Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

A Multi-Object Simulated Annealing Approach to Identify Likely Functional Genomic pairs

M. Hosseinpoor[1,2,] *, H. Parvin[3,4,5]

[1] Department of Computer Engineering, Estahban Branch, Islamic Azad University, Estahban 1979124119, Iran. mj.hosseinpoor@iauest.ac.ir

[2] Young Researchers and Elite Club, Estahban Branch, Islamic Azad University, Estahban, Iran.

[3] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam. parvin@iust.ac.ir

[4] Faculty of Information Technology, Duy Tan University, Da Nang 550000, Vietnam.

[5] Department of Computer Science, Nourabad Mamasani Branch, Islamic Azad University, Mamasani 7351, Iran.

* To whom correspondence should be addressed. E-mails: mj.hosseinpoor@iauest.ac.ir

Abstract: Metaheuristic algorithms have been frequently using to tackle optimization problems, however such algorithms in the analysis of health-related data is not commonly used as developing metaheuristic algorithms that work well on health-related data is a difficult task due to complexity of the health data in particular genomics and epigenetics data. One of the important tasks in genomics is to predict genomic elements that are incorporating together to regulate a disease-related genes. Predicting such elements are important as they can be used to develop a personalized cure. In this study, we present for the first time, a multi-object simulated annealing algorithm to identify enhancer-promoter like interactions from Hi-C (chromosome conformation capture) data. These regulatory elements can potentially play vital roles as promoters and/or enhancers in appearance and exacerbation of the regulation of gene.s To evaluate the efficiency of the proposed method, we applied our proposed method and traditional methods on the Hi-C data from mice and compared together. Our results show that the interacting elements identified by our new method are more likely to be functional. The source code of the method is publicly available.

Keywords: Metaheuristic algorithms, Health data analytics, Multi-object simulated annealing, optimization.

---

1. Introduction

Medical scientists have always been confronted with issues such as prediction, diagnosis, survival, and relapse of the diseases [1]. However, in the field of genetic diagnosis, gene regulation may be the main factor in the incidence of the disease [2] Gene regulation is an activity which could lead to gene expression or bring it to a

© 2021 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

halt. Factors such as silencers, enhancers, and promoters influence on gene regulation which are referred to as regulatory elements.

Tools and technologies known to date can identify the regulatory elements along genome. Also, they are capable of understanding the various interactions among the regulatory elements. For instance, the Hi-C technology was introduced by Aidin *et al* [3], to capture all the chromosome interactions [4, 5, 6]. The Hi-C method altogether with its tools represents three-dimensional interactions between different points of the genome in the regulatory elements. In other words, it determines what region of the genome interacts with the others. The extensive information obtained in this way provides a lot of knowledge concerning genome interactions.

Using Hi-C, many studies have engaged in detecting promoters-enhancers pairs and how they regulate the genes and how they interact with each other [7, 8, 9]. During the past decade, a wide range of tools have been proposed to effectively analysis Hi-C data and detect informative interacting pairs in the genome. One of these tools is Homer which was developed by [10]. This tool is capable of detecting valid interactions among the Hi-C data set [10]. Later, HiCUP was introduced in [11]. Similar to Homer, it can perform the detection of valid interactions. It is also able to execute mapping and Allele specific analysis [11]. The three tools of HiCdat [12], HiClib [13], and Hic-Box [14] that were introduced later were able to carry out the operations of mapping, detection of valid interactions, binning, and correction of systematic noise. Different to previously methods, HiC-Pro that was introduced by [15] is able to do parallel implementation of mapping, quality checking and interaction calling. Note that all the tools mentioned above could enumerate the interactions between different points in the regulatory elements.

The methods mentioned above are only able to detect significant interacting pairs through dividing the genome to fix-bin size fragments, which is a time-consuming task and also identify many non-functional interactions. Here, we used a heuristic based technique based on our previous frameworks [16, 17, 18, 19, 20] to extracts the Hi-C interacting pairs and counts the interactions among them through a bi-object algorithm depending on the simulated annealing. Such interacting pairs have the potential to be functional interactions that play a key role in regulating the gene expression. The rest of this article is organized as follows. In the second section we will introduce the data set used in this study as well as our proposed method. In Section 3 we present our results and discussion. Finally, we present our conclusion and future work in the final section.

2. Methods AND Materials

2.1. Dataset

The Hi-C data set used here is the set of mouse embryonic stem cells (ESCs) [1, 21]. Specifications of the dataset show in table 1. The length of each fragment used in the dataset is 5kb.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

Table 1. The Hi-C Dataset Used.

