MOTIFSIM 2.1: 
An Enhanced Software Platform for Detecting Similarity in Multiple DNA Motif Data Sets 

NGOC TAM L. TRAN and CHUN-HSI HUANG 

ABSTRACT

Finding binding site motifs plays an important role in bioinformatics as it reveals the transcription factors that control the gene expression. The development for motif finders has flourished in the past years with many tools have been introduced to the research community. Although these tools possess exceptional features for detecting motifs, they report different results for an identical data set. Hence, using multiple tools is recommended because motifs reported by several tools are likely biologically significant. However, the results from multiple tools need to be compared for obtaining common significant motifs. MOTIFSIM web tool and command-line tool were developed for this purpose. In this work, we present several technical improvements as well as additional features to further support the motif analysis in our new release MOTIFSIM 2.1.

Keywords: binding site motifs, motif detection tool, motif similarity detection, merged motifs, phylogenetic tree.

1. INTRODUCTION

MOTIFS ARE OFTEN SHORT SEQUENCES of a similar pattern found in sequences of DNA or protein. Binding site motifs play an important role in revealing the transcription factor that controls the gene expression. Many motif finding tools have been developed in the past years such as MEME (Bailey et al., 2006), GLAM2 (Frith et al., 2008), CisFinder (Sharov and Ko, 2009), W-ChIPMotifs (Jin et al., 2009), CompleteMOTIFS (Kuttippurathu et al., 2011), DREME (Bailey, 2011), MEME-ChIP (Machanick and Bailey, 2011), RSAT peak-motifs (Thomas-Chollier et al., 2012), and PScanChIP (Zambelli et al., 2013) among many others. Each tool possesses its unique features for discovering motifs that are undetectable by others. Previous study showed that the results produced by different motif finders for the same data set are diverse (Tran and Huang, 2014). Therefore, using multiple tools for finding motifs is suggested because motifs reported by several different tools are more likely to be biologically significant (Tran and Huang,
However, the results from multiple tools for the same data set require comparing against each other for finding common motifs and those generated by some tools but not by others (Tran and Huang, 2015).

Previous study showed the difficulty of comparing multiple motif data sets, and hence it motivated to develop MOTIFSIM (MOTIF SIMilarity Detection Tool) for automatically detecting similarity in multiple DNA motif data sets (Tran and Huang, 2015). The initial releases of MOTIFSIM provided researchers with a command-line tool for comparing motifs locally and a user-friendly web tool for comparing motifs on-line. The web tool provides convenience for users to save the data sets and experimental results on-line for retrieval. MOTIFSIM web tool and command-line tool accept various input formats and produce multiple results for further analysis. The results include the global significant motifs, the global and local significant motifs, as well as best matches for each motif in every data set (Tran and Huang, 2015). The new version MOTIFSIM 2.1 further supports users with several technical improvements as well as additional features.

2. TECHNICAL IMPROVEMENTS

We present numerous technical improvements for the web tool and the command-line tool as follows.

- **Automatically recognize motif input formats.** The new version can automatically detect motif’s format. In addition, motifs in different formats can be mixed and matched in the same input file and the tools can automatically recognize their formats.
- **Insert motif input on the browser.** In addition to upload and use existing files, the new version allows inserting as many as 20 motif files on the browser for running the web tool.
- **Increase number of motif data sets for comparison.** The initial release of the web tool allowed comparing up to 10 motif data sets simultaneously. The new version allows comparing up to 20 motif data sets concurrently.
- **Option for number of top significant motifs, output file type, and output file format.** MOTIFSIM 2.1 provides more flexibility for users to select the input and output parameters. We added an option for number of top significant motifs. This is a cutoff for the number of top significant motifs to be generated in the results for the global significant motifs as well as the global and local significant motifs. This option also allows users to select as many as 50 top significant motifs. In addition, users can select the output file type and output file format for the results. The output file type option allows selecting the global significant motifs or selects everything otherwise. The output file format option allows selecting a desire output format.
- **HTML and PDF formats with sequence logos.** We added HTML and PDF format options for generating the results. The conversion of HTML to PDF is supported by Prince software package (Prince, 2002). We also added sequence logos for each motif and its reverse complement for these formats. The sequence logos are created by WebLogo software package (Crooks et al., 2004).
- **Combined motifs list.** We added a combined motifs list containing motifs from all data sets in the results. The motifs are in position-specific probability matrices (Li, 2002) and they are in the order of the data sets entered by the user.
- **Additional global significant motifs list.** We include an additional global significant motifs list in the results for further analysis. The list can be generated in HTML, PDF, Text, or in all three formats.
- **Consensus sequences and motif alignment in IUPAC format.** We include the consensus sequences for each motif and its reverse complement in the results. The motif alignment in IUPAC format is also added in the results for better observation.
- **Job submission history.** We added a job submission history to the web tool for users to view and access their submitted jobs. Private jobs can only be accessed by the job’s owner. Public jobs are accessible to everyone.
- **Job search.** Unregistered users can keep submitted jobs and the results private. The results can be retrieved through the Search Job ID page.
- **Email notification.** Registered users of the web tool now receive an email notification when a submitted job is completed and available for download and viewing.
- **Other improvement.** We added the Input and Results sections to the result page of the web tool. Users can view the combined motifs and the results when a job is completed without leaving the page.

