Title

Split Luciferase Biosensors for Detection and Imaging of DNA Sequences and Chromatin Loops in Individual Living Cells

Author

Heath, Nicholas George

Publication Date

2023

Peer reviewed|Thesis/dissertation

Split Luciferase Biosensors for Detection and Imaging of DNA Sequences and Chromatin Loops in Individual Living Cells

By

NICHOLAS GEORGE HEATH

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Dr. David Segal, Chair

_____

Dr. Justin Siegel

_____

Dr. Joel Mackay

Committee in Charge

2023

Dedication

First and foremost, I would like to dedicate this dissertation to my major professor, Dr. Dave Segal, who has been the best mentor any graduate student could ask for. I continuously drew inspiration from Dave's immense curiosity and the joy of seeking and searching for answers to many questions with him via experiments. He is also highly empathetic and level-headed, attributes that have made me feel very supported during graduate school and have encouraged me to strive to be the best I can be.

I would also like to devote this dissertation to my family and to my girlfriend, Devon. Without their support over the course of my time in graduate school, this process would have been much more difficult. Specifically, I would like to thank my dad, Doug, who inspired me to ask questions about the world, spurred my interest in math and science, and galvanized me to pursue these fields from a young age. I would also like to dedicate this work to my mom, Rebecca, who, in the spirit of the Finnish term of my forebears on my maternal side, "sisu," (meaning resolve, persistence, dedication, or tenacity), encouraged me to persevere whenever I felt down. I also continually draw inspiration from her creative writing pursuits. I would also like to thank my brothers Jackson and Luke, who continually support me in everything I do and are my best friends. Finally, I would like to dedicate this work to my incredibly beautiful and exceptionally talented girlfriend, Devon, for supporting and encouraging me along this journey. I couldn't have gotten to this point without all of you.

Table of Contents

Abstract to the Dissertation

A commonly overlooked problem associated with current procedures in molecular biology that seek to gain information about the state of the genome, epigenome, transcriptome, proteome, or metabolome within a particular cell state is the necessary destruction of cell populations to extract genomic material. If a particular cell population is critical or valuable, this is a non-ideal reality of obtaining desired data. While highly sensitive single cell methodologies have evolved in recent years, many procedures still analyze bulk cell populations, often containing a high degree of complexity and heterogeneity. In addition, the information gathered from these types of multiple-omics studies—even at the single cell level—is not gathered in real-time, but after the fact as a snapshot of a particular cell state. One approach to solving these issues with current molecular biology multiple-omics techniques is to use biosensing technology. Biosensors are analytical devices that detect a specific target molecule within a heterogeneous molecular background. These tools operate in real-time or near real-time conditions to provide rapid detection capabilities for a wide range of molecular targets noninvasively (*in situ*) with high sensitivity and specificity.

Our approach to developing a new generation of live cell biosensors is based on a split reporter system, where a signal-producing protein is truncated into distinct segments which have very little to no activity individually but can recombine at a specific interface to produce a full-length, functional version of the protein. To this end, we have developed split luminescent reporter biosensors for both individual DNA sequences and chromatin loops anchored by cohesin and CTCF in living cells. Our platform for development of these biosensors has involved a

combination of *in silico* rational design based on an independently optimized truncation of NanoLuc luciferase (Binary Technology of NanoLuc luciferase or NanoBiT system) and catalytically-inactive Cas9 (dCas9) enzymes from *Streptococcus pyogenes and Staphylococcus aureus* guided to bind target DNA and chromatin loop anchor regions via single guide RNAs (sgRNAs), functional testing in a variety of cell lines, and statistical analysis of signal-to-background data for comparison between on-target and background conditions and determination of sensitivity and specificity. This split reporter platform can be extended to other fluorescent, luminescent, and enzymatic reporters and energy transfer-based systems to yield a powerful modular design, construction, and validation methodology for live cell biosensors. In addition, facilitated by the use of additional biorecognition elements in biosensor design, we envisage application of our live cell biosensing platform to detection of additional molecular targets including various epigenetic marks, other nucleic acids such as RNA or R-loops, and small molecules or ligands. If applied *in vitro*, this platform could feasibly be used to detect pathogen genomic material and environmental contaminants, aiding in monitoring and screening efforts to improve public health across a wide range of fields.

Chapter 1: Introduction to Biosensors as Sensitive Probes for Various Molecular Targets in Living Cells

Introduction to biosensing technology and the unique niche filled by nucleic acid biosensors

In recent years, biosensors have emerged as powerful tools for detecting and quantifying biological molecules with high sensitivity and selectivity, providing valuable insights into the current state of a biological system of interest. Biosensing systems hold great promise for food safety and control, environmental monitoring, biodefense, drug discovery, biomedical diagnosis, and monitoring of treatment and disease progression[1]. Within the medical field, biosensors have immense potential to transform disease diagnosis and to inform and expedite treatment methods. While research on biosensors has spanned several decades, it has gained significant traction in recent years due to increased focus on the versatility and cost-effectiveness of biosensors as analytical devices and on development of next generation biosensing systems. As the demand for accurate and rapid detection methods continues to grow across several fields, so too does the need for further development of innovative next generation biosensor technologies that improve on current design elements to create more user-friendly, accurate, and sensitive platforms.

Specifically, the need for more reliable and sensitive biosensors targeting nucleic acids has become increasingly apparent as society faces complex challenges including emergent infectious diseases like COVID-19, food safety and contamination risks, and invasive pest risks arising from highly interconnected global trade networks. In addition, driven by recent advances in genome engineering and genomics, demand for personalized medical treatments has increased rapidly over the last few years, and future personalized treatments will require extremely precise and

reliable nucleic acid biosensors for rapid disease diagnosis prior to determination of appropriate treatment recommendations. The ability to detect and monitor the state of a biological system such as a living cell in real-time allows for instantaneous decisions and swift actions to save time and resources in many different contexts. For instance, when presented with a patient displaying symptoms of an unknown illness, employing a DNA or RNA biosensor on that patient's saliva or mucus could allow a healthcare professional to detect the broader family or genus classifications of a particular emergent virus subtype and to decide whether to prescribe a certain antiviral or to simply quarantine the patient. Likewise, deploying DNA or RNA biosensors in the agricultural industry at the first signs of a crop disease outbreak could allow for rapid isolation and treatment of the contaminated agricultural field to avoid infection of additional crops. Furthermore, the agricultural industry could apply DNA or RNA biosensors to screen for the presence of new invasive pests that could pose a threat of crop destruction. Finally, within the genome editing field, employing live cell DNA biosensors could theoretically enable rapid, noninvasive detection of gene edits during *ex vivo* editing procedures, such as during *ex vivo* editing of human hematopoietic stem and progenitor cells (HSPCs). Validating these edits noninvasively and instantaneously directly in the HSPCs post-treatment could allow for rapid progression to downstream steps in the procedure, such as reinfusion of the edited cells back into a patient. Such a live cell DNA detection method could then be extended beyond this specific gene editing application to allow for the direct biosensing of any user-defined sequence at single copy with single cell resolution.

Biosensor structure and design

The first biosensor was developed over 55 years ago to monitor blood glucose levels[1]. This novel device combined glucose oxidase with an amperometric oxygen sensor[2]. Since this seminal achievement in the biosensing field, biosensor technology has advanced to the point where it is routinely used in a plethora of diagnostic applications, specifically toward point-of-care analysis of biomarkers[3-6]. Biosensors are analytical devices that detect a specific target molecule within a heterogeneous molecular background by converting a biological or physicochemical signal into an electrical signal[7-8]. Broadly, in the search for answers to questions relating to the quantification of levels of various molecules in cells, biosensors represent attractive alternatives to current *in vitro* methods for quantifying levels of molecules of interest. This is because biosensors offer a sensitive, rapid, and inexpensive way to detect levels of cellular targets in real-time in live cells[9-10]. Biosensors are composed of a bioreceptor element that interacts with a molecular target or bioanalyte, a transducer element that produces a measurable signal due to this interaction, and a reader device that converts the transducer signal into a practical readout for the presence of the target for interpretation by an end user[7-10]. There are a number of bioreceptor element types typically used to target various bioanalytes both *in vitro* and in live cells including natural elements such as antibodies[11], enzymes[12-14], nucleic acids[15-19], and aptamers[11,20] or synthetic elements such as molecularly imprinted polymers (MIPs)[21] or nanostructured binding surfaces such as those present in nanozymes[22-23]. Biosensing signals can be optical (colorimetric, fluorescence, chemiluminescence, or surface plasmon resonance), mass-based (magnetoelectric or piezoelectric), or electrical (voltammetry, impedance, or capacitance) in nature depending on the transducer element selected in the design process[8-9].

Alternative *in vitro* methodologies to DNA biosensing

Currently, there are a wide range of *in vitro* applications for DNA biosensors, such as detection of specific environmental targets and foodborne pathogens[24-26] as well as several medical biosensing applications such as the detection of circulating tumor DNA (ctDNA)[27-29]. In the field of DNA diagnostics, there are many alternative methods to biosensing platforms that allow for quantification of specific DNA sequences, but these methods have mostly been applied *in vitro*—either on non-living samples or samples prepared by extracting DNA from living cells. These methods are typically not applied in real-time to detect DNA in living cells, but to report on the presence of certain DNA sequences retrospectively *in vitro*. In addition, these methods are destructive and invasive if applied to target cell populations of interest to quantify or monitor levels of specific DNA sequences in these populations because they necessitate destruction of at least some portion of a cell population to analyze the DNA content. One of the most commonly used *in vitro* DNA quantification methods is quantitative PCR (qPCR)[30-31]. While qPCR has certainly revolutionized *in vitro* DNA diagnostics with its ability to detect very small quantities of a specific DNA sequence, it requires careful purification procedures that are not amenable to application in point of care (POC) devices[32]. In recent years, many novel next-generation *in vitro* biosensing methods for DNA detection have been developed based on the CRISPR/Cas effector system[33], including methods which take advantage of specific or promiscuous cleavage activities of Cas9 such as SHERLOCK[34-35], SHERLOCKv2[36], DETECTR[37], and CAS-EXPAR[38] as well as other methods which use catalytically-inactive Cas9 (dCas9) as a binding domain for specific DNA sequences combined with split signal transducing proteins such as chemiluminescent, fluorescent or other enzymatic reporters[39]. In addition, a variety of *in vitro* split reporter DNA and RNA biosensing approaches have been described by several previous studies[40-46].

Current methods for nucleic acid biosensing in live cells and their advantages and disadvantages

Because these *in vitro* approaches are necessarily applied outside of cells to detect DNA sequences, they have focused mostly on instantaneous quantification of specific molecular targets and not on understanding their dynamics within a native environment. Furthermore, live cell DNA detection methods in recent years have mostly used monomeric fluorescent proteins and other enzymatic reporters as transducer elements in design with a focus on high-resolution imaging via microscopy rather than focusing on bona-fide biosensing approaches with higher sensitivity and specificity[47-54]. With such designs, unbound and bound forms of the probe appear nearly identical in a plate reader or under the microscope, generally leading to a lower signal-to-background ratio (SBR) when applied as biosensors. Therefore, such "always-on" biosensors require a high local concentration of binding events to distinguish signal from background, so they are mostly limited to imaging highly repetitive elements that can be targeted by one sgRNA or to unique sequences targeted by 20-30 or more sgRNAs[47,49]. Imaging a short sequence present at a single copy has so far remained a difficult challenge in the field.

However, it has recently been demonstrated that a bipartite "turn-on" biosensor design that brings two components with negligible signal output in the absence of a target but higher signal output in its presence can exceed SBRs of such monomeric "always-on" probes and allow for both biosensing and imaging of genomic DNA, RNA, and protein-protein interactions in live cells. Signal output from such systems can occur either by activation of a chromophore by energy transfer from another activated chromophore—as in recent Förster Resonance Energy Transfer (FRET)-based or Bioluminescence Resonance Energy Transfer (BRET)-based

approaches for detection of a variety of targets[55-63]—or by reassembly of a bright fluorescent, chemiluminescent, or enzymatic reporter[39,64]. Many studies in the past have supported the use of split reporters as bipartite transducer elements within platforms used to measure and track the levels of various molecular targets both *in vitro* and in living cells. One such method is Bimolecular Fluorescence Complementation (BiFC), which enables the direct visualization of protein-protein interactions in living cells[65]. Several studies describe using a BiFC-based approach with transcription activator-like effectors (TALEs) as DNA binding domains and split fluorescent proteins as transducer elements for DNA biosensing in live cells[66-67]. Thus, application of such bipartite "turn-on" biosensing approaches to monitoring levels of other molecular targets *in situ* using optical microscopy is a logical progression within the biosensing field that has received increased interest in recent years.

Advantages and disadvantages of luminescent biosensing platforms

In addition to the major issue of accounting for higher background from unbound fluorescent probes leading to a naturally higher cellular auto-fluorescent background signal, fluorescence-based biosensing is plagued by other issues such as cellular phototoxicity and photobleaching of fluorophores and fluorescent proteins[68-70]. All of these issues contribute to a cumulative negative effect on SBR achievable with fluorescent probes. To increase sensitivity, luminescent reporters could offer an attractive alternative to fluorescent reporters in biosensor design as they have negligible auto-luminescent background signal. This is mainly because luminescence represents light produced from a catalytic reaction of an enzyme with its substrate instead of from excitation by incident exogenous light[69,70]. However, despite luminescent reporters having the advantage of

decreased background, the raw signal outputted from fluorescent reporters is brighter than available luminescent reporters[70]. However, a relatively new luciferase, NanoLuc, bridges this gap in signal intensity[71-72]. NanoLuc luciferase offers several advantages over Firefly and Renilla luciferases including enhanced stability, significantly smaller size, and >150-fold enhancement in luminescence output[71-72]. Furthermore, furimazine, the substrate for NanoLuc luciferase, is more stable and exhibits decreased levels of background activity than the substrate for RLuc, coelenterazine[71-72].

Our approach using split reporter reassembly as a bipartite transducer in biosensing platforms for DNA sequences and chromatin loops

In our studies, we hypothesized that by combining bipartite "turn-on" biosensing approaches with common signal reader devices such as luminometers or fluorometers or optical microscopy, we might be able to detect gene sequences and three-dimensional chromatin loops in live cells and track changes in levels of these targets during changes in cell state. We imagined that such approaches to monitoring changes in genomic sequences and architecture of the chromatin landscape could be a viable route to understanding more about early stages of various diseases, which could potentially inform appropriate treatment methodologies. To this end, we designed, optimized, and validated several bipartite "turn-on" biosensors using various transducer components to produce the signal. Namely, we have developed and optimized a luminescence-based, bipartite DNA sequence biosensor based on the NanoLuc Binary Technology (NanoBiT) complementation reporter system[73] recently created for NanoLuc luciferase and dual catalytically inactive Cas9 (dCas9) enzymes from *Streptococcus pyogenes* and *Staphylococcus aureus*. In addition, we developed and tested a different method for light production in a bipartite "turn-on"

DNA biosensor design based on dual species dCas9 enzymes and nonradiative energy transfer via Bioluminescence Resonance Energy Transfer (BRET).

Using both kinetic endpoint and dynamic time-course signal measurement and microscopy, we have demonstrated detection of specific DNA sequences and detection and tracking of chromatin loops in living cells. As a proof-of-concept, we tested these biosensor designs in positive and negative control live cell models that were assessed to either contain or lack the targets, respectively. As an application for our DNA sequence biosensor, we propose a potential method for detecting the presence of both individual repetitive and non-repetitive genomic loci, which could be applied toward genotyping mutant and wild-type cells at a single locus and toward identifying and isolating gene edit positive cells post-genome editing. Similarly, as an application for our chromatin loop biosensor, we propose a novel method to functionally link changes in chromatin interactions to changes in levels of gene expression at a given locus via pairing of live cell biosensing at a loop anchor region formed by a given promoter-enhancer pair with RT-qPCR for a gene of interest.

In this work, we first explore preliminary molecular modeling and initial assays involved in the design and characterization of our plasmid-based and ribonucleoprotein-based single species dCas9-NanoBiT DNA biosensors both *in vitro* and in living cells. Next, we examine the design, characterization, and application of a plasmid-based dCas9-NanoBiT DNA biosensor in living cells. Namely, in this section, we focus on future live cell-based detection and imaging applications for repetitive and non-repetitive endogenous genomic loci within the burgeoning CRISPR-Cas genome editing field. Subsequently, we discuss the impact of a critical

modification to the design of our dCas9-NanoBiT DNA biosensor involving the use of dCas9 enzymes from two separate bacterial species. In this section, we assess various applications of this new dual species dCas9-NanoBiT DNA biosensor design in detection of specific non-repetitive DNA sequences and chromatin loop anchors in living cells. We also compare and contrast various light-producing transducer elements in DNA biosensor design via experimental comparison of the signal-to-background ratios of DNA biosensors containing split luminescent reporter-based and Bioluminescence Resonance Energy Transfer-based transducer elements in live cells. Finally, we conclude this work by describing an exciting potential use case for this technology in the early detection of preleukemic mutations during acute myeloid leukemia tumorigenesis and by examining several future applications of split reporter-based DNA biosensing technology involving the detection and tracking of DNA sequences and three dimensional chromatin contacts *in vitro* and in living cells.

References

1. Morales, M. A., & Halpern, J. M. Guide to selecting a biorecognition element for biosensors. *Bioconjug. Chem.* ***29***(10), 3231–3239 (2018).
2. Giardi, M. T., & Piletska, E. V. (2006). Biotechnological applications of photosynthetic proteins: biochips, biosensors and biodevices. Springer, Boston, MA.
3. Yu, H. L. L., Maslova, A., & Hsing, I.-M. Rational design of electrochemical DNA biosensors for point-of-care applications. *ChemElectroChem* ***4***, 795−805 (2017).
4. Brazaca, L. C., Ribovski, L., Janegitz, B. C., & Zucolotto, V. (2017). Nanostructured materials and nanoparticles for point of care (POC) medical biosensors, in Medical Biosensors for Point of Care (POC) Applications, pp 229−254, Elsevier.
5. Justino, C. I. L., Freitas, A. C., Pereira, R., Duarte, A. C., & Rocha Santos, T. A. P. Recent developments in recognition elements for chemical sensors and biosensors. TrAC, *Trends Anal. Chem*. ***68***, 2−17 (2015).
6. Mittal, S., Kaur, H., Gautam, N., and Mantha, A. K. (2017) Biosensors for breast cancer diagnosis: A review of bioreceptors, biotransducers and signal amplification strategies. *Biosens. Bioelectron*. ***88***, 217−231.
7. Tetyana, P., Morgan Shumbula, P., & Njengele-Tetyana, Z. (2021). Biosensors: Design, development and applications. In S. Ameen, M.S. Akhtar, and H. Shin. (Eds.). *Nanopores*. IntechOpen.
8. Mehrotra, P. Biosensors and their applications - A review. *J. Oral Biol. Craniofac Res.* ***6***(2), 153-9 (2016).
9. Neethirajan, S. *et al*. Biosensors for sustainable food engineering: challenges and perspectives. *Biosensors* ***8***(1), 2 (2018).
10. Ali, J. *et al*. Biosensors: Their fundamentals, designs, types and most recent impactful applications: A review. *J. Biosens. Bioelectron.* ***8***, 235 (2017).
11. Crivianu-Gaita, V. & Thompson, M. Aptamers, antibody scFv, and antibody Fab' fragments: An overview and comparison of three of the most versatile biosensor biorecognition elements. *Biosens. Bioelectron*. ***85***, 32−45 (2016).
12. Amine, A., Mohammadi, H., Bourais, I., & Palleschi, G. Enzyme inhibition-based biosensors for food safety and environmental monitoring. *Biosens. Bioelectron*. ***21***, 1405−1423 (2006).
13. Gaudin, V. Advances in biosensor development for the screening of antibiotic residues in food products of animal origin – A comprehensive review. *Biosens. Bioelectron.* ***90***, 363−377 (2017).
14. Zhao, W. W., Xu, J. J., & Chen, H. Y. Photoelectrochemical enzymatic biosensors. *Biosens. Bioelectron*. ***92***, 294−304 (2017).
15. Li, C., Karadeniz, H., Canavar, E., & Erdem, A. Electrochemical sensing of label free DNA hybridization related to breast cancer 1 gene at disposable sensor platforms modified with single walled carbon nanotubes. *Electrochim. Acta* ***82***, 137−142 (2012).
16. Fan, C., Plaxco, K. W., & Heeger, A. J. Electrochemical interrogation of conformational changes as a reagentless method for the sequence-specific detection of DNA. *Proc. Natl. Acad. Sci. U. S. A.* ***100***, 9134−9137 (2003).
17. Dauphin-Ducharme, P., & Plaxco, K. W. Maximizing the signal gain of electrochemical-DNA sensors. *Anal. Chem.* ***88***, 11654−11662 (2016).

18. Zhu, B., Booth, M. A., Shepherd, P., Sheppard, A., & Travas- Sejdic, J. Distinguishing cytosine methylation using electrochemical, label-free detection of DNA hybridization and ds-targets. *Biosens. Bioelectron.* ***64***, 74−80 (2015).
19. Ferapontova, E. E. DNA electrochemistry and electrochemical sensors for nucleic acids. *Annu. Rev. Anal. Chem.* ***11***, 197−218 (2018).
20. Zhou, W., Jimmy Huang, P.-J., Ding, J., & Liu, J. Aptamer-based biosensors for biomedical diagnostics. *Analyst* ***139***, 2627 (2014).
21. Cieplak, M., & Kutner, W. Artificial biosensors: How can molecular imprinting mimic biorecognition? *Trends Biotechnol.* ***34***, 922−941 (2016).
22. Zhou, Y., Liu, B., Yang, R., and Liu, J. Filling in the gaps between nanozymes and enzymes: Challenges and opportunities. *Bioconjug. Chem.* ***28***, 2903−2909 (2017).
23. Wei, H., & Wang, E. Nanomaterials with enzyme-like characteristics (nanozymes): next-generation artificial enzymes. *Chem. Soc. Rev.* ***42***, 6060 (2013).
24. Meena, R. K., Kamboj, D., Kumar, K., & Karanwal, S. A brief review on aptamer based biosensors for detection of environmental pollution. *Int. J. Curr. Microbiol. App. Sci.* ***7***(10): 1483-1489 (2018).
25. McConnell, E. M., Nguyen, J., & Li, Y. Aptamer-based biosensors for environmental monitoring. *Front. Chem.* ***8***, 434 (2020).
26. Pal, M., Bulcha, M. R., Banu, M. G., & Lema, A.G. Emerging role of biosensors for detection of foodborne pathogens. *Am. J. Microbiol. Res.* ***9***(3), 92-95 (2021).
27. Li, X. *et al*. Liquid biopsy of circulating tumor DNA and biosensor applications. *Biosens. Bioelectron.* ***126***, 596-607 (2019).
28. Yang, X. *et al*. An electrochemiluminescence resonance energy transfer biosensor for the detection of circulating tumor DNA from blood plasma. *iScience* ***24***(9), 103019 (2021).
29. Wang, K. *et al*. Electrochemical biosensors for circulating tumor DNA detection. *Biosensors (Basel)* ***12***(8), 649 (2022).
30. Deepak, S. *et al*. Real-Time PCR: Revolutionizing detection and expression analysis of genes. *Curr. genomics* ***8***(4), 234–251 (2007).
31. Kralik, P., & Ricchi, M. A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything. *Front. Microbiol.* ***8***, 108 (2017).
32. Bonini, A. *et al*. Advances in biosensing: The CRISPR/Cas system as a new powerful tool for the detection of nucleic acids. *J Pharm. Biomed. Anal.* ***192*****,** 113645 (2021).
33. Li, Y., Li, S., Wang, J., & Liu, G. CRISPR/Cas systems towards next-generation biosensing. *Trends Biotechnol.* ***37***(7), 730–743 (2019).
34. Gootenberg, J. S. *et al*. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* ***356***(6336), 438–442 (2017).
35. Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O., & Zhang, F. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat. Protoc.* ***14***(10), 2986–3012 (2019).
36. Gootenberg, J. S. *et al*. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* ***360***(6387), 439–444 (2018).
37. Broughton, J. P. *et al*. CRISPR-Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* ***38***(7), 870–874 (2020).
38. Huang, M., Zhou, X., Wang, H., & Xing, D. Clustered regularly interspaced short palindromic repeats/cas9 triggered isothermal amplification for site-specific nucleic acid detection. *Anal. Chem.* ***90***(3), 2193–2200 (2018).

39. Zhang, Y. *et al*. Paired design of dCas9 as a systematic platform for the detection of featured nucleic acid sequences in pathogenic strains. *ACS Synth. Biol.* ***6***(2), 211–216 (2017).
40. Stains, C. I., Porter, J. R., Ooi, A.T., Segal, D. J. & Ghosh, I. DNA sequence-enabled reassembly of the green fluorescent protein. *J. Am. Chem. Soc.* ***127***, 10782–83 (2005).
41. Ooi, A. T., Stains, C. I., Ghosh, I. & Segal, D. J. Sequence-enabled reassembly of betalactamase (SEER-LAC): a sensitive method for the detection of double-stranded DNA. *Biochemistry* ***45***, 3620–25 (2006).
42. Porter, J. R., Stains, C. I., Segal, D. J., & Ghosh, I. Split beta-lactamase sensor for the sequence-specific detection of DNA methylation. *Anal. Chem.* ***79***(17), 6702–6708 (2007).
43. Ghosh, I., Stains, C. I., Ooi, A.T. & Segal, D. J. Direct detection of double-stranded DNA: molecular methods and applications for DNA diagnostics." *Mol. Biosyst.* ***2***, 551–60 (2006).
44. Stains, C. I., Furman, J. L., Segal, D. J., & Ghosh, I. Site-specific detection of DNA methylation utilizing mCpG-SEER. *J. Am. Chem. Soc.* ***128***(30), 9761–9765 (2006).
45. Furman, J. L. *et al*. Toward a general approach for RNA-templated hierarchical assembly of split-proteins. *J. Am. Chem. Soc.* ***132***(33), 11692–11701 (2010).
46. Zhou, L. *et al*. Tandem reassembly of split luciferase-DNA chimeras for bioluminescent detection of attomolar circulating microRNAs using a smartphone. *Biosens. Bioelectron.* ***173***, 112824 (2021).
47. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* ***155***, 1479–91 (2013).
48. Ye, H., Rong, Z., and Lin, Y. Live cell imaging of genomic loci using dCas9-SunTag system and a bright fluorescent protein. *Prot. Cell* ***8***, 853–55 (2017).
49. Chen, B., Zou, W., Xu, H., Liang, Y. & Huang, B. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag." *Nat. Commun.* ***9***, 5065 (2018).
50. Dreissig, S. *et al.* Live-cell CRISPR imaging in plants reveals dynamic telomere movements. *Plant J.* ***91***, 565–73 (2017).
51. Wu, X., Mao, S., Ying, Y., Krueger, C. J. & Chen, A. K. Progress and challenges for live-cell imaging of genomic loci using CRISPR-based platforms. *Genom. Proteom. Bioinform.* ***17***, 119-128 (2019).
52. Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U S A* ***112***, 11870–75 (2015).
53. Zhang, D. *et al*. CRISPR-Bind: A simple, custom CRISPR/dCas9-mediated labeling of genomic DNA for mapping in nanochannel arrays. Preprint at https://www.biorxiv.org/content/10.1101/371518v1 (2018).
54. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U S A* ***112***, 3002–7 (2015).
55. Boutorine, A. S., Novopashina, D. S., Krasheninina, O. A., Nozeret, K. & Venyaminova, A. G. Fluorescent probes for nucleic acid visualization in fixed and live cells. *Molecules* ***18***, 15357–97 (2013).
56. Dahan, L., Huang, L., Kedmi, R., Behlke, M. A. & Peer, D. SNP detection in mRNA in living cells using allele specific FRET probes." *PloS One* ***8***, e72389 (2013).

57. Didenko, V. V. DNA probes using fluorescence resonance energy transfer (FRET): designs and applications. *BioTechniques* ***31***, 1106–16, 1118 & 1120–21 (2001).
58. Wu, X. *et al*. A CRISPR/molecular beacon hybrid system for live-cell genomic imaging. *Nucleic Acids Res.* ***46***(13), e80 (2018).
59. Mao, S. *et al*. CRISPR/dual-FRET molecular beacon for sensitive live-cell imaging of non-repetitive genomic loci. *Nucleic Acids Res.* ***47***(20), e131. (2019).
60. Pfleger, K. D., Seeber, R. M., & Eidne, K. A. Bioluminescence resonance energy transfer (BRET) for the real-time detection of protein-protein interactions. *Nat. Protoc.* ***1***(1), 337–345 (2006).
61. Machleidt, T. *et al*. NanoBRET—A novel BRET platform for the analysis of protein-protein interactions. *ACS Chem. Biol.* ***10***(8), 1797-1804 (2015).
62. Durrant, D. E. *et al*. Development of a high-throughput NanoBRET screening platform to identify modulators of the RAS/RAF interaction. *Mol. Cancer Ther.* ***20***(9), 1743–1754 (2021).
63. Aird, E. J., Tompkins, K. J., Ramirez, M. P., & Gordon, W. R. Enhanced molecular tension sensor based on bioluminescence resonance energy transfer (BRET). *ACS Sens.* ***5***(1), 34–39 (2020).
64. Luker, K. E. *et al*. Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U S A*, ***101***(33), 12288–12293 (2004).
65. Kerppola, T. K. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu. Rev. Biophys.* ***37***, 465–487 (2008).
66. Hu, H. *et al*. Live visualization of genomic loci with BiFC-TALE. *Sci. Rep.* ***7***, 40192 (2017).
67. Hu, H. *et al*. Visualization of genomic loci in living cells with BiFC-TALE. *Curr. Protoc. Cell Biol.* ***82***, e78 (2019).
68. Bernas, T., Robinson, J. P., Asem, E. K. & Rajwa, B. Loss of image quality in photobleaching during microscopic imaging of fluorescent probes bound to chromatin. *J. Biomed. Opt.* ***10***, 064015 (2005).
69. Tung, J. K., Berglund, K., Gutekunst, C., Hochgeschwender, U. & Gross, R. E. Bioluminescence imaging in live cells and animals. *Neurophotonics* ***3***, 025001 (2016).
70. Choy, G. *et al*. Comparison of noninvasive fluorescent and bioluminescent small animal optical imaging. *Biotechniques* ***35***(5), 1022–1030 (2003).
71. Hall, M. P. *et al*. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* ***7***, 1848–57 (2012).
72. England, C. G., Ehlerding, E. B. & Cai, W. NanoLuc: a small luciferase is brightening up the field of bioluminescence. *Bioconjug. Chem.* ***27***, 1175–87 (2016).
73. Dixon, A. S. *et al*. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.* ***11***, 400–408 (2016).

Chapter 2: A Split Luminescent Reporter Reassembly DNA Sequence Biosensor: Conceptualization, Design, and Preliminary Experiments

INTRODUCTION

Rationale, initial hypotheses, and molecular modeling for split reporter DNA biosensor designs

Before starting any molecular modeling for a DNA sequence biosensor based on catalytically inactive *Streptococcus pyogenes* Cas9 (dCas9) fusions to LgBiT and SmBiT (NanoBiTs), we made initial predictions about the efficiency of NanoLuc luciferase reassembly for several of the fusion protein combinations binding to several possible DNA target site orientations. To predict conditions which might maximize binding efficiency and minimize potential steric hindrance associated with two dCas9-sgRNA-NanoBiT complexes binding within a certain proximity, we primarily needed to consider the unique structural elements of dCas9 as a fusion protein partner, its association with the sgRNA and target DNA helix in living cells, and the structure of its fusions with the NanoBiTs.

The NanoLuc Binary Technology (NanoBiT) complementation reporter system consists of independently optimized truncations of NanoLuc, a small, very bright 19.1 kDa luciferase from the deep-sea shrimp *Oplophorus gracilirostris* which was originally designed to be used in the context of detecting protein-protein interactions in live cells. NanoLuc luciferase is substantially smaller and exhibits blue-shifted emission compared to main competitors Firefly (*Photinus pyralis*) and Renilla (*Renilla reniformis*) luciferases **(Figure 2.1a)**. NanoLuc produces glow-type

luminescence with a signal half-life >2 h and shows ~150-fold greater activity than that of either Firefly or Renilla luciferases[1], making it a particularly advantageous biosensing reporter as it bridges a brightness gap between luminescent and fluorescent reporters that has prevented the widespread use of luminescent reporters for many years. NanoLuc is an oxidoreductase which catalyzes the oxidation of a novel imidazopyrazinone substrate, furimazine, to furimamide **(Figure 2.1b)**. This particular catalysis mechanism typically results in improved luminescence expression for NanoLuc luciferase in mammalian cells ~2.5 million-fold[1-2]. The SmBiT component of the NanoBiT system is a 1.3 kDa peptide with amino acid sequence VTGYRLFEEIL. This particular amino acid sequence was selected from a large library of peptides to have lowest affinity for the rest of the sequence represented by the LgBiT, which is an 18 kDa polypeptide that is a nearly complete set of secondary structures resembling fully functional NanoLuc but shows negligible oxidoreductase activity on the furimazine substrate[3]. In terms of linear amino acid sequence, SmBiT is the final 11 amino acids within the structure of NanoLuc luciferase[3]. As these NanoBiT complementation system reporter subunits have a high dissociation constant (K_D=190 μM) and extremely low catalytic activity, they must be brought into close proximity through a specific molecular interaction in order to reassemble full-length NanoLuc luciferase[3]. This requirement for proximity driven by a specific interaction combined with the availability and ease of use of a live cell furimazine substrate delivery system primarily drove our decision to use the NanoBiT system as a split reporter system in our DNA biosensor design process.

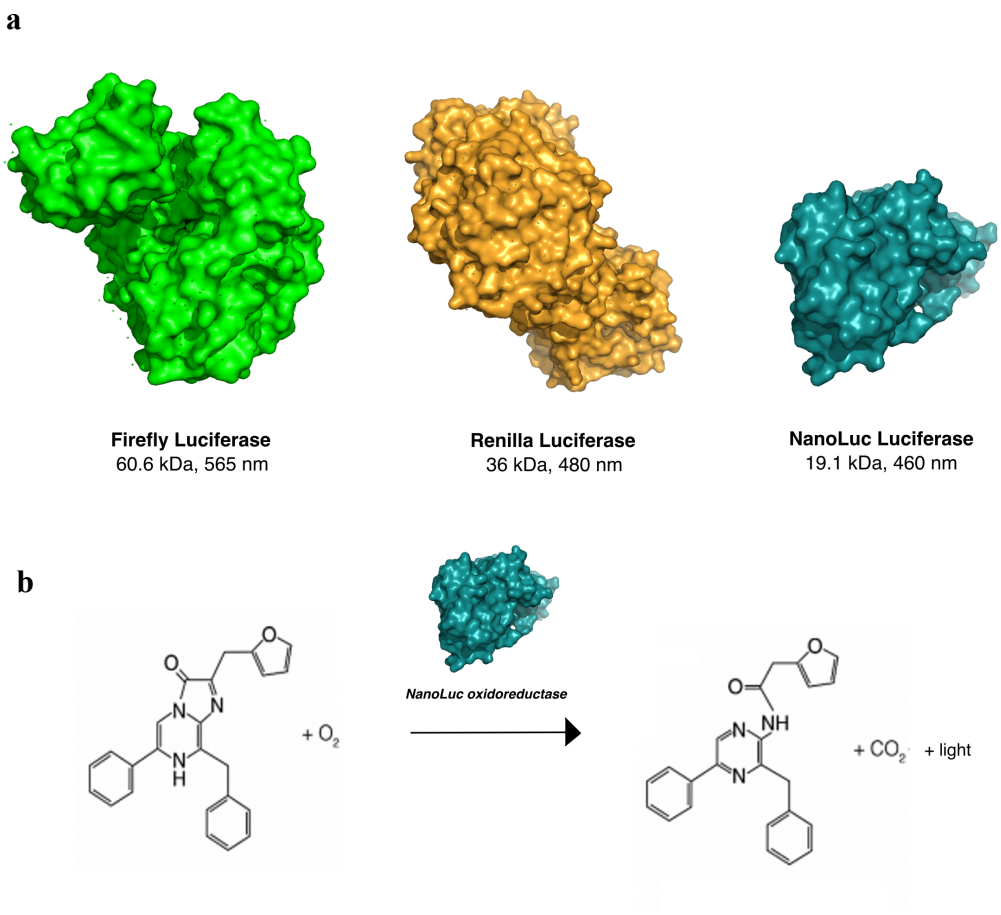


Figure 2.1: Advantageous properties of NanoLuc luciferase as a biosensing reporter

(a) Comparison of molecular weight and relative emission wavelengths for NanoLuc luciferase and two widely available luciferases, Firefly and Renilla, emphasizing the substantially smaller size and blue-shifted emission of NanoLuc. **(b)** The novel oxidation reaction of furimazine to furimamide catalyzed by NanoLuc. The enzyme requires furimazine substrate and oxygen and produces furimamide, CO_2, and glow-type light.

In initial consideration of dCas9 as an endogenous DNA binding domain for DNA sequences, we thought a domain-specific view of the dCas9 enzyme might assist in revealing any domains of the enzyme that might be proximal to termini and thus sterically sensitive to terminal fusions with larger protein domains. The total length covered by the *Streptococcus pyogenes* Cas9-

sgRNA complex on the DNA double helix is ~40 bp (~136 nm) with the complex extending ~7 bp (~2.4 nm) upstream of the target DNA site and ~10 bp (~3.4 nm) downstream of the target DNA site[4,5]. The RuvC nuclease, HNH nuclease, PAM-interacting, and α-helical Rec I and Rec II domains of the Cas9 enzyme associate with the 23 bp target site containing the protospacer adjacent motif (PAM) sequence[4]. The domain-specific structure of Cas9 can be observed from the molecular model shown in **Figure 2.2** below. The portion of the Cas9 enzyme that extends upstream and downstream lengthwise and occupies the back of the structure in this view is the Rec I domain, which binds the crRNA and tracrRNA[6]. On the front portion of Cas9, the Rec II domain and the HNH nuclease domain are visible. The structures of catalytically inactive Cas9 and Cas9 are nearly identical for the purposes of the following models as the former contains two mutations to catalytic sites in RuvC and HNH nuclease domains that do not significantly alter higher order protein structure[5].

Examining the expected fusion protein structure of catalytically inactive Cas9 enzyme with both LgBiT and SmBiT, at first glance it would seem that fusion proteins of LgBiT with dCas9 might be more likely to have a negative effect on dCas9 function depending on the chosen fusion protein linker length[7,8] due to a conformational space that would be more restricted than that of fusion proteins of SmBiT with dCas9. This is mainly because of the relative bulkiness of LgBiT as a fusion partner compared to SmBiT. To model the structure of our N-terminal and C-terminal fusion proteins between dCas9 and LgBiT and dCas9 and SmBiT, the spatial positioning of the N- and C-termini were thus an important consideration. The N-terminal RuvC domain and the C-terminal PAM-interacting domains of dCas9 are both present in nearly the same plane, protruding almost equal distances from both the front and back of the dCas9 structure in **Figure**

2.2. This means any linker loop domains attached to the termini of dCas9 should be expected to protrude in approximately similar directions in space. In addition, the spatial planes occupied by the fusion proteins are quite variable due to the flexibility of the chosen glycine-serine linker peptides[9-10]. Thus, as the PAM-interacting domain is located at the C-terminus of dCas9 and plays a critical role in target specificity of dCas9 as a DNA-binding domain, we expected fusions of a relatively bulky domain like LgBiT to the C-terminus of dCas9 to potentially affect functionality of dCas9 more than N-terminal LgBiT fusions to dCas9 and both N- and C-terminal fusions of SmBiT to dCas9.

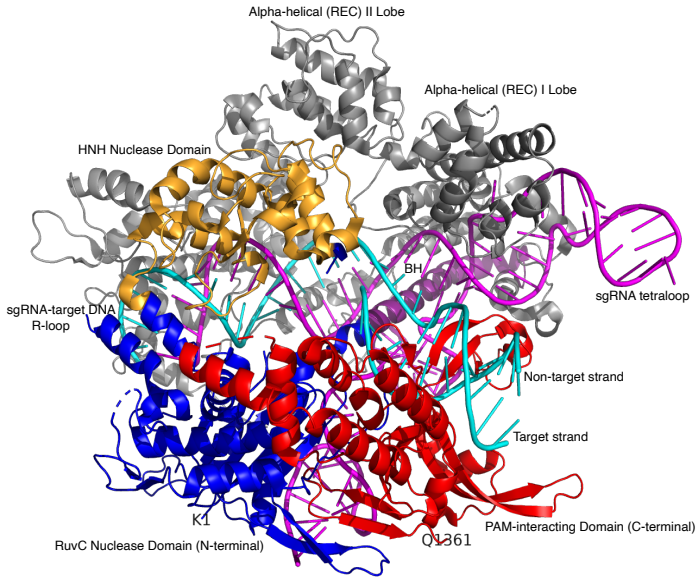


Figure 2.2: Domain substructure of *Sp*Cas9-sgRNA in complex with target DNA

The *Sp*Cas9 enzyme has six major domains: Rec I, Rec II, PAM-interacting (PI), bridge helix (BH), RuvC, and HNH. The α-helical REC lobes are shown in grey, the C-terminal PAM-interacting domain is shown in red, the bridge helix (BH) region is shown in dark purple, and the RuvC and HNH nuclease domains are shown in blue and yellow, respectively. The sgRNA is

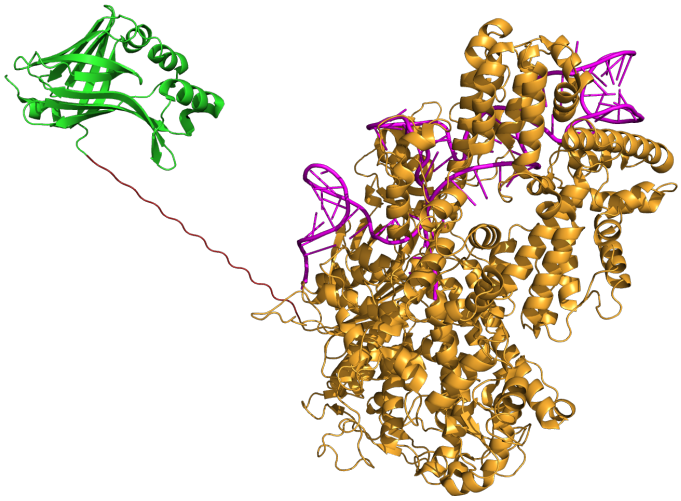
shown in magenta and the target DNA is shown in cyan. Rendered in PyMOL (version 2.5.2) via ray-tracing feature.

In addition, the phasing of the two separate binding events on the target DNA was also considered to be of considerable importance to the efficiency of NanoLuc reconstitution. For tandem target sites with both targets on a single strand, the two dCas9-sgRNA-NanoBiT fusion proteins were predicted to bind to the same side of the DNA. Thus, the Rec I domains of both enzymes were predicted to have the potential to sterically repel one another if the two dCas9s were both close enough in spacing on the helix and in phase due to a spacing that was a multiple of the full helical turn distance. With this in mind, we hypothesized that two turns of B-form DNA (~21 bp) might provide the optimal minimum binding spacing for efficient NanoLuc luciferase reassembly as this spacing might provide proper phasing for tandem target sites and allow the two dCas9-NanoBiT fusion proteins to bind without sterically repelling each other in this orientation. For such tandem sites, we postulated that this predicted in-phase binding requirement should hold true for every multiple of a full helical turn on the helix up to the theoretical maximum linear value for the system.

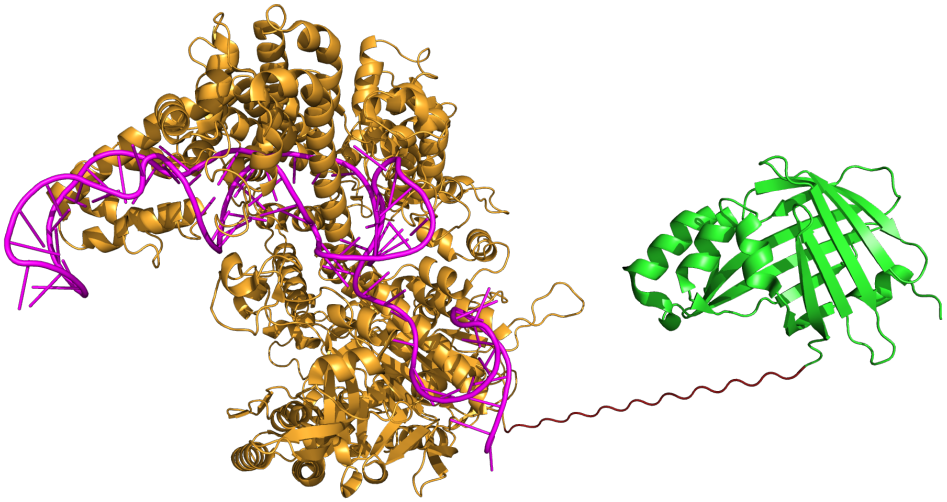
The theoretical maximum linear length such a system could occupy depends on the length of the two 17 amino acid or 20 amino acid flexible linkers plus the lengths of LgBiT and SmBiT. Positionally, where exactly these domains would protrude from the N-terminal RuvC and C-terminal PAM-interacting domains would be dependent on the complete conformational space available to the fusion domains based on the flexibility of their linker domains. Because the average amino acid residue length is ~3.8 Å, the 17 amino acid and 20 amino acid linkers we

selected could be a maximum length of ~65 Å and ~76 Å fully extended, respectively. SmBiT is 11 amino acids (approximate distance from end to end of ~42 Å), while LgBiT is 159 amino acids (approximate distance from end to end of ~600 Å). Thus, hypothetically, if the split reporter biosensor could be completely linear when NanoBiTs were reassembled upon the biosensor binding the DNA, the biosensor could achieve a maximum linear distance of ~787 Å from the edges of DNA-bound dCas9 enzymes. This corresponds to ~22 turns of B-form DNA, or ~232 bp. This was predicted as a theoretical maximum value for spacer development to test a dCas9-NanoBiT DNA biosensor, but we realized that spacers should be much shorter in practice, especially for other target site orientations where LgBiT and SmBiT might not protrude into the same plane of the DNA helix upon biosensor binding to DNA. For instance, when considering inverted target sites with targets on opposite strands, the two dCas9s should bind to opposite sides of the DNA and the Rec I domains would not be sterically interfering unless separated by a multiple of a half helical turn distance. Before starting initial experimental testing of various live cell DNA biosensor designs, we constructed several molecular models of potential designs in PyMOL (version 2.5.2). To begin, we created a fusion protein which depicted full-length, untruncated NanoLuc luciferase fused to catalytically inactive *Streptococcus pyogenes* Cas9 by a flexible 20 amino acid linker peptide containing mostly glycine and serine amino acid residues. This model is shown in **Figure 2.3** below.

a



b



c

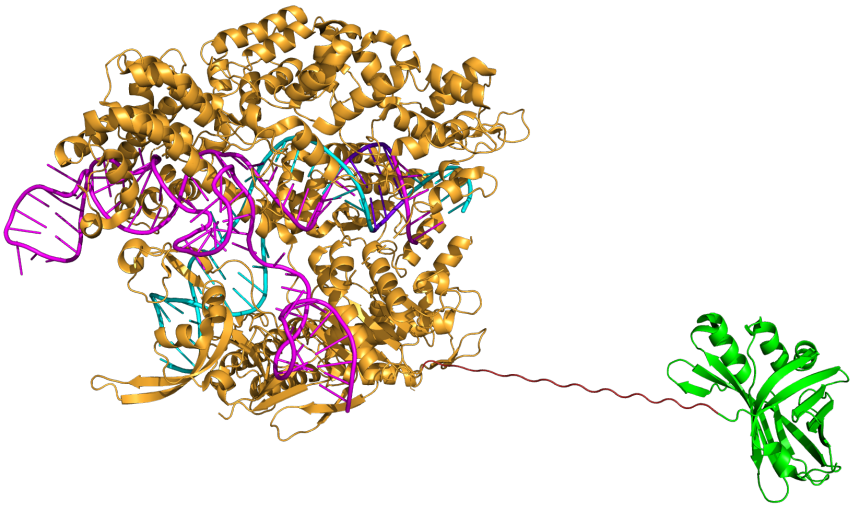


Figure 2.3: NanoLuc—d*Sp*Cas9 structure unbound and bound to a short DNA sequence

(a) Catalytically inactive Cas9 from *Streptococcus pyogenes* (orange) in complex with sgRNA (magenta) fused to NanoLuc luciferase (green) at the N-terminus. Fusion protein structure is depicted with C-terminus on the left and N-terminus on the right. Flexible GTGGSGGSGGSGGSGGSGRP linker peptide is depicted in red. **(b)** To emphasize the sgRNA association with dCas9 within the large α-helical REC lobe, the NanoLuc-dCas9 fusion protein structure is rotated 180° around the vertical axis and further rotated approximately 45° around the horizontal axis. **(c)** NanoLuc-dCas9 structure showing dCas9-sgRNA in complex with target DNA, emphasizing the sharp bending and unwinding of target DNA, thereby flipping DNA nucleotides out of the duplex and toward the guide RNA for sequence interrogation. The sgRNA is shown in magenta, the target DNA is shown in cyan, and the 5'-NGG PAM sequence on the target DNA and the cognate 5'-CCN sequence on the sgRNA are both shown in purple. All models rendered in PyMOL (version 2.5.2) via ray-tracing feature.

Notably, the 20 amino acid linker initially present in the original plasmid used to clone the NanoLuc-dCas9 and NanoBiT-dCas9 DNA constructs was not altered in our subcloning processes, resulting in the first full-length NanoLuc luciferase fusion protein with dCas9 containing a proline residue in the 20th position of the flexible linker. It has been shown that the presence of a proline residue in an alpha helix within its second and subsequent helical turns is strongly correlated with that helix having a kink[11-12], possibly due to its highly constrained phi angle[13]. However, proteins containing a proline residue are highly disfavored from alpha-helix formation[11]. Because the linker sequence contains 60% glycine residues and small polar amino acids serine and threonine to maintain the stability and solubility of the linker in aqueous

solutions by forming hydrogen bonds with water molecules[9-10], thereby reducing the unfavorable interaction between the linker and its fused protein moieties, this linker peptide likely adopts the random coil structure rather than an alpha helical structure[14]. This proline residue is positioned immediately before the start of the catalytically inactive Cas9 sequence. This would only produce a potential kink at the point of the N-terminal linker connection to dCas9 if the linker chosen had been alpha-helix-forming. Since we were able to rule out this alpha helical structure as a reasonable possibility and thus the possibility of a kink complicating the modeling process for our biosensor, this proline residue was not removed from the linker sequence prior to molecular modeling or prior to testing our construct designs in initial experiments and we assumed the entire linker adopted a flexible coil structure.

To generate reasonable models of two dCas9-sgRNA-NanoBiT fusion proteins binding to target DNA sequences, we first generated separate PyMOL models of the individual fusion proteins bound to DNA, then combined these models on variable spacer length DNA target scaffolds to predict which fusion protein orientations would result in the best alignment of LgBiT and SmBiT to reassemble a functional NanoLuc luciferase. The optimal alignment of LgBiT and SmBiT components is shown in **Figure 2.4** below. The LgBiT and SmBiT provide the most efficient reassembly of NanoLuc luciferase when SmBiT aligns just below the C-terminal β-strand of LgBiT. In this orientation, the amino acids at the N-terminus of SmBiT can form appropriate hydrogen bonds to the amino acids in the final β-strand at the C-terminus of LgBiT, forming the complete β-sheet necessary to restore NanoLuc oxidoreductase activity. The individual NanoBiT-dCas9 and dCas9-NanoBiT models are shown in **Figure 2.5** below.

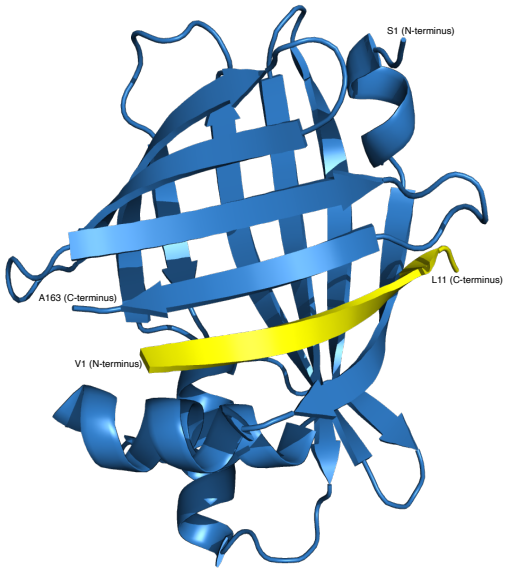
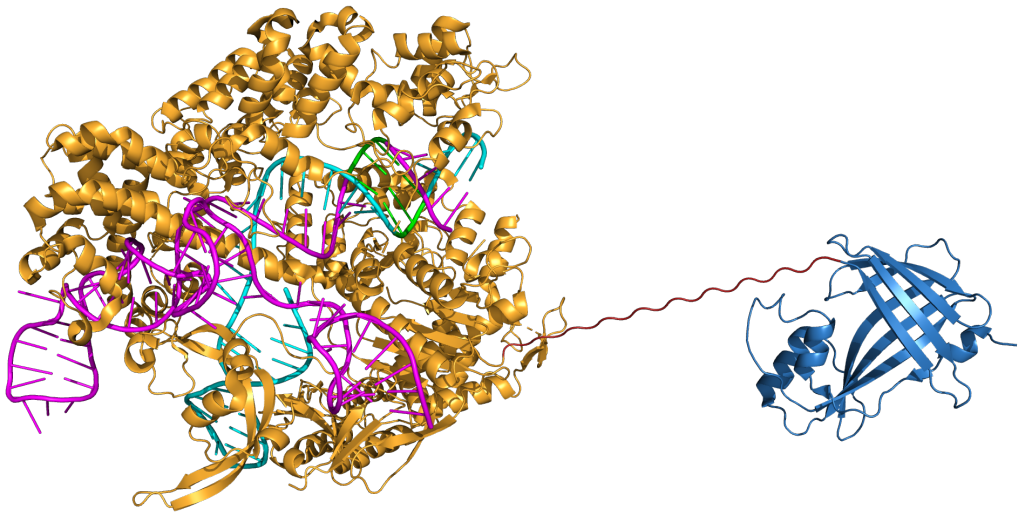


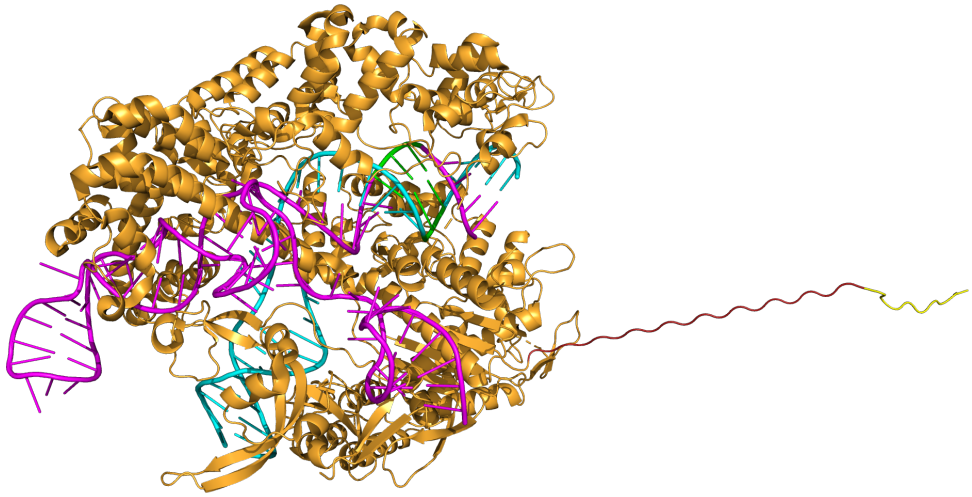
Figure 2.4: Optimal alignment of LgBiT and SmBiT for efficient NanoLuc luciferase reassembly

NanoLuc luciferase is most efficiently reassembled from the LgBiT and SmBiT truncations when the β-strand-containing SmBiT peptide component (yellow) aligns just below the C-terminal β-strand of the LgBiT polypeptide component (blue) with the SmBiT N-terminus proximal to the LgBiT C-terminus. Structure rendered in PyMOL (version 2.5.2) via ray-tracing feature.

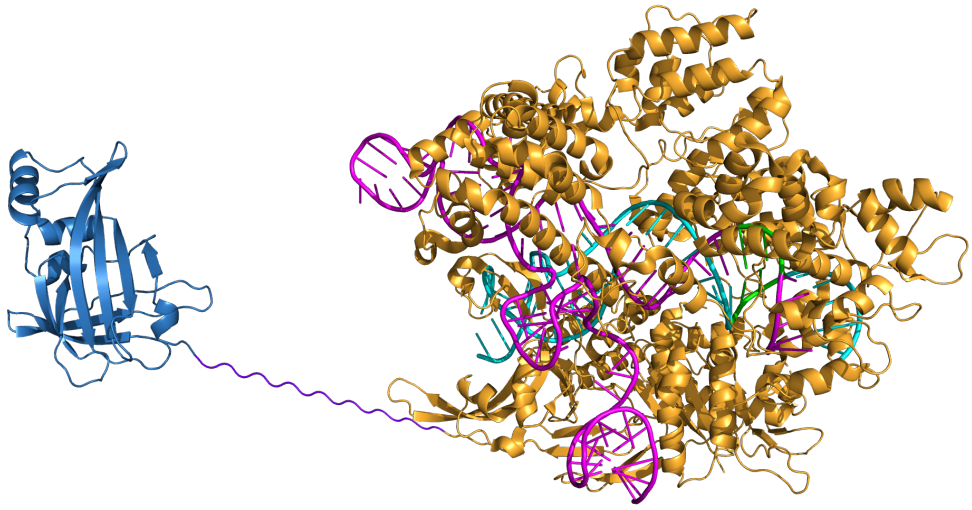
a



b



c



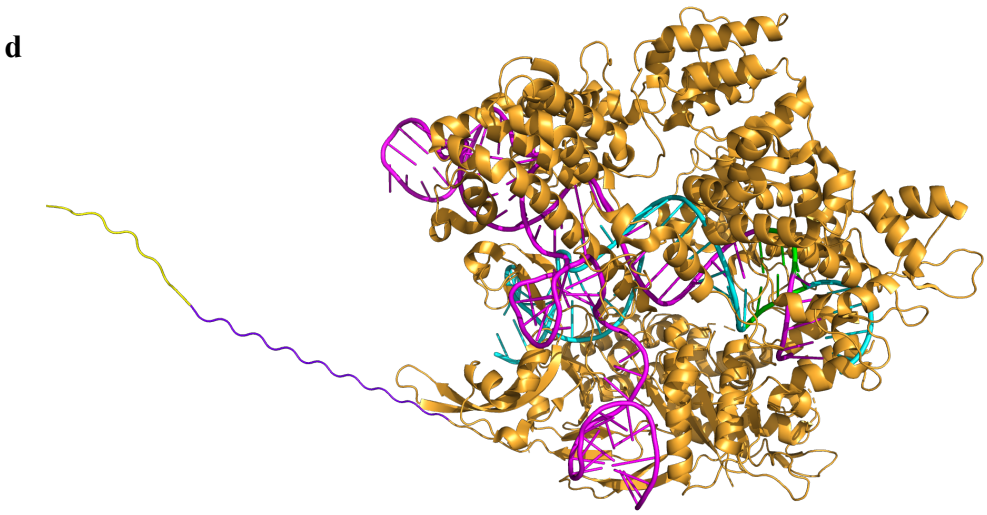


Figure 2.5: Individual split reporter DNA biosensor fusion protein models

(a) LgBiT fused to the N-terminus of dCas9. Flexible GTGGSGGSGGSGGSGGSGRP linker peptide is depicted in red. **(b)** SmBiT fused to the N-terminus of dCas9. Flexible GTGGSGGSGGSGGSGGSGRP linker peptide is depicted in red. **(c)** LgBiT fused to the C-terminus of dCas9. Flexible GGSGGSGGSGGSGGSAS linker peptide is shown in purple. **(d)** SmBiT fused to the C-terminus of dCas9. Flexible GGSGGSGGSGGSGGSAS linker peptide is shown in purple. For all models, the sgRNA is shown in magenta, the target DNA is shown in cyan, and the 5'-NGG PAM sequence on the target DNA and the cognate 5'-CCN sequence on the sgRNA are both shown in green. All models rendered in PyMOL (version 2.5.2) via ray-tracing feature.

We then proceeded to create models for several combinations of dCas9-NanoBiT and NanoBiT-dCas9 fusion proteins binding to 23 bp target DNA sites with various length DNA spacers and orientations between them **(Figure 2.6)**. The three orientations of DNA we were most interested in were tandem (parallel on the same strand), inverted (5'-NGG PAMs oriented inward on

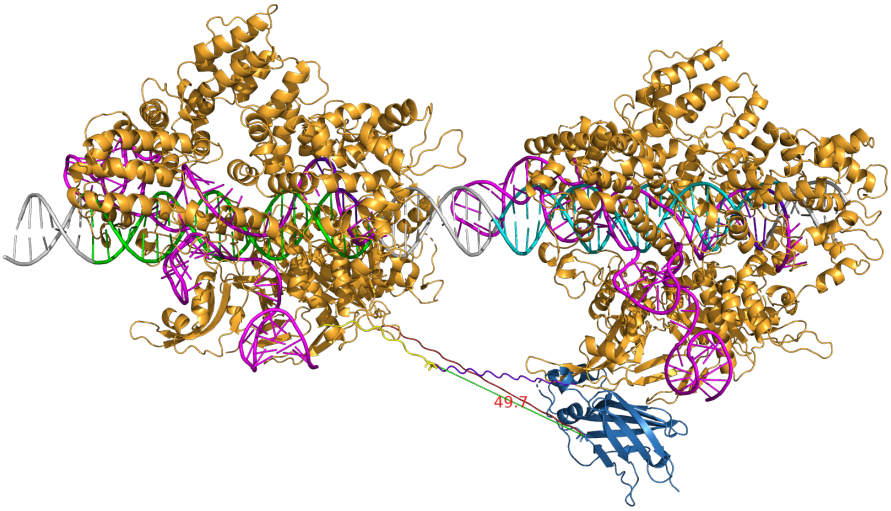
opposite strands) and everted (5'-NGG PAMs oriented outward on opposite strands). Based on the optimal reassembly alignment of the NanoBiTs and the direction which the NanoBiTs would be expected to protrude from dCas9 given the locations of its N- and C-termini, we developed hypotheses for which combinations of fusion protein orientations would result in the highest likelihood that the NanoBiTs would protrude into the same space for a given target DNA orientation.

When both fusion proteins were bound to tandem target DNA sites, we expected combinations of dCas9 fusions with NanoBiTs at opposite termini would allow for the SmBiT and LgBiT to protrude into the same space for efficient NanoLuc reassembly. Given the fusion proteins would bind to the same strand of the target DNA, we hypothesized that when the target sites were separated by a spacer DNA sequence with length equal to a multiple of a helical turn of B-form DNA (~10.5 bp or ~36 Å), the NanoBiTs would be most likely to reassemble properly. One such model of this fusion protein combination, target site orientation, and target site spacing is shown below, with LgBiT-dCas9 and dCas9-SmBiT binding to tandem target DNA sites 10 bp apart **(Figure 2.6a)**. As expected, the NanoBiTs in this model protrude toward each other, which we hypothesized would likely make reassembly more favorable for this condition. Given the relatively linear structure of the linker peptides as shown and their extensive flexibility, even at this relatively short target DNA spacing there appeared to be a possibility of achieving proper NanoLuc reassembly. The possible conformation states of the flexible linker peptides are too extensive to thoroughly model, so we focused on using nearly linear extensions of the coiled linkers in our models. However, for any models depicted, we assumed the linkers could occupy many conformational states.

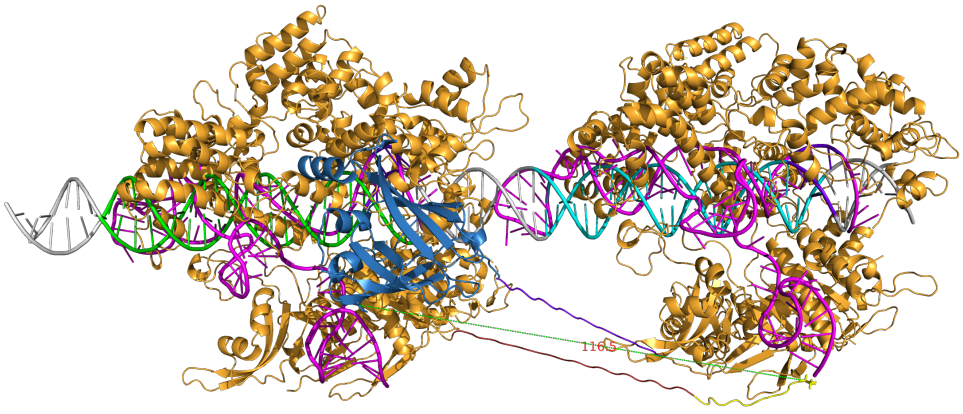
Another such model where NanoBiTs are shown to protrude toward one another when targeting DNA is SmBiT-dCas9 and dCas9-LgBiT binding to tandem target DNA sites 10 bp apart **(Figure 2.6b)**. If the fusion protein combinations of the probe where both LgBiT and SmBiT are fused to the N-terminus of dCas9 **(Figure 2.6c)** or both NanoBiTs are fused to the C-terminus of dCas9 **(Figure 2.6d)** were modeled on the same tandemly oriented target DNA, it appeared that the NanoBiTs would not be able to reassemble efficiently for any possible spacing. If the N-terminus of SmBiT needs to be brought into very close proximity to the C-terminus of LgBiT to restore the complete β-sheet present at the C-terminus of NanoLuc luciferase, the modeled distance between the N-terminal residue in SmBiT and the C-terminal residue in LgBiT is paramount to understanding which fusion protein orientation, target DNA orientation, and target DNA spacing combinations could potentially make NanoLuc reassembly more favorable. If these distances are measured for LgBiT-dCas9 + dCas9-SmBiT **(Figure 2.6a)**, SmBiT-dCas9 + dCas9-LgBiT **(Figure 2.6b)**, LgBiT-dCas9 + SmBiT-dCas9 **(Figure 2.6c)**, and dCas9-SmBiT + dCas9-LgBiT **(Figure 2.6d)** binding to tandem target DNA sites 10 bp apart, they are ~50 Å, ~117 Å, ~152 Å, and ~100 Å, respectively. Thus, when targeting our DNA biosensor to target DNA sites in tandem, LgBiT-dCas9 and dCas9-SmBiT fusion proteins were expected to show the highest luminescence as this combination results in the proper amino acid residues in the NanoBiTs being brought into closest proximity. After scrutinizing these models, we expected one additional helical turn beyond 10 bp spacing to provide more appropriate spacing of the LgBiT and SmBiT components represented in **Figure 2.6a**. We reasoned that extending the grey spacer sequence shown by one full helical turn could accomplish this by directly reducing the spacing between the N-terminus of SmBiT and the C-terminus of LgBiT. Unexpectedly, when

the LgBiT-dCas9 + dCas9-SmBiT fusion protein combination was modeled on tandem target DNA sites 20 bp apart **(Figure 2.6e)**, the distance between the N-terminal residue in SmBiT and the C-terminal residue in LgBiT was very slightly increased to 50.8 Å. This could be due to the modeled distance between the NanoBiTs increasing in other dimensions while the distance in the plane of the double helix length decreased. We imagined that, in the scenario where the same fusion proteins bind to tandem target DNA sites, the distance between the SmBiT and LgBiT could be reduced even further by extending the spacer length up until a certain point. We believed initial experiments testing a large set of various DNA spacer lengths beyond 20 bp for target sites in tandem with the four fusion protein combinations might reveal a more optimal spacing.

a



b



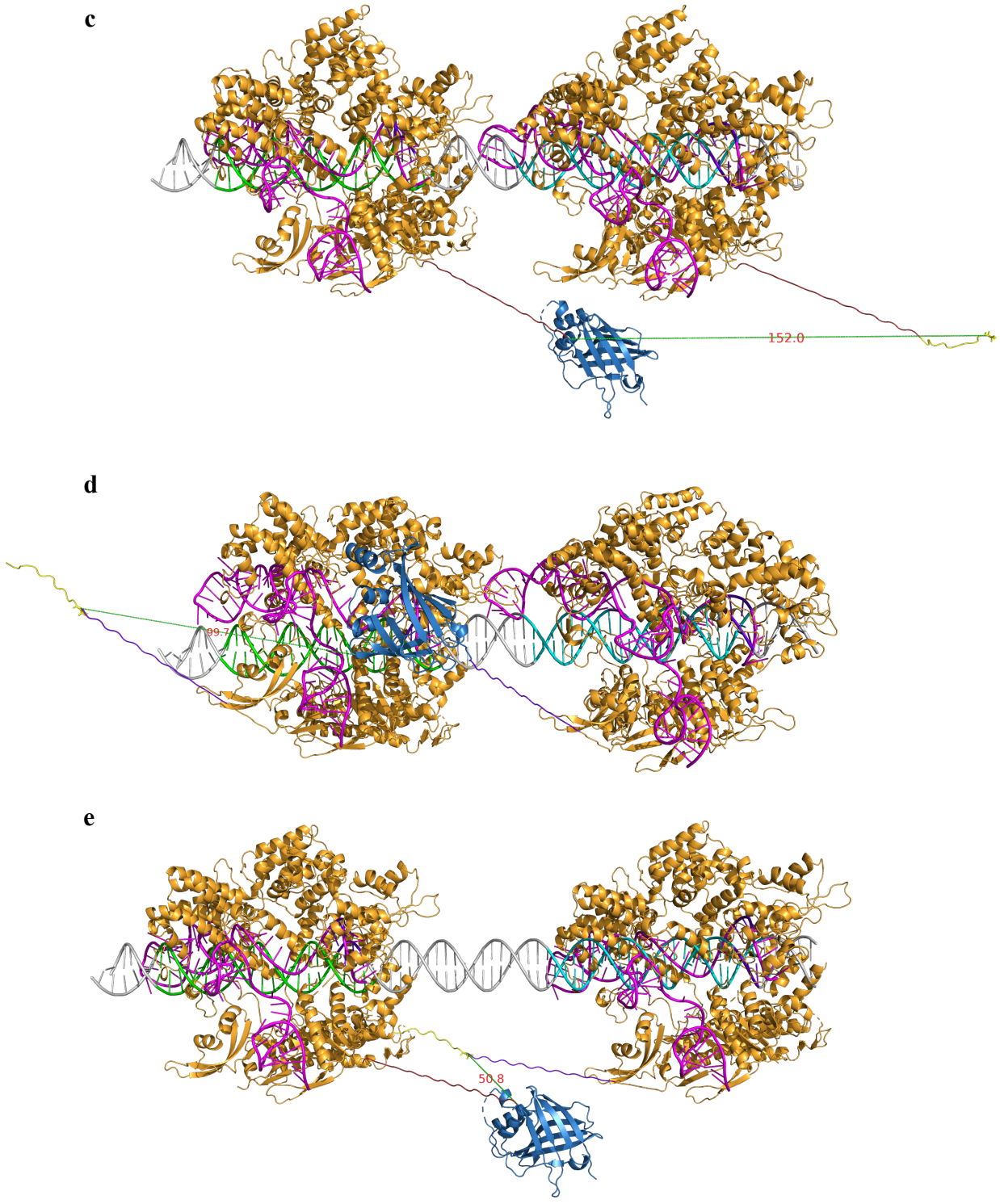


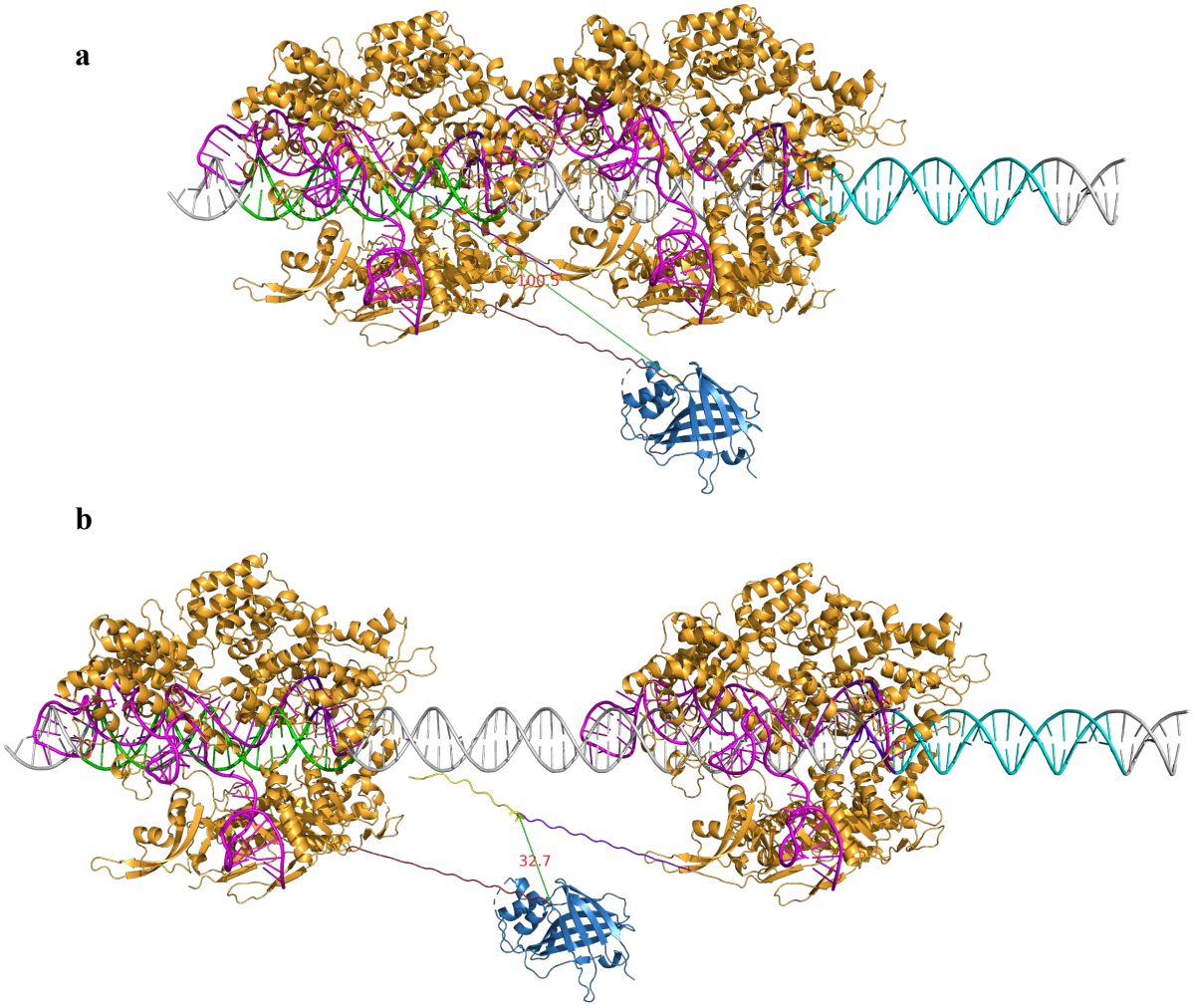
Figure 2.6: Split NanoLuc DNA biosensor modeling on tandem target sites

(a) LgBiT-dCas9 (left) and dCas9-SmBiT (right) binding to tandem target DNA sites 10 bp apart. **(b)** SmBiT-dCas9 (left) and dCas9-LgBiT (right) binding to tandem target DNA

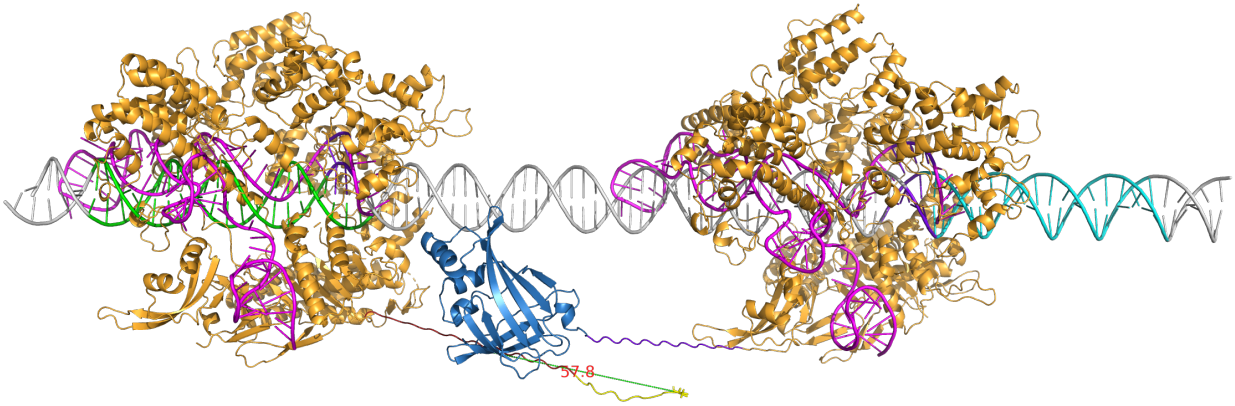
sites 10 bp apart. **(c)** LgBiT-dCas9 (left) and SmBiT-dCas9 (right) binding to tandem target DNA sites 10 bp apart. **(d)** dCas9-SmBiT (left) and dCas9-LgBiT (right) binding to tandem target DNA sites 10 bp apart. **(e)** LgBiT-dCas9 (left) and dCas9-SmBiT (right) binding to tandem target DNA sites 20 bp apart. For all models, dCas9 is shown in orange, the sgRNA is shown in magenta, LgBiT is shown in blue, SmBiT is shown in yellow, the linker peptide on the dCas9 N-terminus is shown in red, the linker peptide on the dCas9 C-terminus is shown in purple, spacer DNA is shown in grey, the upstream target DNA sequence is shown in green, the downstream target DNA sequence is shown in cyan, and the 5'-NGG PAM sequence on the target DNA and the cognate 5'-CCN sequence on the sgRNA are both shown in purple. The observed distance between N-terminus of SmBiT and C-terminus of LgBiT in these models is shown as a dashed green line with red text. All models rendered in PyMOL (version 2.5.2) via ray-tracing feature.

Next, we modeled all four NanoBiT-dCas9 and dCas9-NanoBiT fusion protein combinations binding to inverted target DNA sites. For the inverted target DNA orientation, the target sites are on opposite DNA strands, so we hypothesized that when the target sites were separated by a spacer DNA sequence with length equal to a multiple of one-half helical turn of B-form DNA (~5 bp or ~18 Å), the NanoBiTs would be most likely to reassemble as they might be expected to be in phase on the double helix. We chose a 25 bp spacer DNA sequence for modeling as the 10 and 20 bp tandem DNA modeling showed that the NanoBiTs might be expected to be in closest proximity when separated by around 20-40 bp, and 25 bp is approximately 5 half helical turns. By modeling full-turn multiple spacers on tandem target sites and half-turn multiple spacers on inverted target sites, we could assume the fusion protein orientation that presented the NanoBiTs

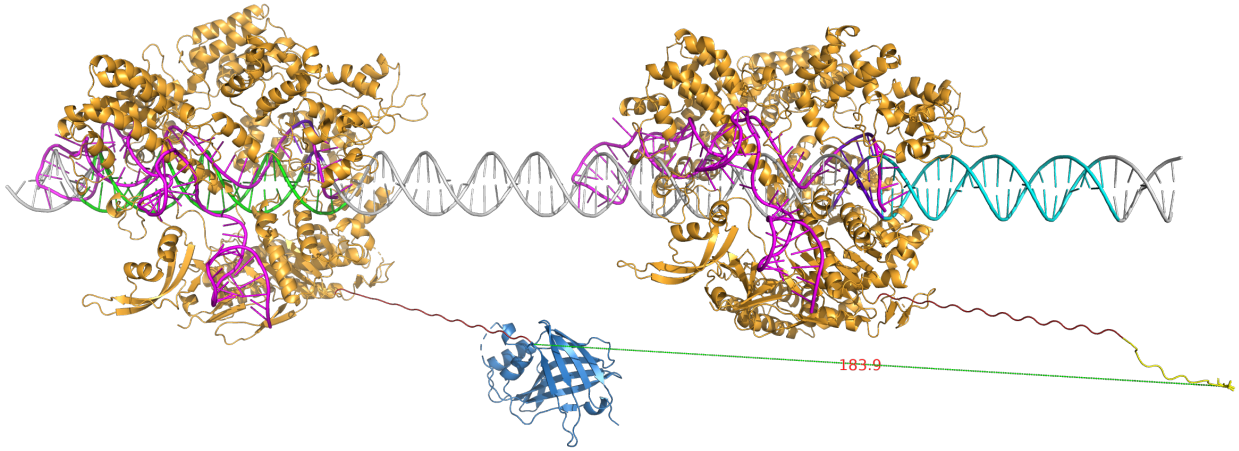
in closer proximity when modeled on tandem target sites could also be expected to work well on inverted and everted target sites as the primary concern was which direction the NanoBiT components protruded when each dCas9-NanoBiT complex bound to DNA. Thus, we modeled LgBiT-dCas9 and dCas9-SmBiT binding to inverted target DNA sites 25 bp apart **(Figure 2.7a)**.



c



d



e

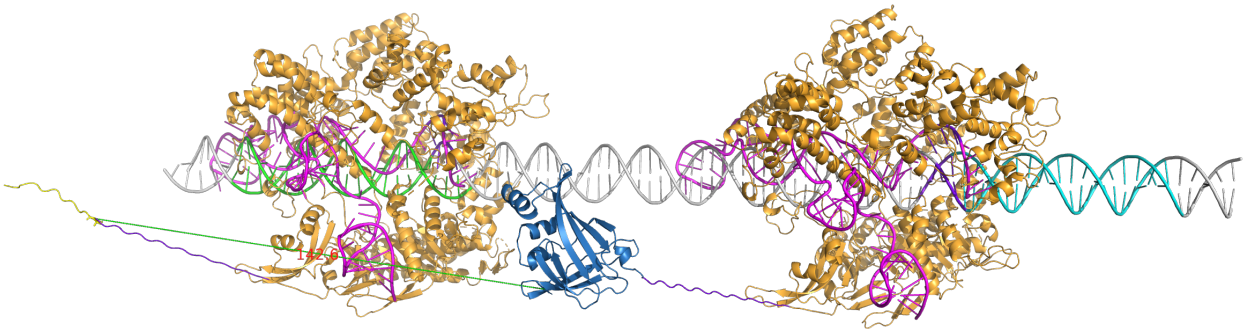


Figure 2.7: Split NanoLuc DNA biosensor modeling on inverted target sites

(a) LgBiT-dCas9 (left) and dCas9-SmBiT (right) binding to inverted target DNA sites 25 bp apart. **(b)** LgBiT-dCas9 (left) and dCas9-SmBiT (right) binding to inverted target DNA sites 45 bp apart. **(c)** SmBiT-dCas9 (left) and dCas9-LgBiT (right) binding to inverted target DNA sites 45 bp apart. **(d)** LgBiT-dCas9 (left) and SmBiT-dCas9 (right) binding to inverted target DNA sites 45 bp apart. **(e)** dCas9-SmBiT (left) and dCas9-LgBiT (right) binding to inverted target DNA sites 45 bp apart. For all models, dCas9 is shown in orange, the sgRNA is shown in magenta, LgBiT is shown in blue, SmBiT is shown in yellow, the linker peptide on the dCas9 N-terminus is shown in red, the linker peptide on the dCas9 C-terminus is shown in purple, spacer DNA is shown in grey, the upstream target DNA sequence is shown in green, the downstream target DNA sequence is shown in cyan, and the 5'-NGG PAM sequence on the target DNA and the cognate 5'-CCN sequence on the sgRNA are both shown in purple. The observed distance between N-terminus of SmBiT and C-terminus of LgBiT in these models is shown as a dashed green line with red text. All models rendered in PyMOL (version 2.5.2) via ray-tracing feature.

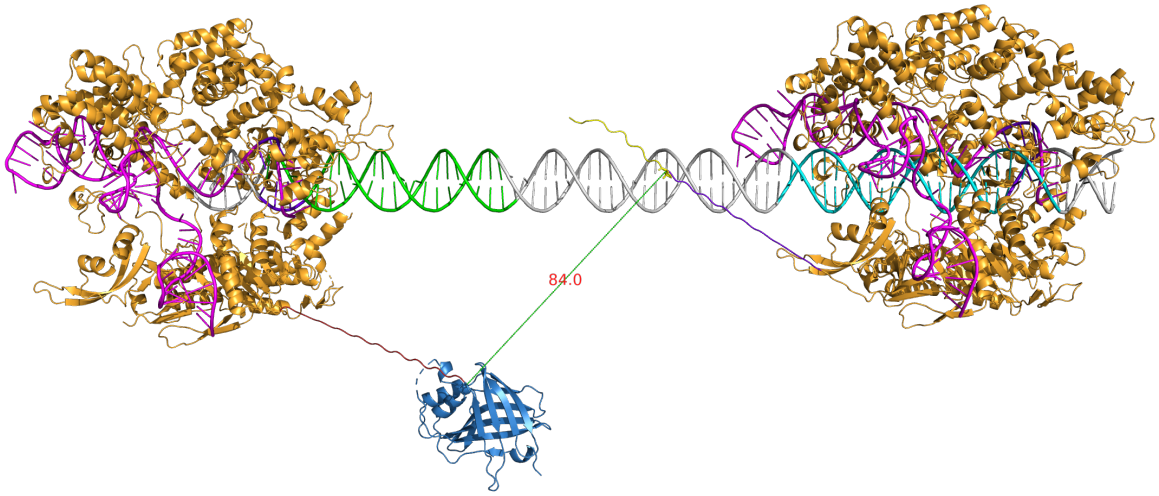
We observed from this model that the dCas9 on the right would be expected to bind the inverted target sequence closer to its 5' end as opposed to its 3' end if it bound to a tandem target sequence. As a result, the two dCas9-NanoBiT fusion proteins are closer in space even when using this 25 bp spacer. In this model, the C-terminal fused SmBiT extends toward the opposing LgBiT-dCas9 complex, and the N-terminus of SmBiT and C-terminus of LgBiT are ~101 Å apart. Thus, we reasoned reassembly might be improved if inverted target sites were 45 bp apart instead of 25 bp apart **(Figure 2.7b)**. In this model, we observed the shortest distance between

N-terminus of SmBiT and C-terminus of LgBiT (~33 Å) of all models up to this point. Thus, this fusion protein orientation, target site orientation, and target site spacing combination was predicted to be one of the most efficient pairings in initial experiments. In the other N-terminal + C-terminal combination of NanoBiTs, SmBiT-dCas9 + dCas9-LgBiT **(Figure 2.7c)**, the NanoBiTs were ~58 Å apart when modeled on the same inverted 45 bp target DNA. When the two N-terminal fusion proteins **(Figure 2.7d)** and the two C-terminal fusion proteins (**Figure 2.7e)** were modeled on inverted target sites 45 bp apart, the distances between the SmBiT N-terminus and the LgBiT C-terminus were found to be ~184 Å and ~143 Å, respectively.

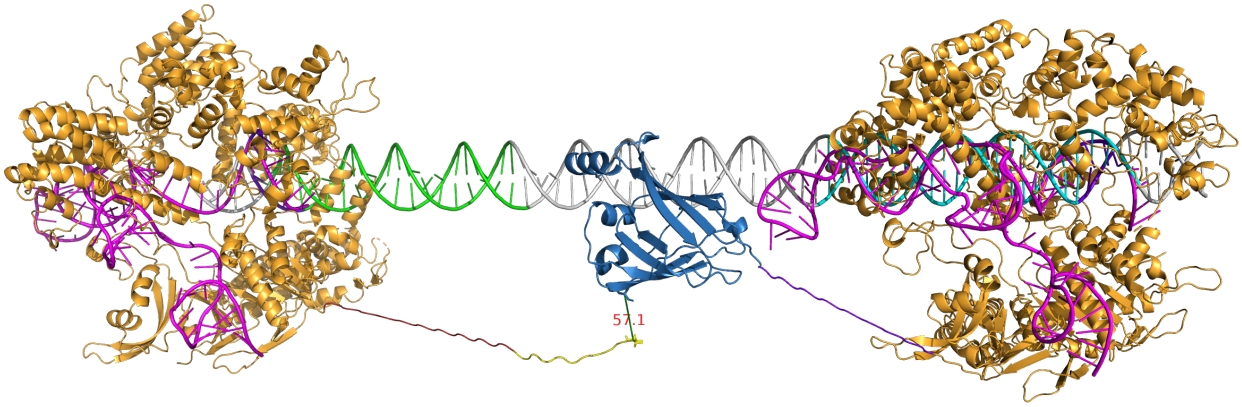
Finally, we modeled the four fusion protein orientations binding to everted target sites. For this modeling, we used only a 25 bp spacer sequence. The rationale for this was that we expected when binding these target sites, the two dCas9-NanoBiT complexes should be farther apart than for the tandem and inverted orientations due to the positioning of the PAM sites on opposing sides of their respective 23 bp target DNA sites (left site PAM is located upstream while right site PAM is located downstream). Also, as for the inverted target sites, the PAM sites are on different strands of the helix. Thus, they might similarly be expected to be in phase when separated by a half-turn multiple distance on the DNA, and 25 bp is approximately 5 half helical turns. As for other models, we started with LgBiT-dCas9 and dCas9-SmBiT binding to everted target sites 25 bp apart **(Figure 2.8a)**, then modeled SmBiT-dCas9 + dCas9-LgBiT, LgBiT-dCas9 + SmBiT-dCas9, and dCas9-SmBiT + dCas9-LgBiT **(Figure 2.8b-d)** binding to the same target DNA scaffold with everted target sites 25 bp apart. These models resulted in distances between SmBiT N-terminus and LgBiT C-terminus of ~84 Å, ~57 Å, ~241 Å, and ~203 Å, respectively. Based on these initial models, we hypothesized that for everted target sites, the

SmBiT-dCas9 + dCas9-LgBiT fusion protein orientation might produce the highest signals in initial experiments, especially for one fewer helical turn in the spacer sequence due to the expected decrease in the distance between the SmBiT N-terminus and the LgBiT C-terminus. Thus, we hypothesized that everted target sites 15 bp apart might provide more ideal spacing for all fusion protein orientations in initial experimental testing of a larger parameter space.

a



b



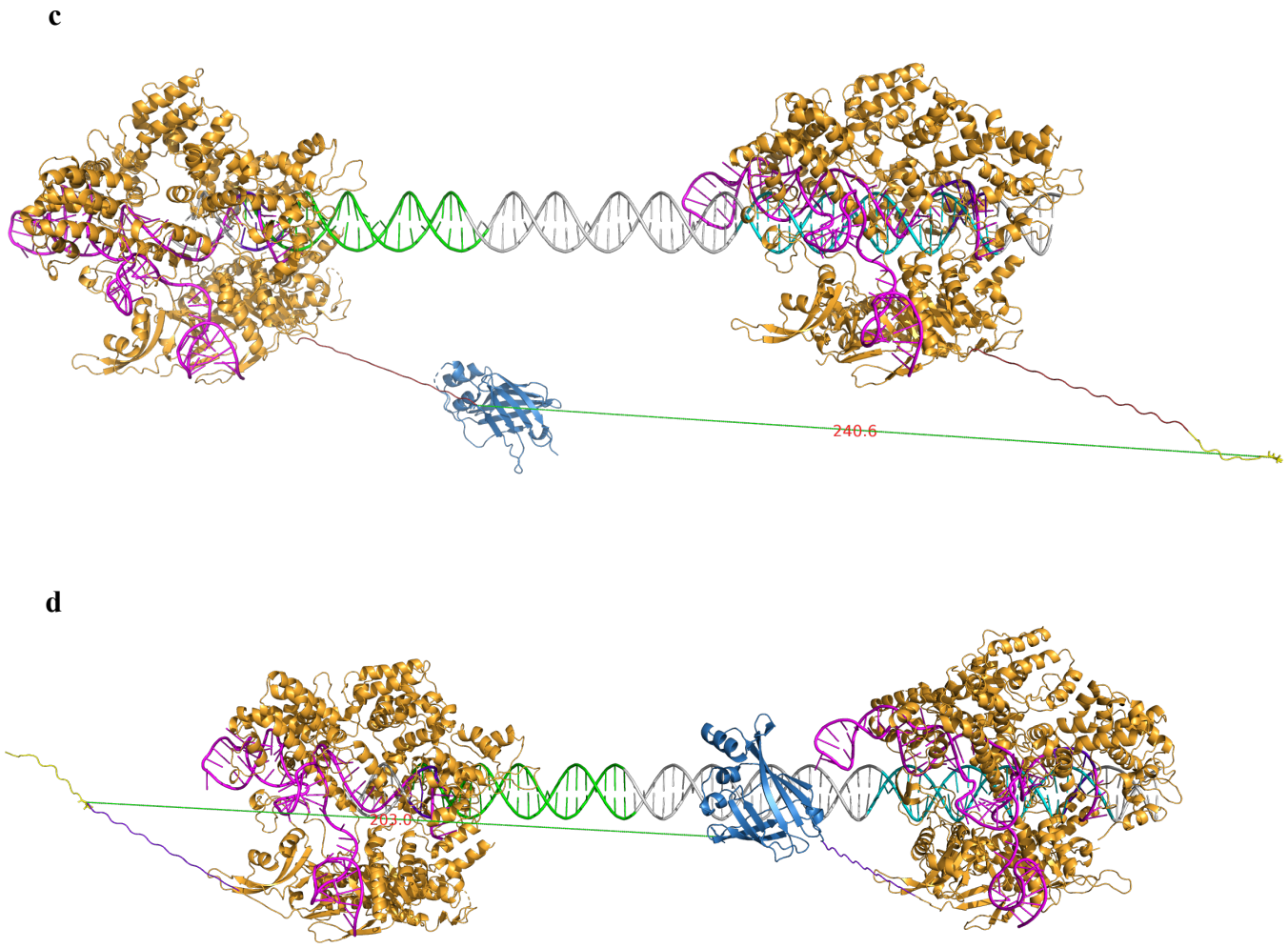


Figure 2.8: Split NanoLuc DNA biosensor modeling on everted target sites

(a) LgBiT-dCas9 (left) and dCas9-SmBiT (right) binding to everted target DNA sites 25 bp apart. **(b)** SmBiT-dCas9 (left) and dCas9-LgBiT (right) binding to everted target DNA sites 25 bp apart. **(c)** LgBiT-dCas9 (left) and SmBiT-dCas9 (right) binding to everted target DNA sites 25 bp apart. **(d)** dCas9-SmBiT (left) and dCas9-LgBiT (right) binding to everted target DNA sites 25 bp apart. For all models, dCas9 is shown in orange, the sgRNA is shown in magenta, LgBiT is shown in blue, SmBiT is shown in yellow, the linker peptide on the dCas9 N-terminus is shown in red, the linker peptide on the dCas9

C-terminus is shown in purple, spacer DNA is shown in grey, the upstream target DNA sequence is shown in green, the downstream target DNA sequence is shown in cyan, and the 5'-NGG PAM sequence on the target DNA and the cognate 5'-CCN sequence on the sgRNA are both shown in purple. The observed distance between N-terminus of SmBiT and C-terminus of LgBiT in these models is shown as a dashed green line with red text. All models rendered in PyMOL (version 2.5.2) via ray-tracing feature.

Statistical mechanical model of probability of split NanoLuc DNA biosensor producing light strictly when bound to one or both target DNA sequences

After modeling which combinations of NanoBiT-dCas9 and dCas9-NanoBiT fusion proteins might be expected to reassemble NanoLuc luciferase most efficiently on various combinations of target DNA orientations and spacings, we were interested in determining additional factors besides structural and spatial elements which could affect efficiency of such a split reporter DNA biosensor. Statistical mechanical models of biochemical systems aim to account for the macroscopic behavior of such systems in terms of kinetics and thermodynamics laws that apply to microscopic components of these systems and the probabilistic assumptions made about them at equilibrium[15,16]. Therefore, we wished to model the probability of luminescence strictly in the bound state to predict the favorability of light production upon binding to one or both target DNA sequences compared to binding events that should not produce detectable luminescence. To this end, based on the low affinity dimerization interface for SmBiT and LgBiT and the binding affinity of each dCas9-sgRNA-NanoBiT complex for its DNA target site, we developed a statistical mechanical model for dCas9-NanoBiT dimerization events that should and should

not reassemble NanoLuc luciferase upon binding to DNA and considered the multitude of different states such a system could occupy, as shown in **Figure 2.9** below.

For such probabilistic predictions, one primary concern was the dissociation constant between SmBiT and LgBiT, shown to be relatively high at 190 μM (k_{on} = 500 M^{-1} s^{-1}, k_{off} = 0.2 s^{-1})[3]. We reasoned that, for a split reporter reassembly biosensing system, a low affinity dimerization surface for the truncations might result in an improved signal-to-background ratio (SBR) due to potential reduction in auto-assembly of the reporter subunits in the possible unbound states of the biosensor. Another important parameter was the binding affinity of each dCas9-sgRNA-NanoBiT complex for its DNA target site, given by the K_D of a specific sgRNA-DNA interaction. There has been a wide range in reported K_D for the d*Sp*Cas9-gRNA complex *in vivo*. The d*Sp*Cas9–sgRNA complex interacts tightly with a perfectly matched DNA site *in vitro*, with a K_D ~1.3 nM by bio-layer interferometry[17]. However, much weaker K_D observations were reported at ~68 nM using microscale thermophoresis[18] and ~105 nM by fitting *in vivo* repression data[19]. In our modeling, we used the latter figure for sgRNA-DNA K_D (**Figure 2.9**).

Our model in **Figure 2.9c** indicates that the probability of luminescence strictly in the bound state of the biosensor, P_{detect_one}, or the fraction of time that the biosensor can be expected to produce signal as a result of bona fide sensing of at least one of the target DNA sequences in all instances when it is bound, is dependent on four main parameters: (1) the dimerization of the two dCas9 monomers, given by the NanoBiT dissociation constant $K_{D\_NanoBiT}$ (190 μM); (2) the binding affinity of each dCas9–sgRNA complex for its DNA site, given by its dissociation constant, K_{DNA} (nM); (3) the fractional saturation of each of the dCas9 proteins with cognate

sgRNAs, λ; and (4) the total concentrations of each dCas9, [C_{total}] (nM). With the assumption that NanoBiTs reassemble as long as both dCas9-sgRNA-NanoBiT monomers bind their DNA target sites (**Figure 2.9** species 3) and should not be impeded much by unproductive binding events, the probability P_{detect_one} can be obtained by dividing the sum of the probabilities of the states where NanoBiTs reassemble when bound (bound and producing light, or P_{BL}) divided by the sum of probabilities of all bound states—which includes those bound and producing light (P_{BL}) and those bound and dark (P_{BD}), or:

$$\frac{P_{BL}}{P_{BL} + P_{BD}} = \frac{\sum P(species\ 3\ through\ species\ 15)}{\sum P(species\ 1\ through\ species\ 15)}$$

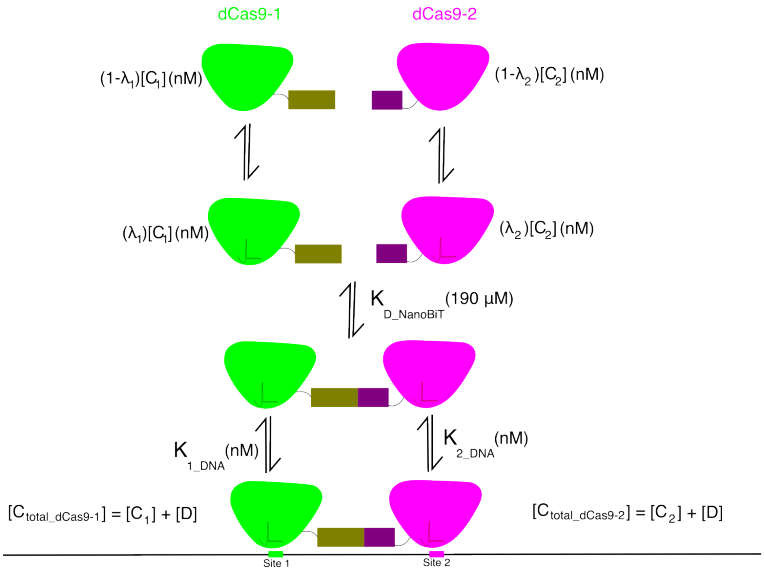
In addition, perhaps a more useful probability that can be derived from this statistical mechanical model is the likelihood of light production from binding of both DNA sequences targeted by our biosensor, P_{detect_both}. This probability is useful for reporting on how well our biosensor might be expected to work for true detection of both sequences of interest, which could have implications for users of the biosensor who desire targeting of longer stretches of sequence covered by both sgRNAs. P_{detect_both} can be obtained by dividing the sum of the probabilities of the states where NanoBiTs reassemble when bound to both target sites (bound and producing light from both, or P_{BL_both}) divided by the sum of probabilities of all bound states, or:

$$\frac{P_{BL\_both}}{P_{BL} + P_{BD}} = \frac{\sum P(species\ 3\ and\ species\ 12\ through\ species\ 15)}{\sum P(species\ 1\ through\ species\ 15)}$$

We then used our model to determine the effect of varying $K_{D\_NanoBiT}$, K_{DNA}, λ, and $[C_{total}]$ on P_{detect_one} and P_{detect_both}. Interestingly, from our models, we observed that at constant K_{DNA}, λ, and C_{total}, P_{detect_one} holds relatively constant when $K_{D\_NanoBiT}$ is decreased, while P_{detect_both} increases at lower $K_{D\_NanoBiT}$ **(Figure 2.9e, 2.9i)**. This means that, when considering bound states only, a higher affinity split reporter should increase the likelihood of detection of both DNA target sites. However, this analysis does not consider the unbound states of our biosensor which would be disfavored by including a lower affinity split reporter. We decided to keep the NanoBiT system as our split luminescent reporter of choice for biosensor design as we theorized that it could still positively affect signal-to-background, especially if background auto-association proved to have a stronger negative effect overall on signal-to-background than the positive effect caused by increasing the favorability of producing light from binding both sequences when including a higher affinity split reporter transducer. In addition, we observed higher values for both P_{detect_one} and P_{detect_both} as the value of K_{DNA} was decreased, up to a plateau in both probabilities starting at ~100-1000 nM K_{DNA} **(Figure 2.9f, 2.9j)**. Thus, we concluded higher affinity sgRNA-DNA interactions than ~100-1000 nM should have little effect on P_{detect_one} and P_{detect_both} while lower affinity sgRNA-DNA interactions should likely be avoided. Furthermore, our models led us to conclude that varying fractional saturation of dCas9 with sgRNA had little effect on both P_{detect_one} and P_{detect_both} overall **(Figure 2.9g, 2.9k)**. Finally, from our models, it was apparent that when varying $[C_{total}]$, both P_{detect_one} and P_{detect_both} were highest at ~0.998 and ~0.499, respectively, when $[C_{total}] = K_{DNA}$ **(Figure 2.9h, 2.9l)**. Thus, higher expression levels of our biosensor might be desirable, at least up to the K_{DNA} for any given sgRNA-DNA pairing.

