Chen MX, Zhu XD, Zhang H *et al.* SMRI: A new method for siRNA design for COVID-19 therapy. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 37(4): 991–1002 July 2022. DOI 10.1007/s11390-021-0826-x

SMRI: A New Method for siRNA Design for COVID-19 Therapy

Meng-Xin Chen[1] (陈梦欣), *Student Member*, *CCF*, Xiao-Dong Zhu[2] (朱晓冬), Hao Zhang[3] (张　浩) Zhen Liu[3,4] (刘　震), and Yuan-Ning Liu[3,*] (刘元宁), *Member*, *CCF*

[1] *College of Software, Jilin University, Changchun 130012, China*

[2] *Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education, Jilin University Changchun 130012, China*

[3] *College of Computer Science and Technology, Jilin University, Changchun 130012, China*

[4] *Graduate School of Engineering, Nagasaki Institute of Applied Science, Nagasaki 851-0193, Japan*

E-mail: chenmx19@mails.jlu.edu.cn; {zhuxd, zhangh}@jlu.edu.cn; liu_zhen@nias.ac.jp; liuyn@jlu.edu.cn

Received July 23, 2020; accepted August 31, 2021.

Abstract First discovered in Wuhan, China, SARS-CoV-2 is a highly pathogenic novel coronavirus, which rapidly spread globally and became a pandemic with no vaccine and limited distinctive clinical drugs available till March 13th, 2020. Ribonucleic Acid interference (RNAi) technology, a gene-silencing technology that targets mRNA, can cause damage to RNA viruses effectively. Here, we report a new efficient small interfering RNA (siRNA) design method named Simple Multiple Rules Intelligent Method (SMRI) to propose a new solution of the treatment of COVID-19. To be specific, this study proposes a new model named Base Preference and Thermodynamic Characteristic model (BPTC model) indicating the siRNA silencing efficiency and a new index named siRNA Extended Rules index (SER index) based on the BPTC model to screen high-efficiency siRNAs and filter out the siRNAs that are difficult to take effect or synthesize as a part of the SMRI method, which is more robust and efficient than the traditional statistical indicators under the same circumstances. Besides, to silence the spike protein of SARS-CoV-2 to invade cells, this study further puts forward the SMRI method to search candidate high-efficiency siRNAs on SARS-CoV-2's S gene. This study is one of the early studies applying RNAi therapy to the COVID-19 treatment. According to the analysis, the average value of predicted interference efficiency of the candidate siRNAs designed by the SMRI method is comparable to that of the mainstream siRNA design algorithms. Moreover, the SMRI method ensures that the designed siRNAs have more than three base mismatches with human genes, thus avoiding silencing normal human genes. This is not considered by other mainstream methods, thereby the five candidate high-efficiency siRNAs which are easy to take effect or synthesize and much safer for human body are obtained by our SMRI method, which provide a new safer, small dosage and long efficacy solution for the treatment of COVID-19.

Keywords COVID-19, SARS-CoV-2, RibonucleicAcid interference

1 Introduction

Coronaviruses are the positive-stranded RNA viruses with the largest RNA viral genomes ranging from 27 kb to 32 kb [1, 2]. The word of coronavirus is derived from its unique crown appearance, which may be attributed to its envelope with long petaloid spines [3]. Coronaviruses can infect various parts of the body, particularly the upper respiratory tract and gastrointestinal tract tissues, liver and the central nervous system (CNS) [2, 4]. For humans, coronaviruses mainly cause respiratory diseases, such as severe acute respiratory syndrome (SARS) and middle east respiratory syndrome (MERS) [2, 5, 6].

In December 2019, a new type of pneumonia was reported in Wuhan, Hubei Province, China [7–9]. Soon after, these patients were found with a novel coronavirus that was subsequently named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by

Regular Paper

The work was supported by Jilin Provincial Key Laboratory of New Biometrics Technology, the National Natural Science Foundation of China under Grant No. 61471181, and the Jilin Province Industrial Innovation Special Fund Project under Grant Nos. 2019C053-2 and 2019C053-6.

*Corresponding Author

the World Health Organization (WHO). Additionally, the new pneumonia was named as coronavirus disease 2019 (COVID-19). SARS-CoV-2 causes typical flu-like symptoms. As reported by a recent study, most patients have fever or cough, a third of patients develop shortness of breath, and other symptoms include diarrhoea, chest pain, confusion, headache, and muscle ache [10]. On March 11th, 2020, the WHO declared COVID-19 as a global pandemic. According to our survey, no vaccine has been developed to successfully prevent the COVID-19 virus infection till March 13th, 2020, and there are limited distinctive clinical drugs available.

RNA interference (RNAi), an endogenous gene expression regulation mechanism, was first found in plants to act as a natural immune mechanism to resist virus infections [11–16]. RNAi is widely employed by many creatures to resist virus infections, such as caenorhabditis elegans, arabidopsis, chlamydomonas, drosophila, filamentous fungi, plants and other organisms [17]. RNAi is mediated by small RNAs, consisting of small interfering RNAs (siRNAs) and microRNAs (miRNAs) [18]. siRNAs, the double-stranded RNA (dsRNA) sequences about 21nt–23nt in length, can down-regulate specific target genes with an accurate complementary sequence and is formed by the cleavage of dsRNA by the Dicer enzyme [19–21]. siRNAs can be used to resist all types of viral genomes, regardless of whether they are double-stranded or single-stranded DNA/RNA viruses. In addition, the use of multiple siRNAs simultaneously can achieve an enhanced antiviral effect [22]. RNAi has shown a potential application prospect in antiviral therapy [23–26].

This study proposes a simple multiple-rule intelligent method (referred to as the SMRI method) for the design of high-efficiency siRNAs for SARS-CoV-2. There are mainly two types of research at present: based on statistical methods and based on machine learning. Compared with the research based on statistical methods, the SMRI method enhances the ability of screening efficient siRNAs by incorporating machine learning methods. Compared with the research based on machine learning, the SMRI method can effectively filter out candidate siRNAs which are difficult to take effect or synthesize by integrating rules. At the same time, compared with the complex neural network model, the complexity of the SMRI method is lower. Different from previous studies [27–35] that only used statistical methods or machine learning methods, the SMRI method integrates the advantages of statistical methods and machine learning methods. Besides, the SMRI method ensures that the candidate siRNAs have a mismatch of over three bases in the Basic Local Alignment Search tool (BLAST) homology analysis of the human genome, so as to avoid silencing the human gene. This is not found in other mainstream methods, thereby the siRNAs designed by the SMRI method are much safer for human body. The advantage of the SMRI method is that it is highly efficient, safe to human and can effectively filter out the candidate siRNAs that are difficult to take effect or synthesize. Moreover, we innovatively propose a new index named siRNA Extended Rules index (SER index) based on integration rules to measure the silencing efficiency of candidate siRNAs. Its advantage lies in that it has stronger and more robust ability to screen high-efficiency siRNA than a single index. The main contributions of this study are as follows.