Supplementary Figures S1–S4 in the Supplementary Materials demonstrate the improvements already described.
3. ADDITIONAL FEATURES

In addition to the technical improvements already described, MOTIFSIM 2.1 provides additional features for further analyzing similar motifs. The global significant motifs as well as every motif in the combined list can be compared with motifs in a database for obtaining similar motifs. In addition, it is often desired to combine similar motifs into new motifs to reduce the number of redundant motifs. MOTIFSIM 2.1 provides such option for combining similar motifs reported for the global significant motifs, the global and local significant motifs, as well as the best matches for each motif. Besides, users can observe the relationship between motifs through the phylogenetic trees. These features are described in the following section.

3.1. Matching motifs with motif database

To match the global significant motifs as well as every motif in the combined list with motifs in a database, we implemented a slightly modified version of our novel algorithm (Tran and Huang, 2015). Instead of comparing motifs with each other in the combined list as in the original algorithm, we compare the global significant motifs and every motif in the combined list with each motif in a database using the same technique as described in the original algorithm. Currently, MOTIFSIM 2.1 supports Jaspar version 2016 (Mathelier et al., 2016), Transfac free version (Matys et al., 2003), and UniPROBE (Newburger and Bulyk, 2009) databases.

**FIG. 1.** Pair-wise merging of two similar motifs. (A) Alignment of two similar motifs CCGCGGCC and SSSCGSSGCS by using similarity percentage (Tran and Huang, 2015). The merged motif is SSCGCSCGCCSS. Motifs are in IUPAC format. (B) Details for merging two motifs in (A). Motifs are in position-specific probability matrix (Li, 2002). Motif CCGCGGCC (left) aligns with motif SSSCGSSGCS by (middle). Merged motif SSCGCSCGCCSS is in the right. The rectangle box shows the overlapping portion between two motifs. The average of corresponding elements between two motifs in rectangle box is equivalent to bold element in the merge motif SSCGCSCGCCSS. The elements which are not in bold in the merged motifs are carried over from motif SSSCGSSGCS. They are in two rows on the top and in two rows at the bottom of the merged motif SSCGCSCGCCSS. (C) Motif logos for the alignment and merged motif in (A).
3.2 Merge similar motifs

To merge similar motifs reported in the results (Tran and Huang, 2015), we merge the motif and its best matches iteratively into new motifs in a pair-wise manner. First, the motif and its most similar motif in the best matches list are merged into the new motif from their best alignment calculated by a similarity score (Tran and Huang, 2015). To merge two motifs from their best alignment, we take the average of the overlapping portion between them and carry over the hanging portions from the left, right, or both sides from the alignment into the new motif. Figure 1 illustrates this process. To ensure the new motif is still within the similarity threshold with its parents, we compare the new motif back with each of its parent by using the similarity percentage (Tran and Huang, 2015). If one of the similarity percentages is out of the threshold (Tran and Huang, 2015), the process stops. Otherwise, the new motif is then merged with the next similar motif in the list. This process goes on until the list is exhausted and it results in the final merged motif SSGCGCGCGCGGCCSS. All merged motifs lie within the similarity percentage with their parents. Pair-wise matching details are also included.