Hi-C Database	Cell type	Number of reads	Number of fragments	Number of tuples
ESCs Mouse	ESCs	7,9450,581	530,922	61,054,972

2.2. Pre-Processing

We used a recently developed tool MHiC [22] for mapping, filtering and interaction calling of Hi-C data. FASTQ paired-end reads were aligned to the mm9 genome and filtered through the HiC-Pro [15] pipeline developed. To remove duplicate reads, filter for valid interactions, and generate Hi-C interaction matrices (5kb and 10kb bin-sizes), we set the following parameters: MIN_FRAG_SIZE = 230; MAX_FRAG_SIZE = 1100; MIN_INSERT_SIZE = 120; and MAX_INSERT_SIZE = 990. All experimental artifacts, such as circularized reads and re-ligations, singletons have been filtered out using HiC-Pro. We also excluded self-interactions from the downstream analyses. For getting real Hi-C interactions, we used a recently developed background correction model, MaxHiC [8] to identify significant Hi-C interactions. Here, we included those significant interactions with P-value < 0.01, read-count >=5.

2.3. The Technique

The simulated annealing (SA) algorithm is a simple and efficient heuristic algorithm which is used in the optimization problems [23, 24, 25]. The method is employed in metallurgy to reach a state in which the energy of solid matter is minimized [23, 25]. This technique includes putting the material at high temperature and then bringing down the temperature, slowly.

The SA algorithm starts at an initial solution and then moves towards the neighboring solutions through a repeating loop. If the neighboring solution would be better than the present solution, the algorithm would set it as the solution (that is, tends to it), otherwise, the algorithm recognizes that as the solution with probability $exp(-\Delta F/T)$, where ΔF is the difference between the objective function of the present solution and the neighboring solution, and T is the parameter dubbed as temperature [23]. In each temperature, several repetitions are performed and then the temperature is slowly reduced [24].

The SA is works in the following manner. In the early steps, the temperature is taken very high to be more likely to accept worse answers. With gradual decrease of temperature, there would remain fewer probabilities for acceptance of worse solutions, and hence the algorithm would converge to a favorite solution [26-28]. In each step, the SA algorithm regards several cases at the neighborhood of the current mode S, and decides probably whether transforms the system from the *S* mode or stays the same. These probabilities lead the system eventually to the mode of lower energy level. Approaching to this solution is happens in the following conditions:

1. The new solution is better than the previous one.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

2. The value of the movement probability function [i] is greater than a random number from the domain [0, 1].

Otherwise, the checker would generate and estimate a new solution. This step by step movement continues until the stop condition for the algorithm being satisfied [26]. The value of the movement probability function is computed in every step from the relation *exp(-ΔF/T)*. The difference between this equation and the object function is the value for the current and the new solution [26,29, 30]. Theoretically, overcoming local optimization, the SA technique is able to find the absolute optimized solution. The main idea in the SA method is to generate a solution at the neighborhood of the current solution and to effectively calculate the variation in the object function *ΔF*. Then, after saving the best solution, the new solution will be accepted if *Δf* <=0. Otherwise, if *Δf* >0 the new solution would be accepted with a probability [24].

2.4. Multi-objective detection of interacting pairs problem

Take the Hi-C dataset with *n* items of data as follows:

$$Hi-C\ Data = \{\xi_1 \sim \xi_{1\le k\le t}, \xi_2 \sim \xi_{1\le k\le t}, \dots, \xi_k \sim \xi_{1\le k\le t}\} \qquad \text{Eq. 1}$$

where ξ is a region of genome in a chromosome. The aim of the problem is to find repetitive items $RE_1 \sim RE_2$, where RE_1 and RE_2 are as follows:

$$RE_2 = \{\xi_m, \xi_{m+1}, \dots, \xi_n\}, m < n,\ 1 \le i, j, m, n \le t \quad , \quad RE_1 = \{\xi_i, \xi_{i+1}, \dots, \xi_j\}, i < j$$

Eq. 2

Here, RE_1 is an interacting pair of the genome in a chromosome defining a potential promoter, and RE_2 is another interacting pair of the genome in a chromosome which defines a potential enhancer. The dataset used in this research is Hi-C which is used to show the interactions among different points of the genome in the cancer patients. Our aim here is to find the interaction $RE_1 \sim RE_2$, where RE_1 and RE_2 are potential promoters and enhancers elements.

2.4.1. Detection of multi-objective interacting pairs

In this section, our proposed method in detecting and extracting interacting pairs from the Hi-C data. The approach is based on the bi-objective simulated annealing algorithm. To explain the objectives of the algorithm, let $RE_1 \sim RE_2$ be a solution to the problem as it was specified in the previous section.

