Western Washington University
Western CEDAR

Scholars Week | Conferences and Events

May 2019

NucJuke: A web tool for re-ranking CRISPR-Cas9 gRNAs based on chromatin accessibility in yeast

Gaea Turman
Western Washinton University

Follow this and additional works at: https://cedar.wwu.edu/scholwk

Part of the Higher Education Commons

Turman, Gaea, "NucJuke: A web tool for re-ranking CRISPR-Cas9 gRNAs based on chromatin accessibility in yeast" (2019). *Scholars Week*. 54.
https://cedar.wwu.edu/scholwk/2019/2019_poster_presentations/54

This Event is brought to you for free and open access by the Conferences and Events at Western CEDAR. It has been accepted for inclusion in Scholars Week by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.

NucJuke: A web tool for re-ranking CRISPR-Cas9 gRNAs based on chromatin accessibility in yeast

Gaea Turman and Dan Pollard
Department of Biology, Western Washington University, Bellingham WA 98225

WESTERN WASHINGTON UNIVERSITY

NucJuke

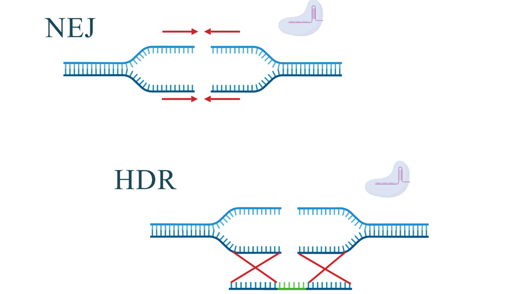
INTRODUCTION

NucJuke is a supplementary gRNA design tool for specific targeting of nucleosome free sites in CRSIPR-Cas9 experiments in yeast.

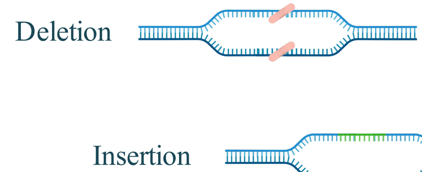
CRISPR-Cas9 is a system for high fidelity genome editing that utilizes: (1) a Cas9 enzyme that can make double stranded breaks to DNA, and (2) a guide RNA (gRNA) that directs Cas9 to a specified location in the genome.

CRISPR has a broad range of applications (Fig. 1) but is not 100% successful at all genomic locations. As a result, researchers often have to design multiple gRNAs to achieve experimental success. A variety of papers have suggested that chromatin accessibility (i.e. how open the DNA is) could have significant influence over the success of CRISPR-Cas9 experiments in *S. cerevisiae* (Wyrick et al., 2015). These have documented that nucleosome presence at a target site (which renders DNA largely inaccessible) can physically prevent Cas9 from making its double stranded break (Fig. 2). As very few gRNA design tools take nucleosome positioning into account, we have created NucJuke to improve the efficiency of this type of experiment in yeast (Hinz et. al, 2016, Horlbeck et. al, 2016). Presented here is data collected from the programmatic validation of NucJuke.

1. Association of gRNA with target DNA, double stranded break produced via Cas9.



2. Non-homologous end joining (NEJ) or homology directed repair (HDR) to insert or delete a sequence at the break site.



3. Desired deletion or insertion completed at targeted location.

Figure 1. A depiction of how CRISPR-Cas9 accomplishes targeted mutagenesis. In a CRISPR experiment DNA can be inserted or removed at a specified location through homology directed repair (HDR) or non-homologous end joining (NEJ) respectively.

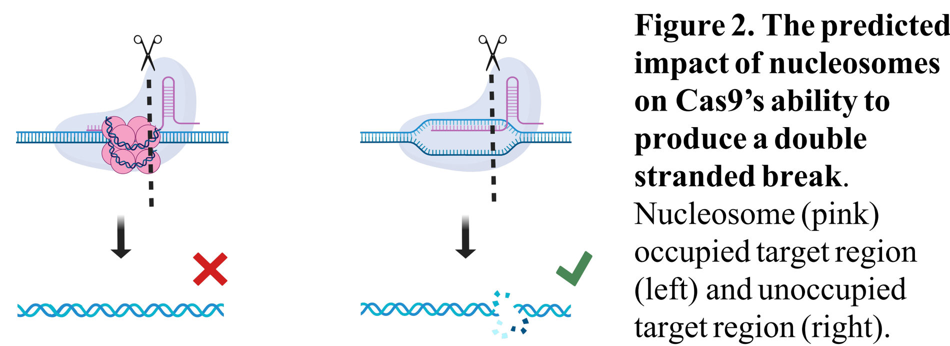


Figure 2. The predicted impact of nucleosomes on Cas9's ability to produce a double stranded break. Nucleosome (pink) occupied target region (left) and unoccupied target region (right).

