



Article

Web-MCOT Server for Motif Co-Occurrence Search in ChIP-Seq Data

Victor G. Levitsky^{1,2,*} , Alexey M. Mukhin¹ , Dmitry Yu. Oshchepkov¹, Elena V. Zemlyanskaya^{1,2} and Sergey A. Lashin^{1,2} 

¹ Department of System Biology, Institute of Cytology and Genetics, 630090 Novosibirsk, Russia

² Department of Natural Science, Novosibirsk State University, 630090 Novosibirsk, Russia

* Correspondence: levitsky@bionet.nsc.ru

Abstract: (1) Background: The widespread application of ChIP-seq technology requires annotation of cis-regulatory modules through the search of co-occurred motifs. (2) Methods: We present the web server Motifs Co-Occurrence Tool (Web-MCOT) that for a single ChIP-seq dataset detects the composite elements (CEs) or overrepresented homo- and heterotypic pairs of motifs with spacers and overlaps, with any mutual orientations, uncovering various similarities to recognition models within pairs of motifs. The first (Anchor) motif in CEs respects the target transcription factor of the ChIP-seq experiment, while the second one (Partner) can be defined either by a user or a public library of Partner motifs being processed. (3) Results: Web-MCOT computes the significances of CEs without reference to motif conservation and those with more conserved Partner and Anchor motifs. Graphic results show histograms of CE abundance depending on orientations of motifs, overlap and spacer lengths; logos of the most common CE structural types with an overlap of motifs, and heatmaps depicting the abundance of CEs with one motif possessing higher conservation than another. (4) Conclusions: Novel capacities of Web-MCOT allow retrieving from a single ChIP-seq dataset with maximal information on the co-occurrence of motifs and potentiates planning of next ChIP-seq experiments.

Keywords: chromatin immunoprecipitation with massively parallel sequencing; transcription factors binding sites prediction; co-binding of transcription factors; composite elements; transcription factor binding sites; motifs conservation; overlap of motifs



Citation: Levitsky, V.G.; Mukhin, A.M.; Oshchepkov, D.Y.; Zemlyanskaya, E.V.; Lashin, S.A. Web-MCOT Server for Motif Co-Occurrence Search in ChIP-Seq Data. *Int. J. Mol. Sci.* **2022**, *23*, 8981. <https://doi.org/10.3390/ijms23168981>

Academic Editors: Yuriy L. Orlov, Anastasia A. Anashkina and Nikolay Kolchanov

Received: 19 July 2022

Accepted: 10 August 2022

Published: 11 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Commonly, two or more transcription factors (TFs) work coordinately to induce transcriptional change [1]. At the genome level, this leads to composite elements (CEs) as specific overrepresented combinations of two motifs in regulatory regions of genes. Two motifs may overlap each other, or a spacer separates them [2]. Chromatin immunoprecipitation followed by massive sequencing (ChIP-seq) specifically for given cells, tissue type and condition maps peaks in a genome and motivates the genome-wide search of binding sites (motifs) not only for the target (Anchor) TF but also for co-binding (Partner) TFs. Several recently developed databases provide uniformly processed genomic profiles for thousands of ChIP-seq experiments [3–5]. Widespread direct interactions of many TFs with DNA in chromatin cause the significant overrepresentation of their motifs in deduced peaks. Conventionally, a *de novo* motif search [6,7] may define among them the motif of Anchor TF, but the enrichment of other motifs only point to many possible Partner TFs. Thus, the CE search task for ChIP-seq data consists in the detection of overrepresented motifs of Partner TFs located nearby the Anchor motifs. The web tool SpaMo represented the first approach to detect the motif co-occurrence in a single ChIP-seq dataset [8]. It searched CEs only with spacers, though later studies indicate that CEs with overlaps of motifs are much more common than CEs with spacers [9,10]. Another approach [11,12] revealed CEs with

spacers and overlaps of motifs through the analysis of a special benchmark collection of multiple consistent ChIP-seq datasets. Unfortunately, due to the impossibility of many datasets being included in such consistent collections, this approach has not been widely applied and it was not implemented as a web tool. Accordingly, up to now, there are no web tools for predicting the significantly enriched co-occurrence of overlapped and spaced motifs in a single ChIP-seq dataset (Table 1).

Table 1. Comparison of various tools for prediction of CEs in ChIP-seq data.