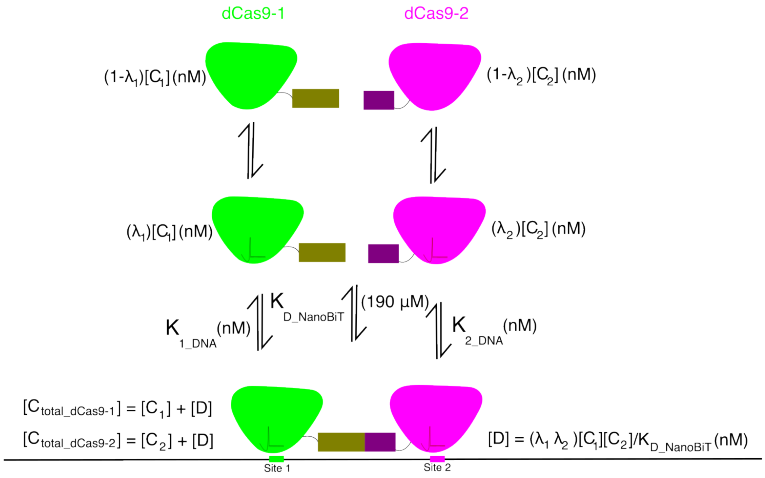
a

Scenario 1: NanoBiTs reassemble in nucleus but not on DNA



b

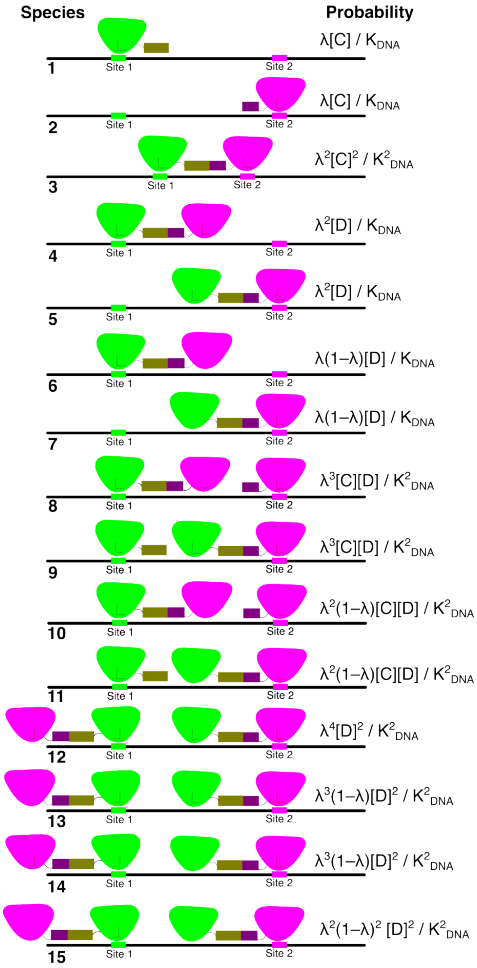
Scenario 2: NanoBiTs reassemble on DNA



c

Parameters:
λ = fractional saturation of dCas9 with gRNA
[C] = concentration of dCas9-NanoBiT monomer
[D] = concentration of dCas9-NanoBiT-NanoBiT-dCas9 dimer
$K_{D\_NanoBiT}$ = dissociation constant of NanoBiTs

Weights:
$K_{DNA} = K_{1\_DNA} = K_{2\_DNA}$
$\lambda = \lambda_1 = \lambda_2$
$C_{total} = C_{total\_1} = C_{total\_2}$



d

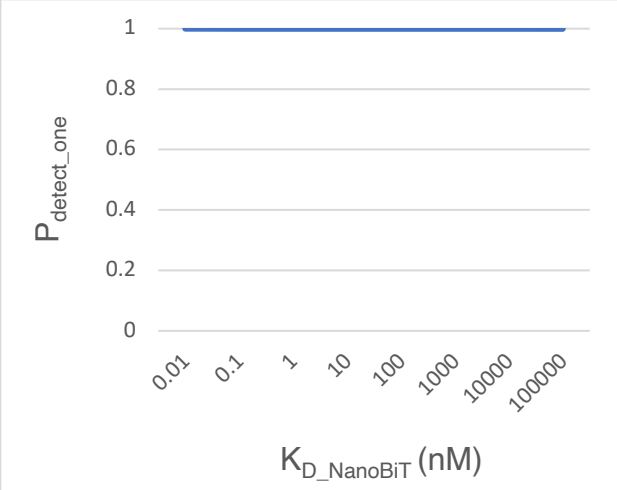
Total probability space $P_{BL} + P_{BD} + P_{UL} + P_{UD} = 1$

$$P_{BL} = P_{BL\_one} + P_{BL\_both}$$

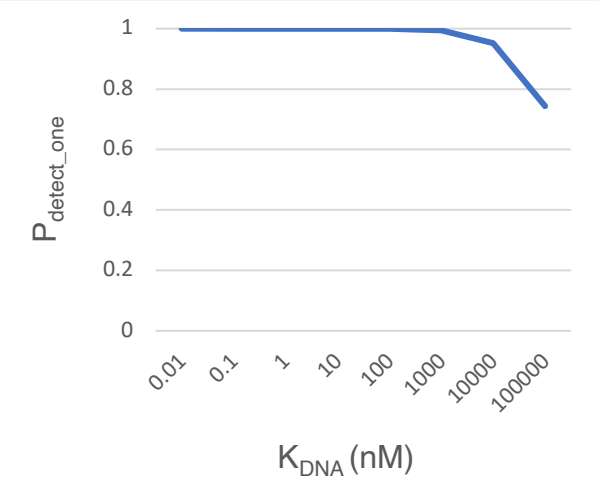
Model probability of detecting at least one DNA sequence while bound : $\frac{P_{BL}}{P_{BL} + P_{BD}}$

Model probability of detecting both DNA sequences while bound : $\frac{P_{BL\_both}}{P_{BL} + P_{BD}}$

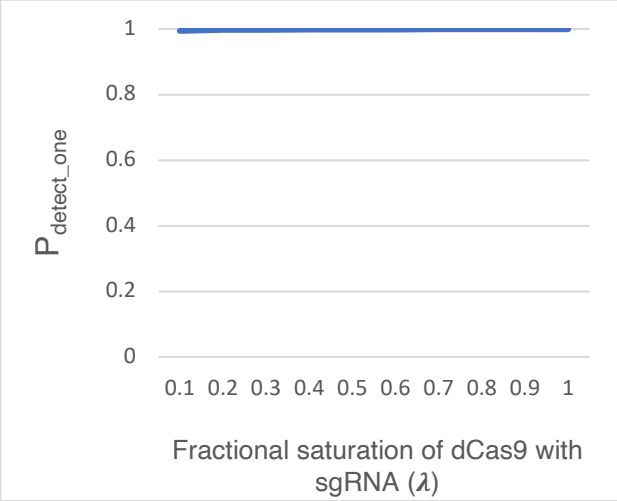
e


1
0.8
0.6
0.4
0.2
0
P detect_one
0.01
0.1
1
10
100
1000
10000
100000
K D_NanoBiT (nM)

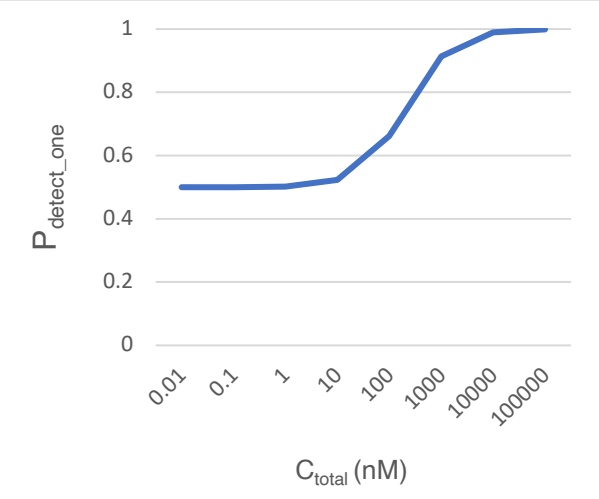
f


1
0.8
0.6
0.4
0.2
0
P detect_one
0.01
0.1
1
10
100
1000
10000
100000
K DNA (nM)

g


1
0.8
0.6
0.4
0.2
0
P detect_one
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Fractional saturation of dCas9 with sgRNA (λ)

h


1
0.8
0.6
0.4
0.2
0
P detect_one
0.01
0.1
1
10
100
1000
10000
100000
C total (nM)

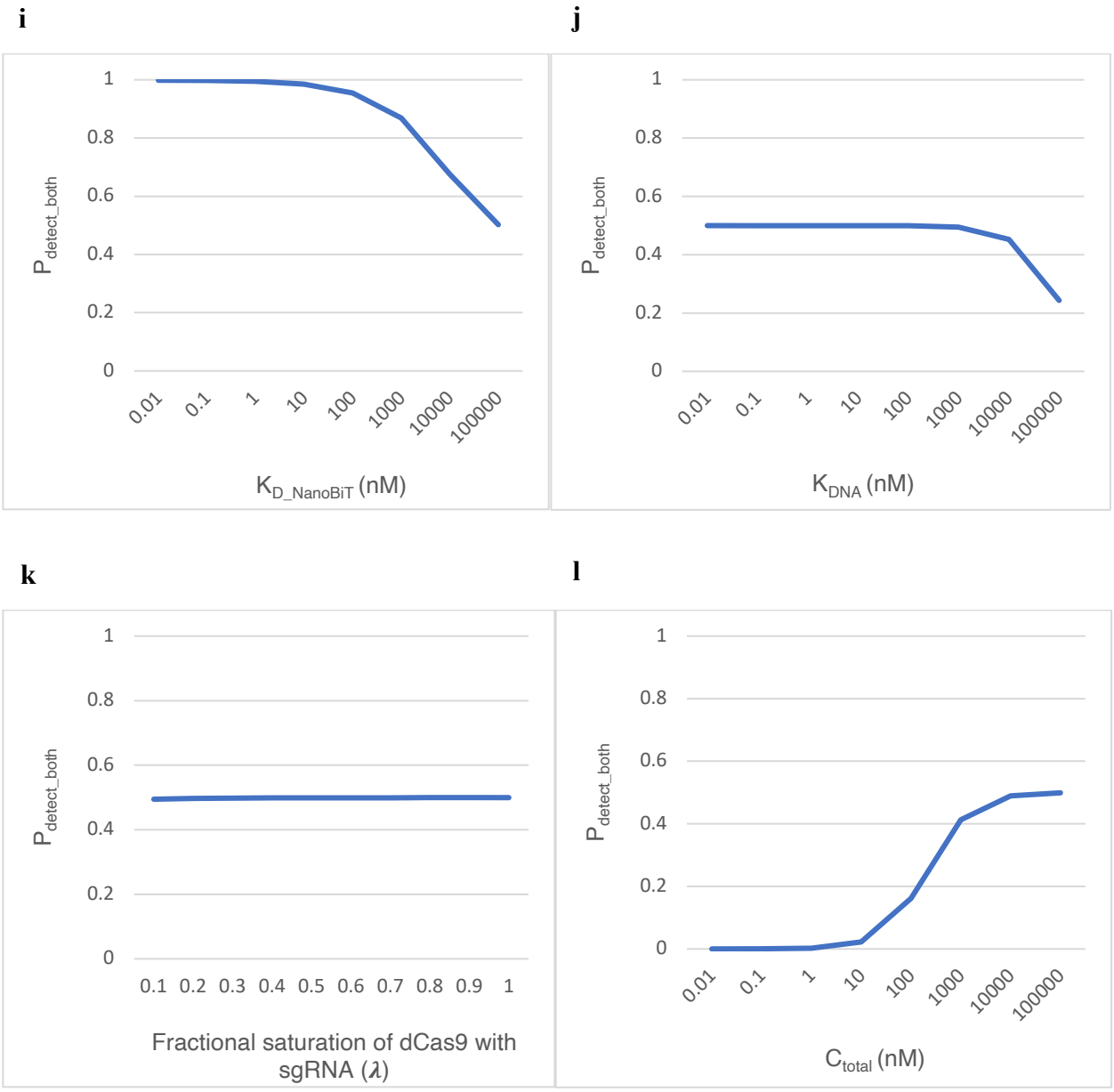


Figure 2.9: Statistical mechanical model of probabilities of split NanoLuc DNA biosensor producing light strictly when bound to one or both target DNA sequences

(a) One possible scenario where NanoBiTs reassemble in the background of the nucleus but not when bound to target DNA. Subscripts 1 and 2 denote two orthogonal dCas9s. **(b)** Another possible scenario where NanoBiTs reassemble NanoLuc luciferase coincident with binding. Subscripts 1 and 2 denote two orthogonal dCas9s. **(c)** Statistical mechanical

model of possible states for dCas9-NanoBiT system binding to target DNA, assuming the same parameters for each dCas9. Model parameters: $K_{D\_NanoBiT}$ = dimerization constant of LgBiT and SmBiT, K_{DNA} = dissociation constant of dCas9-sgRNA complex and DNA, [C] = concentration of dCas9 monomer, [D] = concentration of dCas9 dimer, $[C_{total}]$ = total concentration of one dCas9, and λ = fractional saturation of dCas9 with its cognate sgRNA. **(d)** Determination of the equations representing the probability of light production from binding at least one site on the DNA when the probe is in the bound state, P_{detect_one}, and the probability of light production from binding both DNA target sites when the probe is in the bound state, P_{detect_both}. **(e)-(h)** Modeling the probability of detection of at least one DNA sequence, P_{detect_one}. For this series of plots, one parameter was altered while the others were kept constant: $K_{D\_NanoBiT}$ = 190 μM, K_{DNA} = 105 nM, λ = 1, $[C_{total}]$ = 1.05 x 10^2 nM. For the case where $[C_{total\_1}] = [C_{total\_2}]$, [C] for each dCas9 was calculated as $((\sqrt{(1 + 4[C_{total}]/K_{D\_NanoBiT})})–1)/(2/K_{D\_NanoBiT})$ and [D] was calculated as $[C_{total}] – [C]$[20]. **(i)-(l)** Modeling the probability of detection of both DNA sequences, P_{detect_both}. For this series of plots, one parameter was altered while the others were kept constant: $K_{D\_NanoBiT}$ = 190 μM, K_{DNA} = 105 nM, λ = 1, $[C_{total}]$ = 1.05 x 10^2 nM. For the case where $[C_{total\_1}] = [C_{total\_2}]$, [C] for each dCas9 was calculated as $((\sqrt{(1 + 4[C_{total}]/K_{D\_NanoBiT})})–1)/(2/K_{D\_NanoBiT})$ and [D] was calculated as $[C_{total}] – [C]$[20].

MATERIALS AND METHODS

Preliminary Experiments: Transfection of DNA Biosensor with Plasmid DNA and SBR Analysis in Cell Lysates

Target plasmids containing repeated sgRNA target sites A and B in two orientations (tandem and inverted) with spacer lengths of 20-50 bp were received as a gift from Dr. Jason Low at The University of Sydney. Four directional reporter fusion constructs between the LgBiT and SmBiT of NanoLuc luciferase and d*Sp*Cas9 (D10A, H840A) were generated using the Gibson Assembly method. A positive control full NanoLuc-dCas9 fusion construct was created using overlap extension PCR on LgBiT and SmBiT gBlocks to directionally splice the sequences together followed by Gibson Assembly. The four assembled dCas9-NanoBiT constructs, the dCas9-NanoLuc construct, and pGL4.53 [luc2/PGK] Firefly luciferase vector (Promega Corporation) were transformed into High Efficiency NEB 5-alpha Competent *E. coli* (New England Biolabs) using a standard chemical transformation procedure with heat shock at 42°C and transformed *E. coli* were plated on LB plates containing ampicillin at a final concentration of 100 µg/mL. After an 18 h incubation at 37°C, MiniPreps (QIAGEN) were created for a subset of large colonies. The selected subset of large colonies was screened for recombinant vector + insert using both diagnostic restriction digests and colony PCR.

For the *in vitro* sequence detection assay in cell lysates, 152 total wells of HEK293T cells in 24-well plates (Thermo Fisher Scientific) were transiently transfected with 490 ng total DNA per well using Lipofectamine 3000 (Thermo Fisher Scientific) and subsequently plated in opaque side translucent bottom 96-well microplates (Corning). All 150 wells were transfected with 70 ng

per well of the two target sequence vectors (A and B) and 70 ng per well pGL4.53 [luc2/PGK] Firefly luciferase vector. 144 total wells were transfected with the four possible LgBiT-dCas9, dCas9-LgBiT, SmBiT-dCas9 and dCas9-SmBiT fusion protein combinations in LgBiT:SmBiT molar ratios of 0:1 (SmBiT alone), 1:50, 1:10, 1:4, 1:2, 1:1.33, 1:1, 1.33:1, 2:1, 4:1, 10:1, 50:1, and 1:0 (LgBiT alone), the construct in excess being transfected at 11 fmol per well and the lesser construct being decreased to 8, 5.5, 2.75, 1.1, and 0.2 fmol transfected per well. The 1:1 ratio of LgBiT:SmBiT was also decreased from 11 fmol transfected of each NanoBiT per well to 8 fmol of each, 5.5 fmol of each, 2.75 fmol of each, 1.1 fmol of each, and 0.2 fmol of each to observe differences in background signal at several lower total fusion protein expression levels. 72 of these LgBiT + SmBiT wells were transfected with 5.5 fmol per well sgRNA for target sequence A and 5.5 fmol per well sgRNA for target sequence B (1:1 dCas9-NanoBiT plasmid:sgRNA plasmid molar ratio) and 72 of these LgBiT + SmBiT wells were not transfected with any sgRNA. For wells that did not reach 490 ng total DNA, inert pUC19 vector was transfected to make up the difference. Four wells were transfected with only pUC19 vector as a negative control and four wells were transfected with the full NanoLuc-dCas9 fusion as a positive control. Following the Nano-Glo Dual-Luciferase Reporter Assay System protocol, we normalized NanoLuc luciferase luminescent signals to those of the constant co-transfected Firefly luciferase control reporter.

Preliminary Experiments: 3X-FLAG and HA Epitope Affinity Protein Purification for dCas9-NanoBiT DNA Sequence Biosensor Ribonucleoprotein Production

First, HA-tagged and 3X FLAG-tagged proteins were isolated by affinity purification with epitope antibody-conjugated agarose beads. Before starting, Tris-buffered saline (TBS), 0.1M

glycine, 1X RIPA buffer (Cell Signaling Technology), 1X Dulbecco's phosphate-buffered saline (DPBS) without Ca2+ and Mg2+ (Thermo Fisher Scientific), and elution buffers containing 3X FLAG peptide (Sigma-Aldrich) and HA peptide (Genscript) were prepared and placed on ice. 1X TBS (50 mM Tris-HCl 50 mM NaCl) was supplemented with 1 mM EDTA, 0.02% w/v sodium azide, and 1X protease/phosphatase inhibitor (Cell Signaling Technology) and brought to a final pH of 7.5. EDTA was added to 1X TBS and elution buffers to chelate divalent metal ions and render metalloproteases inactive while sodium azide was added to 1X TBS and elution buffers as an antimicrobial agent to ensure protein stability in the buffer. The 3X-FLAG and HA elution buffers were prepared with 3X FLAG and HA peptides at a final concentration of 400 μg/mL in 1X TBS supplemented with 1 mM EDTA, 0.02% w/v sodium azide, and 1X protease/phosphatase inhibitor (final pH: 8.0).

HEK 239T cells (ATCC) were passaged several times post-thaw until passage 5 and grown until ~80-90% confluent. Cells were split 1:2.5 and plated to 10-cm plates. 16-20 hrs later (based on observed average doubling time for this batch of cells of approximately 12 h), Lipofectamine 3000 (Thermo Fisher Scientific) transfections were prepared using 7 μg total DNA coding for dCas9-NanoBiT, NanoBiT-dCas9, or NanoLuc-dCas9 fusion proteins driven by CMV-promoter + 7 μg pMAX-GFP, 28 μL P3000 reagent and 43.4 μL Lipofectamine 3000 reagent per transfection complex. Briefly, transfection mixtures were incubated 5-10 min to form complexes then the entire volume was slowly added dropwise to cells. Cells were incubated for 24 hrs and GFP expression was checked over several areas as a measure of percent transfection efficiency. All plates showed 90-95% transfection efficiency. Cells were incubated for an additional 48 hrs, then 2 mL 1X TypLE Express Enzyme (Thermo Fisher Scientific) was added to each 10-cm

plate and plates were incubated at 37 °C for 2-3 min until cells detached. 8 mL prewarmed Dulbecco's Modified Eagle Medium (DMEM) + Glutamax (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (Genesee Scientific) and 1X penicillin-streptomycin (Thermo Fisher Scientific) was added to neutralize TypLE reagent and cells were spun down for 5 min at 500 x g. Cell pellets were rinsed once with 10 mL ice-cold 1X DPBS and spun down for 5 min at 500 x g. Cells were lysed by resuspending cell pellet in 1 mL ice-cold 1X RIPA buffer supplemented with 1X protease/phosphatase inhibitor cocktail and incubated on ice for 15 min. Cells were sonicated 3 times with pulsing parameters: 2 s on, 30 s off. Cells were incubated for another 15 min on ice post-sonication and lysates were spun down at 3,000 x g for 10 min at 4 °C. Anti-FLAG affinity matrix (Sigma Aldrich) was thoroughly resuspended, and an aliquot of matrix was prepared that was 10% more than 150 µL x number of plates used in extraction. Allocated Anti-FLAG affinity matrix was centrifuged at 8,000 x g for 1 min and any supernatant was removed so only solid matrix was left in tubes. Unbound FLAG epitopes and potential contaminating proteins were removed from matrix by washing 1X with 1 mL ice-cold 0.1 M glycine and matrices were spun down at 8,000 x g for 1 min then supernatant was removed. The Anti-FLAG affinity matrix was then equilibrated by washing 2-3 times with 1 mL ice-cold 1X TBS, centrifuging at 8,000 x g and removing supernatant after each wash. 150 µL equilibrated Anti-FLAG matrix was added to 1 mL cell lysate and incubated on rotating tube platform overnight at 4 °C. Lysate/matrix mixtures were centrifuged at 8,000 x g for 1 min and supernatants (non-bound total protein cell lysates) were removed and stored at 4 °C for SDS-PAGE. The Anti-FLAG affinity matrix with bound protein was washed 3X with 1 mL ice-cold 1X TBS. To elute 3X-FLAG tagged proteins from matrix, 200 µL ice-cold 3X-FLAG peptide elution buffer with 400 µg/mL 3X FLAG peptide was added and elution tubes were incubated on

rotating tube platform at 4 °C overnight (16 hrs). Elution tubes were centrifuged at 8,000 x g for 2 min and supernatants were saved. Aliquots were saved for short-term storage (1 day to a few weeks) at 4 °C using 4X Protein Stabilizing Cocktail (Pierce/Thermo Fisher Scientific). For longer-term storage (1 month to 1 year), additional aliquots of the protein elutions were stored at -20 °C in 50% glycerol, 50% 3X-FLAG elution buffer.

Preliminary Experiments: SDS-PAGE for dCas9-NanoBiT DNA Sequence Biosensor Ribonucleoprotein Production

Protein concentration was measured using Protein A280 mode on NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) using the following equation to calculate molar absorption coefficients for each fusion protein: $\varepsilon = 5500(nW) + 1490(nY) + 125(nC)$, where nW, nY, and nC refer to the total numbers of tryptophan, tyrosine, and cysteine residues in each protein sequence. The concentration was then calculated using $c = A / \varepsilon L$ based on A values determined spectrophotometrically and L = 1 cm using setting 1 Abs (A) = 1 mg/mL. The protein concentrations were also calculated using the Bradford Assay (Pierce/Thermo Fisher Scientific) and Qubit Protein Assay (Thermo Fisher Scientific). Using an average of these three concentrations, 20 μg of each eluted protein was prepared with 1X Protein Loading Buffer (National Diagnostics) and loaded on a 4-20% Mini-Protean TGX precast SDS-PAGE gel (Bio-Rad) with Precision Plus Protein Dual Color Standards (Bio-Rad). SDS-PAGE was run in 1X NuPAGE MOPS SDS Running Buffer for 2 hrs at 100V to confirm appropriate protein sizes and specificity of the purification procedure. Isopropanol fixing solution was prepared for the gels (10% (v/v) acetic acid, 25% (v/v) isopropanol, 65% Milli-Q H_2O) and gels were fixed by covering with a thin layer of isopropanol and shaking for 30 min at room temperature. Then, the

fixing solution was poured off and gels were stained with 1X Bio-Safe Coomassie Stain (Bio-Rad) for 2 h at room temperature. Finally, the staining solution was poured off and gels were de-stained in 10% (v/v) acetic acid overnight at room temperature.

Preliminary Experiments: *In vitro* Transcription of sgRNAs for dCas9-NanoBiT DNA Sequence Biosensor Ribonucleoprotein Production

To produce sgRNAs for complexing with affinity purified dCas9-NanoBiT, NanoBiT-dCas9, and NanoLuc-dCas9 fusion proteins to produce ribonucleoproteins (RNPs), we used the MEGAscript T7 Kit (Ambion, Invitrogen) for *in vitro* transcription (IVT). 5 μg template DNA containing the sgRNA scaffolds of interest, target sequence, and termination sequence driven by the U6 promoter was linearized downstream of the target sgRNA sequence using 25 units of restriction enzyme SpeI in a 1 hr digest at 37 °C. Template linearization reactions were run on a 1% w/v agarose gel to ensure no visible circular template remained. The template DNA was then treated with proteinase K (100–200 μg/mL), then 1/20th volume 0.5 M EDTA and 0.5% SDS for 30 min at 50°C. This was followed by phenol/chloroform extraction (using an equal volume) and ethanol precipitation. Linearized DNA template was precipitated from solution using 1/10th volume 3M sodium acetate, 3 volumes of 100% ethanol, and 50 μg/mL RNA-grade glycogen (Thermo Fisher Scientific). Ethanol precipitation reactions were incubated for 15 min at room temperature and precipitated DNA was pelleted for 15 min in a tabletop microcentrifuge at top speed. The supernatant was removed, and tubes were spun again briefly. The residual fluid was removed with a very fine-tipped pipet and DNA was resuspended in EB buffer (QIAGEN) at a concentration of 0.5–1 μg/μL. 0.2-0.8 μg linearized DNA templates were used in each IVT reaction. ATP, CTP, GTP, and UTP tubes and buffer solutions were vortexed thoroughly before

preparing IVT reactions. An equimolar ratio of all ribonucleotides (8 μL total) was used with 2 μL T7 RNA polymerase in a 20 μL reaction volume and reactions were incubated at 37 °C for 4 h. Next, IVT reactions were further incubated for 15 min with 1 μL TURBO-DNase to digest the linearized template. IVT sgRNAs were then purified using phenol-chloroform extraction followed by ethanol precipitation. Finally, correct sgRNA size was confirmed for all sets of gRNAs used in *in vitro* and live cell assays on a 1.1% v/v denaturing formaldehyde agarose (1.5% w/v) gel. Briefly, after preparing and adequately cooling a 1.5% w/v agarose gel prepared in 43.5 mL DEPC-treated H_2O, ~1.5 mL 37% formaldehyde was added in the fume hood along with 5 mL 10X MOPS and 1 μg/mL EtBr.

Preliminary Experiments: Complexing Ribonucleoproteins and Activity Measurements

Before using any method to deliver RNPs to cells, we first conduced an assay for activity of the proteins. Briefly, 2 μL of each purified dCas9-NanoBiT, NanoBiT-dCas9, or NanoLuc-dCas9 fusion protein at 0.25 μM was added to 8.5 μL 20 mM HEPES with 150 mM KCl (pH 7.5) and 1 μL of each dCas9-SmBiT fusion protein at 0.25 μM was added to 1 μL of each dCas9-LgBiT fusion protein at 0.25 μM in 8.5 μL 20 mM HEPES with 150 mM KCl (pH 7.5) and these solutions were mixed with 25 μL reconstituted Nano-Glo Live Cell Assay substrate (Promega Corporation) in 96-well white opaque-side translucent bottom assay plates (Thermo Fisher Scientific) to measure luminescence on the SpectraMax M5 Microplate Reader (Molecular Devices) for the NanoLuc-dCas9 fusion protein and all combinations of LgBiT and SmBiT fusion proteins with dCas9. Once protein activity was evaluated, we proceeded to assay activity of RNP complexes with target DNA *in vitro*. Briefly, 2 μL purified dCas9-NanoBiT, NanoBiT-dCas9, and NanoLuc-dCas9 fusion proteins at 0.25 μM were complexed with sgRNAs at 1:1,

1:1.2, 1:2, and 1:3 molar ratios in 8.5 µL 20 mM HEPES with 150 mM KCl (pH 7.5) for 10 min at room temperature. Then, 3 µL target DNA at 0.002 nmol/µL was added and complexes were mixed with 25 µL reconstituted Nano-Glo Live Cell Assay substrate in 96-well white opaque-side translucent bottom assay plates (Thermo Fisher Scientific). NanoLuc luminescent signals were then measured on the SpectraMax M5 Microplate Reader 50 min, 100 min, 150 min, and 200 min after complexation to measure natural luminescence decay of the RNPs.

Preliminary Experiments: RNP Delivery to Cells

To deliver RNPs to cells, two different methods were used: cationic liposome-mediated delivery using Lipofectamine CRISPRMAX reagent (Thermo Fisher Scientific) and electroporation using the Neon Transfection System (Thermo Fisher Scientific). For both CRISPRMAX and electroporation RNP delivery methodologies, 100 ng target DNA plasmids and a 250 ng recombinant GFP (Abcam) protein transfection control were delivered to cells using electroporation on the Neon Transfection System (Thermo Fisher Scientific) 4 h prior to delivery of the RNP complexes. For delivery using CRISPRMAX, cells were plated in 96-well plates approximately 16-20 h prior to starting the procedure. In a 1.5 mL microcentrifuge tube, 105 ng dCas9-NanoBiT, NanoBiT-dCas9, or NanoLuc-dCas9 protein was mixed with 21 ng IVT sgRNA and 0.2 µL Cas9 Plus Reagent in 5 µL Opti-MEM Reduced-Serum Medium (Thermo Fisher Scientific), adding the Cas9 Plus Reagent last. Then, in a second 1.5 mL microcentrifuge tube, 0.3 µL CRISPRMAX Reagent was added to 5 µL Opti-MEM. The entire solution in the first tube was then immediately added to the second tube and the resulting mixture was incubated at room temperature for approximately 5-10 min. The RNP complexes were then added slowly dropwise to cells.

For delivery using Neon electroporation, 0.5-5 pmol of purified dCas9-NanoBiT, NanoBiT-dCas9, or NanoLuc-dCas9 fusion protein was added to 0.5-5 pmol of IVT sgRNA in 10 µL resuspension buffer R. The complexation mixture was incubated for 5-10 min at room temperature, then added to a cell type-specific number of cells and electroporated using cell type-specific sets of pulsing parameters. In initial RNP electroporation assays using plasmid-based target DNA, 1.5 x 10^5 HEK 293T cells were electroporated using 2 1150 V pulses of 20 ms width. In later RNP experiments using multiple cell types at endogenous *MUC4*, *PALB2*, and 8q24 loci, HEK 293T cells were electroporated in the same way but other cell lines were electroporated as follows: 5 x 10^5 HeLa or MCF7 cells were electroporated using 2 1150 V pulses of 20 ms width. 2 x 10^5 K562 or Jlat cells were electroporated using 1 1700 V pulse of 20 ms width. Finally, 1.2 x 10^5 HCT116 cells were electroporated using 1 1300 V pulse of 30 ms width.

In our first experiment in live cells, we varied the amount of the LgBiT-dCas9 fusion protein from 105 ng to 25 ng while adding dCas9-SmBiT at 4-fold and 10-fold molar excesses. All tests were conducted on target site scaffolds with tandem target sites 40 bp apart and with inverted target sites 7 bp apart in this experiment. In the next experiment where 12 different target site scaffolds were tested, 105 ng of dCas9-SmBiT was delivered in 4-fold and 10-fold molar excesses to LgBiT-dCas9. LgBiT-dCas9 and dCas9-SmBiT were delivered alone in these experiments as negative controls and NanoLuc-dCas9 was delivered as a positive control. In the experiments testing response of NanoLuc luminescence to decreasing target DNA concentration, pGL4.53 [luc2/PGK] Firefly luciferase vector (Promega Corporation), essentially random DNA

of approximately the same size with no binding sites with >5 bp homology with the protospacer of either gRNA, was added to the transfection mix. (100-n) ng pGL4.53 [luc2/PGK] was added in conditions where (n) ng of target sequence scaffold was subtracted from the original 100 ng transfected.

Preliminary Experiments: Early Luminescence Microscopy and Image Processing

Transfection experimental setup for microscopy sessions was identical to the setup for microplate reader sessions. In these experiments, low-passage HEK 293T cells were plated in SensoPlate 24 Well F-Bottom, Glass Bottom Black Microplates (Greiner Bio-One) and transfected identically to luminometer-based experiments. For experiments using the Leica DM6000B, instead of imaging whole well populations of adherent cells, cells were split to 1.5 x 10^5 cells/mL images of the cell suspensions were taken on Superfrost Plus Microscope Slides (Thermo Fisher Scientific) with Premium Cover Glass (Thermo Fisher Scientific). An optimized NanoLuc luminescence imaging protocol was developed for use on the Leica DM6000 B Fully Automated Upright Microscope equipped with the Leica DFC9000 GT sCMOS camera and the Exfo X-Cite 120 Fluorescence Illumination System in which cells were placed in a dark box with all light sources covered or off and lamp intensity was set to 0, exposure time was set to 30 s, and sCMOS gain was set to 2.0. The pMAX-GFP transfection normalization control was imaged using an exposure of 150 ms and sCMOS gain of 1.0. The WEKA Segmentation package[21] in Fiji (Image J) was used to delineate boundaries of cell nuclei and then integrate signal intensities within these regions after several training cycles. Raw 16-bit grayscale GFP images were recolored green, brightness was reduced, and contrast was enhanced in Fiji. Raw 16-bit grayscale NanoLuc images were recolored magenta, brightness and contrast were increased, and the

"remove outliers" and "despeckle" noise reduction functions were applied in Fiji (Image J). Following this, scattered speckled noise remained in these images, so the noise was carefully removed around the cell nuclear regions in the GNU Image Manipulation Program (GIMP) using the clone tool with radius 5.0. To merge GFP and NanoLuc images, we took one of two routes: we either directly merged color channels in Fiji (Image J), or if the NanoLuc signal was drowned out by the merge due to its disproportionate dimness, the two separate images were opened in GIMP, making the processed NanoLuc image the upper layer. Then, opacity of the NanoLuc layer was reduced to approximately 95% to visualize the NanoLuc signal.

Preliminary Experiments: IVIS Spectrum Imaging of Cell Clusters

For RNP-based experiments on the IVIS Spectrum Bioluminescence Imaging System, we again split cells to 1.5 x 10^5 cells/mL but suspended them in 7.5 mL Opti-MEM Reduced Serum Medium (Thermo Fisher Scientific) on 100mm Polystyrene Petri Dishes (Thermo Fisher Scientific). We developed an optimized imaging protocol on the IVIS using field of view C (FOV C=13.3 cm), 0 cm specimen height, medium binning, F/Stop of 1, excitation filter set to "block," emission filter set to "open," and exposure set to "auto." Within the IVIS Spectrum LivingImage software, we adjusted the scale of all images to be equal and compared signal-to-background ratios by drawing and integrating circular regions of interest (ROIs) around regions containing cell clusters as judged by presence of luminescent signal. Negative controls in initial IVIS experiments using target site scaffold vectors were cell clusters without target DNA transfected.

Statistical testing

Two-tailed Student's t-tests and Z-tests for signal-to-background analyses were conducted in Microsoft Excel 2016.

RESULTS

After initial molecular modeling to predict structures for various fusion proteins between dCas9 and the NanoBiTs of NanoLuc luciferase, we designed experiments to test feasibility of a split reporter DNA biosensing approach using dCas9 as an endogenous biorecognition element for DNA and the NanoBiT system as a transducer element. We primarily focused on several areas of study for initial proof-of-concept. These included *in vitro* testing of signal-over-background in cell lysates, initial investigation of various live cell microscopy methodologies for luminescence and creating dCas9-NanoBiT DNA sequence biosensor ribonucleoproteins (RNPs) to reduce background noise observed in our early assays using co-transfected plasmid target DNA.

Initial proof of concept: *In vitro* DNA sequence detection in HEK293T cell lysates

To test our hypothesis that target DNA sequences could possibly be detected above background in living cells using two dCas9-NanoBiT or NanoBiT-dCas9 binding events to reconstitute a split NanoLuc luciferase reporter, we first conducted an initial sequence detection assay in cell lysates. For this assay, we transfected dCas9-NanoBiT and NanoBiT-dCas9 fusion constructs directed by two sgRNAs to bind plasmids containing two target sequences, A and B, with 20 bp spacing between them in tandem and inverted orientations. As earlier work demonstrated the viability of using a split reporter reassembly approach to detect specific methylated and

unmethylated DNA sequences *in vitro*[22-27], we felt that an analogous approach could be used in living cells to noninvasively detect DNA sequences. Thus, we first designed an assay using the Nano-Glo Dual-Luciferase Reporter Assay System, which first lyses the cells and subsequently analyzes the luminescence levels from both NanoLuc luciferase and Firefly luciferase in these lysates. We believed this approach would be most comparable to previously demonstrated *in vitro* detection experiments and would yield approximately equal levels of expressed fusion proteins in the lysates to total levels of expressed fusion proteins in living cells. Consequently, we believed this approach would allow us to better understand background levels of NanoLuc luminescence via auto-association of the LgBiT and SmBiT components. We predicted that, in living cells, LgBiT and SmBiT might co-localize to the nucleus and produce luminescence without direct association through dCas9 enzymes binding the target DNA sequence, which would effectively constitute a relevant background value to account for in any subsequent live cell sequence detection assays. To quantify this predicted spontaneous LgBiT-SmBiT background association, we compared normalized luminescent signals of wells transfected with sgRNAs for target DNA sequences A and B to those of corresponding wells not transfected with any sgRNAs. Normalized NanoLuc luminescent signal data from this initial sequence detection assay are presented in **Table 1** below. We defined the signal-to-background ratio (SBR) as the ratio of the average normalized NanoLuc luminescent signal in the conditions transfected with gRNAs to target sites A and B compared to the average normalized NanoLuc luminescent signal in the condition with no sgRNAs transfected. For control conditions, we observed that the wells transfected with pUC19 plasmid alone as a negative control showed an average SBR of 1.0, while the wells transfected with equal moles of NanoLuc-dCas9 as a positive control showed a SBR between 1.0 and 1.12. In conditions where dCas9-LgBiT was co-expressed with dCas9-

SmBiT, the wells transfected with a LgBiT:SmBiT molar ratio of 2:1 showed the highest SBR of 3.01. A paired Student's t-test for sample means (two-tailed, α=0.05) between +sgRNA signals and –sgRNA signals for this orientation resulted in a *p*-value of 0.00946. In conditions where LgBiT-dCas9 was co-expressed with SmBiT-dCas9, the wells transfected with a LgBiT:SmBiT ratio of 1:1 at 8 fmol of each transfected per well showed the highest SBR of 5.96. A paired Student's t-test for sample means (two-tailed, α=0.05) between +sgRNA signals and –sgRNA signals for this orientation resulted in a *p*-value of 0.01352. In conditions where LgBiT-dCas9 was co-expressed with dCas9-SmBiT, the wells transfected with a LgBiT:SmBiT molar ratio of 1:4 showed the highest SBR of 9.19. A paired Student's t-test for sample means (two-tailed, α=0.05) between +sgRNA signals and –sgRNA signals for this orientation resulted in a *p*-value of 0.00135. In conditions where dCas9-LgBiT was co-expressed with SmBiT-dCas9, the wells transfected with a LgBiT:SmBiT molar ratio of 1:1 showed the highest SBR of 6.71. A paired Student's t-test for sample means (two-tailed, α=0.05) between +sgRNA signals and –sgRNA signals for this orientation resulted in a *p*-value of 0.02774. For all four orientations of LgBiT and SmBiT fusion proteins with dCas9, SBRs for all LgBiT:SmBiT molar ratios are shown in **Figure 2.10**. In addition, SBRs for all 1:1 LgBiT:SmBiT transfection ratio conditions for all four orientations are shown in **Figure 2.11**.

Fusion Constructs Transfected	Normalized Signal Range (-sgRNA)	Normalized Signal Range (+sgRNA)	+sgRNA/-sgRNA Ratio Range
pUC19	0.0002-0.0004	0.0002-0.0004	1.00-1.00
NanoLuc-dCas9	38.5-40.6	40.6-43.0	1.00-1.12
SmBiT-dCas9	0.00119-0.00137	0.00120-0.00134	0.88-1.13
dCas9-SmBiT	0.000976-0.00139	0.000959-0.00147	0.99-1.06
LgBiT-dCas9	0.00112-0.00121	0.00111-0.00123	0.92-1.02
dCas9-LgBiT	0.00114-0.00119	0.00106-0.00123	0.90-1.08
dCas9-LgBiT + dCas9-SmBiT	0.225-0.505	0.520-0.677	1.03-3.01
LgBiT-dCas9 + SmBiT-dCas9	0.025-0.063	0.091-0.149	1.44-5.96
LgBiT-dCas9 + dCas9-SmBiT	0.066-0.314	0.405-0.603	1.29-9.19
dCas9-LgBiT + SmBiT-dCas9	0.089-0.248	0.305-0.596	1.23-6.71

Table 1: Columns 1 and 2 show normalized signal ranges for various constructs transiently transfected to HEK293T cells with and without sgRNA in an initial sequence detection assay in cell lysates. NanoLuc luminescence values are normalized to Firefly luciferase luminescence values to correct for variations in transfection efficiency, cell number, or cell viability. The rightmost column shows ranges of signal-to-background ratio, defined as normalized signal in the condition with sgRNA transfected divided by normalized signal in the condition without sgRNA transfected for the four possible combinations of fusion constructs. From top to bottom, rows show normalized NanoLuc luminescence from pUC19 vector expressed alone, NanoLuc-dCas9 expressed alone, SmBiT-dCas9 expressed alone, dCas9-SmBiT expressed alone, LgBiT-dCas9 expressed alone, dCas9-LgBiT expressed alone, dCas9-LgBiT expressed with dCas9-

SmBiT, LgBiT-dCas9 expressed with SmBiT-dCas9, LgBiT-dCas9 expressed with dCas9-SmBiT, and dCas9-LgBiT expressed with SmBiT-dCas9.

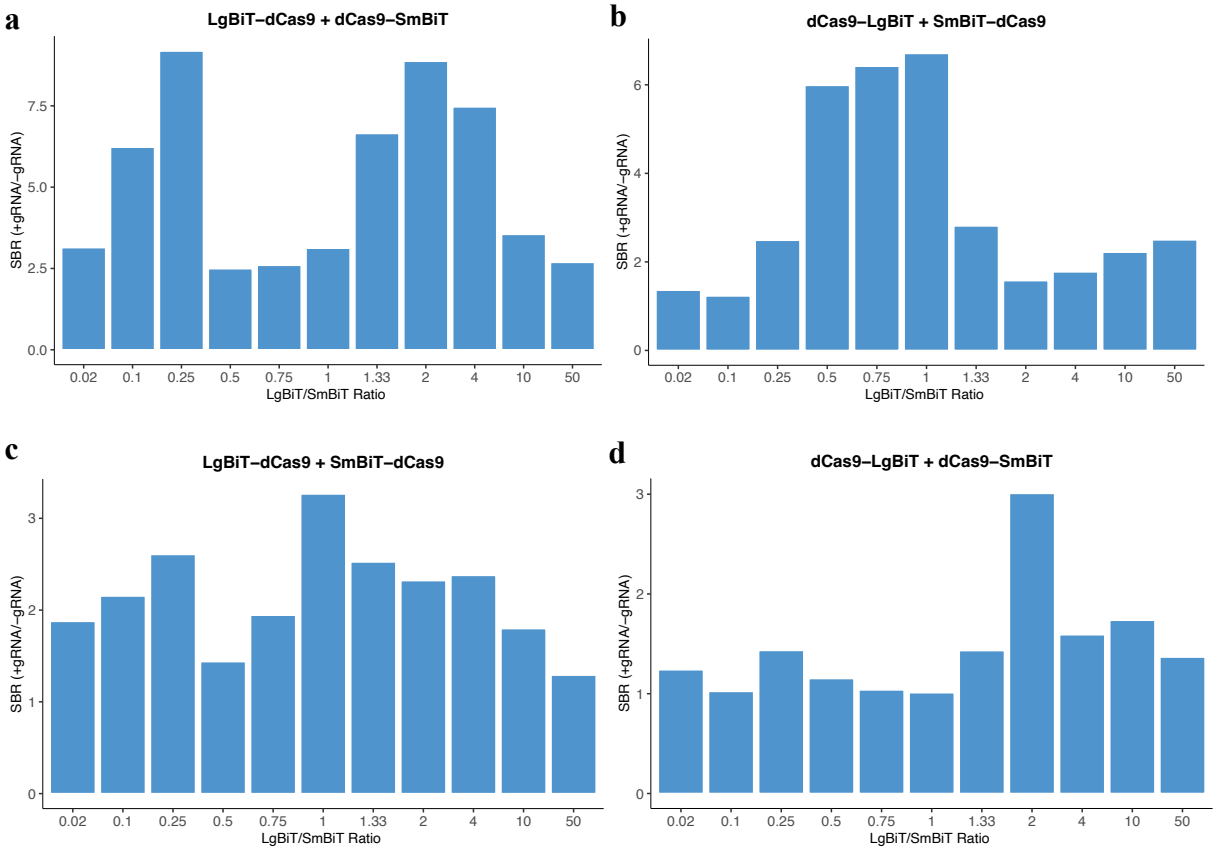


Figure 2.10: Variation of signal-to-background ratio due to variation in LgBiT/SmBiT molar ratio

(a) +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various transfection ratios of LgBiT-dCas9 and dCas9-SmBiT in the initial sequence detection assay. **(b)** +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various transfection ratios of dCas9-LgBiT and SmBiT-dCas9 in the initial sequence detection assay. **(c)** +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various

transfection ratios of LgBiT-dCas9 and SmBiT-dCas9 in the initial sequence detection assay. **(d)** +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various transfection ratios of dCas9-LgBiT and dCas9-SmBiT in the initial sequence detection assay.

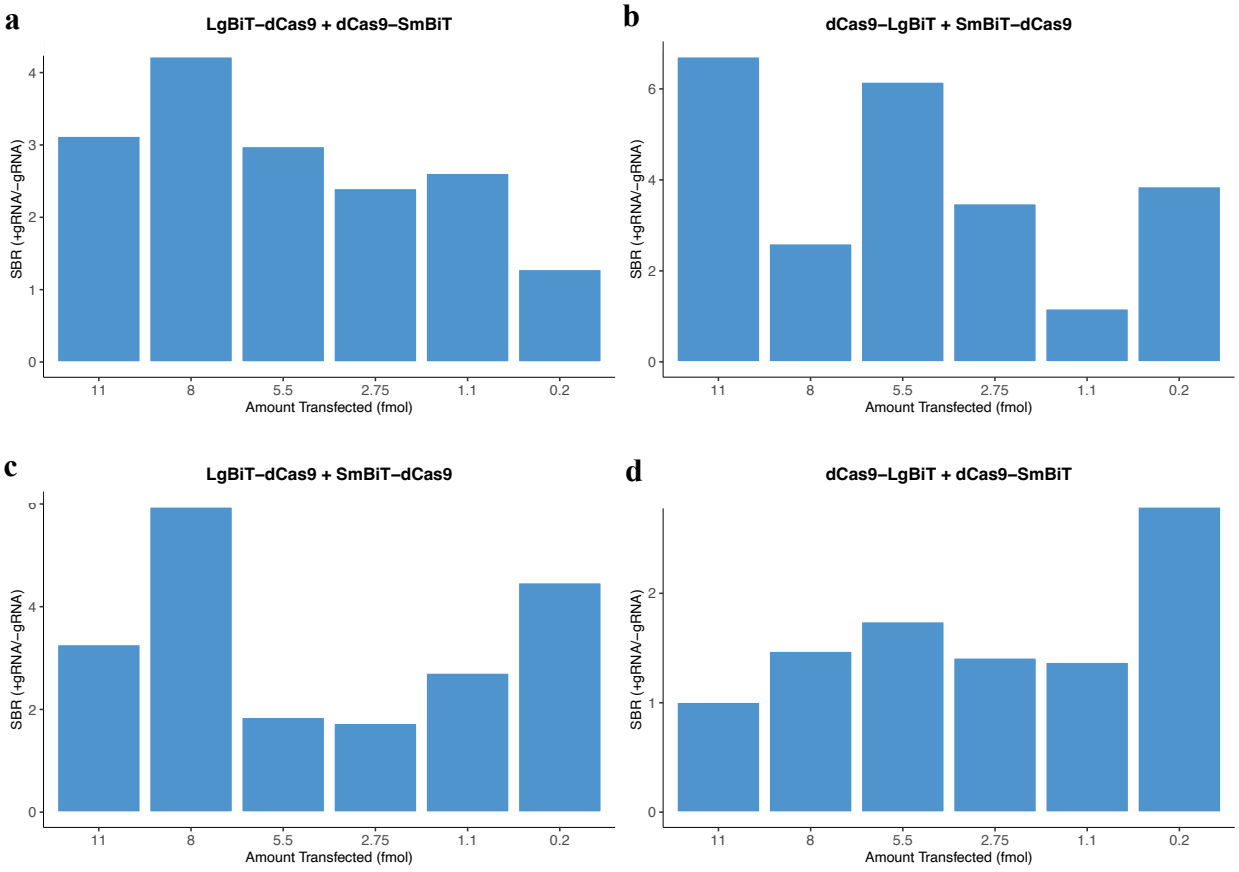


Figure 2.11: Variation of signal-to-background ratio due to variation in 1:1 molar amount LgBiT and SmBiT constructs transfected

(a) +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various 1:1 LgBiT-dCas9 and dCas9-SmBiT plasmid molar amounts transfected in the initial sequence detection assay. **(b)** +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various 1:1 dCas9-LgBiT and SmBiT-dCas9 plasmid molar amounts transfected in the initial

sequence detection assay. **(c)** +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various 1:1 LgBiT-dCas9 and SmBiT-dCas9 plasmid molar amounts transfected in the initial sequence detection assay. **(d)** +sgRNA/-sgRNA normalized NanoLuc signal ratios (signal-to-background) for various 1:1 dCas9-LgBiT and dCas9-SmBiT plasmid molar amounts transfected in the initial sequence detection assay.

Testing an RNP-based DNA biosensor delivery approach in live cells

Due to relatively high background signal in the negative control cell populations with no target DNA transfected in initial assays testing plasmid-based expression of our dCas9-NanoBiT biosensor, we theorized that delivery of the dCas9-NanoBiT fusion proteins and gRNAs as ribonucleoprotein complexes (RNPs) might provide better control of initial nuclear protein concentration and allow it to decrease steadily after administration in contrast to the large increase and slow decrease associated with plasmid-based expression. The steadily decreasing RNPs might therefore provide a strong target signal while reducing the background signal, resulting in more sensitive detection of the DNA target sequence of interest. Thus, we expressed and purified fusion proteins from HEK 293T using immunoprecipitation **(Figure S2.1a)**, complexed them with *in-vitro*-transcribed gRNAs (**Figure S2.1b)**, and validated NanoLuc signal output from the resulting dCas9-NanoBiT RNPs and from the NLuc-dCas9 RNP 1 min post-complexation **(Figure 2.12a)**. Notably, relative signal differences *in vitro* between the dCas9-NanoBiT RNPs binding target DNA and NLuc-dCas9, the LgBiT alone, and the SmBiT alone controls remained largely identical with the exception of the background signal from auto-

association of LgBiT and SmBiT, which was markedly lower relative to all other signals compared to previous plasmid-based delivery experiments **(Figure 2.12a)**.

Based on initial ANOVA and Tukey HSD statistical analysis of data from our first *in vitro* characterizations of the dCas9-NanoBiT DNA biosensing system, we initially chose target sequence plasmids with tandem target sites 40 bp apart and inverted target sites 7 bp apart along with the LgBiT-dCas9 and dCas9-SmBiT fusion protein pairing. Signal output decayed *in vitro* when 560 fmol total LgBiT-dCas9 and dCas9-SmBiT RNPs were mixed with 40 fmol tandem 40-bp and inverted 7-bp target DNA plasmids to the point where 59% and 57% of the original signal was present 200 min after complexation, respectively **(Figure 2.12b-c)**. These results align well with the observed signal half-life for NanoLuc luciferase glow-type luminescence[1]. In complexing the RNPs, we found the ideal complexation ratio for purified fusion protein and gRNA to be 1:1.2. Subsequently, we delivered 560, 280, and 130 fmol total of the RNPs with dCas9-SmBiT and LgBiT-dCas9 fusion proteins in 10:1 and 4:1 molar transfection ratios to HEK 293T cells along with 40 fmol target DNA plasmids with tandem target sites 40 bp apart and inverted target sites 7 bp apart using Lipofectamine CRISPRMAX (**Figure 2.12d**). The range of signal-to-background ratios obtained using this approach was approximately 13-fold to 18-fold, a substantial improvement over earlier plasmid-based delivery experiments.

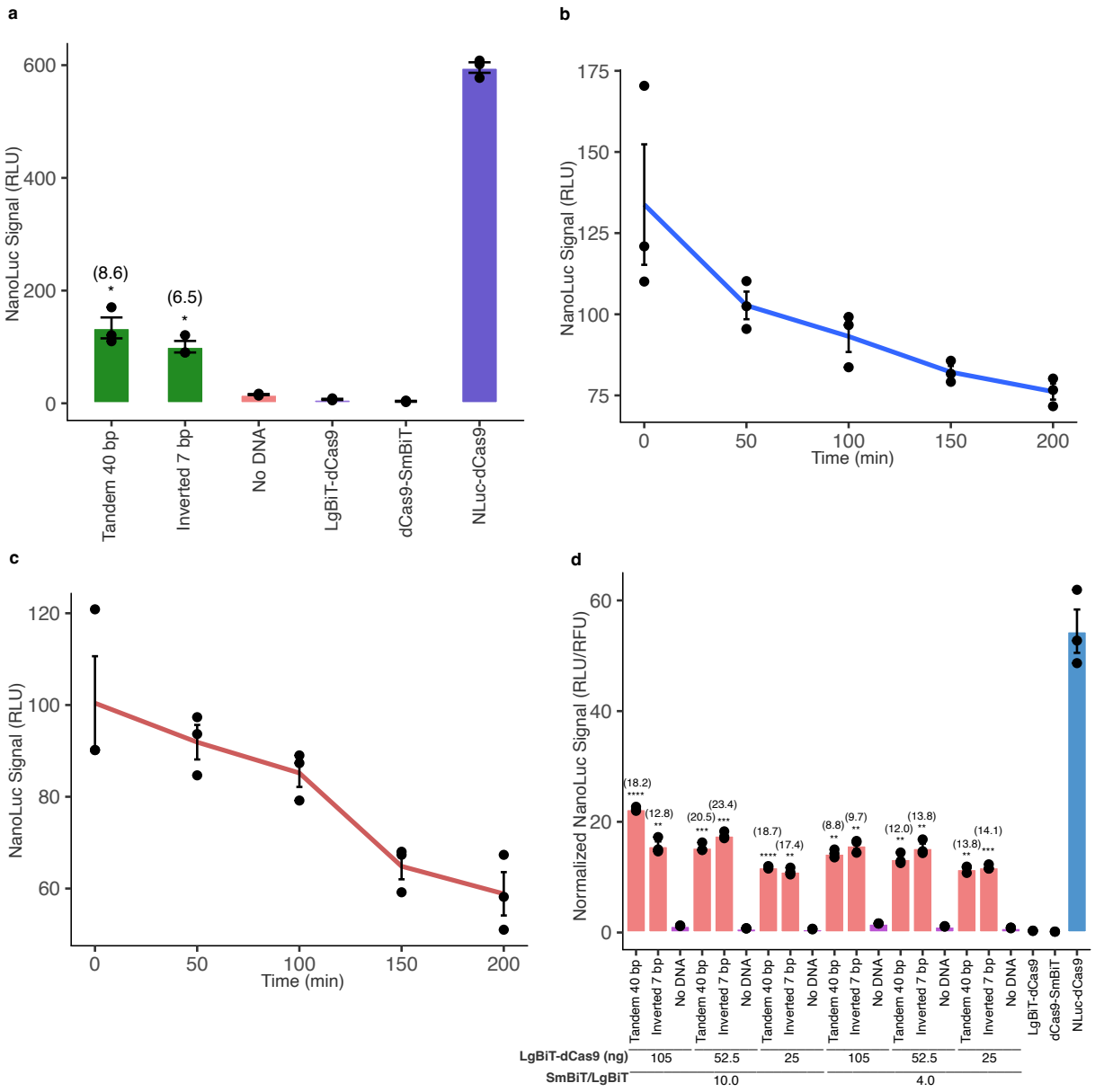


Figure 2.12: Initial RNP DNA biosensor activity assays and optimization of RNP-based delivery

(a) Relative NanoLuc luminescent signals 1 min after complexation of LgBiT-dCas9 and dCas9-SmBiT RNPs. **(b)-(c)** Time course experiments showing luminescent signal decay when LgBiT-dCas9 and dCas9-SmBiT RNPs bind tandem 40-bp (blue line) and inverted 7-bp (red line) target DNA plasmids *in vitro*. **(d)** An initial experiment showing RNP delivery of biosensor

components to live HEK 293T cells. In each condition, dCas9-SmBiT was complexed with IVT gRNA for the upstream target site and LgBiT-dCas9 was complexed with IVT gRNA for the downstream target site. Apparent signal-to-background ratios in **a** and **d** (comparisons made to no DNA background conditions) are listed in parentheses above each biosensing condition. Data in **a** and **d** are presented as the mean ± s.e.m., $n = 3$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

We then tested the RNP-based delivery method using 560 fmol total RNPs on 12 additional DNA target sequence scaffolds (40 fmol each), and the range of signal-to-background ratios was approximately 7.5-fold to 20-fold, underscoring the efficiency of this delivery approach **(Figure 2.13a)**. As we were delivering many copies of the target sequences in transfections, we sought to test the limit of detection for RNP-based delivery of the biosensor. We found that there was a sharper negative response in signal-to-background when target DNA concentration was decreased in RNP transfection compared to plasmid transfection of biosensor components **(Figure 2.13b)**. At the minimum amount of target site DNA transfected of 0.2 fmol, signal-to-background was approximately 6-fold for LgBiT-dCas9 and dCas9-SmBiT RNPs binding the tandem 40-bp target DNA plasmid and 3-fold for LgBiT-dCas9 and dCas9-SmBiT RNPs binding the inverted 7-bp target DNA plasmid. We then tested the same RNP biosensor delivery conditions across five other cell lines, with similar signal-to-background ranges but much lower absolute signals compared to HEK 293T **(Figure 2.13c)**.

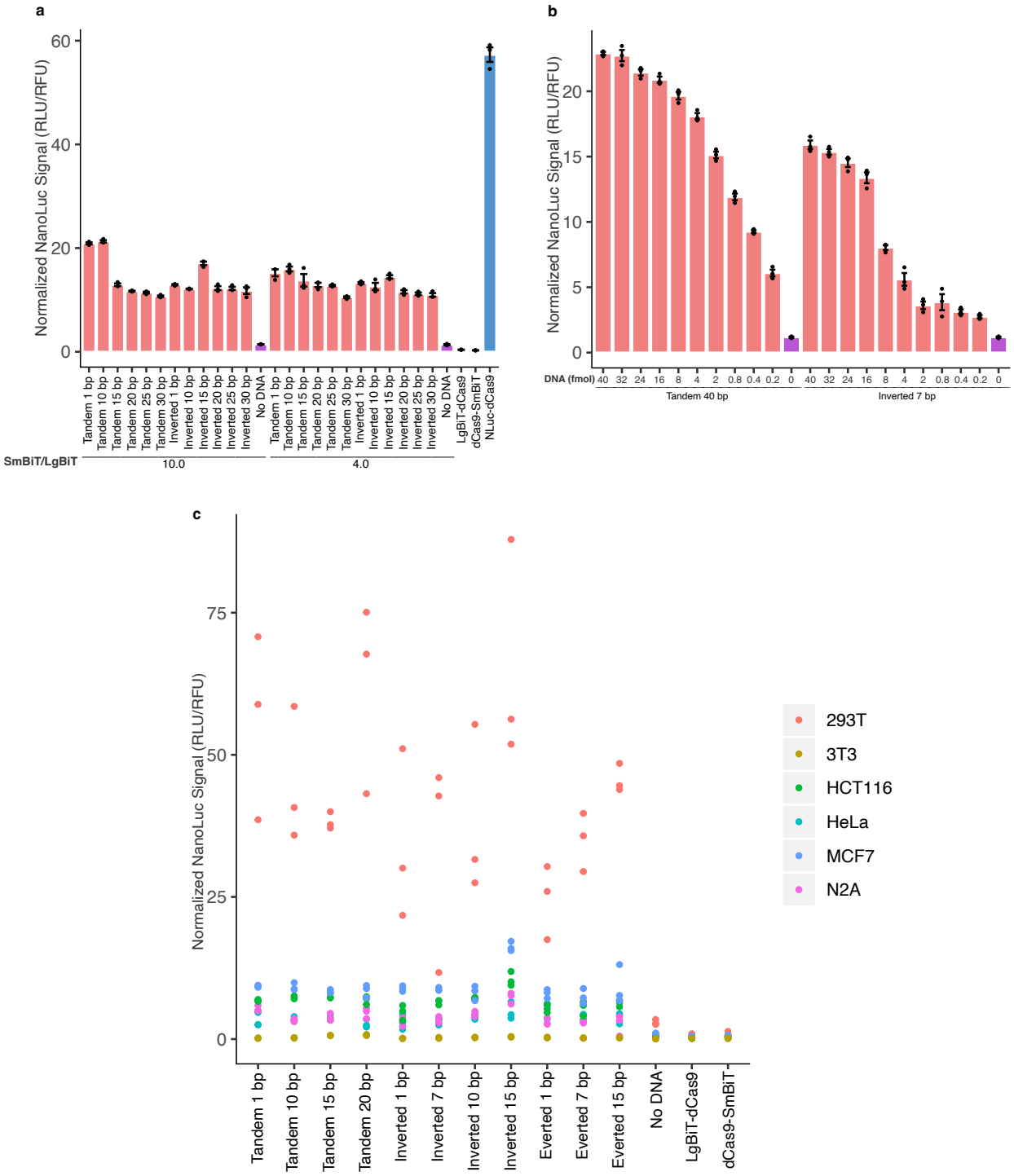


Figure 2.13: RNP-based DNA biosensor delivery in living cells

(a) 12 target sequence scaffolds tested in live cells using the RNP delivery method. In each condition, dCas9-SmBiT was complexed with IVT gRNA for the upstream target site and

LgBiT-dCas9 was complexed with IVT gRNA for the downstream target site and delivered to HEK 293T cells. **(b)** Effect of decreasing target sequence scaffold concentration on NanoLuc signal intensity using RNP-based delivery of biosensor components to live cells. **(c)** A comparison of RNP-based DNA biosensor function across six different cell lines. Apparent signal-to-background ratios in **a-b** (comparisons made to no DNA background conditions) are listed in parentheses above each biosensing condition. Data in **a-b** are presented as the mean ± s.e.m., $n = 3$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

Live single cell biosensor imaging using a standard light microscope and cell cluster imaging using IVIS system

Next, we sought to use orthogonal approaches to measure signal-to-background ratios for our RNP-based DNA biosensor in live cells. In addition to our approach using a luminometer to measure luminescence across whole well cell populations, we envisioned a platform for measurement of luminescence from our biosensor in single cells on relatively common imaging equipment. To this end, we modified an upright fluorescence microscope for imaging the relatively low light intensities associated with NanoLuc and other luminescent reporters. For example, cells were placed in a dark box with all light sources covered or off, and exposure times were lengthened (see Methods). 560 fmol total purified dCas9-SmBiT and LgBiT-dCas9 biosensor proteins were co-transfected in HEK 293T cells along with 40 fmol DNA target plasmids containing either tandem 40 bp or inverted 7 bp target sites and 0.2 fmol pMAX-GFP

plasmid as a normalization control (**Figure 2.14a-b,** respectively). Intensity of signals from these images were compared to those from an auto-association background control without target DNA **(Figure 2.14c)**, a LgBiT-dCas9 fusion construct expressed alone (**Figure 2.14d**) and a full-length NLuc-dCas9 positive control construct (**Figure 2.14e**). As an alternative approach, we also measured the same set of NanoLuc luminescent signals on the PerkinElmer IVIS Spectrum Bioluminescence Imaging System (**Figure 2.14f-j**). GFP images were obtained for all conditions for normalization (**Figure S2.2**). Although the IVIS system may not be as accessible to end-users as a light microscope, it has the advantage of imaging many cells in a culture dish simultaneously, allowing many image-based biosensing experiments to performed with minimal time and effort. For these images, we drew and integrated circular regions of interest (ROIs) around regions containing clusters of cell nuclei within the LivingImage software associated with the IVIS Spectrum, obtaining a comparable range of signal-to-background **(Figure 2.14k)**.

Live cell cluster IVIS imaging of repetitive and unique endogenous genomic sequences

To determine the applicability of our split NanoLuc luciferase DNA biosensor to imaging endogenous copy number DNA sequences, we used a single optimized gRNA, sgMUC4-E3(F+E) from a recent study[28] to direct the biosensor to bind a region of polymorphic 48-bp repeats of copy number between approximately 100 and 400 within exon 2 of the human *MUC4* locus **(Figure 2.15a)**. We then used sgMUC4-E3(F+E) as an anchor gRNA to direct our biosensor to bind the same repetitive region of *MUC4* and constructed four gRNAs with unique spacer lengths and orientations around it. We observed differences in biosensor sensitivity that varied based on cell line and target site configuration **(Figure 2.15b,d)**. For example, signal-to-background peaked at approximately 7.5-fold in HeLa cells **(Figure 2.15f)** and approximately 2-

fold in HEK 293T cells **(Figure 2.15h)**. It should be noted that these peak signal-to-background ratios were obtained with different gRNA pairings in each cell line. However, since the majority of loci within the human genome are non-repetitive, a utility of more profound value would be the detection of such low copy number sequences. To this end, we targeted the non-repetitive region of intron 1 of the human *MUC4* locus with 1-7 pairs of unique gRNAs tiling along the locus with at least 200 bp between pairs to avoid interactions between biosensor components at different binding sites (**Figure 2.15a**). Using this approach, we again observed cell type-specific and target site orientation-specific differences in biosensor sensitivity but also what appeared to be dosage effects relating to number of gRNA pairs transfected **(Figure 2.15c,e)**. Specifically, signal-to-background in HeLa cells peaked at approximately 27-fold using a single pair of gRNAs at a single locus **(Figure 2.15g)** but at approximately 13-fold in 293T cells using two pairs of gRNAs at two loci **(Figure 2.15i)**. Since the differences between signal-to-background in the two different cell lines could be related to dosage of gRNA pairs or intrinsic chromatin structure, we conducted an experiment where each of the seven loci was bound independently and pairwise comparisons in signal-to-background were made **(Figure 2.16)**. The maximum signal-to-background ratios were 52.6-fold at locus 1, 2.47-fold at locus 4, 4.33-fold at locus 4, and 8.36-fold at locus 1 for LgBiT-dCas9 + dCas9-SmBiT, dCas9-LgBiT + SmBiT-dCas9, LgBiT-dCas9 + SmBiT-dCas9, and dCas9-SmBiT + dCas9-LgBiT pairings of fusion proteins, respectively. Locus 1 within the *MUC4* non-repetitive region has a tandem 10-bp target site DNA configuration while locus 4 has a tandem overlapping target site DNA configuration with PAM sites 4 bp apart. This confirms previous results demonstrating signal-to-background dependence on both fusion protein orientation and target site configuration.

Live cell cluster IVIS imaging of single-base changes induced by CRISPR-Cas9 editing

Our main goal in conceiving a split NanoLuc luciferase DNA biosensor was to apply it to detection of various mutations in genomic DNA sequence after targeted genome editing with CRISPR-Cas9. Thus, we created G->T missense single nucleotide polymorphisms (SNPs) at two different loci in two cell lines: within the 8q24 multi-cancer risk locus in HCT116 cells and within the *PALB2* locus in 293 cells **(Figure 2.17a)**. Both SNPs were present within the PAM site of the gRNA used for editing[29]. We confirmed mutant lines were homozygous for the G->T missense mutations by isolating single edited cells by dilution plating then expanding populations and detecting specific alleles by Kompetitive Allele-Specific PCR (KASP). We hypothesized that these mutations could make binding by the gRNA used for editing less efficient or even completely inhibit binding by this gRNA. Thus, we expected signal-to-background within the mutant lines to be lower than signal-to-background within wild-type lines. We observed a result consistent with this hypothesis when we measured signals of wild-type and homozygous mutant 293 cells receiving LgBiT-dCas9 and dCas9-SmBiT RNPs, the gRNA used for editing, and several gRNAs of various orientations and spacer sequences around the gRNA used for editing on the IVIS spectrum. The absolute signals were higher in the mutant cell lines including the no gRNA background signal, resulting in lower signal-to-background for every gRNA pair in the mutant lines **(Figure 2.17b)**. Specifically, the signal-to-background ratios for biosensing conditions with gRNAs 1-5 around the gRNA used for editing were 2.11-fold, 2.03-fold, 1.78-fold, 2.64-fold, and 2.85-fold in wild-type lines compared to 0.79-fold, 1.19-fold, 0.86-fold, 1.30-fold, and 1.36-fold in homozygous mutant lines **(Figure 2.17c)**. In HCT116 cells, the signal-to-background ratios for biosensing conditions with gRNAs 1-4 around the gRNA

used for editing were 3.46-fold, 2.4-fold, 1.64-fold, and 2.2-fold in wild-type cells compared to 1.89-fold, 2.4-fold, 1.51-fold, and 2.62-fold in homozygous mutant lines **(Figure 2.17d)**.

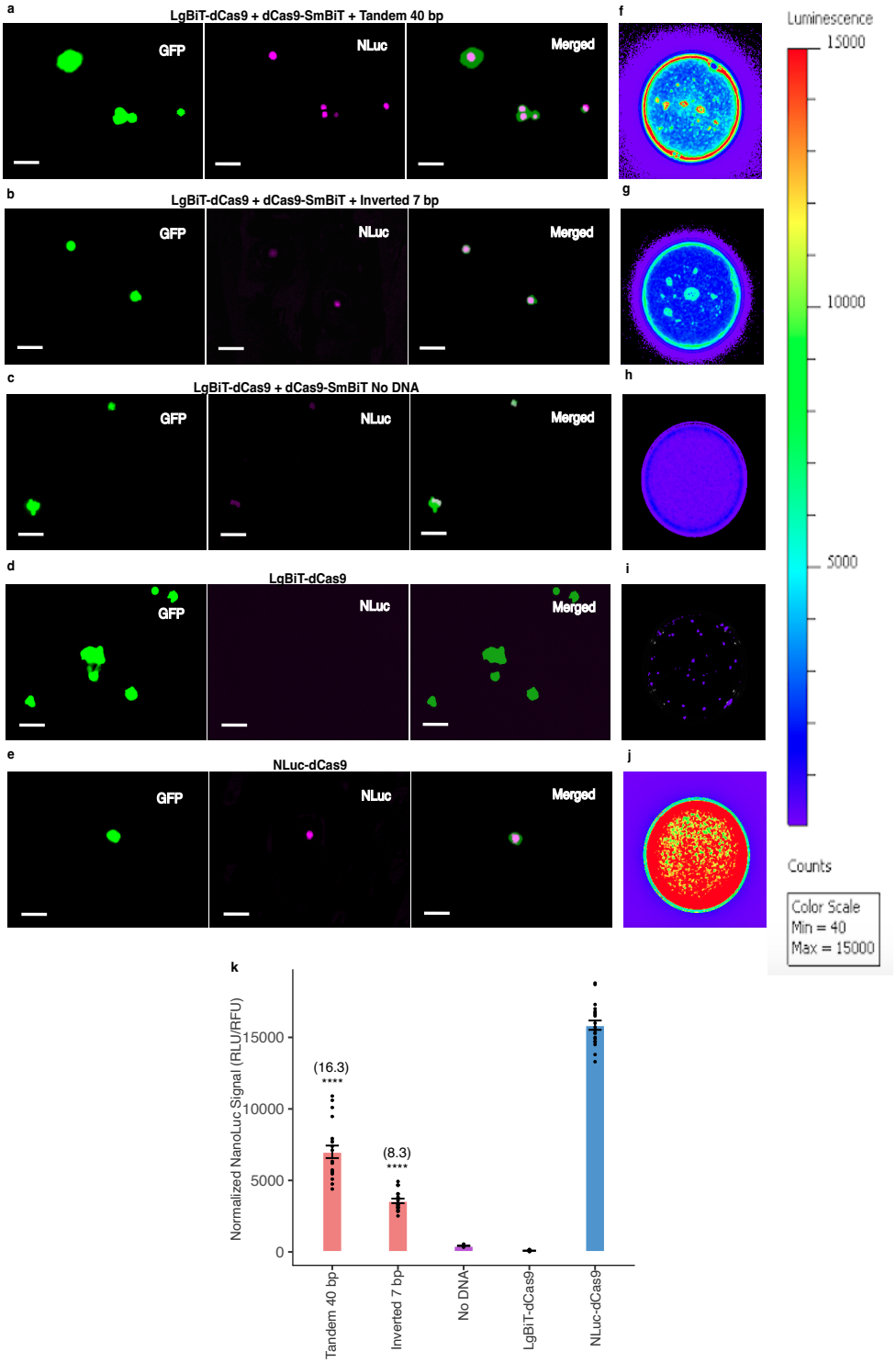


Figure 2.14: Orthogonal measurements of RNP DNA biosensor signal-to-background in live HEK 293T cells

(a)-(e) GFP, NanoLuc, and merged images taken on the Leica DM6000 B upright microscope at 10X magnification. GFP images were taken with 150 ms exposure to excitation light. NanoLuc images were taken with 30 s exposure and gain of 2.0 in a dark box. RNP constructs and DNA target site scaffolds delivered are shown above images. Scale bars = 50 μM. **(f)** Image taken on the IVIS Spectrum Bioluminescence Imaging System of live HEK 293T cell clusters transfected with the same RNPs and target DNA plasmid as shown in **a**. Signal scale is shown at right. **(g)** Image taken on the IVIS Spectrum Bioluminescence Imaging System of live HEK 293T cell clusters transfected with the same RNPs and target DNA plasmid as shown in **b**. Signal scaling shown at right. **(h)** Image taken on the IVIS Spectrum Bioluminescence Imaging System of live HEK 293T cell clusters transfected with the same RNPs but without target DNA as shown in **c**. Signal scaling shown at right. **(i)** Image taken on the IVIS Spectrum Bioluminescence Imaging System of live HEK 293T cells clusters transfected with the LgBiT-dCas9 RNP alone as shown in **d**. Signal scaling shown at right. **(j)** Image taken on the IVIS Spectrum Bioluminescence Imaging System of live HEK 293T cell clusters transfected with NLuc-dCas9 RNP alone as shown in **e**. Signal scaling shown at right. **(k)** Quantification of cell cluster region ROIs for various transfection conditions in IVIS Spectrum LivingImage software. Apparent signal-to-background ratios (comparisons made to no DNA background condition) are listed in parentheses above each biosensing condition. Data in **k** is presented as the mean ± s.e.m., $n = 20$, where n represents the number of distinct cell nuclei quantified; unpaired two-sided Student′s t-test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

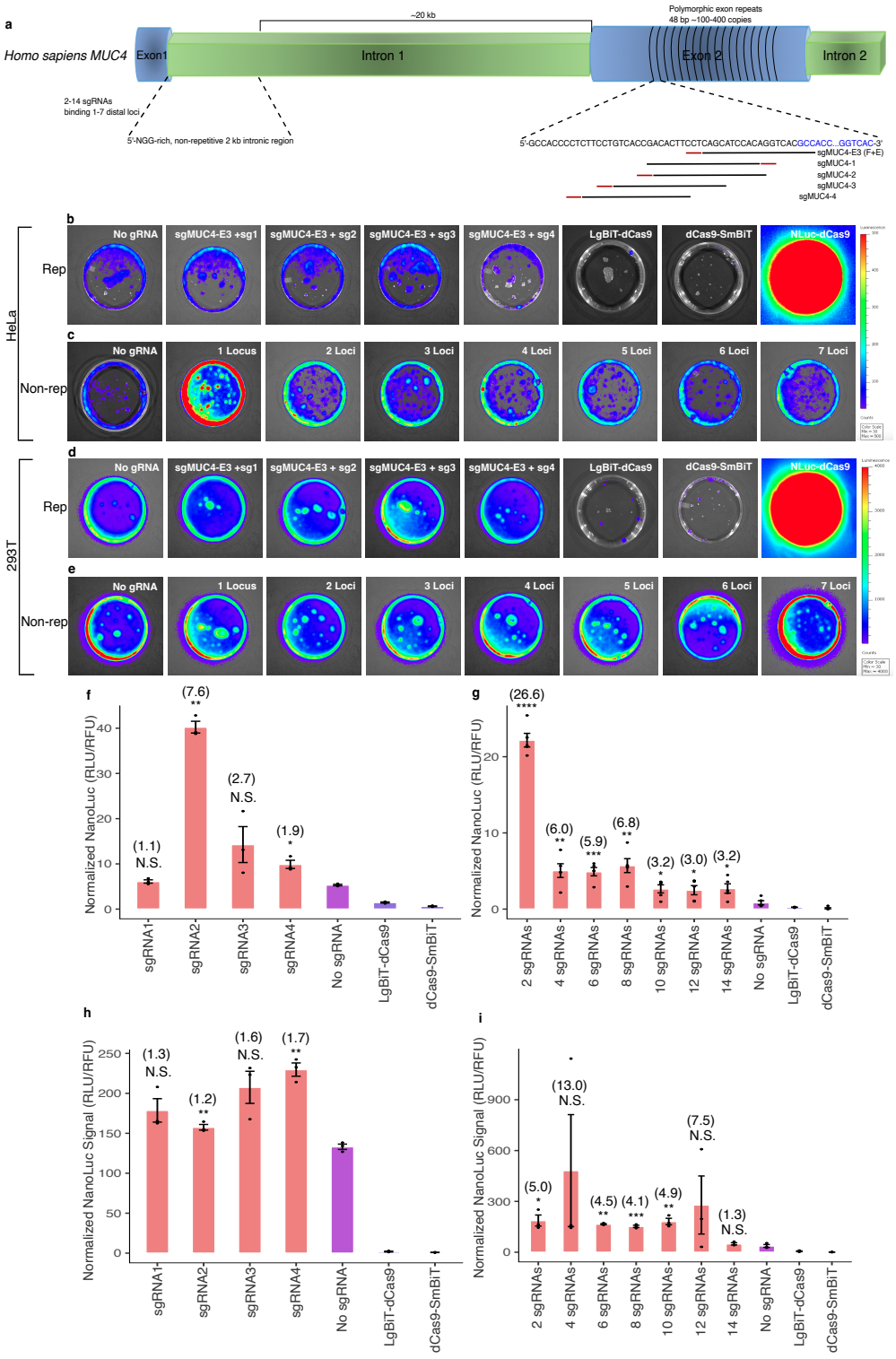

a
Homo sapiens MUC4
Exon1
Intron 1
Exon 2
Intron 2
~20 kb
Polymorphic exon repeats
48 bp ~100-400 copies
2-14 sgRNAs
binding 1-7 distal loci
5'-NGG-rich, non-repetitive 2 kb intronic region
5'-GCCACCCCTCTTCCTGTCACCGACACTTCCTCAGCATCCACAGGTCACGCCACC...GGTCAC-3'
sgMUC4-E3 (F+E)
sgMUC4-1
sgMUC4-2
sgMUC4-3
sgMUC4-4
b
c
d
e
HeLa
293T
Rep
Non-rep
No gRNA
sgMUC4-E3 +sg1
sgMUC4-E3 + sg2
sgMUC4-E3 + sg3
sgMUC4-E3 + sg4
LgBiT-dCas9
dCas9-SmBiT
NLuc-dCas9
1 Locus
2 Loci
3 Loci
4 Loci
5 Loci
6 Loci
7 Loci
f
g
h
i
Normalized NanoLuc (RLU/RFU)
Normalized NanoLuc Signal (RLU/RFU)
sgRNA1
sgRNA2
sgRNA3
sgRNA4
No sgRNA
LgBiT-dCas9
dCas9-SmBiT
2 sgRNAs
4 sgRNAs
6 sgRNAs
8 sgRNAs
10 sgRNAs
12 sgRNAs
14 sgRNAs
(1.1) N.S.
(7.6) **
(2.7) N.S.
(1.9) *
(26.6) ****
(6.0) **
(5.9) ***
(6.8) **
(3.2) *
(3.0) *
(3.2) *
(1.3) N.S.
(1.2) **
(1.6) N.S.
(1.7) **
(5.0) *
(13.0) N.S.
(4.5) **
(4.1) ***
(4.9) **
(7.5) N.S.
(1.3) N.S.

Figure 2.15: Biosensing repetitive and non-repetitive genomic sequences at the human *MUC4* locus

(a) Cartoon visualization of the repetitive and non-repetitive regions of the human *MUC4* locus. **(b)** dCas9-NanoBiT RNP biosensing of the repetitive region of *MUC4* exon 2 in live HeLa cells. **(c)** dCas9-NanoBiT RNP biosensing of the non-repetitive region of *MUC4* intron 1 in live HeLa cells. **(d)** dCas9-NanoBiT RNP biosensing of the repetitive region of *MUC4* exon 2 in live HEK 293T cells. **(e)** dCas9-NanoBiT RNP biosensing of the non-repetitive region of *MUC4* intron 1 in live HEK 293T cells. **(f)** Signal quantification of the DNA biosensor binding the repetitive region of *MUC4* exon 2 in live HeLa cells. **(g)** Signal quantification of the DNA biosensor binding the non-repetitive region of *MUC4* intron 1 in live HeLa cells. Error bars represent s.e.m., $n = 5$. **(h)** Signal quantification of the DNA biosensor binding the repetitive region of *MUC4* exon 2 in live HEK 293T cells. **(i)** Signal quantification of the DNA biosensor binding the non-repetitive region of *MUC4* intron 1 in live HEK 293T cells. Apparent signal-to-background ratios in **f-i** (comparisons made to no sgRNA background conditions) are listed in parentheses above each biosensing condition. Data in **f, h, and i** are presented as the mean ± s.e.m., $n = 3$, where n represents the number of independent experimental technical and biological replicates included in parallel; unpaired two-sided Student′s t-test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

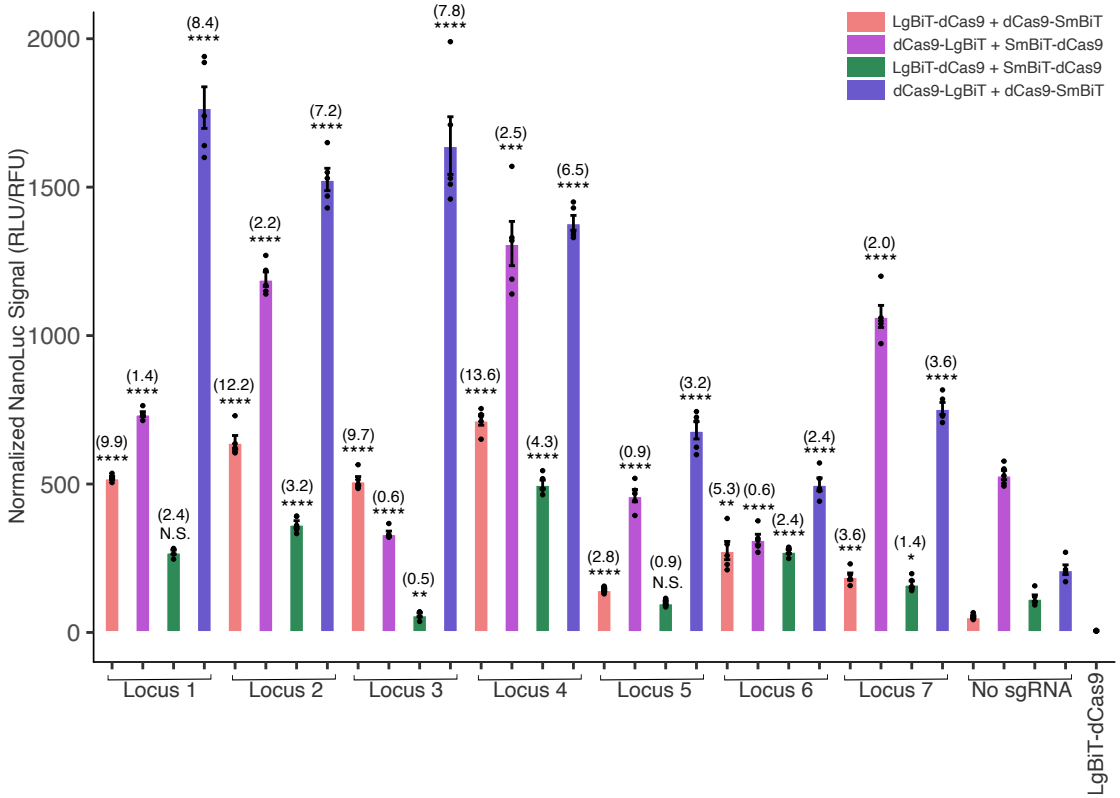


Figure 2.16: RNP DNA biosensor luminescence output variability across seven individual non-repetitive loci at *MUC4*

Normalized luminescence intensities from a DNA biosensing experiment where four orientations of dCas9-NanoBiT RNPs were directed to bind seven individual locations within the non-repetitive region of the human *MUC4* gene. Apparent signal-to-background ratios (comparisons made to no sgRNA background conditions separately for each fusion protein orientation) are listed in parentheses above each biosensing condition. Data is presented as the mean ± s.e.m., $n = 5$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

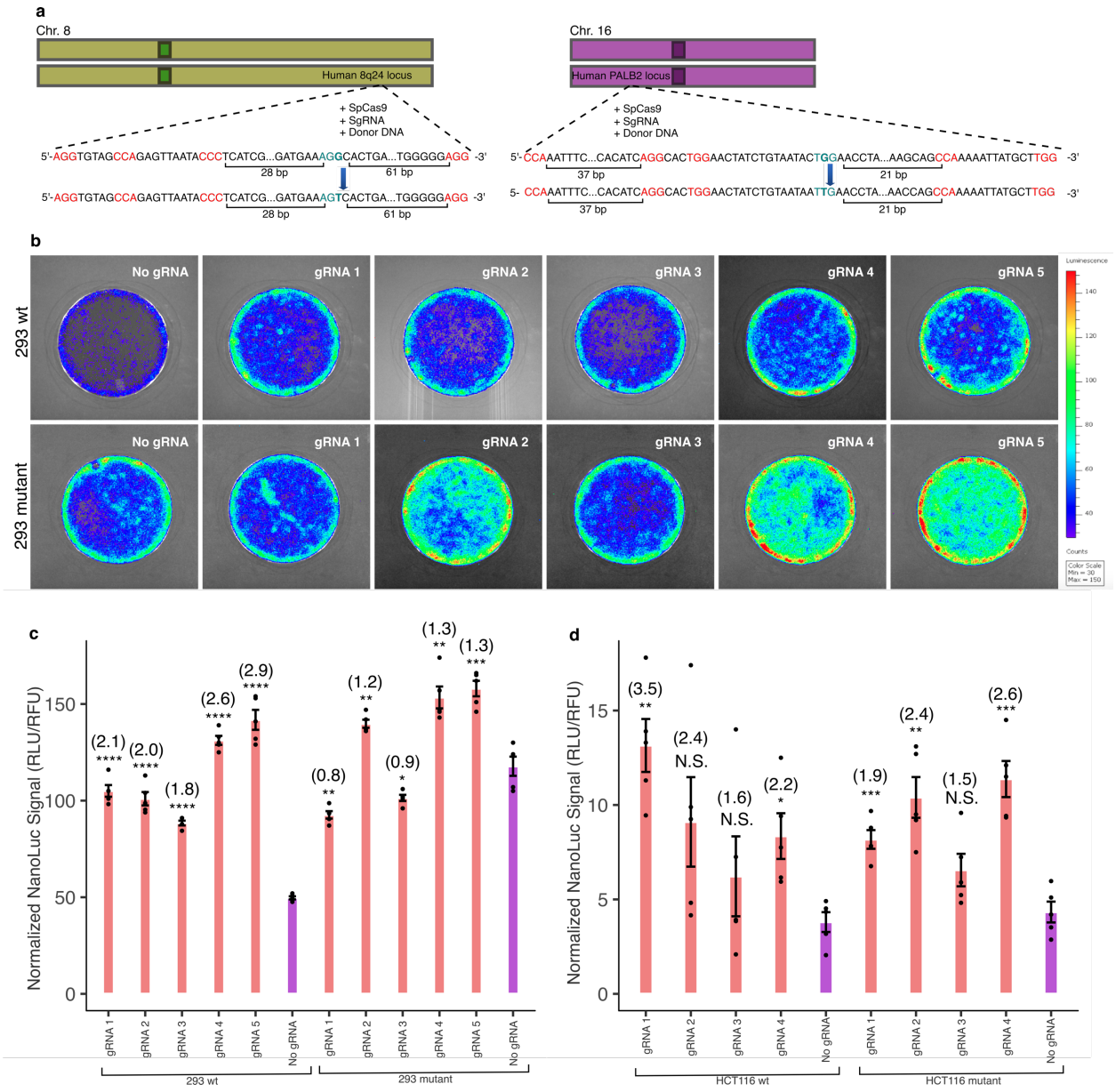


Figure 2.17: Biosensing CRISPR-Cas-induced genome edits in live cells

(a) Cartoon visualization of the editing experiments conducted at the human 8q24 cancer risk and *PALB2* loci. gRNAs used for editing are shown in blue and gRNAs around the site of mutation that were used for detection of mutant cells in biosensing experiments are shown in red. Single base pair edits are shown in bold. **(b)** Images taken on the IVIS Spectrum Bioluminescence Imaging System of the bipartite DNA biosensor applied to

the *PALB2* locus after targeted CRISPR-Cas9 genome editing. Wild type HEK 293 cells transfected with the LgBiT-dCas9 and dCas9-SmBiT RNPs and several gRNAs are compared to HEK 293 cells homozygous for a G->T missense mutation at the *PALB2* locus transfected with the same biosensor RNPs and gRNAs. Both wild type and mutant biosensing conditions are compared to a background condition where the biosensor RNPs are not directed to bind the DNA by gRNAs. **(c)** Signal differences in directed probe binding conditions compared to background conditions for both G->T mutant and wild type HEK 293 cells. **(d)** Application of the DNA biosensor with LgBiT-dCas9 and dCas9-SmBiT RNPs to the 8q24 risk locus after targeted CRISPR-Cas9 genome editing. Signal differences in directed probe binding conditions are compared to background conditions for both G->T homozygous mutant and wild type HCT116 cells. Apparent signal-to-background ratios in **c-d** (comparisons made to no sgRNA background conditions) are listed in parentheses above each biosensing condition. Data in **c-d** are presented as the mean ± s.e.m., $n = 5$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$.

DISCUSSION

In our initial sequence detection assay in cell lysates, we observed variable increases in +sgRNA normalized luminescent signals compared to –sgRNA normalized luminescent signals for all four orientations of LgBiT and SmBiT fusions with dCas9 ($p < 0.05$ using an unpaired two-sided Student′s t-test). We observed the largest increase in signal-over-background of 9.19-fold when

LgBiT-dCas9 was co-expressed with dCas9-SmBiT using a LgBiT:SmBiT molar transfection ratio of 1:4, which provided initial evidence in support of using this specific pairing of fusion protein orientations for live cell DNA biosensing and potentially using a 4-fold molar excess of dCas9-SmBiT. There appeared to be a bimodal distribution in the SBR data for the LgBiT-dCas9 + dCas9-SmBiT pairing where between a 1.33-10-fold excess of one NanoBiT component produced a higher SBR than a 1:1 ratio of constructs, whereas in other fusion construct pairings such as dCas9-LgBiT + SmBiT-dCas9 and LgBiT-dCas9 + SmBiT-dCas9, a 1:1 ratio of constructs produced a higher SBR than any variations in the LgBiT:SmBiT molar transfection ratio. Finally, in the dCas9-LgBiT + dCas9-SmBiT pairing, there was a clear peak in the SBR at 2-fold molar excess LgBiT in transfection. These results informed future assays where we sought to determine differences in the SBR due to variation of LgBiT:SmBiT ratio in live cell assays. In the assays where the LgBiT and SmBiT constructs were transfected at a 1:1 molar ratio and the amount transfected of each component was reduced from 11 fmol to 0.2 fmol, there was a notable negative relationship between the amount transfected and the SBR for the dCas9-LgBiT + dCas9-SmBiT pairing but a clear positive relationship between the amount transfected and the SBR for the three other fusion construct pairings. This indicated that, for the dCas9-LgBiT + dCas9-SmBiT pairing, transfecting lower molar amounts of biosensor components might result in lower background association of the NanoBiTs, at least to a higher degree than observed for the other three fusion construct pairings.

As expected, the ranges of signals over all four possible orientations of dCas9-NanoBiT and NanoBiT-dCas9 fusion constructs was 0.091-0.677, which clearly exceeded the upper range of signals for the SmBiT constructs expressed alone (0.000959-0.00147), the LgBiT constructs

expressed alone (0.00106-0.00123), and pUC19 expressed alone (0.0002-0.0004). We predicted a level of luminescence for the NanoBiT-dCas9 fusions expressed alone at or near instrument background primarily due to auto-luminescence of the NanoLuc substrate furimazine. Consistent with the idea that no NanoLuc luciferase reassembly should be occurring in these wells, these results were similar to results for additional control wells that were not transfected with any DNA but where furimazine was added. Furthermore, the normalized NanoLuc luminescence signal range in conditions where both dCas9-NanoBiT or NanoBiT-dCas9 constructs were co-expressed also fell quite short of the normalized NanoLuc signal range of 38.5-43.0 observed for the condition where an equal molar amount of full NanoLuc enzyme fusion to dCas9 was expressed. This showed that all moles of expressed NanoBiT-containing fusion proteins were not reassembling into functional NanoLuc enzymes in the cell lysates. In addition, we concluded from this initial data that our DNA biosensor produced an appropriate signal output range between positive and negative control conditions. Overall, this initial data demonstrated that our NanoBiT-dCas9 fusion biosensor showed promise toward the goal of detection of specific target DNA sequences in living cells. In future experiments, we hoped to avoid using the target sequence vectors with only two orientations of target sites A and B and four different spacer lengths between them used in this initial study, instead resolving to deconstruct these vectors into individual target site scaffolds and test each scaffold individually to determine the optimal target site orientation and spacing for all four combinations of directional LgBiT and SmBiT fusion constructs.

The rationale for delivering the biosensor components as RNPs was twofold. First, the delivery of the fusion proteins in plasmid form resulted in the production of all possible pairings of fusion

protein and gRNA. We quickly realized that half of these RNP pairings, when bound to target DNA, would not produce a detectable signal. For example, in an experiment delivering LgBiT-dCas9 and dCas9-SmBiT fusion proteins and two different gRNAs to cells, the gRNAs could both associate with LgBiT-dCas9 fusion proteins or both associate with dCas9-SmBiT fusion proteins. These two pairings would direct RNPs with identical NanoBiTs to bind adjacent to one another on the same target DNA. As a result, two LgBiT-dCas9 or two dCas9-SmBiT RNPs would transiently occupy a copy of the target DNA with no NanoLuc luciferase reassembly or signal output. While the actual number of these unproductive assemblies from initial live cell experiments is difficult to predict, these events are not difficult to imagine and were most likely occurring at high enough levels to substantially repress the SBR observed in these assays. Second, as protein expression from the biosensor component plasmids was driven by the constitutive CMV promoter, control of the total concentration of free-floating nuclear RNPs was not possible. Fusion proteins may have been constitutively expressed to a very high level, making auto-association of free-floating nuclear RNPs more favorable and resulting in a measurable increase in the background signal and a reduction in SBR. Third, delivery of system components in plasmid form posed a low risk of spontaneous plasmid integration into the genome. Thus, although plasmid-based delivery could be an appropriate method to deliver our DNA biosensor, we concluded it might be a less productive approach compared to RNP-based delivery. In our initial RNP-based DNA biosensing experiments, we saw a range of normalized signals for our biosensor of 0.049-0.239 RLU/RFU and average normalized signal of 0.116 RLU/RFU in the presence of target DNA compared to a range of normalized signals of 0.015-0.019 RLU/RFU and average normalized signal of 0.016 RLU/RFU in the absence of target DNA. This is a significant difference by unpaired Student's t-test ($p < 0.0001$, two-tailed). From

these results, it is clear that the biosensor detects the presence of DNA in live cells more efficiently when it is delivered in the form of preassembled RNPs. We then moved away from luminometer-based measurement of luminescent signals, using two orthogonal approaches: microscopy and bioluminescence imaging. After specifically modifying these methods for our application, we obtained similar signal-to-background measurements for our biosensor, which further confirmed the efficacy of the RNP-based delivery approach and demonstrated amenability to multiple routes of measurement and data analysis.

