




Check for updates

SOFTWARE TOOL ARTICLE

REVISED RBP-Tar – a searchable database for experimental RBP binding sites

[version 3; peer review: 2 approved, 4 approved with reservations]

Katarina Gresova^{1,2}, Tomas Racek^{1,2}, Vlastimil Martinek^{1,2}, David Cechak^{1,2}, Radka Svobodova^{1,2}, Panagiotis Alexiou ²⁻⁴

¹National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic

²Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czech Republic

³Centre for Molecular Medicine & Biobanking, University of Malta, Msida, Malta

⁴Department of Applied Biomedical science, University of Malta, Msida, Malta

v3 First published: 27 Jun 2023, 12:755
<https://doi.org/10.12688/f1000research.131014.1>
Second version: 12 Aug 2024, 12:755
<https://doi.org/10.12688/f1000research.131014.2>
Latest published: 25 Nov 2024, 12:755
<https://doi.org/10.12688/f1000research.131014.3>

Abstract

Background

RNA-binding proteins (RBPs) play a critical role in regulating gene expression by binding to specific sites on RNA molecules. Identifying these binding sites is crucial for understanding the many functions of RBPs in cellular function, development and disease. Current experimental methods for identifying RBP binding sites, such as ultra-violet (UV) crosslinking and immunoprecipitation (CLIP), and especially the enhanced CLIP (eCLIP) protocol, were developed to identify authentic RBP binding sites experimentally.

Methods

To make this data more accessible to the scientific community, we have developed RBP-Tar (<https://ncbr.muni.cz/RBP-Tar>), a web server and database that utilises eCLIP data for 167 RBPs mapped on the human genome. The web server allows researchers to easily search and retrieve binding site information by genomic location and RBP name.

Use case

Researchers can produce lists of all known RBP binding sites on a gene of interest, or produce lists of binding sites for one RBP on

Open Peer Review

Approval Status

	1	2	3	4	5	6
version 3						
(revision)						
25 Nov 2024						
version 2						
(revision)						
12 Aug 2024						
version 1						
27 Jun 2023						

1. **Xiaoyong Pan** , Shanghai Jiao Tong University, Shanghai, China
2. **Clifford A Meyer**, Harvard T.H. Chan School of Public Health, Dana-Farber Cancer Institute, Boston, USA
3. **Michiel Stock**, Ghent University, Ghent, Belgium
4. **Eric Van Nostrand**, Baylor College of Medicine, Houston, USA
5. **M. J. Nishanth**, St Joseph's University, Bengaluru, India

different genomic loci.

Conclusions


Our future goal is to continue to populate the web server with additional experimental datasets from CLIP experiments as they become available and processed, making it an increasingly valuable resource for the scientific community.

Keywords

RNA Binding Proteins, CLIP, RBP, Web-server



This article is included in the **Bioinformatics** gateway.

6. Zhi-Ping Liu , Shandong University, Jinan, China

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Panagiotis Alexiou (panagiotis.alexiou@um.edu.mt)

Author roles: **Gresova K:** Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Racek T:** Resources, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Martinek V:** Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Cechak D:** Data Curation, Investigation, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Svobodova R:** Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Alexiou P:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic. Computational resources were provided by the ELIXIR-CZ project (LM2018131), part of the international ELIXIR infrastructure. Core Facility Biological Data Management and Analysis of CEITEC Masaryk University, supported by ELIXIR CZ research infrastructure (MEYS Grant No: LM2018131), is gratefully acknowledged for obtaining the scientific data presented in this paper. The Horizon Europe, ERA Chair grant (BioGeMT, ID: 101086768) supplied funding for publication fees.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Gresova K *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Gresova K, Racek T, Martinek V *et al.* **RBP-Tar – a searchable database for experimental RBP binding sites [version 3; peer review: 2 approved, 4 approved with reservations]** F1000Research 2024, 12:755 <https://doi.org/10.12688/f1000research.131014.3>

First published: 27 Jun 2023, 12:755 <https://doi.org/10.12688/f1000research.131014.1>

REVISED Amendments from Version 2

In this update, we have added a p-value column to provide statistical context for evaluating RNA binding protein binding sites. Additionally, we introduced filtering and sorting by p-value to enhance usability, allowing users to prioritize data based on significance.