Tool Name	A Single Dataset of Peaks Is Sufficient	Overlapped Motifs Are Allowed	URL	Reference
SpaMo	Yes	No	https://meme-suite.org/meme/tools/spamo (accessed on 9 August 2022)	[8]
TACO	No	Yes	http://bioputer.mimuw.edu.pl/taco/ (accessed on 9 August 2022)	[12]
MCOT, Web-MCOT	Yes	Yes	https://github.com/AcaDemIQ/mcot-kernel , https://webmcot.sysbio.cytogen.ru (accessed on 9 August 2022)	[10], this study

We recently developed the Motifs Co-Occurrence Tool (MCOT), which integrated the prediction of CEs with spacers and overlaps for a single ChIP-seq dataset [10,13]. In this study, we propose the web tool Web-MCOT promoting this approach. For a given Anchor motif and a dataset of input peaks, Web-MCOT tests the significance of CE overrepresentation for a user-defined Partner motif, or it checks the significance of CEs with a variety of Partner motifs from a library derived from genome-wide sequencing experiments [14,15]. Web-MCOT results allow the sorting of output data; hence, the structure and abundance of top-scoring CEs explain the impacts of Partner TFs supporting the interaction of an Anchor TF with DNA in chromatin.

2. Results

2.1. Input Data

The homepage of Web-MCOT represents the common description of the web server and proposes to proceed to data analysis (Figure 1A, the link ‘Application’). The trigram symbol ‘≡’ in the top left corner (Figure 1A) hides the menu and extends the input or output data page up to the whole width of screen. The application page (Figure 1B) requires the following input data: sequences of peaks (the option ‘Upload or Enter DNA sequences in FASTA format’); a nucleotide frequency matrix of an Anchor motif (the option ‘Upload or Enter Anchor motif’); a single Partner motif or a library of Partner motifs (the option ‘One or Many Partner motif(s) will be tested’).

A *de novo* motif search in peak sequences is the best option since it directly reflects the nucleotide context inherent to the Anchor motifs. Thus, the motifs enriched in ChIP-seq peaks are stored in a number of databases [3–5]. One of them, the Cistrome DB [3], provides ChIP-seq peaks integrated with the results of *de novo* motif search, i.e., enriched motifs as nucleotide frequency matrices. Alternatively, matrices respecting Anchor motifs can be derived indirectly from Hocomoco [14], CIS-BP [16], or JASPAR [17] databases. The ‘Many Partners’ option implies the setting of a public database, i.e., for one Anchor motif, many Partner motifs are tested at once. The Hocomoco database for mammals [14] represents the human/mouse core and full collections of 396/353 and 747/509 motifs, or the Plant Cistrome collection provides 514 motifs for *A. thaliana* [15]. The advanced option (Figure 1B) allows changing the limits of a spacer length. Pressing the ‘RUN’ button starts the calculation and provides a web link to the results. During a running process, the special indicator shows the percentage of completed calculations. The computational time may

vary from several minutes to a couple of hours depending on the number of peaks and their average length. The example input data illustrate the Web-MCOT functionality (Figure 1B, button 'EXAMPLE'). The example page (Figure 1C) contains peaks and the Anchor motif of the ChIP-seq dataset for FoxA2 from mouse liver [18].

A

Menu

- About
- Application
- Help
- Citing
- Authors
- Contacts
- Download & install

MCOT index

Motifs Co-Occurrence Tool

MCOT (Motifs Co-Occurrence Tool) is an approach for transcription factors (TFs) complexes prediction through motifs co-occurrence discovery. MCOT detects overrepresented pairs of TF binding sites (motifs) with an overlap or a spacer (composite elements, CEs) in a single ChIP-seq dataset (Levitsky et al, 2019, 2020). Typically, input data provides from 500 to ~10000 sequences of length from 200 bp to 1 kbp. Each potential CE contains the anchor motif of ChIP-seq target TF and another motif of partner TF. Optionally, one partner motif can be tested, or simultaneously hundreds of partner motifs form a public library are checked. Identical/distinct anchor and partner motifs imply a search of homotypic/heterotypic CEs. Also, MCOT checks whether CEs respect to significantly higher conservation of anchor or partner motif.

Levitsky V, Zemlyanskaya E, Oshchepkov D, Podkolodnaya O, Ignatieva E, Grosse I, Mironova V, Merkulova T. A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. *Nucleic Acids Res.* 2019; 47(21):e139. doi:10.1093/nar/gz280

Levitsky V, Oshchepkov D, Zemlyanskaya E, Merkulova T. Asymmetric conservation within pairs of co-occurred motifs mediates weak direct binding of transcription factors in ChIP-seq data. *Int. J. Mol. Sci.* 2020; 21(17):6023. doi:10.3390/ijms21176023

OPEN APPLICATION

B

The MCOT application discovers motifs co-occurrence with a spacer and with an overlap in a single ChIP-seq dataset

Upload or Enter DNA sequences in FASTA format ?

Upload file Enter sequences

Выберите файл файл не выбран

One or Many Partner motif(s) will be tested ?

One partner Many partners

Advanced options ?

Upload or Enter Anchor motif ?