2.4.1.1. φ Criterion

The criterion φ is defined as the proportion of the sum of all interactions of any item in RE_1 with any item in RE_2 to the sum of all the interactions of the items in RE_1 with any item in Hi-C Data. The value φ falls in the interval [0, 1]. The higher this value, the better φ criteria would be attained.

$$\varphi\ (RE_1 \sim RE_2) = \frac{Number\ of\ interaction\ (RE_1 \sim RE_2)}{Total\ of\ interaction(RE_1)} \qquad \text{Eq. 3}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

2.4.1.2.Ψ Criterion

The Ψ criterion is defined as the size of the interacting pair detected in the genome. As in the following formula, the Ψ of RE_1 is the summation of the size of all ξ belonging to RE_1.

$$\Psi(\mathrm{RE}_1) = Sum(Size(\xi_i, \xi_{i+1}, \dots, \xi_j) \qquad \text{Eq. 4}$$

2.4.1.3. Fitness Function

The fitness function for two potential promoter and enhancer (RE_1, RE_2) in a chromosome is, in accordance with equation (3), the proportion of ratio criterion to the sum of all the sizes of RE_1 and RE_2.

$$Fitness(RE_1 \sim RE_2) = \left(\frac{\varphi(RE_1 \sim RE_2)}{\Psi(RE_1 + RE_2)}\right) \qquad \text{Eq. 5}$$

The value of the fitness function belongs to the interval [0,1]. The higher interaction between these interacting pairs, the closer this value would be to 1. In this formula, for the value of fitness function to be maximum, the ratio criterion (the first objective) should be maximum and the length criterion (the second objective) should be minimum. These conditions tell us that for the minimum length sum of the detected interacting pairs in the genome we will have the most interactions. Also, the fitness function being maximum means that the detected interacting pairs must have strong interaction with each other.

2.5. The proposed method

In this section, we present our suggested algorithm. The algorithm is a multi-objective model based on SA to detect and extract likely functional interacting pairs in the Hi-C dataset. Besides, it is able to enumerate all the interactions among the interacting pairs. Overcoming the local optimization, the proposed method could find the absolute optimal solution as well. The main idea in this method is to generate a solution at the neighborhood of the current solution and to effectively calculate the variation in the object function Δf. Afterwards, saving the best solution, the new solution will be certainly accepted if Δf<=0. Otherwise, if Δf >0, the new solution would be accepted. In Figure 1, the input of the algorithm would be the Hi-C dataset including the information about the inter-genome interactions in each chromosome, the maximum iteration (MaxIt), and the maximum sub-iteration (SubIt). At the pre-processing stage, the operations of sequencing, mapping, and trimming are done on the input dataset. Then, the initial population is created randomly the same extent as MaxIt to detect and extract the interacting pairs**.** At the next stage, each individual from the input population would be assessed via the fitness function. In the next step, 10 neighborhoods (the same number as SubIt) around the best individual are generated from the population of the previous stage before each assessed using the function in formulae 3. After that, the best solution found in each repetition with the most favorite assessment function will be saved. Consequently, the best solutions in each iteration represent the optimal detected regulatory elements in the dataset. Finally, all the detected interacting pairs in the RE set will be displayed as the output (**Figure 1**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

