Computational and Structural Biotechnology Journal 20 (2022) 5680–5689

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

TREAT: Therapeutic RNAs exploration inspired by artificial intelligence technology

Yufan Luo [a,b,1], Liu Liu [a,1], Zihao He [a,1], Shanshan Zhang [c], Peipei Huo [c], Zhihao Wang [c], Qin Jiaxin [a], Lianhe Zhao [a], Yang Wu [a], Dongdong Zhang [e], Dechao Bu [a,d,*], Runsheng Chen [e,f,*], Yi Zhao [a,*]

[a] Research Center for Ubiquitous Computing Systems, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Luoyang Zhongke Information Industry Research Institute, Luoyang, China
[d] Hwa Mei Hospital, University of Chinese Academy of Sciences, China
[e] Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[f] Shenzhen Institute of Nucleic Acid Drug Research, Shenzhen Bay Laboratory Pingshan Translational Medicine Center, Shenzhen 510800, China

ARTICLE INFO

Article history:
Received 30 June 2022
Received in revised form 4 October 2022
Accepted 5 October 2022
Available online 8 October 2022

Keywords:
RNA
Therapeutics
Noncoding

ABSTRACT

Recent advances in RNA engineering have enabled the development of RNA-based therapeutics for a broad spectrum of applications. Developing RNA therapeutics start with targeted RNA screening and move to the drug design and optimization. However, existing target screening tools ignore noncoding RNAs and their disease-relevant regulatory relationships. And designing therapeutic RNAs encounters high computational complexity of multi-objective optimization to overcome the immunogenicity, instability and inefficient translational production. To unlock the therapeutic potential of noncoding RNAs and enable one-stop screening and design of therapeutic RNAs, we have built the platform TREAT. It incorporates 43,087,953 regulatory relationships between coding and noncoding genes from 81 biological networks under different physiological conditions. TREAT introduces graph representation learning with Random Walk Diffusions to perform disease-relevant target screening, in addition to the commonly utilized Topological Degree and PageRank algorithms. Design and optimization of large RNAs or interfering RNAs are both available. To reduce the computational complexity of multi-objective optimization for large RNA, we stratified the features into local and global features. The local features are evaluated on the fixed-length or dynamic-length local bins, whereas the latter are inspired by AI language models of protein sequence. Then the global assessment is performed on refined candidates, thus reducing the enormous search space. Overall, TREAT is a one-stop platform for the screening and designing of therapeutic RNAs, with particular attention to noncoding RNAs and cutting-edge AI technology embedded, leading the progress of innovative therapeutics for challenging diseases. TREAT is freely accessible at https://rna.org.cn/treat.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Over 95 % of the currently known drugs act by targeting proteins in the human body and affect their activity [1]. However, recent advances in RNA engineering have enabled the development of RNA-based therapeutics for a broad spectrum of applications [2]. These novel drugs are of rapid and cost-effective development [3], relatively simple to manufacture, and can act on previously undruggable targets [4]. In 2016, The U.S. Food and Drug Administration (FDA) approved Spinraza (nusinersen), one type of RNA drug approved to treat children and adults with spinal muscular atrophy (SMA) [5]. In 2018, the FDA approved patisiran, the first siRNA drug, to treat hereditary transthyretin amyloidosis [6]. In 2020 and 2021, mRNA vaccines have been developed globally at the forefront of efforts to combat the coronavirus disease (COVID-19 pandemic) [7]. The unprecedented efficacy of these mRNA vaccines has reignited interest in RNA therapeutics [8].

Abbreviations: siRNA, small interfering RNA; mRNA, messenger RNA; ncRNA, Noncoding RNA; PPI, Protein protein interaction; TF, Transcript Factor.

* Correspondence authors at: Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China (Y. Zhao).

E-mail addresses: budechao@ict.ac.cn (D. Bu), crs@ibp.ac.cn (R. Chen), biozy@ict.ac.cn (Y. Zhao).

[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.csbj.2022.10.011
2001-0370/© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Synthetic RNA could be engineered to express therapeutic proteins or manipulate specific genes' expression, making it particularly useful in drug and vaccine development for pathologies with established molecular targets, such as infectious diseases, tumors, and Mendelian disorders [4]. The recent advent of multi-omics technology is allowing scientists to probe the complex, transient and diverse molecular changes that underpin the course of disease and response to treatment, driving the discovery of novel targets in human health and disease [9]. The ability to manipulate those targets, especially those noncoding RNA occupying 85 % of the human genome, shows great promise to open diseases once deemed "undruggable" by small molecules and proteins, thus flaring up new avenues for treating intractable diseases [4].

The early stages of developing RNA therapeutics start with initial steps of RNA target screening and move to the later stages of design and optimization [10]. Those therapeutic RNAs refer to antisense oligonucleotides (ASOs), small interfering RNAs (siRNAs), or large RNAs, such as messenger RNA (mRNA), long noncoding RNA(lncRNA) and circular RNA [11]. Multi-omics integration and network-based analysis are usually employed to fully understand complex biological processes, thus identifying suitable disease targets [12]. Open Targets is an innovative public-private partnership that uses human genetics and genomics data for systematic drug target identification and prioritisation [13]. The Network Analyst platform is a molecular network analysis and visualization platform that integrates Protein–protein interaction (PPI) networks and gene regulatory networks to evaluate the significance of specific targets. [14]. However, most existing tools ignore noncoding RNAs and their heterogeneous regulatory roles. In addition, the spatiotemporal specificity of noncoding RNAs [15] imposes new requirements to regard various physiological conditions.

