TransCRISPR - sgRNA design tool for CRISPR/Cas9 experiments targeting transcription factor motifs

Tomasz Woźniak, Weronika Sura, Marta Kazimierska, Marta Elżbieta Kasprzyk, Marta Podralska, Agnieszka Dzikiewicz-Krawczyk*

Institute of Human Genetics, Polish Academy of Sciences, Poznań, Poland

*correspondence: agnieszka.dzikiewicz-krawczyk@igcz.poznan.pl; Strzeszyńska 32, 60-479 Poznań, Poland

ABSTRACT

Transcription factors (TFs) regulate gene expression via binding to specific sequence motifs in promoters or enhancers. They are involved in various biological processes and their aberrant action may lead to multiple pathological states. To better understand the function of a given TF, it is necessary to identify its crucial binding sites and target genes, which is not a trivial task. CRISPR/Cas9 has been recently successfully applied in the study of TFs. Here, we present an online tool, transCRISPR, created to search for TF binding motifs in the user-provided query and design optimal sgRNAs targeting them. Users can obtain sgRNAs for chosen TFs motifs, for up to tens of thousands of target regions in the human or mouse genome, either for the Cas9 or dCas9 system. TransCRISPR provides user-friendly tables and visualizations, summarizing features of identified motifs and designed sgRNAs such as genomic localization, quality scores, closest transcription start sites, and others. TransCRISPR is available at https://transcrispr.igcz.poznan.pl/transcrispr/ .

INTRODUCTION

Transcription factors (TFs) are a group of proteins that bind to DNA regulatory elements such as promoters or enhancers, affecting target gene expression. TFs can reprogram gene expression profiles in response to internal and external stimuli, and as a result control cellular processes such as cell division, migration, growth, apoptosis, differentiation, etc. (1, 2). TFs recognize specific sequence motifs in DNA and by recruitment and interaction with other factors they can up- or downregulate expression of target genes (3, 4, 5). Transcription factors are a part of complex regulatory networks and any aberrations of their function may have severe consequences, e. g. carcinogenesis (6).

To better understand the function of a given TF, it is necessary to identify its crucial target genes. Although TFs have been comprehensively investigated for years, still their complex regulatory networks remain by large unknown (7). Approaches such as chromatin immunoprecipitation (ChIP) or TF overexpression/knockdown are widely utilized to decipher TFs regulatory mechanisms. ChIP followed by next-generation sequencing (ChIP-seq) can reveal TFs binding sites across the genome and indicate putative target genes (8, 9). It was shown that those binding sites are localized in regulatory sequences like enhancers and promoters, close to the transcription start site (TSS) (7). Moreover, analysis of gene expression upon TF knockdown or overexpression may also reveal thousands of responding genes as well as processes controlled by those TFs (10, 11, 12). However, pinpointing binding sites and target genes crucial for the cell is more complicated.

Clustered Regularly Interspaced Short Palindromic Repeats/Cas9 (CRISPR/Cas9) has become one of the most powerful tools for genome editing and revolutionized genome engineering. Nuclease Cas9 and its catalytically dead form dCas9 are utilized to affect e.g. gene expression by DNA cleavage, blocking transcription factor binding sites and altering epigenetic modifications such as methylation or acetylation, etc. (13, 14). Apart from studying protein-coding genes, pooled CRISPR/Cas9 screens have been also utilized to characterize regulatory elements: promoters, enhancers, etc. Both nuclease Cas9 and dCas9 have been recently used to study transcription factors (TFs), their associations with tumorigenesis and identification of essential target genes (15, 16, 17, 18, 19, 20).

To perform reliable and informative CRISPR/Cas9 experiments, high specificity and efficiency of the approach are required. In answer to this need, numerous online tools have been designed. Those tools give the possibility to design the most optimal single-guide RNAs (sgRNAs) targeting specific sequences and predict their off- and on-target scores to increase their specificity and efficiency (21, 22, 23, 24). Although available tools offer a wide range of possibilities, none of them allows searching for a specific TF binding motif and designing sgRNAs targeting this motif.

Here, we present a highly versatile online tool, transCRISPR, created to identify TFs binding motifs in the sequence of interest and to design sgRNAs with optimal off- and on-target scores. It can be applied both for single target sequences as well as for large lists of genome coordinates, enabling design of sgRNA libraries for genome-wide CRISPR/Cas9 screens.

MATERIALS AND METHODS

Data

Genomic coordinates for coding exons, non-coding exons, introns and transcription start sites (TSS) were downloaded from UCSC using a database interface. Full download command: mysql {genome_name} -h genome-mysql.soe.ucsc.edu -u genome -A -e 'select * from ncbiRefSeqCurated' -NB > {genome_file}. These data were further automatically processed to create specialized .bed files with genes and localization elements for each of available genomes. This approach is dedicated for further development of this tool to include more genomes. Whole genome sequences were downloaded as FASTA files.

Implementation

TransCRISPR is created using Django (with Python programming language). MariaDB is used as a database, Celery with Redis as a query system, Daphne for websocket communication, Nginx as a web server, and Bootstrap based Genetlella for a layout with Highcharts as a data visualization library. Docker with Docker Compose are used for management purposes. Off-targets are calculated using the Cas-OFFinder (25) with a maximum of 4 (standard option) or 3 (rapid option) mismatches and later the CFD score is calculated for each off-target, as well as a cumulative CFD score (25, 26). For on-target value calculation a dockerized version of Azimuth is used (26).