- This study proposes a new model named Base Preference And Thermodynamic Characteristic model (BPTC model) regarding the siRNA silencing efficiency.
- Based on the BPTC model, this study proposes a new efficient and robust index named SER index to extract the rules in methods summarizing the characteristic properties of high-efficiency siRNAs and to measure the silencing efficiency of candidate siRNAs.
- This study proposes a new efficient and safe siRNA design method named the SMRI method for SARS-CoV-2. The SER index is a part of the SMRI method. And the SMRI method uses the SER index to effectively filter out the candidate siRNAs that are difficult to take effect or synthesize. And five potential high-efficiency siRNAs for the treatment of COVID-19 are obtained by the SMRI method.

The rest of this paper is organized as follows. Section 2 summarizes the prediction algorithms of siRNA interference efficiency. Section 3 introduces the specific mechanism of the RNAi therapy in mammals and the factors related to the design of siRNAs with a high silencing efficiency. Section 4 gives the precise definition of the BPTC model, a new model regarding the siRNA silencing efficiency, the SER index, a new index to screen siRNAs with a high silencing efficiency, and the SMRI method to design siRNAs against SARS-CoV-2. Section 5 shows the data source, experimental process and results. Section 6 analyzes the screening ability of the SER index, the contribution of each step in the SMRI method, and the screening ability of the SMRI method, and describes the antiviral process and

advantages of our proposed method. Section 7 concludes the paper.

2 Related Work

The premise of the siRNA formation is to select gene substrings of target genes. Nonetheless, not every siRNA formed by an arbitrary substring of the target gene can induce target gene silencing [36]. Many studies have explored the related factors affecting the design of siRNAs with a high silencing efficiency. These studies are classified as follows.

2.1 Methods Summarizing the Characteristic Properties of High-Efficiency siRNAs

Elbashir *et al.* [27] proposed a set of guidelines aiming at designing the effective siRNAs. These principles include some characteristic attributes of siRNAs with a high silencing efficiency, like the selection of siRNAs with the C and G contents between 30% and 70%.

Reynolds *et al.* [28] identified eight features related to the siRNA silencing efficiency, such as the shortage of reverse repeats. Based on the above-mentioned characteristics, they proposed a principle that significantly improved the predicted silencing efficiency of siRNAs, commonly known as the Reynolds rule.

Amarzguioui and Prydz [29] identified various characteristics of the double-stranded siRNAs with a silencing efficiency of over 70%, like the steadiness of the siRNA ends. Combining these findings, they proposed an effective selective algorithm to design siRNAs targeting the new endogenous human genes.

Hsieh *et al.* [30] detected a significant nucleotide preference at specific positions in siRNAs with a silencing efficiency of over 70%, especially G or C at the 11th base position and C at the nineteenth base position.

Takasaki *et al.* [31] defined a priority score based on single nucleotide's position characteristics to measure gene degradation and developed an effective method to select the target sequence of siRNA.

Katoh and Suzuki [32] proposed that the RNAi activity was dramatically affected by the precise base in every three positions of its siRNAs. Specifically, when the $3n+1$ (4, 7, 10, 13, 16, 19) bases in siRNAs were the corresponding optimized bases, their activities would be enhanced. On account of the above-mentioned observations, they developed an algorithm ("siExplorer") to calculate the silencing efficiency of siRNAs targeting the endogenous human genes.

Researches aiming to summarize the characteristic properties of siRNAs with high silencing efficiency have proposed numerous good siRNAs screening algorithms, such as Elbashir *et al.* [27], Reynolds *et al.* [28], Amarzguioui and Prydz [29], Hsieh *et al.* [30], Takasaki *et al.* [31], and Katoh and Suzuki [32]. These algorithms are effective on the siRNA design to some extent. However, only limited shared properties of siRNAs with a high silencing efficiency are concluded from a single study. As a result, a large number of candidate siRNAs conform to these rules. Moreover, some ineffective or low silencing efficiency siRNAs also comply with these rules, limiting the accuracy of the siRNAs screening algorithm developed by this research.

In this study, inspired by integrated learning, we innovatively propose a new index SER to measure the silencing efficiency of candidate siRNAs based on the methods summarizing the characteristic properties of high-efficiency siRNAs proposed by Reynolds *et al.* [28], Amarzguioui and Prydz [29] and Hsieh *et al.* [30]. According to our analysis, the SER index performs better than any of the above-mentioned three methods [28–30] in screening siRNA. Moreover, the SMRI method further enhances the ability to screen high-efficiency siRNAs by combining with machine learning methods.

2.2 Methods Adopting Machine Learning Algorithms

Vert *et al.* [34] captured the essential factors related to siRNA silencing efficiency with Least Absolute Shrinkage and Selection Operator (LASSO) regression and proposed a linear model Designer of Small Interfering RNA (DSIR) to predict the siRNA silencing efficiency. Their analysis suggested that DSIR achieved a comparable accuracy to other mainstream algorithms. Besides, it had more advantages than other mainstream algorithms in explaining the siRNA efficiency mechanism.

Ichihara *et al.* [35] built a linear regression model on a dataset containing 2 431 siRNAs, and proposed an intuitive algorithm inhibitory-Score (i-Score) to predict siRNAs silencing efficiency. The algorithm only considers the nucleotide preference of each position. Moreover, based on a validation dataset containing 419 siRNAs, the i-Score attained a comparable prediction accuracy compared with other mainstream algorithms where more parameters were used.

Qureshi *et al.* [36] studied a new dataset containing 1 725 viral siRNAs with experimentally verified si-

lencing efficacy. A high-accuracy siRNAs silencing efficiency prediction model was constructed by considering various siRNA characteristic parameters, including binary pattern, thermodynamic properties and secondary structure mono to pentanucleotide frequencies.

Compared with methods aiming to summarize the characteristic properties of siRNAs with a high silencing efficiency, the prediction accuracy of researches modeling by the machine learning methods has a great improvement, such as Vert *et al.*[33], Ichihara *et al.*[34] and Qureshi *et al.*[35]. However, some complex machine learning models are challenging to assist in clarifying the mechanism of siRNA silencing efficiency. In addition, the related factors affecting the siRNA silencing efficiency selected by each model have great influence on the silencing efficiency of siRNAs designed by this model. Moreover, there are many candidate siRNAs that are difficult to take effect or synthesize in the screening results.

In this study, the SMRI method can effectively filter out the candidate siRNAs that are difficult to take effect or synthesize by using the SER index to screen siRNAs. The SER index is proposed in this study to extract the rules in methods summarizing the characteristic properties of high-efficiency siRNAs.

