DEVELOPMENT & APPLICATION OF NOVEL METHODS TO IDENTIFY LOCUS-SPECIFIC TRANSCRIPTIONAL REGULATORS

by

Ipek Selcen

B.S., Molecular Biology and Genetics, Istanbul Technical University, 2016

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics
Bogazici University

2019

ACKNOWLEDGEMENTS

I would like to express my strongest gratitude to my supervisor, Assoc. Prof. Tolga Emre for his endless support and guidance throughout my thesis studies. Along with my master studies, he guided me during tough times and contributed too much on my improvement as a researcher.

I appreciate my lab-mates Nalan Yıldız, Ulduz Sobhi Afshar, Anna Öğmen and Laila Hedaya for their friendship and practical contributions. I would like to thank the members of Sumo Lab members, Bahriye Erkaya, Sezgi Canaslan, Yusuf Tunahan Abaci and especially to Arda Baran Çelen for his endless patience throughout the times when I was stressed. They supported me in all aspects and helped me whenever I needed them the most.

I would like thank Assos. Prof. Umut Sahin for helping me in every way that he can. Besides, I would like to thank him for evaluating my thesisç

I would also like to thank my third jury member Assoc. Prof. Ferruh Ozcan for sparing his valuable time to evaluate, criticize and improve this thesis.

Last but not least, I would like to emphasize my strongest gratitude to my family especially to my sister Selen Selcen. This thesis is dedicated to them for their support throughout my life.

Finally, I would like to emphasize my gratitude to TUBITAK for supporting me financially.

ABSTRACT

DEVELOPMENT & APPLICATION OF NOVEL METHODS TO IDENTIFY LOCUS-SPECIFIC TRANSCRIPTIONAL REGULATORS

Chromatin-associated factors play important roles in the regulation of gene expression. It is essential to identify these factors in order to elucidate their functional roles at the target sites. There are a number of methodologies to identify these chromatin regulators in a locus-specific manner, however, these approaches have their own drawbacks and limitations.

In this study, we aim to develop a new method called “Split-CRISPR-ID” to identify locus-specific transcriptional regulators by combining the flexible genomic locus targeting power of CRISPR proteins with *in vivo* biotinylation-based purification, and proteomic analysis. We set up the system by using two distinct CRISPR proteins fused with the inactive halves of target promiscuous mutant BirA* biotin ligase. We first validated the targeting efficiencies of the CRISPR proteins at various loci on the genome. Then, we created CRISPR-BirA* halves fusion proteins to control the *in vivo* biotinylation event spatiotemporally target specific loci. Our preliminary data shows heterodimerization and activation of BirA* at designed genomic loci. Further studies will provide the proteomic discovery when the method is fully implemented. In parallel, we also aimed to optimize and apply in our laboratory a recently published method for locus-specific protein discovery called “CAPTURE”. CAPTURE is based on the principle of biotinylation and isolation of a nuclease-deficient Cas9 with associated chromatin proteins. Here, we validated the specificity of targeting via various approaches.

We envision that the development and application of such improved locus-specific protein identification methods will help us discover novel locus-specific transcriptional regulators, and contribute to the development of novel therapeutic strategies.

ÖZET

LOKUSA ÖZGÜ TRANSKRİPSİYONEL FAKTÖRLERİN KEŞFİ İÇİN YENİ METOTLARIN GELİŞTİRİLMESİ VE UYGULANMASI

Kromatin ile ilişkili faktörler gen regülasyonunda önemli rol oynarlar. Bu faktörlerin tanımlanması, hedef DNA sekanslarındaki fonksiyonel rollerinin aydınlatılması için önemlidir. Bu kromatin regülatörlerini lokus-spesifik şekilde tanımlamak için belirli birkaç metodoloji vardır, fakat bu metotların bir sürü dezavantajları ve kısıtlamaları bulunmaktadır.

Biz bu çalışmada, lokus-spesifik transkripsiyonel regülatörleri keşfedebilmek için, CRISPR proteinlerinin esnek hedeflenebilirliği ile *in vivo* olarak kullanacağımız biotinilasyonu birleştirerek Split-CRISPR-ID olarak adlandırdığımız yeni bir metot geliştirmeyi hedefledik. Bu sistemi seçici olmayan mutant BirA* biotin enziminin inaktif iki parçasını iki farklı CRISPR proteini ile birleştirerek kurduk. Öncelikle, CRISPR proteinlerinin bağlanma yeterliliklerini çeşitli gen bölgelerinde doğruladık. Sonrasında, amaçladığımız spesifik gen lokusunda *in vivo* biotinilasyonu zamansal ve konumsal açılardan kontrol etmemizi sağlayacak olan füzyon proteinlerimizi oluşturduk. İlk sonuçlarımız BirA* enziminin test ettiğimiz gen lokuslarında birleştiğini ve aktive olduğunu gösteriyor.

Bunun yanında, yakın zamanda yayınlanmış olan CAPTURE metodunu optimize etmeyi amaçladık. Bu method ile spesifik gen lokuslarını güçlü biotin-streptavidin etkileşimi sayesinde izole ettik. Bu bölgelerdeki lokus-spesifik proteinleri western blot yöntemiyle gözlemledik.

Tez kapsamında çalıştığımız bu metotların yeni lokus-spesifik transkripsiyonel regülatörlerin keşfedilmesinde ve bunlara yönelik terapiler geliştirilmesinde yardımcı olacağını düşünüyoruz.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

bp	Base Pairs
g	Gravity
gr	Gram
kb	Kilobase
kDa	Kilodalton
L	Liter
M	Molar
mA	Milliampere
mg	Milligram
min	Minute
ml	Milliliter
mm	Millimeter
mM	Millimolar
ng	Nanogram
s	Second
V	Volt
β	Beta
µg	Microgram
µl	Microliter
^{o}C	Centigrade degree

LIST OF ACRONYMS / ABBREVIATIONS

APS	Ammonium persulfate
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
Cas9	CRISPR-associated endonuclease 9
Cpf1	CRISPR-associated endonuclease in Prevotella and Francisella 1
cDNA	Complementary DNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CO_2	Carbon dioxide
dCas9	Nuclease-deficient Cas9
dCpf1	Nuclease-deficient Cpf1
ddH_2O	Double-distilled water
DMEM	Dubecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ECL	Enhanced chemiluminescent solution
EDTA	Ethylenediaminetetraacetate
EtOH	Ethanol
FBS	Fetal bovine serum
gDNA	Genomic DNA
GFP	Green fluorescent protein
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
LB	Luria-Bertani broth
MAPK	Mitogen-activated protein kinase
$MgCl_2$	Magnesium chloride
mRNA	Messenger RNA
MeOH	Methanol
N-terminal	Amino terminal

C-terminal	Carboxy terminal
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PVDF	Polyvinylidene fluoride
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
QPCR	Quantitative PCR
SDS	Sodium dodecyl sulfate
sgRNA	Single guide RNA
TALEN	Transcription Activator-Like Effector Nucleases
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline with tween
TEMED	Tetramethyl ethylenediamine
tracrRNA	Trans-activating crRNA
Tris	Tris (hydroxymethyl) aminomethane
Tween	Polysorbate 20
Unt	Untransduced

1. INTRODUCTION

The Human Genome Project revealed the relatively permanent sequences that codify who we are as a species (Olson, 1993). Since then, a great deal of research attention has been given to identifying specific alterations in these sequences that result in diverse disease states. However, there is another level of molecular mechanisms that define us, and that was recently brought into the limelight. These so-called "epigenetic" mechanisms determine how our genes are expressed in a much more dynamic fashion while leading to dramatic phenotypic outcomes (Deans & Maggert, 2015; Jaenisch & Bird, 2003). They are responsible for the generation of hundreds of different cell types within an individual organism, although each cell is provided with only a single common blueprint (Cazaly *et al*., 2015; Deans & Maggert, 2015). They are also immensely sensitive to environmental factors (Sutherland & Costa, 2003). For instance, the NASA Twins Study explored the short- and long-term effects of zero gravity aboard the International Space Station and discovered several epigenetic alterations such as transcriptional activity and immune cell DNA methylation status (Garrett-Bakelman *et al*., 2019). Lastly, the epigenetic landscape can also be modulated with etiological consequences in different complex diseases, such as cancer, autoimmunity, and neurodegenerative disorders (Cazaly *et al.*, 2015). Nevertheless, despite the central roles played by epigenetic processes, our knowledge of these processes is still not complete. As the epigenetic landscape differs among individuals and various disease states, a comprehensive and personalized therapeutic approach should be able to account for and possibly exploit these discrepancies (Chatterjee & Ahituv, 2017; Halusková, 2010). Our purpose in this project is to develop a method that can partially address this lack of knowledge and lay the groundwork for the next generation of therapies by identifying locus-specific gene regulators.

1.1. Genome Organization and Transcriptional Regulation

"Epigenetic regulation" is implemented via three distinct modalities, including histone modifications and nucleosome positioning, DNA methylation, and involvement of

other chromatin-associated complexes (Jaenisch & Bird, 2003; Kim *et al*., 2009). Before further exploring these mechanisms, it is vital to comprehend the environment within which they operate. A close-up view of the chromatin reveals a "beads-on-a-string" structure, where the string is the double-stranded DNA molecule and the beads consist of protein assemblies called nucleosomes (Luger *et al*., 1997; Shahid *et al*., 2019). The chromatin achieves its compactness partially by enfolding the DNA around these nucleosome cores, which hampers gene expression by preventing the access of most transcription factors and other chromatin-associated factors (Gaffney *et al*., 2012) (Figure 1.1). However, since gene expression is a highly-regulated process, the positioning of the nucleosomes is also subject to precise and dynamic regulation (Lai & Pugh, 2017). This regulation frequently occurs at the level of histone proteins, whose octameric assemblies constitute the nucleosome core (Manelyte, 2017). Multiple post-translational modifications (PTMs) may alter the affinity of the histones to the negatively-charged DNA backbone, often by occluding the positively-charged amino acid residues, such as in the case of acetylation (Gaffney *et al*., 2012). Other modifications may facilitate the recruitment of chromatin-remodeling complexes, which can also effect changes in nucleosome placement (Lai & Pugh, 2017; Moore *et al*., 2013; Wolffe, 2001). These functional outcomes are predicated on the precisely-orchestrated interplay between a variety of PTMs on several residues on both the histone core and the N-terminal tails, leading to the establishment of a histone code (Shahid *et al*., 2019).

Vincent Allfrey pioneered the efforts to decipher this code by discovering histone modifications such as methylation and acetylation and causally linking them to changes in transcriptional activity (Allfrey *et al*., 1964). In time, this modification repertoire was expanded to include a broad array of PTMs such as phosphorylation, ubiquitylation, and sumoylation (Bannister & Kouzarides, 2011). Interestingly, the neurotransmitter serotonin was recently discovered to act as a histone modifier as well (Farrelly *et al*., 2019). A prominent modification consists of histone 3 lysine 27 trimethylation (H3K27me3), which dramatically inhibits transcriptional activity (King *et al*., 2016). For instance, breast, ovarian and pancreatic cancers, among others, display a loss of H3K27me3, which helps them facilitate disease progression by upregulating cellular oncogenes (Wei *et al*., 2008). On the other hand, H4-Lys16 acetylation and H4-Lys20 trimethylation were reported to be abolished on the repetitive sequences of tumor cells, a phenotype recognized as a cancer hallmark (Fraga *et al*., 2005; Schotta *et al*., 2004). These findings support the idea that

aberrant histone modification patterns may contribute to neoplastic transformation (Fraga *et al*., 2005).

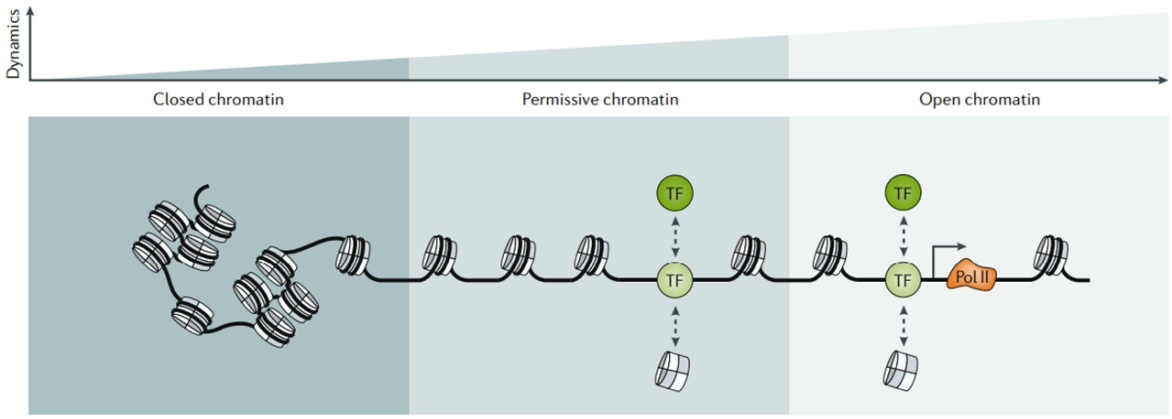


Figure 1.1. Closed chromatin is not accessible by most TFs. Since the permissive chromatin has DNA-linker suitable for TF binding, open chromatin state can be established through chromatin remodelers leading to transcriptional activity. TF, transcription factor; Pol II, RNA polymerase II (from Klemm *et al*., 2019).

Epigenetic regulation also occurs via the transfer of methyl groups onto DNA nucleotides (Jones & Gonzalgo, 1997). DNA methyltransferases (DNMTs) are the enzymes responsible for introducing and maintaining DNA methylation (King *et al*., 2016). This reversible modification can dynamically alter a gene's activity without altering its sequence identity, although, remarkably, methylation pattern is an inheritable property (Kim *et al*., 2009; Robertson, 2005). DNA methylation mostly leads to gene silencing by recruiting compatible transcriptional repressors and limiting transcriptional activator binding (Jin *et al*., 2011). The regulation of gene expression is achieved by the cooperation of DNA methylation and histone modifications (Jaenisch & Bird, 2003; Reid *et al*., 2009). In this context, DNA modifiers such as MeCP2 recognize methyl groups on DNA and recruit histone modifiers such as histone deacetylases (HDACs) to DNA (Fuks *et al*., 2003). This recruitment contributes to gene silencing via enrollment of more histone modifiers and DNA methylases (Jaenisch & Bird, 2003; Kondo, 2009). While methylation occurs mostly on cytosine-guanidine dinucleotides (CpG), promoter CpGs tend to be protected from this modification (Deaton & Bird, 2011; Robertson, 2005). However, changes in the DNA

methylation pattern on promoters and repetitive elements can drive cellular pathogenesis and cancer progression (Kulis & Esteller, 2010). Especially, global hypomethylation, which is observed mostly on repetitive sequences, and promoter hypermethylation of tumor suppressor genes are among the epigenetic hallmarks of cancer (Robertson, 2005). This aberrant DNA methylation profile leads to genomic instability (Jones & Gonzalgo, 1997) and abnormal transcriptional regulator activity (Robertson & A.Jones, 2000) (Figure 1.2.).

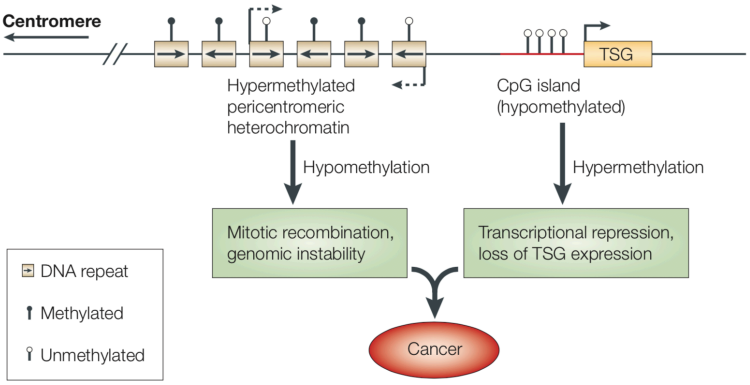


Figure 1.2. The relationship between DNA methylation and cancer. In normal cells, the repeat-rich pericentromeric heterochromatin region is hypermethylated, while the CpG islands of the actively-transcribed tumor suppressor genes (TSG) are hypomethylated. Heterochromatin hypomethylation and CpG island hypermethylation are observed in the early stages of tumorigenesis in cancer (from Robertson, 2005).

Other regulators that control gene expression are chromatin-associated factors including non-coding RNAs (ncRNAs) and chromatin-associated proteins (Gibcus & Dekker, 2012). ncRNAs are functional elements by themselves and are not translated into a peptide sequence (Patil *et al.*, 2014). Some ncRNA subtypes, including microRNAs (miRNAs), small interfering RNAs (siRNAs), piwi-interacting RNAs (piRNAs) and long non-coding RNAs (lncRNAs) are involved in gene regulation at both the transcriptional and post-transcriptional levels (Fernandes *et al.*, 2019; Patil *et al.*, 2014). Promoter-associated RNAs (PARs) and enhancer RNAs (eRNAs) constitute the most recent additions to this family of regulatory ncRNAs (Han *et al.*, 2007; Kim *et al.*, 2010).

Regulatory ncRNAs were shown to play roles in a wide array of epigenetic processes including histone modification (Böhmdorfer & Wierzbicki, 2015), DNA methylation (Aravin *et al*., 2007), heterochromatin formation (Johnson & Straight, 2017) and direct gene silencing (Kaikkonen *et al*., 2011; Zaratiegui *et al*., 2007). This functional diversity arises partially through a cross-talk between ncRNAs and other epigenetic factors and thus is highly context-dependent (Hanly *et al*., 2018). For instance, mir125b, which downregulates hematopoietic differentiation factors, has an oncogenic role in hematopoietic malignancies, whereas it has a tumor suppressive role in chronic lymphocytic leukemia (CLL) patients via BCL2 overexpression (Svoronos *et al*., 2016).

Aside from histones and DNA modifiers, there are two major categories of chromatin-binding protein complexes comprised of chromatin remodelers and transcription factors (TFs) (Lee & Young, 2013; Lucchesi *et al*., 2005; Wolffe, 2001). Chromatin remodelers modulate the chromatin landscape and the affinity of specific TFs to a given locus by displacing nucleosomes, sliding their position across the chromatin, and replacing them with other subtypes in a specific and ATP-dependent manner (Cairns, 2001; Manelyte, 2017). This specificity is conferred by their unique protein domains that recognize and bind to certain histone subtypes and modifications (Hargreaves & Crabtree, 2011; Manelyte, 2017). For instance, the chromatin remodeler SWI/SNF complex has bromodomains and can bind to the acetylated lysine residues on histone tails (Hargreaves & Crabtree, 2011). Similarly, chromatin remodeler CHD family enzymes can recognize and bind to the methylated histone tails (Manelyte, 2017). After chromatin remodelers recognize and read a modification, they can recruit additional modifiers and transcriptional machinery to a specific locus (Gaffney *et al*., 2012; Manelyte, 2017). Functional abnormalities in these enzymes cause chromatin dysregulation and altered gene expression profile (Wolffe, 2001; S. Zhang *et al*., 2017). For instance, the overexpression of Ino80, a SWI/SNF ATPase, was observed in melanoma cells and patients (Zhou *et al*., 2016). This promotes the expression of oncogenic factors via nucleosome replacement and Mediator complex recruitment to the regulatory sequences (Zhou *et al*., 2016).

Transcription factors (TFs) determine which specific gene will be turned “on” or “off” in a cell (Lee & Young, 2013). To do that, TFs bind to cis-acting regulatory DNA

sequences including promoters and/or enhancers and contribute to the recruitment of the transcription machinery (Todeschini *et al*., 2014). These cis-acting regulatory sequences are located at "open" chromatin regions, which are less occupied with nucleosomes compared to their target genes (Lai & Pugh, 2017). Promoters are located at the upstream of their targets and serve as the assembly sites for RNA polymerase complexes and other regulatory factors (Cevher *et al*., 2014; Chatterjee & Ahituv, 2017). Promoter activity is controlled by regions called enhancers that are distally located in nucleosome-depleted regions of the genome (Hao *et al*., 2019). DNA looping allows for the juxtaposition of enhancers and promoters by bringing these distant sites together, which is required for transcription initiation (Chatterjee & Ahituv, 2017; Wu & Hampsey, 1999). This loop is kept in place by the Mediator complex via the formation of a physical bridge that spans the gap between the promoter and the enhancer regions (Cevher *et al*., 2014) (Figure 1.3.). Not surprisingly, misregulation of the Mediator subunits or mutations on the DNA-binding motifs contribute to the malignant transformation of several cell types (Soutourina, 2018). For instance, upregulation of the MED23 subunit promotes cell migration and metastasis through Ras signaling in non-small-cell lung cancer patients (Schiano *et al*., 2014).

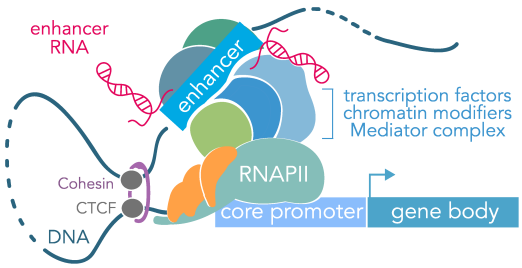


Figure 1.3. The physical bridge between promoter and enhancer created with the help of many regulatory factors including TFs, chromatin modifiers, Mediator complex and one of the ncRNA, enhancer RNAs (red) (from Carullo *et al*., 2019).

TF levels are tightly regulated in order to maintain proper functionality of the cell, and several mechanisms are in place to ensure correct dosage (Engelkamp & Heyningen, 1996; Veitia *et al*., 2013). For instance, eukaryotic TFs recognize shorter genomic sequences compared to prokaryotic ones, which can lead to a greater incidence of off-target binding (Wunderlich & Mirny, 2009). Thus, in eukaryotic organisms, TF overexpression may be

compensated by the buffering effect of these nonfunctional DNA-binding events, precluding any potential overactivation of protooncogenes (Veitia *et al*., 2013). However, in the case of target gene deletion, this stoichiometry is disrupted, leading to an overabundance of rogue transcription factors (Inoue & Fry, 2017). Similarly, chromosomal duplication events may lead to a two-fold increase in TF levels if the genes encoding these proteins happen to be located in the duplicated segments (Veitia *et al*., 2013). In any case, abnormally increased TF to target gene ratio frequently eventuates pathogenesis (Todeschini *et al*., 2014; Veitia *et al*., 2013). The converse case where the TF levels are downregulated may also prove disruptive to a cell's function, especially when these TFs have endogenous tumor suppressor roles (Inoue & Fry, 2017). The tumor suppressor p53 transcription factor gene is silenced during tumorigenesis through deletion of the both alleles, promoter hypermethylation, and/or point mutation (Inoue & Fry, 2017; Shen *et al*., 2018). TF haploinsufficiency, where a single copy of the allele is unable to produce adequate amounts of a protein to achieve the appropriate phenotype, is present in many other diseases including Waardenburg syndrome type II and Li-Fraumeni syndrome (Inoue & Fry, 2017; Seidman & Seidman, 2002). Given these examples, the complexity of these regulatory proteins and their distinct alterations in disease states highlight the importance of their identification for the development of clinical intervention strategies.

1.2. Identification of Chromatin-Associated Proteins

Biological and biochemical techniques are required to address the unknowns concerning *in vivo* DNA-protein interactions and their function. There are several existing methodologies derived from chromatin immunoprecipitation coupled with mass spectrometry (ChIP-MS) to analyze these interactions.

Chromatin immunoprecipitation (ChIP) is an invaluable tool that allows the detection of DNA-protein interactions (Milne *et al*., 2009). ChIP is utilized in the detection of genome-wide DNA-protein interactions and the corresponding amount of a chosen DNA-binding proteins (Zhang *et al*., 2004). In principle, the chromatin is crosslinked with chromatin-associated factors by formaldehyde (Hoffman *et al*., 2015), which makes it possible to detect even temporary interactions (Turner, 2001), and fragmented into small

pieces. A specific antibody is used for the selective isolation of the desired chromatin-associated factor along with its bound DNA fragment and protein interactors (He & Pu, 2010). This protein:DNA complex can then be separated into its components for further analyses (Alekseyenko *et al*., 2015; He & Pu, 2010; Hubner *et al*., 2015). The isolated DNA fragments can be analyzed via several downstream methodologies including quantitative PCR (qPCR) using locus specific primer pairs (Turner, 2001) and/or identified via next generation sequencing (ChIP-seq) (Barski *et al*., 2007). In parallel, the crosslinked protein interactor pool can be probed via Western blotting (Liu *et al*., 2018) and characterized by mass-spectrometry (MS) (Wierer & Mann, 2016).

ChIP methods rely on the action of a specific antibody that binds its target and makes it purification possible (Hoshino & Fujii, 2009). However, ChIP-based chromatin isolation methods require specific antibodies generated for each DNA-binding protein and is limited by the antibody's targeting efficiency (Soleimani *et al*., 2013). Yet, most of the TFs do not have a commercially available antibody, and the existing ones are frequently unable to distinguish between multiple isoforms (Gade & Kalvakolanu, 2012).

One of the initial approaches to overcome this issue resorted to affinity purification subsequent to protein-tagging (Kolodziej *et al*., 2009). As general rules, the protein tags that are used to identify DNA-protein interactions should have strong binding affinity, be small and soluble, and be unaffected by formaldehyde fixation (Kolodziej *et al*., 2009). Biotin tags fulfill all of these criteria even under stringent conditions such as high temperature, high salt concentrations, and non-physiological pH (He & Pu, 2010). In the first locus-specific study based on the PICh method, biotin-streptavidin interaction was utilized to purify biotin-tagged DNA probes targeting the repetitive telomeric regions (Déjardin & Kingston, 2009). Subsequent MS analyses on the isolated telomeric regions pinpointed telomere-specific proteins such as HOT1 (Kappei *et al*., 2013). However, this system targets the copiously-situated telomere repeats and is unable to identify chromatin regulators on a single-locus level (Kappei *et al*., 2013).

In parallel to this PICh method, researchers established the insertional chromatin immunoprecipitation assay (iChIP) in order to identify locus-specific regulators (Hoshino &

Fujii, 2009). For this purpose, they exogenously expressed the FLAG-tagged bacterial LexA protein in a eukaryotic cell line that carried the knock-in of the LexA binding element at the genomic region of interest (Fujita & Fujii, 2011; Hoshino & Fujii, 2009). While the protocol turned out to be a success, inserting an exogenous sequence into the genome is not practical and might lead to undetected disruptions in the chromatin structure and the local proteome (Wierer & Mann, 2016).

Chromatin-associated regulators can be isolated using another method called "ChIP coupled with the selective isolation of chromatin-associated proteins" (ChIP-SICAP) (Rafiee *et al*., 2016). ChIP-SICAP involves sequential purification of the target TF via immunoprecipitation and its bound DNA fragments via biotin-streptavidin interaction (Rafiee *et al*., 2016). As the small DNA fragments are purified along with the attached chromatin-associated proteins, there is no requirement for direct physical interaction between protein partners for identification (Rafiee *et al*., 2016). Contrarily, this method is not locus-specific and requires ChIP-grade antibody purification, which is a limitation as mentioned above and increases material loss during the purification steps (Rafiee *et al*., 2016).