For analyses two queue systems are available: for short calculations and for larger queries. This distinction is important because complex calculations can take up to several days, and it would be undesirable to block calculations that may take a few minutes. Where possible parallelization and other optimizations are used to minimize time required for calculation. Software is freely available online: https://transcrispr.igcz.poznan.pl. All results are kept for 7 days. If an email address is given, an analysis completion message is sent.

Search of the motif positions is performed either with exact search in case of motifs defined as sequences with no IUPAC code or with regular expressions in case of IUPAC code in the sequence. In case of motif matrices, they are converted to IUPAC sequence using a selected rule set and then searched with regular expressions. For each sequence, reverse complementary sequence is generated and used for search on the reverse strand. For each of the found motif positions, potential guides are generated using rules for either Cas9 or dCas9.

In case of search sequences defined as coordinates, respective sequences are selected from downloaded genome files. For guide search sequences extended by 30 nucleotides before and 30 nucleotides after defined coordinates are also selected. This approach is not possible in case of target sequences defined as raw sequences or in FASTA format.

Localization of TF motifs in relation to genes is determined as follows. Firstly, data downloaded from UCSC are sorted and saved to special .bed files containing data from a single chromosome, sorted by given sequence start. This operation is done only once for each genome. Found motifs position on each chromosome is also sorted. Then for each motif position on a given chromosome respective .bed file is being searched. To avoid reading the file multiple times - special mechanism is being used, where firstly there is search for the first localization from .bed file that is overlapping with the motif position. Then in a loop next positions are being read as long as part of them is overlapping with motif position. These data are also added to a buffer that is dedicated to keep them for the next motif position, as localization from .bed files may overlap multiple motifs. In case a motif is on a boundary (e.g. intron-exon, exon-intergenic etc.) or is localized in a position where different transcript variants differ with respect to intron/exon, the following hierarchy is applied: coding exon - non-coding exon - intron - intergenic.

For determination of the closest up- and downstream TSS, data from UCSC are similarly downloaded, sorted and saved to .bed files containing gene localization. During the analysis the TSSs for the closest upstream and downstream gene for each of the sorted motifs are selected and saved.

For each motif and guide, a name is generated. If a target sequence name is given (header in FASTA format or last column in .bed file) this name is used as prefix, in other cases a generic prefix is created.

RESULTS

TransCRISPR is an online tool dedicated to designing sgRNAs targeting transcription factor motifs. This software performs several steps to calculate and display data for given input: 1) processing sequences; 2) processing motifs and finding motifs in sequences; 3) calculating off-targets and on-targets; 4) finding localization and calculating statistics.

Query

To run a query, the user has to provide input data and select available options (Figure 1).

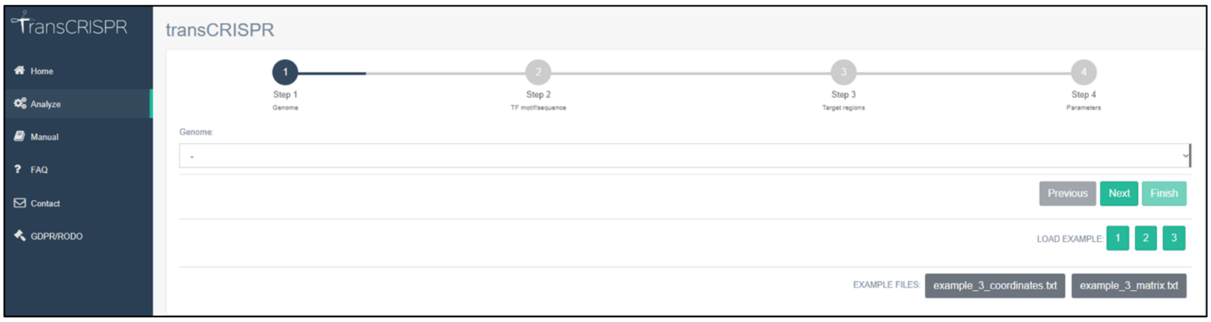


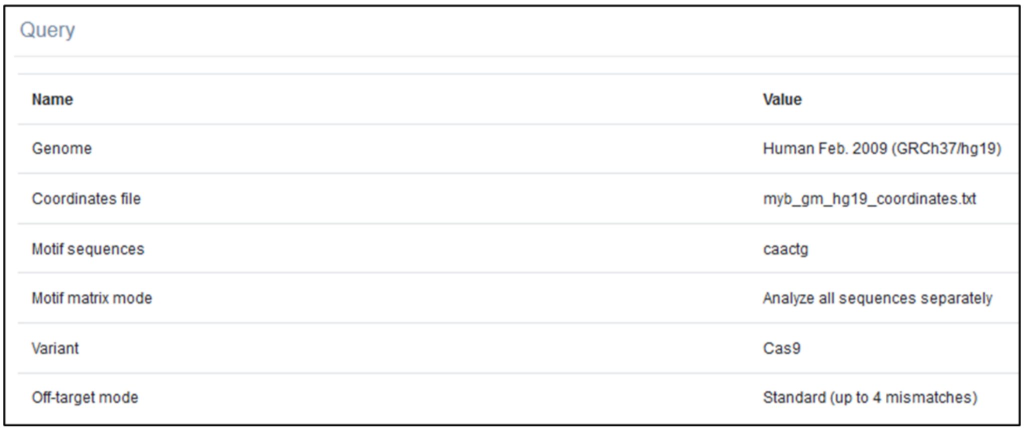
Figure 1. TransCRISPR interface and query steps.