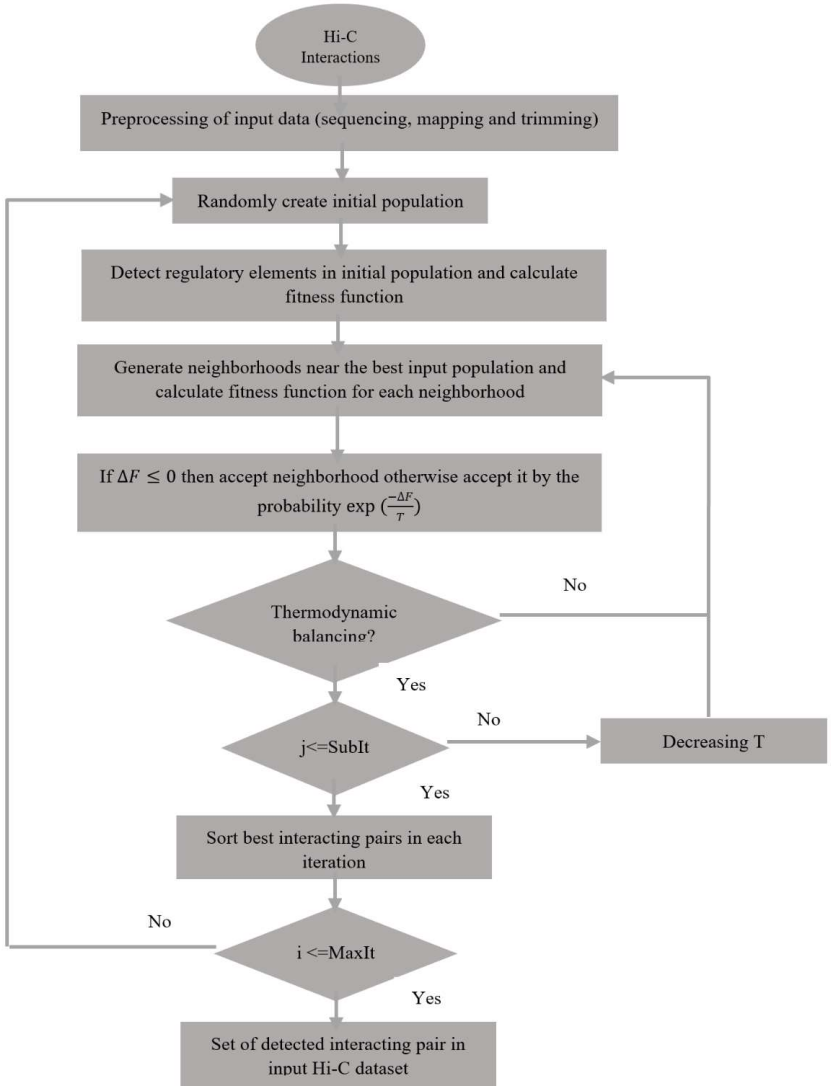


Figure 1. Flowchart of Suggested Algorithm.

2.5.1. Parameters

Table 2 shows the parameters set in the proposed method. According to Table 2, we investigated different values for these parameters and choose those that obtained the best results. The fitness for each individual is calculated via the fitness function. After the computations of the fitness function, the individual from among the input population which has the interacting pairs with the best fitness function is considered as the best actual answer. We repeat the following steps the same number of times as Maxit. First, we create the new population of 20 neighbors close to the resulted present optimal solution. Then, the fitness of each is compared to the optimal solution using the assessment function. If the new solution is better than the best one obtained so far, then we take it as the best new solution [25]. Afterwards, the best solution found will be updated followed by the temperature drop coefficient as the final action. This process would be repeated for the same number of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

iterations and at each iteration, the optimal solution (including the best detected interacting pairs) will be recorded.

Table 2. Parameters of Suggested Algorithm.

Parameters	Values
Number of Max Iteration	1000
Number of Sub iteration	200
Number of Initial Population	25
Number of neighborhoods	20
Initial temperature	10
Temperature drop coefficient	0.99

2.5.2. Individual representation

The structure of each individual of the input population in the problem of discovering and extracting the promoters/enhancers in the Hi-C dataset is as follows. We put in plain words each part of this structure.

Chr name	Left Side	Rand- Seq Left Side	Right Side	Rand- Seq Right Side

ChrName: is the name of the chromosome.
Left Side: determines the beginning part of the genome containing the interacting pair at the left side of the interaction.
Rand-Seq Left Side: is the length of the interacting pair at the left side of the interaction.
Right Side: determines the beginning part of the genome containing the interacting pair at the right side of the interaction.
Rand-Seq Right Side: is the length of the interacting pair at the right side of the interaction.

2.6. Measuring the Performance

The efficiency of the proposed algorithm is evaluated by two decisive factors. The first is the number of interactions between detected interacting pairs, and the second is the value of the fitness function evaluating the rate of inter-genome interactions with regard to the size of genome regions.

2.6.1. The number of inter-genome interactions

Figure 3 shows the plot box related to the interactions detected in the suggested method compared to traditional methods (HiC-Pro [15], HiCUP [11]). As shown in this Figure, the number of detected interactions in our proposed method is more than that in traditional methods. The same holds for the number of average interactions detected in the two methods. The following figure confirms that our suggested method behaves more efficiently than the traditional methods strategy in the detection of likely functional Hi-C interactions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

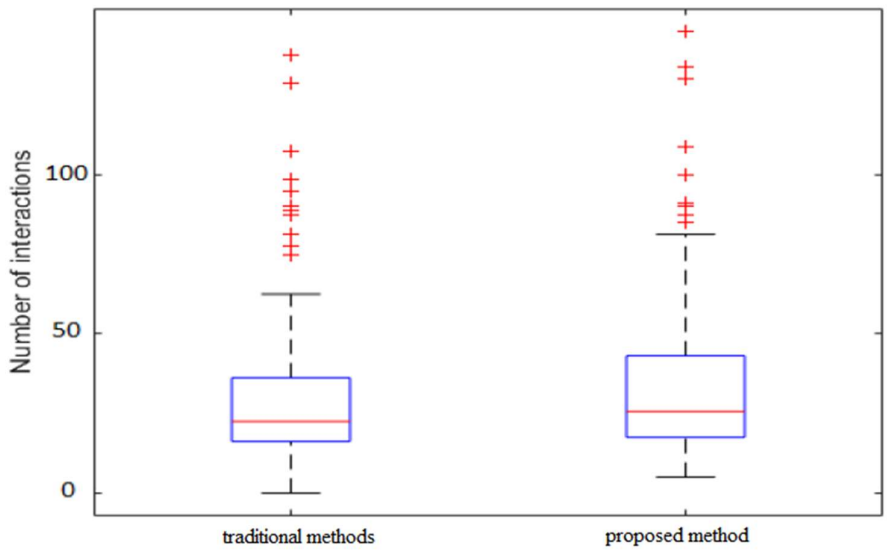


Figure 3. Box plot of the number of interacting reads identified by our proposed method and the traditional methods.