Irrespective of their therapeutic mechanism of action, the instability and immunogenicity of some therapeutic RNAs make it difficult to function on their own. Thus, the candidate RNA needs undergo sequence optimization and modification to enhance RNA stability and reduce immunogenicity. Complex calculations were regarded concerning RNA characteristics, such as translation efficiency, secondary structure, GC content, and off-target possibility. Sequence optimization for long RNAs, such as mRNA, is extremely challenging due to the exponentially large search space. Taking the S protein of COVID-19 as an example, there are a total of 1273 amino acids, and the possible RNA candidates are $2.4*10^{632}$. It is impossible to investigate all combinations to select the optimal one with current computing power. Several tools have been developed for sequence optimization, such as LinearDesign [16], DNA Chisel [17], OPTIMIZER [18], GeneOptimizer [19]. However, the main problems of these tools are limited integrated features, lack of optimization for multiple objectives, and poor stability for the algorithms.

To enables disease-relevant target screening, sequence design and optimization for RNA therapeutics, we have built TREAT platform. This platform fully incorporates 43,087,953 regulatory relationships between coding and noncoding genes from 81 biological networks under different physiological conditions. It introduces three ranking algorithms, degree centrality, PageRank, and Random Walk, to perform disease-relevant target screening. For the multi-objective optimization of large RNA, we stratified the RNA sequence features into the local and global features, to avoid inefficient combinations, thereby reducing the search space. In addition to directly optimizing large RNAs, it also supports designing their interfering RNAs. TREAT is a one-stop platform for the screening and design of therapeutic RNAs, with high quality biological networks integrated, cutting-edge algorithm embedded, friendly interface and exploratory visualization presented.

2. Methods

2.1. Specific regulatory networks of noncoding RNA

The regulatory networks covering noncoding RNAs were collected from ncFANs 2.0 we built previously, which is a functional annotation platform for noncoding genes. We built the noncoding-coding coexpression networks based on the RNA expression data of 50 normal tissues from GTEx [20] and 30 cancer types from TCGA. R package WGCNA [21] was utilized to calculate the Spearman correlation coefficients (Rho), Fisher's asymptotic P-value, and topological overlap measure (TOM). TOM is an approach to measure how close pairs of nodes are in a network [22]. To filter low confidence relationships, RNA pairs were screened with cutoff of adjusted P-value (FDR correction) < 0.05, correlation coefficient > 0.9 and TOM = 0.01.

2.2. Integration of coding and noncoding networks

To completely cover the intracellular molecular relationships under different physiological conditions, we collected PPI from Huri [23] and String database [24], and TF regulatory relationships from GRNdb [25]. These networks are integrated with the above noncoding networks in 2.1 under the same physiological conditions. The original PPI interactions of Huri were retrieved and screened under different tissue conditions according to the tissue-specific score > 0.9. Relationships from GRNdb are retained with confidence = high. Since Huri does not provide cancer-specific relationships, we curated its generic relationships instead. RNAs or proteins with the same Ensembl gene ID, are merged into one node in the integrated network. The relations between two nodes are established if connections exist in at least one of the original networks. Finally, networks for 50 normal tissues and 30 cancer types were constructed.

We also constructed one generic network as an alternative for other pathophysiological studies. This network is integrated from several databases, with the experimental data being pulled and curated. Coding gene-related relationships are from HURI [23], Bioplex [26], PhosphoSitePlus [27], BioGRID [28], InnateDB [29], MINT [30], INSIDER [31], and DIP [32]. TF regulatory relationships are from TRRUST [33], KnockTF [34] and IntACT [35]. log2FC > 1 and p value < 0.01 were utilized to filter KnockTF data, and TF data from BioID and AP-MS experiments in IntACT were retained. miRNA related relationships are from StarBase [36] and miRTarBase [37]. We use the Ensembl database [38] to uniformly convert the IDs in all databases into ensembl IDs.

2.3. Random Walk Diffusions algorithm for target screening

In addition to the degree and PageRank algorithms commonly used in network analysis [14,39], TREAT introduced the Random Walk Diffusions algorithm [40]. This algorithm simulates a walker, starting from the customized differential expressed genes, and randomly moving to adjacent neighbors, or returning to the initial nodes to restart. The algorithm can be formally described as follows:

$$P^{t+1} = (1-\lambda)MP^t + \lambda P^0$$

Where P^0 is the initial probability vector for all the nodes, and P^t is probability vector after t steps of iterations. M is the transfer matrix, normalized by the biological network's adjacency matrix. λ represents the probability of moving to the next node, while $1-\lambda$ represents the probability of restarting the walk from the initial nodes.

When the difference between P^t and P^{t+1} is <1e-6, the random walk is regarded as convergence.

2.4. Multi-objective sequence optimization of large RNA

Eleven features concerning RNA composition, immunogenicity, stability, translation and degradation [41] were collected in TREAT. They were divided into local and global features based on the RNA contextual properties. Firstly, scores summarized from all the local features are calculated at each bin, and taken as a measure to avoid inefficient combinations, thereby reducing the search space. Secondly, the sequences of two adjacent bins would be combined, and the next round of scoring, sorting, and screening would be performed. Finally, the complete sequence combined with all bins will be scored on all the local and global features.