3 Backgroud

3.1 Mechanism of RNAi Therapy in Mammals

The RNAi process in mammals consists of two stages, including an initiation and an execution phase[37]. In the initial phase, a long dsRNA is introduced from an exogenous source or from a transgene or a virus. Then, the long dsRNA is specifically recognized and cleaved by the Dicer enzyme in an energy-consuming process into a double-stranded siRNA. The double-stranded siRNA contains a sense strand and an antisense strand, having the length of 21nt–23nt. The 3' ends of the above two strands contain two overhanging bases[37–39]. In the execution phase, the RNA-induced silencing complex (RISC), which is in charge of unwinding the double-stranded siRNA to form a sense strand and an antisense strand, is formed by the double-stranded siRNA, helicase, and other required enzymes in an energy-consuming way. Then, the RISC unwinds the double strands of siRNA. Afterwards, an endogenous mRNA with a complementary sequence containing the antisense strand is searched under the guidance of the antisense strand. Later, the target mRNA which is cleaved is broken down by the cellular exonucleases, resulting in a post-transcriptional gene silencing[37–39].

3.2 Factors of High-Efficiency siRNAs

Many factors are involved in the success of RNAi.

3.2.1 Secondary Structure of Target Sequence

Schubert *et al.*[40] proposed that the complexity of secondary structure in the target region is negatively correlated with siRNA silencing efficiency. The RNA region can be classified into the stem region, loop region, stem loop region and multi-branch loop region. It is concluded that the siRNA silencing effect is good in the stem region but poor in the multi-branch loop region.

3.2.2 Thermodynamic Stability of siRNA

When siRNA duplexes assemble into RISC, they are transformed into a single-stranded form[41]. The entry of the antisense strand into RISC plays a crucial role in the cleavage of target mRNA. Relatively, the unstable siRNAs antisense strands at 5' end enter RISC advantageously[36]. In addition, the G-C base pair is thermodynamically more stable than the A-U base pair.

3.2.3 Selection of Target Sequence

The base matching degree between siRNA and the target sequence based on the gene complementation principle is crucial to the gene silencing effect of siRNA[42]. Overall, the gene silencing effect of siRNA will be considerably reduced, even if there is only one different base in its antisense strand from the complementary base sequence of the target mRNA sequence.

4 BPTC Model, SER Index and SMRI Method

In this section, Subsection 4.1 introduces the BPTC model regarding the siRNA silencing efficiency. In Subsection 4.2, the SER index for screening siRNAs with high silencing efficiency is proposed based on the BPTC model. Subsection 4.3 illustrates the SMRI method to design siRNAs against SARS-CoV-2, which filters roughly invalid siRNAs by the SER index.

4.1 BPTC Model

In this study, a mathematical model named BPTC model is established to measure the silencing efficiency of siRNA. The factors considered are as follows: the

base preference of each base position in the sense strand and the free energy of the secondary structure in the sense strand. The specific model is as follows:

$$\boldsymbol{w} = \begin{pmatrix} Wa_1 & Wa_2 & \cdots & Wa_{19} \\ Wu_1 & Wu_2 & \cdots & Wu_{19} \\ Wc_1 & Wc_2 & \cdots & Wc_{19} \\ Wg_1 & Wg_2 & \cdots & Wg_{19} \end{pmatrix}. \tag{1}$$

$W_{a_i}, W_{u_i}, W_{c_i}$, and W_{g_i} $(i = 1, 2, 3, \ldots, 19)$ represent the base position weight values at the first to the 19th base positions in the sense strand of siRNA, respectively. W_{a_i} indicates the contribution to the whole siRNA silencing efficiency when the A nucleotide appears at the i-th position. Similarly, W_{u_i}, W_{c_i}, and W_{g_i} can be deduced in the same way.

$$\boldsymbol{e} = \begin{pmatrix} Ea_1 & Eu_1 & Ec_1 & Eg_1 \\ Ea_2 & Eu_2 & Ec_2 & Eg_2 \\ \vdots & \vdots & \vdots & \vdots \\ Ea_{19} & Eu_{19} & Ec_{19} & Eg_{19} \end{pmatrix}.$$

E_{a_i}, E_{u_i}, E_{c_i}, and E_{g_i} $(i = 1, 2, 3, \ldots, 19)$ represent the existence status of a corresponding type of base at the first to the 19th base positions in the sense strand of siRNA, respectively. 0 indicates that the base at the base position is not a corresponding type of bases, while 1 suggests that the base at the base position is a corresponding type of bases.

$$\boldsymbol{W} = \boldsymbol{w} \cdot \boldsymbol{e}.$$

$\boldsymbol{w}$ stands for the base position weight matrix and $\boldsymbol{e}$ denotes the base existence status matrix.

$$SED = \begin{cases} \sum_{i=1}^{4} W_{i,i} + 1, & \text{if } F \geqslant t, \\ \sum_{i=1}^{4} W_{i,i}, & \text{otherwise.} \end{cases}$$

F (kcal/mol) is indicative of the free energy of the secondary structure in the sense strand; t (kcal/mol) represents the threshold of the F's value.

4.2 Model Parameters of the SER Index

After analyzing the different empirical rules for the siRNA design, our present study finds that previous studies mainly summarized the properties of high-efficiency siRNAs from different aspects. However, there are many candidate siRNAs conforming to a single empirical rule, which are often not limited to the effective siRNAs [43].

Therefore, if different rules defined in previous studies are combined to form a new index, the selective advantages of different rules can be integrated to improve the screening efficiency of high-efficiency siRNAs.

By integrating the rules of the three methods proposed by Reynolds *et al.* [28], Amarzguioui and Prydz [29] and Hsieh *et al.* [30] and removing some of their common evaluation rules, a new index SER is proposed to efficiently screen siRNAs based on the above-mentioned ideas.

Here, based on our SER index implementation, the parameters of the BPTC model established in Subsection 3.1 are shown below:

$$t = -0.9, \quad \boldsymbol{w} = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 2 \\ -2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -2 \end{pmatrix}. \tag{2}$$

As shown in (2) with reference to (1) in Subsection 4.1, lines 1–4 in (2) represent the position weight values of A, U, C and G nucleotides. For example, the number in the first row indicates that if the A nucleotide appears at the 1st or 2nd position of the siRNA sequence, its contribution at this position to the silencing efficiency score of the siRNA is all −1. If the nucleotide at positions of the 6th, 13th, 17th or 18th of the siRNA sequence is A, every A at the above base positions will contribute 1 to the siRNA silencing efficiency score. If the A nucleotide appears at the 6th or 13th or 17th or 18th position of the siRNA sequence, its contribution at this position to the silencing efficiency of the siRNA is all 1. If the A nucleotide appears at the 19th position of the siRNA sequence, its contribution at this position to the silencing efficiency score of the siRNA is 2.

4.3 SMRI Method

Also, we propose a new siRNA design method, namely, the SMRI method. In the process of designing siRNAs, we need to screen a large number of candidate

siRNAs to get potential high-efficiency siRNAs. The flow chart of the SMRI method is shown in Fig.1.

As a matter of routine, there are 21 bases in the sense strand and the antisense strand of siRNAs designed by us, including 19 paired and two unpaired bases at the 3' end of each strand. The SMRI method is described in detail as follows.