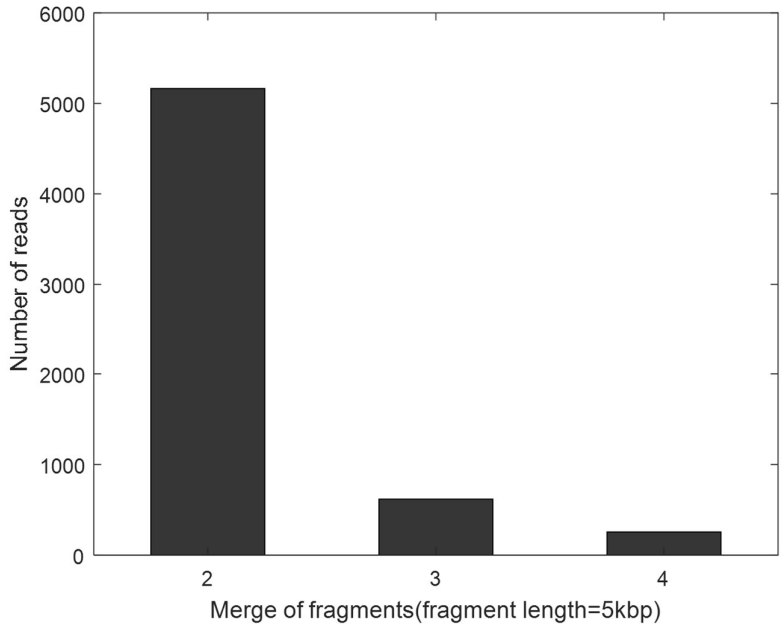
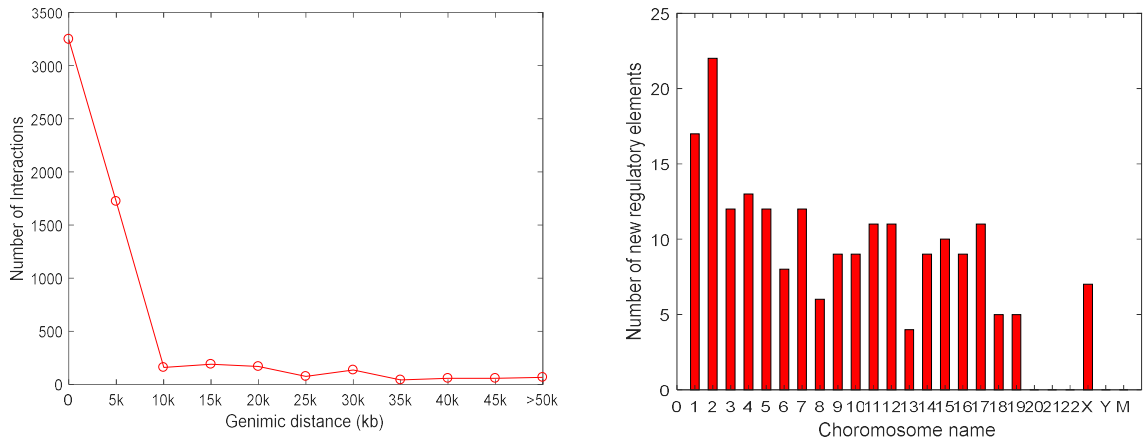


Figure 4. Relationship between the number of reads among detected interacting pairs and the number of merged fragments.

Figure 4 illustrates the relationship between the number of reads counted between the detected interacting pairs and the number of merged fragments. As shown in this figure, the longer the merged fragments, the less the number of counted reads between them would be. Based on the objective 2 in the proposed method, the fitness function would be minimized in the same level as the regions of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

the merged fragments get wider. The fitness function being minimized means that the detected interacting pairs interact weakly with each other.



(a) **(b).**

Figure 5. a) Relationship between the number of interactions between interacting pairs and the distance between them. **b)** The number of interacting pairs in Hi-C dataset per chromosome identified by our proposed method which were not identified by the traditional methods.

Figure 5a, demonstrates the relation between the number of detected interacting pairs and the distances between them. According to this figure, 50% of the detected interactions occur among those interacting pairs being 0kb apart. Also, 30% of the interactions belong to those which are 5kb away. The interacting pairs being 10kb, 15kb, and 20kb away from each other constitute 15% of the interactions, and the rest 5% comprise those interacting pairs positioned more than 25kb apart. As shown in **Figure 5a**, the number of interactions between two interacting pairs has a reverse relationship with the distance between them. **Figure 5b** also illustrates the number of interacting pairs per chromosome in the input Hi-C dataset detected by our proposed method of which the traditional methods have not been able to detect them. Among them, the highest number of the detected interacting pairs is related to chromosome 2, and the least belongs to the two chromosomes 18 and 19. Our proposed method does not detect any new interacting pairs in chromosomes 20, 21, 22, and Y.