Any further responses from the reviewers can be found at the end of the article

Introduction

RNA binding proteins (RBPs) are key players in a broad spectrum of RNA regulation, including all stages of the RNA lifecycle (Gerstberger *et al.* 2014; Gebauer *et al.* 2021). Eukaryotic genomes typically encode hundreds of RBPs. For example, over 1500 human RBPs involved in the maturation, transport, stability and translation of coding and non-coding RNA were recently characterised and manually curated (Gerstberger *et al.* 2014). Each RBP can typically target hundreds of RNAs in a complex coordinated fashion (Hogan *et al.* 2008). The general transcriptomic locations of thousands of RNA binding sites corresponding to hundreds of RBPs have been identified using a family of experimental techniques based on RBP CrossLinking, ImmunoPrecipitation and Sequencing (CLIP-Seq) (Chi *et al.* 2009; Licatalosi *et al.* 2008; Moore *et al.* 2014; Hafner *et al.* 2010). A thorough exploration of tens of RBPs binding characteristics in vitro has shown that RBPs can differentiate their binding sites with context preferences beyond narrowly defined binding sequence motif and secondary structure often involving complex binding configurations (Dominguez *et al.* 2018). Among the experimental techniques available to date, the enhanced CLIP (eCLIP) protocol (Van Nostrand *et al.* 2016) is particularly important, as it significantly reduces required amplification and increases specificity in identifying authentic binding sites. This improves the efficiency and accuracy of RBP binding site identification and allows for a deeper understanding of the role of RBPs in gene regulation.

The availability of a large amount of data points produced from the same experimental technique can be very beneficial for applications such as machine learning, as it allows researchers to train and test models with more confidence. Here we present RBP-Tar (Tomáš Raček 2023), a centralised and searchable database of experimentally identified RBP binding sites that can significantly facilitate the study of the RBP mediated gene regulation. Using RBP-Tar, researchers can quickly and cleanly retrieve RBP binding sites constrained by both genomic location and associated RBP for hundreds of RBPs.

Methods**Implementation****Pipeline for reproducible data download and annotation**

We have developed a reproducible and easy-to-use pipeline for downloading and annotating RBP eCLIP data from the ENCODE (Luo *et al.* 2020) database.

Metadata for eCLIP experiments, as well as additional files containing genomic coordinates are downloaded from the ENCODE database with the following parameters:

```
"status= released&
internal_tags=ENCORE&
assay_title=eCLIP&
biosample_ontology.term_name=K562&
biosample_ontology.term_name=HepG2&
files.file_type=bed+narrowPeak&
type=Experiment
files.analyses.status=released".
```

Following the download, information about the chromosome, start position, end position, strand and p values of reads are extracted for each RBP binding site. Binding sites are filtered by length, excluding ones shorter than 20 and longer than 100 nucleotides. As a last step, genomic sequences of binding sites are retrieved.

The described pipeline is implemented as a set of python scripts and is freely available at [GitHub](#).

Using this pipeline, data for 168 RBPs on two cell lines (K562, HepG2) were downloaded. In total, 42 MB of data, representing more than 400 thousand binding sites, were thus processed.

Web server

Here we present RBP-Tar, a web server that can access the above-curated dataset of RBP binding sites (<https://rbp-tar.biodata.ceitec.cz/>) and was built with Python (RRID:SCR_008394), the web development framework Flask, and a simple SQLite database (RRID:SCR_017672). The application's source codes can be found on the project's [GitHub](#) page, along with the requirements and instructions for the deployment if a user wants to run the application locally.

The web user interface allows searching and filtering based on the start and end position of the binding site, strand, chromosome, p value and protein name. It offers the download of the filtered data based on the search done by the user. Due to the size of the dataset, the view is limited to 10,000 results. However, the whole dataset can be conveniently downloaded as a gzipped CSV (14 MB) (https://rbp-tar.biodata.ceitec.cz/download_all).

Operation

The RBP-Tar web server can take as input any of the following user-provided parameters: [Start min, Start max] denote the limits of the low genomic coordinate of the locus of interest. Similarly, [End min, End max] denote the limits of the high coordinate. [Strand] and [Chromosome] can be used to narrow down the search to only one strand and a specific chromosome and [-log10(p-value) min, -log10(p-value) max] allows for filtering based on the significance of binding sites. Using these combinations of parameters, a user can easily search for binding sites on their favourite gene, exon, or even a whole chromosome. The last parameter is [Protein name], which brings out a drop-down menu of all the RBPs in the database. If this parameter is not set, all RBPs are queried.

Results are shown as a table with the [Chromosome, Start, End, Strand] genomic location of the binding site, followed by the [-log10(p-value)] p-value of the binding site, [Protein name] of the associated RBP and the [Sequence] of the binding site contains the genomic sequence of the binding site. Results can be seen online in a table format or downloaded as a CSV file with one button click. We expect most users to download the results and use them for further downstream analyses.