2.5. Ensemble scoring and boosting framework

The optimization of multiple objectives relies on principles defined by the multiple features. However, conflicts may exist between different features. In order to effectively integrate different principles, we calculate an ensemble score for each sequence to evaluate the deviation from all the principles. This score ranges between 0 and 1, with 0 representing perfect compliance and 1 representing the divergence. This ensemble score was calculated by summing the boosting score of each single feature, using the following formula:

$$ES = \frac{\sum_{i=1}^{n} B_i * DS_i}{10 * n}$$

Where ES is the ensemble score, B_i is the boosting value for feature i (a user-customized integer between 1 and 10), DS_i is the divergence score for feature i, and n is the number of features the user customed to conduct the analysis.

3. Results

3.1. Framework of TREAT

3.1.1. Components

TREAT is designed for the screening, design and optimization of therapeutic RNAs. It consists of two computational components of "target screening" and "drug design and optimization" (Fig. 1). The "target screening" component focuses on identifying and ranking disease-relevant targets from existing biomedical data sources. It utilizes multiple biological networks under different physiological states and three different algorithms to rank the importance of targets for the concerned disease. In order to facilitate understanding and screening of the candidate targets, TREAT also collected gene expression profiles of 50 normal tissues and 30 cancer tissues in GTEx and TCGA databases. The "drug design and optimization" component integrates multiple features on RNA composition, immunogenicity, stability, translation and degradation, and proposes a hierarchical multi-objective optimization strategy for the input RNA sequence. Design and optimization of large RNAs or the interfering RNAs are both available.

3.1.2. Implementation

TREAT is implemented in Python, with the database stored in MySQL. The web application is built using the Flask microframework, with REST API embedded to send and receive data. The curated heterogeneous networks under different physiological conditions for disease-relevant targets screening are available at the download page. Three algorithms for target evaluation are introduced from third party tools, including degree centrality, PageRank and Random Walk. The TREAT web server is hosted on an elastic cloud server from the Ali cloud, running on a CentOS Linux system with 4 CPU and 16 GB of memory. Analyzing tasks submitted by users are scheduled by Python package multiprocessing. The user could asynchronously retrieve their result using a task id.

3.2. Disease-relevant target screening

3.2.1. Integrated tissue-specific and generic networks

We curated intracellular regulatory and interactive relationships between coding and noncoding genes from multiple databases. These relationships mainly include RNA-RNA co-expression, Protein-Protein interactions, and TF regulations. Through filtering, unifying, and merging the nodes and edges in different networks (see methods section), we finally construct 80 specific networks (Fig. 2) from 50 normal tissues and 30 cancer types, and one generic network (Table S1). The generic network play a complementary role to the specific networks. These networks contain a total of 43,087,953 coding-noncoding relationships. IDs and gene names from different databases were uniformly converted to Ensembl IDs and gene symbols.

3.2.2. Target screening and mechanisms characterization

Target screening begins with identifying a possible therapeutic target, and follows by characterization of the molecular mechanisms addressed by this target (Fig. 3). An acceptable target should be efficacious, safe, meet clinical requirements and be "druggable". In order to better utilize the above-integrated network for discovering disease-relevant targets, we integrated degree centrality, PageRank and Random Walk algorithms for assessing the target importance (Fig. 1). A ranked list would be generated through the target screening analysis. Moreover, to facilitate understanding and screening of the candidate targets, we collected gene expression profiles of 50 normal tissues and 30 cancer tissues in GTEx and TCGA databases, which could be visualized by clicking on 'Normal Tissue' and 'Cancer Tissue' at the concerned gene (Fig. 3). Functional enrichment of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) is performed on the target ranking result to characterize the possible molecular mechanisms.

3.3. Drug design and optimization

3.3.1. Large RNA design and optimization

For the therapeutic RNAs with large sizes, such as mRNAs and lncRNAs, their anionic charge, susceptibility to RNases and special structures, make it difficult for therapeutic RNAs to enter cells efficiently and function on their own [4]. Thus, the candidate RNA needs to undergo sequence optimization and modification to enhance RNA stability and reduce immunogenicity [16]. TREAT supports the design and optimization of large RNAs, regarding their sequence composition, immunogenicity, stability, translation and degradation. It builds a framework that can integrate more features, boost specific features, and give a deterministic output in a controlled space of search (Table 1).

TREAT collects 11 influential features, such as GC content [42], secondary structure [43], codon usage [44], bicodon usage [45], Codon Adaptation Index (CAI) [46], and degradation score [47] (Fig. 4B). These features are stratified into local and global features. The entire sequence is first optimized using local features to generate different optimized candidates, and then finally assessed on the global features (Fig. 4A). The local bins are divided by the fixed-length or dynamic-length stragedy, whereas the latter is inspired by AI language models of the protein sequence [48].

