Stony Brook University

Research Proposal

Frederick Brunn
SBU ID: 108724486

2017-03-30
CSE 300

Optimizing sgRNA-specific SpCas9-HFx specificity in CRISPR-Cas nucleases

CRISPR (clustered, regularly interspaced, short palindromic repeat) is a genome-editing technique utilizing a class of novel RNA-guided proteins. Previous genome-editing tools have tended to rely on the protein's structure solely in order to identify the location at which to induce mutation, which limited their use as they required significant redesign to target different locations. CRISPR proteins bind to sites matching an arbitrary RNA string provided it, allowing for faster and more easier alteration. By contrast, the procedure only requires the design of an RNA string in order to determine its target sites and modifications. CRISPR editing has the potential to enable cost-effective genetic treatments that are currently infeasible for deployment in healthcare.

Our new research aims to continue research into mapping a prominent family of CRISPR protease variants, the SpCas9-HFx family as first explored by Kleinstiver et al. In their article posed SpCas9-HF1 as "a high-fidelity variant harbouring alterations designed to reduce non-specific DNA contacts." (Kleinstiver et al. 490) We aim to explore the orthologues of SpCas-HF1 via performing genome-scale knockout testing, as per Shalem et al., and comparing the specificity achieved by use of each of the chosen orthologues in order to determine the ideal choice for this procedure. We believe that this will form the basis of a practical understanding of the reasons for differentiation between the SpCas-HFx orthologues' efficacy, as well as the basis of a predictive model to choose the optimal SpCas-HFx orthologue for a novel procedure.

Our proposed research relies on a great deal of previously performed CRISPR research. The application we propose to use was first accomplished in 2014 by Shalem et al., who used a

mainstream CRISPR variant to accomplish genome-scale knockout screening. Consistency was measured, with CRISPR falling behind the established RNA-interference method on all counts: “First, we plotted the P values for the top 100 hits using either RIGER (Fig. 4A) or redundant siRNA activity (RSA) (fig. S9) scoring. Lower P values for the GeCKO versus shRNA screen indicate better scoring consistency among sgRNAs. Second, for the top 10 RIGER hit genes, 78 ± 27% of sgRNAs targeting each gene ranked among the top 5% of enriched sgRNAs, whereas 20 ± 12% of shRNAs targeting each gene ranked among the top 5% of enriched shRNAs (Fig. 4B).” (Shalem et al. 86) Genome-scale knockout testing represents a potential medical application, whose improvement would apply to potentially many medical uses.

The high-fidelity CRISPR family was first designed for the heightened specificity achieved by the designed protease, SpCas9-HF1. The researchers remarked on the notable specificity of the protease: “The SpCas9-HF1 variant characterized in this report reduces all or nearly all genomewide off-target effects to undetectable levels as judged by GUIDE-seq and targeted next generation sequencing...” (Kleinstiver et al. 493) They go on to note that their design approach might be extensible to further orthologues within the HF family: “Overall, our results demonstrate that the approach of mutating non-specific DNA contacts is highly effective at increasing SpCas9 specificity and suggest it might be extended to other naturally occurring and engineered Cas9 orthologues [38, 39, 40, 41, 42], as well as other CRISPR-associated nucleases [43, 44].” (Kleinstiver et al. 494) This possibility forms the basis of our experimentation- we aim to repeat their process to create other high fidelity Cas9 orthologues and compare their overall effectiveness.

The experiment will primarily take place in an appropriately stocked sterile biological laboratory environment. HF orthologues will be synthesized and altered in novel ways that

follow logically from the changes made to define SpCas9-HF1 in Kleinstiver et al., as well as two controls: SpCas-HF1 and a mainstream CRISPR variant. 40 GeCKO experiments per orthologue will be prepared, in order to provide a meaningful sample size without requiring undue expenditure of laboratory resources. The GeCKO experiments will be run without any further major modification from the procedure in Shalem et al. The experiments' results will then be analyzed using the statistical method proposed in Shalem et al., ensuring a fair and quantifiable comparative analysis between experimental results by orthologue. Finally, we will analyze the structure of the off-site mutations on selected samples, searching for patterns between different orthologues' performance at different areas, so that we may determine some basis by which to choose a high-fidelity variant for a given sgRNA sequence while optimizing for specificity.

The high-fidelity CRISPR family promises great improvements in specificity when used instead of a mainstream CRISPR variant, yet the family remains tragically unexplored. Kleinstiver et al. found that while SpCas9 acted on 2-25 off-target sites on seven of the eight samples (the last of which being a control sample chosen for its low likelihood of SpCas9 errors), SpCas9-HF1 produced an error on only one sample of eight, and matched the control sample activity otherwise. (Kleinstiver et al. 492) This represents an improvement in specificity on the order of multiple scales of magnitude. Yet a review of CRISPR research reveals that nothing more has been explored beyond the first high-fidelity variant by Kleinstiver.

Our research aims to explore the other high-fidelity CRISPR variants and compare them in genome-scale knockout testing. Although this proposed research measures the specificity for just one CRISPR application, it is hoped that detailed sample analysis can result in a better understanding of the orthologue-specific behavior of the high-fidelity CRISPR family and form

the basis of a predictive model to determine which orthologue will have highest specificity for a given change. Moreover, as different variants will have different affinities and specificities for different desired changes (Kleinstiver et al. 494) possession of a wide and diverse array of CRISPR-based tools will make it more likely that a given change will have a high specificity method by which to be executed. Once it is possible to understand and design the change that needs to be made with CRISPR, the only remaining work to be done is in developing highly specific tools that can execute the change effectively- as such, this research has the potential to finally push CRISPR into the realm of *in situ* human experimentation.

Works Cited

Shalem, Sanjana, et al. “Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells.” *Science*, vol. 343, Issue 6166, 03 Jan 2014, pp. 84-87., doi: 10.1126/science.1247005. Accessed 24 Mar. 2017.

Kleinstiver, Benjamin P., et al. “High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects.” *Nature*, vol. 529, 28 January 2016, pp. 490-495., doi:10.1038/nature16526. Accessed 30 Mar. 2017.