Year : 2018

CHROMATIN INDUCIBLE TARGETING (CIT) : A NOVEL SYSTEM TO MANIPULATE GENE EXPRESSION AT EXPANDED CAG/CTG REPEATS

Yang Bin

Yang Bin, 2018, CHROMATIN INDUCIBLE TARGETING (CIT) : A NOVEL SYSTEM TO MANIPULATE GENE EXPRESSION AT EXPANDED CAG/CTG REPEATS

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_150017DB44732

UNIL | Université de Lausanne

Faculté de biologie et de médecine

Ecole de biologie

CHROMATIN INDUCIBLE TARGETING (CIT): A NOVEL SYSTEM TO MANIPULATE GENE EXPRESSION AT EXPANDED CAG/CTG REPEATS

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Bin YANG

Master de la Fudan University

Jury

Prof. Laurent Lehmann, Président
Prof. Vincent Dion, Directeur de thèse
Prof. Nouria Hernandez, Co-directeur
Prof. Françoise Stutz, expert
Prof. Liliane Michalik, expert

Lausanne September 2018

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof. Laurent	**Lehmann**
Directeur·trice de thèse	Monsieur	Prof. Vincent	**Dion**
Co-directeur·trice	Madame	Prof. Nouria	**Hernandez**
Expert·e·s	Madame	Dre Liliane	**Michalik**
	Madame	Prof. Françoise	**Stutz**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Bin Yang

Master of Science Fudan University, Chine

intitulée

Chromatin Inducible Targeting (CIT): a novel system to manipulate gene expression at expanded CAG/CTG repeats

Lausanne, le 10 septembre 2018

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Laurent Lehmann

Acknowledgement

First and foremost, I would like to thank my Ph.D. supervisor, Professor Vincent Dion. During my whole Ph.D. time, he shows enormous passion for research, great patient for educating students and excellent talent to manage people. All the training I received from him is extremely valuable for my further career. I am really appreciated to all his efforts for my education.

Secondly, I would like to thank my co-thesis supervisor and my mentor Professor Nouria Hernandez. All the discussion and suggestion in the seminars and my defense are very helpful. In the meantime, as my mentor, Nouria pays attention to my future career and provide support about it. I am really lucky to have her as my co-supervisor and mentor.

I want to thank my committee members, Françoise Stutz, Liliane Michalik, Laurent Lehmann. I am looking forward to the discussion on my defense and I am sure they will give me helpful feedback and suggestions. Joachim Lingner, Winship Herr, Didier Trono, Jesper Svejstrup and Ana Claudia Marques also help me greatly with project direction suggestion. I am very lucky to receive all these assistance from them.

Thirdly, all the Dion lab members are my great treasure. Lorène is an amazing technician who gives me a lot of help for my project as well as my outside lab Suisse life. Alicia provided inspiring discussion and extensively help with my project. Oscar helped me with experiments, and he is always willing to spend time discussing scientific ideas. Cinzia helped with IF work for my project. Gustavo constantly gives inspiring comments and suggestions. Flavia helped me a lot with data analysis. All the other members also give helpful discussions about daily project progress and presentations. Nathalie Clerc is our sweet and professional superwoman. I would never have such wonderful abroad life without her. I am also appreciated with Bystricky and Wilson's lab generously sharing reagent with us.

Most importantly, I would like to show my love to my parents. They are the coolest mama and papa in the world for me. No matter I got little progress or a big failure, they are always supportive and caring. Without their love and encouragement, I will never be able to have the faith of finishing my study. I love you forever.

Abstract

Epigenetic modifications have drawn significant attention due to their crucial roles in development and disease. The recent emergence of epigenome editing tools provides an attractive strategy for manipulating chromatin structure and gene expression. They have been shown to successfully activate and silence targeting genes in conjunction with chromatin modifying enzymes. A major gap in knowledge pertains to how epigenome editing differs in efficiency at distinct DNA contexts. The focus of this thesis was to build a novel system, chromatin inducible targeting (CIT), to test the effect of sequence context on the efficiency of chromatin modifying enzymes to control gene expression. Specifically, I tested the effect of CAG/CTG repeat expansion on the ability of HDAC5, HDAC3, and DNMT1 targeting to modify chromatin structure.

CIT can be divided into three major components. First, a GFP-based reporter monitors gene expression. It contains an intron carrying a varying number of CAG/CTG triplets. Second, I adapted the ParB-INT targeting system such that any protein of interest can be recruited within 300bp of the expanded CAG/CTG repeat tract. Third, an ABA-based chemical inducible proximity system that allows for spatiotemporal and reversible targeting of proteins to chromatin. Notably, CIT is also well suited to ask whether chromatin modifying enzymes work locally to regulate gene expression or *in trans*.

I found that HDAC5 targeting silences the GFP reporter while decreasing local histone acetylation. Interestingly, HDAC5 preferentially silences the reporter with the shorter repeat tract, probably because of the lower levels of histone acetylation present at the expanded repeat tract before targeting. HDAC5 is thought to deacetylate histones by recruiting another HDAC, HDAC3. Surprisingly, however, HDAC3 targeting increased GFP expression, and its effect is insensitive to the size of the repeat tract. This effect is controversial to the current models of HDAC3 function that deacetylation of histone tails by HDAC3 in gene body improves transcriptional output. Moreover, I found that Dnmt1 targeting has a similar effect on gene silencing as HDAC5 targeting: it is more efficient in the context of shorter CAG/CTG repeats.

CIT provides a novel strategy to optimize the efficiency of epigenome editing in a highly controlled and flexible manner. Our data uncover novel mechanisms of gene regulation by these chromatin modifiers and guides their use in manipulating chromatin structure. CIT is suitable for screening and can be adapted to study the effect of virtually any sequences on epigenome editing.

Sommaire

Les modifications épigénétiques ont attiré une grande attention en raison de leurs rôles cruciaux dans le développement et dans plusieurs pathologies. L'émergence récente d'outils d'édition de l'épigénome constitue une stratégie intéressante pour manipuler la structure de la chromatine et l'expression des gènes. Il a été démontré qu'il est possible d'activent et de bloquer avec succès l'expression des gènes en recrutant des enzymes modifiant la chromatine. Une lacune majeure dans la compréhension de ce processus concerne la façon dont l'édition de l'épigénome diffère en efficacité dans des contextes d'ADN distincts. L'objectif de cette thèse était de construire un nouveau système de ciblage de la chromatine inductible (CIT), pour tester l'effet du contexte de la séquence sur l'efficacité des enzymes modifiant la chromatine à contrôler l'expression des gènes. Spécifiquement, j'ai testé l'effet de l'expansion de répétition CAG / CTG sur la capacité du ciblage de HDAC5, HDAC3, et DNMT1 de modifier la structure de la chromatine. CIT peut être divisé en trois principaux composants. Tout d'abord, un rapporteur basé sur la GFP pour quantifier l'expression des gènes. Il contient un intron portant un nombre variable de triplets CAG / CTG. Deuxièmement, j'ai adapté le système de ciblage ParB-INT de telle sorte que toute protéine d'intérêt puisse être recrutée à moins de 300 pb de l'expansion de triplets. Troisièmement, notre outil contient un système de proximité inductible chimique à base d'ABA permettant un recrutement spatiotemporel et réversible des protéines à la chromatine. Notamment, CIT est également bien adapté pour déterminer si les enzymes modifiant la chromatine travaillent localement pour réguler l'expression des gènes ou en *trans*. J'ai trouvé que le ciblage HDAC5 fait taire le rapporteur GFP tout en diminuant l'acétylation locale des histones. Fait intéressant, HDAC5 diminue l'expression du rapporteur GFP de manière préférentielle avec la répétition la plus courte, probablement en raison des niveaux inférieurs d'acétylation des histones présents près des expansion avant le ciblage. HDAC5 est censé de désacétyler les histones en recrutant une autre HDAC, HDAC3. Étonnamment, cependant, le recruitment de HDAC3 au rapporteur a augmenté son expression peu importe la taille de la région répétée. Cet effet est en désaccord avec les modèles actuels de la fonction de HDAC3 qui stipulent que la désacétylation des queues d'histones par HDAC3 dans le corps de gènes améliore la transcription. De plus, j'ai trouvé que le ciblage de Dnmt1 a un effet similaire à HDAC5 sur l'expression génique: il réduit le niveau de GFP plus efficacement dans le contexte de répétitions CAG / CTG plus courtes. CIT fournit une nouvelle stratégie pour optimiser l'efficacité de l'édition de l'épigénome d'une manière hautement contrôlée et flexible. Nos données révèlent de nouveaux mécanismes de régulation des gènes par ces modificateurs de la chromatine et guident leur utilisation dans la manipulation de la structure de la chromatine. Le CIT convient au criblage et peut être adapté pour étudier l'effet de pratiquement toutes les séquences sur l'édition de l'épigénome.

List of Abbreviations

ABA abscisic acid

ABI ABA insensitive

acH3 acetylated histone H3

BER base excision repair

Cas CRISPR-associated

ChIP Chromatin immunoprecipitation

CIP chemical inducible proximity

CIT chromatin inducible targeting

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

dCas9 catalytic dead Cas9

DM1 Myotonic Dystrophy Type 1

Dnmt1 DNA methyltransferase 1

Dnmt3a DNA methyltransferase 3a

Dnmt3b DNA methyltransferase 3b

Dnmt3L DNA methyltransferase 3L

DSBR double-strand break repair

ESC embryonic stem cells

FKBP FK506 binding protein

FrB FKBP-Rap binding domain

FRDA Friedreich's ataxia

FXS Fragile X syndrome

GGR global genomic repair

HD Huntington Disease

HDAC5 Histone deacetylase 5

HDAC3 Histone deacetylase 3

HP1 heterochromatin protein 1

INT ANCHOR system DNA segment contains *parS* site

iPSC induced pluripotent stem cells

KRAB Kruppel-associated box domain

LacO lactose operon

LIP light inducible proximity

LSD1 lysine-specific demethylase 1

MMR mismatch repair

Nco-R nuclear receptor co-repressor

NER nucleotide excision repair

NSC neural stem cell

PAM protospacer adjacent motif

ParB ANCHOR system DNA binding protein

PYL pyrabactin resistance 1-like

qPCR quantitative PCR

RAN-translation Repeat-associated non-ATG translation

Rap rapamycin

SCA Spinocerebellar ataxia

S-DNA Slipped-strand DNA

sgRNA single guide RNA

SKD super KRAB domain

SID mSin interaction domain

SMRT silencing mediator of retinoic acid and thyroid hormone receptor

SP-PCR small pool PCR

TALEs transcription activator-like effectors

TCR transcription-coupled repair

TET Ten-eleven translocation dioxygenases

TNR trinucleotide repeat

tracrRNA trans-activating crRNA

VEGFA vascular endothelial growth factor A

VP16 (Vmw65) Herpes simplex virus protein vmw65

ZFs zinc fingers

ZFNs zinc finger nucleases

Table of Contents

List of Figures

List of Tables

Chapter I

Introduction

The human genome is highly organized in the nucleus and packaged into a structure called chromatin. About 147bp of negatively charged DNA wrapped around a positively charged histone protein complex containing two copies of each H2A, H2B, H3 and H4, to form one nucleosome (Margueron and Reinberg, 2010; Rando and Chang, 2009; Shahbazian and Grunstein, 2007; Venkatesh and Workman, 2015). Nucleosomes are further packaged into a complex, and controversial three-dimensional (3D) chromatin structure. Based on distinct compaction levels based on the density of DNA staining during interphase, chromatin is further grouped into two forms: euchromatin and heterochromatin. Euchromatin is lightly stained and is transcriptionally highly active, accompanied by typical histone marks like acetylation in histone H3 tail. Heterochromatin, by contrast, shows more condensed staining, low transcriptional activity and more DNA methylation at CpG dinucleotides. Heterochromatic histone marks include trimethylation of histone H3 at lysine 9 and 27, among others. Nevertheless, how chromatin modifying enzymes alter gene regulations is still unclear. It has been shown chromatin structure is tightly correlated with cancer and neurological disorders (Beltran et al., 2008; Dion and Wilson, 2009; Egger et al., 2004; Feinberg, 2007; Portela and Esteller, 2010; Robertson, 2005; Robertson and Wolffe, 2000). The hypothesis is manipulating chromatin structure, and gene expression in a controllable manner may contribute to the alleviation of disease symptoms. However, having chromatin modifying enzymes to influence gene expression at will is not trivial. This is the main reason I am interested in epigenome editing and gene regulation.

I will focus on newly developed epigenome editing tools and inducible proximity systems to evaluate their applicability to manipulating chromatin and gene expression. I will also describe a novel inducible chromatin targeting system that I have developed to study the relationship between sequence context and epigenome editing as well as to uncover novel mechanisms regulating gene expression.

1. Epigenome editing tools

Chromatin modifications have long been known to correlate with gene expression tightly. However, the mechanisms of gene regulation by chromatin structures were difficult to assess, especially through loss-of-function studies. This is because chromatin modifying enzymes loss of activity often has pleiotropic effects. The recent rise of epigenome editing tools, with their potential for spatial and temporal regulation as well as reversibility, has created a significant opportunity for both basic research and translational studies. The discovery and development of customizable sequence-specific DNA binding peptides make it possible to recognize endogenous DNA sequences and bring chromatin modifying enzymes to a locus of choice and catalyze histone or DNA modifications (Groote et al., 2012).

Three primary epigenome editing techniques are currently in use are zinc finger proteins (ZFs), transcription activator-like effectors (TALEs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) proteins (Waryah et al., 2018). In the following sections, these tools will be discussed from their history, structural properties, mechanisms of action, as well as their applications and limitations.

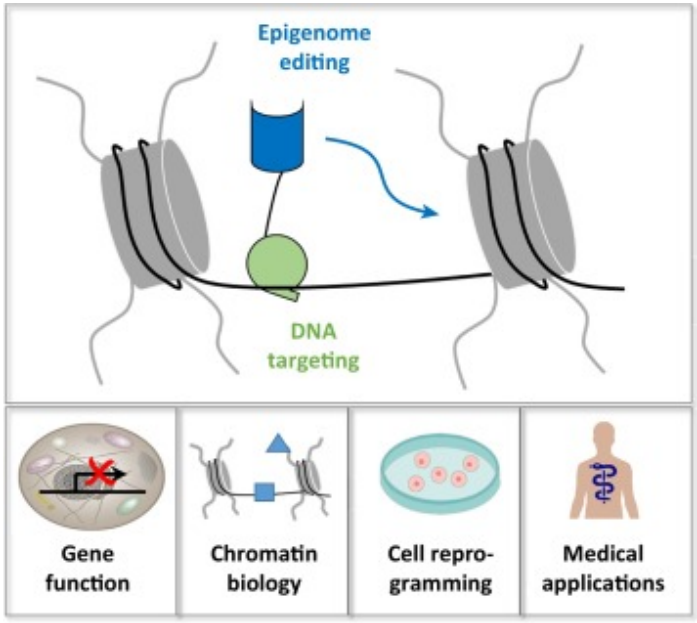


Figure II. 1. Epigenome editing tools applications. Epigenome editing tools can be used to study gene function, chromatin biology, cell reprogramming and further medical fields. **(**Derived from Kungulovski, G. et al. Trends Genet, 2016)

1.1 Zinc Finger proteins (ZFs)

1.1.1 Discovery, working mechanism, and design

A. Klug and colleagues first reported zfs in 1985 (Miller et al., 1985). They found a small protein containing repetitive zinc-binding domains with 7 to 11 zinc atoms that could bind to the 5S RNA genes in *Xenopus Laevis* oocytes extracts (Miller et al., 1985). This small protein Zif268 can bind to DNA tightly and RNA as well (Blancafort et al., 1999; Moore et al., 2001). The ZFs bind to the major groove of DNA via an α-helix of the ZFs with the N-terminal close to the DNA and the C-terminal located away from the binding site (Elrod-Erickson et al., 1996; Fairall et al., 1993; Pavletich and Pabo, 1991). Each zinc finger protein can recognize three bps of DNA so that individual finger can be assembled to match more extended target sequences (Segal et al., 1999). After zinc finger nucleases (ZFNs) screening, the efficiency of FokI cleavage driven by ZFs recognition are successfully reach 15.7% (Wang et al., 2013). The high efficiency of targeting has been observed in human embryonic stem cells *OCT4* gene, which on-target efficiency can reach 94% (Hockemeyer et al., 2009). In one word, ZFs can be engineered to target DNA locus with high efficiency.

1.1.2 ZFs in gene activation

Since ZFs or multiple ZFs complexes can not modulate epigenome themselves, gene activators or repressors need to be linked to ZFs to do the work. Synthetic transcription activation domain VP64 had been used as effector domains in conjunction with ZFs to achieve an erbB-2-luciferase reporter activated transcription in HeLa/tet-off cell line (Beerli et al., 2000; Beltran et al., 2007; Liu et al., 2001; Zhang et al., 2000). Human histone acetyltransferase P300 core fused to ZFs also managed to target *ICAMI* promoter and activate gene transcription in HEK293T cells (Hilton et al., 2015). All of those show ZFs can be fused to transcription activators and induce gene activation.

