

CLINICAL REPORT

A preliminary computational outputs versus experimental results: Application of sTRAP, a biophysical tool for the analysis of SNPs of transcription factor-binding sites

Shirin Moradifard¹ | Reza Saghiri¹ | Parastoo Ehsani² | Fatemeh Mirkhani¹ |
Mina Ebrahimi-Rad¹ 

¹Biochemistry Department, Pasteur Institute of Iran, Tehran, Iran

²Molecular Biology Department, Pasteur Institute of Iran, Tehran, Iran

Correspondence

Mina Ebrahimi-Rad, Biochemistry Department, Pasteur Institute of Iran, Tehran, Iran.

Email: minebrad@pasteur.ac.ir

Abstract

Background: In the human genome, the transcription factors (TFs) and transcription factor-binding sites (TFBSs) network has a great regulatory function in the biological pathways. Such crosstalk might be affected by the single-nucleotide polymorphisms (SNPs), which could create or disrupt a TFBS, leading to either a disease or a phenotypic defect. Many computational resources have been introduced to predict the TFs binding variations due to SNPs inside TFBSs, sTRAP being one of them.

Methods: A literature review was performed and the experimental data for 18 TFBSs located in 12 genes was provided. The sequences of TFBS motifs were extracted using two different strategies; in the size similar with synthetic target sites used in the experimental techniques, and with 60 bp upstream and downstream of the SNPs. The sTRAP (http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi) was applied to compute the binding affinity scores of their cognate TFs in the context of reference and mutant sequences of TFBSs. The alternative bioinformatics model used in this study was regulatory analysis of variation in enhancers (RAVEN; <http://www.cisreg.ca/cgi-bin/RAVEN/a>). The bioinformatics outputs of our study were compared with experimental data, electrophoretic mobility shift assay (EMSA).

Results: In 6 out of 18 TFBSs in the following genes *COL1A1*, *Hb $\zeta\psi$* , *TF*, *FIX*, *MBL2*, *NOS2A*, the outputs of sTRAP were inconsistent with the results of EMSA. Furthermore, no p value of the difference between the two scores of binding affinity under the wild and mutant conditions of TFBSs was presented. Nor, were any criteria for preference or selection of any of the measurements of different matrices used for the same analysis.

Conclusion: Our preliminary study indicated some paradoxical results between sTRAP and experimental data. However, to link the data of sTRAP to the biological functions, its optimization via experimental procedures with the integration of expanded data and applying several other bioinformatics tools might be required.

KEYWORDS

EMSA, SNP, sTRAP, transcription factor, transcription factor-binding site

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Molecular Genetics & Genomic Medicine* published by Wiley Periodicals, Inc.

1 | INTRODUCTION

In recent years, the increasing access to high-throughput data of sequencing have explained the pathology of several diseases by the analysis of variations in the noncoding regions of the genome, transcription factors-binding sites (TFBSs) being among them (ENCODE Project Consortium, 2012; MacArthur et al., 2017; Maurano, Wang, Wang, Kutuyavin, & Stamatoyannopoulos, 2012).

The specific binding of transcription factors (TFs) to their target-binding sites is a critical component of gene regulation at transcription and expression levels, a hallmark of several biological processes, including development, differentiation, and evolution, to name a few (Lai et al., 2019; Lambert et al., 2018; Savinkova et al., 2013). Several single-nucleotide polymorphisms (SNPs), affecting TFBSs might be potentially involved in either destruction or creation of the new TFBSs, resulting in a genetic disease or a phenotypic trait (Chorley et al., 2008; Kumar, Ambrosini, Ambrosini, & Bucher, 2016; Rana, Coshic, Coshic, Goswami, & Tyagi, 2017).

The binding difference of a TF for the reference and alternate alleles might be linked to the emergence of the diseases. So, access to a bioinformatics tool with a capacity of such prediction would be very valuable in creating the related hypothesis on the issue. However, it is a challenging task in the functional genomic analysis. Previous efforts have proposed several computational models and tools to compute the impacts of the SNPs on the binding affinity of the TFs; however, due to the shortness and degenerateness of TFBSs, some of approaches were found to be impractical (Boyle et al., 2012; Chowdhary et al., 2012; Mathelier & Wasserman, 2013; Riva, 2012).

Manke et al. introduced a new biophysical model dedicated to predict the impacts of SNPs of the target TFBSs on their related TFs binding affinities (Manke, Heinig, Heinig, & Vingron, 2010). The authors represented the sTRAP web tool with the potential capability to compare the wild and mutant motifs of TFBSs in interaction with their cognate TFs and to quantify the difference between binding activity scores of TF for the allelic sequences. The tool is sequence based and takes advantage of the application of position weight matrices (PWMs), frequently used model, to compute TF–TFBS-specific interaction (Zhao, Granas, & Stormo, 2009; Zhao, Ruan, Pandey, & Stormo, 2012) and the fixed-length TFBS models for such prediction.