The advent of engineered DNA-binding proteins enables us to label and purify a specific genomic locus without the necessity of inserting exogenous DNA sequences into the genome. The first method that utilized engineered proteins to purify a specific genomic locus was enChIP. For this, researchers reengineered the peptide sequence of a FLAG-tagged Transcription-activator like (TAL) protein to recognize and bind to telomeric repeats, enabling their purification (Fujita *et al*., 2013). However, as the target specificity is built into the amino acid structure of the TAL protein, a new variant needs to be engineered each time a novel sequence is targeted, rendering the process cumbersome (Scholze & Boch, 2010). With the advancements in clustered regularly interspaced short palindromic repeats (CRISPR) technology, enChIP was coupled with a nuclease-deficient version of Cas9 (dCas9) protein used as a cargo delivery system (Fujita & Fujii, 2013). The CRISPR system is advantageous compared to the TAL system, because the central CRISPR enzyme Cas9 is able to target genomic sequences under the guidance of an RNA molecule (sgRNA) that can be easily engineered and synthesized for each new target (Fujita & Fujii, 2013). However,

widespread off-target binding events of the Cas9 protein can result in false positive identifications and lowers the system's specificity (Zhang *et al*., 2015).

In the meantime, BioID method has been developed to increase the selective purification and the identification of possible interactors of a target protein (Roux *et al*., 2012). BioID utilizes a promiscuous BirA* biotin ligase for the *in vivo* biotinylation of proteins in close proximity (<10 nm) on their lysine residues without necessitating the presence of a biotin-acceptor-tag. The biotinylated target can then be pulled down in a specific manner using streptavidin as with the previously mentioned methods. The major advantage of this method is the ability to detect relatively weak and temporary interactions in their native states thanks to its dependence on proximity as the sole criterion (Roux et al., 2012). A more recent method named "CasID" combines the BioID system with the cargo delivery feature of dCas9 to target specific genomic loci with greater pulldown efficiency to identify the neighboring protein partners (Schmidtmann et al., 2016) (Figure 1.4.). Protein purification from repeat regions has been successful, yet the method suffers from multiple drawbacks including the likely inability to target a single-locus, off-target dCas9 binding, and nonspecific biotinylation caused by constitutive BirA* activity (Roux et al., 2018, 2012; Schmidtmann et al., 2016).

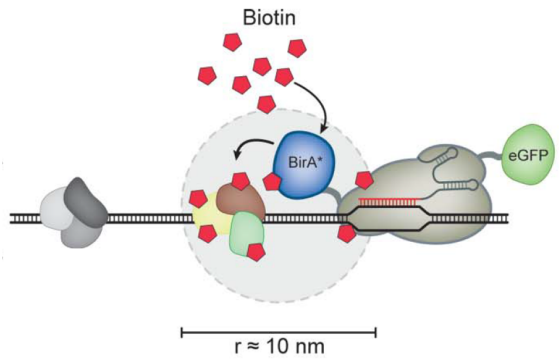


Figure 1.4. Diagram showing CasID method where promiscuous BirA* fused to dCas9 fused is targeted to a specific locus to identify transcription regulatory proteins by biotin-tagging (Adapted from Schmidtmann *et al*., 2016).

Two different groups came up with the Split-BioID method to overcome this nonspecific labeling problem and thus fine-tune the activity of promiscuous BirA* (De Munter *et al*., 2017; Schopp *et al*., 2017). This method employs a BirA* version that was generated by splitting the protein into two inactive halves from either E140/Q141(De Munter et al., 2017) or E256/G257 (Schopp et al., 2017).Each half is linked to a pair of known interactor proteins whose partners are to be identified. Only when these interactors come within close proximity the BirA* fractions come together and get reconstituted and enzymatic activity activated; therefore, significantly limiting background biotinylation (Figure 1.5.). Moreover, the system can also be temporally controlled, as BirA* activity relies on the external provision of biotin (De Munter *et al*., 2017; Schopp *et al*., 2017). In this way, the known interactor pool of the two partners can be extended to include heretofore unidentified proteins following streptavidin pulldown and MS analysis.

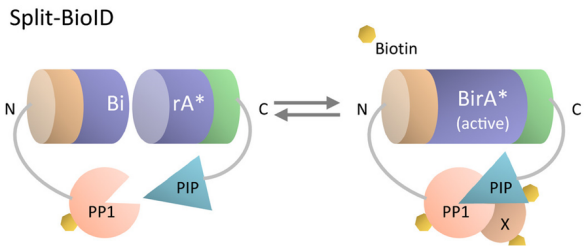


Figure 1.5. The diagram of Split-BioID method. Two inactive halves of promiscuous BirA* protein are fused to proteins, PP1 and PIP. With physical interaction of fused partners, inactive BirA* halves heterodimerizes and gets activated. In the presence of biotin, interaction partners of PP1 and PIP get biotinylated (from De Munter *et al*., 2017)

Recently, the CAPTURE method was established to isolate long-range chromatin interactions and the adjacent regulatory proteins via biotinylated dCas9 (Liu *et al*., 2017). This method employs coexpression of dCas9 that carries a biotin-acceptor-tag and BirA biotin ligase which can ordinarily only recognize and biotinylate proteins that have a biotin-acceptor-tag in contrast to the promiscuous version (Liu *et al*., 2017). The method was validated on telomeric regions and tested on cis-regulatory elements (CREs) to identify interactions between promoters, enhancers and their regulators (Liu *et al*., 2017). However,

this system comes with the potential drawback of dCas9 off-target binding (Liu *et al*., 2018, 2017).

The new tools developed for the identification of chromatin-associated regulators provide us with extremely robust and novel information when applied properly. While existing methodologies have been developed to understand chromatin-templated events, many of them have significant drawbacks. Therefore, it is necessary to develop new methods without these limitations. For this purpose, we should consider each preceding method with its strengths and weaknesses. In this study, our aim is to identify locus-specific transcriptional regulators by establishing a new, flexible method. We will combine the targeting power of CRISPR with *in vivo* biotinylation, while alleviating the off-target effects by simultaneously employing two distinct CRISPR proteins that have different target recognition sequences: Cas9 and Cpf1(Cas12a). We also propose to mitigate nonspecific biotinylation that was observed in the earlier methodologies by implementing a spatiotemporally controlled Split-BioID system.

2. AIM OF STUDY

In this study, our ultimate purpose is to develop a method called Split-CRISPR-ID that will enable the identification of locus-specific transcriptional regulators, and therefore pave the way to elucidate their roles on critical genomic target regions. To do that, we aim to induce the biotinylation of locus-specific proteins using a promiscuous mutant biotin ligase BirA* which is split into two halves and carried by nuclease-deficient CRISPR proteins. We also aim to optimize a recently published method (CAPTURE) in our laboratory, and to be able to ultimately apply it to our regions of interest.

In the first part of this study, for the Split-CRISPR-ID method, our aim is to generate two sets of protein fusion pairs that are the combination of nuclease-deficient CRISPR proteins, dCas9 and dCpf1, and a split version of the mutant BirA* biotin ligase enzyme. By doing so, we aim to spatiotemporally control biotin-tagging activity specifically at target regions. Then we aim to validate the sgRNA targeting efficiencies using ChIP-qPCR on four different target genomic regions. Our next step is to validate the biotinylation activity on these target regions by streptavidin ChIP followed by qPCR. The ultimate aim of this method is to isolate and identify locus-specific critical transcriptional regulators via mass spectrometry in the future.

In the second part of the project, we aim to establish stable cells expressing biotin-acceptor-tag carrying dCas9, BirA and engineered sgRNA, the components of the CAPTURE method. We will do this by validating the dCas9 targeting by streptavidin-bead pulldown followed by CAPTURE-qPCR and CAPTURE-Western blotting.

3. MATERIALS

3.1. General Kits, Enzymes and Reagents

Table 3.1. List of general kits, enzymes and reagents.

BCA Protein Assay Kit	Life Technologies, USA
RPMI MEDIUM 1640	Thermo Scientific, USA
DMEM	Gibco, LifeTechnologies, USA
FBS	Gibco, LifeTechnologies, USA
Non-Essential Amino Acids	Gibco, LifeTechnologies, USA
Penicillin / Streptomycin	Gibco, LifeTechnologies, USA
Trypsin	Gibco, LifeTechnologies, USA
Plasmid Miniprep Plus Purification Kit	GMBio Lab, Taiwan
ZymoPURE Plasmid Miniprep Kit	Zymo Research, USA
ZymoPURE II Plasmid MidiPrep Kit	Zymo Research, USA
PCR Purification and Gel Extraction Kit	Macherey Nagel, Germany
Q5 High Fidelity DNA Polymerase	New England Biolabs, USA
Deoxynucleotide (dNTP) Solution Mix	New England Biolabs, USA
T4 PNK	New England Biolabs, USA
T4 DNA Ligase	New England Biolabs, USA
Taq DNA Ligase	New England Biolabs, USA
T5 Exonuclease	New England Biolabs, USA
Phusion High Fidelity DNA Polymerase	New England Biolabs, USA
My*Taq* DNA Polymerase	Bioline, UK
Protease Inhibitor Cocktail Tablets	Roche, Switzerland
BsmBI	Thermo Scientific, USA
FastDigest Esp3I	Thermo Scientific, USA
FastAP	Thermo Scientific, USA
NheI-High Fidelity	New England Biolabs, USA

Table 3.1. List of general kits, enzymes and reagents (cont.)

BmtI	New England Biolabs, USA
BamHI-High Fidelity	New England Biolabs, USA
XhoI	New England Biolabs, USA
PstI	New England Biolabs, USA
EcoRV	New England Biolabs, USA
NcoI-HF	New England Biolabs, USA
DpnI	New England Biolabs, USA
BstXI	New England Biolabs, USA

3.2. Biological Materials

3.2.1. Bacterial Strains

For storage, expansion, and transformation of all plasmids, *Escherichia Coli* STBL3 bacterial strain was used since it has low recombinase activity and prevents recombination of lentiviral vectors. For Home-made Gibson Assembly reaction, *Escherichia Coli* DH5a bacterial strain was used since Gibson Assembly buffer is incompatible with STBL3 bacterial strain.

3.2.2. Cell Lines

K-562 (human erythroleukemic cell line, kindly provided by Dr. Tolga Sutlu), MCF-7 (human breast cancer cell line, kindly provided by Dr. Devrim Gozuacik), MDA-MB-231 (human metastatic breast cancer cell line, kindly provided by Dr. Devrim Gozuacik), HEK293FT (human embryonic kidney cell line, kindly provided by Dr. Nesrin Ozoren) were mainly used in the experiments.

3.2.3. Plasmids

pCW-Cas9 (Addgene plasmid # 50661, a gift from Eric Lander & David Sabatini), pCW-dCas9-puro, pCW-Cas9-blast (Addgene plasmid # 83481), WN10151 (Addgene plasmid # 80441), pCw-dCpf1-blast, pCW-BirN-dCas9-puro, pCW-BirC-dCas9-puro, pCW-BirN-dCpf1-blast, pCW-BirC-dCpf1-blast, pEF1a-FB-dCas9-puro (Addgene plasmid # 100547), pEF1a-BirA-V5-neo (Addgene plasmid # 100548), pSLQ1651-sgRNA(F+E)-sgGal4 (Addgene plasmid # 100549), pSLQ1651-sgTelomere were used for the generation of stable cell lines. pLKO5.sgRNA.EFS.GFP backbone (Addgene plasmid #57822, a gift from Benjamin Ebert) and lenti_gRNA-puro (Addgene plasmid # 84752), were used for the expression of all sgRNAs used in this study.

3.3. Primers

Table 3.2. Primers used in this study.

Primer ID	Sequence (5'to 3')	Application
BirC-Kozak-F	GTGAACCGTCAGATCGCCTGGAGAATT GGCTAGCCACCATGATGGGAGGACTGG CTCCTTACCTG	PCR
BirC-linker-R	CGAATTCGCCAGAACCAGCAGCGGAGC CAGCGGATCCAGCCTTCTCTGCGCTTCT CAGG	PCR
BirC-dCas9-R	CGTGGTCCTTATAGTCCATGGTGGCGCT AGCGAATTCGCCAGAACCAGCAG	PCR
BirC-dCpf1-R	ATGGTGGCGGTACCAAGCTTAAGTTTAA AGCTAGCGAATTCGCCAGAACCAGC	PCR

Table 3.2. Primers used in this study (cont.).

BirN-Kozak-F	AGTGAACCGTCAGATCGCCTGGAGAATTGGCTAGCCACCATGATGAAGGACAACACCGTGCCC	PCR
BirN-linker-R	GAATTCGCCAGAACCAGCAGCGGAGCCAGCGGATCCAGCCTCCTGCTCGAACAGCTCCA	PCR
BirN-dCas9-R	CGTGGTCCTTATAGTCCATGGTGGCGCTAGCGAATTCGCCAGAACCAGCAGCGGA	PCR
BirN-dCpf1-R	ATGGTGGCGGTACCAAGCTTAAGTTTAAAGCTAGCGAATTCGCCAGAACCAGCAG	PCR
dCpf1-pS2-sgRNA-1F	CACCGAGTGATCCGCCTGCTTTGGC	sgRNA
dCpf1-pS2-sgRNA-1R	GTAGGCCAAAGCAGGCGGATCACTC	sgRNA
dCpf1-pS2-sgRNA-2F	CACCGCAGGCCTACAATTTCATTAT	sgRNA
dCpf1-pS2-sgRNA-2R	GTAGATAATGAAATTGTAGGCCTGC	sgRNA
dCpf1-pS2-sgRNA-3F	CACCGCGGCCATCTCTCACTATGAA	sgRNA
dCpf1-pS2-sgRNA-3R	GTAGTTCATAGTGAGAGATGGCCGC	sgRNA
dCpf1-pS2-sgRNA-4F	CACCGCCCTGGCGGGAGGGCCTCTC	sgRNA
dCpf1-pS2-sgRNA-4R	GTAGGAGAGGCCCTCCCGCCAGGGC	sgRNA
dCpf1-Mapk-sgRNA-1F	CACCGTCTGTCTCCTCTCTCTAAAA	sgRNA
dCpf1-Mapk-sgRNA-1R	GTAGTTTTAGAGAGAGGAGACAGAC	sgRNA
dCas9-Mapk-sgRNA-1F	CACCGGCATGCGGATTGGTTCCTCG	sgRNA
dCas9-Mapk-sgRNA-1R	AAACCGAGGAACCAATCCGCATGCC	sgRNA
dCas9-Mapk-sgRNA-2F	CACCGGACTAAAAGGAGACACTGAG	sgRNA
dCas9-Mapk-sgRNA-2R	AAACCTCAGTGTCTCCTTTTAGTCC	sgRNA
dCas9-Mapk-sgRNA-3F	CACCGACACCCAGGGAGACAGACGT	sgRNA
dCas9-Mapk-sgRNA-3R	AAACACGTCTGTCTCCCTGGGTGTC	sgRNA

Table 3.2. Primers used in this study (cont.).

dCas9-Mapk-sgRNA-4F	CACCGAGGCAGGCAATCGGTCCGAG	sgRNA
dCas9-Mapk-sgRNA-4R	AAACCTCGGACCGATTGCCTGCCTC	sgRNA
dCas9-pS2-sgRNA-1F	CACCGTCCAGAGTAGCTAGGATTAC	sgRNA
dCas9-pS2-sgRNA-1R	AAACGTAATCCTAGCTACTCTGGAC	sgRNA
dCas9-pS2-sgRNA-2F	CACCGCAGGCGGATCACTTAAAGTC	sgRNA
dCas9-pS2-sgRNA-2R	AAACGACTTTAAGTGATCCGCCTGC	sgRNA
dCas9-pS2-sgRNA-3F	CACCGTAGGACCTGGATTAAGGTC	sgRNA
dCas9-pS2-sgRNA-3R	AAACGACCTTAATCCAGGTCCTAC	sgRNA
dCas9-pS2-sgRNA-4F	CACCGTGATAGACAGAGACGACATG	sgRNA
dCas9-pS2-sgRNA-4R	AAACCATGTCGTCTCTGTCTATCAC	sgRNA
dCas9-pS2-sgRNA-5F	CACCGAAATTGTAGGCCTGGCGCAG	sgRNA
dCas9-pS2-sgRNA-5R	AAACCTGCGCCAGGCCTACAATTTC	sgRNA
dCas9-pS2-sgRNA-6F	CACCGCAGTATTTACCCTGGCGGGA	sgRNA
dCas9-pS2-sgRNA-6R	AAACTCCCGCCAGGGTAAATACTGC	sgRNA
U6-sequencing-primer	GATACAAGGCTGTTAGAGAGATAATT	Sequencing
LNCX	AGCTCGTTTAGTGAACCGTCAGATC	Sequencing
sgTelomere-F	GGAGAACCACCTTGTTGGTTAGGGTT AGGGTTAGGGTTAGTTTAAGAGCTAT GCTGGAAACAGCA	PCR
sgTelomere-R	CTAGTACTCGAGAAAAAAAGCACCGA CTCGGTGCCAC	PCR
U6-PCR-fwd	GAGGGCCTATTTCCCATGATTCC	PCR
ColonyPCR-dCpf1-F	CTCTGAGCTTCTACGGCGAG	PCR
ColonyPCR-dCpf1-R	TCACCACGGCCTTCTTCTTC	PCR
ColonyPCR-dCas9-F	AACCGCCCTGATCAAAAAGT	PCR
ColonyPCR-dCas9-R	TCTTGGACTTGCCCTTTTCC	PCR
Mapk-target-1F	GCCCATGAGTTCTGCCCC	ChIP-qPCR
Mapk-target-1R	CAAGAAGTGTGCATGCGGAT	ChIP-qPCR
Mapk-target-2F	GGTAACTTCTATAAGGGTGGATCC	ChIP-qPCR

Table 3.2. Primers used in this study (cont.).

Mapk-target-2R	GCACTGGGGCAGAAACCG	ChIP-qPCR
Mapk-target-3F	AAGTTTATCAGCGTGTGGACCA	ChIP-qPCR
Mapk-target-3R	CGGGTCGCGTCTCTTAAGA	ChIP-qPCR
Mapk-target-4F	GAACCAATCCGCATGCACAC	ChIP-qPCR
Mapk-target-4R	GAGGAGGAAGGAAGACGCC	ChIP-qPCR
pS2-target-1F	TGGCCAGGCTAGTCTCAAAC	ChIP-qPCR
pS2-target-1R	GGGAAGGGAAGGAGCTCATG	ChIP-qPCR
pS2-target-2F	GCGCCAGGCCTACAATTTC	ChIP-qPCR
pS2-target-2R	TCATAGTGAGAGATGGCCGGA	ChIP-qPCR
pS2-target-3F	CCGTGAGCCACTGTTGTCAC	ChIP-qPCR
pS2-target-3R	AAGTCCCTCTTTCCCATGGG	ChIP-qPCR
ChIP-Telomere-F	GGTTTTTGAGGGTGAGGGTGAGGGTG AGGGTGAGGGT	ChIP-qPCR
ChIP-Telomere-R	TCCCGACTATCCCTATCCCTATCCCTA TCCCTATCCCTA	ChIP-qPCR
ChIP-36B4-F	CAGCAAGTGGGAAGGTGTAATCC	ChIP-qPCR
ChIP-36B4-R	CCCATTCTATCATCAACGGGTACAA	ChIP-qPCR
RPS12-F	TGTTTGGCTTCTGTCGCGCT	ChIP-qPCR
RPS12-R	TCCTTCTTTGCGGAGGCGTG	ChIP-qPCR
HIRA-target-1F	TAGGGTGTCAGCTGCGCA	ChIP-qPCR
HIRA-target-1R	TAGAAGATCGGGGCCGTTTG	ChIP-qPCR

3.4. Chemicals

Table 3.3. List of chemicals used in this study.

Ethidium Bromide	Merck, USA
EDTA	Merck, USA
Hydrochloric Acid	Merck, USA
Sodium Chloride	Merck, USA
Bromophenol Blue Indicator	Merck, USA
Tween 20	Merck, USA
Methanol	Merck, USA
Acetic Acid	Merck, USA
LB Broth	Merck, USA SigmaAldrich, USA Caisson Laboratories, USA
Glycine	Merck, USA
DMSO	Merck, USA
Ethanol	Merck, USA
Nonidet P40	Fluka, USA
Hexadimethrine bromide	Sigma Aldrich, USA
Chloroquine diphosphate	Applichem, Germany
Ammonium persulfate	Applichem, Germany
SDS	Applichem, Germany
Ampicillin	Applichem, Germany
Agar bacteriology grade	AppliChem, Germany
Acrylamide	Sigma Aldrich, USA
Sodium Deoxycholate	Sigma Aldrich, USA
N,N'-Methylenbisacrylamide	Sigma Aldrich, USA
DEPC-treated water	Fisher Scientific, USA
Agarose	Sigma Aldrich, USA

Table 3.3. List of chemicals used in this study (cont.).

Glycerol	Merck, USA
Chloroquine	Merck, USA
TEMED	Merck, USA
Tris-Cl	Merck, USA
Calcium Chloride	Merck, USA

3.5. Buffers and Solutions

Table 3.4. List of buffers and solutions used in this study.

Solutions	Ingredients
10x SDS Running Buffer	1% SDS 1.92M Glycine 250 mM Tris-Base
10x Transfer Buffer	1.92 M Glycine 250 mM Tris-Base
1x Transfer Buffer	10% 10X Transfer Buffer 20% Methanol 70% ddH_2O
4x Protein Loading Dye	200 mM Tris-Cl (pH: 6.8) 8% SDS 40% Glycerol 4% β-mercaptoethanol 50 mM EDTA 0.8% Bromophenol Blue

Table 3.4. List of buffers and solutions used in this study (cont.).

Radioimmunoprecipitation (RIPA) Buffer (500 ml)	50 mM TRIS pH 8.0 1% NP40 0.5% NaDOC 0.1% SDS 150 mM NaCl 2 mM EDTA
4x Laemmli Buffer	200 mM Tris-HCl (pH 6.8) 8% SDS 40% glycerol 4% 2-mercaptoethanol 50 mM EDTA 0.08% bromophenol blue in ddH_2O
7% SDS-PAGE Gel, Resolving (8ml)	4.4 ml ddH_2O 1.4 ml 40% Acrylamide:bisacrylamide 2ml 1.5M Tris, pH: 8.8 80 µl 10% SDS 80 µl 10% APS 8 µl TEMED
6% SDS-PAGE Gel, Stacking (5ml)	2.9 ml ddH_2O 0.75 ml 40% Acrylamide:bisacrylamide 1.25 ml 0.5 M Tris, pH: 6.8 50 µl 10% SDS 50 µl 10% APS 5 µl TEMED
10x TBS	200 mM Tris-Cl (pH: 7.6) 1.5 M NaCl
1x TBS-T	50 mM Tris-Base (pH: 7.4) 150 mM NaCl 0.1% Tween 20
Paraformaldehyde (4%)	1 g Paraformaldehyde 25 ml distilled water
5% Blocking Solution	5% BSA (w/v) in 1X TBS-T

Table 3.4. List of buffers and solutions used in this study (cont.).

5x Isothermal Reaction Mix (300 µl)	75 µl 2M Tris-HCl (pH 7.5) 7.5 µl 2M $MgCl_2$ 120 µl 10 mM dNTP 15 µl 1M DTT 0.075 g PEG-8000 30 µl 50 mM NAD^+ 52.2 µl nuclease-free water
Assembly Master Mix (60 µl)	16 µl 5x Isothermal Reaction Mix 1 µl T5 exonuclease (0.32 U/µl) 1 µl Phusion HF DNA Polymerase (2 U/µl) 8 µl *Taq* DNA Ligase (40 U/µl) 34 µl nuclease-free water
High Salt Wash Buffer	50 mM N-(2-hydroxyethyl)piperazine-N-2-ethanesulfonic acid (HEPES; pH 7.5) 1 mM EDTA (pH 8.0) 0.5 M NaCl 1% Triton X-100 0.1% sodium deoxycholate (NaDOC) Store at 4˚C for up to 6 months
LiCl Wash Buffer	10 mM Tris·Cl (pH 8.1) 1 mM EDTA (pH 8.0) 250 mM LiCl 0.5% NP-40 0.5% sodium deoxycholate (NaDOC) Store at 4˚C for up to 6 months
TE Buffer	10 mM Tris·Cl (pH 7.5) 1 mM EDTA (pH 8.0) Store at room temperature for up to 6 months
Cellular Lysis Buffer	5 mM PIPES pH8 85 mM KCl 0.5% NP40
Nuclear Lysis Buffer	50 mM Tris pH8 10 mM EDTA pH8 0.2% or 1% SDS

Table 3.4. List of buffers and solutions used in this study (cont.).

IP Dilution Buffer	16.7 mM Tris pH 8 1.2 mM EDTA pH8 167 mM NaCl 0.01% SDS 1.1% Triton X-100

3.6. Antibodies

Table 3.5. List of antibodies used in this study.

Antibody	Source	Dilution	Supplier
Anti-HA	Rabbit	1:1000	3724S, Cell Signaling Technology
Anti-FLAG M2	Mouse	1:2000	F1804-200UG, Sigma Aldrich
Anti-beta-tubulin	Mouse	1:2000	D3U1W, Cell Signaling Technology
Anti-V5	Mouse	1:1000	Sc-81594, Santa Cruz
Anti-Cas9	Mouse	1:2000	7A9-3A3, Cell Signaling Technology
Anti-TRF2	Mouse	1:500	Sc-47693, Santa Cruz
Anti-Ref-1 (APEX)	Mouse	1:500	Sc-17774, Santa Cruz
Anti-Vinculin	Sheep	1:2000	Af6896, R&D systems
Anti-GATA-1	Mouse	1:500	Sc-265, Santa Cruz
Anti-SP1	Rabbit	1:2000	Sc-59, Santa Cruz
Anti-rat IgG, HRP-linked	Goat	1:1000	7077S, Cell Signaling Technology
Anti-sheep IgG, HRP-linked	Donkey	1:5000	A3415, Sigma Aldrich
Anti-rabbit IgG, HRP-linked	Goat	1:2000	7074S, Cell Signaling Technology
Anti-mouse IgG, HRP-linked	Horse	1:2000	7076S, Cell Signaling Technology

3.7. Disposable Labware

Table 3.6. List of disposable labware used in this study.