SORTING ALGORITHM

NucJuke utilizes two, single base pair resolution nucleosome positioning maps generated by the Widom lab (Widom et. al, 2012):

- a unique map (in which nucleosomes aren't allowed to have overlap by more than 40bp)
- a redundant map (in which nucleosomes are allowed to overlap arbitrarily)

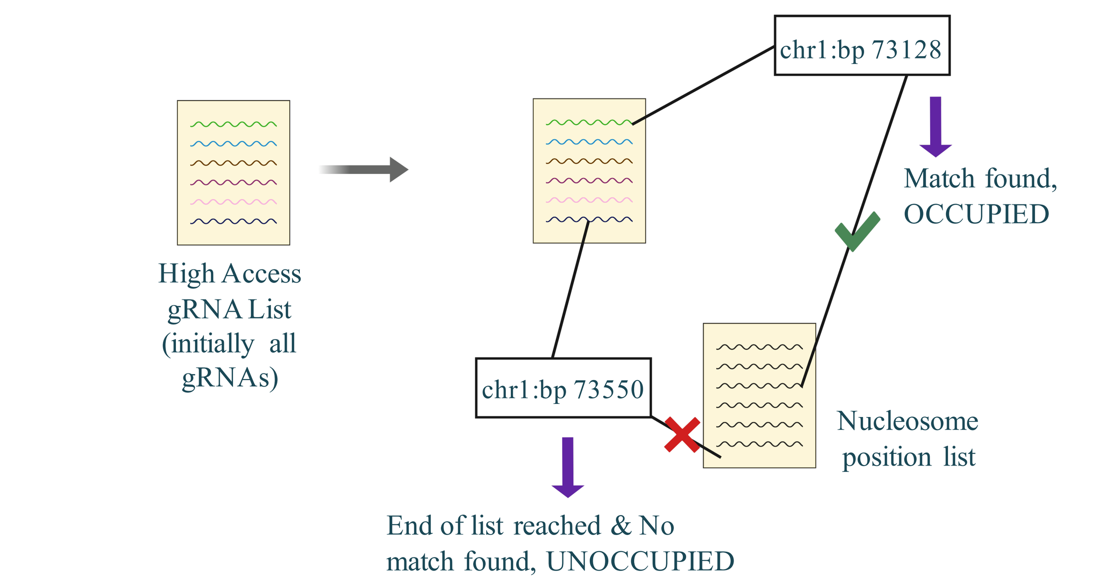
Each of these has information about nucleosomes position and the strength of their association as represented by a nucleosome center positioning (NCP) score corrected against background noise (NCP score: noise ratio).

NucJuke has three sorting settings that apply these nucleosome maps in different ways:

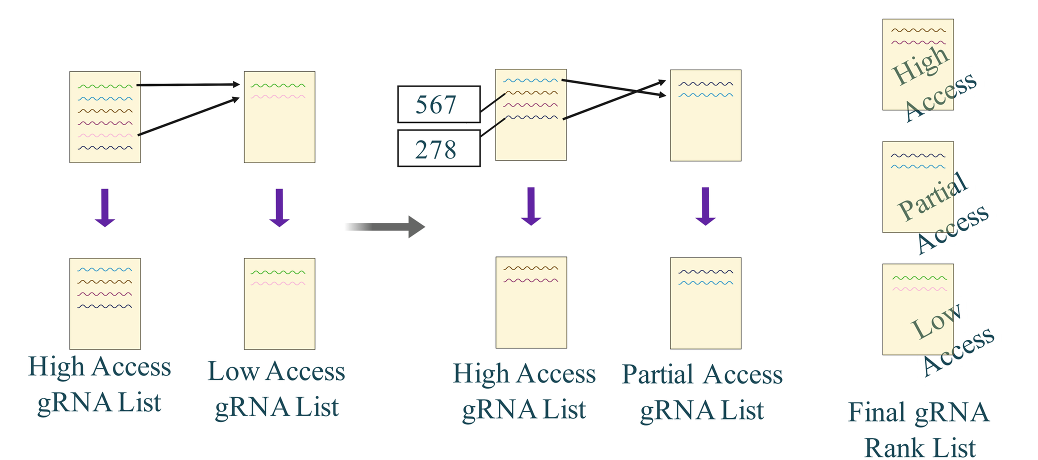
(1) Dynamic sorting – treats the nucleosomes in the unique map as fully occupying and treats those in the redundant map as partially occupying. First, gRNAs with positions overlapping with fully occupying nucleosomes from the unique map are sorted (maintaining original order), then gRNAs overlapping with the partial nucleosomes map are sorted according to the product of their NCP score:noise ration and original rank (Fig. 3).