1.1.3 ZFs in gene silencing

ZFs have been widely used in conjunction with DNA methyltransferases to induce local methylation and transcription silencing. One early report in 1997 showed ZFs fused to the bacterial DNA methyltransferase M.SssI successfully methylated the 5' side of the ZF binding site (Xu and Bestor, 1997). After this, several

ZFs-DNA methyltransferases fusions, for instance, using the catalytic domain of DNA methyltransferase 3a (DNMT3a) and DNMT3b, all increased DNA methylation levels locally and repress transcription (Li et al., 2007; Nunna et al., 2014). Beyond DNA methyltransferases, histone methyltransferases G9a and SUV39-H1 also efficiently silenced their target genes when recruited via synthetic ZFs (Snowden et al., 2002; Falahi et al., 2013). A ZF fusion with Kruppel-associated box domain (KRAB) recruited cofactor KAP1 and resulted in the loss of histone acetylation, a gain in histone methylation, and a long-range gene silencing (Groner et al., 2010; Stolzenburg et al., 2012). Interestingly, a genome-wide study on zinc finger fusion protein with artificial transcription factors (ATFs) super KRAB domain (SKD) targeting to the human *SOX2* promoter in MCF7 breast cancer cells revealed thousands of off-target binding sites by ChIP-seq (Grimmer et al., 2014). Majority of the off-target genes showed no expression differences. Also, ATF-SKD induced transcription repression did not affect histone modifications like H3K4me3 and H3K9ac (Grimmer et al., 2014). Above all, ZFs shows usage widely in silencing gene with a variety of factors combined.

1.1.4 ZFs in the clinical trial

ZFs show potential clinical usage thanks to its small size, high expression level and toleration of variable chromatin contexts (Beltran et al., 2008; Gregory et al., 2013). Clinical trials have been conducted using ZFs for diabetic neuropathy and completed in 2016 (Eisenstein, 2012). Sangamo Biosciences conducted this clinical phase 2 trial with ZF proteins linked to the P65 transcription activator to induce the expression of the vascular endothelial growth factor A (VEGFA) in diabetic neuropathy patients. Unfortunately, the trial failed because the treated group failed to display significant improvements compared to placebo (Eisenstein, 2012).

1.1.5 ZFs limitations

Despite the ZFs' high efficiency of epigenome editing capabilities, the off-target effects remain a significant concern (Grimmer et al., 2014). Off-target rates for ZFs can reach 8%, which is 18 fold more than the empty vector with no ZFs. Its off-target binding sites can be located in other promoters and thus enact changes in gene expression (Huisman et al., 2015). Optimization could significantly increase the

specificity of ZFs and help epigenome editing become safe enough for clinical use (Hurt et al., 2003). However, newer technologies appear to be more flexible in design and more comfortable to use, and thus ZF technologies are being phased out.

1.2 Transcription Activator-Like Effectors (TALEs)

1.2.1 Discovery, working mechanism and design

TALEs are another customizable DNA binding domains used for epigenome editing. TALEs are proteins secreted by *Xanthomonas spp* bacteria that alter gene expression in their host plants (Joung and Sander, 2013). Then how TALEs recognize DNA was uncovered in 2009 by the Bonas group (Boch et al., 2009). One TALE repeat can bind one base of DNA via amino acids 12 and 13. This recognition site is also located in the major groove of DNA, similar to ZF-mediated binding, thereby making base-specific contacts with the target sequence (Deng et al., 2012). Changing amino acids at position 12 and 13 changes the specificity of binding, allowing for recognition of a wide array of sequences through multiplexing of TALE repeats. This protein-DNA recognition can be widely used for genome editing and epigenome engineering by assembling specific TALE repeats based on targeting sequences. Using a puromycin expression driven by TALEN cleavage and recombination, the efficiency of TALENs on-target cleavage varies between different tissues and genomic locations. It shows similar efficiency as ZFs, reaching more than 90% on-target cleavage with a mean of 22.2% (Hockemeyer et al., 2011). An endogenous EGFP transgene disruption test with 48 engineered TALENs in human cells also revealed the highest on-targeting cleavage rate was around 60-70% (Reyon et al., 2012). In conclusion, TALEs could be another epigenome editing tool with the relatively high efficiency of recognizing on-targeting sequences.

1.2.2 TALEs in gene activation

Similar to the applications described above with ZFs, TALEs are used to activate gene expression by fusing them to synthetic transcription activation domains like VP16 and VP64 (Bultmann et al., 2012; Konermann et al., 2013; Zhang et al., 2011). TALEs themselves can activate transcription. At promoter regions of *SOX2, KLF4, c-MYC*, and *OCT4* targeted by TALEs in 293FT cells, TALEs were able to increase *Sox2* and *KLF4* transcription by 5.5 fold and 2.2 fold respectively (Zhang et

al., 2011). This may indicate that epigenome editing by TALEs depends on the sequence context. Studies using TALE-VP16 fusions showed that multiple TALEs have the better efficiency to promote gene transcription (Maeder et al., 2013a; Perez-Pinera et al., 2013a). Ten-Eleven Dioxygenase-1 (TET1) had been shown in conjunction with TALEs and optogenetic tools, two light inducible complementary fusion protein constructs (Konermann et al., 2013), to demethylate DNA and increase the levels of Ascl1 mRNA by 2.5 fold in rat striatal neural stem cell (Lo et al., 2017). Thus, TALEs, like ZFs, can efficiently modify chromatin and DNA to activate gene expression.

1.2.3 TALEs in gene repression

TALEs have been used for epigenome silencing by fusing them to KRAB and mSin interaction domain (SID). They were shown to repress endogenous *SOX2* expression in HEK293FT cells (Cong et al., 2012). SID can further assemble into SID4X, which in turn reduces H3K9 acetylation and silences transcription (Konermann et al., 2013). DNA methyltransferase 3a (DNMT3a) fused to TALEs was similarly shown to induce DNA methylation at the Ascl1 gene in dorsal root ganglion-derived neural stem cell (NSC), which resulted in gene silencing (Lo et al., 2017). TALEs-lysine-specific demethylase 1 (LSD1) fusion reduced both H3K4 methylation and H3K27 acetylation, presumably due to crosstalk with HDACs (Lee et al., 2006; Shi et al., 2004). The TALEs-LSD1 fusion was capable of reducing H3K4me2 and H3K27ac levels and further repressing genes by targeting their enhancer regions in K562 erythroleukemia cells (Mendenhall et al., 2013). So TALEs show not only the gene activation ability but also silencing gene with chromatin modifying enzymes cooperation.

1.2.4 TALEs in the clinical trial

There has been no report of clinical trials involving TALE-mediated epigenome editing, only one patient test with its nuclease version. What has been done is to transplant TALEN edited universal CAR19 T cell to leukemia patient (Qasim et al., 2015). TALEN induced T cell receptor alpha constant chain expression disruption and CD52 gene silencing in T cell transplantation showed no significant toxicity in

this patient. It provides an early proof-of-concept for the strategy, but a large-scale study remains to be conducted to confirm or refute clinical efficacy and safety.

1.2.5 TALEs limitations

TALEs targeting showed off-target effects due to a particular subset of TALE repeat by high throughput sequencing or protein-binding microarray (Guilinger et al., 2014; Juillerat et al., 2014; Meckler et al., 2013; Rogers et al., 2015). However, the off-targeting effect is not always the case depend on TALE length and targeting loci. TALE fused to histone demethylase LSD1 successfully removed enhancer-associated modification to downregulate gene expression but not in control TALE construct treatment group in K562 cells (Mendenhall et al., 2013). In a genome-wide ChIP-seq study, TALEs-VP64 activated *IL1RN* or *HBG1/2* have been shown 31 off-target binding sites and four off-target binding sites respectively. However, no significant changes in gene expression were detected on these off-target sites (Polstein et al., 2015). They also found all off-target sites contain a GC-rich 3' end that overlapped with TALEs recognition motif. Off-target effects could be reduced by engineering longer TALENs (between 15 to 19 repeats) to achieve significantly increased specificity (Rinaldi et al., 2017). TALEs also appears closely related to sequence or chromatin contexts like DNA methylation (Chen et al., 2013; Cong et al., 2012). Using a different TALE repeat N* with no extensive involvement into DNA major groove, TALEN can successfully disrupt endogenous methylated *XPC1* in HEK293H cells (Valton et al., 2012). Above all, TALEs also obsess off-target effect like ZFs, and its DNA context sensitivity may affect targeting efficiency.

Compare to ZFs, the modular assembly of TALEs improves the rapidity and simplicity of design and production. However, the resulting TALEs DNA sequence is highly repetitive, which promotes deletion and recombination when propagating them in bacteria and after viral delivery (Holkers et al., 2013).

1.3 Catalytic dead CRISPR-associated (dCas9)

1.3.1 Discovery, working mechanism and design

CRISPR-Cas9 has been developed into a widely used genome engineering tool in the last five years. Nakata and colleagues first discovered it in 1987, but it was not

extensively used until it was shown to be an RNA-targeted nuclease that its genome editing potential was uncovered (Ishino et al., 1987; Jinek et al., 2012). This was quickly followed by genome editing in mammals (Cong et al., 2013; Mali et al., 2013a). Cas9 is guided by a noncoding trans-activating crRNA (tracrRNA) and mature crRNA to target DNA and perform cleavage. This is dramatically different from the protein-DNA recognition mechanisms of ZFs and TALEs. CRISPR-Cas9 is widely applicable in different models including yeast (Jacobs et al., 2014), zebrafish (Chang et al., 2013; Hruscha et al., 2013; Hwang et al., 2013), Drosophila, mice, and human (Cho et al., 2013). This efficient and versatile nuclease can be turned into a targeting system thanks to 4 point mutations that inactivate the nuclease domain without affecting the efficiency of targeting (dCas9) (Jinek et al., 2012; Qi et al., 2013)(Qi et al., 2013). Protospacer adjacent motifs (PAMs) are short sequences adjacent to protospacer, and they have shown to be critical for Cas9 spacer acquisition and interference (Mojica et al., 2009; Shah et al., 2013). Further understanding about Protospacer Adjacent Motif (PAM) and single guide RNA development make CRISPR-Cas9 system efficiently designed for any target DNA sequence for engineering.

1.3.2 dCas9 in gene activation

CRISPR-Cas9 based targeting systems have become the standard tool for altering gene expression (Braun et al., 2017; Hilton et al., 2015; Kwon et al., 2017; Maeder et al., 2013b; Mali et al., 2013b; Morita et al., 2016, 2016; Perez-Pinera et al., 2013b; Vojta et al., 2016). Synthetic transcriptional activator VP16 or VP64 fused to catalytic dead Cas9 has moderate gene activation capabilities with one sgRNA guidance. Using multiple sgRNAs can further increase activation rates (Konermann et al., 2013; Maeder et al., 2013b; Mali et al., 2013b; Perez-Pinera et al., 2013b). Epigenome editing using a fusion between dCas9 and the histone methyltransferase PRDM9 increased H3K4me3 and intercellular adhesion molecule 1 (*ICAM1*), Pas association domain-containing protein 1 (*RASSF1a*) and epithelial cell adhesion molecule (*EpCAM*) gene expression in HEK293T and A549 cells (Cano-Rodriguez et al., 2016). This report also indicates the importance of local chromatin modifications in altering gene expression. Ten-eleven translocation dioxygenases (TET) acts as DNA demethylase also showed gene activation ability in conjunction with dCas9 (Liu

et al., 2016; Morita et al., 2016). dCas9-TET1 fusion targeting endogenous hypermethylated *BDNF* gene promoter and *MyoD* enhancer showed demethylation and gene activation in neurons and myoblast respectively (Liu et al., 2016). Using two copies of bacteriophage MS2 RNA elements as a linker for the TET1 catalytic domain (TET-CD) with sgRNA, dCas9-TET-CD and sgRNA-MS2-TET-CD can significantly upregulate transcription in *RANKL, MAGEB2* and *MMP2* gene in HEK293FT and HeLa cells (Xu et al., 2016). SunTag was reported in 2014 which is a ten copies repeating GCN4 peptide array for recruiting multiple proteins by an antibody (Tanenbaum et al., 2014). With amplification effect by SunTag inclusion, dCas9-TET1-CD can efficiently demethylate 7 out of 9 loci in ESCs, cancer cell lines, primary neural precursor cells and mouse fetuses (Morita et al., 2016). One recent publication using dCas9 fused to Tet1 successfully reversed the hypermethylation within CGG repeats and the upstream promoter region of Fragile X syndrome induced Pluripotent Stem Cells (iPSCs) and the derived neurons (Figure II.2, Liu et al., 2018). This correction of the chromatin state accompanied by increases of the *FMR1* mRNA and FMRP upregulation in iPSC and transgenic mice as well as electrophysiological abnormalities rescue by MEA assays in editing FXS neurons. Interestingly, the acetyltransferase domain of P300 works more efficiently than V64 to activate genes, and this gene upregulation ability bias only exists in dCas9 epigenetic engineering, not ZFs nor TALEs (Hilton et al., 2015). On the one hand, it showed that histone acetylation is sufficient to activate gene transcription, at least in the loci *IL1RN, MYOD* and *OCT4* in HEK293T cells Hilton and colleagues looked at. On the other hand, this also indicates different epigenome engineering tools have a variable binding affinity or

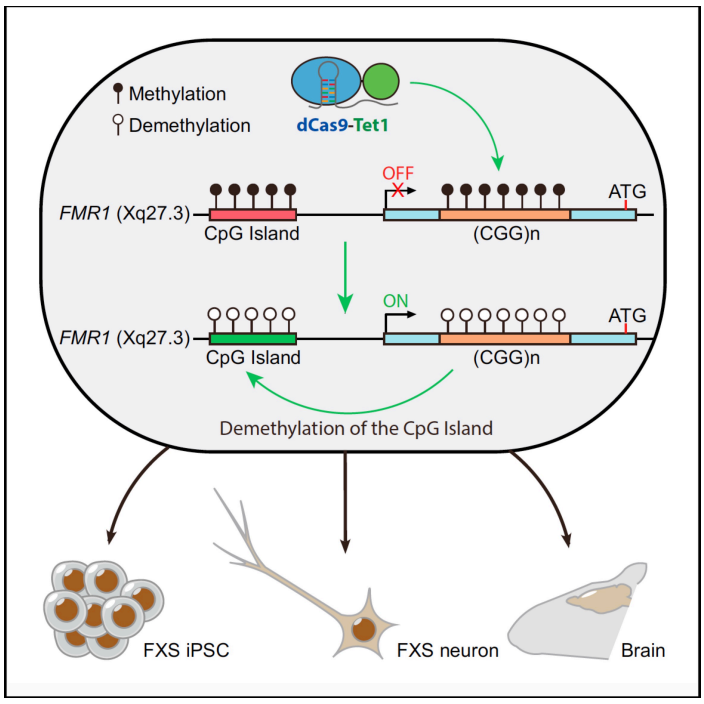


Figure II. 2. dCas9-Tet1 reactivates *FMR1* gene expression. Tet1 driven by dCas9 reverses hypermethylation in FMR1 gene and reactivates transcription in FXS iPSC. The FMR1 gene reactivation remains in iPSC derived FXS neuron and *in vivo*. (Derived from Liu et al., 2018)

protein folding properties. Above all, dCas9 can be used as an epigenome editing tool to activate gene expression.

1.3.3 dCas9 in gene silencing

dCas9 can induce gene silencing without the help of any transcriptional repressor in some circumstances due to steric hindrance (Qi et al., 2013). More efficient silencing is the use of dCas9 fusion proteins linked to KRAB or Mxi1, a mammalian transcriptional repressor domain (Gilbert et al., 2013). CTCF DNA recognition site can be methylated by dCas9 linked to the *de novo* DNA methyltransferase 3a (DNMT3a) and further blocked CTCF binding (Liu et al., 2016). The SunTag-mediated oligomerization strategy was applied to dCas9-DNMT3a in HEK293T with equal success (Huang et al., 2017). Unlike transient transcription alternation as reported for KRAB, dCas9-DNMT3a can achieve long-term gene silencing to 25 days in K-562 (Amabile et al., 2016) and 23 days in HEK293 cells (Vojta et al., 2016). Using a different effector, histone deacetylase 3 (HDAC3), dCas9 could increase or decrease the expression of target genes in murine neuroblastoma cell line, N2a (Kwon et al., 2017). The authors proposed that this be due to the chromatin context. All of those indicate dCas9 can be used to manipulate chromatin structure and gene silencing by fusing to chromatin modifying enzymes.

1.3.4 Cas9 in clinical trial and limitation

The primary concern for translational application is the same as for the ZFs and TALEs: the off-target effects. sgRNAs can bind genomic locations with multiple mismatches and often require delicate screening and optimization. Several trials focusing on targeting sequence optimization of the sgRNA, and modification of the Cas9 enzyme to better bind the sgRNA or to alter PAM requirements as well as the use of Cas9 variants or nickase variants have been shown to reduce off-target effects (Doench et al., 2014, 2016; Hruscha et al., 2013; Hsu et al., 2013; Kleinstiver et al., 2015, 2016; Mali et al., 2013b; Slaymaker et al., 2016). With the benefit of convenient design to virtually any locus of interest, Cas9-derived genome editing and epigenome engineering has exciting potential for novel therapeutic approaches in cancers and neurodegenerative diseases (Cinesi et al., 2016; Gori et al., 2015; Kick

et al., 2017; Lin et al.; Liu et al.; Long et al., 2014; Niu et al., 2014; Park et al., 2015; Pinto et al., 2017).

All the clinical trials, for now, are nuclease version Cas9, not the dCas9. Clinical studies using Cas9-mediated gene editing to target several types of cancer have been initiated in China, including HPV-related Cervical intraepithelial Neoplasial, Leukemia and Lymphoma (Baylis and McLeod, 2017; Kick et al., 2017). It is likely that Cas9-mediated treatment of two types of blood disorders, β-thalassemia and sickle cell disease, may start their clinical trials this year in Europe and the United States (Baylis and McLeod, 2017). It is still early to speculate the outcome of these trials. If Cas9 proved to be efficient and safe to apply, it would give hope to cure a wide range of devastating diseases.

In conclusion, epigenome editing tools have been widely involved in driving chromatin modifying enzymes to specific genomic loci for gene regulation. These targeting tools provide a convenient way to tackle the interplay between chromatin structure and gene regulation and make it possible to manipulate epigenome at will.

Thus far, the majority of the studies use a continuous expression of the ZFs, TALEs, or dCas9 fusions. It is therefore unclear how permanent the changes to the epigenome might be. The key is to use reversible systems, such as many chemical induced and light-induced proximity systems.