The present study addresses the analysis of the binding affinity variations of putative TFs due to SNPs introduced in their TFBSs. The study's objective was to check for compliance between the data predicted by sTRAP and those of experimental approaches in the literature.

For any model (biophysical or bioinformatics) to become a predictive tool, some validation against wet-lab data are

required (Cooper et al., 2018). If the analysis is properly conducted with limiting measurement uncertainties, the model would be capable and functional in a true prediction. Otherwise, something might be missing in the model, which should be introduced in its structure. However, the experimental procedures are not the exceptions of this rule. They also need to be validated by other approaches (Cooper et al., 2018). The challenge of the biophysical model versus the experimental approach might bring two important kinds of outcomes; the high compatibility or a great contradiction between their outputs. The results of both circumstances would be of worth to be reported because the knowledge on the power or inabilities of the model might enable the user to design his project and get an accurate conclusion.

2 | METHODS

2.1 | Data collection

In this study, a literature review was done to provide the experimental data by collecting the eligible studies with the required information, relevant to our commitment. Those investigated the impact of the SNPs in TFBSs on the binding affinities of related TFs, by experimental approaches such as EMSA, were selected for preliminary analysis. Then among them, the articles focused on the nuclear extract or cell extract as the source of TFs for their analysis were excluded from our study. While, the data of the articles with the application of recombinant or synthetic TFs were included in our implementation. The articles of Mann et al. (2001) and Savinkova et al. (2013), among the eligible studies, obtained by literature review, were the only sources for extraction of the candidate TFBSs motifs for our analysis. They reported the functional analysis of Sp1 and TBP (TATA-Binding Protein/TATA Box) TFs binding affinity by using electrophoretic mobility shift assay (EMSA) technique, recombinant TFs, and synthetic DNA target sites, respectively. The latter group also found that the experimental results of TBP/TATA, highly correlated with those predicted by *in silico* prediction approach based on PWMs, they used in their study ($r = .822, \alpha < 10^{-7}$).

2.2 | TF affinity analysis

The bioinformatics analysis was performed using the wild and mutant DNA sequences of the selected TFBS motifs from the sources mentioned above. sTRAP (http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi; Thomas-Chollier et al., 2011), the computational and biophysical tool, was applied to evaluate the DNA motifs to assess the impact of SNPs in the TFBS on TF-binding affinity. The used input consisted

of the DNA sequences in FASTA format (Thomas-Chollier et al., 2011), the length of the motifs in the bioinformatics analysis was considered to be as long as the synthetic target sites used in the experimental techniques, to avoid any deviation in the predictions. In the other strategy, 60 bp upstream and downstream of the SNPs were included in the evaluation. In the motif analysis, the highest score among those obtained by using different matrices, was considered as the related binding energy.

2.3 | Analysis with RAVEN

Also, we utilized the regulatory analysis of variation in enhancers (RAVEN) (<http://www.cisreg.ca/cgi-bin/RAVEN/a>; Manke et al., 2010) as an alternative tool, due to its application together with sTRAP by Thomas-Chollier et al. (2011) in their study.

3 | RESULTS

3.1 | The outputs of bioinformatics versus experimental data

The 18 TFBSs with SNPs (located in 12 genes), experimentally analyzed by the other researchers, were included in our study and scored against the wild TFBSs, using bioinformatics tools; sTRAP and RAVEN. One out of 18 (in *COL1A1* gene), was the target for the Sp1 transcription factor although, the remained 11 cases were those being the cognate binding sites for TBP transcription factor, inside several genes (Table 1).

One of TFBSs (in *Hb β* gene), in turn, corresponded with seven different mutant forms, due to the different contents and diverse SNPs. In the four out of six mutant TFBSs in the following genes, the binding sites could not be detected using RAVEN bioinformatics tool for SNP analysis (*COL1A1*, *Hb $\zeta\psi$* , *TF*, *FIX*). The results of RAVEN for other SNPs analysis were consistent with those produced by the experimental procedure. However, the data of 2 TFBSs out of 14 (inside *MBL2*, *NOS2A* genes) were contradictory to those produced by sTRAP. Concerning the TFBSs of *TF* and *FIX* genes, there were inconsistent scores in sTRAP analysis, when two different matrices were applied for binding affinity prediction of TBP for each of the target sites.

