Supplementary Information

CRISPR-RT: A web application for designing CRISPR-C2c2 crRNA with improved target specificity

Houxiang Zhu¹, Emily Richmond² and Chun Liang¹,²,*

¹Department of Biology and ²Department of Computer Science and Software Engineering, Miami University, Oxford 45056, USA

HZ: zhuh4@miamioh.edu
ER: richmoet@miamioh.edu
CL: liangc@miamioh.edu

* To whom correspondence should be addressed.
**Supplementary Figure S1.** The configuration and architecture of the CRISPR-C2c2 system. The CRISPR array of CRISPR-C2c2 is first transcribed into pre-crRNA. Then, the pre-crRNA is processed by C2c2 into mature crRNAs without attaching to trans-activating crRNAs. The mature crRNA binds to C2c2 and guides C2c2 to target a specific single-strand RNA at a proper target site.
Supplementary Figure S2. The parameter setting page of CRISPR-C2c2. First, users input an RNA sequence in FASTA format into the text area. Second, users select a reference transcriptome. Third, users set up the PFS and crRNA requirements for the CRISPR-C2c2 system. Fourth, users choose an off-target setting (either “Basic settings” or “Specific settings”).
Supplementary Figure S3. The result web interfaces of CRISPR-RT. (A) The detailed information of all the target candidates in the user input RNA sequence. (B) User input RNA sequence viewer. (C) The detailed information of target sites in the transcriptome for each target candidate. (D) Visualization of on- and off-targets with gene and transcript annotations in JBrowse.
Supplementary Figure S4. The basic process of crRNA design based on the target candidate sequence (protospacer). The target complementarity sequence of crRNA is labelled in purple color. The stem-loop sequence of crRNA comes from previous research (Abudayyeh et al., 2016). The designed crRNA binds to C2c2 to form a crRNA-C2c2 complex for RNA targeting.
Supplementary Figure S5. The flowchart of CRISPR-RT backend pipeline that processes user input RNA sequence and displays results in the interfaces. All of input parameters, including the RNA sequence, are passed to the main Perl script. The main Perl script invokes other specific Perl scripts or commands to execute specific functions. Finally, the results will be displayed in DataTables and JBrowse in the web interfaces.
Supplementary Methods

CRISPR-RT is essentially composed of many web interfaces and a backend pipeline. Web interfaces are implemented by PHP and JavaScript code, which are used to accept user inputs and display the results interactively. The backend pipeline is implemented by Perl code, which is used to process user input data and generate multiple result files. The same strategy was applied in our previously published web application CT-Finder (Zhu et al., 2016), which helps to design Cas9 gRNAs.

After setting up proper parameters and clicking the “Find targets!” button in the parameter setting page of CRISPR-C2c2 (Supplementary Figure S2), all of the parameters stored in PHP code are passed to the main Perl script, which invokes other specific Perl scripts or commands to execute specific functions. The default parameters are set based on Abudayyeh et al.’s research. As shown in Supplementary Figure S5, first, the Perl script for target candidate search is called to find target candidates of specified length with the PFS in the input RNA sequence. Second, Bowtie2 (Langmead and Salzberg, 2012) is used to map each target candidate sequence to the reference transcriptome, which is extracted from the genome by RSEM (Li and Dewey, 2011) using Ensembl or Phytozome gene annotation, to search for on- and off-target sites within the transcriptome. Bowtie2 is particularly good at aligning short reads to long genomes or transcriptomes, which has been used for Cas9 gRNA design in previous study (Heigwer et al., 2014; Zhu et al., 2016). RSEM is a quite popular package for analyzing RNA-Seq data (Haas et al., 2013; Konermann et al., 2015; Shalek et al., 2013). Third, the Perl scripts for result filtration based on the input parameters are invoked to filter out the off targets that do not meet the requirements set by users. Since the seed region is more sensitive to mismatches than the non-seed region, they are handled separately. Next, RSEM is used to convert the
transcriptome mapping result to a genome mapping result, which can be displayed properly in JBrowse (Skinner et al., 2009). Then, the main Perl script separates the file storing target sites of all target candidates into many single files that store target sites for each target candidate respectively. The main Perl script also processes the file of target candidates and the files of target sites for each target candidate to generate Ajax format files, which are required by DataTables (a table plug-in for jQuery). After getting all of those files, the PHP and JavaScript code is used to display detailed information of target candidates and corresponding target sites in DataTables. The on- and off-target sites can be visualized in JBrowse.

References