(2) Strict sorting – treats the nucleosomes in both maps as fully occupying and resorts them according to original rank.

(2) Lenient sorting – uses only the strict nucleosome map to resort gRNAs and resorts them according to original rank; it is the most permissive sorting method.



Check gRNA positions against known nucleosome positions, if there is a match then the gRNA is partially or fully occupied (has partial or low accessibility).



If matched against the strict nucleosome map, maintain design order and categorize as lowly accessible (left). If matched against the lenient nucleosome map, reorder based on the total occupancy score and categorize as partially accessible (right).

Figure 3. Pictographic representation of algorithm scoring/dynamic re-ranking mechanism. Low access gRNAs (overlapping with strict map nucleosomes) are pulled first and then partial access gRNAs (overlapping with lenient map nucleosomes) are pulled/re-ranked.

TOOL VALIDATION

gRNAs designed for 5001 genes => re-ranked using dynamic sorting algorithm

Two data sets collected:

(1) For all gRNAs designed for a particular gene, the proportion of gRNAs that were partially occupied and fully occupied

(2) For the gRNAs originally ranked 1-10, the number of gRNAs that were partially occupied and fully occupied (Fig. 4).

<u>Using this information allowed us to make some assessment of the frequency with which using NucJuke would actually matter to a researcher – if we're only ever re-ranking 1 originally top 10 gRNA, the tool would have less utility than if we frequently re-ranked 3-4 originally top 10 gRNAs.</u>

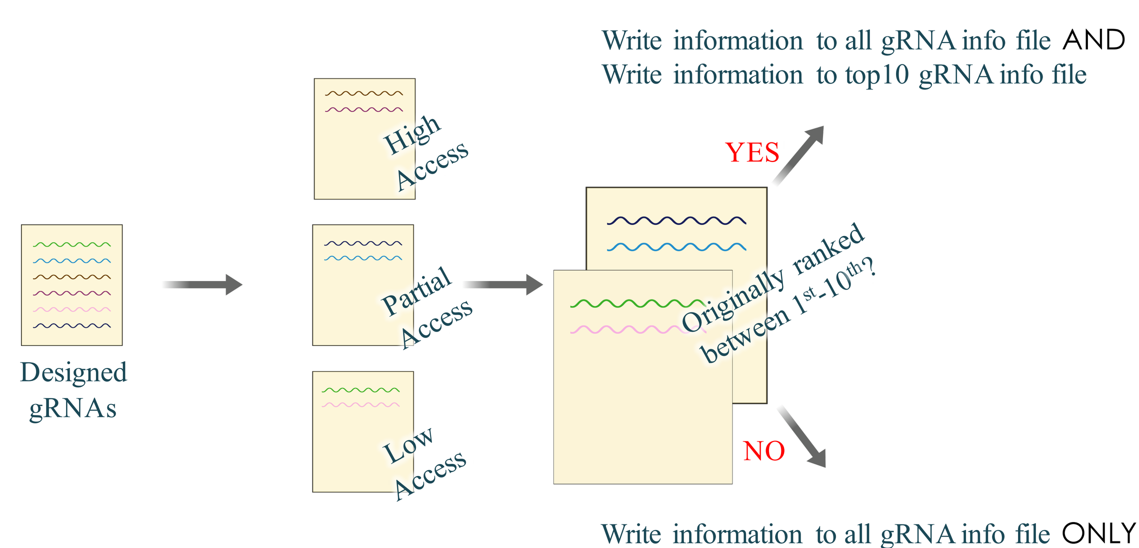


Figure 4. Pictographic representation of programmatic validation. For every coding gene in *S. cerevisiae*: (1) Design gRNAs (2) Re-rank gRNAs dynamically (3) For partial and low accessibility gRNAs, tally the number of gRNAs originally ranked between one and ten AND tally the total number of partially/fully occupied gRNAs (for all gRNAs).