3.3 Phylogenetic trees

MOTIFSIM 2.1 provides an option for generating the phylogenetic tree for observing the relationship between motifs. The phylogenetic tree is built by using \texttt{hclust} function in R (R Core Team, 2016). This function implements the hierarchical clustering algorithm. The distance matrix, which is used to feed into \texttt{hclust} for

![FIG. 2. Merging of a motif and its best matches](image)

Motifs are in IUPAC format. (A) Motif GTCGCG and its five best matches from highest to lowest. (B) Pair-wise merging of motif GTCGCG and its best matches. Merging starts with motif GTCGCG and its first best match CGGCYBCGCG. The merged motif GBCGCGCGCGC is subsequently merged with the second best match in the list. The process goes on until the list is exhausted and it results in the final merged motif SSGCGCGCGCGGCSS. All merged motifs lie within the similarity percentage with their parents. Pair-wise matching details are also included.

### Table 1. ChIP-Seq Data Sets

| ChIP-seq data set | Mark | Species/tissue  | GEO accession  |
|-------------------|------|----------------|----------------|
| DM721             | H3K27ac (H3 lysine 27 acetylation) | Mouse/liver    | GSM851275      |
| DM01              | H3K4me1 (histone H3 lysine 4 monomethylation) | Mouse/liver    | GSM722760      |

The data sets were generated from ChIP-Seq experiments on mouse liver tissue (Tran and Huang, 2014).
building the tree, contains the best similarity scores (Tran and Huang, 2015) between motifs. To generate the phylogenetic tree for all motifs in the combined list, MOTIFSIM 2.1 builds the distance matrix containing the best similarity scores between motifs and then feeds it into hclust for generating the tree. The phylogenetic tree for the global significant motifs and their best matches is generated by using a subset of this distance matrix, which contains only the best similarity scores between the global significant motifs and their best matches.

3.4. Using MOTIFSIM 2.1

MOTIFSIM 2.1 web tool and command-line tool were designed for simple use. Detailed examples for running both tools can be found in the Supplementary Materials. Further instructions can be found in the user manual on the tool’s website.

Table 2. Motif Data Sets Used in Case Studies

| Case study | Motif data set | Motif input format | Number of motifs | Motif finder | ChIP-seq data set |
|------------|----------------|--------------------|-----------------|--------------|------------------|
| 1          | CisFinder_DM721_Cluster | Position-specific scoring matrix | 153 | CisFinder | DM721 |
| 2          | DREME_DM01 | Output from MEME | 51 | DREME | DM01 |
| 3          | MEME-CHIP_DM01 | Output from MEME | 9 | MEME-CHIP | DM01 |
| 4          | PScanChIP_DM01 | Jaspar | 27 | PScanChIP | DM01 |
| 5          | RSAT_peak-motifs_DM01 | TRANSFAC-like | 40 | RSAT peak-motifs | DM01 |

The data sets came from experiments in Tran and Huang (2014).

Table 3. Top 10 Global and Local Significant Motifs in Case Study 1

| No. | Data set no. | Motif ID | Motif name | Motif logo |
|-----|--------------|----------|------------|------------|
| 1   | 1            | 108      | C108       | ![Motif Logo](image1) |
| 2   | 1            | 25       | C025       | ![Motif Logo](image2) |
| 3   | 1            | 70       | C070       | ![Motif Logo](image3) |
| 4   | 1            | 78       | C078       | ![Motif Logo](image4) |
| 5   | 1            | 23       | C023       | ![Motif Logo](image5) |
| 6   | 1            | 104      | C104       | ![Motif Logo](image6) |
| 7   | 1            | 18       | C018       | ![Motif Logo](image7) |
| 8   | 1            | 65       | C065       | ![Motif Logo](image8) |
| 9   | 1            | 84       | C084       | ![Motif Logo](image9) |
| 10  | 1            | 54       | C054       | ![Motif Logo](image10) |

Motifs are listed by ID, name, and logos.
3.5. Case studies

We present the application of MOTIFSIM 2.1 in two case studies in the following sections. The data sets used in the case studies were produced by several motif finders, including CisFinder, DREME, MEME-ChIP, PScanChIP, and RSAT peak motifs for the same peak data. The peak data sets were generated from ChIP-Seq data sets produced by ChIP-Seq experiments on mouse liver tissue for two marks, H3 lysine 27 acetylation (H3K27ac) and histone H3 lysine 4 monomethylation (H3K4me1) (Tran and Huang, 2014). The ChIP-Seq data sets are given in Table 1 and the motif data sets are in Table 2. Since different motif finders implement different algorithms and possess unique features for detecting motifs, the results reported by them vary for the same ChIP-Seq data set. In particular, four motif finders given in Table 2 report different number of motifs for the same ChIP-Seq data set DM01. These numbers differ significantly from one tool to another. Thus, it is useful to identify common motifs reported by these tools because these motifs are more significant. MOTIFSIM can identify such motifs and reported them as the global significant motifs. It can also identify the global and local significant motifs as well as best matches for every motif in each data set.

