search function, the user can limit the search to specific fields.

‘Antibody’ entries. Each antibody may have multiple entries consisting of the precise sample used (which cell line or tissue), the PubMed ID of the publication (PMID), and notes on the experimental procedure (Figure 1) - *de facto* resembling individual ChIP experiments.

Database interface and use

The interface of the database website by default displays all entries as a list, including the gene symbol, full gene name (‘Description’), and species. Further, each entry has a ‘Show Primers’ button, displaying the number of primer pairs available for the respective genes. By clicking the ‘Show Primers’ button, the database will display all primer sequences for the gene alongside their melting temperatures (T_m), the antibodies used, the tissues or cell lines used (‘Sample’), the PubMed ID (PMID) of the publication reporting the primers, and ‘Notes’. The PMID is a clickable link which opens the PubMed website with the original article in a new browser window, to allow the user to quickly assess the complete information on the experimental setup and materials used for this ChIP experiment. The ‘Notes’ section contains additional information on the experiment, e.g. if cells were treated with drugs, a gene knockdown was

performed, or alternate/outdated gene names were used in the manuscript.

ChIPprimersDB has a search function that reacts in real time to the users input. By default, the search is performed for every field for each database entry (Figure 2). To allow more specific searches, we also implemented an ‘Advanced Search’ option to restrict the search to certain fields, allowing the user to search for e.g. genes or antibodies only. All entries can be easily downloaded by clicking on the ‘Download’ button, which provides a .csv file that can be viewed instantly for instance in Microsoft Excel. To enable users to submit their own published primer sequences, we also designed a submission form to streamline this process. Users can easily enter all the required information and submit the entry for revision by a site administrator, which will check if all quality measures outlined above are met. This will allow this repository to grow organically and for users to get credit for designing quality ChIP primers by citations from the database users.

DISCUSSION

ChIPprimersDB serves multiple use cases frequently emerging when planning ChIP experiments: The direct link to the original publications on the PubMed website allows users

to quickly compare the data from different groups and decide on an experimental strategy and a specific antibody, which yielded the best results. Choosing the right antibody is especially important, as there are significant differences in how antibodies perform in ChIP experiments, with only a fraction performing well. Further, searches for a specific antibody allow the users to pick primers for genes that can be used as positive controls to verify successful ChIP experiments. Another classic use case is to retrieve primers for genes of interest. Even if the database does not have primers for certain genes for the same antibody intended for use, the primers listed for individual gene regions are a good starting point and will perform well in many cases.

The web interface of ChIPprimersDB offers researchers a fast and convenient way to search for quality ChIP-primer sequences, compare published datasets and also submit own primer sequences to foster continuous growth. Together, this repository serves as an important tool for researchers planning ChIP experiments.

FUNDING

National Cancer Institute grants R01 [CA125970, CA161870 to J.W.H.]; Department of Defense grant [W81XWH-15-1-0578 to J.W.H.]; Bankhead-Coley Cancer Research Program [to J.W.H.]; Florida Academic Cancer Center Alliance [to J.W.H.]; Research to Prevent Blindness [to J.W.H.]; Inc. Senior Scientific Investigator Award [to J.W.H.]; 2015 RRF/Kayser Global Pan-American Award [to J.W.H.]; Alcon Research Institute [to J.W.H.]; Sylvester Comprehensive Cancer Center [to J.W.H.]; AACR-Ocular Melanoma Foundation Fellowship in honor of Robert C. Allen, MD [15-40-39-KURT to S.K.]; a generous gift from Dr Mark J. Daily [to J.W.H.]; The Bascom Palmer Eye Institute also received funding from a National Institute of Health Core [P30EY014801]; Department of Defense Grant [W81XWH-13-10048]; Research to Prevent Blindness Unrestricted Grant. Funding for open access charge: Sylvester Comprehensive Cancer Center.

Conflict of interest statement. None declared.