Upload file Enter motif Choose motif

Выберите файл файл не выбран

Database of Partner motifs ?

Select the database with the proven recognition models of known transcription factors

Hocomoco v.11 Human core collection

RUN CLEAN

EXAMPLE

C

The MCOT application discovers motifs co-occurrence with a spacer and with an overlap in a single ChIP-seq dataset

Upload or Enter DNA sequences in FASTA format ?

Upload file Enter sequences

ChIP-seq data for FOXA2 supported by ocrevus.analysis

Upload or Enter Anchor motif ?

Upload file Enter motif Choose motif

Enter Anchor motif

```
>TGTTTACWYWRB 1-
TGTTTACWYWRB,BestGuess:Foxa2/MA0047.2/Jaspar(0.974):7.957761
-3002.664598 0 T:2871.0(64.44%),B:254123.1(13.86%),P:1e-1304
0.048 0.004 0.001 0.947
0.237 0.013 0.723 0.027
```

One or Many Partner motif(s) will be tested ?

One partner Many partners

Advanced options ?

Database of Partner motifs ?

Select the database with the proven recognition models of known transcription factors

Hocomoco v.11 Mouse core collection

RUN CLEAN

EXAMPLE

Figure 1. Basic Web-MCOT pages: (A) homepage, (B) application, and (C) example.

2.2. Output Data

Figure 2 shows the Web-MCOT basic output table for the example FoxA2 dataset. The table rows respect second motifs in CEs, since all CEs have the same first Anchor motif. Hence, the first column 'Motif name' lists the names of the second motif. The name 'Anchor' would mean homotypic CEs, otherwise the names of Partner motifs respect heterotypic CEs. Next, the common header 'CE significance, $-\text{Log}_{10}(P\text{-value})$ ' joins five columns showing significances of CEs without respect to motif conservation respecting five computation flows. The next column, 'Similarity, $-\text{Log}_{10}(P\text{-value})$ ', marks the significances of similarity between Anchor and Partner motifs. The column 'CE histogram' displays icons for distributions of the abundance of CEs as a function of mutual orientation and location of motifs. These icons are links to larger-sized histograms. Figure 3 shows this histogram for the example CE FoxA2/HNF1B. For each Partner motif, among all CEs with overlaps of motifs, we find the most common one and draw a CE logo; see icons in the column 'CE logo' (Figure 2). They are links for CE logos. Figure 4A shows the logo for the CE example FoxA2/HNF1B. This logo explains how two motifs with a specific orientation and overlap are located in almost the same place. Finally, we provide links to heatmaps that show the abundance of CEs with various ratios of conservation between Anchor and Partner motifs. Five columns with the common header 'Asymmetry heatmap, per mille' contain links to these heatmaps for five computation flows. The example heatmap in Figure 4B shows that in overlapped locations of FoxA2 and HNF1B motifs, the second one possesses higher conservation than the first. The output table allows sorting the rows in ascending or descending order, all columns containing text or numerical content. We marked this option with the up/down tooltip arrows appearing next to the column headings. Thus, the example (Figure 2) represents the sorting for the column 'Asymmetry, $-\text{Log}_{10}(P\text{-value})$ ', 'Overlap'.

Four links above the output table (Figure 2) allow download input and all output data. Thus, the link 'Download $P\text{-value}$ table' (Supplementary Data S1) lists significances for the example CEs FoxA2/HNF1B; the link 'Download histogram data' (Supplementary Data S2) contains the source data for histograms of the abundance for CEs between the Anchor FoxA2 and all Partner motifs, so that for each histogram, various orientations, overlaps and spacers are compared. The link 'Download additional data' contains the rest output data, e.g., the list predicted CEs FoxA2/HNF1B for all peaks of the example FoxA2 dataset (Supplementary Data S3). The manual page (the link Help, Figure 1A) describes all output data in detail.

Download Input files
Download P-value table
Download Histogram data
Download additional data

Motif Name	CE significance, $-\text{Log}_{10}(P\text{-value})$					Similarity $-\text{Log}_{10}(P\text{-value})$	Asymmetry, $-\text{Log}_{10}(P\text{-value})$					CE histogram	CE logo	Asymmetry heatmap, per mille				
	Full	Partial	Overlap	Spacer	Any		Full	Partial	Overlap ↑ 1	Spacer	Any			Full	Partial	Overlap	Spacer	Any
HXA9_MOUSE.H11M0.0.B	0	196.96	196.33	0	70.63	0	0	-43.59	-42.33	-51.11	-227.91							
HNF1B_MOUSE.H11M0.0.A	112.92	115.55	135.75	0	47.64	0	0	-18.9	-25.91	-49.6	-176.07							
AHR_MOUSE.H11M0.0.B	0	0	0	0	0	0	0	0	0	0	0							
AIRE_MOUSE.H11M0.0.C	30.41	26.44	46.36	0	0	0	0	0	0	0	0							
ALX1_MOUSE.H11M0.0.B	0	22.37	22.37	0	0	0	0	0	0	0	0							