2. Chemical and light-induced proximity system

To gain a better understanding and to manipulate gene expression, multiple approaches have been developed to control cell in a switchable way like light inducible proximity (LIP) and chemical inducible proximity (CIP). The systems that have been developed to alter gene expression are listed in Table I.1 and I.2. Invariably, they are composed of two same or different components that dimerize with each other upon addition of a small molecule or by the light of a specific wavelength. One component binds chromatin, and the other is fused to some transcriptional activator or repressor.

Light induction for protein-protein interactions was developed within the past ten years with different systems using different wavelengths of light (Kennedy et al., 2010; Levskaya et al., 2009; Yazawa et al., 2009). The Zhang group further adapts light induced dimerization into transcription control usage with TALEs DNA binding motif fusion (Konermann et al., 2013). In Gao et al., they compared the transcriptional efficiency of LIP and CIP systems side by side and found that LIP systems are less potent than CIP ones in activating transcription, reaching a maximum induction of 126 fold (Gao et al., 2016).

CIP systems have been used widely to study transcriptional activation, signaling transduction, protein-protein interactions, apoptosis, glycosylation, splicing, and mouse development (Belshawl et al., 1996; Graef et al., 1997; Gruber et al., 2006; Ho et al., 1996; Holsinger et al., 1995; Kohler and Bertozzi, 2003; Liu et al., 2007; Mootz and Muir, 2002; Stankunas et al., 2003; Stanton et al., 2018). The original CIP system is the FK1012 system (Holsinger et al., 1995; Spencer et al., 1993). It involves the chemical inducer FK1012 with two ligand-binding domains of FKBP12 (Holsinger et al., 1995). More recently, Zhang and colleagues produced a split Cas9 and dCas9 constructs fused to FKBP (FK506 binding protein) and FKBP rapamycin binding (FRB) domains and the addition of rapamycin-induced the dimerization of a protein of interest (POI) fused to FRB (Zetsche et al., 2015). This reconstruction of Cas9 functioned as a normal Cas9 and enabled the experimenters

to restrict the activity to specific tissues by putting each fragment under different tissue-specific promoter while maintaining a low off-target rate. Further results from Qi and colleagues showed that multiple CIP systems could be combined and used for programming complex transcription *in vitro* and *in vivo* (Gao et al., 2016). Taking advantage of the reversible properties of Frb-FKBP-Rap and Pyl1-Abi1-ABA, targeting a truncation of heterochromatin protein 1 alpha (HP1α) to *Oct4* in ESC induced H3K9me3-dependent gene silencing. This silencing remained for multiple cell generations after HP1α targeting was relieved (Hathaway et al., 2012). A recent study using FKBP-FrB-Rap system linked to dCas9 led to the recruitment of mSWI/SNF (BAF) and activated bivalent gene transcription in mouse ESCs in a reversible way (Braun et al., 2017). Frb-FKBP dimerized dCas9 with HP1 and histone methyltransferase Suv39h1 induced H3K9me3 deposit, and the effect can be reversed by washing out inducer rapamycin (Braun et al., 2017). As a potent immunosuppressant, Rap inhibits mammalian target of Rap, mTOR, and showed toxicity in the beta cell (Barlow et al., 2012, 2013). The Crabtree lab developed another CIP system based on abscisic acid (ABA), a phytohormone (Liang et al., 2011). ABA insensitive 1 (ABI) can form a dimer with pyrabactin resistance 1-like (PYL) in the presence of ABA into the cell medium. Here again, the dimerization is reversible. It is non-toxic, and ABA is reasonably prized. Also, it is an entirely exogenous system, reducing the potential for side effects in mammalian cells.

CIP system has been used to build a safety switch for adoptive cell therapy (Di Stasi et al., 2011). Human caspase 9 fused to FK-binding protein became active by adding AP1903 in T cells with haploidentical stem-cell transplants patients. This iCasp9 suicide system improved cellular therapies safety in graft-versus-host disease. This being said, the spatiotemporal characteristic of CIP system may have significant benefit for understanding cellular functions and even for clinical usage.

In conclusion, with its power for spatiotemporal control of cellular processes, CIP and LIP systems may contribute significantly to the study of chromatin structure and gene regulation. To take one step further, it could be interesting to apply the CIP system to one type of human diseases to investigate their molecular mechanisms.

Table I.1. Chemical-induced proximity system comparison

<table>
<tr><th colspan="2">Proteins</th><th>Ligands</th><th>Toxicity</th><th>Receptors in human</th></tr>
<tr><td>FKBP</td><td>FRB</td><td>Rapamycin</td><td>Immunosuppression
(Kang et al., 2008)</td><td>mammalian target of rapamycin (mTOR)
(Kang et al., 2008)</td></tr>
<tr><td>FKBP</td><td>FKBP</td><td>FK1012</td><td rowspan="3">Immunosuppression
(Kang et al., 2008)</td><td rowspan="4">CaM-dependent phosphatase calcineurin (CaN)
(Kang et al., 2008)</td></tr>
<tr><td>FKBP</td><td>Calcineurin</td><td>FK506</td></tr>
<tr><td>FKBP</td><td>CyP</td><td>FKCsA</td></tr>
<tr><td>Calcineurin</td><td>CyP</td><td>Cyclosporine A</td><td>Immunosuppression
(Kang et al., 2008)</td></tr>
<tr><td>PYL</td><td>ABI1</td><td>Abscisic Acid</td><td>Non-toxic for human
(Liang et al., 2011)
Organ toxicity in rat
(Celik et al., 2007)</td><td rowspan="3">No report</td></tr>
<tr><td rowspan="2">GID1</td><td rowspan="2">GA1</td><td>GA</td><td rowspan="2">Organ toxicity in rat
(Celik et al., 2007)
Teratogenic to Xenopus laevis embryos
(Boga (Pekmezekmek et al., 2009)</td></tr>
<tr><td>Gibberellin3-AM</td></tr>
<tr><td rowspan="2">FKBP F36V</td><td rowspan="2">FKBP F36V</td><td>AP1903</td><td>Pancytopenia
(Zhou et al., 2016)</td><td rowspan="2">FKBP</td></tr>
<tr><td>AP20187</td><td>Non-toxic
(Je et al., 2009)</td></tr>
<tr><td>IAA17</td><td>TIR1</td><td>Auxin</td><td>Non-toxic
(Ester et al., 2009)</td><td>No report</td></tr>
</table>

Table I.2. Light-induced proximity system comparison

Proteins		Ligands	Toxicity	Receptors in human
GI	FKF1	450nm blue light (Yazawa et al., 2009)	No report	No report
CIB1	CRY2	488nm blue light (Kennedy et al., 2010)		
PhyB	PIF3	650nm red light (Levskaya et al., 2009)		
		750nm red light (Levskaya et al., 2009)		

3. Potential applications - Trinucleotide repeat diseases

Tandem repeats are common in genomes from yeast to human. At least 20 neurological diseases are caused by trinucleotide repeat (TNR) expansion (Table I.3). Since the first report of a CGG/CCG repeat expansion causing Fragile X syndrome (FXS) (Oberlé et al., 1991; Verkerk et al., 1991), it has been shown that Huntington Disease (HD), Myotonic Dystrophy Type 1 (DM1) and Friedreich's ataxia (FRDA) are due to CAG, CTG, and GAA expansions, respectively (Brook et al., 1992; Campuzano et al., 1996; MacDonald et al., 1993). Decades of research focus on understanding the molecular mechanisms by which trinucleotide repeats cause the diseases. These diseases are caused by the expression of a toxic RNA and/or a toxic protein. Therefore, preventing their expression, for example by manipulating gene expression, or by reducing repeat sizes may be suitable therapeutic avenues. Currently, the knowledge of these phenomena is not well characterized and prevents the rational development of effective treatments.

Table I.3. Neurological disease caused by TNR expansion

Disease name	repeats	Repeat location
Huntington disease (HD)	CAG	Exon
Spinocerebellar ataxia 1 (SCA1)	CAG	Exon
Spinocerebellar ataxia 2 (SCA2)	CAG	Exon
Spinocerebellar ataxia 3 (SCA3)	CAG	Exon
Spinocerebellar ataxia 6 (SCA6)	CAG	Exon
Spinocerebellar ataxia 7 (SCA7)	CAG	Exon
Spinocerebellar ataxia 12 (SCA12)	CAG	5'-UTR
Spinocerebellar ataxia 17 (SCA17)	CAG	Exon
Spinal and bulbar muscular atrophy (SBMA)	CAG	Exon
Dentatorubral–pallidoluysian atrophy (DRPLA)	CAG	Exon
Huntington disease-like 2 (HDL2)	CTG	3'-UTR
Spinocerebellar ataxia 8 (SCA8)	CTG	3'-UTR
Myotonic dystrophy 1 (DM1)	CTG	3'-UTR
Fuchs endothelial corneal dystrophy (FECD)	CTG	Intron
Fragile X syndrome (FXS)	CGG	5'-UTR
Fragile X–associated tremor/ataxia syndrome (FXTAS)	CGG	5'-UTR
Fragile X–associated primary ovarian insufficiency (FXPOI)	CGG	5'-UTR
FRA7A	CGG	Intron
Fragile XE mental retardation syndrome (FRAXE MR)	GCC	5'-UTR
Friedreich's ataxia	GAA	Intron

Data were obtained from (Dion and Wilson, 2009; Mirkin, 2007; Schmidt and Pearson, 2016; Zhao and Usdin, 2015)

3.1 Mechanism of repeat instability

3.1.1 Trinucleotide repeats form a non-B structure that causes instability

Expanded TNRs are unstable. Below 35 units, the repeats loci tend to be stable, which would not cause disease. However, once the numbers of repeat units reach 35, the TNR expansion derived neurological disease appears (Castel et al., 2010). Worse still, longer repeats cause more severe phenotypes and passing down unstable expanded repeats to the offspring causes a more severe disease phenotype and an earlier age of onset, a phenomenon coined anticipation (McMurray, 2010; Mirkin, 2007).

The subset of trinucleotide repeats can form aberrant secondary DNA structures during transcription, DNA repair, and translation (Castel et al., 2010; Mirkin, 2007). The structures include mismatched hairpins, slipped-strands, R-loops, G-quadruplex, et al. (Mirkin, 2007; Usdin et al., 2015). The current models stipulate that the DNA repair machinery recognize these non-B structures as DNA lesions and initiate repair. The repetitive nature of the sequences, as well as their ability to form non-B DNA structure, make the repair error-prone, leading to instability. One of the first observation of non-B DNA structures formed by CAG/CTG repeats was done *in vitro* using chemical probing (Kohwi et al., 1993). They showed that even a single AGC could form unusual DNA structures independently of flanking sequences. Another milestone occurred when McMurray and colleagues in 1995 showed that CAG/CTG and CGG/CCG repeats could form hairpins *in vitro* in a length-dependent and sequence-specific manner (Gacy et al., 1995). Melting temperature profiles revealed that CAG, CGG and AT repeats could form stable hairpins, but AAG and AC cannot. *In vivo* evidence came first in 2010 from the Leffak lab. They used isogenic HeLa cells with CAG/CTG multiple repeats inserted after the Myc replication origin to uncover hairpins formation in 102 CAG/CTG repeats cells. Using ZFN_{CTG} and ZFN_{CAG}, secondary structure formed cell lines would be cleaved and showed repeat length change by small pool PCR (SP-PCR). Liu and colleagues revealed 102 CAG/CTG repeats cells formed hairpins but hardly in 12 repeats counterpart (Liu et al., 2010). Moreover, Axford et al. showed that they could precipitate slipped DNA from the expanded repeat locus in DM1 patient-derived lymphoblastoid cell lines

(Axford et al., 2013). Similar hairpin forming feature with CGG repeat was also observed *in vitro* (Nadel et al., 1995; Usdin and Woodford, 1995).

Slipped-strand DNA (S-DNA) is also extensively studied. Using electron microscopy as well as chemical and enzymatic probing followed by gel electrophoresis, slipped strands are readily visible under microscopy (Pearson et al., 1998, 2002). R-loops are DNA: RNA hybrids formed during transcription. They preferentially form at GC rich sequences, including CAG and CGG repeats (Lin et al., 2010; Loomis et al., 2014; Reddy et al., 2011, 2014). This structure also depends on sequence context since R-loops are not detected when AGG or AAG repeats are studied.

3.1.2 DNA repair is involved in TNR instability through processing of the non-B structure

Multiple lines of evidence show that virtually all DNA repair pathways influence trinucleotide repeat instability. The involvement of mismatch repair (MMR), nucleotide excision repair (NER), and base excision repair (BER) has a particularly important function in this process.

Mismatch Repair (MMR)

The primary substrates for the mismatch repair machinery are base mismatches and small DNA loops that lead to insertion and deletions (INDLs). They are typically formed during DNA replication (Kunkel and Erie, 2005). Depending on whether the lesions are single base mismatches or loops of mismatched DNA, MMR complexes MutSα (composed of MSH2/MSH6) or MutSβ (MSH2/MSH3) are called into play respectively to recognize and bind the bases and then recruit MutL proteins to repair the mismatches with the collaboration of several other enzymes. The primary procedure follows from recognition, intermediate binding, excision, amplification and DNA ligation. Both MutSα and MutSβ complexes with components involved in MutL indicate the role to destabilize CAG/CTG repeats from *E.coli*, yeast to HD and DM1 mouse models (Broek et al., 2002; Dandelot and Tomé, 2017; Foiry et al., 2006; Jaworski et al., 1995; Manley et al., 1999; Pinto et al., 2013; Schmidt and Pearson, 2016; Schweitzer and Livingston, 1997; Tomé et al., 2009, 2013).

Nucleotide Excision Repair (NER)

NER can be divided into two sub-pathways: transcription-coupled repair (TCR) in transcriptionally active regions and global genomic repair (GGR) that can take care of lesions in both silent and active loci (Hanawalt, 2002). GGR proteins are typically maintained in a relatively low abundance (Usdin et al., 2015). In contrast, TCR targets lesions in the transcribed strands of genes. The repair process also starts with the recognition of the lesion, which is different between the TCR and GGR pathways, but they both merge at the level of TFIIH recruitment. The lesion is then excised, creating a DNA gap that is then filled in by a polymerase and ligated. The NER pathway has significant impacts on repeat instability both in vivo and in cells. For example, Xpa knockout mice carrying 145 CAGs at the SCA1 locus showed dramatically reduced levels of instability in neuronal tissues, but not in peripheral organs compared to Xpa+/+ mice (Hubert et al., 2011). Besides, RNAi knockdown of XPA reduced high contraction frequencies in human cells (Lin et al., 2006). Moreover, Lin et al. found that the NER and MMR pathway components appear to crosstalk to cause contractions during a transcription-coupled NER-like reaction (Lin et al., 2006). The same is true if ERCC1 or XPG are knocked down (Lin and Wilson, 2007). RNAi knockdown of CSB, which recognizes the stalled DNA polymerase that triggers TCR, in human cells also decreased contraction frequencies (Lin and Wilson, 2007). By contrast, knockdown of XPC in these cells (Lin and Wilson, 2007) or knockout of Xpc in an HD mouse model did not affect repeat instability (Dragileva et al., 2009). These results implicate the TCR pathway in repeat instability and suggest that GGR has only a minor role if any.

Base Excision Repair (BER)

Base excision repair (BER) is the primary pathway that repairs oxidized bases (Banerjee et al., 2011; Zharkov, 2008). Loss of OGG1, a glycosylase that excises the common oxidized base 8-oxoguanine, decreased somatic expansions in a transgenic HD mouse model (Kovtun et al., 2007). This presumably worked by increasing the repair rates within the repeat tract. Moreover, oxidative stress induced by H_2O_2 treatment in human HD fibroblasts resulted in an MSH2 enrichment at the CAG repeats (Kovtun et al., 2004). These results indicate that MMR and BER pathways show crosstalk during TNR instability.

3.2 TNR expansion affects chromatin structure and gene expression

One intriguing observation is that changes in chromatin structure accompany repeat expansion through mechanisms that remain unclear (Dion and Wilson, 2009). Specifically, expansions tend to acquire heterochromatic marks. The first correlations between long trinucleotide repeat and heterochromatic-like chromatin marks were observed in FXS (Hornstra et al., 1993; Oberlé et al., 1991; Sutcliffe et al., 1992). Indeed, in FXS patient cells, the *FMR1* gene is abnormally methylated at the DNA level, which is thought to silence the gene. This is not seen in healthy individuals. Meanwhile, the hypermethylation in the *FMR1* promoter in male patient cells was even more significant than that observed in the inactive X of normal female cells (Stöger et al., 1997). ChIP experiments in FXS patient cells showed the accumulation of H3K9 methylation, a heterochromatic mark, together with a decrease in euchromatic histone marks H3K4me and H3K4ac (Coffee et al., 1999, 2002). As predicted by the model whereby chromatin structure changes decrease FMR1 expression, reversal of this heterochromatic state using histone deacetylase and DNA methyltransferase inhibitors led to a transient reactivation of the expanded repeat. (Abel and Zukin, 2008; Biacsi et al., 2008; Didonna and Opal, 2015).

Heterochromatinization of expanded repeats is not only restricted to CGG repeats. In DM1 patient cells, the expanded CTG tract is associated with the loss of a DNase I hypersensitive site in the nearby promoter region of the *SIX5* gene (Otten and Tapscott, 1995). Taking advantage of methylation-sensitive restriction enzymes, Steinbach et al. showed that abnormal CpG methylation is mapped in the flanking sequence of *DMPK* gene in DM1 patients with the congenital form of the disease (Steinbach et al., 1998). This was further confirmed by deep sequencing of bisulfite-treated DNA (Barbé et al., 2017). Further observations show that in DM1 patients fibroblasts, the CTG expanded repeat region sees an increase in antisense transcription, loss of CTCF binding, decreased H3K4 methylation, and an increase in H3K9me (Cho et al., 2005).