Step 1. The BLAST analysis is performed on the target gene fragment of all viral strains to obtain the conservative sequences of the target gene fragment.

Step 2. According to the principle of gene complementation, 19 consecutive nucleotides are searched on the target gene fragment, and the corresponding siRNAs double strands are obtained.

Step 3. The candidate siRNAs whose sense strand with a G/C content is beyond the range of 36%–53% are excluded.

Step 4. The candidate siRNAs whose 5′ end free energy is higher than the 3′ end free energy of the sense strand are excluded.

Step 5. The values of SER proposed in Subsection 3.1 for all the candidate siRNAs are calculated. According to the SER value, all candidate siRNAs are sorted in descending order, and candidate siRNAs with the SER index value less than 4 are excluded.

Step 6. Candidate siRNAs are scored according to

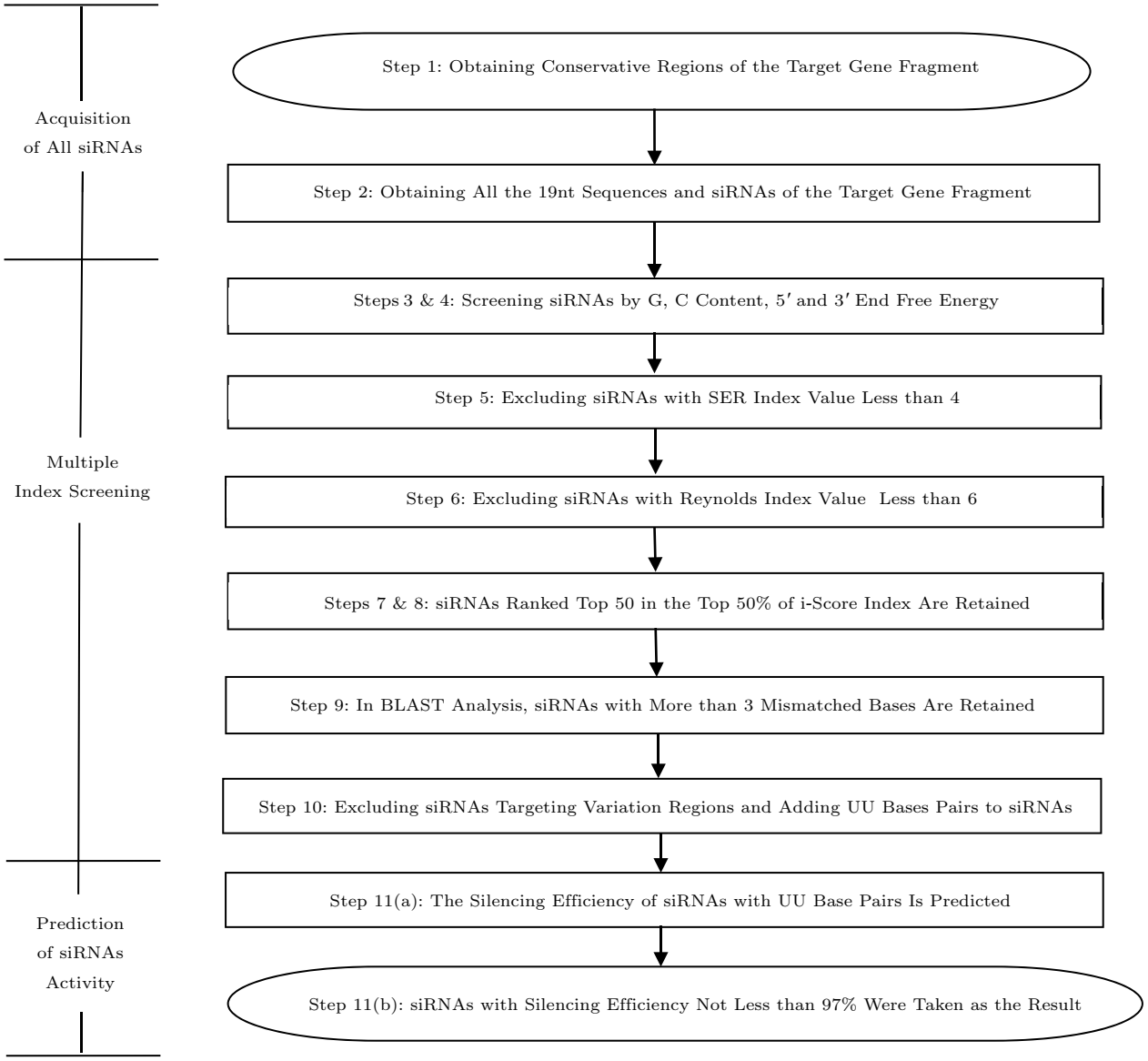


Fig.1. Flow chart of the SMRI method for screening siRNAs.

Reynolds *et al.*'s rule [28], and candidate siRNAs with Reynolds index value less than 6 are excluded.

Step 7. According to the i-Score method proposed by Ichihara *et al.* [34], the candidate siRNAs are scored and sorted in descending order according to the i-Score value, and then the top 50% siRNAs are screened out.

Step 8. If more than 50 siRNAs are selected in the above steps, we rank them in a descending order according to the i-Score value, and retain the top 50 siRNAs; otherwise, we retain all the candidate siRNAs.

Step 9. All the candidate siRNAs are confirmed to have more than three base mismatches in BLAST homology analysis of human genome to avoid silencing human genes.

Step 10. Candidate siRNAs whose target sequences are not in the conserved region of the target gene fragment are excluded. Meanwhile, overhanging UU base pairs are added to the 3' end of the sense strand and the antisense strand.

Step 11. The silencing efficiency of candidate siRNAs is predicted, and the candidate siRNAs with the predicted silencing efficiency equal to or greater than 97% are retained. In this study, 97% is used as the threshold to screen an appropriate number of high-efficiency siRNAs.

5 Experiment and Results

5.1 Data Sources

Altogether 98 samples① are obtained by searching the 2019 Novel Coronavirus Information Database (2019nCoVR) of National Genomics Data Center [44] to identify the whole-genome nucleotide sequences of SARS-CoV-2 till February 2020.

5.2 Selection of the Target Sequences

According to the latest research, Hoffmann *et al.* [45] proved that SARS-CoV-2 utilized the spike protein to bind to angiotensin-converting enzyme2 (ACE2), the SARS-CoV receptor [46], thus entering the host cells. The S gene sequence is a sequence that encodes the spike protein located near the 3' end of the whole genome of SARS-CoV-2. Considering that virus mutation at the site on the target sequence can cause RNAi failure, the target sequences are selected from the highly conserved regions in the S gene sequence of the SARS-CoV-2 genome.

5.3 Acquisition of Conserved Regions in S Gene of SARS-CoV-2

The S gene sequence starts at 21563 of the SARS-CoV-2 genome and ends at 25381 of the SARS-CoV-2 genome. According to the variation annotation of SARS-CoV-2 of 2019nCoVR from National Genomics Data Center till February 15th, 2020, the variation regions in the S gene of SARS-CoV-2's cDNA are exhibited in Table 1. The distribution of variation regions in S gene can be seen in the full version of Table 1②.