Start min:
Start max:
End min:
End max:
Strand:
Chromosome:
Protein name:

Chromosome	Start	End	Strand	Protein name	Sequence
chr16	31187085	31187118	+	AATF	GGCCTTGACTGGGCCGTTGCCACACTGGCAC
chr16	31190303	31190397	+	APOBEC3C	GGTGGCAGTGGTGGTGGTGGCCGAGGAGGATTCCAGTGGAGGTGGCGGTGGAGGAC
chr16	31191081	31191101	+	APOBEC3C	GGCTTTGGCCCTGGCAAGAT
chr16	31191023	31191081	+	APOBEC3C	GGGCCGCGCGGGGACCGTGGAGGCTCCGAGGGGGCGGGGTGGTGGGGACAGAGGT
chr16	31190964	31190999	+	APOBEC3C	GGTAACACGGGGATGATCGTCGTGGTGGCAGAGG
chr16	31191453	31191496	+	APOBEC3C	AGGTTCTGGAACAGCTTTTGTCTGTACCCAGTGTACCTC
chr16	31191498	31191550	+	APOBEC3C	TATTTTGAACCTTCAATTCCTGATACCCAAGGGTTTTTTGTGTGCGGAC
chr16	31191550	31191587	+	APOBEC3C	TATGTAATTGTAACATACCTCTGGTTCCCATTA
chr16	31190107	31190141	+	APOBEC3C	GGTGGCAATGGTCGTGGAGGCCGAGGGCGAGGAG
chr16	31190777	31190841	+	APOBEC3C	ATGCAACCATGTGAAGGCCCTAAACCAGATGGCCAGGAGGGGACCAAGTGGCTCTCACATC

Showing 1 to 10 of 334 entries

Figure 1. Extracting binding sites of all RNA-binding proteins on one locus (use case 1).

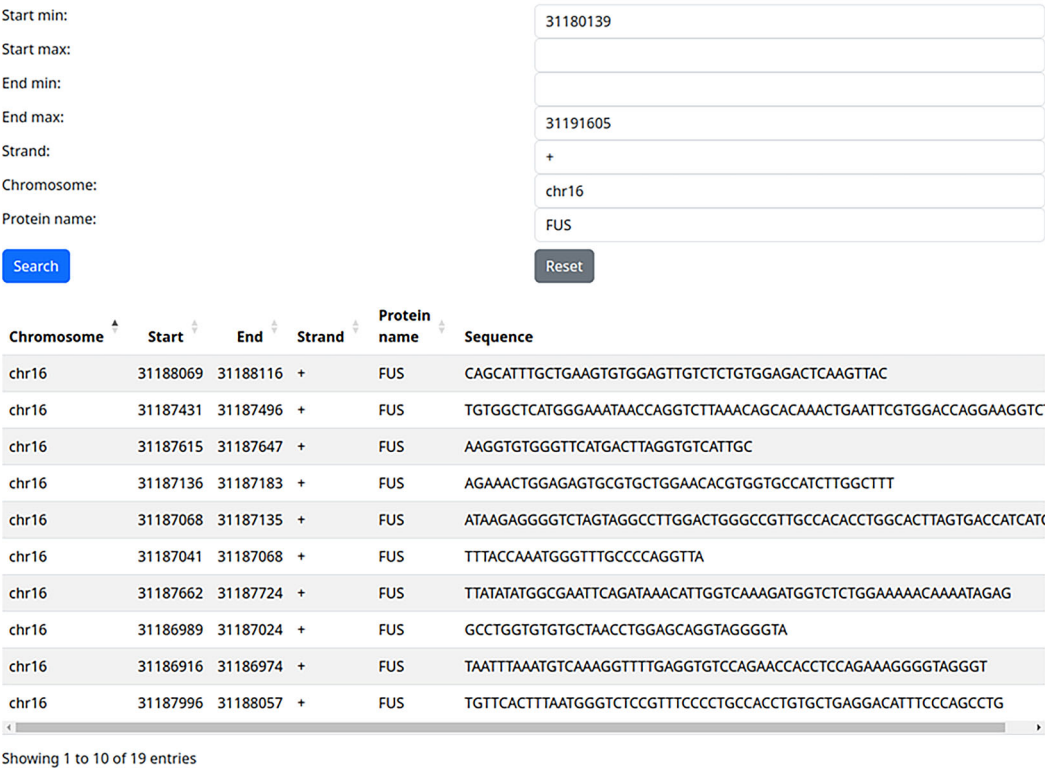


Figure 2. Fused in Sarcoma (FUS) self-targeting identified by filtering by gene location and protein name.