We also realized that introducing DNA target sites on plasmids diluted biosensor components in transfection, provided DNA targets that were only transiently available for binding in the nucleus, and resulted in target sequence copy numbers that were likely much higher than those observed for genomic loci. Thus, we designed new gRNAs to target endogenous DNA binding sites on genomic DNA in live cells instead of introducing DNA target plasmids in transfection. We theorized that this approach would allow us to investigate the critical question of whether our dCas9-NanoBiT biosensor could be sensitive enough to detect extremely low copy numbers. One consequence of removing DNA target site vectors from the transfection was that it necessitated a new definition of the auto-association background condition. We thus employed another auto-association condition where the biosensor was directed to bind genomic target sites not predicted to be proximal in the nucleus or target sites predicted not to exist within a given genome. In an analogous fashion to our preliminary assays using target DNA vectors, we first assessed whether signal output was in the expected range for our biosensor. Directing the biosensor to bind a repetitive region of the human *MUC4* locus in HeLa cells, normalized luminescent signals for all pairings of NanoBiT-dCas9 fusion proteins in biosensing conditions was in the range of 5.54-42.83 RLU/RFU, which again exceeded the upper range of normalized signals for dCas9-SmBiT

expressed alone (0.52-0.77 RLU/RFU) and for dCas9-LgBiT expressed alone (1.24-1.63 RLU/RFU) but was below the lower range of normalized signals for the NLuc-dCas9 fusion protein (1422.23-1951.68 RLU/RFU). Thus, we determined that our dimeric DNA biosensor produced expected signal output between positive and negative controls on endogenous copy number sequences. As before, we next compared our biosensing condition with gRNAs delivered in transfection to our background auto-association condition with no supplied gRNAs. We saw a range of normalized signals for the auto-association condition of 5.09-5.61 RLU/RFU in HeLa cells, again demonstrating that assembly of free floating NanoBiT fusion proteins occurred at a lower level compared to the endogenous DNA biosensing condition. Furthermore, the average normalized signal across all 12 biosensing condition wells in HeLa cells was ~17.63 RLU/RFU whereas the average normalized signal across 3 auto-association wells in HeLa cells was ~1.46 RLU/ RFU ($p < 0.05$ for this difference in means, unpaired, two-tailed Student's *t*-test). In addition, we observed large differences between biosensing conditions and background conditions at the repetitive region of *MUC4* in 293T cells ($p < 0.0001$ for this difference in means, unpaired, two-tailed Student's *t*-test). Taken together, these differences in signal intensities for targeted DNA binding conditions compared to undirected conditions using the dCas9-NanoBiT biosensor suggested NanoLuc reassembly was occurring in target cell nuclei upon targeted binding of the *MUC4* repetitive region. We then tested our biosensor on a non-repetitive portion of the human *MUC4* locus. Comparing our biosensing condition with gRNA to our undirected auto-association condition without gRNA in HeLa cells, normalized signal ranges were 0.96-21.31 RLU/RFU and 0.42-1.76 RLU/RFU, respectively. Average normalized signals were ~6.53 RLU/RFU and ~0.83 RLU/RFU for the same two conditions, respectively ($p < 0.0001$ for this difference in means, unpaired, two-tailed Student's t-test). Furthermore,

comparing biosensing conditions to background auto-association conditions in 293T cells, normalized signal ranges were 31.59-1142.48 RLU/RFU and 26.4-53.64 RLU/RFU, respectively. Average normalized signals were ~213.77 RLU/RFU and ~37.01 RLU/RFU for the same two conditions, respectively ($p < 0.01$ for this difference in means, unpaired, two-tailed Student's t-test). Thus, we concluded that the detection of endogenous level copy number sequences was reliable and consistent using an RNP-based biosensor and further probing of its use in this form should be warranted.

One pertinent application for our split NanoLuc luciferase DNA biosensor that we imagined would require high sensitivity was isolation of mutant cells from a population of cells after genome editing. To investigate the feasibility of this application, we conducted CRISPR-Cas9 editing experiments at two genomic loci in HCT116 and HEK 293 cells with the goal of using our DNA biosensor to detect the difference in copy number of a specific sequence between wild-type and homozygous mutant cells. Using difference in SBR as a primary endpoint, we found that SBR was higher across several sites bound by gRNA pairs around the original Cas9 cut site in wild-type HEK 293 cells compared to HEK 293 cells that were homozygous mutants for a single-base pair change in the PAM site of the editing gRNA target sequence. This effectively demonstrated differentiation between binding two and zero copies of the target sequence, as HEK 293 cells have two copies of chromosome 16 with no commonly reported abnormalities[30]. In HCT116 cells, only one gRNA pair with slightly overlapping protospacer sequences with PAM sites 28 bp apart showed reliable detection of the target sequence. We hypothesized that mutating the PAM site in both cell lines would create a condition where Cas9 would not be able to recognize the original target site[31]. The fact that all gRNA pairs showed higher SBR in wild-type compared to mutant HEK 293 cells, yet this seemingly gRNA-independent effect was not

observed in HCT116 cells may be due to intrinsic differences in chromatin structure between cell lines at the edited loci. If this is the case, then future experiments using this RNP-based biosensor should be designed to facilitate interactions with more ideal orientation and spacing of DNA target sites given biosensor fusion protein orientations.

SUPPLEMENTARY INFORMATION

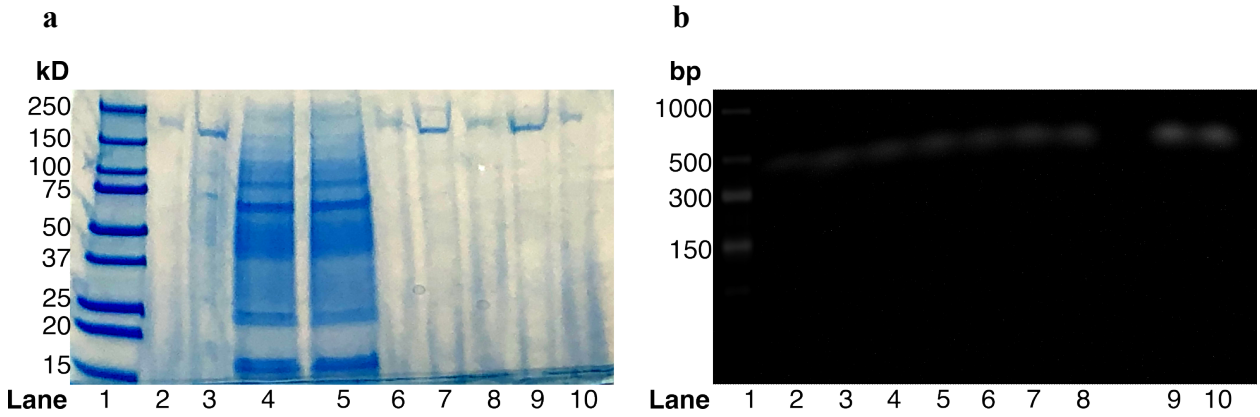


Figure S2.1: Ribonucleoprotein production via affinity purification and *in vitro* transcription

(a) Coomassie blue-stained SDS-PAGE gel from affinity purification of dCas9-NanoBiT, NanoBiT-dCas9 and NanoLuc-dCas9 fusion proteins. From left to right: Lane 1—Precision Plus Protein Standards; Lane 2—LgBiT-dCas9 batch 2; Lane 3—dCas9-SmBiT batch 2; Lane 4—LgBiT-dCas9 batch 1 total protein lysate; Lane 5—dCas9-SmBiT batch 1 total protein lysate; Lane 6—LgBiT-dCas9 batch 2; Lane 7—dCas9-SmBiT batch 2; Lane 8—dCas9-LgBiT batch 1; Lane 9—SmBiT-dCas9 batch 1; Lane 10—NLuc-dCas9 batch 1. **(b)** Denaturing formaldehyde (1.1% v/v) 1.5% (w/v) agarose gel of IVT gRNAs for *in vitro* and live cell DNA biosensing experiments using RNPs. From left to right: Lane 1—Low range ssRNA ladder; Lane 2—IVT gRNA repetitive *MUC4* 1; Lane 3—IVT gRNA repetitive *MUC4* 2; Lane 4—IVT gRNA non-repetitive *MUC4* 1; Lane 5—IVT gRNA non-repetitive *MUC4* 2; Lane 6—IVT gRNA 293 PALB2 1; Lane 7—IVT gRNA 293 PALB2 2; Lane 8—IVT gRNA HCT116 8q24 1; Lane 9—IVT gRNA JL 1; Lane 10—IVT gRNA JL 2.

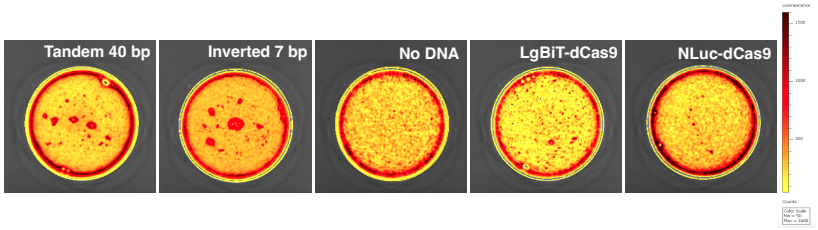


Figure S2.2: IVIS GFP Images

IVIS GFP images used for normalization of images shown in Figure 2.15f-j.

References

1. Hall, M. P. *et al*. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol. **7***, 1848–57 (2012).
2. England, C. G., Ehlerding, E. B. & Cai, W. NanoLuc: a small luciferase is brightening up the field of bioluminescence. *Bioconjug. Chem. **27***, 1175–87 (2016).
3. Dixon, A. S. *et al*. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol. **11***, 400–408 (2016).
4. Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature **513***(7519), 569–573 (2014).
5. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature **507***(7490), 62–67 (2014).
6. Palermo, G. *et al*. Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q. Rev. Biophys. **51***, e91 (2018).
7. van Rosmalen, M., Krom, M., & Merkx, M. Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins. *Biochemistry **56***(50), 6565–6574 (2017).
8. Händel, E. M., Alwin, S., & Cathomen, T. Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther*. ***17***(1), 104–111 (2009).
9. Chen, X., Zaro, J. L., & Shen, W. C. Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev. **65***(10), 1357–1369 (2013).
10. Reddy Chichili, V. P., Kumar, V., & Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Sci*. ***22***(2), 153–167 (2013).
11. Wilman, H. R., Shi, J., & Deane, C. M. (2014). Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins **82***(9), 1960–1970 (2014).
12. von Heijne, G. Proline kinks in transmembrane alpha-helices. *J Mol. Biol. **218***(3), 499–503 (1991).
13. Morgan, A. A., & Rubenstein, E. Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PloS one **8***(1), e53785 (2013).
14. Krieger, F., Möglich, A., & Kiefhaber, T. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc. **127***(10), 3346–3352 (2005).
15. Batterman, R. W. (1998). Why equilibrium statistical mechanics works: Universality and the renormalization group. *Philos. Sci. **65***(2): 183–208 (1998).
16. Darrigol, O. (2018). *Atoms, Mechanics, and Probability: Ludwig Boltzmann's Statistico-Mechanical Writings—an Exegesis*, Oxford: Oxford University Press.
17. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol*. ***34***(3), 339–344 (2016).
18. Dong et al. Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature **546***(7658), 436–439 (2017).

19. Farasat, I. & Salis, H. M. A biophysical model of CRISPR/Cas9 activity for rational design of genome editing and gene regulation. *PLoS Comput. Biol.* ***12***(1), e1004724 (2016).
20. Hao, N., Shearwin, K. E., & Dodd, I. B. Programmable DNA looping using engineered bivalent dCas9 complexes. *Nat. Commun.* ***8***(1), 1628 (2017).
21. Arganda-Carreras, I. *et al*. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* ***33***(15), 2424–2426 (2017).
22. Stains, C. I., Porter, J. R., Ooi, A.T., Segal, D. J. & Ghosh, I. DNA sequence-enabled reassembly of the green fluorescent protein. *J. Am. Chem. Soc.* ***127***, 10782–83 (2005).
23. Ooi, A. T., Stains, C. I., Ghosh, I. & Segal, D. J. Sequence-enabled reassembly of betalactamase (SEER-LAC): a sensitive method for the detection of double-stranded DNA. *Biochemistry* ***45***, 3620–25 (2006).
24. Ghosh, I., Stains, C. I., Ooi, A.T. & Segal, D. J. Direct detection of double-stranded DNA: molecular methods and applications for DNA diagnostics." *Mol. Biosyst.* ***2***, 551–60 (2006).
25. Stains, C. I., Furman, J. L., Segal, D. J., & Ghosh, I. Site-specific detection of DNA methylation utilizing mCpG-SEER. *J. Am. Chem. Soc.* ***128***(30), 9761–9765 (2006).
26. Porter, J. R., Stains, C. I., Segal, D. J., & Ghosh, I. Split beta-lactamase sensor for the sequence-specific detection of DNA methylation. *Anal. Chem.* ***79***(17), 6702–6708 (2007).
27. Zhang, Y. *et al*. Paired design of dCas9 as a systematic platform for the detection of featured nucleic acid sequences in pathogenic strains. *ACS Synth. Biol.* ***6***(2), 211–216 (2017).
28. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* ***155***, 1479–91 (2013).
29. Coggins, N. B., Stultz, J., O'Geen, H., Carvajal-Carmona, L. G. & Segal, D. J. Methods for scarless, selection-free generation of human cells and allele-specific functional analysis of disease-associated SNPs and variants of uncertain significance. *Sci. Rep.* ***7***, 15044 (2017).
30. Lin, Y., *et al*. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* ***5***, 4767 (2014).
31. Jiang, F. & Doudna, J.A. CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–29 (2017).

Chapter 3: Imaging Unique DNA Sequences in Individual Cells Using A CRISPR-Cas9-Based, Split Luciferase Biosensor

Chapter 3 is a reprint of the published manuscript:

"Imaging Unique DNA Sequences in Individual Cells Using a CRISPR-Cas9-Based, Split Luciferase Biosensor." Heath Nicholas G., O'Geen Henriette, Halmai Nicole B., Corn Jacob E., Segal David J. Front Genome Ed. 2022; 4: 867390. DOI: 10.3389/fgeed.2022.867390.

NGH conducted experiments and wrote the manuscript. DS edited the manuscript and co-drafted the primary funding grant. JC co-drafted the primary funding grant. HO'G assisted in conception and design of experiments. NBH created CRISPR-edited cell lines which were provided for use in experiments.

ABSTRACT

An extensive arsenal of biosensing tools has been developed based on the clustered regularly interspaced short palindromic repeat (CRISPR) platform, including those that detect specific DNA sequences both *in vitro* and in live cells. To date, DNA imaging approaches have traditionally used full fluorescent reporter-based fusion probes. Such "always-on" probes differentiate poorly between bound and unbound probe and are unable to sensitively detect unique copies of a target sequence in individual cells. Herein we describe a DNA biosensor that provides a sensitive readout for such low-copy DNA sequences through proximity-mediated reassembly of two independently optimized fragments of NanoLuc luciferase (NLuc), a small, bright luminescent reporter. Applying this "turn-on" probe in live cells, we demonstrate an application not easily achieved by fluorescent reporter-based probes, detection of individual endogenous genomic loci using standard epifluorescence microscopy. This approach could enable detection of gene edits during *ex vivo* editing procedures and should be a useful platform for many other live cell DNA biosensing applications.

INTRODUCTION

A critical bottleneck in the gene editing process is the ability to identify and isolate individual cells with desired edits within a population of treated cells. Current approaches typically require time-consuming and labor-intensive single cell isolation followed by isogenic population expansion[1-3] and downstream *in vitro* analysis of DNA sequence[4-7] in a portion of the expanded population. However, cell types that exhibit low efficiencies of transfection, editing, single cell isolation, or population expansion can make these procedures particularly challenging[8-12]. Furthermore, homology directed repair (HDR) can exhibit extremely low efficiency in certain cell types[13].

A promising alternative strategy to validate gene edits could be the direct biosensing of user-defined sequences at single copy with single cell resolution. In recent years, the CRISPR/Cas gene editing system has been modified for imaging endogenous genomic loci, but the vast majority of current approaches utilize full-length fluorescent reporter-based probes, such as dCas9-GFP[14-21]. However, each full reporter sensor molecule produces a signal whether bound to its target DNA or not, resulting in a high fluorescent background that negatively impacts the signal-to-background ratio (SBR). For this reason, such "always-on" sensors must rely on obtaining a high local concentration of probes to distinguish signal from background, limiting their use to highly repetitive elements that can be targeted by one sgRNA or to unique sequences targeted by 20-30 or more sgRNAs[14,16]. Imaging a short sequence present at a single copy is so far not possible.

However, “turn-on” DNA biosensors offer the possibility that signal could be produced only after binding of one or more subunits to a target sequence. Light production in such a system can occur either by activation of a chromophore by energy transfer from another activated chromophore or by reassembly of a bright reporter. However, recent efforts to apply Förster Resonance Energy Transfer (FRET) to sense DNA[22-26] have still required more than three unique sgRNAs, while split reporter DNA or RNA biosensing has been described mainly by previous studies *in vitro*[27-31] with several studies describing using transcription activator-like effectors (TALEs) as DNA binding domains and split fluorescent proteins for DNA biosensing in live cells[32-33].

In addition to higher background from unbound probes, fluorescence-based biosensing is plagued by issues with phototoxicity and photobleaching as well as a naturally high cellular auto-fluorescent background[34-36], all of which negatively impact the SBR. To counteract these negative effects on biosensing SBR and increase sensitivity for an underlying physicochemical target, luminescent reporters could offer an attractive alternative in biosensing experiments. Cellular luminescent background signal is essentially nonexistent due to the fact that light is produced from a catalytic reaction of an enzyme with its substrate instead of from excitation by incident exogenous light[35]. While luminescent reporters have the advantage of lower background, one historical advantage of fluorescent reporters is that they have remained brighter than available luminescent reporters[37]. However, a relatively new luciferase, NanoLuc (NLuc; Promega Corporation, Madison, WI) bridges this gap in signal intensity. NLuc offers several advantages over Firefly (FLuc) and Renilla (RLuc) luciferases including enhanced stability, significantly smaller size, and >150-fold enhancement in luminescence output[37-38].

Furthermore, the substrate for NLuc, furimazine, is more stable and exhibits decreased levels of background activity than the substrate for RLuc, coelenterazine,[37-38].

We developed a split luciferase DNA biosensor based on the NanoLuc Binary Technology (NanoBiT; Promega Corporation, Madison, WI) complementation reporter system created for NLuc[39] and catalytically inactive Cas9 (dCas9) from *Streptococcus pyogenes*. Due to the high dissociation constant (K_D=190 μM) and extremely low catalytic activity of the NanoBiT complementation reporter system subunits—termed LgBiT and SmBiT—they must be brought into close proximity in order to reassemble full-length NLuc. Thus, our approach was to direct two dCas9-sgRNA-NanoBiT complexes to two target DNA sites with a specific DNA orientation and spacing. Initially, we achieved a maximum of 8-fold increase in signal in live populations of cells transfected with the biosensor and various target DNA scaffolds compared to populations transfected with the biosensor but no target DNA. Subsequently, we tested the sensitivity of the biosensor on specific endogenous genomic DNA sequences across multiple cell lines and compared the signal-to-background of this approach to a common fluorescence-based method. Finally, we were able to detect single copy genome edits induced by CRISPR-Cas9 with single cell resolution through distinct differences in signal intensity between homozygous mutant and wild-type cells.

MATERIALS AND METHODS

Construction of Directional dCas9-NanoBiT and dCas9-NanoLuc Fusion Proteins

The directional fusion constructs containing the LgBiT and SmBiT of NLuc (Promega Corporation) fused to catalytically inactive Cas9 (D10A and H840A double mutant) were generated using the Gibson Assembly method (New England Biolabs). We used an improved version of the pCDNA3-dCas9 containing two nuclear localization signals, an N-terminal 3× Flag epitope tag and [$(GGS)_5$] flexible linker sequences and well as two separate multiple cloning sites at the N- and C-termini of dCas9 (vector map shown in **Supplementary Methods 1**). The LgBiT and SmBiT were each cloned onto the N- and C-termini of dCas9 using two separate multiple cloning sites in the modified pCDNA3-dCas9 vector (see **Supplementary Methods 1** for sequences). Overnight N- and C- terminal double restriction digests of sets of flanking restriction sites (XbaI, KpnI) and (NheI, NotI) produced the necessary vector backbones for subsequent Gibson Assembly. LgBiT and SmBiT inserts were ordered as gBlocks Gene Fragments (Integrated DNA Technologies) containing approximately 45 bp homologous sequences with the doubly-digested dCas9 vectors upstream and downstream of the two cut sites. A positive control NLuc-dCas9 fusion construct was created using overlap extension PCR on LgBiT-dCas9 and SmBiT-dCas9 gBlocks to directionally splice the sequences followed by the Gibson Assembly method again using the N-terminal doubly digested dCas9 vector. The four assembled dCas9-NanoBiT constructs, the dCas9-NLuc construct, and pGL4.53 [luc2/PGK] Firefly luciferase vector (Promega Corporation) were separately transformed into 5-alpha Competent *E. coli* (New England Biolabs) using a standard chemical transformation procedure

with heat shock at 42°C and transformed *E. coli* were plated on LB plates containing ampicillin at a final concentration of 100 µg/mL. After an 18 h incubation at 37°C, MiniPreps (QIAGEN) were created for a subset of large, well-separated colonies. The selected subset of large colonies was screened for recombinant vector and insert using both diagnostic restriction digests and colony PCR. Clones positive for the four NanoBiT inserts, the full NanoLuc insert, and the luc2 insert using both methods were subsequently sequenced to confirm exact sequences were present.

Construction of sgRNA Expression Plasmids

The sgRNA expression vector backbone was obtained from Addgene (Addgene #41824) and was linearized using a restriction digest with AflII. Two 19-bp sgRNA target sequences common throughout several genomes but not present in the human genome were selected using CRISPRscan and the UCSC genome browser (see **Supplementary Methods 2** for sequences). Each sgRNA sequence was incorporated into two 60mer oligonucleotides that contained homologous sequences to the sgRNA expression vector for subsequent Gibson assembly. After oligonucleotide annealing and extension, the PCR-purified (PCR purification kit; QIAGEN) 100 bp dsDNA was inserted into the AflII linearized sgRNA expression vector using Gibson assembly.

Construction of sgRNA Target Site Vector Scaffolds

Scaffolds containing the two sgRNA target sequences in tandem, inverted, and everted orientations were created using two separate plans. The first plan consisted of a series of overlap extension PCRs on ssDNA oligonucleotides (Integrated DNA Technologies) followed by PCR

purification using the MinElute PCR Purification Kit (QIAGEN). The resulting target sequence scaffold oligonucleotides were then subjected to a final amplification with 2X GoTaq Green Master Mix (Promega Corporation) to create poly-dT tails and cloned into the PCR4TOPO vector using the Topo TA Cloning Kit for Sequencing (Invitrogen). The second plan consisted of a series of targeted blunt-end double restriction digests on cloned scaffolds from the first plan, PCR-purification (removing oligonucleotides <~70 bp) again using the MinElute PCR purification kit (QIAGEN), and re-ligation using excess T4 DNA ligase (New England Biolabs). See **Supplementary Methods 3** for sequences.

Plasmid-Based DNA Biosensor Testing in Live HEK 293T Cells

HEK 293T cells were originally purchased from ATCC and maintained in Dulbecco's Modified Eagle Medium (Life Technologies) supplemented with 10% FBS and 1X Penicillin/Streptomycin at 37 °C under 5% CO2. In the first experiment, which sought to determine the optimal molar transfection ratio of LgBiT to SmBiT fusion constructs, 25,000 low passage HEK 293T cells per well were seeded in 96-well white opaque-side microplates (Thermo Fisher Scientific) approximately 20 h before transfection. These cells were then transiently transfected with 100 ng total DNA per well using the Lipofectamine 3000 transient transfection protocol (Invitrogen). Each well was co-transfected with 16.67 ng/well of plasmid expressing each dCas9-NanoBiT fusion construct, 16.67 ng/well of plasmid expressing each of two sgRNAs, 16.67 ng/well of plasmids containing the target sequence, and 16.67 ng/well pMAX-GFP plasmid. With these methods, cells were typically transfected at approximately 90-95% efficiency. We tested various LgBiT:SmBiT molar transfection ratios with the construct in excess being transfected at 16.67 ng/well and the lesser construct being decreased by specific amounts to form desired molar

transfection ratios. 33 of the LgBiT + SmBiT wells were transfected with the tandem PAMs 10 bp apart target sequence scaffold and 33 of the LgBiT + SmBiT wells were identically transfected but without any target DNA. For wells that did not reach 100 ng total DNA, pUC19 vector was transfected to make up the difference. In this experiment, signals were measured 24 h post-transfection. In our next experiment, several molar excesses of sgRNA to dCas9-NanoBiT fusion constructs (1:1, 1.2:1, 2:1, 5:1, and 20:1) were delivered to cells using the same method as described above, holding the molar amount of sgRNA constant but decreasing the molar amount of dCas9-NanoBiT fusion proteins. We then held the 20-fold molar excess sgRNA parameter constant and progressively decreased the amount of target DNA transfected, making up the difference with pGL4.53 [luc2/PGK] Firefly luciferase vector (Promega Corporation), essentially random DNA with no binding sites with >5 bp homology with the protospacer of either sgRNA. All fluorescent signals were measured on the SpectraMax M5 Microplate Reader (Molecular Devices) with high PMT sensitivity setting and 100 reads/well before taking any luminescent readings. After adding 25 µL furimazine substrate (Promega Corporation) reconstituted at a 1:19 volumetric ratio with Nano-Glo LCS Dilution Buffer (Promega Corporation) according to the Nano-Glo Live Cell Assay System protocol to each well, luminescent signals were measured on the SpectraMax M5 Microplate Reader with 1 s integration and high PMT sensitivity setting. The ideal delivery parameters were used with the same Lipofectamine 3000 transfection protocol for comparing all orientations of PAM orientation, spacer length, and dCas9-NanoBiT fusion construct pairing.

Luminescence Microscopy and Image Processing

Transfection experimental setup for microscopy sessions was identical to the setup for microplate reader sessions except that inert pUC19 plasmid was added to the transfection mix to account for the amount of plasmid lost by eliminating target DNA and sgRNAs in background transfection conditions. In addition, an auto-association background condition without target DNA (mouse cell lines transfected with sgRNA to locus 1) was included in the measurements of the non-repetitive region of *MUC4* intron 1 in addition to the no sgRNA background condition as locus 1 sgRNA has no matches with 100% homology within the mouse genome. With these methods, cells were typically transfected at approximately 90-95% efficiency. In these experiments, low-passage HEK 293T, HeLa, MCF7, HCT116, K562, and JLat cells were plated in SensoPlate 24 Well F-Bottom, Glass Bottom Black Microplates (Greiner Bio-One). Next, instead of imaging whole well populations of adherent cells, we split the cells to $1.5x10^5$ cells/mL and took images of the cell suspensions on Superfrost Plus Microscope Slides (Thermo Fisher Scientific) with Premium Cover Glass (Thermo Fisher Scientific) combined at a 1:1 volumetric ratio with reconstituted furimazine substrate (Promega Corporation). The Nano-Glo Live Cell Reagent furimazine is nontoxic and nonlytic when delivered to live cells. An optimized NLuc imaging protocol was developed for use on the Leica DM6000 B fully automated upright microscope equipped with the Leica DFC9000 GT sCMOS camera and the Exfo X-Cite 120 Fluorescence Illumination System in which cells were placed in a dark box with all light sources blocked and lamp intensity was set to 0, exposure time was set to 30 s, and sCMOS gain was set to 2.0. The GFP signal produced from pMAX-GFP transfection was imaged using an exposure time of 150 ms and sCMOS gain of 1.0.

Post-processing was applied to better visualize the images. Raw 16-bit grayscale GFP images were recolored using the "green" look up table (LUT), brightness was increased by 50%, and contrast was decreased by 50% in Fiji. Raw 16-bit grayscale NLuc images were recolored using the "red hot" LUT, which displays areas of highest intensity as white and areas of lower or average intensity as varying shades of red. Then, brightness was increased to 100% and contrast was decreased by approximately 50%. Subsequently, the subtract background function was applied in Fiji (Image J) with radius 5.0 and "create background (don't subtract)" option applied to reduce diffuse background and artifacts from the imaging process. To merge GFP and NLuc images, we directly merged color channels in Fiji (Image J).

Quantitation was then performed on the original unprocessed images. The WEKA Segmentation package[45] in Fiji (Image J) was used to segment cells using 50 ROI traces of the nuclear NLuc signals and 50 ROI traces of the background outside of cells as a training data set for nuclear boundaries. This trained WEKA segmentation model was then applied to each NLuc image to determine boundaries of nuclei. Each 8-bit segmented image outputted from the WEKA model was binarized using the auto-threshold function, and the analyze particles function was applied to create ROIs for each nuclear area. These ROIs were then overlaid onto the original unprocessed images. Then, the mean intensity (equivalent to integrated intensity of each nucleus divided by area of each nucleus) after 30 s of total light collection was calculated and recorded for each segmented nuclear area using ImageJ (Fiji). Any cells that were positive for GFP signal but negative for NLuc signal were omitted from final statistical analysis.

Statistical Testing

Two-tailed Student's t-tests for signal-to-background analyses were conducted in Microsoft Excel 2016. Two-way ANOVA and pairwise Tukey's HSD post-hoc tests were conducted in R (version 4.0.3) on combinatorial signals from initial biosensing experiments in live cells. Statistics shown in all box-and-whisker plots were computed in R (version 4.0.3).

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files). Source data are provided with this paper.

Code Availability

The Trainable WEKA Segmentation model used within this study is included in this published article as a .model file executable directly within ImageJ (Fiji).

RESULTS

Construction and optimization of a split luciferase DNA sequence biosensor

To design a live cell DNA sequence biosensor, we fused two independently optimized protein fragments of NLuc—LgBiT and SmBiT—to a catalytically inactive Cas9 from *S. pyogenes* (dCas9) (**Fig. 1a**). We constructed five fusion protein plasmids: two in which the LgBiT and SmBiT were fused to the carboxy-terminus of dCas9 (dCas9-LgBiT and dCas9-SmBiT), two in which they were fused to its amino-terminus (LgBiT-dCas9 and SmBiT-dCas9) and one in which full-length NLuc was fused to the amino-terminus of dCas9 (NLuc-dCas9) **(Fig. 1c,**

Supplementary Methods 1). For target sites, we produced 33 plasmids each harboring one copy of a DNA target site scaffold containing two SpCas9 sgRNA target sites in three orientations with 1-50 base pair (bp) spacer sequences between them in tandem, inverted, and everted orientations (**Fig. 1b, Supplementary Methods 2**). The two sgRNAs were chosen to have no homology within the human genome and minimal off-targets (**Supplementary Methods 3)**. To determine optimal conditions for this biosensor, different molar ratios of dCas9- and sgRNA-expressing plasmids and different molar amounts of target DNA plasmids were transiently transfected into HEK 293T cells with a range of incubation times post-transfection **(Fig. 1d)**. A common reporter for transfection efficiency, pMAX-GFP, was co-transfected in all conditions. Signal-to-background peaked when we used a 10:1 ratio of LgBiT-dCas9 to dCas9-SmBiT fusion proteins, a 20:1 ratio of sgRNA:total NanoBiT plasmid, and a 24 h incubation time between transfection and signal measurement (**Supplementary Figure 1**). In addition, we found very little dependence of signal-to-background on molar amount of target DNA transfected (**Supplementary Figure 1**). Hypothesizing that fusion protein orientation and target DNA orientation might have a synergistic effect on signal output, we conducted a two-way ANOVA assuming there was an interaction between these two variables. Significant variation in the efficiency of NLuc reassembly was observed across conditions **(Fig. 1e)**, with fusion protein orientation and target DNA orientation being associated with significant differences in luminescent signal output ($p < 0.0001$ and $p < 0.05$, respectively, two-way ANOVA, **Supplementary Table 1**). The relationship between signal output and fusion protein orientation was also shown to depend on target DNA orientation and vice versa ($F(96, 264) = 2.064$, $p < 0.0001$, two-way ANOVA, **Supplementary Table 1**) indicating that these results are affected by an interaction between fusion protein and target DNA orientations. In addition, the LgBiT-dCas9

+ dCas9-SmBiT protein configuration produced the highest set of luminescent signals, ($p < 0.0001$ for three pairwise comparisons, Tukey HSD, **Supplementary Table 2**).

Live single-cell imaging of repetitive and unique endogenous genomic sequences

After optimizing delivery conditions for our DNA sequence biosensor in live cells using a luminometer to measure luminescence across whole well cell populations, we sought to investigate the feasibility of detecting luminescence in single cells on relatively common imaging equipment. To this end, we modified an upright fluorescence microscope to capture the relatively low light intensities associated with NLuc and other luminescent reporters. Cells were placed in a dark box with all light sources blocked and exposure times were lengthened. To determine the applicability of our split luciferase biosensor to imaging endogenous DNA sequences, we first compared its sensitivity to that of both a previously described dCas9-EGFP fluorescent probe[14] and the NLuc-dCas9 probe from our study. We used a single optimized sgRNA, sgMUC4-E3(F+E)[14] to direct these probes to bind a region of polymorphic 48-bp repeats of copy number between approximately 100 and 400 within exon 2 of the human *MUC4* locus **(Fig. 2a)**. We found that both full reporter probes had comparable signals when binding the tandem repeats compared to a background condition with no sgRNA in HEK 293T and HeLa cells **(Supplementary Figure 2)**. We then used sgMUC4-E3(F+E) as an anchor sgRNA and constructed four sgRNAs with unique spacer lengths and orientations around it to direct our split luciferase probe to bind the same repetitive region of *MUC4* in HEK 293T cells (see **Supplementary Methods 4** for target sequences and construction methods).

Subsequently, we compared signals of pairwise combinations of each the four unique sgRNAs and sgMUC4(F+E) to signals of an identical transfection without gRNA in HEK 293T cells. We observed variable sensitivity for the *MUC4* tandem repeats based on molar amount of probe transfected and target site configuration **(Fig. 2b-c, Supplementary Figure 8a)**. Overall gRNA-directed signal was greater at higher concentrations of probe, but with accompanying greater background signal in cells lacking gRNA. However, we found that signal-to-background was greatly improved to approximately 5.5-7-fold by reducing the amount of probe delivered from 10 fmol to 1 fmol. To assess the diagnostic power of our probe to detect true positives for presence of a single repetitive genomic locus, we used Receiver Operating Characteristic (ROC) analysis. We found that sensitivity and specificity of the probe increased at lower concentrations in transfection, from ~0.89 to ~0.94 and from ~0.82 to 1.0, respectively. **(Fig. 2b-c)**.

Since the majority of loci within the human genome are non-repetitive, a more significant application would be the potential detection of such low copy number, unique genomic sequences. To this end, we targeted the non-repetitive region of intron 1 of the human *MUC4* locus with 1-3 pairs of unique sgRNAs tiling along the locus with at least 200 bp between pairs to avoid interactions between probe components at different binding sites (**Fig. 3a**, **Supplementary Figure 3, Supplementary Methods 4** for target sequences and construction methods). We observed strong cell type-specific differences in biosensor sensitivity based on the amount of probe transfected **(Fig. 3b, Supplementary Figures 3-5)**. Specifically, signal-to-background in HeLa cells peaked at approximately 1.3-fold using a single pair of sgRNAs and 10 fmol probe in transfection but at 7-fold using 0.1 fmol probe in transfection. Furthermore, to

assess the diagnostic power of our probe to detect true positives for presence of a single non-repetitive genomic locus, we used Receiver Operating Characteristic (ROC) analysis. We found area under the curve (AUC) was 0.608 at 10 fmol probe transfected to HeLa cells, whereas AUC increased to 0.992 when 1 fmol was transfected **(Fig. 3c)**. Likewise, using a single pair of sgRNAs and 10 fmol probe in MCF7 cells, signal-to-background peaked at 4.5-fold, but was increased to 7.6-fold by reducing amount of probe transfected to 0.1 fmol. Whereas AUC was 0.877 at 10 fmol transfected to MCF7 cells, AUC increased to 0.983 when 1 fmol was transfected **(Fig. 3d)**. At 10 fmol probe transfected in HCT116, K562, 293T, and Jlat cells, AUC was 0.622, 0.734, 0.841, and 0.856, respectively (**Supplementary Figure 3c)**.

Live single-cell biosensor imaging of single-base changes induced by CRISPR-Cas9 editing

One of the most pertinent applications for our split luciferase biosensor is the detection of various mutations in genomic DNA sequence after targeted genome editing with CRISPR-Cas9. Thus, we created G>T missense single nucleotide polymorphisms (SNPs) at two different loci in two cell lines: within the 8q24 multi-cancer risk locus in HCT116 cells and within the PALB2 locus in HEK 293 cells **(Fig. 4a)**. Both SNPs were present within the PAM site of the sgRNA used for editing[40] (**Supplementary Methods 5)**. In a previous study, mutants were confirmed to be homozygous for the G>T mutations by dilution plating followed by detection of specific alleles by Kompetitive Allele-Specific PCR (KASP) in expanded populations[40]. We hypothesized that these mutations should completely inhibit binding by the sgRNA used for editing, making luminescence within mutant cell clones lower than luminescence within wild-type cell clones transfected with the same probe components. To investigate this hypothesis, we

transfected wild-type and homozygous mutant clones of both cell lines with the dCas9-NanoBiT probe carrying the sgRNA used for editing along with 4-5 sgRNAs flanking it in various orientations at various distances. We observed reduced luminescence in the mutant clones compared to the wild-type clones when 0.1 fmol probe was delivered to cells **(Fig. 4b-c, Supplementary Figures 6 and 7)**.

Specifically, luminescence in the HEK 293 wild-type clones was approximately 2.8-11.5-fold higher across all five sgRNA pairs tested at *PALB2* compared to HEK 293 mutant clones **(Fig. 4d)**. Likewise, in HCT116, luminescence in the wild-type clones was approximately 1.2-9.8-fold higher across all four sgRNA pairs tested at the 8q24 poly-cancer risk locus compared to mutant clones **(Fig. 4d)**.

DISCUSSION

Traditionally, fluorescence signal-to-background analysis involves specialized confocal fluorescent microscopes that can resolve bright foci and compare these signals to the background nucleoplasm at high resolution. We envisioned a platform for measurement of luminescence in single cells on relatively common imaging equipment normally geared toward fluorescence. Our hope was that this might reduce the need for expensive, specialized imaging equipment for luminescence and ultimately serve to lower the barriers to entry for this technique. Unlike imaging for traditional fluorescent DNA probes, the imaging resolution of this technique is insufficient to consistently distinguish bright foci from the larger regions of signal accumulation within the nucleoplasm. However, a range of signal intensity across the nuclear regions is visible in many nuclei. It is possible the whiter signal areas represent the approximate location of the

probe within the nucleus and the remainder of the signal accumulation area represents free-floating, complemented NanoLuc. In addition, our no sgRNA background condition represents signal resulting from complemented NLuc that is not bound to DNA for a given transfection condition. Thus, we expect differences in our SBR metric are primarily due to total complemented NanoLuc minus free-floating complemented NanoLuc and the majority of signal in on-target conditions should be produced through bright luminescent foci on the DNA.
When optimizing delivery conditions, we found using 20-fold molar excess sgRNA plasmid in transfection compared to dCas9-NanoBiT plasmid resulted in an increase in signal-to-background compared to other tested ratios. This result can partially be explained by the shorter nuclear lifetime of cellular RNAs compared to both cellular DNA and proteins[41]. Since RNA molecules are degraded much faster than their DNA and protein counterparts, transient plasmid transfection-based delivery of this biosensor may require higher initial amounts of DNA template for the sgRNA to reach a steady-state level of transcription in cells. These factors may also explain our ideal incubation time before measurement of luminescence post-transfection of 24 h. Plasmid transcription, mRNA degradation, and mRNA translation show exquisite temporal control in cells[41], and a 24 h incubation time likely resulted in fairly stable levels of both the dCas9-NanoBiT fusion proteins and available sgRNAs, allowing for high rates of sgRNA-fusion protein association and DNA binding in cells.

In all cases where <10 fmol of probe plasmid was delivered to cells, we observed differences between on-target conditions and background conditions at both repetitive and non-repetitive regions of *MUC4* that were statistically significant ($p < 0.01$ and $p < 0.0001$, respectively, unpaired Student's t-test, two-tailed). These marked differences in signal intensities indicated

NLuc reassembly was occurring in target cell nuclei upon binding of the probe to these regions of the *MUC4* locus. In addition, ROC analysis showed our probe to be an excellent discriminator of true and false positive detection events when transfected at low concentrations at endogenous *MUC4* with high area-under-the-curve for both repetitive and non-repetitive sequences. Thus, our split luciferase probe can detect low copy number sequences with high sensitivity and specificity and optimal signal cutoff points can be selected using statistics such as Youden's J to maximize these parameters. However, given variable signal-to-background ratios across six cell lines tested, the performance of this probe is moderately cell-type specific. A number of factors may be expected to vary to some degree across different cell lines, including transfection efficiency, sgRNA and fusion protein decay rates, uptake efficiency of the luminescent substrate, or attenuation rate of the resulting signal. In addition, we observed a relatively high degree of variability within each transfection condition. Additional optimization of probe design and methodology in future studies, such as fine-tuning linker composition and length and exploring alternatives to transient transfection, could reduce variability, increase complementation efficiency, and improve the utility of this approach. The copy number of a given genomic locus may also vary across cell lines, with higher copy numbers potentially resulting in more robust signal output.

In addition, we envisioned that we could apply our probe to isolate mutant cells from a population of genome-edited cells via detection of SNPs induced by editing experiments at two genomic loci in HEK 293 and HCT116 cells. We found that luminescent signals were higher across several sites bound by sgRNA pairs around the original Cas9 cut site in wild-type HEK 293 and HCT116 cells compared to homozygous mutant HEK 293 and HCT116 cells. This

effectively demonstrated differentiation between binding two and zero copies of the target sequence, as HEK 293 cells have two copies of chromosome 16 and HCT116 cells have two copies of chromosome 8 with no commonly reported abnormalities[42-43]. We hypothesized that mutating the PAM site in both cell lines would create a condition where Cas9 would not be able to recognize the original target site[44]. The fact that all sgRNA pairs produced higher signals in wild-type compared to mutant cells shows that our split luciferase probe can detect very minor differences in genomes across individual cells, including differences in SNP copy number. For mutant cells to be expected to glow brighter than wild-type so desired mutants can be more easily isolated in practice, guide RNA design for the probe could be specifically altered. In addition, ROC analysis of this data showed our probe to be an excellent discriminator of true and false positives for SNP detection with high area-under-the-curve for edits at both *PALB2* and 8q24. Thus, our probe can also detect SNPs with high sensitivity and specificity. However, it is worth noting that mismatches within the PAM site are less permissive to binding than those within the sgRNA hybridization region. SNPs occurring in the protospacer region should be evaluated for their effect on specificity of the probe. For screening edited cells from wild-type cells within an edited cell population, perhaps the most pertinent metric is the minimum fraction of edited cells required to be screened in order to make a positive call. This is also equal to 1 minus the probability of correctly identifying a positive clone within the mutant cell population or 1 - sensitivity. For the 293 edited cells, since we observed a sensitivity of ~0.91 within an isogenic mutant population, we would need to screen a minimum of 9% of cells within this population to detect a true positive mutant clone. Likewise, for the HCT116 edited cells, since we observed a sensitivity of ~0.88, we would need to screen a minimum of 12% of cells within this population to find a true positive mutant clone. In practice, the minimum fraction of cells

required to be screened from an edited cell population would be much higher, as this population is a mixture of both homozygotes and heterozygotes and the ratio of mutant to wild-type clones would depend on the editing efficiency.

In practice, signal detection with this probe could potentially be used as a precursor to manual single cell isolation after gene editing, which could allow users of gene editing techniques to save valuable time and resources during the single cell cloning process. As we transfect at a high density and replate cells at a low density for imaging, this technique sets the stage for the next step, which would be to isolate the cells with the detected edit. It is conceivable that after transfecting an edited cell population with our probe and taking a subset of this edited population for imaging, single mutant cells could be isolated by manual separation using specialized cloning cylinders based on differences in luminescent signal intensity. Then, these isolates could be clonally expanded to produce an isogenic cell population. We conclude that this split luciferase probe should be a broadly useful platform for many live cell DNA biosensing applications that require low copy number resolution and minimal destruction of highly valuable cell populations, including identification and isolation of mutant cells from a population of cells that has undergone a genome editing procedure, real-time identification of cells harboring new driver mutations or broad chromosomal rearrangements such as inversions or translocations during neoplasia, or more generally, *in situ* genotyping of heterozygotes and homozygotes at a defined locus

FIGURES

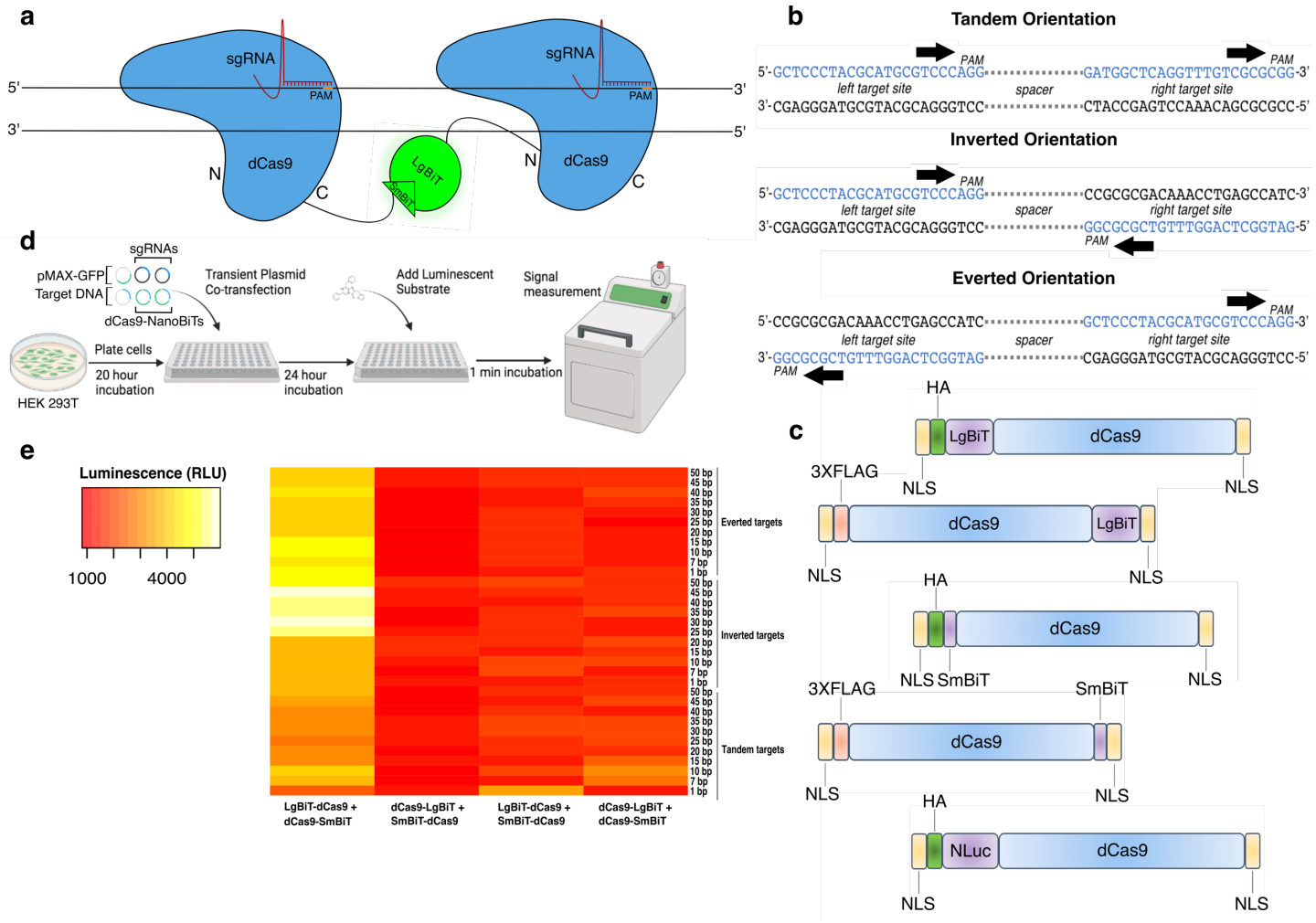


Figure 1: Design and characterization of a split luciferase DNA biosensor

(a) A cartoon depiction of sequence-dependent reassembly of NanoLuc luciferase. **(b)** Schematic of target site designs with PAM sites in tandem (parallel on the same strand), inverted (PAMs oriented inward on opposite strands) and everted (PAMs oriented outward on opposite strands). **(c)** Cartoon representation of dCas9-NanoBiT and full-length dCas9-NanoLuc fusion constructs. **(d)** Depiction of experimental process for initial luminometer-based DNA biosensing assays labeling all co-transfected plasmids. **(e)** A heat map showing variation in luminescent signal intensity between four possible orientations of dCas9-NanoBiT fusion proteins

across 33 DNA target site spacings and orientations. Sequential scale ranges from lowest signals of the set (red) to highest signals of the set (white).

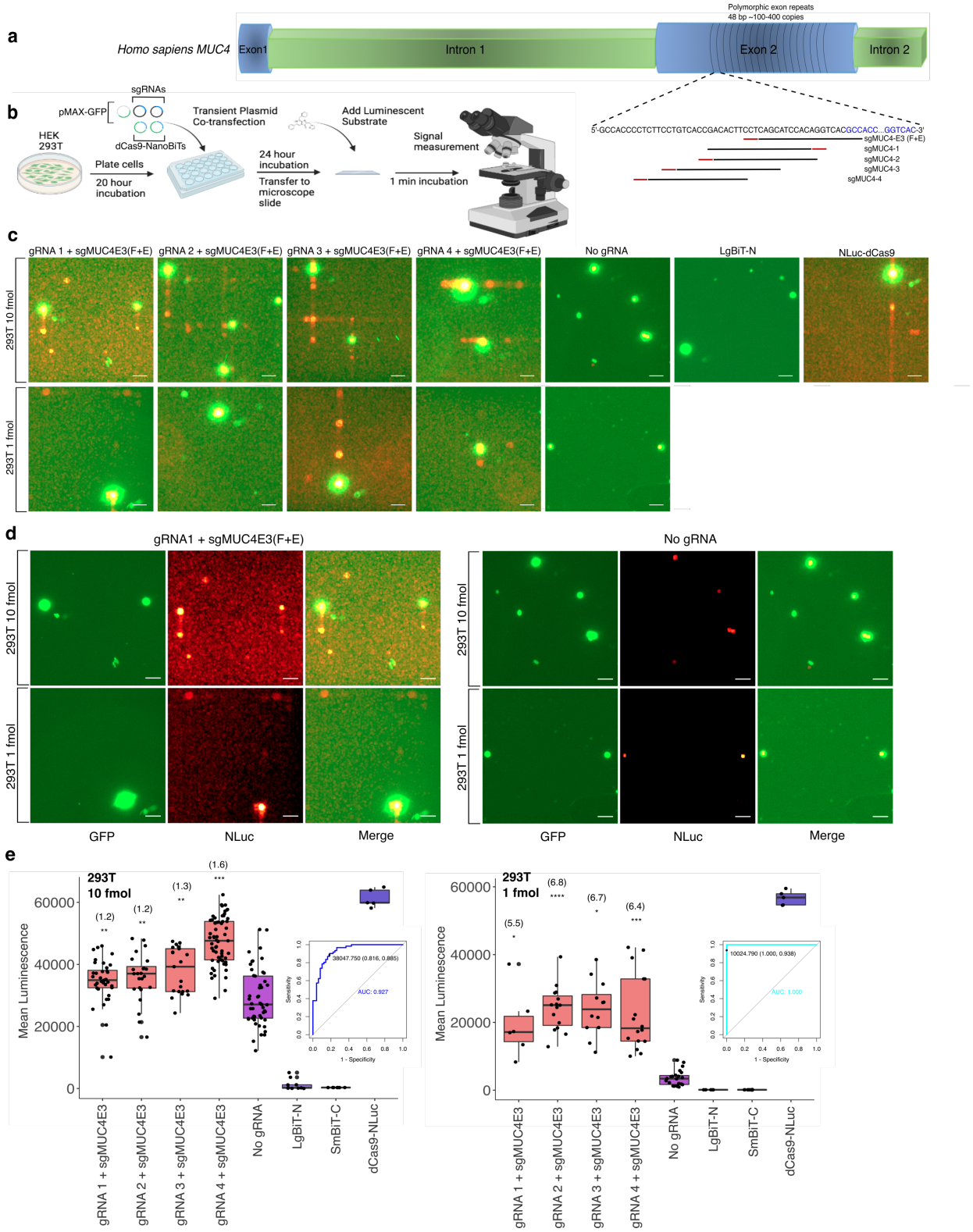


Figure 2: Biosensing repetitive genomic sequences at the human *MUC4* locus

(a) Cartoon visualization of the repetitive region of exon 2 within the human *MUC4* locus showing sgRNA design strategy using the binding site for sgMUC4-E3(F+E) as an anchor point. **(b)** Depiction of experimental process for all live cell DNA biosensing assays conducted using microscopy, labeling all co-transfected plasmids. **(c)** Merged GFP fluorescence (green) and NLuc luminescence (red) images taken on the Leica DM6000B upright microscope at 10X magnification depicting dCas9-NanoBiT biosensing of the repetitive region of *MUC4* exon 2 in live HEK 293T cells using two different amounts of dCas9-NanoBiT plasmids in transfection (10 and 1 fmol). Scale bars = 50 μM. **(d)** Individual GFP fluorescence (green), NLuc luminescence (red), and merged images taken on the Leica DM6000B upright microscope at 10X magnification depicting dCas9-NanoBiT biosensing using sgRNA 1 paired with sgMUC4E3(F + E) at the repetitive region of MUC4 exon 2 compared to no sgRNA controls using two different amounts of dCas9-NanoBiT plasmids in transfection (10 and 1 fmol). Scale bars = 50 μM. **(e)** Signal quantification for images taken of the split luciferase probe binding the repetitive region of *MUC4* exon 2 in live HEK 293T cells. Apparent signal-to-background ratios are listed in parentheses above each biosensing condition. $5 < n < 61$, where n represents the number of unique cells quantified; unpaired two-sided Student′s t-test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$. Boxes show the median and interquartile range (IQR) and whiskers show dispersion from the IQR that is equal to the lesser of the 1st or 3rd quartiles plus 1.5xIQR or the distance from the 1st or 3rd quartiles to the minimum or maximum points, respectively. Receiver Operating Characteristic (ROC) curves representing biosensing results using sgRNA 4 paired with sgMUC4E3(F+E) at 10 fmol and 1 fmol transfected within the repetitive region of *MUC4* in HEK 293T are shown. False positives were determined by signals due to auto-assembly (no sgRNA). The signal threshold for distinguishing true positives from false positives that maximized Youden's J Statistic (sensitivity + specificity – 1) is shown as a point on the ROC curve along with corresponding specificity and sensitivity values in parentheses.

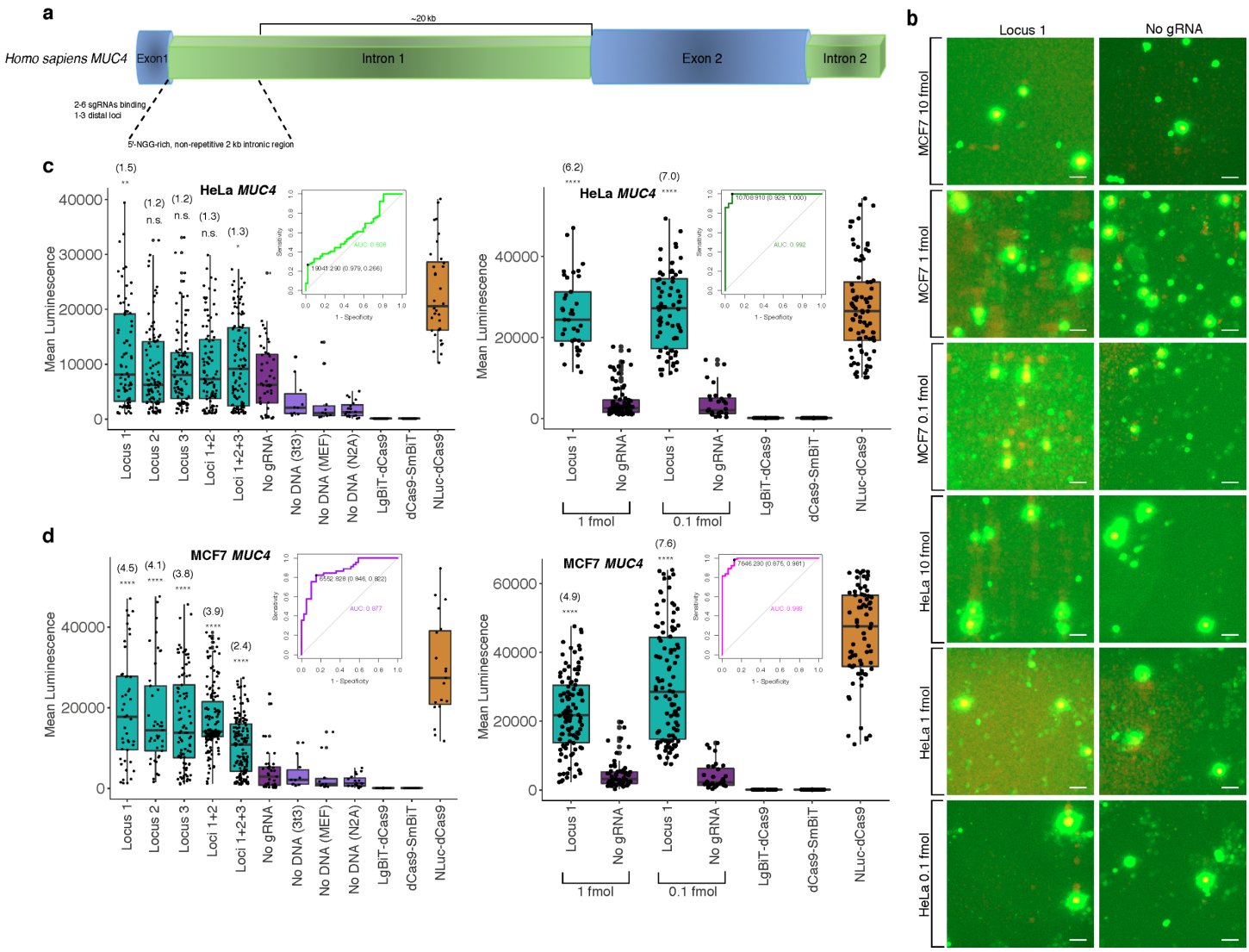


Figure 3: Biosensing non-repetitive genomic sequences at the human *MUC4* locus

(a) Cartoon visualization of the non-repetitive region within intron 1 of human *MUC4* showing sgRNA design strategy. **(b)** Merged fluorescence (green) and luminescence (red) images taken on the Leica DM6000B upright microscope at 10X magnification depicting dCas9-NanoBiT biosensing of a single locus within non-repetitive *MUC4* intron 1 in live MCF7 and HeLa cells. Scale bars = 50 µM. **(c)** Signal quantification for images taken of the dCas9-NanoBiT probe binding to several combinations of loci in the non-repetitive region of *MUC4* intron 1 in HeLa cells at 10 fmol probe transfected (left) and for images taken of the probe binding to a single non-repetitive locus within *MUC4* intron 1 at 1 and 0.1 fmol probe transfected in HeLa cells (right). Apparent signal-to-background ratios (no sgRNA=background condition) are listed in parentheses. $32 < n < 150$ for HeLa

cells at 10 fmol and $28 < n < 100$ for HeLa cells at the two lower concentrations, where n represents the number of unique cells quantified; unpaired two-sided Student′s t-test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$. Receiver Operating Characteristic (ROC) curves representing biosensing results at *MUC4* locus 1 at 10 fmol (left) and 1 fmol (right) probe transfected in HeLa cells are shown. **(d)** Signal quantification for images taken of the dCas9-NanoBiT probe binding to several combinations of loci in the non-repetitive region of *MUC4* intron 1 in MCF7 cells (left) and for images taken of the probe binding to a single non-repetitive locus within *MUC4* intron 1 at two lower concentrations in MCF7 cells (right). Apparent signal-to-background ratios are listed in parentheses. $19 < n < 159$ for MCF7 cells at 10 fmol and $32 < n < 106$ for MCF7 cells at the two lower concentrations, where n represents the number of unique cells quantified; unpaired two-sided Student′s t-test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$. Receiver Operating Characteristic (ROC) curves representing biosensing results at *MUC4* locus 1 at 10 fmol (left) and 1 fmol (right) probe transfected in MCF7 cells are shown. Boxes show median and IQR and whiskers show dispersion from IQR.

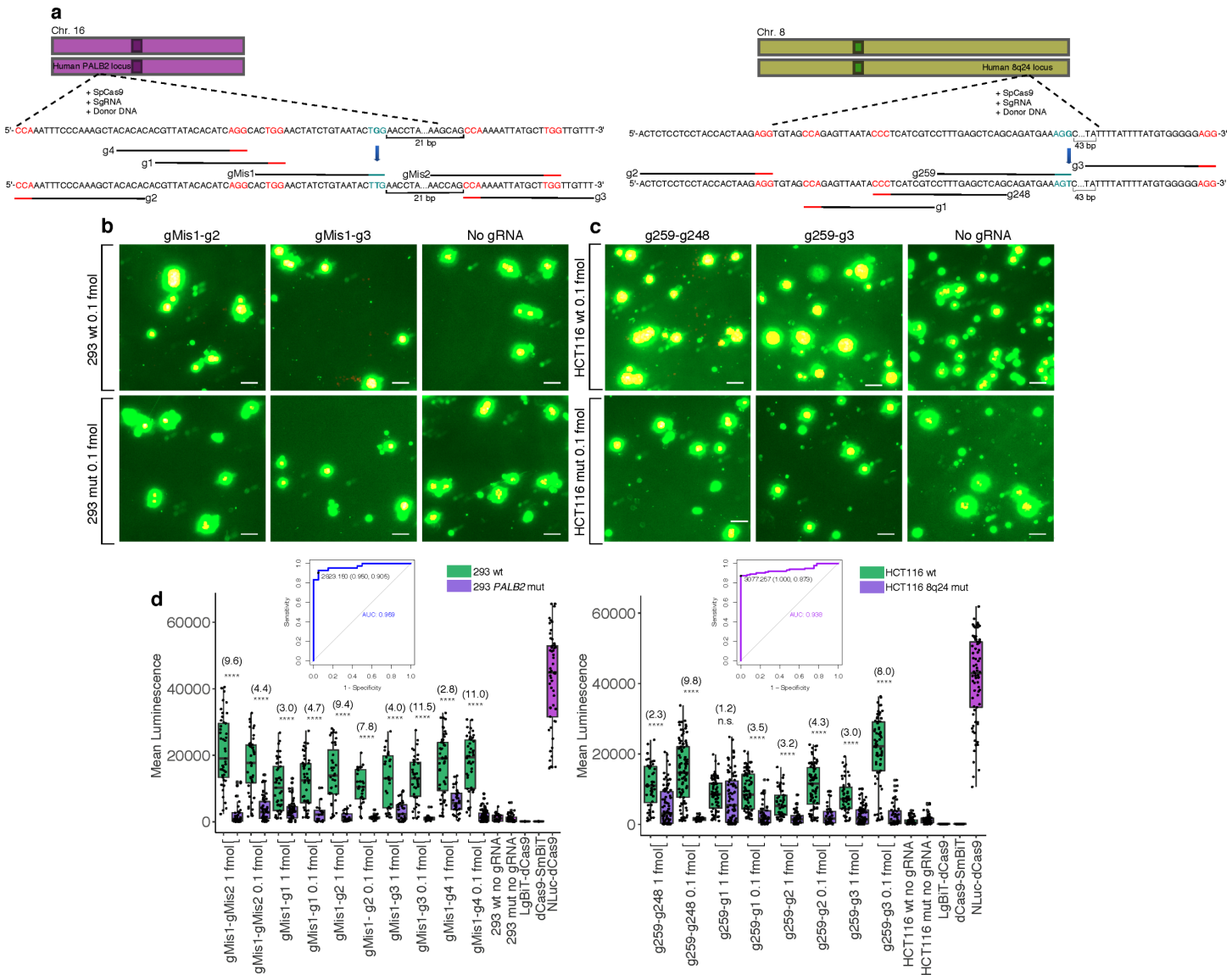


Figure 4: Biosensing CRISPR-Cas-induced genome edits in live cells

(a) Cartoon visualization of the CRISPR-Cas editing experiments conducted at the human 8q24 poly-cancer risk locus and at *PALB2*. sgRNAs used for editing have blue PAM sites while sgRNAs around the site of mutation used for detection of mutant cells in biosensing experiments have red PAM sites. Single base pair edits are shown in bold. **(b)** Merged fluorescence (green) and luminescence (red) images taken on the Leica DM6000B upright microscope at 10X magnification of the dCas9-NanoBiT probe applied to the *PALB2* locus in wild-type and confirmed homozygous mutant cells after targeted CRISPRCas9 genome editing. Scale bars = 50 µM. **(c)**

Merged images taken on the Leica DM6000B upright microscope at 10X magnification of the dCas9-NanoBiT probe applied to the 8q24 poly-cancer risk locus in wild-type and confirmed homozygous mutant cells after targeted CRISPR-Cas9 genome editing. Scale bars = 50 μM. **(d)** Signal quantification for the above images as well as several other sgRNA pairs. Apparent signal-to-background ratios (mutant cell signal=background) are listed in parentheses. Data are presented such that boxes show the median and interquartile range (IQR) and whiskers show dispersion from the IQR that is equal to either the 1st or 3rd quartiles minus/plus 1.5xIQR or the distance from the 1st or 3rd quartiles to the minimum or maximum points, respectively. $26 < n < 55$ for 293 wild-type, $20 < n < 84$ for 293 mutant, $51 < n < 102$ for HCT116 wild-type, and $32 < n < 86$ for HCT116 mutant, where n represents the number of unique cells quantified; unpaired two-sided Student′s t-test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$. Receiver Operating Characteristic (ROC) curves representing biosensing results within CRISPR-Cas edited lines are shown. For 293 cells, the ROC represents sgRNA pair gMis1-g3 and 0.1 fmol probe delivered and for HCT116 cells, the ROC represents sgRNA pair g259-g248 and 0.1 fmol probe delivered. False positives were determined using signals in homozygous mutant lines.

SUPPLEMENTARY INFORMATION for Imaging Unique DNA Sequences in Individual Cells Using a CRISPR-Cas9-Based, Split Luciferase Biosensor

Fig. S1: Optimization of plasmid-based delivery

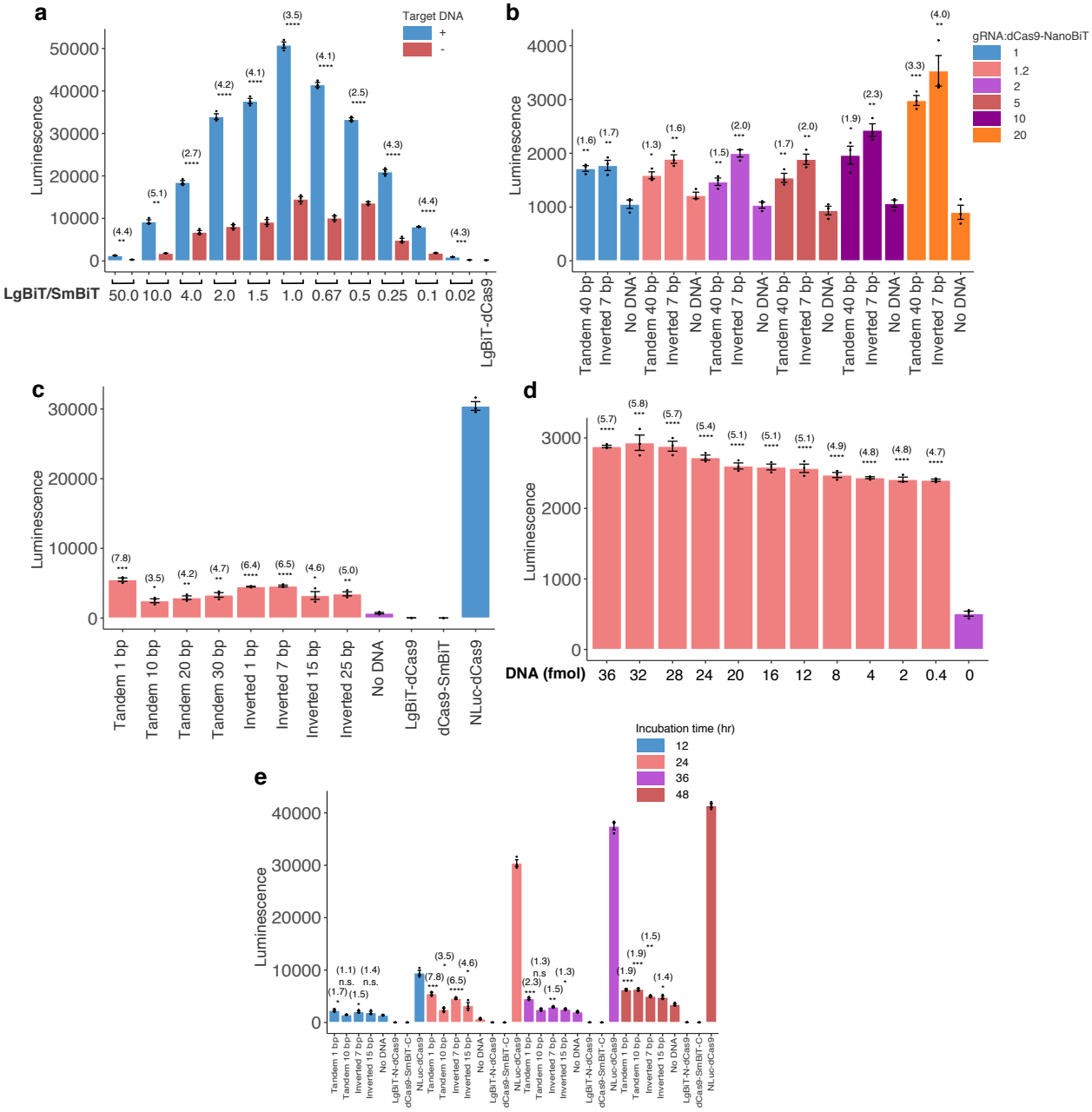


Fig. S1: (a) To determine the optimal ratio of LgBiT:SmBiT in terms of signal production, plasmids expressing LgBiT-dCas9 and dCas9-SmBiT fusion proteins were co-transfected in ratios ranging between 50:1 and 1:50 with or without target DNA plasmids. For simplicity, this initial experiment used only tandemly oriented target sites with 10-bp spacers, as this design was expected to bring the luciferase subunits into close proximity based on the idea that spacers that are a multiple of the helical turn length should foster target site alignment on the double helix. Relative NLuc signal intensity is shown across indicated molar transfection ratios of LgBiT-dCas9

to dCas9-SmBiT with (blue bars) or without (red bars) DNA target plasmids in HEK 293T cells. **(b)** Signal intensities of tandem 40-bp and inverted 7-bp DNA targets compared to no DNA controls over 1:1, 1:1.2, 1:2, 1:5, 1:10, and 1:20 fusion protein:gRNA molar transfection ratios. **(c)** Relative signal intensities using targets of indicated spacing and orientation. gRNAs plasmids were transfected at 20-fold molar excess to dCas9-NanoBiT fusion constructs. **(d)** The dependence of target plasmid concentration was assayed using fixed ratios of the dCas9-NanoBiT and gRNA plasmids. **(e)** The dependence of incubation time post-transfection was assayed using fixed ratios of all plasmids in the indicated configurations. Apparent signal-to-background ratios in **a-e** (comparisons made to no DNA background conditions) are listed in parentheses above each biosensing condition. Data in **a-e** are presented as the mean ± s.e.m., $n = 3$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. In transfections where the amount of one dCas9-NanoBiT interaction partner was decreased, an equal amount of inert pUC19 DNA was included in the transfection mix.

Fig. S2: Signal-to-background of full reporter probes

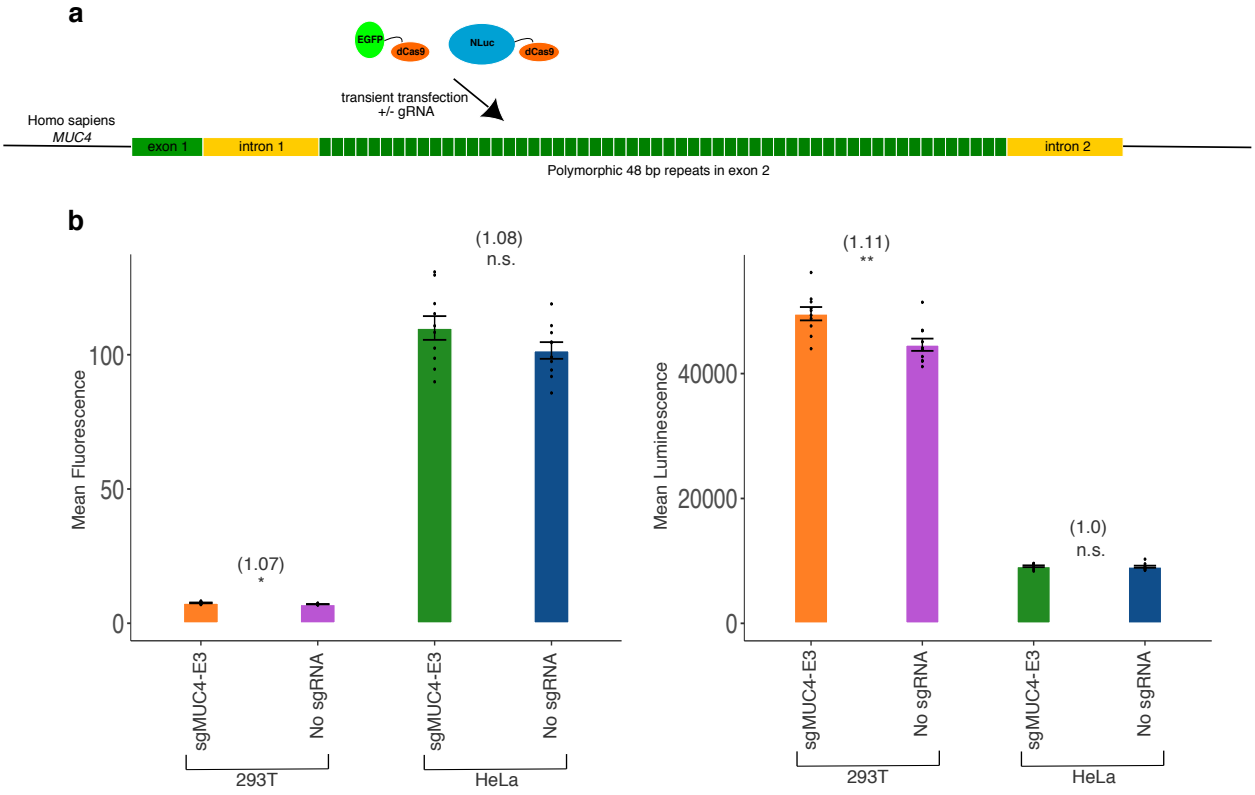


Fig. S2: (a) Cartoon representation of a transient transfection experiment to directly compare signal-to-background of full reporter probes to split reporter probes (background condition without sgRNA transfected as reported in the main text). Experimental conditions are identical to those in Fig. 2 targeting the non-repetitive region of exon 2 within the human *MUC4* locus. **(b)** On-target signal compared to no gRNA background conditions for dCas9-EGFP fluorescent probe shown in two cell lines. **(c)** On-target signal compared to no gRNA background conditions for NLuc-dCas9 luminescent probe shown in two cell lines. Apparent signal-to-background ratios in **a-b** (comparisons made to no sgRNA background conditions) are listed in parentheses. Data in **a-b** are presented as the mean ± s.e.m., $n = 10$, where n represents where n represents the number of unique cells quantified; unpaired two-sided Student's t-test, *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$, n.s. = not significant.

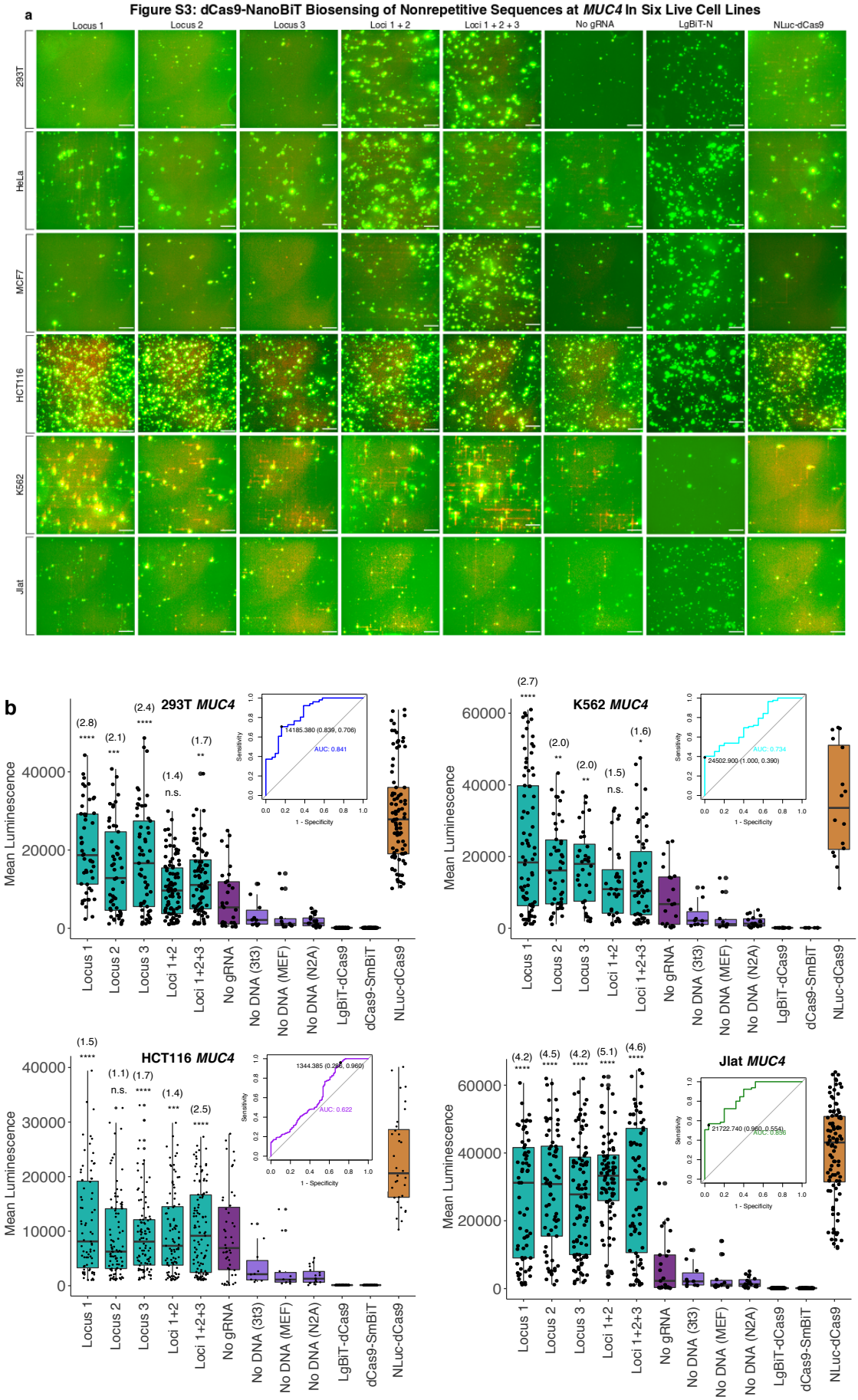


Figure S3: dCas9-NanoBiT Biosensing of Nonrepetitive Sequences at MUC4 In Six Live Cell Lines
a
Locus 1
Locus 2
Locus 3
Loci 1 + 2
Loci 1 + 2 + 3
No gRNA
LgBiT-N
NLuc-dCas9
293T
HeLa
MCF7
HCT116
K562
Jlat
b
293T MUC4
K562 MUC4
HCT116 MUC4
Jlat MUC4
Mean Luminescence
Sensitivity
1 - Specificity
AUC: 0.841
AUC: 0.734
AUC: 0.622
Locus 1
Locus 2
Locus 3
Loci 1+2
Loci 1+2+3
No gRNA
No DNA (3t3)
No DNA (MEF)
No DNA (N2A)
LgBiT-dCas9
dCas9-SmBiT
NLuc-dCas9

Fig. S3: dCas9-NanoBiT biosensing images of three loci individually and in combinations of two and three within the nonrepetitive region of intron 1 of the *MUC4* gene in six human cell lines at 10 fmol probe transfected. The two suspension lines (Jlat and K562) were electroporated with the Invitrogen Neon Transfection System and the other four adherent lines were transfected with Lipofectamine 3000. **(a)** Panel of images representing merged GFP and NLuc channels at 10X magnification taken on the Leica DM6000 B upright microscope. Scale bars=200 μM. **(b)** Signal quantification of above images for four cell lines not displayed in the main text. Apparent signal-to-background ratios (comparisons made to no sgRNA background conditions) are listed in parentheses above each biosensing condition. Error bars represent s.e.m., with $31 < n < 94$ for 293T cells, $112 < n < 439$ for HCT116 cells, $8 < n < 82$ for K562 cells, and $25 < n < 106$ for Jlat cells, where n represents the number of unique cells quantified; unpaired two-sided Student′s t-test, $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$. Receiver Operating Characteristic (ROC) curves representing biosensing results at locus 1 within *MUC4* in all four additional cell lines are shown. False positives were determined by signals due to auto-assembly (no sgRNA). The signal threshold for distinguishing true positives from false positives that maximized Youden's J Statistic (sensitivity + specificity – 1) is shown as a point on the ROC curve along with corresponding specificity and sensitivity values for this threshold in parentheses.

Fig. S4: dCas9-NanoBiT Biosensing of a Single Nonrepetitive Locus at *MUC4* at Reduced Probe Concentrations in MCF7 Cells

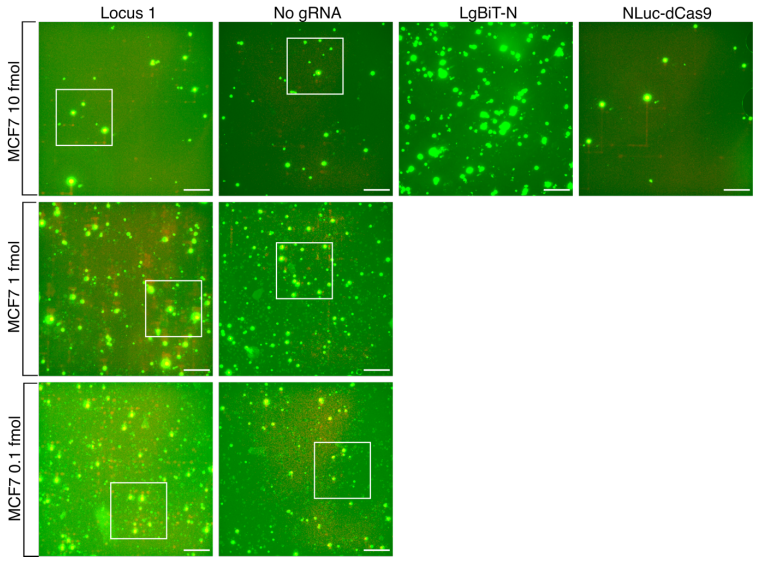


Fig. S4: Panel of images depicting merged GFP (fluorescence) and NLuc (luminescence) channels at 10X magnification taken on the Leica DM6000 B upright microscope. Images represent dCas9-NanoBiT biosensing of a single locus within the nonrepetitive region of intron 1 of the *MUC4* gene in MCF7 cells at all three amounts tested for transfection (10 fmol, 1 fmol, and 0.1 fmol). Controls with no gRNA transfected, LgBiT-dCas9 alone transfected, and NLuc-dCas9 probe transfected are shown for comparison. Scale bars=200 µM. Boxes are shown in the above images which depict approximately 437 µM x 437 µM square sections and represent the image sections shown in Fig. 3b.

Fig. S5: dCas9-NanoBiT Biosensing of a Single Nonrepetitive Locus at *MUC4* at Reduced Probe Concentrations in HeLa Cells

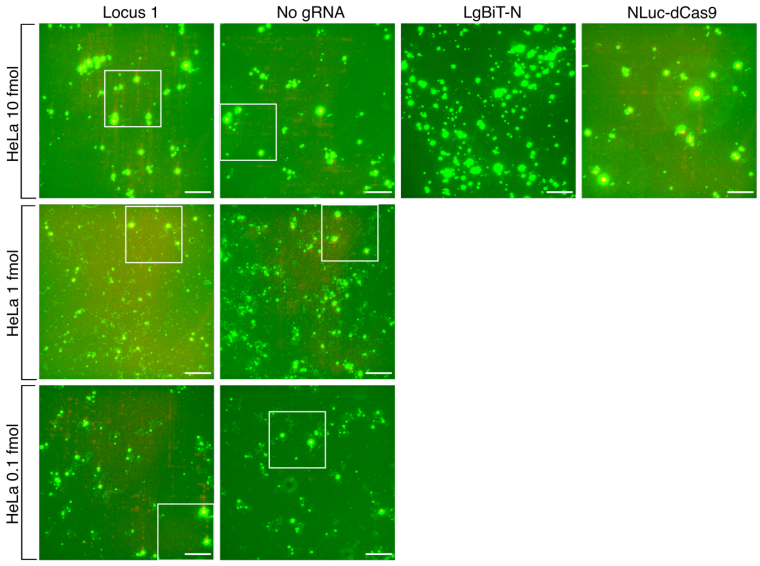


Fig. S5: Panel of images depicting merged GFP (fluorescence) and NLuc (luminescence) channels at 10X magnification taken on the Leica DM6000 B upright microscope. Images represent dCas9-NanoBiT biosensing of a single locus within the nonrepetitive region of intron 1 of the *MUC4* gene in HeLa cells at all three amounts tested for transfection (10 fmol, 1 fmol, and 0.1 fmol). Controls with no gRNA transfected, LgBiT-dCas9 alone transfected, and NLuc-dCas9 probe transfected are shown for comparison. Scale bars=200 µM. Boxes are shown in the above images which depict approximately 437 µM x 437 µM square sections and represent the image sections shown in Fig. 3b.

Fig. S6: dCas9-NanoBiT Biosensing of CRISPR-Induced SNP at *PALB2* Locus in HEK 293 Cells

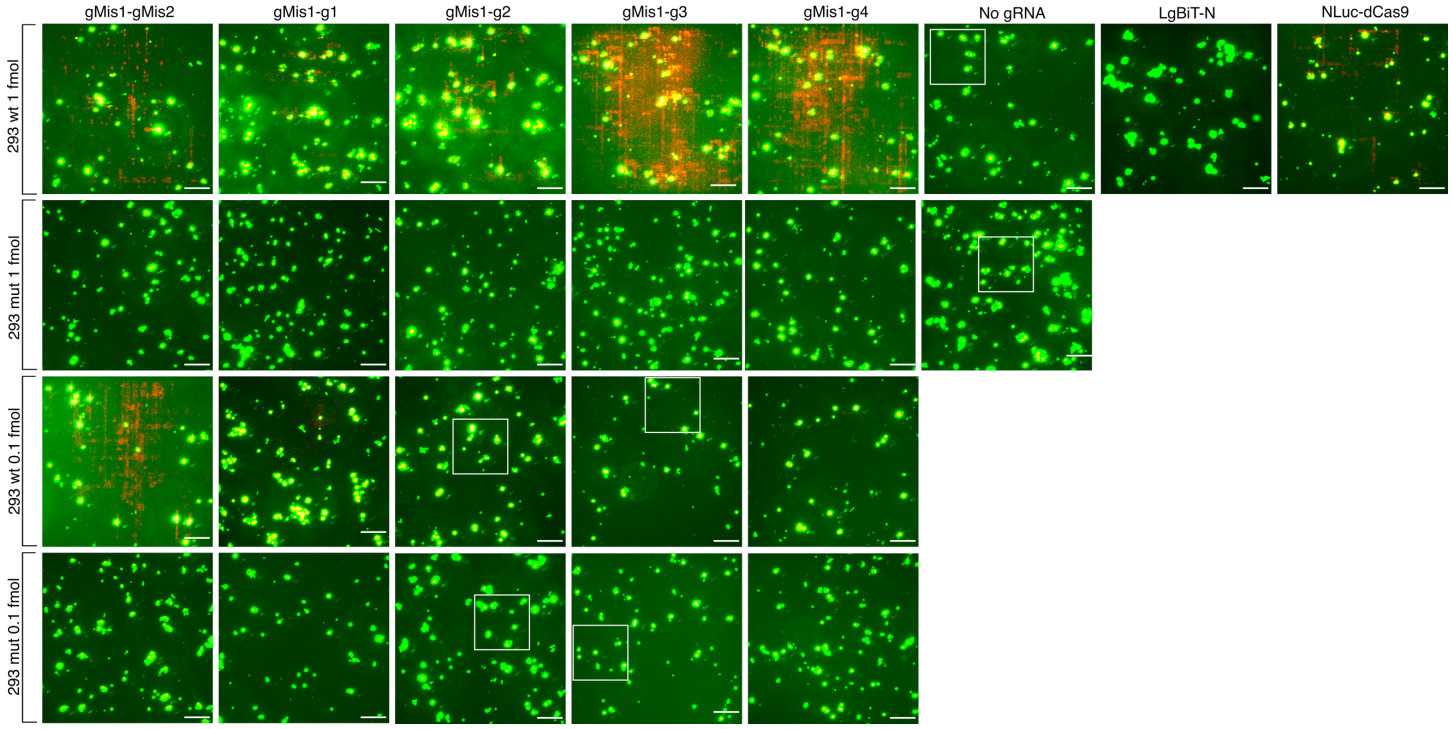


Fig. S6: Panel of images representing merged GFP (fluorescence) and NLuc (luminescence) channels at 10X magnification taken on the Leica DM6000 B upright microscope. Two reduced amounts (1 and 0.1 fmol) of the dCas9-NanoBiT probe were applied to the *PALB2* locus after targeted CRISPR-Cas9 genome editing. Wild-type HEK 293 cells expressing the LgBiT-dCas9 and dCas9-SmBiT constructs and several gRNA pairs are compared to HEK 293 cells homozygous for a G->T missense mutation within an SpCas9 PAM site at the *PALB2* locus expressing the same probe components. Controls with no gRNA transfected, LgBiT-dCas9 alone transfected, and NLuc-dCas9 probe transfected are shown for comparison. Boxes are shown in the above images which depict approximately 437 µM x 437 µM square sections and represent the image sections shown in Fig. 4b. Scale bars=200 µM.

Fig. S7: dCas9-NanoBiT Biosensing of CRISPR-Induced SNP at 8q24 Poly-Cancer Risk Locus in HCT116 Cells

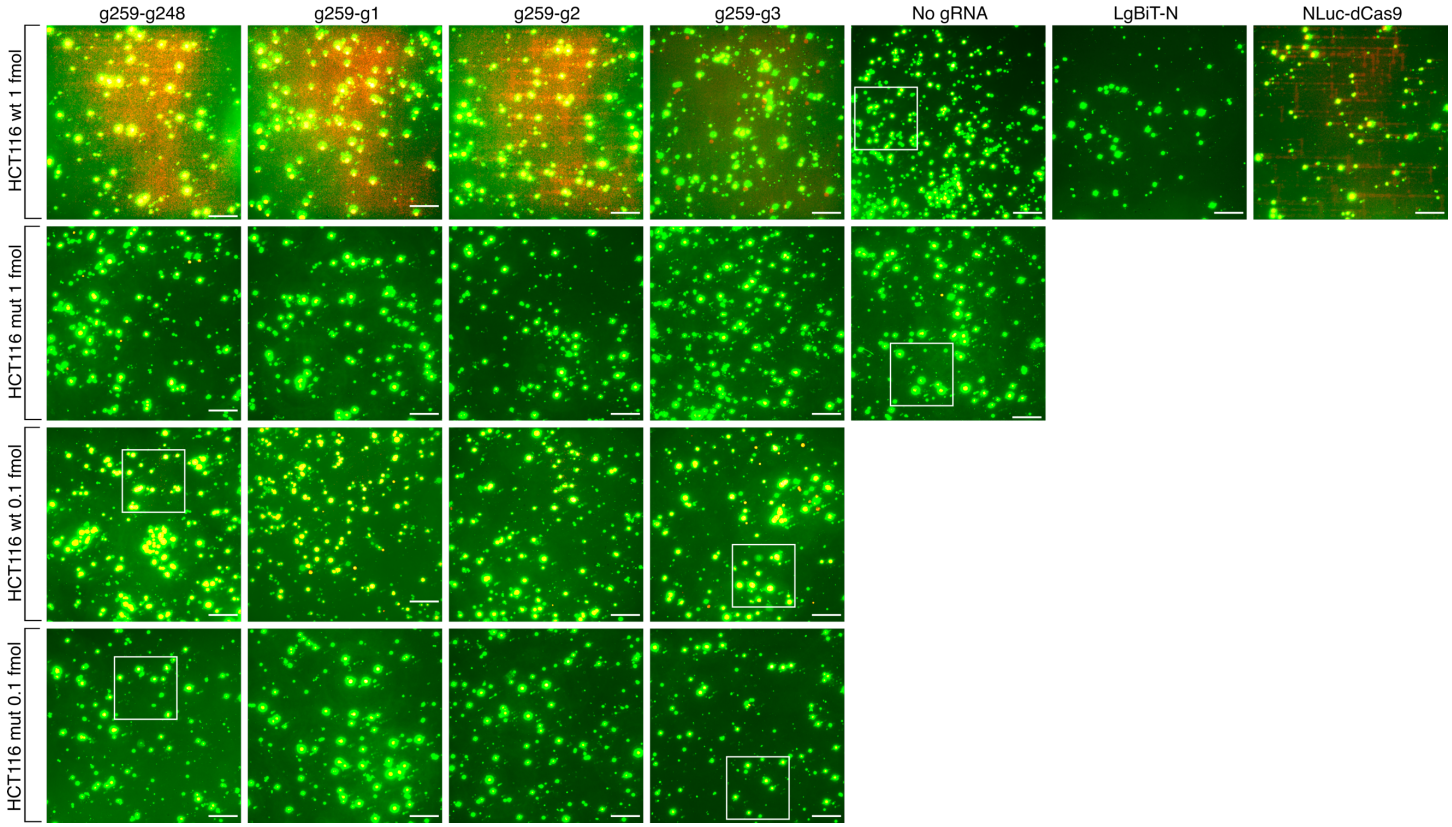


Fig. S7: Panel of images representing merged GFP (fluorescence) and NLuc (luminescence) channels at 10X magnification taken on the Leica DM6000 B upright microscope. Two reduced amounts (1 and 0.1 fmol) of the dCas9-NanoBiT probe were applied to the 8q24 poly-cancer risk locus after targeted CRISPR-Cas9 genome editing. Wild-type HCT116 cells expressing the LgBiT-dCas9 and dCas9-SmBiT constructs and several gRNA pairs are compared to HCT116 cells homozygous for a G->T missense mutation within an SpCas9 PAM site at the 8q24 poly-cancer risk locus expressing the same probe components. Controls with no gRNA transfected, LgBiT-dCas9 alone transfected, and NLuc-dCas9 probe transfected are shown for comparison. Boxes are shown in the above images which depict approximately 437 µM x 437 µM square sections and represent the image sections shown in Fig. 4c. Scale bars=200 µM.

Fig. S8: dCas9-NanoBiT Biosensing at Nonrepetitive *MUC4*

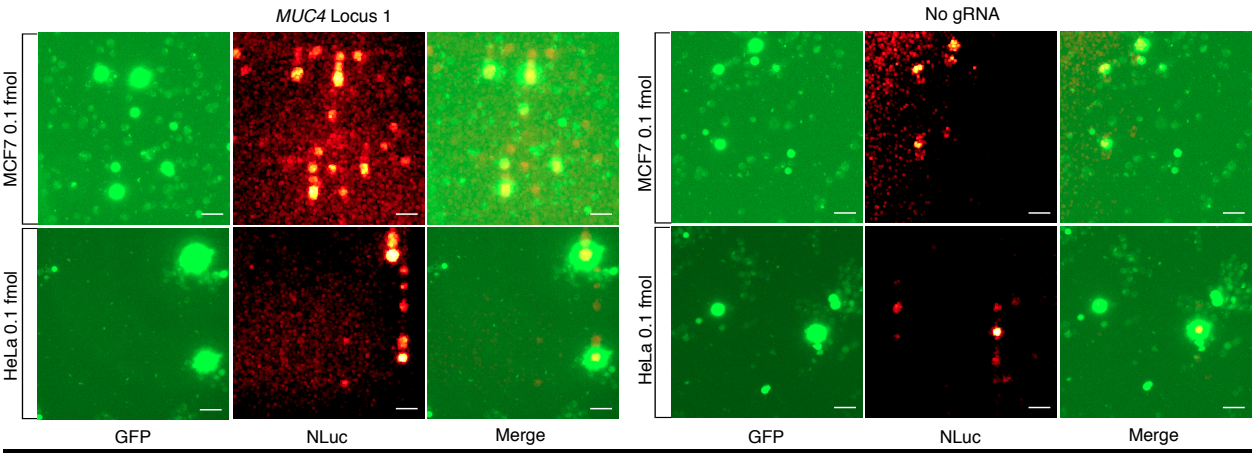


Fig. S8: Panels of images representing individual GFP (fluorescence) and NLuc (luminescence) image channels and corresponding merged images of approximately 437 µM x 437 µM square sections at 10X magnification taken on the Leica DM6000 B upright microscope. Panels depict biosensing images of a single pair of sgRNAs directed to bind at locus 1 of the nonrepetitive region of *MUC4* intron 1 compared to no sgRNA controls for HeLa and MCF7 cells at 0.1 fmol transfected. Scale bars=50 µM.