The experimental data, and in silico prediction values (Savinkova et al., 2013) for two certain TFs in interaction with their target TFBSs, versus those obtained from sTRAP and RAVEN in our study, are represented in Table 1. The whole details about the analysis process are categorized in Table S1. The workflow of the study also summarized in Figure 1.

4 | DISCUSSION

The most of high-throughput genomic data, with predictions based on SNPs variation, are still prone to error due to several reasons, including the bias mediated by the sequence context. In such cases, there would be a need to confirm the predicted results using orthogonal technology. Otherwise, the false positive and negative results would be inevitable, making the estimation of variants and their linkage to a disease, impractical (Cooper et al., 2018; Kamali et al., 2015).

However, this is not restricted to SNP predictions, the other estimations such as epigenetics, fusion proteins, and so on, would also be the cases of error profiling. As an example, the gold standard verification technology, for the high-throughput data, next-generation sequencing (NGS), is Sanger sequencing, specifically in the cases with quality scores <Q500 (Cooper et al., 2018; Park et al., 2014; Strom et al., 2014). Nonetheless, several other studies have substituted the alternative techniques including; targeted next-generation sequencing, and mass spectrometric, for Sanger sequencing (Cooper et al., 2018; Sikkema-Raddatz et al., 2013).

There are several bioinformatics datasets in the literature to predict the target sequences for the microRNAs, or vice versa (Kumar, Wong, Wong, Tizard, Moore, & Lefèvre, 2012; Piriyaopongsa, Bootchai, Bootchai, Ngamphiw, & Tongshima, 2012; Agarwal, Bell, Nam, & Bartel, 2015); this is while some other databases have provided information on the binding microRNAs to the studied sequences, validated by experimental procedures. This might stand for the other example of bioinformatics data verification using experimental procedures (Huang et al., 2020; Karagkouni et al., 2018).

Although sTRAP, an online web tool constructed on the ChIP-seq database, is fast and easily applied, in some cases, its results were paradoxical to the experimental data. We have looked at some TFs binding affinities by introducing SNPs inside their target-binding sites, using sTRAP and RAVEN. Our results provided supporting evidence that at least in the case of Sp1 and TBP, sTRAP performance in six out of 18 SNPs was not consistent with those from documented experimental procedures (Mann et al., 2001; Savinkova et al., 2013). This might be indicative of sTRAP limitations in predicting the impacts of some of the SNPs in TFBSs and scoring the binding energies of their related TFs. The consequence of such restriction is to create some problems in quantifying the influence of the particular SNPs on human health and disease, estimating the functionalities of SNPs to waken or enhance the binding affinity of TFs, and testing related hypotheses based on SNPs variations in TFBSs.

Moreover, lack of any information on the probability (p value) between the binding energies of a TF for wild and mutant TFBSs, and also the absence of any cutoff of the significant differences makes the comparison between the two scores impractical. These features might imply the other

TABLE 1 Functional analysis of transcription factors (Sp1 and TBP) binding affinities to the target sites to score the SNPs impact, based on “EMSA,” in silico prediction analysis, “sTRAP,” and RAVEN approaches

TFBS/Gene	Matrix ID	TF	Target site W/M	Affinity W/M			EMSA p value/ α	RAVEN ^e score W/M
				sTRAP ^a W/M	EMSA	Prediction value -ln [K _D] ^d		
<i>COL1A1</i>	M00008	Sp1	agggaatGGGGGGGGGATGaggcct/agggaatTGTGGGGGGGATGaggcct	6.70/5.85	0.095/0.039 (μM) ^b	—	Mann et al. (2001)	TF not recognized for TFBS
	M00196							
	M00933							
	M00931							
	M00932							
<i>Hb ζ₂</i>	M00980	TBP	ctgccacaccaCATTATTgaaaaa/ctgccacaccCACATTATCagaaaa	4.14/2.67	15.70/16.00 ^c	17.72/18.28	Savinkova et al. (2013)	TF not recognized for TFBS
	M00471							
<i>MBL2</i>	M00980	TBP	catctatttcTAZATAgcctgcacc/catctatttcTACATAgcctgcacc	4.83/5.24, 2.59/0.98	17.39/16.66 ^c	19.68/18.57	Savinkova et al. (2013)	6.851(82.3%)/2.186(73.1%)
	M00471							
<i>TF</i>	M00471	TBP	gccggcccTTTATAgcgcgggggca/gccggcccTTTATAgTgcgcggggca	0.19/1.52	16.45/17.47 ^c	18.91/19.43	Savinkova et al. (2013)	TF not recognized for TFBS
	M00980							
<i>NOS2A</i>	M00980	TBP	atgggtgagTATAAAATAcctctgg/atgggtgagTATAAAATAcCctctgg	0.09/0.09	20.14/20.25 ^c	19.85/20.06	Savinkova et al. (2013)	9.534 (87.7%)/11.614 (91.8%)
	M00471							
<i>FIX</i>	M00980	TBP	acagctcagcTTGTACTTTggtaaca/acagctcagcTTCTACTTTggtaaca	2.63/3.04, 0.33/0.33	14.49/14.51 ^c	18.24/17.75	Savinkova et al. (2013)	TF not recognized for TFBS
	M00471							

Abbreviations: M, Mutant; SNPs, single-nucleotide polymorphisms; W, Wild Type.