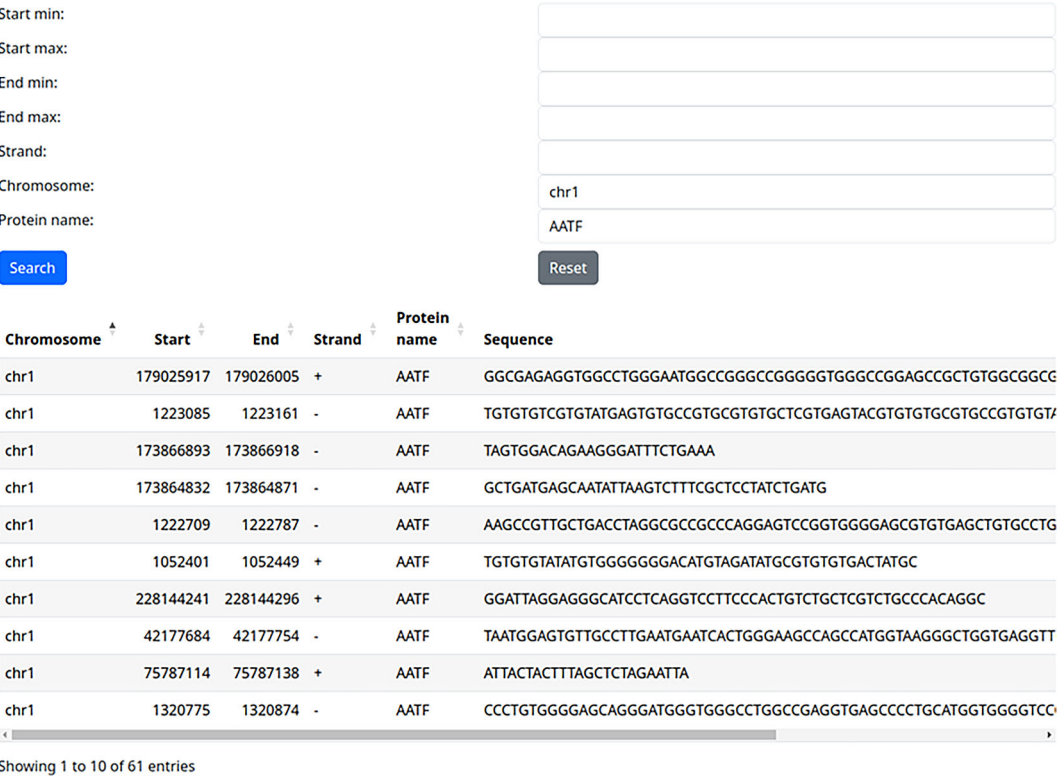


Figure 3. Extracting all binding sites of a single protein on a single chromosome (use case 2).

Use cases

Use case 1: all known RBP binding sites on the gene of interest

The first and potentially most common use case would be the query of all known RBP binding sites on a gene of interest. For example, we can query our web server with the coordinates of the Fused in Sarcoma (FUS) gene (chr16: 31180139-31191605, +) and leave the protein field empty. After this search, all 345 known RBP binding sites on this gene are returned and can be easily downloaded in a CSV file for further analysis (Figure 1).

In fact, the gene we used encodes the RBP FUS, which plays important roles, among others, in neurodegeneration and cancer progression. We can use the filter [Protein Name] to identify the potential self-targeting of FUS on itself. Indeed, we can thus identify 19 potential FUS self-targeting binding sites identified *via* eCLIP (Figure 2). Of course, the biological relevance of this type of finding is left to the users, as is the further validation of high-throughput derived RBP binding sites.

Use case 2: training/testing dataset for one RBP

A potential user may want to develop an RBP binding site machine learning tool that would be able to predict binding sites based on a sequence. It is important to make sure that training and testing sets for their machine learning method are not overlapping. Using our web server, they can download all binding sites for a specific RBP, for example, on chromosome 1, and use them as a training set (Figure 3). Then, do the same for chromosome 2 and use it as an independent testing set, thus ensuring that the training and testing sets do not overlap.

Conclusion

Recent advancements in experimental techniques, such as eCLIP, have greatly expanded our understanding of RBP binding preferences and their role in gene regulation. The development of centralised and searchable databases of experimentally identified RBP binding sites allows researchers to access and analyse the binding preferences of RBPs easily. This information can be used to identify known binding sites on genes of interest and aid in training machine-learning models for RBP binding site prediction. This paper presents RBP-Tar, a centralised web server that can retrieve RBP Target sites with location and RBP constraints. RBP-Tar has been designed to be easily accessible by non-experts. It is still confined to a single source of data, which is helpful for avoiding experimental design effects, but makes its scope limited. We plan to expand the web server with other sources of data, as well as ways for the user to be able to take into account provenance and experimental variation.

Data availability

Source data

All data used in RBP-Tar has been downloaded from ENCODE and Ensembl projects in November 2024. RBP eCLIP metadata were downloaded from https://www.encodeproject.org/metadata/?status=released&internal_tags=ENCORE&assay_title=eCLIP&biosample_ontology.term_name=K562&biosample_ontology.term_name=HepG2&files.file_type=bed+narrowPeak&type=Experiment&files.analysis.status=released. A list of all downloaded files containing genomic coordinates can be found here: https://github.com/ML-Bioinfo-CEITEC/rbp_encode_eclip/blob/main/csv/coord_links.txt. The reference genome was downloaded from http://ftp.ensembl.org/pub/release-97/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.toplevel.fa.gz

Columns 'chr', 'start', 'end', 'strand' and 'pValue' from downloaded files containing genomic coordinates can be found here: https://github.com/ML-Bioinfo-CEITEC/rbp_encode_eclip/tree/main/csv. The whole dataset (curated and including sequence) can be downloaded from <https://rbp-tar.biodata.ceitec.cz/> as a zipped CSV (18 MB) (https://rbp-tar.biodata.ceitec.cz/download_all).

