MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231069

Data Visualization of CRISPR-Cas9 Guide RNA Design Tools

Yatish Jain[a,b], Fathima Afra Mohamed Izzath[c], Laurence OW Wilson[a,b], Denis C Bauer[a,b,d,1]

[a] *Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, New South Wales, Sydney, Australia*
[b] *Department of Biomedical Sciences, Macquarie University, New South Wales, Sydney, Australia*
[c] *School of Science, RMIT University, Melbourne 3000, Australia*
[d] *Applied BioSciences, Faculty of Science and Engineering, Macquarie University, New South Wales, Sydney, Australia*

Abstract. With the advancement of genomic engineering and genetic modification techniques, the uptake of computational tools to design guide RNA increased drastically. Searching for genomic targets to design guides with maximum on-target activity (efficiency) and minimum off-target activity (specificity) is now an essential part of genome editing experiments. Today, a variety of tools exist that allow the search of genomic targets and let users customize their search parameters to better suit their experiments. Here we present an overview of different ways to visualize these searched CRISPR target sites along with specific downstream information like primer design, restriction enzyme activity and mutational outcome prediction after a double-stranded break. We discuss the importance of a good visualization summary to interpret information along with different ways to represent similar information effectively.

Keywords. Visualization, UX design, CRISPR-Cas9, gRNA, knock-out, design tools

1. Introduction

The discovery of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) revolutionized the field of genome editing and genetic engineering. With the ability to make breaks in DNA at precise locations, the applications of genome editing are enormous, including but not limited to, gene expression regulation, epigenetic modification, cell imaging, functional gene screening, therapeutic drug development, and gene diagnosis [1].

Guide RNA (gRNA) provides specificity (precise cleavage) to the CRISPR-cas9 (CRISPR-associated protein 9) system by binding to the targeted DNA via base complementarity. The CRISPR-Cas complex nuclease will then create a double-stranded break in the RNA/DNA hybrid complex. The cleaved DNA is subsequently repaired by different pathways leading to different outcomes I.e., knock-out or knock-in [2]. Non-homologous end joining (NHEJ pathway), an error-prone repair pathway that introduces mutations and disrupts gene function, leads to gene knock-out. Similarly, Homology

[1] Corresponding Author: Denis Bauer, email: Denis.Bauer@csiro.au.

directed repair (HDR) pathway uses the template DNA to repair cleavage and hence integrates a new sequence at a specific target called knock-in. Designing a custom gRNA with maximum on-target activity and minimum potential off-target activity is the first step of any genome editing experiment [3].

Visualizations and UX design are important elements for any data presentation. With so many case-specific CRISPR-Cas9 design tools and differing underlying algorithms, data visualization becomes even more important. Using the right set of visualizations to match the data characteristics, tasks and the intended audience is important to remove uncertainty from implied interpretations [4]. Here we present an overview of different visualizations used in guide RNA design tools, their advantages, their shortcomings, and the minimum set of visualization components needed to deliver results effectively.

2. Methods

Guide RNA design pipeline typically consists of three steps: Identify all CRISPR interference targets, filter targets based on scoring metrics/off-targets/orientation, and lastly, predict knockdown efficacies of mismatch gRNA [5]. In this visualization and UX design review we focus on the first two steps of identifying all targets and filtering out bad targets using different criteria.

These design tools let users narrow down the search space by identifying the best targets within user-specified locations. Thus, the main aim of the visualization for these design tools is to present a small number of optimal targets that the user can utilize in their genome editing experiments. A good visualization hence provides a summary of the derived information to give a go/no-go test point for a researcher designing genome editing experiments. For example, missed/incorrect strand information or off-target activity can result in expensive downstream experimentation without any benefits.

In general, while working with CRISPR-cas9 gRNA design tool, good visualization should minimally convey the following information (important features):

- Location of targets along with upstream and downstream sequence information.
- Targets strand/orientation.
- Represent the scoring metric used and visually represent better targets.
- Make visualization interactive to show more information about the targets.
- Demonstrate adherence to specific rules like Protospacer Adjacent Motif (PAM) sequence presence.

With the same objective of identifying candidate guide RNA sequences with high specificity and low off-target activity, these design tools use similar datasets. We used the same sequence string to design targets on different design tools. Differences in the underlying algorithms and scoring metrices were evident with differences in ideal candidates generated. Effective visualization can save a lot of time during the design process by providing a filtered view of efficient targets. CHOPCHOP [6], CRISPOR [7], E-CRISP [8], GT-SCAN [9], CRISPRSCAN [10], and CCTOP [11] all generate a filtered view to quickly select targets to examine further (Figures 1-6).

3. Results

CHOPCHOP (Figure 1), CRISPOR (Figure 2), GT-SCAN (Figure 4), and CCTOP (Figure 6) adhere to all 5 important features and display the scoring metric within visualization, which screens out poor targets easily. CHOPCHOP (Figure 1), and CCTOP (Figure 6) further improved the visualization by ranking the targets, which makes the better targets stand out. CRISPRSCAN (Figure 5) does not offer the important feature list but its interactions between visualization and table display make it nonetheless comparatively easy to identify better targets. In contrast, the dynamically created visualization by E-CRISP (Figure 3) does not convey additional information other than the coordinates of all targets, relying on the accompanying table to convey information. Interestingly, both CCTOP and E-CRISP use color coding and mini visualizations within the table, which makes important information stand out and easier to find.