2.6.2. Fitness Function Value

Table 3 demonstrates three new interacting pairs detected by our proposed method. The interacting pairs in row 1 is the outcome of merging the regions 155,580,000-155,585,000 and 155,585,000-155,590,000 in chromosome 1. The corresponding region interacts with the region 155,565,000-155,570,000 from chromosome 1. The fitness function gives the value 1.10E-05 for the resulted interacting pair. As shown in the second and third row of this table, this value of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

fitness function is more favorable than the values of those interacting pairs identified by the traditional methods.

Table 3. Three new interacting pairs detected by our proposed method.

		Left Side				Right Side				
Number	Method	Chromosome Number	Start	End	Number of Read	Chromosome Number	Start	End	Number of Interactions	Fitness Value
1	Propose	1	155,565,000	155,570,000	145	1	155,580,000	155,590,000	24	1.10E-05
2	Fix-bin	1	155,565,000	155,570,000	145	1	155,580,000	155,585,000	9	6.21E-06
3	Fix-bin	1	155,565,000	155,570,000	145	1	155,585,000	155,590,000	15	1.03E-05
4	Propose	2	174,235,000	174,240,000	110	2	174,285,000	174,300,000	28	1.27E-05
5	Fix-bin	2	174,235,000	174,240,000	110	2	174,285,000	174,290,000	11	1.00E-05
6	Fix-bin	2	174,235,000	174,240,000	110	2	174,290,000	174,295,000	10	9.09E-06
7	Fix-bin	2	174,235,000	174,240,000	110	2	174,295,000	174,300,000	7	6.36E-06
8	Propose	10	42,220,000	42,225,000	521	10	42,240,000	42,260,000	52	3.99E-06
9	Fix-bin	10	42,220,000	42,225,000	521	10	42,240,000	42,245,000	14	2.69E-06
10	Fix-bin	10	42,220,000	42,225,000	521	10	42,245,000	42,250,000	17	3.26E-06
11	Fix-bin	10	42,220,000	42,225,000	521	10	42,250,000	42,255,000	6	1.15E-06
12	Fix-bin	10	42,220,000	42,225,000	521	10	42,255,000	42,260,000	15	2.88E-06

The row 4 shows the interacting pairs resulted from the merge of the three regions (174,285,000-17,4290,000, 174,290,000-174,295,000, and 174,295,000-174,300,000) of chromosome 2. The merged region interacts with the region 174,235,000-174,240,000 of the same chromosome. The fitness value of this new interacting pairs is also better than the values of those detected by the traditional methods (the rows 5, 6, and 7).

The row 8 of the table is dedicated to the result of merging the four regions of 42,240,000-42,245,000, 42,245,000-42,250,000, 42,250,000-42,255,000, and 42,255,000-42,260,000 from chromosome 10. The resulting region interacts with the region 42,220,000-42,225,000 of chromosome 10. The value attributed to the fitness function at the detected interacting pairs is much better than all the values obtained from the traditional methods (the rows 9, 10, 11, and 12).

The diagram in **Figure 6**, outlines the average value of fitness function in the proposed method along with enumeration of their interactions per chromosome

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

compared to the traditional methods. As shown in this figure, the average fitness per chromosome is much better in our proposed compared to the traditional methods for determining the interacting pairs. As it was explained earlier in Section 2, the fitness function being better means that the number of interactions among the detected interacting pairs is more than that in the traditional approaches.

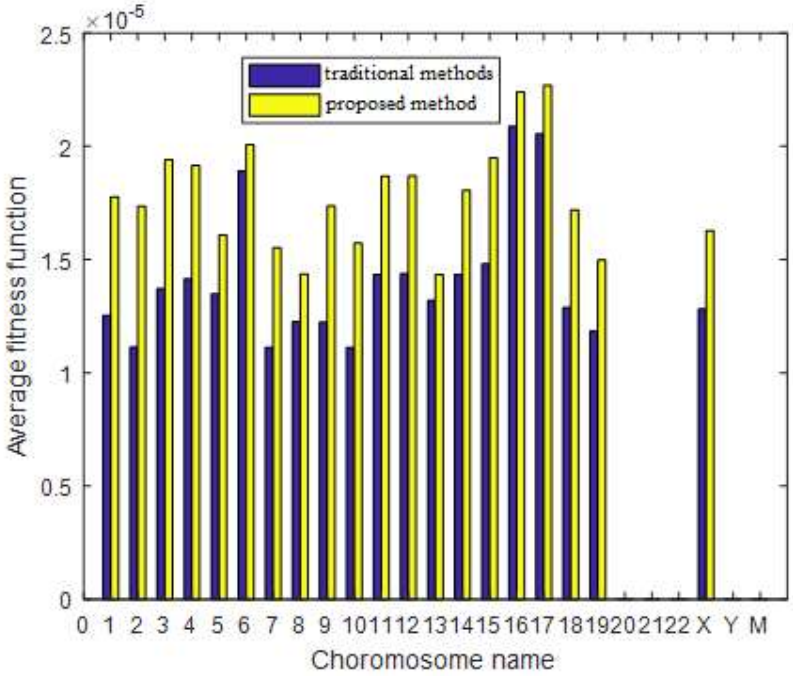


Figure 6. Diagram comparing average fitness values and interactions count per chromosome in suggested method and the traditional methods.