^asTRAP model has applied different matrices to score the binding affinity of transcription factors (TFs), each of which applies diverse frames of transcription-binding sites (TFBSs) for its computation. The score of the matrix with the highest energy in scanning TFBS consisting of target SNP has been presented. Excluding *MBL2*, *TF*, and *FIX*, for the others, different matrices gave the contradictory results.

^bSp1 results: The concentrations of radiolabeled competitor at 50% inhibition for “S” and “s” alleles, respectively; the lower concentration showed a higher binding affinity.

^cTBP results: Equilibrium Dissociation Constant (−ln [K_D]) which characterizes the binding affinity of TFs for TFBSs; the higher values showed the more affinities.

^d−ln [K_D] prediction value for TBP/TATA.

^eRegulatory Analysis of Variation in Enhancers (RAVEN) is a Web-based application utilized for detection and characterization of regulatory sequence variation.

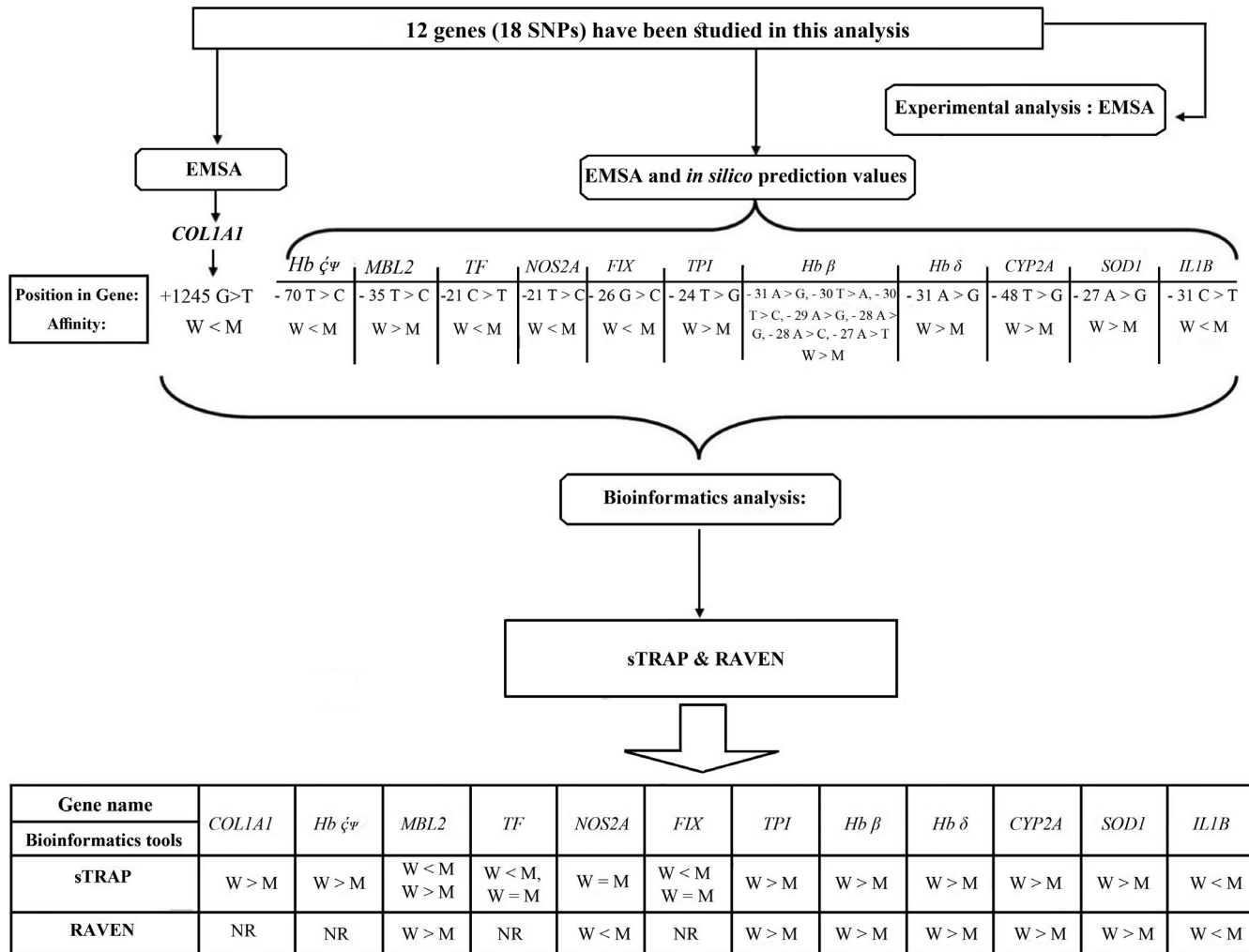


FIGURE 1 Workflow demonstrates the whole process in this study, consisting of experimental (EMSA) and bioinformatics data (sTRAP and RAVEN). W > M: The affinity is increased in the wild-type sequences (W) versus mutant sequences (M). M > W: The affinity increased in mutant sequences (M) versus wild-type sequences (W). W = M: There were no differences between two sequences. NR, not recognized

restrictions of sTRAP performance. Such an explanation would also lead to the uncertainty in concluding that the results of the 12 out of 18 studied allelic variations exactly matched with the results of the experimental approaches. Nevertheless, there is a “log-ratio ranking” of the affinities, which might not be properly responsive to the mentioned limitation, especially when the TF of the search is not among the highly ranked TFs.

The log ratio would rank the several TFs in a comparative model based on the higher absolute values of the difference between their binding affinities for the allelic variants of a specific motif. However, concerning an individual TF, the log ratio does not provide the probability value between two scores of the energy of TF for its binding sites (reference and mutant sequences). This finding of our study corresponded with the results of the other study (Macintyre, Bailey, Bailey, Haviv, & Kowalczyk, 2010). As an example, Skuse et al. (2014) used sTRAP to investigate if the sequence harboring rs237887 SNP, associated with social cognitive behavior, is

a TF-binding site and could induce altered gene expression. They reported the members of E26 transformation specific family of TFs being among those with significantly different binding affinities for their allelic motifs, ranked as top 11 candidate TFs. They established the hypothesis accordingly on the role of rs237887 SNP in the disease due to the altered TFs binding affinity, it makes. Their finding was according to the ranking of TFs based on the log ratio. However, as mentioned before, the relative logarithms do not provide a strong statistical tool to show the actual significant difference between the two values.