Case study 1 presents the use of MOTIFSIM 2.1 for identifying similar motifs in a single data set, whereas Case study 2 identifies similar motifs in multiple data sets. In both case studies, we used the same input parameters for comparing motifs. These parameters include top 10 significant motifs, 5 best matches for each motif, and 75% or greater for similarity cutoff. The option All was used for selecting both output file type and output file format. In addition, the results were further compared with motifs in the UniPROBE

| ID | Motif name | Logo | Logo | Motif name |
|----|------------|------|------|------------|
| 108 | C108       | ![Logo](image1) | ![Logo](image2) | C125 |
|    | C021       | ![Logo](image3) | ![Logo](image4) |     |
|    | C053       | ![Logo](image5) | ![Logo](image6) |     |
|    | C017       | ![Logo](image7) | ![Logo](image8) |     |
|    | C070       | ![Logo](image9) | ![Logo](image10) |     |
| 23  | C023       | ![Logo](image11) | ![Logo](image12) | C004 |
|    | C104       | ![Logo](image13) | ![Logo](image14) |     |
|    | C054       | ![Logo](image15) | ![Logo](image16) |     |
|    | C003       | ![Logo](image17) | ![Logo](image18) |     |
|    | C008       | ![Logo](image19) | ![Logo](image20) |     |

The best matched motifs are listed in the order of similarity from highest to lowest.
database for mouse. We also generated the phylogenetic trees to observe the relationship between motifs as well as combined similar motifs reported in the results.

3.5.1. Case study 1: ChIP-Seq data set DM721 for H3K27ac (H3 lysine 27 acetylation). In this case study, we identified similar motifs in a single data set produced by CisFinder given in Table 2. This tool reported 153 cluster motifs. We ran MOTIFSIM 2.1 on this data set using the input parameters

| Motif | Alignment | Combined motif | Format | Direction | Position no. | Overlap |
|-------|-----------|----------------|--------|-----------|--------------|--------|
| C108  | --RBACWGASAWASVY | ASACAGASA WASMBSK | Original motif | Backward | 1 | 14 |
| C125  | SKASWYASAGRWSMS | WASMBSK | Original motif | Backward | 1 | 14 |

Combined motif

| C021  | MKAGMVAGAKGAMK-- | ASACAGASA TAMSKS | Original motif | Forward | 1 | 14 |

| C053  | RMACRGAGAAAYCCTGKY | AABCBK | Original motif | Forward | 1 | 16 |

Combined motif

| C017  | MRYVBGTCTCTCTCTGTHK-- | MRCSKSTHT CTCTSTTVT | Reverse complement | Forward | 3 | 16 |

Merging begins with motif C108 and its first best match C125. The combined motif is subsequently merged with the second best match C021. The process stops with the final merged motif VDSTSTSTBTCTBSTVT. All merged motifs fall within similarity threshold with their parents. The pair-wise alignment of motifs, matching format of each motif, matching direction, matching position, and the number of overlaps are included.

database for mouse. We also generated the phylogenetic trees to observe the relationship between motifs as well as combined similar motifs reported in the results.

3.5.1. Case study 1: ChIP-Seq data set DM721 for H3K27ac (H3 lysine 27 acetylation). In this case study, we identified similar motifs in a single data set produced by CisFinder given in Table 2. This tool reported 153 cluster motifs. We ran MOTIFSIM 2.1 on this data set using the input parameters

| Motif | Alignment | Combined motif | Format | Direction | Position no. | Overlap |
|-------|-----------|----------------|--------|-----------|--------------|--------|
| C023  | --------------SSGGSDG | SSSGGSSGSSGGS | Reverse complement | Backward | 2 | 14 |
| C004  | SSSGGSSGSSGSS | SSSGGSSGSSGSS | Reverse complement | Forward | 6 | 14 |

Combined motif

| C104  | SSSGGSSGGG | SSSGGSSGSSGSS | Reverse complement | Forward | 6 | 14 |

| C054  | SSSGGSSGGG | SSSGGSSGSSGSS | Reverse complement | Forward | 6 | 14 |

Combined motif

| C008  | GSCBCVSSCCVSSCCBCC CCSCCCSCSCSCSCSCSCSC | GSCBCVSSCCSC | Reverse complement | Forward | 7 | 16 |