2.6.3. Validating the Method and Results

This section compares all detected interacting pairs with the regions reported as promoters by Stefan *et al.* along with their interactions with other regions [1, 21]. Our results illustrated in **Figure 7** demonstrates that 70% of the interacting pairs recognized by our proposed method overlap the enhancer-promoter pairs reported in [21], and the rest 30% does not. **Figure 7** shows the percentage of overlapping and non-overlapping of the new interacting pairs detected by the suggested method with significant enhancer-promoter regions identified in Stefan *et al*. This proves the ability of our proposed method in detection and extraction of the interacting pairs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

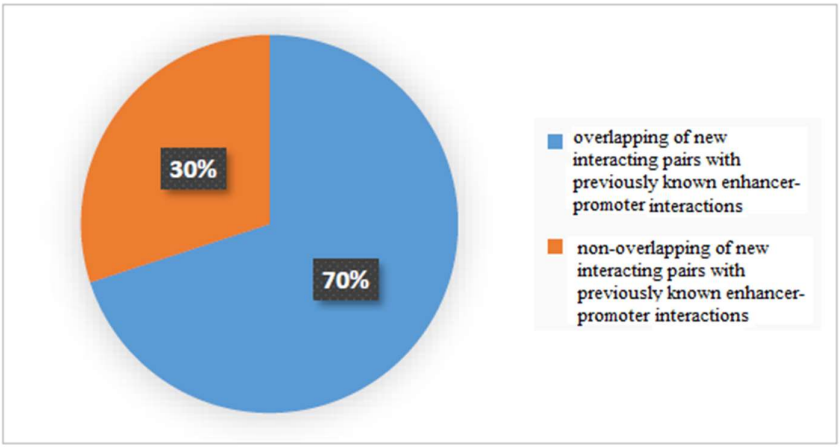


Figure 7. Percentage of overlapping and non-overlapping interacting pairs with enhancer-promoter pairs.

So, here the regions detected in the proposed method were compared with the significant promoters reported in Stefan *et al*. Then, according to Figure 7, the discovered areas of the proposed method was able to match 70% with significant promoters. This indicates that the proposed method correctly detects promoter regions.

3. Conclusion

Here, we proposed a multi-objective approach to identify likely functional genomic interacting pairs. The average value of the fitness function related to the suggested algorithm per chromosome is much better than the traditional method in identifying genomic interacting pairs. Indeed, the suggested algorithm derives those interacting pairs that were not identified by the traditional methods. This "being better" of fitness values entails that the number of interactions among the extracted interacting pairs by the proposed method are more than of the interactions identified by the traditional methods.

As the final point, the detected interacting pairs in our method with the ESCs data went under comparison with the significant enhancer-promoter regions reported by Stefan *et al* [1, 21]. The comparison proved that 70% of the interacting pairs recognized via our suggested method have overlapped the Stephan's enhancer-promoter pairs. This percentage represents the appropriate successfulness of the suggested approach. This means that our technique has strongly enumerated the detected and extracted regulatory elements.

Authors' contributions: MH developed the concept. MH wrote the manuscript; MH and HP revised the manuscript. HP and MH provided the data. MH carried out tool implementation. MH, HP generated all figures and tables. MH, HP performed statistical analyses. All authors have read and approved the final version of the paper.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

Ethics approval and consent to participate: Not applicable.

Consent for publication: All authors have read and approved the final version of the paper.

Availability of data and material: All data are publicly available.

Funding: This work was supported by grants to HP.

Competing interests: The authors declare that they have no financial or non-financial competing interests.