Western Blotting Paper	Whatman, UK
Parafilm	Brand, Germany
Scalpel	Swann-Morton, UK
Centrifuge Tubes, 15 ml	Vwr, USA
Centrifuge Tubes, 50 ml	Vwr, USA
Serological pipette, 5 ml	Tpp, Switzerland
Serological pipette, 10 ml	Tpp, Switzerland
Serological pipette, 25 ml	Tpp, Switzerland
Pipette Tips, filtered	Biopointe, USA
Pipette Tips, bulk	Biopointe, USA
Microcentrifuge tubes	Axygen, USA
PCR Tubes, 0.2 ml	Axygen, USA
Medical Gloves	Vwr, USA
Syringe Filters	Sartorius, Germany
Cryovial	Tpp, Switzerland
Cell Culture Flasks, 75 cm^2	Tpp, Switzerland
Cell Culture Plates, 6-well	Tpp, Switzerland
Insulin Syringes, 1 ml	Vwr, USA
Syringe, 10 ml	Becton Dickinson, USA
PS Test Tubes, 5 ml	Becton Dickinson, USA Isolab, Germany
Glass Pasteur Pipette, 230 mm	Witeg, Germany
Chemiluminescent Detection Film	Roche, Switzerland

3.8. Equipment

Table 3.7. List of equipment used in this study.

Micropipettes	Axygen, USA
Heat Block	Cleaver Scientific EL-01, UK
Pipettor	Greiner Labopet 240, Germany
Vortex	Vwr, USA
Vertical Electrophoresis	Cleaver Scientific Omni-page Mini, UK
Centrifuges	J2-21, Beckman Coulter, USA Allegra X-22, Beckman Coulter, USA 5415R, Eppendorf, USA
Rotor	Beckman Coulter JA-14, USA
Cell Culture Incubator	Binder C-150, Germany
Documentation System	GelDoc XR System, Bio-Doc, Italy
Flow Cytometer	FACSCalibur, Becton Dickinson, USA
Freezers	-20°C, Ugur UFR 370 SD, Turkey -80°C, Thermo Scientific TS368, USA -150°C, Sanyo MDF1156, Japan
Microplate Reader	680, Biorad, USA
Microscopes	Inverted Microscope, Nikon Eclipse TS100, Japan Fluoroscence Microscope, Observer Z1, Zeiss, Germany
Thermal Cycler	Antarus MyCube ANT101, USA
Power Supply	Vwr, USA
Shaker	VIB Orbital Shaker, InterMed, Denmark Vwr Mini Orbital Shaker, USA
Spectrophotometer	NanoDrop 1000, USA
Stirrer - Heater	Dragonlab MS-H-S, China
Rotator	Grant Bio Multifunctional Rotator PTR-35, UK

Table 3.7. List of equipment used in this study (cont.).

Refrigerator	4°C, Ugur USS 374 DTKY, Turkey
Refrigerated Vapor Trap	Thermo Scientific SPD111V, USA
Oil-free Gel Pump	Thermo Scientific Savant VLP110, USA
Vacuum Pump Oil Filter	Thermo Scientific VPOF110, USA
Carbon dioxide Tank	Genc Karbon, Turkey
Ice Maker	Brema, Italy Scotsman Inc. AF20, Italy
Autoclave	Model ASB260T, Astell, UK
Dishwasher	Miele Mielabor G7783, Germany
Stella	Raytest, Germany
Freezing Container	Nalgene, USA
Oven	Nüve KD200, Turkey
GenePulser XCell Electroporation Systems	BioRad, USA

4. METHODS

4.1. Cell Culture

4.1.1. Maintenance of Cell Lines

MCF-7, MDA-MB-231, HEK293FT cell lines were grown in DMEM-high glucose media supplemented with 10% FBS, 1% penicillin/streptomycin antibiotics and 1% non-essential amino acid solution. K562 cell line was grown in RPMI media 10% FBS, 1% penicillin/streptomycin antibiotics and 1% non-essential amino acid solution. Cells were maintained in a 37°C incubator with 5% CO_2. Cells were passaged every two days for the maintenance.

For adherent cell passage, old cell medium was aspirated and cell pellet was washed with 1X PBS. After PBS was aspirated, 0.05% trypsin-EDTA was added onto the cells, and they were incubated at 37°C for 3 minutes. After incubation, twice amount of DMEM was added to the cells to inactivate trypsin. Then, the cell suspension was transferred to 15 ml falcon tubes and centrifuged at 500 x g for 2 minutes. Medium was discarded, and the cell pellet was resuspended in fresh DMEM supplemented with 10% FBS. Cells were mixed well to prevent clump formation and seeded on plates or flasks.

To passage suspension cells, medium-cell suspension was transferred to a 50 ml falcon tube and centrifuged at 300 x g for 5 minutes. Old media was aspirated, and the cell pellet was resuspended in warm RPMI 1640 medium. Total cell number was counted in every passage, and the culture was split to about 0.4 X 10^6 per ml fresh RPMI 1640 medium.

For long-term storage of the cells, they were centrifuged at 500 x g for 2 minutes and the old medium was discarded. The cell pellet was resuspended in the cell freezing medium, which is DMEM medium or RPMI 1640 medium supplemented with 10% DMSO and 20%

FBS. This cell suspension was aliquoted into 1 ml cryovials as 2 x 10^6 cell per ml and put into Nalgene Freezing Container that is filled with isopropanol and stored at -80°C for 2-3 days. Then, cryovials were transferred into -150°C freezer for long-term storage.

To open new cells from -150°C stock, cryovials were quickly transferred to the 37°C water bath. After cell suspension was defrost completely, it was transferred to 15 ml falcon tubes with warm 9 ml growth media and centrifuged at 500 x g for 2 minutes. Freezing medium was discarded and the cell pellet was resuspended in regular DMEM medium or RPMI 1640 medium. Cell suspensions were transferred to the T75 culture flasks and the additional medium was added to the cells for the maintenance.

4.1.2. Transfection of HEK293FT Cells via Calcium Phosphate Method

HEK293FT cells were used in the lentivirus production. For transfection, cells were seeded into 10 cm plates at 80% confluency 1 day prior to the transfection. Next day, the cell medium was aspirated and replaced with 9 ml fresh medium supplemented with chloroquine at a final concentration of 25 µM. Cells were incubated at 37°C, while preparing the 1 ml transfection mixture. In microcentrifuge tubes, autoclaved double distilled water was mixed with 10 µg of the corresponding plasmid, 7.5 µg psPAX2 and 4 µg p-VSV-G helper plasmids. 62.5 µl of 2 M $CaCl_2$ (125 mM) was added to the solution drop by drop and mixture was incubated for 5 minutes. 1X Hepes Buffered Saline (HBS) solution was added to the mixture very slowly and dropwise manner. This solution was mixed very well until bubbles were observed, and incubated at the room temperature for 10 minutes. After incubation, the DNA mixture was added onto the cells dropwise and cells were incubated at 37°C. After 6 hours, the medium of the cells was aspirated, and 11 ml medium was added onto the cells.

4.1.3. Lentivirus Harvest from Cell Media

Lentivirus containing cell media of HEK293FT cells were harvested 48-72 hours later from the calcium phosphate transfection. Transfection efficiency of the plasmids that

contain fluorescent marker such as pLKO-5-EFS-sgRNA-GFP was observed by fluorescent microscopy. Lentiviruses produced for the plasmids that only have an antibiotic selection marker were directly harvested. Collected medium was filtered through 0.45 µM filters and aliquoted as 1 ml solutions in microcentrifuge tubes. Lentiviruses were stored at -80°C for later use.

4.1.4. Lentiviral Transduction

For transduction of adherent cells, 200.000 cells were seeded into 6-well plates one day before the transduction. Next day, lentivirus aliquots were defrosted from -80°C and mixed with polybrene at a final concentration of 4 µg/ml. The media of the cells were aspirated, and the lentivirus-containing medium was added onto the cells. After 6-8 hours, the lentivirus-containing medium was removed, and the regular medium was added onto the cells.

For transduction of suspension cells, 125.000 cells were centrifuged, and the old medium was discarded. The cell pellet was resuspended in lentivirus-containing medium mixed with polybrene at a final concentration of 8 µg/ml. Cell suspension was seeded into 24-well plate and centrifuged at 1,000 x g for 30 minutes at 32°C. After 6-8 hours, cells were resuspended in lentivirus containing medium and transferred into a 15 ml falcon. Centrifuged at 300 x g for 5 minutes and the medium was discarded. Cell pellet was resuspended in regular medium.

Lentivirus transduction was continued with antibiotic selection to the generate stable cell lines.

4.1.5. Analysis of Transduction Efficiency by Flow Cytometry

Transduction efficiency of fluorescent marker carrying cells are analyzed by BD Accuri 6 machine. Cells were trypsinized and collected into FACS tubes. Cells were centrifuged at 500 x g for 2 minutes to discard cell medium. Cells were washed with PBS

twice and centrifuged again. Cell pellet was resuspended in 500 µl 1X PBS, and vortexed well to obtain single cells and to prevent clump formation. FL-1 channel was used to determine self-fluorescence of the untransduced cells and threshold was set. Based on this threshold, the percentage of transduced cells was determined.

4.1.6. Generation of Stable Cell Lines by Antibiotic Selection

For the generation of pCW-dCas9-puro, pCW-BirN-dCas9-puro, pCW-BirC-dCas9-puro, pCW-dCpf1-blast, pCW-BirN-dCpf1-blast, and pCW-BirC-dCpf1-blast stable MCF-7 and K562 cells, cells were transduced with lentiviruses. After 72 hours of transduction, 2 µg/ml puromycin was added to puromycin carrier cells, and 5 µg/ml blasticidin S was added to blasticidin carrier cells. Selection was continued until all untransduced cells were dead.

4.1.7. Doxycycline Treatment of Inducible Stable Cell Lines

For the induction of dCas9, BirN-dCas9, BirC-dCas9, dCpf1, BirN-dCpf1, BirC-dCpf1 proteins, stable cell lines were treated with 2 µg/ml doxycycline. Cells were seeded one day before the experiment at 80% confluency. Doxycycline is a light-sensitive agent and the treatments were performed in the dark. Doxycycline was refreshed every 24 hours since its half-life is short.

4.1.8. Generation of Stable Cell Lines of by Electroporation

In CAPTURE Method, K562 cells were electroporated to create stable cell lines (Liu *et al*., 2017). This method was optimized by using pLKO5.EFS.GFP empty plasmid and use to deliver pEF1a-FB-dCas9-puro and pEF1a-BirA-V5-neo constructs, followed by selection with neomycin and puromycin. K562 cells were collected and centrifuged at 500 x g for 5 minutes. Cell medium was discarded and the cell pellet was resuspended in Opti-MEM medium as 5 x 10^6 cell per 1 ml. 20 µg FB-dCas9 and 20 µg BirA carrying plasmids were mixed with Opti-MEM medium containing the K562 cells in a microcentrifuge tube. Mixture was transferred into electroporation 4 mm cuvette, and electroporation was

performed as a single pulse at 875 V, infinite Ω, and 50 μF by using X-Cell GenePulser Electroporator (Bio-Rad). After electroporation, cells were incubated for 10 minutes without any movement. Then, cells were transferred into 15-ml falcon and centrifuged at 300 x g for 5 minutes. Cell pellet was resuspended in 1 ml RPMI 1640 medium supplemented with 20% FBS, 1% penicillin/streptomycin antibiotics and 1% non-essential amino acid. Cells were transferred into 96-well plates (100 μl/well), immediately. After 72 hours, cells were treated with 1 μg/ml puromycin and 600 μg/ml neomycin. Fresh medium supplemented with 1 μg/ml puromycin and 600 μg/ml neomycin was added onto the cells every two days for 2 weeks until the colonies become apparent in wells. Single cell-derived colonies transferred into 12-well plates and antibiotic treatment was continued for 2 weeks. Later, 1 x 10^6 cells were harvested and, western blot analysis were performed on selected colonies to confirm dCas9 and BirA expression. All validated colonies were frozen immediately, and kept at -150°C for later use.

4.2. Molecular Cloning

4.2.1. Plasmid Preparation

Bacterial culture of the plasmids were prepared from glycerol stock and inoculated into 10 ml LB-Ampicillin solution and grown overnight at 37°C. Plasmid DNA was isolated by using the GeneMark Plasmid Miniprep Purification Kit according to the manufacturer's instructions. In addition, ZymoPURE II Plasmid Midiprep kit was used to obtain higher concentrations and remove endotoxins for cell culture usage. All plasmid DNAs were stored at -20°C.

4.2.2. Generation of CRISPR Protein Constructs by Restriction-Ligation

For the generation of lentiviral dCpf1 plasmid, pCW-blast, which is a doxycycline-inducible system, was chosen as a backbone. First, pCW-Cas9-blast and WN10151 plasmids were digested with NheI and BamHI restriction enzymes at 16°C overnight. Digested samples were loaded on 1% agarose gel and, electrophoresis was performed at 110 V for 45

minutes. The gel was visualized under UV light by using BioRad GelDoc. pCW-blast and dCpf1 insert were cut from gel on a UV box and extracted by Nucleospin Gel and PCR Clean-up Kit (Macherey Nagel) according to manufacturer's instructions. Purified samples were used in the ligation reaction using a molar ratio of 1:3 vector to insert. In the ligation reaction, 1 µl T4 DNA ligase, 1 µl 10X T4 DNA ligase buffer and ddH_2O was used to complete the reaction to 10 µl. 5 µl of the ligation product was transformed into 50 µl competent cells. Single colonies were chosen, and plasmid isolation was performed. Correct colonies were confirmed by Sanger Sequencing.

In order to add 3x HA to dCpf1, pCW-dCpf1-blast plasmid was digested with BamHI restriction enzyme to linearize the construct at the C-terminal end of dCpf1. Linearized plasmid was purified by using Nucleospin Gel and PCR Clean-up Kit (Macherey Nagel) to use it in home-made Gibson assembly reaction (See Section 4.2.3.). 3x HA coding sequence, which was used as an insert in the Gibson assembly reaction, was ordered as a oligo-duplex from Macrogen.

4.2.3. Generation of CRISPR-BirA* Halves Fusion Proteins by Home-made Gibson Assembly

Gibson assembly is a quick and easy method that provides all-in-one cloning (Gibson et al., 2009). This method does not necessitate a specific restriction site, and assemblies are independent of sequences. The assembly mixture of this cloning method was prepared in our lab and called "home-made". This reaction requires 30-40 bp complementary sequences at both sites of the insert and the backbone. PCR reaction was utilized to insert these complementary sequences to the reaction components. A typical primer for this reaction was 60 bp in length (30 bp homology sequences of the vector, 30 bp the an insert) (Gibson et al., 2009).

To generate CRISPR-BirA* halves fusion proteins, home-made Gibson assembly reaction was optimized by combining all the components of a reaction. Two sequential PCR reactions were performed to prepare inserts, which are BirA* halves. In the first PCR

reaction, BirA* protein was split into two halves, called BirN and BirC. The second PCR reaction was to insert complementary sequences to the both ends of the BirA* halves. In parallel, pCW-dCas9-puro and pCW-dCpf1-blast plasmids were linearized using NheI restriction enzyme, which is located at the N-terminal sites of CRISPR proteins. Then, four different reaction mixture were prepared by combining the linearized vectors with PCR products of BirA* halves to create BirN-dCas9, BirC-dCas9, BirN-dCpf1, and BirC-dCpf1. The DNA molecules were equimolar amounts (1:1) in each reaction. The reaction was held at 50°C for 1 hour, and followed by transformation of the competent bacteria (see Section 4.2.8).

4.2.4. Restriction Enzyme Digestion of sgRNA Expression Plasmids

Three μg pLKO5.sgRNA.EFS.GFP and lenti_gRNA-puro plasmids were digested with FastDigest Esp3I restriction enzyme. The reaction was prepared with 1 μl FastDigest Esp3I, 1 μl 10X FastDigest Buffer and 1 μl Fast AP (alkaline phosphates) enzyme. Volume of the reaction was completed to 60 μl by using nuclease-free double distilled water. Restriction digestion reaction was performed for 2 hours at 37°C, followed with incubation at 65°C for 10 minutes to inactivated enzymes by heat.

4.2.5. Preparation, Annealing and Phosphorylation of sgRNA Oligos

sgRNA oligos were designed by using Benchling CRISPR tool, and ordered as antisense and sense oligos from Macrogen (S. Korea) oligo synthesis services. The primers were suspended in appropriate volumes of nuclease-free distilled water to obtain 100 μM oligo suspensions, which were stored at -20°C for long-term. For oligo annealing and phosphorylation reaction, 1 μl of sense and antisense oligos were mixed with 1 μl T4 polynucleotide kinase enzyme (NEB) and 1 μl of 10X T4 DNA ligase buffer (NEB). The total volume of the reaction was completed to 10 μl with nuclease free distilled water. For phosphorylation reaction, the mixture was incubated at 37°C for 45 minutes in a thermal cycler. After phosphorylation, samples were incubated at 95°C for 5 minutes and the temperature was decreased to 25°C with a °C/min rate. After incubation at 25°C for 5 min,

oligo phosphorylation and annealing reaction was completed and samples were stored at -20°C.

In the cloning of sgTelomere (Chen *et al*., 2013), pSLQ1651-sgGal4(F+E) plasmid was used as a backbone after digested with BstXi and XhoI restriction enzymes. In parallel, PCR reaction was performed in which pSLQ1651-sgGal4(F+E) plasmid was used as a template to amplify sgTelomere sequence along with the scaffold. The PCR reaction condition was 98°C for 30 sec; 98°C for 10 sec, 62°C for 10 sec, 72°C for 10 sec, and 30 cycles; 72°C for 5 minutes. PCR product was run on 1% agarose gel and, the 150 bp PCR amplicon was isolated from the gel (see Figure E.1.). Gel isolated sgTelomere amplified with its scaffold was digested with BstXI and XhoI restriction enzymes. Then, pSLQ1651 plasmid backbone and sgTelomere insert was ligated. The sequence was confirmed by Sanger sequencing.

4.2.6. Agarose Gel Electrophoresis of Digested Plasmid

All digested and undigested plasmids were loaded into 0.8% agarose gel to check the success of restriction digestion reaction of the samples. Electrophoresis was performed at 100 V for 1 hour, and agarose gel was incubated in 1X TAE supplemented with ethidium bromide for 10 minutes. The gel was visualized under UV light by using Biorad Geldoc. In the case of successful digestion, bands that have correct size were cut from the gel and purification was performed by using Nucleospin Gel and PCR Clean-up kit (Macherey Nagel) according to manufacturer's instructions. Extracted digested plasmid DNA was stored at -20°C. Similarly, PCR products were run in different concentrations of agarose gel based on their product size. PCR products were purified either by PCR clean-up or gel extraction protocols of Nucleospin Gel and PCR Clean-up kit (Macherey Nagel).

4.2.7. Ligation Reaction of Digested Plasmid and Insert

In regular ligation reaction, 1:3 (vector:insert) molar ratio of sample were calculated based on the 50 ng digested plasmid DNA, which was used as a backbone. For the cloning

of sgRNAs, 50 ng plasmid DNA is ligated with 2 μl of 1:200 diluted annealed oligo duplexes, which were diluted in double distilled water. For the reaction, T4 DNA ligase enzyme (NEB) was used. Plasmid DNA and diluted oligo was mixed with 1 μl T4 DNA ligase buffer (NEB) and 1 μl T4 DNA ligase was added. The total volume of the reaction was completed to 10 μl with double distilled water. The mixture was incubated at 16°C for overnight in a thermal cycler, followed by heat inactivation at 65°C for 10 minutes. Ligation products were used in the transformation directly.

4.2.8. Transformation of Ligation Products into Competent Bacteria

Ligation products were transformed into recombinase deficient Stbl3 *E. coli* strain which were prepared home-made by $CaCl_2$ method. 5 μl of ligation product was added into the competent bacterial cells, and this mixture was incubated on ice for 30 minutes. Then, competent cells were incubated at 42°C for 45 seconds followed by the incubation on ice for 5 minutes. For the recovery of heat-shocked competent bacterial cells, 0.95 ml of LB without any antibiotics was added to the cells, and incubated at 37°C for 1 hour on an orbital shaker. Bacterial culture was spinned down, and 800 μl of LB medium was discarded. The pellet was resuspended in the remaining LB medium, and spread on antibiotic containing LB-agar plates. Plates were incubated at 30°C or 37°C overnight. Next day, colony numbers were compared to the no insert controls.

4.2.9. Validation of sgRNA Oligo Cloning via Polymerase Chain Reaction

Single bacterial colony was picked from the agar plates, and the micropipette tip was touched to both LB and Bacterial Lysis Buffer, which was boiled for 15 minutes. 20 μl Colony PCR Reaction was prepared by using My*Taq* DNA Polymerase, and added 2 μl of the boiled bacterial lysis buffer (Figure 4.1.). In the reaction, U6-fwd was used as a universal forward primer, while the reverse primer was the sgRNA oligo itself.

95°C	5 min	
95°C	30 sec	29x
60-56°C	30 sec	
72°C	30 sec	
72°C	5 min	
12°C	∞	

Figure 4.1. Cycles of Colony PCR Reaction is indicated.

The PCR products were run on 2% agarose gel to validate the single amplification. The sequence of PCR-validated clones were analyzed by Sanger sequencing.

4.2.10. Validation of sgRNA Oligo Cloning via Sanger Sequencing

All of the sequencing samples were analyzed by Macrogen Europe Incorporation. Single bacterial colonies were picked from the agar plates, and inoculated overnight. Following day, plasmid DNAs were isolated by using GeneMark Plasmid Miniprep Purification Kit. Plasmid DNA concentrations were measured and 2 μg DNAs were prepared for the sequencing reaction. Primers that are used in the sequencing is indicated in the Primers section.

4.3. Western Blotting

4.3.1. Preparation of Protein Samples

Culture medium was discarded, and cells washed with 1X PBS. Cells were trypsinized and collected in 15 ml tubes. Trypsinized cells were centrifuged at 500 g for 2 minutes and the supernatant was discarded. The cell pellet was washed with ice-cold 1X PBS twice and centrifuged again. The supernatant was discarded and the cell pellet was resuspended in cold RIPA buffer (150 µl per $1x10^6$ cells) supplemented with protease inhibitor cocktail. Resuspended cell pellet was vortexed vigorously 10 times for 1 second and incubated on ice for 30 minutes by vortexing every 10 minutes. Then, cell lysate was

centrifuged at 14,000 x g for 15 minutes, and the supernatant was transferred into a new 1,5 ml microcentrifuge tube. Total protein samples were stored at -20°C. For long-term storage, they were kept at -80°C.

4.3.2. Calculation of Protein Concentration by BCA Assay

Protein concentration was measured by the Pierce BCA Protein Assay Kit (Thermo scientific). BSA samples in different concentrations were prepared and used as a standard solution based on the instructions of the manufacturer. Working Reagent was prepared with Reagent A and Reagent B as 50:1 ratio. Working reagent and the samples were mixed in 96-well plate. In each well, 200 µl Working Reagent was mixed with 25 µl protein and standard samples. RIPA buffer was used as the blank solution for protein samples, and 1x PBS was used as the blank solution for standard solutions. Plate was covered and incubated at 37°C for 30 minutes in dark. Later, the absorbance of each samples were measured at 562 nm by using a plate reader. Absorbance of the blank solutions were subtracted from the samples, and the standard curve was drawn by using Microsoft Office Excel by fitting a linear regression line to the datapoints. Protein concentrations were calculated by using a formula obtained from the standard curve.

4.3.3. Protein Sample Preparation and SDS-PAGE Electrophoresis

Protein samples were mixed with 4X Loading Sample Buffer, and boiled at 95°C for 5 minutes. Samples were centrifuged at 14,000 x g for 10 minutes to clear samples from possible impurities. Samples were loaded into SDS the wells of SDS-PAGE gel. SDS-PAGE gels were prepared based on the sample size that will be analyzed. Usually, 8 % resolving gel was prepared and loaded into Biorad gel cassette, and the gel surface was covered with 100% isopropanol immediately. After the solidification of the resolving gel, isopropanol was discarded completely. 6% stacking gel was prepared, and loaded onto the resolving gel, and the comb was placed. After solidification, protein samples were loaded into the gel. The samples were run at 80 V for 30 minutes, and at 100 V for about 1.5 hours in 1x Running Buffer.

4.3.4. Membrane Blotting and Transfer of Proteins

When the running is completed, proteins were transferred to PVDF membrane. First, Whatmann papers and PVDF membrane were cut in equal sizes with the SDS-PAGE gel. PVDF membrane was incubated in 100% Methanol for 20 seconds, and activated. Membrane was incubated in ddH_2O for 2 minutes to remove methanol. Membrane and the gel were incubated in 1X Transfer buffer for 30 minutes on shaker. In transfer, sandwich was prepared with the Whattmann papers, PVDF membrane and the gel, which were soaked in 1X Transfer Buffer. The excess transfer buffer was removed from the sandwich with roller, and the sandwich was placed into the Biorad Semi-Dry TransBlot. Transfer of proteins was performed at constant 25 V for 30 minutes. After transfer, membrane was blocked in 5% BSA dissolved in 1x TBS-T for 1 hour at room temperature, followed by wash steps in 1X TBS-T three times for 5 minutes.

4.3.5. Primary and Secondary Antibody Incubations

After washing, membrane was cut according to the sizes of the proteins by using protein ladder as a reference. Primary antibodies were diluted in 5% BSA solution dissolved in 1X TBS-T. Membranes were incubated in primary antibodies overnight at 4°C on an orbital shaker. The following day, membranes were washed with 1X TBS-T three times for 5 minutes. Secondary antibodies were prepared 5% BSA solution dissolved in 1X TBS-T. Membranes were incubated in secondary antibodies at room temperature for 1 hour on an orbital shaker. When the incubation is over, membranes were again washed with 1X TBS-T three times for 5 minutes.

4.3.6. Chemiluminescence Detection of Protein Bands

WesternBright ECL chemilumiscent substrate mixture was prepared to visualize protein bands. Membranes were incubated in the substrate mixture for a short time at dark and then visualized by using Syngene image processor (Syngene International Limited).

4.4. Chromatin Immunoprecipitation (ChIP)

4.4.1. Isolation of DNA by Immunoprecipitation

5-10 million cells were prepared for each chromatin immunoprecipitation. First, the cells were crosslinked by using in a 1% final concentration of formaldehyde with gently shaking for 10 minutes in room temperature. Formaldehyde crosslinking was quenched by 0.125 M final concentration of glycine with gently shaking for 5 minutes in room temperature. Cells were washed twice with 10 ml ice-cold PBS. In the usage of adherent cells, cells were scraped with 5 ml ice-cold PBS, collected in 15 ml falcons, and centrifuged at 2000 x rpm for 2 min at 4°C. Cell pellet was resuspended in cellular lysis buffer supplemented with protease inhibitor (typically 1 ml for 10 cm cell plate), incubated 5 min on ice, and centrifuged at 1200 x rpm for 2 min at 4°C. The supernatant was discarded and the pellet was resuspended in nuclear lysis buffer supplemented with protease inhibitor (typically 300 µl for 1 x 10^7 cells). Cell suspension was transferred into a 0.5 ml thin-wall PCR tubes. Previously, sonication conditions were optimized specific to a cell line using QSonica machine. Based on a cell line, sonication was performed to obtain 100-500 bp chromatin. Sonicated lysates were cleared by spinning at 14000 x rpm for 10 min at 4°C. 20% of the sonicated lysate was separated as 'input control'. Rest of the soluble chromatin lysate was transferred into a new microcentrifuge tube, and diluted ten-fold with IP dilution buffer supplemented with protease inhibitor cocktail. Lysates were incubated overnight at 4°C with 50 µl of Protein G magnetic beads pre-bound with 5 µg target antibody for 45 minutes at room temperature, or only beads. On the following day, bead-bound immune complexes were washed 2 times each for 2 min at room temperature with shaking using 0.8 ml of the following: IP Dilution Buffer, High Salt Buffer with 500 mM NaCl, LiCl Buffer and TE. Input controls were treated with RNase A (0.1 µg/µl final concentration) for 45 minutes at 38°C, and boiled for 10 minutes along with the ChIP DNA samples. Both input controls and ChIP DNA samples were treated with Proteinase K (200ng/µl final concentration) for 30 minutes at 55°C, and boiled again for 10 minutes. Beads were collected by magnetic stands, and the soluble chromatin fragments were isolated either with silica-based columns or phenol-chloroform.