DATA & ANALYSIS

The Results:

- Most gRNAs fall into the low accessibility category (across all genes including top 10s), least gRNAs fall into the highly accessible category
- Interquartile ranges for proportion of gRNAs in all categories show very tight spread of data

The programmatic analysis for NucJuke suggests high utility for researchers completing CRISPR-Cas9 experiments in yeast. gRNAs are sorted into the lowly or partially accessible categories more often than they are into the highly accessible category, and this is consistent for originally top 10 ranked gRNAs. These would be the most likely gRNAs used by an experimenter originally, so if they are actually low/partial accessibility, it suggests our tool might have an appreciable impact on the work of other researchers doing CRISPR-Cas9 experiments.

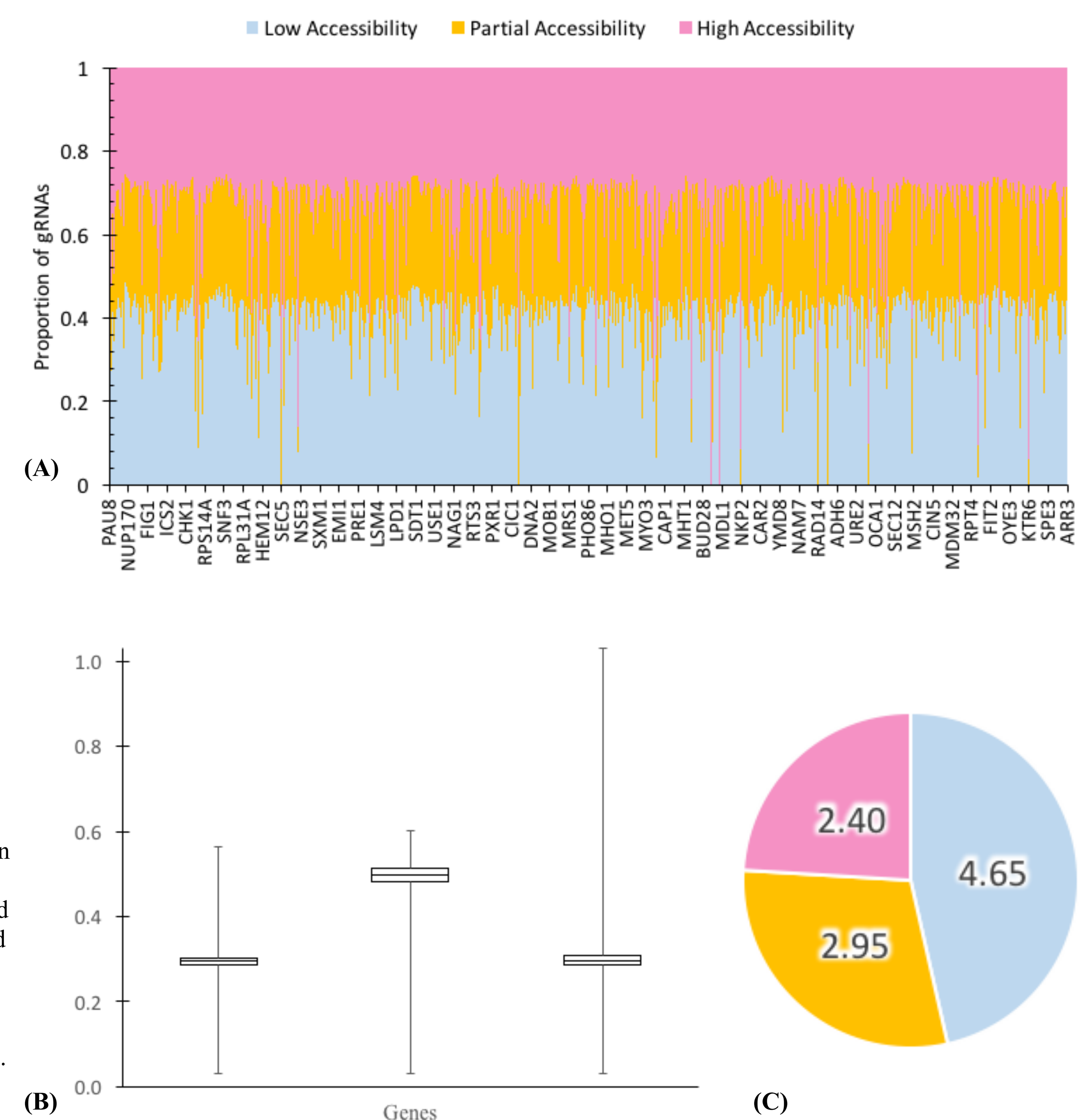


Figure 5. Analysis of top10 and all gRNA data. (A) Relative proportion of total gRNAs falling into each accessibility category for all genes, x-axis depicts every 100th gene name. Mean proportions for low, partial and high accessibility respectively were 0.459, 0.264, and 0.277. **(B)** Box and whisker plot for proportion of gRNAs falling into each accessibility category for all genes, interquartile ranges for partial, low, and high accessibility respectively were 0.015, 0.032, and 0.021. The minimum across all categories was 0, and the maximums were 0.53, 0.57, and 1.00. **(C)** Pie chart depiction of average number of top10 gRNAs falling into each accessibility category.

FUTURE WORK

Summer Wet Lab Validation, Procedure from Ryan et. al (2016):

1. Design and order oligos for gRNA targets (completing five experiments for each accessibility category)
2. Construct gRNA plasmids (15 total)
3. Perform transformation experiments and determining transformation efficiencies (TEs)

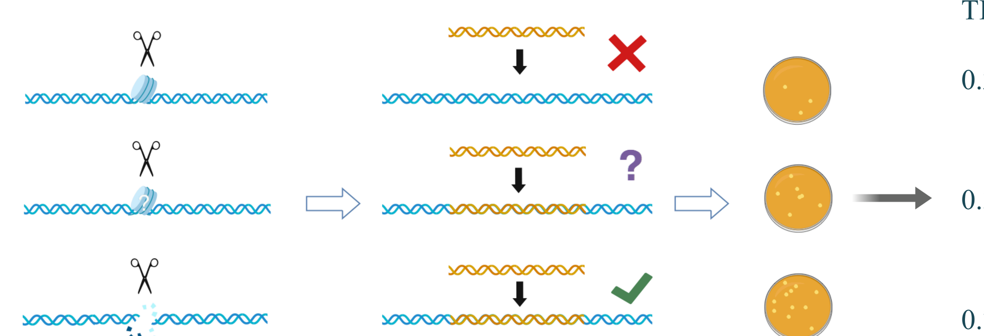


Figure 6. Diagram of wetlab experimental workflow. Expected transformation efficiency (TE) will be baselined against transformation experiments in which CRISPR-Cas9 isn't used.