Table 1. Variation Regions in the S Gene of SARS-CoV-2's cDNA

Position in Genome	Original Bases of Variation Regions
21644	T
21656	T
21707	C
21949	A
22224	C
22303	T
22586	T
22622	A
22652	G
22661	G
23403	A
23569	T
23605	T
24034	C
24325	A
24947	G
24990	C
25060	A
25153	C
25347	A

5.4 Design of siRNAs

In this study, the target sequences of siRNAs are selected from the conserved regions in the S gene sequence of the SARS-CoV-2 genome, and the SMRI method described in Section 3 is employed to design siRNAs based on the SARS-CoV-2 sequence NC_045512.2 obtained from Genbank. When step 10 in SMIR method is finished, altogether 19 safe and potentially efficient siRNAs are obtained, as shown in Table 2.

In this study, the siRNA scales are adopted to predict the prediction efficiency of the obtained siRNAs ac-

①https://workdrive.zohopublic.com.cn/file/lk8pq67c30baa9b3b487f9e4e2f9616018633, June 2022.
②https://workdrive.zoho.com.cn/file/lk8pqd1108f616c8a449b9fe04ddadb038eae, June 2022.

Table 2. Candidate siRNAs Obtained After Step 10 in the SMRI Method

Number	Position in Genome	Target Sequence	Sense	Antisense	Predicted Silencing Efficiency (%)
1	21785	GGTACTAAGAGGTTTGATA	GGUACUAAGAGGUUUGAUAuu	UAUCAAACCUCUUAGUACCuu	95
2	22103	GGAAAACAGGGTAATTTCA	GGAAAACAGGGUAAUUUCAuu	UGAAAUUACCCUGUUUUCCuu	90
3	22180	GCACACGCCTATTAATTTA	GCACACGCCUAUUAAUUUAuu	UAAAUUAAUAGGCGUGUGCuu	96
4	22454	GAAACAAAGTGTACGTTGA	GAAACAAAGUGUACGUUGAuu	UCAACGUACACUUUGUUUCuu	95
5	22844	GATTTTACAGGCTGCGTTA	GAUUUUACAGGCUGCGUUAuu	UAACGCAGCCUGUAAAAUCuu	97
6	23128	AACTGTTTGTGGACCTAAA	AACUGUUUGUGGACCUAAAuu	UUUAGGUCCACAAACAGUUuu	95
7	23138	GGACCTAAAAAGTCTACTA	GGACCUAAAAAGUCUACUAuu	UAGUAGACUUUUUAGGUCCuu	95
8	23438	GCAGATCAACTTACTCCTA	GCAGAUCAACUUACUCCUAuu	UAGGAGUAAGUUGAUCUGCuu	96
9	23517	GGGCTGAACATGTCAACAA	GGGCUGAACAUGUCAACAAuu	UUGUUGACAUGUUCAGCCCuu	95
10	24230	GCAGGTGCTGCATTACAAA	GCAGGUGCUGCAUUACAAAuu	UUUGUAAUGCAGCACCUGCuu	95
11	24421	CCAAAATGCACAAGCTTTA	CCAAAAUGCACAAGCUUUAuu	UAAAGCUUGUGCAUUUUGGuu	98
12	24577	GACATATGTGACTCAACAA	GACAUAUGUGACUCAACAAuu	UUGUUGAGUCACAUAUGUCuu	94
13	24655	GTGTGTACTTGGACAATCA	GUGUGUACUUGGACAAUCAuu	UGAUUGUCCAAGUACACACuu	89
14	24863	CACTGGTTTGTAACACAAA	CACUGGUUUGUAACACAAAuu	UUUGUGUUACAAACCAGUGuu	97
15	25124	GAGGTTGCCAAGAATTTAA	GAGGUUGCCAAGAAUUUAAuu	UUAAAUUCUUGGCAACCUCuu	98
16	25172	GGAAAGTATGAGCAGTATA	GGAAAGUAUGAGCAGUAUAuu	UAUACUGCUCAUACUUUCCuu	94
17	25229	GGCTTGATTGCCATAGTAA	GGCUUGAUUGCCAUAGUAAuu	UUACUAUGGCAAUCAAGCCuu	97
18	25230	GCTTGATTGCCATAGTAAT	GCUUGAUUGCCAUAGUAAUuu	AUUACUAUGGCAAUCAAGCuu	92
19	25239	CCATAGTAATGGTGACAAT	CCAUAGUAAUGGUGACAAUuu	AUUGUCACCAUUACUAUGGuu	94

cording to the method proposed by Matveeva *et al.* in 2007 [47]. The input parameters of the method include partial duplex stability, nucleotide position dependence preference and total G/C content of the siRNAs duplex. Besides, linear regression fitting is adopted to achieve a faster calculation speed. It is found that siRNA scales achieved the same performance in identifying the efficient and inefficient siRNAs as that of the best method identified at the moment [47]. Moreover, the siRNA scales [47] are utilized to predict the silencing efficiency of 19 candidate siRNAs obtained from the BLAST analysis, as presented in Table 2.

According to Table 2, the prediction silencing efficiency of almost all the screened candidate siRNAs is greater than 90%, indicating that our SMRI method is quite effective. The candidate siRNAs with the predicted silencing efficiency greater than or equal to 97% are finally selected as the designed siRNAs for SARS-CoV-2, as shown in Table 3. In previous studies, it is generally considered that siRNAs with silencing efficiency larger than 70% are high-efficiency siRNAs. Therefore, these five siRNAs are mostly high-efficiency. However, the actual silencing efficiency of these five siRNAs needs to be synthesized and tested experimentally.

6 Experimental Analysis

In this subsection, we analyze the performance of the SER index to screen siRNAs with high silencing efficiency in Subsection 6.1. And we explore the rationality and the performance to screen siRNAs with high silencing efficiency of the SMRI method in Subsection 6.2 and Subsection 6.3 respectively. Last but not least, we introduce the antiviral process and advantages of the SMRI method in Subsection 6.4.

6.1 SER Index Screening Performance Analysis

As shown in Fig.2, we screen siRNAs with four different indexes based on 3 802 21nt siRNAs on the S gene sequence of SARS-CoV-2, and then predict the silencing efficiency of the screened siRNAs with the siRNA scales [47]. Thereafter, the average value of the predicted silencing efficiency of siRNAs is used as the standard to measure the screening effect of each index. In the chart, the abscissa represents the sample capacity, and the ordinate stands for the average value of the predicted sample silencing efficiency.