4.4.2. Quantitative PCR

Isolated ChIP-DNA and input DNA samples are diluted 15-fold and 50-fold, respectively. PikoReal instrument (Thermo Scientific) was used to amplify diluted DNA samples in triplicates. Different master mixes consisting of 65-fold diluted primers, enzyme and nuclease-free water were prepared separately for target-specific and control primer pairs due to prevent contamination or unwanted amplification. For general real time PCR, protocol was followed as 40 cycles of 95°C for 15 s, 60°C for 1 minutes.

The calculations and the real time PCR protocol were different in the case of real time PCR amplifying telomeric regions. Each real-time PCR reaction for the amplification of telomere and a single copy gene was prepared in a separate 96-well plates. In the calculation of chromatin-associated protein enrichment to the telomeric region, a single copy gene was chosen, 36B4. Primer concentrations in the master mix were prepared based on a modified protocol (Cawthon, 2002). The final primer concentrations in the master mix were: ChIP-Telomere-F 270 nM; ChIP-Telomere-R 900 nM; ChIP-36B4-F 300 nM; ChIP-36B4-R 500 nM. During the amplification by thermo cycler, telomere real time PCR protocol was followed as 18 cycles of 95°C for 15 s, 54°C for 2 min. On the other hand, 36B4 amplification protocol was followed as 30 cycles of 95°C for 15 s, 58°C for 1 min.

ΔΔCT method was used to calculate fold-differences of each ChIP obtained from real-time PCR amplification. The result of each ChIP was normalized to corresponding data of input samples, and the graph was obtained.

4.5. CAPTURE

4.5.1. Isolation of DNA

1 x 10^7 cells expressing FB-dCas9/BirA and sgRNA were collected, and centrifuged. The cells were resuspended in 10 ml ice-cold PBS. Cells were treated with 1% final concentration of formaldehyde to fix protein-DNA complexes, and incubated for 10 minutes

at room temperature with rotation. 1 M glycine was used to quench the cross-linking, and incubated for 5 minutes at room temperature with rotation. Cells were washed twice with 10 ml ice-cold PBS, and centrifuged. Cell pellet was resuspended in 1 ml nuclear extraction buffer, and incubated for 30 minutes at 4°C with rotation. Then, cells were centrifuged at 2,500 x g for 5 minutes at 4°C to pellet nuclei. Nuclear pellets were resuspended in 400 µl nuclear lysis buffer, and sonicated in an ice water bath using a QSonica at optimal sonication conditions for each cells. 44 µl 10% Triton X-100 was added to sonicated nuclei lysate, and the lysate was centrifuged 10 minutes at 16,100 × g at 4°C. 400 µl supernatant was transferred into a new microcentrifuge tube, and added 24 µl 5 M NaCl. In parallel, MyOneTM Streptavidin C1 Dynabeads were mixed well, and 20 µl beads per IP were transferred into a new microcentrifuge tube. Beads were collected on a magnet stand, and washed twice with 1 ml RIPA 0.3 buffer. Beads were again collected on a magnet stand, and sonicated chromatin was added to the beads. Tubes were incubated on a rotator shaker at 4°C for overnight. On the following day, beads were collected on a magnet stand, and washed twice by adding 1 ml 2% SDS. Suspension was vortexed well for 15 seconds and beads were collected. Beads were washed twice for each with 1 ml high salt wash buffer, 1 ml LiCl wash buffer and 1 ml TE buffer. beads were collected and resuspended in 100 µl SDS elution buffer. Incubate tubes at 65°C for 6 hours. Then, beads were collected on a magnetic stand, and supernatant was transferred into a new microcentrifuge tube. Beads were rinsed with 50 µl SDS elution buffer, and supernatant was added to SDS eluted chromatin. Eluted chromatin was treated with 1 µl of 0.5 µg/µl RNase A and 1 µl 20 mg/ml Protease K, and incubated incubate at 37°C for 2 hours. ChIP DNA was isolated in 15 µl elution buffer using Zymo ChIP DNA Cleaner and Concentrator kit. Isolated DNA sequences were validated via qPCR.

4.5.2. CAPTURE-Western blotting

1 x 10^8 cells expressing FB-dCas9/BirA and sgRNA were collected, and aliquoted into 15-ml falcons with 5 x 10^7 cells per tube. The cells were centrifuged at 4°C, and resuspended in 10 ml ice-cold PBS. Cells were treated with 2% final concentration of formaldehyde to fix protein-DNA complexes, and incubated for 10 minutes at room temperature with rotation. 2 M glycine was used to quench the cross-linking, and incubated

for 5 minutes at room temperature with rotation. Cells were washed twice with 2 ml ice-cold PBS, and centrifuged. Cell pellet was resuspended in 2 ml cold cell lysis buffer, and incubated for 30 minutes at 4°C with rotation. Then, cells were centrifuged at 2,500 x g for 5 minutes at 4°C to pellet nuclei. Nuclear pellets were resuspended in 500 µl cell lysis buffer, and treated with 1 µl 0.5 µg/µl RNase A. Cell suspension was incubated at 37°C for 30 minutes with rotation. Then, cells were centrifuged at 2,500 x g for 5 minutes at 4°C to pellet nuclei. Nuclear pellets were carefully resuspended in 400 µl SDS nuclear lysis buffer, and incubated for 10 minutes at room temperature. 1.2 ml urea buffer was added, and the cell suspension was mixed by inverting. Then, cells were centrifuged at 16,100 x g for 25 minutes at room temperature. This step was repeated. The pellet was carefully resuspended in 400 µl cell lysis buffer, and 1.2 ml cell lysis buffer was added to dilute SDS. Then, cells were centrifuged at 16,100 x g for 25 minutes at room temperature. This step was repeated. The pellet was resuspended in 400 µl IP binding buffer without NaCl. The suspension was transferred into a twin-wall microcentrifuge tubes, and sonicated with optimal conditions. Supernatant were pooled in a new 1.5 ml microcentrifuge tube. 5 M NaCl was added to sonicated chromatin to a final concentration of 150 mM, and mixed well. In parallel, 50 µl MyOneTM Streptavidin C1 Dynabeads were collected per IP. The beads were washed twice with IP binding buffer and collected on a magnetic stand. Beads were resuspended in 25 µl IP binding buffer and combined with the sonicated chromatin. suspension was incubated at °C for overnight on a rotator shaker. The following day, beads were collected on a magnetic bead, and resuspended in 1 ml IP binding buffer. They were incubated for 10 minutes at 4°C. this step was repeated four times. Then, beads were collected on a magnetic stand, and resuspended in 30 µl RIPA buffer and 10 µl 4x loading sample buffer. incubated at 95°C for 20 minutes. Beads were collected completely, and the supernatant was loaded on a SDS-PAGE gel for western blot analysis.

5. RESULTS

5.1. Experimental Design

To identify locus-specific chromatin-associated regulators, we decided to develop a method called Split-CRISPR-ID, which is expected to be more specific and sensitive compared to the previous methods. In this method, we combine the power of CRISPR-based chromatin loci targeting, which brings specificity, and temporally optimized biotin tagging via Split-BioID, which brings sensitivity (Figure 5.1.).

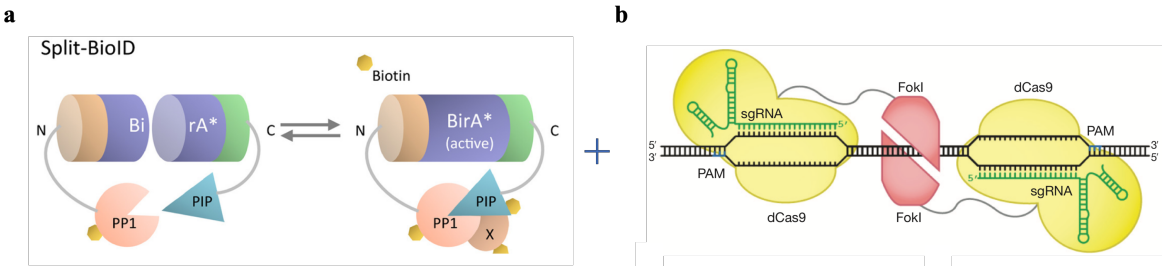


Figure 5.1. Split-CRISPR-ID method was created employing the idea of the Split-BioID (De Munter *et al.*, 2017) and the dual-targeting of CRISPR dCas9 proteins (Guilinger *et al.*, 2014).

We decided to fuse two inactive halves of BirA* biotin ligase (BirN and BirC) with nuclease-deficient versions of two distinct CRISPR proteins, dCas9 and dCas12a (dCpf1). We fused both BirN and BirC fragments to both dCas9 and dCpf1 proteins at the N-terminal site. A Gly-Ser rich flexible and soluble linker (Chen *et al.*, 2013), which is inserted in between the BirA* halves and CRISPR proteins, is expected to facilitate the motion of these BirN and BirC fragments in the cellular environment. When the CRISPR proteins bind to their targeted regions through their specific sgRNAs, which we design in a certain proximity, inactive BirN and BirC halves move around with the help of this linker, and expected to get activated by heterodimerizing when they come close to each other, as shown in (De Munter *et al.*, 2017; Schopp *et al.*, 2017). The proteins that are in close proximity to the activated

BirA* will get biotinylated *in vivo*, which will make it possible to isolate and characterize them using streptavidin-bead pulldown followed by mass spectrometry.

Cas9 and Cpf1 proteins bind different sequences with their corresponding sgRNAs that recognize target sequences through the protospacer adjacent motif (PAM) (Zetsche *et al*., 2015). The combinatorial implementation of these two separate PAM recognition modalities effectively increases the target recognition length and decreases the possibility of mismatching events. In dual-targeting, there are four different sgRNA binding orientations, which are PAM-in, PAM-out, PAM-left and PAM-right (Zhang *et al*., 2017) (Figure 5.2.). Based on the previous studies, PAM-out orientation gives the highest efficiency in the N-terminal fusion strategies (Guilinger *et al*., 2014; Zhang *et al*., 2017).

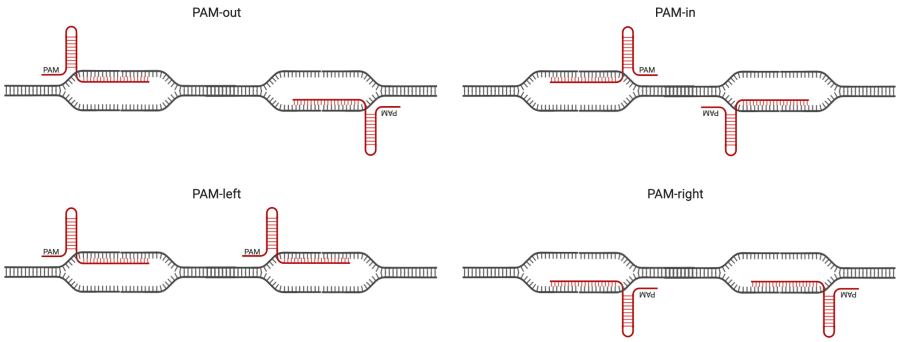


Figure 5.2. Schematic representation of the possible sgRNA orientations (created using BioRender).

In this method, heterodimerization and activation of the BirA* halves (called BirN and BirC) are completely dependent upon the dual-targeting by CRISPR proteins and the characteristics of Gly-Ser rich flexible linkers (Figure 5.3.).

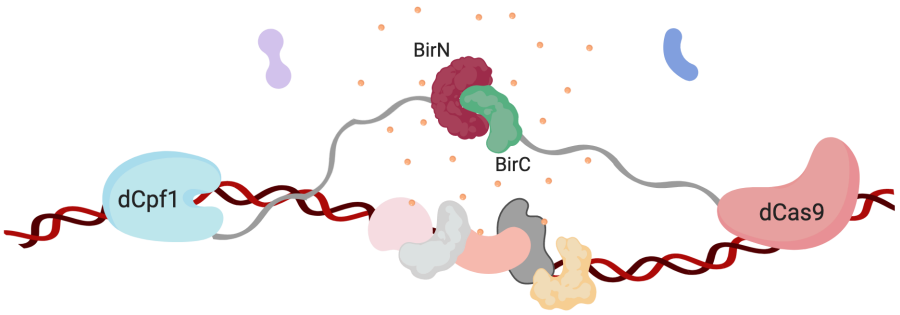


Figure 5.3. The model of Split-CRISPR-ID method represented on a specific locus (created using BioRender).

5.2. Generation of Fusion Proteins

For the application of dual CRISPR targeting, we decided to use lentiviral and doxycycline (dox) inducible systems. We used four different expression plasmid backbones to create sgRNAs using pLKO5.sgRNA.EFS.GFP (Heckl *et al*., 2014) and lenti_gRNA-puro (Kim *et al*., 2017), and fusion proteins BirN-dCas9, BirC-dCas9, BirN-dCpf1, and BirC-dCpf1 using pCW-puro and pCW-blast. We fused BirN and BirC to both dCas9 and dCpf1 at the N-terminal sites, because C-terminal fusions may require longer linkers, which is known for SpCas9 (Guilinger *et al*., 2014; Tsai *et al*., 2014). The linker sequence that we chose was GSAGSAAGSGEF, which was flexible, soluble and working efficiently in dual-CRISPR targeting approach (Guilinger *et al*., 2014).

As described previously, there are two publications where BirA* protein was split into two halves from different sites E140/Q141 and E256/G257 (De Munter *et al*., 2017; Schopp *et al*., 2017) (Figure 5.4.). One of these groups showed that E256/257 split site is more efficient compared to the other one in terms of the reconstitution and activation of BirA* (Schopp *et al*., 2017). They indicated that the alpha-helices connecting the C-terminal domain to the central domain are critical for the enzymatic activity of BirA* (Schopp *et al*., 2017). Thus, we decided to split and subclone BirA* protein into two parts (BirN and BirC) at E256/G257 using PCR.

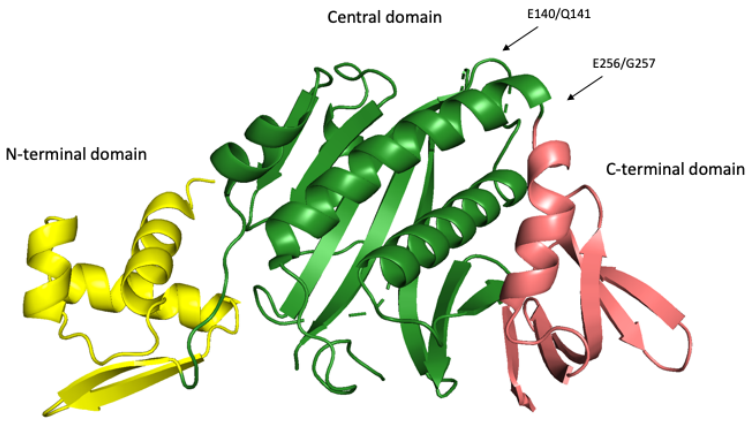


Figure 5.4. Crystal structure of BirA biotin ligase from *Escherichia coli* showing split sites (E140/Q141) and (E256/G257) are indicated (PDB 1BIB).

We used pCDNA3.1-MCS-BirA(R118G)-HA plasmid as a template for the generation of BirA* halves. We performed two sequential PCR reactions to prepare these BirA* halves, BirN and BirC, for the home-made Gibson assembly reaction (see Methods) (Figure 5.5.). In the first PCR reaction, we split BirA* into two halves and added 30 bp homology arms of destination vectors, pCW-puro and pCW-blast, and Gly-Ser rich flexible linkers at the 3' end to both of them. In second PCR reaction, we used the first PCR products as a template and added second 30 bp homology arms of destination vectors at the 3' ends (Figure 5.5.). We isolated each PCR products from agarose gel (Figure 5.6).

For the dCpf1 fusion proteins, we first generated pCW-dCpf1-blast construct, which does not contain any tag, and inserted 3x HA epitope-tag at the C-terminal site of the dCpf1 protein via home-made Gibson assembly (see Methods). Later, we linearized our pCW-dCpf1-blast with 3x HA tag by using NheI restriction enzyme in order to use the plasmid in home-made Gibson assembly reaction. On the other hand, we also digested pCW-dCas9-puro plasmid carrying 3x FLAG tag, which was created in our lab, with NheI restriction enzyme to linearize the vector to use it in home-made Gibson assembly (see Methods), which we prepared the reaction mixture by combining all ingredients (see Materials).

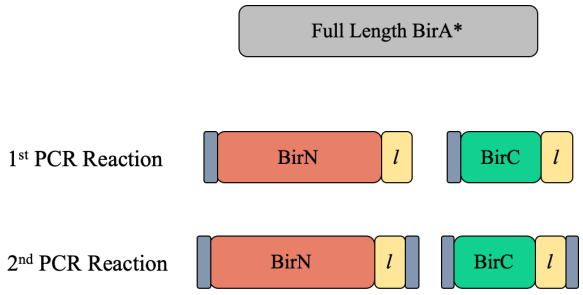


Figure 5.5. Schematic representation of PCR reactions to prepare BirA* halves for home-made Gibson assembly. 1st PCR reaction splits BirA* into two parts, BirN and BirC, and adds first homology arms at the 5' end (blue) and the Gly-Ser rich flexible linker at the 3' end (yellow). 2nd PCR reaction adds second homology arms at the 3'end of the protein parts.

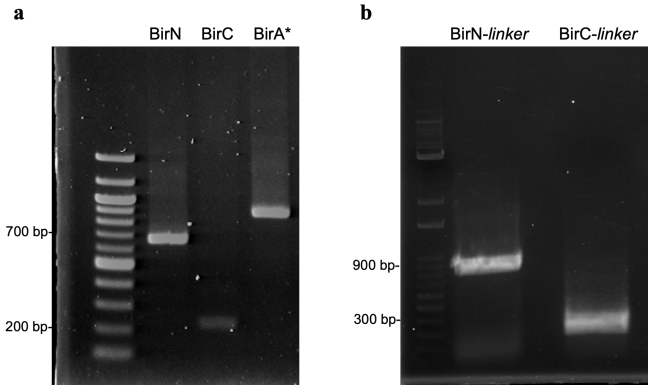


Figure 5.6. BirA* is split into two halves via PCR reaction. (a) The sizes of BirN, BirC and full lentgh BirA* is indicated. (b) 3' linkers and homology arms added to BirN and BirC fragments added after the 2nd PCR reaction.

BirN and BirC final PCR products were ligated with pCW-dCas9-puro and pCW-dCpf1-blast linearized plasmids in 1:1 vector to insert molar ratio. We created all pCW-BirN-dCas9-puro, pCW-BirC-dCas9-puro, pCW-BirN-dCpf1-blast and pCW-BirC-dCpf1-blast constructs using Gibson assembly that we prepared the reaction mixture ourselves (Figure 5.7.). We confirmed the plasmid sequences by Sanger sequencing, and concluded that both our sequential PCR reaction and Gibson assembly reaction were successful.

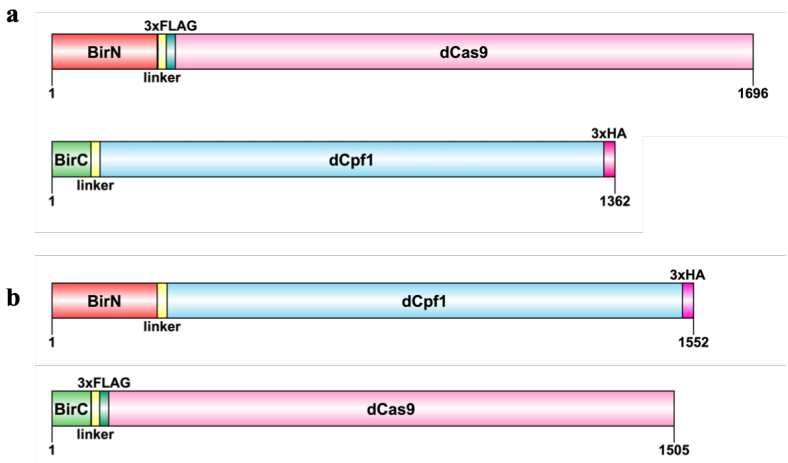


Figure 5.7. (a) Diagrams of BirA* fragments and CRISPR fusion proteins were constructed using Domain Graph (DOG) software (Ren *et al*., 2009), version 2.0. (b) Schematic representation of plasmid constructs consisting of fusion proteins.

In order to detect the expressions of these fusion proteins in HEK293FT cells, we performed western blot analysis. Here, we observed the shift in the molecular weights of the fusion proteins (Figure 5.8).

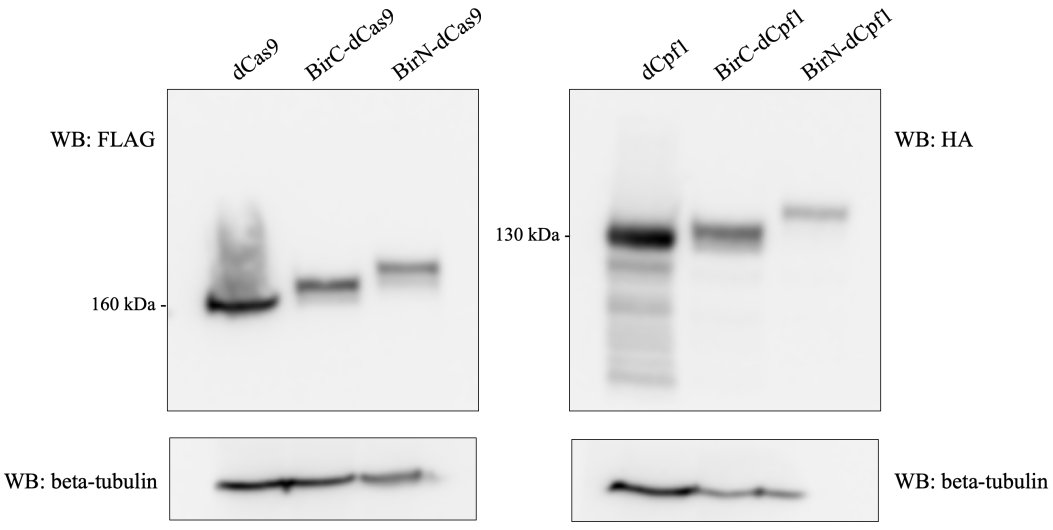


Figure 5.8. Fusion protein expressions were detected by western blotting.

5.3. 3D Model of Split-CRISPR-ID

To be able to estimate the optimal distances (in base pairs) between the gRNAs targeting the two halves of BirA*, we wanted to create a model using PyMOL tool to draw the actual dimensions of Cas9, Cpf1 and BirA* halves (Figures 5.9. and 5.10.). We found the longest dimensions of these proteins on chromatin as follows; Cas9 is 10.7 nm, dCpf1 is 11.5 nm, BirN is 6.4 nm and BirC is 2.5 nm. In addition, Gly-Ser rich flexible linkers (Chen *et al*., 2013) were calculated to be about 2 nm in their extended form. Thus, we designed our distinct sgRNA pairs in 80-150 bp distance (Figure 5.11.).

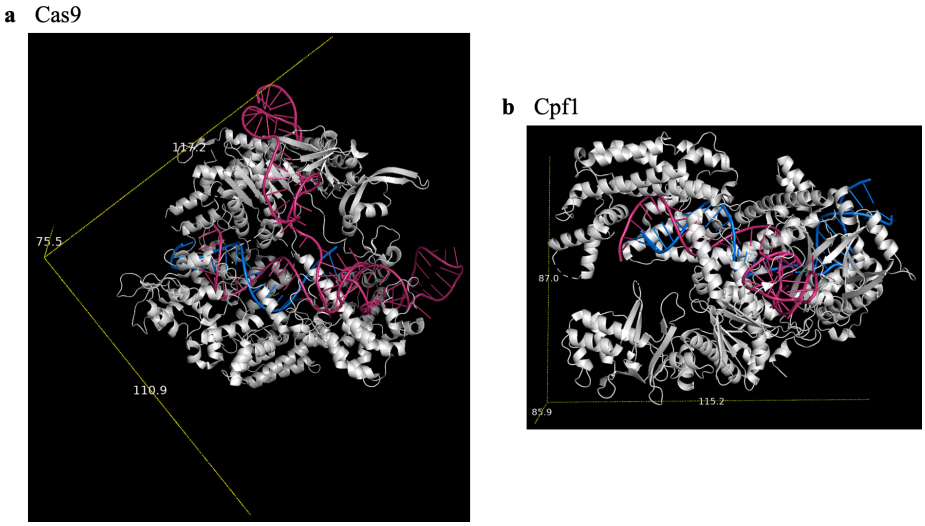


Figure 5.9. (a) 3D structure of dCas9 in complex with gRNA (pink) and target DNA (blue) is represented with cartoon. (b) 3D structure of dCpf1 in complex with crRNA (pink) and target DNA (blue) is represented with cartoon.

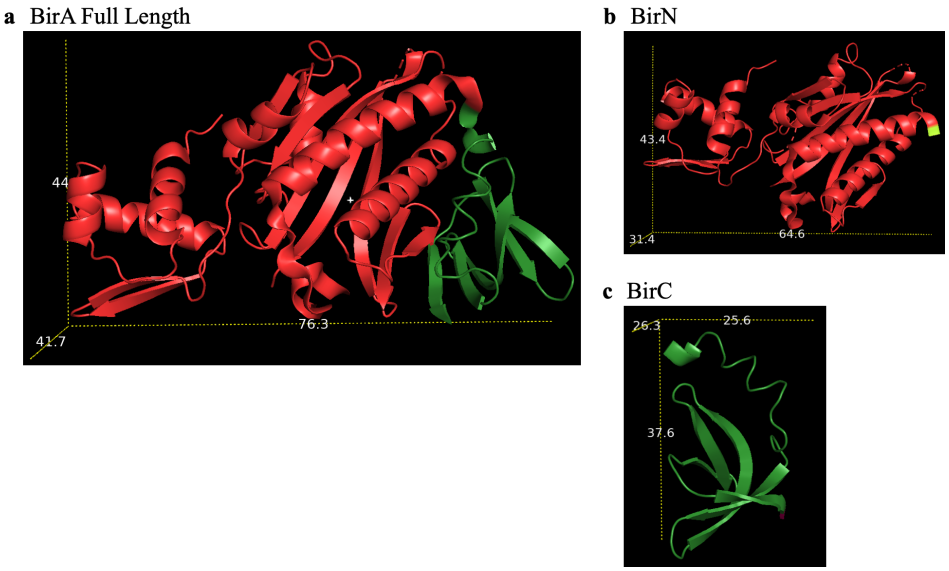


Figure 5.10. (a) Crystal structure of Promiscuous BirA* biotin ligase. (b) N-terminal region of BirA* (BirN) (c) C-terminal region of BirA* (BirC).

