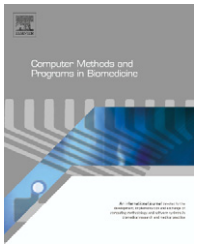
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 85 (2007) 203–209

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Batch RNAi selector: A standalone program to predict specific siRNA candidates in batches with enhanced sensitivity

*Shyamala Iyer, Kerry Deutsch, Xiaowei Yan, Biaoyang Lin**

The Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103, USA

ARTICLE INFO

Article history:
Received 15 July 2006
Received in revised form
12 November 2006
Accepted 14 November 2006

Keywords:
RNA interference (RNAi)
Short interfering RNAs (siRNA)
FASTA
SSEARCH

ABSTRACT

RNA interference (RNAi) is a popular and effective method for silencing gene expression. siRNAs should be gene-specific and effective to achieve specific and potent gene silencing. However, most currently available siRNA design programs are web-based programs that either require each sequence be submitted individually, making large-batch analyses difficult to conduct, or only provide limited options for searching off-target candidates (e.g. NCBI-BLAST). We have developed a stand-alone, enhanced RNAi design program that overcomes these shortcomings. We have implemented WU-BLAST, FASTA and SSEARCH homology searches for siRNA candidates to improve gene specific siRNA selection and to identity siRNA candidates that could lead to off-target gene silencing. We also included many new features such as siRNA score calculation and calculation of siRNA internal stability to help select highly potent siRNAs. This program is freely available for academic and commercial use (ftp://ftp.systemsbiology.net/pub/blin/SiRNA/), and can be installed and run on any Linux machine. Our program automates the search for siRNAs and the resulting data files including a list of siRNA primers with scores and database search results for each siRNA candidate are stored locally for easy retrieval and inspection when needed.

© 2006 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

RNA interference (RNAi) is a popular and effective method for silencing gene expression in a variety of biological systems including human cells (reviewed [1,2]). RNA interference (RNAi) is triggered by double stranded RNAs that are cleaved into short 21–23 nucleotides (nt) duplexes by an RNase III type enzyme Dicer in the RNA induced silencing complex (RISC) [3]. There are two ways in which RNA silencing can be delivered: using chemically synthesized short interfering RNAs (siRNAs) that are usually 21–23 nt or using short hairpin RNAs (shRNAs) that are usually 70 nt. shRNA consists of two oligos around 30 nt forming a stem structure and a loop region. The vector for making an shRNA construct is usually transcribed from a single promoter, therefore a long oligonucleotide is needed to create an shRNA structure [4]. Recently, an ingenious way of putting dual and opposite promoters into vectors was developed so that siRNA can be readily generated from 21 to 23 nucleotides cloned into such vectors rather than chemically synthesized, thus simplifying the design and reducing the cost of making RNA silencing constructs, especially when creating an RNAi library consisting of RNAi silencing constructs for thousands of genes [5].

The challenge is to design siRNAs with high specificity and high potency. There are many web-based programs available to predict siRNA or shRNA sequences [6–9]. The most

* *Corresponding author*. Tel.: +1 206 732 1297; fax: +1 206 732 1299.
E-mail address: blin@systemsbiology.org (B. Lin).
0169-2607/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2006.11.004

commonly used programs for siRNA designs (such as siRNA Target Finder at Ambion Inc., siDesign at Dharmacon Inc., siRNA Target Finder at GenScript Inc., OptiRNAi [10], Si-RNA DEsign [11] and the Gene-specific siRNA Selector [12], etc.) are all web server-based. They either require that each individual sequence be pasted or uploaded, making large-batch analyses difficult to conduct, or they only provide limited options for searching off-target candidates (e.g. NCBI-BLAST). The NCBI-BLAST search algorithm has been shown to fail to identify close sequence similarities [13] and siRNA candidates designed using these programs might have off-target silencing effects. Snove and Holen studied 359 published siRNA sequences and found that 75% of them have potential off-target silencing as most of them were designed using popular NCBI-BLAST searches [14].