In Step 1, the reference genome is selected. Currently, one can operate on two versions of human genome assembly: GRCh38/hg38 and GRCh37/hg19, and mouse genome assembly GRCm39/mm39. However, we plan to add other genomes, especially those of model organisms, in near future.

TransCRISPR offers many ways of submitting queries. In Step 2 transcription factor motifs might be entered directly in the window as FASTA or coma separated format, or as a motif matrix in various formats, including output files of programs analyzing transcription factors (e. g. JASPAR, TRANSFAC). The motifs might be also uploaded as a sequence or matrix file. Motifs provided as a sequence can contain A, C, G, T nucleotides or IUPAC codes. Next, in Step 3 the target regions where TF motifs will be searched for are provided. Target sequence may be pasted directly in the window as a text (in FASTA or coma separated format) or as genomic coordinates; those can be uploaded as files as well. In the last step, one may choose from various modes of motifs and guides search.

If the motif sequences were entered directly in the form of “Motif sequences” or in the form of “Motif sequence file”, they will be all searched separately. However, if they were entered in the form of a motif matrix (“Motif matrix” or “Motif matrix file”), the user can choose criteria according to which motifs will be generated from the matrix. The first and recommended option is to use a set of D. R. Cavener rules for consensus sequence (27). The nucleotide having over 50% frequency at the specific position and simultaneously having the frequency more than twice higher than the second most common nucleotide is regarded as the consensus nucleotide. If this criterium is not met, two nucleotides whose sum of the frequencies exceeds 75% are regarded as consensus nucleotides; if none of these criteria is met, N is assigned to the position. Alternatively, the user can set that motifs will be generated including all nucleotides which cover at least x% (5%, 10%, 15%, 20%, 25%) at a given position, or the most frequent nucleotides which together cover at least 80%, 85% or 90% at the given position in the motif.

Next, one can choose between the Cas9 and dCas9 variants which define how the sgRNAs are searched in respect to the motifs. In the Cas9 variant, only those guides that lead to the cut within the TF motifs (taking into account that the cut occurs 3 nt upstream of PAM) are designed. In the dCas9 variant, any guides that overlap with at least one nucleotide of TF motifs are considered. As a standard, for all found sgRNAs off-targets up to 4 mismatches are analyzed. To reduce the analysis time, the “rapid” option can be chosen which includes only off-targets with up to 3 mismatches. The user can optionally enter their e-mail address to be informed when calculations are done. Moreover, during the analysis, the user is informed about its progress and current step, as well as about the queue of tasks. For convenience, the details of the query might be checked later (Figure 2). Results are available on the website for seven days.



Query

Name	Value
Genome	Human Feb. 2009 (GRCh37/hg19)
Coordinates file	myb_gm_hg19_coordinates.txt
Motif sequences	caactg
Motif matrix mode	Analyze all sequences separately
Variant	Cas9
Off-target mode	Standard (up to 4 mismatches)

Figure 2. Summary of query details available on the result page.

Analysis results

On the results page, the user is first informed about the number of found motifs and sgRNAs, an average number of guides per motif, and average on- and off-target scores, which are additionally presented on histograms. Distribution of guides per motif and their genomic localization are presented on pie charts. These charts can be viewed full screen, printed or downloaded in various formats. Information about found motifs includes: motif sequence, their position in the provided genomic sequence and localization in relevance to genes (whether it is in coding or non-coding exon, intron or between genes), including the genomic positions of transcription start sites (TSSs) of the closest up- and downstream gene. Next, sgRNAs found for the motifs are presented in the table, where their sequence, relative position of PAM in the sequence and targeted DNA strand are shown, together with the calculated on- and off-target scores. It is possible to view the details of the most significant off-targets. Information about the genomic localization of motifs and closest TSSs is not available if target regions were provided as sequences; genomic coordinates are required to obtain the full characteristics .

The results may be additionally filtered according to several parameters. The user may choose to include only motifs within a specific genomic localization. For Cas9, we recommend to include only motifs in non-coding exons, introns and intergenic, since cutting within motifs located in coding exons will likely disrupt the protein and hence the results will not be conclusive for the TF binding. For dCas9, we recommend excluding motifs localized in the window -200 nt/+100 nt relative to TSS as this region is the most effective for CRISPR interference. Thus, it would be difficult to determine whether the observed effect is a result of blocking the TF binding or of interfering with gene transcription due to dCas9-induced mechanisms. sgRNAs can be filtered based on the on- and off-target threshold. The user may a priori remove sgRNAs having off-targets with 0 or 1 mismatches, regardless of their CFD score. We recommend that the cut-off of 30 is used for CFD off-target score. It is also possible to filter out motifs for which no guides could be designed.