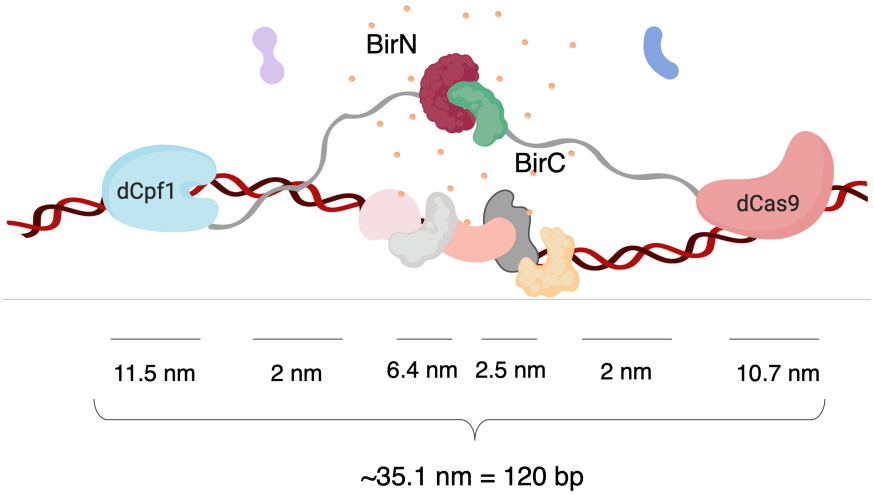


Figure 5.11. Actual dimensions of the Split-CRISPR-ID components. The dimensions were calculated based on the crystal structures of the proteins.

5.4. Generation of Stable Cell Lines that Contain FLAG-dCas9 and HA-dCpf1

To test our method on various loci, we decided to create several sgRNAs, which enable us to select the targetable sequences efficiently. In order to test the guiding efficiencies of these sgRNAs, we used FLAG-dCas9 and HA-dCpf1 stable cancer cell lines created by lentiviral transduction. The expressions of these proteins are controlled by a doxycycline inducible promoter (Figure 5.12.).

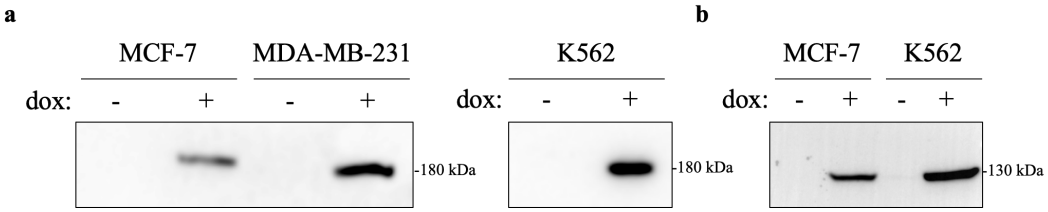


Figure 5.12. (a) Stably doxycycline inducible FLAG-dCas9 expressing MCF-7 , MDA-MB-231 breast cancer cells lines and K562 erythroleukemic cell line were created. (b) Stably doxycycline inducible HA-dCpf1 expressing MCF-7 and K562 cell lines were created.

5.5. Design, Cloning and Expression of sgRNAs in Stable Cell Lines

We designed our sgRNAs for CRISPR-targeting by using Benchling's (https://benchling.com) CRISPR tool. This tool ranks candidate target sites based on their on-target and off-target scores, which indicates the efficiency of guiding. We chose target sites with the lowest number of off-target and highest on-target scores for our further experiments. We listed all target sites used in this study and their scores in Table C.1. (see Appendix).

We chose several target sites, which are mostly well-studied, including Telomeric regions, pS2/TFF1 promoter, MAPK promoter and HIRA promoter to test our method as described below. We used pLKO5.sgRNA.EFS.GFP plasmid for dCas9 sgRNAs and

lenti_gRNA-puro plasmid for dCpf1 sgRNAs, both of which are also lentiviral (Heckl *et al*., 2014; Kim *et al*., 2017). All of the plasmids express sgRNAs under U6 promoter in a constitutive manner, however; the following gRNA scaffold sequences are different and specific to each of the CRISPR proteins due to the differences in the sgRNA secondary structures (Zetsche *et al*., 2015). To clone sgRNAs under hU6 promoter, we digested our plasmids with Esp3I (BsmBI) restriction enzyme and isolated the plasmid backbone from agarose gel for the further cloning steps (Figure 5.13).

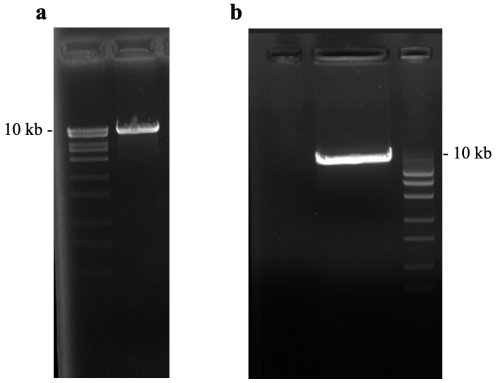


Figure 5.13. (a) lenti_gRNA-puro and (b) pLKO5.EFS.GFP plasmids were digested with Esp3I (BsmBI) enzyme.

We designed and ordered sgRNA coding sequences as sense and anti-sense DNA oligos, which carry Esp3I (BsmBI) restriction site overhangs. We annealed complementary oligos, and cloned these oligo-duplexes into either lenti_gRNA-puro or pLKO5.sgRNA.EFS.GFP plasmids, followed by the transformation into competent bacterial cells. Later, we isolated plasmid DNAs from single colonies and confirmed the results by Sanger sequencing (see Appendix B section).

5.5.1. Telomere Targeting

We chose one of the popular target region, telomeric repeats, to initially test our novel method. The main reason that we chose this region is that the telomeres are repetitive genomic regions, and their isolation provides us with more material to study with, when

targeted with a single sgRNA (Figure 5.14.), in terms of abundance than a single-copy locus in a single cell (Fujita *et al*., 2013; Liu *et al*., 2017; Schmidtmann *et al*., 2016). In addition, some of the binding factors are already known (Liu *et al*., 2017).

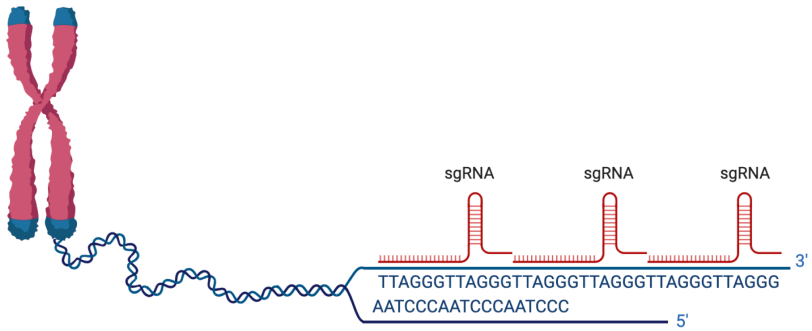


Figure 5.14. Schematic representation of the telomere. The sgRNA target sequence (purple hairpin) and PAM sequence (pink) indicated (created using BioRender).

We cloned the sgRNA targeting telomeric region into pSLQ1651-sgGal4(F+E) plasmid backbone (Chen *et al*., 2013; Liu *et al*., 2018, 2017). We amplified sgTelomere with the scaffold sequence, which is especially modified for telomere targeting, by using PCR reaction (see Appendix E section). We ligated the backbone and sgTelomere, and transformed the ligation product into competent bacterial cells. Later, we isolated plasmid DNAs from single colonies and confirmed the results by Sanger sequencing (see Methods and Appendix B section).

We first validated the sgTelomere targeting MCF-7 cells stably expressing FLA-dCas9. We performed ChIP experiment performed by immunoprecipitation with anti-FLAG antibody. Targeting telomeric repeats was efficient, when we compared the results with negative region amplification and only beads control (Figure 5.15.).

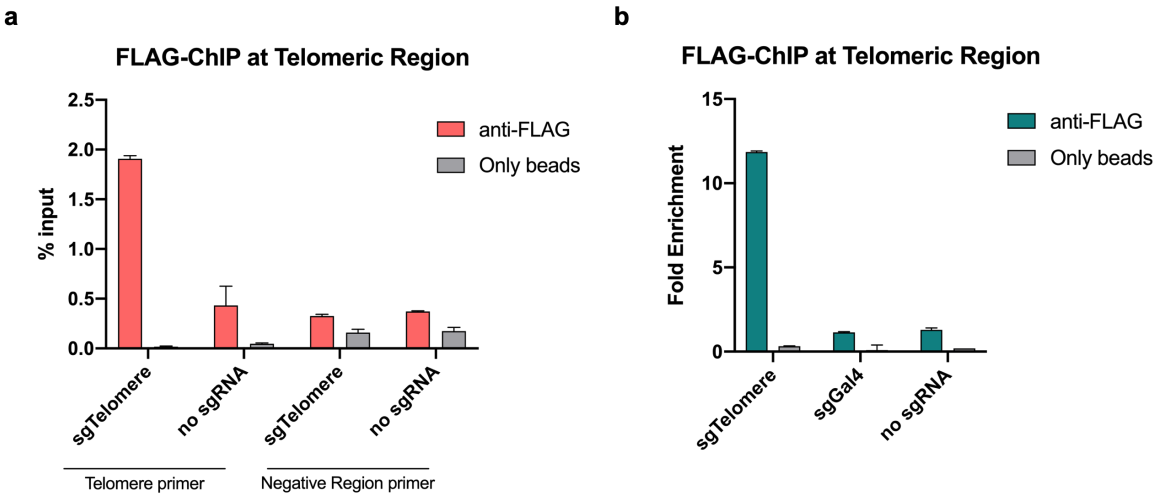


Figure 5.15. anti-FLAG-ChIP performed at Telomeric repeats followed by qPCR in MCF-7 cells stably expressing FLAG-dCas9. 3xFLAG-dCas9 is targeted to Telomeric repeats using sgTelomere. (a) 3xFLAG-dCas9 is enriched at Telomeric repeats compared to the negative region. (b) 3xFLAG-dCas9 is enriched at Telomeric repeats compared to non-targetted sgRNA (sgGal4) and the negative region.

After confirming the telomere targeting, we first tested the biotinylation in our system using BirN-dCas9 and BirC-dCas9 fusion proteins. Since the multiple binding events occur with a single sgRNA on telomeric repeats (Figure 5.14.), we expected to resume the biotinylation event on telomeres. To test the biotinylation and the targeting efficiencies, we performed streptavidin-beads pulldown and immunoprecipitations performed by anti-FLAG and anti-Cas9 antibodies (Figure 5.16.). We observed that the targeting efficiency of BirN-dCas9 and BirC-dCas9 was confirmed by immunoprecipitation, and the biotinylation was activated through heterodimerization, which was proven by biotin pulldown performed with streptavidin beads (Figure 5.16.).

5.5.2. MAPK Promoter Targeting

We found that MAPK and HIRA promoter have high copy numbers, 11 and 9 copies, in K562 cell line (by canSAR V4.0) (Tym *et al.*, 2016). Increase in the copy number of a gene provides us with a higher number of starting material, as in telomeric repeats. In

canSAR V4.0 database (Tym *et al*., 2016)., we chose the genes with gain of function, and rank them based on the copy number.

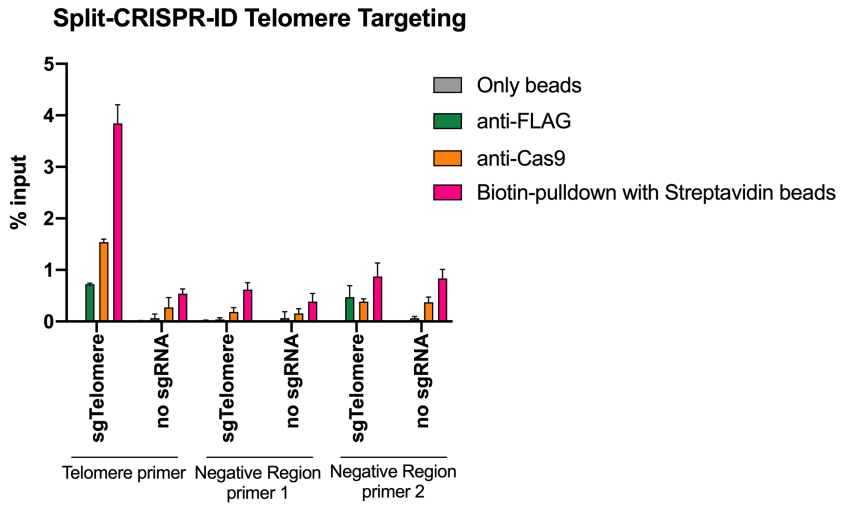


Figure 5.16. Upon 50 µM biotin treatment, reconstitution and activation of BirA* was tested using BirN-dCas9 and BirC-dCas9 fusion proteins expressed in HEK293FT cells. Streptavidin-ChIP performed at Telomeric repeats followed by qPCR. BirN-dCas9 and BirC-dCas9 are enriched and biotinylation is activated at Telomeric repeats compared to the negative regions. Error bars depict standard deviation of the mean.

We again cloned several sgRNA coding sequences into both pLKO5.sgRNA.EFS.GFP and lenti_gRNA-puro plasmids. In this cloning, we optimized a new colony PCR method (see Methods) to detect sgRNA sequences in single colonies obtained after transformation (see Appendix E Section). We used a universal U6-fwd primer and the specific reverse primer, which is the oligo of sgRNA itself. The only possibility of this PCR reaction to work was to have a sgRNA sequence inside (Figure 5.17.). We validated these colones by PCR (see Appendix E section) and Sanger sequencing (see Appendix B section).

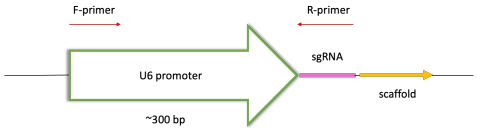


Figure 5.17. Schematic representation of the colony PCR reaction that we optimized to amplify U6 promoter in the presence of sgRNA.

We tested the efficiencies of subclones sgRNAs at MAPK promoter in K562 cells stable expressing FLAG-dCas9 and HA-dCpf1. We performed ChIP experiment by immunoprecipitation with anti-FLAG and anti-HA antibodies. We observed that these sgRNAs were also targeting their regions efficiently (Figure 5.18.).

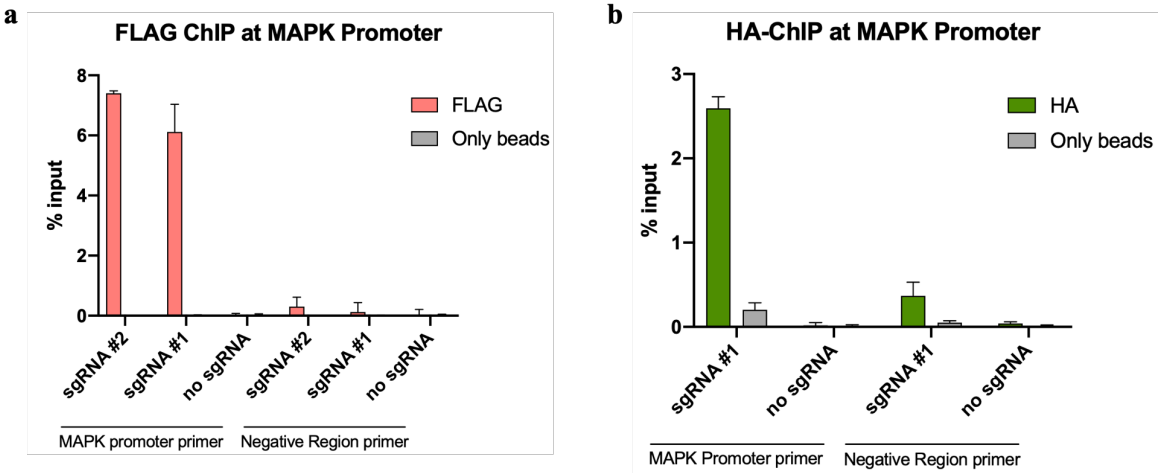


Figure 5.18. (a) anti-FLAG-ChIP performed at MAPK promoter in FLAG-dCas9 stable K562 cells. (b) anti-HA-ChIP performed at MAPK promoter in HA-dCpf1 stable K562 cells. dCas9 and dCpf1 are enriched at MAPK promoter compared to the negative region controls. Error bars depict standard deviation of the mean.

After confirming the targeting efficiencies of these sgRNAs, we tested *in vivo* biotinylation at this loci using BirN-dCas9 and BirC-dCpf1. We used single sgRNA targeting MAPK promoter as a negative control (Figure 5.19).

5.5.3. HIRA Promoter Targeting

The other gene with high copy number that we found using canSAR V4.0 in K562 cells was HIRA, which is a histone chaperon.

We cloned only one pair of sgRNAs into the sgRNA expression vectors for the dual-targeting of dCas9 and dCpf1 fusions. We confirmed the sequences of the clones by Sanger sequencing (see Appendix B section).

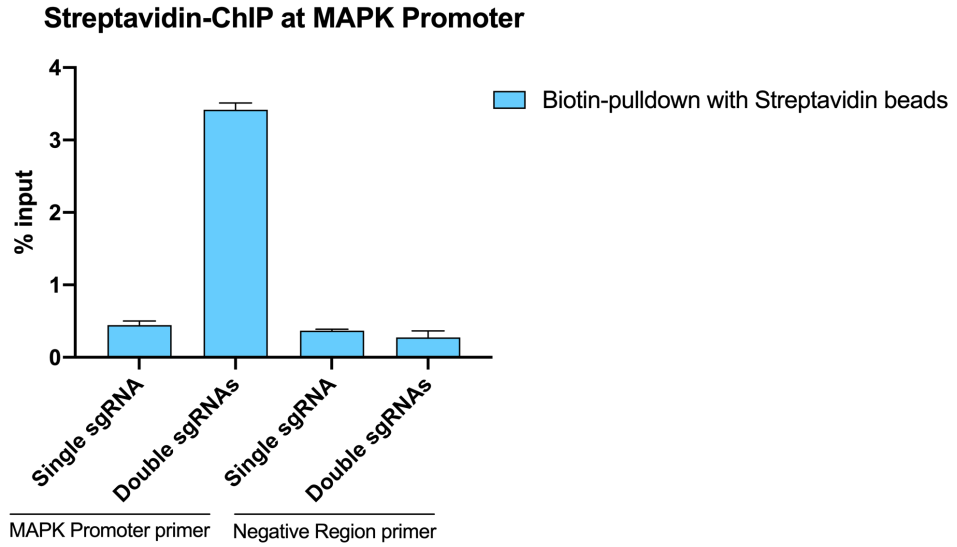


Figure 5.19. BirN-dCas9 and BirC-dCpf1 targeted to MAPK promoter in K562 cells. *In vivo* biotinylation is analyzed by Streptavidin-pulldown followed by qPCR. Biotinylation is activated and enriched at MAPK promoter with double sgRNA targeted samples, but not in single sgRNA and negative region controls. Error bars depict standard deviation of the mean.

We performed ChIP followed by qPCR to test the binding efficiencies of these sgRNAs to HIRA promoter in K562 cells stably expressing either FLAG-dCas9 or HA-dCpf1. The targeting of these proteins was successful (Figure 5.20).

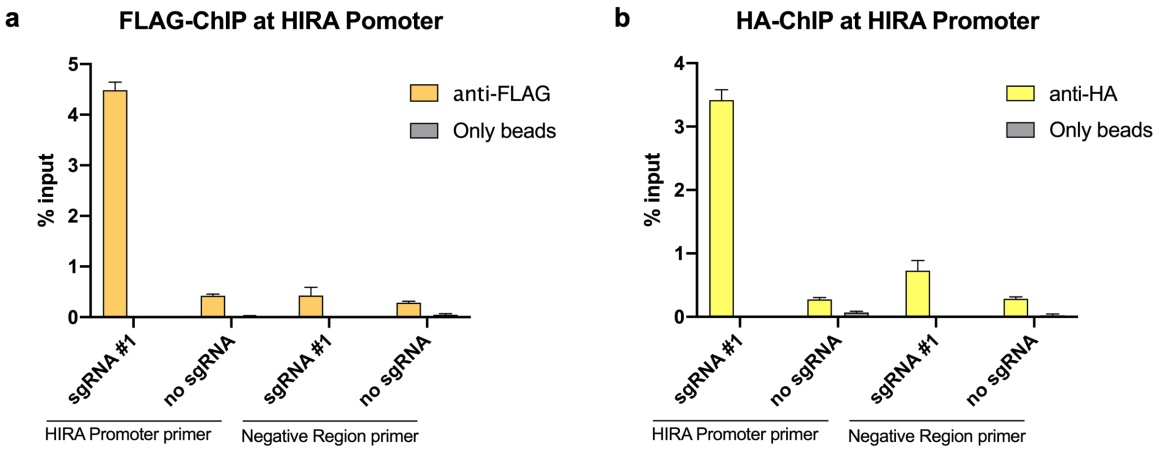


Figure 5.20. (a) anti-FLAG-ChIP is performed at HIRA promoter in K562 cells stably expressing 3xFLAG-dCas9. (b) anti-HA-ChIP is performed at HIRA promoter in K562 cells stably expressing 3xHA-dCpf1. dCas9 and dCpf1 are enriched at HIRA promoter compared to the negative region controls. Error bars depict standard deviation of the mean.

5.5.4. pS2/TFF1 Promoter Targeting

Estrogen plays a critical role in the initiation and progression of breast cancers (Kim *et al.*, 2000). Estrogen molecules activate gene expression by binding to intracellular estrogen receptor (ER), which dimerizes on estrogen response elements (EREs) (Kim et al., 2000). pS2/TFF1 is an estradiol inducible gene, which is expressed actively only in breast cancer cells (Rio *et al*., 1987). The promoter of this gene contains three EREs, and is a well-studied model to study estradiol-induced gene regulation with many transcriptional regulators known to interact, therefore providing us with test region to develop our method (Reid *et al*., 2009).

We chose well-studied ERE located at the -400 position as our target (Figure 5.21.). For this, we generated several sgRNAs for both dCas9 and dCpf1. Again, we used pLKO5.sgRNA.EFS.GFP for dCas9 targeting and lenti_gRNA-puro for dCpf1 targeting. We clones these sgRNAs either by using regular restriction digestion followed by ligation or by home-made Gibson assembly. The clones were validated by Sanger sequencing (see Appendix B section).

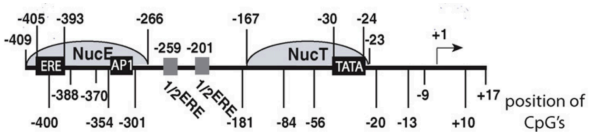


Figure 5.21. Schematic representation of pS2/TFF1 gene promoter. There are three estrogen response elements (EREs), one of which is protected by nucleosome binding. Upon estradiol treatment, nucleosomes are replaced and the transcription is activated (from Reid *et al*., 2009).

We tested the targeting efficiencies of gRNAs on pS2/TFF1 promoter using FLAG-dCas9 and HA-dCpf1 stable cells (Figure 5.12.). To do that, we performed ChIP using anti-FLAG and anti-HA antibodies, followed by qPCR (Figure 5.22.). Targeting a single-copy locus pS2/TFF1 promoter was efficient, when we compared the results with negative region amplification and only beads control (Figure 5.22.).

We targeted pS2/TFF1 promoter using BirN-dCas9 and BirC-dCpf1 fusion proteins to test *in vivo* biotinylation event. To test biotinylation, we performed Streptavidin-ChIP in MCF-7 cells stably expressing these fusion proteins and sgRNAs, which are sgRNA #6 with sgRNA #21, and utilized to transport BirA* halves to pS2 promoter. We used single sgRNA targeting pS2 promoter as a control (Figure 5.23.). Our preliminary data showed that targeting a single-copy locus pS2/TFF1 was efficient, and the *in vivo* biotinylation was successful (Figure 5.23.).

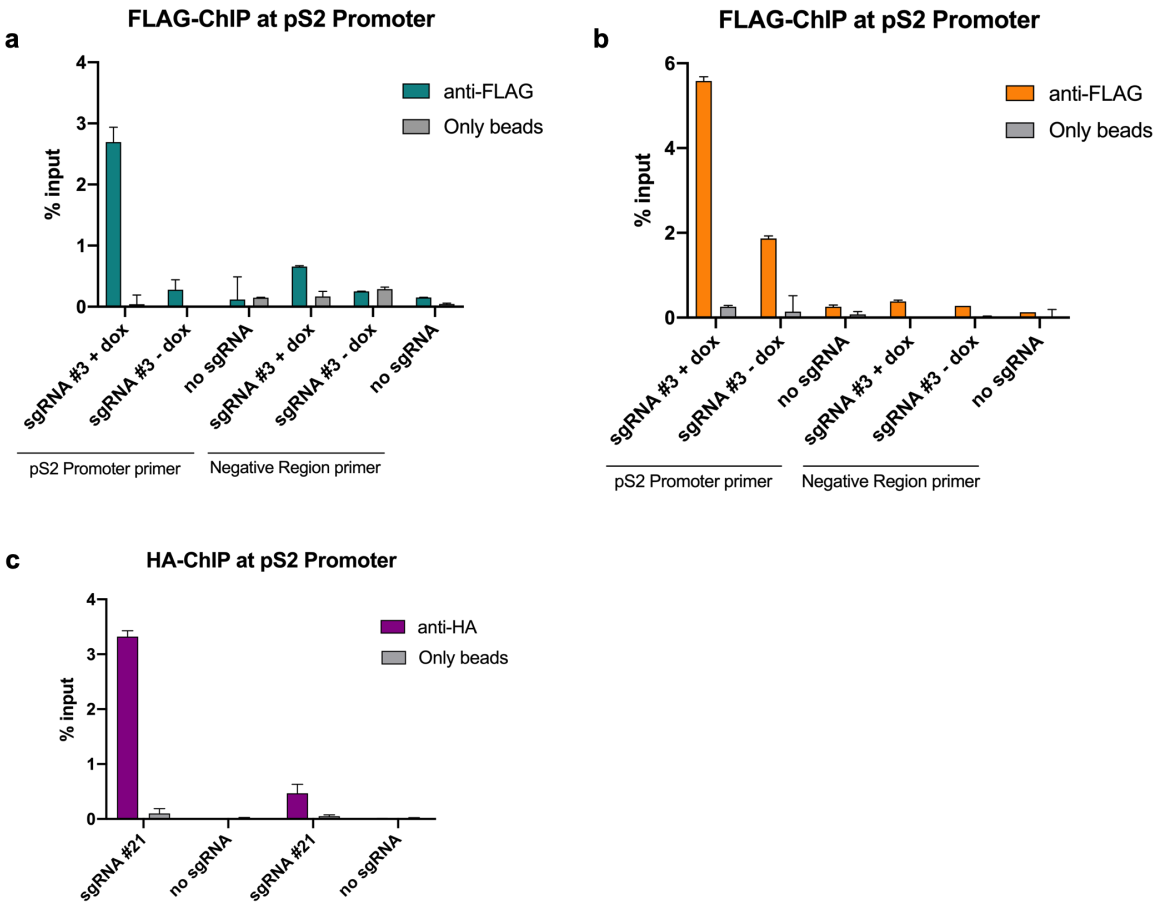


Figure 5.22. (a) anti-FLAG-ChIP performed at pS2 promoter in MCF-7 cells stably expressing FLAG-dCas9. (b) anti-FLAG-ChIP performed at pS2 promoter in MDA-MB-231 cells expressing FLAG-dCas9. (c) anti-HA-ChIP performed at pS2 promoter in MCF-7 cells expressing HA-dCpf1. dCas9 and dCpf1 are enriched at pS2/TFF1 promoter compared to the negative region controls. Error bars depict standard deviation of the mean.