As part of our goal to create an siRNA library for hundreds to thousands of differentially expressed genes that we identified using high throughput technologies (e.g. cDNA arrays, massively parallel signature sequencing, etc.), we felt the need to have a stand-alone program that can design gene specific siRNAs for hundreds of genes simultaneously and that is flexible enough so that parameters can be modified easily. Also, it would be convenient to be able to store the results automatically and locally, so that one is able to access the data and to correlate it with wet bench experimental results. To minimize the probability of designing an siRNA that has cross-gene and off-target silencing efforts, we wanted to have the option to use more sensitive sequence similarity searches such as FASTA or SSEARCH, which are more sensitive in picking up homologous sequences than the standard NCBI-BLAST search commonly used in the web-based siRNA design programs. Finally, a good siRNA design program should be able to determine both high specificity and potency of the siRNA oligo for silencing a specific gene and give users options to specify search options to balance search speed, specificity and potency (Fig. 1).

2. Methods

2.1. *Overall design and data flow of our RNAi selection program*

Our program, Batch RNAi Selector, is built upon the Apache web server siRNA program [12] with many improvements and user-defined options (Table 1). Fig. 2 shows a data flow overview of our siRNA program. Our program is designed for batch searches, allowing multiple sequences or GI numbers to be processed simultaneously. The input for this program can be a list of accession numbers or sequences in FASTA format and the output is an ordered and prioritized list of siRNA oligos based on RNAi silencing scores. The output files (siRNA candidate list with scores and free energy values, homology search results by NCBI-BLAST, WU-BLAST, FASTA or SSEARCH) are saved in a subfolder named by accession number, current date and time.

UniGene database is a non-redundant set of gene-oriented clusters (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene). We search against the UniGene database so that we do not have to worry about redundant entries. In contrast, in commonly used web-based RNAi design programs, the database used was often the whole Genbank databases that contain redundant entries. Please note that even the nr (non-redundant) subset of the Genbank contains many redundant entries. Another advantage of using the UniGene database is that it also increases the speed in homology search, especially when SSEARCH or FASTA with word size = 1 search option is used.

Table 1 – Batch RNAi selector program options

Options	Accepted values
-file	Input file with accession numbers or FASTA sequences
-format	File format (*seq* for sequences, *ginumber* for accession numbers)
-FASTAdb	Database to retrieve sequence in FASTA format given a GI accession number
-siRNA_len	Length of siRNA sequence
-dna_start	Number of residues to ignore from start codon
-dna_end	Number of residues to ignore from termination codon
-min_GC	Minimum GC content
-max_GC	Maximum GC content
-border	Flanking residue border (0: no AATT; 1: AATT; 2: AA; 3: TT)
-expect	Expect value for NCBI-BLAST
-word	Word size for NCBI-BLAST
-mismatch	Mismatch value for NCBI-BLAST
-organism	Organism (human, mouse or rat)
-blast_num	Number of siRNA candidates selected for NCBI-BLAST, WU-BLAST and FASTA
-FASTA_check	Y or N value for conducting FASTA searches
-FASTA_ktup	ktup value for FASTA searches
-FASTA_evalue	E value for FASTA searches
-wublast	Y or N value for conducting WU-BLAST searches
-wublast_word	Word size for WU-BLAST
-fasta_t -T	Use multiple processors for FASTA. Specify the number of processors with option -T
-Ssearch	Y or N value for conducting SSEARCH searches
-Ssearch_t -T	Use multiple processors for SSEARCH. Specify the number of processors with option -T
-ensembl	Y or N value for the Ensembl database. Default is N, and the UniGene Hs.seq.uniq database is used

Download English Version:

https://daneshyari.com/en/article/467281

Download Persian Version:

https://daneshyari.com/article/467281

Daneshyari.com