SUPPLEMENTARY TABLES

Table S1:

Two-way ANOVA	Df	Sum Sq	F value	Pr(>F)
Fusion Protein Orientation	3	28557	302.92	<2e-16
Target DNA Orientation	32	1498	1.49	0.0494
FP Orientation:DNA Orientation	96	6227	2.06	2.94e-06
Residuals	264	8296		

Key: Df: degrees of freedom for each variable**; Sum sq:** sum of squares**; F value:** the test statistic from the F-test**; Pr(>F):** p-value of the F statistic

Table S2:

Tukey HSD	diff	lwr	upr	p adj
LC+SN-LC+SC	-0.9021779	-2.9621672	1.15781142	0.66988987
LN+SC-LC+SC	16.3056095	14.2456202	18.3655988	7.57E-14
LN+SN-LC+SC	-6.4615939	-8.5215832	-4.4016046	2.33E-13
LN+SC-LC+SN	17.2077874	15.1477981	19.2677767	7.57E-14
LN+SN-LC+SN	-5.559416	-7.6194053	-3.4994267	1.45E-10
LN+SN-LN+SC	-22.767203	-24.827193	-20.707214	7.57E-14

Key: LN: LgBiT-dCas9 fusion protein**; LC:** dCas9-LgBiT fusion protein**; SN:** SmBiT-dCas9 fusion protein**; SC:** dCas9-SmBiT fusion protein**; diff:** difference between means of the two groups**; lwr**, **upr:** the lower and the upper end point of the confidence interval at 95% (default)**; p adj:** p-value after adjustment for the multiple comparisons

Experiment 1: Process for Creation of dCas9-NanoBiT Fusion Constructs

gBlocks for initial dCas9-NanoBiT cloning scheme

NLS-HA-LgBiT (Nfus):

TCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGAGGATCGAACCCTTGCCACCATG
CCCAAGAAGAAGAGGAAGGTGGGAGGCTCCGGAGGAAGCTACCCATACGATGTCCCAGACTACG
CGGGTGGCGGGTCCGGCGGTGGATCCATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGGAA
CAGACAGCCGCCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGCTGCAGAA
TCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGTCCGGAGCGGTGAAAATGCCCTGAAGATCG
ACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGACCAAATGGCCCAGATCGAAGAGGTG
TTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCCCTATGGCACACTGGTA
ATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTT
CGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGC
CTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAGTGGTACCGGAGGGAGTGG
TGGAAGCGGCGGTTCTGGTGGCTCAG

NLS-HA-SmBiT (Nfus):

TCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGAGGATCGAACCCTTGCCACCATG
CCCAAGAAGAAGAGGAAGGTGGGAGGCTCCGGAGGAAGCTACCCATACGATGTCCCAGACTACG
CGGGTGGCGGGTCCGGCGGTGGATCCATGGTGACCGGCTACCGGCTGTTCGAGGAGATTCTCGGT
ACCGGAGGGAGTGGTGGAAGCGGCGGTTCTGGTGGCTCAG

LgBiT-NLS (Cfus):

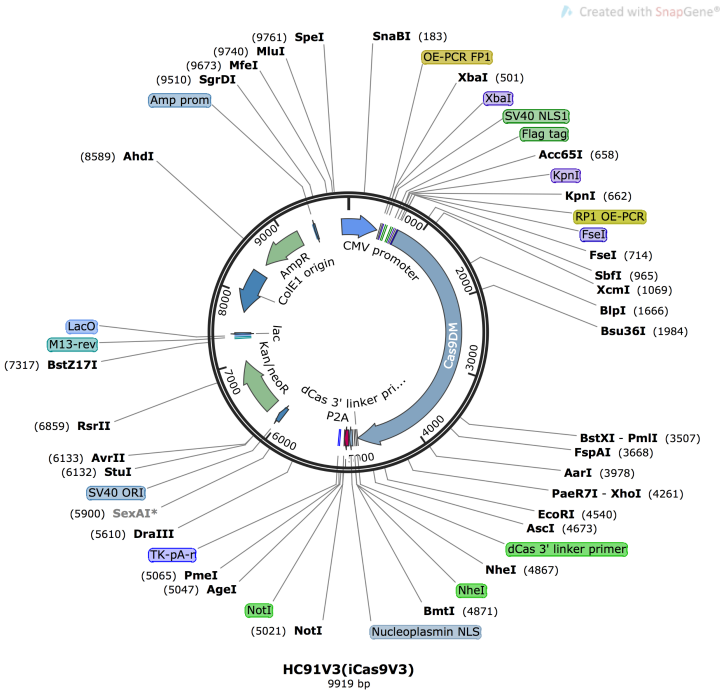
TAGTGGAGGTTCAGGAGGATCCGGGGGGAGCGGAGGGAGCGCTAGCGTCTTCACACTCGAAGATT
TCGTTGGGGACTGGGAACAGACAGCCGCCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTG
TCCAGTTTGCTGCAGAATCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGTCCGGAGCGGTGAA
AATGCCCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGACCAAATGGC
CCAGATCGAAGAGGTGTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCC
CTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGTATG
AAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACAA

AATTATCGACGAGCGCCTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAGCG
GTGGAGGCTCCGGAGGTGGATCTAAAAGGCCGGCGGCCACGAAAAAGGCCGGTCAGGCAAAAAA
GAAAAAGGGTGGTAGTGGAAGCGGAGCGGCCGCATGAAAGGGTTCGATCCCTACCGGTTAGTAAT
GAGT

SmBiT-NLS (Cfus):

TAGTGGAGGTTCAGGAGGATCCGGGGGGAGCGGAGGGAGCGCTAGCGTGACCGGCTACCGGCTG
TTCGAGGAGATTCTGGGTGGAGGCTCCGGAGGTGGATCTAAAAGGCCGGCGGCCACGAAAAAGG
CCGGTCAGGCAAAAAAGAAAAAGGGTGGTAGTGGAAGCGGAGCGGCCGCATGAAAGGGTTCGAT
CCCTACCGGTTAGTAATGAGT

HC91V3 (iCas9V3) vector map:



Overlap Extension PCR Primers to create NLuc-dCas9 Fusion Construct:

FP 1 (LgBiT-N gBlock): 5'-TCCATAGAAGACACCGGGAC

RP 1 (LgBiT-N gBlock w/ 5' homology to SmBiT-N gBlock): 5'-
CGAACAGCCGGTAGCCGGTCACACTGTTGATGGTTACTCGGAAC

FP 2 (SmBiT-N gBlock w/ 5' homology to LgBiT-N gBlock): 5'-
GTTCCGAGTAACCATCAACAGTGTGACCGGCTACCGGCTGTTCG

RP 2 (SmBiT-N gBlock): 5'-CTGAGCCACCAGAACCGCCGC

Final verified protein sequences:

NLuc-dCas9:

MPKKKRKVGGSGGSYPYDVPDYAGGGSGGGSMVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQ
NLAVSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILPYGTLVIDGVT
PNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLITPDGSMLFRVTINSVTGYRLFEEILGTGGS
GGSGGSGGSGGSGRPMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGDGGSGGSGGSGGSGGSASGGGSGGGSKRPAATKKAGQAKKKKGGSGSGATNFSLLKQAGD
VEENPGPAAA*

KEY: SV40 NLS, HA epitope, dCas9 (D10A H840A), NLuc, Nucleoplasmin NLS, P2A, variable length flexible linkers

LgBiT-dCas9:

MPKKKRKVGGSGGSYPYDVPDYAGGGSGGGSMVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQ
NLAVSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILPYGTLVIDGVT
PNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLITPDGSMLFRVTINSGTGGSGGSGGSGGSG
GSGRPMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP
TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINA
SGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG
FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY
VDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK
KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSG
GSGGSGGSGGSASGGGSGGGSKRPAATKKAGQAKKKKGGSGSGATNFSLLKQAGDVEENPGPAAA*

KEY: SV40 NLS, HA epitope, dCas9 (D10A H840A), LgBiT, Nucleoplasmin NLS, P2A, variable length flexible linkers

SmBiT-dCas9:

MPKKKRKVGGSGGSYPYDVPDYAGGGSGGGSMVTGYRLFEEILGTGGSGGSGGSGGSGGSGGSGRPMDK
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY
TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKL
VDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE
DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN
RLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG
GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII
HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGS
GGSASGGGSGGGSKRPAATKKAGQAKKKKGGSGSGATNFSLLKQAGDVEENPGPAAA*

KEY: SV40 NLS, HA epitope, dCas9 (D10A H840A), SmBiT, Nucleoplasmin NLS, P2A, variable length flexible linkers

dCas9-LgBiT:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTGGSGGSGGSGGSGGSGRPM
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR
KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA

KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ
RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSG
GSGGSGGSASVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSGENALKIDIHV
IIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILPYGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKITV
TGTLWNGNKIIDERLITPDGSMLFRVTINSGGGSGGGSKRPAATKKAGQAKKKKGGSGSGAAA*

KEY: SV40 NLS, 3xFlag epitope, dCas9 (D10A H840A), LgBiT, Nucleoplasmin NLS, variable length flexible linkers

dCas9-SmBiT:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTGGSGGSGGSGGSGGSGRPM
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR
KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL

GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ
RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSG
GSGGSGGSASVTGYRLFEEILGGGSGGGSKRPAATKKAGQAKKKKGGSGSGAAA*

KEY: SV40 NLS, 3xFlag epitope, dCas9 (D10A H840A), SmBiT, Nucleoplasmin NLS, variable length flexible linkers

Experiment 2: Process for Creation of DNA Target Site Plasmids

gBlock 1 Sequence with tandem A and B target sites:

GGGTTTGCTGCTCATCTATACTTTCACAATCTTGAGCTGCAGGGCAAAGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGTGAAAAGGAACCCG
GGGATGGAGGAAGGGACATAGGGAGATGGCTCAGGTTTGTCGCGCGGTATGTAGCATGGCCCGGGAAGTACAGTAGAGCTCCCTACGCATGCGTCCCAG
GTGCTACTTACATATTCTCCCGGGTAAATTAATTCTTATGAGATGGCTCAGGTTTGTCGCGCGGCTAGTAGCCCGGGGCATTGTGCTCCCTACGCATGC
GTCCCAGGATCTAATCATATCCCGGGATGAAGGTCTATGATGGCTCAGGTTTGTCGCGCGGTATGCTGAATAATTGAGCCCGGGATAGTGAAATTTATG
ATGCTCCCTACGCATGCGTCCCAGGTGCTTTTCCCGGGTGCACAAGATGGCTCAGGTTTGTCGCGCGGAATATAATAATATGTAGATGGTCCCGGGTAG
GTTGTTATACATTTACTGAGCTCCCTACGCATGCGTCCCAGGTTTGTAGAAGGCTAGGGGAACAGGTTAGTTTGAGGGAATTCTAATGGATCCTTCTAT
GGG

1. **PCR and OE-PCR on gBlock 1 to generate spacers from 6 bp to 50 bp. Red indicates mispriming), Green indicates Target Site A, Blue indicates Target Site B**

6 bp spacer:

FP1: 5′ – CTTGAGCTGCAGGGCAAA (Tm 68) (Taq 58)

RP1: 5′ – CGACAAACCTGAGCCATCTCCCTGCCTGGGACGCATGC (Tm 72) (Taq 63)

FP2: 5′ – GCATGCGTCCCAGGCAGGGAGATGGCTCAGGTTTGTCG (Tm 70) (Taq 61)

RP2: 5′ – TTCCCGGGCCATGCTACA (Tm 72) (Taq 62)

-FP1 and RP1 generate 64bp product, FP2 and RP2 generate 64bp product

-FP1 and RP2 with round 1 products as templates generate 92bp product

Final seq: 5′ –
CTTGAGCTGCAGGGCAAAGAGCTCCCTACGCATGCGTCCCAGGCAGGGAGATGGCTCAGGTTTGTCGCGCGGTATGTAGCATGGCCCGGGAA

10 bp spacer:

FP1: 5′ – CTTGAGCTGCAGGGCAAA (Tm 68) (Taq 58)

RP1: 5′ – CAAACCTGAGCCATCTCCCTCGCTGCCTGGGACGCAT (Tm 74) (Taq 65)

FP2: 5′ – ATGCGTCCCAGGCAGCGAGGGAGATGGCTCAGGTTTG (Tm 68) (Taq 59)

RP2: 5′ – TTCCCGGGCCATGCTACA (Tm 72) (Taq 62)

-FP1 and RP1 generate 66bp product, FP2 and RP2 generate 66bp product

-FP1 and RP2 with round 1 products as templates generate 96bp product

Final seq: 5′ –
CTTGAGCTGCAGGGCAAAGAGCTCCCTACGCATGCGTCCCAGGCAGCGAGGGAGATGGCTCAGGTTTGTCGCGCGGTATGTAGCATGGCCCGGGAA

15 bp spacer:

FP1: 5′ – CTTGAGCTGCAGGGCAAA (Tm 68) (Taq 58)

RP1: 5′ – CAAACCTGAGCCATCTCCCTATACACGCTGCCTGGGACG (Tm 72) (Taq 65)

FP2: 5′ – CGTCCCAGGCAGCGTGTATAGGGAGATGGCTCAGGTTTG (Tm 67) (Taq 59)

RP2: 5′ – TTCCCGGGCCATGCTACA (Tm 72) (Taq 62)

-FP1 and RP1 generate 69bp product, FP2 and RP2 generate 68bp product

-FP1 and RP2 with round 1 products as templates generate 101bp product

Final seq: 5′ –
CTTGAGCTGCAGGGCAAAGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGGGAGATGGCTCAGGTTTGTCGCGCGGTATGTAGCATGGCCCGGG
AA

20 bp spacer:

FP1: 5′ – GGGATAGTGAAATTTATGAT (Tm 54) (Taq 45)

RP1: 5′ – GGGACCATCTACATATTATTATATT (Tm 57) (Taq 48)

-FP1 and RP1 produce 111bp product

Final seq: 5′ –
GGGATAGTGAAATTTATGATGCTCCCTACGCATGCGTCCCAGGTGCTTTTCCCGGGTGCACAAGATGGCTCAGGTTTGTCGCGCGGAATATAATAATAT
GTAGATGGTCCC

25 bp spacer:

FP1: 5′ – CTTGAGCTGCAGGGCAAA (Tm 68) (Taq 58)

RP1: 5′ – GAGCCATCTCCCTATGTCCCTCTATACACGCTGCCTGGG (Tm 64) (Taq 58)

FP2: 5′ – CCCAGGCAGCGTGTATAGAGGGACATAGGGAGATGGCTC (Tm 68) (Taq 61)

RP2: 5′ – TTCCCGGGCCATGCTACA (Tm 72) (Taq 62)

-FP1 and RP1 generate 73bp product, FP2 and RP2 generate 74bp product

-FP1 and RP2 with round 1 products as templates generate 111bp product

Final seq: 5′ –
CTTGAGCTGCAGGGCAAAGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGAGGGACATAGGGAGATGGCTCAGGTTTGTCGCGCGGTATGTAGC
ATGGCCCGGGAA

30 bp spacer:

FP1: 5′ – CTAGTAGCCCGGGGCATT (Tm 66) (Taq 60)

RP1: 5′ – GGGCTCAATTATTCAGCATA (Tm 61) (Taq 51)

-FP1 and RP1 produce 116bp product

Final seq: 5′ –
CTAGTAGCCCGGGGCATTGTGCTCCCTACGCATGCGTCCCAGGATCTAATCATATCCCGGGATGAAGGTCTATGATGGCTCAGGTTTGTCGCGCGGTAT
GCTGAATAATTGAGCCC

35 bp spacer:

FP1: 5′ – CTTGAGCTGCAGGGCAAA (Tm 68) (Taq 58)

RP1: 5′ – GCCATCTCCCTATGTCCCTTCCTTTTTCACTATACACGCTGCCTG (Tm 66) (Taq 57)

FP2: 5′ – CAGGCAGCGTGTATAGTGAAAAAGGAAGGGACATAGGGAGATGGC (Tm 71) (Taq 63)

RP2: 5′ – TTCCCGGGCCATGCTACA (Tm 72) (Taq 62)

-FP1 and RP1 generate 78bp product, FP2 and RP2 generate 73bp product

-FP1 and RP2 with round 1 products as templates generate 121bp product

Final seq: 5′ –
CTTGAGCTGCAGGGCAAAGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGTGAAAAAGGAAGGGACATAGGGAGATGGCTCAGGTTTGTCGCGC
GGTATGTAGCATGGCCCGGGAA

40 bp spacer:

FP1: 5′ – GGCCCGGGAAGTACAGTAGA (Tm 67) (Taq 61)

RP1: 5′ – ACAATGCCCCGGGCTACTAG (Tm 69) (Taq 62)

-FP1 and RP1 produce 126bp product

Final seq: 5′ –
GGCCCGGGAAGTACAGTAGAGCTCCCTACGCATGCGTCCCAGGTGCTACTTACATATTCTCCCGGGTAAATTAATTCTTATGAGATGGCTCAGGTTTGT
CGCGCGGCTAGTAGCCCGGGGCATTGT

50 bp spacer:

FP1: 5′ – CTTGAGCTGCAGGGCAAA (Tm 68) (Taq 58)

RP1: 5′ – TTCCCGGGCCATGCTACA (Tm 72) (Taq 62)

-FP1 and RP1 produce 136bp product

Final seq: 5′ – CTTGAGCTGCAGGGCAAAGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGTGAAAAGGAACCCGGGGATGGAGGA
AGGGACATAGGGAGATGGCTCAGGTTTGTCGCGCGGTATGTAGCATGGCCCGGGAA

gBlock 2 with inverted A and B target sites:

GGGTTTGCTGCTCATCTATACTTTCACAATCTTGAGCTGCAGGGCAAAGACTACAATGGGATTAATAAATTGTACTCTAA
AGGATATTGAAAACTTGTGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGTGAAAAGGAACCCGGGGATGGAGGA
AGGGACATAGGGACCGCGCGACAAACCTGAGCCATCTATGTAGCATGGCCCGGGAAGTACAGTAGAGCTCCCTACGCATG
CGTCCCAGGTGCTACTTACATATTCTCCCGGGTAAATTAATTCTTATGACCGCGCGACAAACCTGAGCCATCCTAGTAGC
CCGGGGCATTGTGCTCCCTACGCATGCGTCCCAGGATCTAATCATATCCCGGGATGAAGGTCTATCCGCGCGACAAACCT
GAGCCATCTATGCTGAATAATTGAGCCCGGGATAGTGAAATTTATGATGCTCCCTACGCATGCGTCCCAGGTGCTTTTCC
CGGGTGCACAACCGCGCGACAAACCTGAGCCATCAATATAATAATATGTAGATGGTCCCGGGTAGGTTGTTATACATTTA
CTGAGCTCCCTACGCATGCGTCCCAGGTTTGTAGAAGGCTAGGGGAACAGGTTAGTTTGAGGGAATTCTAATGGATCCTT
CTATGGG

2. **PCR and OE-PCR on gBlock 2 to generate spacers from 6 bp to 50 bp. Red indicates mispriming), Green indicates Target Site A, Blue indicates Target Site B**

6 bp spacer:

FP1: 5′ – ACTCTAAAGGATATTGAAAACTTGTGA (Tm 63) (Taq 53)

RP1: 5′ – AGGTTTGTCGCGCGGTCCCTGCCTGGGACGCATG (Tm 69) (Taq 60)

FP2: 5′ – CATGCGTCCCAGGCAGGGACCGCGCGACAAACCT (Tm 73) (Taq 64)

RP2: 5′ – ACTGTACTTCCCGGGCCA (Tm 68) (Taq 61)

-FP1 and RP1 generate 70bp product, FP2 and RP2 generate 70bp product

-FP1 and RP2 with round 1 products as templates generate 106bp product

Final seq: 5′ –

ACTCTAAAGGATATTGAAAACTTGTGAGCTCCCTACGCATGCGTCCCAGGCAGGGACCGCGCGACAAACCTGAGCCATCTATGTAGCATGGCCCGGGAA
GTACAGT

10 bp spacer:

FP1: 5′ – ACTCTAAAGGATATTGAAAACTTGTGA (Tm 63) (Taq 53)

RP1: 5′ – GTTTGTCGCGCGGTCCCTCGCTGCCTGGGACGCAT (Tm 74) (Taq 65)

FP2: 5′ – ATGCGTCCCAGGCAGCGAGGGACCGCGCGACAAAC (Tm 73) (Taq 64)

RP2: 5′ – ACTGTACTTCCCGGGCCA (Tm 68) (Taq 61)

-FP1 and RP1 generate 73bp product, FP2 and RP2 generate 72bp product

-FP1 and RP2 with round 1 products as templates generate 110bp product

Final seq: 5′ –
ACTCTAAAGGATATTGAAAACTTGTGAGCTCCCTACGCATGCGTCCCAGGCAGCGAGGGACCGCGCGACAAACCTGAGCCATCTATGTAGCATGGCCCG
GGAAGTACAGT

15 bp spacer:

FP1: 5′ – ACTCTAAAGGATATTGAAAACTTGTGA (Tm 63) (Taq 53)

RP1: 5′ – GTTTGTCGCGCGGTCCCTATACACGCTGCCTGGGAC (Tm 66) (Taq 60)

FP2: 5′ – GTCCCAGGCAGCGTGTATAGGGACCGCGCGACAAAC (Tm 73) (Taq 64)

RP2: 5′ – ACTGTACTTCCCGGGCCA (Tm 68) (Taq 61)

-FP1 and RP1 generate 78bp product, FP2 and RP2 generate 73bp product

-FP1 and RP2 with round 1 products as templates generate 115bp product

Final seq: 5′ –
ACTCTAAAGGATATTGAAAACTTGTGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGGGACCGCGCGACAAACCTGAGCCATCTATGTAGCATG
GCCCGGGAAGTACAGT

20 bp spacer:

FP1: 5′ – GGGATAGTGAAATTTATGAT (Tm 54) (Taq 45)

RP1: 5′ – GGGACCATCTACATATTATTATATT (Tm 57) (Taq 48)

-FP1 and RP1 produce 111bp product

Final seq: 5′ –
GGGATAGTGAAATTTATGATGCTCCCTACGCATGCGTCCCAGGTGCTTTTCCCGGGTGCACAACCGCGCGACAAACCTGAGCCATCAATATAATAATAT
GTAGATGGTCCC

25 bp spacer:

FP1: 5′ – ACTCTAAAGGATATTGAAAACTTGTGA (Tm 63) (Taq 53)

RP1: 5′ – GCGGTCCCTATGTCCCTCTATACACGCTGCCTGG (Tm 60) (Taq 55)

FP2: 5′ – CCAGGCAGCGTGTATAGAGGGACATAGGGACCGC (Tm 65) (Taq 59)

RP2: 5′ – ACTGTACTTCCCGGGCCA (Tm 68) (Taq 61)

-FP1 and RP1 generate 79bp product, FP2 and RP2 generate 80bp product

-FP1 and RP2 with round 1 products as templates generate 125bp product

Final seq: 5′ –
ACTCTAAAGGATATTGAAAACTTGTGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGAGGGACATAGGGACCGCGCGACAAACCTGAGCCATCT
ATGTAGCATGGCCCGGGAAGTACAGT

30 bp spacer:

FP1: 5′ – CTAGTAGCCCGGGGCATT (Tm 66) (Taq 60)

RP1: 5′ – GGGCTCAATTATTCAGCATA (Tm 61) (Taq 51)

-FP1 and RP1 produce 116bp product

Final seq: 5′ –
CTAGTAGCCCGGGGCATTGTGCTCCCTACGCATGCGTCCCAGGATCTAATCATATCCCGGGATGAAGGTCTATCCGCGCGACAAACCTGAGCCATCTAT
GCTGAATAATTGAGCCC

35 bp spacer:

FP1: 5′ – ACTCTAAAGGATATTGAAAACTTGTGA (Tm 63) (Taq 53)

RP1: 5′ – GGTCCCTATGTCCCTTCCTTTTTCACTATACACGCTGCCT (Tm 63) (Taq 56)

FP2: 5′ – AGGCAGCGTGTATAGTGAAAAAGGAAGGGACATAGGGACC (Tm 64) (Taq 58)

RP2: 5′ – ACTGTACTTCCCGGGCCA (Tm 68) (Taq 61)

-FP1 and RP1 generate 87bp product, FP2 and RP2 generate 88bp product

-FP1 and RP2 with round 1 products as templates generate 135bp product

Final seq: 5′ –
ACTCTAAAGGATATTGAAAACTTGTGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGTGAAAAAGGAAGGGACATAGGGACCGCGCGACAAACC
TGAGCCATCTATGTAGCATGGCCCGGGAAGTACAGT

40 bp spacer:

FP1: 5′ – GGCCCGGGAAGTACAGTAGA (Tm 67) (Taq 61)

RP1: 5′ – ACAATGCCCCGGGCTACTAG (Tm 69) (Taq 62)

-FP1 and RP1 produce 126bp product

Final seq: 5′ –
GGCCCGGGAAGTACAGTAGAGCTCCCTACGCATGCGTCCCAGGTGCTACTTACATATTCTCCCGGGTAAATTAATTCTTATGACCGCGCGACAAACCTG
AGCCATCCTAGTAGCCCGGGGCATTGT

50 bp spacer:

FP1: 5′ – ACTCTAAAGGATATTGAAAACTTGTGA (Tm 63) (Taq 53)

RP1: 5′ – ACTGTACTTCCCGGGCCA (Tm 68) (Taq 61)

-FP1 and RP1 produce 150bp product

Final seq: 5′ - ACTCTAAAGGATATTGAAAACTTGTGAGCTCCCTACGCATGCGTCCCAGGCAGCGTGTATAGTGAAAAGGAACCCGGGGA
TGGAGGAAGGGACATAGGGACCGCGCGACAAACCTGAGCCATCTATGTAGCATGGCCCGGGAAGTACAGT

Everted A & B Sites — Target B Rev followed by Target A Fwd

3. **PCR and OE-PCR on gBlock 2 to generate spacers from 6 bp to 50 bp. Red indicates mispriming), Green indicates Target Site A, Blue indicates Target Site B**

6 bp spacer:

FP1: 5′ - CCGGGTAAATTAATTCTTATGA (Tm 60) (Taq 49)

RP1: 5′ - GCATGCGTAGGGAGCACATAGGATGGCTCAGGTTTG (Tm 61) (Taq 53)

FP2: 5′ - CAAACCTGAGCCATCCTATGTGCTCCCTACGCATGC (Tm 69) (Taq 61)

RP2: 5′ - ATCTAATCATATCCCGGGATGA (Tm 65) (Taq 54)

-FP1 and RP1 generate 66bp product, FP2 and RP2 generate 66bp product

-FP1 and RP2 with round 1 products as templates generate 96bp product

Final seq: 5′ -

CCGGGTAAATTAATTCTTATGACCGCGCGACAAACCTGAGCCATCCTATGTGCTCCCTACGCATGCGTCCCAGGATCTAATCATATCCCGGGATGA

10 bp spacer:

FP1: 5′ - CCGGGTAAATTAATTCTTATGA (Tm 60) (Taq 49)

RP1: 5′ - ATGCGTAGGGAGCACAATACTAGGATGGCTCAGGTT (Tm 59) (Taq 54)

FP2: 5′ - AACCTGAGCCATCCTAGTATTGTGCTCCCTACGCAT (Tm 63) (Taq 56)

RP2: 5′ - ATCTAATCATATCCCGGGATGA (Tm 65) (Taq 54)

-FP1 and RP1 generate 68bp product, FP2 and RP2 generate 68bp product

-FP1 and RP2 with round 1 products as templates generate 100bp product

Final seq: 5′ -
CCGGGTAAATTAATTCTTATGACCGCGCGACAAACCTGAGCCATCCTAGTATTGTGCTCCCTACGCATGCGTCCCAGGATCTAATCATATCCCGGGATG
A

15 bp spacer:

FP1: 5′ - CCGGGTAAATTAATTCTTATGA (Tm 60) (Taq 49)

RP1: 5′ - CGTAGGGAGCACAATGCCCTACTAGGATGGCTCAGG (Tm 57) (Taq 53)

FP2: 5′ - CCTGAGCCATCCTAGTAGGGCATTGTGCTCCCTACG (Tm 67) (Taq 59)

RP2: 5′ - ATCTAATCATATCCCGGGATGA (Tm 65) (Taq 54)

-FP1 and RP1 generate 70bp product, FP2 and RP2 generate 71bp product

-FP1 and RP2 with round 1 products as templates generate 105bp product

Final seq: 5′ –

CCGGGTAAATTAATTCTTATGACCGCGCGACAAACCTGAGCCATCCTAGTAGGGCATTGTGCTCCCTACGCATGCGTCCCAGGATCTAATCATATCCCG
GGATGA

20 bp spacer:

FP1: 5′ – CCGGGTAAATTAATTCTTATGA (Tm 60) (Taq 49)

RP1: 5′ – ATCTAATCATATCCCGGGATGA (Tm 65) (Taq 54)

-FP1 and RP1 produce 110bp product

Final seq: 5′ –
CCGGGTAAATTAATTCTTATGACCGCGCGACAAACCTGAGCCATCCTAGTAGCCCGGGGCATTGTGCTCCCTACGCATGCGTCCCAGGATCTAATCATA
TCCCGGGATGA

25 bp spacer:

FP1: 5′ – TTCTCCCGGGTAAATTAATTCTTATGA (Tm 67) (Taq 55)

RP1: 5′ – GGGAGCACAATGCCCCGGCCCGGGCTACTAGGATGG (Tm 67) (Taq 60)

FP2: 5′ – CCATCCTAGTAGCCCGGGCCGGGGCATTGTGCTCCC (Tm 76) (Taq 66)

RP2: 5′ – GACCTTCATCCCGGGATATGATTAGAT (Tm 71) (Taq 60)

-FP1 and RP1 generate 81bp product, FP2 and RP2 generate 80bp product

-FP1 and RP2 with round 1 products as templates generate 125bp product

Final seq: 5′ –
TTCTCCCGGGTAAATTAATTCTTATGACCGCGCGACAAACCTGAGCCATCCTAGTAGCCCGGGCCGGGGCATTGTGCTCCCTACGCATGCGTCCCAGGA
TCTAATCATATCCCGGGATGAAGGTC

30 bp spacer:

FP1: 5′ – GATGGAGGAAGGGACATAGG (Tm 65) (Taq 57)

RP1: 5′ – CCGGGAGAATATGTAAGTAGCA (Tm 64) (Taq 56)

-FP1 and RP1 produce 120bp product

Final seq: 5′ –
GATGGAGGAAGGGACATAGGGACCGCGCGACAAACCTGAGCCATCTATGTAGCATGGCCCGGGAAGTACAGTAGAGCTCCCTACGCATGCGTCCCAGGT
GCTACTTACATATTCTCCCGG

35 bp spacer:

FP1: 5′ – GATGGAGGAAGGGACATAGG (Tm 65) (Taq 57)

RP1: 5′ – AGCTCTACTGTACTTCCCGGGCCGGGCCATGCTACATAGAT (Tm 68) (Taq 59)

FP2: 5′ – ATCTATGTAGCATGGCCCGGCCCGGGAAGTACAGTAGAGCT (Tm 66) (Taq 61)

RP2: 5′ – CCGGGAGAATATGTAAGTAGCA (Tm 64) (Taq 56)

-FP1 and RP1 generate 83bp product, FP2 and RP2 generate 83bp product

-FP1 and RP2 with round 1 products as templates generate 125bp product

Final seq: 5′ –
GATGGAGGAAGGGACATAGGGACCGCGCGACAAACCTGAGCCATCTATGTAGCATGGCCCGGCCCGGGAAGTACAGTAGAGCTCCCTACGCATGCGTCC
CAGGTGCTACTTACATATTCTCCCGG

40 bp spacer:

FP1: 5′ – ATCCCGGGATGAAGGTCTAT (Tm 66) (Taq 57)

RP1: 5′ – TTGTGCACCCGGGAAAA (Tm 69) (Taq 57)

-FP1 and RP1 produce 126bp product

Final seq: 5′ –
ATCCCGGGATGAAGGTCTATCCGCGCGACAAACCTGAGCCATCTATGCTGAATAATTGAGCCCGGGATAGTGAAATTTATGATGCTCCCTACGCATGCG
TCCCAGGTGCTTTTCCCGGGTGCACAA

50 bp spacer:

FP1: 5′ – TTTTCCCGGGTGCACAAC (Tm 69) (Taq 59)

RP1: 5′ – GTTCCCCTAGCCTTCTACAAACC (Tm 67) (Taq 60)

-FP1 and RP1 produce 134bp product

Final seq: 5′ – TTTTCCCGGGTGCACAACCGCGCGACAAACCTGAGCCATCAATATAATAATATGTAGATGGTCCCGGGTAGGTTGTTATA
CATTTACTGAGCTCCCTACGCATGCGTCCCAGGTTTGTAGAAGGCTAGGGGAAC

Experiment 3: Process for Creation of gRNAs

JL gRNAs oligos for annealing to create JL1 and JL2 gRNAs

JL gRNA1

gRNA1 = GCTCCCTACGCATGCGTCCC
DNA target site 1/A (fwd) = GCTCCCTACGCATGCGTCCC**AGG** (no 100% match within hg38 and lowest off-targets (CFD score, CRISPRscan))

JL gRNA2

gRNA2 = GATGGCTCAGGTTTGTCGCG
DNA target site 2/B (fwd) = GATGGCTCAGGTTTGTCGCG**CGG** (no 100% match within hg38 and second lowest off-targets (CFD score, CRISPRscan))

Insert_F: TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC***G****NNNNNNNNNNNNN NNNNNNN*

Insert_R: GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*NNNNNNNNNNNNN NNNNNN****C***

JL gRNA1 F
TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC***G****CTCCCTACGCATGCGTCCC*

JL gRNA1 R
GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*GGGACGCATGCGTAGGGAG****C***

JL gRNA2 F
TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC***G****ATGGCTCAGGTTTGTCGCG*

JL gRNA2 R
GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*CGCGACAAACCTGAGCCAT****C***

Experiment 4: Process for Creation of gRNAs for MUC4 DNA Biosensing

Repetitive region in exon 2:

MUC4 repetitive DNA region—48 bp repeat:

5′-GCCACCCCTCTTCCTGTCACCGACACTTCCTCAGCATCCACAGGTCAC~GCC-3′
3′-CGGT***G****GG****G****A**GAA*GG*ACAGTGGCTGTGAAGGAGTCGTAGGTGTCCAGTG~CGG-5′

sgMUC4-E3(F+E): GGCGTGACCTGTGGATGCTGAGG

MUC4 gRNA tgt 1: GACACTTCCTCAGCATCCACAGG

-Everted overlapping, PAMs 10 bp apart
-CFD:110.89

MUC4 gRNA tgt 2: GGTGGATGCTGAGGAAGTGTCGG

-Tandem overlapping, PAMs 6 bp apart
-CFD:163.22

MUC4 gRNA tgt 3: GGTGAGGAAGTGTCGGTGAC*AGG*

-Tandem overlapping, PAMs 13 bp apart
-CFD:70.68

MUC4 gRNA tgt 4: GAAGTGTCGGTGACAGGAAG**A**GG

-Tandem overlapping by 1 bp, PAMs 19 bp apart
-CFD:118.16

MUC4 gRNA tgt 5: GGTGTCGGTGACAGGAAGAG***GG****G*

-Tandem 1 bp
-CFD:122.54

MUC4 gRNA tgt 6: GGCGGTGACAGGAAGAGGGGTGG

-Tandem 4 bp
-CFD:227.72

Selected gRNAs 1-4 for experiments

Non-repetitive region in intron 1:

MUC4 non-repetitive DNA region with Cas9 target sites:

ATGAAGGGGGCACGCTGGAGGAGGGTCCCCTGGGTGTCCCTGAGCTGCCTGTGTCTCTGCCTCCTTCCGCATGT
GGTCCCAGGTAAGTGATGGAGACAGCAGATGAGGCTGGCTGCGGGGAGCACTTGGGGGAGGTGGGAGCTGTCAG
AGAAAGAGGTCCGGGGAGACAGAGAGAGAGAGAGAGAGAATAGGGGAAAGGGAGACAGCGAAGAGGAAGAGAAG
GGAGAGAAAAAGAGGGAGAGGGAAAGGAGAAAGAGATGAATGGGACAACATGGGGGGAAGGTGGAGAGAGACCC
AGAGAGGGAAAGAAGAGGAAGAGAAGAGGGAGAGAGAAAGAAGAGTGGAGGCCGTGCGCGG**TGG**CTCATGCCTG
TAATCCCAGCACTTTCGGAGG**CCA**AGGCAGGAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCGACAT
GGTGAAA**CCC**CGTCTCTACTAAATATACAAAAATTAGCCGGTCGTGGTGGG**CCC**CACCTGTAATTCCAGCTACT
CAGGAGTCTGAGGCAGGAGAATCACTTGAA**CCT**GGGAGGTGGAGGTTGCAGTGAGCCAAGATCGCGCCACTGCA
CTCCAG**CCT**GGGAGAGAGAGCGAGACTCTGTCTCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAA
TAAATATAAAATAAAATAAAAAATAAGAAGAGGAGAAAAGTGGGGAAG**AGG**AGGCATGAACTGGCAGATACGGG
ACAAGATCTGAGGGGAAAGACAGAGGGAGAATGCTCGAAAGAGAGAGAAAAGAGAACAGAGGGCCA*GAGAGCAG*
*C*C***CGG****CGATGTCTGG*AAGGATCCATGGTGAGAG**CCC**AGGCTTACTCGCAGAGAGAAAGACAGGCAGAGCCAGAG
CAAGAGGAACAGAGTCAAGGAGAAAGATGTACACCCTTGTGTACAGAGCT**GGG**GGTAGAGGGGATGCCAGGAAA
GCTGGGTGATGGAGACGGAAGAAAACTCATGTAAAGCTGCA**GGG**TGAGAGGACGAGACAGGTGAGACGCAGACA
AACTGAGGACCCTGGGAATGGAGAGAGGAGAAGATCGGGAGACAGCAGCAAGCAAGGGAAGCGACAAGG**AGG**AG
AGGGGCAGGCCGGCCGGGAGGGTGGTGCGGAGGAGGCGGCCAGGGCGCAGA**GGG**CCGGGAGGTGCTGGCCGTGG
GCTTCTTA**CCT**CTGAGCTCGGGTTTAAAAGCCTCCATTTGGGTCACGGCCTTGCCTGGGGCTCGTAGCCCCGGC
ATTGGCCTT**GGG**CTCCTCCGTGTACAGAGCTGGGAGGGGAGGGATGCCAGGCCTGTGGGAGATGTTCCCTC**GGG**
GGCCCCCGTCCTCTTCCCCACACTTT**CCA**GGCTGTCCCTCTGGCTTCAGGACCAAGTTTTATTCTGTGTTTCTG
GGTGTCTGAGTCTTTGGGGGAGAGTCT**GGG**GTCCAGAGTTCAAGCTGGGGTTAGAGTCTCAGCTCCTGCCCTGC
CTCTCAGC**AGG**CTAAGAACAGTCGCCGAGGGAAAATATTTCTTGGGCGCATATTTGAGGAGCTTCCT**GGG**AGTG
AGTCAGAAGGCGAGTGCCGTTTAAAGGCTGCAAGAGAAGCCATGC**TGG**TGAAGCGGACCCTTCCACCTCGGGAT
GTTTCAGGACTAGGCTGA**GGG**CAAAGGAAACTGCCACCACCT**CCC**TACACCTCCCCACCCTCCAGCACCCCCAC
CCCACCCTGGCCACACAAC**CCC**GCTCCAGTGCTCATCCCACCGTGAGGACGTGGAGGCCGGAAGGAGCCGCCAC

ACGGCCCTGCCCTGCAGATGTGGTTGA**AGG**AGTCTCCACGGGAATCATGACTCCCAGAGCGAGGCTGGGGCTTG

GGGCGCC**GGG**GAGGCAGCTTGGATTTAGGAGCCCCAGGGCCAAGTCTTTGCCGTGAACTGTTC**TGG**CCCCTGTG

ACCAGGCCCTGC**CCC**GTGTCTCCCCAGGGCCCCGGTCCCCTGTGTAAAAAGCAGTGGTGAACGGTTGGACCTCC

TGACGCCCAAGTTCTTGAGTTTCCAAATCTGTGATTTAAAGCTGAGCCCAAATGTGCTGGGTACCAGCTGGACA

CTCAGCTCCATG**TGG**AGCCAGGAAGTGGGGTCTGTGGAGAGGAGCGCAGAGGGGCAAGACCT**GGG**GTGGGCGTG

GAAAAGCACGGGGGCGTGACCCGGAGAAGGAGTGAAGGACTGT**TGG**TGTGCAAGGGCGTCTCCATGACGA**CCC**G

AAGAAGCTAGGCATGTCGTGGAGCGCT*GAGTCCTTTGCGTCGCTAAGGG*GACCAAGTGGAGCTGGGCCAGGAGA

GGAGA**TGG**TCGTGGCTGGGAGATGGCACCCACACATCTGACCGGGCATGACCAGGGCCT**TGG**CAGGAAAAGCAG

TCACCAAGGGCGGGTGGGCAGCCCCCACCCCCACA**GGG**CAGCTGCTGGAGGACTGGCAGCCAGCCAGCCCCGTT

CCTTTTGGCTCCCTGAAG**GGG**TTTACAGATGACCTGCCTATACTTGAGTCTAGGGTCTGTTTGCACACTTGC**CG**

GCAGGACCCTCACCCAGGCTGGGTCACACTGAAGCCCAGGCCAGAGGAAAAACACA**GGG**TTTCCACAAAGGAGC

TGCCGCAATGAGGGTTTCCTTAAGGAACAGCCC**TGG**CTCTCAAGGGTTAAAGGATAAGGCACAGCAGACAGAGG

TGGGCTAGACA**AGG**ACAGATGGAAATTTGGTGTCTACTGGTCGCCCCAGGCAGGAATGACTCAGA**AGG**AAGCCT

GGCCGTCCTGGTTCCATG**CCA**CAGGGAAAGGCAACTGGGTCGAAATAGGCCTTGGTCTCCAGCACTATCAGTGA

CCCCAGGGAGGTGACAGGC**TGG**AGCAAGTGCAGGGCAGGCAGGGGAGGGGACGCCGGCCACAGCGCACTCCACG

GGGAA**GGG**TCTTTATGGGCCCCTCCTCGGAGAACC**CCC**GGTCTATCTGTCAGTCTGGGACAGGCCACCTCAACT

TGCCACCGAGGACA**CCA**AAACTCTCCACAGACCCCTCTGCCCCTCTGGGAAACCCCACTGTGCTCCAGGACACT

CAAAAGGAAAGGATCCCTGGACAAGAGGTCCTGCCAGGAACATCAGCCAAATTTTGGCCAACGACCAGCAAGGT

GCACAGGGAAGAGCAGGGGCTGAAACTCAGAGGTCCAGCATCAGCGACGCCCT**TGG**CAGCCCAGGGAACACAGG

CAACG**CCT**TTTGGCTCTGGAGTCTTAGGCTCTTCATCGGCAAACTGAGCCCAGGGGGAAGGGGCTACTACGTAG

GGTTGTCATG**AGG**ATGAAACGAGACAGCATCTGGTGTAAAGTAGAAAAGGCATAAA**GGG**CCGGGCGCGGTGGCT

CACGCTGTAATCCCAGCACTTTTGGAGGCCCAGGC**GGG**TGGATCACCTGAGGTCAGGAGTTCAAGA**CCA**GCTTG

GCCAACCCTGTCTCCACTAAAAATAAAAAATTTTGCCGGGCGTGGTGGCGAGCGCCTGTAATTCCAGCTACTCG

GGAGGCTGAGGTAGGAGAATGGCTTGAA**CCT**GGGAGGCAGAGGTTGCAGGGAGCCGAAATGGCAGCACTCTAGC

TTGGGTGACAGAGCAAGACTCTGTCTAAAAAAAAAAGAAAAG**CCA**TAAAGACGTGTTTGAGAAAGAGGCCTGGG

AAGACGGGGGAAGGAGGGTGATTGAACCCGGAATGGCACTTGTGT**CGG**CCCAGGGTCATATCCCTTCATCTAAG

GAT**CCT**CGTGCCTCTAAAAAGCCACCCCGTGCTTCCTGTGGGTTTGCAAGGGCTGGCTTGGTGTATTCAGAATG

TGGCTTGCTGCATGAACGGACCCCGAGGGCCATGGCCCTAGAGCAGGGGCTCGCTCCAGCGGACAGCTCTGCCT

CACCGCTCCCTGCCTGTGAGTCCCGCCACGCCCTTGGTTTCTGGGCTCAGCCGTGGAGGCAGAGGCTGG**CCT**GG

CAGAGGCTGGCCTGGCAGTGCTTGACACGCAAGTGATTTGTGTCTTCATTGCTAAGGACAAGAGGCAATGAG**AG**

GACAAGAAGTGGTTGGCCTTTTGTACGCTCAACGGGTGGTTTTGCTACTCTGTGTCTTTTCTCTGATTTCACGG

TGCTGTTAAGTGCTTAAAATATGCACATCGTGTAGCTCACAGAGCCACTTCTCTGAAGGCCAGGACAGAGACCT

TATAGGCTCTCTCTCCCCCTAGTTTCAGCCTTTTACCTTAAATATACGTCTTTCTTACTGCTAGGCTGAGTTCC

CG**CCC**CAGCATGTTCTGAGAAATTGAGTCAAAATAACTGAGTCTGTTGGCACCTCATCGACGATTTCTTCATAG

ACGGTTTTTTTATTGTTGCTGTTGTTGTTGGTTTTTTGGGTTTGTTTGTTTGTTTTTTGAGACAGAGTTTCTCT
CTGTCCCCC**AGG**CTGCAGTGCAGTGGCGTGGTCTCAGCTCAGTGCAGCCTCTGCCTCCCGGGTTCAAGAGATTC
TCCTGCCTCAGCCTCCCGAGTAGCTGGGATTATAGACGCCCAACACCACAGCGGCTAATGTTTGTATTTTTAGT
AGAGATGGGGTTTCACCATGTTGGCC**AGG**CTGGTCTCGAACTCCTGACCTCAGGTGATCCGCTCGCCTCGGCTC
CCAAAGTGC**TGG**GATTATAGGCGTGAGCTACTGTGCCTGGCCCTACTTCATAGAGGTTTAAATGCCTTTTCACC
CTTTTCCTGGAGACTCTGAAGAAGTCTCAGGAACTGGGCATTTGTGTTGCACGTG**AGG**CCTTGCAATGGCGGCC
CTGCTTGGAGGAAGGGCACTGG**CCT**GGGTTGCCCGCAGCTCCACTCCCCGTGTATGTGTTTAGGGACCACAGAG
GACAGACATCGACTCTCTGTAGAGATGCCGCCCCGCCC**AGG**TTGCAGTTTAGGTTCCAAAAGTCCAGTGGCCAG
TGGATTTTG**GGG**GAATTTGGAATAAGAAACAGCCTAGACTTTGGAGTTGTTCATTCACTTGCAGAATTTCTACT
CATGCCAGCTGCTCTGGACAGGAAGATGAATGCGTCACAGTTCCTGCTTTTCAAAGCTCTCTAAGTTAAGTGAC
TTGTTTAAGATCATAGAACCCATAAGTGAGGCAGCTGGGACTAGAACCC**AGG**TCTCCTGACTCACTGCAGCACA
CAGCCTTTCGGCAATCT**CCA**AACCAGCCCAGCCCACCGACGGAGGGAAGAACAGAAGCATTCACACA**CCC**TGCT
GAGACAGCCATTCATTCATTCATTTGTTAATTAAACCACCATTTAGGAAACGCCTGCCTTAAGTTCCTGACATT
GTTCTAGGACACAGCACTGGATGCACACAGTGAAGAGTGAAACAGACGTGGCCCAGTCTCT**TGG**CACTAAAATC
TTGGTGCAGACAGACATCAAATAATTACGGAAATGTTCTCAACTGCACATGTGGTAAATGCAGTGTGGAAAAGT
ACAGGGTGTGCTGAGAGCTGCATTTCGAA**TGG**CCAGAGAGTAGGGGAGGTGCATCTGACTGACAAGTCAGGAAG
GGCCCTGTG**AGG**AACCGTTCTGCGGGGAGCTGAGGCCTGAGGCTGAGGACAGCCAGGTGGAGAAGGTGCCAGGC
CTGAGCAGGCAGAGGCGGAGCTCATGGAGAGGCAGGAAAGAGCTTGGCCCCTTGGAGGACTTGAAAGAGAAGGC
AGG

gRNAs from low to high CFD

gRNA1: 1.62 w/ tandem 10 bp nearby site; tandem overlapping PAMs w/ gRNA4, everted 7 bp with gRNA7

gRNA2: 1.79 w/ tandem overlapping, PAMs 17 bp apart nearby site

gRNA3: 2.94 w/ tandem overlapping, PAMs 15 bp apart nearby site

gRNA4: 3.20 w/ tandem 9 bp nearby site; tandem overlapping PAMs w/ gRNA1, everted 8 bp w/ gRNA7

gRNA5: 3.50 w/ tandem overlapping, PAMs 4 bp apart nearby site

gRNA6: 4.13 w/ everted overlapping, PAMs 15 bp apart nearby site

gRNA7: 4.82; everted 9 bp with gRNA1, everted 8 bp w/ gRNA4

gRNA8: 5.26 w/ tandem 12 bp nearby site

gRNA9: 6.29 w/ tandem overlapping, PAMs 8 bp apart nearby site

gRNA10: 6.55 w/ everted PAMs overlapping nearby site

gRNA11: 6.83 w/ tandem overlapping, PAMs 7 bp apart nearby site; everted overlapping, PAMs 8 bp apart w/ gRNA10

gRNA12: 7.25 w/ tandem 3 bp nearby site

gRNAs 1-3, 5, 9-10, and 12 from this list selected to bind loci 1-7 at non-repetitive *MUC4*

Experiment 5: Process for creation of gRNAs for 8q24 and PALB2 editing and edit biosensing

8q24 risk locus (+) chr8:127,400,950-127,401,200

tctcagctccctatccataaaacagagggacgaataaactctcctcctaccacta

agaggtgtagccagagttaaTACCCTCATCGTCCTTTGAGCTCAGCAGATGAAAGGCACTGAGA

AAAGTACAAAGAATTTTTATGTGCTATTGACTTTATTTTATTTTATGTGGGGGAGGGAGCCGGC

CCCAGCTGGAAAGCTGCTTTCTCTGAATCAAAGGGCAGGAACCCAGCAAGTTTCTCAGGATTGG

GGCC

Editing sgRNA

g259 (G->T edit): CTTTGAGCTCAGCAGATGAAAGG

sgRNAs adjacent to sgRNA used for editing

g248 (inverted overlapping): CTGAGCTCAAAGGACGATGAGGG
8q24gRNA1 (inverted 0 bp): GACGATGAGGGTATTAACTCTGG CFD 5.24
8q24gRNA2 (tandem 28 bp): actctcctcctaccactaagagg
8q24gRNA3 (tandem 41 bp): TATTTTATTTTATGTGGGGGAGG

Palb2 locus (+) chr16:23,624,025-23,624,175

CCAAATTTCCCAAAGCTACACACACGAGATTATACACATCAGGCACTGGAACTATCTGTAATAC

TGGAACCTAAATAAAACAAAGCAGCCAAAAATTATGCTTGGTTGTTTCATTTTTGTTTAATCCA

GATTTTCCAAAATTTATCACATT

Editing sgRNA

gPalbMis1 (G->T missense): ACTGGAACTATCTGTAATAC**TGG**

sgRNAs adjacent to sgRNA used for editing

gPalbMis2 (tandem 15 bp): AAGCAGCCAAAAATTATGCT**TGG**

Palb2gRNA1 (tandem overlapping, PAMs 15 bp apart):

AGATTATACACATCAGGCAC**TGG** CFD: 23.05

Palb2gRNA2 (everted 21 bp): GTGTGTAGCTTTGGGAAATT**TGG** CFD 28.08

Palb2gRNA3 (inverted 21 bp): AAACAACCAAGCATAATTTT**TGG**

Palb2gRNA4 (tandem 21 bp): **CACACGAGATTATACACATCAGG**

References

1. Giuliano, C. J., Lin, A., Girish, V. & Sheltzer, J. Generating single cell-derived knockout clones in mammalian cells with CRISPR/Cas9. *Curr. Protoc. Mol. Biol.* **128**, e100 (2019).
2. Mathupala, S. & Sloan, A. A. An agarose-based cloning-ring anchoring method for isolation of viable cell clones. *BioTechniques* **46**, 305–307 (2009).
3. Hu, P., Wenhua Zhang, Xin, H., and Deng, G. Single cell isolation and analysis. *Front. Cell Dev. Biol.* **4**, 116 (2016).
4. Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P. & Pruett-Miller, S. M. A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.* **8**, 888 (2018).
5. Ren, C., Xu, K., Segal, D. J. & and Zhang, Z. Strategies for the enrichment and selection of genetically modified cells. *Trends Biotechnol,* **37**, 56–71 (2019).
6. Bauer, D. E., Canver, M. C. & Orkin, S. H. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J. Vis. Exp.* **95,** e52118 (2015).
7. Vouillot, L., Thélie, A., and Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)* **5**, 407–15 (2015).
8. Li, X. *et al.* Highly efficient genome editing via CRISPR-Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. *Nucleic Acids Res.* **46**, 10195–215 (2018).
9. Tamm, C., Kadekar, S., Pijuan-Galitó, S. & Annerén, C. Fast and efficient transfection of mouse embryonic stem cells using non-viral reagents. *Stem Cell Rev.* **12**, 584–91 (2016).
10. Zhang, Z. *et al.* CRISPR/Cas9 genome-editing system in human stem cells: current status and future prospects. *Mol. Ther. Nucleic Acids* **9**, 230–41 (2017).
11. Bruenker, H-G. 558. High efficiency transfection of primary cells for basic research and gene therapy. *Mol. Ther.* **13**, S215 (2006).
12. Modarai, S. R. *et al.* Efficient delivery and nuclear uptake is not sufficient to detect gene editing in CD34+ cells directed by a ribonucleoprotein complex. *Mol. Ther. Nucleic Acids* **11**, 116–29 (2018).
13. Liu, M. *et al.* Methodologies for improving HDR efficiency. *Front. Genet.* **9**, 691 (2019).
14. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–91 (2013).
15. Ye, H., Rong, Z., and Lin, Y. Live cell imaging of genomic loci using dCas9-SunTag system and a bright fluorescent protein. *Protein Cell* **8**, 853–55 (2017).
16. Chen, B., Zou, W., Xu, H., Liang, Y. & Huang, B. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag." *Nat. Commun.* **9**, 5065 (2018).
17. Dreissig, S. *et al.* Live-cell CRISPR imaging in plants reveals dynamic telomere movements. *Plant J.* **91**, 565–73 (2017).
18. Wu, X., Mao, S., Ying, Y., Krueger, C. J. & Chen, A. K. Progress and challenges for live-cell imaging of genomic loci using CRISPR-based platforms. *Genomics Proteomics Bioinformatics* **17**, 119-128 (2019).
19. Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U S A* **112**, 11870–75 (2015).
20. Zhang, D. *et al*. CRISPR-Bind: A simple, custom CRISPR/dCas9-mediated labeling of genomic DNA for mapping in nanochannel arrays. Preprint at https://www.biorxiv.org/content/10.1101/371518v1 (2018).
21. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U S A* **112**, 3002–7 (2015).
22. Boutorine, A. S., Novopashina, D. S., Krasheninina, O. A., Nozeret, K. & Venyaminova, A. G. Fluorescent probes for nucleic acid visualization in fixed and live cells. *Molecules* **18**, 15357–97 (2013).

23. Dahan, L., Huang, L., Kedmi, R., Behlke, M. A. & Peer, D. SNP detection in mRNA in living cells using allele specific FRET probes." *PloS One* **8**, e72389 (2013).
24. Didenko, V. V. DNA probes using fluorescence resonance energy transfer (FRET): designs and applications. *BioTechniques* **31**, 1106–16, 1118, 1120–21 (2001).
25. Wu, X., *et al*. A CRISPR/molecular beacon hybrid system for live-cell genomic imaging. *Nucleic Acids Res.* **46**, e80 (2018).
26. Mao, S., Ying, Y., Wu, X., Krueger, C. J. & Chen, A. K. CRISPR/dual-FRET molecular beacon for sensitive live-cell imaging of non-repetitive genomic loci. *Nucleic Acids Res.* gkz752 (2019).
27. Stains, C. I., Porter, J. R., Ooi, A.T., Segal, D. J. & Ghosh, I. DNA sequence-enabled reassembly of the green fluorescent protein. *J. Am. Chem. Soc.* **127**, 10782–83 (2005).
28. Ooi, A. T., Stains, C. I., Ghosh, I. & Segal, D. J. Sequence-enabled reassembly of betalactamase (SEER-LAC): a sensitive method for the detection of double-stranded DNA. *Biochemistry* **45**, 3620–25 (2006).
29. Ghosh, I., Stains, C. I., Ooi, A.T. & Segal, D. J. Direct detection of double-stranded DNA: molecular methods and applications for DNA diagnostics." *Mol. Biosyst.* **2**, 551–60 (2006).
30. Zhang, Y. *et al.* Paired design of dCas9 as a systematic platform for the detection of featured nucleic acid sequences in pathogenic strains. *ACS Synth. Biol.* **6**, 211–16 (2017).
31. Zhou, L. *et al*. Tandem reassembly of split luciferase-DNA chimeras for bioluminescent detection of attomolar circulating microRNAs using a smartphone. *Biosens. Bioelectron.* **173**, 112824 (2021).
32. Hu, H. *et al*. Live visualization of genomic loci with BiFC-TALE. *Sci. Rep.* **7**, 40192 (2017).
33. Hu, H. *et al*. Visualization of genomic loci in living cells with BiFC-TALE. *Curr. Protoc. Cell Biol.* **82**, e78 (2019).
34. Bernas, T., Robinson, J. P., Asem, E. K. & Rajwa, B. Loss of image quality in photobleaching during microscopic imaging of fluorescent probes bound to chromatin. *J. Biomed. Opt.* **10**, 064015 (2005).
35. Tung, J. K., Berglund, K., Gutekunst, C., Hochgeschwender, U. & Gross, R. E. Bioluminescence imaging in live cells and animals. *Neurophotonics* **3**, 025001 (2016).
36. Choy, G. *et al*. Comparison of Noninvasive Fluorescent and Bioluminescent Small Animal Optical Imaging." *BioTechniques* (2003). https://doi.org/10.2144/03355rr02.
37. Hall, M. P. *et al*. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–57 (2012).
38. England, C. G., Ehlerding, E. B. & Cai, W. NanoLuc: a small luciferase is brightening up the field of bioluminescence. *Bioconjug. Chem.* **27**, 1175–87 (2016).
39. Dixon, A. S. *et al*. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.* **11**, 400–408 (2016).
40. Coggins, N. B., Stultz, J., O'Geen, H., Carvajal-Carmona, L. G. & Segal, D. J. Methods for scarless, selection-free generation of human cells and allele-specific functional analysis of disease-associated SNPs and variants of uncertain significance. *Sci. Rep.* **7**, 15044 (2017).
41. Schwanhäusser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337–42 (2011).
42. Lin, Y., *et al*. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, article number 4767 (2014).
43. Roschke, A.V., Stover, K., Tonon, G., Schäffer, A. A., & Kirsch, I. R. Stable karyotypes in epithelial cancer cell lines despite high rates of ongoing structural and numerical chromosomal instability. *Neoplasia* **4**, 19–31 (2002).
44. Jiang, F. & Doudna, J.A. CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–29 (2017).
45. Arganda-Carreras, I. *et al*. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* **33**, 2424–26 (2017).

Chapter 4: Toward Improvement of Signal-to-Background Ratio for Our Split Luciferase Live Cell DNA Sequence Biosensor

INTRODUCTION

Proposed inefficiencies in original DNA sequence biosensor design

After carefully considering the design of the DNA biosensor used in our first study, we realized that there were several apparent inefficiencies within the design which could limit maximum achievable signal-to-background ratio (SBR). In our initial statistical mechanical model, we assumed that NanoBiTs should reassemble as long as both dCas9-sgRNA-NanoBiT monomers bind to their DNA target sites (**Figure 2.9** species 3). However, this is not accurate for a d*Sp*Cas9-NanoBiT DNA biosensor such as that employed in our first study, which is plagued by several major disadvantages associated with using a single dCas9 enzyme with a single type of cognate sgRNA. First, half of the monomers in a d*Sp*Cas9-NanoBiT DNA biosensor would be paired with sgRNAs of the same sequence specificity and would inhibit DNA detection via competition for a single binding site. Furthermore, two LgBiT-dCas9 fusion proteins, two dCas9-LgBiT fusion proteins, two SmBiT-dCas9 fusion proteins, or two dCas9-SmBiT fusion proteins could bind adjacent to one another on the DNA, which would result in an unproductive binding event and no reassembly of NanoLuc luciferase. We had not accounted for these unproductive binding events in our statistical mechanical model and we realized that we could further optimize our split reporter DNA biosensor design by making alterations to minimize unproductive assemblies of dCas9-sgRNA-NanoBiT monomers on the DNA.

MATERIALS AND METHODS

Construction of Directional d*Sp*Cas9-NanoBiT, NanoBiT-d*Sp*Cas9, d*Sa*Cas9-NanoBiT and NanoBiT-d*Sa*Cas9 Fusion Proteins

The directional fusion constructs containing the LgBiT and SmBiT of NanoLuc luciferase (Promega Corporation) fused to catalytically inactive Cas9 enzymes from *Streptococcus pyogenes* and *Staphylococcus aureus* were generated using the Gibson Assembly method (New England Biolabs). For d*Sp*Cas9 fusion constructs, we used an improved version of the pCDNA3-d*Sp*Cas9 vector containing two nuclear localization signals, an N-terminal 3× Flag epitope tag and [$(GGS)_5$] flexible linker sequences and well as two separate multiple cloning sites at the N- and C-termini of dCas9 and cloned HaloTag, NanoLuc, LgBiT, and SmBiT onto the N- and C-termini of d*Sp*Cas9 using two separate N- and C-terminal multiple cloning sites as described in our previously published study[1]. Overnight N- and C- terminal double restriction digests of sets of flanking restriction sites (XbaI, KpnI) and (NheI, NotI) produced the necessary vector backbones for subsequent Gibson Assembly. HaloTag, NanoLuc, LgBiT and SmBiT inserts were ordered as double stranded DNA fragments or gBlocks Gene Fragments (Azenta Life Sciences and Integrated DNA Technologies) containing approximately 45 bp homologous sequences with the doubly-digested d*Sp*Cas9 vectors upstream and downstream of the two cut sites. NLuc-d*Sp*Cas9 fusion construct was created in our previously published study[1] using overlap extension PCR on LgBiT-d*Sp*Cas9 and SmBiT-d*Sp*Cas9 gBlocks to directionally splice the sequences followed by the Gibson Assembly method again using the N-terminal doubly digested d*Sp*Cas9 vector.

For d*Sa*Cas9 fusion constructs, we used a pCDNA-KRAB-d*Sa*Cas9 vector containing two nuclear localization signals, an N-terminal 3× FLAG epitope tag, [$(GGS)_5$] flexible linker sequences, and two separate multiple cloning sites at the N- and C-termini of d*Sa*Cas9. We cloned HaloTag, NanoLuc luciferase, LgBiT and SmBiT onto the N- and C-termini of d*Sa*Cas9 using two separate N- and C-terminal multiple cloning sites in the

pCDNA-KRAB-d*Sa*Cas9 vector. An overnight N-terminal double restriction digest using restriction sites KpnI and FseI produced the necessary vector backbone with the KRAB domain removed for subsequent N-terminal Gibson Assembly reactions. For C-terminal constructs, this overnight N-terminally double digested vector was further processed to create an appropriate vector backbone with resealed N-terminal multiple cloning site. Two more short oligos were ordered (Integrated DNA Technologies) to insert a short 9 nucleotide adapter sequence, adding 3 amino acids between KpnI and FseI for C-terminal constructs, resulting in a 16 amino acid flexible linker between the 3X FLAG epitope and the N-terminus of d*Sa*Cas9 (See **Extended Experimental Procedures** for sequences). Then, a double digest using NheI and NotI was used to create a final linearized vector backbone for subsequent C-terminal Gibson Assembly reactions. HaloTag, NanoLuc, LgBiT, and SmBiT inserts were ordered as double stranded DNA fragments or gBlocks Gene Fragments (Azenta Life Sciences and Integrated DNA Technologies) containing approximately 45 bp sequences homologous to the doubly-digested dCas9 vectors upstream and downstream of the two cut sites.

For both d*Sp*Cas9 and d*Sa*Cas9 fusion constructs, the assembled d*Sp*Cas9-NanoBiT, NanoBiT-d*Sp*Cas9, d*Sa*Cas9-NanoBiT, NanoBiT-d*Sa*Cas9, d*Sp*Cas9-NanoLuc, NanoLuc-d*Sp*Cas9, d*Sp*Cas9-HaloTag, HaloTag-d*Sp*Cas9, d*Sa*Cas9-NanoLuc, NanoLuc-d*Sa*Cas9, d*Sa*Cas9-HaloTag, and HaloTag-d*Sa*Cas9 constructs were separately transformed into 5-alpha High Efficiency Competent *E. coli* (New England Biolabs) using a standard chemical transformation procedure with heat shock at 42°C and transformed *E. coli* were plated on LB plates containing ampicillin at a final concentration of 100 μg/mL. After an 18 h incubation at 37°C, MiniPreps (QIAGEN) were created for a subset of large, well-separated colonies. The selected subset of large colonies was screened for recombinant vector and insert using both diagnostic restriction digests and colony PCR. Clones positive for the inserts by diagnostic digest and colony PCR were subsequently sequenced to confirm exact sequences were present.

Construction of sgRNA Expression Cassettes and Plasmids

The sgRNA expression vector backbone was obtained from Addgene (Addgene #41824) and was linearized using a restriction digest with AflII. Fourteen 19-bp d*Sp*Cas9 sgRNA target sequences used at non-repetitive *MUC4* in our previous study[1] were combined with 13 additional 19-bp d*Sp*Cas9 sgRNA target sequences and 7 additional d*Sa*Cas9 sgRNA target sequences at non-repetitive human *MUC4* (see **Extended Experimental Procedures** for sequences). These d*Sp*Cas9 and d*Sa*Cas9 sgRNA target sequences were selected using a combination of design tools including CHOPCHOP, CRISPRscan, CRISPick, CRISPOR, and the UCSC genome browser to maximize specificity for the on-target loci.

For d*Sp*Cas9 sgRNAs, each sgRNA sequence was incorporated into two 60mer oligonucleotides that contained homologous sequences to the sgRNA expression vector for subsequent Gibson assembly. After oligonucleotide annealing and extension, the PCR-purified (PCR purification kit; QIAGEN) 100 bp dsDNA was inserted into the AflII linearized sgRNA expression vector using Gibson assembly. Assembly reactions were separately transformed into 5-alpha High Efficiency Competent *E. coli* (New England Biolabs) using a standard chemical transformation procedure with heat shock at 42°C and transformed *E. coli* were plated on LB plates containing kanamycin at a final concentration of 100 µg/mL. After an 18 h incubation at 37°C, MiniPreps (QIAGEN) were created for a subset of large, well-separated colonies. Clones were subsequently sequenced to confirm exact sequences were present. For d*Sa*Cas9 sgRNAs, 456 bp gBlocks Gene Fragments were designed (Integrated DNA Technologies) which contained all elements needed for expression of d*Sa*Cas9 sgRNAs in cells, including a functional U6 promoter, d*Sa*Cas9 sgRNA scaffold, 5' G for enhanced transcription from the U6 promoter, specific 21-bp d*Sa*Cas9 sgRNA target sequence, and a TTTTTT termination signal (see **Extended Experimental Procedures** for gBlock sequences and design principles). These gBlocks were then amplified using Q5 Hot Start High-Fidelity DNA polymerase (New England Biolabs) to produce sufficient quantities for transient transfection into cells.

Dual dCas9 Species dCas9-NanoBiT DNA Biosensing Assays Using Tecan Spark Multimode Plate Reader

For all assays using the Tecan Spark Multimode Plate Reader, HEK 293T cells were originally purchased from ATCC and maintained in Dulbecco's Modified Eagle Medium (Life Technologies) supplemented with 10% FBS and 1X Penicillin/Streptomycin at 37 °C under 5% CO_2. Low passage HEK 293T cells were seeded at 2 x 10^4 cells per well in 96-well opaque white translucent bottom assay plates (Thermo Fisher Scientific) approximately 20 h prior to transfection. 100 ng total DNA was transiently transfected in each well using the Lipofectamine 3000 protocol (Thermo Fisher Scientific), consisting of 29 ng (5 fmol) LgBiT fusion construct, 31 ng SmBiT fusion construct (5 fmol), 13 ng d*Sp*Cas9 sgRNA expression plasmid (5 fmol), 1.41 ng d*Sa*Cas9 sgRNA expression cassette (5 fmol), and 25.59 ng pMAX-GFP. 3 biological and technical replicate wells were included for each combination of fusion protein orientation, target site orientation, and target site spacing. With these methods, HEK 293T cells were typically transfected with approximately 90-95% efficiency. 24 h post-transfection, Nano-Glo Live Cell Substrate (furimazine) was diluted 1:20 in Nano-Glo LCS Dilution Buffer (Promega Corporation) and 25 μL reconstituted Nano-Glo Live Cell Substrate was added to each well of the 96-well assay plates. 10 min after addition of the luminescent substrate, luminescence (photon counts/s, 1500 ms integration time) was read using the Tecan Spark with a band pass (BP) filter centered at the peak emission wavelength of 460 nm with a band pass range of 200 nm from 360 nm-560 nm. For non-targeting sgRNAs, *PALB2 Sp*Cas9 sgRNA 1 from our previous study[1] and *MUC4 Sa*Cas9 sgRNA 1 were transfected. For conditions where no sgRNAs were transfected, 14.41 ng inert pUC19 vector was added to the transfection mix. In conditions where 1 fmol and 0.1 fmol of total dCas9-NanoBiT biosensor plasmids were transfected, the amounts of LgBiT fusion construct, SmBiT fusion construct, and d*Sp*Cas9 sgRNA plasmids, and d*Sa*Cas9 sgRNA expression cassettes were reduced by 10-fold and 100-fold, respectively.

Dual dCas9 Species NanoBRET DNA Biosensing Assays Using Tecan Spark Multimode Plate Reader

In NanoBRET DNA biosensing assays using the Tecan Spark Multimode Plate Reader, low passage HEK 293T cells were seeded at 1.5 x 10^4 cells per well in 96-well opaque white translucent bottom assay plates (Thermo Fisher Scientific) approximately 20 h prior to transfection. 100 ng total DNA was transiently transfected in each well using the Lipofectamine 3000 protocol (Thermo Fisher Scientific), consisting of 30.33 ng (5 fmol) HaloTag fusion construct, 32 ng NanoLuc fusion construct (5 fmol), 13 ng d*Sp*Cas9 sgRNA expression plasmid (5 fmol), 1.41 ng d*Sa*Cas9 sgRNA expression cassette (5 fmol), and 25.59 ng pMAX-GFP. Eight biological and technical replicate wells were included for each combination of fusion protein orientation, target site orientation, and target site spacing. With these methods, HEK 293T cells were typically transfected with approximately 90-95% efficiency. Approximately 20 h post-transfection, the eight replicates for each condition were split into two groups of four and cells were washed once with 25 µL DPBS, then detached from 96-well plates using 25 µL 0.05% Trypsin-EDTA at room temperature. Trypsin-EDTA was neutralized with 75 µL DMEM + 10% FBS + 1X penicillin/streptomycin and cells were centrifuged for 3 min at 125 x g. Then, cells were counted and 100 µL cell suspension was plated in each well of a new 96-well plate along with 0.1 µL of 0.1 mM HaloTag 618 ligand in 100% DMSO for 4 NanoBRET wells (100 nM final concentration HaloTag 618 ligand) or 0.1 µL of 100% DMSO for 4 vehicle alone wells (0.1% final concentration DMSO) for each transfection condition. After allowing cells to reattach and grow for 24 h, a 5X solution of NanoBRET Nano-Glo Substrate (furimazine) was prepared by diluting the stock NanoBRET Nano-Glo Substrate (Promega Corporation) 100-fold in Opti-MEM Reduced Serum Medium with no phenol red (Thermo Fisher Scientific). Then, 25 µL diluted NanoBRET Nano-Glo Substrate was added to each well of the 96-well assay plates. Plates were shaken on a tabletop rotating platform for 1 minute at room temperature then placed back in the incubator at 37 °C under 5% CO_2 for 9 min. 10 min after addition of the luminescent substrate, luminescence (total photon counts, 1000 ms integration time) was read using the Tecan Spark with a band pass (BP) filter centered

approximately at the peak emission wavelength of 460 nm with a band pass range of 85 nm from 415 nm-500 nm. Fluorescence from HaloTag 618 ligand (total photon counts, 1000 ms integration time) was read immediately following NanoLuc luminescence measurements using a long pass (LP) filter starting at 610 nm (near the peak fluorescence emission wavelength of 618 nm) and ending at 700 nm. For non-targeting sgRNAs, *PALB2* d*Sp*Cas9 sgRNA 1 from our previous study[1] and *MUC4* d*Sa*Cas9 sgRNA 1 were transfected. For conditions where no sgRNAs were transfected, 14.41 ng inert pUC19 vector was added to the transfection mix.

RESULTS

Modifying DNA sequence biosensor design elements for improved sensitivity

Because the use of two orthogonal dCas9 enzymes allows the DNA binding specificity of each dCas9 component of the dual species dCas9-NanoBiT complex to be independently programmed by its specific cognate sgRNA, unproductive assemblies in the bound state should not occur when using such a biosensor design strategy. Instead, each fusion protein would be paired only with its specific cognate sgRNA, resulting in more equalization between planned molar transfection ratios and real molar biosensor levels in cells available to produce on-target signal. As a result, we hypothesized that using a system with two orthogonal dCas9 enzymes as opposed to a d*Sp*Cas9-NanoBiT DNA biosensor containing only one type of dCas9 species might increase signal-to-background for our DNA biosensor. To create a dual species dCas9-NanoBiT DNA biosensing system, we chose the well-studied *Streptococcus pyogenes* (*Sp*) and *Staphylococcus aureus* (*Sa*) dCas9s. Specifically, we created 8 directional fusion constructs **(Figure 4.1)** between LgBiT and SmBiT of NanoLuc luciferase and catalytically inactive Cas9 enzymes from *Streptococcus pyogenes* (d*Sp*Cas9) and *Staphylococcus aureus* (d*Sa*Cas9).

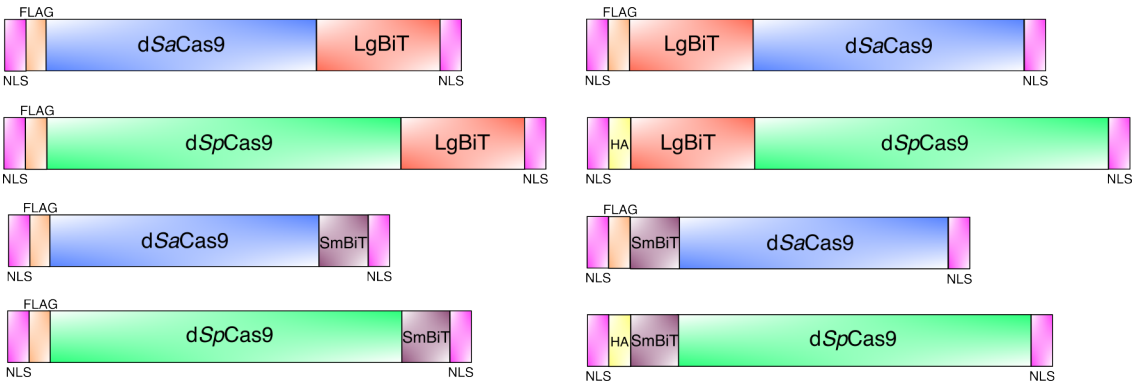


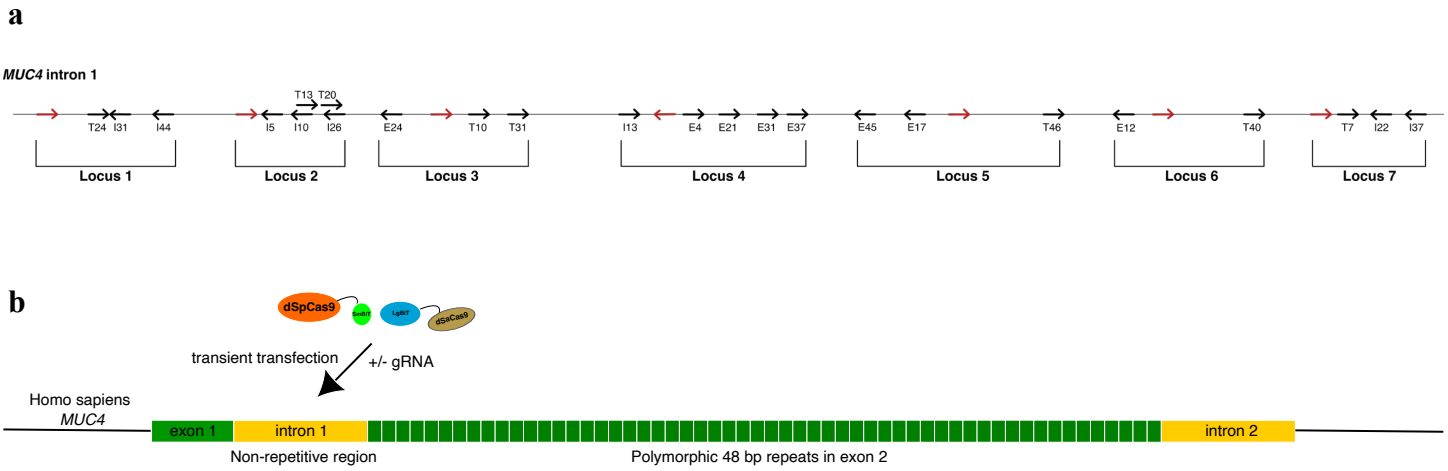
Figure 4.1: Directional d*Sp*Cas9-NanoBiT, d*Sa*Cas9-NanoBiT, NanoBiT-d*Sp*Cas9, and NanoBiT-d*Sa*Cas9 fusion constructs

A cartoon representation of the C-terminal fusion constructs between two orthogonal dCas9 enzymes and LgBiT and SmBiT of NanoLuc luciferase (left) and the N-terminal fusion constructs between two orthogonal dCas9 enzymes and LgBiT and SmBiT of NanoLuc luciferase (right). All C-terminal d*Sp*Cas9-NanoBiT fusion proteins, C-terminal d*Sa*Cas9-NanoBiT fusion proteins, and N-terminal NanoBiT-d*Sa*Cas9 fusion proteins have 3X FLAG epitopes, while N-terminal NanoBiT-d*Sp*Cas9 fusion proteins have HA epitopes. All constructs have two nuclear localization signals (NLS).

Conceptualization and testing of new DNA sequence biosensor designs at endogenous *MUC4* in HEK 239T cells

After creating directional fusion constructs for our dual species dCas9-NanoBiT DNA biosensor, we next created 24 new d*Sp*Cas9 and d*Sa*Cas9 sgRNA pairings with specific spacing and orientation between them at the non-repetitive intron 1 of human *MUC4* **(Figure 4.2a)**. We then transiently transfected HEK 293T cells with all possible combinations of dCas9-NanoBiT fusion protein constructs and sgRNA pairings and measured luminescent signals using a luminometer (**Figure 4.2b-c, see Extended Experimental Procedures for sgRNA sequences)**. To establish background luminescence levels, we sought to measure the signal when transducer

elements were not proximal because they bound at distant regions of the genome, leading to little NanoLuc luciferase reassembly aside from reassembly due to auto-association of the NanoBiTs. To this end, we transfected a sgRNA pair to direct our biosensor to bind distant regions of the genome (*PALB2* on chromosome 16 and *MUC4* on chromosome 3) and compared on-target biosensing signals to the average signal from this background condition. We observed a clear optimal fusion protein orientation in these initial experiments, LgBiT-d*Sa*Cas9 + d*Sp*Cas9-SmBiT, where there was a tendency toward higher signal-to-background ratios compared to other fusion protein orientations **(Figure 4.2c)**. Specifically, signal-to-background peaked at 8.28-fold when this fusion protein pair was directed to bind tandem target sites with a 46 bp spacing between them at endogenous *MUC4* **(Figure 4.2c)**. However, signal-to-background of other target site orientation and spacing combinations, such as inverted target sites 44 bp apart, everted target sites 4 bp apart, everted target sites 37 bp apart, and all other tandem target site spacings showed signal-to-background ratios in the range of 5.5-fold to 8-fold. The lowest range of signal-to-background ratios (1.06-fold to 2.03-fold) was observed in the pairing of d*Sp*Cas9-LgBiT + d*Sa*Cas9-SmBiT fusion proteins **(Figure 4.2c)**.



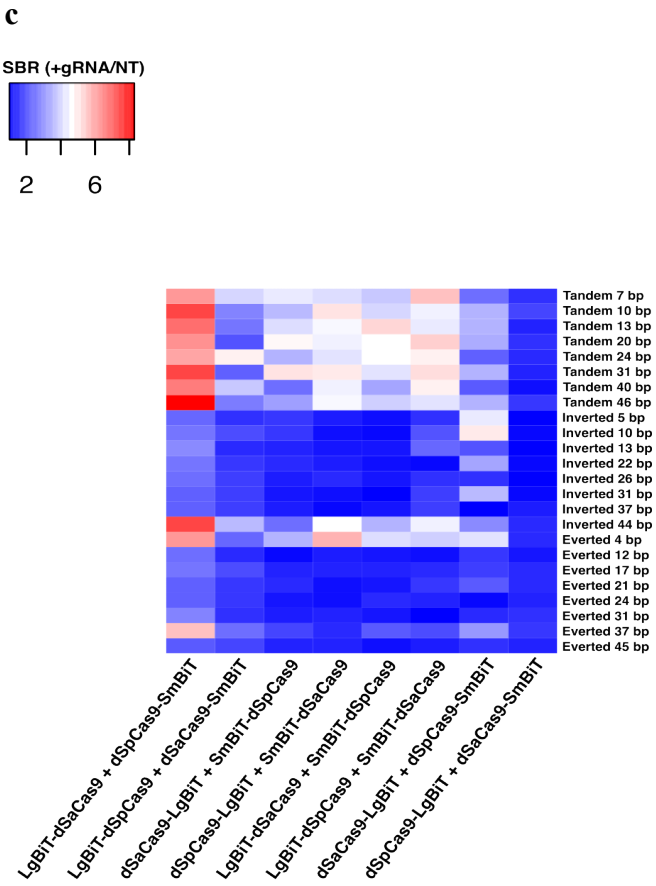


Figure 4.2: Evaluation of signal-to-background of dual species dCas9-NanoBiT DNA biosensor at non-repetitive *MUC4* intron 1

(a) Schematic showing the binding locations and orientations for sgRNAs directed to bind seven loci within the non-repetitive *MUC4* intron 1 region. Orientations are given for d*Sp*Cas9 sgRNAs (black) relative to the d*Sa*Cas9 sgRNA (red) within each locus. Each cluster of sgRNA binding sites was separated from other clusters by between ~100 and 700 bp genomic distance. **(b)** Cartoon depiction of experimental setup for testing our dual dCas9 species DNA biosensor at the non-repetitive intron 1 of human *MUC4*. Unlike previous experiments, the repetitive region of exon 2 was not targeted in these assays with the dual dCas9 species DNA biosensor. **(c)** A heatmap representing signal-to-background ratio (SBR)—defined as normalized NanoLuc luminescence in biosensing conditions with on-target sgRNA pairs transfected divided by the normalized NanoLuc luminescence in biosensing conditions using non-

targeting sgRNA pairs—for 8 fusion protein orientations of the dual dCas9 species DNA biosensor and 24 different orientations and spacing combinations for *MUC4* sgRNA pairings.

As these results were obtained by transfection of higher molar quantities of dual species dCas9-NanoBiT biosensor plasmids and we observed a positive response in signal-to-background at reduced biosensor plasmid levels in initial transfection assays using a single species dCas9-NanoBiT DNA biosensor, we wanted to observe whether a similar effect could be observed using a dual species dCas9-NanoBiT biosensor. Thus, we tested lower dCas9-NanoBiT plasmid molar quantities in transfection using the fusion protein orientation that showed the highest signal-to-background in our initial assays using our dual dCas9 species DNA biosensor, LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT. We discovered that, in stark contrast to previous results observed for a single species d*Sp*Cas9-NanoBiT DNA biosensor, signal-to-background ratios progressively decreased with decreasing molar quantity in transfection **(Figure 4.3a-b)**. Specifically, when 1 fmol and 0.1 fmol dual species LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor plasmids were transfected with sgRNAs for tandem target DNA sites 46 bp apart in HEK 293T cells, signal-to-background dropped from 9.0-fold at 10 fmol transfected to 5.9-fold at 1 fmol transfected and 3.4-fold at 0.1 fmol transfected **(Figure 4.3a-b)**. Similar trends were observed when transfecting the dual species LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor plasmids with sgRNAs for tandem target sites 10 bp apart, tandem target sites 13 bp apart, tandem target sites 31 bp apart, and tandem target sites 40 bp apart **(Figure 4.3a-b)**.

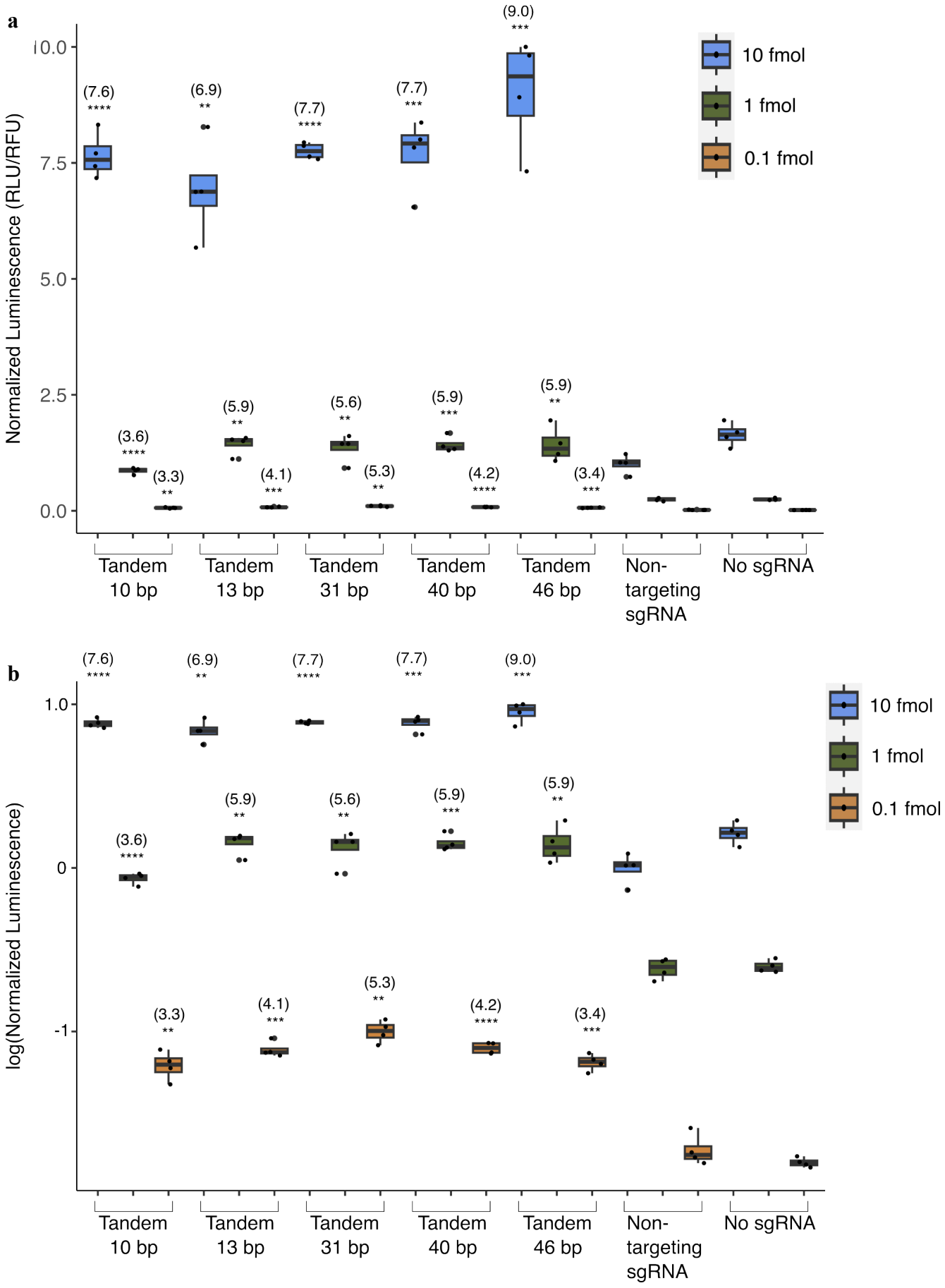

a
Normalized Luminescence (RLU/RFU)
10.0
7.5
5.0
2.5
0.0
10 fmol
1 fmol
0.1 fmol
(7.6) ****
(3.6) ****
(3.3) **
(6.9) **
(5.9) **
(4.1) ***
(7.7) ****
(5.6) **
(5.3) **
(7.7) ***
(5.9) ***
(4.2) ****
(9.0) ***
(5.9) **
(3.4) ***
Tandem 10 bp
Tandem 13 bp
Tandem 31 bp
Tandem 40 bp
Tandem 46 bp
Non-targeting sgRNA
No sgRNA
b
log(Normalized Luminescence)
1.0
0
-1
10 fmol
1 fmol
0.1 fmol
(7.6) ****
(3.6) ****
(3.3) **
(6.9) **
(5.9) **
(4.1) ***
(7.7) ****
(5.6) **
(5.3) **
(7.7) ***
(5.9) ***
(4.2) ****
(9.0) ***
(5.9) **
(3.4) ***
Tandem 10 bp
Tandem 13 bp
Tandem 31 bp
Tandem 40 bp
Tandem 46 bp
Non-targeting sgRNA
No sgRNA

Figure 4.3: Signal-to-background response to reduction in molar amount of LgBiT-d*Sa*Cas9 + d*Sp*Cas9-SmBiT DNA biosensor in transfection

(a) Luminescent signals taken of the dual species LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor binding to several combinations of loci in the non-repetitive region of *MUC4* intron 1 in 293T cells at 10 fmol, 1 fmol, and 0.1 fmol probe transfected. Apparent signal-to-background ratios (non-targeting sgRNA pair background condition) are listed in parentheses. Data are presented as the mean ± s.e.m., $n = 4$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. **(b)** Log-transformation of normalized luminescence for the same data shown in **a** to better visualize the color scheme and full interquartile range for conditions where 1 fmol and 0.1 fmol dual species LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor plasmids were transfected. Apparent signal-to-background ratios (non-targeting sgRNA pair background condition) are listed in parentheses. Data are presented as the mean ± s.e.m., $n = 4$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

Conceptualization and testing of a bipartite DNA sequence biosensor design based on resonance energy transfer

After designing a bipartite DNA biosensor based on split NanoLuc luciferase and an orthogonal dCas9 system, we decided to test other biosensor transducer element designs. We wanted to directly compare signal-to-background measurements of other transducer elements with signal-to-background measurements obtained using split NanoLuc luciferase to assess several drawbacks associated with using this transducer element. Despite the substantially lower auto-luminescent background compared to auto-fluorescent background when applied in cells, the absolute signal from a live cell split luminescent reporter is several orders of magnitude dimmer compared to the absolute signal observed in live cell systems employing full fluorescent reporters[2,3].

Therefore, we imagined a hybrid system that could simultaneously benefit from the high absolute brightness of the signal in fluorescent systems and the low cellular background signal in luminescent systems.

In initial conceptualization of such a biosensor, we hypothesized that it should still contain a bipartite transducer element ideally with minimal signal produced in the unbound state and high signal produced in the bound state, resulting in a true "turn-on" system with relatively low signal from nonspecific association in cells. Thus, we considered a limited set of transducer element options which had a different "turn-on" mechanism for light production. One such transducer element that we thought might be readily combined with our orthogonal biorecognition element using d*Sp*Cas9 and d*Sa*Cas9 as DNA binding domains to provide a sensitive biosensing platform was Förster resonance energy transfer (FRET). FRET is a physical process by which energy is transferred nonradiatively from an excited donor fluorophore to an acceptor fluorophore via intermolecular long-range dipole–dipole coupling[4,5]. FRET is highly dependent on distance between donor and acceptor fluorophores, varying with the inverse of the sixth power of intermolecular separation[4]. The FRET process is maximally efficient when donor and acceptor are positioned within the Förster radius, which is defined as the distance at which half the excitation energy of the donor is transferred to the acceptor[4]. This distance commonly ranges from 3-6 nm[4], and this relatively short distance means FRET is typically used to screen for protein-protein interactions in cells[6]. However, several Förster resonance energy transfer (FRET)-based DNA biosensors have recently been reported in the literature[7-10]. The majority of reported FRET pairs to date have used fluorescent proteins or fluorophores as donor and acceptor chromophores. However, if we made use of FRET pairs containing a fluorescent protein or fluorophore as a donor chromophore in our DNA biosensor designs, such DNA biosensors would be plagued by drawbacks of systems employing fluorescence-based transducer elements such as the requirement for exogenous excitation light to be entrained on cells and fluorophore photobleaching even after short exposure times. Another more subtle problem with FRET systems employing fluorophores or fluorescent proteins as donor and acceptor chromophores is that the background autofluorescence from the donor fluorescent protein or fluorophore could potentially cause a low level of

acceptor fluorescent protein or fluorophore excitation, potentially resulting in negative impacts to the signal-to-background ratio from unintended background signal.

However, a new resonance energy transfer-based system was recently created called the NanoBRET system (Promega), which makes use of a luminescent signal from the donor chromophore, NanoLuc luciferase, and a spectrally optimized fluorophore, HaloTag 618 ligand, as the acceptor chromophore[11]. As with other FRET systems, it is designed to have maximal overlap between the donor emission and acceptor excitation spectra to maximize the efficiency of bioluminescence resonance energy transfer (BRET). Similar to FRET, BRET is typically used to detect protein-protein interactions in cells[12]. Efficient FRET pairs must be selected such that they have little overlap between the donor excitation and acceptor excitation spectra to avoid simultaneously activating both fluorescent moieties and little overlap between the donor emission and acceptor emission spectra to avoid drowning out both signals within the same spectral detection range. One major advantage of the NanoBRET system is that it is not plagued by the former potential problem with excitation signals, as there is no exogenous excitation light required to produce glow-type luminescence from a luciferase enzyme[11]. Essentially, in the NanoBRET process, a low-background, glow-type luminescent signal is amplified to produce a bright fluorescent signal without the need for exogenous excitation light. In addition, there is a great degree of spectral separation between NanoLuc luciferase emission (~460 nm) and HaloTag 618 ligand emission (~618 nm)[11], thus avoiding the latter potential problem with emission signals. Therefore, we reasoned that a dual dCas9 species DNA biosensor where either d*Sp*Cas9 or d*Sa*Cas9 was fused to NanoLuc and either d*Sa*Cas9 or d*Sp*Cas9 was fused to HaloTag—a self-labeling protein tag which can be used to deliver a covalently-bound fluorescent ligand[13]—might therefore offer a more sensitive signal-to-background compared to our dual dCas9 species split NanoLuc luciferase DNA biosensor.

Thus, we created 8 directional fusion constructs **(Figure 4.4)** between HaloTag and NanoLuc luciferase and d*Sp*Cas9 and d*Sa*Cas9. After creating DNA constructs for this dual dCas9 species NanoBRET DNA biosensor,

we transiently transfected them along with the 24 new d*Sp*Cas9 and d*Sa*Cas9 sgRNA pairings with specific spacing and orientation between them at the non-repetitive intron 1 of human *MUC4* to HEK 293T cells and again measured luminescent signals using a luminometer (**Figure 4.5a-b, see Extended Experimental Procedures for sgRNA sequences)**. We observed a clear optimal fusion protein orientation in these initial experiments, HaloTag-d*Sp*Cas9 + d*Sa*Cas9-NanoLuc, where there was a clear tendency toward both higher NanoBRET efficiency and signal-to-background ratios **(Figure 4.5a-b)**. Specifically, signal-to-background peaked at 4.76-fold when this fusion protein pair was directed to bind everted target sites with a 24 bp spacing between them at endogenous *MUC4* **(Figure 4.5b)**. However, signal-to-background of other target site orientation and spacing combinations within this fusion protein orientation, such as tandem target sites 24 bp apart, inverted target sites 10 bp apart, inverted target sites 22 bp apart, and all other everted target site spacings showed signal-to-background ratios in the range of 2.58-fold to 3.95-fold. The lowest range of signal-to-background ratios (1.02-fold to 1.57-fold) was observed in the pairing of d*Sa*Cas9-HaloTag + d*Sp*Cas9-NanoLuc fusion proteins **(Figure 4.5b)**.

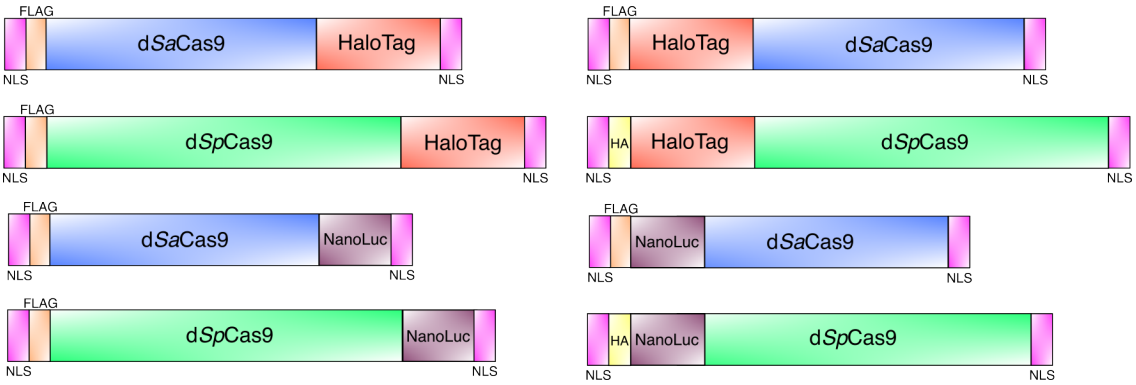


Figure 4.4: Directional d*Sp*Cas9-HaloTag, d*Sa*Cas9-HaloTag, HaloTag-d*Sp*Cas9, HaloTag-d*Sa*Cas9, d*Sp*Cas9-NanoLuc, d*Sa*Cas9-NanoLuc, NanoLuc-d*Sp*Cas9, and NanoLuc-d*Sa*Cas9 fusion constructs

A cartoon representation of the C-terminal fusion constructs between two orthogonal dCas9 enzymes and HaloTag and NanoLuc luciferase (left) and the N-terminal fusion constructs between two orthogonal dCas9

enzymes and HaloTag and NanoLuc luciferase (right). All C-terminal fusion proteins, N-terminal NanoLuc-d*Sa*Cas9 fusion proteins, and N-terminal HaloTag-d*Sa*Cas9 fusion proteins have 3X FLAG epitopes, while N-terminal NanoLuc-d*Sp*Cas9 and HaloTag-d*Sp*Cas9 fusion proteins have HA epitopes. All constructs have two nuclear localization signals (NLS).

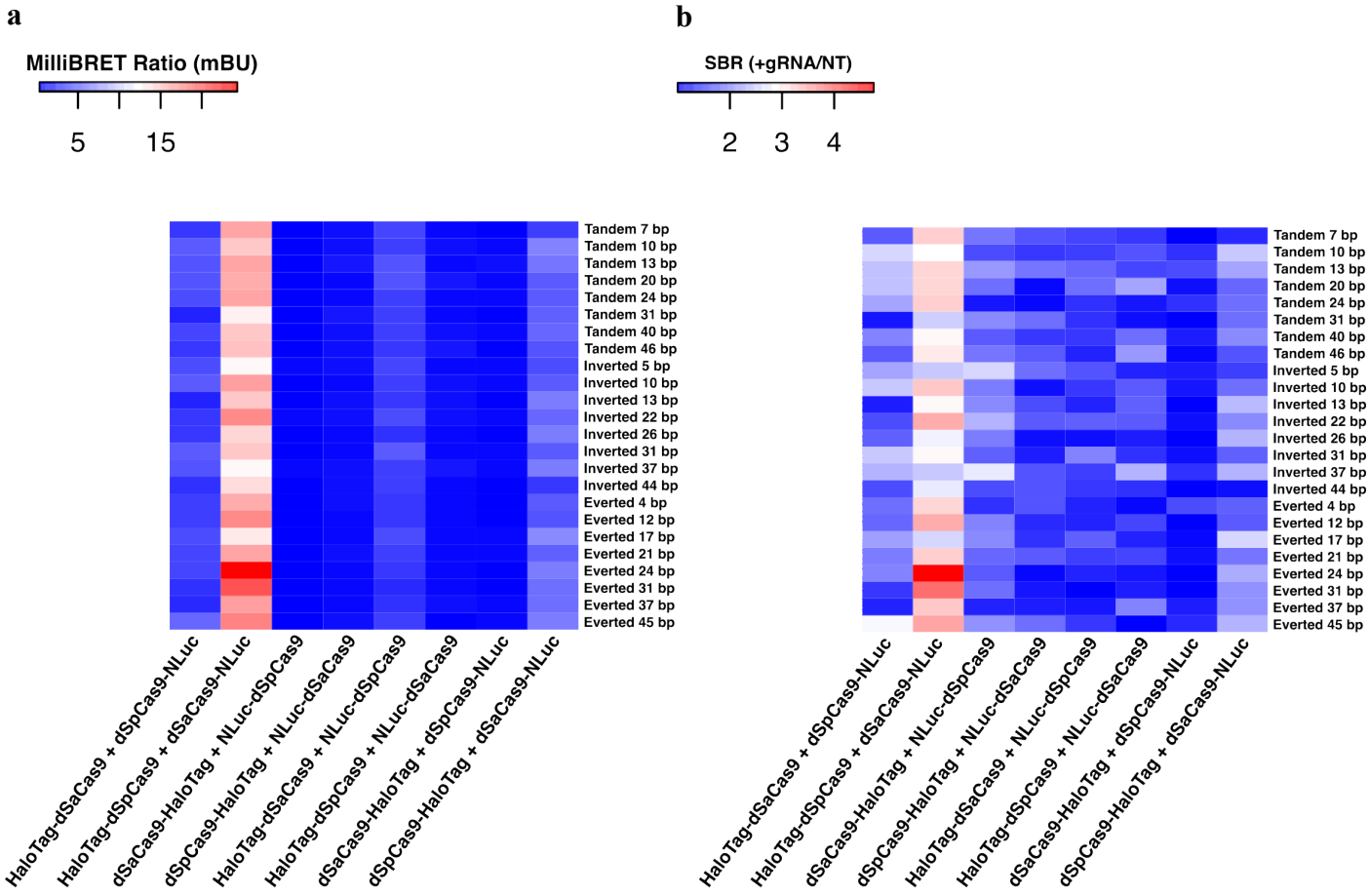


Figure 4.5: Evaluation of BRET efficiency and signal-to-background of NanoBRET DNA biosensor at non-repetitive *MUC4*

(a) A heatmap representing the NanoBRET efficiency given by the milliBRET ratio (mBU)—defined as the HaloTag 618 ligand fluorophore emission value (RFU, 618 nm) divided by the NanoLuc luciferase donor luminescence (RLU, 460 nm) multiplied by 1000—for 8 fusion protein orientations of the NanoBRET DNA biosensor and 24 different orientation and spacing combinations for *MUC4* sgRNA pairings. **(b)** A heatmap representing signal-to-background ratio (SBR)—defined as normalized

milliBRET ratio in NanoBRET biosensing conditions with sgRNA pairs transfected divided by the normalized milliBRET ratio in background conditions using non-targeting sgRNA pairs—for 8 fusion protein orientations of the NanoBRET DNA biosensor and 24 different orientation and spacing combinations for *MUC4* sgRNA pairings.