Moreover, inserting arrays of CD2 transgenes containing CTG or GAA expansions in the mouse genome led to silencing independently of the site of integration (Saveliev et al., 2003). By contrast, wild-type CD2 transgene arrays

displayed position-effect variegation. These data suggest that expansions cause the changes in chromatin structure.

Similarly, FRDA cells show hypoacetylation of H3 and H4 and H3K9 hypermethylation (Herman et al., 2006a). Moreover, several different histone deacetylase inhibitors could revert the acetylation loss and promote FXN expression in FRDA derived cells (Herman et al., 2006a). FRDA patients lymphoblast also underpinned three novel CpG methylations compare to non-affected patients (Greene et al., 2007). One of the three sites is known to be crucial for transcription initiation. It provides further evidence that the triplet repeat is causing heterochromatic structure influence not only transcription elongation but also initiation (Greene et al., 2007). All of these tell us TNR tract can produce a correlation with heterochromatic structure formation.

3.3 TNR disease treatment strategy

Until now, there is no cure for any of the expanded repeat diseases. Palliative treatments for expanded TNR disorders mainly focus on alleviating the symptoms. Importantly, the longer the trinucleotide repeat is, the earlier the age of onset. This was noted in HD patients as early as 1993 (Andrew et al., 1993; Duyao et al., 1993; Lee et al., 2012; Stine et al., 1993). About 50% variation in the age of onset can be attributed to repeat size in the blood (Holmans et al., 2017). The rest is attributed to repeat interruptions (Goldberg et al., 1995), *cis* elements (Warby et al., 2009), and/or *trans* acting genetic factors (Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium, 2015), all of which could be acting via altering repeat expansion rates. This association between age of onset and repeat length is not limited to HD but has also been reported for other expanded repeat disorders (Filla et al., 1996; Harley et al., 1993; Koide et al., 1994; Orr et al., 1993). However, since all of them are a repeat-driven problem, would that be more effective and sufficient to focus on repeat itself? In another word, can removal of repeat expansion reverse disease phenotype?

3.3.1 Genome editing may be possible to provide efficient treatment

The most obvious approaches to reducing repeat size are either to induce repeat contractions or to remove the track altogether using gene editing tools. The

Wilson group first tried this with ZFNs recognizing CAG/CTG repeats to induce DSBs. They found that this treatment dramatically increased the frequencies of contractions in a human cell line systems (Mittelman et al., 2009). It was unclear whether the cells also accumulated expansions because the assay could only detect contractions. Another approach used to contract the repeat tract in HD-derived iPSCs was by homologous recombination using a donor DNA with a short repeat tract. Although the frequencies of correction were extremely low, corrected cells showed improved cell survival and pathogenic HD signaling pathways recovery (caspase activity, cadherin, TGF-b, and BDNF). Importantly, the corrected cells were readily differentiated into neurons (An et al., 2012). These experiments open the door to a cell-based therapy. Our lab has shown that targeting the Cas9 nickase to CAG repeats within a GFP reporter can induce contractions without a significant increase in expansions (Cinesi et al., 2016). If targeting the Cas9 nickase to CAG repeats *in vivo* can also induce a bias towards contraction, then this strategy could prove to be useful in correcting the mutation leading to these devastating diseases.

Another possible approach of correcting expanded TNRs is to remove the TNR region altogether. AAV virus-mediated Cas9, and *HTT* exon1 specific sgRNA were injected into HD140Q-KI mice striatum at the age of 3 or 9 months (Yang et al.). They achieved permanent suppression of endogenous *mHTT* expression with HTT aggregates depletion and neuropathology recovery. A self-inactivating system, KamiCas9, with an additional sgRNA to block Cas9 translation, has been used to suppress mHTT expression and reverse aggregates formation in Ki140CAG HD mice (Merienne et al., 2017). To tackle the mutated HD allele-specific editing goal, 6 SNPs in *HTT* gene promoter regions were screened and applied to engineer *mHTT* exon 1 removal altogether specifically but not in normal *HTT* in HD patients fibroblasts and Bac HD mice which contains modified human HD allele and SNPs. Dabrowska and colleagues showed to use Cas9 nickase to excise all the CAG repeats from HD patient-derived fibroblasts can inactivate HTT expression in variable repeat length cells (Dabrowska et al., 2018). Not only in HD system, but similar Cas9 based gene editing to repeat expansion also achieved expansion tract removal and pathological recovery in myogenic capacity, nuclocytoplasmid distribution, RNA-binding capability in DM1 patient-derived myoblasts cells (van Agtmaal et al., 2017). Taken

together, removing CAG/CTG repeats from HD and DM1 locus showed promising improvement in molecule level and part of pathology status.

Expanded GAA repeats were excised, rather than contracted, using ZFNs that flanked the expansion in FRDA patient fibroblasts and iPSC-derived neurons. The GAA repeat in these patients is found within the intron of FRDA and leads to silencing (Campuzano et al., 1996; Filla et al., 1996). Thus, excision of the repeat region led to an increase in FRDA expression along with disease-associated biomarkers normalization and aconitase activity and intracellular ATP levels increase (Li et al., 2015). Ouellet and colleagues took advantage of Cas9 to excise GAA repeat expansion from *FRDA* gene and increase frataxin gene transcription and protein level in YG8R and YG8sR mouse models and their derived cell lines (Ouellet et al., 2017).

Taking advantage of CRISPR/Cas9-based gene editing, Park et al. induced a DSB within the expanded CGG repeat of FXS-derived iPSCs. In those clones that contained a contraction, they saw an increase in the *FMR1* mRNA and the up-regulation of FMRP. This was presumably due to the loss of heterochromatic marks in the promoter of *FMR1* (Park et al., 2015).

Thus, it is becoming feasible to precisely edit the repeat region. It remains to be seen whether these approaches will work *in vivo* and when during disease development they would be efficient and safe to use.

3.3.2 Epigenome editing may also provide a potential treatment

As discussed previously, chromatin structure changes have been observed in TNR expansion causing diseases mice models, and patients derived cells (Dion and Wilson, 2009). It would be logical to suspect chromatin modifying enzymes may take action during disease onset and progression based on those observations. Moreover, here histone deacetylases and DNA methyltransferases involvement in TNR disorders will be discussed.

3.3.2.1 Histone deacetylases (HDACs)

HDACs appear to play a central role in neurodegeneration and thus have been considered for treatment of multiple expanded TNR disorders (Butler and Bates,

2006; Soragni and Gottesfeld, 2016). For instance, HDAC inhibitors, TSA and SAHA, have been shown to reduce cell death and/or neurodegeneration in mouse motor neuron-neuroblastoma fusion cell lines (MN-1), Drosophila, and transgenic Huntington mouse model R6/2 (Hockly et al., 2003; McCampbell et al., 2001; Steffan et al., 2001). HDAC3 selective inhibitor, RGFP966, can minimize cognitive defects and suppress somatic repeat expansion in Hdh^{Q111} transgenic mice (Suelves et al., 2017). FRDA is another prime candidate for HDAC inhibition treatment. As described previously, FRDA is an autosomal recessive neurodegenerative disease due to GAA triplet expansion located in the first intron of FXN gene, which ends up with declined frataxin production (Yandim et al., 2013). Gottesfeld and colleagues demonstrated that one specific HDAC inhibitor, a 2-aminobenzamide derived compound, can increase levels of frataxin mRNA and protein in lymphoblastoid and primary lymphocytes (Herman et al., 2006b). Another HDAC inhibitor, 109, also showed to lead an increase in chromatin accessibility and *FXN* transcription production (Chutake et al., 2016). All these promising HDACi as drug treatment share a slow-on/slow-off kinetic profile and target class I HDAC primarily (Soragni and Gottesfeld, 2016). Compound 109 has been tested in FRDA patient fibroblasts, iPSCs, and neurons and proved successful in increasing frataxin expression at concentrations that were not significantly toxic (Soragni et al., 2014). In the context of DM1, a new flow cytometry-based screening in Hela cells revealed two more HDAC inhibitors, ISOX, and vorinostat, that can dramatically increase *MBNL1* mRNA level (Zhang et al., 2017). MBNL1 is a splicing factor that gets sequestered by mRNAs with an expanded CUG repeat, leading to mis-splicing of mRNAs *in trans* (Echeverria and Cooper, 2012). The two HDAC inhibitors also up-regulated MBNL1 protein levels and partially reversed the splicing defect in DM1 patient-derived fibroblasts (Zhang et al., 2017).

From all of these losses of function approaches, they failed to answer whether HDACs work locally to affect TNR or affect other protein transcription *in trans*. (Jia et al., 2016).

3.3.2.2 DNA methyltransferase (DNMTs)

There are five paralogues of DNA methyltransferases (DNMTs) in the human genome, DNMT1, DNMT2, DNMT3a, DNMT3b, and DNMT3L (Lyko, 2018). DNMT1,

3a, and 3b are cytosine-5 DNMTs, which catalyze and maintain DNA methylation. Their action closely tracks with gene silencing. By contrast, Dnmt2 targets RNA (Goll et al., 2006) and Dnmt3L (Bourc'his et al., 2001) does not possess catalytic activity. Aberrant DNA methylation and, by extension, DNMTs activity, are found in several expanded TNR diseases. For instance, as early as 1993, Knight et al. described aberrant CpG island hypermethylation at the FRAXE mental retardation locus upon CGG expansion (Knight et al., 1993). The Usdin group found three hypermethylated CpG residues in the FXN gene promoter region only in FRDA patient-derived cells (Greene et al., 2007). They also showed that expanded GAA repeats could influence transcription initiation (Kumari et al., 2011; Punga and Bühler, 2010). CpG hypermethylation was also found around the CTG expansion in DM1 cells, where it correlates with the congenital form of the disease (Barbé et al., 2017). In a knock-in mouse model of SCA1, CpG methylation levels at three sites around the CAG repeat tract were found to be sensitive to DNMT1 levels. Although there was a correlation between high levels of CpG methylation at these sites and instability in the germlines, there did not seem to be a drastic effect on the pathogenesis of the disease (Dion et al., 2008).

Treating patients with HDAC inhibitors and/or DNMT inhibitors is likely to lead to serious side effects given that HDACs and DNMTs are involved in regulating gene expression genome-wide. A more precise approach, therefore, will be to use epigenome editing. However, as seen above, the chromatin context can influence the efficiency of epigenome editing and expanded repeats come with an altered chromatin state. Thus, it is crucial to understand how repeat expansion affects the ability of chromatin modifying enzymes to modify gene expression locally.

3.4 How to study TNRs with an inducible targeting system

To understand how sequence contexts affecting epigenome editing, it is necessary to be able to target chromatin modifies to a locus that differs only by the size of the repeat tract. Classically targeting systems have used bacterial operators like LacO, TetO, and LexA. For instance, inserting arrays of LacOs allows visualizing chromatin dynamics, gene expression, and chromatin structure (Belmont, 2001; Carpenter et al., 2005; Rafalska-Metcalf and Janicki, 2007; Verschure et al., 2005).

However, the LacO array is prone to breakage (Jacome and Fernandez-Capetillo, 2011) and a repetitive sequence inserted near a CAG repeat may influence the latter's stability (Blackwood et al., 2010). Thus, we concluded that a repetitive array, like LacO or TetO, would not be ideal as a targeting system to study expanded CAG/CTG repeats.

The Bystricky lab described a non-repetitive alternative in 2014 (Saad et al., 2014). It involves a chromosome partitioning system form *B. cenocepacia,* in which a sequence of roughly 1kb, INT, that contains four binding sites for dimers of ParB (Dubarry et al., 2006). This initial binding initiates the oligomerization of the ParB protein, leading to the recruitment of up to 200 molecules, at least *in vitro* (Khare et al., 2004). After optimization for usage in yeast, INT-ParB system was used to monitor DSB repair in live cells (Saad et al., 2014). The authors found no increase in fragility and no effect on gene expression. We concluded that this was a better system to use for our purpose.

The problem remained that directly comparing the overexpression of a protein to a constitutive targeting at the expanded repeat would require two independent cell lines. This is not ideal and can be overcome by the use of a CIP system. We opted for the exogenous ABA-based CIP (Liang et al., 2011).

Finally, we needed an efficient and scalable way of monitoring the effect of targeting chromatin modifying enzymes on gene expression and repeat instability. One of the main challenges when working with expanded repeats is to size them accurately and rapidly. The gold standard, but tedious, a method to determine repeat length is small-pool polymerase chain reaction (SP-PCR) (Monckton et al., 1995). Since PCR amplification is biased towards shorter repeats, in SP-PCR the template DNA is diluted to just a few genomes per reaction. Since many reactions are then necessary to have a following number of allele measured, this is not an efficient way of screening the effect of chromatin modifiers. More recently, new long-read sequencings methods have been applied to tandem repeats (Bustos et al., 2016; Liu et al., 2017; Loomis et al., 2013; McFarland et al., 2015; Tsai et al., 2017). However,

the number of quantifiable events remains low and the cost high. Thus this is not suitable for our purposes.

Some plasmid-based instability assays exist that are average throughput and labor-intensive (Claassen and Lahue, 2007; Cleary and Pearson, 2003; Farrell and Lahue, 2006). These would have the added issues that they cannot measure both expansions and contractions at the same time and that the chromatin structure may be entirely different than if the sequence is integrated stably in the genome. A better approach, therefore, is to use a chromosomal reporter developed by the Wilson lab. The first generation was based on the observation that inserting a large CAG repeat within the intron of selectable genes like APRT and HPRT decreased their expression (Gorbunova et al., 2003; Lin et al., 2006). This works because the CAG repeat acts as an alternative exon whose strength depends on the size of the repeat, which acts as a splicing enhancer (Blencowe, 2000; Elrick et al., 1998; Hong and Li, 2002; Yeakley et al., 1996). When this CAG exon is included, it also includes 38bp downstream of the repeat tract that throws off the reading frame. Thus, starting with a large repeat tract, and thus an inactive gene, massive contractions can be readily quantified by selecting for the activity of the selectable marker. Here again, one caveat is that the assay only measures one class of events, in this case, rare massive contractions. Attempts to modify the assay such that expansions can be detected had limited impact and was impractical (Lin et al., 2005).

The latest generation of the assay is GFP-based. It contains an inducible promoter, and a GFP split into two exons with the intron of the mouse Pem1 gene inserted in between (Santillan et al., 2014). It was initially designed as a contraction assay, but our lab has shown that it can monitor both expansions and contractions at the same time within only a few days (Cinesi et al., 2016). The system will be introduced in detail in Chapter III Results.

The goal of this thesis was to build an inducible targeting assay to study the effect of repeat expansion on epigenome editing. I describe herein the engineering of the system and the use of this synthetic biology approach for uncovering novel mechanisms by which chromatin modifiers affect gene expression.

Chapter II

Method

Plasmids

The plasmids used to make the cell lines or transient transfection are found in Table II.1.

Cell culture conditions

The majority of the cell lines, including all the parental lines, used (Table II.1) were genotyped by Microsynth, AG (Switzerland) and found to be HEK293.2sus. They were free of mycoplasma as assayed by the Mycoplasma check service from GATC Biotech. The cells were maintained in DMEM containing 10% FBS, penicillin, and streptomycin, as well as the appropriate selection markers at the following concentrations 15 µg ml^{-1} blasticidin, 1µg ml^{-1} puromycin, 150µg ml^{-1} hygromycin, 400 µg ml^{-1} G418, 400 µg ml^{-1} zeocin at 37 °C with 5% CO_2. The ABA concentration used was 500 µM, unless otherwise indicated. Dox was used at a concentration of 2 µg ml^{-1}

Cell line construction

A schematic of cell line construction is found in Figure III. 7 and the lines are listed in Table S1. For each line, single clones were picked and verified for expression of ParB-ABI and PYL-fusions by western blotting using the protocol described before (Cinesi et al., 2016). Proteins were extracted in RIPA Buffer and scrapped from the plate. Centrifugation removed the cell debris. The supernatant was collected and further quantified using the Pierce BCA Protein Assay Kit (ThermoScientific). Proteins were then mixed with loading buffer and run onto a Tris-glycine 10% SDS PAGE gels. The proteins were then transferred to nitrocellulose membrane (Axonlab). The membranes were blocked using the Blocking Buffer for Fluorescent Western Blotting (Rockland) and then primary antibodies were added overnight. Membranes were then washed with PBS + 0.1% Tween at room temperature followed by the addition of the secondary antibody (1:2000). After washing with PBS, the fluorescent signal was detected using an Odyssey Imaging System (Li-CoR). All antibodies used are found in Table II.2. Besides, repeat sizes were verified using oVIN-0459 and oVIN-0460 with the UNG-based PCR reaction described before (Aeschbach and Dion, 2017) and then Sanger-sequenced by Microsynth AG (Switzerland).

PaB-ABI (pBY-008), PYL (pAB-NEO-PYL), PYL-DNMT1 (pAB(EXPR-PYL-DNMT1-NEO)), PYL-HDAC5 (pAB(EXPR-PYL-HDAC5-NEO)) and PYL-HDAC3 (pAB(EXPR-PYL-HDAC3-NEO)) were randomly inserted. The GFP-reporter cassette was inserted using the Flp-mediated method as the manufacturer's instructions (https://www.thermofisher.com/ch/en/home/references/protocols/proteins-expression-isolation-and-analysis/protein-expression-protocol/flp-in-system-for-generating-constitutive-expression-cell-lines.html#prot3 Thermo Scientific). Single colonies were picked and screened for zeocin sensitivity to ensure that the insertion site was correct.