References

1. Novo, C. L., et al. (2018). "Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition." Cell Rep 22(10): 2615-2627.
2. Tang H., L. T., Qiu T., Park Y. (2012). "Segmentation of Heart Sounds Based on Dynamic Clustering." Biomedical Signal Processing and Control: 509–516.
3. Lieberman-Aiden E, v. B. N., Williams L, Imakaev M, Ragoczy T, Telling A, et al (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science: 289–293.
4. Charalampos L, S. K., Panagiotis N, Iannis A, Aristotelis T (2017). "HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking." BMC Genomics: 3387-3403.
5. Chepelev I, W. G., Wangsa D, Tang Q and Zhao K (2012). "Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization." Cell Research: 490-503.
6. Jin F., L. Y., Dixon JR, Selvaraj S, Ye Z, Lee AY,Yen CA, Schmitt AD, et al (2013). "A high- resolution map of the three-dimensional Chromatin interactome in human cells." Nature letter 290-294.
7. Hwang Y-C, L. C.-F., Valladares O, Malamon J, Kuksa PP, Zheng Q, et al (2015). "HIPPIE: a high-throughput identification pipeline for promoter interacting enhancer elements." Bioinformatics: 1290–1292.
8. Alinejad-Rokny H., Ghavami R., Rabiee HR., Rezaei N., Tam KT., Forrest A (2020). “MaxHiC: robust estimation of chromatin interaction frequency in Hi-C and capture Hi-C experiments”, bioRxiv, 2020.
9. Sean W, R. M., Katherine SP (2016). "Enhancer–promoter interactions are encoded by complex genomic signatures on looping chromatin." Nature Genetics: 488–496.
10. Heinz S, B. C., Spann N, Bertolino E, Lin YC, Laslo P, et al (2010). "Simple combinations of lineage-determining transcription factors prime cisregulatory elements required for macrophage and B cell identities." Mol Cell: 576–589.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

11. Wingett S, E. P., Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, et al (2015). "HiCUP: pipeline for mapping and processing Hi-C data." F1000Res: 1310-1322.
12. Schmid Marc W, S. M., Grob S, Grossniklaus U (2015). "HiCdat: a fast and easy-to-use Hi-C data analysis tool. BMC Bioinformatics. ;16:277." BMC Bioinformatics: 16:277.
13. Imakaev M, F. G., McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al (2012). "Iterative correction of Hi-C data reveals hallmarks of chromosome organization." Nat Meth: 999–1003.
14. R, K. (2016). HiC-Box, https://github.com/rkoszul/HiC-Box.
15. Servant N, V. N., Lajoie BR, Viara E, Chen CJ, Vert JP, et al (2015). "HiC-Pro: an Optimized and Flexible pipeline for Hi-C Data Processing." Genome Biol, 259-270.
16. Hosseinpoor M., Parvin H., Nejatian S., Rezaie V., Bagherifard K., Dehzangi A. (2019). “Proposing a novel community detection approach to identify cointeracting genomic regions”, Mathematical Biosciences and Engineering, 17(3): 2193-2217.
17. Javanmard R., JeddiSaravi K. (2013). “Proposed a new method for rules extraction using artificial neural network and artificial immune system in cancer diagnosis”, Journal of Bionanoscience, 7(6): 665-672.
18. Parvin H., MirnabiBaboli M. (2015). “Proposing a classifier ensemble framework based on classifier selection and decision tree”, Engineering Applications of Artificial Intelligence, 37: 34-42.
19. Moore, J.E., Pratt, H.E., Purcaro, M.J. et al (2020). A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. Genome Biol 21, 17.
20. Lee, B.H., Rhie, S.K. (2021). “Molecular and computational approaches to map regulatory elements in 3D chromatin structure”, Epigenetics & Chromatin 14**,** 14.
21. Stefan S, M. F., et al (2015). "The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements." Genome Research: 582-597.
22. Khakmardan S., Rezvani M., Pouyan A. Fateh M. (2020). “MHiC, an integrated user-friendly tool for the identification and visualization of significant interactions in Hi-C data”, BMC genomics, 21(1): 1-10.
23. Anwar, F., Waters, S. A., Davenport, M. P., & Ebrahimi, D. (2016). Source of CpG depletion in the HIV-1 genome. Molecular biology and evolution, 33(12), 3205-3212.
24. Hansen, P. B. (1992). Simulated Annealing. Syracuse University, School of Computer and Information Science, Electrical Engineering and Computer Science.
25. Konishi, T., et al. (2020). "Using simulated annealing for locating array construction." Information and Software Technology 126: 106346.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2021 doi:10.20944/preprints202107.0358.v1

26. Kang, M., et al. (2020). "Computationally efficient optimization of wavy surface roughness in cooling channels using simulated annealing." International Journal of Heat and Mass Transfer 150: 119300.
27. Ahmadinia, M., Meybodi, M. R., Esnaashari, M., H. (2013). Energy-efficient and multi-stage clustering algorithm in wireless sensor networks using cellular learning automata. IETE Journal of Research, 59(6), 774-782.
28. Ahmadinia, M., Ahangarikiasari, H. (2014). Data aggregation in wireless sensor networks based on environmental similarity: A learning automata approach. Journal of Networks, 9(10), 2567.
29. Niu, H., Khozouie, N., Parvin, H., Beheshti, A., & Mahmoudi, M. R. (2020). An ensemble of locally reliable cluster solutions. Applied Sciences, 10(5), 1891.
30. Esmaeili, L., Minaei-Bidgoli, B., & Nasiri, M. (2012). Hybrid recommender system for joining virtual communities. Research Journal of Applied Sciences, Engineering and Technology, 4(5), 500-509.