Merging starts with motif C023 and ends with the merged motif GSCBCVSSCCSCSCSCSCSCSCSCSCSCSCSCSC. All merged motifs fall within similarity threshold with their parents. Pair-wise matching information is included.
already described. Table 3 shows the top 10 significant motifs reported by the tool. The five best matches for the first and fifth significant motifs are given in Table 4. These best matches are not only similar to their top significant motif but they are also similar to each other. In particular, motif C125 and motif C070 share the same motif Sox7_secondary in UniPROBE database for mouse as the first best match for motif C125 and the second best match for motif C070. Likewise, motif C021 and motif C070 share the same motif Sox12_secondary in UniPROBE database for mouse as the first best match for motif C012 and as the second best match for motif C070. In addition, motif C053 and motif C071 share an identical motif Gli1_v016060_primary in UniPROBE database for mouse as the first best match for motif C053 and the third best match for motif C071. Thus, by analyzing these similar motifs, it is useful for determining whether they are redundant motifs. MOTIFSIM 2.1 also provides the option for combining similar motifs. Tables 5 and 6 show the merging for motif C108 and its five best matches as well as for motif C023 and its best matches, respectively. The detailed merging results can be found in the user manual on the tool’s website. We further matched each motif in the data set with motifs in UniPROBE database for mouse. Table 7 shows the first best match in the database for each top 10 significant motifs. The detailed matching results for each motif with the database can be found in the user manual on the tool’s website. To observe the relationship between motifs, we generated a phylogenetic tree shown in Figure 3 for all motifs in the data set. In this figure, the most similar pair of motifs by similarity score is placed in one cluster. The cluster is joined with the next similar motif. Similar clusters are joined until they form a complete phylogenetic tree. The motif is labeled by concatenating its ID with its name for easy differentiation, as the same motif may appear multiple times in the combined list because it is reported by multiple motif finders.

3.5.2. Case study 2: ChIP-Seq data set DM01 for H3K4me1 (histone H3 lysine 4 monomethylation). This case study demonstrates the use of MOTIFSIM 2.1 for identifying similar motifs in multiple data sets generated by four different tools including DREME, MEME-ChIP, PScanChIP, and

| Motif ID | Motif name | Motif logo | Motif format | UniPROBE database matching |
|----------|------------|------------|--------------|---------------------------|
| 108      | C108       |            | Reverse complement | UP00101 Sox12_secondary |
| 25       | C025       |            | Reverse complement | UP00080 Gata5_secondary |
| 70       | C070       |            | Reverse complement | UP00011 Irf6_secondary |
| 78       | C078       |            | Reverse complement | UP00011 Irf6_secondary |
| 23       | C023       |            | Original motif | UP00539 Gli1_v016060_secondary |
| 104      | C104       |            | Original motif | UP00539 Gli1_v016060_secondary |
| 18       | C018       |            | Original motif | UP00028 Tcfap2e_secondary |
| 65       | C065       |            | Original motif | UP00080 Gata5_secondary |
| 84       | C084       |            | Reverse complement | UP00099 Asc12_primary |
| 54       | C054       |            | Original motif | UP00043 Bcl6b_secondary |

The first best match in the database for each significant motif is included. Motif ID, motif name, motif logo, and motif format are included.
RSAT peak-motifs for the same ChIP-Seq data set DM01 given in Table 2. These motif finders report different number of motifs. Thus, it is useful to identify common motifs reported by them. MOTIFSIM 2.1 identifies these common motifs as the global significant motifs. It also identifies the global and local significant motifs, as well as best matches for each motif in the combined motif list. The top 10 global significant motifs are given in the Supplementary Table S1. Each global significant motif and its best matches were reported by at least two motif finders. The top 10 global and local significant motifs are also given in the Supplementary Table S2. In this table, the ninth global and local significant motif, *Motif 25*, and its five best matches were reported by all four tools. However, the fifth global and local significant motif, $ssCkGGYCCCs$, and its five best matches were reported by only one tool, which is RSAT peak-motifs. The motif $ssCkGGYCCCs$ and its best matches are given in the Supplementary Table S3. This observation allows users to determine whether these similar motifs are redundant motifs. The analysis can be carried out further for any motif and its best matches.