Targeting assays

For targeting assays involving transient transfections, cells were plated onto poly-D-lysine-coated 12 well plates with 600'000 per well on the first day. Constructs were transfected using Lipofectamine 2000 or Lipofectamine 3000 (Thermofisher Scientific) with 1 µg per well on the same day. 6 hours after transfection, the medium was replaced with one containing dox and ABA or DMSO. 48h after the transfection, the cells were split, and dox with ABA or DMSO were added again. On the fifth day, samples were detached from the plate with PBS + 1 mM EDTA for flow cytometry analysis.

In the case of stable cell lines, cells were seeded at a density of 400'000 per well in 12-well plates. The media includes dox and ABA or DMSO. The medium was changed 48 hours later, and the cells were resuspended in 500µl PBS + 1 mM EDTA for Accuri (BD) flow cytometry analysis with 12517 events recorded.

For dCas9 targeting, CIT0 and CIT40 cells were seeded into coated 12 well plates to a density of 600’000 cells per well. dCas9-KRAB (pBY-040), dCas9-HDAC5 NT (pBY-038), or dCas9-HDAC5 CD (pBY-039) were transfected using Lipofectamine 2000 (Life technology) together with the sgRNAs in a 1:1 ratio. 6 hours later, the medium was changed for one containing dox. The medium was changed again on the third day, and fresh dox was added. The cells were then recovered and analyzed by flow cytometry.

Flow cytometry and analysis

We used an Accuri C6 flow cytometer from BD and measured the fluorescence in at least 12517 events for each treatment. The raw data was exported as FCS files and analyzed using FlowJo version 10.0.8r1. The statistical analysis was done using R studio version 3.4.0. All R scripts are attached in the Appendix.

Chromatin immunoprecipitation

For chromatin immunoprecipitation, CIT cells were treated as for the targeting experiments except that the number of cells used was 10 times higher and 10cm plates were used. At the end of the five days, formaldehyde was added to the medium to a final concentration of 1% and incubated with gentle shaking for 10 minutes at room temperature. The samples were then quenched with 0.125M PBS-glycine for 5 minutes at room temperature. Samples were then centrifuged, the supernatant was discarded, and the cell pellets were washed with ice-cold PBS twice. The samples were split into 10 million cells per aliquot and either used immediately or stored -75 °C for later use.

Sonication was done using a Bioruptor in 1 ml of cell lysis buffer (10 mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.25% Sodium lauroyl sarcosinate, protease inhibitor Complete EDTA free (Roche)) for 25 to 30 min. DNA shearing was visualized by agarose gel electrophoresis after crosslink reversal and RNase treatment. 20% of sonicated supernatant was used per IP with 3 µg anti-Flag (M2, Sigma), anti-PAN acetylated H3 (Merck), or anti-IgG (3E8, Santa Cruz Biotechnology) in 800 µl IP dilution buffer (1.25% Triton-X, 1 mM EDTA pH 8.0, 0.5 mM EGTA, 16.25 mM Tris-HCl, 137.5 mM NaCl, 1x protease inhibitor Complete EDTA free (Roche)) with 50 µl blocked 50% slurry of Protein G Sepharose 4 Fast Flow beads (GE healthcare). The samples were incubated at 4 °C overnight and then washed with progressively more stringent conditions (wash 1: 0.1% SDS, 1% Trition, 20mM Tris-HCl, 2mM EDTA, 300mM NaCl; wash 2: 0.1% SDS, 1% Trition, 20mM Tris-HCl, 2mM EDTA, 500mM NaCl; wash3: 1% NP-40, 1% Na-deoxycholate, 10mM Tris-HCl, 1mM EDTA, 250mM LiCl; wash4: 10mM Tris-HCl,

1mM EDTA). After the IP, the samples were de-crosslinked and purified using a QIAquick PCR purification kit (Qiagen) and analyzed using a qPCR.

Quantitative PCR

qPCR was performed with the FastStart Universal SYBR Green Master (Roche) using a 7900HT Fast Real-Time PCR System in a 384-Well Block Module (Applied Biosystems™). Primers used to detect enrichment at the INT sequence and *ACTA1* gene are listed below and in Table S3.

oVIN-0969 (INT)
5' TGAATACCATGCGCTCTA
oVIN-0970 (INT)
5' GCCGTTCGTGGCAGAGAT
oVIN-1075 (*ACTA1*)
5' AGCGCGGCTACAGCTTCAC
oVIN-1076 (*ACTA1*)
5' CAGCCGTGGCCATCTCTT

Ct values were analyzed using the SDS Software v2.4. The reported enrichments were obtained using the ΔCt method:

$$\Delta Ct[\text{normalized ChIP}] = (Ct[\text{ChIP}] - (Ct[\text{input}] - \log_2(\text{input dilution factor}))$$

$$\%\ \text{input} = 2^{-\Delta Ct[\text{normalized ChIP}]}$$

ΔΔCt[ABA PYL INT ChIP] = Ct[normalized ABA PYL INT ChIP] - Ct[normalized DMSO PYL INT ChIP]

$$\text{Fold difference} = 2^{-\Delta\Delta Ct[\text{ABA PYL INT ChIP}]}$$

Small-pool PCR

CIT40BYH5 and CIT89BYD cells were treated as for the targeting experiments but split, and the medium replaced every three days for one month. Cells were harvested, and genomic DNA was extracted using NucloSpin Tissue Kit (Macherey Nagel).

Small Pool PCR was performed as previously described (Aeschbach and Dion, 2017). Seven PCR reactions per sample, plus one negative control, were set up using the Phusion U Green Hot Start DNA Polymerase kit (Thermo Fisher Scientific) as follow: 1X reaction buffer, 0.2mM dNTP mix (without dTTP), 0.4mM dUTP, 0.5µM oVIN-460, 0.5µM oVIN-1425, 3% DMSO, 0.1U Uracil-DNA N-Glycosylase heat labile (Roche #11775367001), 0.4U Phusion Taq, 1ng of DNA, and H_2O to 10µL. The PCR program used consist of 20°C for 10 min, 95°C for 7 min, followed by 35 cycles at 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min 30 s, at the end a final extension of 72°C for 10 min. After the PCR run finish, 100µg ml^{-1} of Proteinase K was added to each reaction and incubated at 37°C for 1 hour.

PCR products were separated on a 2% agarose gel. The gel was washed twice for 20 min in alkaline transfer buffer (0.4M NaOH, 1M NaCl). The gel was then transferred overnight onto a charged membrane (Zeta-Probe GT Genomic Tested, Bio-Rad) using capillary action in alkaline transfer buffer. After transfer, the membrane was washed with a neutralization buffer (1.5M NaCl, 0.5M Tris base, pH 7.4) for 5 min, and added into a hybridization cylinder along with pre-warmed Ultrahyb buffer (Thermo Scientific). Salmon sperm DNA was added and the membrane, which was then incubated for at least one hour at 52°C.

The probe for hybridization was prepared using T4 PNK (New England BioLabs) as follow: 1X Reaction buffer, 50pmol of oVIN-100, 10mM [γ-^{32}P] ATP, 10U T4 PNK, and H_20 to 25µL. The probe was incubated at 37°C for 1 hour, followed by 10 minutes at 65°C. The probe was then added directly to the membrane and incubated at 52°C for 2 hours. After incubation, the membrane was washed four times for 20 min with a washing buffer (0.5X SSC, 0.1% SDS). Finally, the membrane was washed with 2X SSC, exposed to a phosphoscreen for 24 hours, and revealed with a Typhoon scannersphoimager.

Cas9 T7E1 test

CIT40 cells were seeded into coated 12 well plates to a density of 600'000 cells per well. Cas9 (pcDNA3.3-TOPO-hCas9) with sgRNAs were transfected in a 1:1 ratio using Lipofectamine 2000 (Life technology) for 6h followed after which the medium was changed and included dox. Cells were harvest 72h after transfection, and genomic DNA was extracted using the NucloSpin Tissue Kit (Macherey Nagel). Genomic regions targeted by the sgRNAs were amplified with oVIN-0755 and oVIN-0970, which listed in Table II.3. PCR products were purified with the NucloSpin Gel and PCR Clean-up Kit (MN) and quantified using the Nanodrop spectrophotometer. 200ng of purified PCR product was mixed with NEB buffer 2 and ddH_2O to 19µl, followed by 5 min 95°C, then a slow 2°C per second decrease from 95 to 85 °C followed by 0.1°C per second decrease in temperatures to 25°C and then paused at 15°C. The reaction mix was incubated with 1µl of T7 endonuclease I (NEB) for 15min at 37 °C. Treated samples were visualized on a 1% agarose gel.

Statistics

When analyzed flow cytometry data between ABA and DMSO treatment, we could not be certain they were normally distributed. I performed paired Student's t-test for all the mean values of each sample. The same statistics analysis was applied to different repeat length POI cell lines comparison. All the statistical analysis has been done in R studio 3.4.0. We used p-value < 0.05 as significant difference exist.

Table II.1. Cell line using and construction

Cell line name	Parent cell line	Transgenes	Plasmid used	Integration method	Resistance marker	Reference
T-REX Flp-in	HEK293	pFRT/*lac*Zeo		-	Blasticidin Zeocin	Thermo Fisher
$GFP(CAG)_0$	T-REX Flp-in	$GFP(CAG)_0$	$pGFP(CAG)_0$	Flp-mediated integration	Blasticidin Hygromycin	(Cinesi et al., 2016; Santillan et al., 2014)
$GFP(CAG)_{101}$	T-REX Flp-in	$GFP(CAG)_{101}$	$pGFP(CAG)_{101}$	Flp-mediated integration	Blasticidin Hygromycin	(Cinesi et al., 2016; Santillan et al., 2014)
$GFP(CAG)_0B$	$GFP(CAG)_0$	$GFP(CAG)_0$	$pGFP(CAG)_0$	Flp-mediated integration	Blasticidin Puromycin Hygromycin	This study
		ParB-ABI*	pBY-008	Random integration		
$GFP(CAG)_{101}B$	$GFP(CAG)_{101}$	$GFP(CAG)_{101}$	$pGFP(CAG)_{101}$	Flp-mediated integration	Blasticidin Puromycin Hygromycin	This study
		ParB-ABI*	pBY-008	Random integration		
CIT0	T-REX Flp-in	GFP-INT-CAG_0	pVIN-221	Flp-mediated integration	Blasticidin Hygromycin	This study
CIT40	T-REX Flp-in	GFP-INT-CAG_{40}	pVIN-117	Flp-mediated integration	Blasticidin Hygromycin	This study
CIT40B	CIT40	GFP-INT-CAG_{40}	pVIN-117	Site specific insertion	Blasticidin Puromycin Hygromycin	This study
		ParB-ABI*	pBY-008	Random integration		

*: Contains 3xHA tag and a NLS.
†: Contains 3xFLAG and a NLS.

Continued

Cell line name	Parent cell line	Transgenes	Plasmid used	Integration method	Resistance marker	Reference
HEKB	T-REX	ParB-ABI*	pBY-008	Random integration	Blasticidin Zeocin Puromycin	This study
HEKBY	HEKB	ParB-ABI*	pBY-008	Random integration	Blasticidin Zeocin Puromycin Neomycin	This study
		PYL†	pAB-NEO-PYL	Random integration		
CIT40BY	CIT40B	ParB-ABI*	pBY-008	Random integration	Blasticidin Puromycin Neomycin Hygromycin	This study
		PYL†	pAB-NEO-PYL	Random integration		
		GFP-INT-CAG_{40}	pVIN-117	Flp-mediated integration		
HEKBYD	HEKB	ParB-ABI*	pBY-008	Random integration	Blasticidin Zeocin Puromycin Neomycin	This study
		PYL-Dnmt1†	pAB(EXPR-PYL-DNMT1-NEO)	Random integration		
CIT16BYD	HEKBYD	ParB-ABI*	pBY-008	Random integration	Blasticidin Puromycin Neomycin Hygromycin	This study
		PYL-Dnmt1†	pAB(EXPR-PYL-DNMT1-NEO)	Random integration		
		GFP-INT-CAG_{16}	pBY-050	Flp-mediated integration		
CIT89BYD	HEKBYD	ParB-ABI*	pBY-008	Random integration	Blasticidin Puromycin Neomycin Hygromycin	This study
		PYL-Dnmt1†	pAB(EXPR-PYL-DNMT1-NEO)	Random integration		
		GFP-INT-CAG_{89}	pBY-018	Flp-mediated integration		

Continued

Cell line name	Parent cell line	Transgenes	Plasmid used	Integration method	Resistance marker	Reference
HEKBYH5	HEKB	ParB-ABI*	pBY-008	Random integration	Blasticidin Zeocin Puromycin Neomycin	This study
		PYL-HDAC5†	pAB(EXPR-PYL-HDAC5-NEO)	Random integration		
CIT16BYH5	HEKBYH5	ParB-ABI*	pBY-008	Random integration	Blasticidin Puromycin Neomycin Hygromycin	This study
		PYL-HDAC5†	pAB(EXPR-PYL-HDAC5-NEO)	Random integration		
		GFP-INT-CAG_{16}	pBY-050	Flp-mediated integration		
CIT59BYH5	HEKBYH5	ParB-ABI*	pBY-008	Random integration	Blasticidin Puromycin Neomycin Hygromycin	This study
		PYL-HDAC5†	pAB(EXPR-PYL-HDAC5-NEO)	Random integration		
		GFP-INT-CAG_{89}	pBY-018	Flp-mediated integration		
HEKBYH3	HKEB	ParB-ABI*	pBY-008	Random integration	Blasticidin Zeocin Puromycin Neomycin	This study
		PYL-HDAC3†	pAB(EXPR-PYL-HDAC3-NEO)	Random integration		
CIT16BYH3	HEKBYH3	ParB-ABI*	pBY-008	Random integration	Blasticidin Puromycin Neomycin Hygromycin	This study
		PYL-HDAC3†	pAB(EXPR-PYL-HDAC3-NEO)	Random integration		
		GFP-INT-CAG_{16}	pBY-050	Flp-mediated integration		
CIT89BYH3	HEKBYH3	ParB-ABI*	pBY-008	Random integration	Blasticidin Puromycin Neomycin Hygromycin	This study
		PYL-HDAC3†	pAB(EXPR-PYL-HDAC3-NEO)	Random integration		
		GFP-INT-CAG_{89}	pBY-018	Flp-mediated integration		

Table II.2. Antibodies used

Epitope	Company	Catalog number	Dilution	Assay
FLAG	Sigma-Aldrich	F1804-5MG	3 µg per IP 1:1000	ChIP WB
IgG	Santa Cruz Biotechnology	sc-69786	3 µg per IP	ChIP
Pan-acetylation of H3	Merck	#06-599	3 µg per IP	ChIP
HA	Roche	11 867 423 001	1:1000	WB
Histone H3	Abcam	ab1791	3 µg per IP	ChIP
Actin	Sigma-Aldrich	A2066-.2ML	1:2000	WB
PolII	Abcam	ab26721	3 µg per IP	ChIP

Table II.3. Primers used

Name	Sequence	Locus	Purpose	Reference
oVIN-0969	5'- TGAATACCATGCGCTCTA-3'	INT	ChIP-Q-PCR	This study
oVIN-0970	5'- GCCGTTCGTGGCAGAGAT-3'	INT	ChIP-Q-PCR	This study
oVIN-1075	5'- AGCGCGGCTACAGCTTCAC-3'	Actin	ChIP-Q-PCR	This study
oVIN-1076	5'- CAGCCGTGGCCATCTCTT-3'	Actin	ChIP-Q-PCR	This study
oVIN-0459	5'- AAGAGCTTCCCTTTACACAACG-3'	GFP reporter	CAG repeat amplification	(Cinesi et al., 2016)
oVIN-0460	5'- TCTGCAAATTCAGTGATGC-3'	GFP reporter	CAG repeat amplification	(Cinesi et al., 2016)
oVIN-1425	5'-GACCTCATACGAAGATAGGCTT-3'	GFP reporter	CAG repeat amplification	This study
oVIN-0100	5'-AGCAGCAGCAGCAGCAGCAGCAGCAGC-3'	GFP reporter	CAG repeat probe	This study
oVIN-0755	5'-GACCTCATACGAAGATAGGCTT-3'	GFP reporter	INT amplification	This study
oVIN-0970	5'-GCCGTTCGTGGCAGAGAT-3'	GFP reporter	INT amplification	This study

Chapter III

Results

1. Targeting system design and construction

The inducible targeting system can be divided into three major parts. It contains a GFP reporter assay to monitor CAG repeat instability and expression, a ParB-INT protein-DNA binding system and the ABA-induced proximity system (Figure III. 1 A-C). Each component was engineered *in vitro* and stably integrated into Flp-in T-REX (HEK293) cells.

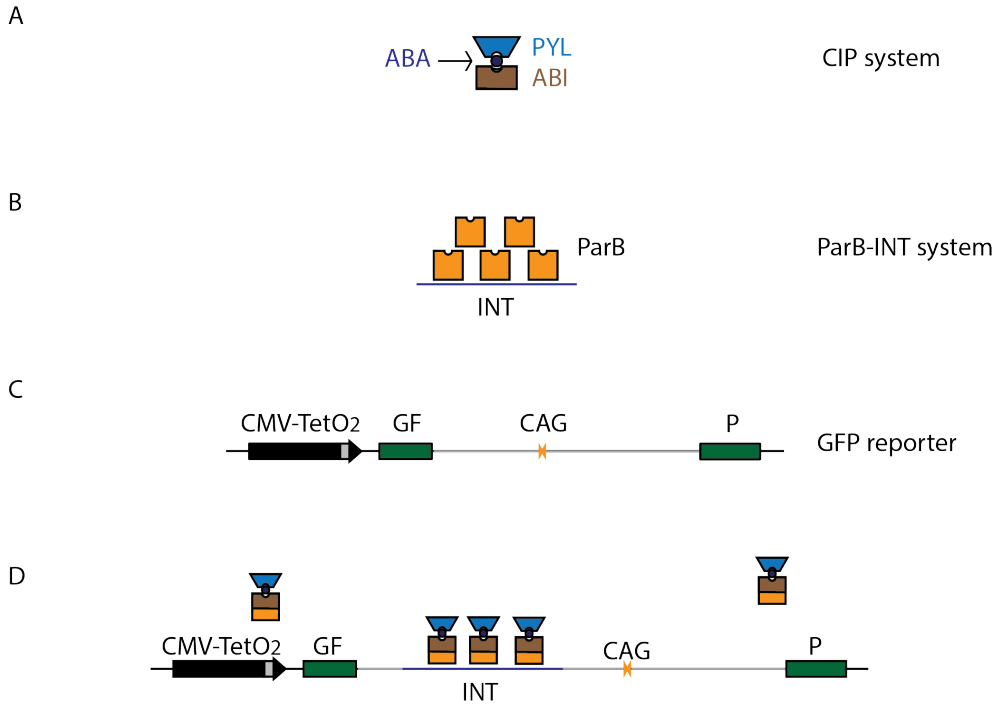


Figure III. 1. Three components of the inducible targeting system.