Software availability

RBP-Tar

Software: <https://rbp-tar.biodata.ceitec.cz/>

Source code: <https://github.com/sb-ncbr/RBP-Tar/releases/tag/v1.0.2>

Archived source code at the time of publication: <https://doi.org/10.5281/zenodo.14135388> (Raček 2024)

License: MIT

eCLIP RBP

Source code: https://github.com/ML-Bioinfo-CEITEC/rbp_encode_eclip/tree/v1.0.3

Archived source code at the time of publication: <https://doi.org/10.5281/zenodo.14082502> (Grešová & Raček 2024)

License: [Apache 2.0](#)

References

- Chi SW, Zang JB, Mele A, *et al.*: **Argonaute HITS-CLIP Decodes microRNA-mRNA Interaction Maps.** *Nature*. 2009; **460**: 479–486.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dominguez D, Freese P, Alexis MS, *et al.*: **Sequence, Structure, and Context Preferences of Human RNA Binding Proteins.** *Mol. Cell* 2018; **70**(5): 854–67.e9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gerstberger S, Hafner M, Tuschl T: **A Census of Human RNA-Binding Proteins.** *Nat. Rev. Genet.* 2014; **15**(12): 829–845.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gebauer F, Schwarzl T, Valcárcel J, *et al.*: **A-binding proteins in human genetic disease.** *Nat. Rev. Genet.* 2021; **22**: 185–198.
[Publisher Full Text](#)
- Grešová K, Raček T: **"ML-Bioinfo-CEITEC/rbp_encode_eclip: v1.0.3 (v1.0.3)."** 2024.
[Publisher Full Text](#)
- Hafner M, Landthaler M, Burger L, *et al.*: **Transcriptome-Wide Identification of RNA-Binding Protein and microRNA Target Sites by PAR-CLIP.** *Cell*. 2010; **141**(1): 129–141.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hogan DJ, Riordan DP, Gerber AP, *et al.*: **Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System.** *PLoS Biol.* 2008; **6**(10): e255.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Licatalosi DD, Mele A, Fak JJ, *et al.*: **HITS-CLIP Yields Genome-Wide Insights into Brain Alternative RNA Processing.** *Nature* 2008; **456**(7221): 464–469.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Luo Y, Hitz BC, Gabdank I, *et al.*: **New developments on the Encyclopedia of DNA Elements (ENCODE) data portal.** *Nucleic Acids Res.* 2020 Jan 8; **48**(D1): D882–D889.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Moore MJ, Zhang C, Gantman EC, *et al.*: **Mapping Argonaute and Conventional RNA-Binding Protein Interactions with RNA at Single-Nucleotide Resolution Using HITS-CLIP and CIMS Analysis.** *Nat. Protoc.* 2014; **9**(2): 263–293.
[Publisher Full Text](#)
- Nostrand V, Eric L, Pratt GA, *et al.*: **Robust Transcriptome-Wide Discovery of RNA-Binding Protein Binding Sites with Enhanced CLIP (eCLIP).** *Nat. Methods* 2016; **13**(6): 508–514.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Raček T: **"sb-ncbr/RBP-Tar: v1.0.2 (v1.0.2)."** 2024.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:



Version 3

Reviewer Report 16 January 2025

<https://doi.org/10.5256/f1000research.175053.r354325>

© 2025 Liu Z. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Zhi-Ping Liu 

Shandong University, Jinan, Shandong, China

In this paper, the authors present the development of RBP-Tar, a searchable database and web server dedicated to experimental RNA-binding protein (RBP) binding sites. RBP-Tar currently focuses on eCLIP data for 167 RBPs mapped on the human genome, though its scope has the potential to be expanded to include a broader range of CLIP data to enhance its utility.

To address the current limitation of focusing solely on eCLIP data, the authors suggest exploring the inclusion of other CLIP data types in future updates. Additionally, to clarify the purpose and content of this paper, the authors recommend refining the title and abstract to more accurately reflect the scope and contributions of their work. Furthermore, to provide context for their work, the authors need include a discussion of existing databases similar to RBP-Tar. This section should summarize the features and specificity of RBP-Tar compared to these existing resources. If possible, a comparison of the overlapping information with these databases would be beneficial to highlight the unique aspects of RBP-Tar. For the web server component of RBP-Tar, the authors need consider incorporating submission and download options for original data and query results. These features would enhance the usability and accessibility of the server, making it a more comprehensive resource for researchers studying RNA-binding proteins.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets

and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, Systems Biology.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 16 January 2025