Similar motifs reported in the results for the global significant motif, the global and local significant motif, as well as for each motif in this case study were combined into new motifs. The detailed merging

**FIG. 3.** A phylogenetic tree for all cluster motifs in the data set. The tree was created by using a distance matrix consisting of best similarity scores between motifs (Tran and Huang, 2015). Motif ID is concatenated with motif name at the label of the tree.
results can be found in the user manual on the tool’s website. In addition, we further compared the global significant motifs, the global and local significant motifs, and each motif in the combined motif list with motifs in the UniPROBE database for mouse to obtain similar motifs. The Supplementary Tables S4 and S5 show the first best match in the database for each global significant motif as well as for each global and local significant motif, respectively. The detailed matching results with the database can be observed in the user manual on the tool’s website. In addition, the relationship between motifs for the global significant motifs and their best matches, as well as for all motifs in the combined list can be further observed through the phylogenetic trees in the Supplementary Figures S11 and S12.

4. CONCLUSION

MOTIFSIM 2.1 web tool and command-line tool contain several technical improvements as well as additional features to further support the motif analysis. The new version allows combining similar motifs. It also supports the comparisons for the global significant motifs as well as every motif with motifs in a database. In addition, the relationship between motifs can be observed through the phylogenetic trees. MOTIFSIM 2.1 web tool and command-line tool including user manuals, test data sets, and test results are freely available at http://motifsim.org

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Education Graduate Assistance in Areas of National Need (Grant P200A130153 to N.T.L.T.). The web tool’s infrastructure was supported by an AWS in Education Research Grant Award to N.T.L.T.

AUTHORS’ CONTRIBUTIONS

N.T.L.T. carried out the technical improvements, designed and implemented the additional features, and prepared the article. C.-H.H. directed the research.

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

Bailey, T. 2011. DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 27, 1653–1659.
Bailey, T., Williams, N., Misleh, C., et al. 2006. MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373.
Crooks, G.E., Hon, G., Chandonia, J.M., et al. 2004. WebLogo: A sequence logo generator. *Genome Res.* 14, 1188–1190.
Frith, M., Saunders, N., Kobe, B., et al. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol.* 4, e1000071.
Jin, V.X., Apostolos, J., Nagisetty, N.S., et al. 2009. W-ChIPMotifs: A web application tool for *de novo* motif discovery from ChIP-based high-throughput data. *Bioinformatics* 25, 3191–3193.
Kuttipurathu, L., Hsing, M., and Liu, Y. 2011. CompleteMOTIFs: DNA motif discovery platform for transcription factor binding experiments. *Bioinformatics* 27, 715–717.
Li, H. 2002. Computational approaches to identifying transcription factor binding sites in yeast genome. *Methods Enzymol.* 350, 484–495.
Machanick, P., and Bailey, T. 2011. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics.* 27, 1696–1697.
Mathelier, A., Fornes, O., Arenillas, D.J., et al. 2016. JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110–D115.
Matys, V., Fricke, E., Geffers, R., et al. 2003. TRANSFAC®: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31, 374–378.

Newburger, N., and Bulyk, M. 2009. UniPROBE: An online database of protein binding microarray data on protein–DNA interactions. *Nucleic Acids Res.* 37, D77–D82.

Prince. 2002. Available at: www.princexml.com Accessed March 19, 2016.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/ Last viewed on March 19, 2016.

Sharov, A., and Ko, M. 2009. Exhaustive search for over-represented DNA sequence motifs with CisFinder. *DNA Res.* 16, 261–273.

Thomas-Chollier, M., et al. 2012. RSAT peak-motifs: Motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* 40, e31.

Tran, N.T.L., and Huang, C.-H. 2014. A survey of motif finding Web tools for detecting binding site motifs in ChIP-Seq data. *Biol. Direct.* 9, 1–22.

Tran, N.T.L., and Huang, C.-H. 2015. MOTIFSIM: A web tool for detecting similarity in multiple DNA motif datasets. *BioTechniques.* 59, 26–33.

Zambelli, F., Pesole, G., and Pavesi, G. 2013. PscanChIP: Finding over-represented transcription factor-binding site motifs and their correlations in sequences from ChIP-Seq experiments. *Nucleic Acids Res.* 41, W535–W543.

Address correspondence to:
Ngoc Tam L. Tran

*Department of Computer Science and Engineering*
*University of Connecticut*
*Fairfield Way, U-4155*
*Storrs, CT 06269*

*E-mail: ngoc.tran@uconn.edu*