A. Abscisic acid (ABA)-inducible proximity system. PYL can bind to ABI when ABA is present.
B. ParB-INT protein DNA binding assay. Bacteria-derived ParB can bind to INT in a sequence-dependent manner and nucleate the oligomerization of ParB at the locus.
C. GFP reporter. Doxycycline addition can induce transcription along the locus and thus GFP expression.
D. Illustration of the CIT system structure. INT sequence is integrated into GFP reporter intron where ParB could bind. PYL forms a dimer with ParB-ABI fusion protein when ABA is present in the system.

We took advantage of the GFP-based reporter assay published by Santillan and colleagues in 2014 and further broadened its usage to monitor CAG repeat instability (Cinesi et al., 2016; Santillan et al., 2014) (Figure III. 2). This GFP reporter consists of a $TetO_2$ CMV inducible promoter with a mouse Pem1 intron flanked by two exons coding for GFP. One DM1 patient-derived CTG repeats containing fragment was cloned into the Pem1 intron in the CAG orientation, including 72 bp upstream and 93 bp downstream of patient-derived material. As being a splicing signal enhancer, CA dinucleotide influence GFP splicing with a reverse correlation between CAG repeat length and GFP expression (Tacke and Manley, 1999). A similar strategy based on the HPRT mini-gene showed that as CAG repeats increase in length, they promote the inclusion of an alternative 'CAG exon' that contains the repeat tract minus the first CAG and 38bp downstream of the CAG repeat. mRNAs containing the CAG exon are degraded, presumably via nonsense-mediated decay (Gorbunova et al., 2003). In conclusion, GFP intensity can act as a readout of CAG repeats instability and/or expression and is easily detected by flow cytometry.

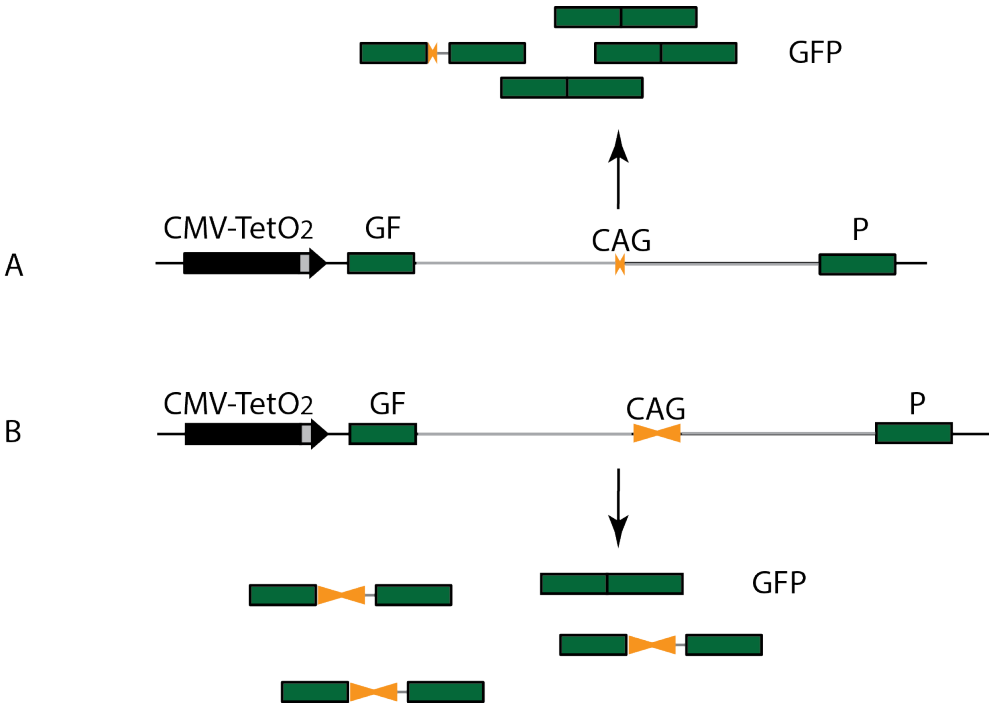


Figure III. 2. GFP reporter. The report contains an inducible CMV $TetO_2$ promoter (black box) and a split GFP gene (exons in green, the intron in grey) with a varying repeat size (orange).

A. After the induction of transcription by the addition of dox, the short repeat construct produces high levels of GFP.
B. Expanded CAG repeats tend to be included into mRNA during transcription. This CAG repeat containing transcript would end up with degradation. The longer the CAG repeat, the lower the amount of GFP produced.

To build an inducible targeting system with as little interference with the repeat tract as possible, we opted for a non-repetitive bacterial-derived system developed by the Saad et al. in yeast (Saad et al., 2014) and adapted it here for mammalian use. This system (Figure III. 1B), called ParB-INT, was first used to visualize the dynamics in double-strand break resection. Initially, P31 from *Burkholderia cenocepacia* J231 binds to chromosome 2 via a p*arS* DNA motif to activate mitotic segregation in bacteria (Dubarry et al., 2006). Saad and colleagues named the optimized DNA binding sequence INT and DNA binding protein as ParB. ParB initially binds to 4 nucleation sites within INT and leads to oligomerization beyond INT itself such up to two hundred ParB molecules can be recruited to a single site *in vitro* (Khare et al., 2004). The relatively small size of INT (< 1kb) and the absence of fragility seen in yeast or toxicity make ParB-INT ideal for our purposes (Mariamé et al., 2018).

We chose sequence optimized ParBc2 for use in human cells. The INT fragment was cloned in the intron of the GFP reporter construct 289 bp downstream of the first GFP exon and 308 bp upstream of the CAG repeat. Using the Flp-in integration method, we generated stable cell lines in Flp-in T-Rex (HEK293) cells, called chromatin Inducible Targeting (CIT), harboring the GFP-INT construct containing varying numbers of CAG repeats.

To test the effect of inserting the INT sequencing within the Pem1 intron of the GFP construct, we monitored GFP expression in GFP$(CAG)_0$, which do not have an INT sequence, and CIT0 cells by flow cytometry. This was done in the presence of dox to activate transcription of the reporter. We found that there was a minimal effect (Figure III. 3) and conclude that INT sequence insertion has a minor effect, if any, on GFP expression.

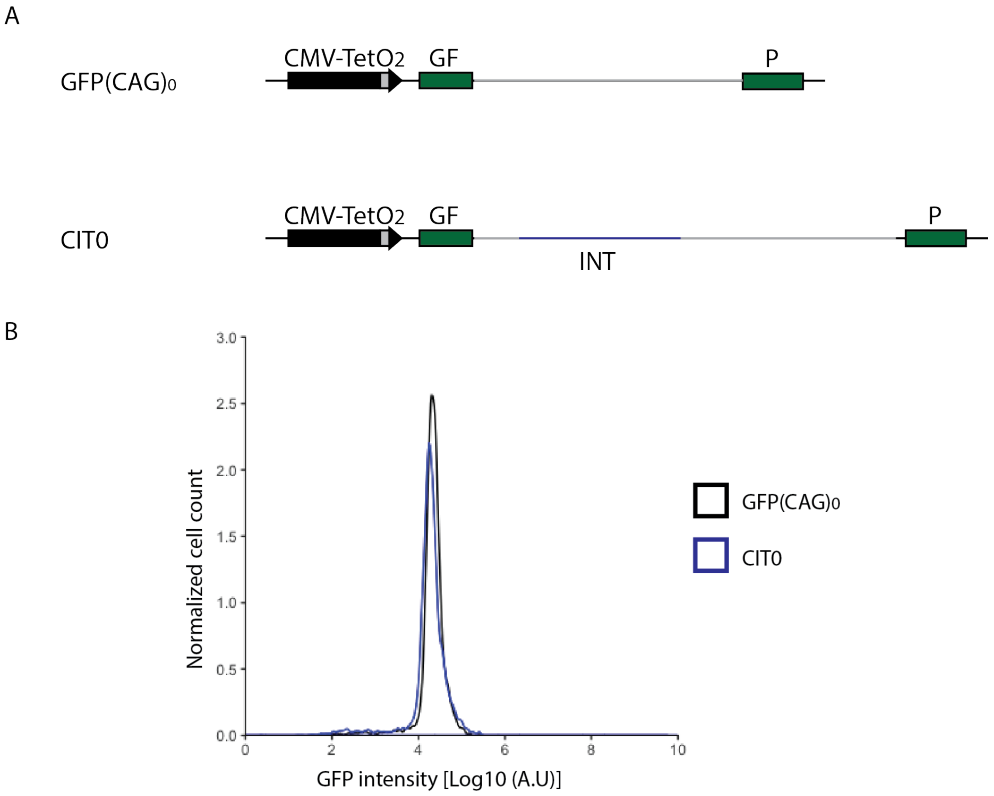
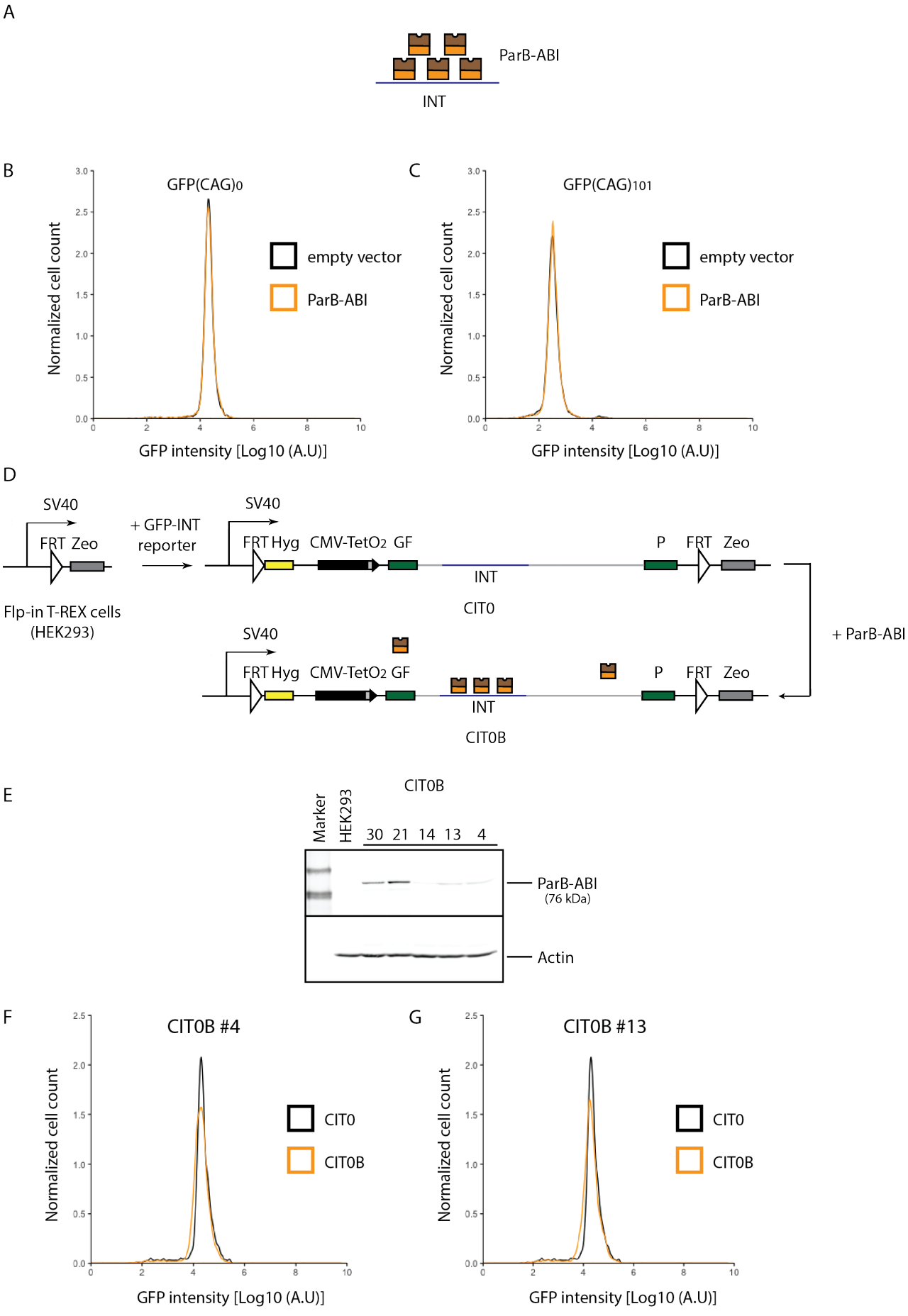


Figure III. 3. INT has a minor effect on GFP expression. Comparison between GFP(CAG)$_0$ and CIT0 cells.

A. Illustration of GFP constructs with and without the INT sequence. INT sequence (blue) is located in GFP reporter intron (grey) in CIT0 cell line.
B. Density plot of GFP in cells with and without INT. Black: GFP(CAG)$_0$ cells, blue: CIT0 cells. The x-axis indicates log10 scale GFP intensity. Y-axis is normalized cell count.

We next tested whether the ParB-ABI fusion by itself interfered with GFP expression. We, therefore, have transfected plasmid expressing the fusion in GFP$(CAG)_0$ and GFP$(CAG)_{101}$ cells and measured GFP produced five days later in the presence of dox. We found that ParB-ABI had no detectable effect on GFP expression in the conditions (Figure III. 4ABC). However, when both the INT sequence and the ParB-ABI proteins are stably integrated into the cells' genome to create CITnB, which n is the number of repeats, GFP expression decreased with increasing amount of the fusion. This is evidence that ParB-ABI is indeed binding to the INT in human cells. However, the effect on expression suggests that ParB-ABI binding to INT may block transcription or alter splicing, which is in contrast to what was found in yeast (Saad et al., 2014).


A
ParB-ABI
INT
B
GFP(CAG)0
Normalized cell count
empty vector
ParB-ABI
GFP intensity [Log10 (A.U)]
C
GFP(CAG)101
Normalized cell count
empty vector
ParB-ABI
GFP intensity [Log10 (A.U)]
D
SV40
FRT Zeo
Flp-in T-REX cells
(HEK293)
+ GFP-INT
reporter
SV40
FRT Hyg CMV-TetO2 GF
INT
CIT0
P FRT Zeo
+ ParB-ABI
SV40
FRT Hyg CMV-TetO2 GF
INT
CIT0B
P FRT Zeo
E
CIT0B
Marker
HEK293
30 21 14 13 4
ParB-ABI
(76 kDa)
Actin
F
CIT0B #4
Normalized cell count
CIT0
CIT0B
GFP intensity [Log10 (A.U)]
G
CIT0B #13
Normalized cell count
CIT0
CIT0B
GFP intensity [Log10 (A.U)]

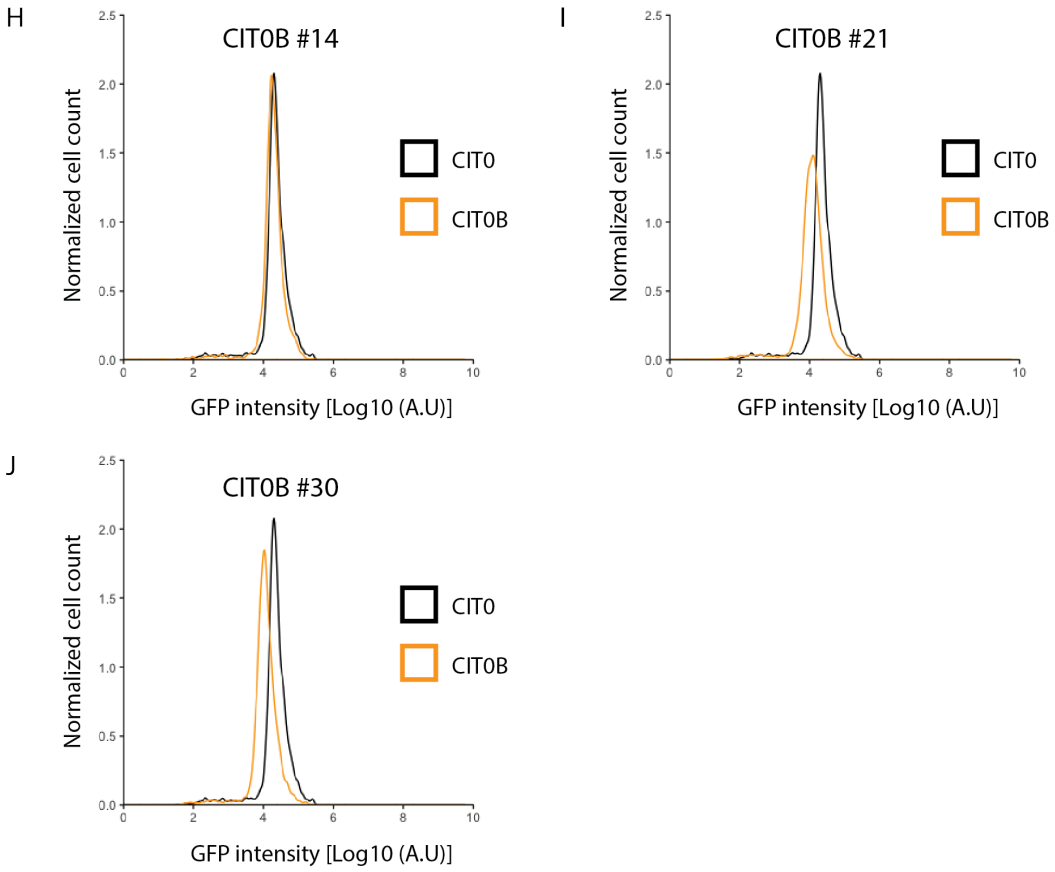


Figure III. 4. ParB expression and its effect on GFP expression.