In case several sgRNAs are found per motif, the user may wish to include only a given number of the best guides. For this purpose, the maximum number of sgRNAs based on the off- or on-target score can be requested. The order in which the guides are displayed can be also changed (default by on-target score; off-target or position are possible). In some instances, TF motifs may overlap partially and in such a situation some sgRNAs may be duplicated in the motif view. To obtain the list of nonredundant sgRNAs, the user should switch to the “Unique guides” view. There, the information about the sequence and the position of sgRNAs together with on- and off-target scores and targeted motifs are provided.

The user may download the analysis results in several formats (xslx, csv, tsv, bed). In the Excel file, separate sheets provide results per motifs or per unique guides. It is also possible to download the track to visualize motifs and guides together with their on- and off-target scores coded by colors in the UCSC Genome Browser. Moreover, by clicking Display in Genome Browser the user is directly taken to the Genome Browser with this track loaded.

Detailed explanation of preparing the query and analyzing results is provided in the manual available on the transCRISPR webpage, also as a downloadable pdf. To get familiar with transCRISPR and available options, users are advised to run one of the preloaded examples.

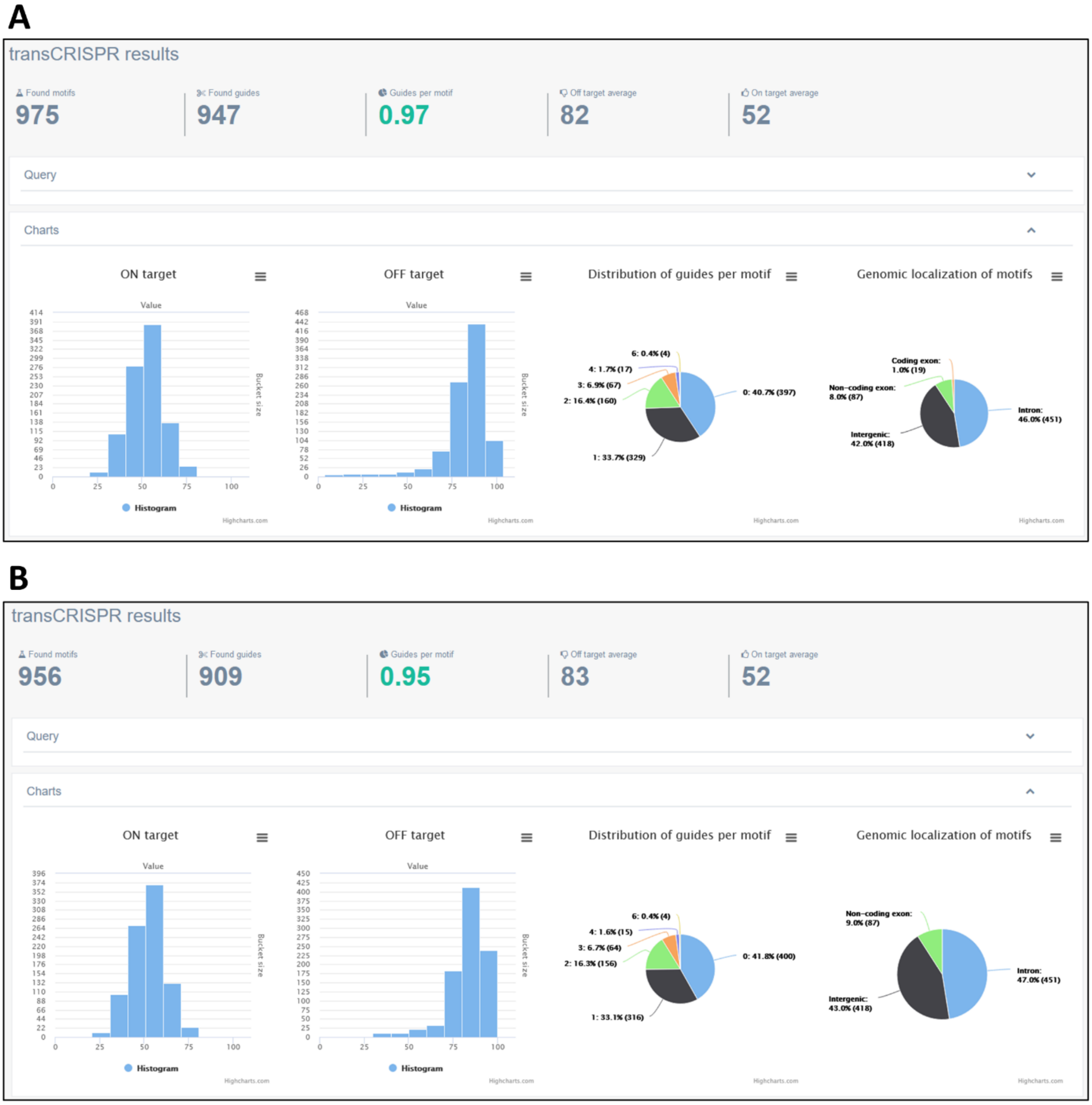
Example query and results

We used transCRISPR to find motifs and sgRNAs within the ChIP peaks for the MYB transcription factor in human GM12878 cells. For this purpose, we chose GRCh37/hg19 genome assembly and pasted the most common MYB motif sequence (CAACTG) directly into the motif window. Next, we used the coordinates of MYB ChIP peaks retrieved from UCSC Table Browser, track ENCODE 3 TFBS, table GM12878 MYB, and uploaded them as a coordinates file (3748 peaks). At first, we chose Cas9 variant and standard off-target mode (up to 4 mismatches) (Figure 2).