Figure 2. Basic output table of Web-MCOT for the example FoxA2 ChIP-seq dataset [18]. The first column lists the names of the second motifs in CEs. The common header ‘CE significance, $-\text{Log}_{10}(P\text{-value})$ ’ joins columns showing the significances of CEs without respect to motif conservation for five computation flows. The column ‘Similarity, $-\text{Log}_{10}(P\text{-value})$ ’ shows the significance of similarity of motifs. The common header ‘Asymmetry, $-\text{Log}_{10}(P\text{-value})$ ’ joins columns with the significances of asymmetry within CEs for five computation flows. The columns ‘CE histogram’ and ‘CE logo’ contain icons for histograms of CE abundance as a function of mutual orientation and location of the motifs, and for logos for the most common CE structural type with an overlap of motifs. The last five columns with the common header ‘Asymmetry heatmap, per mille’ respect five flows and show the links to heamaps of CEs abundance as a function of the ratios of conservation between Anchor and Partner motifs.

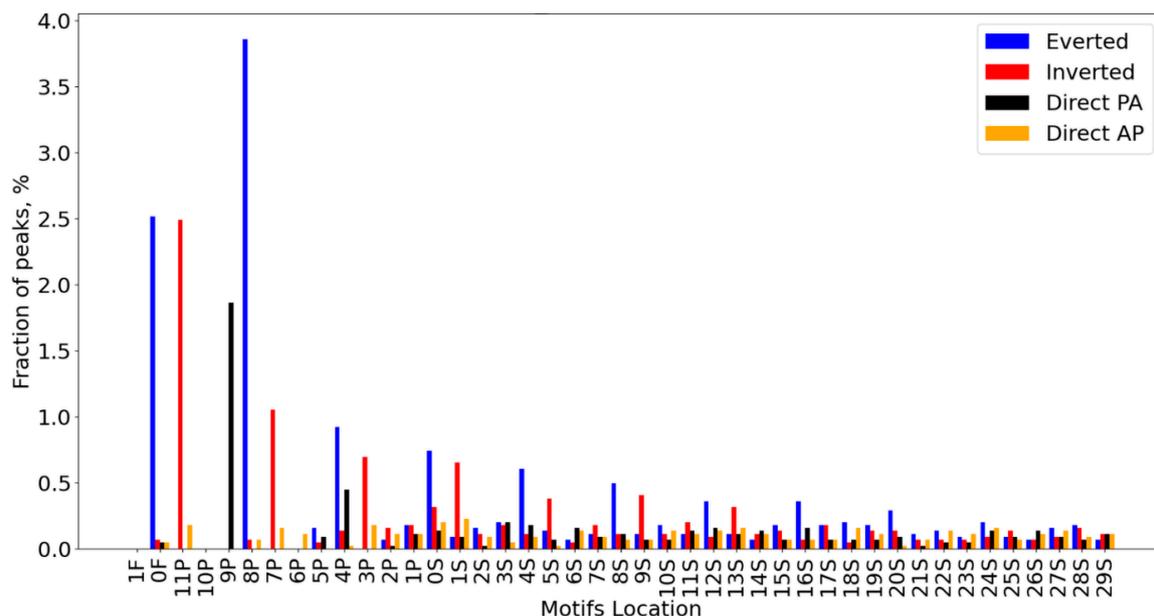


Figure 3. Distribution of structural variants of predicted CEs with various orientations, overlaps and spacers of the Anchor FoxA2 and Partner HNF1B motifs for the FoxA2 dataset [18]. Colors denote different mutual orientations. The letters in labels near axis X from left to right mean full ('F') or partial overlaps ('P'), and spacer length ('S'). The numbers preceding these letters denote the distance between nearest borders of two motifs, the length of overlap and the length of spacer, respectively. Axis Y shows the percentage of peaks containing CEs variants specific in mutual orientation and location. FoxA2 and HNF1B motifs were derived with the Homer tool [6] and the mouse Hocomoco library [14], respectively.