Furthermore, although the default threshold for the hit-based method used by sTRAP is normally set on 5, the ranking of the TFs by the tool is mostly performed based on the threshold 0, being less stringent than 5. The outcome of this might be the highly ranked TFs with minimal affinity binding to the query motif, due to higher log ratio value only. The issue mentioned here was also experienced in our analysis, as we had to adjust the threshold on 0 to have a list of TFs

in sTRAP output. So, the TFs with higher binding affinity for our motifs had lower scores of ranking as a result of the lower value of log ratio. This held true in most of the SNPs we analyzed.

In the view of the TFBSs inside *MBL2*, *TF*, and *FIX* genes in our analysis, there were discrepancies between the predicted scores, using mainly sTRAP, in the context of two different matrices applied for binding affinity prediction of TBP to any of the target sites (Table 1; Table S1). Such contradictory data might confuse the users to decide which result to consider. Besides, there is no *p* value between the two binding energies of a TF in such circumstances.

Manke et al. (2010) reported the investigation of 20 different SNPs in TFBSs, previously examined experimentally by Andersen et al. (2008), to evaluate the biophysical model, sTRAP, they had introduced. However, the Anderson et al. had examined a mix of either the nuclear extract, or cell extract, or the recombinant proteins as the sources of the putative TFs to analyze their binding energy.

Since cell and nuclear extract consist of a combination of several TFs, binding to the same motifs so, there might be a coincident activation of them. Therefore, the results of such an investigation might not indicate the binding energy of an individual target TF. Besides, this is not consistent with sTRAP, which exclusively computes the data of ChIP-seq and applies the individual TFs (Thomas-Chollier et al., 2011). We concluded that the protocol used by the authors to show the validity of sTRAP data might need to be designed more precisely. Nonetheless, extensive analysis of TFBSs as the cis-acting elements are required to establish any corroborated correlation between the computational and experimental results. Although a biophysical model does not describe the biological systems 100% due to several parameters, there would be an absolute need for its approval; and validation by experimental procedures to find the level and degree of its discrepancies and deviations from the results of the wet-lab experiments (Cooper et al., 2018). Without such challenges, sTRAP or any other biophysical model will not be known in terms of its performance. However, there might be a need, for extended analyses designed by several computational models and more integrated experimental data of SNPs analysis, for this purpose. Nevertheless, this is not a barrier to prevent researchers from looking at sTRAP in practice and experiment in a preliminary analysis. On the other side, sTRAP is one of the few accessible biophysical models to estimate the binding energy of the TFs for the wild and mutant sequences of the target TFBSs, web-based, free of charge, able to produce numerical scores for the analysis, user-friendly, specifically helping for the biologists with no need for strong background in mathematics and complex formulas, and no requirement for bioinformatics training to use such models. Although sTRAP has not been updated since its