Table 3. siRNAs Designed for SARS-CoV-2

Number	Position in Genome	Sense	Antisense	Predicted Silencing Efficiency (%)
5	22844	GAUUUUACAGGCUGCGUUAuu	UAACGCAGCCUGUAAAAUCuu	97
11	24421	CCAAAAUGCACAAGCUUUAuu	UAAAGCUUGUGCAUUUUGGuu	98
14	24863	CACUGGUUUGUAACACAAAuu	UUUGUGUUACAAACCAGUGuu	97
15	25124	GAGGUUGCCAAGAAUUUAAuu	UUAAAUUCUUGGCAACCUCuu	98
17	25229	GGCUUGAUUGCCAUAGUAAuu	UUACUAUGGCAAUCAAGCCuu	97

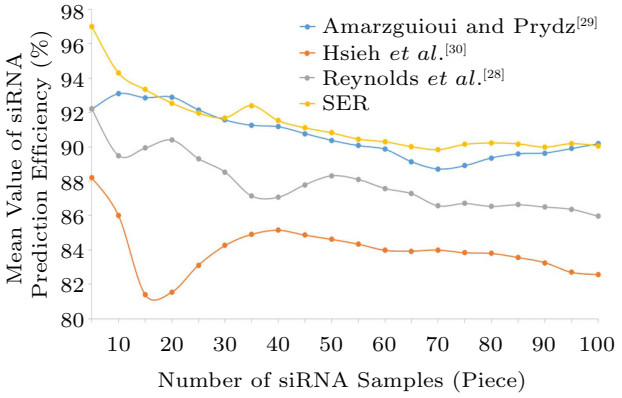


Fig.2. Screening performance chart for different indicators.

To measure the ability of the proposed new index SER in screening the high-efficiency siRNAs, our method is compared with those proposed by Reynolds *et al.*[28], Amarzguioui and Prydz[29], and Hsieh *et al.*[30] based on 3 802 21nt siRNAs on the S gene sequence of SARS-CoV-2. Four methods are used to sort all the candidate siRNAs in the descending order, then multiple samples were taken, and the siRNA scales[47] are utilized to predict the sample silencing efficiency. The comparison results of the screening abilities of the four methods are displayed in Fig.2.

As shown in Fig.2, the prediction efficiency and the silencing efficiency of siRNAs screened by our SER index decrease with the increase of the number of predicted siRNAs, indicating that the index attains efficient screening performance. Therefore, the relatively inefficient siRNAs with a lower ranking in the number of predicted siRNAs reduces the overall average prediction efficiency. In addition, compared with the siRNAs screened by other indexes, the siRNAs screened by the SER index have evident advantages in the prediction silencing efficiency of the top 10 siRNAs. Moreover, the siRNAs screened by our SER index exhibit the highest prediction efficiency on the whole, which indicates that our SER index outperforms the other three methods in screening high-efficiency siRNAs.

6.2 Rationality Analysis of SMRI

The siRNA scales[47] are utilized to predict the silencing efficiency of candidate siRNAs in the screening process of the SMRI method. As shown in Fig.3, the abscissa represents the number of execution steps in the SMRI method. Steps 1 and 2 in the SMRI method do not screen the candidate siRNAs. As a result, they are not drawn in Fig.3.

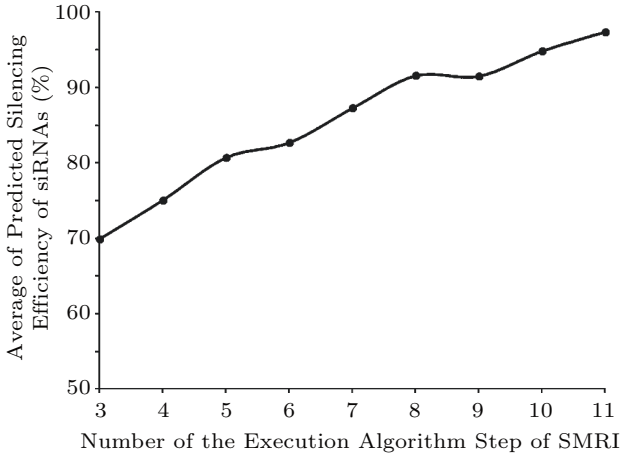


Fig.3. Trend chart of average prediction efficiency of candidate siRNAs in the screening process using the SMRI method.

It is observed from Fig.3 that, after screening the three indexes, the prediction efficiency of candidate siRNAs increases. Besides, the selection steps of our index effectively utilize the screening ability of each index and successfully screen out the high-efficiency siRNAs layer by layer. Firstly, the SER index with more evaluation features is used to preliminarily screen the candidate siRNAs. Secondly, the Reynolds rule[28] is employed to more rigorously and precisely screen the candidate siRNAs. Thirdly, the i-Score method[34] with a high screening capability is utilized to screen siRNAs with high silencing efficiency. By adopting this order, our proposed SMRI method achieves excellent performance in screening siRNAs.

6.3 Screening Performance Analysis of SMRI

The whitehead siRNA selection web server[48] is an siRNA design tool developed by the Whitehead Institute at MIT, which has the characteristics of flexible customization. As shown in Table 4, we simply record this method as Whitehead. In the present study, this tool is adopted to retrieve the S genome sequence of SARS-CoV-2 with a typical AA(N19)UU format, in which N indicates any nucleotide[29]. Candidate siRNAs which have a mismatch of over three bases in the BLAST homology analysis of the human genome were retained. After the BLAST analysis, nine candidate siRNAs are obtained. The remaining settings are the same as the default settings of the Whitehead[48]. Then, the siRNA scales[47] are utilized to predict the siRNAs silencing efficiency and calculate the average value to obtain the items presented in Table 4. Additionally, Designer of Small Interfering RNA (DSIR)[33] and inhibitory-Score (i-Score)[34]

are the precise second-generation siRNA design methods. This study uses these two methods to screen all the candidate siRNAs with the same siRNA sequence length as siRNAs obtained after step 10 in the SMRI method on the S gene of SARS-CoV-2, respectively. Then, as shown in Table 4, the average predicted silencing efficiency of siRNAs screened by DSIR and i-Score with the same number of siRNAs obtained after step 10 using the SMRI method is calculated.

Table 4. Comparison of the Screening Ability for Effective siRNAs of Different siRNA Design Algorithms

Method	Sample Size	Average Value of Predicted Interference Efficiency (%)
Whitehead	9	74.555 6
i-Score	19	93.947 3
DSIR	19	95.052 6
SMRI	19	94.842 1

It is obviously observed from Table 4 that, our method outperforms the typical siRNA design method Whitehead in searching for target sequences conforming to the AA(N19)UU format [29], and the performance of our method is comparable to that of the advanced high-precision siRNA design methods, such as i-Score and DSIR. Moreover, considering the off-target effect, siRNAs designed by the SMRI method do not target the human genome, and are safe for human body, which cannot be achieved by siRNAs designed by the other three methods.