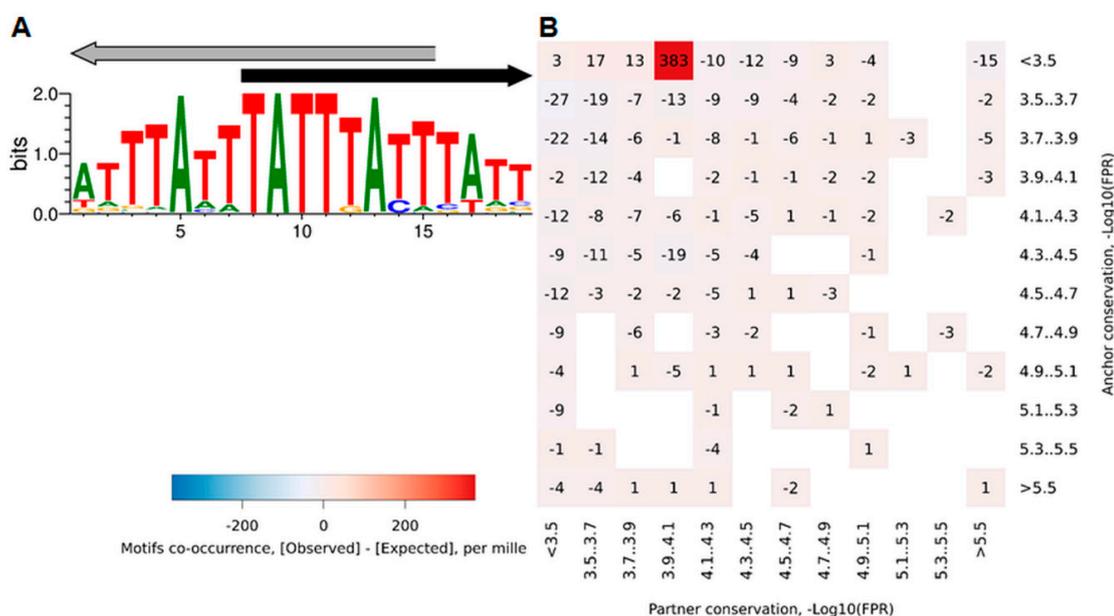


Figure 4. Graphical description of the most common CEs with an overlap of the Anchor FoxA2 and Partner HNF1B motifs for the FoxA2 dataset [18]. **(A)** Logo of the most common CE structural type. Black/grey arrows show the location and orientation of Anchor/Partner motifs. **(B)** Heatmap visualization of relationship of motifs conservation in CEs. Axes X/Y show ranges of conservation level of Partner/Anchor motifs. The color implies the per-mille measure for difference between observed and expected abundances of CEs with specific conservation of motifs (see Section 4.3).

2.3. Architecture

The web server contains the frontend, work processes, and backend parts. The frontend part builds Single Page Application using JavaScript language, HTML, CSS and Vue.js framework. The work-processes part performs kernel and additional calculations. The C++ kernel maps CEs in peaks and computes their significances. Kernel source code is available at the stand-alone MCOT site [19]. Additionally, R script generates heatmaps of CEs abundance with various conservation of motifs; Python code draws histograms of structural variants of predicted CEs with various orientations, overlaps and spacers; and we used the standard WebLogo library [20] to develop the library [21] producing the logo for CEs. The backend part is developed in Python language with Flask framework to build REST API web-service, Celery and Redis databases organize the task query and uWSGI application server.

3. Discussion

Our previous study [10] confirmed that MCOT is a methodologically novel approach for CE prediction, which outperforms in the recognition performance other available tools (see Table 1). Here we constructed the Web-MCOT server detecting significantly enriched co-occurred motifs taking into account their overlapping and spacing in a single ChIP-seq dataset. The server extends advantages of previously developed tools [8,11,12] for motif co-occurrence prediction in ChIP-seq data. Web-MCOT is similar to the popular web tool SpaMo [8], since for given Anchor motif it requires only a single ChIP-seq dataset, tests the co-occurrence of spaced motifs, and reveals Partner motifs with top-ranked significances of co-occurrence by processing a library of Partner motifs. The advantage of Web-MCOT consists in analysis of the co-occurrence of motifs with overlaps, and in analysis of co-occurrence of motifs with various conservations in pairs. The significant co-occurrence of motifs with an overlap may suggest synergistic or antagonistic mechanisms of their cooperation. The systematic difference in the conservation between two co-occurring motifs proposes that the TF with more conserved motif potentiates TF-DNA interaction of another TF of a pair [13]. Various visualization opportunities of Web-MCOT including histograms depicting abundance of CEs with various mutual orientations, overlaps and spacers of motifs, and the logos for the most common CE structural types with an overlap of motifs may further clarify the interaction between two respective TFs and genomic DNA.