DISCUSSION

We expected using a DNA biosensor containing both d*Sp*Cas9 and d*Sa*Cas9 as DNA binding domains instead of only d*Sp*Cas9 as a DNA binding domain to positively affect signal-to-background in our initial assays due to reduction in unproductive pairings of dCas9-NanoBiT fusion proteins and sgRNAs. In our early DNA biosensing assays using microscopy to measure signals produced when we transiently transfected d*Sp*Cas9-NanoBiT biosensor plasmids to sense DNA sequences within the non-repetitive intron 1 of *MUC4* in live cells, we observed very low signal-to-background between 1.2-fold and 1.5-fold in HeLa cells at 10 fmol transfected biosensor plasmids, but higher signal-to-background of 6.2-fold and 7.0-fold when the amount of biosensor plasmids transfected was reduced to 1 fmol and 0.1 fmol, respectively[1]. Likewise, in MCF7 cells, signal-to-background was increased from 2.4-4.5-fold at 10 fmol biosensor plasmids transfected to 4.9-fold and 7.6-fold at 1 fmol and 0.1 fmol biosensor plasmids transfected, respectively[1]. Using a dual species d*Sa*Cas9-LgBiT and d*Sp*Cas9-SmBiT DNA biosensor, we initially transfected 10 fmol biosensor plasmids and saw a range of 1.06-fold to 8.28-fold signal-to-background at the non-repetitive intron 1 of *MUC4* in HEK 293T cells using a luminometer to measure average luminescence in whole wells of cells. This range exceeded the range observed using the lower quantities (between 0.1 fmol and 10 fmol) of our single species d*Sp*Cas9-NanoBiT DNA biosensor in individual live cells. As seen in our previously published data, luminescent signal population averages across whole wells of cells measured via luminometer are often lower than average signals of groups of brightly glowing cells visualized via low light microscopy[1]. Thus, it is possible signal-to-background

measurements of cell population average signals using the luminometer for our dual species dCas9-NanoBiT DNA biosensor might be lower than microscopy-based measurements for our single species dCas9-NanoBiT DNA biosensor. Since we observed a higher range of signal-to-background ratios using a luminometer to measure whole well average luminescent signals for our dual dCas9 species DNA biosensor than we observed using microscopy to measure average signals of groups of brightly glowing single cells for our single dCas9 species d*Sp*Cas9-NanoBiT DNA biosensor, we concluded that this dual dCas9 species DNA biosensor seemed to be producing higher signal-to-background than our single dCas9 species d*Sp*Cas9-NanoBiT DNA biosensor overall. We conclude that signal-to-background ratios observed for our dual dCas9 species DNA biosensor could potentially be improved by measuring signals in groups of bright single cells visualized via low light microscopy.

Informed in initial assays by a positive response in signal-to-background at reduced biosensor plasmid levels in transfection using a single dCas9 species DNA biosensor, we tested lower molar plasmid quantities in transfection with our dual dCas9 species DNA biosensor and discovered that the opposite relationship between signal-to-background and molar quantities of LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT biosensor plasmids was true for our dual dCas9 species DNA biosensor. Contrary to our expectations, we did not observe a notable increase in signal-to-background at lower molar amounts of biosensor plasmids for our dual dCas9 species DNA biosensor as we did for our single dCas9 species DNA biosensor. In fact, we observed a notable decrease in signal-to-background at lower molar amounts of biosensor plasmids for our dual dCas9 species LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor. One potential explanation for these contrasting results could stem from the nature of the molecular species available for binding an endogenous DNA target site in live cells when using a DNA biosensor that contains one dCas9 species compared to using a DNA biosensor that contains two dCas9 species. As previously noted, in the case of a single dCas9 species DNA biosensor, half of the available d*Sp*Cas9-NanoBiT and NanoBiT-d*Sp*Cas9 monomers would be paired with sgRNAs of the same sequence specificity and would inhibit DNA detection via competition for a single binding site. In addition, two LgBiT-

dCas9 fusion proteins, two dCas9-LgBiT fusion proteins, two SmBiT-dCas9 fusion proteins, or two dCas9-SmBiT fusion proteins could bind adjacent to one another on the DNA, which would result in an unproductive binding event and no reassembly of NanoLuc luciferase. This means that a significant proportion of the fusion proteins would be involved in unproductive binding events on the DNA which would not produce detectable signal-over-background. In our previously reported study, we observed an increase in on-target signal and a decrease in background signal when the amount of biosensor plasmids was reduced from 10 fmol to 1 fmol and 0.1 fmol[1]. This result suggests that both unproductive assemblies of d*Sp*Cas9-NanoBiT and NanoBiT-d*Sp*Cas9 fusion proteins and background auto-association events may have been reduced at lower biosensor plasmid quantities in transfection and lower expression levels in cells at the time of measurement, both contributing to an increase in the signal-to-background ratio.

In contrast, in the case of our dual dCas9 species DNA biosensor, we expected that there should still be background auto-association events but there should be no unproductive assemblies occurring on the target DNA due to the specific pairing of different dCas9 species with their cognate sgRNA. Thus, we expected the dual dCas9 species DNA biosensor to exhibit higher sensitivity for target DNA compared to our single dCas9 species DNA biosensor even at higher molar quantities in transfection and higher expression levels in cells. Indeed, signal-to-background of our dual dCas9 species DNA biosensor exceeded signal-to-background of our single dCas9 species DNA biosensor at an equal molar quantity of 10 fmol biosensor plasmids in transfection. However, when the molar quantity used in transfection was reduced by 10-fold and 100-fold in transfection for our dual dCas9 species DNA biosensor, we observed a progressive decrease in the background auto-association signal but a slightly larger progressive decrease in the on-target biosensing signal. This result suggests that background auto-association events were reduced due to lower biosensor molar quantities in transfection and lower expression levels at the time of measurement, but on-target binding efficiency was not improved much by lowering the quantity of biosensor plasmids transfected. As there were no unproductive assemblies occurring on the target DNA for our dual dCas9 species DNA biosensor, on-target signal was likely not being suppressed as

it seemed to be in the case of our single dCas9 species DNA biosensor at higher biosensor molar quantities in transfection. Thus, it is possible the on-target signal was already being produced at the highest levels at 10 fmol biosensor plasmids transfected and simply dropped upon lowering biosensor quantities in transfection due to lower expression levels in the nucleus. This result indicates that raising the molar amount of biosensor transfected to the maximum recommended by a given transfection protocol for a specific plate area might improve the signal-to-background further by increasing expressed biosensor fusion proteins and sgRNAs available for producing on-target signal.

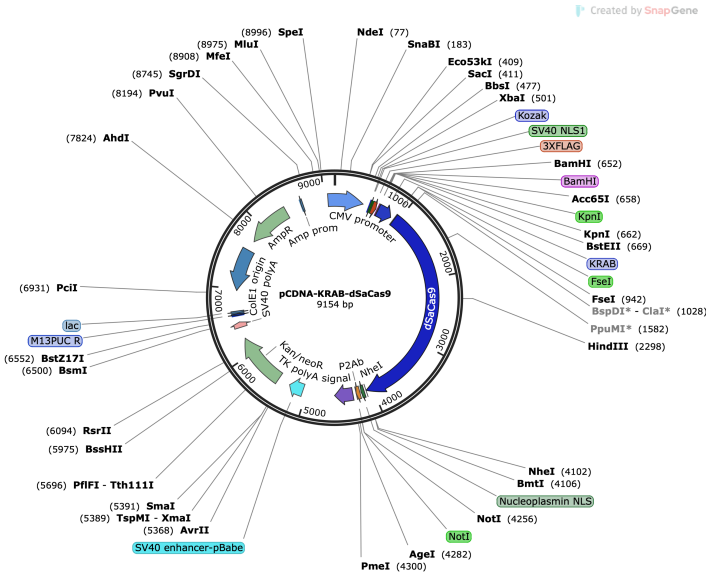
In designing a dual dCas9 species NanoBRET-based DNA biosensor, we reasoned that due to the major advantages associated with the NanoBRET process, which essentially amplifies a glow-type luminescent signal that already has very low-background to produce a much brighter fluorescent signal that is easier to detect using common equipment for measuring fluorescence without the need for exogenous excitation light, a NanoBRET-based DNA biosensing approach might therefore offer a more sensitive signal-to-background compared to our split NanoLuc luciferase DNA biosensing approach. However, one pitfall of using such a system could be that it might be more distance-constrained at distances greater than the Förster radius compared to the broad ranges of readily targetable distances between two DNA target sites we observed with both our single and dual dCas9 species split NanoLuc luciferase DNA biosensors (1-50 bp or ~0.34-17 nm). Indeed, we observed a negative relationship between target site distance and signal-to-background using a dual dCas9 species NanoBRET DNA biosensor when all four fusion protein orientations where NanoLuc and HaloTag were fused to d*Sp*Cas9 and d*Sa*Cas9 on opposite termini were directed to bind tandem and inverted target site orientations. This result could be due to the inward direction of protrusion of fusion protein linkers and linked NanoLuc and HaloTag domains from bound dCas9 enzymes toward the intervening spacer sequence for opposite termini fusion protein orientations on all target sites as observed in initial molecular modeling in PyMOL. However, this negative relationship between target site distance and signal-to-background was not strictly true for everted target site orientations, where often a slight increase in signal-to-background was observed at slightly larger distances.

However, there appeared to be an optimal distance of between 10-30 bp for our dual dCas9 species NanoBRET DNA biosensor across all target sites.

Besides being distance-constrained, FRET and NanoBRET processes are also orientation-constrained, requiring near perfect alignment of donor and acceptor to foster efficient energy transfer via BRET[4,11]. It is likely that a closer interacting protein pair with much lower separation between them would make for better fusion proteins for HaloTag and NanoLuc in the NanoBRET process than d*Sp*Cas9 and d*Sa*Cas9, as flexible linkers attaching HaloTag and NanoLuc to a tightly interacting protein pair might be expected to result in more appropriate alignment and distance well within the Förster radius. In contrast to the wider range of signal-to-background ratios observed using the dual dCas9 species split NanoLuc luciferase DNA biosensor at endogenous non-repetitive *MUC4* (1.06-fold to 8.28-fold), a narrower range of signal-to-background ratios was observed using the dual dCas9 species NanoBRET DNA biosensor on the same endogenous non-repetitive *MUC4* target sites (1.02-fold to 4.76-fold). Thus, we concluded that a NanoBRET DNA biosensing approach, while having several theoretical advantages to a split reporter reassembly DNA biosensing approach and producing readily detectable signal over background, was less sensitive for DNA sequences in live cells than our split NanoLuc luciferase DNA biosensing approach.

Experiment 1: Creation of Fusion Proteins

pCDNA-KRAB-d*Sa*Cas9 vector map:



Oligos for KRAB removal for cloning C-terminal constructs:

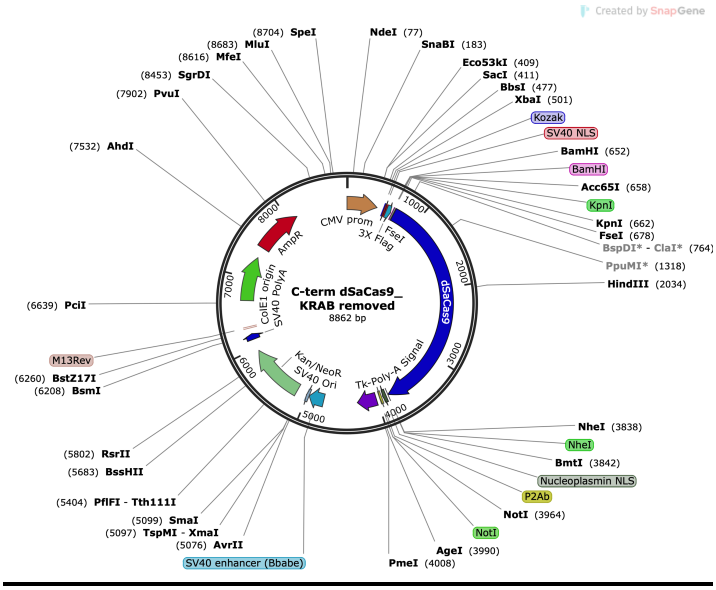
F-oligo:

5’-GACGATGACGATAAGGGTGGCGGGTCCGGCGGTGGATCCGGTACCAGCCTGAGCGGCCGG-3’

R-oligo:

5’-TGATGCCGATGGCCAGGCCCAGGATGTAGTTCCGCTTTGGCCGGCCGCTCAGGCTGGTAC-3’

pCDNA-d*Sa*Cas9 vector map (C-terminal construct cloning):



gBlocks used in N- and C-terminal cloning of HaloTag, NanoLuc, LgBiT and SmBiT with d*Sp*Cas9 and d*Sa*Cas9

NLS-HA-LgBiT (N-terminal fusion), NLS-HA-SmBiT (N-terminal fusion), LgBiT-NLS (C-terminal fusion), SmBiT-NLS (C-terminal fusion) gBlocks and overlap extension PCR primers for d*Sp*Cas9 fusion proteins shown in previously published study[1]

d*Sp*Cas9-NLuc (C-terminal fusion):

TAGTGGAGGTTCAGGAGGATCCGGGGGGAGCGGAGGGAGCGCTAGCATGGTCTTCACACTCGAA
GATTTCGTTGGGGACTGGGAACAGACAGCCGCCTACAACCTGGACCAAGTCCTTGAACAGGGAGG
TGTGTCCAGTTTGCTGCAGAATCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGTCCGGAGCGG
TGAAAATGCCCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGACCAAA
TGGCCCAGATCGAAGAGGTGTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCC
TGCCCTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGT
ATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAAC
AAAATTATCGACGAGCGCCTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAG

TGTGACCGGCTACCGGCTGTTCGAGGAGATTCTCGGTGGAGGCTCCGGAGGTGGATCTAAAAGGC
CGGCGGCCACGAAAAAGGCCGGTCAGGCAAAAAAGAAAAAGGCGGCCGCATGAAAGGGTTCGAT
CCCTACCGGTTAGTAATGAGT

d*Sp*Cas9-HaloTag (C-terminal fusion):

TAGTGGAGGTTCAGGAGGATCCGGGGGGAGCGGAGGGAGCGCTAGCGCAGAAATCGGTACTGGC
TTTCCATTCGACCCCCATTATGTGGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCG
CGCGATGGCACCCCTGTGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATC
ATCCCGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGAC
AAACCAGACCTGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTG
GGTCTGGAAGAGGTCGTCCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAG
CGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGAC
GAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGCT
GATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGCTGACTGA
AGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTT
CCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGG
ACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCAC
CGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGT
CTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGAC
GCTCGAGATTTCCGGCGGTGGAGGCTCCGGAGGTGGATCTAAAAGGCCGGCGGCCACGAAAAAG
GCCGGTCAGGCAAAAAAGAAAAAGGCGGCCGCATGAAAGGGTTCGATCCCTACCGGTTAGTAAT
GAGT

HaloTag-d*Sp*Cas9 (N-terminal fusion):

TCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGAGGATCGAACCCTTGCCACCATG
CCCAAGAAGAAGAGGAAGGTGGGAGGCTCCGGAGGAAGCTACCCATACGATGTCCCAGACTACG
CGGGTGGCGGGTCCGGCGGTGGATCCATGGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATT
ATGTGGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTG
CTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCG
ACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTA
TTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGT
CCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGT
CAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGC

CCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGCTGATCATCGATCAGAACG
TTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATT
ACCGCGAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCA
ATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGGACTGGCTGCACCAGTC
CCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCG
CCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAG
AAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCCGGC
GGTACCGGAGGGAGTGGTGGAAGCGGCGGTTCTGGTGGCTCAG

LgBiT-d*Sa*Cas9 (N-terminal fusion):

GACGATGACGATAAGGGTGGCGGGTCCGGCGGTGGATCCGGTACCATGGTCTTCACACTCGAAGA
TTTCGTTGGGGACTGGGAACAGACAGCCGCCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTG
TGTCCAGTTTGCTGCAGAATCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGTCCGGAGCGGTG
AAAATGCCCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGACCAAATG
GCCCAGATCGAAGAGGTGTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTG
CCCTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGTAT
GAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACA
AAATTATCGACGAGCGCCTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAGT
GGTGGTAGTGGAGGTTCAGGAGGATCCGGGGGGGAGCGGAGGGAGCGGCCGGCCAAAGCGGAACT
ACATCCTGGGCCTGGCCATCGGCATCA

SmBiT-d*Sa*Cas9 (N-terminal fusion):

GACGATGACGATAAGGGTGGCGGGTCCGGCGGTGGATCCGGTACCATGGTGACCGGCTACCGGCT
GTTCGAGGAGATTCTCGGTGGTAGTGGAGGTTCAGGAGGATCCGGGGGGGAGCGGAGGGAGCGGC
CGGCCAAAGCGGAACTACATCCTGGGCCTGGCCATCGGCATCA

d*Sa*Cas9-LgBiT (C-terminal fusion):

AGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCGCTAGCGGTGGTAGTGGAGGTTCAG
GAGGATCCGGGGGGAGCGGAGGGAGCATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGGAA
CAGACAGCCGCCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGCTGCAGAA
TCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGTCCGGAGCGGTGAAAATGCCCTGAAGATCG
ACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGACCAAATGGCCCAGATCGAAGAGGTG
TTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCCCTATGGCACACTGGTA

ATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTT
CGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGC
CTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAGCGGTGGAGGCTCCGGAGG
TGGATCTAAAAGGCCGGCGGCCACGAAAAAGGCCGGTCAGGCAAAAAAGAAAAAGGCGGCCGCA
TGAAAGGGTTCGATCCCTACCGGTTAGTAATGAGTTTAAAC

d*Sa*Cas9-SmBiT (C-terminal fusion):

AGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCGCTAGCGGTGGTAGTGGAGGTTCAG
GAGGATCCGGGGGGAGCGGAGGGAGCGTGACCGGCTACCGGCTGTTCGAGGAGATTCTGGGTGG
AGGCTCCGGAGGTGGATCTAAAAGGCCGGCGGCCACGAAAAAGGCCGGTCAGGCAAAAAAGAAA
AAGGCGGCCGCATGAAAGGGTTCGATCCCTACCGGTTAGTAATGAGTTTAAAC

HaloTag-d*Sa*Cas9 (N-terminal fusion):

GACGATGACGATAAGGGTGGCGGGTCCGGCGGTGGATCCGGTACCATGGCAGAAATCGGTACTGG
CTTTCCATTCGACCCCCATTATGTGGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCC
GCGCGATGGCACCCCTGTGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACAT
CATCCCGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGA
CAAACCAGACCTGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCT
GGGTCTGGAAGAGGTCGTCCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAA
GCGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGA
CGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGC
TGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGCTGACTG
AAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGC
TTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACAT
GGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCC
ACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGG
GTCTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCG
ACGCTCGAGATTTCCGGCGGTGGTAGTGGAGGTTCAGGAGGATCCGGGGGGAGCGGAGGGAGCG
GCCGGCCAAAGCGGAACTACATCCTGGGCCTGGCCATCGGCATCA

NLuc-d*Sa*Cas9 (N-terminal fusion):

GACGATGACGATAAGGGTGGCGGGTCCGGCGGTGGATCCGGTACCATGGTCTTCACACTCGAAGA
TTTCGTTGGGGACTGGGAACAGACAGCCGCCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTG

TGTCCAGTTTGCTGCAGAATCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGTCCGGAGCGGTG
AAAATGCCCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGACCAAATG
GCCCAGATCGAAGAGGTGTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTG
CCCTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGTAT
GAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACA
AAATTATCGACGAGCGCCTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAGT
GTGACCGGCTACCGGCTGTTCGAGGAGATTCTCGGTGGTAGTGGAGGTTCAGGAGGATCCGGGGG
GAGCGGAGGGAGCGGCCGGCCAAAGCGGAACTACATCCTGGGCCTGGCCATCGGCATCA

d*Sa*Cas9-HaloTag (C-terminal fusion):

AGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCGCTAGCGGTGGTAGTGGAGGTTCAG
GAGGATCCGGGGGGAGCGGAGGGAGCATGGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCAT
TATGTGGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTG
CTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCG
ACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTA
TTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGT
CCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGT
CAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGC
CCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGCTGATCATCGATCAGAACG
TTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATT
ACCGCGAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCA
ATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGGACTGGCTGCACCAGTC
CCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCG
CCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAG
AAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCCGGC
GGTGGAGGCTCCGGAGGTGGATCTAAAAGGCCGGCGGCCACGAAAAAGGCCGGTCAGGCAAAAA
AGAAAAAGGCGGCCGCATGAAAGGGTTCGATCCCTACCGGTTAGTAATGAGTTTAAAC

d*Sa*Cas9-NLuc (C-terminal fusion):

AGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCGCTAGCGGTGGTAGTGGAGGTTCAG
GAGGATCCGGGGGGAGCGGAGGGAGCATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGGAA
CAGACAGCCGCCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGCTGCAGAA
TCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGTCCGGAGCGGTGAAAATGCCCTGAAGATCG

ACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGACCAAATGGCCCAGATCGAAGAGGTG
TTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCCCTATGGCACACTGGTA
ATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTT
CGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGC
CTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAGTGTGACCGGCTACCGGCTG
TTCGAGGAGATTCTCGGTGGAGGCTCCGGAGGTGGATCTAAAAGGCCGGCGGCCACGAAAAAGG
CCGGTCAGGCAAAAAAGAAAAAGGCGGCCGCATGAAAGGGTTCGATCCCTACCGGTTAGTAATG
AGTTTAAAC

Final verified protein sequences:

NLuc-d*Sp*Cas9, LgBiT-d*Sp*Cas9, SmBiT-d*Sp*Cas9, d*Sp*Cas9-LgBiT, d*Sp*Cas9-SmBiT sequences shown in previously published study[1]

d*Sp*Cas9-NLuc:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTGGSGGSGGSGGSGGSGRPM
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR
KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ
RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE

VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSG
GSGGSGGSASMVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSGENALKIDI
HVIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILPYGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKI
TVTGTLWNGNKIIDERLITPDGSMLFRVTINSVTGYRLFEEILGGGSGGGSKRPAATKKAGQAKKKKG
GSGSGAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sp*Cas9 (D10A H840A), NanoLuc, Nucleoplasmin NLS, variable length flexible linkers

HaloTag-d*Sp*Cas9:

MPKKKRKVGGSGGSYPYDVPDYAGGGSGGGSMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPV
LFLHGNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVI
HDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTL
PMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFW
GTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGGTGGSGGSGGSGG
SGGSGRPMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK

SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGS
GGSGGSGGSGGSASGGGSGGGSKRPAATKKAGQAKKKKGGSGSGATNFSLLKQAGDVEENPGPAAA
*

KEY: SV40 NLS, HA epitope, d*Sp*Cas9 (D10A H840A), HaloTag, Nucleoplasmin NLS, P2A, variable length flexible linkers

d*Sp*Cas9-HaloTag:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTGGSGGSGGSGGSGGSGRPM
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR
KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKG
ASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ
RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ

AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSG
GSGGSGGSASAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWRNIIPHVAPTH
RCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFHWAKRNPERVKGIA
FMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPFLNPV
DREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAV
DIGPGLNLLQEDNPDLIGSEIARWLSTLEISGGGGSGGGSKRPAATKKAGQAKKKKGGSGSGAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sp*Cas9 (D10A H840A), HaloTag, Nucleoplasmin NLS, variable length flexible linkers

LgBiT-d*Sa*Cas9:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTMVFTLEDFVGDWEQTAAY
NLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDH
HFKVILPYGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLITPDGSMLFRVTIN
SGGSGGSGGSGGSGGSGRPKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR
GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVH
NVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNAD
LYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFT
NLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLS
LKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKY
GLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSL
EAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHIL
NLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSIN
GGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEI
ETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDN
DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKI
KYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKC
YEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRII
KTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGASGGGSGGGSKRPAATKKAGQAKKKKGGSGSG
ATNFSLLKQAGDVEENPGPAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, LgBiT, Nucleoplasmin NLS, P2A, variable length flexible linkers

SmBiT-d*Sa*Cas9:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTMVTGYRLFEEILGGSGGSG
GSGGSGGSGRPKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEED
TGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN
NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIK
DITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE
LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELA
REKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN
NPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRI
SKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRR
KWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEI
FITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINK
SPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNA
HLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKI
SNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIK
KYSTDILGNLYEVKSKKHPQIIKKGASGGGSGGGSKRPAATKKAGQAKKKKGGSGSGATNFSLLKQA
GDVEENPGPAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, SmBiT, Nucleoplasmin NLS, variable length flexible linkers

HaloTag-d*Sa*Cas9:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTMAEIGTGFPFDPHYVEVLG
ERMHYVDVGPRDGTPVLFLHGNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRF
MDAFIEALGLEEVVLVIHDWGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTT
DVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEE
YMDWLHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTL
EISGGGSGGSGGSGGSGGSGRPKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRS

KRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLK
VQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYN
ADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKP
EFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTH
NLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAII
KKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCL
YSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKK
HILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVK
SINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESM
PEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDK
DNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSK
CYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPR
IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGASGGGSGGGSKRPAATKKAGQAKKKKGGSGS
GATNFSLLKQAGDVEENPGPAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, HaloTag, Nucleoplasmin NLS, variable length flexible linkers

NLuc-d*Sa*Cas9:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTMVFTLEDFVGDWEQTAAY
NLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDH
HFKVILPYGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLITPDGSMLFRVTIN
SVTGYRLFEEILGGSGGSGGSGGSGGSGRPKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANV
ENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALL
HLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVK
EAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRS
VKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRV
TSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKG
YTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIK
VINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKIS

YETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNL
DVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEK
QAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLN
GLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKD
NGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENY
YEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENM
NDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGASGGGSGGGSKRPAATKKAGQAKK
KKGGSGSGATNFSLLKQAGDVEENPGPAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, NanoLuc, Nucleoplasmin NLS, variable length flexible linkers

d*Sa*Cas9-LgBiT:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTSLSGRPKRNYILGLAIGITSV
GYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS
GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAE
LQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF
GWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVF
KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQ
SSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDL
SQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTN
ERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKV
LVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN
ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKK
PNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ
YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYR
FDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVI
GVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
ASGGSGGSGGSGGSGGSMVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSG
ENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILPYGTLVIDGVTPNMLNYFGRPYEGIA
VFDGKKITVTGTLWNGNKIIDERLITPDGSMLFRVTINSGGGSGGGSKRPAATKKAGQAKKKKAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, LgBiT, Nucleoplasmin NLS, variable length flexible linkers

d*Sa*Cas9-SmBiT:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTSLSGRPKRNYILGLAIGITSV
GYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS
GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAE
LQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF
GWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVF
KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQ
SSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDL
SQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTN
ERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKV
LVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN
ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKK
PNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ
YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYR
FDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVI
GVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
ASGGSGGSGGSGGSGGSVTGYRLFEEILGGGSGGGSKRPAATKKAGQAKKKKAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, SmBiT, Nucleoplasmin NLS, variable length flexible linkers

d*Sa*Cas9-HaloTag:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTSLSGRPKRNYILGLAIGITSV
GYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS
GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAE
LQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF
GWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVF
KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQ
SSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDL
SQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTN

ERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKV
LVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN
ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKK
PNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ
YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYR
FDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVI
GVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
ASGGSGGSGGSGGSGGSMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWRN
IIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFHWAKR
NPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDH
YREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARLAK
SLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGGGGSGGGSKRPAATKKAGQAKKKKAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, HaloTag, Nucleoplasmin NLS, variable length flexible linkers

d*Sa*Cas9-NLuc:

MPKKKRKVGGSGGSDYKDHDGDYKDHDIDYKDDDDKGGGSGGGSGTSLSGRPKRNYILGLAIGITSV
GYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELS
GINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAE
LQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF
GWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVF
KQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQ
SSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDL
SQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTN
ERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKV
LVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN
ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKK
PNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ
YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYR
FDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVI

GVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
ASGGSGGSGGSGGSGGSMVFTLEDFVGDWEQTAAYNLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSG
ENALKIDIHVIIPYEGLSADQMAQIEEVFKVVYPVDDHHFKVILPYGTLVIDGVTPNMLNYFGRPYEGIA
VFDGKKITVTGTLWNGNKIIDERLITPDGSMLFRVTINSVTGYRLFEEILGGGSGGGSKRPAATKKAGQ
AKKKKAAA*

KEY: SV40 NLS, 3X FLAG epitope, d*Sa*Cas9, NanoLuc, Nucleoplasmin NLS, variable length flexible linkers

Experiment 2: Creation of d*Sp*Cas9 sgRNA plasmids and d*Sa*Cas9 sgRNA cassettes

Oligos for d*Sp*Cas9 sgRNA cloning:

Insert_F: TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC***G****NNNNNNNNNNNNNNNNNNN*

Insert_R: GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC*NNNNNNNNNNNNNNNNNNN****C***

Target protospacer sequence is denoted in green in F oligo and red in R oligo

Design notes for self-contained dSaCas9 gRNA gBlocks:

d*Sa*Cas9 tracrRNA from Ran *et al.* 2015[14]:

5'-
GUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAA
CUUGUUGGCGAGAUUUUUU-3'

Sense DNA strand for synthesis:

5'-
GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTT
GTTGGCGAGATTTTTT-3'

Final d*Sa*Cas9 sgRNA expression cassette gBlock design:

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCG
GGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTT
AGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACG
TAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCA
TATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG

AAACACCGNNNNNNNNNNNNNNNNNNNNGTTTTAGTACTCTGGAAACAGAATCTACTAAAA CAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTTCTAGACCCAGCTT TCTTGTACAAAGTTGGCATTA

Key:

U6 promoter, 5' G for enhanced U6 transcription, 21 nt d*Sa*Cas9 sgRNA target sequence, dSaCas9 sgRNA scaffold, termination signal.

Note: Primer set for gBlock amplification with Q5 Hot Start High-Fidelity Polymerase is underlined

Experiment 3: Creation of non-repetitive *MUC4* sgRNAs for d*Sp*Cas9 and d*Sa*Cas9

MUC4 intron 1 non-repetitive region with **d*Sp*Cas9** and **d*Sa*Cas9** target sites (hg19 chr3 195534754-195538017):

ATAAAATAAAATAAAAAATAAGAAGAGGAGAAAAGTGGGGAAG**AGG**AGGCATGAACTGGCAGATACGGGACAAG

ATCTGAGGGGAAAGACAGAGG**GAGAAT**GCTCGAAAGAGAGAGAAAAGAGAACAGAGGG**~~CCA~~***~~GAGAGCAGC~~*** ***~~CCGG~~***

~~CGATGTCTGG~~AAGGATCCATGGTGAGAG**CCC**AGGCTT**ACTCGC**AGAGAGAAAGACAGGCAGAGCCAGAGCAAGA

GGAA**CAGAGT**CAAGGAGAAAGATGTACACCCTTGTGTACAGAGCT**GGG**GGTAGA**GGGGAT**GCCAGGAAAGCTGG

GTGATGGAGACGGAAGAAAACTCATGTAAAGCTGCA**GGG**TGAGAGGACGAGACAGGTGAGACGCAGACAAACTG

AGGACCCTGGGAATGGAGAGAGGAGAAGATCGGGAGACAGCAGCAAGCAAGGGAAGCGACAAGG**AGG**AGAGGGG

CAGGCCGGCCGGGAGGGTGGTGCGGAGGAGGCGG**CCA**GGGCGCAGAGGG**CCG**GGAGGTGCTGGCCGTGGGCTTC

TTACCTCTGAGCTCGGGTTTAAAAGCCTCCATTTGGGTCACGG**~~CCT~~**TG***~~CCT~~******~~GGGGCTCGTAGCCCGGGCCATTGG~~***

CCTT**GGG**CTCCTCCGTGTACAGAGCTGGGAGGGGAGGGATGCCAGGCCTGTGGGAGATGTTCCCTC**GGG**GGCCC

CCGTCCTCTTCCCCACACTTT**CCA**GGCTGT**CCC**TCTGGCTTCAGGA**CCA**AGTTTTATTCTGTGTTTCTGGGTGT

CTGAGTCTTTGGGGGAGAGTCT**GGG**GTCCAGAGTTCAAGCTGGGGTTAGAGTCTCAGCTCCTGCCCTG**CCT**CTC

AGC**AGG**CTAAGAACAGTCGCCGA**GGG**AAAATATTTCTTGGGCGCATATTTGAGGAGCTTCCT**GGGAGT**GAGTCA

GAAGGCGAGTGCCGTTTAAAGGCTGCAAGAGAAGCCATGC**TGG**TGAAGCGG**ACCCTT**CCACCT**CGGGAT**GTTTC

AGGACTAGGCTGA**GGG**CAAAGGAAACTGCCACCACCT**CCC**TACACCTCCCCACCCTCCAGCACCCCCACCCCAC

CCTGGCCACACAAC**CCC**GCTCCAGTGCTCATCCCACCGTGAGGACGTGGAGGCCGGAAGGAGCCGCCACACGGC

CCTGCCCTGCAGATGTGGTTGA**AGGAGT**CTCCACGGGAATCATGACTCCCAGAGCGAGGCTGGGGCTTGGGGCG

CC**GGG**GAGGCAGCTTGGATTTAGGAGCCCCAGGGCCAAGTCTTTGCCGTGAACTGTTC**TGG**CCCCTGTGACCAG

GCCCTGCCCCGTGTCTCCCCAGGGCCCCGGTCCCCTGTGTAAAAAGCAGTGGTGAACGGTTGGACCTCCTGACG
CCCAAGTTCTTGAGTTTCCAAATCTGTGATTTAAAGCTGAGCCCAAATGTGCTGGGTACCAGCTGGACACTCAG
CTCCATGTGGAGCCAGGAAGTGGGGTCTGTGGAGAGGAGCGCAGAGGGGCAAGACCTGGGGGTGGGCGTGGAAAA
GCACGGGGGCGTGACCCGGAGAAGGAGTGAAGGACTGTTGGTGTGCAAGGGCGTCTCCATGACGACCCGAAGAA
GCTAGGCATGTCGTGGAGCGCTGAGTCCTTTGCGTCGCTAAGGGGACCAAGTGGAGCTGGGCCACGACAGGAGA
TGGTCGTGGCTGGGAGATGGCACCCACACATCTGACCGGGCATGACCAGGGCCTTGGCAGGAAAAGCAGTCACC
AAGGGCGGGTGGGCAGCCCCCACCCCCACAGGGCAGCTGCTGGAGGACTGGCAGCCAGCCAGCCCCGTTCCTTT
TGGCTCCCTGAAGGGGTTTACAGATGACCTGCCTATACTTGAGTCTAGGGTCTGTTTGCACACTTGCCGGCAGG
ACCCTCACCCAGGCTGGGTCACACTGAAGCCCAGGCCAGAGGAAAAACACAGGGTTTCCACAAAGGAGCTGCCG
CAATGAGGGTTTCCTTAAGGAACAGCCCTGGCTCTCAAGGGTTAAAGGATAAGGCACAGCAGACAGAGGTGGGC
TAGACAAGGACAGATGGAAATTTGGTGTCTACTGGTCGCCCCAGGCAGGAATGACTCAGAAGGAAGCCTGGCCG
TCCTGGTTCCATGCCACAGGGAAAGGCAACTGGGTCGAAATAGGCCTTGGTCTCCAGCACTATCAGTGACCCCA
GGGAGGTGACAGGCTGGAGCAAGTGCAGGGCAGGCAGGGGAGGGGACGCCGGCCACAGCGCACTCCACGGGGAA
GGGTCTTTATGGGCCCCTCCTCGGAGAACCCCCGGTCTATCTGTCAGTCTGGGACAGGCCACCTCAACTTGCCA
CCGAGGACACCAAAACTCTCCACAGACCCCTCTGCCCCTCTGGGAAACCCCACTGTGCTCCAGGACACTCAAAA
GGAAAGGATCCCTGGACAAGAGGTCCTGCCAGGAACATCAGCCAAATTTTGGCCAACGACCAGCAAGGTGCACA
GGGAAGAGCAGGGGCTGAAACTCAGAGGTCCAGCATCAGCGACGCCCTTGGCAGCCCAGGGAACACAGGCAACG
CCTTTTGGCTCTGGAGTCTTAGGCTCTTCATCGGCAAACTGAGCCCAGGGGGAAGGGGCTACTACGTAGGGTTG
TCATGAGGATGAAACGAGACAGCATCTGGTGTAAAGTAGAAAAGGCATAAAGGGCCGGGCGCGGTGGCTCACGC
TGTAATCCCAGCACTTTTGGAGGCCCAGGCGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCTTGGCCAA
CCCTGTCTCCACTAAAAATAAAAAATTTTGCCGGGCGTGGTGGCGAGCGCCTGTAATTCCAGCTACTCGGGAGG
CTGAGGTAGGAGAATGGCTTGAACCTGGGAGGCAGAGGTTGCAGGGAGCCGAAATGGCAGCACTCTAGCTTGGG
TGACAGAGCAAGACTCTGTCTAAAAAAAAAAGAAAAGCCATAAAGACGTGTTTGAGAAAGAGGCCTGGGAAGAC
GGGGGAAGGAGGGTGATTGAACCCGGAATGGCACTTGTGTCGGCCCAGGGTCATATCCCTTCATCTAAGGATCC
TCGTGCCTCTAAAAAGCCACCCCGTGCTTCCTGTGGGTTTGCAAGGGCTGGCTTGGTGTATTCAGAATGTGGCT
TGCTGCATGAACGGACCCCGAGGGCCATCCCCTCCAGCGGACAGCTCTGCCTCACCG
CTCCCTGCCTGTGAGTCCCGCCACGCCCTTGGTTTCTGGGCTCAGCCGTGGAGGCAGAGGCTGGCCTGGCAGAG
GCTGGCCT

***Sp*Cas9 gRNAs from low to high CFD**

gRNA1: 1.62 w/ tandem 10 bp nearby site (gRNA 1-1)

gRNA2: 1.79 w/ tandem overlapping, PAMs 17 bp apart nearby site (gRNA 2-1)

gRNA3: 2.94 w/ tandem overlapping, PAMs 15 bp apart nearby site (gRNA 3-1)

gRNA4: 3.50 w/ tandem overlapping, PAMs 4 bp apart nearby site (gRNA 4-1)

gRNA5: 6.29 w/ tandem overlapping, PAMs 8 bp apart nearby site (gRNA 5-1)

gRNA6: 6.55 w/ inverted PAMs overlapping nearby site (gRNA 6-1)

gRNA7: 7.25 w/ tandem 3 bp nearby site (gRNA 7-1)

_dSa_Cas9 gRNAs from high to low efficiency (colored to correspond to d_Sp_Cas9 counterparts)

gRNA6: ATCTGAGGGGAAAGACAGAGG**GAGAAT** **inverted 44 bp** w/ d*Sp*Cas9 g6, **tandem 24 bp** w/ d*Sp*Cas9 g6-1; Eff: 0.8

gRNA2: TCTGGGGTCCAGAGTTCAAGC**TGGGGT** **tandem 31 bp** w/ d*Sp*Cas9 g2, **tandem 10 bp** w/ d*Sp*Cas9 g2-1; Eff: 0.61

gRNA4: CTCGGGTTTAAAAGCCTCCAT**TTGGGT** **tandem 13 bp** w/ d*Sp*Cas9 g4, **tandem 20 bp** with d*Sp*Cas9 g4-1; Eff: 0.6

gRNA5: ACAGGCAACGCCTTTTGGCTC**TGGAGT** **tandem 31 bp** w/ d*Sp*Cas9 g5, **tandem 40 bp** w/ d*Sp*Cas9 g5-1; Eff: 0.52

gRNA3: AGCCCTGGCTCTCAAGGGTTA**AAGGAT** **tandem 46 bp** w/ d*Sp*Cas9 g3, **tandem 64 bp** w/ d*Sp*Cas9 g3-1; Eff: 0.39

gRNA1: TCCACGACATGCCTAGCTTCT**TCGGGT** **everted 4 bp** w/ d*Sp*Cas9 g1, **everted 37 bp** w/ d*Sp*Cas9 g1-1; Eff: 0.35

gRNA7: CAAGGGCTGGCTTGGTGTATT**CAGAAT** **tandem 7 bp** w/ d*Sp*Cas9 g7, **tandem overlapping PAMs 1 bp apart** w/ d*Sp*Cas9 g7-1; Eff: 0.24

Spacings covered with d*Sp*Cas9 gRNAs 1-7, 1-1 to 7-1, and SaCas9 gRNAs 1-7:

Tandem 7, 10, 13, 20, 24, 31, 40, 46 bp

Inverted 44 bp

Everted 4, 37 bp

Additional d*Sp*Cas9 gRNAs for more inverted + everted spacings:

~~gRNA8~~: ~~GGGGCTACGAGCCCCAGGCA~~**~~AGG~~** inverted 5 bp w/ SaCas9 gRNA4

gRNA9: *ATGCCGGGGCTACGAGCCCC**AGG*** inverted 10 bp w/ SaCas9 gRNA4

gRNA10: GAAGGACTGTTGGTGTGCAA**GGG** inverted 13 bp w/ SaCas9 gRNA1

~~gRNA11~~: ~~GCTCTAGGGCCATCGCCCTC~~**~~GGG~~** inverted 22 bp w/ SaCas9 gRNA7

gRNA12: GAGGAGCCCAAGGCCAATGC**CGG** inverted 26 bp w/ SaCas9 gRNA4

~~gRNA13~~: ~~GACATCGCCGGGCTGCTCTC**TGG**~~ inverted 31 bp w/ SaCas9 gRNA6

gRNA14: CTGGAGCGAGCCCCTGCTCT**AGG** inverted 37 bp w/ SaCas9 gRNA7

gRNA15: GCCAAGGGCGTCGCTGATGC**TGG** everted 12 bp w/ SaCas9 gRNA5

gRNA16: TCATTGCGGCAGCTCCTTTG**TGG** everted 17 bp w/ SaCas9 gRNA3

gRNA17: TAAGGGGACCAAGTGGAGCT**GGG** everted 21 bp w/ SaCas9 gRNA1

gRNA18: CAGAAACACAGAATAAAACT**TGG** everted 24 bp w/ SaCas9 gRNA2

~~gRNA19~~: ~~AAGTCGAGCTGGGCCAGGAC**AGG**~~ everted 31 bp w/ SaCas9 gRNA1

gRNA20: CTGTGTTTTTCCTCTGGCCT**GGG** everted 45 bp w/ SaCas9 gRNA3

Note: Text colors, highlighting, underlining, italics, and strikethrough denote positions of gRNA target sequences in *MUC4* genomic sequence above

References

1. Heath, N. G., O'Geen, H., Halmai, N. B., Corn, J. E., & Segal, D. J. Imaging unique DNA sequences in individual cells using a CRISPR-Cas9-Based, split luciferase biosensor. *Front. Genome Ed.* ***4***, 867390 (2022).
2. Tung, J. K., Berglund, K., Gutekunst, C., Hochgeschwender, U. & Gross, R. E. Bioluminescence imaging in live cells and animals. *Neurophotonics* **3**, 025001 (2016).
3. Choy, G. *et al*. Comparison of noninvasive fluorescent and bioluminescent small animal optical imaging. *BioTechniques* ***35***(5), 1022-1026 & 1028-1030 (2003).
4. Sekar, R. B., & Periasamy, A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol*. ***160***(5), 629–633 (2003).
5. Rajdev, P., & Ghosh, S. Fluorescence resonance energy transfer (FRET): A powerful tool for probing amphiphilic polymer aggregates and supramolecular polymers. *J. Phys. Chem.* ***123***(2), 327–342 (2019).
6. Margineanu, A. *et al.* Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM). *Sci. Rep.* ***6***, 28186 (2016).
7. Sato, Y., Nakao, M., & Kimura, H. Live-cell imaging probes to track chromatin modification dynamics. *Microscopy* ***70***(5), 415-422 (2021).
8. Didenko, V. V. DNA probes using fluorescence resonance energy transfer (FRET): designs and applications. *BioTechniques* **31**, 1106–16, 1118 & 1120–21 (2001).
9. Wu, X. *et al*. A CRISPR/molecular beacon hybrid system for live-cell genomic imaging. *Nucleic Acids Res.* ***46***(13), e80 (2018).
10. Mao, S., Ying, Y., Wu, X., Krueger, C. J. & Chen, A. K. CRISPR/dual-FRET molecular beacon for sensitive live-cell imaging of non-repetitive genomic loci. *Nucleic Acids Res.* ***47***(20): e131 (2019).
11. Machleidt, T. *et al*. NanoBRET—A novel BRET platform for the analysis of protein-protein interactions. *ACS Chem. Biol.* ***10***(8), 1797-1804 (2015).
12. Pfleger, K. D., Seeber, R. M., & Eidne, K. A. Bioluminescence resonance energy transfer (BRET) for the real-time detection of protein-protein interactions. *Nat. Protoc.* ***1***(1), 337–345 (2006).
13. Los, G. V. *et al*. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol.* ***3***(6), 373–382 (2008).
14. Ran, F. A. *et al*. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* ***520***(7546), 186–191 (2015).

Chapter 5: A Split Luciferase Biosensing Platform for Detection and Imaging of Chromatin Loops in Individual Living Cells

ABSTRACT

Eukaryotic cells regulate higher-order chromatin architecture, gene expression, and gene recombination via compaction of the 2-meter-long genome into chromatin loops and topologically associating domains (TADs) that operate within a nuclear space with ~10 μM diameter. To facilitate this ~2 x 10^5-fold compaction of chromatin, it has been proposed that chromatin loop structures are created by a loop extrusion process mediated by cohesin, a ring-shaped DNA-entrapping adenosine triphosphatase (ATPase) complex, and CCCTC-binding factor (CTCF), a zinc finger transcription factor. While chromatin architecture has been thoroughly characterized for many eukaryotic genomes using various *in vitro* techniques such as 3C-based methods, noninvasive biosensing tools to probe the landscape of billions of three-dimensional chromatin contacts in real-time are lacking. Using a dual dCas9 species DNA biosensor based on a split NanoLuc luciferase luminescent reporter transducer element, we directly detected chromatin loops in live cells using luminescence quantification in a luminometer and directly visualized chromatin looping at the *MYC* TAD in live cells using high-resolution, low light live-cell imaging. Our biosensing platform should provide a useful methodology for noninvasive, real-time detection of known or novel loops and for monitoring looping dynamics upon alterations in cell state.

INTRODUCTION

As changes in chromatin structure are implicated in disease pathogenesis, we theorized that chromatin loops anchored by cohesin and CTCF represent a biologically significant target for detection and tracking in live cells

using a DNA biosensor. Thus, after designing, testing, and optimizing our dual dCas9 species DNA biosensor for individual DNA sequences in live cells, we wanted to investigate whether it could be repurposed to target native chromatin loops. Although probes based on zinc finger proteins, TALE, and CRISPR-Cas9 combined with fluorescent reporters have recently been developed to track chromatin dynamics in live cells using super-resolution microscopy and other microscopy-based techniques[1-8], these tools are plagued by several limitations, including high auto-fluorescent background, cellular phototoxicity, and rapid photobleaching of fluorophores and other fluorescent reporters, requiring a large number of target-bound probe complexes to reliably distinguish the target-specific foci from the background[9]. In addition, *in vitro* methods such as ChIA-PET, Hi-C, and HiChIP provide a time- and population-averaged representation of the landscape of chromatin contacts in cells, requiring destruction of often valuable cell populations and retrospective analysis of changes in chromatin interactions[10-12]. While these methods have achieved impressive kilobase resolution for chromatin interactions, sensitive detection methods to probe point interactions at even higher resolution are lacking[13-14]. Such a niche could be filled by application of DNA biosensors to quantitate three-dimensional chromatin contacts noninvasively at base-pair resolution in individual live cells and across populations of live cells in real-time. We also hypothesized that a biosensor for chromatin loops could be applied toward monitoring changes in looping frequency at a given locus as a cell responds to certain stimuli or upon other changes in cell state.

Thus, we applied our single dCas9 species d*Sp*Cas9-NanoBiT DNA biosensor from our previous study[15] and our dual dCas9 species LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor to detecting and tracking three-dimensional chromatin contacts in live cells. The idea for application of our DNA biosensor designs to biosensing chromatin loops is based on a loop extrusion model for formation of chromatin loops and topologically associated domains (TADs)[16,17]. This model describes the interaction of distal DNA sequences in *cis* to form insulated neighborhoods, regions typically displaying robust contacts within each neighborhood but weaker contacts between different neighborhoods[16,17,18]. Mechanistically, the loop extrusion model describes a process where chromatin loops form via cohesin initially extruding a loop on DNA bidirectionally until it forms

a complex with CTCF bound to its cognate binding site on the DNA[16-19]. Then, cohesin continues to extrude a loop through its ring-shaped structure by sliding an opposing DNA sequence toward the first bound CTCF until encountering a second CTCF bound to the other DNA strand in an appropriate orientation, where CTCF proteins homodimerize to form a loop anchor[16-19]. Recent studies have shown that formation of TADs is highly dependent on individual CTCF site DNA orientation, with tandem and convergent CTCF sites typically forming boundaries within TADs and divergent CTCF sites forming boundaries between TADs[20,21]. We hypothesized that the loop extrusion mechanism might create a relatively short bridge or anchor region at the base of a chromatin loop, increasing the likelihood of the components of our split NanoLuc transducer element being brought close enough together to produce a detectable signal over background at targetable sites around a cohesin-CTCF homodimer-anchored loop.

In this study, we employed our previously described d*Sp*Cas9-NanoBiT DNA biosensor and a new DNA biosensor containing orthogonal dCas9 enzymes from two bacterial species (NanoBiT-d*Sa*Cas9 and d*Sp*Cas9-NanoBiT) to target chromatin loop anchors formed by cohesin and CTCF homodimers, focusing on interactions between the *MYC* promoter and various cell type-specific super enhancer regions in cancer cell lines. We demonstrate detection of insulating chromatin loops at TAD boundaries at up to 10.5-fold above background and intra-TAD loops between the *MYC* promoter and various cell type-specific super enhancers at up to 6.1-fold above background in common cancer cell lines using a live cell DNA biosensor composed of LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT fusion proteins combined with various sgRNA pairs that bridge across loop anchor regions. In addition, we directly visualized chromatin looping at the *MYC* TAD in live cells using high-resolution, low light live-cell imaging. We show that when one of the two looping loci containing one of the CTCF binding sites was deleted, the on-target biosensing signal decreased to near background levels. Finally, we targeted RAD21, a crucial member of the cohesin complex, for degradation using an auxin-inducible degron (AID) system and found that the chromatin loop biosensing signal decreased at a higher rate to close to background levels when HCT116-RAD21-mAC OsTIR1(F74G) cells were treated with 5-Ph-IAA for 105 min

compared to untreated HCT116-RAD21-mAC OsTIR1(F74G) cells. Together, these results demonstrate probable roles of cohesin and CTCF in the output of luminescence from our biosensor, providing a reliable method for the detection of cohesin and CTCF homodimer-anchored chromatin loops in live cells.

MATERIALS AND METHODS

Chromatin Loop Biosensing Using the Tecan Spark Multimode Plate Reader

For chromatin loop biosensing assays using the Tecan Spark Multimode Plate Reader, HCT116 and K562 cells were originally purchased from ATCC and maintained in McCoy's 5A Medium (Thermo Fisher Scientific) and RPMI 1640 (Thermo Fisher Scientific), respectively, both supplemented with 10% FBS and 1X Penicillin/Streptomycin at 37 °C under 5% CO_2. Low passage HCT116 cells were seeded at 2 x 10^4 cells per well in 96-well opaque white translucent bottom assay plates (Thermo Fisher Scientific) approximately 20 h prior to transfection. 100 ng total DNA was transiently transfected in each well using the Lipofectamine 3000 protocol (Thermo Fisher Scientific), consisting of 29 ng (5 fmol) LgBiT-d*Sa*Cas9 fusion construct, 31 ng d*Sp*Cas9-SmBiT fusion construct (5 fmol), 13 ng d*Sp*Cas9 sgRNA expression plasmid (5 fmol), 1.41 ng d*Sa*Cas9 sgRNA expression cassette (5 fmol), and 25.59 ng pMAX-GFP plasmid. 4 biological and technical replicates were included for each *cis*-interacting loop anchor sgRNA pair transfected. With these methods, HCT116 cells were typically transfected with approximately 70-85% efficiency. Low passage K562 cells were counted, and 4 x 10^5 cells were washed once with 1X DPBS and resuspended in 20 µL Resuspension buffer R (Thermo Fisher Scientific). 500 ng total DNA was transiently transfected via electroporation (1450 V, 10 ms pulse width, 3 pulses) for each 4 x 10^5 K562 cell suspension using the Neon Transfection System (Thermo Fisher Scientific), consisting of 180 ng (30 fmol) LgBiT-d*Sa*Cas9 fusion construct, 192 ng d*Sp*Cas9-SmBiT fusion construct (30 fmol), 81 ng d*Sp*Cas9 sgRNA expression plasmid (30 fmol), 8.75 ng d*Sa*Cas9 sgRNA expression cassette (30 fmol), and 38.25 ng pMAX-GFP plasmid. 4 biological and technical replicate

electroporations were included for each *cis*-interacting loop anchor sgRNA pair by electroporating each condition 4 separate times using two 10 μL Neon tips (Thermo Fisher Scientific) and plating each replicate in a single well of a 24-well cell culture plate (Olympus) with 500 μL RPMI 1640 + 10% FBS with no antibiotics. With these methods, K562 cells were typically transfected with >95% efficiency. 24 h post-electroporation, cells were resuspended in 400 μL fresh RPMI 1640 + 10% FBS with no antibiotics and 100 μL of each of the four replicates was added to a well of a 96-well opaque white translucent bottom assay plate (Thermo Fisher Scientific). For each cell line, 24 h post-transfection, Nano-Glo Live Cell Substrate (furimazine) was diluted 1:20 in Nano-Glo LCS Dilution Buffer (Promega Corporation) and 25 μL reconstituted Nano-Glo Live Cell Substrate was added to each well of the 96-well assay plates. For single time point assays, 10 min after addition of the luminescent substrate, luminescence (total photon counts, 1500 ms integration time) was read using the Tecan Spark with a band pass (BP) filter centered at the peak emission wavelength of 460 nm with a band pass range of 200 nm from 360 nm-560 nm. For time-course biosensing assays, 10 min after addition of luminescent substrate, wild-type HCT116 or HCT116-RAD21-mAID2-Clover cells were treated with either DMSO (untreated) or 1 μM 5-Ph-IAA, marking time point 0. Then, luminescence (total photon counts, 1500 ms integration time) was read every 5 min for 85-105 min starting at time point 0 using the Tecan Spark with a band pass (BP) filter centered at the peak emission wavelength of 460 nm with a band pass range of 200 nm from 360 nm-560 nm. For non-targeting sgRNAs, *PALB2* d*Sp*Cas9 sgRNA 4 from our previous study[15] and *MYC* promoter d*Sa*Cas9 sgRNA d were transfected. For all transfections, in conditions where no sgRNAs were transfected and in transfections of rapamycin-inducible FRB-SmBiT + LgBiT-FKBP fusion proteins (Promega Corporation), 14.41 ng inert pUC19 vector was added to the transfection mix. For positive control NanoLuc-d*Sp*Cas9 and negative control LgBiT-dSaCas9 alone conditions, 32.41 ng inert pUC19 vector was added to the transfection mix. For negative control SmBiT-d*Sp*Cas9 alone condition, 42 ng inert pUC19 vector was added to the transfection mix.

Chromatin Loop Biosensing Using Low Light Microscopy on Andor Dragonfly 200 Spinning Disc Confocal Microscope and Image Processing

For microscopy sessions on the Andor Dragonfly 200 Multi-modal Confocal System, low passage HCT116 cells were plated in 35 mm diameter CELLview 4 quadrant glass bottom imaging dishes (Greiner Bio-One) and transfected identically to experiments using the Tecan Spark Multimode Plate Reader, except that Lipofectamine 3000 (Thermo Fisher Scientific) transfections were scaled up for 24-well plate growth area (1.9 cm^2). Briefly, 2 x 10^5 or 2.5 x 10^5 cells were plated in each quadrant of the imaging dishes and 500 ng total DNA was transfected in each quadrant, consisting of 180 ng LgBiT-d*Sa*Cas9 fusion construct (30 fmol), 192 ng d*Sp*Cas9-SmBiT fusion construct (30 fmol), 81 ng d*Sp*Cas9 sgRNA expression plasmid (30 fmol), 8.75 ng d*Sa*Cas9 sgRNA expression cassette (30 fmol), and 38.25 ng pMAX-GFP plasmid. 4 biological and technical replicate quadrants were included for each *cis*-interacting loop anchor sgRNA pair or control condition transfected. 36 h post-transfection, Nano-Glo Live Cell Substrate (furimazine) was diluted 1:20 in Nano-Glo LCS Dilution Buffer (Promega Corporation) and 200 μL reconstituted Nano-Glo Live Cell Substrate was added to 300 μL McCoy's 5A Medium (Thermo Fisher Scientific) in each quadrant of the imaging dishes. 15 min after addition of the luminescent substrate, GFP fluorescence and NanoLuc luminescence were measured on the Andor Dragonfly Spinning Disc Confocal Microscope. An optimized luminescence imaging protocol was developed for use on the Andor Dragonfly 200 equipped with the Andor iXon Ultra 888 EMCCD camera in which cells were placed in an imaging chamber with temperature control at 37 ºC and all major exogenous light sources in the room were blocked. Using the widefield imaging modality, NanoLuc luminescence was subsequently imaged at 63X magnification with oil immersion using the DAPI channel with 405 nm laser set to minimum intensity (0.002%), shutter turned off, exposure time set to 15 s, and EM gain set to 300. We also changed settings in the Andor image acquisition software to 1X magnification under "Confocal Limit" and to 100% pass under "Image Splitter." GFP fluorescence was imaged using an exposure time of 2 ms and EM gain of 300.

Minimal post-processing was applied to better visualize the images. Raw 16-bit grayscale GFP images were recolored using the "green" look up table (LUT) in Fiji and brightness was reduced to approximately 50% of maximum for each image. Raw 16-bit grayscale NanoLuc luminescence images were recolored using the "red" LUT, brightness was increased to 100%, and contrast was increased to approximately 50% of maximum for each image. To merge GFP fluorescence and NanoLuc luminescence images, we directly merged color channels in Fiji. The WEKA Segmentation package[22] in Fiji was then used to segment cell nuclei using 25 ROI traces of the nuclear luminescence signals and 25 ROI traces of the background outside of cells as a training data set for nuclear boundaries. This trained WEKA segmentation model was then applied to each NanoLuc image to determine boundaries of nuclei. Each 8-bit segmented image outputted from the WEKA model was binarized using the "auto-threshold" function, and the "analyze particles" function was applied to create ROIs for each nuclear area. These ROIs were then overlaid from the ROI manager onto unprocessed images for quantification. Then, the mean intensity (equivalent to integrated intensity of each nucleus divided by area of each nucleus) after 15 s of total light collection was calculated and recorded for each segmented nuclear area using Fiji. Any cells that were positive for GFP fluorescence but negative for NanoLuc luminescence were omitted from final statistical analysis.

4C-seq

4C-seq was conducted as described previously[23-26]. For each replicate, approximately 5 million cells were crosslinked in 2% formaldehyde and 10% FBS in PBS for 10 min while rotating at room temperature. Glycine was added to a final concentration of 0.125 M to quench crosslinking reaction and cells were centrifuged at 500 x g for 5 min. Cells were washed twice with PBS, transferred to an eppendorf tube, and lysed on ice for 20 min with 1 mL lysis buffer (50 mM Tris-HCl (pH 7.5), 10 mM NaCl, 5 mM EDTA, 0.5% NP-40 and 1% Triton X-100) supplemented with 1X protease inhibitors (Roche). Cells were then washed once with 450 µL of 1.2X DpnII buffer and resuspended in 500 µL 1.2X DpnII buffer and 15 µL of 10% SDS (final: 0.3% SDS) and

incubated in a thermoshaker at 37 °C for 1 h while shaking at 750 RPM. Then, 75 µL of 20% Triton X-100 (final: 2.5% Triton-X-100) was added, nuclear aggregates were resuspended with a pipet, and tubes were incubated at 37°C for 1 h while shaking at 750 RPM. A 5 µL aliquot of undigested control was stored, and nuclei were incubated overnight with 250 U of DpnII restriction enzyme (New England Biolabs). A fresh 250 U of DpnII was added the following morning, and samples were digested for an additional 3 h. After this, digestion was checked for completion by running 5 µl of sample on a 0.6% agarose gel next to undigested lysates. DpnII was inactivated by incubating the reaction tubes for 20 min at 65 °C, and a proximity ligation reaction was performed in a 7 mL volume with 3200 U of T4 DNA ligase (New England Biolabs) for 16 h at room temperature followed by an overnight period at 16 °C and finally a 500 U T4 DNA ligase spike-in and another 36 h incubation at room temperature. After this, ligation was checked for completion by running 5 µl of sample on a 0.6% agarose gel next to the digested control. Cross-links were then reversed at 65 °C overnight after adding 300 µg proteinase K. Samples were then treated with 300 µg RNase A for 30 min at 37 °C, and DNA was purified using MagMax DNA binding beads (Thermo Fisher Scientific) and a magnetic bead-based protocol described in a previous reference[24]. A second restriction digest was performed overnight while shaking at 500 RPM at 37 °C in a 500 µL reaction with 200 U of NlaIII (New England Biolabs). After this, digestion was checked for completion by running 5 µl of sample on a 0.6% agarose gel next to the samples from the first ligation. NlaIII was inactivated at 65 °C for 25 min, and a proximity ligation reaction (5 ng/µL DNA concentration) was performed in a 5 mL volume with 2000 U of T4 DNA ligase for 16 h at room temperature followed by an overnight period at 16 °C and finally a 400 U T4 DNA ligase spike-in and another 36 h incubation at room temperature. After ligation, DNA was again purified using MagMax DNA binding beads (Thermo Fisher Scientific)

To generate the 4C–seq library, several viewpoint-specific inverse PCR primer sets targeting the *MYC* promoter conserved CTCF binding site (see **Extended Experimental Procedures** for all viewpoint and primer sequences) were compared for efficiency and the primer pair that produced the strongest product was selected

for the 1st round PCR in library preparation. 200 ng of prepared 4C template was amplified with 16 PCR cycles in 4 separate 50 µL reactions using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs). Next, these 4 reactions were mixed and 50 µL of 1st round PCR product was purified using ChargeSwitch PCR Clean-Up Kit (Thermo Fisher Scientific) to remove primer dimers. 5 µL purified 1st round PCR product was used as a template in a 2nd round of PCR for 20 cycles in a 50 µL reaction using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) to add Illumina P5/P7 flow cell adapters, Illumina sequencing primer sites, and unique 6 bp indexes to all replicates. These 2nd round reactions were then purified with the QIAquick PCR purification kit (QIAGEN), the purity was measured with a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific), the concentration was more accurately determined using the Qubit dsDNA HS Assay (Thermo Fisher Scientific), and all replicates were run on the Bioanalyzer 2100 (Agilent) to check for any remaining primer dimers and determine average library size. After final quality control and Element library circularization, the 4C-seq libraries were sequenced on the Element Biosciences AVITI for 150 bases in paired-end read mode at 300 cycles.

Statistical Testing

Two-tailed Student's *t*-tests for signal-to-background analyses and Pearson's correlation analyses were conducted in Microsoft Excel 2016. Statistics shown in all box-and-whisker and time-series plots were computed using R (version 4.3.0).

RESULTS

Modeling the structure of cohesin and CTCF homodimerization-mediated loop anchors

To better understand the feasibility of designing a biosensor for chromatin loops, we modeled the structure of a chromatin loop anchored by cohesin and a CTCF homodimer. Although to date, a native chromatin loop

structure has not been captured *in situ* using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy as it is a highly dynamic and often transient assembly involving multiple protein complexes, reasonable models can be made for such a structure for the purposes of biosensor design via combination of known crystal and NMR structures and a full predicted structure from AlphaFold. Starting with the published crystal structure of ZnFs 4-10 of CTCF bound to 28-mer DNA[27] **(Figure 5.1a)** and the discovery of an N-terminal segment of CTCF that directly interacts with the SA2-SCC1 subcomplex of cohesin and is specifically required for cohesin positioning at CTCF binding sites and for the formation of CTCF-anchored loops[28] **(Figure 5.1b)**, we began to model the structure of a cohesin and CTCF homodimerization-mediated chromatin loop anchor. From the SA2-SCC1-CTCF complex structure, a critical stretch of amino acids from positions 222-232 in CTCF—particularly Y226 and F228—associate closely with several amino acids in cohesin subunits SCC1 and SA2, including S334, I337, R338, and L341 of SCC1 and Y297, R298, and W334 of SA2[28]. Previous literature showed that SA2 contains an 86-amino-acid motif termed the stromalin conservative domain[29,30] or conserved essential surface (CES)[31,32] which coincides with the CTCF binding pocket. Thus, we needed to align this stretch of amino acids in the SA2-SCC1-CTCF structure (CES) with the correct stretch of amino acids in the full CTCF structure, then further align this hybrid structure to the structure of ZnFs 4-10 bound to DNA. We reasoned this sequence of alignments would allow us to accurately position the ZnF domains of CTCF and bound DNA in relation to the cohesin subunit structure. As the N-terminus of CTCF was missing in the crystal structure of ZnFs 4-10 and the SA2-SCC1-CTCF complex but the idea that it is an unstructured loop domain is thoroughly supported by the literature[27,28,33], we used a full CTCF structure predicted by AlphaFold to locate the N-terminal amino acids in CTCF that interact with cohesin and create a more accurate molecular model of the positioning of CTCF relative to cohesin subunits and to bound DNA. AlphaFold predicted the N-terminal and C-terminal domains missing from all published structures to be unorganized loop domains, consistent with the literature **(Figure 5.1c)**. The absence of the N- and C-terminal loop domains in published structures of CTCF could have occurred because the innate flexibility in proteins can often result in the absence of loops and even entire domains in structures determined by x-ray diffraction or NMR spectroscopy[34].

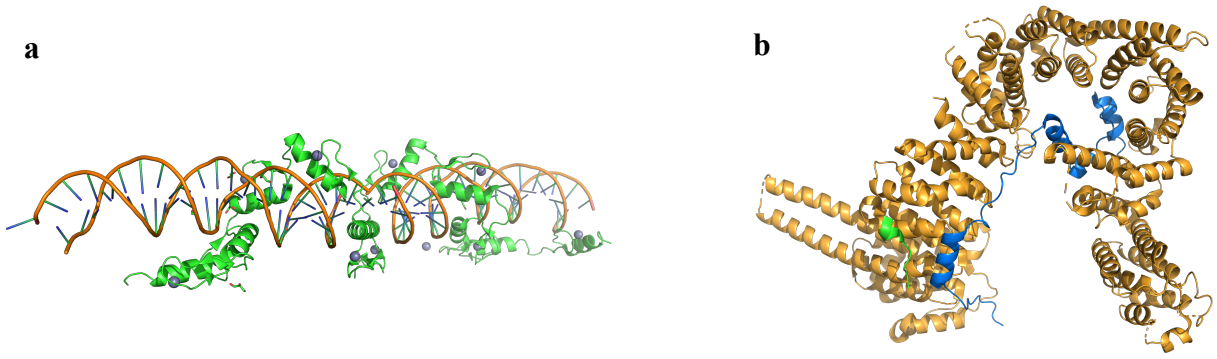
It has recently been shown that directional CTCF binding to cognate binding sites is achieved through insertion of ZnF 3, ZnFs 4-7, and ZnFs 9-11 into the major groove along cognate CTCF binding sites[35]. All CTCF binding sites are composed of a 19 bp core sequence that has a specific directionality and is bound by ZnFs 3-7[35]. However, some CTCF binding sites also include an additional upstream 10 bp sequence bound by ZnFs 9-11 separated from the core sequence by a 5-6 bp intervening sequence bound by ZnF 8[35]. With this in mind, we first aligned the structures which both contained the stretch of amino acids from positions 222-232 in the N-terminus of CTCF **(Figure 5.1d)** then aligned ZnFs 4-10 in this aligned structure with ZnFs 4-10 in the crystal structure where these ZnFs are bound to 28-mer DNA, removing the redundant ZnFs 4-10 and loop domains predicted by AlphaFold from the combined complex for clarity **(Figure 5.1e)**.

Then, we created a full loop anchor model with two *cis*-interacting DNA sites. To do this, we reconciled the remarkable fact that the cohesin complex, having a diameter of only ~50 nm[36,37] can discriminate between approach of comparatively tiny DNA-bound CTCF protein (~3-5 nm) from the N-terminal or C-terminal side during extrusion of a chromatin loop[37,38] with the well described preference for CTCF-CTCF homodimers in CTCF-anchored chromatin loops to be in the convergent orientation[39,40]. As there is substantial evidence that CTCF proteins homodimerize *in vivo*[41] and the N-termini of two homodimerizing CTCF proteins must align in loop anchors of the preferred convergent orientation[37], we first modeled the CTCF homodimer in the case where N-terminal CTCF domains align fairly closely, at ~15 Å apart **(Figure 5.2a-b)**. In the case where ZnF domains were aligned ~15 Å apart, the largest distance between *cis*-interacting DNA sequences was found to be ~40 Å (~4 nm) while the smallest distance was found to be ~10 Å (~1 nm). Often, side chain interactions between residues at homodimer interfaces can be <4.5 Å[42] and a large protein such as CTCF (727 amino acids) likely adopts a relatively tightly associated 3-state with dimer intermediate (3DSI) homodimerization interface[43] with the main interacting domain being the unstructured N-terminal loop domain[44,45]. Thus, we also modeled the case of a very tight dimerization interface where N-terminal CTCF domains were ~4.5 Å apart **(Figure**

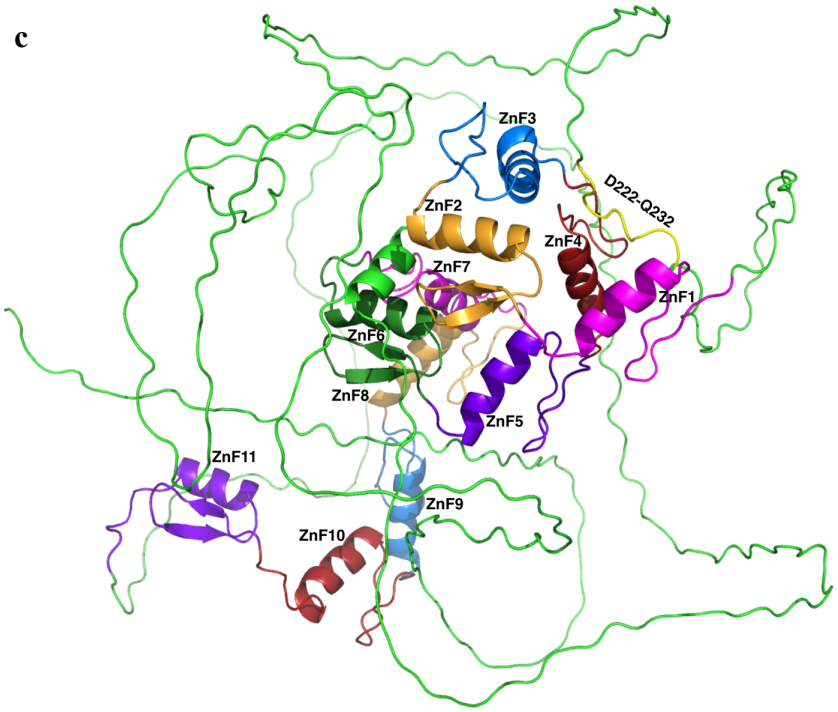
5.3a-b). In these tight dimerization interface models, we found that the largest distance between cis-interacting DNA sequences was ~30 Å (~3 nm) while the smallest distance was found to be ~7 Å (~0.7 nm). Thus, perhaps unsurprisingly, we found distance between *cis*-interacting DNA sequences was positively correlated with distance between homodimerized CTCF proteins in our models. However, depending on the degree to which negatively charged *cis*-interacting phosphodiester backbones are neutralized and shielded via bound CTCF ZnF domains with Zn2+ cations, CTCF-CTCF homodimers dimerizing with such tight interfaces might be somewhat destabilized by electrostatic repulsion between DNA backbones[46]. Thus, this results in some uncertainty regarding the actual expected separation for two *cis*-interacting DNA sites in a cohesin-CTCF-CTCF-anchored loop. The true distance between two *cis*-interacting DNA sequences might be less than ~7 Å or more than ~40 Å depending on the error in our models.

Currently, there are two main schools of thought regarding how cohesin extrudes loops and anchors loops with CTCF. Either cohesin subunits bind to each interacting DNA site, bridging them together, or the ring structure of cohesin encloses both interacting DNA sites. Currently, the latter idea is highly favored and supported by substantial evidence[47-50]. Therefore, while still debated, the idea that monomeric cohesin extrudes loops is currently preferred. The SMC1 and SMC3 domains of cohesin contain antiparallel coiled coil domains that allow for dimerization, creating a hinge-like structure and completing a ring structure bridged on the opposite terminus by SCC3 and SCC1/RAD21 domains[51]. Since the *cis*-interacting DNA sequences shown in the models in **Figure 5.2** and **Figure 5.3** are stabilized via homodimerization of CTCF proteins and anchored by attachment of CTCF proteins to cohesin SA2 and SCC1/RAD21 domains, if these sequences were to be extended toward the SMC1/SMC3 hinge, they would be inserted through the cohesin ring structure starting from the attachment point of CTCF to cohesin at the SCC1/RAD21 and SA2 domains. On both sides of the loop anchor moving away from the CTCF binding sites, the distance between *cis*-interacting DNA sequences is likely to increase at a certain rate as they become unanchored by cohesin-CTCF-DNA interactions. On the side of the loop anchor closest to the insertion point of the *cis*-interacting DNA sequences to the cohesin ring, the

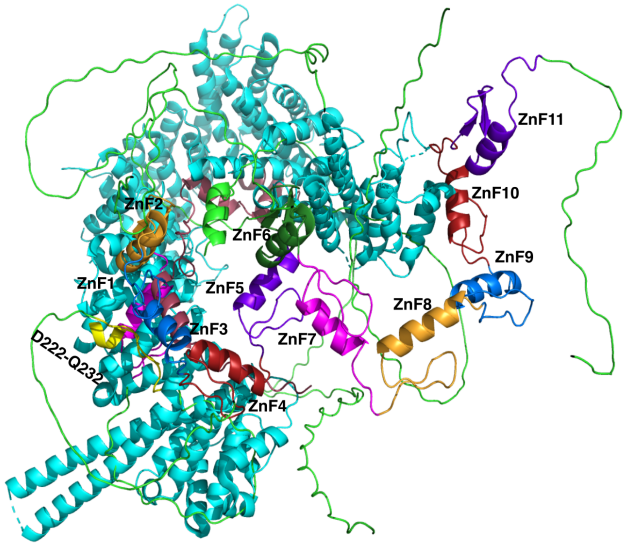
distance between them would increase, but only to an upper bound of ~50 nm. However, if these sequences were to be extended away from the SMC1/SMC3 hinge, it is possible the distance between them would increase more quickly due to lack of constraint. Thus, we concluded that the ~50 nm cohesin ring diameter formed an upper bound for the distance between interacting DNA sites within a reasonably short distance from the two CTCF binding sites. To date, the relationship of distance between *cis*-interacting DNA sequences and spacing from the core CTCF binding sites at loop anchor regions has not been determined. For all models, to ensure we measured distances across the loop anchor as close as possible to the more predictably anchored core CTCF binding sites and immediately adjacent sequence instead of the more unpredictably anchored extensions from the loop anchor in each direction on the DNA, we measured distances between *cis*-interacting DNA sequences across the loop anchor immediately upstream of the non-core 10 bp CTCF binding module **(Figure 5.2, Figure 5.3)**. Thus, we established from our models that the degree of separation between CTCF monomers within the homodimer structure created a lower bound for the distance between interacting DNA sites of ~0.7-4 nm, well within the range of interaction distances at which our LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT probe might operate given by results of our previous studies showing signal-to-background production at linear DNA distances of ~3.4-17 nm for single and dual dCas9 species DNA biosensors[15]. In addition, our model showed that proximal to the two CTCF binding sites, any two *cis*-interacting DNA sequences in a loop anchor are likely not separated by a distance near the cohesin ring diameter but are instead more proximal within the loop anchor structure.



c


ZnF3
D222-Q232
ZnF2
ZnF4
ZnF7
ZnF1
ZnF6
ZnF8
ZnF5
ZnF11
ZnF9
ZnF10

d


ZnF11
ZnF10
ZnF2
ZnF6
ZnF9
ZnF1
ZnF5
ZnF8
ZnF3
ZnF7
D222-Q232
ZnF4

e

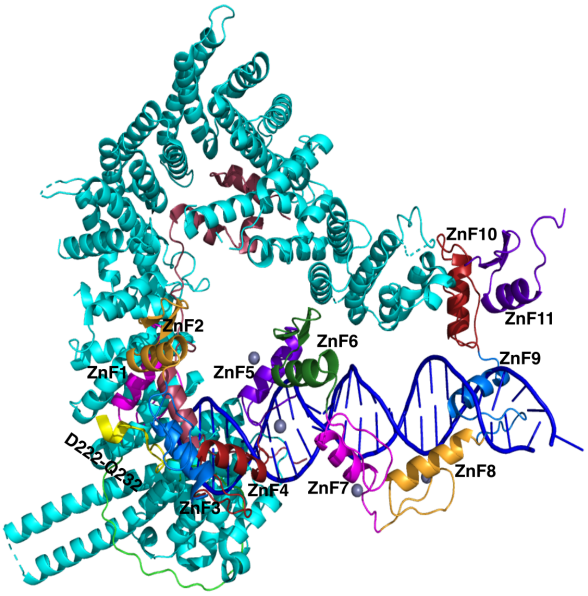


Figure 5.1: Conceptualization of cohesin and CTCF homodimer-mediated loop anchor models

(a) Crystal structure of ZnFs 4-10 of CTCF (green) bound to 28-mer DNA (orange). **(b)** Crystal structure of cohesin subunits SA2 (orange) and SCC1/RAD21 (blue) in association with the N-terminal loop domain of CTCF. CTCF amino acids D222-Q232 are shown in green, which interact with S334, I337, R338, and L341 of SCC1 and Y297, R298, and W334 of SA2. **(c)** Full CTCF structure predicted by AlphaFold. CTCF amino acids D222-Q232 are highlighted in yellow within the N-terminal loop domain. ZnFs 1-11 in CTCF are labeled. **(d)** Alignment of the structures which both contain the stretch of amino acids from positions 222-232 in the N-terminus of CTCF (yellow) to produce a full CTCF-SCC1-SA2 complex. Cohesin subunits SA2 (teal) and SCC1/RAD21 (maroon) are shown attached to CTCF via intermolecular interactions between specific amino acids in SCC1/RAD21 and CTCF. **(e)** Alignment of the CTCF-SCC1-SA2 complex with ZnFs 4-10 with 28-mer CTCF binding site. Redundant ZnFs 4-10 and CTCF loop domains predicted by AlphaFold were removed after alignment. 28-mer DNA is shown in dark blue.

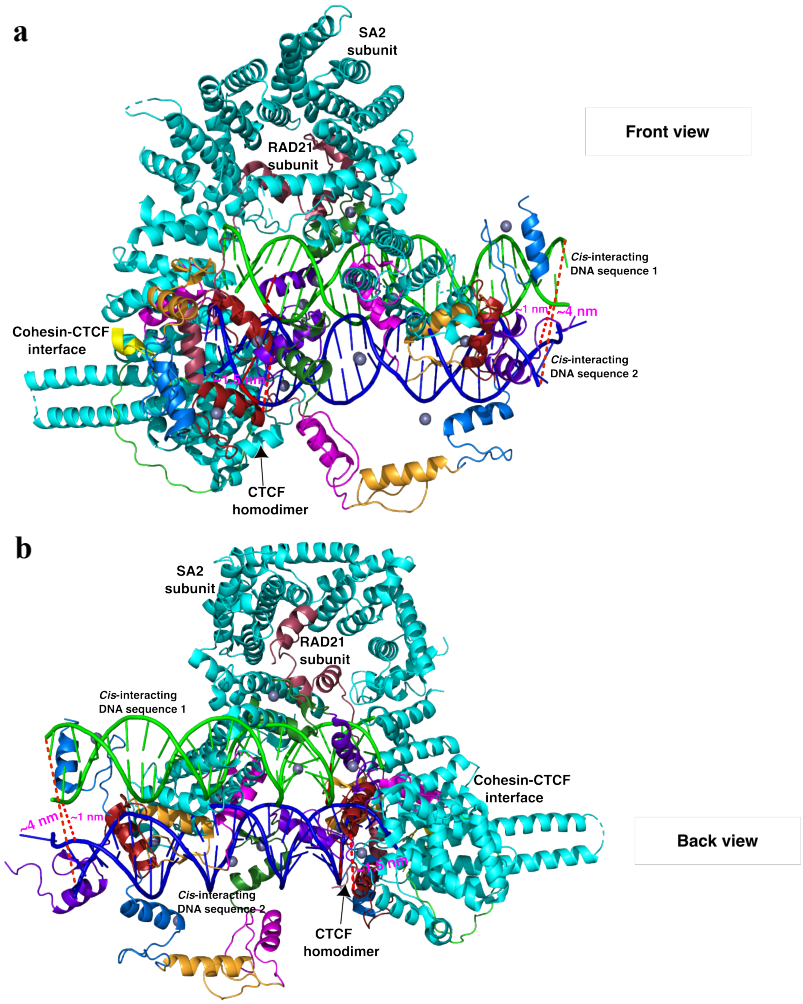


Figure 5.2: Cohesin and CTCF-mediated loop anchor models: Loose (15 Å) homodimerization interface

(a) Composite loop anchor model containing cohesin subunits SA2 (teal) and SCC1/RAD21 (maroon), homodimerized CTCF proteins with N- and C-terminal loop domains removed (multicolored) showing a 15 Å dimerization interface, and two *cis*-interacting DNA sites (dark blue and green). **(b)** This structure is rotated 180 degrees around the vertical axis to depict a view from the back of the loop anchor. For both models, the cohesin-CTCF interface is highlighted in yellow and the CTCF homodimer interface at ZnF 4 (red) is highlighted. Distances between *cis*-interacting DNA sites are shown between positions ~11 bp upstream of the core 19 bp consensus CTCF binding site (upstream of sequence associated with ZnF 9) on the most distant and the most proximal pairings of strand positions, suggesting a specific

range of expected distances between sgRNA binding sites for a DNA biosensor repurposed for chromatin loops of ~1-4 nm.

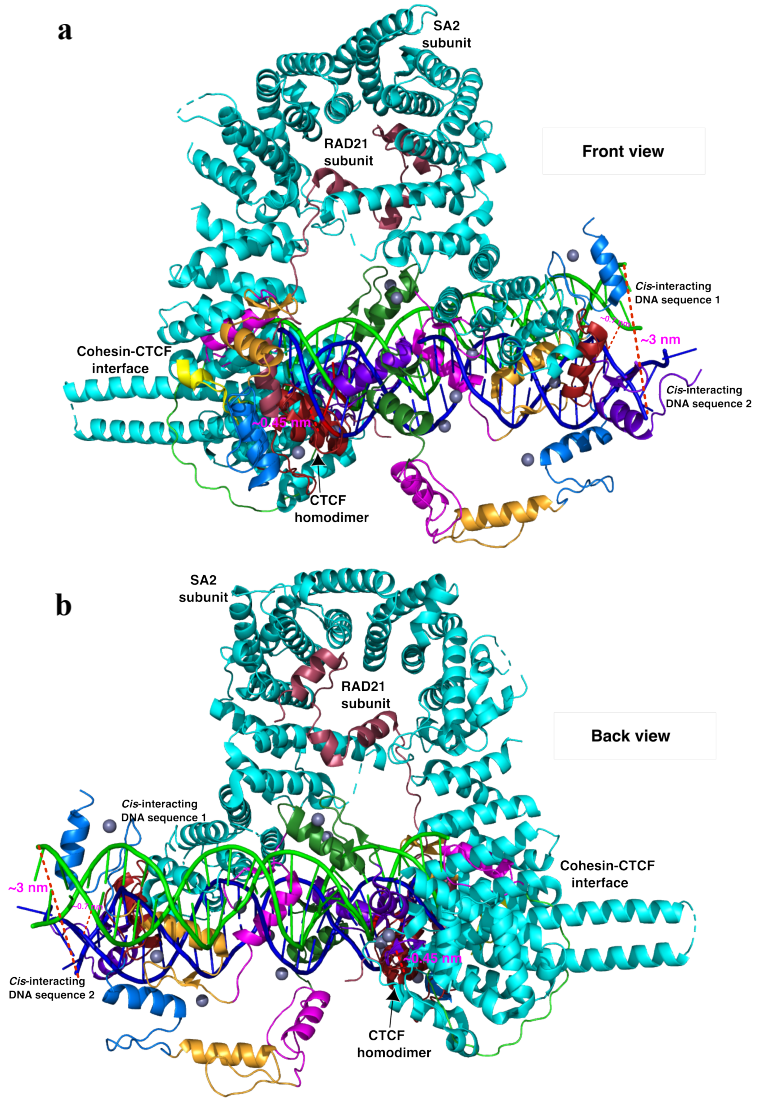


Figure 5.3: Cohesin and CTCF-mediated loop anchor models: Tight (4.5 Å) homodimerization interface

(a) Composite loop anchor model containing cohesin subunits SA2 (teal) and SCC1/RAD21 (maroon), homodimerized CTCF proteins with N- and C-terminal loop domains removed (multicolored) showing a 4.5 Å dimerization interface, and two *cis*-interacting DNA sites (dark blue and green). **(b)** This structure is rotated 180 degrees around the vertical axis to depict a view from the back of the loop anchor. For both models, the cohesin-CTCF interface is highlighted in yellow and the CTCF homodimer interface at

ZnF 4 (red) is highlighted. Distances between *cis*-interacting DNA sites are shown between positions ~11 bp upstream of the core 19 bp consensus CTCF binding site (upstream of sequence associated with ZnF 9) on the most distant and the most proximal pairings of strand positions, suggesting a specific range of expected distances between sgRNA binding sites for a DNA biosensor repurposed for chromatin loops of ~0.7-3 nm.

Biosensing chromatin loops between the *MYC* promoter and cell type-specific super enhancers using a single dCas9 species DNA biosensor

Based on a relatively short distance between *cis*-interacting DNA sequences in our cohesin-CTCF homodimer-anchored chromatin loop model at positions adjacent to the homodimerized CTCF proteins, detection of chromatin loop anchors using DNA biosensors appeared to be feasible. Thus, we first tested the functionality of the single dCas9 species d*Sp*Cas9-NanoBiT DNA biosensor from our previous study[15] for detection of chromatin loops. We sought out a region with well-characterized chromatin interactions in cell lines highly amenable to transient transfection so we could noninvasively deliver our biosensor and rapidly evaluate its sensitivity for chromatin loop anchors. Thus, we chose the chromatin interactions between the *MYC* promoter region and various cell type-specific super enhancer regions that have been thoroughly characterized previously across many common cancer cell lines[26,52-58] as a primary focal region. For our studies, we specifically picked two major interactions between a strongly conserved CTCF binding site upstream of the *MYC* promoter and distal super enhancer regions in HCT116 and K562 cells **(Figure 5.4a)**, shown to be strongly interacting regions in these cell lines from HiChIP and ChIA-PET data sets[26,52,53]. The loop between the *MYC* promoter and a super enhancer region ~1.85 Mb downstream in K562 cells and the loop between the *MYC* promoter and a super enhancer ~0.53 Mb upstream in HCT116 are shown by corner dots on Hi-C contact heatmaps of the *MYC* TAD (hg38 chr8:126837700-129737800) generated from publicly available K562 and HCT116 Hi-C data from previous studies[59,60] **(Figure 5.4b,c)**. The TAD boundary or insulated neighborhood loop is also shown at the

apex of the Hi-C contact map for both cell lines. In addition, we confirmed these loops were present in HCT116 and K562 cells using 4C-seq with a non-blind viewpoint centered on the conserved CTCF binding site upstream of the *MYC* promoter **(Figure S1, Figure S2)**.

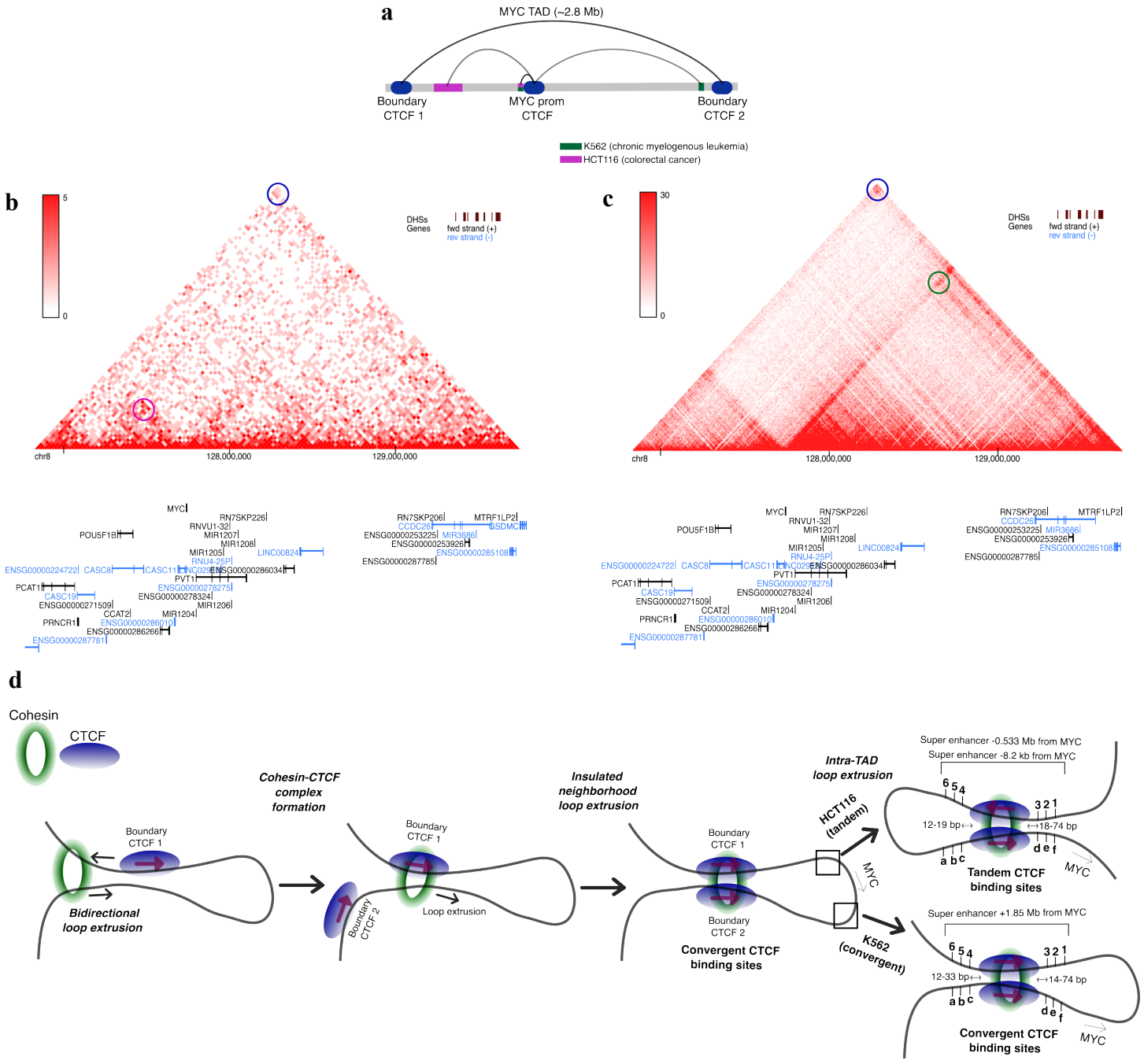


Figure 5.4: *MYC* promoter-super enhancer loops in HCT116 and K562 cells

(a) Cartoon schematic of the loops targeted with both single dCas9 species and dual dCas9 species DNA biosensors. Magenta and green rectangles represent super enhancer regions in HCT116 and K562 cells, respectively. Blue ovals represent all CTCF binding sites targeted that are not located within super enhancer regions. **(b)** Hi-C contact map at 25 kb resolution showing the *MYC* TAD in HCT116 cells. The TAD boundary and super enhancer interactions (corner dots) are highlighted in blue and magenta, respectively. The intensity of each pixel represents the normalized number of contacts between a pair of loci. Maximum intensity is indicated in the scale at right. **(c)** Hi-C contact map at 10 kb resolution showing the *MYC* TAD in K562 cells. The TAD boundary and super enhancer interactions (corner dots) are highlighted in blue and green, respectively. The intensity of each pixel represents the normalized number of contacts between a pair of loci. Maximum intensity is indicated in the scale at right. **(d)** Cartoon representation of the favored extrusion model for the formation of chromatin loops at the *MYC* TAD boundary region and at intra-TAD loops between the *MYC* promoter and super enhancer regions in HCT116 and K562. Representative sgRNA design features for biosensing experiments including 5'-3' positioning and spacing from CTCF binding sites are shown for intra-TAD loop anchor regions (not drawn to scale). sgRNAs were given labels a-f at the *MYC* promoter binding sites and labels 1-6 at the cell type-specific super enhancer binding sites.

From our cohesin and CTCF-mediated loop anchor models, we developed a rational approach for directing specific sgRNA pairs to CTCF homodimer binding sites. This approach was based on expected differences in the presentation of LgBiT and SmBiT transducer elements across chromatin loop anchors of various orientations. Such chromatin loop anchor structures were consistent with the loop extrusion model, where cohesin bidirectionally extrudes a loop until it encounters CTCF bound to its cognate binding site, at which point it keeps extruding the opposite strand from the bound CTCF until a second CTCF binding site is encountered, where, in an event thought to be driven by homodimerization between CTCF proteins dictated by

CTCF binding site orientation, a loop anchor forms[16-18,39,40,49,50]. In our initial sgRNA design process, we considered both the traditional "hairpin" model for the structure of convergent or tandem cohesin-CTCF-anchored loops and the "coil" model proposed for loops with tandemly oriented CTCF binding sites[18,40]. As a general principle for positional pairing of sgRNAs at loop anchors, we realized that based on the "hairpin" model, the 5' sequence of the upstream genomic site should be proximal to the 3' sequence of the downstream genomic site and the 3' sequence of the upstream genomic site should be proximal to the 5' sequence of the downstream genomic site in our loop anchor models. Hence, if sgRNAs were numbered or lettered consecutively from upstream to downstream and spaced approximately evenly from one another, sgRNA 1-6/a-f of the upstream genomic site should logically pair with sgRNA 6-1/f-a of the downstream genomic site when adopting the traditional "hairpin" model for structure of convergent or tandem cohesin-CTCF-anchored loops **(Figure 5.5a-b)**. One notable exception to this principle would be if the "coil" model proposed for tandem sites is considered, which assumes that where CTCF binds cohesin, they must align in parallel[40]. In this model, the 5' sequence of the upstream genomic site should be proximal to the 5' sequence of the downstream genomic site and the 3' sequence of upstream genomic site should be proximal to the 3' sequence of the downstream genomic site. Thus, if sgRNAs were numbered or lettered consecutively from upstream to downstream and spaced approximately evenly from one another, sgRNA 1-6/a-f of the upstream genomic site should logically pair with sgRNA 1-6/a-f of the downstream genomic site when adopting the "coil" model proposed for tandem CTCFs **(Figure 5.6a)**.

Based on known asymmetric binding of CTCF to its cognate binding sites[35], we left a 12-33 bp gap between the 5' edge of the core 19 bp CTCF binding site and the nearest sgRNA binding site for all CTCF sites targeted (see **Extended Experimental Procedures** for sequences). On this side of some CTCF binding sites, there is an additional upstream 10 bp sequence bound by CTCF ZnFs 9-11 separated from the core sequence by a 5-6 bp intervening sequence bound by ZnF 8[35]. The *MYC* promoter CTCF binding site is one such instance of a CTCF binding site with this additional upstream binding module. Thus, we expected a minimum of 16 bp of space

from the 5' side of the core CTCF binding site to be required to minimize interference with native binding of CTCF to such two-module binding sites. However, for one sgRNA, *MYC* promoter sgRNA 3, we left a 12 bp spacing from the 5' edge of the core CTCF binding site. While we were concerned the closer spacing between this sgRNA and CTCF might reduce signal-to-background for its pairing with any super enhancer region sgRNAs due to potential interference with CTCF binding, d*Sa*Cas9 sgRNA PAM sites were highly restrictive in the region surrounding the *MYC* promoter CTCF binding site. We also left a 14-40 bp gap between the 3' edge of the core 19 bp CTCF binding site and the nearest sgRNA binding site for all CTCF sites targeted (see **Extended Experimental Procedures** for sequences). On this side of all CTCF binding sites, accumulated structural information thus far has indicated lack of a specific DNA recognition function for ZnF domains 1-2 near the N-terminus of CTCF[27,28,35]. Thus, we predicted that there may be no need for a spacer sequence between the 3' edge of the CTCF binding site and the nearest sgRNA binding site, but we still included spacer sequences on this side of the CTCF binding site to ensure we did not disrupt native CTCF binding. We also attempted to space each sgRNA apart by a certain distance when possible given limitations in the frequencies of 5'-NGG and 5'-NNGRRT PAM sites within each *cis*-interacting genomic region. While there was one instance when two adjacent sgRNA binding sites were immediately adjacent, all other sgRNA binding sites had spacings of at least 6 bp and at most 109 bp in tandem. We hoped this design rule would minimize competition of sgRNAs for binding sites at one *cis*-interacting genomic region and allow us to assess the relationship between adjacent sgRNA distance from the CTCF binding site and signal-to-background. Finally, we attempted to equalize the distances between sgRNAs in positions where they might be expected to be located across the loop anchor structure with the shortest lateral distance according to either the "coil" or "hairpin" models (i.e. sgRNA 1-2 spacing at the super enhancer approximately equal with sgRNA a-b and sgRNA e-f spacing at the *MYC* promoter for the ~0.53 Mb *MYC* promoter-super enhancer loop in HCT116 cells, which can be modeled as either "coil" or "hairpin" structures). We hoped this design rule would allow us to assess the relationship between distance of sgRNA pairs from the CTCF homodimer and signal-to-background.

In addition to understanding effects of adjacent sgRNA spacing from CTCF binding sites and sgRNA pair distance from the CTCF homodimer on signal-to-background, we were also interested in determining effect of the sgRNA strand polarity on signal-to-background. To be consistent in our analysis and comparisons between conditions, we designed all sgRNAs to bind to the negative strand. For tandemly oriented CTCF binding sites such as the ~0.53 Mb *MYC* promoter-super enhancer loop in HCT116 cells, this principle resulted in both sets of six sgRNAs binding to the opposite strand polarity from that of CTCF, as both CTCF binding sites are on the positive reference strand for this chromatin loop. Conversely, for convergently oriented CTCF binding sites such as the ~1.85 Mb *MYC* promoter-super enhancer loop in K562 cells, this principle resulted in the set of sgRNAs for the upstream interacting genomic site binding to the opposite strand polarity from that of CTCF and the set of sgRNAs for the downstream interacting genomic site binding to the same strand polarity as that of CTCF, as the upstream CTCF binding site is on the positive reference strand and the downstream CTCF binding site is on the negative strand for this chromatin loop.