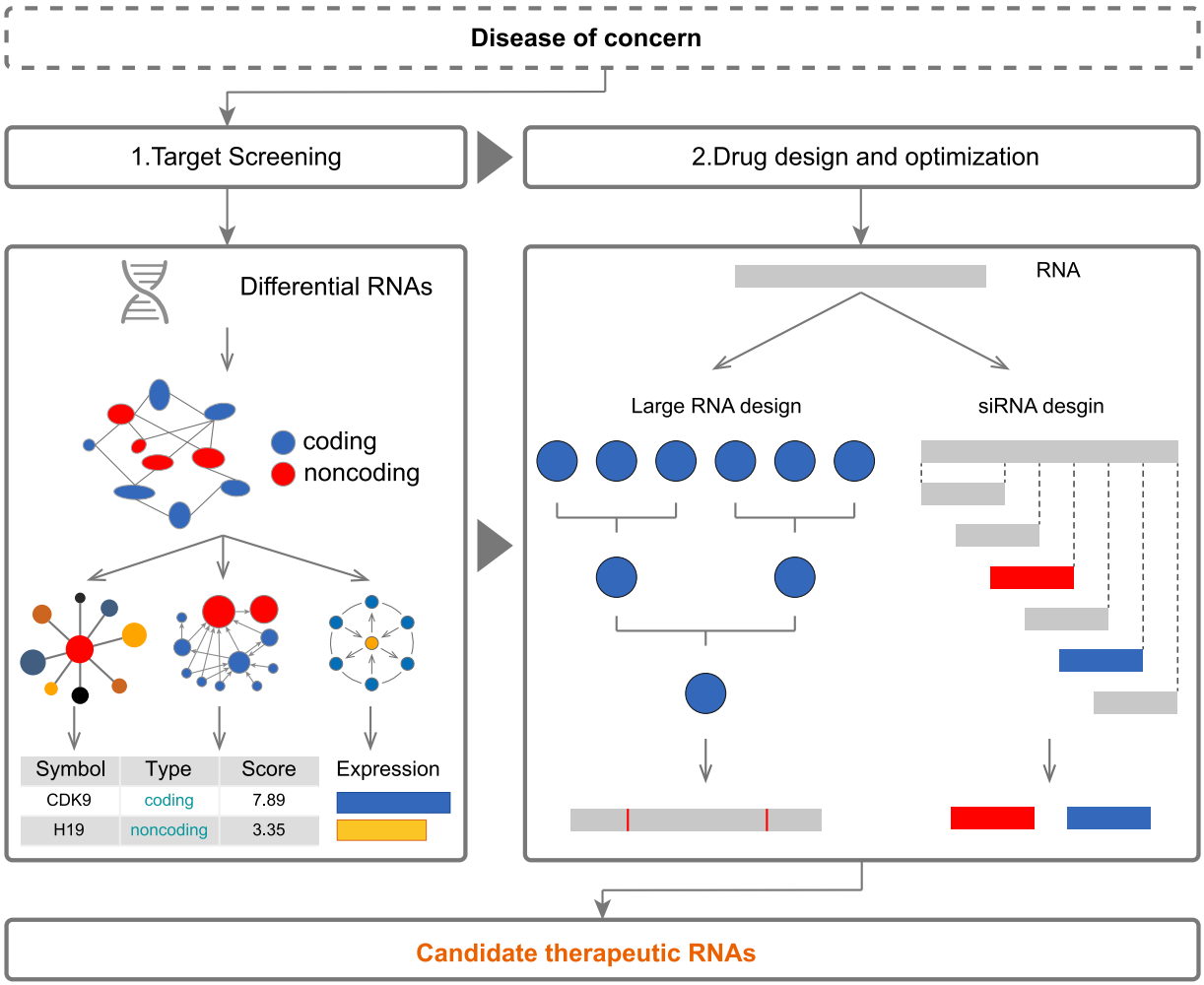

Fig. 1. Framework of TREAT. TREAT consists of two computational components: "target screening" and "drug design and optimization". The "target screening" component focuses on identifying and ranking disease-relevant targets. The "drug design and optimization" component focuses on the design and optimization of large RNAs and siRNAs. The grey bar represents an RNA sequence. The blue circle on the left represents the local bins, while the red lines below represent the modified bases on the sequence. The red and blue rectangular on the right represent candidate siRNAs with different scores. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3.2. RNA degradation characteristics

To protect the RNA from degradation and thus maximize the effectiveness of therapeutic RNAs, we included several features concerning RNA degradation to conduct large RNA design and optimization. A global degradation score was calculated to help assess an RNA sequence's degradation risk, enabling RNA sequence optimization for enhanced in-solution stability [47]. In addition to the global degradation score, a base score to evaluate degradation risk at each base was also provided. Since ψ modification provides a general technology that can be applied to stabilize specific nucleotides [49], this single base score is beneficial in guiding the posting nucleoside modifications. Scientists may generate their preferred candidates through some ways of engineering technology.

3.3.3. Specific features boosting

Users may have different concerns during large RNA design and optimization. In order to better integrate multiple features, we define a boosting weight for each feature. It could balance or exaggerate specific features when we sum all the feature scores together. By manipulating the web service, users can customize their weights of different features to try different directions of optimization. This helps generate diverse sequence candidates with different characteristics. For example, when the boosting value of the CAI feature is set to 5 times than default, the constraints of the CAI Principe will be heightened so that the sequence will be optimized to have more translational efficacy.