4. Materials and Methods

4.1. Algorithm

Web-MCOT uses the previously described algorithm [10,13] to compute for a single ChIP-seq dataset and pairs of motifs the significances of co-occurrence for their locations with overlaps and with spacers. Web-MCOT takes the following input data: DNA sequences of ChIP-seq peaks, an Anchor motif, and a Partner motif or a name of Partner motifs library (Figure 5A). The homotypic and heterotypic pairs of motifs are considered, i.e., containing the same (Anchor/Anchor) or distinct (Anchor/Partner) motifs. Web-MCOT applies the recognition model of Position Weight Matrix for mapping motifs in sequences. For each heterotypic CE, Web-MCOT computes the significance of similarity between Anchor and Partner motifs, $P\text{-value}(A,P)$ [10]. The criterion $P\text{-value}(A,P) < 0.05$ means that a CE may be false-positive prediction. For each motif, Web-MCOT computes five recognition profiles respecting five ranges of conservation levels (CL) deduced from estimates of the false-positive rate (FPR), $CL = -\text{Log}_{10}(\text{FPR})$. FPRs are estimated as recognition rates for the whole-genome dataset of promoters. For each pair of motifs and each foreground sequence, Web-MCOT performs a permutation procedure to generate the background sequences. Next, the tool counts CEs in the foreground and background datasets and classifies CEs according to the location, orientation, similarity and conservation of participant motifs.

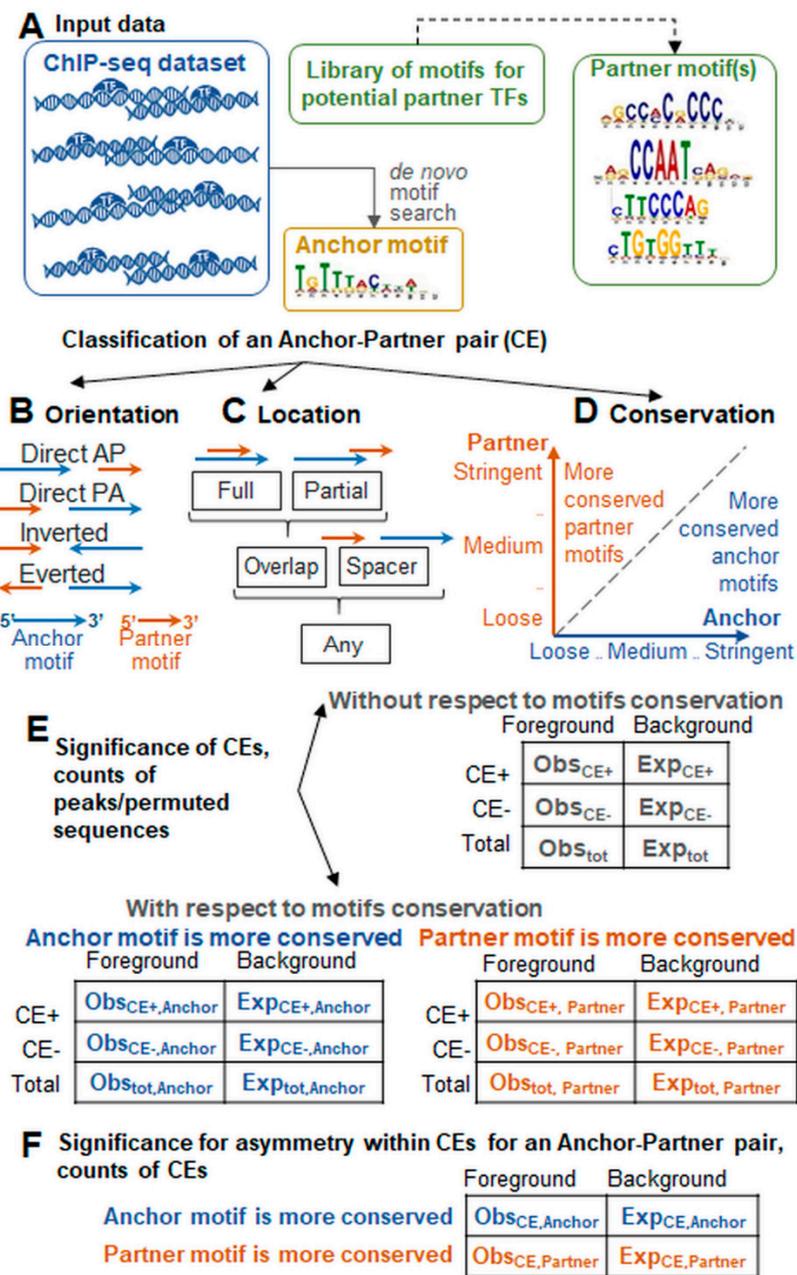


Figure 5. Web-MCOT workflow details. (A) Input data comprise a dataset of ChIP-seq peaks; a nucleotide frequency matrix for an Anchor motif; and a nucleotide frequency matrix for a Partner motif or designation of a public library of partner motifs. (B–D) show CEs classification according to orientations, locations (overlaps and spacers), and ratios of motifs conservation. (E) 2 × 2 contingency tables for computation of significances of enrichment for three CE types: without taking into account relationships of motifs conservation, and with an Anchor/Partner motif possessing higher conservation than Partner/Anchor motif (notations CE+/CE– respect sequences with/without CEs). (F) 2 × 2 contingency table for computation of significances of asymmetry within CEs.