Thus, we transfected our single dCas9 species DNA biosensor to HCT116 and K562 cells and assessed signal-to-background across 1-f, 2-e, 3-d, 4-c, 5-b, and 6-a sgRNA pairings in addition to the 1-a, 2-b, 3-c, 4-d, 5-e, and 6-f sgRNA pairings for the tandem ~0.53 Mb *MYC* promoter-super enhancer loop in HCT116 cells **(Figure 5.5c, 5.6b)** and assessed signal-to-background across 1-f, 2-e, 3-d, 4-c, 5-b, and 6-a sgRNA pairings for the convergent ~1.85 Mb *MYC* promoter-super enhancer loop in K562 cells **(Figure 5.5d)**. According to our models, we hypothesized that these sgRNA pairings might increase the likelihood of LgBiT and SmBiT transducer elements protruding into the same region of the ~0.7-4 nm gap between *cis*-interacting DNA backbones at loop anchors. Along with these *MYC* promoter-super enhancer sgRNA pairings, we transfected LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT biosensor components and a pMAX-GFP reporter for transfection efficiency in K562 and HCT116 cells and measured fluorescence and luminescence using a multimode plate reader 48 h post-transfection. To establish background luminescence levels, we sought to measure the signal when transducer elements were not proximal because they have not detected a chromatin loop, leading to little

NanoLuc luciferase reassembly. To this end, in our following experiments, we transfected sgRNA pairs to direct our biosensor to bind non-interacting regions of the genome or did not transfect any sgRNAs at all. We observed ranges of signal-to-background of 1.07-fold to 2.45-fold in HCT116 cells and 1.36-fold to 1.88-fold in K562 cells for sgRNA pairings consistent with the shortest distances between *cis*-interacting DNA sequences assuming the "hairpin" model **(Figure 5.5c-d)**. In addition, pairing sgRNA 2 at the super enhancer with sgRNA e at the *MYC* promoter in HCT116 cells and pairing sgRNA 5 at the super enhancer with sgRNA b at the *MYC* promoter in K562 cells resulted in the highest signal-to-background ratios, at 2.45-fold and 1.88-fold, respectively **(Figure 5.5c-d)**. We then observed a range of signal-to-background of 2.01-fold to 2.66-fold in HCT116 cells for sgRNA pairings consistent with the shortest distances between *cis*-interacting DNA sequences assuming the "coil" model **(Figure 5.6b)**. For the "coil" model sgRNA pairings, sgRNA 3 at the super enhancer combined with sgRNA c at the *MYC* promoter in HCT116 cells resulted in the highest signal-to-background ratio, at 2.66-fold **(Figure 5.6b)**.

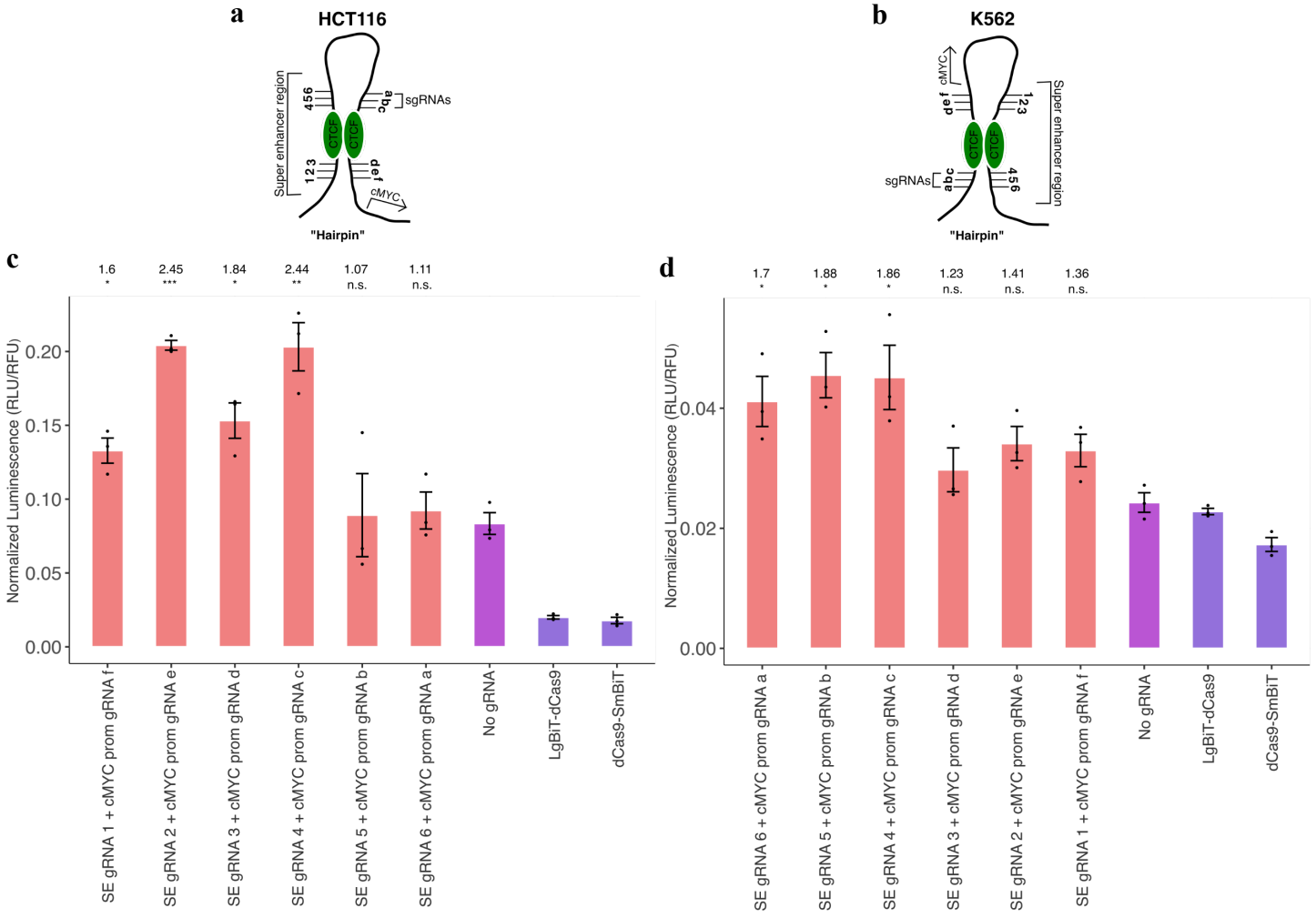


Figure 5.5: Targeting promoter-super enhancer chromatin loops at the *MYC* locus using a single dCas9 species d*Sp*Cas9-NanoBiT DNA biosensor: "Hairpin" loop model

(a) A cartoon representation of the loop anchor formed between a HCT116 cell type-specific super enhancer and the conserved CTCF site at the *MYC* promoter assuming the "hairpin" model for loop structure. **(b)** A cartoon representation of the loop anchor formed between a K562 cell type-specific super enhancer and the conserved CTCF site at the *MYC* promoter assuming the "hairpin" model for loop structure. In **a-b**, sgRNAs were given labels a-f at the *MYC* promoter binding sites and labels 1-6 at the cell type-specific super enhancer binding sites. **(c)** Initial biosensing results in HCT116 cells where six pairs of sgRNAs targeting areas tiling upstream and downstream of a CTCF binding site within a large super enhancer ~0.533 Mb upstream of the *MYC* promoter and a highly conserved CTCF binding site at the

MYC promoter were transfected and luminescence and fluorescence were measured using a multimode plate reader 48 h post-transfection. **(d)** Initial biosensing results in K562 cells where six pairs of sgRNAs targeting areas tiling upstream and downstream of a CTCF binding site within a relatively small super enhancer ~1.85 Mb downstream of the *MYC* promoter and a highly conserved CTCF binding site at the *MYC* promoter were transfected and luminescence and fluorescence were measured using a multimode plate reader 48 h post-transfection. Apparent signal-to-background ratios in **d-e** (comparisons made to background condition where no sgRNA was transfected to cells) are listed in parentheses above each biosensing condition. Data in **d-e** are presented as the mean ± s.e.m., $n = 3$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

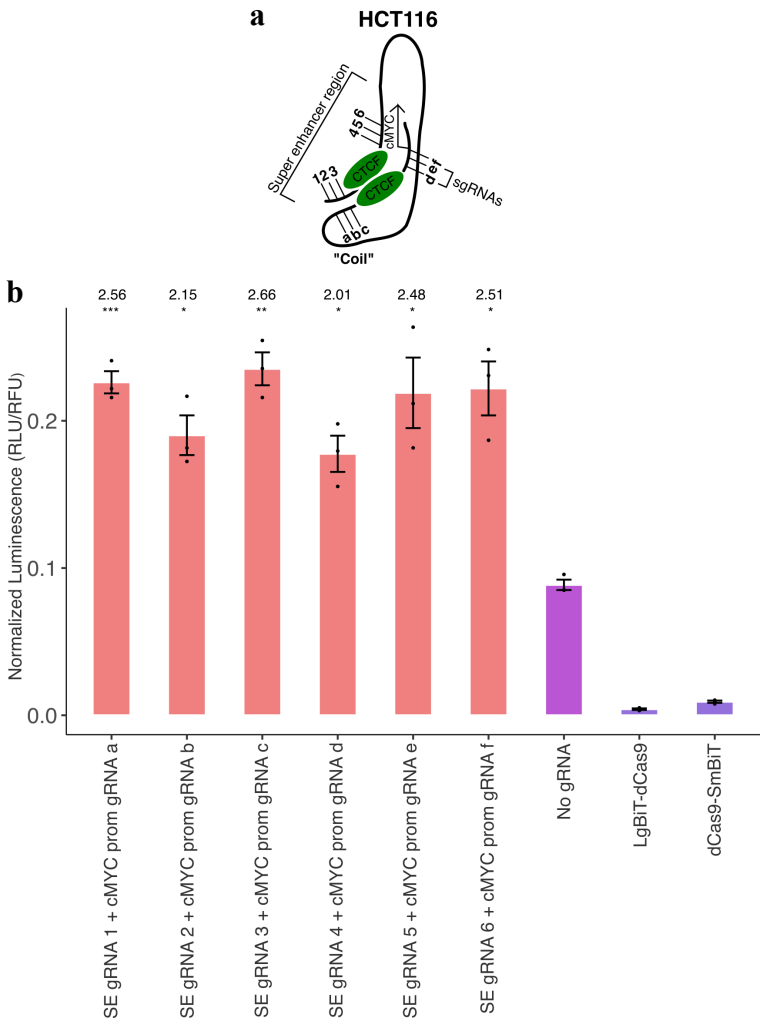


Figure 5.6: Targeting promoter-super enhancer chromatin loops at the *MYC* locus using a single dCas9 species d*Sp*Cas9-NanoBiT DNA biosensor: "Coil" loop model

(a) A cartoon representation of the loop anchor formed between a HCT116 cell type-specific super enhancer and the conserved CTCF site at the *MYC* promoter assuming the "coil" model for loops anchored by tandemly oriented CTCF binding sites. **(b)** Initial biosensing results in HCT116 cells where six pairs of sgRNAs targeting areas tiling upstream and downstream of a CTCF binding site within a large super enhancer ~0.533 Mb upstream of the *MYC* promoter and a highly conserved CTCF binding site at the *MYC* promoter were transfected and luminescence and fluorescence were measured using a multimode plate reader 48 h post-transfection. Apparent signal-to-background ratios (comparisons made to background condition where no sgRNA was transfected to cells) are listed in parentheses above each

biosensing condition. Data are presented as the mean ± s.e.m., $n = 3$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

Biosensing chromatin loops between the *MYC* promoter and cell type-specific super enhancers using a dual dCas9 species DNA biosensor

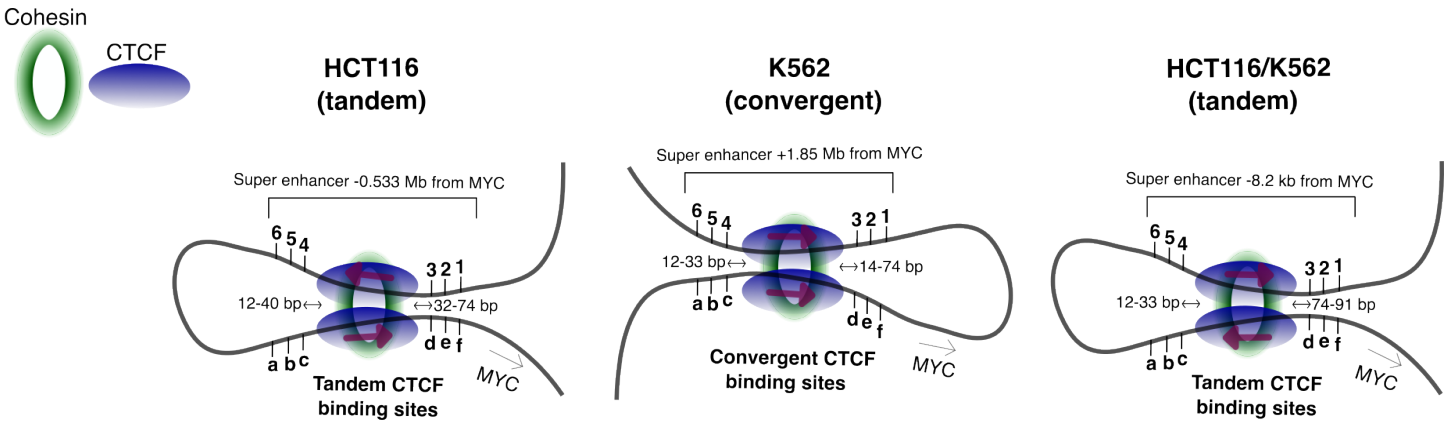
In earlier experiments, we observed higher sensitivity for our dual dCas9 species DNA biosensor at endogenous *MUC4* compared to our single dCas9 species DNA biosensor. Thus, we next applied the dual species LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor developed earlier to reduce unproductive auto-assembly of the biosensor in the nucleus to target the same *MYC* promoter-super enhancer chromatin loops in HCT116 and K562 cells **(Figure 5.4d)**. Like our experiments testing our single dCas9 species DNA biosensor, we transfected our dual dCas9 species DNA biosensor components along with all possible combinations of six d*Sa*Cas9 sgRNAs and six d*Sp*Cas9 sgRNAs targeting promoter and enhancer loci at locations we hypothesized to be in close proximity at loop anchors according to the "hairpin" or "coil" loop anchor structures consistent with the loop extrusion model. We observed a range of signal-to-background between 3.32-fold and 4.96-fold when we directed our dual dCas9 species DNA biosensor to bind the ~0.53 Mb *MYC* promoter-super enhancer chromatin loop in HCT116 cells **(Figure 5.7a)**. When we directed our dual dCas9 species DNA biosensor to bind the ~1.85 Mb *MYC* promoter-super enhancer chromatin loop in K562 cells, we observed a range of signal-to-background of 1.08-fold to 6.08-fold **(Figure 5.7b)**. In addition, pairing sgRNA 5 at the super enhancer with sgRNA b at the *MYC* promoter in HCT116 cells and pairing sgRNA 5 at the super enhancer with sgRNA a at the *MYC* promoter in K562 cells resulted in the highest signal-to-background ratios, at 4.96-fold and 6.08-fold, respectively **(Figure 5.7a-b)**.

There is a well-known negative relationship between frequency of *cis* contacts formed between two genomic sites and distance between them[14,59,60]. Since the other chromatin loops targeted were fairly long-range *cis* interactions at between ~0.53 Mb and ~1.85 Mb away from the conserved *MYC* promoter CTCF, we designed another set of sgRNAs for a relatively small super enhancer much closer to the conserved *MYC* promoter CTCF. This small super enhancer is located ~8.2 kb upstream of the *MYC* promoter CTCF and the local *cis* interaction between this super enhancer and the *MYC* promoter CTCF is common to both HCT116 and K562 cells. Thus, we transfected our dual dCas9 species DNA biosensor components along with a set of six d*Sp*Cas9 and six d*Sa*Cas9 sgRNAs targeting this ~8.2 kb *cis* interaction in HCT116 cells with all pairwise combinations of sgRNAs. We observed a range of signal-to-background between 2.13-fold and 5.01-fold when we directed our dual dCas9 species DNA biosensor to bind chromatin loops between the ~8.2 kb super enhancer and the *MYC* promoter in HCT116 cells **(Figure 5.7c)**. Since this closer super enhancer is unique to both HCT116 and K562 cells, we applied a set of 12 sgRNA pairs which produced highest signal-to-background ratios in HCT116 cells in biosensing experiments targeting the same *cis* interaction in K562 cells. We observed a range of signal-to-background of 1.74-fold to 5.86-fold when we directed our dual dCas9 species DNA biosensor to bind the same ~8.2 kb chromatin loop in K562 cells **(Figure 5.7d)**. For this ~8.2 kb *cis* interaction in both cell lines, signal-to-background was highest when sgRNA 1 at the super enhancer was paired with sgRNA a at the *MYC* promoter, at 5.01-fold in HCT116 cells and 5.86-fold in K562 cells **(Figure 5.7c-d)**.

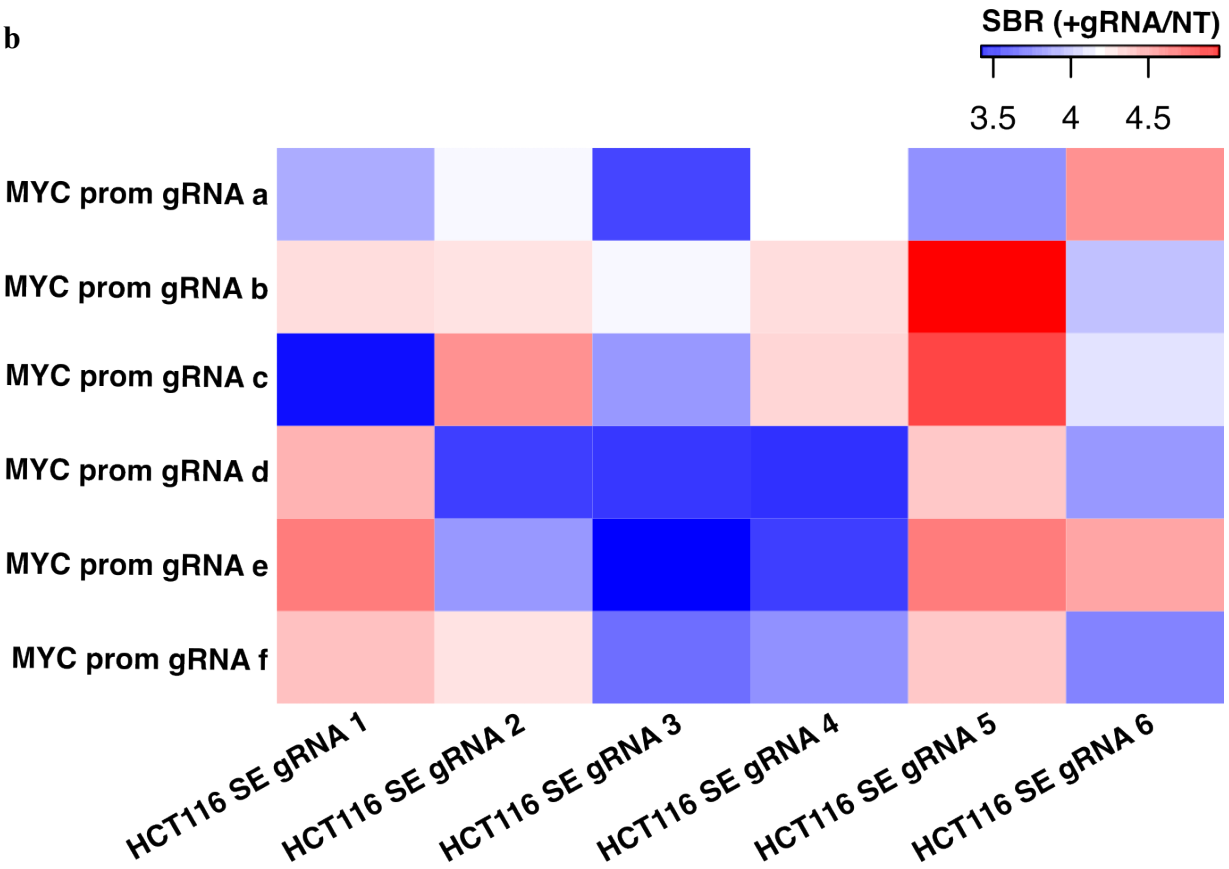
In addition to measuring luminescence from our dual dCas9 species DNA biosensor using a luminometer, we used the Andor Dragonfly 200 Multi-modal Confocal System to gather high resolution images of HCT116 cells after transfection with our biosensor and sgRNAs targeting the ~0.53 Mb *MYC* promoter-super enhancer intra-TAD loop in HCT116 cells. We compared hundreds of individual nuclear luminescence signals from this on-target condition to individual nuclear signals from a non-interacting background condition where a sgRNA pair between *PALB2* on chromosome 16 and the *MYC* promoter on chromosome 8 was transfected, from a background condition where no sgRNAs were transfected, and from a positive control condition where an

equimolar amount of a NanoLuc-d*Sp*Cas9 fusion protein with sgRNAs to bind each known *cis*-interacting region individually was transfected to establish expected signal output if the enzyme were to be reassembled perfectly at the loop anchors **(Figure 5.8a-c)**. For our on-target condition, we used the pair of sgRNAs from our luminometer-based intra-TAD chromatin loop biosensing experiments which produced the highest signal-to-background for this particular chromatin loop in HCT116 cells, super enhancer sgRNA 5 and *MYC* promoter sgRNA b. The nuclei of many cells transfected with our biosensor showed bright luminescent punctae while cells transfected with non-interacting sgRNA pairs showed no luminescent punctae and diffuse signal in the nucleus **(Figure 5.8a-b)**. Analysis of individual nuclear signals showed a signal-to-background ratio of 3.71-fold for the pairing of super enhancer sgRNA 5 and *MYC* promoter sgRNA b compared to our non-targeting background condition **(Figure 5.8c)**. In addition, comparison of our theoretical maximum biosensing signal positive control and the biosensing signal in the on-target condition for this ~0.53 Mb intra-TAD chromatin loop in HCT116 cells revealed ~57% of maximum theoretical signal output for our biosensor.

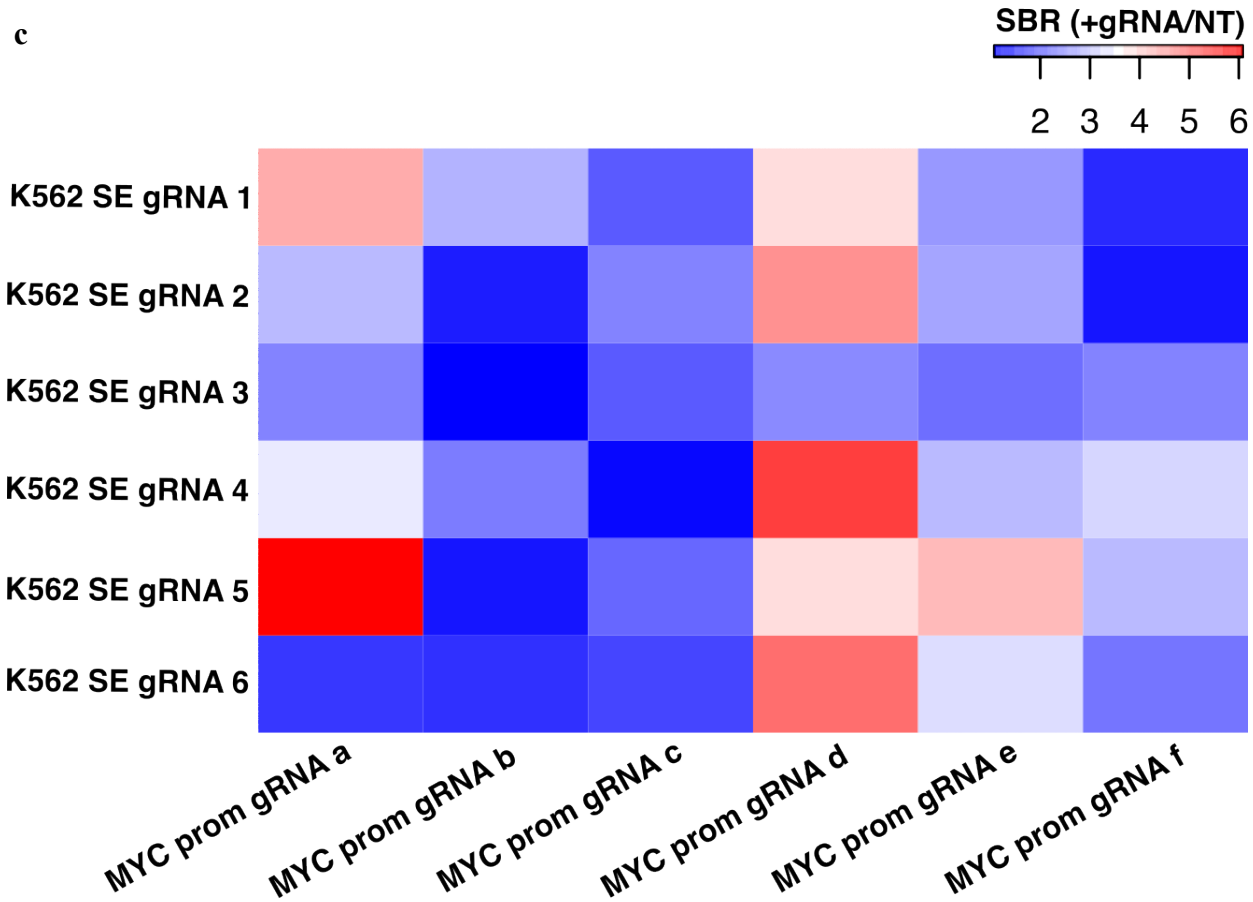
a



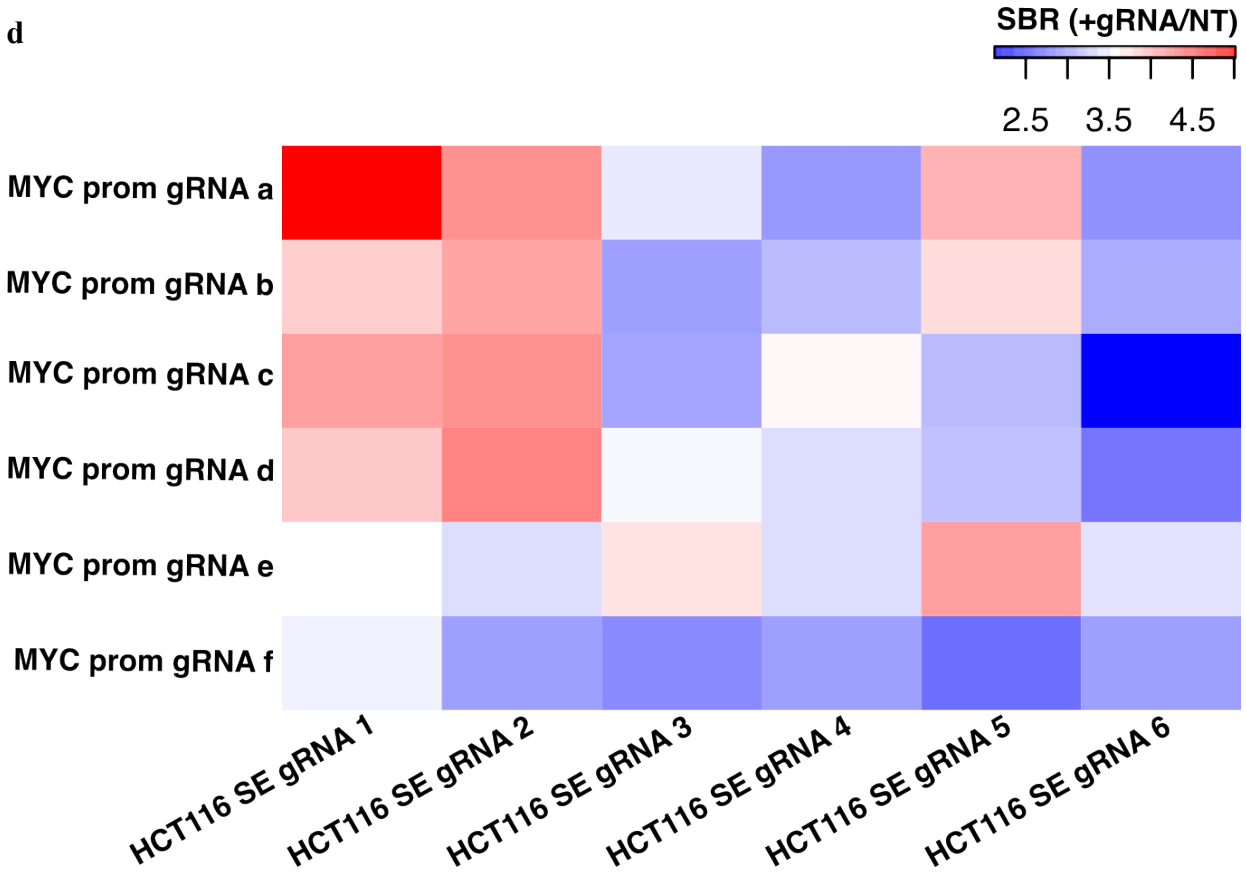
b


SBR (+gRNA/NT)
3.5
4
4.5
MYC prom gRNA a
MYC prom gRNA b
MYC prom gRNA c
MYC prom gRNA d
MYC prom gRNA e
MYC prom gRNA f
HCT116 SE gRNA 1
HCT116 SE gRNA 2
HCT116 SE gRNA 3
HCT116 SE gRNA 4
HCT116 SE gRNA 5
HCT116 SE gRNA 6

c


SBR (+gRNA/NT)
2
3
4
5
6
K562 SE gRNA 1
K562 SE gRNA 2
K562 SE gRNA 3
K562 SE gRNA 4
K562 SE gRNA 5
K562 SE gRNA 6
MYC prom gRNA a
MYC prom gRNA b
MYC prom gRNA c
MYC prom gRNA d
MYC prom gRNA e
MYC prom gRNA f

d


SBR (+gRNA/NT)
2.5
3.5
4.5
MYC prom gRNA a
MYC prom gRNA b
MYC prom gRNA c
MYC prom gRNA d
MYC prom gRNA e
MYC prom gRNA f
HCT116 SE gRNA 1
HCT116 SE gRNA 2
HCT116 SE gRNA 3
HCT116 SE gRNA 4
HCT116 SE gRNA 5
HCT116 SE gRNA 6

e

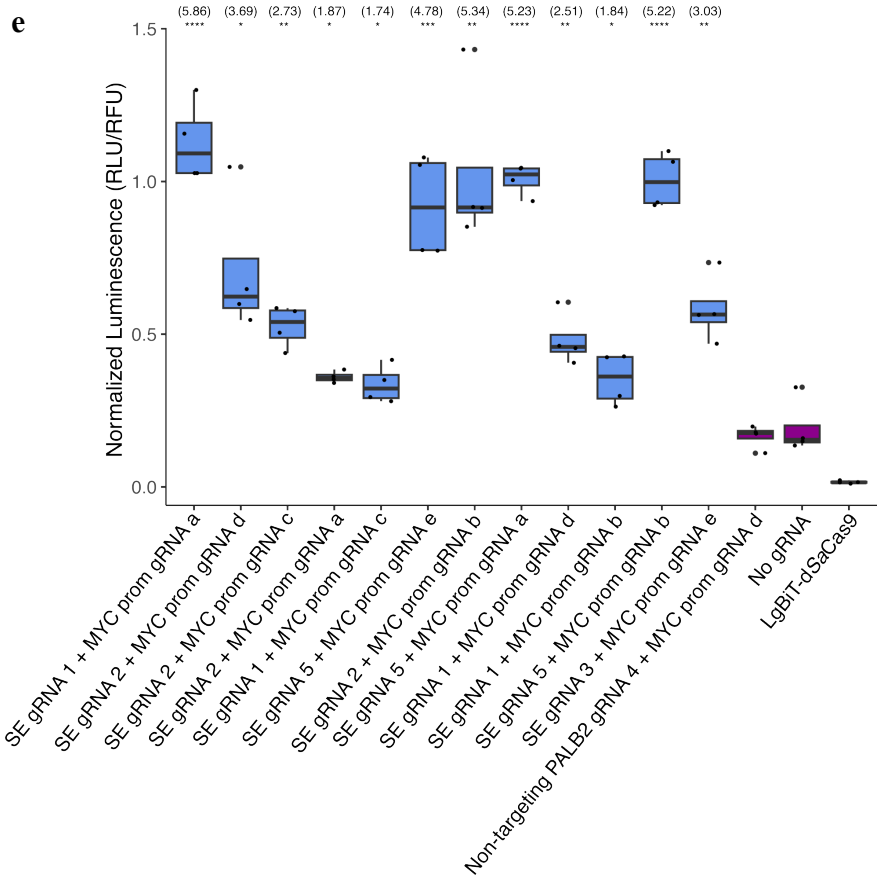
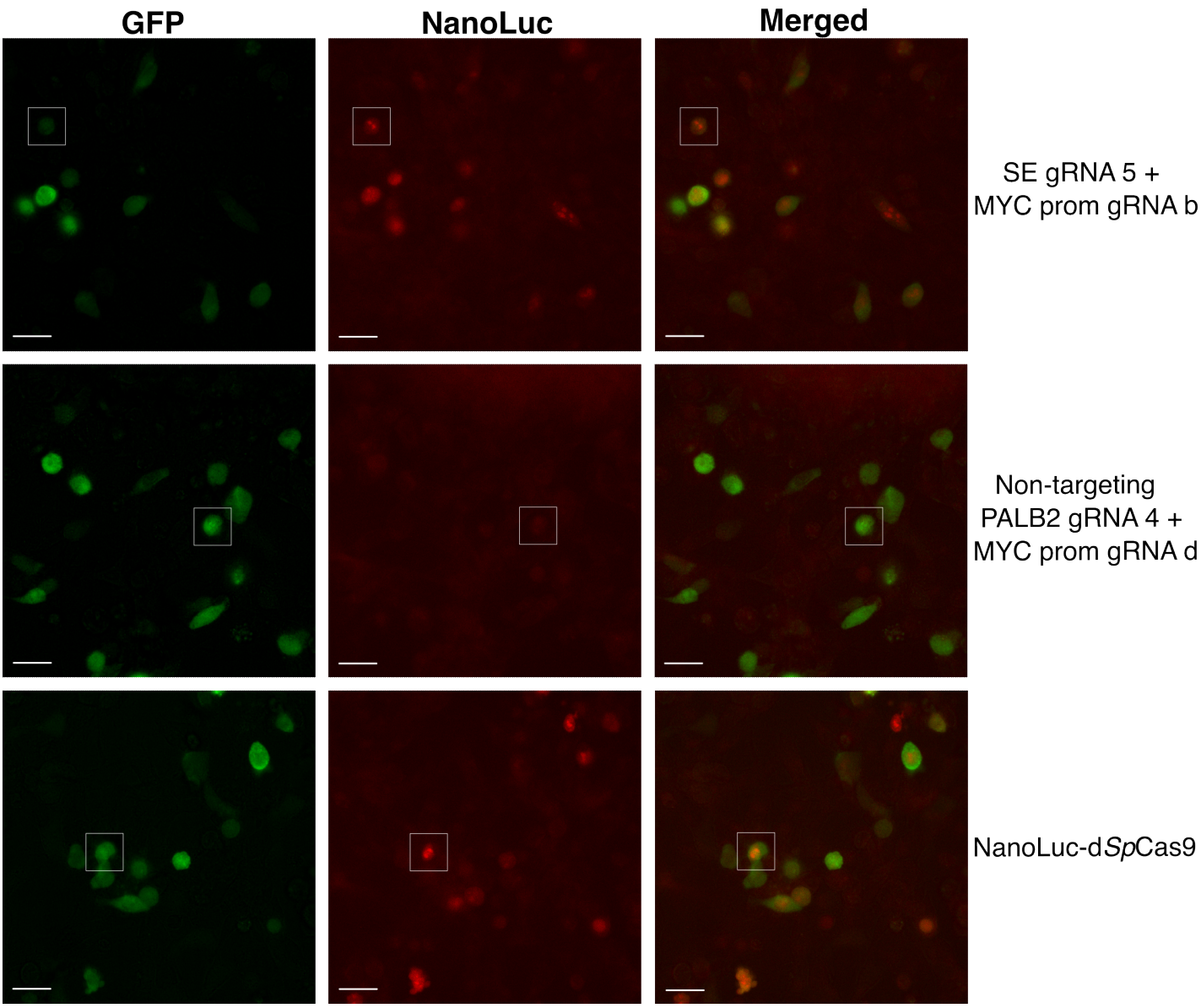

(5.86) (3.69) (2.73) (1.87) (1.74) (4.78) (5.34) (5.23) (2.51) (1.84) (5.22) (3.03)
Normalized Luminescence (RLU/RFU)
1.5
1.0
0.5
0.0
SE gRNA 1 + MYC prom gRNA a
SE gRNA 2 + MYC prom gRNA d
SE gRNA 2 + MYC prom gRNA c
SE gRNA 2 + MYC prom gRNA a
SE gRNA 1 + MYC prom gRNA c
SE gRNA 5 + MYC prom gRNA e
SE gRNA 2 + MYC prom gRNA b
SE gRNA 5 + MYC prom gRNA a
SE gRNA 1 + MYC prom gRNA d
SE gRNA 5 + MYC prom gRNA b
SE gRNA 1 + MYC prom gRNA b
SE gRNA 5 + MYC prom gRNA e
SE gRNA 3 + MYC prom gRNA d
Non-targeting PALB2 gRNA 4 + MYC prom gRNA e
No gRNA
LgBiT-dSaCas9

Figure 5.7: Targeting promoter-super enhancer chromatin loops at the *MYC* locus using a dual dCas9 species LgBiT-d*Sa*Cas9 + d*Sp*Cas9-SmBiT DNA biosensor

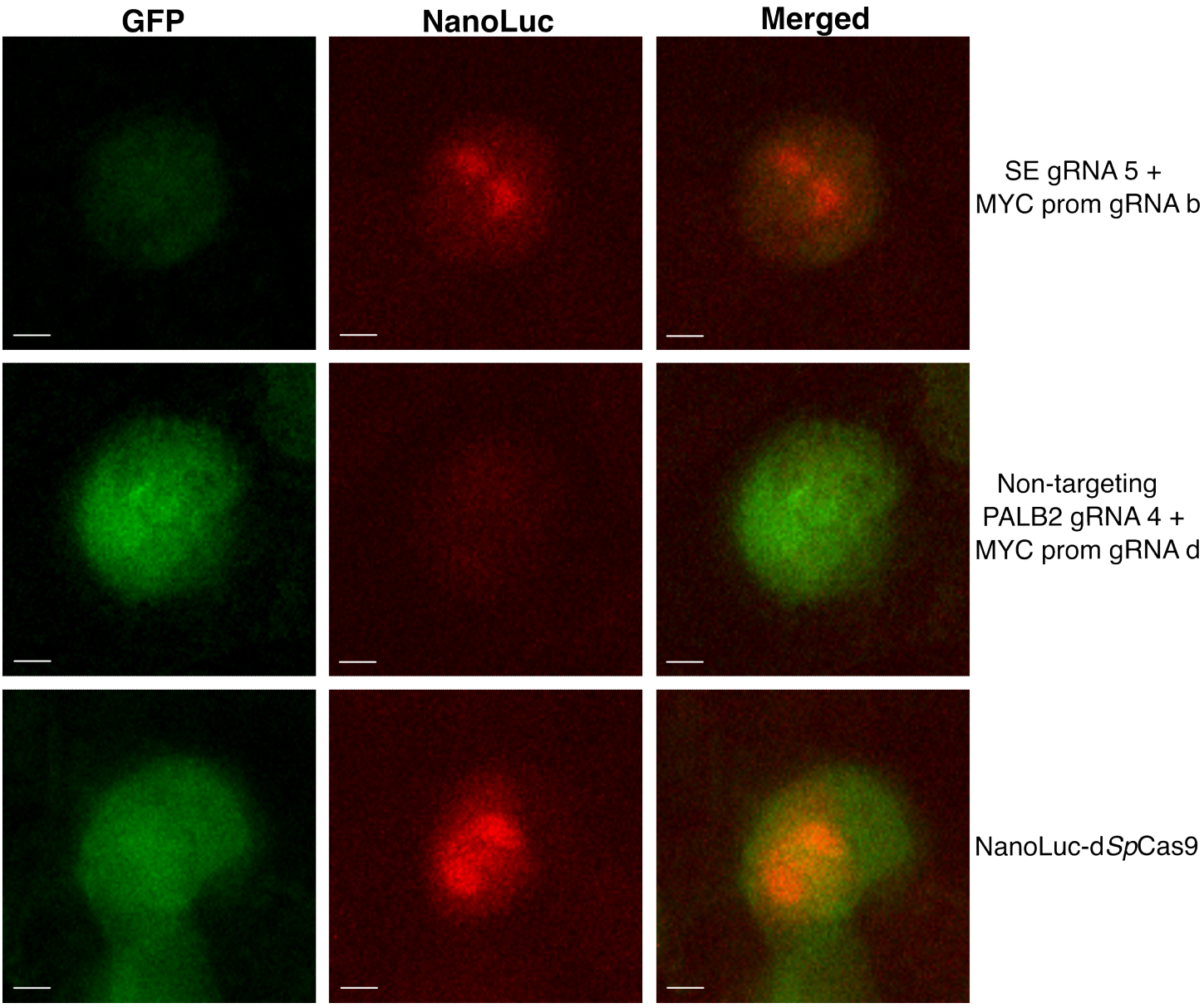
(a) Cartoon representation of intra-TAD chromatin loops between the *MYC* promoter and super enhancer regions in HCT116 and K562. Representative sgRNA design features for biosensing experiments including 5'-3' positioning and spacing from CTCF binding sites are shown for each of the loop anchor regions (not drawn to scale). sgRNAs were given labels a-f at the *MYC* promoter binding sites and labels 1-6 at the cell type-specific super enhancer binding sites. **(b)** Heatmap representing chromatin loop biosensing results in HCT116 cells where 36 pairs of sgRNAs tiling along a CTCF binding site within a large super enhancer ~0.533 Mb upstream of the *MYC* promoter and tiling along a highly conserved CTCF binding site at the *MYC* promoter were transfected and luminescence and fluorescence were measured using a multimode plate reader 24 h post-transfection. **(c)** Heatmap representing chromatin loop biosensing results in K562 where 36 pairs of sgRNAs tiling along a CTCF binding site within a relatively small super enhancer ~1.85 Mb downstream of the *MYC* promoter and tiling along a highly conserved CTCF binding site at the *MYC* promoter were transfected and luminescence and fluorescence were measured using a multimode plate reader 24 h post-transfection. **(d)** Heatmap representing chromatin loop biosensing results in HCT116 cells where 36 pairs of sgRNAs tiling along a CTCF binding site within a relatively small super enhancer ~8.2 kb upstream of the *MYC* promoter and tiling along a highly conserved CTCF binding site at the *MYC* promoter were transfected and luminescence and fluorescence were measured using a multimode plate reader 24 h post-transfection. Apparent signal-to-background ratios in **b-d** (comparisons made to non-targeting (NT) background condition where sgRNAs targeting non-interacting genomic regions selected based on public Hi-C data sets were transfected) are shown for each sgRNA pair, $n = 4$, where n represents the number of independent experimental technical replicates included in parallel. For **b-d**, individual signal-to-background ratio (SBR) scales are shown at top right for each heatmap. **(e)** Biosensing results in K562 cells where 12

sgRNA pairs with the highest signal-to-background ratios in HCT116 cells for the ~8.2 kb *MYC* promoter-super enhancer chromatin loop were targeted to the same loop anchor region. These 12 sgRNA pairs were transfected and luminescence and fluorescence were again measured using a multimode plate reader 24 h post-transfection. Apparent signal-to-background ratios in **e** are listed above each sgRNA pair. Data in **e** are presented such that boxes show the median and interquartile range (IQR) and whiskers show dispersion from the IQR that is equal to the lesser of the 1st or 3rd quartiles plus or minus 1.5xIQR or the distance from the 1st or 3rd quartiles to the minimum or maximum points. Data in **e** are presented as the mean ± s.e.m., $n = 4$, where n represents the number of independent experimental technical replicates included in parallel; unpaired two-sided Student′s t-test, *$p < 0.05$; **p < 0.01; ***$p < 0.001$; ****$p < 0.0001$.

a


GFP
NanoLuc
Merged
SE gRNA 5 +
MYC prom gRNA b
Non-targeting
PALB2 gRNA 4 +
MYC prom gRNA d
NanoLuc-d*Sp*Cas9

b


GFP
NanoLuc
Merged
SE gRNA 5 +
MYC prom gRNA b
Non-targeting
PALB2 gRNA 4 +
MYC prom gRNA d
NanoLuc-d*Sp*Cas9

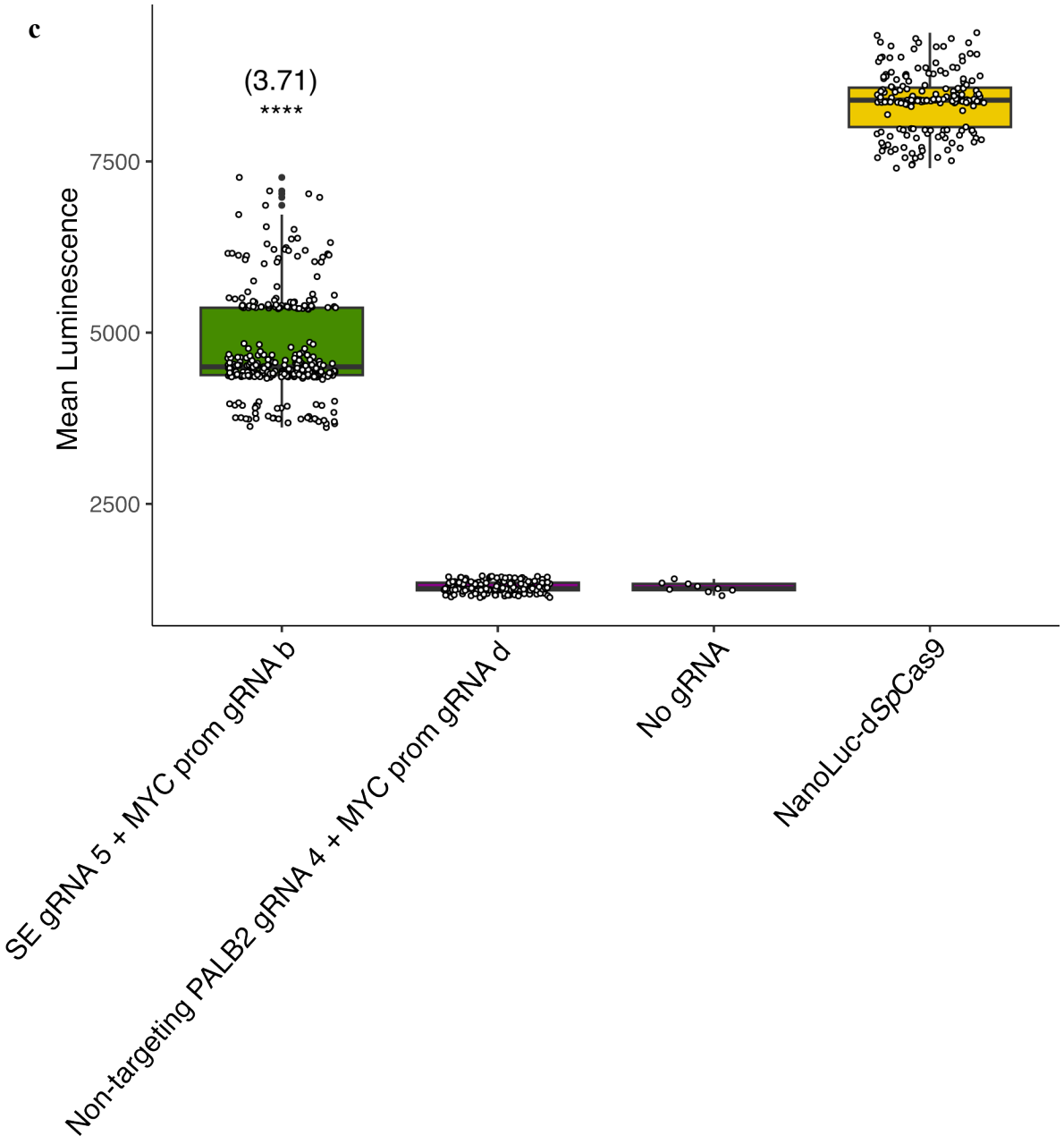


Figure 5.8: Imaging *MYC* promoter-super enhancer loops in live cells using low-light luminescence microscopy

(a) Individual GFP fluorescence (green), NanoLuc luminescence (red), and merged images taken on the Andor Dragonfly 200 Multi-modal Confocal System at 63X magnification depicting a biosensing experiment targeting the intra-TAD loop between the *MYC* promoter region and the super enhancer region ~0.53 Mb upstream from the *MYC* promoter in HCT116 cells. Super enhancer sgRNA 5 paired with *MYC* promoter sgRNA b is compared to a non-targeting sgRNA control (known non-interacting

genomic region—*PALB2* region and *MYC* promoter region from ENCODE Hi-C data) and a NanoLuc-d*Sp*Cas9 control. Scale bars = 25 μM. **(b)** 25 μM x 25 μM white panels from individual GFP fluorescence (green), NanoLuc luminescence (red), and merged images in **a** magnified. Scale bars = 3 μM. **(c)** Signal quantification (36 h post-transfection) for luminescence and fluorescence images taken of the dual dCas9 species DNA biosensor directed to bind at known interacting regions within a large super enhancer ~0.53 Mb upstream of the *MYC* promoter and at a highly conserved CTCF binding site at the *MYC* promoter. Apparent signal-to-background ratio for the ~0.53 Mb super enhancer sgRNA 5 + *MYC* promoter b sgRNA pair is listed in parentheses. For ~0.53 Mb super enhancer sgRNA 5 + *MYC* promoter sgRNA b, non-targeting sgRNA, no sgRNA, and NanoLuc-d*Sp*Cas9 conditions, $n = 325$, $n = 143$, $n = 9$, and $n = 193$ unique nuclei quantified, respectively. Comparisons between group means made using unpaired two-sided Student′s *t*-test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$. Boxes show the median and interquartile range (IQR) and whiskers show dispersion from the IQR that is equal to the lesser of the 1st or 3rd quartiles plus or minus 1.5xIQR or the distance from the 1st or 3rd quartiles to the minimum or maximum points.

Biosensing insulated neighborhood chromatin loops at the *MYC* TAD boundary region using a dual dCas9 species DNA biosensor

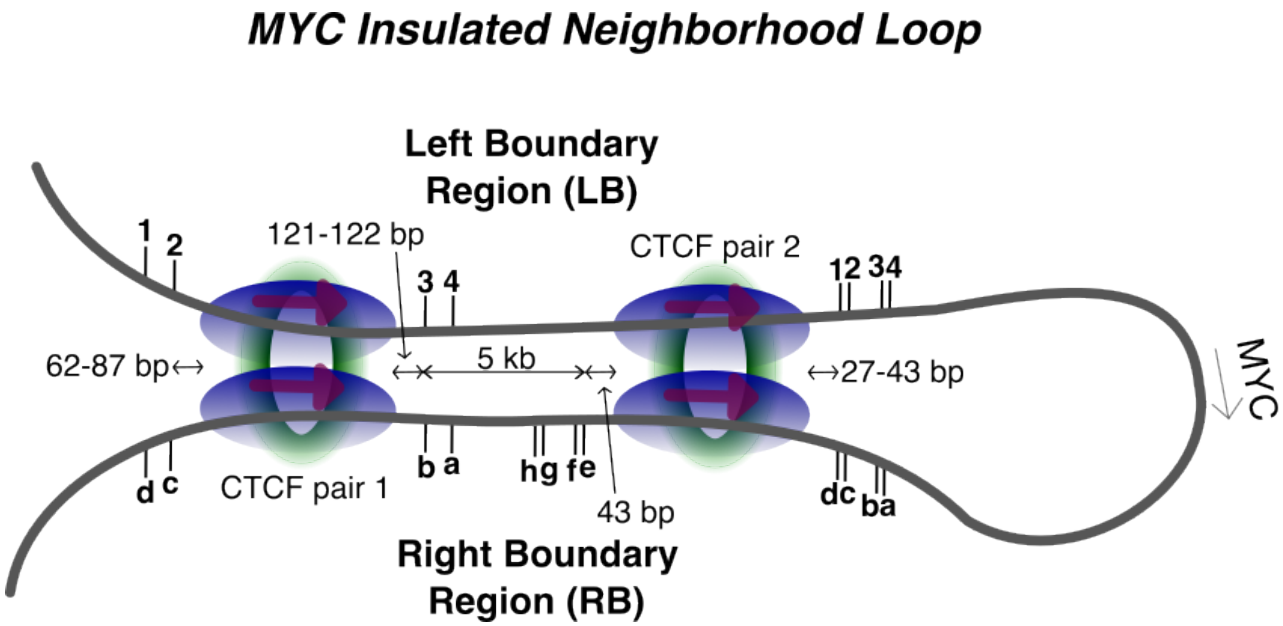
As TAD boundaries are enriched in convergently oriented CTCF binding sites[61,62], often representing the strongest chromatin contact regions[19,63,64] that insulate super enhancer regions from surrounding TADs[65], we wanted to examine whether strong TAD boundary loops with convergently oriented CTCF proteins could be detected at a higher level above background than convergently oriented intra-TAD loops using our dual dCas9 species DNA biosensor. Thus, we designed a set of four d*Sp*Cas9 and d*Sa*Cas9 sgRNAs for each of two convergent TAD boundary region CTCF pairs separated by ~2.8 Mb **(Figure 5.9a)**. These two CTCF pairs anchor an insulated neighborhood (hg38 chr8:126,837,755-129,737,754) that contains *MYC* and all interacting

super enhancers from earlier experiments. To get an idea of contact strength for these *MYC* TAD boundary interactions, we analyzed publicly available ENCODE CTCF ChIA-PET and CTCF ChIP-seq data for both convergent CTCF pairs[40,66,67]. The upstream TAD boundary interaction (hg38 chr8:126,869,146-129,682,463) has a ChIA-PET interaction score of 300. Within this upstream TAD boundary interaction, the upstream CTCF has peak raw ChIA-PET and peak raw ChIP-seq signals of 26 and 51, respectively, and the downstream CTCF has peak raw ChIA-PET and peak raw ChIP-seq signals of 139 and 92, respectively **(Figure 5.9b-c)**. Similarly, the downstream TAD boundary interaction (hg38 chr8:126,876,320-129,688,015) has a ChIA-PET score of 200. Within this downstream TAD boundary interaction, the upstream CTCF has peak raw ChIA-PET and peak raw ChIP-seq signals of 119 and 195, respectively, and the downstream CTCF has peak raw ChIA-PET and peak raw ChIP-seq signals of 121 and 169, respectively **(Figure 5.9b-c)**. Thus, these loop anchors represent strongly interacting regions defining the boundary of the *MYC* TAD within the human genome.

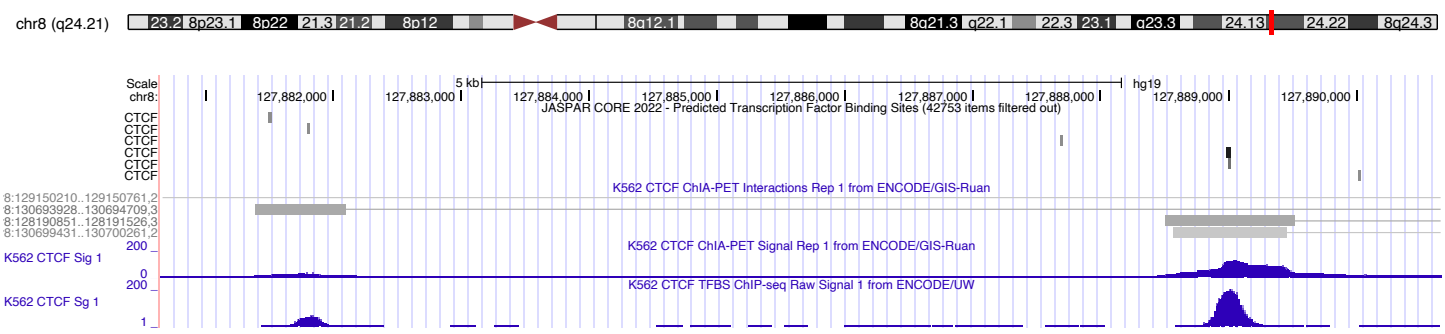
In earlier experiments, we only tested sgRNAs bound to a single strand (-). However, in testing our dual dCas9 species DNA biosensor at the *MYC* TAD boundary loop anchors, we designed both sets of d*Sp*Cas9 and d*Sa*Cas9 sgRNAs with combinations of (+) and (-) strand polarities. When we directed our dual dCas9 species DNA biosensor to bind the downstream TAD boundary chromatin loop in K562, we combined pairings of sgRNAs with approximately equal spacing from the CTCF binding site as much as possible given PAM site frequency constraints in this region (**Figure 5.9a)**. In addition, we tested combinations of equally spaced sgRNAs on both strands and on either side of the loop anchor **(Figure 5.9a)**. We observed a range of signal-to-background of 3.26-fold to 8.99-fold when we directed our dual dCas9 species DNA biosensor to bind the upstream TAD boundary chromatin loop in K562 **(Figure 5.9d)**. In addition, signal-to-background was highest for the upstream TAD boundary interaction in K562 when sgRNA 1 at the left boundary region (- strand, -153 bp from CTCF) was paired with sgRNA c at the right boundary region (- strand, +62 bp from CTCF), at 8.99-fold **(Figure 5.9d)**. For the innermost pairings of sgRNAs at the downstream TAD boundary chromatin loop anchor in K562, we observed a range of signal-to-background of 4.93-fold to 10.59-fold **(Figure 5.9e)**. In

addition, signal-to-background was highest for the innermost combinations of sgRNAs at the downstream TAD boundary interaction in K562 when sgRNA 2 at the left boundary region (+ strand, +49 bp from CTCF) was paired with sgRNA e at the right boundary region (+ strand, +41 bp from CTCF), at 10.59-fold **(Figure 5.9e)**. For the outermost pairings of sgRNAs at the downstream TAD boundary chromatin loop anchor in K562, we observed a range of signal-to-background of 3.12-fold to 8.33-fold **(Figure 5.9f)**. Finally, for the outermost combinations of sgRNAs at the downstream TAD boundary interaction in K562, signal-to-background was highest when sgRNA 3 at the left boundary region (+ strand, +179 bp from CTCF) was paired with sgRNA g at the right boundary region (+ strand, +166 bp from CTCF), at 8.33-fold **(Figure 5.9f)**.

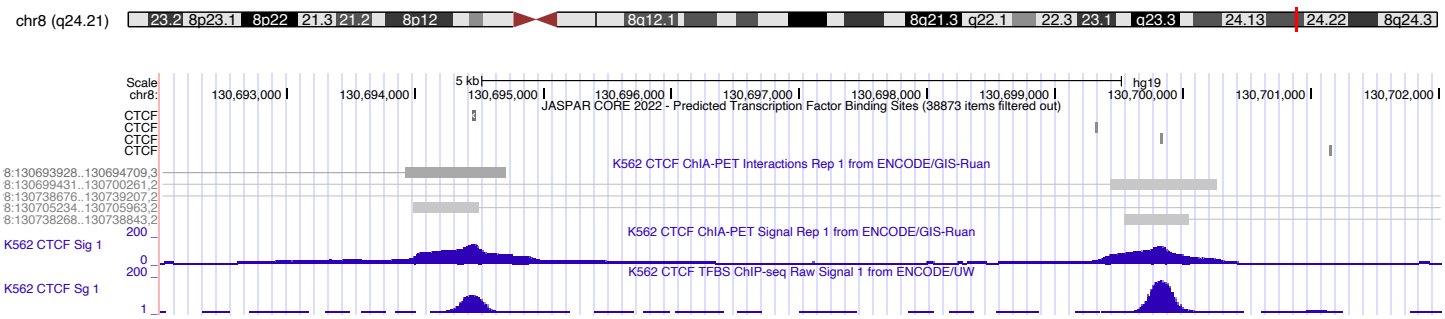
a



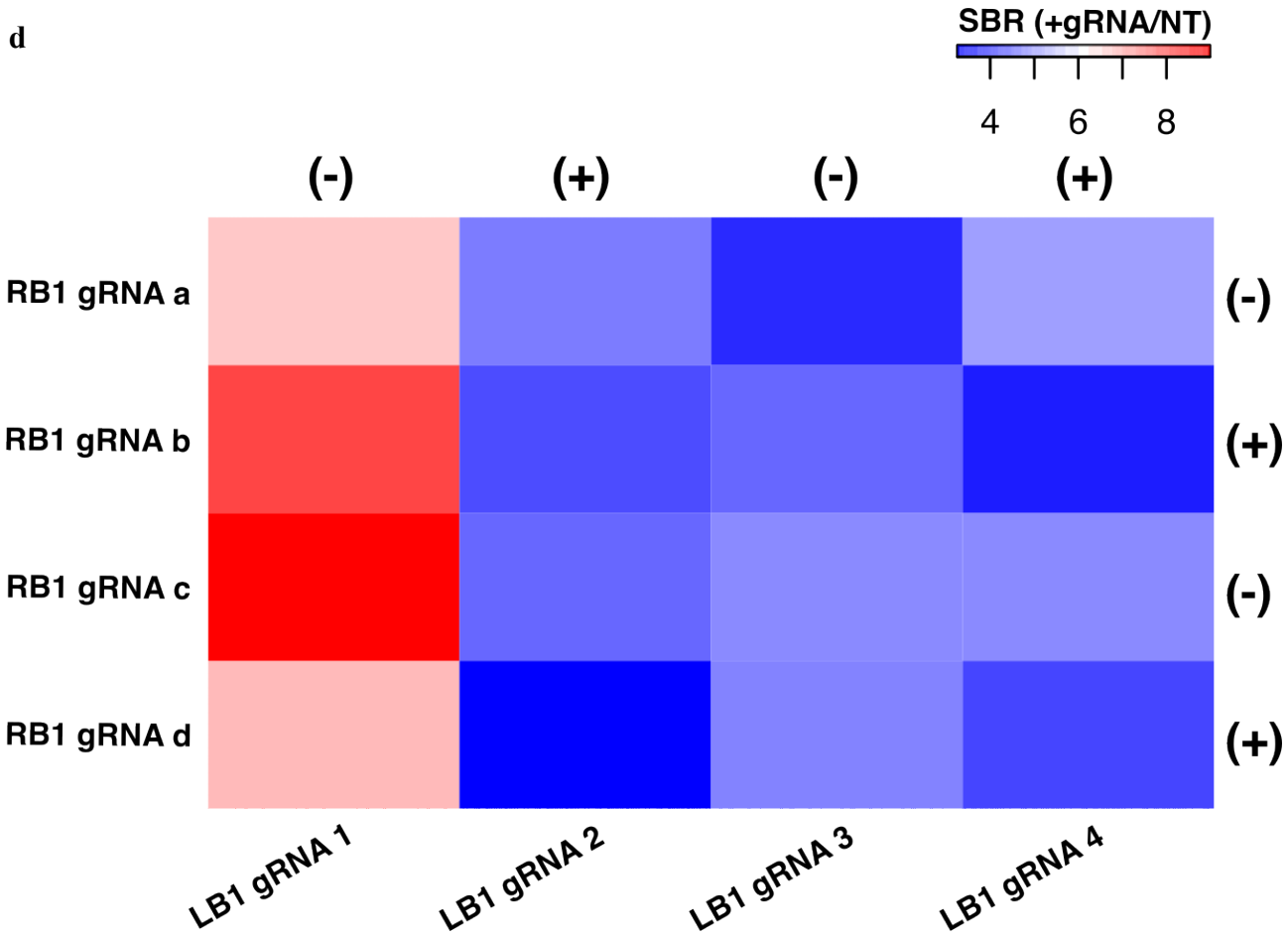
b



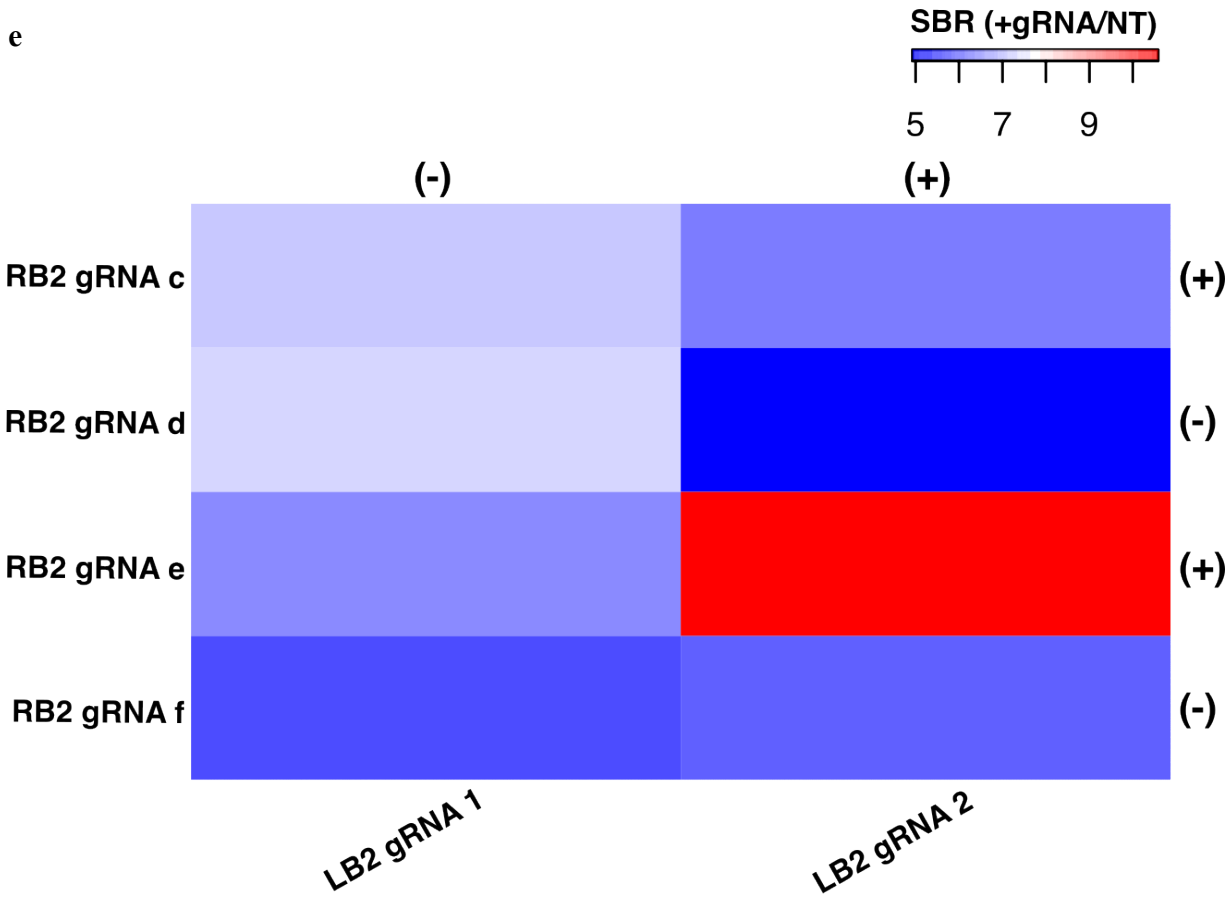
c


chr8 (q24.21)
23.2 8p23.1 8p22 21.3 21.2 8p12 8q12.1 8q21.3 q22.1 22.3 23.1 q23.3 24.13 24.22 8q24.3
Scale
chr8:
5 kb
130,693,000
130,694,000
130,695,000
130,696,000
130,697,000
130,698,000
130,699,000
130,700,000
hg19
130,701,000
130,702,000
JASPAR CORE 2022 - Predicted Transcription Factor Binding Sites (38873 items filtered out)
CTCF
CTCF
CTCF
CTCF
K562 CTCF ChIA-PET Interactions Rep 1 from ENCODE/GIS-Ruan
8:130693928..130694709,3
8:130699431..130700261,2
8:130738676..130739207,2
8:130705234..130705963,2
8:130738268..130738843,2
K562 CTCF ChIA-PET Signal Rep 1 from ENCODE/GIS-Ruan
200
0
K562 CTCF Sig 1
K562 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
200
1
K562 CTCF Sg 1

d


SBR (+gRNA/NT)
4
6
8
(-)
(+)
(-)
(+)
RB1 gRNA a
RB1 gRNA b
RB1 gRNA c
RB1 gRNA d
(-)
(+)
(-)
(+)
LB1 gRNA 1
LB1 gRNA 2
LB1 gRNA 3
LB1 gRNA 4

e


SBR (+gRNA/NT)
5
7
9
(-)
(+)
RB2 gRNA c
RB2 gRNA d
RB2 gRNA e
RB2 gRNA f
(+)
(-)
(+)
(-)
LB2 gRNA 1
LB2 gRNA 2

f

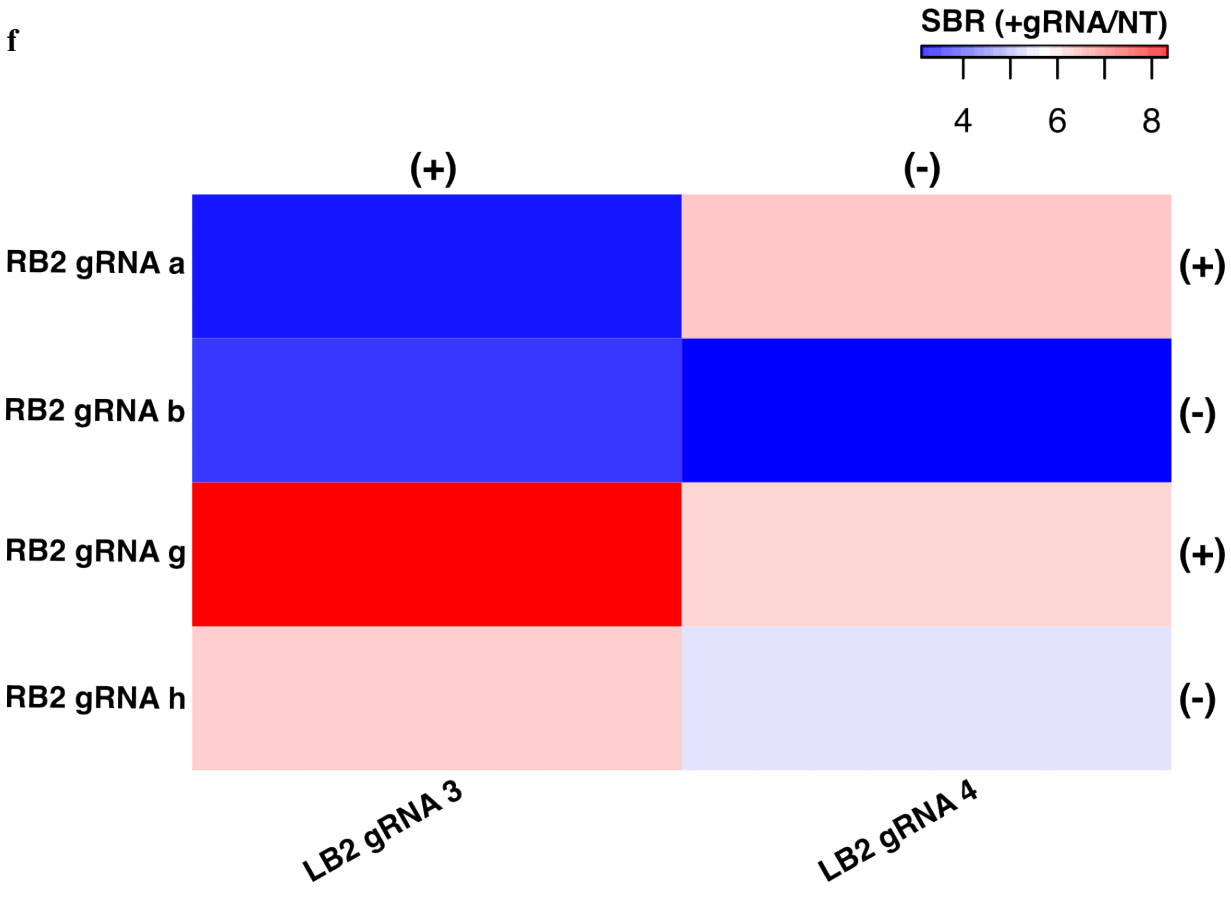

SBR (+gRNA/NT)
4
6
8
(+)
(-)
RB2 gRNA a
RB2 gRNA b
RB2 gRNA g
RB2 gRNA h
(+)
(-)
(+)
(-)
LB2 gRNA 3
LB2 gRNA 4

Figure 5.9: Targeting TAD boundary region chromatin loops at the *MYC* TAD using a dual dCas9 species LgBiT-d*Sa*Cas9 + d*Sp*Cas9-SmBiT DNA biosensor

(a) Detailed cartoon representation of the *MYC* TAD boundary region in K562. Representative sgRNA design features for biosensing experiments including 5'-3' positioning, spacing from CTCF binding sites, and genomic separation of CTCF pairs are shown for each of the two CTCF pairs that anchor the boundary region (not drawn to scale). sgRNAs were given labels 1-4 at the *MYC* TAD left boundary region binding sites and either labels a-d (upstream CTCF pair 1) or a-h (downstream CTCF pair 2) at the *MYC* TAD right boundary region binding sites. **(b)** UCSC genome browser track set showing K562 CTCF ChIA-PET and CTCF ChIP-seq signal tracks (aligned read density) from the ENCODE project at the *MYC* TAD left boundary region (LB). The core predicted JASPAR transcription factor binding sites track for CTCF is shown at the top. **(c)** UCSC genome browser track set showing K562 CTCF ChIA-PET and CTCF ChIP-seq signal tracks (aligned read density) from the ENCODE project at the *MYC* TAD right boundary region (RB). The core predicted JASPAR transcription factor binding sites track for CTCF is shown at the top. **(d)** Heatmap representing chromatin loop biosensing results where 16 pairs of sgRNAs tiling along the upstream pair of convergent CTCF binding sites (CTCF pair 1) within the ~2.8 Mb *MYC* TAD boundary region were transfected in K562 cells. Luminescence and fluorescence were measured using a multimode plate reader 24 h post-transfection. sgRNA strand polarity is shown at top and right. **(e)** Heatmap representing chromatin loop biosensing results where 8 pairs of sgRNAs tiling along the downstream pair of convergent CTCF binding sites (CTCF pair 2) within the ~2.8 Mb *MYC* TAD boundary region were transfected in K562 cells. Luminescence and fluorescence were measured using a multimode plate reader 24 h post-transfection. These innermost sgRNAs are all between 27 bp and 49 bp from the CTCF binding site. sgRNA strand polarity is shown at top and right. **(f)** Heatmap representing chromatin loop biosensing results where 8 pairs of sgRNAs tiling along the downstream pair of convergent CTCF binding sites (CTCF pair 2) within the ~2.8 Mb

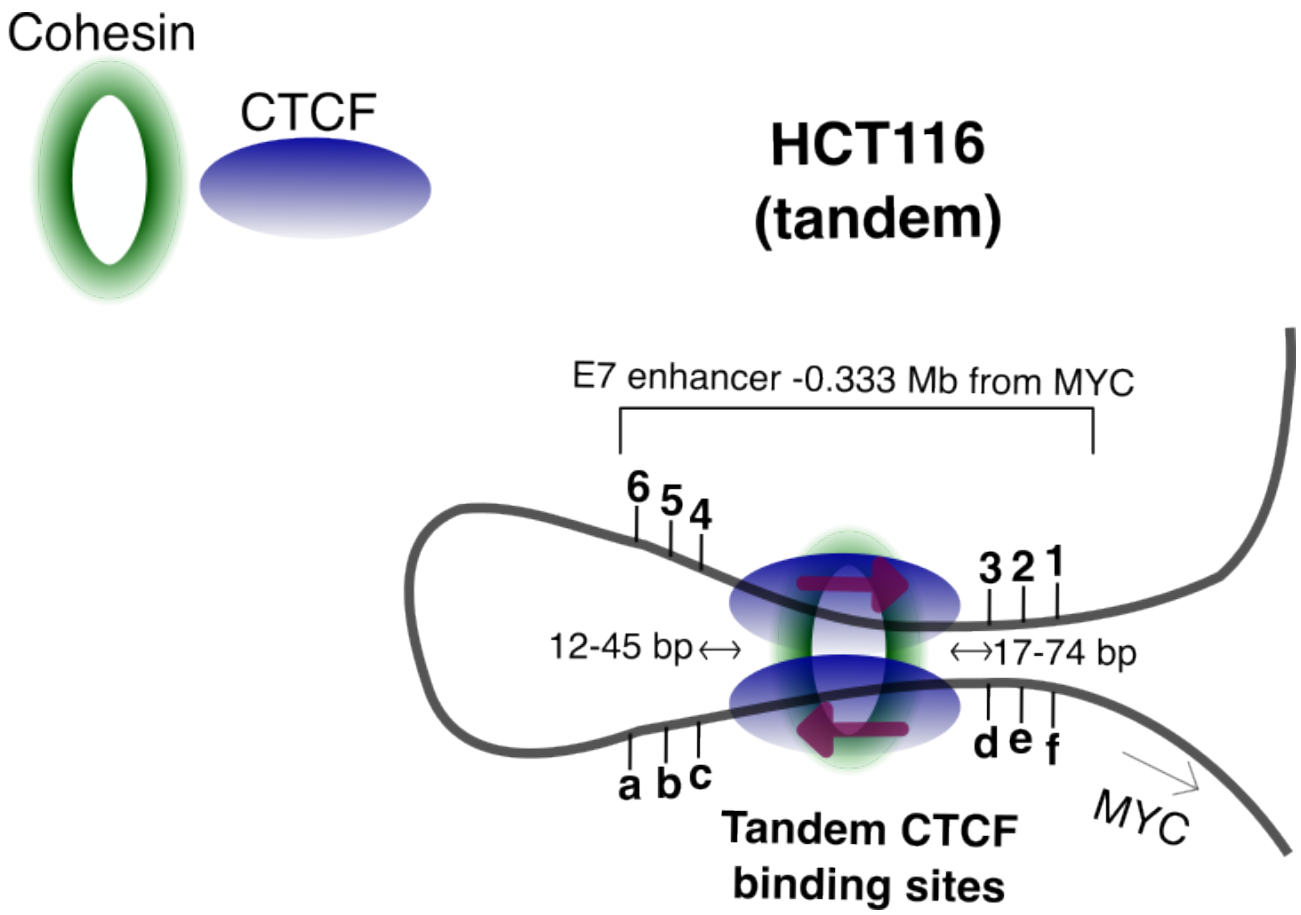
MYC TAD boundary region were transfected in K562 cells. Luminescence and fluorescence were measured using a multimode plate reader 24 h post-transfection. These outermost sgRNAs are all between 156 bp and 192 bp from the CTCF binding site. sgRNA strand polarity is shown at top and right. Apparent signal-to-background ratios in **d-f** (comparisons made to non-targeting (NT) background condition where sgRNAs targeting non-interacting genomic regions selected based on public Hi-C data sets were transfected) are shown for each sgRNA pair, $n = 4$, where n represents the number of independent experimental technical replicates included in parallel. For **d-f**, individual signal-to-background ratio (SBR) scales are shown at top right for each heatmap.

Biosensing an interaction between the *MYC* promoter and the E7 enhancer and using a dual dCas9 species DNA biosensor and effect of E7 enhancer deletion on signal-to-background

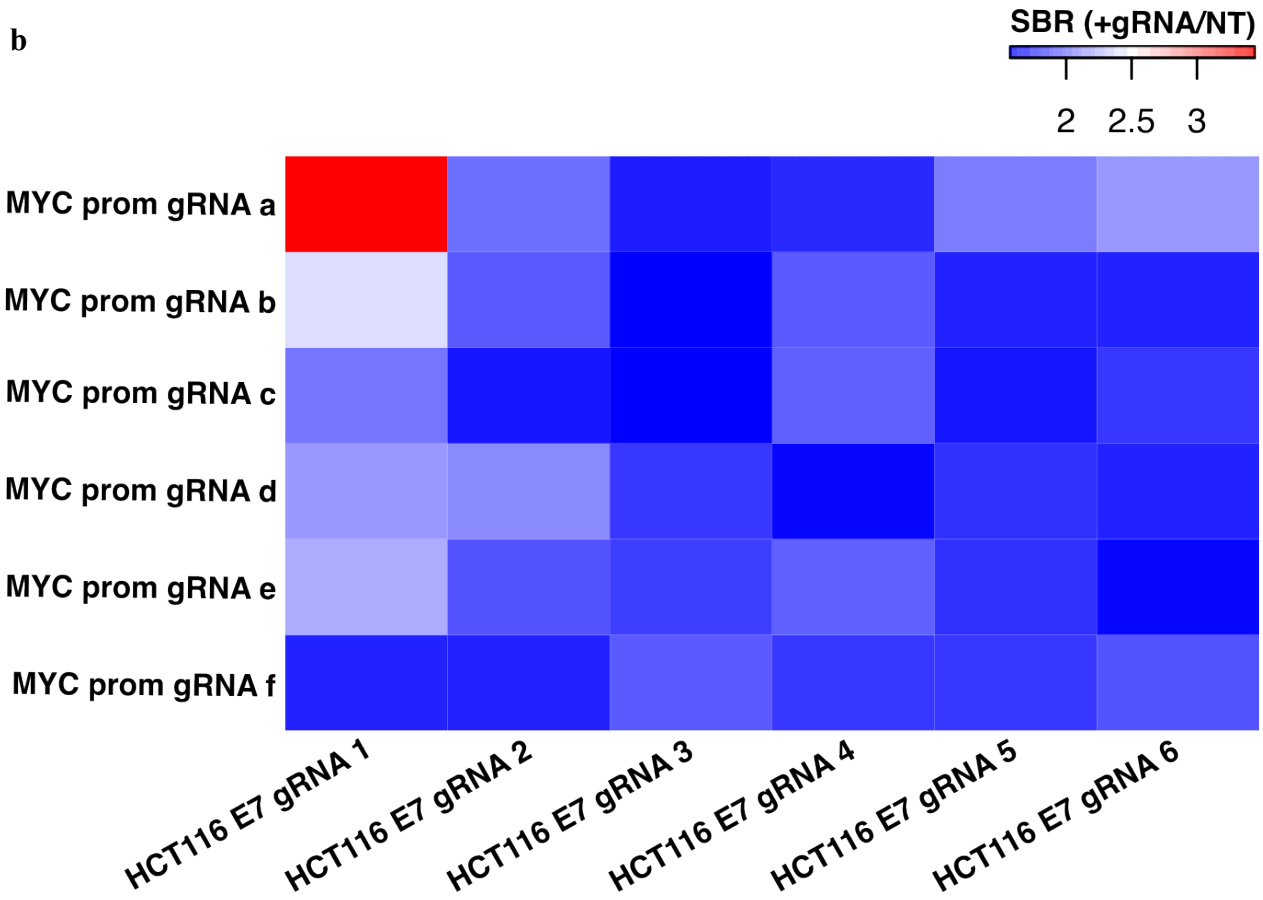
To determine whether the presence of an intact CTCF binding site was sufficient to produce luminescence from our dual dCas9 species chromatin loop biosensor, we next directed our LgBiT-d*Sa*Cas9 and d*Sp*Cas9-SmBiT DNA biosensor to a chromatin loop between the conserved *MYC* promoter CTCF binding site and the well-characterized E7 enhancer region ~0.34 Mb upstream of the *MYC* promoter **(Figure 5.10a)**. An HCT116 cell line with the E7 enhancer region deleted was provided as a kind gift from the Peggy Farnham lab. Like our experiments testing our dual dCas9 species DNA biosensor at super enhancer-*MYC* promoter chromatin loops, we transfected our dual dCas9 species DNA biosensor components along with all possible combinations of six d*Sa*Cas9 sgRNAs and six d*Sp*Cas9 sgRNAs targeting the *MYC* promoter and E7 enhancer loci at locations we hypothesized to be proximal at this ~0.34 Mb loop anchor in both wild-type HCT116 cells and in an HCT116 cell line where the E7 enhancer region was deleted[68]. In addition, using 4C-seq with a non-blind viewpoint centered on the conserved CTCF binding site at the *MYC* promoter, we confirmed that the ~0.34 Mb *MYC* promoter-E7 enhancer loop domain originally present in wild-type HCT116 cells was reduced to near

background levels in the E7 enhancer deletion HCT116 cells **(Figure S1d, Figure S3)**. In wild-type HCT116 cells, we observed a range of signal-to-background of 1.58-fold to 3.44-fold **(Figure 5.10b)**. However, when directing our dual dCas9 species DNA biosensor to the same loop anchor regions in the E7 enhancer deletion HCT116 cells, we observed a range of signal-to-background of 0.99-fold to 1.16-fold **(Figure 5.10c)**, showing E7 enhancer deletion directly reduced luminescence to near background levels, which were determined by signal from a pair of sgRNAs directed to bind at non-interacting genomic sites.

a



b


SBR (+gRNA/NT)
2
2.5
3
MYC prom gRNA a
MYC prom gRNA b
MYC prom gRNA c
MYC prom gRNA d
MYC prom gRNA e
MYC prom gRNA f
HCT116 E7 gRNA 1
HCT116 E7 gRNA 2
HCT116 E7 gRNA 3
HCT116 E7 gRNA 4
HCT116 E7 gRNA 5
HCT116 E7 gRNA 6

c

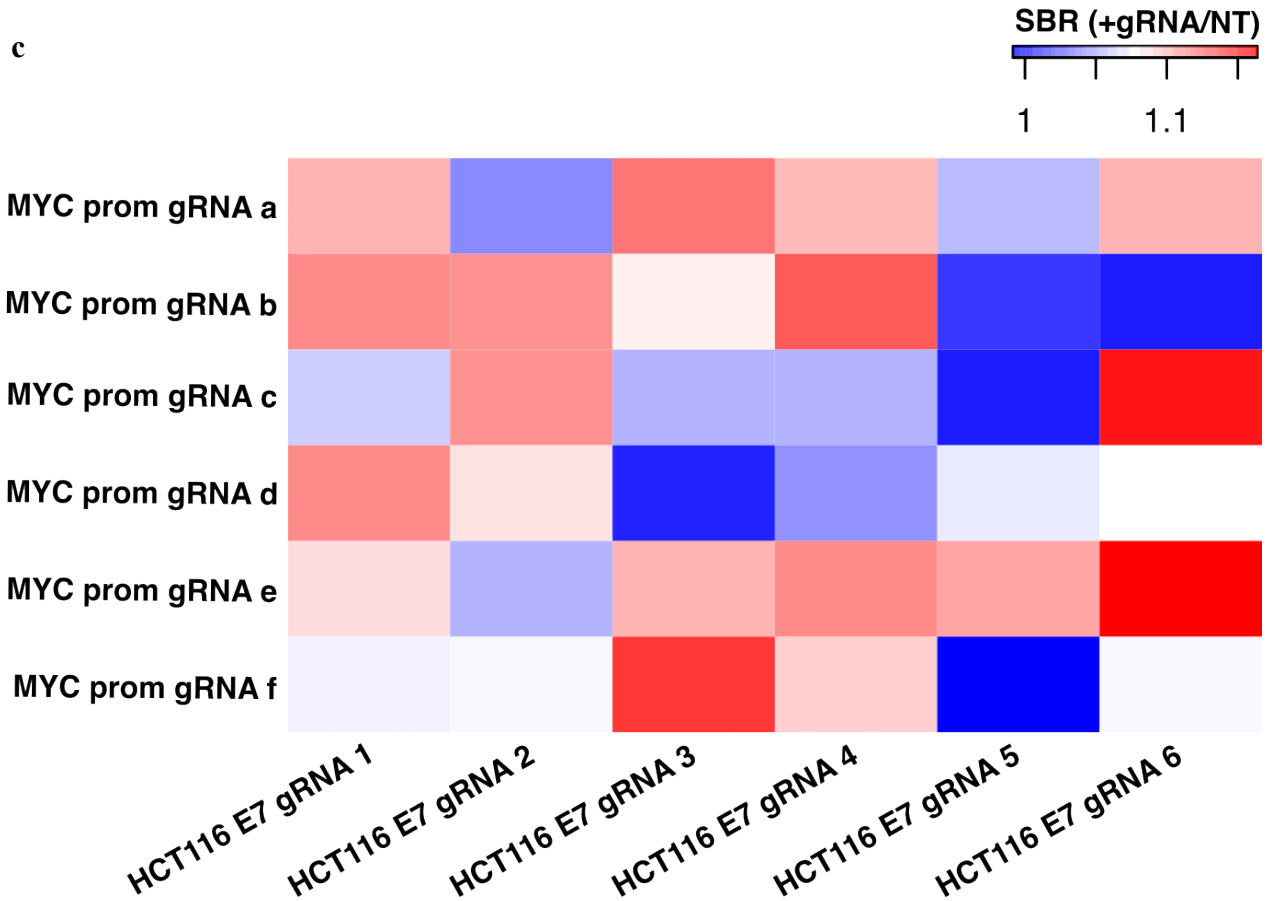

SBR (+gRNA/NT)
1
1.1
MYC prom gRNA a
MYC prom gRNA b
MYC prom gRNA c
MYC prom gRNA d
MYC prom gRNA e
MYC prom gRNA f
HCT116 E7 gRNA 1
HCT116 E7 gRNA 2
HCT116 E7 gRNA 3
HCT116 E7 gRNA 4
HCT116 E7 gRNA 5
HCT116 E7 gRNA 6

Figure 5.10: Targeting the *MYC* promoter-E7 enhancer chromatin loop using a dual dCas9 species LgBiT-d*Sa*Cas9 + d*Sp*Cas9-SmBiT DNA biosensor

(a) Cartoon representation of intra-TAD chromatin loops between the *MYC* promoter and the E7 enhancer region in HCT116 cells. Representative sgRNA design features for biosensing experiments including 5'-3' positioning and spacing from CTCF binding sites are shown for each of the loop anchor regions (not drawn to scale). sgRNAs were given labels a-f at the *MYC* promoter binding sites and labels 1-6 at the E7 enhancer binding sites. **(b)** Heatmap representing chromatin loop biosensing results in wild-type HCT116 cells where 36 pairs of sgRNAs tiling along a CTCF binding site within the E7 enhancer ~0.333 Mb upstream of the *MYC* promoter and tiling along a highly conserved CTCF binding site at the *MYC* promoter were transfected and luminescence and fluorescence were measured using a multimode plate reader 24 h post-transfection. **(c)** Heatmap representing chromatin loop biosensing results in an E7 enhancer deletion HCT116 cell line where the same 36 pairs of sgRNAs were targeted to the ~0.333 Mb *MYC* promoter-E7 enhancer chromatin loop. These sgRNA pairs were transfected and luminescence and fluorescence were again measured using a multimode plate reader 24 h post-transfection. Apparent signal-to-background ratios in **b-c** (comparisons made to non-targeting (NT) background condition where sgRNAs targeting non-interacting genomic regions selected based on public Hi-C data sets were transfected) are shown for each sgRNA pair, $n = 4$, where n represents the number of independent experimental technical replicates included in parallel. For **b-c**, individual signal-to-background ratio (SBR) scales are shown at top right for each heatmap.

Comparison of looping strength and CTCF binding frequency for loops targeted with our dual dCas9 species DNA biosensor as measured by Hi-C, 4C-seq, CTCF ChIA-PET, and CTCF ChIP-seq

Given our results indicate higher signal-to-background ratios were observed for nearly every pairing of sgRNAs at the *MYC* TAD boundary regions compared to sgRNA pairings at the intra-TAD *MYC* promoter-super enhancer and *MYC* promoter-E7 enhancer loops in K562, we were interested in examining potential correlations between our loop biosensing signal-to-background measurements and strength of these loops as revealed by existing *cis* contact frequency data sets such as CTCF ChIA-PET and Hi-C data sets. In K562 cells, while the CTCF pair involved in the upstream TAD boundary interaction has peak CTCF ChIA-PET signals of 26 and 139, the CTCF pair involved in the downstream TAD boundary interaction has peak CTCF ChIA-PET signals of 119 and 121. For intra-TAD loops in K562 cells, the CTCF pair involved in the ~8.2 kb super enhancer-*MYC* promoter interaction has peak CTCF ChIA-PET signals of 128 and 173 while the CTCF pair involved in the +1.85 Mb super enhancer-*MYC* promoter interaction has peak CTCF ChIA-PET signals of 173 and 61. These CTCF ChIA-PET data show low to moderate positive correlation with average and peak signal-to-background across these four chromatin loops (Pearson's R=0.38 and R=0.51, respectively). Thus, looking at the average of CTCF ChIA-PET peak signals for these K562 loop anchors, the ~8.2 kb super enhancer-*MYC* promoter interaction was predicted to be the strongest or most frequent interaction compared to any of the other loops targeted in K562 cells. In addition, the downstream TAD boundary interaction was predicted to be stronger than the upstream TAD boundary interaction in K562 cells.

Next, we compared normalized Hi-C contact frequencies for the loops targeted in K562. We found the corner dot for the 5 kb bins containing the -8.2 kb super enhancer CTCF site and the *MYC* promoter CTCF site has a score of 13.04, the corner dot for the 5 kb bins containing the +1.85 Mb super enhancer CTCF site and the *MYC* promoter CTCF site has a score of 1.83, the corner dot for the 5 kb bins containing the upstream TAD boundary interaction CTCF pair has a score of 1.61, and the corner dot for the 5 kb bins containing the downstream TAD boundary interaction CTCF pair has a score of 2.69. These Hi-C data show low to moderate positive correlation

with average and peak signal-to-background across these four chromatin loops (Pearson's R=0.30 and R=0.54, respectively). In addition, we wanted to examine potential correlations between our loop biosensing signal-to-background measurements and strength of intra-TAD *MYC* loops as revealed by our 4C-seq data sets. Comparing average normalized peak 4C-seq signals and total number of uniquely mapped interacting segments within 1 kb on either side of the two loop anchor CTCF binding sites across three replicates for the loops targeted in K562, the -8.2 kb super enhancer-*MYC* promoter CTCF pair had average normalized peak 4C-seq signal of ~1,135 and 6 unique mapped interactions of average normalized 4C-seq signal ~621 while the +1.85 Mb super enhancer-*MYC* promoter CTCF pair had average normalized peak 4C-seq signal of ~177 and 7 unique mapped interactions of average normalized 4C-seq signal ~126. For HCT116 intra-TAD loops, the -8.2 kb super enhancer-*MYC* promoter CTCF pair had average normalized peak 4C-seq signal of ~914 and 6 unique mapped interactions of average normalized 4C-seq signal ~578, the -0.53 Mb super enhancer-*MYC* promoter CTCF pair had average normalized peak 4C-seq signal of ~189 and 5 unique mapped interactions of average normalized 4C-seq signal ~150, and the -0.34 Mb E7 enhancer-*MYC* promoter CTCF pair had average normalized peak 4C-seq signal of ~129 and 6 unique mapped interactions of average normalized 4C-seq signal ~119. Our average normalized 4C-seq data for all interactions within 1 kb upstream and downstream of the CTCF homodimer show low to moderate positive correlation with average and peak signal-to-background across these four chromatin loops (Pearson's R=0.45 and R=0.35, respectively). However, our average normalized peak 4C-seq signal show moderate positive correlation with average and peak signal-to-background across these four chromatin loops (Pearson's R=0.45 and R=0.40, respectively). Therefore, based on normalized Hi-C contact frequencies and normalized 4C-seq signals, the -8.2 kb super enhancer-*MYC* promoter interaction was also predicted to be the strongest or most frequent interaction compared to any of the other loops targeted.

In addition, we wanted to determine if there was a relationship between strength of CTCF ChIP-seq peaks at a loop anchor and signal-to-background in our biosensing experiments. While the upstream TAD boundary interaction had peak CTCF ChIP-seq signals of 51 and 92, the downstream TAD boundary interaction had peak

CTCF ChIP-seq signals of 195 and 169 in K562 cells. For K562 intra-TAD loops, the CTCF pair involved in the -8.2 kb super enhancer-*MYC* promoter interaction had peak CTCF ChIP-seq signals of 92 and 130 while the CTCF pair involved in the +1.85 Mb super enhancer-*MYC* promoter interaction had peak CTCF ChIP-seq signals of 130 and 77. For HCT116 intra-TAD loops, the CTCF pair involved in the -8.2 kb super enhancer-*MYC* promoter interaction had peak CTCF ChIP-seq signals of 90 and 119, the CTCF pair involved in the -0.53 Mb super enhancer-*MYC* promoter interaction had peak CTCF ChIP-seq signals of 105 and 119, and the CTCF pair involved in the -0.34 Mb E7 enhancer-*MYC* promoter interaction had peak CTCF ChIP-seq signals of 144 and 119. These CTCF ChIP-seq data show low positive correlation with average and peak signal-to-background across these four chromatin loops (Pearson's R=0.27 and R=0.29, respectively). Looking at the average of CTCF ChIP-seq peak signals, the downstream TAD boundary interaction was predicted to show the strongest CTCF binding.

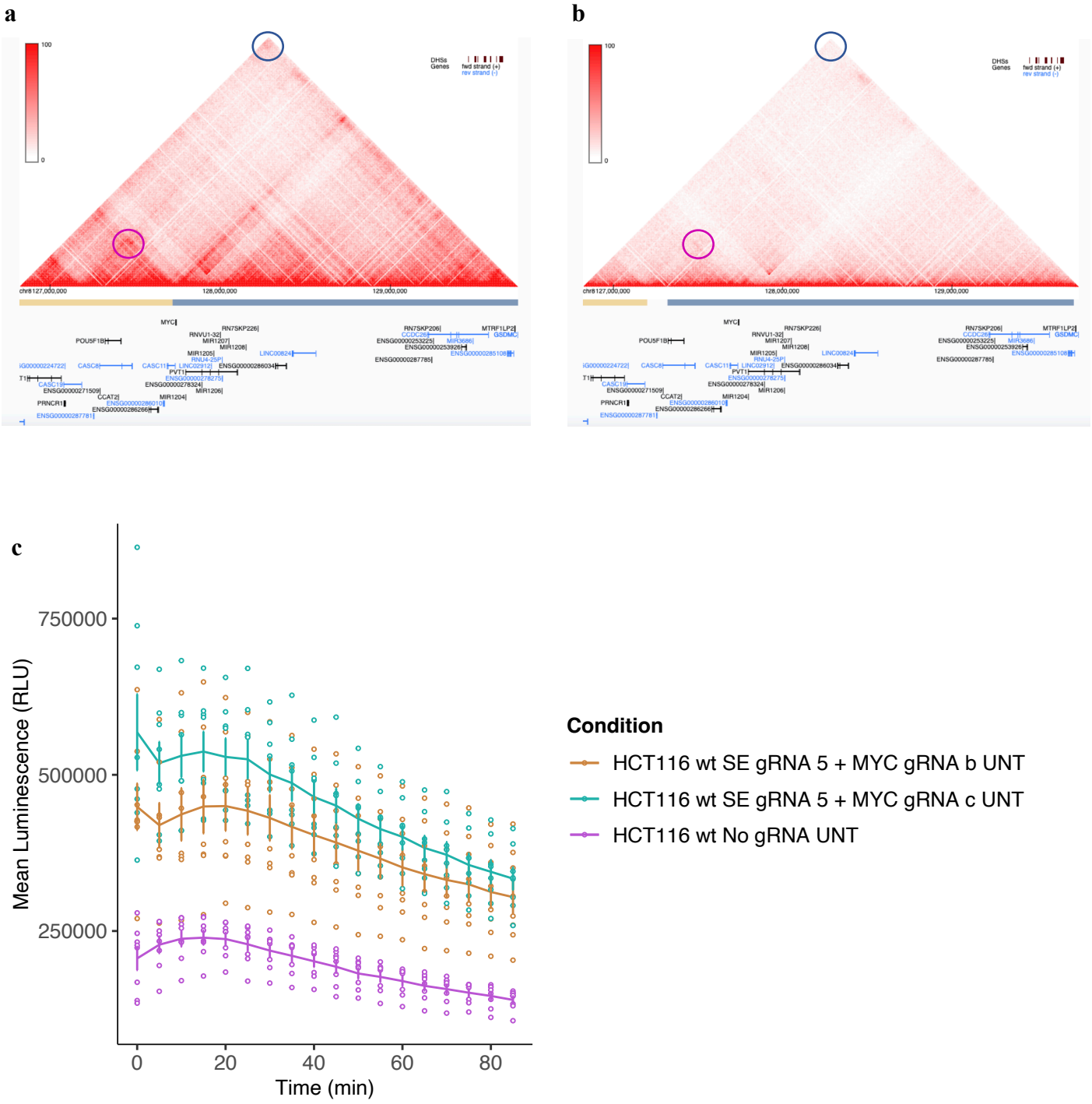
To get a better idea of the strength of various predictors for average and peak signal-to-background for our dual dCas9 species DNA biosensor, we also determined coefficients of multiple correlation for combinations of Hi-C, 4C-seq, CTCF ChIA-PET and CTCF ChIP-seq data sets with average and peak signal-to-background ratios across several chromatin loops. For the five intra-TAD loops measured via 4C-seq in HCT116 and K562 cells which also have CTCF ChIP-seq peak data, these two predictors show high positive correlation with average and peak signal-to-background for our biosensor when combined (coefficient of multiple correlation R=0.67 and R=0.99, respectively). However, for the four K562 loops for which both ChIA-PET and Hi-C data were available, these two predictors show only moderate positive correlation with of average and peak signal-to-background for our biosensor when combined (coefficient of multiple correlation R=0.38 and R=0.56, respectively). In addition, we found that Hi-C combined with CTCF ChIP-seq peak data showed moderate to high positive correlation with average and peak signal-to-background for our biosensor (coefficient of multiple correlation R=0.57 and R=0.74, respectively) while ChIA-PET combined with CTCF ChIP-seq peak data

showed high positive correlation with average and peak signal-to-background for our biosensor when combined (coefficient of multiple correlation R=0.79 and R=0.93, respectively).