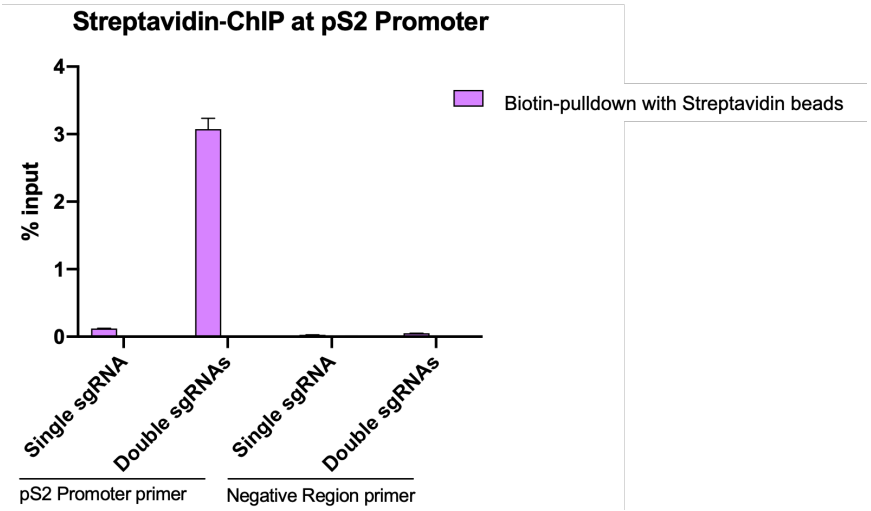


Figure 5.23. BirN-dCas9 and BirC-dCpf1 targeted to pS2 promoter in MCF-7 cells. *In vivo* biotinylation is analyzed by Streptavidin-ChIP followed by qPCR. Biotinylation is activated and enriched at pS2/TFF1 promoter with double sgRNA targeted samples, but not in single sgRNA and negative region controls. Error bars depict standard deviation of the mean.

5.6. Optimization of CAPTURE method

As a side project, we optimized CAPTURE method (Liu *et al.*, 2017) in our lab. This project mostly overcomes the non-specific biotinylation by using the BirA biotin ligase, which recognizes the biotin-acceptor-tags, yet it still has off target binding issue of dCas9, which needs to be considered. This method employs three components: biotin-tag carrying dCas9 (FB-dCas9), the original, non-promiscuous BirA biotin ligase and a specific sgRNA.

5.6.1. Optimization of the Electroporation Conditions for K562 Cell Line

dCas9 and BirA are introduced to the cells with mammalian expression plasmids, which are not suitable for the production of viruses (Liu *et al*., 2018). For that reason, we optimized electroporation of K562 cells by using a plasmid carrying a GFP marker, which we used to measure the transfection efficiency by using flow cytometry. The GFP signal

found higher than a threshold in K562 electroporated cells, when it was compared to the parental cells, K562 (Figure 5.24.).

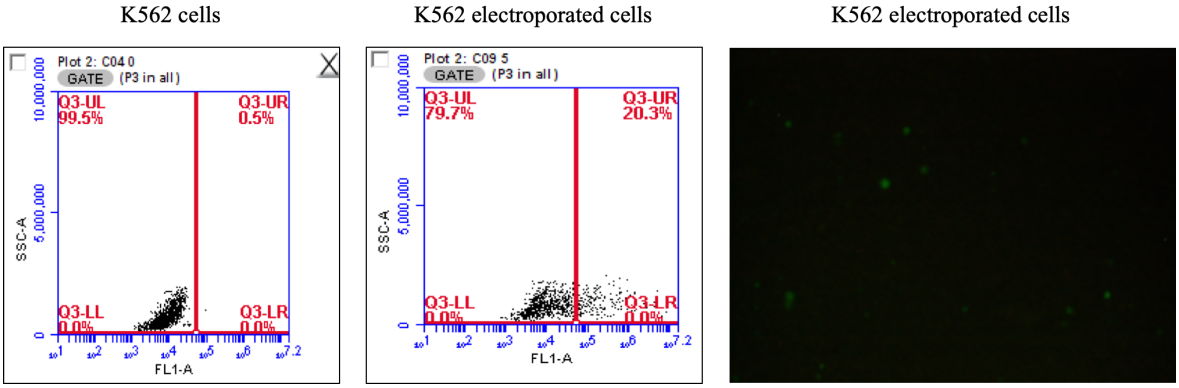


Figure 5.24. The electroporation conditions of K562 cells were optimized with GFP carrying plasmid. The intensities of GFP expressing cells were measured by flow cytometry. Fluorescent image is also indicated.

We applied the optimal electroporation conditions on K562 cells to create dCas9 with a biotin-acceptor-tag and BirA stable single cell-derived colony (see Methods). After electroporation, we seeded cells on a 96-well plate as 100 cells/well. After three days of electroporation, we treated cells with suitable antibiotics, puromycin and neomycin, to select single cell-derived colonies (see Methods).

5.6.2. Validation of the FB-dCas9 and BirA expression in K562 Cell Line

We observed 24 colonies after 4 weeks of treatment, and performed Western blot analysis to detect protein expressions (Figure 5.25.). We observed that some of the colonies are expressing only dCas9, only BirA or both of them. We selected the Colony 1 for further analysis and called is as CAPTURE-stable K562 cells (Figure 5.25.).

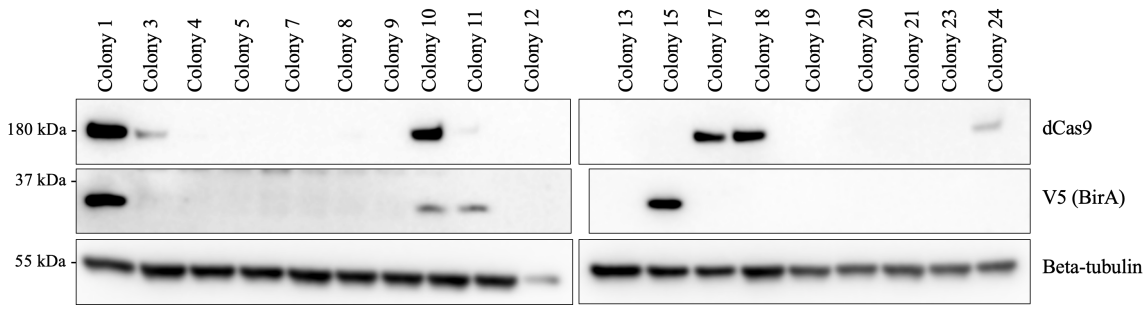


Figure 5.25. K562 dCas9-BirA stable cells were generated by electroporation followed by antibiotic selection, and protein expressions were confirmed by Western blotting.

5.6.3. CAPTURE-qPCR of FB-dCas9/BirA stable K562 Cell Line

We transduced the CAPTURE Colony 1 K562 cells with MAPK promoter targeting sgRNAs, and performed CAPTURE-ChIP-qPCR by pulling down dCas9 with anti-Cas9, anti-FLAG and Streptavidin-beads. We tested the ChIP DNA fragment by using MAPK specific primer pairs, and two negative region primers: 1 (RPS12 gene promoter) and 2 (XBP1 gene promoter) (Figure 5.26.).

5.6.4. CAPTURE-Western Blot of FB-dCas9/BirA stable K562 Cell Line

After validating CAPTURE method on MAPK promoter, we aimed to perform immunoblotting to confirm Telomeric-protein pulldown by sgTelomere targeting. We wanted to test the system by sgTelomere stable K562 cells, yet we had a problem with sgTelomere transduction. Thus, we transfected 1 x 10^8 HEK293FT cells with FB-dCas9, BirA, and sgTelomere. We fixed and collected the samples on the following day, and performed the western blot analysis (see Methods) (Figure 5.27a).

We also performed CAPTURE-western blot using FB-dCas9/BirA stable K562 Colony 1 cells (Figure 5.25.). In this experiment, we transduced these cells with MAPK targeting sgRNAs, and performed pulldown using Streptavidin beads. We used the cell line without sgRNA expression (no sgRNA) as a negative control (Figure 5.27b).

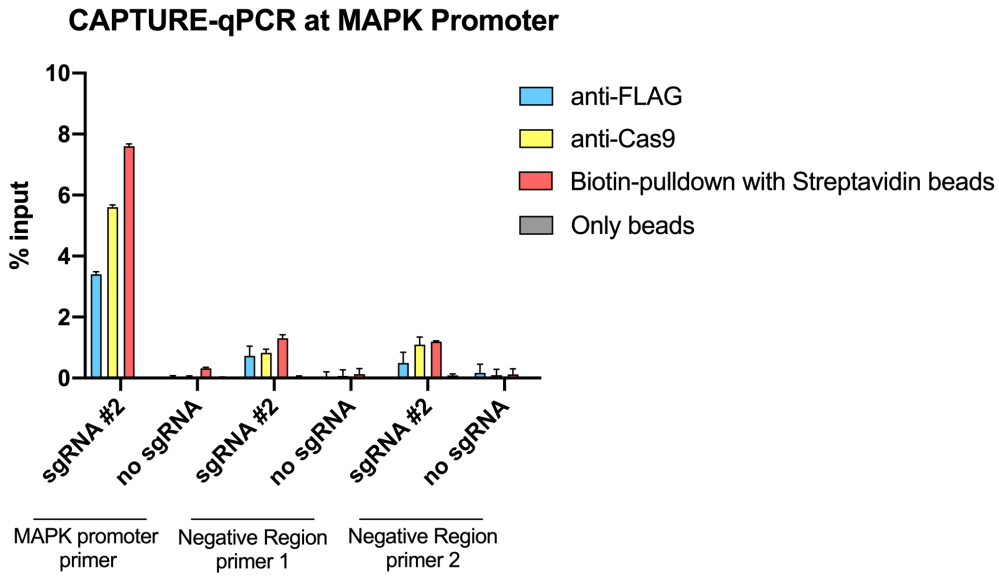


Figure 5.26. CAPTURE-qPCR performed with FB-dCas9 and BirA K562 cell line colony 1 at MAPK promoter. This experiment was performed with 3 different pulldown via FLAG antibody, Cas9 antibody and Streptavidin beads. Negative region 1: RPS12 gene promoter. Negative region 2: XBP1 gene promoter

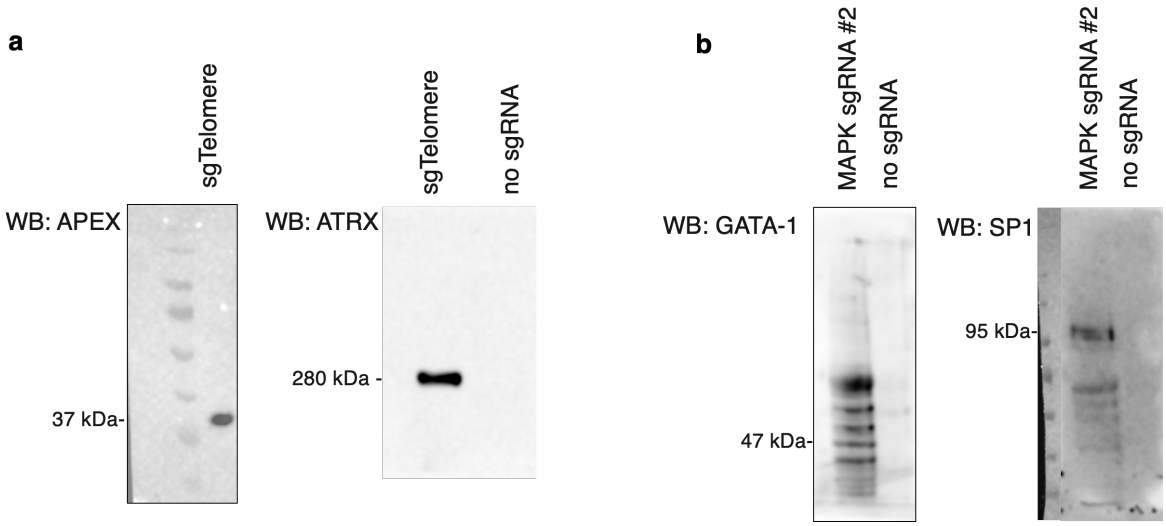


Figure 5.27. *In vivo* biotinylated dCas9 pulldown with Streptavidin beads followed by western blot. (a) Western blot shows APEX and ATRX enrichment in sgTelomere-targeted region. (b) Western blot shows GATA-1 and SP1 enrichment in MAPK-targeted region but not in no sgRNA control.

6. DISCUSSION

Mistargeting of transcriptional regulators and their malfunctions cause disease including diabetes, autoimmunity, developmental problems, and cancer (Chatterjee & Ahituv, 2017). Thus, identification of these regulators at critical genomic loci is a prerequisite to further research that aims to disclose their functional nature (Cazaly *et al*., 2015). There are several methods to isolate chromatin-associated factors; however, current methods that are based on ChIP-MS present difficulties in focusing on a specific locus, as it is more challenging to "label" and selectively isolate a genomic region of interest compared to a protein (Soldi & Bonaldi, 2014; Wierer & Mann, 2016). The CRISPR/Cas9 gene editing platform has been adapted to overcome this problem (Fujita & Fujii, 2013; Schmidtmann *et al*., 2016). The modified platform employs a sgRNA molecule that can escort to a locus of interest catalytically inactive variants of the CRISPR proteins, which can be isolated using immunoprecipitation or affinity pulldown (Fujita & Fujii, 2013; Schmidtmann *et al*., 2016). As the target locus also comes down along with the CRISPR protein, its protein interactors can be analyzed in the subsequent steps using standard techniques such as Western blotting and MS (Liu *et al.*, 2017).

However, all current methods have their own limitations that need to be overcome to prevent false positive proteomic identifications and increases the extent of discovery (Wierer & Mann, 2016). In order to overcome such limitations, we aimed to develop a novel method, which can combine the power of the preceding methods and go beyond those in terms of specificity and sensitivity. We decided that the most promising isolation strategy would involve the cargo delivery feature of CRISPR proteins along with affinity purification based on the biotin-streptavidin interaction. A major limitation of the previous CRISPR-based protein purification methods was off-target binding of these nucleases to undesired loci (Dominguez *et al*., 2016; Schmidtmann *et al*., 2016; Zhang *et al*., 2017) (Figure 5.1.). To solve this issue, we decided to use dual-targeting by distinct catalytically-inactive CRISPR proteins, dCas9 and dCpf1 (Figure 5.3.). The combinatorial implementation of the two CRISPR proteins will serve to drastically reduce the off-target identifications, as the

activation of the system depends on both proteins docking to their target sites which was demonstrated in other contexts including gene editing (Guilinger *et al*., 2014).

Another component of the previous methods was *in vivo* biotinylation, which requires the expression of bacterial biotin ligases such as BirA* (He & Pu, 2010; Kim *et al*., 2009). Another limitation of the previous methods was that the constitutive activity of this protein can lead to promiscuous labeling of all proteins within 10-30 nm range, rendering proper identification of desired interactions difficult due to high background signal (Roux *et al.,* 2012). Two groups demonstrated that the split version of BirA* solves the non-specific biotinylation problem (De Munter *et al.*, 2017; Schopp *et al.*, 2017) (Figure 5.4.). In this study, we propose a novel method that employs two distinct CRISPR proteins fused with the N-terminal and C-terminal halves of BirA* (Figure 5.3.). These CRISPR proteins are targeted to a genomic region of interest with the help of designer sgRNA molecules, leading to reconstitution of the intact and functional biotin ligase that can label the interacting partners in the vicinity (Figure 5.3). By using this split system whose activation requires two separate binding events, we are aiming to confine the labeling radius so that only the true partners will be identified.

As part of the optimization process, we designed two separate sets of proteins, where the N-terminus and the C-terminus of BirA* is alternately linked to either dCas9 or dCpf1. The resulting product sets were named BirN-dCas9 and BirC-dCpf1, and BirN-dCpf1 and BirC-dCas9 (Figure 5.7.). We devised a cloning strategy to generate the required constructs, which included the two fusion proteins that are carried in different lentiviral plasmids with different antibiotic-selectable markers. We chose a doxycycline-inducible promoter to control the timing of their expression, which will enable us to indirectly control the timing of their binding to the DNA. After verifying that the cloning was successful, we created stable cell lines expressing our fusion proteins and assessed the protein levels via Western blotting (Figure 5.8.).

The CRISPR proteins used in this study, dCas9 and dCpf1, recognize distinct PAM sequences and interact with sgRNAs that possess different secondary structures known as the gRNA scaffold (Zetsche *et al*., 2015; Zhang *et al*., 2015). For this reason, we had to

design a separate sgRNA for each CRISPR protein. One consideration is that, while the individual PAM sequences for each of these proteins are ubiquitous (NGG for Cas9 and TTTN for Cpf1) (Zetsche *et al*., 2015; Zhang *et al*., 2015), both sequences need to be present within a targeted genomic region for this system to work (Figure 5.3.). The short length of these sequences makes it possible to find them and design the sgRNA molecules accordingly on many target loci, including those that we are planning to study. However, as the relative locations of the PAM sequences and the docking sites vary, the lengths of the amino acid linkers that connect the BirA* fragments may need to be adjusted correspondingly, so that the biotin ligase can be successfully reconstituted at the desired locus (Figure 5.11). Our system is flexible in that this adjustment can be made relatively easily, as we can change the linker length through a simple PCR reaction to target a wide number of loci with different spacer needs. In our experiments, we used the PyMOL software to calculate these relative distances and estimate the proper linker length for our target loci (Figures 5.9. and 5.10.).

To validate the system within a cellular setting, we chose four different target loci on the chromatin. Our first target was telomeric repeats, as they are short and extensive stretches (5'-TTAGGG-3') that contain the Cas9 PAM sequence (NGG) (Chen *et al*., 2013; Schmidtmann *et al*., 2016) (Figure 5.14.). We aim to use this region as control by targeting with only dCas9 proteins that are linked to the two BirA* halves. This serves to maximize the possibility of dCas9 binding and yields us the ideal conditions to test whether the BirA* fragments are being reconstituted properly and creating the functional BirA* protein. As the telomeric repeats enable a great number of binding events (Figure 5.14.), we had a clear read-out showing high levels of biotinylation compared to background levels indicating that the reconstitution process was successful (Figure 5.16.). The second region of interest for us is the pS2/TFF1 promoter in breast cancer cells due to its high activity in cancer cells compared to normal mammary cells (Kim *et al.*, 2000; Reid *et al.*, 2009). This promoter was well-characterized along with its interacting partners in previous mass spectrometry analyses by other groups (Kim *et al.*, 2000; Reid *et al.*, 2009; Jeong *et al.*, 2012). Our other two target regions are MAPK and HIRA promoters, which we chose due to their high copy numbers in K562 leukemia cells. We designed several sgRNAs for both dCas9 and dCpf1 targeting these regions. Therefore, it serves as a positive control for us to see whether our system is able to identify the expected interactors that were reported in the literature at a single-copy locus.

Another part of the optimization process is verifying that our sgRNA constructs are able to escort the CRISPR proteins to the desired loci. To do this, we performed ChIP followed by qPCR analysis after targeting our four target regions using dCas9 and dCpf1. Our results confirmed that our sgRNA molecules can successfully target all four of our targets and can be used for further experiments (Figures 5.15., 5.18., 5.20., and 5.22.). We were able to show that the BirA* can reconstitute at three different loci, Telomere repeats, MAPK and pS2/TFF1 promoters, and resume *in vivo* biotinylation (Figures 5.16., 5.19. and 5.23). As a further step, we plan to test our method with different sgRNA sets, and prove that the system works for the identification of locus-specific regulators. To do that, we will perform western blot analysis with known proteins, and move forward with the mass spectrometry to discover new ones.

As a side project, we also attempted to optimize the CAPTURE protocol in our laboratory. This method employs three components: a biotin-tag carrying dCas9 (FB-dCas9), the BirA biotin ligase and a specific sgRNA molecule (Liu *et al.*, 2018, 2017). While off-target binding remains a concern, this method overcomes the non-specific biotinylation by only attaching biotin to the dCas9 on its biotin-tag. Research group that has developed the method showed that the method was successful when targeting the telomeric regions (Liu *et al.*, 2018, 2017). We tested this method on the MAPK promoter in K562 cells (Figure 5.26.). Our results confirmed that the method can also be applied to lower copy number regions (Figure 5.26.). For the detection of proteins, we performed Western blot analysis with sgTelomere targeted dCas9 isolation. We were able to isolate APEX and ATRX proteins by streptavidin-bead pulldown in sgRNA targetted cells, but not in no sgRNA negative control (Figure 5.27a). This result confirms that we were able to detect the isolated proteins on target loci via western blot. We performed streptavidin-beads pulldown with K562 Colony 1 cells to validate the system. We targeted MAPK promoter, and were able to detect GATA-1 and SP1 proteins in our western blot result (Figure 5.27b).

In this project, we attempt to establish the Split-CRISPR-ID method to isolate locus-specific chromatin-associated factors. As part of this, we successfully generated the necessary fusion proteins and validated their expression by immunoblotting. We selected four different regions to test our method and confirmed the ability of our sgRNA constructs

to efficiently target them and specific reconstitution of BirA* activity. Moving forward, we plan to validate our method by performing further experiments on these target loci. We also successfully isolated some of the known chromatin-associated proteins by CAPTURE method, and detected them with western blotting. The next step is to purify the proteins that are bound on our target genomic regions, and identify the novel factors with mass spectrometry. After this, we will be able to use it to identify novel regulators for disease-associated genes.

REFERENCES

Alekseyenko, A. A., McElroy, K. A., Kang, H., Zee, B. M., Kharchenko, P. V, & Kuroda, M. I. (2015). BioTAP-XL: Cross-linking/Tandem Affinity Purification to Study DNA Targets, RNA, and Protein Components of Chromatin-Associated Complexes. *Current Protocols in Molecular Biology*, *109*, 21.30.1-32.

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation And Methylation Of Histones And Their Possible Role In The Regulation Of RNA Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(5), 786–794.

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., & Hannon, G. J. (2007). Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science*, *316*(5825), 744–747.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., … Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, *129*(4), 823–837.

Böhmdorfer, G., & Wierzbicki, A. T. (2015). Control of Chromatin Structure by Long Noncoding RNA. *Trends in Cell Biology*, *25*(10), 623–632.

Cairns, B. R. (2001). Emerging roles for chromatin remodeling in cancer biology. *Trends in Cell Biology*, *11*(11), S15-21.

Carullo, N. V. N., Day, J. J., Carullo, N. V. N., & Day, J. J. (2019). Genomic Enhancers in Brain Health and Disease. *Genes*, *10*(1), 43.

Cawthon, R. M. (2002). Telomere measurement by quantitative PCR. *Nucleic Acids Research*, *30*(10), e47.

Cazaly, E., Charlesworth, J., Dickinson, J. L., & Holloway, A. F. (2015). Genetic Determinants of Epigenetic Patterns: Providing Insight into Disease. *Molecular Medicine (Cambridge, Mass.)*, *21*(1), 400–409.

Cevher, M. A., Shi, Y., Li, D., Chait, B. T., Malik, S., & Roeder, R. G. (2014). Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nature Structural & Molecular Biology*, *21*(12), 1028–1034.

Chatterjee, S., & Ahituv, N. (2017). Gene Regulatory Elements, Major Drivers of Human Disease. *Annual Review of Genomics and Human Genetics*, *18*(1), 45–63.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., … Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, *155*(7), 1479–1491.

Chen, X., Zaro, J. L., & Shen, W.-C. (2013). Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews*, *65*(10), 1357–1369.

De Munter, S., Görnemann, J., Derua, R., Lesage, B., Qian, J., Heroes, E., … Bollen, M. (2017). Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions. *FEBS Letters*, *591*(2), 415–424.

Deans, C., & Maggert, K. A. (2015). What do you mean, "epigenetic"? *Genetics*,

199(4), 887–896.

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–1022.

Déjardin, J., & Kingston, R. E. (2009). Purification of Proteins Associated with Specific Genomic Loci. *Cell*, *136*(1), 175–186.

Dominguez, A. A., Lim, W. A., & Qi, L. S. (2016). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature Reviews Molecular Cell Biology*, *17*(1), 5–15.

Engelkamp, D., & Heyningen, V. van. (1996). Transcription factors in disease. *Current Opinion in Genetics & Development*, *6*(3), 334–342.

Farrelly, L. A., Thompson, R. E., Zhao, S., Lepack, A. E., Lyu, Y., Bhanu, N. V., … Maze, I. (2019). Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature*, *567*(7749), 535–539.

Fernandes, J., Acuña, S., Aoki, J., Floeter-Winter, L., Muxel, S., Fernandes, J. C. R., … Muxel, S. M. (2019). Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Non-Coding RNA*, *5*(1), 17.

Fraga, M. F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., … Esteller, M. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature Genetics*, *37*(4), 391–400.

Fujita, T., Asano, Y., Ohtsuka, J., Takada, Y., Saito, K., Ohki, R., & Fujii, H. (2013). Identification of telomere-associated molecules by engineered DNA-binding molecule-

mediated chromatin immunoprecipitation (enChIP). *Scientific Reports*, *3*, 3171.

Fujita, T., & Fujii, H. (2011). Direct identification of insulator components by insertional chromatin immunoprecipitation. *PloS One*, *6*(10), e26109.

Fujita, T., & Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochemical and Biophysical Research Communications*, *439*(1), 132–136.

Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., & Kouzarides, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *The Journal of Biological Chemistry*, *278*(6), 4035–4040.

Gade, P., & Kalvakolanu, D. V. (2012). Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. *Methods in Molecular Biology (Clifton, N.J.)*, *809*, 85–104.

Gaffney, D. J., McVicker, G., Pai, A. A., Fondufe-Mittendorf, Y. N., Lewellen, N., Michelini, K., … Pritchard, J. K. (2012). Controls of Nucleosome Positioning in the Human Genome. *PLoS Genetics*, *8*(11), e1003036.