6.4 Antiviral Process and Advantages of SMRI

In this study, in order to maintain the reliability of the efficacy of siRNAs in the case of virus mutation, the target region of designed siRNAs selects the conserved regions in the S gene of SARS-CoV-2. As mentioned in Section 2, if the designed siRNAs were applied to the patients, they cleaved the RNA sequences of SARS-CoV-2, resulting in the functional silencing of the SARS-CoV-2 S gene. This further blocks the translation of the spike glycoprotein, which is utilized by SARS-CoV-2 for cellular invasion, thereby significantly reducing the threat and harm induced by the virus to the human body. In addition, an enhanced antiviral effect can be acquired by using multiple siRNAs simultaneously [22]. With the cooperation of both the immune system and other drugs, patients will overcome the residual SARS-CoV-2 and regain their health hopefully. Compared with other therapies, RNAi is safe and only specific to SARS-CoV-2. Through the BLAST analysis on the target sequences of siRNAs, the siRNAs designed using our method have more than three mismatched bases with the human genome. This minimizes the miss target effect of siRNAs and ensures the safety and reliability of siRNAs for the human body. Most importantly, compared with other siRNA design algorithms, siRNAs designed by our new method have a high silencing efficiency, which will render them with strong antiviral effect.

7 Conclusions

Compared with the existing research, this study proposed a new mathematical model named BPTC model regarding the siRNA silencing efficiency, a new index SER to extract the rules in methods summarizing the characteristic properties of high-efficiency siRNAs and to efficiently screen high-efficiency siRNAs based on the BPTC model. It also presented a new, efficient method named SMRI to efficiently design siRNAs using the SER index to effectively filter out the candidate siRNAs that are difficult to take effect or synthesize. That is, the SER index is part of the SMRI method. Moreover, this study successfully designed five potentially efficient siRNAs using the SMRI method. Findings in this study provided an effective support for the subsequent research and development of RNAi for the treatment of COVID-19.

Moreover, there are several interesting directions in the research of RNAi we want to expand upon, such as developing new siRNA interference efficiency prediction algorithms using deep learning methods, and explaining the related factors of effective RNAi using machine learning methods.

References

[1] Saif L J, Wang Q, Vlasova A N *et al.* Coronaviruses. In Diseases of Swine (11th edition), Zimmerman J J, Karriker L A, Ramirez A, Schwartz K J, Stevenson G W, Zhang J Q (eds.), 2019, pp.488-523. DOI: 10.1002/9781119350927.ch31.

[2] Cui J, Li F, Shi Z L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 2018, 17(3): 181-192. DOI: 10.1038/s41579-018-0118-9.

[3] Li F. Structure, function, and evolution of coronavirus spike proteins. *Annual Review of Virology*, 2016, 3(1): 237-261. DOI: 10.1146/annurev-virology-110615-042301.

[4] Richman D D, Whitley R J, Hayden F G. Clinical Virology (3rd edition). ASM Press, 2009.

[5] Drosten C, Günther S, Preiser W *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *The New England Journal of Medicine*, 2003, 348: 1967-1976. DOI: 10.1056/NEJMoa030747.

[6] Yin Y, Wunderink R G. MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*, 2017, 23(2): 130-137. DOI: 10.1111/resp.13196.

[7] Shereen M A, Khan S, Kazmi A *et al.* COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, 2020, 24: 91-98. DOI: 10.1016/j.jare.2020.03.005.

[8] Zhu N, Zhang D, Wang W *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 2020, 382(8): 727-733. DOI: 10.1056/NEJMoa2001017.

[9] Wang C, Horby P W, Hayden F G *et al.* A novel coronavirus outbreak of global health concern. *The Lancet*, 2020, 395(10223): 470-473. DOI: 10.1016/S0140-6736(20)30185-9.

[10] Chen N, Zhou M, Dong X *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, 2020, 395(10223): 507-513. DOI: 10.1016/S0140-6736(20)30211-7.

[11] Lindbo J A, Silva-Rosales L, Proebsting W M *et al.* Induction of a highly specific antiviral state in transgenic plants: Implications for regulation of gene expression and virus resistance. *The Plant Cell*, 1993, 5(12): 1749-1759. DOI: 10.1105/tpc.5.12.1749.

[12] Vance V, Vaucheret H. RNA silencing in plants—Defense and counterdefense. *Science*, 2001, 292(5525): 2277-2280. DOI: 10.1126/science.1061334.

[13] Montgomery M K, Xu S, Fire A. RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95(26): 15502-15507. DOI: 10.1073/pnas.95.26.15502.

[14] Voinnet O, Pinto Y M, Baulcombe D C. Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(24): 14147-14152. DOI: 10.1073/pnas.96.24.14147.

[15] Nishikura K. A short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. *Cell*, 2001, 107(4): 415-418. DOI: 10.1016/S0092-8674(01)00581-5.

[16] Tijsterman M, Ketting R F, Plasterk R H A. The genetics of RNA silencing. *Annual Review of Genetics*, 2002, 36(1): 489-519. DOI: 10.1146/annurev.genet.36.043002.091619.

[17] Lippman Z, Martienssen R. The role of RNA interference in heterochromatic silencing. *Nature*, 2004, 431(7006): 364-370. DOI: 10.1038/nature02875.

[18] Siomi H, Mikiko C, Siomi M C. On the road to reading the RNA-interference code. *Nature*, 2009, 457(7228): 396-404. DOI: 10.1038/nature07754.

[19] McManus M T, Sharp P A. Gene silencing in mammals by small interfering RNAs. *Nature Reviews Genetics*, 2002, 3(10): 737-747. DOI: 10.1038/nrg908.

[20] Thomas M, Greil J, Heidenreich O. Targeting leukemic fusion proteins with small interfering RNAs: Recent advances and therapeutic potentials 1. *Acta Pharmacologica Sinica*, 2006, 27(3): 273-281. DOI: 10.1111/j.1745-7254.2006.00282.x.

[21] Ichim T E, Li M, Qian H *et al.* RNA interference: A potent tool for gene-specific therapeutics. *American Journal of Transplantation*, 2004, 4(8): 1227-1236. DOI: 10.1111/j.1600-6143.2004.00530.x.

[22] Qureshi A, Tantray V G, Kirmani A R *et al.* A review on current status of antiviral siRNA. *Reviews in Medical Virology*, 2018, 28(4): Article No. e1976. DOI: 10.1002/rmv.1976.

[23] Shi Y, Yang D H, Xiong J *et al.* Inhibition of genes expression of SARS coronavirus by synthetic small interfering RNAs. *Cell Research*, 2005, 15(3): 193-200. DOI: 10.1038/sj.cr.7290286.

[24] Barouch-Bentov R, Einav S. Turning up your nose for a flaviviral encephalitis cure. *Cell Host & Microbe*, 2018, 23(4): 427-429. DOI: 10.1016/j.chom.2018.03.014.

[25] Wu C J, Chan Y L. Antiviral applications of RNAi for coronavirus. *Expert Opinion on Investigational Drugs*, 2006, 15(2): 89-97. DOI: 10.1517/13543784.15.2.89.