Time-course biosensing to measure the effect of cohesin subunit RAD21 degradation on signal-to-background using an auxin-inducible degron system

Next, to determine whether the presence of RAD21, a core component of the cohesin complex, was sufficient to produce luminescence in our chromatin loop biosensing system, we directed our dual dCas9 species DNA biosensor to the same ~0.53 Mb *MYC* promoter-super enhancer loop in HCT116 cells examined previously and targeted RAD21 for degradation using the second-generation auxin-inducible degron 2 (AID2) system[69], an improved version of the originally described auxin-inducible degron (AID) technology[70]. It was previously demonstrated that loss of cohesin subunit RAD21 eliminated all loop domains immediately after auxin treatment of HCT116 cells harboring a RAD21-AID endogenous system **(Figure 5.11a-b)**[71]. In addition, using 4C-seq with a non-blind viewpoint centered on the conserved CTCF binding site upstream of the *MYC* promoter, we confirmed that loop domains originally present in HCT116-RAD21-mAC OsTIR1(F74G) cells were reduced to near background levels after a 120 min treatment with 5-Ph-IAA ligand **(Figure S4, Figure S5)**. Then, we tracked luminescence over time after addition of 5-Ph-IAA ligand to cell culture medium in untreated wild-type HCT116 cells **(Figure 5.11c)**, wild-type HCT116 cells treated with 1 μM 5-Ph-IAA **(Figure 5.11d)**, untreated HCT116-RAD21-mAC OsTIR1(F74G) cells **(Figure 5.11e)**, and HCT116-RAD21-mAC OsTIR1(F74G) cells treated with 1 μM 5-Ph-IAA **(Figure 5.11e)**. Using this AID2 system for RAD21, we found that the luminescent signal decreased at a much higher rate to close to background levels when HCT116-RAD21-mAC OsTIR1(F74G) cells were treated with 1 μM 5-Ph-IAA for 105 min compared to untreated HCT116-RAD21-mAC OsTIR1(F74G) cells over the same time-course **(Figure 5.11e)**. Approximately 67% of the initial luminescent signal at t_0 was degraded in the first 5 minutes upon treatment with 1 μM 5-Ph-IAA while approximately 4% of the initial luminescent signal at t_0 was degraded in untreated

cells **(Figure 5.11e)**. These results are consistent with the extremely short half-life of ~12 min for RAD21-mAID-Clover reported upon treatment of HCT116-RAD21-mAC OsTIR1(F74G) cells with 1 μM 5-Ph-IAA[69].



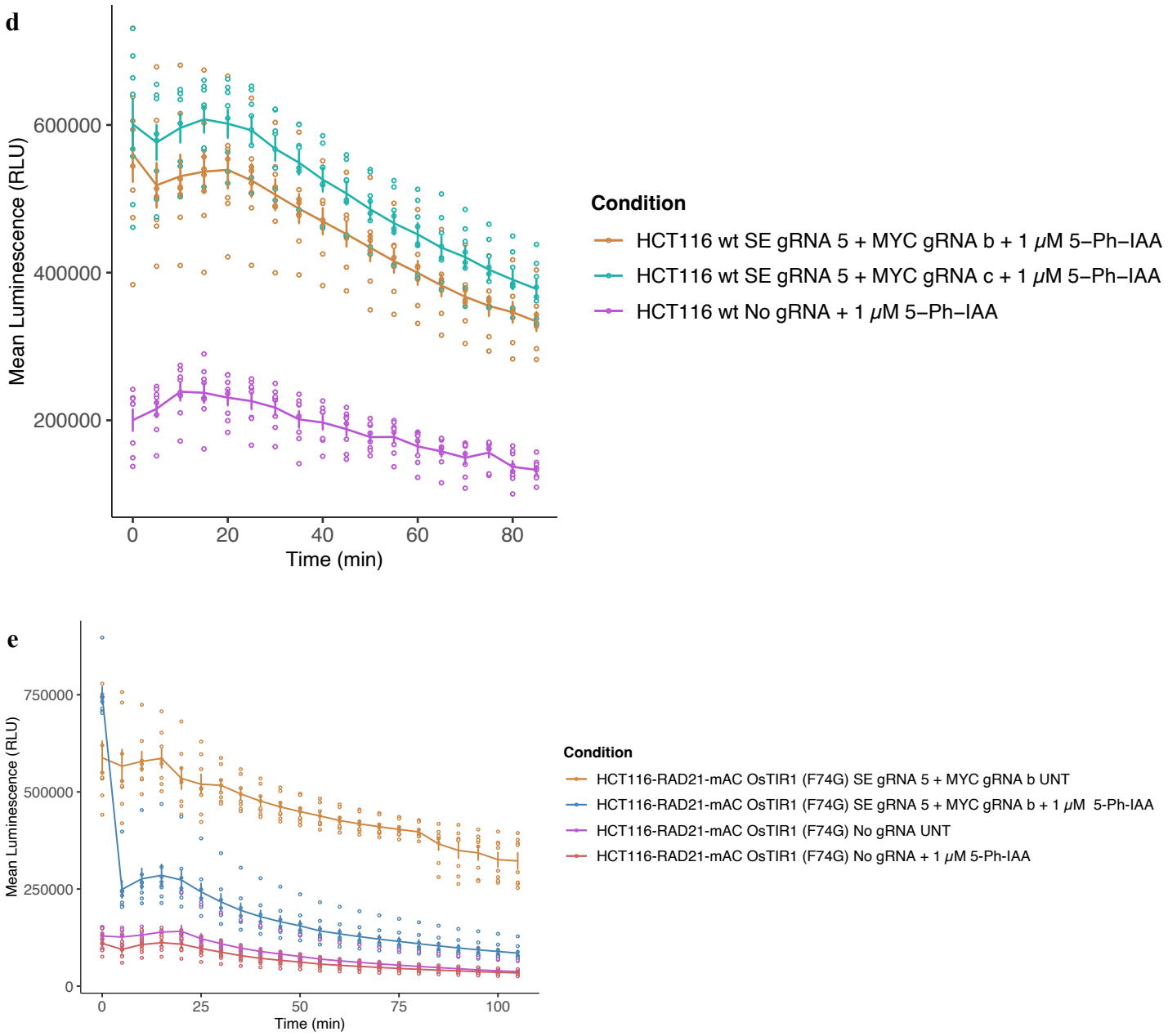


Figure 5.11: Time-course chromatin loop biosensing using an auxin-inducible degron (AID) system for endogenous RAD21 in HCT116 cells

(a) Hi-C contact map at 10 kb resolution showing the *MYC* TAD in untreated HCT116-RAD21-mAC cells. The TAD boundary and super enhancer interactions (corner dots) are highlighted in blue and magenta, respectively. The intensity of each pixel represents the normalized number of contacts between a pair of loci. Maximum intensity is indicated in the scale at right. **(b)** Hi-C contact map at 10 kb resolution showing the *MYC* TAD in HCT116-RAD21-mAC cells subjected to a 6 h auxin treatment. The location

of the depleted TAD boundary and super enhancer interactions (corner dots) are highlighted in blue and magenta, respectively. The intensity of each pixel represents the normalized number of contacts between a pair of loci. Maximum intensity is indicated in the scale at right. **(c)** Time-course biosensing results measured for 85 min in untreated wild-type HCT116 cells transfected with two different *MYC* promoter-super enhancer sgRNA pairs to measure natural degradation of NanoLuc luciferase luminescence. A condition where no sgRNA pairs were transfected is shown for comparison. Luminescence in 8 technical replicate wells in a 96-well plate were measured every 5 min for 85 min for each transfection condition. **(d)** Time-course biosensing results measured for 85 min in wild-type HCT116 cells transfected with two different *MYC* promoter-super enhancer sgRNA pairs and treated with 5-Ph-IAA to demonstrate the lack of 5-Ph-IAA ligand effect on wild-type cells. A 5-Ph-IAA treatment condition where no sgRNA pairs were transfected is shown for comparison. Luminescence in 8 technical replicate wells in a 96-well plate were measured every 5 min for 85 min for each transfection condition. **(e)** HCT116-RAD21-mAC OsTIR1(F74G) cells treated with 1 μM 5-Ph-IAA for 105 min were compared to untreated HCT116-RAD21-mAC OsTIR1(F74G) cells over the same 105 min time-course. Both treated and untreated conditions were transfected with LgBiT-d*Sa*Cas9 + d*Sp*Cas9-SmBiT + super enhancer gRNA 5 + *MYC* promoter gRNA b. 5-Ph-IAA-treated and untreated conditions where no sgRNA pairs were transfected are shown for comparison. Luminescence in 8 technical replicate wells in a 96-well plate were measured every 5 min for 105 min for each transfection condition.

DISCUSSION

In this study, we hypothesized that directing our dual dCas9 species DNA biosensor to chromatin loop anchors could be a reasonable approach if the expected distance to bridge a loop anchor at the base region defined by CTCF homodimerization was short enough to allow for efficient protrusion of NanoBiT interaction partners into the gap between *cis*-interacting genomic regions. Our results demonstrate that our dual dCas9 species DNA biosensor can detect the presence of chromatin loops in living cells at up to 10.59-fold above background levels. Thus, we believe that our models that predict a relatively short ~0.7-4 nm distance between *cis*-interacting DNA sequences at the base of a loop anchor immediately adjacent to homodimerized CTCFs are accurate and this makes loop anchor regions highly amenable to being targeted by noninvasive DNA biosensors in real-time in live cells. Our proposed loop anchor models should allow for design and development of similar DNA biosensors with a wide range of bioreceptor and transducer elements that precisely target chromatin loops at their CTCF-proximal anchor regions in the future.

We originally targeted our single dCas9 species split NanoLuc luciferase DNA biosensor to cohesin and CTCF homodimer-anchored loops using combinations of sgRNAs predicted to result in higher likelihood of NanoLuc luciferase reassembly assuming validity of various abstract models for chromatin loop structures such as the "hairpin" and "coil" models. These chromatin loop structures could both be consistent with the favored loop extrusion model for chromatin loop formation. Results from testing our single dCas9 species DNA biosensor in both HCT116 and K562 cell lines indicate that pairings of sgRNAs closest to the CTCF binding sites resulted in higher signal-to-background for both "hairpin" and "coil" sgRNA pairings. This included pairing sgRNA 4 at the super enhancer with sgRNA c at the *MYC* promoter in both HCT116 and K562 cells for targeting the "hairpin" model and pairing sgRNA 3 at the super enhancer with sgRNA c at the *MYC* promoter in HCT116 cells for targeting the "coil" model. This relationship did not appear to hold true when we tested our dual dCas9 species DNA biosensor on the same loops, where the highest signal-to-background ratios were generally observed for pairings of sgRNAs more distal from the CTCF homodimer. A notable exception was in the case

of pairing sgRNA 4 at the +1.85 Mb super enhancer and sgRNA d at the *MYC* promoter in K562 cells, a pairing immediately adjacent to the CTCF homodimer. This sgRNA pairing produced the second highest signal-to-background ratio of 5.45-fold at this chromatin loop in K562 cells. Notably, the sgRNAs for this pair were separated from the nearest CTCF site by 33 bp and 74 bp, respectively, which likely provided enough space to minimize steric hindrance effects with CTCF that could negatively affect efficiency of biosensor binding. For the chromatin loops targeted between the -0.53 Mb super enhancer and *MYC* promoter in HCT116 cells and between the -8.2 kb super enhancer and the *MYC* promoter in HCT116 and K562 cells, sgRNA pairs that produced the highest signal-to-background ratios had 151 bp/167 bp and 148 bp/126 bp spacing from the nearest CTCF sites, respectively. In addition, for the E7 enhancer-*MYC* promoter loop targeted in HCT116 cells, the sgRNA pair that produced the highest signal-to-background ratios had 209 bp/219 bp spacing from the nearest CTCF sites. Finally, for the two CTCF homodimers targeted at the *MYC* TAD boundary region, sgRNA pairs that produced the highest signal-to-background ratios had 153 bp/62 bp and 49 bp/41 bp spacing from the nearest CTCF sites. Conversely, we observed some of the lowest signal-to-background ratios when sgRNA pairs with <20 bp spacing from at least one CTCF binding site were directed to bind chromatin loop anchors. Given this striking data, we conclude that a reasonable spacing of at least 30-40 bp should be left between CTCF and the nearest sgRNA when designing sgRNA pairings to target chromatin loops in live cells using our DNA biosensor in order to maximize signal-to-background. However, signal-to-background of the dual dCas9 species DNA biosensor might be positively affected by increasing distance from the CTCF sites up to at least 200-225 bp. Additional future testing should be done to determine if there is a maximum allowable distance threshold for targeting cohesin and CTCF homodimer-anchored chromatin loops.

After analyzing results using our single dCas9 species DNA biosensor, we conclude that it is possible that the higher and less variant range of signal-to-background ratios observed for pairings of sgRNAs if tandem loops were formed consistent with the “coil” model (2.01-fold to 2.66-fold) compared to the lower and more variant range of signal-to-background ratios observed for pairings of sgRNAs if tandem loops were formed consistent

with the “hairpin” model (1.07-fold to 2.45-fold), may provide evidence that the “coil” model could be a preferred structure for such tandem chromatin loops. For several other tandemly oriented loops targeted with our dual dCas9 species DNA biosensor, including the -8.2 kb super enhancer-*MYC* promoter loop in HCT116 and K562 cells and the -0.34 Mb E7 enhancer-*MYC* promoter loop in HCT116 cells, signal-to-background was maximum using a combination of sgRNAs that would be expected to be in closest proximity assuming the “coil” model at these loop anchors. These combinations included pairing sgRNA 1 at the -8.2 kb super enhancer region and -0.34 Mb E7 enhancer region with sgRNA a at the *MYC* promoter. Conversely, for the tandem -0.53 Mb super enhancer-*MYC* promoter loop in HCT116 cells, signal-to-background was maximum using a combination of sgRNAs—sgRNA 5 at the super enhancer region paired with sgRNA b at the *MYC* promoter—that would be expected to be in closest proximity assuming the “hairpin” model at this loop anchor. However, there is an alternative nearby CTCF site in this -0.53 Mb super enhancer region that could also be used to form the same loop anchor. If the loop anchor is formed using this alternate CTCF, the same super enhancer sgRNA would be at the 2nd position relative to the CTCF, and when paired with *MYC* promoter sgRNA b would be a combination that would be expected to be proximal assuming a “coil” model. Further analysis is required to determine if this pattern holds true for other tandem loop anchors targeted with “coil” model-specific sgRNA pairings.

Our single and multiple correlation analysis using individual Hi-C, 4C-seq, CTCF ChIA-PET and CTCF ChIP-seq data sets and average and peak signal-to-background data sets for the two TAD boundary loops and five intra-TAD loops targeted in this study showed low to moderate correlation between the individual *cis* interaction data sets and our signal-to-background data. However, we observed moderate to high correlation between our signal-to-background data and each of the *cis* interaction frequency data sets when combined with CTCF ChIP-seq data. Thus, an efficient approach to determining promising known looping regions to target using our biosensor could involve using CTCF ChIP-seq data in combination with either Hi-C, 4C-seq, or CTCF ChIA-PET data to locate such loop anchor regions throughout the genome. Based on the interaction

frequencies determined via Hi-C, 4C-seq, and ChIA-PET, the -8.2 kb super enhancer-*MYC* promoter interaction was predicted to be the strongest or most frequent interaction compared to any of the other loops targeted and the downstream TAD boundary interaction was predicted to be the second strongest or most frequent. However, the downstream TAD boundary interaction showed a higher range of signal-to-background (3.12-fold to 10.59-fold) than the -8.2 kb super enhancer-*MYC* promoter interaction (1.74-fold to 5.86-fold). This could be due to individual loop-specific differences in CTCF binding affinity, as revealed by the higher average CTCF ChIP-seq peak signal for the downstream TAD boundary interaction relative to the -8.2 kb super enhancer-*MYC* promoter interaction. Increased CTCF binding frequency revealed by higher average ChIP-seq peak signal for a given CTCF pair could increase the likelihood that a given loop will be bound by both CTCF proteins to allow homodimerization and loop anchor formation at any given time point. In addition, within the two *MYC* TAD boundary loop anchors targeted, the downstream TAD boundary interaction was predicted to be stronger or more frequent than the upstream TAD boundary interaction in K562 cells based on Hi-C, CTCF ChIA-PET, and CTCF ChIP-seq data sets. Indeed, our results showed that the range of signal-to-background (3.12-fold to 10.59-fold) at the downstream TAD boundary interaction was slightly higher than the range of signal-to-background at the upstream TAD boundary interaction (3.26-fold to 8.99-fold), possibly indicating that *cis* contact frequency data sets are more predictive of biosensing signal-to-background within a specific region of defined length than they are of biosensing signal-to-background between different regions separated by large genomic distances. Notably, as is the case for a Hi-C contact matrix, where entries of raw contact counts would ideally be proportional to the true pairwise contact frequency of various loci but are not exactly comparative due to "one-dimensional" experimental Hi-C biases[13], signal-to-background in our biosensing experiments should not be expected to be exactly proportional to true pairwise contact frequency. Chromatin accessibility, nucleosome occupancy, and sgRNA binding affinity and specificity at genomic regions forming a chromatin loop anchor can all affect the observed biosensing signal-to-background for this loop anchor region. Such experimental biases are locus-specific effects and tend to influence contact frequency between that locus and other loci, making comparisons of contact frequencies between two loops observed via biosensing especially

difficult. Additional normalization factors could be considered in the future for such biosensing data sets to make signals between loops more directly comparable.

In our earliest experiments, we only tested sgRNAs bound to a single strand (-). However, in testing our dual dCas9 species DNA biosensor at the *MYC* TAD boundary loop anchors, we designed both sets of d*Sp*Cas9 and d*Sa*Cas9 sgRNAs with combinations of (+) and (-) strand polarities. From our results, sgRNA pairings that produced the highest signal-to-background at loop anchors were always same strand sgRNA pairings ((+) strand with (+) strand or (–) strand with (–) strand) as opposed to opposite strand pairings. However, this effect could be related to a number of other factors such the distances of these sgRNA pairs from the CTCF sites, individual sgRNA binding affinities and specificities, and other experiment-specific parameters. We propose that several other TAD or sub-TAD boundary loops within the *MYC* TAD could be targeted to further explore this effect due to their high *cis* contact frequency determined in our 4C-seq analysis, such as the +0.92 Mb super-enhancer-*MYC* promoter and +1.95 Mb super enhancer-*MYC* promoter loops in K562 and HCT116 cells.

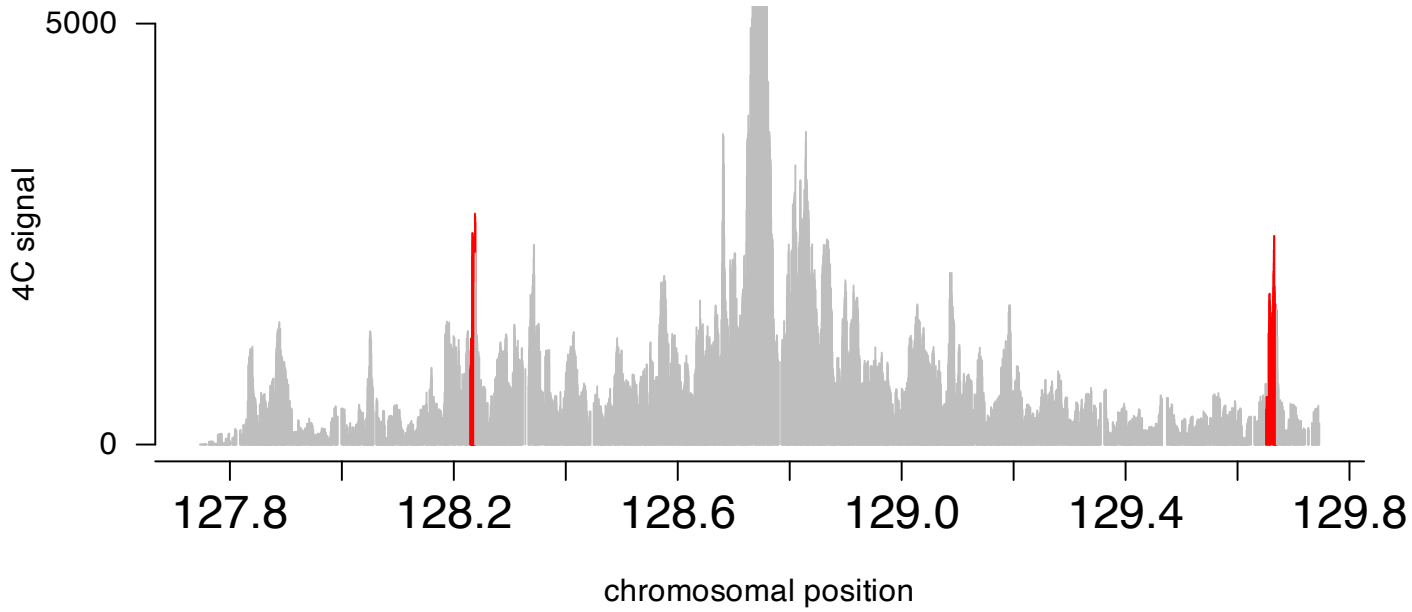
To demonstrate the specificity of CTCF binding in generating loops detected using our chromatin loop biosensor, we required a condition where the ability of CTCF to bind was greatly diminished or completely inhibited. We found the E7 enhancer deletion HCT116 cell line[68] provided a convenient model for testing this idea. In our 4C-seq experiments in the HCT116 E7 deletion cells, we found a notable decrease in the 4C-seq signal to near background levels at and around the ~2.6 kb deleted region compared to 4C-seq signal at the same region in wild-type HCT116 cells. Likewise, in our biosensing experiments comparing the signal for the E7 enhancer-*MYC* promoter interaction in wild-type HCT116 cells to the signal for this same interaction in E7 enhancer deletion HCT116 cells, we found that deletion of the E7 enhancer drove biosensing signal for this interaction to near background levels (0.99-fold to 1.16-fold) compared to biosensing signals 1.58-fold to 3.44-fold brighter than background levels for this interaction in wild-type HCT116 cells. This result strongly

implicates the deleted CTCF binding site targeted at the E7 enhancer in the production of luminescent signal from our biosensor.

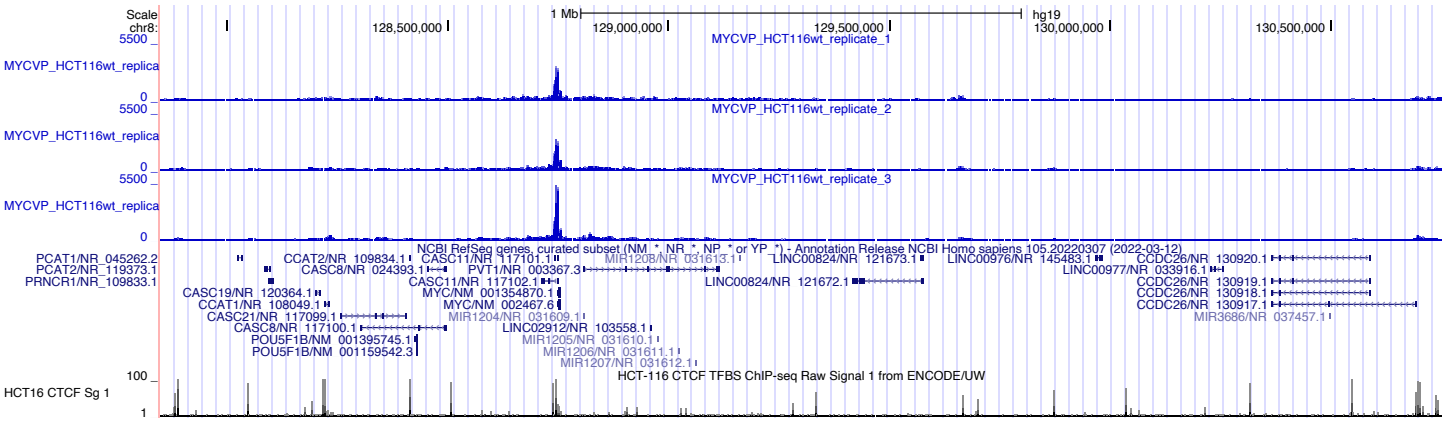
In experiments targeting the critical RAD21 subcomponent of the cohesin complex for degradation using a novel 5-Ph-IAA ligand, we found that when RAD21 was degraded, the resulting signal outputted by our chromatin loop biosensor dropped rapidly within 5-10 minutes. Thus, RAD21 degradation—and cohesin complex dissolution—was sufficient to decrease signal produced by our chromatin loop biosensor. This is presumably because RAD21-mAID was degraded with half-life ~12 min by 5-Ph-IAA[69], causing rapid loop dissolution for the ~0.53 Mb super enhancer-*MYC* promoter loop we targeted. Indeed, our independent 4C-seq analysis using the *MYC* promoter CTCF binding site as a viewpoint showed that raw 4C-seq signal was reduced to near background levels across the ~2.8 Mb *MYC* TAD and notable 4C-seq peaks were ablated as a result of 5-Ph-IAA treatment. However, an important caveat is that even though we observed a precipitous drop within the biosensing signal within the first 5 minutes of treatment of the HCT116-RAD21-mAC OsTIR1(F74G) cells with 5-Ph-IAA, RAD21 may or may not be degraded for this condition to be true. Given other factors within the cell culture medium and broader cell environment likely changed during incubation within the multimode plate reader, we cannot be certain that RAD21 degradation was a necessary condition for biosensing signal to drop. Hence, RAD21 degradation and cohesin complex destruction is a sufficient, but not a necessary feature for decrease in biosensing signal.

SUPPLEMENTARY INFORMATION for A Split Luciferase Biosensing Platform for Detection and Imaging of Chromatin Loops in Individual Living Cells

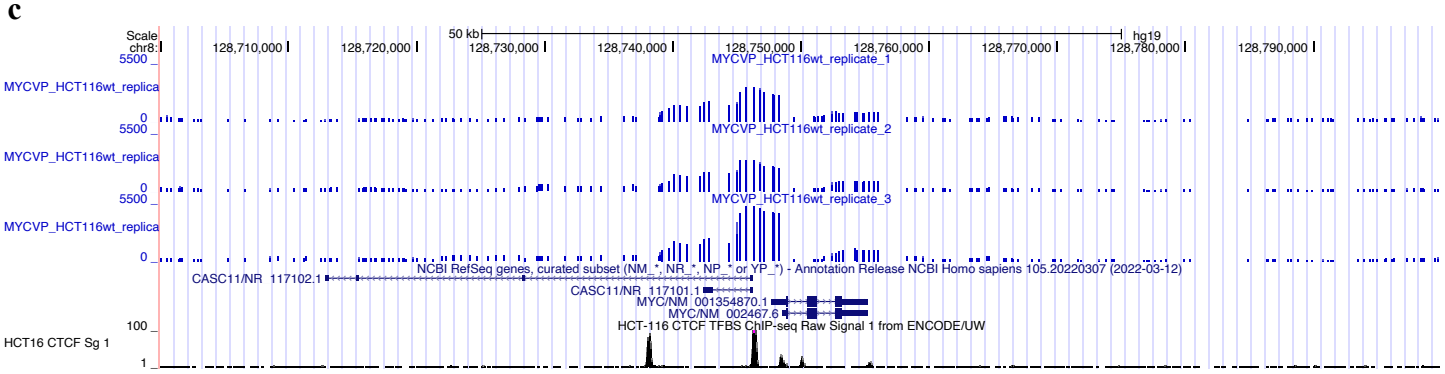
a



b



c



d

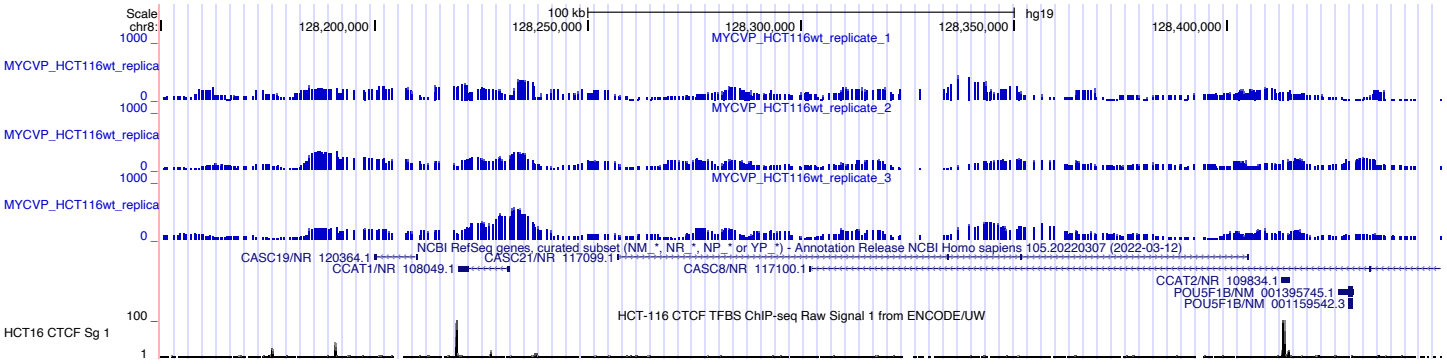
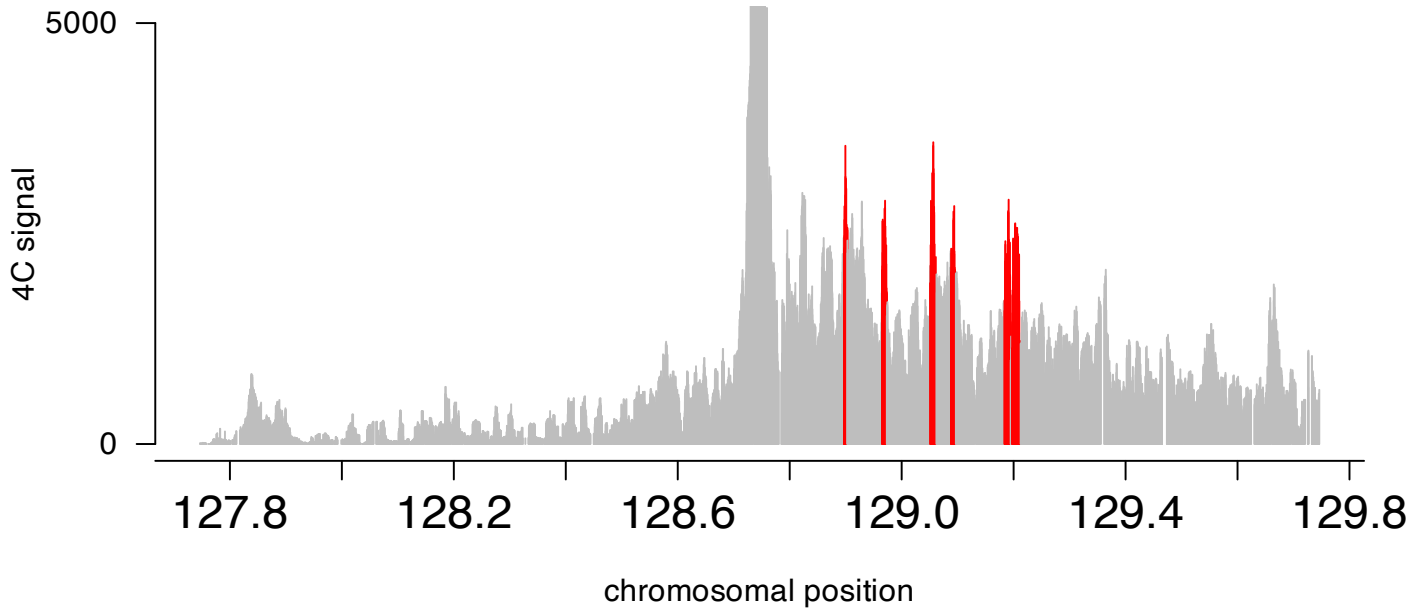


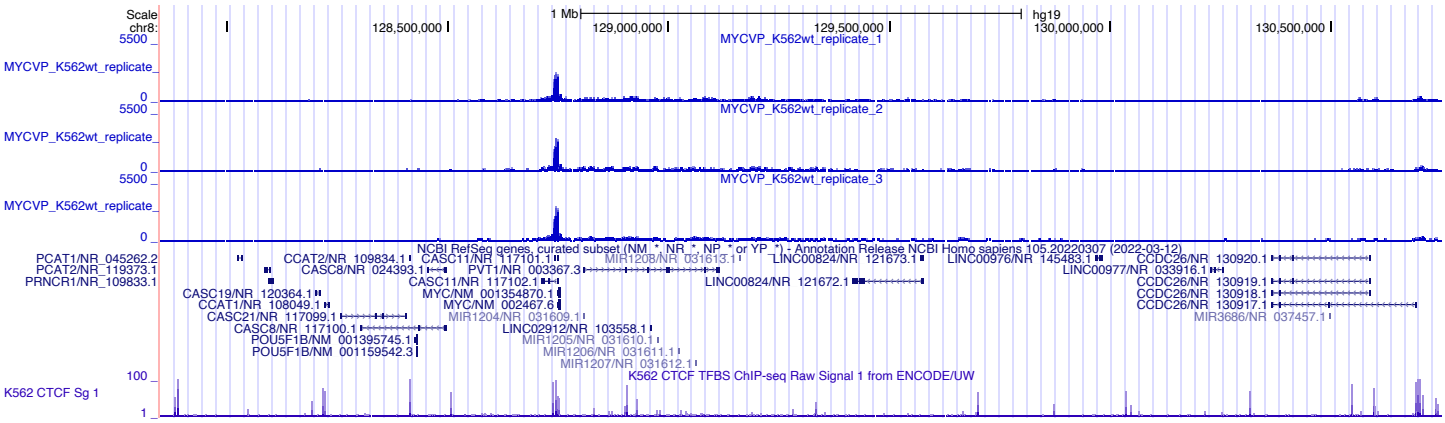
Figure S1: 4C-seq in wild-type HCT116 cells with viewpoint centered at conserved *MYC* promoter CTCF

(a) 4C peak calls (red, peakC R package) shown for 4C signal plotted by genomic position for a triplicate 4C-seq experiment in wild-type HCT116 cells with anchor at the conserved *MYC* promoter CTCF binding site (hg19 chr8:128,746,370). **(b)** UCSC genome browser tracks at the ~2.8 Mb *MYC* TAD region (hg19 chr8:127,850,000-130,750,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(c)** UCSC genome browser tracks showing a 100 kb region surrounding the *MYC* promoter region (hg19 chr8:128,700,000-128,800,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(d)** UCSC genome browser tracks showing a 300 kb region surrounding the -0.53 Mb super enhancer and -0.34 Mb E7 enhancer regions that form loops with the *MYC* promoter CTCF anchor region (hg19 chr8:128,150,000-128,450,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom).

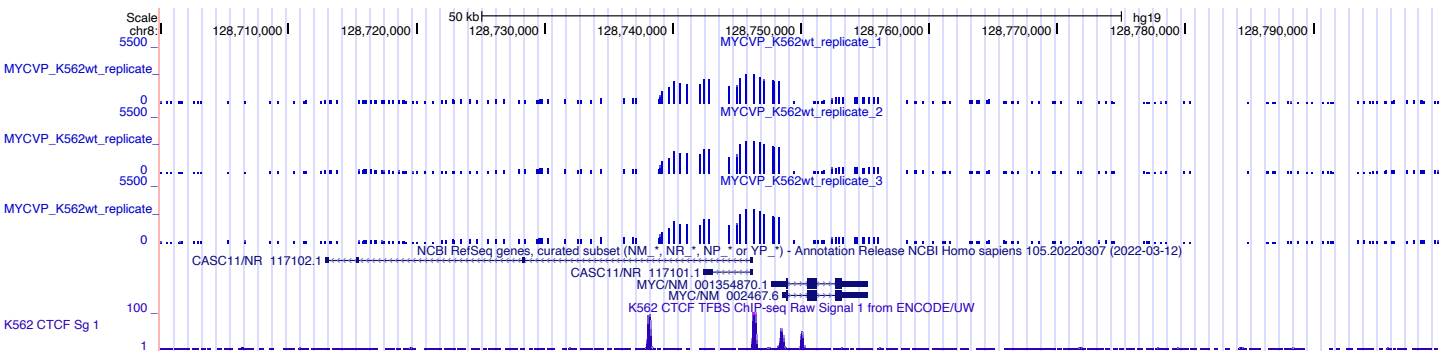
a


5000
0
4C signal
127.8
128.2
128.6
129.0
129.4
129.8
chromosomal position

b


Scale
chr8:
1 Mb
hg19
128,500,000
129,000,000
129,500,000
130,000,000
130,500,000
MYCVP_K562wt_replicate_1
MYCVP_K562wt_replicate_2
MYCVP_K562wt_replicate_3
NCBI RefSeq genes, curated subset (NM_* NR_* NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
PCAT1/NR_045262.2
PCAT2/NR_119373.1
PRNCR1/NR_109833.1
CCAT2/NR_109834.1
CASC11/NR_117101.1
MIR1208/NR_031613.1
LINC00824/NR_121673.1
LINC00976/NR_145483.1
CCDC26/NR_130920.1
CASC8/NR_024393.1
PVT1/NR_003367.3
LINC00977/NR_033916.1
CASC11/NR_117102.1
LINC00824/NR_121672.1
CCDC26/NR_130919.1
CASC19/NR_120364.1
MYC/NM_001354870.1
CCDC26/NR_130918.1
CCAT1/NR_108049.1
MYC/NM_002467.6
CCDC26/NR_130917.1
CASC21/NR_117099.1
MIR1204/NR_031609.1
MIR3686/NR_037457.1
CASC8/NR_117100.1
LINC02912/NR_103558.1
POU5F1B/NM_001395745.1
MIR1205/NR_031610.1
POU5F1B/NM_001159542.3
MIR1206/NR_031611.1
MIR1207/NR_031612.1
K562 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
K562 CTCF Sg 1

c


Scale
chr8:
50 kb
hg19
128,710,000
128,720,000
128,730,000
128,740,000
128,750,000
128,760,000
128,770,000
128,780,000
128,790,000
MYCVP_K562wt_replicate_1
MYCVP_K562wt_replicate_2
MYCVP_K562wt_replicate_3
NCBI RefSeq genes, curated subset (NM_* NR_* NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
CASC11/NR_117102.1
CASC11/NR_117101.1
MYC/NM_001354870.1
MYC/NM_002467.6
K562 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
K562 CTCF Sg 1

d

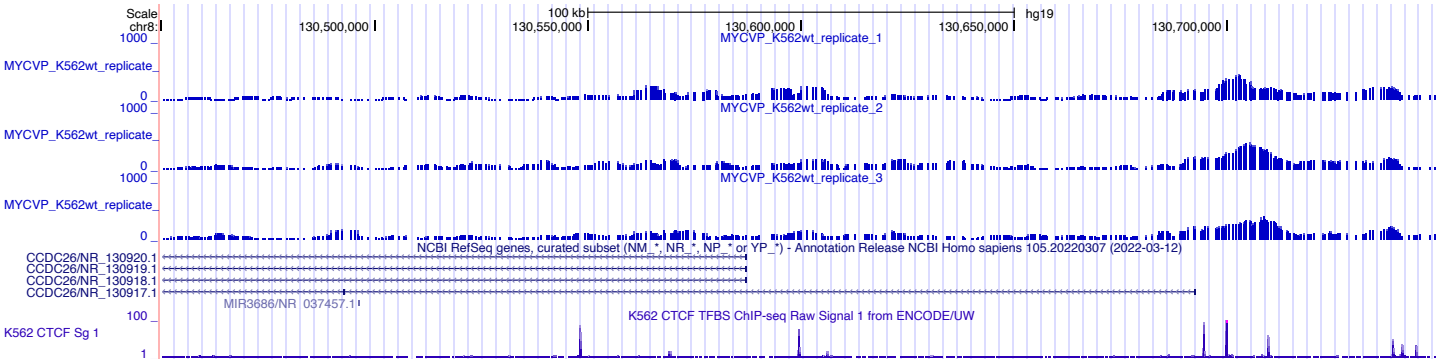
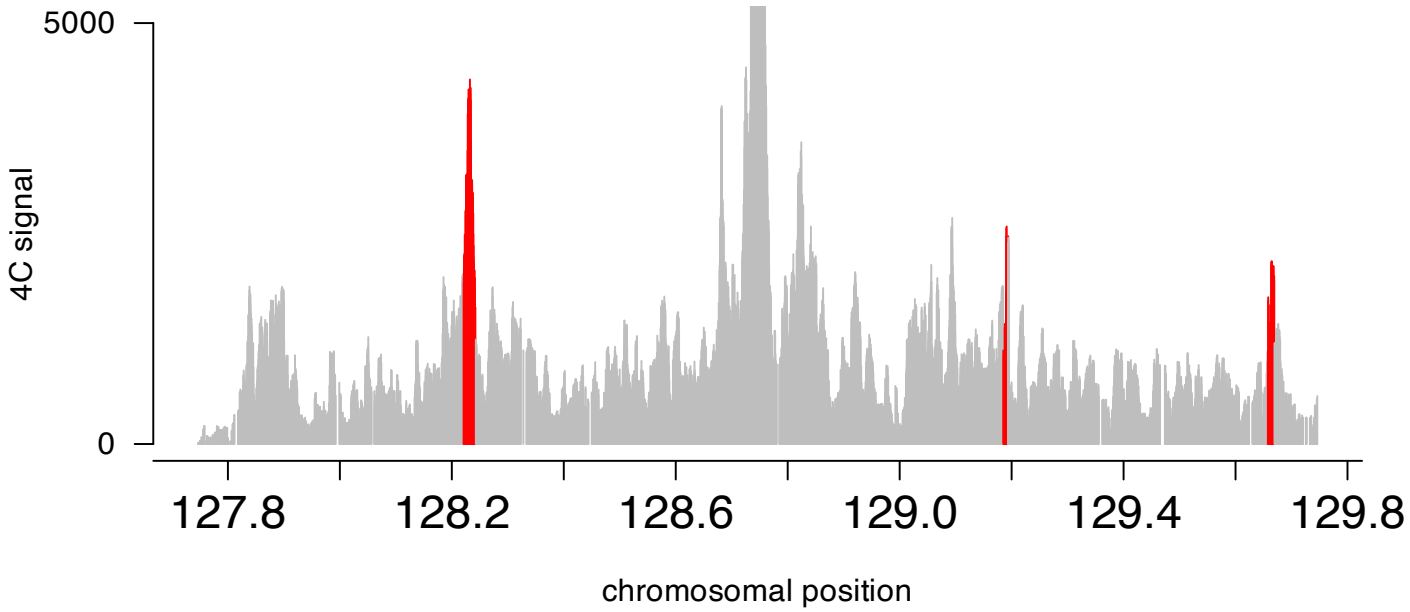


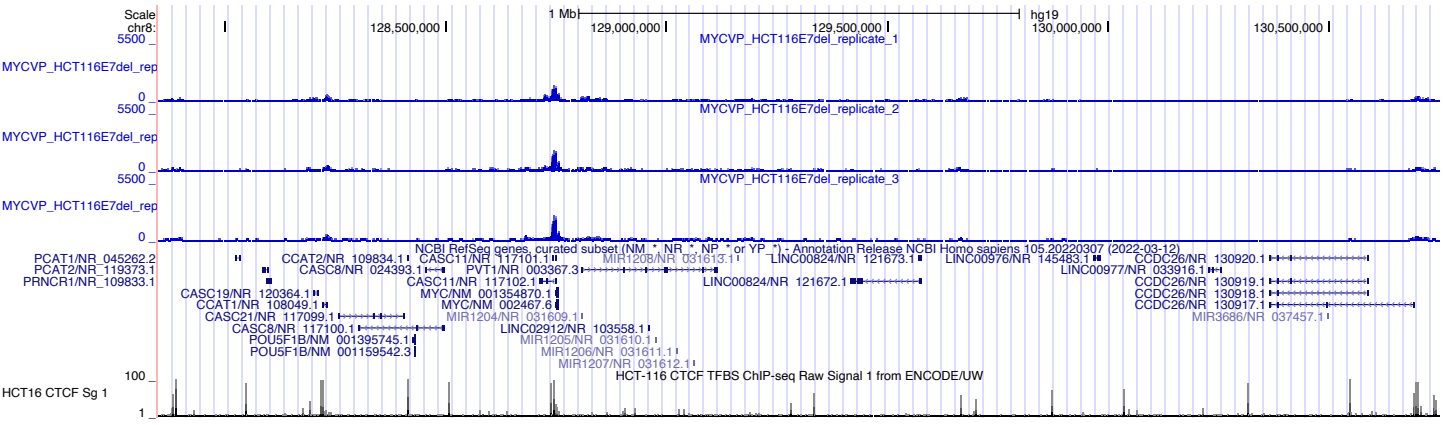
Figure S2: 4C-seq in wild-type K562 cells with viewpoint centered at conserved *MYC* promoter CTCF

(a) 4C peak calls (red, peakC R package) shown for 4C signal plotted by genomic position for a triplicate 4C-seq experiment in wild-type K562 cells with anchor at the conserved *MYC* promoter CTCF binding site (hg19 chr8:128,746,370). **(b)** UCSC genome browser tracks at the ~2.8 Mb *MYC* TAD region (hg19 chr8:127,850,000-130,750,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE K562 CTCF ChIP-seq data (bottom). **(c)** UCSC genome browser tracks showing a 100 kb region surrounding the *MYC* promoter region (hg19 chr8:128,700,000-128,800,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE K562 CTCF ChIP-seq data (bottom). **(d)** UCSC genome browser tracks showing a 300 kb region surrounding the +1.85 Mb and +1.95 Mb super enhancer regions that form loops with the *MYC* promoter CTCF anchor region (hg19 chr8:130,450,000-130,750,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE K562 CTCF ChIP-seq data (bottom).

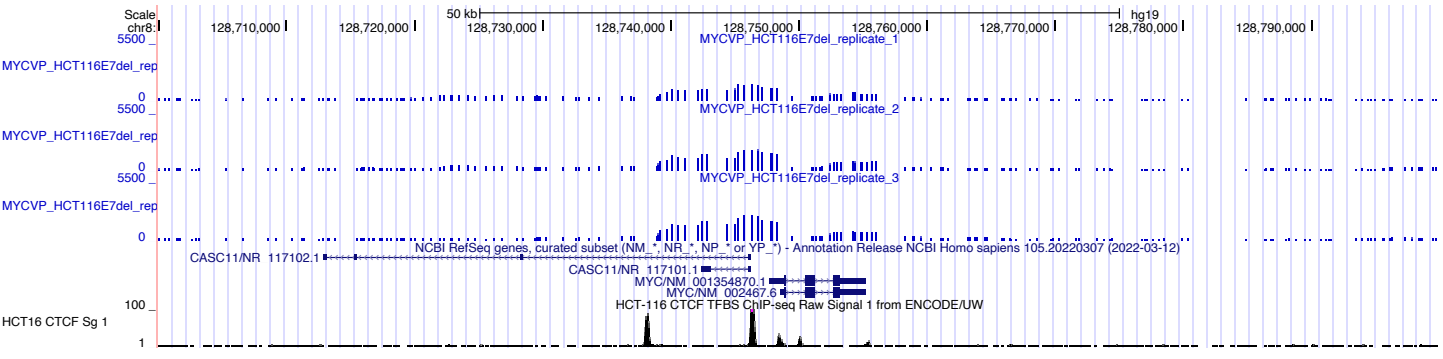
a


5000
4C signal
0
127.8
128.2
128.6
129.0
129.4
129.8
chromosomal position

b


Scale
chr8:
1 Mb
hg19
128,500,000
129,000,000
129,500,000
130,000,000
130,500,000
MYCVP_HCT116E7del_replicate_1
MYCVP_HCT116E7del_replicate_2
MYCVP_HCT116E7del_replicate_3
NCBI RefSeq genes, curated subset (NM_* NR_* NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
HCT-116 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
HCT16 CTCF Sg 1

c


Scale
chr8:
50 kb
hg19
128,710,000
128,720,000
128,730,000
128,740,000
128,750,000
128,760,000
128,770,000
128,780,000
128,790,000
MYCVP_HCT116E7del_replicate_1
MYCVP_HCT116E7del_replicate_2
MYCVP_HCT116E7del_replicate_3
NCBI RefSeq genes, curated subset (NM_* NR_* NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
HCT-116 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
HCT16 CTCF Sg 1

d

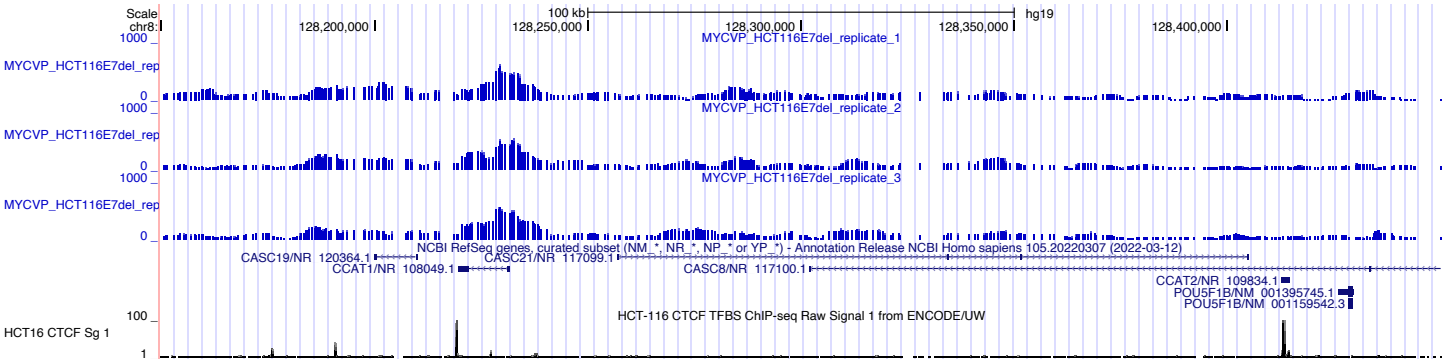
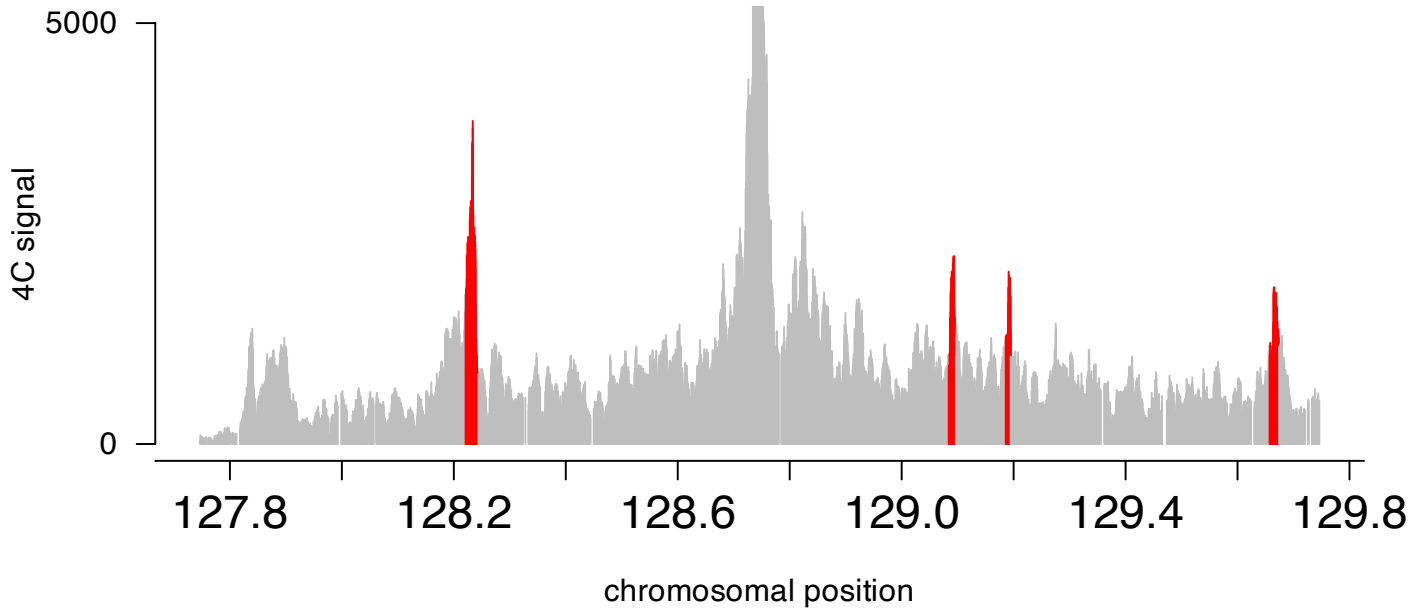


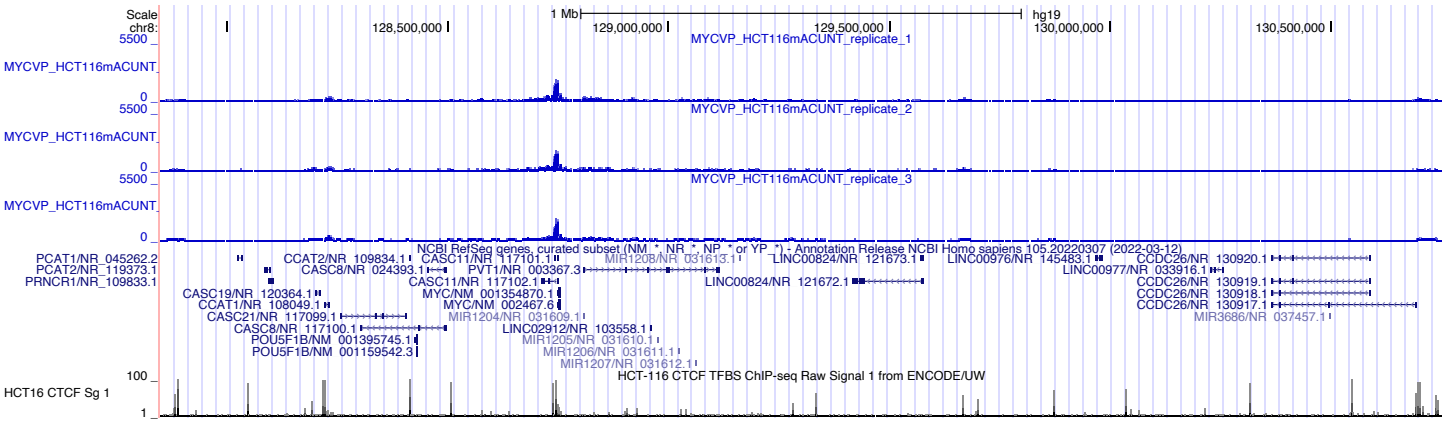
Figure S3: 4C-seq in E7 enhancer deletion HCT116 cells with viewpoint centered at conserved *MYC* promoter CTCF

(a) 4C peak calls (red, peakC R package) shown for 4C signal plotted by genomic position for a triplicate 4C-seq experiment in E7 enhancer deletion HCT116 cells with anchor at the conserved *MYC* promoter CTCF binding site (hg19 chr8:128,746,370). **(b)** UCSC genome browser tracks at the ~2.8 Mb *MYC* TAD region (hg19 chr8:127,850,000-130,750,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(c)** UCSC genome browser tracks showing a 100 kb region surrounding the *MYC* promoter region (hg19 chr8:128,700,000-128,800,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(d)** UCSC genome browser tracks showing a 300 kb region surrounding the -0.53 Mb super enhancer and -0.34 Mb E7 enhancer regions that form loops with the *MYC* promoter CTCF anchor region (hg19 chr8:128,150,000-128,450,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom).

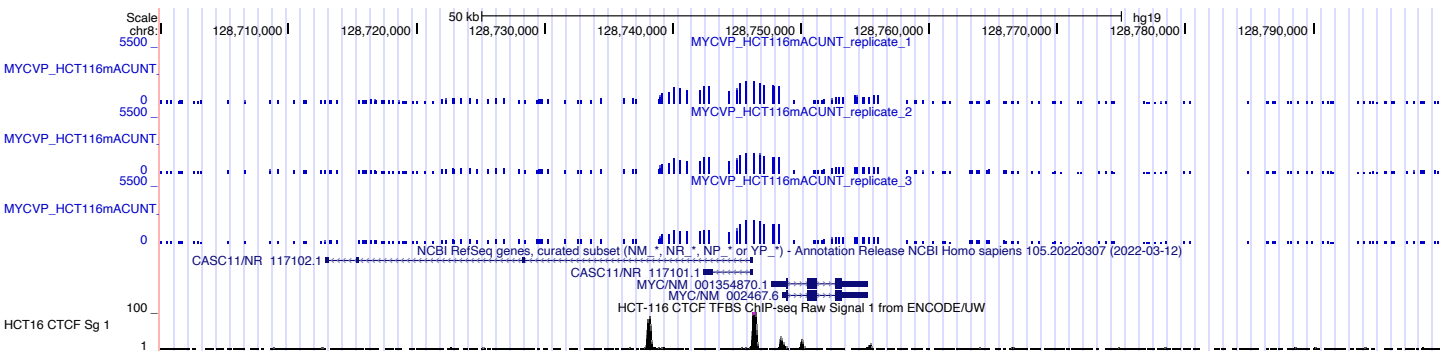
a


5000
0
4C signal
127.8
128.2
128.6
129.0
129.4
129.8
chromosomal position

b


Scale
chr8:
1 Mb
hg19
128,500,000
129,000,000
129,500,000
130,000,000
130,500,000
MYCVP_HCT116mACUNT_replicate_1
MYCVP_HCT116mACUNT_replicate_2
MYCVP_HCT116mACUNT_replicate_3
NCBI RefSeq genes, curated subset (NM_* NR_* NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
PCAT1/NR_045262.2
PCAT2/NR_119373.1
PRNCR1/NR_109833.1
CCAT2/NR_109834.1
CASC8/NR_024393.1
CASC11/NR_117102.1
CASC11/NR_117101.1
MIR1208/NR_031613.1
PVT1/NR_003367.3
LINC00824/NR_121673.1
LINC00824/NR_121672.1
LINC00976/NR_145483.1
LINC00977/NR_033916.1
CCDC26/NR_130920.1
CCDC26/NR_130919.1
CCDC26/NR_130918.1
CCDC26/NR_130917.1
MIR3686/NR_037457.1
CASC19/NR_120364.1
CCAT1/NR_108049.1
CASC21/NR_117099.1
CASC8/NR_117100.1
MYC/NM_002467.6
MYC/NM_001354870.1
MIR1204/NR_031609.1
LINC02912/NR_103558.1
MIR1205/NR_031610.1
MIR1206/NR_031611.1
MIR1207/NR_031612.1
POU5F1B/NM_001395745.1
POU5F1B/NM_001159542.3
HCT-116 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
HCT16 CTCF Sg 1

c


Scale
chr8:
50 kb
hg19
128,710,000
128,720,000
128,730,000
128,740,000
128,750,000
128,760,000
128,770,000
128,780,000
128,790,000
MYCVP_HCT116mACUNT_replicate_1
MYCVP_HCT116mACUNT_replicate_2
MYCVP_HCT116mACUNT_replicate_3
NCBI RefSeq genes, curated subset (NM_* NR_* NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
CASC11/NR_117102.1
CASC11/NR_117101.1
MYC/NM_001354870.1
MYC/NM_002467.6
HCT-116 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
HCT16 CTCF Sg 1

d

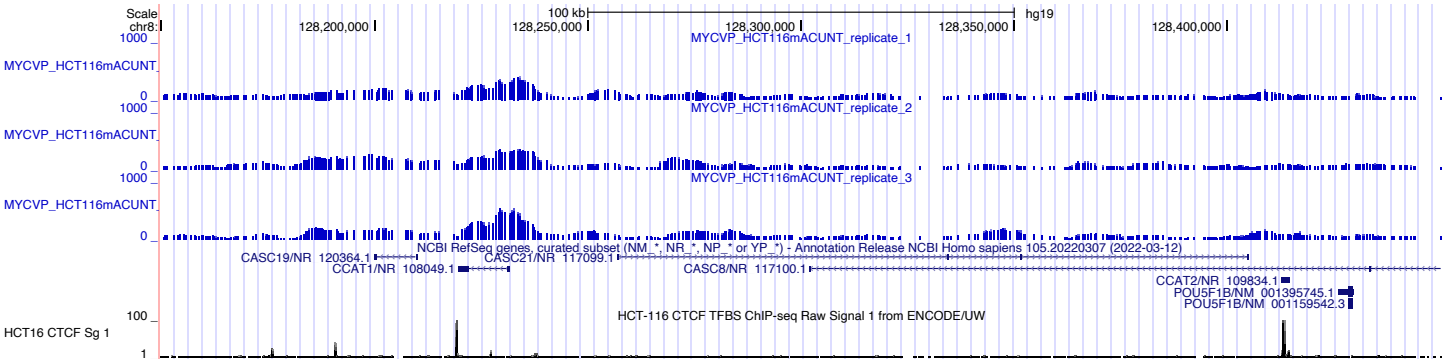
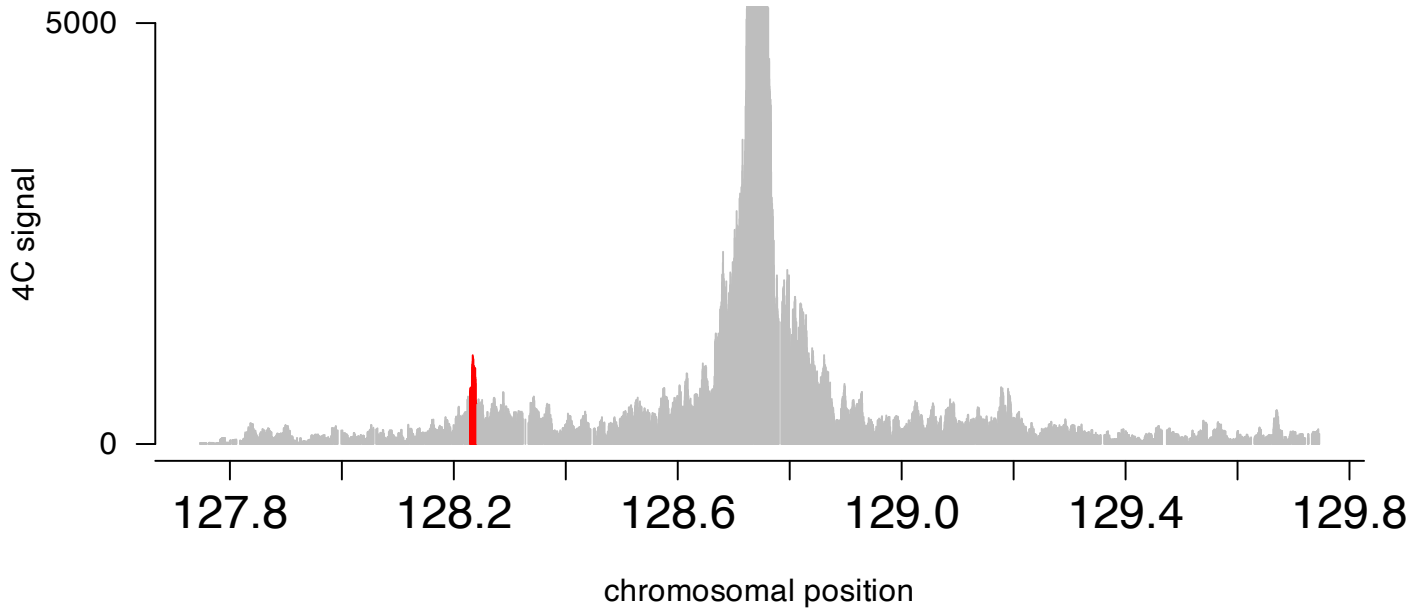


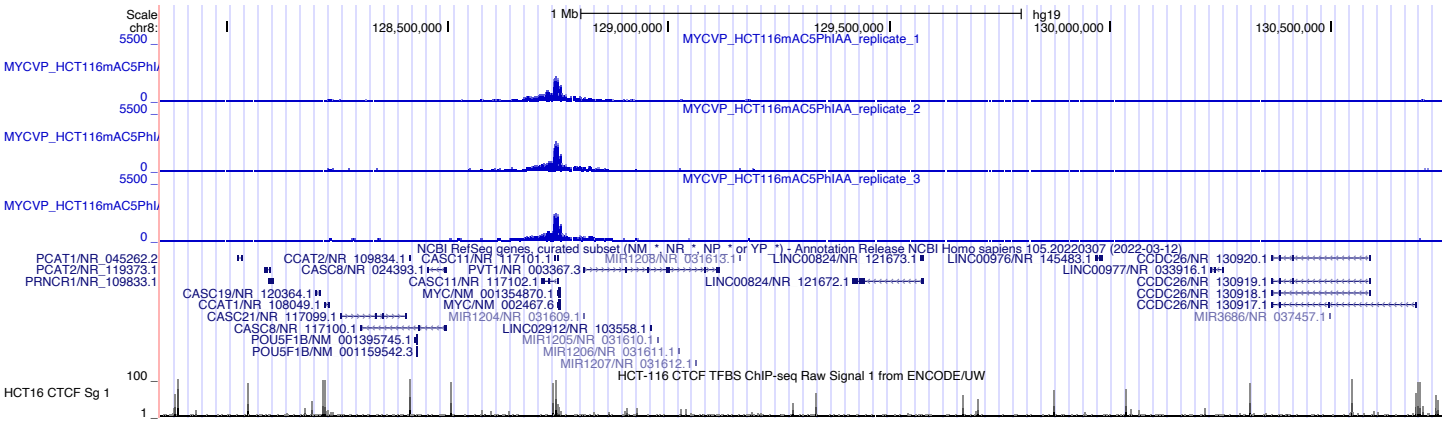
Figure S4: 4C-seq in untreated HCT116-RAD21-mAC OsTIR1(F74G) cells with viewpoint centered at conserved *MYC* promoter CTCF

(a) 4C peak calls (red, peakC R package) shown for 4C signal plotted by genomic position for a triplicate 4C-seq experiment in HCT116-RAD21-mAC OsTIR1(F74G) cells with anchor at the conserved *MYC* promoter CTCF binding site (hg19 chr8:128,746,370). **(b)** UCSC genome browser tracks at the ~2.8 Mb *MYC* TAD region (hg19 chr8:127,850,000-130,750,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(c)** UCSC genome browser tracks showing a 100 kb region surrounding the *MYC* promoter region (hg19 chr8:128,700,000-128,800,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(d)** UCSC genome browser tracks showing a 300 kb region surrounding the -0.53 Mb super enhancer and -0.34 Mb E7 enhancer regions that form loops with the *MYC* promoter CTCF anchor region (hg19 chr8:128,150,000-128,450,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom).

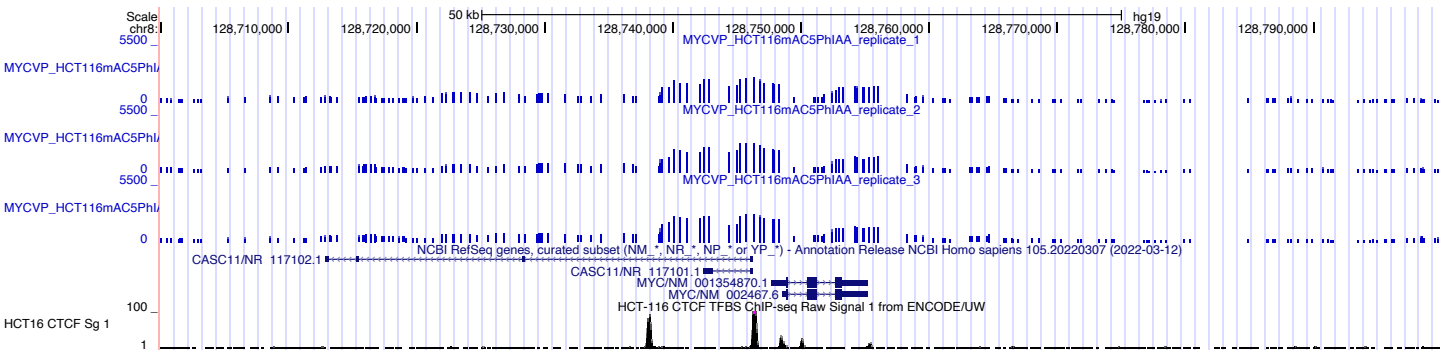
a


5000
0
4C signal
127.8
128.2
128.6
129.0
129.4
129.8
chromosomal position

b


Scale
chr8:
1 Mb
hg19
128,500,000
129,000,000
129,500,000
130,000,000
130,500,000
MYCVP_HCT116mAC5PhIAA_replicate_1
MYCVP_HCT116mAC5PhIAA_replicate_2
MYCVP_HCT116mAC5PhIAA_replicate_3
NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
HCT-116 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
HCT16 CTCF Sg 1

c


Scale
chr8:
50 kb
hg19
128,710,000
128,720,000
128,730,000
128,740,000
128,750,000
128,760,000
128,770,000
128,780,000
128,790,000
MYCVP_HCT116mAC5PhIAA_replicate_1
MYCVP_HCT116mAC5PhIAA_replicate_2
MYCVP_HCT116mAC5PhIAA_replicate_3
NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-12)
HCT-116 CTCF TFBS ChIP-seq Raw Signal 1 from ENCODE/UW
HCT16 CTCF Sg 1

d

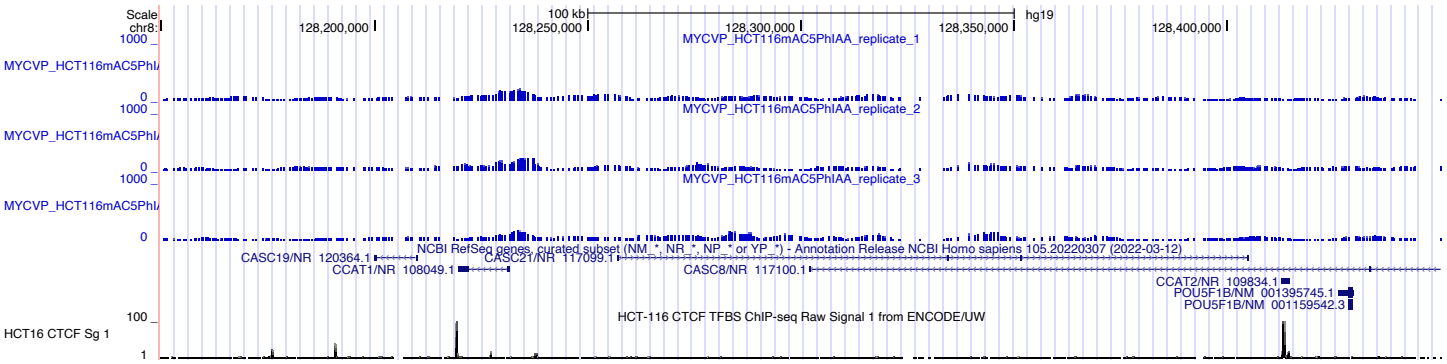


Figure S5: 4C-seq in 120 min 5-Ph-IAA-treated HCT116-RAD21-mAC OsTIR1(F74G) cells with viewpoint centered at conserved *MYC* promoter CTCF

(a) 4C peak calls (red, peakC R package) shown for 4C signal plotted by genomic position for a triplicate 4C-seq experiment in 120 min 5-Ph-IAA-treated HCT116-RAD21-mAC OsTIR1(F74G) cells with anchor at the conserved *MYC* promoter CTCF binding site (hg19 chr8:128,746,370). **(b)** UCSC genome browser tracks at the ~2.8 Mb *MYC* TAD region (hg19 chr8:127,850,000-130,750,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(c)** UCSC genome browser tracks showing a 100 kb region surrounding the *MYC* promoter region (hg19 chr8:128,700,000-128,800,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom). **(d)** UCSC genome browser tracks showing a 300 kb region surrounding the -0.53 Mb super enhancer and -0.34 Mb E7 enhancer regions that form loops with the *MYC* promoter CTCF anchor region (hg19 chr8:128,150,000-128,450,000) displaying 4C-seq signal as custom tracks for three replicates (top) compared to a track displaying ENCODE HCT116 CTCF ChIP-seq data (bottom).

Experiment 1: HCT116 and K562 *MYC* TAD Boundary and *MYC* Promoter-Super Enhancer Loop Biosensing sgRNA Designs (Based on CTCF Homodimerization and CTCF/Cohesin-Mediated Loop Extrusion Model)

CTCF orientation key: + strand; - strand

Conserved CTCF site upstream of MYC:

TCCCCCGCTGGAAACCTTGCACCTCGGACGCTCCTGCTCCTGCCCCCACCTGACCCCCGCCCTCGTTGACATCC
AGGCGCGATGATCTCTG**CTGCCAGTAGAGGGCACAC**TTACTTTACTTTCGCAAACCTGAACGCGGGTGCTGCCC
AGAGAGGGGGCGGAGGGAAAGACGCTTTGCAGCAAAATCCA

Region 1: Upstream of *MYC* promoter:
hg19 chr8:128745485-128746829

Top conserved CTCF site upstream of *MYC* (polarity: + strand):
hg19 chr8:128746351-128746370

dSpCas9 sgRNAs:

CTATTAAAAAAAATCACTTATTATTCACCAGCCCAATATTTTAAAAGTAAAAATAATAAGCCAAggccaggagc
gatgactcgcacttgtattcccagcagtttcagaggcaaaggccgaaggatcgctttaaccgaggagtttgaga
ccagcctgggcaacatgaccagactgcctctctacaaaaagtttaaaaaattaaccgggtgtggtggtgcactg
cactcccagctactgggctggggtatcaggctgaggtaggaggtttgctttgagcccgggggggatcgaggctgc
agtgagctttgattgtgccactgcactccagcctgggtgacagaaggagaccctgtctcaaaaataataagaaT
AATAATTAATAATAATAGGCCAAACCAAATACCCATCACCTTCTGCTGTGCCTCCCCTTTCCCCAATAAATCCA
GTGTCTTGCTTTCAAATTTTGTGGTTAAAAAAGATGATGAGTTTCTAAGACGTGGGGGCTAAAGCTTGTTTGGC
CGTTTTAGGGTTTGTTGGAATTTTTTTTTCGTCTATGTACTTGTGAATTATTTCACGTTTGCCATTACCGGTTC
TCCATAGGGTGATGTTCATTAGCAGTGGTGATA*GGTTAATTTTCACCATCTCTTATGCGGTTGAATAGTCACCT*
*CTGAACCACTTTTTCCTCCA*GTAACT*CCTCTTTCTTCGGACCTTCTGCA*GCCAACCTGAAAGAATAACAAGGAG
*GTGGCTGGAAACTTGTTTTAAGGAAC*CGCCTGTCCTTCCCCCGCTGGAAACCTTGCACCTCGGACG*CTCCTGCT*
*CCTGCCCCCA**CCT**GACCCC*CGCCCTCG*TTGACA*TCCAGGCGCGATGATCTCTG**CTGCCAGTAGAGGGCACAC***TT
*ACTTTACTTTCGCAAA**CCT**GAACGCGGGTGCTGCCCAGA*GAGGGGGCGGAGGGAAAGACGCTTTGCAGCAAAAT
*CCAGCATAGCGAT*TGGTTGCTCCCCGCGTTTGCGGCAAAGGCCTGGAGGCAGGAGTAATTT*GCAATCCTTAAAG*
*CTGAATTGTGCAGTG*CATCGGATTTGGAAGCTACTATATTCACTTAACACTTGAACGCTGAGCTGCAAACTCAA
CGGGTAATAACCCATCTTGAACAGCGTACATGCTATACACGCACCCCTTTCCCCCGAATTGTTTTCTCTTTTGG
AGGTGGTGGAGGGAGAGAAAAGTTTACTTAAAATGCCTTTGGGTGAGGGACCAAGGATGAGAAGAATGTTTTTT
GTTTTTCATGCCGTGGAATAACACAAAATAAAAAATCCCGAGGGAATATACATTATATATTAAATATAGATCAT
TTCAGGGAGCAA

Lowest to highest CFD (Moreno-Mateos 2015), - strand:

gRNA1: GGAGCAACCAATCGCTATGCTGG CFD: 3.15
gRNA2: TGTCAACGAGGGCGGGGGTCAGG CFD: 15.55
gRNA3: TGCAGAAGGTCCGAAGAAAGAGG CFD: 39.43
gRNA4: TGGAGGAAAAAGTGGTTCAGAGG CFD: 435.35
gRNA5: TCTGGGCAGCACCCGCGTTCAGG CFD: not shown
gRNA6: CACTGCACAATTCAGCTTTAAGG CFD: not shown

gRNA-Cas – tandem 6 bp – gRNA-Cas – tandem 109 bp – gRNA-Cas – 19 bp – **CTCF site** – 18 bp – gRNA-Cas – tandem 35 bp – gRNA-Cas – tandem 43 bp – gRNA-Cas

Total: 142 bp to distal grey site, 166 bp to distal pink site (from ends of CTCF binding site)

dSaCas9 sgRNAs:

CTATTAAAAAAAATCACTTATTATTCACCAGCCCAATATTTTAAAAGTAAAAATAATAAGCCAAggccaggagc
gatgactcgcacttgtattcccagcagtttcagaggcaaaggccgaaggatcgctttaaccgaggagtttgaga
ccagcctgggcaacatgaccagactgcctctctacaaaaagtttaaaaaattaaccgggtgtggtggtgcactg
cactcccagctactgggctggggtatcaggctgaggtaggaggtttgctttgagcccgggggggatcgaggctgc
agtgagctttgattgtgccactgcactccagcctgggtgacagaaggagaccctgtctcaaaaataataagaaT
AATAATTAATAATAATAGGCCAAACCAAATACCCATCACCTTCTGCTGTGCCTCCCCTTTCCCCAATAAATCCA
GTGTCTTGCTTTCAAATTTTGTGGTTAAAAAAGATGATGAGTTTCTAAGACGTGGGGGCTAAAGCTTGTTTGGC
CGTTTTAGGGTTTGTTGGAATTTTTTTTTCGTCTATGTACTTGTGAATTATTTCACGTTTGCCATTACCGGTTC
TCCATAGGGTGATGTTCATTAGCAGTGGTGATA*GGTTAA*TTTTCACCATCTCTTATGCGGTTGAATAGTCACCT
CTGAACCACTTTTTCCTCCAGTA**ACTCCT***C**TTTCTTCGGA*CCTTCTGCAGCCAACCTGAAAGAAT*AACAAGGAG*
*GTGGCTGGAAACTTGTTTTAAGGAACCGCCTGTCCT*TCCCCCGCTGGAAACCTTGCACCTCGGACGCTCCTGCT
CCTGCCCCCACCTG**ACCCCC**GCCCTCGTTGACATCCAGGCGCGATGATCTCTG**CTGCCAGTAGAGGGCACAC**TT
ACTTTACTTTCGCAAACCTGAACGCGGGTGCTGCCCAGAGAGGGGGCGGAGGGAAAGACGCTTTGCAGCAAAAT
CCAG*CATAGCGATTGGTTGCTCCCC*GCGTTTGCGGCAAAGGCCTGGAGGCAGGAGTAATTTGCAATCCTTAAAG
CTGAATTGTGCAGTGCATCGGATTTGGAAGCTACTAT**ATTCAC**TTAACACTTGAACGCTGAGCTGCAAACTCAA
CGGGTAATA**ACCCAT**CTTGAACAGCGTACATGCTATACACGCACCCCTTTCCCCCGAATTGTTTTCTCTTTTGG
AGGTGGTGGAGGGAGAGAAAAGTTTACTTAAAATGCCTTTGGGTGAGGGACCAAGGATGAGAAGAATGTTTTTT
GTTTTTCATGCCGTGGAATAACACAAAATAAAAAATCCCGAGGGAATATACATTATATATTAAATATAGATCAT
TTCAGGGAGCAA

Lowest to highest efficiency (Najm 2018), - strand:

gRNA2: CGCCTGGATGTCAACGAGGGCGGGGGT Score: 0.66
gRNA3: CTGCAGAAGGTCCGAAGAAAGAGGAGT Score: 0.6
gRNA4: ATAGCATGTACGCTGTTCAAGATGGGT Score: 0.54
gRNA5: GGGGAGCAACCAATCGCTATGCTGGAT Score: 0.47
gRNA6: AGCTCAGCGTTCAAGTGTTAAGTGAAT Score: 0.38

+ strand:

gRNA1: TTTTCACCATCTCTTATGCGGTTGAAT Score: 0.71

gRNA-Cas – inverted 31 bp – gRNA-Cas – tandem 113 bp – gRNA-Cas – 12 bp – **CTCF site** – 74 bp – gRNA-Cas – tandem 87 bp – gRNA-Cas – tandem 20 bp – gRNA-Cas

Total: 194 bp to distal pink site; 233 bp to distal grey site (from ends of CTCF binding site)

Region 2: Large HCT116 super enhancer peak 0.526 Mb upstream of *MYC* promoter CTCF
hg19 chr8:128218546-128220766

Top CTCF site peak (polarity: + strand)
hg19 chr8:128220222-128220241

dSpCas9 sgRNAs:

GGTTTGTAACTGGTGAGAGGAAGTATTGAGCAATTTCTTGTAAAGTTACAATAGCCAGATCTGCAAAGGAAACC
CCTCCCAGCCTGGAATTCCAGAGTATTTCCTAGAGCTTGAAAGGTAGCCAAAAAATGAGCAGGCATTTAAAAGG
GAAGGGAATGGGGGGAAACCTGGAGTGAAGAACATGAATCCATCTGTCTTTCCCATCCACTACCCCATCCGCCT
CTTTCATAAATAGTTTTAAGAGGAACTAATAGACCTCTACTTAATATCTTCTCTCGTACCTTCTCACTTTTGTC
TTCCTCATGCCTTCTCTTCTAGTGTGGCTCAGTGCTTAGTGCAGGAGAAACCAGTAACCCTGAATTGCAGCTCT
TCCTTTCTCACTACATGAATtcaatttcctcatctgtcaaatgagtataccaattcatatctcgaaagctgttg
ctgtgagaatcagataagcataacctcacagcttatgtctattagaacagcacttggcacatggtaaacactcC
AAAGTATTTGTTAAATGAATGAATAGATTAAAAGGTGGCATGTTTTGTACTAAACTGTTCAATGATAGTGTAAA
ACCATTTGGTCATAATGCGGAAAGGGAAGTAAGGCGGAATTCCTTTAATCTGTGTTTTACGCAGGTTCCAAAGG
AGCGTGGTGGAGAGAAGGATGCAGATAGTCTGGGTGAGAGCTAGAGGCTGGAGTCAGCAGGAAGGACTGAGGCC
GTTGGTGCTTGGGGAGTGAGGGCTCCTTTCTGCTCTGTCCTAGGCTAAGTTCCCCACCCATTCCTTCTTGAGAT
CTACCTCAAACACAAATCCCTCAA**TTGACCACAGGGGGCGCCC**CTTCTATGAATTTGGCGCTGATAGCTGTGAT
CTGCCCAGCACAGTGGGGAAAACACAAAATTTACAGATCAGGCATGTCCGGGCTCAGATTCCTA**CTCCAGCACC**
TGGTGGCCAAGGGACCCCAACTGTTAAATAGGCATGGTGATGCCTGCTTTCCAAGCCTGTTgggaaagagagag
gggagcggggaggaatgggagagagagagagaCTGAGCAAGCATGCCAAGACttaatatacttatatttatatt
aaaagaaataaatatCAGATGATTACAATTTGGTTGAACTAAGATACACAGTAGAATATGGAACTAATATCCAA
TATCACAAAGTATTCTAGCGAGCCTTCCTACAGAAAGAATTGTGGGTGGCTGGGGAGTAGGCATTAGCTACTAT
GTGAGTGCAGAGAATACTCAGCCTTCTTCCAGATGGTGAGCTAAAGTTCAAAGATCAAGTCACGTACACACCTT
CTTTCTCATCCCAGGTCCTAGTCTGCTTGAATTCAAATGGCCATCCACACCTTGCCTGAAATACTTGCAATAAT
TAAGATACGGCTTTCTGCCTGCTTGGGGTTTGGTCCACAATTCCCTTAAGAGGCCTCATTTCAATTAGGACTCA
CACATCCCTTCAACAGTAATTTTGTGTCAGGCTTGGTTAGCAACTCAAGGCTCAAGCATAAATGGGACAGAATT
CTTTTCCTTTTGAAACTCA**CCA**ATATAGTGATTGTAGCAACTAGCTACATTGTTTTTGTTTTTTTTTT**CCCCCT**
CAATTctaagcactatgcaaaggctttaaagcagtggtcccaagcctttttggcaccagggaccagttttgtgg
aagacaattttgtgaaaagacaaaatgtggaagaccgtggactgggatggtttggggatgattcaagcacatta
catttgttgtgcactgtgtttctattattattacattgtattatataatgaaataattatacaactcaccataa
tgtagaatcagtggaagccctgagcttgtttcctgcaactagacactcccatctaggggtgatgggagacggtg
acaggtcattaggcattagattctcataaggagcgcacaacctagatccctcgcatgtgcagttcatgacaggg
tttgtgctgctatgagaatttaatgccactgctgatctgacaggaggtggagctcaggcagtaaggtgagcaat

ggggagcagctgtaaataacgctgatctcactcacccactgctcacctcctgctgtgtggcccagttcctaaca
ggccacaaaatggtacctgtctgtgtccccagggttggggaccactgccttaaaggccttcatctcattcagtt
ttcatcaaaattctgtgtggtaggtactctcattagacccattttatgggtaaggaactgaggtaaaattggtt
atataacttgcctaaaataagtcaagtctctgatgagagggccaggattcaagttcaagcagtctgactccaaa
atcTCAAAGCACTTCT

Lowest to highest CFD (Moreno-Mateos 2015), - strand:

gRNA1: AGTTGCTACAATCACTATATTGG CFD: 12.69
gRNA2: GACACAAAATTACTGTTGAAGGG CFD: 44.36
gRNA3: AGTTGCAGGAAACAAGCTCAGGG CFD: 44.67
gRNA4: cttccactgattctacattatgg CFD: not shown
gRNA5: ccccaaaccatcccagtccacgg CFD: not shown
gRNA6: catagtgcttagAATTGAGGGGG CFD: not shown

gRNA-Cas – tandem 65 bp – gRNA-Cas – tandem 26 bp – gRNA-Cas – 32 bp – **CTCF site** – 40 bp – gRNA-Cas – tandem 85 bp – gRNA-Cas – tandem 0 bp – gRNA-Cas

Total: 526 bp to distal pink site; 167 bp to proximal grey site (upstream of CTCF binding site)

Region 3: HCT116 E7 enhancer 0.333 Mb upstream of *MYC* promoter CTCF hg19 chr8:128412627-128415224

Top CTCF peak (polarity: + strand)
hg19 chr8:128413089-128413107

d*Sp*Cas9 sgRNAs:

TGGAGGCCCTAGCAATTCCGAGGTGATCAGGTGGACTTTCCTGGATGTTCTGGGTCTTGACCTGATTGCTGAAA
AATGAATACAAATTCAGAGAAGAAGAAAGCTAGTATGAGACTACCAAATGATCATCAGACATTTCTTGAACACC
AATTAAATTGCTAGGTATGCTAAAGTTTGCAAAACTGGTATAGACACCAAGAGGGAGGTATCAACAGAGACT**CC**
CCAAGAgctaagaggaaaccaccttggactggaagtcaagagccaaaattctagactctactctgtaattaact
a**cct**atttgaatttggaaaaatcaccatcaactttcccagcctcgttctctgcatctgggaaatgaaggcgtcg
tccaaatgattacaagcttctttt**cct**gctcttattgcatgattccaCTTCCACAGCCCTCCAGCATTTTTTAG
CAGCTGCATCGCTCCATA**GAGCCtgcagagggcacta**gactgggaattagaaaa**cct**gatttcccttccagctc
cacctctgaccaattgcctgaccctggtcaaattgcttaacctcttcctatctcagctccctatccataaaaca
gagggacgaataaactctcctcctaccactaag**agg**tgtag**cca**gagttaaTA**CCC**TCATCGTCCTTTGACCTC
AGCAGATGAA**AGG**CACTGAGAAAAGTACAAAGAATTTTTATGTGCTATTGACTTTATTTTATTTTATGTGGGGG
AGGGAGCCGGCCCCAGCTGGAAAGCTGCTTTCTCTGAATCAAAGGGCAGGAACCCAGCAAGTTTCTCAGGATTG
GGGCCTTAGACTGGGCTGTGTATACAGACAGTGCCAGCCAACCCCACAGTTCAGTTTCCTTTAACCTGGTGCTC
CAGGCAATAACTGTGCAACTCTGCAATTTAACAATGTGTTCTTTGTCCCACAACTGTTCTCGTTTCTCAACTGC
CCAGGTAATATGTTTGGGCCTGTAGGAAGAGTCAAATAGTTAATAAGGGAAGGGTTTGGCATGCCCTACGTAAG
TTCTACCAGCAAGTCCCAACAAGAAGGCATTCTGTGTCTCCTGATTCCTGACCTACCCCCAAAATGTACAAATG
TACAAGGAATGAGCCCACTTTCCCAGCAGGCTGTAATACCAGTTTGGCCTATATCAATGCATTGGTGAGCTGTG

TTTTGTTTATGGTTTTATGCCATCTATTTTCCCATGGATATTATGTTTTCTAAAGAGCCCTTAAGTTTACGTCA
GCTTTTAAAGCTACCAGCAGCACCATTTCAGTTCATATTAAGCCCTTAATATGGTATGAATAGGAGAGCTATTA
GACTAAAGAGCCATAATCATCCCTGAGGAAAACATCCATCACCAACATTTATGTGGTCCCTGAACTTCTAAAAG
GTGTCATCTCTCTGGGGTGTATCTGGTGAGAGCTTTCTCTGGGAGATGCCAAAAAGCCAATGCATTAGATGAAG
CTTAGAAGGGCATTTTCTAACCATTACAAATTGCCTAGTCTAGCATCTCAATTTCATCTACGTGAAGAGCCTTA
ATTAAATTTGTTGGGGTTTGATCCTTTATCCCCAGATGTGGCGCTGACAGAGATTGCTTACATAAATAATGTGT
GCTCCAAGTGCTTGCCAGGCTCCTGGCTCAGCTGGGACAGCTGTAGCTTTTTGAATGTCATTCCCAAGATATCC
TGCAGGTGTTCAGCTTCCCCTGTTCTACTCTGGGAAGAGAGCCGTGGGCAACATCAGCCCAGAAGACACTAGTG
CATCTTGACAGAGCATTCCCACTGGACAGGTGCACTACCTGCAGACACCTCCAACAGTGAAAGCACCCCCCGGC
TGCAAGTATTGTCTCGCCAAACGGAAGGCATTTCTTGTATGAGTCTTGAATCTGATATTCTTTAATAAGACTTT
CTCTAAAACTGTATACACACACGCACCTGAAAACGTTATCAGGCTTTTACTTATTTGACAGTAATATTCCAGGA
AATGGTCCAAGTAGCAAAATCTTCTCTCAGTGGCTTTCTCGTTCCTCATGTCTGACTTGCTTTCCTGCTGTCTA
CTGGAGCCAACAAGAGAGGCATTCTCACTCTGGAAGTTGTTAGTTCATTACAGCCTAGGTTGATCTCATCTGGC
TCTGGTTGGGGGTTTTATTCTGAAGATTGGCTGGTAAAAGCCTCCACAAAGTCTTAGCTAGAACATTCATAGCC
CTGTGTTGACTTCCGCCTCAACTCTGGCTTGAACTTAGCTGACGGTCCTATCAGTTTGCTCAAAGGAGCTTTTT
TGCTGCAggcaacagaaaggagtggaaagaacacactcattggaacccaggagatccaggctcgcctctggctc
taggacttgctcatcatgtgatatcagatacatgacttcatcgctgagtctcagtttccttatctatgaaccgg
ggatattcataccttcttcacaggACAAGAGAGGCACATACAAGTTCATATCTGTCCCTCTATTTCAAGTGATT
CTCCCACATCAGAACCAGTTTGGCTTCCCCGGCCTGGTCTTCTTTCAAAGGATGCACAAGGCCATTTACAACTG
GCCTGAGG

Lowest to highest CFD (Moreno-Mateos 2015), - strand:

gRNA1:TGGTTTCCTCTTAGCTCTTG**GGG** CFD: 0
gRNA2:TGATTTTTCCAAATTCAAAT**AGG** CFD: 0
gRNA3:TGGAATCATGCAATAAGAGC**AGG** CFD: 0
gRNA4:GTGGAGCTGGAAGGGAAATC**AGG** CFD: 0.015

8q24 designs from Coggins *et al* 2017 and Heath *et al* 2022:

g248 (- strand, Coggins 2017):CTGAGCTCAAAGGACGATGA**GGG**
g259 (+ strand, Coggins 2017):**CT**TTGAGCTCAGCAGATGAA**AGG**
gRNA1 (- strand, Heath 2022):GACGATGAGGGTATTAACTC**TGG**
gRNA2 (+ strand, Heath 2022):actctcctccta**cca**ctaag**agg**
gRNA3 (+ strand, Heath 2022):TATTTTATTTTATGTGGGGG**AGG**

gRNA-Cas – tandem 54 bp – gRNA-Cas – tandem 74 bp – gRNA-Cas – 45 bp – **CTCF site** – 17 bp – gRNA-Cas – everted 84 bp – gRNA-Cas (8q24 gRNA2) – inverted 5 bp – gRNA-Cas(8q24 gRNA1) – tandem overlapping – gRNA-Cas (8q24 g248) – everted overlapping – gRNA-Cas (8q24 g259) – tandem 41 bp – gRNA-Cas (8q24 gRNA3)

Total: 239 bp to distal green site; 219 bp to distal grey site (from ends of CTCF binding site)

Region 4: Smaller K562 super enhancer 1.853 Mb downstream of *MYC* promoter CTCF

hg19 chr8:130598258-130600266

Top CTCF peak (polarity: - strand)
hg19 chr8:130599463-130599481

dSpCas9 sgRNAs:

GAATCTCAGATTAAGAAACATTCTGCAAAAATCTCACCAGTACTTTTCTAAAATGTCATTGTCATAAAAGACAG
AAAACTGTTCCAGATTTAAAAAGATGAGACACTAATAAAGTCAATGTGTGATCCAGTGTTTTCTTTTACCATAT
AGCATAAAACTGGAACAATTGGTGAAATTCAACTGAGGTCTGTAGATTAGGTAATAGTTTTGTATTAATGTTAA
TTTCCCAGTTTTGATGATTAACAATAGTGCTTATGTAACAAAATGCATTTTCAGAAAGTACACATCGAAGTCTT
TAGGGATAAAGAACATTATGTCTATACATTTCAAATAATTCAGGAAAAAATATGGCATTATATGAAAATATATG
GTATCATATAAAAATATATGGCATATATAATAGATGGTACATATATATTAAATTAAGTAAAGTTGGTAAATTAT
TAACACGAGGGAAACTCCTTGTACTATTCTGCCAATGTTTTAAAAAGTCTGAAATTTTGTCAAAATAAAAAGTT
AAAAGCAGTAGCAAGAAAGAGGTACAATGTGGATAACTTTAATCAGCCCAGAACCAACACAAAGCATAGCGAAG
AGGTTTGGCTGCACTTTCCAGAAACTCCAGATGCCAAACATGTTTCACCCATTTTTTTATCCCTGTCTTCCAGG
CTTCTGATATTCATGACAAGATGGGAGATGATACTAAATGGGCCCATTTCTTCAGTGATGGTTTCCCATAACTA
AGCAGTGTCGTTAGAACCACCTATATTATTGCCGAGCCAGCTGGAGTCTTAAAAGTTTATTTCCTAGAACACAT
AGTCCACATAATCTCTTCCATTACTGCTGGAAGGCAGCAGGGTACCTCTCCTGTAGTGCCTCATGCCACCACAA
ATAGATCATGCAGAAATCAGGTGGCAATGGGGTCTCTCCTGCTATTTGCACCATATGATGGCTGTTTGTTGTGC
TGGACTCTTCATTTGCCACTTGAGCCACGGTGTTCAGCTCCTATCCGCCCTGGAAAAAAACACTGGCAAAGTCA
GCTGGGCCAGGATGCTTTGCATTTTGAGGAATCATGAAATGGCCAGTCCCAGCTAATGGATAAAAGGTAACTCA
CCATCCCCACAGTGACCCAAGGACTATACTGGACAATAGAGACAAAAATGCATCCCCTCCTCCCAACTGGTGAA
GACAATCTAGAGACAGTGCAAGAGTGCTCCCTTCTGCCCAGAGCTCAAATTGCCCTAACTTCACACAGTTTCAC
CTCAAGAGAGTTTGAGATGAACCCTTCTGCTTCTGGAGCACACATCTTAAATGCTATCCCATAGTCGCCCCTTA
AGCCTGCACTTAAGATTTCACCTTTATGATCTCATTTCATCTCTTCAACAACCTCAATAGCTAAGTAGAGCAGA
TATCATCACACCCTCTAACAGGTGATTTGAAACAGGTGCCTCAATAGTAAATTACTTACCCAACGTCAGGAAGA
CTCCCCACAGAAAGCCTCATCTCCGTATTTAAAATTCAGTGGCTGCAAAGAATGGCTTTAAATTCTCCTGCACC
CCGAAAAGTATAGGCAAGATTCTGCAATTCTGGGTCATCCTCAGTGATGAAGGGGCTGGGGAGGAATGGAAAAC
TGGCAGAGGGACTATCTACAGTCATTATAAGTCAATACAGGCTTTATTAGCTTTCCTGTTGAGATCTGACCTAC
CTAGCGCTATCTTCTTGGTTTCTCATGAATCTACTGTTCTTCACTACCTGCCCTCAGTAGGAGCTGCCAGAATG
GAAATGACTATCAAATCTGAAGAGAAAACAGACGAGACTTGTGCCACAGAGAGCATCGATCCAGGACTACGGGG
TTCAAGGCTCTCAGCTCACTCAGACCATGGGGGTAGGGAAGAGGAGGGGAGGACAGGTGCTTCTTCAGGTACAA
GTGAATTTCCTATCCCTGGAGTTTACAAGCAGAAGCAAGCATAGTAATCCCACCTCTCAGTCTGCACACCTGAG
AAACACACACA

Lowest to highest CFD (Moreno-Mateos 2015), - strand:

gRNA1: TGTTTCAAATCACCTGTTAGAGG CFD: 16.70
gRNA2: GTTCATCTCAAACTCTCTTGAGG CFD: 19.87
gRNA3: GATTGTCTTCACCAGTTGGGAGG CFD: 20.61
gRNA4: TCTGCTCTACTTAGCTATTGAGG CFD: not shown
gRNA5: TCCTTGGGTCACTGTGGGGATGG CFD: not shown
gRNA6: CCTCAAAATGCAAAGCATCCTGG CFD: not shown

gRNA-Cas - tandem 45 bp - gRNA-Cas - tandem 35 bp - gRNA-Cas - 14 bp - **CTCF site** - 33 bp - gRNA-Cas - tandem 103 bp - gRNA-Cas - tandem 11 bp - gRNA-Cas

Total: 193 bp to pink site; 140 bp to grey site (from ends of CTCF binding site)

Region 5: Very small HCT116/K562 super enhancer 8.2 kb upstream of *MYC* promoter CTCF
hg19 chr8:128737780-128738469

Top CTCF peak (polarity: + strand)
hg19 chr8:128738136-128738154

d*Sp*Cas9 sgRNAs:

CTTTGATATTTTCTGTGCCAGCCCATTTTGTGAAATTGCTCATCATGATTCATTAAATTCATTTCTTATTTACT
Atttttaaatttttatacatgcaggggctcaagtgaggatttctttcatgtatacattgcatagtcgtcaagt
ctggtcattaaatgtattcattatccaaatagtgaacattgttaaattgatttcatgatccactaaggggtcat
catttgccattttaaaaCTCTGACAGTATGAGCTTCTCCCTAGCCCAGTTCCTGTTACCATCTTCCCATTCTTC
CCTTCCTTCTTCAATTCAGATAGGATTTTCCTCCAGAGGGATTATAAAGTTGCGAGGAAAGCGCCTGCAGGGGG
TGCTGTTCCACACTGTTGTTGAAGTGTGGTTTGGTTTTTATTTCGTTGCATTTGCTTTTCGGTCAATGAGGGCA
ATTCATCTGGAATGACCCCCATCCTCGTCACCCTTGCTCCAACGATGTTGGGGCCCAGCTCATCAACAAGGACA
CCTGAACAGAGCCCTACCCATTGATGGAACCGAAGCAAGGGCAAGGAAGAGTTCTCAACCCTTCTCTCTATATA
CGATTAAAACTGGGTTAGGCTAGGTGTGCCCTCAGCTCAGAAGCTCTCTCTAATAGcattccttcactaagcac
ttacagagtgcctaccacgtgcca