establishment in 2011, the named characterizations have made it the tool of choice among the existing computational models for many researchers in their ongoing projects to formulate the hypothesis to link the TFBSs alleles to the diseases by the prediction results (Cavalli et al., 2019; Huber et al., 2019; Skuse et al., 2014; Thormann et al., 2018). Such a frequent application of sTRAP makes its validation against experimental procedures valuable.

5 | SUMMARY

The computational tool, sTRAP, in a user-friendly manner, is capable to scan the TFBSs allele and predict the binding energy of TFs for their target sequences, simply using DNA sequence context. It is a practical biophysical tool which can be easily applied by even nonexperienced users.

Taken as a whole, sTRAP as a biophysical tool that takes advantage of multiple models requires being validated through experimental data and empirical measurements for assessing limitations and confidence. There would be a need to check the quality of the performance of the bioinformatics tool to accept the accuracy of its prediction (Cooper et al., 2018). So, a large scale of experimental data integrated with biophysical tool might be a prerequisite for sTRAP optimization and validations to precisely score the SNPs variations in TF-TFBSs interactions. However, due to the complex scenario of TF activities in vivo (cross-talking of TFs and coincident activation of them, cross-talking of signal transduction pathways, numerous numbers of TFBS for an individual TF, existence of nonproductive interactions of genomic binding of TFs, chromatin modifications, cell type-specific TFs, ...; Adelaja & Hoffmann, 2019; Deplancke, Alpern, Alpern, & Gardeux, 2016; Huang et al., 2018; Keilwagen, Posch, Posch, & Grau, 2019; Mullen et al., 2011; Naidu, Kostov, Kostov, & Dinkova-Kostova, 2015; Xin & Rohs, 2018), the results of neither EMSA nor sTRAP, analyzing individual TF-binding energy could be an actual representation of the fate of nucleotide variations of TFBSs in vivo. Of note, the outcome of invalid estimations of a bioinformatics model might result in incorrect conclusions, and improper design of downstream experiments (Hayden, 2015). Our data indicated some limitations of sTRAP in the prediction of binding energy variations due to some SNPs inside TFBS in the human genome. To link an SNP to a disease or a phenotypic trait in a hypothesis, there might be a need to use sTRAP together with other bioinformatics models and validate their data by experimental sets.

Due to limitations in access to further experimental results in literature, this study has been presented as a preliminary analysis on the comparison of the experimental results and sTRAP data on the analysis of functional SNPs in the noncoding sequences of the human genome. However, for

more comprehensive results, there would be a need to expand the study using several computational models and integrating more experimental data for analysis.

To the best of the authors' knowledge, although several researchers have integrated "sTRAP" results in their studies and have compared them with the data obtained from other bioinformatics tools, this is the first report outlining the validation of the data of sTRAP by those of experimental approaches.

The data reported here add new information regarding sTRAP performance and might open a new window to the restrictions and capabilities of the biophysical tool, which requires being confirmed with an increased number of SNP analysis against the experimental sets.

ACKNOWLEDGMENTS

We would like to thank Pasteur Institute of Iran for supporting this study. This research did not receive any grants and was conducted using the authors' personal funds.

CONFLICT OF INTEREST

There was no conflict of interest to disclose.