After the calculations were finished, the upper panel showed the summary of motifs and guides (Figure 3A) – there were 975 found motifs and 947 guides targeting them, which gives 0.97 guides per motif. The detailed information about the number of sgRNAs per motif presented on the pie chart below shows that for 40.7% of the motifs no sgRNAs could be designed. The remaining motifs were targeted mostly by 1 or 2 sgRNAs; for some up to 6 sgRNAs were designed. The average on- and off-target scores are above 50 which indicates an overall good quality of designed sgRNAs. Detailed information provided on histograms shows that the majority of sgRNAs have off-target scores above 70, only a few are below 50. The predicted cutting efficiency is medium as the majority of sgRNAs have the on-target score around 50. Looking at the genomic localization of identified motifs, we observe that the vast majority is localized in the introns or non-coding exons, much less in intergenic regions and only a few in coding exons.

As mentioned previously, results can be filtered based on various criteria. We used this option to exclude motifs present in coding regions and guides presenting off-target scores below 30. As a result, we obtained 956 motifs and 909 guides (0.95 guides per motif) with the average off-target score increased to 83. We can see on charts that indeed, now there are no motifs in coding exons and no guides with off-target score below 30 (Figure

3B). Figure 4. shows various modes of presenting results by transCRISPR – they can be displayed as a list at the result page (Figure 4A), as a list downloaded in xslx format (Figure 4B and C) or visualized in Genome Browser (Figure 4D). Colors of sgRNAs bars depict on- and off-target values.



transCRISPR results

Found motifs 975 | Found guides 947 | Guides per motif 0.97 | Off target average 82 | On target average 52

Query

Charts

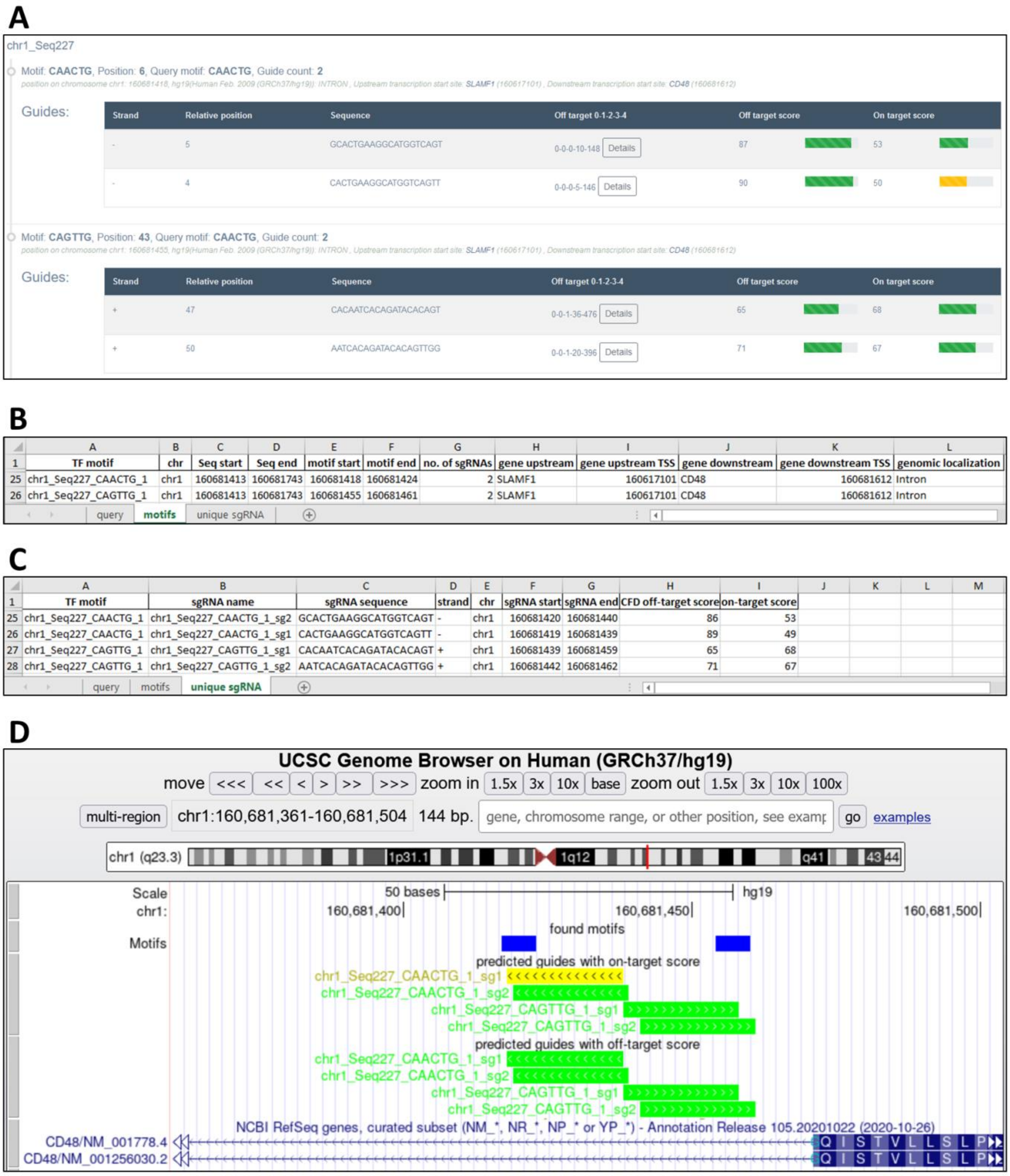
transCRISPR results

Found motifs 956 | Found guides 909 | Guides per motif 0.95 | Off target average 83 | On target average 52