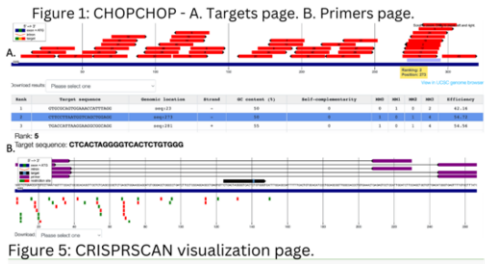

Figure 1: CHOPCHOP - A. Targets page. B. Primers page.

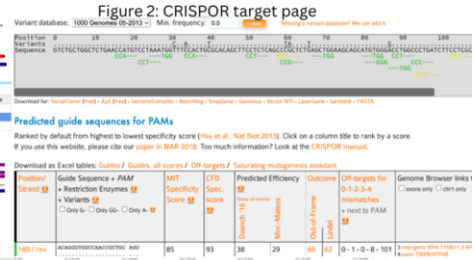

Figure 2: CRISPOR target page

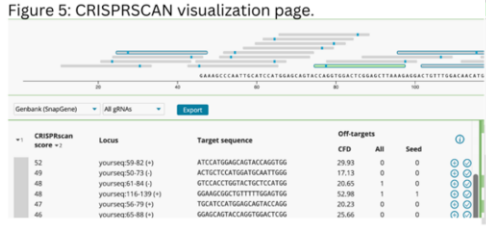

Figure 5: CRISPRSCAN visualization page.

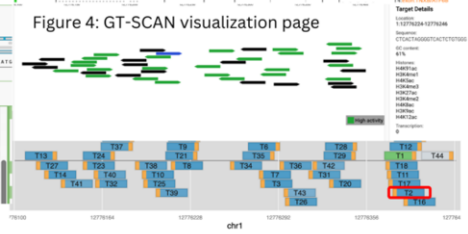

Figure 4: GT-SCAN visualization page

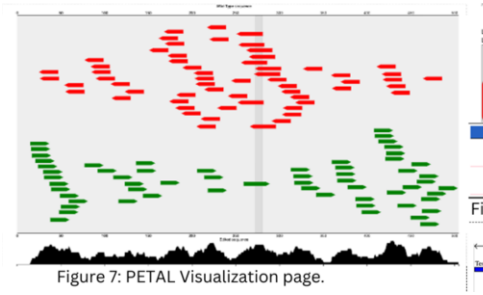

Figure 7: PETAL Visualization page.

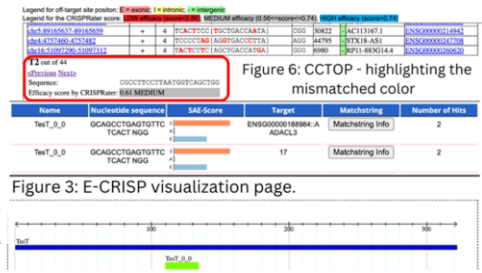

Figure 6: CCTOP - highlighting the mismatched color

Figure 3: E-CRISP visualization page.

Among the tools reviewed, CHOPCHOP and CRISPOR are the two tools that perform end-to-end analysis for guide RNA design. These tools first identify the best targets available within the specified region and then provide additional details like available primers, restriction enzyme activity, and target site location for each target. CHOPCHOP uses a very effective primer design visualization for targets where available primers are depicted on either side of the target and the restriction enzyme activity along with target site location are color coded with legend. CRISPOR provides additional

details like cloning and expression of guide RNA, PCR to amplify on target site, Restriction sites for PCR product etc. in a free-form text format for each target. However, when exploring several targets this information can become overwhelming. Comparing CHOPCHOP's primers design visualization with CRISPOR free-form text underlines how crucial effective visualization is for summarizing complex information.

Similarly, an error in the visualization can quickly lead to incorrect inferences. For example, CCTOP uses the CRISPRater score to rank guides color coded with red, grey and blue to represent low, medium and high efficiency, respectively. Figure 6 highlights a discrepancy between the CRISPRater score of 0.61 in the table, classifying the target as a medium (grey) efficiency, compared to the corresponding visualization, which shows a blue target representing high efficiency. Users only looking at the visualization would have considered target T2 as a high efficiency target.

4. Discussion

The main challenge with the visualization of gRNA design tools is to depict all the available guides in a specified search space. The ideal visual space for guides should specify coordinates and sequence information. To allow visualization of big search spaces, guides can be placed on a zoomed-out pane with coordinates that display sequence information after zooming in. Such visualization panes are available via a feature-viewer package [12] or can be customized using D3 libraries [13]. For example, PETAL [14] (Figure 7) identifies all valid targets around the edited sequence and then allows users to interact with the valid targets to generate oligo and adapter sequences for downstream prime editing experiments. PETAL displays coverage of targets within sequence region in zoom and pan panel. Zooming onto a specific area also shows the sequence within that region.