A. Illustration of ParB-INT targeting. ParB-ABI fusion proteins (brown and orange) bind to INT sequence (blue).

B-C. Transient transfection of the ParB-ABI constructs to GFP$(CAG)_0$ (B), GFP$(CAG)_{101}$ (C) hardly affects GFP intensity compare to the transfection of empty vectors.

D. Cell line construction. Flp-in T-REX cells were stably transfected with GFP-INT reporter in a site-specific manner to construct CIT0 cell lines. CIT0 cell line was further stably transfected with the ParB-ABI construct to achieve CIT0B cell line.

E. ParB-ABI expression level in CIT0B clones. Numbers below CIT0B indicate clone numbers.

F-J. Stably expressed ParB-ABI construct in CIT0 induces GFP intensity shift. As ParB-ABI expression levels shown in E, clones with more ParB-ABI expression show greater GFP intensity difference compare to CIT0 control.

Since showed ParB-INT binding could affect GFP intensity, it was essential to determine whether any further effect was seen upon targeting of a control protein. Given that the presence of ABA drives the targeting, we first tested whether ABA by itself modified GFP expression. We found that adding up to 500μM of ABA to GFP$(CAG)_0$ and GFP$(CAG)_{101}$ cells did not affect GFP expression (Figure III. 5). We used 500μM of ABA in all our experiments because we found that this induced dimerization by co-immunoprecipitation (not shown) and is in line with the concentrations used by Liang et al. (Liang et al., 2011).

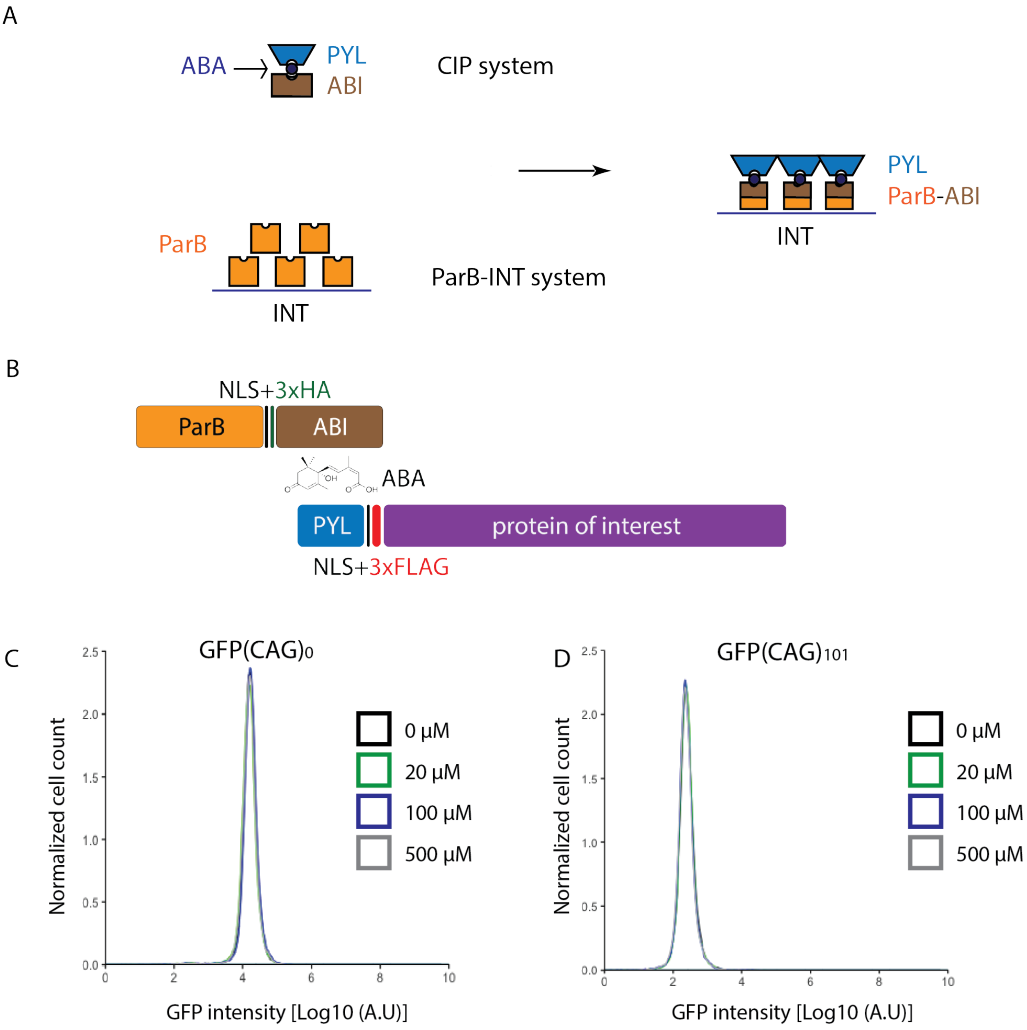


Figure III. 5. ABA does not affect GFP expression.

A. Cartoon of CIT. The fusion of ABI (brown) with ParB (orange) make CIP system and ParB-INT system into inducible. The presence of ABA brings PYL to INT sequence showing on the right.

B. Schematics of ABA-induced dimerization and the tags found on ParB-ABI and PYL-fusions.

C-D. ABA titration test in $GFP(CAG)_0$ (C) and $GFP(CAG)_{101}$ (D). None of the ABA concentration tested change GFP intensity compare to the untreated group.

Next, we introduced PYL alone or PYL fusion proteins (all containing 3 FLAG tags and an NLS) into the genome of HEK293 T-Rex FlpIN cells already containing ParB-PYL (HEKB cells) to generate HEKBY cells. To these, we introduced the GFP-INT constructs with either 16 CAGs or an expanded repeat (CITnBY cells – Figure III. 6A). This was done in collaboration with Alicia Borgeaud. This strategy allows us to compare the effect on targeting next to different repeat lengths directly. We used the same strategy to generate PYL-HDAC5 (CITnBYH5), PYL-HDAC3 (CITnBYH3) and PYL-DNMT1 (CITnBYD) expressing cells. All cell lines were verified with western blot for ParB-ABI and PYL levels, the repeat tract was sequenced, and the cells were tested for Zeocin sensitivity to ensure that the integration site was correct for the Flp-IN insertions. All stable cell lines generated are listed in Table II.1 and Figure III. 7.

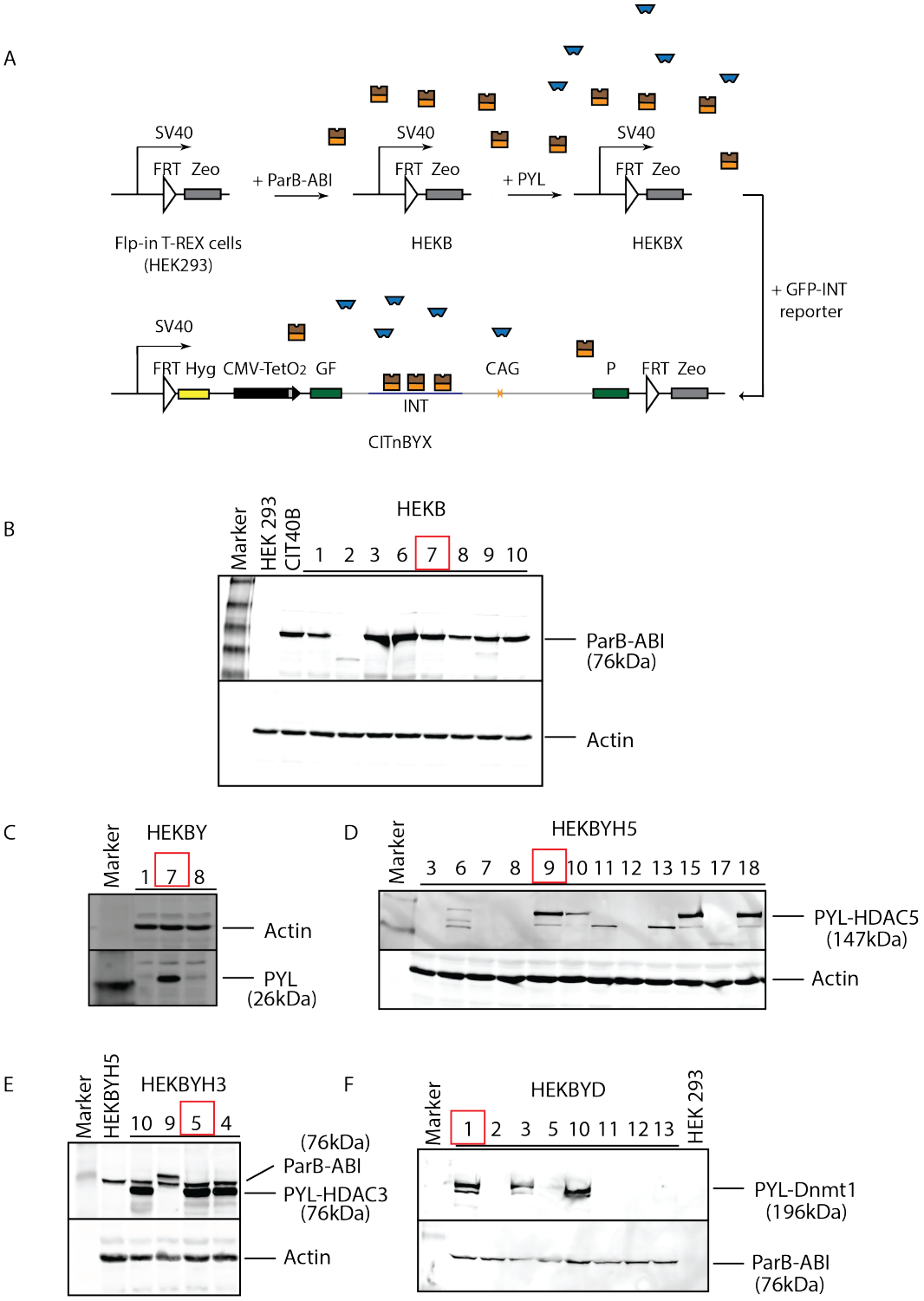


Figure III. 6. Cell line construction with protein expression analysis.

A. CIT cell lines construction. ParB-ABI, PYL/PYL-protein of interest and GFP reporter constructs are stably transfected into Flp-in T-REX cells in a sequential manner.

B. HEKB clones and their levels of ParB-ABI. Actin is shown as a loading control. The chosen clone is shown with a red box.

C. PYL expression in HEKBY clones. Actin is shown as a loading control. The chosen clone is shown with a red box.

D. PYL-HDAC5 expression in HEKBYH5 clones. Actin is shown as a loading control. The chosen clone is shown with a red box.

E. PYL-HDAC3 expression in HEKBYH3 cells. Actin is shown as a loading control. The chosen clone is shown with a red box.

F. PYL-DNMT1 expression in HEKBYD cells. Actin is shown as a loading control. The chosen clone is shown with a red box.

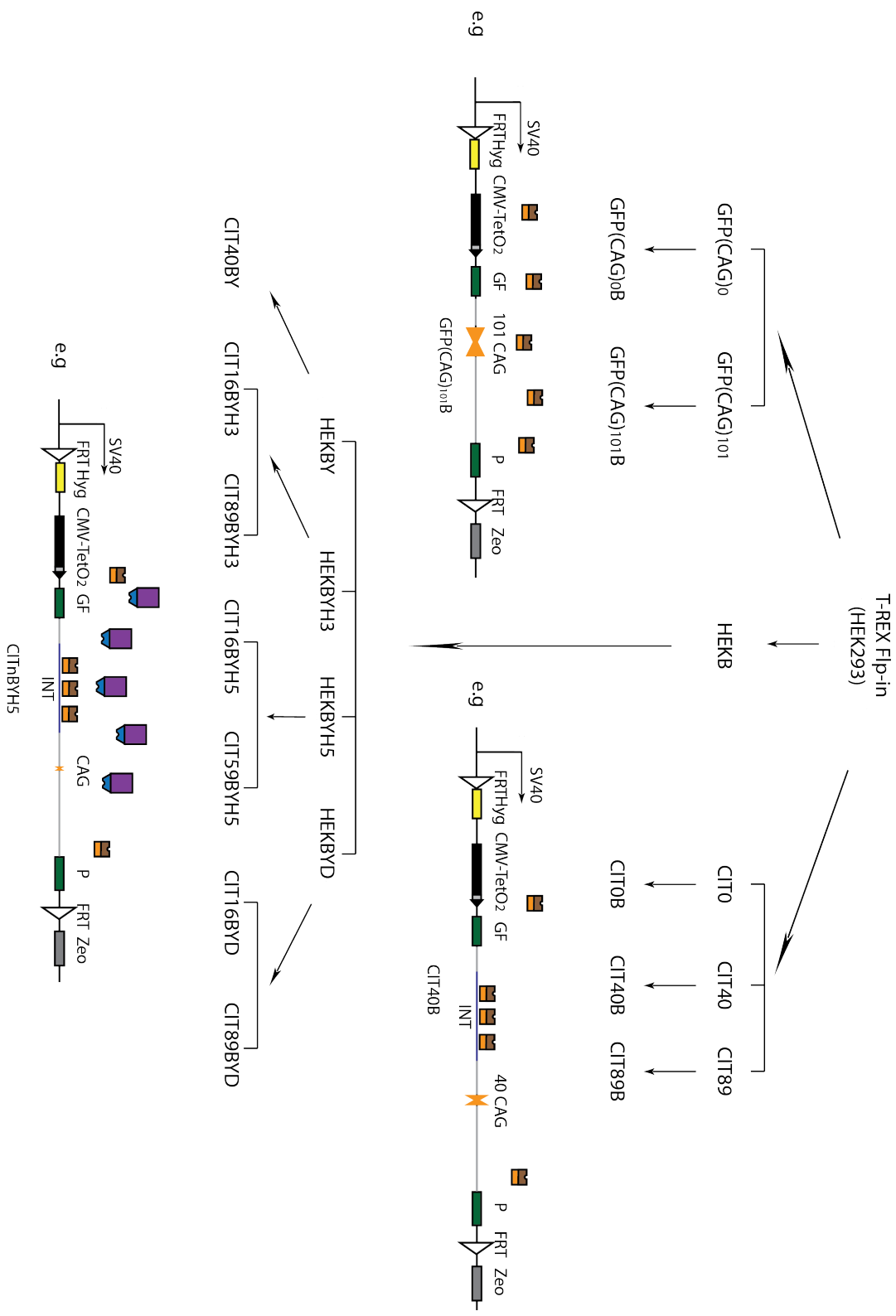


Figure III. 7. Stable cell line generation process with example cell line cartoons.

2. PYL targeting does not change GFP expression in CIT system

I next determined the effect of targeting PYL alone to the INT locus (Figure III. 8A). The working procedure is showing in Figure III. 8B. We found that there was no significant effect on the amount of GFP produced with or without ABA in CIT40BY cells. As expected, PYL was efficiently recruited to the INT sequence by ChIP upon addition of ABA by over nine-fold in CIT40BY cells but remained unchanged and at deficient levels at the ACTA1 locus, showing the specificity and robustness of the targeting (Figure III. 9). These results led me to conclude that targeting of PYL does not interfere further with GFP expression.

Due to unforeseen issues, the PYL control cells with 16 CAG and 89 CAG are on-going as of this writing. Therefore, I use below CIT40BY cells as a comparison.