ORCID

Mina Ebrahimi-Rad  <https://orcid.org/0000-0002-4208-2206>

REFERENCES

- Adelaja, A., & Hoffmann, A. (2019). Signaling crosstalk mechanisms that may fine-tune pathogen-responsive NFκB. *Frontiers in Immunology*, *10*, <https://doi.org/10.3389/fimmu.2019.00433>
- Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, *4*, e05005.
- Andersen, M. C., Engström, P. G., Lithwick, S., Arenillas, D., Eriksson, P., Lenhard, B., ... Odeberg, J. (2008). In silico detection of sequence variations modifying transcriptional regulation. *PLoS Computational Biology*, *4*, e5.
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., ... Snyder, M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*, *22*(9), 1790–1797. <https://doi.org/10.1101/gr.137323.112>
- Cavalli, M., Baltzer, N., Umer, H. M., Grau, J., Lemnian, I., Pan, G., ... Wadelius, C. (2019). Allele specific chromatin signals, 3D interactions, and motif predictions for immune and B cell related diseases. *Scientific Reports*, *9*(1), 1–14. <https://doi.org/10.1038/s41598-019-39633-0>
- Chorley, B. N., Wang, X., Campbell, M. R., Pittman, G. S., Noureddine, M. A., & Bell, D. A. (2008). Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutation Research*, *659*(1–2): 147–157. <https://doi.org/10.1016/j.mrrev.2008.05.001>
- Chowdhary, R., Tan, S. L., Pavesi, G., Jin, J., Dong, D., Mathur, S. K., ... Bajic, V. B. (2012). A database of annotated promoters of genes associated with common respiratory and related diseases. *American Journal of Respiratory Cell and Molecular Biology*, *47*(1), 112–119. <https://doi.org/10.1165/rcmb.2011-04190C>
- Cooper, C. I., Yao, D., Sendorek, D. H., Yamaguchi, T. N., P'ng, C., Houlihan, K. E., ... Boutros, P. C. (2018). Valection: Design optimization for validation and verification studies. *BMC Bioinformatics*, *19*(1), 1–11. <https://doi.org/10.1186/s12859-018-2391-z>
- Deplancke, B., Alpern, D., & Gardeux, V. (2016). The genetics of transcription factor DNA binding variation. *Cell*, *166*(3), 538–554. <https://doi.org/10.1016/j.cell.2016.07.012>
- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. <https://doi.org/10.1038/nature11247>
- Hayden, E. C. (2015). Journal buoys code-review push. *Nature*, *520*(7547), 276–277.
- Huang, H.-Y., Lin, Y.-C.-D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., ... Yu, Y. (2020). miRTarBase 2020: Updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Research*, *48*(D1), D148–D154.
- Huang, Q., Ma, C., Chen, L., Luo, D., Chen, R., & Liang, F. (2018). Mechanistic insights into the interaction between transcription factors and epigenetic modifications and the contribution to the development of obesity. *Frontiers in Endocrinology*, *9*, 370.
- Huber, R., Kirsten, H., Näkki, A., Pohlers, D., Thude, H., Eidner, T., ... Kinne, R. W. (2019). Association of human fos promoter variants with the occurrence of knee-osteoarthritis in a case control association study. *International Journal of Molecular Sciences*, *20*(6), 1382.
- Kamali, A. H., Giannoulatou, E., Chen, T. Y., Charleston, M. A., McEwan, A. L., & Ho, J. W. (2015). How to test bioinformatics software? *Biophysical Reviews*, *7*(3), 343–352. <https://doi.org/10.1007/s12551-015-0177-3>
- Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., ... Skoufos, G. (2018). DIANA-TarBase v8: A decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Research*, *46*(D1), D239–D245.
- Keilwagen, J., Posch, S., & Grau, J. (2019). Accurate prediction of cell type-specific transcription factor binding. *Genome Biology*, *20*(1), 9.
- Kumar, A., Wong, A.-K.-L., Tizard, M. L., Moore, R. J., & Lefèvre, C. (2012). miRNA_Targets: A database for miRNA target predictions in coding and non-coding regions of mRNAs. *Genomics*, *100*(6), 352–356. <https://doi.org/10.1016/j.ygeno.2012.08.006>
- Kumar, S., Ambrosini, G., & Bucher, P. (2016). SNP2TFBS—a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Research*, *45*, D139–D144.
- Lai, X., Stigliani, A., Vachon, G., Carles, C., Smaczniak, C., Zubieta, C., ... Parcy, F. (2019). Building transcription factor binding site models to understand gene regulation in plants. *Molecular Plant*, *12*(6), 743–763. <https://doi.org/10.1016/j.molp.2018.10.010>
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., ... Weirauch, M. T. (2018). The human transcription factors. *Cell*, *172*(4), 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>
- MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., ... Morales, J. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, *45*(D1), D896–D901.
- Macintyre, G., Bailey, J., Haviv, I., & Kowalczyk, A. (2010). is-rSNP: A novel technique for in silico regulatory SNP detection. *Bioinformatics*, *26*, i524–i530.