Once all the guides are presented in the visualization, the next challenge is to filter targets using scoring metrics visually. Scoring metrics can be used to color code all guides within the visualization along with directional arrows representing orientation. Another layer of data can be added to these guides by creating a text window while hovering over any guide. The hover-over text can contain off-target information, scoring metric score, GC content etc. Visualizing downstream information like target site location (intron/exon/intergenic), restriction enzyme activities and primers availability present further challenges. This information needs to be presented within the visualization, both upstream and downstream of the selected target. The available primers on either side of the target can be presented similar to available guides with orientations. Target site locations and restriction enzyme sites can be included within this visualization using color-coded scheme with legend.

Although visual aids assist in making inferences quickly, errors/mismatches in visualization as shown in CCTOP's Figure 6 can lead to downstream challenges. It is important to test visualizations while developing to make sure there is no data loss or misinformation presented. The use of visualization regression testing [15], as well as multiview visualization [16] should be incorporated as a part of the design process to avoid visualization mistakes.

5. Conclusions

Visualizations are important to precisely convey insights generated from any dataset. Specifically, for the human genome where the search space is vast, an efficient visualization of CRISPR-cas9 gRNA design tools aids the ideal target selection reducing the time, effort, and potential for error considerably. Incorporating 5 important visualization features would further streamline the design process. In general, visualizations are widely used and there are a range of libraries to make displaying items on a genomic scale easily possible. However, more testing and quality control measures need to be implemented to ensure high quality insights can be generated.

References

[1] Li H, Yang Y, Hong W, Huang M, Wu M, Zhao X. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. Signal Transduct and Target Ther. 2020 Jan;5(1):1, doi:10.1038/s41392-019-0089-y.

[2] Kim H, Kim JS. A guide to genome engineering with programmable nucleases. Nat Rev Genet. 2014 May;15(5):321-34, doi:10.1038/nrg3686.

[3] Wilson LOW, O'Brien AR, Bauer DC. The current state and future of CRISPR-Cas9 gRNA design tools. Front Pharmacol. 2018 Jul;9:749, doi: 10.3389/fphar.2018.00749.

[4] Weiskopf D. Uncertainty Visualization: Concepts, Methods, and Applications in Biological Data Visualization. Front Bioinforma. 2022;2, doi: 10.3389/fbinf.2022. 793819.

[5] van Gestel J, Hawkins JS, Todor H, Gross CA. Computational pipeline for designing guide RNAs for mismatch-CRISPRi. STAR Protoc. 2021 Jun;2(2):100521, doi:10.1016/j.xpro.2021.100521.

[6] Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 2019 Jul;47(W1):W171-4, doi: 10.1093/nar/gkz365.

[7] Concordet JP, Haeussler M. CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 2018 Jul;46(W1):W242-5, doi:10.1093/nar/gky354.

[8] Heigwer F, Kerr G, Boutros M. E-CRISP: Fast CRISPR target site identification. Nat Methods. 2014 Feb;11(2):122-3, doi:10.1038/nmeth.2812.

[9] O'Brien A, Bailey TL. GT-Scan: Identifying unique genomic targets. Bioinform. 2014 Sep;30(18):2673-5, doi: 10.1093/bioinformatics/btu354.

[10] Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ. CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 2015 Oct;12(10):982-8, doi:10.1038/nmeth.3543.

[11] Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. PLoS One. 2015 Apr;10(4):e0124633, doi: 10.1371/journal.pone.0124633.

[12] Paladin L, Schaeffer M, Gaudet P, Zahn-Zabal M, Michel PA, Piovesan D, Tosatto SC, Bairoch A. The Feature-Viewer: A visualization tool for positional annotations on a sequence. Bioinform. 2020 May;36(10):3244-5, doi:10.1093/bioinformatics/btaa055.

[13] Bostock M, Ogievetsky V, Heer J. D3 data-driven documents. IEEE Trans Vis Comput Graph. 2011 Nov;17(12):2301-9, doi: 10.1109/TVCG.2011.185.

[14] Adikusuma F, Lushington C, Arudkumar J, Godahewa GI, Chey YCJ, Gierus L, Piltz S, Geiger A, Jain Y, Reti D, Wilson LO. Optimized nickase- and nuclease-based prime editing in human and mouse cells. Nucleic Acids Res. 2021 Oct;49(18):10785-95, doi: 10.1093/nar/gkab792.

[15] Opmanis R, Kikusts P, Opmanis M. Visualization of Large-Scale Application Testing Results. Balt J Mod Comput. 2016;4(1).

[16] Hammad M, Otoom AF, Hammad M, Al-Jawabreh N, Seini RA. Multiview visualization of software testing results. Int J Comput Digit Syst. 2020 Jan;9(1), doi:10.12785/ijcds/090105.