<https://doi.org/10.5256/f1000research.175053.r354315>

© 2025 Nishanth M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



M. J. Nishanth

St Joseph's University, Bengaluru, India

The authors have developed RBP-Tar, a database and tool that can be used to identify the known target motifs of RBPs on human genes or chromosomal loci. This would be a useful tool for scientific community to identify RBP target motifs. Following are the technical queries/suggestions for the authors:

1. The authors state that the target motifs < 20 bases in length or >100 bases are excluded from the result-list. Why are the motifs < 20 bases excluded? The rationale for this parameter needs to be given.
2. It may be beneficial to add an 'evidence' column in the result/output table. This column could have the reference of the primary study/experiment which forms the evidence for the particular target site being displayed in the result table. This would be helpful for the users if further validation of the target motifs is needed.
3. The authors could consider adding a summary of the existing software tools in the realm of RBP target motif identification and related bioinformatic analyses. They could elaborate on the utility of RBP-Tar in the context of the existing software tools, which would give a broader perspective of the usefulness of RBP-Tar. A few of the existing relevant tools that could be discussed: RBP Map (<https://rbpmap.technion.ac.il/>), Transite (<https://transite.mit.edu/>), oRNAmEnt (<https://rnabiology.ircm.qc.ca/oRNAmEnt/>), and ATtRACT (<https://attract.cnice.es/>).
4. The authors need to detail the statistical tests and analyses on which the p-value calculations are based.

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular biology, gene regulation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 10 January 2025

<https://doi.org/10.5256/f1000research.175053.r349169>

© 2025 Van Nostrand E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Eric Van Nostrand

Baylor College of Medicine, Houston, TX, USA

This article by Gresova et al build a searchable web tool for the ENCODE eCLIP data that profiles RNA binding protein interactions.

Since this is a named review, I can specifically comment - as one other reviewer notes, some of the filtering choices here are done without really justifying why they are appropriate - i.e., why are peaks shorter than 20 or longer than 100nt excluded? That's something that should either be justified with analysis showing that it improves signal-to-noise, or not done. (In particular, the CLIP per peak caller used in ENCODE tends to err on the side of calling short peaks, so I highly suspect filtering peaks shorter than 20 is removing a significant amount of real signal)

Similarly, we reported the fold-enrichment versus input because we found that to often be more informative than p-value (which is often more a function of transcript abundance rather than binding signal), but that is not reported here for an undescribed reason.

Since the entire website is based off of the ENCODE eCLIP data, it seems appropriate to cite that manuscript (Van Nostrand EL, et al., 2020 - PMID [32728246](#) [Ref-1]) rather than the broader ENCODE overview Luo et al publication.

If the idea is to make it extremely user-friendly, it seems like having an ability to copy-paste a UCSC browser-style region (e.g. "chr1:xxx-yyy") would make it simpler to use

References

1. Van Nostrand EL, Freese P, Pratt GA, Wang X, et al.: A large-scale binding and functional map of human RNA-binding proteins. *Nature*. 2020; **583** (7818): 711-719 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: eCLIP, RNA binding protein studies

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 09 January 2025

<https://doi.org/10.5256/f1000research.175053.r349176>

© 2025 Stock M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Michiel Stock**

Ghent University, Ghent, Belgium

This article presents RBP-tar, a small database with a web interface that enables users to access RNA-binding proteins in humans. Users can select regions on different chromosomes and view a list of all binding proteins. Additionally, there are various filter options, such as filtering by protein name for self-binding. Users can also directly download a CSV file containing the entire database, which I always appreciate. The database is designed to facilitate the development of machine learning models.

The short article appears to be well explained. However, I find one aspect lacking: the comparison and relationship with similar databases, such as RBP2GO or EuRPDB. This should be addressed in a designated section.

I would also like some information on how this will be maintained. Will more RBPs be added?

Furthermore, I would appreciate some additional clarification regarding what the p-values refer to.

minor:

- "p values" => "p-values" (clarify what the p-values are for).
- what is exactly the " $[-\log_{10}(\text{p-value})]$ p-value of the binding site"?

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics and machine learning.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Version 2

Reviewer Report 05 September 2024

<https://doi.org/10.5256/f1000research.169056.r313335>

© 2024 Meyer C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Clifford A Meyer

Harvard T.H. Chan School of Public Health, Dana-Farber Cancer Institute, Boston, USA

The revision is not satisfactory as most of the reviewers' suggestions for improving the website were not implemented. We hope that these will be implemented in future versions.

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epigenetics, Chromatin analysis, Gene regulation, Single cell

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 19 Nov 2024

Panagiotis Alexiou

Thank you for the continued feedback on our manuscript and webserver resource. We appreciate the reviewer's commitment to ensuring that the resource becomes more robust and user-friendly. In response to the earlier suggestions, we have implemented the following improvements in our webserver, as outlined below:

1. Inclusion of p-values: We have added a p-value column to our data, providing an additional layer of statistical context to aid users in evaluating the significance of RNA binding protein binding sites.
2. Filtering and sorting functionality by p-value: To enhance usability, we have introduced the functionality to filter and sort binding site data based on p-values. This improvement allows users to more easily interpret and prioritize results according to their significance.