Query

Charts

Figure 3. Statistics of transCRISPR results for the example query in Cas9 mode. The screenshots present initial results of analysis described in the example (A) and results after filtering by genomic localization of motifs and off-target scores (B).



chr1_Seq227

Motif: **CAACTG**, Position: **6**, Query motif: **CAACTG**, Guide count: **2**
position on chromosome chr1: 160681418, hg19(Human Feb. 2009 (GRCh37/hg19)): INTRON , Upstream transcription start site: SLAMF1 (160617101) , Downstream transcription start site: CD48 (160681612)

Guides:

Strand	Relative position	Sequence	Off target 0-1-2-3-4	Off target score	On target score
-	5	GCACTGAAGGCATGGTCAGT	0-0-0-10-148 Details	87	53
-	4	CACTGAAGGCATGGTCAGTT	0-0-0-5-146 Details	90	50

Motif: **CAGTTG**, Position: **43**, Query motif: **CAACTG**, Guide count: **2**
position on chromosome chr1: 160681455, hg19(Human Feb. 2009 (GRCh37/hg19)): INTRON , Upstream transcription start site: SLAMF1 (160617101) , Downstream transcription start site: CD48 (160681612)

Guides:

Strand	Relative position	Sequence	Off target 0-1-2-3-4	Off target score	On target score
+	47	CACAATCACAGATACACAGT	0-0-1-36-476 Details	65	68
+	50	AATCACAGATACACAGTTGG	0-0-1-20-396 Details	71	67

	A	B	C	D	E	F	G	H	I	J	K	L
1	TF motif	chr	Seq start	Seq end	motif start	motif end	no. of sgRNAs	gene upstream	gene upstream TSS	gene downstream	gene downstream TSS	genomic localization
25	chr1_Seq227_CAACTG_1	chr1	160681413	160681743	160681418	160681424	2	SLAMF1	160617101	CD48	160681612	Intron
26	chr1_Seq227_CAGTTG_1	chr1	160681413	160681743	160681455	160681461	2	SLAMF1	160617101	CD48	160681612	Intron

query | **motifs** | unique sgRNA

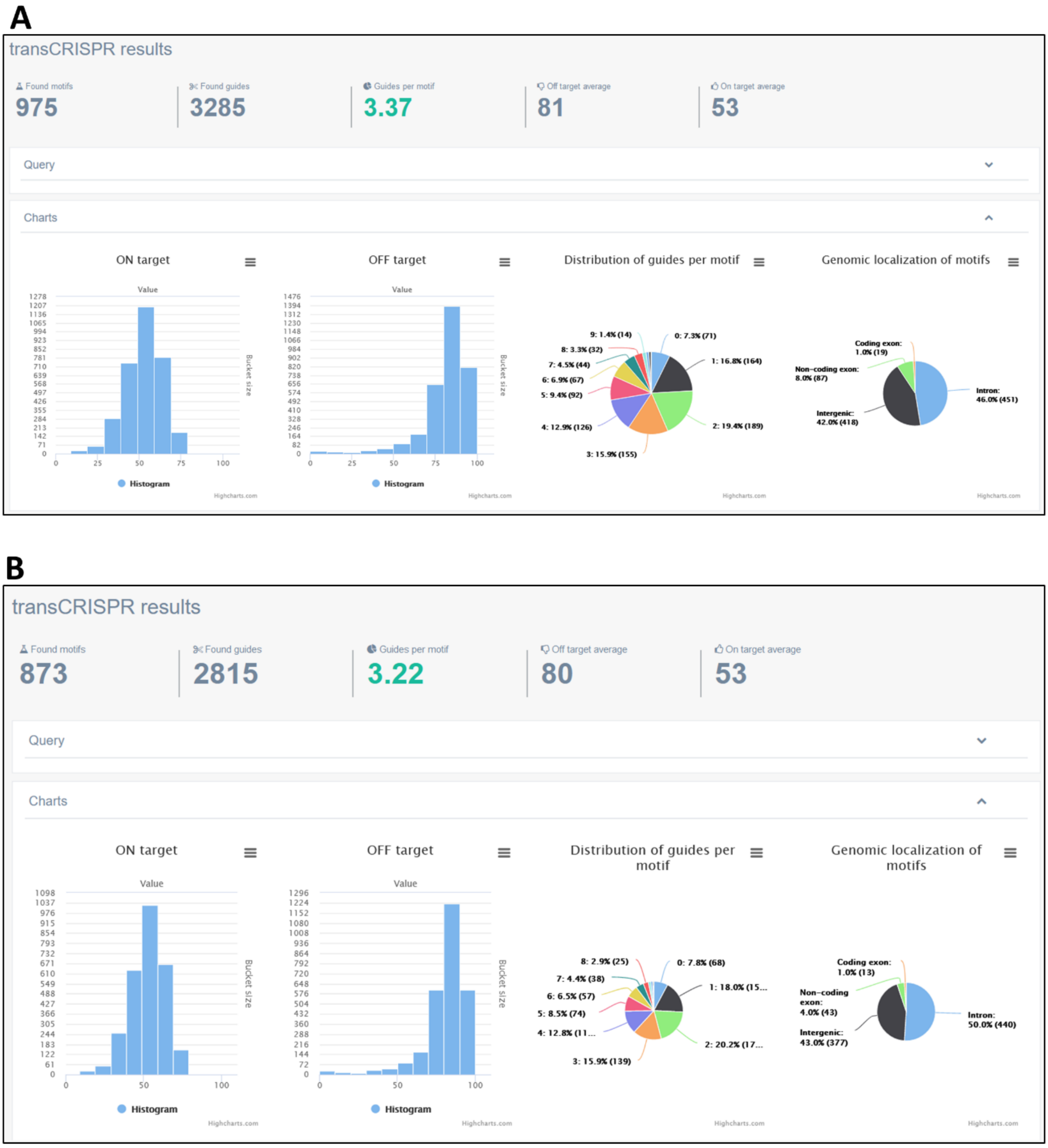
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	TF motif	sgRNA name	sgRNA sequence	strand	chr	sgRNA start	sgRNA end	CFD off-target score	on-target score				
25	chr1_Seq227_CAACTG_1	chr1_Seq227_CAACTG_1_sg2	GCACTGAAGGCATGGTCAGT	-	chr1	160681420	160681440	86	53				
26	chr1_Seq227_CAACTG_1	chr1_Seq227_CAACTG_1_sg1	CACTGAAGGCATGGTCAGTT	-	chr1	160681419	160681439	89	49				
27	chr1_Seq227_CAGTTG_1	chr1_Seq227_CAGTTG_1_sg1	CACAATCACAGATACACAGT	+	chr1	160681439	160681459	65	68				
28	chr1_Seq227_CAGTTG_1	chr1_Seq227_CAGTTG_1_sg2	AATCACAGATACACAGTTGG	+	chr1	160681442	160681462	71	67				