4.2. Mutual Orientations and Mutual Locations

Four types of mutual orientation of motifs include Direct AP (Direct PA) types respecting shifted location of Anchor/Partner (Partner/Anchor) motifs in the same DNA strand; the opposite DNA strand allows Everted and Inverted types (Figure 5B).

Five types of mutual location of motifs comprise ‘Full overlap’ denoting one motif entirely covering another, ‘Partial overlap’ meaning all other overlaps, ‘Overlap’ respecting

any overlap, ‘Spacer’ representing spacing of motifs, and ‘Any’ designating either Overlap or Spacer (Figure 5C). Hence, we use five separate computation flows: Full, Partial, Overlap, Spacer, and Any. For all flows, we count sequences with/without CEs in foreground and background datasets and compute the significance *P-value* (CE) with Fisher’s exact test (Figure 5E).

4.3. Conservation of Motifs

Estimates of CL divide all CEs into two classes of those with more conserved Anchor or Partner motifs (Figure 5D). Besides two separate significances {*P-value* (CE)} for these classes, we compute the significance of asymmetry within CEs (Figure 5E,F). We assigned to this significance $-\text{Log}_{10}(\textit{P-value})$ signs ‘+’/‘−’ in the cases of the enrichment toward an Anchor/Partner motif. To draw the asymmetry heatmap for an Anchor/Partner pair, we consider CLs in the ranges $\text{CL} < 3.5$, $3.5 \leq \text{CL} < 3.7$, etc., up to $5.3 \leq \text{CL} < 5.5$ and $\text{CL} \geq 5.5$, and we count CEs with specific CLs for foreground and background datasets. The heatmap shows the per-mille measure, $\{1000 \times \text{Obs}_{i,j}/\text{Obs}\} - \{1000 \times \text{Exp}_{i,j}/\text{Exp}\}$. Here, $\text{Obs}_{i,j}$ and $\text{Exp}_{i,j}$ mean counts of CEs for specific CLs for the foreground and background datasets, respectively; Obs and Exp denote corresponding total counts; indices *i* and *j* refer to Anchor and Partner motifs.

5. Conclusions

In this study, we propose the web server Web-MCOT, which predicts pairs of co-occurring spaced or overlapped motifs in a single ChIP-seq dataset. Web-MCOT requires DNA sequences of ChIP-seq peaks, a motif of the target TF (Anchor motif) and a motif of potential Partner TF (or a name of a public library of potential Partner motifs). Web-MCOT checks the significance of overrepresentation for pairs of overlapped and spaced motifs and visualizes the most significant pairs of motifs, taking into account various orientations and conservations of participants. Web-MCOT results may uncover various mechanisms of interaction between a target TF and genomic DNA, facilitating the planning of future ChIP-seq experiments.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms23168981/s1>.

Author Contributions: Conceptualization V.G.L., E.V.Z. and D.Yu.O.; methodology V.G.L., D.Yu.O. and S.A.L.; software V.G.L. and A.M.M.; investigation V.G.L.; writing—original draft preparation V.G.L.; writing—review and editing, V.G.L. and E.V.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The algorithm development was supported by Russian Science Foundation project 21-14-00240. The web server creation and development were performed using computational resources of the “Bioinformatics” Joint Computational Center supported by State Budget Projects FWNR-2022-0006 and FWNR-2022-0020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Web-MCOT is available online at <https://webmcot.sysbio.cytogen.ru/> (accessed on 18 July 2022).