3.3.4. Small interfering RNAs design

One class of RNA therapeutics requires the delivery of small RNA molecules, which can reduce gene expression via RNA-induced silencing complex (RISC)-mediated mRNA degradation. For small interfering RNAs design, the input RNA sequence is traversed using a sliding window of 21nt, generating N-23 candidate sequences, where N is the length of input RNA. Guide sequence and passenger sequence are obtained for each sequence according to the principle of complementary base pairing. The 2–8 bases at the 5′ end are regarded as the seed region. Each sequence is filtered according to user-defined rules, such as TM score [50], GC content [51], and specific sequence structures. Further, these sequences are Blast against human RefSeq genes [52] to identify possible off-targets.

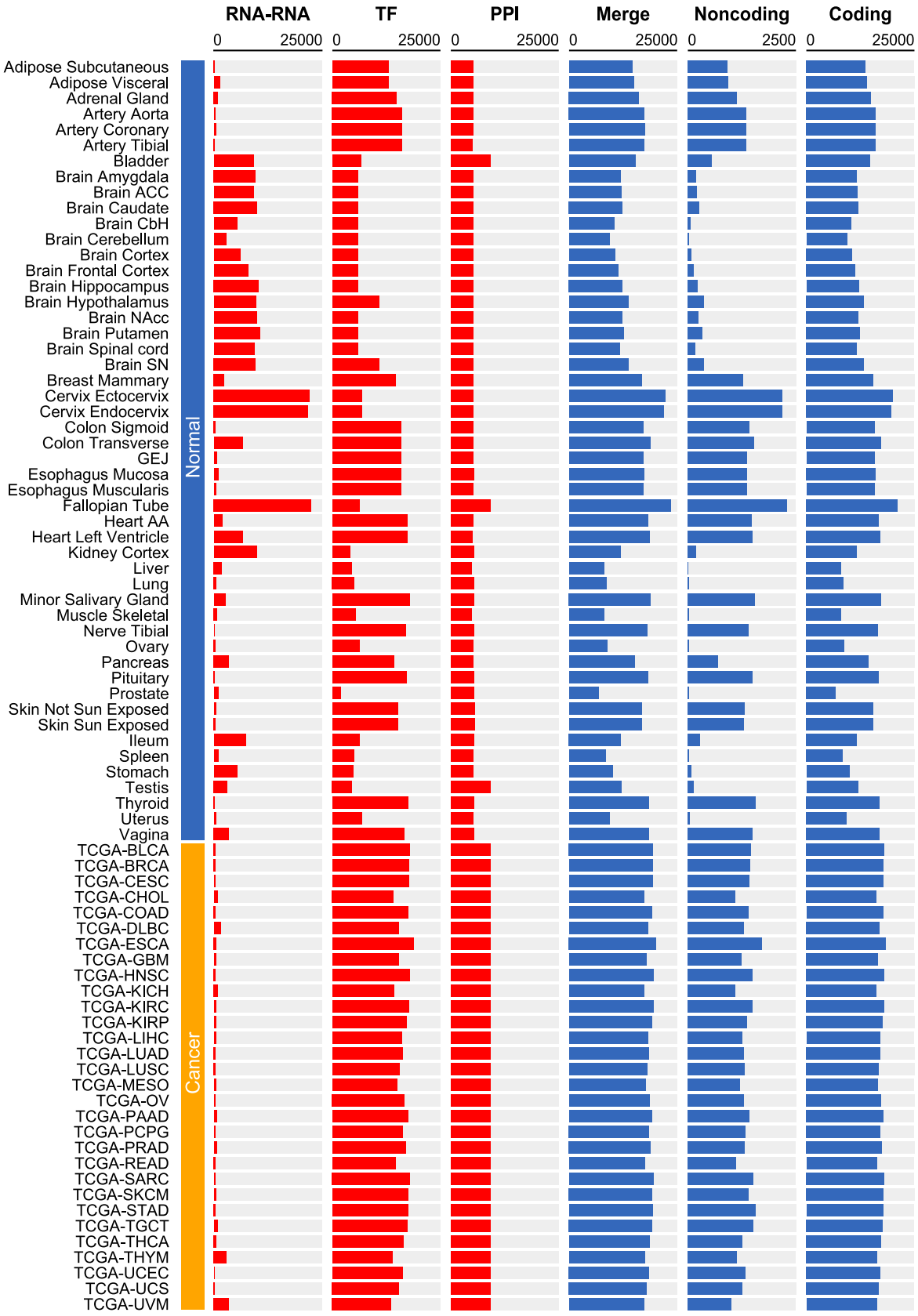

Fig. 2. Integrated tissue-specific networks. TREAT integrates 80 tissue-specific biological networks, including 50 normal tissue-specific networks and 30 cancer tissue-specific networks. The red bar length was proportional to the number of edges in the network, while the blue bar represents the number of nodes. Three different types of relationships are curated: 1.RNA-RNA: relationships between RNAs; 2.TF: transcript factor regulation relationships; 3.PPI: protein-protein interactions'. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