Garrett-Bakelman, F. E., Darshi, M., Green, S. J., Gur, R. C., Lin, L., Macias, B. R., … Turek, F. W. (2019). The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. *Science (New York, N.Y.)*, *364*(6436), eaau8650.

Gibcus, J. H., & Dekker, J. (2012). The context of gene expression regulation. *F1000 Biology Reports*, *4*, 8.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345.

Guilinger, J. P., Thompson, D. B., & Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnology*, *32*(6), 577.

Halusková, J. (2010). Epigenetic studies in human diseases. *Folia Biologica*, *56*(3), 83–96.

Han, J., Kim, D., & Morris, K. V. (2007). Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proceedings of the National Academy of Sciences*, *104*(30), 12422–12427.

Hanly, D. J., Esteller, M., & Berdasco, M. (2018). Interplay between long non-coding RNAs and epigenetic machinery: emerging targets in cancer? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1748).

Hao, N., Shearwin, K. E., & Dodd, I. B. (2019). Positive and Negative Control of Enhancer-Promoter Interactions by Other DNA Loops Generates Specificity and Tunability. *Cell Reports*, *26*(9), 2419-2433.e3.

Hargreaves, D. C., & Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Research*, *21*(3), 396–420.

He, A., & Pu, W. T. (2010). Genome-wide location analysis by pull down of in vivo biotinylated transcription factors. *Current Protocols in Molecular Biology*, *Chapter 21*, Unit 21.20.

Heckl, D., Kowalczyk, M. S., Yudovich, D., Belizaire, R., Puram, R. V, McConkey, M. E., … Ebert, B. L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnology*, *32*(9), 941–946.

Hoffman, E. A., Frey, B. L., Smith, L. M., & Auble, D. T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. *The Journal of Biological Chemistry*, *290*(44), 26404–26411.

Hoshino, A., & Fujii, H. (2009). Insertional chromatin immunoprecipitation: A method for isolating specific genomic regions. *Journal of Bioscience and Bioengineering*, *108*(5), 446–449.

Hubner, N. C., Nguyen, L. N., Hornig, N. C., & Stunnenberg, H. G. (2015). A Quantitative Proteomics Tool To Identify DNA–Protein Interactions in Primary Cells or Blood. *Journal of Proteome Research*, *14*(2), 1315–1329.

Inoue, K., & Fry, E. A. (2017). Haploinsufficient tumor suppressor genes. *Advances in Medicine and Biology*, *118*, 83–122.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33*(S3), 245–254.

Jin, B., Li, Y., & Robertson, K. D. (2011). DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? *Genes & Cancer*, *2*(6), 607.

Johnson, W. L., & Straight, A. F. (2017). RNA-mediated regulation of heterochromatin. *Current Opinion in Cell Biology*, *46*, 102–109.

Jones, P. A., & Gonzalgo, M. L. (1997). Altered DNA methylation and genome instability: a new pathway to cancer? *Proceedings of the National Academy of Sciences of the United States of America*, *94*(6), 2103–2105.

Kaikkonen, M. U., Lam, M. T. Y., & Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research*, *90*(3), 430–440.

Kappei, D., Butter, F., Benda, C., Scheibe, M., Draškovič, I., Stevense, M., … Buchholz, F. (2013). HOT1 is a mammalian direct telomere repeat-binding protein contributing to telomerase recruitment. *The EMBO Journal*, *32*(12), 1681–1701.

Kim, H. K., Song, M., Lee, J., Menon, A. V., Jung, S., Kang, Y.-M., … Kim, H. (2017). In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nature Methods*, *14*(2), 153–159.

Kim, J. K., Samaranayake, M., & Pradhan, S. (2009). Epigenetic mechanisms in mammals. *Cellular and Molecular Life Sciences : CMLS*, *66*(4), 596–612.

Kim, J, Petz, L. N., Ziegler, Y. S., Wood, J. R., Potthoff, S. J., & Nardulli, A. M. (2000). Regulation of the estrogen-responsive pS2 gene in MCF-7 human breast cancer cells. *The Journal of Steroid Biochemistry and Molecular Biology*, *74*(4), 157–168.

Kim, Jonghwan, Cantor, A. B., Orkin, S. H., & Wang, J. (2009). Use of in vivo biotinylation to study protein–protein and protein–DNA interactions in mouse embryonic stem cells. *Nature Protocols*, *4*(4), 506–517.

Kim, Jongsook, Petz, L. N., Ziegler, Y. S., Wood, J. R., Potthoff, S. J., & Nardulli, A. M. (2000). Regulation of the estrogen-responsive pS2 gene in MCF-7 human breast cancer cells. *The Journal of Steroid Biochemistry and Molecular Biology*, *74*(4), 157–168.

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., … Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182–187.

King, A. D., Huang, K., Rubbi, L., Wang, Y., Pellegrini, M., & Correspondence, G. F. (2016). Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Reports*, *17*, 289–302.

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, *20*(4), 207–220.

Kolodziej, K. E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., & Strouboulis, J. (2009). Optimal use of tandem biotin and V5 tags in ChIP assays. *BMC Molecular Biology*, *10*, 6.

Kondo, Y. (2009). Epigenetic cross-talk between DNA methylation and histone modifications in human cancers. *Yonsei Medical Journal*, *50*(4), 455–463.

Kulis, M., & Esteller, M. (2010). DNA Methylation and Cancer. In *Advances in genetics* (Vol. 70, pp. 27–56).

Lai, W. K. M., & Pugh, B. F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nature Reviews. Molecular Cell Biology*, *18*(9), 548–562.

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, *152*(6), 1237–1251.

Liu, X., Zhang, Y., Chen, Y., Li, M., Shao, Z., Zhang, M. Q., & Xu, J. (2018). CAPTURE:

In Situ Analysis of Chromatin Composition of Endogenous Genomic Loci by Biotinylated dCas9. *Current Protocols in Molecular Biology*, *123*(1), e64.

Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., … Xu, J. (2017). In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*, *170*(5), 1028-1043.e19.

Lucchesi, J. C., Kelly, W. G., & Panning, B. (2005). Chromatin Remodeling in Dosage Compensation. *Annual Review of Genetics*, *39*(1), 615–651.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260.

Manelyte, L. (2017). Chromatin remodelers, their implication in cancer and therapeutic potential. In *J Rare Dis Res Treat* (Vol. 2).

Milne, T. A., Zhao, K., & Hess, J. L. (2009). Chromatin immunoprecipitation (ChIP) for analysis of histone modifications and chromatin-associated proteins. *Methods in Molecular Biology (Clifton, N.J.)*, *538*, 409–423.

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *38*(1), 23–38.

Olson, M. V. (1993). The human genome project. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(10), 4338–4344.

Patil, V. S., Zhou, R., & Rana, T. M. (2014). Gene regulation by non-coding RNAs. *Critical Reviews in Biochemistry and Molecular Biology*, *49*(1), 16–32.

Rafiee, M.-R., Girardot, C., Sigismondo, G., & Krijgsveld, J. (2016). Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Molecular Cell*, *64*(3), 624–635.

Reid, G., Gallais, R., & Métivier, R. (2009). Marking time: The dynamic role of chromatin and covalent modification in transcription. *The International Journal of Biochemistry & Cell Biology*, *41*(1), 155–163.

Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., & Yao, X. (2009). DOG 1.0: illustrator of protein domain structures. *Cell Research*, *19*(2), 271–273.

Rio, M. C., Bellocq, J. P., Gairard, B., Rasmussen, U. B., Krust, A., Koehl, C., … Chambon, P. (1987). Specific expression of the pS2 gene in subclasses of breast cancers in comparison with expression of the estrogen and progesterone receptors and the oncogene ERBB2. *Proceedings of the National Academy of Sciences*, *84*(24), 9243–9247.

Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics*, *6*(8), 597–610.

Robertson, K. D., & A.Jones, P. (2000). DNA methylation: past, present and future directions. *Carcinogenesis*, *21*(3), 461–467.

Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology*, *196*(6), 801–810.

Schiano, C., Casamassimi, A., Rienzo, M., de Nigris, F., Sommese, L., & Napoli, C. (2014). Involvement of Mediator complex in malignancy. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1845*(1), 66–83.

Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F., Leonhardt, H., Schmidtmann, E., … Leonhardt, H. (2016). Determination of local chromatin composition by CasID. *Nucleus*, *7*(5), 476–484.

Scholze, H., & Boch, J. (2010). TAL effector-DNA specificity. *Virulence*, *1*(5), 428–432.

Schopp, I. M., Amaya Ramirez, C. C., Debeljak, J., Kreibich, E., Skribbe, M., Wild, K., & Béthune, J. (2017). Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nature Communications*, *8*(1), 15690.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., … Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development*, *18*(11), 1251–1262.

Seidman, J. G., & Seidman, C. (2002). Transcription factor haploinsufficiency: when half a loaf is not enough. *The Journal of Clinical Investigation*, *109*(4), 451–455.

Shahid, Z., Simpson, B., & Singh, G. (2019). Genetics, Histone Code. In *StatPearls*. StatPearls Publishing.

Shen, L., Shi, Q., & Wang, W. (2018). Double agents: genes with both oncogenic and tumor-suppressor functions. *Oncogenesis*, *7*(3), 25.

Soldi, M., & Bonaldi, T. (2014). The ChroP Approach Combines ChIP and Mass Spectrometry to Dissect Locus-specific Proteomic Landscapes of Chromatin. *Journal of Visualized Experiments*, (86).

Soleimani, V. D., Palidwor, G. A., Ramachandran, P., Perkins, T. J., & Rudnicki, M. A.

(2013). Chromatin tandem affinity purification sequencing. *Nature Protocols*, *8*(8), 1525–1534.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology*, *19*(4), 262–274.

Sutherland, J. E., & Costa, M. (2003). Epigenetics and the Environment. *Annals of the New York Academy of Sciences*, *983*(1), 151–160.

Svoronos, A. A., Engelman, D. M., & Slack, F. J. (2016). OncomiR or Tumor Suppressor? The Duplicity of MicroRNAs in Cancer. *Cancer Research*, *76*(13), 3666–3670.

Todeschini, A.-L., Georges, A., & Veitia, R. A. (2014). Transcription factors: specific DNA binding and specific gene regulation. *Trends in Genetics*, *30*(6), 211–219.

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., … Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology*, *32*(6), 569–576.

Turner, B. (2001). ChIP with Native Chromatin: Advantages and Problems Relative to Methods Using Cross-Linked Material. In *Mapping Protein/DNA Interactions by Cross-Linking*. Institut national de la santé et de la recherche médicale.

Tym, J. E., Mitsopoulos, C., Coker, E. A., Razaz, P., Schierz, A. C., Antolin, A. A., & Al-Lazikani, B. (2016). canSAR: an updated cancer research and drug discovery knowledgebase. *Nucleic Acids Research*, *44*(D1), D938–D943.

Veitia, R. A., Bottani, S., & Birchler, J. A. (2013). Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends in Genetics*, *29*(7), 385–393.

Wei, Y., Xia, W., Zhang, Z., Liu, J., Wang, H., Adsay, N. V, … Hung, M.-C. (2008). Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. *Molecular Carcinogenesis*, *47*(9), 701–706.

Wierer, M., & Mann, M. (2016). Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Human Molecular Genetics*, *25*(R2), R106–R114.

Wolffe, A. P. (2001). Chromatin remodeling: why it is important in cancer. *Oncogene*, *20*(24), 2988–2990.

Won Jeong, K., Chodankar, R., Purcell, D. J., Bittencourt, D., & Stallcup, M. R. (2012). Gene-specific patterns of coregulator requirements by estrogen receptor-α in breast cancer cells. *Molecular Endocrinology (Baltimore, Md.)*, *26*(6), 955–966.

Wu, W. H., & Hampsey, M. (1999). Transcription: Common cofactors and cooperative recruitment. *Current Biology : CB*, *9*(16), R606-9.

Wunderlich, Z., & Mirny, L. A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. *Trends in Genetics*, *25*(10), 434–440.

Zaratiegui, M., Irvine, D. V., & Martienssen, R. A. (2007). Noncoding RNAs and Gene Silencing. *Cell*, *128*(4), 763–776.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., … Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, *163*(3), 759–771.

Zhang, L., Zhang, K., Prändl, R., & Schöffl, F. (2004). Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. *Biochemical and Biophysical*

Research Communications, *322*(3), 705–711.

Zhang, S., Zhou, B., Wang, L., Li, P., Bennett, B. D., Snyder, R., … Hu, G. (2017). INO80 is required for oncogenic transcription and tumor growth in non-small cell lung cancer. *Oncogene*, *36*(10), 1430–1439.

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids*, *4*, e264.

Zhang, Y., Qian, L., Wei, W., Wang, Y., Wang, B., Lin, P., … Lou, C. (2017). Paired Design of dCas9 as a Systematic Platform for the Detection of Featured Nucleic Acid Sequences in Pathogenic Strains. *ACS Synthetic Biology*, *6*(2), 211–216.

Zhou, B., Wang, L., Zhang, S., Bennett, B. D., He, F., Zhang, Y., … Hu, G. (2016). INO80 governs superenhancer-mediated oncogenic transcription and tumor growth in melanoma. *Genes & Development*, *30*(12), 1440–1453.

APPENDIX A: VECTORS

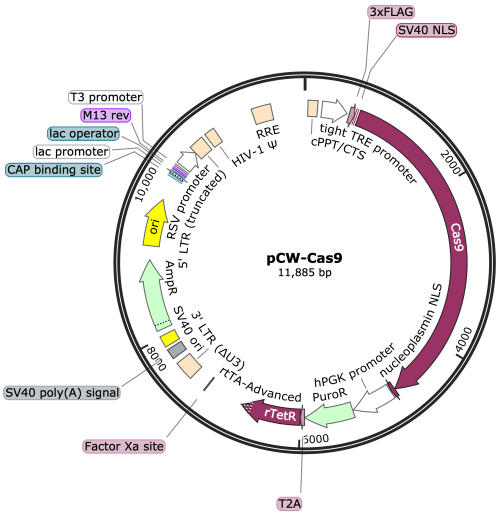


Figure A.1. Vector map of pCW-Cas9 (Addgene: 50661) plasmid that is used to generate pCW-dCas9 plasmid (from Wang *et al*., 2014).

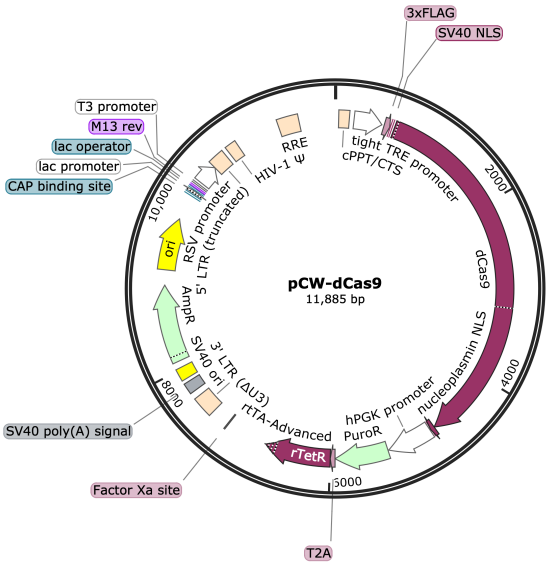


Figure A.2. Vector map of pCW-dCas9 plasmid that is used to generate 3X FLAG-dCas9 stable breast cancer cell lines (Ogmen A., unpublished).

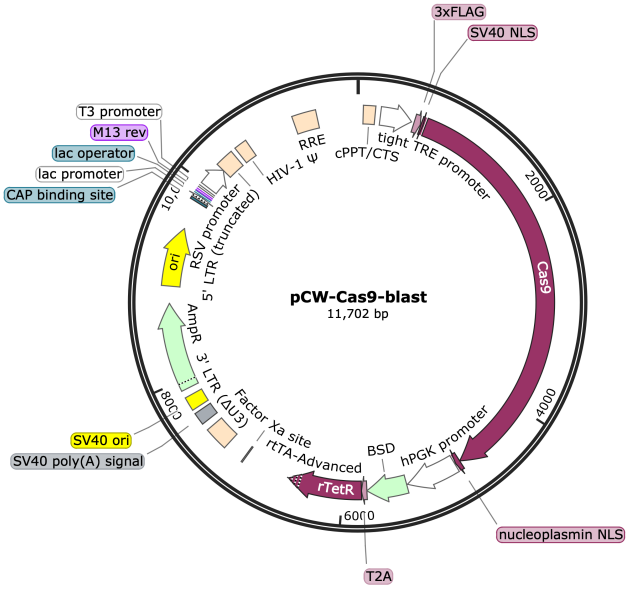


Figure A.3. Vector map of pCW-Cas9-blast (Addgene: 83481) plasmid that is used to generate pCW-dCpf1-blast plasmid (Habashy *et al*., unpublished).

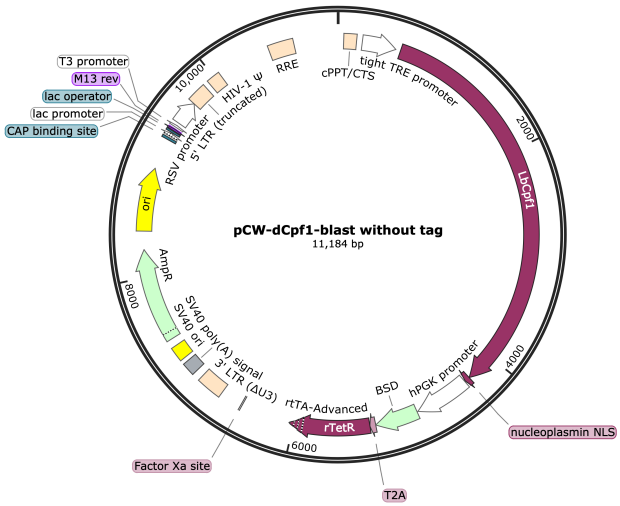


Figure A.4. Vector map of pCW-dCpf1-blast (no HA tag) plasmid that is used to generate pCW-dCpf1-blast plasmid.

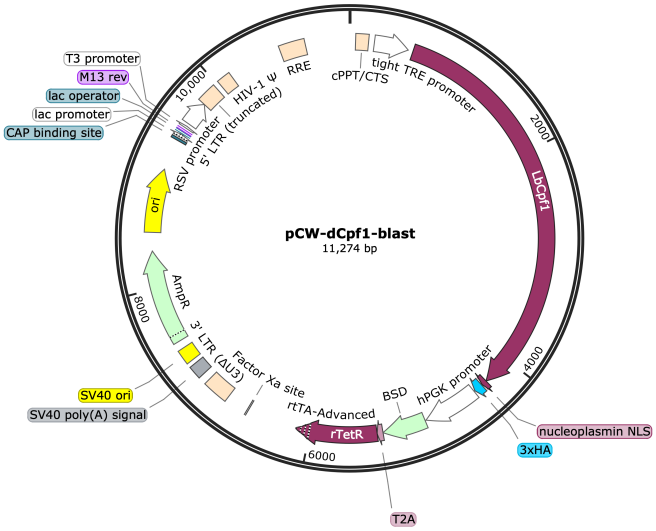


Figure A.5. Vector map of pCW-dCpf1-blast plasmid carrying a 3xHA tag that is used to generate dCpf1-3xHA stable cancer cell lines.

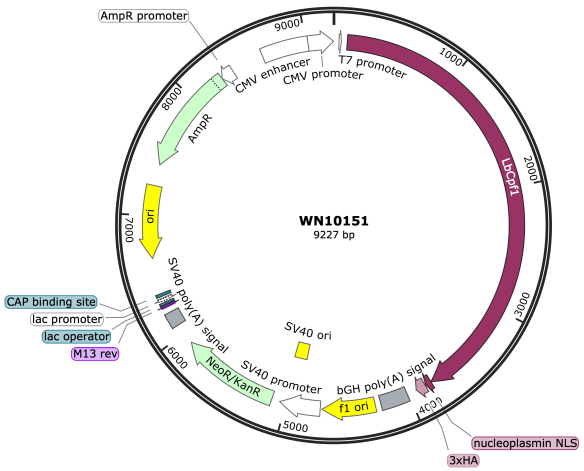


Figure A.6. Vector map of WN10151 plasmid (Addgene: 80441) that is used to generate pCW-dCpf1-blast plasmid (from Toth *et al*., 2016).

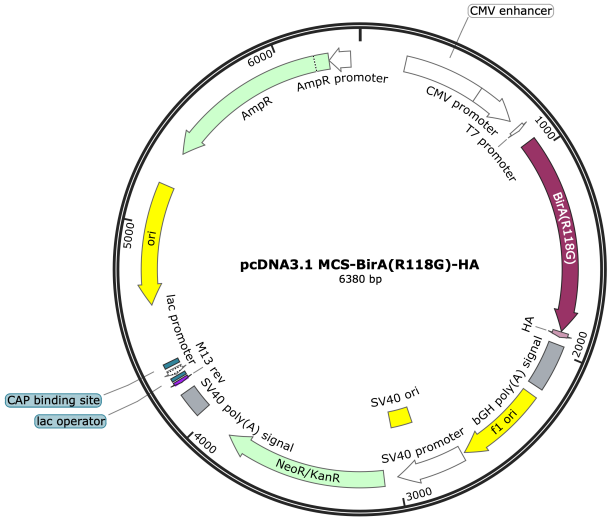


Figure A.7. Vector map of pcDNA3.1 MCS-BirA(R118G)-HA plasmid (Addgene: 36047) that is used to generate BirA N-terminal and C-terminal fragments (from Roux *et al*., 2012).

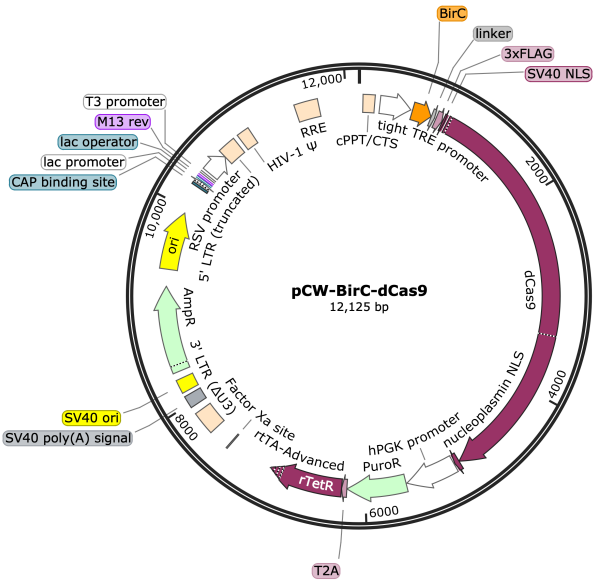


Figure A.8. Vector map of pCW-BirC-dCas9 plasmid that is used to generate BirC-dCas9 stable cancer cell lines.

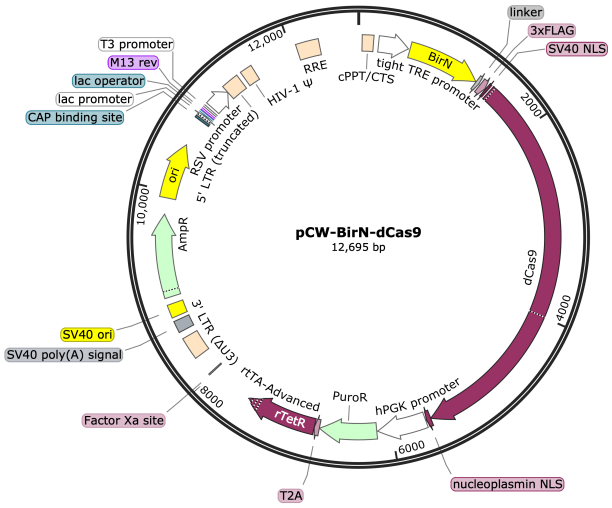


Figure A.9. Vector map of pCW-BirN-dCas9 plasmid that is used to generate BirN-dCas9 stable cancer cell lines.

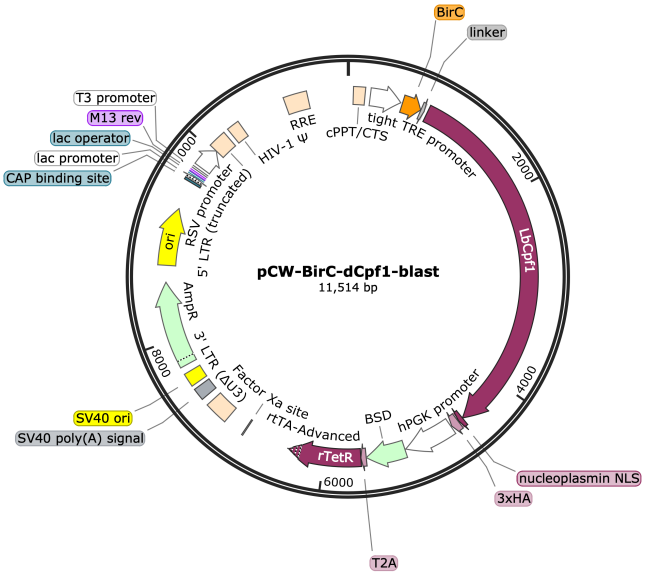


Figure A.10. Vector map of pCW-BirC-dCpf1-blast plasmid that is used to generate BirC-dCpf1 stable cancer cell lines.

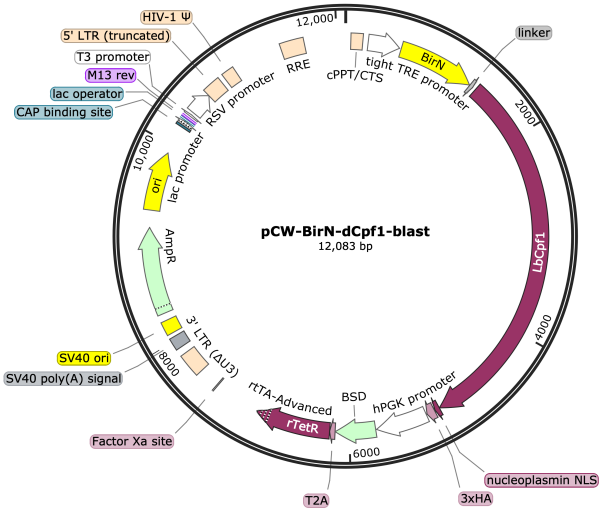


Figure A.11. Vector map of pCW-BirN-dCpf1-blast plasmid that is used to generate BirN-dCpf1 stable cancer cell lines.