- For each accessibility level, complete transformations and determine if NucJuke helps improve experimental efficiency. Will be attempting to insert antibiotic resistance gene that will confer ability to grow on selective media, e.g. when insertion is completed successful, colonies will grow on an antibiotic plate
- As we expect Cas9 (represented by scissors in Fig. 2, left) to be unable to make double stranded breaks where there are strongly associated nucleosomes, we anticipate seeing fewer transformed cells (colonies) when we plate our low accessibility gRNA target transformants on selective media.
- For highly accessible gRNA target transformants, would expect Cas9 will be able to cut/gene would be sucessfully inserted resulting in more colonies growing on selective media.
- In the absence of CRISPR-Cas9, some cells will still experience gene insertion so the difference between these experimental conditions will be baselined against a treatment in which CRISPR is absent.

WORKS CITED / ACKNOWLEDGEMENTS

1. Widom, Jonathan, et al. "A Map of Nucleosome Positions in Yeast at Base-Pair Resolution." Nature News, Nature Publishing Group, 3 June 2012
2. Chen, Yunru, et al. "Using Local Chromatin Structure to Improve CRISPR/Cas9 Efficiency in Zebrafish." PLOS ONE, Public Library of Science, Aug. 2017.
3. Hinz, John M., et al. "Nucleosomes Impede Cas9 Access to DNA in Vivo and in Vitro." Journal of Biological Chemistry, 25 Nov. 2016.
4. Horlbeck, Max A, et al. "Nucleosomes Impede Cas9 Access to DNA in Vivo and in Vitro." ELife, ELife Sciences Publications Limited, 17 Mar. 2016.
5. Ryan, Owen W, et al. "CRISPR-Cas9 Genome Engineering in Saccharomyces Cerevisiae Cells." Cold Spring Harbor Laboratory., U.S. National Library of Medicine, 1 June 2016
6. Wyrick, John J., et al. "Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro." ACS Publications, 8 Nov. 2015.

…and a special thanks to Chayse Jones for helping NucJuke find its name

<u>*Work Supported by National Science Foundation Award MCB-1518314 to Dan Pollard</u>