Lowest to highest CFD (Moreno-Mateos 2015), - strand:

gRNA1:GAGCTGGGCCCCAACATCGTTGG CFD: 1.66, Doench 2016 efficiency score: 52 (distance from CTCF: +107 bp)

gRNA2:GGCAAATGATGACCCCTTAGTGG CFD: 3.45, Doench 2016 efficiency score: 53.54 (distance from CTCF: -126 bp)

gRNA3:TGGTAACAGGAACTGGGCTAGGG CFD: 5.41, Doench 2016 efficiency score: 60.58 (distance from CTCF: -74 bp)

gRNA4:AATCCTATCTGAATTGAAGAAGG CFD: not predicted, Doench 2016 efficiency score: 55.7 (distance from CTCF: -33 bp)

gRNA5:TCGTTGGAGCAAGGGTGACGAGG CFD: not predicted, Doench 2016 efficiency score: 65.83 (distance from CTCF: +91 bp)

gRNA6:CTTCGGTTCCATCAATGGGTAGG CFD: not predicted, Doench 2016 efficiency score: 46.32 (distance from CTCF: +155 bp)

gRNA-Cas – tandem 29 bp – gRNA-Cas – tandem 18 bp – gRNA-Cas – 33 bp – **CTCF site** – 91 bp – gRNA-Cas – tandem overlapping – gRNA-Cas – tandem 25 bp – gRNA-Cas

Total: 126 bp to pink site; 155 bp to red site (from ends of CTCF binding site)

Designs for *MYC* TAD Boundary Region Loop Anchors

Two strong CTCF-CTCF interactions for same TAD boundary

1. **chr8: 127881391-130694709 (ChIA-PET score: 300)**
2. **chr8: 127888565-130700261 (ChIA-PET score: 200)**

Region 6: Left TAD boundary region for *MYC* insulated neighborhood

hg19 chr8:127,880,433-127,890,996

Top CTCF peaks (polarities: + strand)

hg19 chr8:127,881,798-127,881,816 AND chr8:127,888,996-127,889,014

d*Sa*Cas9 sgRNAs:

gttgtaaactttatgcctccttcttcatttaaggatcctagtggttacactgggcccactgcctattctagtat
aatctccctaccacagatgcagctgatttgcaaagttgattccatctgcaaCCTTAATTAGCTCTGGGAGAAGA
TTCCACTCAGAGTTGGGCAGGAGGTAGCTTGAGAGGCAACTGGCCAGGGAATGTAAAAGAGGGCTCAGAAGCAT
TCTTGTCTTTATACTATACATTTGCTCTAACCTACTTCACAACTTTGGCAAGAAGTATATCTTGTTGAAAAACA
TTACTCTGAGACTACAACAGAAAATTCAATATGTAGCAATTGCCATCAAAAGGACAAAACAAACATGCATTAAA
TATTTAAACTACAGCATAAGAAACATCTATTATTAAATGCAACATAGGTTATCAGTGCAATAAATAACAATAAT
GAATAGTTCTCAATGTTCTGTAGGAATGAGctttatgtcaaacaatgagctgagtattttacatacagattatc
ttttaatttttaaaacaactatgaggtaggtgctattgatatttacttttttaatttttttttttagatgaagtc
ttgctctgtcacccaggctggaatgcagtggcacgatctcggctcactgcaacctccgccttccaggttcaagt
gattctcctgcctcagcctcccaagtagctgggactacaagcatgtgccaccacgcccagctaatttttttgtat
ttttagtagatacagggtttcaccgtgttagccaggatggcctcgatctcctgacttcatgagctgcctgcctc
ggcctcccgaagtgttggctcattacaggcatgagccaccactactggcctgatatttacatttttaaacaaga
aaatatagacttaaagaggcttggtatgttgcaaaaggtcacaaagctgtatatggtgaggaggaatttgaacc
tgggtttgattccatcacccgaacccaaccactaggctccactgCCActttaggagttcagcggagagaacaat
cagtgagacctgagtggctaagaaagccttcctgaaggtggtgatatgtgaacACCCGTTTAAAAGGTTGCTCT
GCTGGAATCCTCTTAAAATCTAAGACTAAAGAAAGAACCAGAACTTATTTACCCTAGAGCTACCAAAACCATGT
CACCCCCAGGGCCTGCCTGTGCATTAGGGCAGTGTGTATTAATGAAACATGCTGCTTTTATTTCAACTTCCCAC
TTCCACCTTGGTGAGGGAATTAAAAAAGAAATCACCTTTGTTTTCTTTAGGCCTGTAGATTGGCTTTTTTTTTT
TTTTTTTTGCCCTTCAAGAAACATTTCCAAGCATGGCCAGAAGAGGGTGTGCTTGGGTGTAGGGGAAGGACAGG
TGGGACGTGCTCAGTGTGGGATATGCTCTCAGCATGCTTTAGGGGCTGTTTTAAGGAACATAAATTAATAAGGT
AGAAATTACTTGTATCCCATTAGAGACTCTGCAGATCACGCAAGTCATTCATTCCTTATTAATCATTCACTTAG
TAGGCACTATGAATATGGAAATCAGCAAGACATGGCCCTTTTGAAGATCTAAGAGTCAAAAGAAGAAAATAAAT
AGAAATCATCTTCTATGAACAAACTCAGGCACAGTTATAGCAAACatgtattaagcacagacaaagcatcaggt
ggattaaacatgtggtgctgatattaatcttacactgtagataatattattcacgttccatgaaagaagatcta
aggttggggttgcccaaagccccacagcaagcaagcagattagttaggattcaaatgcagtcctgcctgactcc
aaaacatactttgtactaaaccatgtgataccccCTCATAGCCTTCTGGAGcagggatcagcaaagttttgctt
aaggagccaaacagcagatactttaggctttgccagtgaagagacaaatgaaagatgttatgtagatacatata

taTACATTTAAAATGAAACCAGTCAATAATGTAAAAATTATTTTCAGCTCACAAGCCATAGCAAGGCAGCAAGA
GTTTTCCTATAGACCACAGTTCCAGGGCAATTTCTCTCTTCCATATtcttccatattgtatgtagtgtgtatat
atatgtgtatacatatatacatatatgtatatACGtatatatacgtatacacgtgtatatatgcgtatgtgtat
atatacatgtatatatgcgtatgtgtatatatactacatatatttacttatagagatgtttatatataaacatg
caatatgtatattatactacctatctacatatacagtatgtgtatatCTCCACATATACCTTATCTGACATTCT
AATGTGTGTGTATACATATTGTTACTTCTGCTAGGATGTGATATCTTTGAAGAAAAAGATCATCTAAGCCACCT
TTGCTTCTCTAGAGAGTGCTACATAAATTCTTGCCTATAGTAATGCCTAATAATGATAGTGTTGGAACAGGTGT
AATCCATTATTATTTACCGAAAAGACTCCATGGAGAAGGCCATGAATATTTGTTTTATATTGCACCAAGTCACT
AATTCCGTGTGAGGATAAGATGTGGTTTTGCTGCTTGAAGTGGTTTTGGGAGAGGCTCTAGAATTTCAGAAGTT
GATTGATTTCCAGGCATCGTGACTCCTGGGCTGAGCAGCACAACAGTTTAATGCTTAAGTTCCTGAGTTTCAGA
GATTCACAGATCTGACTCCTAATTTTTTCTTCTGCTACTTTTCTTTCTTGAGAGGGCATGGTTCATAGGTTACA
AATATCTGTGACAGTAATGATGAAAAGTTCatattccctgtgttcttcctgtggtcaggaactggtacggtagc
ttacatgtgtgaaatcatttaattcttatatcaactctcaaaggtaggtcatatcattgcccccatttgttaga
aaaacagaggcacagagaggttaagtaatttgcttaagattacacagctaggaagtgtcagaacgagattcaaa
ccagatgagttggctacagaacctgtactttattattatTAAAACTCCTATGATATTTCTTCTCCAGGGAAATT
TATATGAAAAAGCTGCAATTTTAATAGAACTGGAATGGAGTTTCCCAAACCTTGGGTTTGGAACCTATGGTGGC
TCTATCTTTGATCTTTTTCTCTCTAAAATGATGAGCACGATAAAAGAGCCACATCATCTGCTTAGAAAATATTT
TTATTTTTGAAAAACTTCATTGAGCAAGTGGTAGATGCTGTTGGAGCGTGGAACTCCACACACCTTTCTGTTCT
TTCTGTTCTTCTATTAGCAGTTAGTTAAGTCTCCCCTCATGGAGACTTCATACTTACAGTGGTTTAAGAAGTGA
GAAACCATGGGGTGTTCAAAGGAGAGCCATTGTGGCAAGAAGTCTAGAGATAGTTATGGGCAAGTGATCAGTCT
AAAGAAGAGAAGACTCAGGGCCCATGGCAGTGACCCCCAACCTCCTGGCAGATGGATGCTTCACTGAGGCAGCC
AAGCATCGTTGTTaagaccttgattgtctcagacccaactccacactctgagctgtgtgcgtcaggggggaacta
cttacaaagcctcactttgtttattagcaaaatggagacactaatacaaattccaaaccactaatttgcatatc
taaaatccccaaatacctgagaacagaaaacattttcataagcccatcacaaactcatctttgacctaaactgg
catgagacgggagactggagacttactatgatatttctttatcccacagtctgtgaatatgcacatattttgtt
gcaaaatattaatgtgtttgattacagcatgctatcccagaccctgctggtggggttaaacaagatacaataca
tatactatatattcttaaataataaaaatagttatgaatttcaaaatttggcctaaaagttttggataagggat
agtggccctgtgctaactacctcattaggtcttggggtggatcaagtagaataaaatatggaaagggcttagca
gagttagtagcacatcataaatcaacaataaatgttTCATTTCCTAAATAGCTATTACATTCTATGAATTAGGT
ATATTAGTTGAGCAGTTTTGCTCtatcgtcccattgattttactcttaagaatactgtaatacaactttgttta
gagtcctccactattgtctcctgagatttttttttcttccaagatgccttctgaatgttttgaatctggtgct
ttcatgattttcaaaaaatctccataaatttttttaagacatagaaggcttatattccttattccaaagactc
ttaaatagtttgactgaaatattgagagaatttttatcacactacctgaaacagcctgtctcacacatgcataa
gcaacaacgtcgtgtctgctgcccagccaagcacaatcagaagtgtcttcaattgttatgaatgttcacagata
cagtgtaaacacatatgcatctttagaatagatgaTTTGCAGTACTGTACTACAAAAAGTTGTATTTAGATGGA
ATTCAAAGAGACCACCTACAGCATGCTGTAAAACAAACAACAATAAaataacttcctcacaagtagatttgttc
caaaatggattgggctatgtttggtattattaaattccaattctccagcaatggggactcttgagtctggataa
atgcttattgggcatattgtagctggacttcaaatttcacaaagcattgtgcttgagatgacctttgagaacct
tctaactctcagattctatgaGACTTGAGGAAAGAAGAGACAATCTAAGAGAATGGGGAAACGCAGATGAAACA
TCCTTGACCTCCCAGGTTTTGCTTGATTCTTGGAATATTCTCTTCGGAGGGGTTGGATTTGAAGGTCGAGTGGA
GTTCACTTCAGTCATATATTCTGCCCTGCTTCTGTCCAAAACACAAAGTATAGGAAAAAAGCATTGAGTTGTGA
AACGTAATTATGATAAAGAAAAAACTTTACATTTCACATAAAATTGTTAGATCAACTTGGCAACATATGCAGAA
GCAGTGCGGAAGCAACTCAGTCTTTAAATTAGCTACTGCTAATGCATACTGGCATTCTACGGTGAAAAGGGTTT
AAGAGTGAACACGACACACCCCTTCAcctctttccatgccagagctctgaatcgaccaagagatgcaacggggc
tgatagttggaggatagaatgccagctctattactgacaaacactacaaagcaaaactacagtgggctttccaa
gagtccccagaattagtctggcttttctgcccagtcactgaaagggctgaaaaaATCAAAACAAGACAAAACCA
CGATAAATATCTTTCTCCTTGAGGTAGGACCTTTACCTACTGGTAGAGAACACAGGTGCTCTCTGCAAGCCATT
ATGGCCTATAAAGAGGAGCACTGTAGAATGCTTATCTTCATCCTAAACTTTACAGTGAGATGACCTCGAAGGCC
CTCTATGGAGAGGAGTAAGCAGGTGGACAGCGGTATTTAGGGAAAATCTGTCTTTGTGGAGAAGTTCCCTACCT
AATTAACCTTCCCTTGTTTGAAGCCTTTTTTGGTGGGACCTGGGGGTGGAAGAGCCCATCTCATTTATCGTTTG

TCCTTTGGAAGCATTCCACTCTCCTAAATGGAATTTTATATTGCACACTTTCACATACAAATTGTACAAAACCC
TGACCCTTATGTTTGGAAGAAGAGAAATAATGTATTCTATTAAAAGGAGGTCCCCCCATTTAATTGCACAATGA
TGTGACGGTGAAGCACCCCTAGAGAAGTGTCCTTTAGATACATCTGTGCAGAAGGaactagctgcgtgaactgg
gacaagtcatgtcaccgctcctggcctcatttcccttggggattaagcccagtggggcactggaggctgccttg
gtcccttccagccctcacattctCTCACACTCGGCTCTTCTGTTTCCAAGAACAATAATGACTGCAGAAAATAT
CATTTTGCCTTAGATTGTGATCTTCGTGACTTCATCTGTGTGAACAGACAAACATTAAAAAAACAGCTGAAGCA
TAACACAGGATCTAAGGCCATGTCCAGGGGCCCTTCCTTAAAGCCTCCAGTTGAGATAAACTTTTAAACAAACA
GCATTTGTGTGTTGAGAGAGGTGGGAGAAGGTGGTACGGTATGAAGTGGAGTCTTTATTCAGGGTCTAAGCTCT
CGTTTCCTGGGACAGTGAGAACTTTGTCTTTTCGGAAAGCTTGCCTGAAGTGTTTGCCTGTTTTATAGCTTGAG
GTCAGGGAAGAAGAACAAGATATATCTGAGCCGCTTCCTCCCTGAGCCTTTCATCCCTCCCTGCTACCTAAGTC
TGCAGCCTCTGCACCCTTTGGCTAGATTCCTGCTCAGGAGAACCAGGTCAactcgcacacatgtgcacatacac
acatacaaacacatatgcacacaACAGAACAATCGCACTCCTTTCTACCACTGAAATGAGGCTGATTCCCCTGT
GGAGGAAGGGATGGTTTGCTACATGGCCCTCACACAGTCCTCActccttaatgacatccttgcctcacccggct
tcactgccccctctctggaggctgctacagtactgtgtacatttccagtctactttctttgcctggaagggtgct
aattttgttttcacacatgcatagcacttctagttcatgagctctctgaatgcaaggcctgtatctgatttact
tttgattccctggtccgaggccttagtatataaagagccatcagctcgtgatttttgaataaatTTCAAATTAA
GGAACATAGTAAACACCTTTTGGGATCTCTGCTTATCTCACAACCACCTCCAATTAACTGCTCATTAATAGGGG
TGGACGTCTGGCAGCCGTGTTTCAACTCTCTAAGACTCTGTATGAACCAATCAGAGTCCCTCTAGCAGGTTGAG
AGAGAACGCTTAGCCTTTCCCACAGCAGGAACGGGGATGTGAGGTGTCTGTCCTTCACCAAGTCTCCAGAGAAA
CAATGTAAGCTCTGCAGAAGATATAGAACTGggccgggggcggtggctcacgcctgtaatcccagcactttggg
agacccaggcgggtggatcacgaggtcaggagatcgagactacggtgaaacccgtctctgctaaaaatacaaaa
acagctgggggcagtggcaggtgcctgtagtcccagctactcgggagactgaagcaggagagtggcatgaactc
gggaggcggagcttgcagtgagccgagatcgtgccactgcactccagcctgggcgacagagcgagactccgtct
caaaacaaaaacaaaaacaaaaaaaCAACTGAAGACCTTTTATTAACAAAAATAATGCAAGGTTGTTATAACAA
ATCAGACCACGCATACCCATATGATGGAGGAACCAATGTTcgagaatggctaagaatattttataaaagaaaaa
taggaagcgggtaggggcttatcctaccagattaaaacagtgtggtagtgagtataaatacggtatgttggcag
aacagataggatcccgctacatataggcacttactcgataataaatgtgccacttcagtggggaaatgacagat
tgttattttcttaatggtacagagacaattgattccacatttgggtgaaaagctagcttaaatggagtataaaa
ataaatccatatgaactaaacataaacataaaatcattattaaaaatagtattaaagactatactttgaggtgg
aggcagctttctaagcaagatagaaaacacagaagctataaaagacaaactattagcagcttcagttacattaa
aaaattaataatttttaaaattggaaagagatcccataaagaaataagccaggtaattattagaaagtatttgc
aactcatatgatgcacaagaggatattatccccaggatgcaaggagtttctacaaattaagaactaaaagatag
acaaccgaacagaataatgAATGAGGGCATAGTTTTAAGTCTCCCTCCCTCTCTTTTTATAAAACACTAATTTT
TGTTCAAAGCAATAAGCAAAAGTTTAATTTGTTCTGTGTCCTCTATATGTAAAGTCCATAAAGTTCTGCGGGAA
TCATCTCCTTTTAAATATGAGCTGCTCAATTTCGACACTGAATTCAGCTAACGTATATTAACCAAGAGAAATCA
AGAGTGCTTTGAATTCCCACTCAAATTCATTTGACGGCTATTTGCCTTACCACGCAGGGCTTTTATAAAGCCCC
AGGAACGGGGTGTGCAAACAGTGGTACCTAACTCAATTATTTCTCTTCCACctttcttttctttctttctttct
ttttttttAATTGACTAGTCTTCTTCAGCAGCTGTTTCAaaggaagggagggaggaaagtaggaaggagggGG
TGCAGGCAGCCTGGCCGAATTGGTCCATTCTGTTCCTGGTGCGGTTCCTCCTC**TAGCCACATGATGGCGGCA**TG
AGCATTGCCGTACTTCTGCGCTCTCAAGTCTAGCGCTTTAA**ATCCCC**AGTTTTAAAGCCAGGGGCGC**C**GGGGTG
TGTTTGTGTTTGTGCCTGCAAGTCAAGGCCATGCCCAGACTTAGCACTCTCAGCTGCTCTGTAATGCAAGTCGA
CAGGAAAATCTACAAGGGCTTCACGGTCTTCTTTCTAGAAAAATCCCAACAAGGGTTTGATAGTAAAAAGTACT
TCGTAGTCATTTTCAAATAGGGGAGAATTTCCTGAAGTATGCAGCATCTGCTCTGCCGGGTTTGCAATCTGTGC
ACAGGTATTGCAGTTAATGAGCCACCGGCTTGCCGCTTCCAGGGCCAGGCTGGAGCAGTCAGCTGGTGGAATGA
TGGGAACTGCAATCGGCAGCTCTGCACCAACGTGCCCCGGGTTGGAAAATGTGATTTGTGACAGAGAAGCTTGG
GAATAATAATGAAAACATAATTGCTCCAGGTTGGTTGCTCTCTTTCAATTGCAAAGCCGTCTCACCTCTTGTGA
GAGGAACTGTAGTTCTAAGCGTTCATTTCTCTAGAACAACTAGAGGCACACCCAGCAACTGTCCTTCTGCCCCC
AGTGAAACTTGGATTTTTTTTTTTTTGAACTCCAGAGAAGAGCTGTCCCCCAAATTCAAGAAGAATCGTACTTA
CTTCATCTCCAGGATGCCTGGGGCAGCTTGTCTGTTCAAGCAGAATCTCCCAAGTCATGAATGTGGCTTTGAAA
CCTATTAGTCCCATCCTCTGATGCCTGGGACCAGGCTATTATATTGTCGGGTTGGTTTATTGTTGTTGTTGTTG

TTTGTTTTTGCCTCATGAAAATAAAAAAAATTAAACCCCTGCGTATGAATGAATTTTCATCGATTTTTCATTTG
GTCATCCTACAGAAGAGGCAGAGGATGTTGACACAAACTTACAGACTGGGACTCCAAGGATCAGCCAGATGAAA
TGCTATGTGCAGAGTCTTCATTTCCCagctcagtggtggacagcagaggactctaaggtgggttgccctgattg
gagtccccttcagccactttaagttaacggatgtctctgagctccaagtttctcatctgtaaaatgaagagaa
agatgtcaattgcttgggttttgagaaagactgaaagagatcatagatgtaaaatacttagcgtggggcccgga
acCTCTCCTGTGGCTGGGAGTGAGTGGGCATGTTGCACATTGAGAACCTCTTTACACAATTGCCCATCCTCCTA
GCCCTGCAGTCTGTGCAGACATTTCCATGCCAATCTCCACGGTATTTGTGCAAGTAATTCTTTCTGCCATGGGG
CTTCTCTGGCTGACTTCAGAATGTTGCTCAGCTTTAAAGGATGGACTgagaggttaatgctggcggaggcaccc
taaccaaaggggactggatctgttggataaatgcatccgctcttcatttccccaagggtgatggttccgagac
atgtctcaaaagaggggacactgagccacagttactcccaggaatggcccaacacacagcatatttctacatt
catgtttctttttcctgttactcagcggtgcttcactcttgttccttagaatcacctcataaatgaactacctt
gacatgccctgacccaggctctttatttggaaacccaggctaggacaGGAATGAAGAGAAGCCAGACCTTCTGA
CTCGAAGGTTAGGAGTCTTTTGATTTacacacacatgcgcacgcacacagacacacacacTTTGTTTTGTGTCC
ATGGCAGGCAGGTGGAGGAAGGGGGAGTTTGGCCAGCTCTAAGGTCTTTCTAGATCATCCCATACATTGCTGGG
GACCTCCAAGGGACAGACAGTGTCTTCATCAGGTCCTTTAAGGACACAGGTTCCTGTGCATTGATGGCATTAGT
GCTGAAGCCCAATTGCCCCAGCAGCAGCAACCAGCTTTGATCTGGGGCTGGCATCA

First CTCF, highest to lowest efficiency (Najm 2018), - strand:

gRNA1:ATGAATGACTTGCGTGATCTGCAGAGT Najm 2018 efficiency score: 77
(distance from CTCF: +121 bp)

gRNA2:CCTAATGCACAGGCAGGCCCTGGGGGT Najm 2018 efficiency score: 71
(distance from CTCF: -153 bp)

First CTCF, highest to lowest efficiency (Najm 2018), + strand:

gRNA3:TTCCCACTTCCACCTTGGTGAGGGAAT Najm 2018 efficiency score: 82
(distance from CTCF: -87 bp)

gRNA4:ATCATTCACTTAGTAGGCACTATGAAT Najm 2018 efficiency score: 62
(distance from CTCF: +157 bp)

gRNA-Cas – everted 41 bp - gRNA-Cas – 87 bp - **CTCF site** – 121 bp - gRNA-Cas – everted 11 bp - gRNA-Cas

Total: 153 bp to red site; 157 bp to pink site (from ends of CTCF binding site)

Second CTCF, highest to lowest efficiency (Najm 2018), - strand:

gRNA1:GGCGCCCCTGGCTTTAAAACTGGGGAT Najm 2018 efficiency score: 65
(distance from CTCF: +43 bp)

gRNA2:TTTTTACTATCAAACCCTTGTTGGGAT Najm 2018 efficiency score: 55
(distance from CTCF: +192 bp)

Second CTCF, highest to lowest efficiency (Najm 2018), + strand:

gRNA3:TCTTTCTAGAAAAATCCCAACAAGGGT Najm 2018 efficiency score: 81 (distance from CTCF: +179 bp)

gRNA4:CAGTTTTAAAGCCAGGGGCGCCGGGGT Najm 2018 efficiency score: 79 (distance from CTCF: +49 bp)

gRNA-Cas – 33 bp - **CTCF site** – 43 bp - gRNA-Cas – everted overlapping - gRNA-Cas – tandem 104 bp - gRNA-Cas – everted overlapping

Total: 33 bp to blue site; 192 bp to blue text site (from ends of CTCF binding site)

Region 7: Right TAD boundary region for *MYC* insulated neighborhood
hg19 chr8: 130692569-130700970

Top CTCF peaks (polarities: - strand)
hg19 chr8: 130694449-130694467 AND chr8: 130699821-130699839

d*Sp*Cas9 sgRNAs:

AGGGGGCTGAAAGGAGTCATTGACTAATCTGGTAACATCACAACTGACCCCAGCATACAAAATCTAAGTATCCC
CTCTATGAACATCATTTTAATTTCCAGGAGTACTTGATGCATTAATCATAGCTCTATTATAAGCAGAAAATTCA
GGCTCATGctctctcttgctctctgtctctctctctctttctGTTAAAGTCCTTACTTTAAGACTCTGCATTAT
ACTGTAAGCTATATGCTCTTTGAGCATCAACTGTCACTGTACTTATTACACTGTAGTGTCATAATCTGTCTATG
TGTCCTGAACTCAACAGAACTCTTGAAAGACAAAAACCATGTGTCATCCATGGAAACAATATAGTATAGCTATT
AAGAATACCACGTCCCAactgttaatgggtacaggtttattttgggggggatgaggcatgttatgagacatgttc
taaaattgactgtgatgtttgctcatctctgtcaaatatactaaaaaccattgaattatacactataaataggc
aaattatggagtacatgacttatcactcagtaaagctgttttgaaaaaGAAAAAGGATGCTactgcctcccttt
aaattctggcttcactacttactgtagcaagttacataatttctctgtgcctcagtttcttcatctgtaaactg
ggttaaagcaatacctacctcaaagggacactaactaagaggaattaaattggcaattatttttaaagccatta
acaacagtttctggaacataTAGAGCAGTATgcaaggatagttatttatttttcacctctattctcagaacatg
gtttagtgcctgacacatactgagtaaatgataaagatttttaaaatgtataaCCCAGCTTATAAGTACCTACA
GGAACTCAAATATGGGGCACGTTAGTTGCTATGCCTTACTCAATAGAAAGAGGACTGAACTAAGTTGCCAAATG
ACTTGGGTTTGAggctgggtacagtggctcacacctgtaatcccagcattttaggaggccaaggcaggcagatc
acttgaggtcaggagttcgagatcagcctggccaacatagtaaaactccgtctctactaaaaatacagaaatta
gccaggcatggtgcctgtaatcccagctactcaggaggctgaggcaggagaattacttgaacccgggagacgga
ggttacaagatcataccactgcactccaacctgggtggcagagcaaggctccatcAatacatacatacatacat
acatacatacatacatgacttgggtttgagttctgatcaccacattagagcaggctgacttgaaaaaaccactt
tcccatttaccctgagactctgcatcctcatctgccgaaactgacaattttgcttctaccgaactgaaagggta
gtttcaaagatgaaatgaatcaatgtgtatgaaagctcttttaaaaccatcatgtgcaaatataaatTCCTCCG
AGAAATTGGTCTTGTCCTAAAGATGGCCCGGGTTGCCCTCTTACCCTAAAAGTCTCCTTATTCACTCCCAGCCC
GCCTCACTCGTTCATCATTGCAGGAAAATGTGCCAAAGATCAAAAACAGGTGAGATCCGACCCCAGGGACAGAG
CAAAGAAGTAACTGGACTGTAGGCACCAGCTGGTCCTGTTCTGCCCTGCCCTCGTAGAACTGTGAACAGCTAAG
AAGTGCTGTT**CCT**TTAAACTTTGCCCCTTTGTGGCTCACAGATGAGTAGATTC**AGG**CAGCAGCTGTCAGCTCTG
CAGGCAGCAGAGAATCTTTCCAGAGCAGCATTCCCAGCAGTGTCAGGATGGGAGGATGAAATTAGGTGAGCCAA
GCAACCTGAAAGTCCTCATTCAAAAGCCTC**TACTGCCATCTACTGCCTC**TCTGAAGAACTCATTGGCAAAAAGC

CAAGAAAAGACTGCAATGGAGGTGGGGTAGAGAGAAGCCTTAGCAATAGACTGAATTGTCTGgatcaattagtc
caccttcattctccagctggggaaaatgagacccaaagaggtaaagaggcttgtctgaagtcactctgctagc
ccacggcagCCCATCATCTCTTGGATACACATATTAGCTCTGCCCCACAACACACTGTTTCTGTCTGAATATAT
AAACTATCAATAAAGCTGGCTTAGAGCAAAGAGGTTTGTGTGGTTGTTATattgagacagggtctcactatgt
tgacgaggctggtcttgaactcctggcctcaagtgatcctccagcatcagcctcccaaagtgttggggtgacag
gcgtaagccaccacacccagccAGAGGAAAGAGCTTCTGCCAGCTATCCTTTTCCTCGCCATTTTTTGGTTCCA
CCTTCTCTGCTTTGCACAAGAAGCCCCAGACTGGACCTGTTGTCTTATCCATCTAGAACCCAGTGTCTCATACT
GTCCAGCAGTCCCTTTCAGGGAGAAAACATTGAGCCGTTCACTTGAAAGCATTTGCTTTCTTTTTTTTTTTTT
TAAAGAACATGAAGAATAGCACAAATCCTGTTTAGTAGACACTCCACACCCCCATACTACAGAAAGCAAACTAA
GGCTCAGGGACATCAAAGAATTGGCTTCTAATGtgcctgacattaggatcacctgggtatcttttacaacctcc
aacgctcaggccacacccagacctattgaatcagaatctctgggttagaaaacaggcatcagtagtttttgaag
ctctcccacacccaggtgattccaacatttaggtaggcttgggaaccactgACTTACGTTCTCCTAATTAGTAG
GTAAAATGGCAGGGACCCAAGTCTTTCAACTTCCAGGCCACTGAAGAAAagtgatgtccatattttagctgttt
gcatcacccttgaagatttttgccacttccacacactaactatattaatatgtgaagtgaccccatttttttta
cttaaagcaatgtatttttaaaagaacactgttactctcgggaatggaaaaacaatctcttgtcacaaaACAAA
AATCAGCCTTTAAAGACACACACTAAAACCCAAAAGTTAAAATGAAAATGTACATCCATGTATGAAATAAAGTT
ATCTCACTTTCTAATTTGGGGGAGCCATTACAAAGAAAGATACCCTCTCTGCTCCTGATTAGACTGAAAGAAAT
GTGGCAAAAACACATTTTGGGACAGAGAGTATCTGTGCTCACTCCGCGCTTCTCCATGTGTACGCACCATTTTC
TCTCTCTGTCATTCTGTCTCTCTCTAGCCTCGCTACTAAGCTCGGGAGGAACTCCCATCAGCTTTCCCTGTGTT
GTAAGACCTAAGCCCAGGACTGCGGGTTCTGGCCAAATTGTAATTGTGACAGCATCCTGCCCAGGACTTACCCG
ACAGCGGAGAACTGTCACTCATTACCAAGAGCAACAGACGTTCTTTCGGCATTGTCCTGGCCCTCTCTTGATAG
GCTGGCCCGCGTGTTGGGGTGGAACCGTGGCACCACCAGCTCTTTGTGACCCCACCTGTGACTGCAGCCCAACT
CCATGACCCCATTTCCGCTTCCTACAGGTGAGTGTCATCACGCTGTTTGCTCAGAACACAGCTGTTTGGCAAGG
GCACGCAGTAGATGTGGAGTCCGGGCCCCAACTGCTCTGGGCAGTGTCCTTTTCAATGCCACGGACCCACTGGT
GGCTGACTGGGGCTGTCTTCATTGACAAAAAAAAAAAGTTAGTTAACAATTACTCCTGAAAACATCATTTGTGC
CTGTTTTTAAATTATATGGTAATACATCTTTTGAACAACCATTTGAAAGAAAGCTGGTTATTCCATTTTTTCCC
ACTTAGAATCATCACCTTTACTCCTTTAAGAGAAAGAACTTTTTAAAActgattaccacattaatatatggtca
atttagatatttaagtaaaagaagaaaagcaataagaaaaaaattgaaagcacttgtagtcccaccacctatag
aagaccactgttaacCATATGCAATCATTCTTGTCTCTTTATTGTGTACTTTGTGCATTTTTATACAAATTAAG
TTCATACAAAACATACTACTTGTATCATTTGTCTAACATAGCAATATATTACTATCATCTTTCTATCATTCAAT
GTTCCTAAATGAAATCATTTATAAGGAATGTAGAATATTGTTTTCAAATGACTAAGATGAAGAGAGTGAACGAA
TTAATGTATTGAGGaccctaaaagttagatattataatgcccattttaaagatgtgaaaatgaaggtaagtgag
cagccagttaactgtagaacagagatgcccaccaggtctgcgtggtttcaggcctaaactccttcTCCCTTACC
CGTGTTTCCTCCGCATGCCTTGTCATAAGACTCACCTTGGAAATCTGCATTATCAACATGTAATACATTTTATT
ATAAACTGAATTCCAATTGAGAACCTACAAAAAGCTACCTGAGGTCAGATCATGATTCAGAAGAAACAAGTTAA
CAAGATTGTGCCTCTGATGGAGCCCAGACATGGTCAGATATGGTTTCTCACAACCTTACATGTCAagaggaaag
aattccagactagatgtcaggagatccgggttgtattctcacctcctccagtcaccagccaaaaagtcatggcc
acatcattcaatttatctgggcctcagttttctaatctatatattgggataacacaaccttccctgtcaagctg
aatagattctaaatagcaaaggaaataatgAAGCAATATACAAACATAAGAGACCTTTATGAATCCTACTACAG
ATCCATCAGCGGTGAATTTTCCAGCCCTCTAAAACACATCAAAGAATTGTAAACAGAGAAGGAAACTCTGagcc
caactaattttagagacagagagactgacttcctgagaaggtaggcagcttgcccaaagattcagagtgagtta
ggaacccagctaacactagaagccagtgtctGGATTCACACAGGAATATTCCTTTTATAAGACCACCTTTACGG
TGATAAATGTCCCAGACGCGTTCACATAAACACGTTGACAGCAATTCAGTTTAGTTCCGTTTATTTCAGTTAAG
TAAAGATAAGCTGAGCATTTGTTTGCCATGTACAATGACCTCTTCAGTCAAGGAATTCCAGCTCTTGTTTTTGT
CTTAACTTCTTTCAACTTTGAGGGTTTCTTTGCATGTTCCTTACCTAACAAGACCAGCCTGAAAAGAGTAGAAA
AAAACATAAAGAAACTTGGAGTCTGAATTTTCCACATGCTGGGAGCTCTCACACCCAATTCCAGATACAGCACT
CTCCTGAATGTGACCTGAAGTGGTTTCCTCAATTGATCACATCCCAGGAATCTGGAGAGCTTCCAGGGCAGGGA
AAATTAGTTACAGAGTCACAGAGTGGTGCGGCCTGTCACATGAAAGTCACTTCCAGCTTGGAACGGATTGCACA
TTCCTGCCTTCTGAGTGGTGTTTCAGATGACTGGGATCTGGCCTGTCTGCAGACAGGCCCCAAAGCAGCAGAAG
GAAGCCAGCGGGCAGGGGCATTCCAACACCGTTCCCGAGACCCGCAGTCACAGGAAAGGAGAACTTCCTGAGGC

TCCTCCAAGCCTTGCCTCAACTTAATGGTGTTTGAGAAAAACATCAAGTTTCCATTTACCTCCATTTTGAAACA
CTGAAAGCATTTGTCATTAAATAGCATGATGACTCTCCTGCCTCAACTCAGAGCTTTGTCAAACTGCTCACATG
GCATAAACAAGGATGATGGGGTTTCCTACACATATATCAGAACTTCtaaatgcttcttatatgccaagtaccat
tatataccatgccttggctcatctaatccttacagacaacctcaggagagaggtgctctatatcactaccacag
ccaaggaacctggggctcaggatggttaagtaacttgtacaaggtcaaaaattggcaaatgacaatgcacgaac
ttaacacagatctgtctgactccaaatcccAGGTCTACACTGGCTGGATAAATATGAGGCAGAggccacatgta
gtagctcacgcctgtaatcccagcaccttggaaggccgaggcgagtggctaacctgaggtcaggagttcaagac
tagcttggccaacatggtgaaaccccgtctctactaaaaatacaaaaactagctgaacgtggtggtgcatgcct
gtaatcccagctacttgggagactgagacaggagaattgcttgaacccgggaggcggaggttgcagtgagccaa
gatcgcaccattgcactccagcctgggcgacaaaagcgaaactccatctcagtaaaaaaaTAGGTATATATGAG
GCAGAATGACTTTCAAATAATACACAATGGTTTCAGAATCTGGCCTCTGTTTTGTTGATTGCAAGGCCCTGTCA
TATTTCCAAGTCACCTATTGCTGGGTTTCACAGCTTTAATGGTTTGAAAAGTCTTCCTTATGTCAAAATGTTTC
AACAGGCTTCGTTCCCGCCCTCTGCACCCCACAGTCCTCTTCCTTCCCAAACTCTCCTCTGAGAAGTCTGTCTC
ATCCCCCATTCCCATGAAATACATCTGCTGACATCACAAAGGGATTCTTGTCTAACTTCCATCTCTCTAATTGA
AGTTTAAGCAAAATGTTTCTTGACCTAATCCAAGTAAAAGAGCATGATTGTCCTCTATTTCATGTTATTCTCAC
AGTTACAATATGTGCTTACATCTGCTGGTGCAAAATATGATTAACTTTTTTTTTCTAAAACCCTAGGGCTAGGA
GGAGCTTGGAAATTCTATCTTTTACTTCCCCTTGCCTTCAGACCGATCCCTGTATGTCAACAAAACATGAAAAC
ACCCTGTGAAATCCATTAGTAGCCTCACTTTATATACTAGGAAATCGAGGCAGGGTGATGGATCTGAAGTCCCT
CAAGCAGATTGTGGCTAGAGCAAGAAGGGAAGCATGGACCTGGGTCTCTCAGAGGAATCCCTGTCTTCCCATAC
AGCCTTTTTCCATAGCATCTATGCACTAACTATAAGCTATCGACCCTATAAAGGTCCACGGACATATGAGAACA
TTATCTTCTTCCAGGCATTCAGTTTCATTTATCCGACATTTCCTGCACTATCCTAAAGATATAGAGGCAGGTAG
ATTAATTCACTATGCCCTAGAAGTCCTGGTCAAGGTCTCAGAATCAAGAACTACCTCTCTAACAGACAGGCATT
GAGCACCCTCTGTAGGCAAGGCAGAGTCTTGCAAGCTGGCACCAATGATTGTCCTGCATCTGCCTGGTCCACAT
ACTCCCTGGACGTCTGTGTGACAGGCTCTGTAGAAGACAAGCATAaagcagacctgagggtgagtcccagccca
gctctgccactcaccagcggtgtggccatgggcaggttacctccatggactagacctgggctttctcatcttta
agttggtgttaataacaAGTTGGGGTTAAATGCTGTCCCTCACTCAATACCTGtgtccacatcttaatccttag
gacctctgaacgtgaccatatttgataaaaaggtcttggcagatgtaattatgtaaaggttcttgagatgagat
catcctatattatctaggttgtctctaaatccagtgacaaatgtccttatgaccgcagaaagacagacaccagg
agaagaagaaaagtccacgcaaagaaagatgaaggtagaggtaggagttatccagccacaagccacagaatgcc
tggaacctccctaagctgaaagaggctaagacaaagtcccctctagagccatcagagtgagtgtggccctgcca
gcaccttgagtttgtactctggcctccaggactctaagcaaaaaaatttctgggggatttttaagccatgaag
ctttgtggtcatttgttaaaacagccctaggaaactagtaTCCACCTATTTCATGGTATGTTAAGCATGGAACC
TATTCCAGAGTAAGTACTTCATCATATTATTGGTGTTATTCTTGCCCTTGCCATGCTCCAAGTTAACTTAAGAC
CCCTTTTATTCTGACGATACCTCCTGGAGACCACCCACCCATCCAAAGCTCCAGAATCAAGGTTTTAGTATTTT
GATTTGGGGTTGCCTATTTGTTAGGAGGAGGGGCTGCACATACCTATTTCCTATTTTTCAAAATAATAGCCCCA
AATTTTGATTCAAGTGAAGCTGGATAGACCCTGAAGGGTGTATTTGTTAAAAATTTCACAGCTCATTCTGACAC
AAAACTGGTACTGATATTCAACCAATGCACACTGACCAGCATCCTCTCTGATTTGTCTGGGACTTGGTCCTTTG
GTGAAATATTTGCCCTCTCTCCGCACACCAGCACTGCTCT

First CTCF, Lowest to highest CFD (Moreno-Mateos 2015), - strand:

gRNA2:GACAATTCAGTCTATTGCTAAGG CFD: 2.93, Doench 2016 efficiency score: 49 (distance from CTCF: +62 bp)

gRNA4:CACAAAGGGGCAAAGTTTAAAGG CFD: not predicted, Doench 2016 efficiency score: 27.17 (distance from CTCF: -145 bp)

First CTCF, Lowest to highest CFD (Moreno-Mateos 2015), + strand:

gRNA1:GAAGTCACTCTGCTAGCCCA**CGG** CFD: 2.56, Doench 2016 efficiency score: 71 (distance from CTCF: +156 bp)

gRNA3:GCTCACAGATGAGTAGATTC**AGG** CFD: 4.83, Doench 2016 efficiency score: 49 (distance from CTCF: -122 bp)

gRNA-Cas – everted 0 bp – gRNA-Cas – 122 bp – **CTCF site** – 62 bp – gRNA-Cas – everted 71 bp – gRNA-Cas

Total: 145 bp to red site; 156 bp to pink site (from ends of CTCF binding site)

Second CTCF, Lowest to highest CFD (Moreno-Mateos 2015), - strand:

gRNA1:TTGATTCTGAGACCTTGACC**AGG** CFD: not predicted, Doench 2016 efficiency score: 63 (distance from CTCF: -27 bp)

gRNA2:ACTTAAAGATGAGAAAGCCC**AGG** CFD: not predicted, Doench 2016 efficiency score: 60 (distance from CTCF: +183 bp)

gRNA3:GTCCAGGGAGTATGTGGACC**AGG** CFD: not predicted, Doench 2016 efficiency score: 54 (distance from CTCF: +43 bp)

gRNA4:TATGTCCGTGGACCTTTATA**GGG** CFD: not predicted, Doench 2016 efficiency score: 47 (distance from CTCF: -156 bp)

Second CTCF, Lowest to highest CFD (Moreno-Mateos 2015), + strand:

gRNA5:TGCCCTAGAAGTCCTGGTCA**AGG** CFD: not predicted, Doench 2016 efficiency score: 60 (distance from CTCF: -39 bp)

gRNA6:TACCTCCATGGACTAGACCT**GGG** CFD: not predicted, Doench 2016 efficiency score: 69 (distance from CTCF: +166 bp)

gRNA7:TGCCTGGTCCACATACTCCC**TGG** CFD: not predicted, Doench 2016 efficiency score: 54 (distance from CTCF: +41 bp)

gRNA8:ATAAGCTATCGACCCTATAA**AGG** CFD: not predicted, Doench 2016 efficiency score: 44 (distance from CTCF: -168 bp)

gRNA-Cas – everted overlapping – gRNA-Cas – everted 94 bp – gRNA-Cas – everted overlapping – gRNA-Cas – 27 bp – **CTCF site** – 41 bp – gRNA-Cas – everted overlapping – gRNA-Cas – everted 100 bp – gRNA-Cas – everted overlapping – gRNA-Cas

Total: 168 bp to yellow text site; 183 bp to blue text site (from ends of CTCF binding site)

Experiment 2: 4C-seq viewpoint, primer design, and library primers

4C viewpoint fragment upstream of *MYC* promoter (283 bp):

GATCTCTGCTGCCAGTAGAGGGCACACTTACTTTACTTTCGCAAACCTGAACGCGGGTGCTGCCCAGAGAGGGGGCGGAGGGAAAGACG
CTTTGCAGCAAAATCCAGCATAGCGATTGGTTGCTCCCCGCGTTTGCGGCAAAGGCCTGGAGGCAGGAGTAATTTGCAATCCTTAAAGC
TGAATTGTGCAGTGCATCGGATTTGGAAGCTACTATATTCACTTAACACTTGAACGCTGAGCTGCAAACTCAACGGGTAATAACCCATC
TTGAACAGCGTACATG

Creating "fake" fragment for primer design using Primer3:

AGGCCTGGAGGCAGGAGTAATTTGCAATCCTTAAAGCTGAATTGTGCAGTGCATCGGATTTGGAAGCTACTATATTCACTTAACACTTG
AACGCTGAGCTGCAAACTCAACGGGTAATAACCCATCTTGAACAGCGTACATGGATCTCTGCTGCCAGTAGAGGGCACACTTACTTTAC
TTTCGCAAACCTGAACGCGGGTGCTGCCCAGAGAGGGGGCGGAGGGAAAGACGCTTTGCAGCAAAATCCAGCATAGCGATTGGTTGCTC
CCCGCGTTTGCGGCAA

FP (non-reading primer—NlaIII): 5'-GCAAACTCAACGGGTAATAA-3' T_m 55.4
RP (reading primer—DpnII): 5'-GTAAGTGTGCCCTCTACTGG-3' T_m 54.9

Targets (Primer3): 139,8
Included region (Primer3): 89,93

Other primer sets for efficiency test on 4C template:

Primer3 outputted the same RP for all 8 primer pairs

Other FPs:

FP2: AATAACCCATCTTGAACAGC
FP3: GGGTAATAACCCATCTTGAA
FP4: GGGTAATAACCCATCTTGAAC
FP5: ACGGGTAATAACCCATCTTG
FP6: CGGGTAATAACCCATCTTG
FP7: TAATAACCCATCTTGAACAGC

2-step library prep primer sets:

1st round:

VP_RP_readp (1st_r):
TACACGACGCTCTTCCGATCTGTAAGTGTGCCCTCTACTGG

VP_FP_nonreadp (1st_r):

ACTGGAGTTCAGACGTGTGCTGCAAACTCAACGGGTAATAA

VP_FP_nonreadp + 10 bp spacer (1st_r):

ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAAACTCAACGGGTAATAA

2nd round:

Universal FP (2nd_r):
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

RP_1 (2nd_r):
CAAGCAGAAGACGGCATACGAGATCTTTTGGTGACTGGAGTTCAGACGTGTGCT

RP_2 (2nd_r):
CAAGCAGAAGACGGCATACGAGATTAGTTGGTGACTGGAGTTCAGACGTGTGCT

RP_3 (2nd_r):
CAAGCAGAAGACGGCATACGAGATCCGGTGGTGACTGGAGTTCAGACGTGTGCT

RP_4 (2nd_r):
CAAGCAGAAGACGGCATACGAGATATCGTGGTGACTGGAGTTCAGACGTGTGCT

RP_5 (2nd_r):
CAAGCAGAAGACGGCATACGAGATTGAGTGGTGACTGGAGTTCAGACGTGTGCT

RP_6 (2nd_r):
CAAGCAGAAGACGGCATACGAGATCGCCTGGTGACTGGAGTTCAGACGTGTGCT

RP_7 (2nd_r):
CAAGCAGAAGACGGCATACGAGATGCCATGGTGACTGGAGTTCAGACGTGTGCT

RP_8 (2nd_r):
CAAGCAGAAGACGGCATACGAGATAAAATGGTGACTGGAGTTCAGACGTGTGCT

RP_9 (2nd_r):
CAAGCAGAAGACGGCATACGAGATTGTTGGGTGACTGGAGTTCAGACGTGTGCT

RP_10 (2nd_r):
CAAGCAGAAGACGGCATACGAGATATTCCGGTGACTGGAGTTCAGACGTGTGCT

RP_11 (2nd_r):
CAAGCAGAAGACGGCATACGAGATAGCTAGGTGACTGGAGTTCAGACGTGTGCT

RP_12 (2nd_r):
CAAGCAGAAGACGGCATACGAGATGTATAGGTGACTGGAGTTCAGACGTGTGCT

RP_13 (2nd_r):
CAAGCAGAAGACGGCATACGAGATTCTGAGGTGACTGGAGTTCAGACGTGTGCT

RP_14 (2nd_r):
CAAGCAGAAGACGGCATACGAGATGTCGTCGTGACTGGAGTTCAGACGTGTGCT

RP_15 (2nd_r):
CAAGCAGAAGACGGCATACGAGATCGATTAGTGACTGGAGTTCAGACGTGTGCT

Note: Indexes for RP 1-15 in this set = indexes for RP 28-42 in Illumina RNA PCR set

References

1. Xue, Y., & Acar, M. Live-cell imaging of chromatin condensation dynamics by CRISPR. *iScience* ***4***, 216–235 (2018).
2. Gu, B. *et al*. Opposing effects of cohesin and transcription on CTCF organization revealed by super-resolution imaging. *Mol. Cell* ***80***(4), 699–711.e7 (2020).
3. Brandão, H. B., Gabriele, M., & Hansen, A. S. Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Curr. Opin. Cell Biol.* ***70***, 18–26 (2021).
4. Liu, Y. *et al*. Visualizing looping of two endogenous genomic loci using synthetic zinc-finger proteins with anti-FLAG and anti-HA frankenbodies in living cells. *Genes Cells* ***26***(11), 905–926 (2021).
5. Geng, Y., & Pertsinidis, A. Simple and versatile imaging of genomic loci in live mammalian cells and early pre-implantation embryos using CAS-LiveFISH. *Sci. Rep.* ***11***(1), 12220 (2021).
6. Clow, P. A. *et al*. CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus. *Nat. Commun.* ***13***(1), 1871 (2022).
7. Gabriele, M. *et al*. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* ***376***(6592), 496–501 (2022).
8. Viushkov, V. S., Lomov, N. A., Rubtsov, M. A., & Vassetzky, Y. S. Visualizing the genome: experimental approaches for live-cell chromatin imaging. *Cells* ***11***(24), 4086 (2022).
9. Chaudhary, N., Im, J. K., Nho, S. H., & Kim, H. Visualizing live chromatin dynamics through CRISPR-based imaging techniques. *Mol. Cells* ***44***(9), 627–636 (2021).
10. Fullwood, M. J., Han, Y., Wei, C. L., Ruan, X., & Ruan, Y. Chromatin interaction analysis using paired-end tag sequencing. *Curr. Protoc. Mol. Biol.* ***Ch. 21***, Unit–21.15.25 (2010).
11. Lieberman-Aiden, E. *et al*. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* ***326***(5950), 289–293 (2009).
12. Mumbach, M. R. *et al*. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* ***13***(11), 919–922 (2016).
13. Lajoie, B. R., Dekker, J., & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* ***72***, 65–75 (2015).
14. Dekker, J., Marti-Renom, M. A., & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* ***14***(6), 390–403 (2013).
15. Heath, N. G., O'Geen, H., Halmai, N. B., Corn, J. E., & Segal, D. J. Imaging unique DNA sequences in individual cells using a CRISPR-Cas9-Based, split luciferase biosensor. *Front. Genome Ed.* ***4***, 867390 (2022).
16. Phillips, J. E., & Corces, V. G. CTCF: master weaver of the genome. *Cell* ***137***(7), 1194–1211 (2009).
17. Alipour, E., & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* ***40***(22), 11202–11212 (2012).
18. Sanborn, A. L. *et al*. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *PNAS* ***112***(47), E6456-E6465 (2015).
19. Ghirlando, R., & Felsenfeld, G. CTCF: making the right connections. *Genes Dev.* ***30***(8), 881–891 (2016).
20. de Wit, E. *et al*. CTCF binding polarity determines chromatin looping. *Mol. Cell* ***60***, 676–684 (2015).
21. Guo, Y. *et al*. CRISPR inversion of CTCF sites alters genome topology and promoter/enhancer function. *Cell* ***162***(4), 900–910 (2015).
22. Arganda-Carreras, I. *et al*. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* ***33***(15), 2424–2426 (2017).
23. Brettmann, E. A., Oh, I. Y., & de Guzman Strong, C. High-throughput Identification of Gene Regulatory Sequences Using Next-generation Sequencing of Circular Chromosome Conformation Capture (4C-seq). *J. Vis. Exp.* (140), 58030 (2018).
24. Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E., & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* ***170***, 17–32 (2020).

25. Kloetgen, A. *et al*. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet.* ***52***(4), 388–400 (2020).
26. Schuijers, J. *et al.* Transcriptional dysregulation of MYC reveals common enhancer-docking mechanism. *Cell Rep.* ***23***(2), 349–360 (2018).
27. Hashimoto, H. *et al*. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell* ***66***(5), 711–720.e3 (2017).
28. Li, Y. *et al*. The structural basis for cohesin-CTCF-anchored loops. *Nature* ***578***(7795), 472–476 (2020).
29. Pezzi, N. *et al*. STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. *FASEB J.* ***14***, 581–592 (2000).
30. Orgil, O. *et al*. A conserved domain in the Scc3 subunit of cohesin mediates the interaction with both Mcd1 and the cohesin loader complex. *PLoS Genet.* ***11***, e1005036 (2015).
31. Hara, K. *et al*. Structure of cohesin subcomplex pinpoints direct shugoshin–Wapl antagonism in centromeric cohesion. *Nat. Struct. Mol. Biol.* ***21***, 864–870 (2014).
32. Roig, M. B. *et al*. Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Lett.* ***588***, 3692–3702 (2014).
33. Davidson, I. F. *et al*. CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* ***616***(7958), 822–827 (2023).
34. Petoukhov, M. V., Eady, N. A., Brown, K. A., & Svergun, D. I. Addition of missing loops and domains to protein models by x-ray solution scattering. *Biophys. J.* ***83***(6), 3113–3125 (2002).
35. Yin, M. *et al*. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. Cell Res. ***27***(11), 1365-1377 (2017).
36. Yatskevich S, Rhodes J, Nasmyth K. Organization of chromosomal DNA by SMC complexes. *Annu. Rev. Genet*. ***53***(1), 445–482 (2019).
37. Hansen, A. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* ***11***(1), 132-148 (2020).
38. Nagy, G. *et al*. Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. *BMC Genom.* ***17***(1), 637 (2016).
39. Nichols, M. H. & Corces, V. G. A CTCF code for 3D genome architecture. *Cell* ***162***(4), 703–705 (2015).
40. Tang, Z. *et al*. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* ***163***(7), 1611–1627 (2015).
41. Yusufzai, T.M., Tagami, H., Nakatani, Y., & Felsenfeld, G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* ***13***, 291–298 (2004).
42. Brinda, K. V., Kannan, N., & Vishveshwara, N. Analysis of homodimeric protein interfaces by graph-spectral methods. *Protein Eng.* ***15***(4), 265-277 (2002).
43. Karthikraja, V., Suresh, A., Lulu, S., Kangueane, U., & Kangueane, P. Types of interfaces for homodimer folding and binding. *Bioinformation* ***4***(3), 101-111 (2009).
44. Fursenko D.V., Georgiev P.G., & Bonchuk A.N. Study of the N-Terminal domain homodimerization in human proteins with zinc finger clusters. *Dokl Biochem. Biophys.* ***499***(1), 257-259 (2021).
45. Bonchuk, A. *et al*. N-terminal domain of the architectural protein CTCF has similar structural organization and ability to self-association in bilaterian organisms. *Sci Rep*. ***10***(1), 2677 (2020).
46. Kornyshev, A. A., & Leikin, S. Electrostatic interaction between helical macromolecules in dense aggregates: an impetus for DNA poly- and meso-morphism. *Proc. Natl. Acad. Sci. U S A* ***95***(23), 13579–13584 (1998).
47. Gruber, S., Haering, C.H. & Nasmyth, K. Chromosomal cohesin forms a ring. *Cell* ***112***(6): 765–77 (2003).
48. Peters, J. M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev*. ***22***(22): 3089–114 (2008).

49. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J., & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* ***366***(6471), 1345–1349 (2019).
50. Davidson, I.F. *et al*. DNA loop extrusion by human cohesin. *Science* ***366***(6471), 1338–1345 (2019).
51. Losada, A., Hirano, M., & Hirano, T. Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev*. ***12***(13), 1986–1997 (1998).
52. Heidari, N., *et al*. Genome-wide map of regulatory interactions in the human genome. *Genome Res.* ***24***, 1905–1917 (2014).
53. Hnisz, D. *et al*. Super-enhancers in the control of cell identity and disease. *Cell* ***155***(4), 934–947 (2013).
54. Wang, D. *et al*. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* ***474***(7351), 390–394 (2011).
55. Lin, C. Y. *et al*. Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* ***151***(1), 56–67 (2012).
56. Frietze, S. et al. Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.* ***13***(9), R52 (2012).
57. Pope, B. D. *et al*. Topologically associating domains are stable units of replication-timing regulation. *Nature* ***515***(7527), 402–405 (2014).
58. Becket, E. *et al*. Identification of DNA methylation-independent epigenetic events underlying clear cell renal cell carcinoma. *Cancer Res*. ***76***(7), 1954–1964 (2016).
59. Rao, S. S. P. *et al*. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* ***159***(7), 1665–1680 (2014).
60. Cremer, M. *et al*. Cohesin depleted cells rebuild functional nuclear compartments after endomitosis. *Nat. Commun.* ***11***(1), 6146 (2020).
61. Fudenberg, G. *et al*. Formation of chromosomal domains by loop extrusion. *Cell Rep.* ***15***(9), 2038–2049 (2016).
62. Nanni, L., Ceri, S., & Logie, C. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol.* ***21***(1), 197 (2020).
63. Beagan, J. A., & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat. Genet.* ***52***(1), 8–16 (2020).
64. Szabo, Q., Bantignies, F., & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* ***5***(4), eaaw1668 (2019).
65. Gong, Y. *et al*. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* ***9***(1), 542 (2018).
66. Hnisz, D. *et al*. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* ***351***(6280), 1454–1458 (2016).
67. Wang, J. *et al*. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res.* ***41***, D171–D176 (2013).
68. Tak, Y. G. *et al*. Effects on the transcriptome upon deletion of a distal element cannot be predicted by the size of the H3K27Ac peak in human cells. *Nucleic Acids Res.* ***44***(9), 4123–4133 (2016).
69. Yesbolatova, A. *et al*. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* ***11***(1), 5701 (2020).
70. Natsume, T., Kiyomitsu, T., Saga, Y., & Kanemaki, M. T. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep.* ***15***(1), 210–218 (2016).
71. Rao, S. S. P. *et al*. Cohesin loss eliminates all loop domains. *Cell* ***171***(2), 305–320.e24 (2017).

Chapter 6: The Unmet Need for Deploying Biosensing Platforms for Detection of Specific DNA Sequences in Early Cancer Screening

An exciting potential future use case for bipartite "turn-on" live cell biosensors for DNA, chromatin loops, DNA methylation, and histone acetylation in understanding tumorigenesis during the very early preleukemic stage of acute myeloid leukemia

A particular cancer subtype with substantial symptom burden[1], poor clinical outcome[2], and well-established cell and murine models[3] is acute myeloid leukemia (AML), an aggressive malignancy of the bone marrow characterized by uncontrolled proliferation of undifferentiated myeloid lineage cells[4]. It is also the most common acute leukemia in adults, with over 20,000 new cases in the United States in 2015, approximately 50% of which resulted in patient death[5]. AML has been shown to arise through a clonal process of mutation acquisition where early-stage preleukemic mutations occur predominantly in genes which regulate CpG methylation, such as DNA methyltransferase 3A (*DNMT3A*) and Ten-eleven translocated 2 (*TET2*) while late-stage mutations occur predominantly in various oncogenes and tumor suppressor genes[4,6,7]. Recently, multiple studies have revealed that AML pathogenesis involves the accumulation of mutations within human stem and progenitor cells (HSPCs), creating preleukemic HSPCs (pre-LHSPCs), an ancestral population of undifferentiated cells that have recently been implicated in AML tumor formation, treatment resistance, and tumor regeneration[6]. Thus, these pre-LHSPCs represent a very promising cell population for early treatment in AML development[4]. Studies using expanded patient cohorts have shown that *DNMT3A* and *TET2* loss-of-function mutations likely contribute to AML pathogenesis at the early preleukemic stage even before late-stage mutations occur[8]. Indeed, it has been shown that *DNMT3A* and *TET2* mutations are associated with high preleukemic burden in AML, meaning these mutations confer a growth advantage to HSPCs[4]. As a result, over time, there will be more HSPCs with these mutations as they outcompete normal HSPCs, and these preleukemic, self-renewing cells will have a high probability of going on to form mature AML through additional acquisition of mutations and uncontrolled proliferation[4]. *DNMT3A* is an enzyme that catalyzes the

transfer of a methyl group to the 5' carbon of cytosine, involved in DNA silencing[9]. *TET2* is an enzyme that oxidizes 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), a function proposed to be implicated in un-silencing methylated DNA through a multistage pathway[10]. Thus, these mutations have seemingly opposite functions, yet both represent the most highly enriched mutations in pre-LHSPCs in next generation sequencing-based studies[8]. Mutations in *DNMT3A* occur in approximately 20% of all AML patients with the most common abnormality being a missense mutation at amino acid position 882 (R882C), while mutations in *TET2* occur in approximately 7-10% of adult AML patients and 1.5-4% of pediatric AML patients and are highly variable, with the most common abnormality being a nonsense mutation immediately following R1216[11].

This complex interplay between accumulated mutations, altered chromatin structure, and altered gene expression resulting from epigenomic changes during acute myeloid leukemia (AML) pathogenesis remains poorly understood despite recent advances. In particular, exact mechanisms to explain why loss-of-function of the most highly associated *de novo* loss-of-function mutations in preleukemic AML, *DNMT3A* and *TET2*, leads to aberrant hematopoietic stem and progenitor cell (HSPC) function and malignancy are not well characterized[4,12,13]. These *de novo* mutant HSPCs define an early preleukemic stage of AML that has posed a challenge for treatments particularly because these preleukemic HPSCs (pre-LHPSCs) have been shown to underlie AML initiation, treatment resistance, and malignancy resurgence after treatment[4,12,13]. In the field of AML oncology, recurring questions include which genes are altered epigenetically by mutations in *DNMT3A* and *TET2,* what epigenetic changes are observed in these genes, and when these changes occur after induction of the preleukemic stage. In the search for answers to these questions, next generation live cell chromatin loop, DNA methylation, and histone acetylation biosensors represent attractive alternatives to current *in vitro* methods for quantifying levels of these cellular targets. We propose a potential method to use these live cell biosensors to report on real-time changes in levels of important targets involved in AML pathogenesis during the critical stages of cancer initiation and progression. This approach could inform appropriate treatment intervention for AML patients earlier than normal treatments that usually do not happen until after an oncogenic

driver mutation has been acquired and diagnosed through biopsies and other procedures later in the progression of the disease.

Live cell biosensors offer a sensitive, rapid, and inexpensive way to detect levels of cellular targets in real-time in live cells. To date, live cell biosensing approaches have focused mostly on instantaneous quantification of specific molecular targets and not on understanding their dynamics within a native environment. Furthermore, live cell biosensing methods have mostly used monomeric fluorescent proteins and other enzymatic reporters as transducer elements in design[14]. With such designs, unbound and bound forms of the probe appear nearly identical under the microscope, generally leading to a lower signal-to-background ratio (SBR). Our group and others have recently demonstrated that bipartite "turn-on" biosensor designs that bring two components with negligible signal output in the absence of a target but high signal output in its presence can exceed SBRs of such monomeric "always-on" probes in live cell biosensing of genomic DNA sequences[15-18]. Thus, application of such a bipartite "turn-on" biosensing approach to monitoring levels of other molecular targets *in situ* using a standard fluorescence and luminescence plate reader or optical microscopy is a logical progression within the biosensing field. Using bipartite biosensors and optical microscopy, we hypothesize that tracking dynamic changes in chromatin looping, DNA methylation, and histone acetylation at specific genes during initiation of the preleukemic stage of AML in mutant *DNMT3A* and *TET2* pre-LHSPC models could shed light on major questions relating to pathogenic mechanisms for these preleukemic mutations. To investigate our hypothesis, we propose several core workflows.

Because of the promising preliminary data for our dual dCas9 species split NanoLuc luciferase DNA biosensor, we believe further investigation of its utility within the specific contexts of AML diagnostics and AML pathogenesis is warranted. To this end, an interesting approach would be to investigate feasibility of various cell sorting methods using live cell DNA biosensing to isolate cells with different genotypes at a particular genomic locus. If successful, this approach could allow for creation of isogenic single and double mutant *DNMT3A* and

TET2 pre-LHSPC models for functional characterization of these preleukemic AML mutations. Using a recently published HSPC genome editing approach with high efficiency[19], one could first use *Sp*Cas9, sgRNAs to either *DNMT3A* or *TET2* or both, and donor DNA templates to introduce targeted mutations at these genes, then transfect the DNA biosensor components to the edited cell populations, incubate the cells for 24-48 h, and employ either microscopy combined with single cell cloning cylinders or FACS to isolate single cells homozygous and heterozygous for mutations at these genes. Since additional data using our dual dCas9 species split NanoLuc luciferase DNA biosensor in single cells would be needed to be able to confidently call copy number differences, one could also create these pre-LHSPC models using standard single cell isolation, isogenic population expansion, and genome edit validation methods[20-27]. The main difference between our proposed single cell isolation technique using live cell DNA biosensing and traditional methods is the potential ability to determine genotype by SBR differences before clonal expansion as opposed to having no knowledge of genotype until many single cell isolates are clonally expanded together and analyzed. If our proposed method works, it could save valuable time and resources during this process.

Currently, complete mechanisms to explain why loss-of-function mutations in *DNMT3A* and *TET2* lead to anomalous HSPC function and tumorigenesis are poorly characterized. To be complete, pathogenic mechanisms will need to identify specific genes affected by these mutations, what the nature of their epigenetic changes are, and when these changes occur. In addition, potential routes by which these mechanisms lead to progression of preleukemic AML to later-stages should be identified. We propose that using bipartite live cell biosensors to track dynamic changes in levels of chromatin loops, DNA methylation, and H3K9 acetylation *in situ* during time course experiments of early AML pathogenesis could shed light on its progression to later stages and pathogenic mechanisms for *DNMT3A* and *TET2* mutations, which could lead to development of more informed early-stage AML treatments. The primary goal of these future experiments should be to determine the functional contribution of different *DNMT3A* and *TET2* mutation copy numbers within a wild-type background in changing chromatin looping, DNA methylation, and H3K9 acetylation profiles at various oncogenes and

tumor suppressor genes compared to the status of these profiles in wild-type HSPCs. We hypothesize that there will be changes in chromatin looping frequencies, DNA methylation, and H3K9 acetylation which activate expression of oncogenes and silence tumor suppressor genes, thus causing aberrant HSPC proliferation and hindering differentiation and apoptosis even prior to additional mutation acquisition. This pattern would be consistent with recent evidence showing broad-scale changes in chromatin structure and aberrant CpG methylation are important biomarkers for carcinogenesis in AML and other cancers[7-10,28-31].

For future DNA methylation biosensor designs, we propose to use several bioreceptor elements based on fusion proteins containing the conserved methyl-CpG binding domain (MBD) from the MBD protein family flanked on either side by dCas9 bound to adjacent DNA sequences. In addition to the MBD, MBD family proteins contain other catalytic or protein–protein association domains[32,33]. We propose that the same transducer elements used for chromatin loop biosensing be tested for DNA methylation biosensing. We hypothesize that the likelihood of the transducer components being brought close enough together to produce detectable SBR will increase when a particular combination of dCas9-sgRNA-transducer component and MBD-transducer component binds to a stretch of methylated genomic DNA. To validate functionality of designs, one could use known differentially methylated locations in the genome, perhaps first using recent cell models of differential methylation at the *HER2* promoter created in our lab[34]. In these experiments, one could transfect biosensor components and sgRNAs to the *HER2* promoter and measure SBRs in cell populations with a stably methylated *HER2* promoter compared to those with a hypo-methylated *HER2* promoter. One could also test these biosensor designs at other known hypo-methylated and hyper-methylated regions of the genome. More specifically, this could involve transfection of biosensor components, incubation of cells, and analysis of SBR in populations transfected with and without sgRNAs. The expected result from these experiments would be an increase in SBR in populations of cells transfected with both bipartite DNA methylation biosensor components and sgRNA to the known hyper-methylated regions compared to populations of cells transfected with the biosensor components and no sgRNA. In addition, one could set up a negative control experiment to ensure that regions of

the genome known to be hypo-methylated produce SBRs on par with that of the biosensor components transfected without sgRNAs, indicating expected lack of transducer element proximity. Together, both positive and negative validation experiments could support feasibility of this approach for detection of changes in DNA CpG methylation in real-time in live cells.

For future H3K9 acetylation biosensor designs, we propose one bioreceptor element design making use of immunobiosensing using a recently developed single-chain variable fragment (scFv)—a fusion between the light and heavy chain portions of the variable region of antibodies—which has high specificity towards H3K9ac[35]. This scFv could be fused to similar light-producing transducer elements used in previous designs—split NanoLuc luciferase (NanoBiT) fragments or NanoLuc and HaloTag 618 ligand for NanoBRET. A recent analogue of this design has been developed using fluorescence-resonance energy transfer (FRET) to transduce the recognition signal[36] and this approach offers one of the most versatile and easily reprogrammable bioreceptor elements[37]. One could also develop a second histone acetylation biosensor based on the fragment antigen-binding (Fab') unit—one of two "half" fragments of an antibody, each one consisting of the light variable, light constant, heavy variable, and heavy constant regions and including original disulfide bridge thiol groups—recently developed for H3K27ac[38,39]. The thiol functional groups of the Fab' fragments allow for easy immobilization onto biosensor surfaces and thus amenability to many design considerations[37]. A third design for histone acetylation biosensors could make use of specific bromodomains such as the second bromodomain from BDF1 (BDF1-B), which requires the presence of acetyl-lysine for binding, as bioreceptor elements. While histone acetyltransferases (HATs) are the writers of histone acetylation, bromodomains are the readers[40]. They offer a higher specificity approach to histone acetylation biosensing with lower variability between manufactured batches, albeit with lower affinity for their acetylated lysine targets[41].

Our first proposed experiment to understand how chromatin loops, DNA methylation, and H3K9 acetylation might change on a locus-specific level as a result of preleukemic initiation in AML would involve a reverse

genetic approach: using *DNMT3A* and *TET2* single and double heterozygous, single mutant, and double mutant pre-LHSPC models, one could look for a definite transition to a preleukemic stage using several known cell surface markers that can separate leukemic cells from normal HSPCs, TIM3 and CD99[42-44]. One would first assess which combinations of mutations in either *DNMT3A*, *TET2*, or both produce the TIM3+; CD99+ cell surface marker combination, which would highlight necessary mutation dosage to induce a preleukemic transition. This could involve addition of two distinct antibodies labeled with differently colored fluorophores separately targeting CD99 and TIM3 cell surface markers to the culture medium of *DNMT3A* and *TET2* single and double heterozygous, single mutant, and double mutant pre-LHSPC models and FACS to isolate preleukemic populations. After isolating dual-color fluorescent preleukemic populations of HSPCs by FACS, one could conduct experiments to analyze their chromatin looping interactions, DNA methylation, and H3K9 acetylation at individual loci using proposed bipartite biosensor designs for these targets.

To test our core hypothesis that changes in chromatin contacts will occur such that oncogenes are activated and tumor suppressor genes are silenced early in the preleukemic state, one could develop a comprehensive sgRNA library targeting contacts between promoters of known oncogenes such as *c-MYC* and *KRAS* and their characterized enhancer regions and between promoters of known tumor suppressor genes such as *TP53* and *BEX1* and their characterized enhancer regions. Likewise, for CpG methylation and H3K9 acetylation, one could develop comprehensive sgRNA libraries targeting CpG islands in promoters and gene bodies of known oncogenes and known tumor suppressor genes. Subsequently, one could conduct a set of time course experiments where plasmid biosensor components and promoter-enhancer and CpG island sgRNA libraries are transfected to sorted pre-LHSPCs and wild-type HPSCs and signals are collected at multiple time points post-transfection to visualize changes in specific chromatin contacts, mCpG levels, and H3K9ac levels over time. All transfection experiments should compare signals between at least three but ideally five to twenty separate wells and between individual cells within each well for high statistical power in comparing signals between transfection conditions. After inducing the preleukemic stage in HSPCs, we expect that increases in frequency

of promoter-enhancer contacts, promoter H3K9 acetylation levels, and gene body methylation levels might be observed for screened oncogenes along with decreases in promoter methylation levels. Conversely, we expect the opposite results to be observed for tumor suppressor genes given their opposing contribution to tumorigenesis. These expectations would be revealed by increases in SBR for promoter-enhancer contact, promoter H3K9 acetylation, and gene body methylation biosensing experiments and decreases in SBR for promoter methylation biosensing experiments on oncogenes, and opposite results for tumor suppressor genes. These results could uncover potential early epigenetic activation and silencing mechanisms for *DNMT3A* and *TET2* mutations in AML even prior to acquisition of common AML driver mutations, which is the stage at which most AML malignancies are currently diagnosed. In parallel with bipartite biosensor development and testing in *DNMT3A* and *TET2* mutant pre-LHSPC models, one could conduct Chromosome Conformation Capture-based methods such as 3C or 4C-seq, whole genome or gene-targeted bisulfite sequencing, and H3K9 ChIP-seq or ChIP-qPCR using harvested portions of pre-LHSPC and wild-type HSPC populations at several time points after initial sorting of pre-LHSCs. If these bipartite biosensing approaches are not successful, these *in vitro* methods would ensure changes in chromatin loops, DNA CpG methylation, and H3K9 acetylation could be monitored over time after inducing the preleukemic state in mutant HSPC models.

References

1. Niscola, P., Tendas, A., Scaramucci, L., Giovannini, M., & De Sanctis, V. Pain in blood cancers. *Indian J. Palliat. Care* ***17***(3), 175–183 (2011).
2. Estey, E. & Dohner, H. Acute myeloid leukaemia. *Lancet* ***368***, 1894–1907 (2006).
3. Almosailleakh, M. & Schwaller, J. Murine Models of Acute Myeloid Leukaemia. *Int. J. Mol. Sci.* ***20***(2), 453 (2019).
4. Corces, M. R., Chang, H. Y., & Majeti, R. Preleukemic hematopoietic stem cells in human acute myeloid leukemia. *Front. Oncol. 7*, 263 (2017).
5. De Kouchkovsky, I. & Abdul-Hay, M. Acute myeloid leukemia: A comprehensive review and 2016 update. *Blood Cancer J.* ***6***(7), e441 (2016).
6. Shlush, L. I. *et al*. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* ***506***(7488), 328–333 (2014).
7. Kavianpour, M., Ahmadzadeh, A., Shahrabi, S. & Saki, N. Significance of oncogenes and tumor suppressor genes in AML prognosis. *Tumour Biol*. ***37***(8), 10041-52 (2016).
8. Sato, H., Wheat, J. C., Steidl, U., & Ito, K. DNMT3A and TET2 in the pre-leukemic phase of hematopoietic disorders. *Front. Oncol.* ***6***, 187 (2016).
9. Sandoval, J. E., Huang, Y. H., Muise, A., Goodell, M. A. & Reich, N. O. Mutations in the DNMT3A DNA methyltransferase in acute myeloid leukemia patients cause both loss and gain of function and differential regulation by protein partners. *J Biol. Chem.* ***294***(13), 4898-4910 (2019).
10. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* ***30***(7), 733–750 (2016).
11. Yohe, S. Molecular genetic markers in acute myeloid leukemia. *J Clin. Med.* ***4***(3), 460–478 (2015).
12. Hansen, Q., Bachas, C., Smit, L., & Cloos, J. Characteristics of leukemic stem cells in acute leukemia and potential targeted therapies for their specific eradication. *Cancer Drug Resist*. ***5***(2), 344–367 (2022).
13. Zhang, X. *et al*. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat. Genet.* ***48***(9), 1014–1023.
14. Ali, J., Najeeb, J., Ali, M.A., Aslam, M.F., & Raza, A. Biosensors: Their fundamentals, designs, types and most recent impactful applications: A review. *J. Biosens. Bioelectron.* ***8***, 235 (2017).
15. Sato, Y., Nakao, M., & Kimura, H. Live-cell imaging probes to track chromatin modification dynamics. *Microscopy* ***70***(5), 415-422 (2021).
16. Wu, X. *et al*. A CRISPR/molecular beacon hybrid system for live-cell genomic imaging. *Nucleic Acids Res.* ***46***, e80 (2018).
17. Mao, S., Ying, Y., Wu, X., Krueger, C. J. & Chen, A. K. CRISPR/dual-FRET molecular beacon for sensitive live-cell imaging of non-repetitive genomic loci. *Nucleic Acids Res.* ***47***(20), e131 (2019).
18. Heath, N. G., O'Geen, H., Halmai, N. B., Corn, J. E., & Segal, D. J. Imaging unique DNA sequences in individual cells using a CRISPR-Cas9-Based, split luciferase biosensor. *Front. Genome Ed.* ***4***, 867390 (2022).
19. Gundry, M. C. *et al*. Highly efficient genome editing of murine and human hematopoietic progenitor cells by CRISPR/Cas9. *Cell Rep.* ***17***(5), 1453–1461 (2016).
20. Giuliano, C. J., Lin, A., Girish, V. & Sheltzer, J. Generating single cell-derived knockout clones in mammalian cells with CRISPR/Cas9. *Curr. Protoc. Mol. Biol.* ***128***, e100 (2019).
21. Mathupala, S. & Sloan, A. A. An agarose-based cloning-ring anchoring method for isolation of viable cell clones. *BioTechniques* ***46***, 305–307 (2009).
22. Hu, P., Wenhua Zhang, Xin, H. & Deng, G. Single cell isolation and analysis. *Front. Cell Dev. Biol.* ***4***, 116 (2016).
23. Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P. & Pruett-Miller, S. M. A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.* ***8***, 888 (2018).
24. Ren, C., Xu, K., Segal, D. J. & and Zhang, Z. Strategies for the enrichment and selection of genetically modified cells. *Trends Biotechnol.* ***37***, 56–71 (2019).

25. Bauer, D. E., Canver, M. C. & Orkin, S. H. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *JoVE* ***95***, e52118 (2015).
26. Vouillot, L., Thélie, A., and Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3* ***5***, 407–15 (2015).
27. Basu, S., Campbell, H. M., Dittel, B. N., & Ray, A. Purification of specific cell population by fluorescence activated cell sorting (FACS). *JoVE* ***41***, 1546 (2010).
28. Zhou, Y. *et al*. Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. *Nat. Commun.* ***10***(1), 1522 (2019).
29. Costello, J. F. *et al*. Aberrant CpG-island methylation has non-random and tumor-type-specific patterns. *Nat. Genet.* ***24***, 132–138 (2000).
30. Kalari, S., & Pfeifer, G. P. Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. *Adv. Genet.* ***70***, 277–308 (2010).
31. Pfeifer, G. P. Defining driver DNA methylation changes in human cancer. *Int. J. Mol. Sci.* ***19***(4), 1166 (2018).
32. Clouaire, T., & Stancheva, I. Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *CMLS* ***65***(10), 1509–1522 (2008).
33. Du, Q., Luu, P. L., Stirzaker, C. & Clark, S. J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics 7*(6), 1051-73 (2015).
34. O'Geen, H. *et al*. Ezh2-dCas9 and KRAB-dCas9 enable engineering of epigenetic memory in a context-dependent manner. *Epigenetics Chromatin* ***12***, 26 (2019).
35. Sato, Y. *et al*. Genetically encoded system to track histone modification in vivo. *Sci. Rep*. ***3***, 2436 (2013).
36. Chung, C. I. *et al*. Intrabody-based FRET probe to visualize endogenous histone acetylation. *Sci. Rep.* ***9***(1), 10188 (2019).
37. Crivianu-Gaita, V. & Thompson, M. Aptamers, antibody scFv, and antibody Fab' fragments: An overview and comparison of three of the most versatile biosensor biorecognition elements. *Biosens. Bioelectron*. ***85***, 32–45 (2016).
38. Hayashi-Takanaka, Y. *et al*. Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res*. ***39***(15), 6475–6488 (2011).
39. Kimura, H., Hayashi-Takanaka, Y., Stasevich, T. J., & Sato, Y. Visualizing posttranslational and epigenetic modifications of endogenous proteins in vivo. *Histochem. Cell Biol*. ***144***(2), 101–109 (2015).
40. Marmorstein, R., & Zhou, M. M. Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb. Perspect. Biol*. ***6***(7), a018762 (2014).
41. Bryson, B. D., Del Rosario, A. M., Gootenberg, J. S., Yaffe, M. B., & White, F. M. Engineered bromodomains to explore the acetylproteome. *Proteomics* ***15***(9), 1470–1475 (2015).
42. Jan, M. *et al*. Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. *Proc. Natl. Acad. Sci*. ***108***(12), 5009–5014 (2011).
43. Kikushige, Y. *et al*. TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. *Cell Stem Cell 7*, 708–17 (2010).
44. Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C., & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl. Acad. Sci*. U S A ***111***(7), 2548–2553 (2014).

Chapter 7: Future Directions for Split Reporter DNA and Chromatin Loop Biosensors

Future Directions in Development and Application of Split Reporter DNA Sequence Biosensors

While we have demonstrated that our DNA biosensors can detect DNA sequences in living cells with high sensitivity and specificity, our modular designs likely still lack optimization for several parameters. It is possible the designs that have been extensively tested up to this point could be hindered due to competition for binding sites on chromatin with endogenous factors, inaccessibility of binding sites on silenced chromatin, or for other reasons. In addition, cell type variability in results for our DNA biosensors thus far is a major concern and may limit their application to cell types that are highly amenable to transient transfection. Thus, we conclude that future optimization of our DNA biosensors—particularly our dual dCas9 species DNA biosensor—is warranted to further maximize signal-to-background. For future development of new DNA biosensor designs, we strongly believe that proposing several design strategies increases the likelihood that any single design will be successful in detecting DNA at high signal-to-background. In a broader sense, we are very encouraged by the data gathered for our various DNA biosensor designs and we are excited about future possibilities offered by our live cell split reporter DNA biosensing technology. Specifically, we are most excited about the applications in quantifying and tracking levels of specific DNA targets across a range of contexts including noninvasive live cell genotyping and detection of *ex vivo* gene editing modifications to DNA. In addition to identification of specific sequences in single cells, other very promising applications of our DNA biosensors include physical isolation of individual cells from complex populations based on presence of a particular mutation or based on presence of a particular number of copies of a DNA sequence.

Our DNA biosensors have shown the ability to accurately differentiate between cells containing different SNPs at a single locus. Therefore, in addition to being a useful platform for noninvasive determination of genotype in live cells, we hypothesize that our DNA biosensors may also be useful for identifying cells with different copy

number variants (CNVs) at a particular genomic locus. Provided our DNA biosensing platform produces sufficient SBR differences between different CNVs and that the user can identify copy number SBR cutoffs for a particular set of CNVs with high confidence, it could allow for identification of cells of variable copy number within a population of cells. To determine whether we could see differences in CNVs using our DNA biosensor, it is most intuitive to consider whether we observed significant differences in SBR at repetitive genomic sites with many copies of the same sequence compared to non-repetitive genomic sites with only a single copy of a sequence. Perhaps counterintuitively, our DNA biosensor has generally shown higher SBR at non-repetitive genomic sites compared to repetitive genomic sites. This effect could be explained by possible steric hindrance effects at repetitive genomic sites, as repeats tested at *MUC4* were only ~48 bp. Repeats of such short sequences may cause dCas9-NanoBiT biosensor elements to sterically interfere with one another between repeats, and longer repeat lengths might be needed to minimize such interference of dCas9-NanoBiT biosensors binding to adjacent repeats. Thus, further analysis is needed to determine whether a progressive increase in SBR is observed with increases in mutation copy number in cells transfected with our DNA sequence biosensor. To become confident in copy number calls using our DNA sequence biosensor, one would need to gather additional SBR data across many different copy number genomic loci to account for interlocus variability in efficiency of biosensor reassembly due to variable gRNA binding efficiency to dCas9 target sites, differences in spacing and orientation between the two dCas9 target sites, and accessibility of chromatin, among other factors. This could involve obtaining mutant models of various copy numbers or identifying cell lines in which copy numbers naturally vary using COSMIC or RefCNV, creation of large sets of gRNAs to target specific loci where mutations of variable copy number exist, and identical biosensing experiments to those conducted during our biosensor characterization experiments involving transfection of biosensor components, cell incubation, and downstream statistical analysis of SBR. Then, this complete data set could be used to determine whether appropriate high-confidence cutoffs exist for calling mutant cells from wild-type. In addition, if our hypothesis is true, a user of our biosensor could potentially use this approach in practice to isolate cells that are heterozygous (1 mutation copy for a diploid), from cells that have a homozygous mutation present (2 mutation

copies for a diploid) in certain genes post-genome editing, allowing for increased efficiency in the genome editing process for correction of genes involved in serious diseases. This type of analysis would be highly dependent on the chromosomal copy number of the chromosome where a given mutation resides within a particular cell type.

In addition, if different sequences or different copy numbers can be differentiated with high confidence within a population of cells, a logical progression would be toward isolation methodologies for cells showing a higher SBR from those showing a lower SBR according to predetermined SBR cutoff ranges. We never attempted to use our DNA biosensors to physically separate cells from a complex mixture of cells with different genotypes or CNVs at a given locus. For this to be possible in practice, one could use a combination of manual microscopy for SBR determination, precise algorithmic determination of variable SBR cell positions, and robotic isolation using specialized single cell cloning cylinders. For such an isolation technique to work based on SBR cutoffs, a large data set with both non-targeting and no sgRNA background conditions could be used to train the algorithm to call cells negative for a particular sequence for each cell line of interest. Intriguingly, it is possible to imagine application of our dual dCas9 species NanoBRET DNA biosensor in an experimental sorting technique using red fluorescence output where one could try to sort successful gene edited clones—or even simply cells with different genotypes—based on acceptor HaloTag 618 ligand fluorescent signal differences at specific genomic sites. Specifically, it is conceivable that Fluorescence-Activated Cell Sorting (FACS) could be used for determination of cells producing red fluorescent biosensing signals beyond a certain cutoff threshold and for isolation of single cells into plates where these cells could be expanded to produce homogenous populations for downstream applications. However, this process might be mechanically tedious since, prior to sorting, one would need to introduce the NanoLuc luminescent substrate to the cell population to induce BRET and red fluorescence in the cells. This critical initial step would need to be conducted in at least the first 5-10 minutes before starting the FACS process to allow for sufficient time for uptake of the luminescent substrate.

Another exciting potential future application of our DNA biosensing system could be the direct application to detection of the formation or destruction of "perfect" junction regions on genomic DNA that form frequently in cells as a result of chromosomal rearrangement processes such as insertions, deletions, inversions, or translocations. Functionally, one could design several sgRNAs for each of the two expected junction regions for a given insertion, deletion, inversion, or translocation to verify whether the rearrangement of interest has occurred. Each of the two expected genomic junction regions could be targeted with our DNA biosensor and sgRNA pairs could be optimized for a particular region in wild-type cells. There are two possible ways to design such experiments: targeting the formation of a new junction region at a given genomic position that is different than the sequence at this position in wild-type cells or targeting the destruction of a particular wild-type genomic sequence. For the former, one would use sgRNAs for the new expected sequence that would be present if the junction of interest forms successfully, and for the latter, one would use sgRNAs targeting the wild-type sequence that should be destroyed when a junction forms. As the wild-type sequence is known in all cases and mutations in the wild-type sequence are typically of interest to a wide range of disease-related research questions, the latter methodology is both easier to apply in practice and perhaps more versatile. After transfection of biosensor components and optimized sgRNAs for the wild-type sequence, a researcher targeting the destruction of this wild-type sequence due to chromosomal rearrangements would expect to compare a statistically significant level of signal-to-background in wild-type cells to very little signal-to-background in the cells where a chromosomal rearrangement is observed at that genomic location. One drawback of screening based on sgRNAs targeting the wild-type sequence is that it would not inform a researcher about what type of mutation or genomic rearrangement has occurred. However, such a method could be used in particularly exciting applications such as noninvasively screening for mutations and genomic rearrangements formed at mutational "hotspots" in cancer cells in real-time.

While we have applied these DNA biosensing methods strictly *ex vivo* to monitor the presence of specific DNA sequences in live cells, perhaps a more useful application for these live cell DNA biosensors would be their

direct application to monitoring mutation of specific DNA sequences *in vivo*. Many tumors in research animals are monitored using bioluminescence[1,2]. This is typically done using delivery of a luminescent substrate to tumor cells which are modified to express luciferase. Most of these luciferase-expression systems in tumor cells are designed with the goal of tracking the size of a particular tumor over time. However, it is also possible to imagine using our split luciferase DNA biosensor to track mutations in these animals over time. Given that many *in vivo* bioluminescence monitoring applications for research animals with tumors involve Firefly luciferase and its luciferin substrate, our NanoLuc luciferase and furimazine substrates could even be used in coordination with current monitoring methods. In theory, using this DNA biosensor in live animals could allow for detection and tracking of novel mutations that might contribute to tumor initiation and progression. Additional experiments using our dual dCas9 species split NanoLuc luciferase biosensor *in vivo* will be required in order to determine whether its application to detecting mutations in live animals could work in practice.

In addition, given our promising early *in vitro* data from cell lysates, we believe purified ribonucleoprotein versions of our DNA biosensors could be used to detect DNA sequences *in vitro*. Over the last decade, the cancer screening and detection field has seen a large increase in the types of screening technologies available for early tumor detection. One relatively new approach of particular interest within the scientific community is using circulating tumor DNA (ctDNA) to track and monitor tumorigenesis and tumor progression[3]. For this methodology, blood samples—referred to as liquid biopsies—are taken from patients to screen for circulating DNA from many types of tumors that have not yet metastasized, including breast cancer, colorectal cancer, spinal chordomas, and head and neck cancer[4-7]. Currently, a wide range of techniques such as digital PCR, real-time PCR, mass spectrometry, next-generation sequencing, and long-read nanopore-based DNA sequencing are used to detect ctDNA from blood samples[8,9]. In addition, ctDNA is also present within other non-blood bodily fluids such as urine, saliva, sputum, stool, cerebrospinal fluid (CSF), and pleural fluid[10]. However, there is no unified standard for ctDNA detection as it is present at extremely low concentrations in blood[11]. Therefore, ctDNA detection requires highly sensitive assays and false positives are common with current methods[11]. As

our preliminary results demonstrate our DNA biosensor is highly sensitive and specific and can detect as low as 190 amol of a DNA sequence in cells, we believe it could potentially be used to detect the 5% of cell-free DNA in the blood that originates in the tumor[9]. Indeed, recent studies have produced such biosensors for ctDNA[12,13] and many possible bioreceptor elements can be envisaged for development of such biosensors such as complementary DNA (cDNA), peptide nucleic acid (PNA) and anti-5methylcytosines[14]. Purified dCas9 protein could provide a bioreceptor element with very high affinity for specific DNA sequences *in vitro*. Electrochemical biosensors represent a very promising platform for ctDNA detection due to their unique advantages such as high sensitivity, high specificity, low cost, and good portability[15]. However, it is not difficult to imagine chemiluminescent or fluorescent biosensors being successfully applied to detect ctDNA in the future. Additional *in vitro* experiments using our dual dCas9 species split NanoLuc luciferase or NanoBRET DNA biosensors will be required in order to determine whether their application to ctDNA detection from liquid biopsies could work in practice.

Future Directions in Development and Application of Split Reporter Chromatin Loop Biosensors

Because the anchor regions of chromatin loops—especially those formed by *cis* promoter-enhancer interactions—are typically formed at areas of active chromatin with high accessibility[16,17], it is less likely that our dual dCas9 species DNA biosensor would be hindered due to competition for binding sites on chromatin with endogenous factors, inaccessibility of binding sites on silenced chromatin, or for other reasons when targeted to many chromatin loop anchors than it would be when targeted to individual non-interacting sites on genomic DNA, which can be in either euchromatic or heterochromatic regions. However, future optimization of the base platform for our dual dCas9 species DNA biosensor is necessary to further maximize SBR for its specific application to chromatin loops. We believe the data we have gathered for the chromatin loop biosensor application of our DNA sequence biosensing platform is very encouraging and we are excited about future

possibilities offered by a live cell split reporter chromatin loop biosensing technology. Specifically, we are most excited about the possibility of using our chromatin loop biosensor to predict changes in gene expression at a given locus by pairing live cell biosensing at a loop anchor region formed by a given promoter-enhancer pair with RT-qPCR for the gene of interest.

Enhancers are gene-distal sequences that increase the transcription of a gene by *cis*-interaction with its promoter region[18,19]. To functionally link changes in promoter-enhancer chromatin interactions to resulting variation in gene expression over time, we suggest creating a platform where one could pair our real-time, noninvasive chromatin loop biosensor to track levels of verified promoter-enhancer looping pairs over time with RT-qPCR done on mRNA samples gathered at various time points to quantify gene expression for the gene containing the promoter of interest over time. In this way, changes in the frequency of promoter-enhancer chromatin loops could be linked to changes in gene expression for a given gene promoter-enhancer pair. This platform would be especially useful for functional studies relating to the effects of changes in cell state on chromatin architecture and gene expression. One type of cell state change that could be easily replicable in experiments designed to test live cell biosensors would be changes in conditions within the extracellular environment. For example, recent evidence has shown that promoter-enhancer interactions can change considerably during hypoxic conditions. Notably, hypoxia has been shown to increase certain chromatin interactions between putative enhancer elements and the MALAT1 locus in breast cancer cells[20]. Thus, we believe that an intriguing application of our biosensor would be in tracking changes to chromatin loops in breast cancer cells as a result of induction of hypoxic conditions using sgRNA pairs for the MALAT1 promoter and putative enhancer elements.

In addition, our chromatin loop biosensor could be used to monitor changes in loops not related to promoter-enhancer interactions, such as those at TAD boundaries. Evidence in the literature has revealed dramatic changes in TAD boundaries during cell state transitions such as those observed during the transition from pluripotency to a determined cell lineage[21]. Specifically, loop domains formed by chromatin are less frequent,

closer-range, and "looser" such that they are more permissive to inter-TAD interactions during pluripotency but become more numerous, longer-range, and "tighter" such that they allow fewer inter-TAD interactions upon differentiation and cell lineage determination[21]. Thus, we imagine application of our chromatin loop biosensor to tracking the formation of new long-range loop domains or to tracking the dissolution of short-range inter-TAD loop domains during cell differentiation. This concept could have wide ranging utility for differentiation within a number of stem cell subtypes and for differentiation of neural progenitor cells (NPCs) or neural stem cells (NSCs) to functional neurons. One potential drawback of such an application it that it would require some preexisting knowledge of loop domains in the pluripotent cell type of interest and time course imaging of these loop domains during differentiation. This naturally necessitates high amenability to transient transfection procedures in stem cell lines which are normally more difficult to transfect and requires a very stable biosensing signal over the course of a number of days or weeks.

In addition to a platform to link real-time changes in promoter-enhancer interactions and changes in TAD boundaries to changes in gene expression, our chromatin loop biosensor could be used to screen for novel enhancer elements for a given promoter region. In recent years, high-throughput approaches have been utilized to identify or validate enhancers. The first approach is to use techniques to search genomes for marks associated with enhancer activity, such as EP300 ChIP-seq, H3K27ac ChIP-seq, H3K4me1 ChIP-seq, DNase I hypersensitivity (DHS), and others[22]. Other assays that measure enhancer activity are *in vitro* luciferase assays and massively parallel reporter assays (MPRAs), which rely on sequencing-based quantification of thousands of RNA barcodes in parallel, each associated with a different candidate enhancer[23-25]. In addition, strategies using CRISPR-Cas9-based approaches to identify novel enhancer elements have been described[26]. If changes in gene expression of a gene of interest were to first be measured in a cell population, one could design a large enough sgRNA library targeting sites across the genome and apply each of these sgRNAs paired with an optimized sgRNA targeting the promoter of the gene of interest and measure any statistically significant increase in the biosensing signal-to-background ratio. As most sgRNAs in this large library would produce a signal near

background levels when paired with the promoter sgRNA, and our dual dCas9 species DNA biosensor has shown a signal-to-background of between 1.08-fold and 6.1-fold when applied to promoter-enhancer pairs across two cell lines, we hypothesize that putative enhancer candidates could be identified based on statistically significant spikes in the signal-to-background ratio when analyzing the resulting data.

References

1. Close, D. M., Xu, T., Sayler, G. S., & Ripp, S. In vivo bioluminescent imaging (BLI): noninvasive visualization and interrogation of biological processes in living animals. *Sensors (Basel)* ***11***(1), 180–206 (2011).
2. Kim, S. *et al*. Optimizing live-animal bioluminescence imaging prediction of tumor burden in human prostate cancer xenograft models in SCID-NSG mice. *Prostate* ***79***(9), 949–960 (2019).
3. Bettegowda, C. *et al*. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med*. ***6***(224), 224ra24 (2014).
4. Addanki, S., Meas, S., Sarli, V. N., Singh, B., & Lucci, A. Applications of circulating tumor cells and circulating tumor DNA in precision oncology for breast cancers. *Int. J. Mol. Sci.* ***23***(14), 7843 (2022).
5. Arisi, M. F., Dotan, E., & Fernandez, S. V. Circulating tumor DNA in precision oncology and its applications in colorectal cancer. *Int. J. Mol. Sci.* ***23***(8), 4441 (2022).
6. Mattox, A. K. *et al*. The mutational landscape of spinal chordomas and their sensitive detection using circulating tumor DNA. *Neuro-oncol. Adv.* ***3***(1), vdaa173 (2020).
7. Egyud, M. *et al*. Plasma circulating tumor DNA as a potential tool for disease monitoring in head and neck cancer. *Head & Neck*, ***41***(5), 1351–1358 (2019).
8. Saha, S., Araf, Y. & Promon, S.K. Circulating tumor DNA in cancer diagnosis, monitoring, and prognosis. *J. Egypt Natl. Canc. Inst.* ***34***, 8 (2022).
9. Marcozzi, A. *et al.* Accurate detection of circulating tumor DNA using nanopore consensus sequencing. *npj Genom. Med.* ***6***, 106 (2021).
10. Peng, M., Chen, C., Hulbert, A., Brock, M. V., & Yu, F. Non-blood circulating tumor DNA detection in cancer. *Oncotarget* ***8***(40), 69162–69173 (2017).
11. Franczak, C. *et al*. Technical considerations for circulating tumor DNA detection in oncology, *Expert Rev. Mol. Diagn.* ***19***(2), 121-135 (2019).
12. Yang, X. *et al*. An electrochemiluminescence resonance energy transfer biosensor for the detection of circulating tumor DNA from blood plasma. iScience ***24***(9), 103019 (2021).
13. Miao, P., Chai, H., & Tang, Y. DNA hairpins and dumbbell-wheel transitions amplified walking nanomachine for ultrasensitive nucleic acid detection. ACS Nano. ***16***(3), 4726-4733 (2022).
14. Li, X. *et al*. Liquid biopsy of circulating tumor DNA and biosensor applications. *Biosens. Bioelectron.* ***126***, 596-607 (2019).
15. Wang, K. *et al*. Electrochemical biosensors for circulating tumor DNA detection. *Biosensors (Basel)* ***12***(8), 649 (2022).
16. Kadauke, S., & Blobel, G. A. Chromatin loops in gene regulation. *Biochimica et Biophysica Acta* ***1789***(1), 17–25 (2009).
17. Yuan, J. *et al*. Open chromatin interaction maps reveal functional regulatory elements and chromatin architecture variations during wheat evolution. *Genome Biol.* ***23***(1), 34 (2022).
18. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. Enhancers: five essential questions. *Nat. Rev. Genet.* ***14***(4), 288–295 (2013).
19. Panigrahi, A., & O'Malley, B. W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* ***22***(1), 108 (2021).
20. Stone, J.K. *et al*. Hypoxia induces cancer cell-specific chromatin interactions and increases MALAT1 expression in breast cancer cells. *J Biol. Chem*. ***294***(29), 11213-11224 (2019).
21. Pękowska, A. *et al*. Gain of CTCF-anchored chromatin loops marks the exit from naive pluripotency. *Cell Syst*. *7*(5), 482-495.e10 (2018).
22. Shlyueva, D., Stampfel, G., & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* ***15***, 272–286 (2014).
23. Patwardhan, R.P. *et al*. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* ***27***, 1173–1175 (2009).

24. Patwardhan, R.P. *et al*. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* ***30***, 265–270 (2012).
25. Melnikov, A. *et al*. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* ***30***, 271–277 (2012).
26. Klein, J. C., Chen, W., Gasperini, M., & Shendure, J. Identifying novel enhancer elements with CRISPR-based screens. *ACS Chem. Biol.* ***13***(2), 326–332 (2018).