query | motifs | **unique sgRNA**

UCSC Genome Browser on Human (GRCh37/hg19)

Figure 4. Different modes of viewing transCRISPR results. The screenshots present the same motifs and guides (A) directly on the results page; (B) and (C) in the downloaded xslx file (information on motifs – B, information on unique guides - C); and (D) as tracks in Genome Browser.

Changing the Cas9 variant in the query to dCas9, yielded 3285 guides (3.37 guides per motif) (Figure 5). This is expected, as the rules for sgRNA design are broader in this option. In line with this, the number of guides per motif was more diversified, with the prevalence of 1-4 sgRNAs per motif, and maximum of 8 sgRNA per motif. Average on- and off-target values as well as their distribution on histograms look quite similar to the ones from Cas9 mode. When the filters recommended for dCas9 mode were applied, i.e. motifs localized between -200 nt and +100 nt relative to TSS were excluded, the number of found motifs went down to 873 and the number of designed guides decreased to 2815. Other statistics changed accordingly (Fig. 5B).



A

transCRISPR results

Found motifs 975
Found guides 3285
Guides per motif 3.37
Off target average 81
On target average 53

B

transCRISPR results

Found motifs 873
Found guides 2815
Guides per motif 3.22
Off target average 80
On target average 53

Figure 5. Statistics of transCRISPR results for the example query in Cas9 mode. The screenshots present initial results of analysis described in the example (A) and results after filtering by localization of motifs relative to TSS (B).

DISCUSSION AND FUTURE PLANS

We designed the transCRISPR tool to facilitate the study of transcription factor binding sites. This is a unique tool that enables design of sgRNAs targeting a particular sequence in the region of interest. Although created with transcription factor motifs in mind, we envisage that the users may find it useful in other cases when targeting a specific DNA sequence motif is required.

We plan to further develop our tool. First, we are going to expand the list of available reference genomes, including especially those from most common model organisms. We are also going to include more Cas9 variants, recognizing various PAM sequences. This will enable more comprehensive design of sgRNAs targeting specific motifs. We welcome all suggestions for improvement and development from the users via the contact details provided on the webpage.

AUTHORS' CONTRIBUTIONS

TW wrote the analysis source code and website. WS, MK, MEK, MP and ADK performed tests. WS and MK wrote the manuscript and prepared figures. MEK and MP wrote the manual and prepared figures. ADK conceived and supervised the study, and revised the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY

The tool is available freely at https://transcrispr.igcz.poznan.pl/transcrispr/

FUNDING

This work was supported by National Science Centre in Poland (Grant no. 2016/23/D/NZ1/01611) and by the First Team programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund (grant no. POIR.04.04.00-00-5EC2/18-00). AD-K was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 952304.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Babu,M.M., Luscombe,N.M., Aravind,L., Gerstein,M. and Teichmann,S.A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.,* **14,** 283-291.

2. Spitz,F. and Furlong,E.E.M. (2012) Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics,* **13,** 613-626.

3. Gill,G. (2001) Regulation of the initiation of eukaryotic transcription. *Essays Biochem.,* **37,** 33-43.

4. Becskei,A. (2020) Tuning up transcription factors for therapy. *Molecules,* **25,** 1902. doi: 10.3390/molecules25081902.