REFERENCES

- Mullis,K., Faloona,F., Scharf,S., Saiki,R., Horn,G. and Erlich,H. (1992) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Biotechnology*, **24**, 17–27.
- Mullis,K., Faloona,F., Scharf,S., Saiki,R., Horn,G. and Erlich,H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.*, **51**, 263–273.
- Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350–1354.
- Bartlett,J.M. and Stirling,D. (2003) A short history of the polymerase chain reaction. *Methods Mol. Biol.*, **226**, 3–6.
- Higuchi,R., Fockler,C., Dollinger,G. and Watson,R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, **11**, 1026–1030.
- Heid,C.A., Stevens,J., Livak,K.J. and Williams,P.M. (1996) Real time quantitative PCR. *Genome Res.*, **6**, 986–994.
- Cui,W., Taub,D.D. and Gardner,K. (2007) qPrimerDepot: a primer database for quantitative real time PCR. *Nucleic Acids Res.*, **35**, D805–D809.
- Gubelmann,C., Gattiker,A., Massouras,A., Hens,K., David,F., Decouttere,F., Rougemont,J. and Deplancke,B. (2011) GETPrime: a gene- or transcript-specific primer database for quantitative real-time PCR. *Database (Oxford)*, **2011**, bar040.
- Han,S. and Kim,D. (2006) AtRTPrimer: database for Arabidopsis genome-wide homogeneous and specific RT-PCR primer-pairs. *BMC Bioinformatics*, **7**, 179.
- Hu,Y., Sopko,R., Foos,M., Kelley,C., Flockhart,I., Ammeux,N., Wang,X., Perkins,L., Perrimon,N. and Mohr,S.E. (2013) FlyPrimerBank: an online database for *Drosophila melanogaster* gene expression analysis and knockdown evaluation of RNAi reagents. *G3 (Bethesda)*, **3**, 1607–1616.
- Lefever,S., Vandesompele,J., Speleman,F. and Pattyn,F. (2009) RTPrimerDB: the portal for real-time PCR primers and probes. *Nucleic Acids Res.*, **37**, D942–D945.
- Pattyn,F., Robbrecht,P., De Paepe,A., Speleman,F. and Vandesompele,J. (2006) RTPrimerDB: the real-time PCR primer and probe database, major update 2006. *Nucleic Acids Res.*, **34**, D684–D688.
- 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.
- Spandidos,A., Wang,X., Wang,H., Dragnev,S., Thurber,T. and Seed,B. (2008) A comprehensive collection of experimentally validated primers for Polymerase Chain Reaction quantitation of murine transcript abundance. *BMC Genomics*, **9**, 633.
- Spandidos,A., Wang,X., Wang,H. and Seed,B. (2010) PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Res.*, **38**, D792–D799.
- Wang,X. and Seed,B. (2003) A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res.*, **31**, e154.
- Wang,X., Spandidos,A., Wang,H. and Seed,B. (2012) PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.*, **40**, D1144–D1149.
- Zeisel,A., Yizhaky,A., Bossel Ben-Moshe,N. and Domany,E. (2013) An accessible database for mouse and human whole transcriptome qPCR primers. *Bioinformatics*, **29**, 1355–1356.
- Ye,J., Coulouris,G., Zaretskaya,I., Cutcutache,I., Rozen,S. and Madden,T.L. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, **13**, 134.
- Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
- Johnson,D.S., Mortazavi,A., Myers,R.M. and Wold,B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**, 1497–1502.
- Mikkelsen,T.S., Ku,M., Jaffe,D.B., Issac,B., Lieberman,E., Giannoukos,G., Alvarez,P., Brockman,W., Kim,T.K., Koche,R.P. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
- Kibbe,W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.*, **35**, W43–W46.
- Landt,S.G., Marinov,G.K., Kundaje,A., Kheradpour,P., Pauli,F., Batzoglou,S., Bernstein,B.E., Bickel,P., Brown,J.B., Cayting,P. *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.*, **22**, 1813–1831.