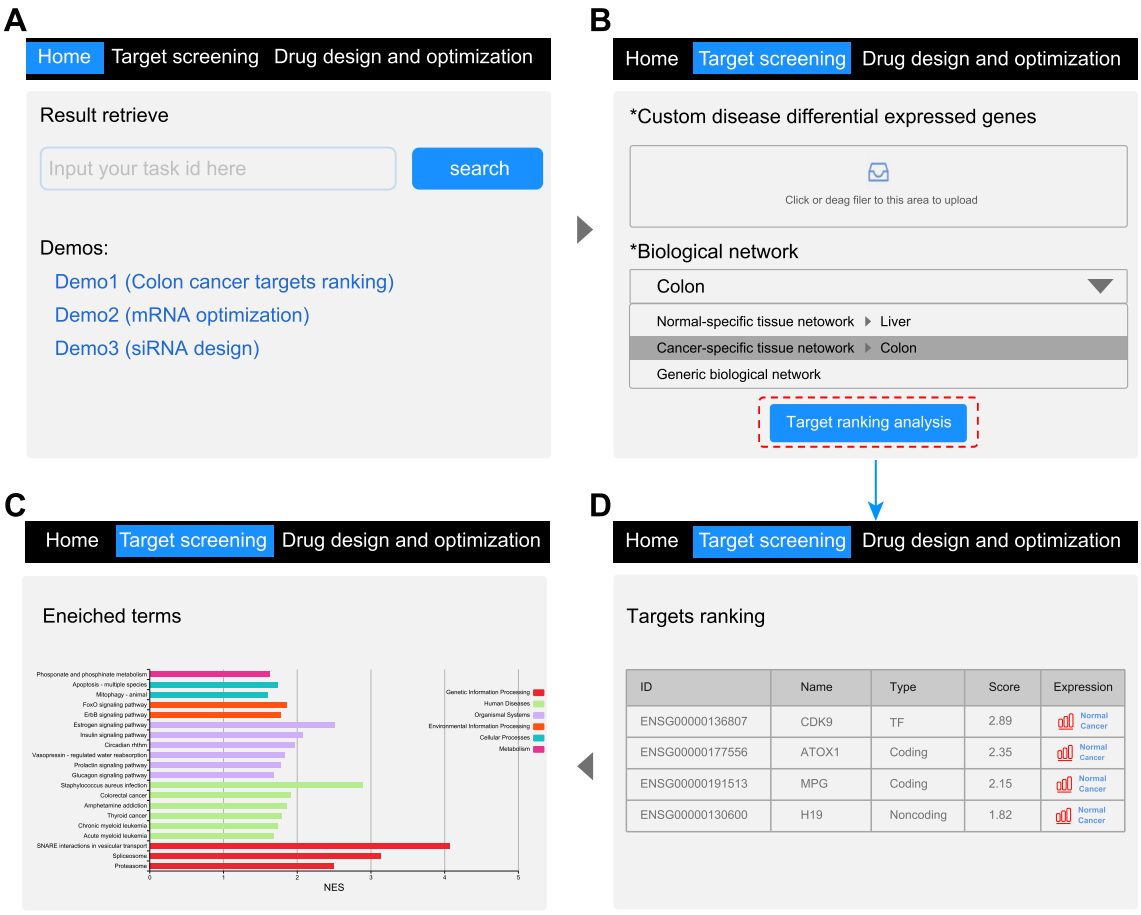

Fig. 3. Steps for target screening analysis. A.Homepage of TREAT. B.Input the differential expressed genes and biological network to start target ranking analysis. C. Targets ranking result. D. Enriched terms for the ranking result.

Table 1
Comparison with other platforms or tools.

Study	Tool	Features	Feature Boosting	Optimization algorithm	Output determinism	PMID
Ours	TREAT	10 + features	Yes	Ensemble score with feature stratification	deterministic	–
He Zhang, et al., 2020 [16]	LinearDesign	MFE, CAI	No	Dynamic Programming	deterministic	–
Valentin Zulkower et al, 2022 [17]	DNA Chisel	10 + features	Yes	Stochastic search, Exhaustive search	Non-deterministic	32647895
Pere Puigbò et al., 2007 [18]	OPTIMIZER	Codon Usage	No	Monte Carlo algorithm	deterministic	17439967
David Raab et al., 2010 [19]	GeneOptimizer	Codon Usage, DNA motifs, GC content, Repetitions, Homologies	No	Ensemble score	deterministic	21189842

3.4. Interface and visualization

3.4.1. Input and output

TREAT is a one-stop portal to systematically select and optimize therapeutic RNAs, working in a powerful and user-friendly interface. Firstly, the user could start an analysis from a concerning disease using the "target screening" component, with the differential genes customized as the prior knowledge. Then, the user choose a biology network associated with this disease to perform target ranking using three embedded algorithms (Fig. 3). The specific expression of genes in different tissues or cancers is available to assess the importance of a target. Secondly, the determined RNA could be modified for further optimization. Directly optimizing the RNA or designing its interfering RNA are both available at the "drug design and optimization" component. It takes one RNA sequence as the input, and generate multiple candidate therapeutic large RNAs or siRNAs (Fig. 5).

3.4.2. Exploratory visualization for sequence optimization

Exploratory visualization is a heuristic approach to knowledge discovery. To help explore in different directions across the huge sequence traversal space when performing optimization, TREAT visualized the RNA and associated contextual information in an interactive graph, allowing flexible and interactive editing of the target sequence (Fig. 5B). Alternative choices are provided when clicking on a nucleotide. Evaluations of different features and a

B

Feature	Type	Coding / noncoding	Local / Global	Scoring
Restriction site	Composition	Both	Local	Y
Avoid motif	Immunogenicity	Both	Local	Y
GC content	Stablity	Both	Local	Y
MFE	Stablity	Both	Global	N
Secondary structure	Stablity	Both	Global	N
Codon usage	Translation	Coding	Local	N
Bicodon usage	Translation	Coding	Local	Y
CAI	Translation	Coding	Local	Y
Half life	Degradation	Both	Global	N
Degradation score	Degradation	Both	Global	Y
Degradation base score	Degradation	Both	Local	N

Fig. 4. Multi-objective optimization Framework. A. Hierarchical multi-objective optimization framework for large RNA. B. Features integrated in the large RNA optimization.