[26] Thakur N, Qureshi A, Kumar M. VIRsiRNAdb: A curated database of experimentally validated viral siRNA/shRNA. *Nucleic Acids Research*, 2012, 40(D1): D230-D236. DOI: 10.1093/nar/gkr1147.

[27] Elbashir S, Harborth J, Weber K, Tuschl T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods*, 2002, 26(2): 199-213. DOI: 10.1016/S1046-2023(02)00023-3.

[28] Reynolds A, Leake D, Boese Q, Scaringe S, Marshall W S, Khvorova A. Rational siRNA design for RNA interference. *Nat. Biotechnol.*, 2004, 22(3): 326-330. DOI: 10.1038/nbt936.

[29] Amarzguioui M, Prydz H. An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun.*, 2004, 316(4): 1050-1058. DOI: 10.1016/j.bbrc.2004.02.157.

[30] Hsieh A C, Bo R, Manola J, Vazquez F, Bare O, Khvorova A, Scaringe S, Sellers W R. A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: Determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res*, 2004, 32(3): 893-901. DOI: 10.1093/nar/gkh238.

[31] Takasaki S, Kotani S, Konagaya A. An effective method for selecting siRNA target sequences in mammalian cells. *Cell Cycle*, 2004, 3(6): 790-5. DOI: 10.4161/cc.3.6.892.

[32] Katoh T, Suzuki T. Specific residues at every third position of siRNA shape its efficient RNAi activity. *Nucleic Acids Res.*, 2007, 35(4): Article No. e27. DOI: 10.1093/nar/gkl1120.

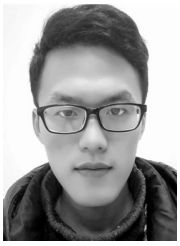
[33] Vert J P, Foveau N, Lajaunie C *et al.* An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics*, 2006, 7(1): Article No. 520. DOI: 10.1186/1471-2105-7-520.

[34] Ichihara M, Murakumo Y, Masuda A *et al.* Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic Acids Research*, 2007, 35(18): Article No. e123. DOI: 10.1093/nar/gkm699.

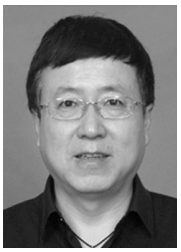
[35] Qureshi A, Thakur N, Kumar M. VIRsiRNApred: A web server for predicting inhibition efficacy of siRNAs targeting human viruses. *Journal of Translational Medicine*, 2013, 11: Article No. 305. DOI: 10.1186/1479-5876-11-305.

[36] Pei Y, Tuschl T. On the art of identifying effective and specific siRNAs. *Nature Methods*, 2006, 3(9): 670-676. DOI: 10.1038/nmeth911.

[37] Geley S, Müller C. RNAi: Ancient mechanism with a promising future. *Experimental Gerontology*, 2004, 39(7): 985-998. DOI: 10.1016/j.exger.2004.03.040.
[38] Martinez J, Patkaniowska A, Urlaub H, Lührmann R, Tuschl T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, 2002, 110(5): 563-574. DOI: 10.1016/s0092-8674(02)00908-x.
[39] Novina C D, Sharp P A. The RNAi revolution. *Nature*, 2004, 430(6996): 161-164. DOI: 10.1038/430161a.
[40] Schubert S, Grünweller A, Erdmann V A *et al*. Local RNA target structure influences siRNA efficacy: Systematic analysis of intentionally designed binding regions. *Journal of Molecular Biology*, 2005, 348(4): 883-893. DOI: 10.1016/j.jmb.2005.03.011.
[41] Preall J B, Sontheimer E J. RNAi: RISC gets loaded. *Cell*, 2005, 123(4): 543-545. DOI: 10.1016/j.cell.2005.11.006.
[42] Holen T, Amarzguioui M, Wiiger M T, Babaie E, Prydz H. Positional effects of short interfering RNAs targeting the human coagulation trigger tissue factor. *Nucleic Acids Res.*, 2002, 30(8): 1757-1766. DOI: 10.1093/nar/30.8.1757.
[43] Yiu S M, Wong P W H, Lam T W *et al*. Filtering of ineffective siRNAs and improved siRNA design tool. *Bioinformatics*, 2005, 21(2): 144-151. DOI: 10.1093/bioinformatics/bth498.
[44] Zhao W M, Song S H, Chen M L *et al*. The 2019 novel coronavirus resource. *Hereditas* (*Beijing*), 2020, 42(2): 212-221. DOI: 10.16288/j.yczz.20-030. (in Chinese)
[45] Hoffmann M, Kleine-Weber H, Schroeder S *et al*. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 2020, 181(2): 271-280. DOI: 10.1016/j.cell.2020.02.052.
[46] Li W, Moore M J, Vasilieva N *et al*. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 2003, 426(6965): 450-454. DOI: 10.1038/nature02145.
[47] Matveeva O, Nechipurenko Y, Rossi L, Moore B, Saetrom P, Ogurtsov A Y, Atkins J F, Shabalina S A. Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Res.*, 2007, 35: Article No. e63. DOI: 10.1093/nar/gkm088.
[48] Yuan B, Latek R, Hossbach M, Tuschl T, Lewitter F. siRNA selection server: An automated siRNA oligonucleotide prediction server. *Nucleic Acids Res.*, 2004, 32(suppl_2): W130-W134. DOI: 10.1093/nar/gkh366.



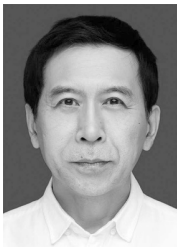
Meng-Xin Chen received his B.S. degree in chemical engineering and technology from Changzhou University, Changzhou, in 2019. He is currently studying for his M.S. degree in College of Software of Jilin University, Changchun. His research interests are machine learning, bioinformatics, and RNA interference.



Xiao-Dong Zhu received his B.S. degree in computer science and technology from Peking University, Beijing, his M.S. degree in computer application and technology and his Ph.D. degree in communication information system, both from Jilin University, Changchun. He is currently a professor of computer science of Jilin University, Changchun. His research interests include software engineering, RNA interference, pattern recognition, and machine learning techniques for biometrics.

Hao Zhang received his B.S. degree in computer applications, M.S. degree in biomechanics and Ph.D. degree, all from Jilin University, Changchun. He is currently a professor of computer science of Jilin University, Changchun. His current research interests include machine learning, bioinformatics, and RNA interference.

Zhen Liu received his B.S. degree, M.S. degree, both in computer applications, from Jilin University, Changchun, and his Ph.D. degree in information science from Tohoku University, Sendai, in 1998. He is a professor in the Graduate School of Engineering, Nagasaki Institute of Applied Science, Nagasaki. His current research interests include RNA interference, big data and data mining, and embedded computing. He is a member of IEEE Computer Society.



Yuan-Ning Liu received his B.S. degree, M.S. degree in computer application and Ph.D. degree in communication information system from Jilin University, Changchun. He is currently a professor of computer science of Jilin University, Changchun. His current research interests are software engineering, RNA interference, pattern recognition, and bioinformatics.