We recognize that many of the enhancements suggested in previous rounds, such as expanding the dataset scope and incorporating broader quality control metrics, require a significant overhaul and are part of our longer-term development roadmap.

Our current updates reflect initial, impactful steps toward addressing reviewer concerns, and we remain committed to further expanding the features and datasets in future versions.

Competing Interests: No competing interests were disclosed.

Reviewer Report 23 August 2024

<https://doi.org/10.5256/f1000research.169056.r313334>

© 2024 Pan X. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Xiaoyong Pan 

Shanghai Jiao Tong University, Shanghai, Shanghai, China

The authors addressed my comments.

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 18 December 2023

<https://doi.org/10.5256/f1000research.143817.r224195>

© 2023 Meyer C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Clifford A Meyer

Harvard T.H. Chan School of Public Health, Dana-Farber Cancer Institute, Boston, USA

This paper describes a database of RNA binding protein binding sites based on 2 ENCODE cell lines. The website allows users to search for data by querying a gene or genomic interval.

The scope of the database is very limited. It is not clear why the data is restricted to eCLIP data and not CLIP data in general. Although the authors make the argument that eCLIP data is of good quality, quality will vary from sample to sample and there are likely to be high quality CLIP data sets as well as low quality eCLIP ones.

It appears that the processed peaks are downloaded from ENCODE, but the methods used by ENCODE to identify these peaks are not described, and statistics such as p-values, false discovery rates and enrichment relative to background are not provided. Quality controls are also not provided, so it is not possible to know if the experiment worked.

As processed RBP data is downloaded from ENCODE this resource does not add much to what is already available on the ENCODE website. A resource that processed the raw read data from other data resources, such as GEO, would be more valuable. There are over 2,000 eCLIP samples and over 6,000 CLIP samples in GEO.

The search features are useful, although it is hard to interpret the results. Some ranking of the results from most significant to least is needed. It would also be helpful to enable filtering based on some binding sites statistics.

Overall, the resource is limited in terms of the data included and the website features. It would be helpful to include quality control metrics to let users know which samples are more informative.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epigenetics, Chromatin analysis, Gene regulation, Single cell

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 11 Jul 2024

Panagiotis Alexiou

This paper describes a database of RNA binding protein binding sites based on 2 ENCODE cell lines. The website allows users to search for data by querying a gene or genomic interval. The scope of the database is very limited. It is not clear why the data is restricted to eCLIP data and not CLIP data in general. Although the authors make the argument that eCLIP data is of good quality, quality will vary from sample to sample and there are likely to be high quality CLIP

data sets as well as low quality eCLIP ones.

We agree that the scope of the database is limited at this point and we aim to widen it in the future updates of the app. ENCODE/ENCORE has a comprehensive eCLIP experimental list, and only 3 iCLIP libraries (https://www.encodeproject.org/encode-matrix/?type=Experiment&status=released&internal_tags=ENCORE). We chose eCLIP to keep the experimental procedure more consistent in case of different methods introducing different biases. In a future version, we will explore using different datasets from multiple sources.

It appears that the processed peaks are downloaded from ENCODE, but the methods used by ENCODE to identify these peaks are not described, and statistics such as p-values, false discovery rates and enrichment relative to background are not provided. Quality controls are also not provided, so it is not possible to know if the experiment worked.

Methods used by ENCODE to identify peaks from eCLIP data are described here: <https://www.encodeproject.org/eclip/> while the detailed description of data processing pipeline can be downloaded from here: https://www.encodeproject.org/documents/739ca190-8d43-4a68-90ce-1a0ddfffc6fd/@@download/attachment/eCLIP_analysisSOP_v2.2.pdf. This pipeline ensures that only high-quality peaks are identified.

As processed RBP data is downloaded from ENCODE this resource does not add much to what is already available on the ENCODE website. A resource that processed the raw read data from other data resources, such as GEO, would be more valuable. There are over 2,000 eCLIP samples and over 6,000 CLIP samples in GEO.

We agree that a resource that processed the raw read data from other data resources, such as GEO, would be more valuable, we will consider adding other data resources in the future updates of the webserver. The addition of our app to what is already available on the ENCODE website is advanced filtering of binding sites based on the genome position.

Competing Interests: No competing interests were disclosed.