- Manke, T., Heinig, M., & Vingron, M. (2010). Quantifying the effect of sequence variation on regulatory interactions. *Human Mutation*, *31*, 477–483.
- Mann, V., Hobson, E. E., Li, B., Stewart, T. L., Grant, S. F., Robins, S. P., ... Ralston, S. H. (2001). A COL1A1 Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. *The Journal of Clinical Investigation*, *107*, 899–907.
- Mathelier, A., & Wasserman, W. W. (2013). The next generation of transcription factor binding site prediction. *PLoS Computational Biology*, *9*, e1003214.
- Maurano, M. T., Wang, H., Kutuyavin, T., & Stamatoyannopoulos, J. A. (2012). Widespread site-dependent buffering of human regulatory polymorphism. *PLoS Genetics*, *8*(3). <https://doi.org/10.1371/journal.pgen.1002599>
- Mullen, A. C., Orlando, D. A., Newman, J. J., Lovén, J., Kumar, R. M., Bilodeau, S., ... Young, R. A. (2011). Master transcription factors determine cell-type-specific responses to TGF- β signaling. *Cell*, *147*(3), 565–576. <https://doi.org/10.1016/j.cell.2011.08.050>
- Naidu, S. D., Kostov, R. V., & Dinkova-Kostova, A. T. (2015). Transcription factors Hsf1 and Nrf2 engage in crosstalk for cytoprotection. *Trends in Pharmacological Sciences*, *36*, 6–14.
- Park, M.-H., Rhee, H., Park, J. H., Woo, H.-M., Choi, B.-O., Kim, B.-Y., ... Koo, S. K. (2014). Comprehensive analysis to improve the validation rate for single nucleotide variants detected by next-generation sequencing. *PLoS ONE*, *9*(1). <https://doi.org/10.1371/journal.pone.0086664>
- Piriyaopongsa, J., Bootchai, C., Ngamphiw, C., & Tongsima, S. (2012). microPIR: An integrated database of microRNA target sites within human promoter sequences. *PLoS ONE*, *7*(3), e33888–e33888.
- Rana, M., Coshic, P., Goswami, R., & Tyagi, R. K. (2017). Influence of a critical single nucleotide polymorphism on nuclear receptor PXR-promoter function. *Cell Biology International*, *41*, 570–576.
- Riva, A. (2012). Large-scale computational identification of regulatory SNPs with rSNP-MAPPER. *BMC Genomics*, *13*(Suppl. 4). <https://doi.org/10.1186/1471-2164-13-S4-S7>
- Savinkova, L., Drachkova, I., Arshinova, T., Ponomarenko, P., Ponomarenko, M., & Kolchanov, N. (2013). An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein. *PLoS ONE*, *8*, e54626.
- Sikkema-Raddatz, B., Johansson, L. F., de Boer, E. N., Almomani, R., Boven, L. G., van den Berg, M. P., ... Sinke, R. J. (2013). Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Human Mutation*, *34*(7), 1035–1042. <https://doi.org/10.1002/humu.22332>
- Skuse, D. H., Lori, A., Cubells, J. F., Lee, I., Conneely, K. N., Puura, K., ... Young, L. J. (2014). Common polymorphism in the oxytocin receptor gene (OXTR) is associated with human social recognition skills. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(5): 1987–1992.
- Strom, S. P., Lee, H., Das, K., Vilain, E., Nelson, S. F., Grody, W. W., & Deignan, J. L. (2014). Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genetics in Medicine*, *16*(7), 510–515. <https://doi.org/10.1038/gim.2013.183>
- Thomas-Chollier, M., Hufton, A., Heinig, M., O'keeffe, S., El Masri, N., Roider, H. G., ... Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nature Protocols*, *6*, 1860.
- Thormann, V., Rothkegel, M. C., Schöpflin, R., Glaser, L. V., Djuric, P., Li, N., ... Meijnsing, S. H. (2018). Genomic dissection of enhancers uncovers principles of combinatorial regulation and cell type-specific wiring of enhancer–promoter contacts. *Nucleic Acids Research*, *46*(6), 2868–2882. <https://doi.org/10.1093/nar/gky051>
- Xin, B., & Rohs, R. (2018). Relationship between histone modifications and transcription factor binding is protein family specific. *Genome Research*, *28*(3), 321–333. <https://doi.org/10.1101/gr.220079.116>
- Zhao, Y., Granas, D., & Stormo, G. D. (2009). Inferring binding energies from selected binding sites. *PLoS Computational Biology*, *5*(12): e1000590. <https://doi.org/10.1371/journal.pcbi.1000590>
- Zhao, Y., Ruan, S., Pandey, M., & Stormo, G. D. (2012). Improved models for transcription factor binding site identification using nonindependent interactions. *Genetics*, *191*(3), 781–790. <https://doi.org/10.1534/genetics.112.138685>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Moradifard S, Saghiri R, Ehsani P, Mirkhani F, Ebrahimi-Rad M. A preliminary computational outputs versus experimental results: Application of sTRAP, a biophysical tool for the analysis of SNPs of transcription factor-binding sites. *Mol Genet Genomic Med*. 2020;8:e1219. <https://doi.org/10.1002/mgg3.1219>