Fig. 5. Steps for RNA drug design and optimization. A. Input for the large RNA design and optimization. B.Output for the large RNA design and optimization. C. Input for the siRNA design and optimization. D.Output for the siRNA design and optimization.

new ensemble score would be calculated for the new sequence in real-time. Facilitated by the exploratory visualization, the user could edit and continuously optimize an RNA sequence until founding the preferred one.

4. Discussion

Accompanied by RNA generation, purification, and delivery technology breakthroughs, RNA-based therapeutics are developed rapidly, and the RNA drug approval rate has also been recently accelerated. RNA therapeutics comprise a rapidly expanding category of drugs, emerging as suitable treatment options for many challenging diseases. However, although mRNA-based drugs and vaccines have been developed and broadly used, these RNAs only involve 2 %-3% of the human genome. The vast noncoding region, which accounts for 97 % of the human genome, has not been included in the scope of current RNA drug development. To unlock the therapeutic potential of noncoding RNAs, we have built TREAT, which enables us to identify, and selectively disrupt disease-relevant noncoding RNAs. The target screening and drug design engine embedded in TREAT could help scientists rapidly expand their pipelines of RNA therapeutics in multiple disease areas.

The first step in discovering a drug is identifying the molecular origin of a disease and the potential targets for intervention. With the development of omics technology, target identification is mainly based on principles of molecular biology. Scientists usually integrate available molecular networks based on muti-omics data to build and rank target-disease associations. TREAT fully incorporates 81 heterogeneous regulatory networks covering noncoding RNA and three advanced algorithms to perform target ranking. Despite the huge number of noncoding genes in the human genome, the number of those integrated into TREAT is still incomparable with coding genes. We will continuously update the scope of the noncoding networks and introduce more cutting-edge methods, such as artificial intelligence algorithms.

To overcome the high complexity of sequence optimization for large RNA, we stratified the features into local features and global features. Local features are calculated on each bin, while global features are calculated only on the final refined candidates, thus reducing the search space. Moreover, the ensemble score summarized could better balance different features and resolve conflicts between different features. However, this computational strategy still has a certain probability of missing some candidates, and its computational complexity still needs improvement. We would try more computing technology such as indexing and linear programming to improve the algorithm further. Correspondingly, we would continue collecting features at multiple omics levels for RNA sequence optimization and siRNA design.

The downstream RNA engineering technologies, even the tissues or species to be investigated, may have different preferences for RNA sequence characteristics. So it is difficult to define a universal strategy for ranking candidate RNA sequences. To be compatible with the optimization requirements of different platforms, TREAT integrates multiple sequence features and supports customized adjusting of their weights. An interface for interactive editing and real-time evaluation of the concerned RNA sequence is also available. These customization and exploratory visualization make the platform more flexible. In the future, we

will conduct more web service development on the interactive editing of therapeutic RNAs, to best inspire users' innovation in conducting the design and optimization. Moreover, to improve prediction performance on specific experimental platforms, such as mRNA vaccines for human, TREAT are going to integrate more testing data from several common experimental platforms and optimize TREAT's design parameters and algorithms based on these benchmark data. It is expected TREAT will become more flexible, automated and intelligent in the future.

5. Conclusions

In this study, we propose a platform TREAT for the screening, design and optimization of therapeutic RNAs, with particular attention to noncoding RNAs. It consists of two component of "target selection" and "drug design and optimization", with heterogeneous biological networks, powerful algorithms embedded, and a friendly interface for exploratory search. We believe that TREAT would be a great source and intelligent tool for RNA therapeutics and provide essential clues toward novel drug innovation for challenging diseases.

CRediT authorship contribution statement

Yufan Luo: Methodology, Software, Writing – original draft. **Liu Liu:** Validation, Data curation. **Zihao He:** Data curation, Writing – original draft, Methodology, Validation. **Shan Zhang:** Investigation. **Peipei Huo:** Visualization, Software. **Zhihao Wang Methodolog:** Software. **Qin Jiaxin:** . **Lianhe Zhao:** Software, Formal analysis. **Yang Wu:** Investigation. **Dongdong Zhang:** Investigation. **Dechao Bu:** Methodology, Software, Conceptualization. **Runsheng Chen:** Conceptualization. **Yi Zhao:** Conceptualization, Methodology, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (No: ZYYCXTD-C-202006); Zhejiang Provincial Natural Science Foundation of China (LY20C060001); National Natural Science Foundation of China (32070670); Innovation Fund of Institute of Computing and Technology, CAS (E161080).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.10.011.