Reviewer Report 18 December 2023

<https://doi.org/10.5256/f1000research.143817.r224199>

© 2023 Pan X. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Xiaoyong Pan

Shanghai Jiao Tong University, Shanghai, Shanghai, China

1. In the paper, the author claim they download narrowPeak metadata from https://www.encodeproject.org/metadata/?status=released&internal_tags=ENCORE&assay_title=eCLIP&..., and the link summary is provided in the file https://github.com/ML-Bioinfo-CEITEC/rbp_encode_eclip/blob/main/csv/coord_links.txt. In the ENCODE dataset, the majority of RNA-binding proteins (RBPs) have biological replicate(s) or technical replicate(s), with each replicate having an associated narrowPeak file. Additionally, there is another merged peak file containing reproducible peaks as determined by entropy-ordered peaks between two replicates. However, it is observed that only one file per RBP from K562 or HepG2, which is both the merged peak file. For eCLIP experiment on K562 against QKI, I just find ENCF190XSX that is from replicate 1. A careful examination of the data is recommended to ensure accuracy and completeness.
2. After narrowPeak download, the author filter the peak by length, excluding ones shorter than 20 and longer than 100 nucleotides. It's important to note that must short (rarely more than 4–6-nucleotide-long) sequence elements within RNA targets that are recognized and bound by RNA-binding proteins (Gerstberger, S et al. (2014)¹). In my perspective, employing a filter based on fold change and p-value, in accordance with the eCLIP-seq Processing Pipeline of ENCODE (https://www.encodeproject.org/documents/739ca190-8d43-4a68-90ce-1a0ddfffc6fd/@@download/attachment/eCLIP_analysisSOP_v2.2.pdf) is powerful.
3. If this job is for the benefit of biologists, why not use transcripts as a reference? Because RBP binds to transcripts. And it's too simplistic for systems biologists, the author can refer to the POSTAR3 data.

References

1. Gerstberger S, Hafner M, Tuschl T: A census of human RNA-binding proteins. *Nat Rev Genet.* 2014; **15** (12): 829-45 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: The authors need solve some issues before published

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 11 Jul 2024

Panagiotis Alexiou

Reviewer Comment:

In the paper, the author claim they download narrowPeak metadata from https://www.encodeproject.org/metadata/?status=released&internal_tags=ENCORE&assay_title=eCLIP&biosample_, and the link summary is provided in the file https://github.com/ML-Bioinfo-CEITEC/rbp_encode_eclip/blob/main/csv/coord_links.txt. In the ENCODE dataset, the majority of RNA-binding proteins (RBPs) have biological replicate(s) or technical replicate(s), with each replicate having an associated narrowPeak file. Additionally, there is another merged peak file containing reproducible peaks as determined by entropy-ordered peaks between two replicates. However, it is observed that only one file per RBP from K562 or HepG2, which is both the merged peak file. For eCLIP experiment on K562 against QKI, I just find ENCF190XSX that is from replicate 1. A careful examination of the data is recommended to ensure accuracy and completeness.

Author Response: We would like to thank the reviewer for the insightful finding. We have carefully examined the downloaded metadata and adjusted the processing pipeline to ensure the merged peak file is used for every RBP. In the methods, section "Pipeline for reproducible data download and annotation", we have added the parameter "files.analyses.status=released" that is ensuring all available files are downloaded. Updated data have been uploaded to the RBP-Tar webserver and the manuscript has been updated with the current number of binding sites and size of files. We have provided new figures with up-to-date views of the web server.

Reviewer Comment:

After narrowPeak download, the author filter the peak by length, excluding ones shorter than 20 and longer than 100 nucleotides. It's important to note that must short (rarely more than 4–6-nucleotide-long) sequence elements within RNA targets that are recognized and bound by RNA-binding proteins (Gerstberger, S et al. (2014) [1](#)). In my perspective, employing a filter based on fold change and p-value, in accordance with the eCLIP-seq Processing Pipeline of ENCODE (https://www.encodeproject.org/documents/739ca190-8d43-4a68-90ce-1a0ddfffc6fd/@@download/attachment/eCLIP_analysisSOP_v2.2.pdf) is powerful.

Author Response: We appreciate the reviewer's perspective on filtering criteria. Our decision to filter based on length aimed to exclude potential artifacts and prioritize biologically relevant binding sites. While fold change and p-value are valuable parameters, our focus was on ensuring the inclusion of robust binding sites within the specified length range. We will consider incorporating additional filtering parameters in future iterations leading to a new version of the web server.

Reviewer Comment:

If this job is for the benefit of biologists, why not use transcripts as a reference? Because RBP binds to transcripts. And it's too simplistic for systems biologists, the author can refer to the POSTAR3 data.

Author Response: Our choice to focus on genomic coordinates was driven by the goal of providing a comprehensive resource for accessing RBP binding sites across the genome. While transcripts offer valuable context, our approach allows for flexibility in analyzing binding sites within different genomic regions and across multiple transcripts. However, we recognize the importance of incorporating transcript-level information and will explore integrating such data in future updates of the webserver to enhance the utility of our tool for systems biologists.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research