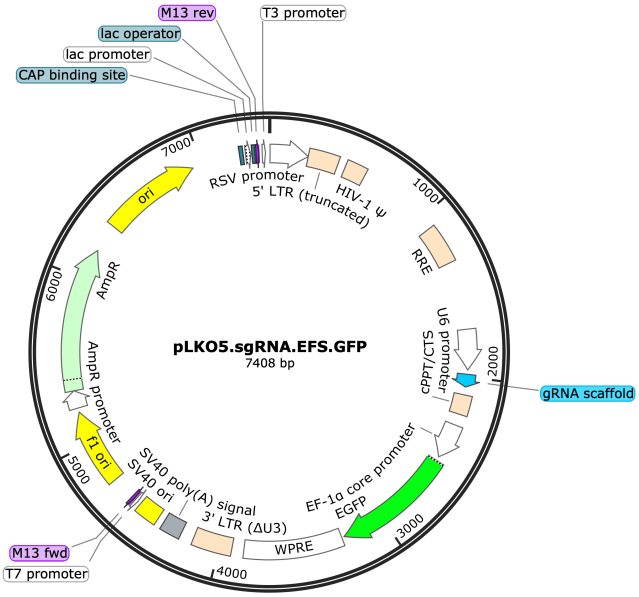


Figure A.12. Vector map of pLKO5.sgRNA.EFS.GFP plasmid (Addgene: 57822) that is used to express Cas9 customizable sgRNAs in dCas9 stable cancer cell lines (from Heckl *et al.*, 2014).

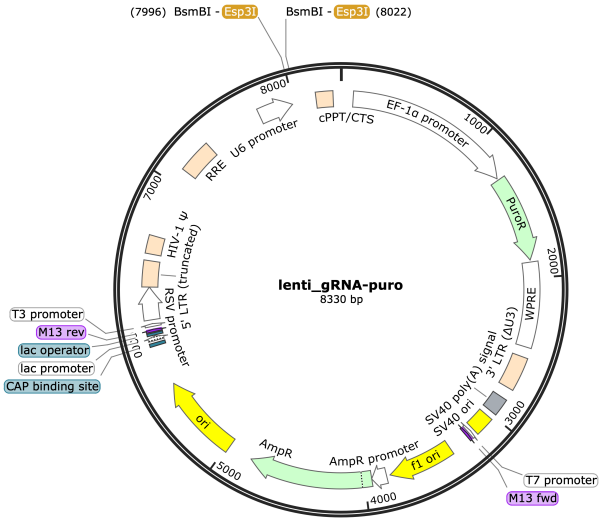


Figure A.13. Vector map of lenti_gRNA-puro plasmid (Addgene: 84752) that is used to express Cpf1 customizable sgRNAs in dCpf1 stable cancer cell lines (from Kim *et al*., 2016).

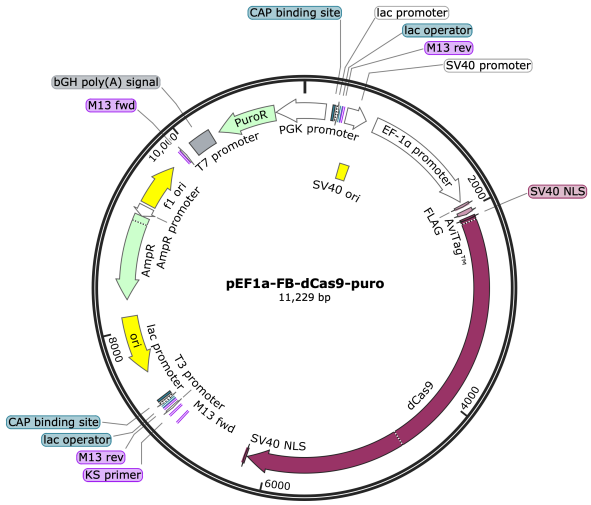


Figure A.14. Vector map of pEF1a-FB-dCas9-puro plasmid (Addgene: 100547) that is used to generate dCas9 stable K562 cells by electroporation (from Liu *et al*., 2017).

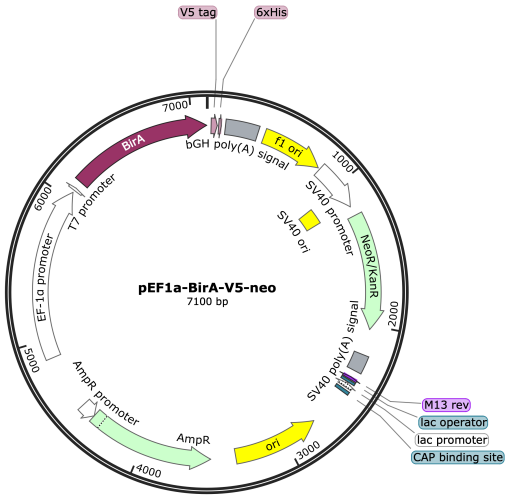


Figure A.15. Vector map of pEF1a-BirA-V5-neo plasmid (Addgene: 100548) that is used to generate BirA stable K562 cells by electroporation (from Liu *et al*., 2017).

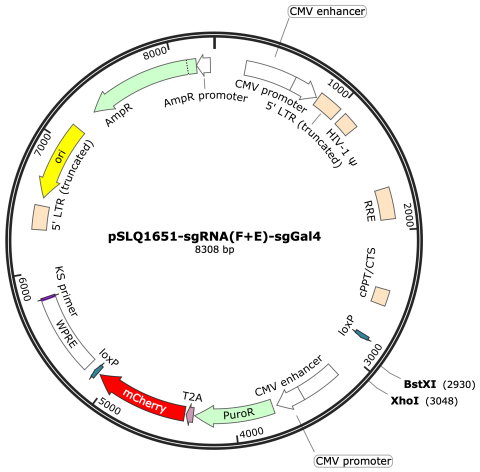


Figure A.16. Vector map of pSLQ1651-sgRNA(F+E)-sgGal4 plasmid (Addgene: 100549) that is used to generate sgRNA plasmid targeting Telomere (from Liu *et al*., 2017).

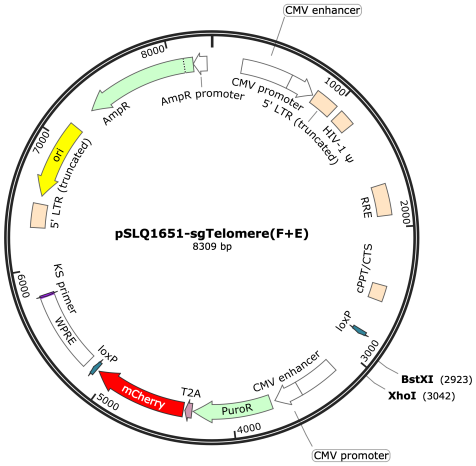


Figure A.17. Vector map of pSLQ1651-sgTelomere plasmid that is used to express Cas9 customizable sgRNAs in dCas9 stable K562 cell lines.

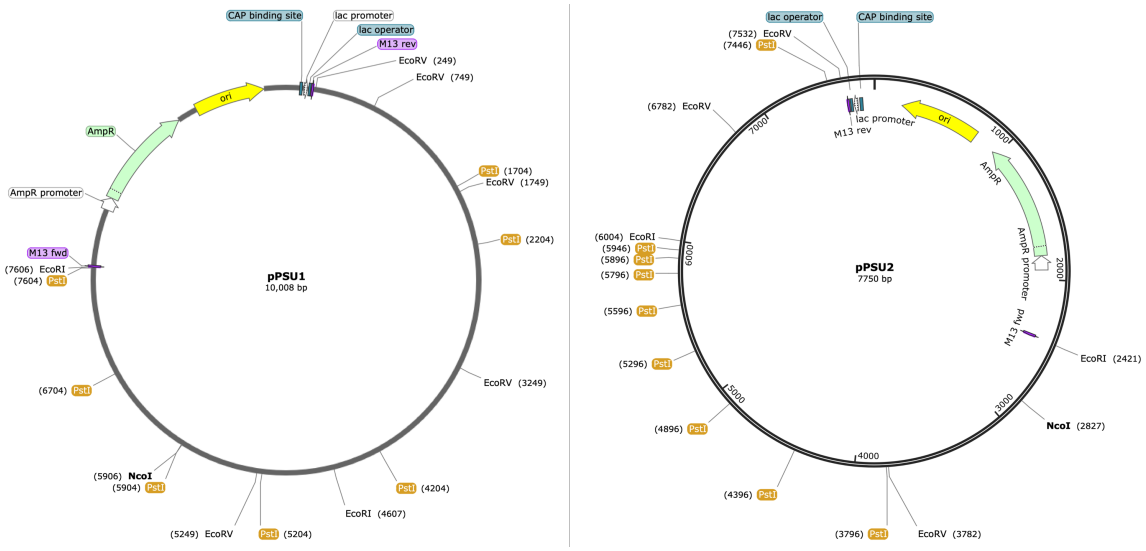
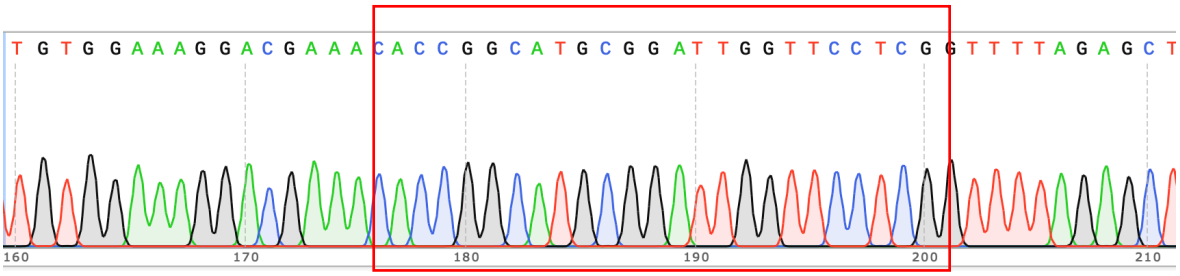
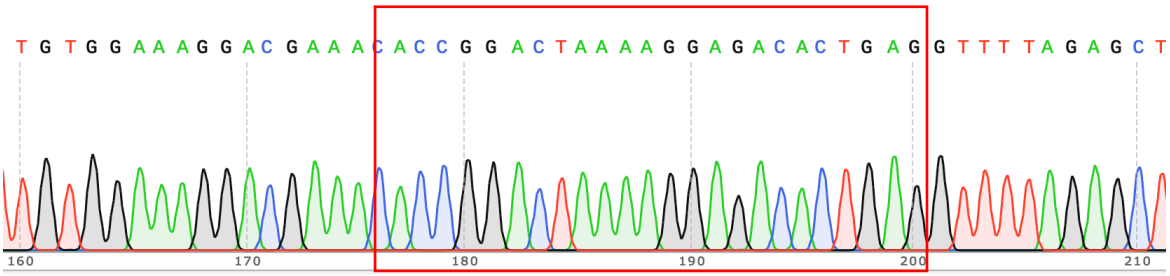


Figure A.18. Vector map of pPSU1 (Addgene: 89439) and pPSU2 (Addgene: 89566) plasmids that are used to generate 100bp and 1 kp DNA Ladders (Henrici *et al.*, 2017).

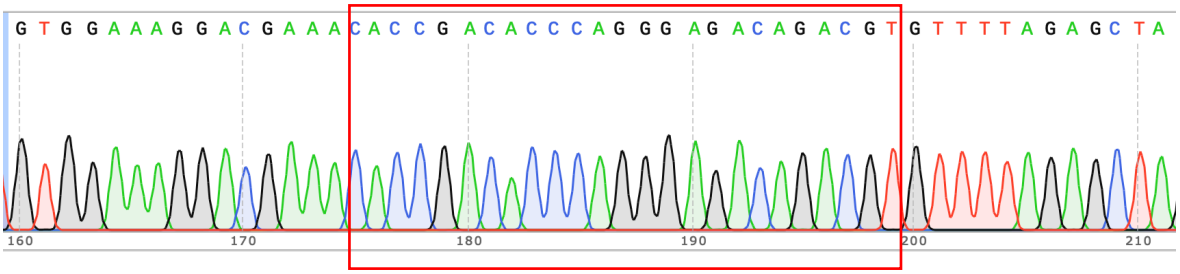
APPENDIX B: SANGER SEQUENCING RESULTS



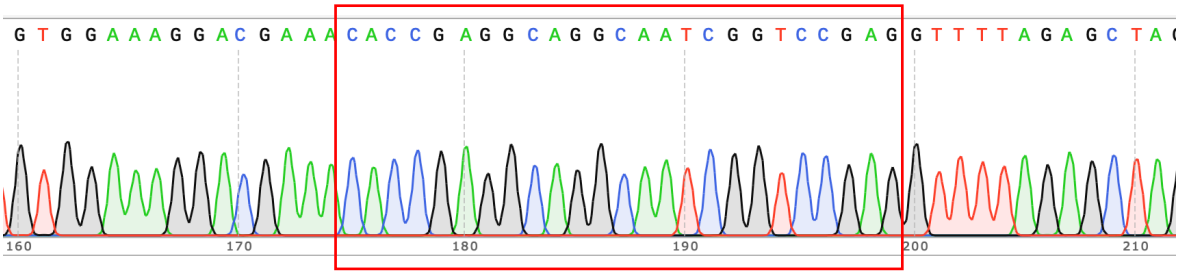
sgRNA-1



sgRNA-2



sgRNA-3



sgRNA-4

Figure B.1. Sanger sequencing result of MAPK1 promoter targeting sgRNA constructs cloned into pLKO5.sgRNA.EFS.GFP plasmid.

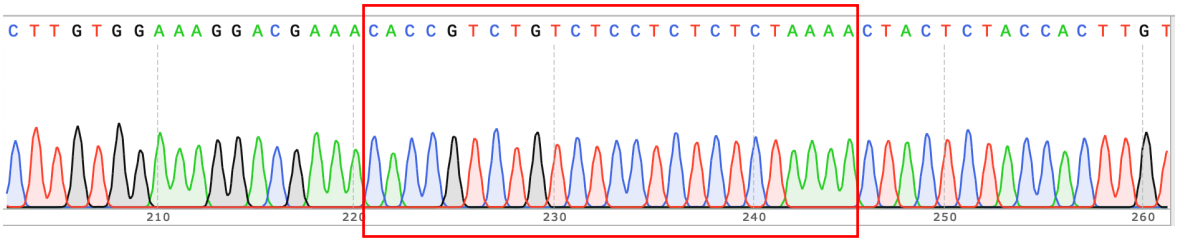


Figure B.2. Sanger sequencing result of MAPK1 promoter targeting sgRNA construct cloned into lenti_gRNA-puro plasmid.

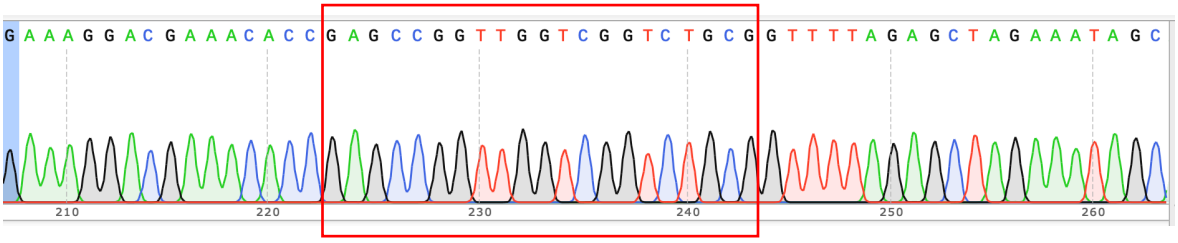


Figure B.3. Sanger sequencing result of HIRA promoter targeting sgRNA construct cloned into pLKO5.sgRNA.EFS.GFP plasmid.

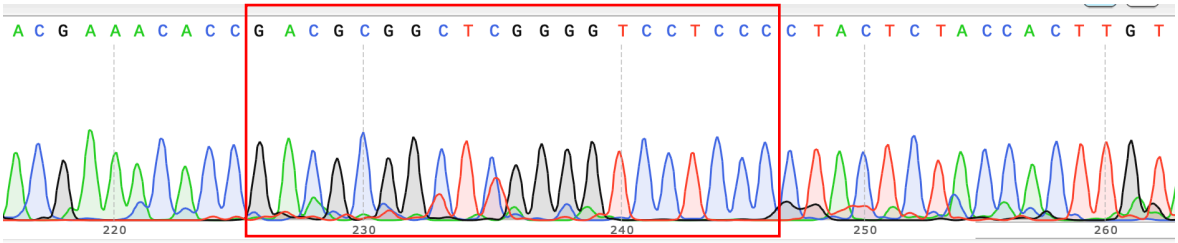


Figure B.4. Sanger sequencing result of HIRA promoter targeting sgRNA construct cloned into lenti_gRNA-puro plasmid.

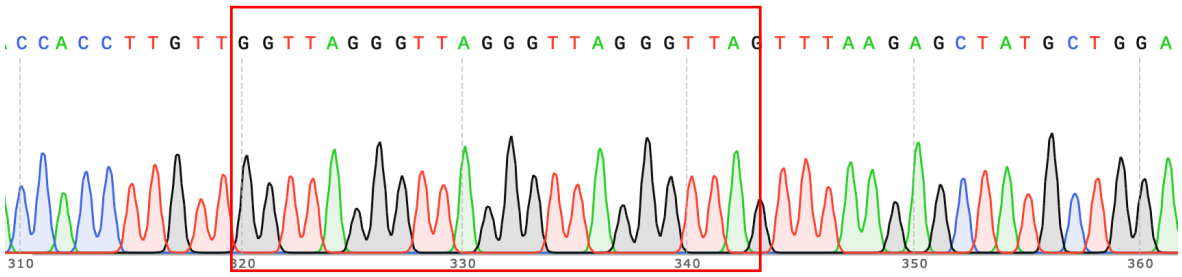


Figure B.5. Sanger sequencing result of pSLQ1651-sgTelomere-c2 sgRNA construct.

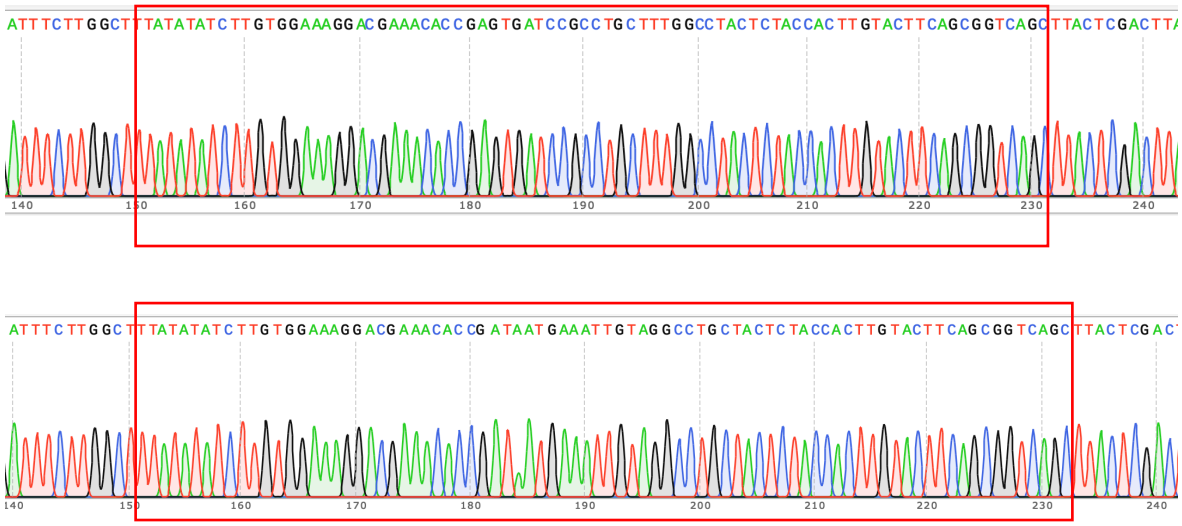


Figure B.6. Sanger sequencing result of TFF1/pS2 promoter targeting sgRNA constructs cloned into lenti_gRNA-puro plasmid by home-made Gibson assembly.

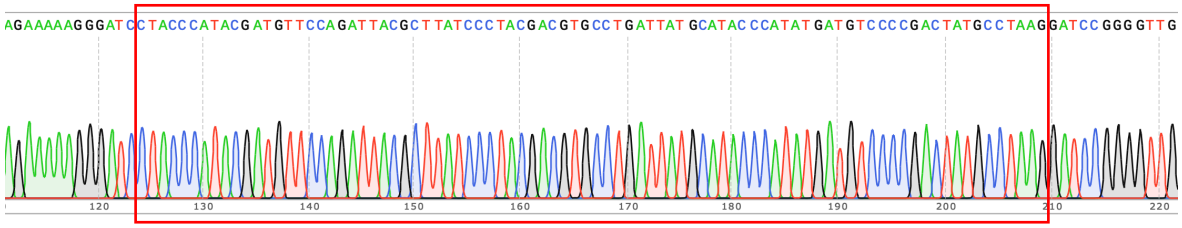


Figure B.7. Sanger sequencing result of 3xHA cloned into pCW-dCpf1-blast plasmid by home-made Gibson assembly

APPENDIX C: TARGET SITES

Table C.1. Target sites of dCas9 that are used in this study.

sgRNA Construct	Target Site	Score	Target Site Sequence (5'-3')
sgTelomere	Telomere	N/A	GTTAGGGTTAGGGTTAGGGTTA
MAPK1-sgRNA-1	MAPK1 promoter	77.1%	GCATGCGGATTGGTTCCTCG
MAPK1-sgRNA-2	MAPK1 promoter	75.1%	GACTAAAAGGAGACACTGAG
MAPK1-sgRNA-3	MAPK1 promoter	74.5%	ACACCCAGGGAGACAGACGT
MAPK1-sgRNA-4	MAPK1 promoter	66.7%	AGGCAGGCAATCGGTCCGAG
pS2-sgRNA-1	TFF1/pS2 promoter	70.4%	GTCCAGAGTAGCTAGGATTAC
pS2-sgRNA-2	TFF1/pS2 promoter	85.9%	GCAGGCGGATCACTTAAAGTC
pS2-sgRNA-3	TFF1/pS2 promoter	45.2 %	GTAGGACCTGGATTAAGGTC
pS2-sgRNA-4	TFF1/pS2 promoter	44.0%	GTGATAGACAGAGACGACATG
pS2-sgRNA-5	TFF1/pS2 promoter	79.9%	GAAATTGTAGGCCTGGCGCAG
pS2-sgRNA-6	TFF1/pS2 promoter	46.0%	GCAGTATTTACCCTGGCGGGA
HIRA-sgRNA-1	HIRA promoter	49.5%	GCAGGCGGATCACTTAAAGTC

Table C.2. Target sites of dCpf1 that are used in this study.

sgRNA Construct	Target Site	Score	Target Site Sequence (5'-3')
MAPK1-sgRNA-1	MAPK1 promoter	72.2%	TCTGTCTCCTCTCTCTAAAA
pS2-sgRNA-3	TFF1/pS2 promoter	49.1%	CGGCCATCTCTCACTATGAA
pS2-sgRNA-4	TFF1/pS2 promoter	49.2%	CCCTGGCGGGAGGGCCTCTC

Table C.2. Target sites of dCpf1 that are used in this study (cont.).

pS2-sgRNA-12	TFF1/pS2 promoter	97.9%	CAGGCCTACAATTTCATTAT
pS2-sgRNA-21	TFF1/pS2 promoter	89.2%	CAGGCCTACAATTTCATTAT
HIRA-sgRNA-1	HIRA promoter	49.9%	GCAGGCGGATCACTTAAAGTC

APPENDIX D: TRANSFECTION OF HEK293FT CELLS WITH sgRNA PLASMID

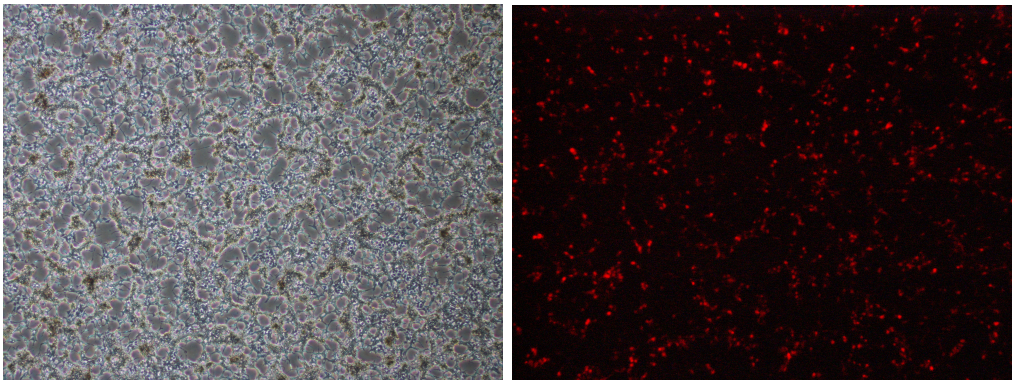


Figure D.1. Fluorescence microscopy images of HEK293FT cells that are transfected with pSLQ1651-sgTelomere plasmid.

APPENDIX E: sgRNA CLONING AND VALIDATIONS

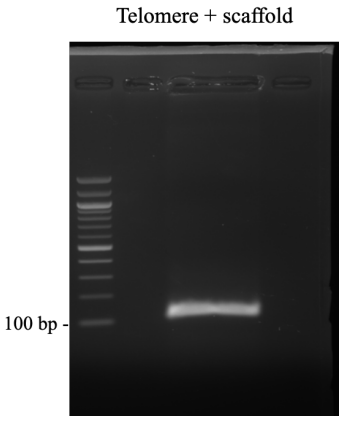


Figure E.1. sgTelomere coding sequence amplified with its scaffold is indicated.

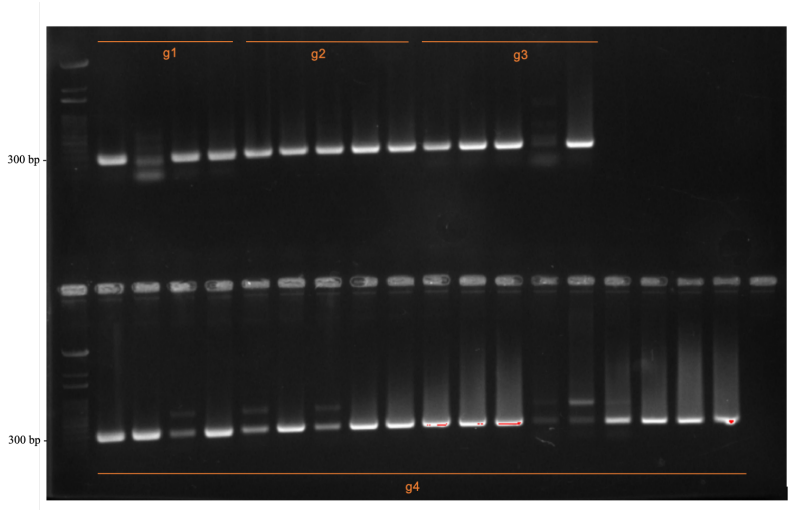


Figure E.2. Colony PCR result is shown for 32 colonies carrying the MAPK promoter targeting four different sgRNAs cloned into pLKO5.EFS.GFP plasmid.

DEVELOPMENT & APPLICATION OF NOVEL METHODS TO IDENTIFY LOCUS-SPECIFIC TRANSCRIPTIONAL REGULATORS

APPROVED BY:

Assoc. Prof. N. C. Tolga Emre .....
(Thesis Supervisor)

Assoc. Prof. Umut Şahin .....

Assoc. Prof. Ferruh Özcan .....

DATE OF APPROVAL: 23.07.2019