References

[1] Santos R et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov 2017;16(1):19–34.
[2] Sahin U, Karikó K, Türeci Ö. mRNA-based therapeutics–developing a new class of drugs. Nat Rev Drug Discov 2014;13(10):759–80.
[3] Damase TR et al. The Limitless Future of RNA Therapeutics. Front Bioeng Biotechnol 2021;9.
[4] Paunovska K, Loughrey D, Dahlman JE. Drug delivery systems for RNA therapeutics. Nat Rev Genet 2022;23(5):265–80.
[5] Kulkarni JA et al. The current landscape of nucleic acid therapeutics. Nat Nanotechnol 2021;16(6):630–43.
[6] Adams D et al. Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. N Engl J Med 2018;379(1):11–21.
[7] Chaudhary N, Weissman D, Whitehead KA. mRNA vaccines for infectious diseases: principles, delivery and clinical translation. Nat Rev Drug Discov 2021;20(11):817–38.
[8] Kim YK. RNA therapy: rich history, various applications and unlimited future prospects. Exp Mol Med 2022;54(4):455–65.
[9] Paananen J, Fortino V. An omics perspective on drug target discovery platforms. Brief Bioinform 2020;21(6):1937–53.
[10] Yu AM, Choi YH, Tu MJ. RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. Pharmacol Rev 2020;72(4):862–98.
[11] Winkle M et al. Noncoding RNA therapeutics – Challenges and potential solutions. Nat Rev Drug Discov 2021;20(8):629–51.
[12] Bodein A et al. Interpretation of network-based integration from multi-omics longitudinal data. Nucleic Acids Res 2022;50(5):e27.
[13] Ochoa D et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. Nucleic Acids Res 2021;49(D1):D1302. D1310.
[14] Zhou G et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res 2019;47(W1):W234. W241.
[15] Zhao L et al. NONCODEV6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. Nucleic Acids Res 2021;49(D1): D165. D171.
[16] Zhang H, LinearDesign,, et al. Efficient Algorithms for Optimized mRNA. Sequence Design 2022:arXiv.
[17] Zulkower V, Rosser S. DNA Chisel, a versatile sequence optimizer. Bioinformatics 2020;36(16):4508–9.
[18] Puigbò, P., et al., *OPTIMIZER: a web server for optimizing the codon usage of DNA sequences.* Nucleic Acids Res. 2007. **35**(Web Server issue). W126-31.
[19] Raab D et al. The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. Syst Synth Biol 2010;4(3):215–25.
[20] . Science 2020;369(6509):1318–30.
[21] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 2008;9:559.
[22] Ravasz E et al. Hierarchical organization of modularity in metabolic networks. Science 2002;297(5586):1551–5.
[23] Luck K et al. A reference map of the human binary protein interactome. Nature 2020;580(7803):402–8.
[24] Szklarczyk D et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47(D1):D607. D613.
[25] Fang L et al. GRNdb: decoding the gene regulatory networks in diverse human and mouse conditions. Nucleic Acids Res 2021;49(D1):D97–d103.
[26] Huttlin EL et al. The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell 2015;162(2):425–40.
[27] Hornbeck PV et al. 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. Nucleic Acids Res 2019;47(D1):D433. D441.
[28] Oughtred R et al. The BioGRID interaction database: 2019 update. Nucleic Acids Res 2019;47(D1):D529. D541.
[29] Breuer K et al. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res 2013;41(Database issue):D1228. D1233.
[30] Licata L et al. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res 2012;40(Database issue):D857. D861.
[31] Meyer MJ et al. Interactome INSIDER: a structural interactome browser for genomic studies. Nat Methods 2018;15(2):107–14.
[32] Salwinski L et al. The Database of Interacting Proteins: 2004 update. Nucleic Acids Res 2004;32(Database issue):D449. D451.
[33] Han H et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 2018;46(D1):D380. D386.
[34] Feng C et al. KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. Nucleic Acids Res 2020;48(D1):D93–d100.
[35] Göös H et al. Human transcription factor protein interaction networks. Nat Commun 2022;13(1):766.
[36] Li JH et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 2014;42(Database issue):D92. D97.
[37] Huang HY et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res 2020;48(D1):D148. D154.
[38] Howe KL et al. Ensembl 2021. Nucleic Acids Res 2021;49(D1):D884. D891.
[39] Bu D et al. FangNet: Mining herb hidden knowledge from TCM clinical effective formulas using structure network algorithm. Comput Struct Biotechnol J 2021;19:62–71.
[40] Ruiz C, Zitnik M, Leskovec J. Identification of disease treatment mechanisms through the multiscale interactome. Nat Commun 2021;12(1):1796.
[41] Leppek K et al. Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. Nat Commun 2022;13(1):1536.
[42] Courel M et al. GC content shapes mRNA storage and decay in human cells. Elife 2019;8.
[43] Mauger DM et al. mRNA structure regulates protein expression through changes in functional half-life. Proc Natl Acad Sci U S A 2019;116 (48):24075–83.

[44] Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol 2018;19(1):20–30.
[45] Gamble CE et al. Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. Cell 2016;166(3):679–90.
[46] Sharp PM, Li WH. The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1987;15(3):1281–95.
[47] Leppek, K., et al., *Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics.* bioRxiv. 2021.
[48] Lin, Z., et al., *Language models of protein sequences at the scale of evolution enable accurate structure prediction.* 2022.
[49] Karikó K et al. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity 2005;23(2):165–75.
[50] Ui-Tei K et al. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. Nucleic Acids Res 2004;32 (3):936–48.
[51] Safari F, Rahmani Barouji S, Tamaddon AM. *Strategies for Improving siRNA-Induced Gene Silencing Efficiency*. Adv. Pharm Bull 2017;7(4):603–9.
[52] Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 2007;35(Database issue):D61. D65.