5. Lambert,S.A., Jolma,A., Campitelli,L.F., Das,P.K., Yin,Y., Albu,M., Chen,X., Taipale,J., Hughes,T.R. and Weirauch,M.T. (2018) The human transcription factors. *Cell,* **175,** 598-599.

6. Bushweller,J.H. (2019) Targeting transcription factors in cancer - from undruggable to reality. *Nat. Rev. Cancer.,* **19,** 611-624.

7. Yu,C.P., Kuo,C.H., Nelson,C.W., Chen,C.A., Soh,Z.T., Lin,J.J., Hsiao,R.X., Chang,C.Y. and Li,W.H. (2021) Discovering unknown human and mouse transcription factor binding sites and their characteristics from ChIP-seq data. *Proc. Natl. Acad. Sci. U. S. A.,* **118,** e2026754118. doi: 10.1073/pnas.2026754118.

8. O'Geen,H., Frietze,S. and Farnham,P.J. (2010) Using ChIP-seq technology to identify targets of zinc finger transcription factors. *Methods Mol. Biol.,* **649,** 437-455.

9. Ouyang,Z., Zhou,Q. and Wong,W.H. (2009) ChIP-seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.,* **106,** 21521-21526.

10. Fan,J., Zeller,K., Chen,Y.C., Watkins,T., Barnes,K.C., Becker,K.G., Dang,C.V. and Cheadle,C. (2010) Time-dependent c-myc transactomes mapped by array-based nuclear run-on reveal transcriptional modules in human B cells. *PLoS One,* **5,** e9691.

11. Leplat,C., Nicaud,J.M. and Rossignol,T. (2018) Overexpression screen reveals transcription factors involved in lipid accumulation in yarrowia lipolytica. *FEMS Yeast Res.,* **18,** 10.1093/femsyr/foy037.

12. Menssen,A. and Hermeking,H. (2002) Characterization of the c-MYC-regulated transcriptome by SAGE: Identification and analysis of c-MYC target genes. *Proc. Natl. Acad. Sci. U. S. A.,* **99,** 6274-6279.

13. Dai,X., Chen,X., Fang,Q., Li,J. and Bai,Z. (2018) Inducible CRISPR genome-editing tool: Classifications and future trends. *Crit. Rev. Biotechnol.,* **38,** 573-586.

14. Vojta,A., Dobrinić,P., Tadić,V., Bočkor,L., Korać,P., Julg,B., Klasić,M. and Zoldoš,V. (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.,* **44,** 5615-5628.

15. Borys,S.M. and Younger,S.T. (2020) Identification of functional regulatory elements in the human genome using pooled CRISPR screens. *BMC Genomics,* **21,** 107-020-6497-0.

16. Canver,M.C., Smith,E.C., Sher,F., Pinello,L., Sanjana,N.E., Shalem,O., Chen,D.D., Schupp,P.G., Vinjamur,D.S., Garcia,S.P., et al. (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature,* **527,** 192-197.

17. Han,R., Li,L., Ugalde,A.P., Tal,A., Manber,Z., Barbera,E.P., Chiara,V.D., Elkon,R. and Agami,R. (2018) Functional CRISPR screen identifies AP1-associated enhancer regulating FOXF1 to modulate oncogene-induced senescence. *Genome Biol.,* **19,** 118-018-1494-1.

18. Kazimierska,M., Podralska,M., Żurawek,M., Woźniak,T., Kasprzyk,M.E., Sura,W., Łosiewski,W., Ziółkowska-Suchanek,I., Kluiver,J., van den Berg,A., et al. (2021) CRISPR/Cas9 screen for functional MYC binding sites reveals MYC-dependent vulnerabilities in K562 cells. *bioRxiv,* 2021.08.02.454734.

19. Kim,Y.W. and Kim,A. (2017) Deletion of transcription factor binding motifs using the CRISPR/spCas9 system in the β-globin LCR. *Biosci. Rep.,* **37,** BSR20170976. doi: 10.1042/BSR20170976. Epub 2017 Jul 20.

20. Korkmaz,G., Lopes,R., Ugalde,A.P., Nevedomskaya,E., Han,R., Myacheva,K., Zwart,W., Elkon,R. and Agami,R. (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotech,* **34,** 192-198.

21. Aach,J., Mali,P. and Church,G.M. (2014) CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes. *bioRxiv,* 005074.

22. Concordet,J.P. and Haeussler,M. (2018) CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.,* **46,** W242-W245.

23. Heigwer,F., Kerr,G. and Boutros,M. (2014) E-CRISP: Fast CRISPR target site identification. *Nat. Methods,* **11,** 122-123.

24. Montague,T.G., Cruz,J.M., Gagnon,J.A., Church,G.M. and Valen,E. (2014) CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.,* **42,** W401-7.

25. Bae,S., Park,J. and Kim,J.S. (2014) Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics,* **30,** 1473-1475.

26. Doench,J.G., Fusi,N., Sullender,M., Hegde,M., Vaimberg,E.W., Donovan,K.F., Smith,I., Tothova,Z., Wilen,C., Orchard,R., et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.,* **34,** 184-191.

27. Cavener,D.R. (1987) Comparison of the consensus sequence flanking translational start sites in drosophila and vertebrates. *Nucleic Acids Res.,* **15,** 1353-1361.