Acknowledgments: The work was performed on the equipment of the Bioinformatics Shared Access Center with the support of State Budget Projects FWNR-2022-0006 and FWNR-2022-0020.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Morgunova, E.; Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* **2017**, *47*, 1–8. [[CrossRef](#)] [[PubMed](#)]
2. Kel, O.V.; Romaschenko, A.G.; Kel, A.E.; Wingender, E.; Kolchanov, N.A. A compilation of composite regulatory elements affecting gene transcription in vertebrates. *Nucleic Acids Res.* **1995**, *23*, 4097–4103. [[CrossRef](#)] [[PubMed](#)]
3. Zheng, R.; Wan, C.; Mei, S.; Qin, Q.; Wu, Q.; Sun, H.; Chen, C.-H.; Brown, M.; Zhang, X.; Meyer, C.A. Cistrome Data Browser: Expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **2019**, *47*, D729–D735. [[CrossRef](#)] [[PubMed](#)]
4. Hammal, F.; de Langen, P.; Bergon, A.; Lopez, F.; Ballester, B. ReMap 2022: A database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res.* **2022**, *50*, D316–D325. [[CrossRef](#)] [[PubMed](#)]
5. Kolmykov, S.; Yevshin, I.; Kulyashov, M.; Sharipov, R.; Kondrakhin, Y.; Makeev, V.J.; Kulakovskiy, I.V.; Kel, A.; Kolpakov, F. GTRD: An integrated view of transcription regulation. *Nucleic Acids Res.* **2021**, *49*, D104–D111. [[CrossRef](#)] [[PubMed](#)]
6. Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y.C.; Laslo, P.; Cheng, J.X.; Murre, C.; Singh, H.; Glass, C.K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **2010**, *38*, 576–589. [[CrossRef](#)] [[PubMed](#)]
7. Bailey, T.L. STREME: Accurate and versatile sequence motif discovery. *Bioinformatics* **2021**, *37*, 2834–2840. [[CrossRef](#)] [[PubMed](#)]
8. Whittington, T.; Frith, M.C.; Johnson, J.; Bailey, T.L. Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res.* **2011**, *39*, 98. [[CrossRef](#)] [[PubMed](#)]
9. Jankowski, A.; Szczurek, E.; Jauch, R.; Tiuryn, J.; Prabhakar, S. Comprehensive prediction in 78 human cell lines reveals rigidity and compactness of transcription factor dimers. *Genome Res.* **2013**, *23*, 1307–1318. [[CrossRef](#)] [[PubMed](#)]
10. Levitsky, V.; Zemlyanskaya, E.; Oshchepkov, D.; Podkolodnaya, O.; Ignatieva, E.; Grosse, I.; Mironova, V.; Merkulova, T. A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. *Nucleic Acids Res.* **2019**, *47*, e139. [[CrossRef](#)] [[PubMed](#)]
11. Guo, Y.; Mahony, S.; Gifford, D.K. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.* **2012**, *8*, e1002638. [[CrossRef](#)] [[PubMed](#)]
12. Jankowski, A.; Prabhakar, S.; Tiuryn, J. TACO: A general-purpose tool for predicting cell-type-specific transcription factor dimers. *BMC Genomic.* **2014**, *15*, 208. [[CrossRef](#)] [[PubMed](#)]
13. Levitsky, V.; Oshchepkov, D.; Zemlyanskaya, E.; Merkulova, T. Asymmetric conservation within pairs of co-occurred motifs mediates weak direct binding of transcription factors in ChIP-seq data. *Int. J. Mol. Sci.* **2020**, *21*, 6023. [[CrossRef](#)] [[PubMed](#)]
14. Kulakovskiy, I.V.; Vorontsov, I.E.; Yevshin, I.S.; Sharipov, R.N.; Fedorova, A.D.; Rumynskiy, E.I.; Medvedeva, Y.A.; Magana-Mora, A.; Bajic, V.B.; Papatsenko, D.A.; et al. HOCOMOCO: Expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.* **2018**, *46*, D252–D259. [[CrossRef](#)] [[PubMed](#)]
15. O'Malley, R.C.; Huang, S.C.; Song, L.; Lewsey, M.G.; Bartlett, A.; Nery, J.R.; Galli, M.; Gallavotti, A.; Ecker, J.R. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **2016**, *165*, 1280–1292. [[CrossRef](#)] [[PubMed](#)]
16. Weirauch, M.T.; Yang, A.; Albu, M.; Cote, A.G.; Montenegro-Montero, A.; Drewe, P.; Najafabadi, H.S.; Lambert, S.A.; Mann, I.; Cook, K.; et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **2014**, *158*, 1431–1443. [[CrossRef](#)] [[PubMed](#)]
17. Castro-Mondragon, J.A.; Riudavets-Puig, R.; Rauluseviciute, I.; Lemma, R.B.; Turchi, L.; Blanc-Mathieu, R.; Lucas, J.; Boddie, P.; Khan, A.; Manosalva Pérez, N.; et al. JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **2022**, *50*, D165–D173. [[CrossRef](#)] [[PubMed](#)]
18. Wederell, E.D.; Bilenky, M.; Cullum, R.; Thiessen, N.; Dagpinar, M.; Delaney, A.; Varhol, R.; Zhao, Y.; Zeng, T.; Bernier, B.; et al. Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Res.* **2008**, *36*, 4549–4564. [[CrossRef](#)] [[PubMed](#)]
19. MCOT, Stand-Alone Version. Available online: <https://github.com/AcaDemIQ/mcot-kernel> (accessed on 9 August 2022).
20. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)] [[PubMed](#)]
21. WebLogo Library. Available online: <https://github.com/academiq/weblogo> (accessed on 9 August 2022).