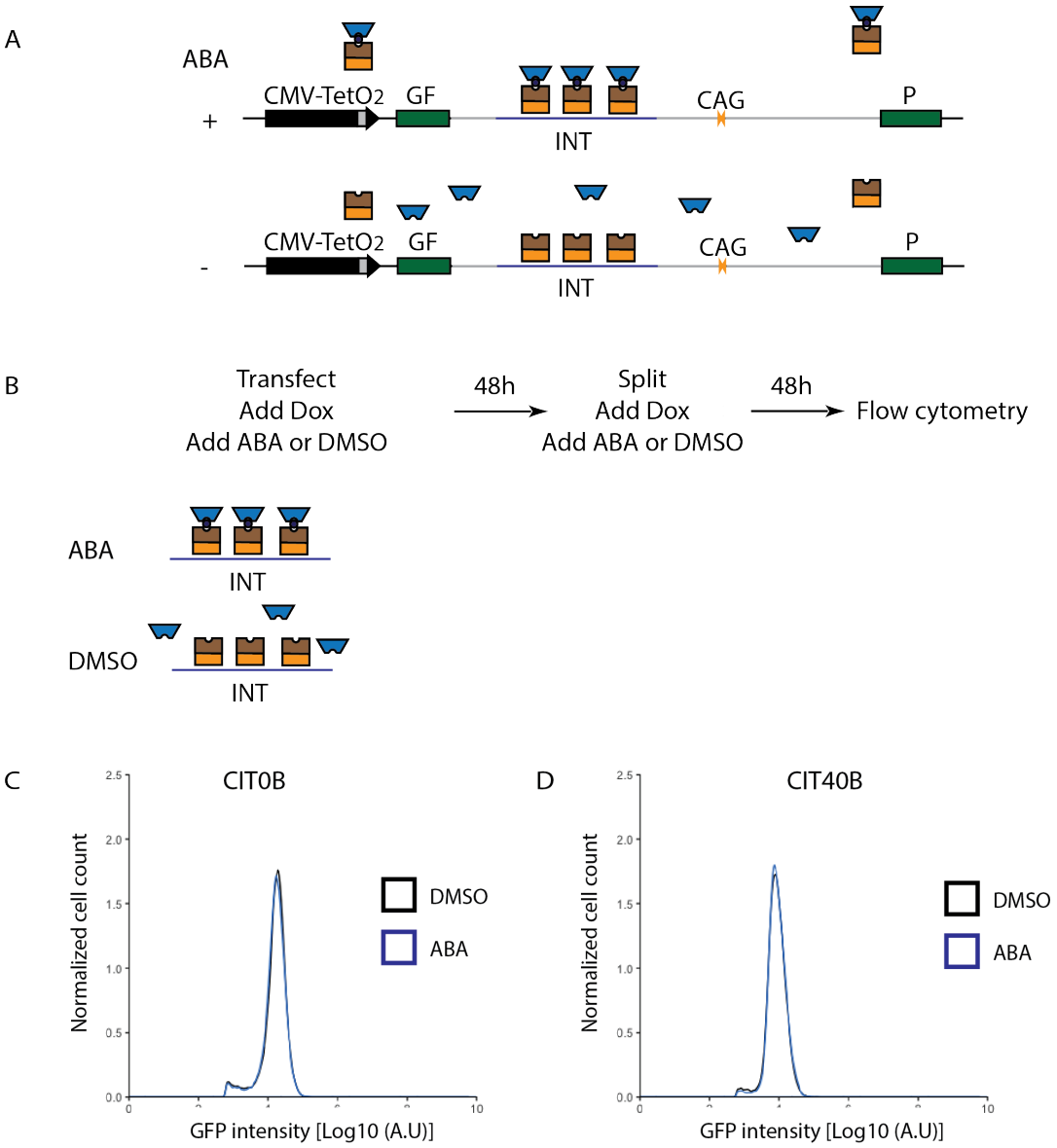


Figure III. 8. Transient transfection of PYL does not change GFP intensity upon targeting in CIT0B and CIT40B cells.

A. Illustration of PYL targeting. With ABA addition, PYL (blue) forms a dimer with ParB-ABI protein (brown and orange) and further target to INT sequence in GFP reporter. With no ABA presence, the PYL targeting would not occur.

B. Experimental design. Cells were transfected with PYL construct and drugs being added. After 48 hours, drugs were changed with cell medium. On the fifth day, cells passed through the flow cytometer and GFP intensity was recorded.

C-D. PYL targeting showed no GFP intensity shift in CIT0B (C) and CIT40B (D) cell lines compared to DMSO–treated cells.

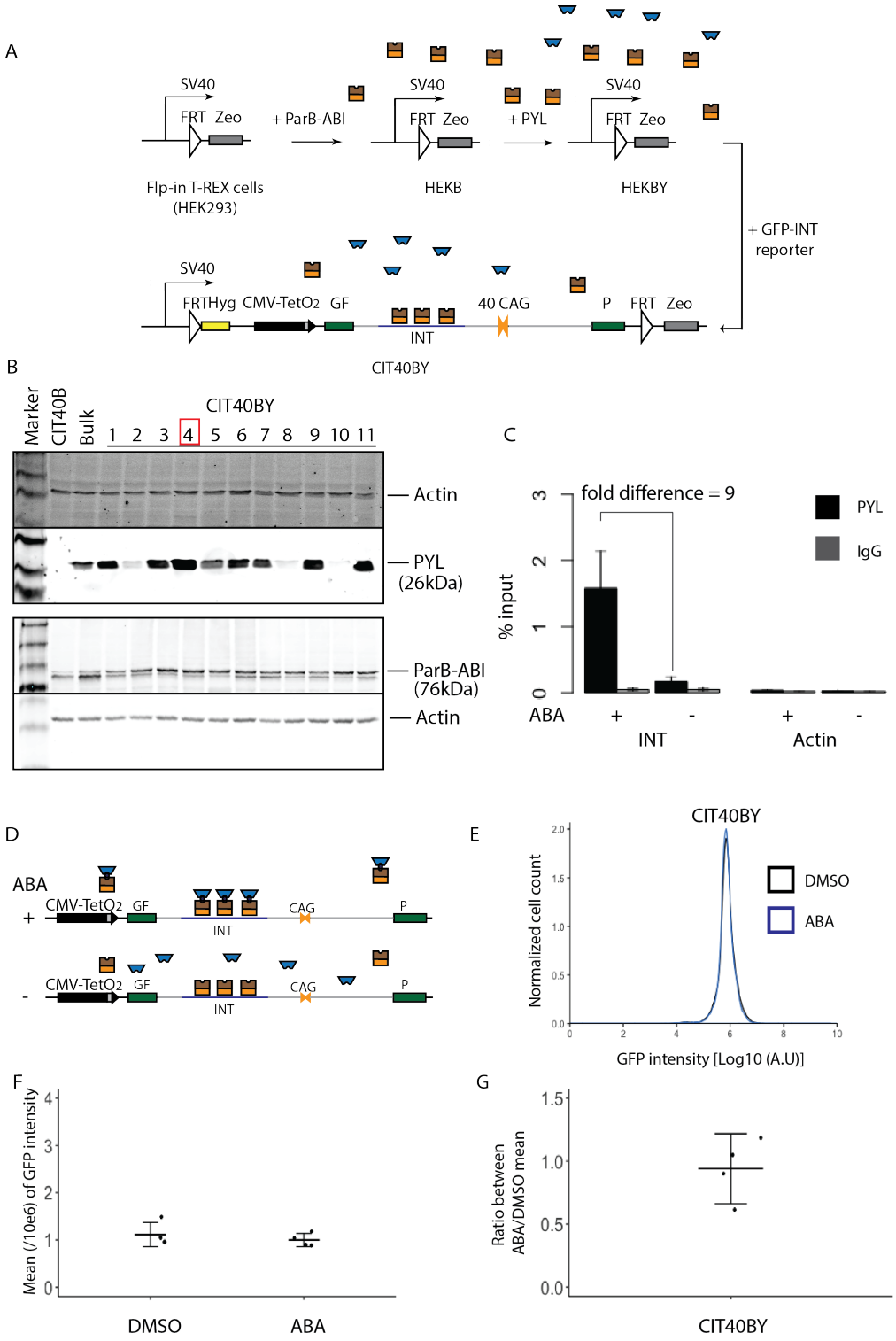


Figure III. 9. PYL targeting in CITnBY cells do not affect GFP expression.

A. Cell line construction. ParB-ABI, PYL and GFP reporter constructs are stably transfected into Flp-in T-REX cells in a sequential manner.

B. Protein expression pattern in the cells.

C. ChIP-qPCR experiment at INT and Actin locus upon ABA addition or not using an antibody against PYL construct. N=4. Error bars stand for standard error.

D. Illustration of PYL targeting and non-targeting scenarios.

E. Typical flow cytometry profile for targeting of PYL.

F-G Quantification of data in E for CIT40BY cells treated with ABA or DMSO. N=4. The error bars represent the standard error and the mean.

3. Histone deacetylase 5 (HDAC5) targeting decreases GFP expression

HDAC5 was found to promote CAG repeat expansion in a plasmid-based assay for repeat instability in immortalized astrocytes (Gannon et al., 2012). The mechanism of action remains unclear, and Gannon et al. postulated that it might not be because of local changes in chromatin structure. CIT is ideal to test this hypothesis directly. If indeed targeting of HDAC5 acts locally, we expected to find a shoulder on the dark side of the GFP curve that corresponds to expansions (Cinesi et al., 2016). We first tested the hypothesis by transient transfection of PYL-HDAC5 in CIT40B cells (Figure III. 10A). We found that a bulge appears in the GFP distribution upon ABA addition (Figure III. 10D), which prompted us to sort individual cells with less GFP, expecting to find expansions. However, all clones generated in this experiment had 40 repeats. Besides, we found that the culture lost the characteristic shoulder, as expected if this effect depended on expansions when the targeting was released, and the GFP was degraded over a time course of 3 weeks (Alicia Borgeaud Master Thesis). Together these results suggest that the GFP intensity decrease seen upon targeting HDAC5 is not due to CAG instability.

To control for efficient HDAC5 targeting efficiency when adding ABA, I performed ChIP-qPCR after transient transfection to determine if the INT sequence is enriched. Unfortunately, I could never manage to achieve ChIP-qPCR signal over background level or co-immunoprecipitating PYL and ParB-ABI in these conditions (data not shown). This is likely due to a low abundance of HDAC5 being introduced into the cell by transient transfection.

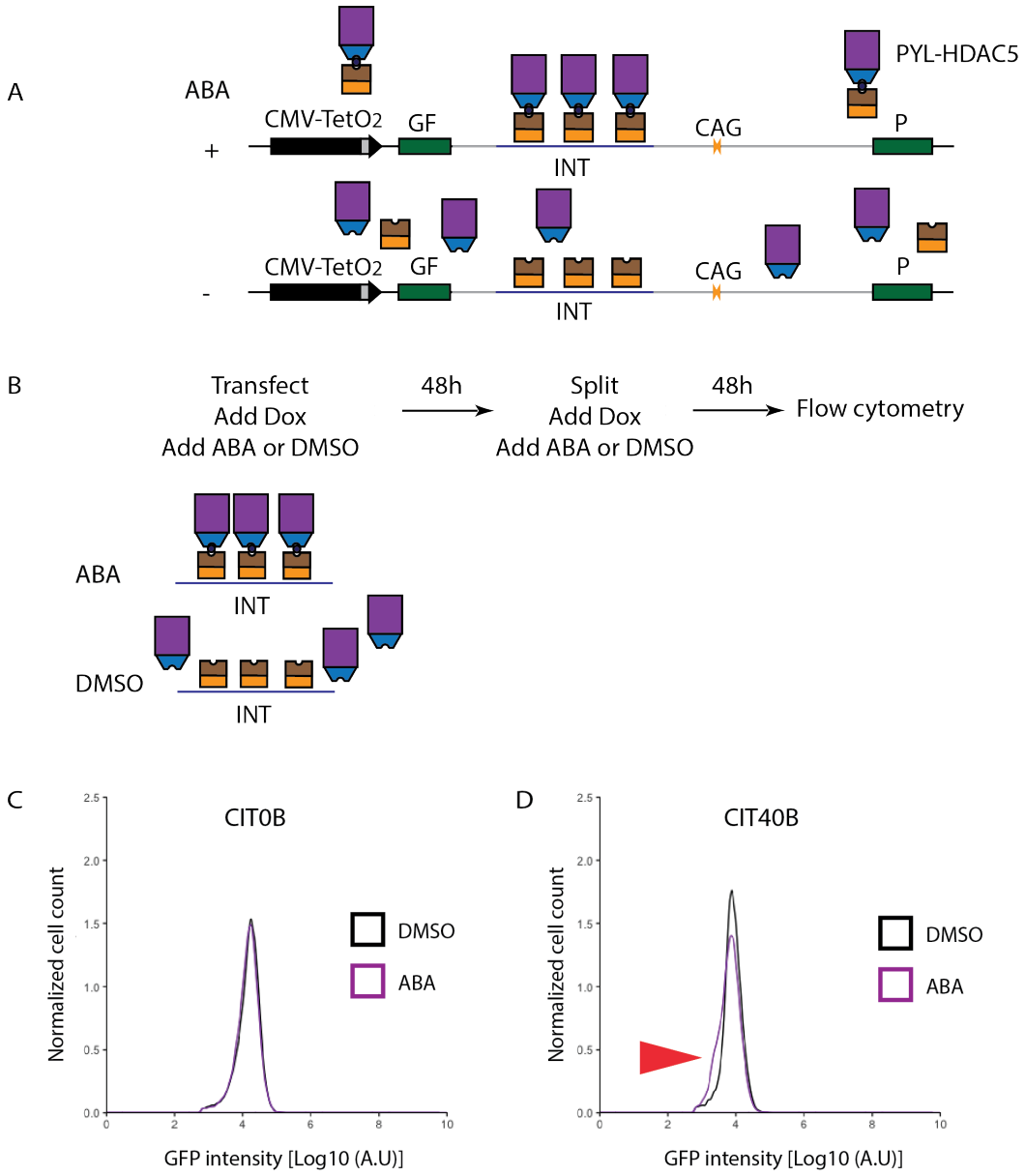


Figure III. 10. Transient transfection and targeting of PYL-HDAC5 downregulate GFP expression.

A. Illustration of PYL-HDAC5 targeting. With ABA addition, PYL-HDAC5 (blue and purple) forms dimer with ParB-ABI protein (brown and orange) and further target to INT sequence in GFP reporter. With no ABA presence, the PYL-HDAC5 targeting would not occur.

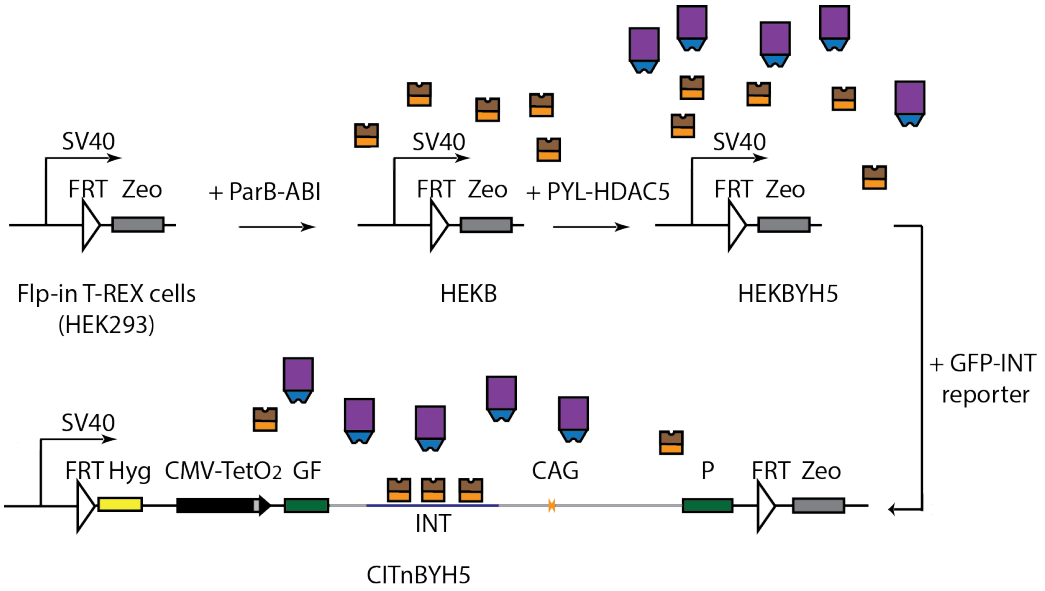
B. Experimental design. Cells were transfected with PYL-HDAC5 construct and drugs being added. After 48 hours, drugs were changed with cell medium. On the fifth day, cells passed through the flow cytometer and GFP intensity was recorded.

C. Effect of targeting PYL-HDAC5 on GFP expression in CIT0B.

D. Effect of targeting PYL-HDAC5 on GFP expression in CIT40B cells. Arrowhead indicates the change in expression seen upon targeting.

To overcome some of the limitations of transient transfections, I generated CIT16BYH5 and CITBY59H5 stable cell lines (Figure III. 11AB).

A



B

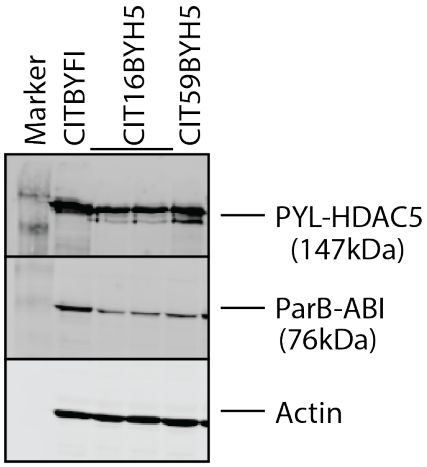


Figure III. 11. CITnBYH5 stable cell line making.

A. Cell line construction. ParB-ABI, PYL-HDAC5 and GFP reporter constructs are stably transfected into Flp-in T-REX cells in a sequential manner.

B. Transgene expression in CITnBYH5 cells.

In CIT16BYH5 cells, I found a robust enrichment of 39 fold in targeting of PYL-HDAC5 upon addition of ABA by ChIP-qPCR (Figure III. 12B). This was specific to the INT locus because the levels of PYL-HDAC5 remained unchanged at the ACTA1 locus. We conclude that PYL-HDAC5 is robustly targeted upon ABA addition in our stable cell lines.

To test the effect of PYL-HDAC5 targeting on GFP expression, I added dox as well as ABA or DMSO to CIT16BYH5 cells (Figure III. 12C). I found over a three-fold decrease in expression upon ABA addition compared to DMSO controls (Figure III. 12DE – $P < 0.01$).

The HDAC5 targeting dependent silencing I observed may occur due to HDAC5 deacetylase activity. To test this hypothesis, I quantified acetylated histone H3 locally at the INT insertion using ChIP-qPCR, using an antibody recognizing pan-acetylated histone H3 in CIT16BYH5 cells, I found a8n eight-fold decrease in the acetylation levels upon ABA addition. ChIP using an antibody against histone H3 indicated that this acH3 enrichment was not due to changes in H3 levels at this locus (Figure III. 12F). This indicates that targeting HDAC5 to the INT sequence led to deacetylation of H3 and suggested that the PYL-HDAC5 construct is functional.

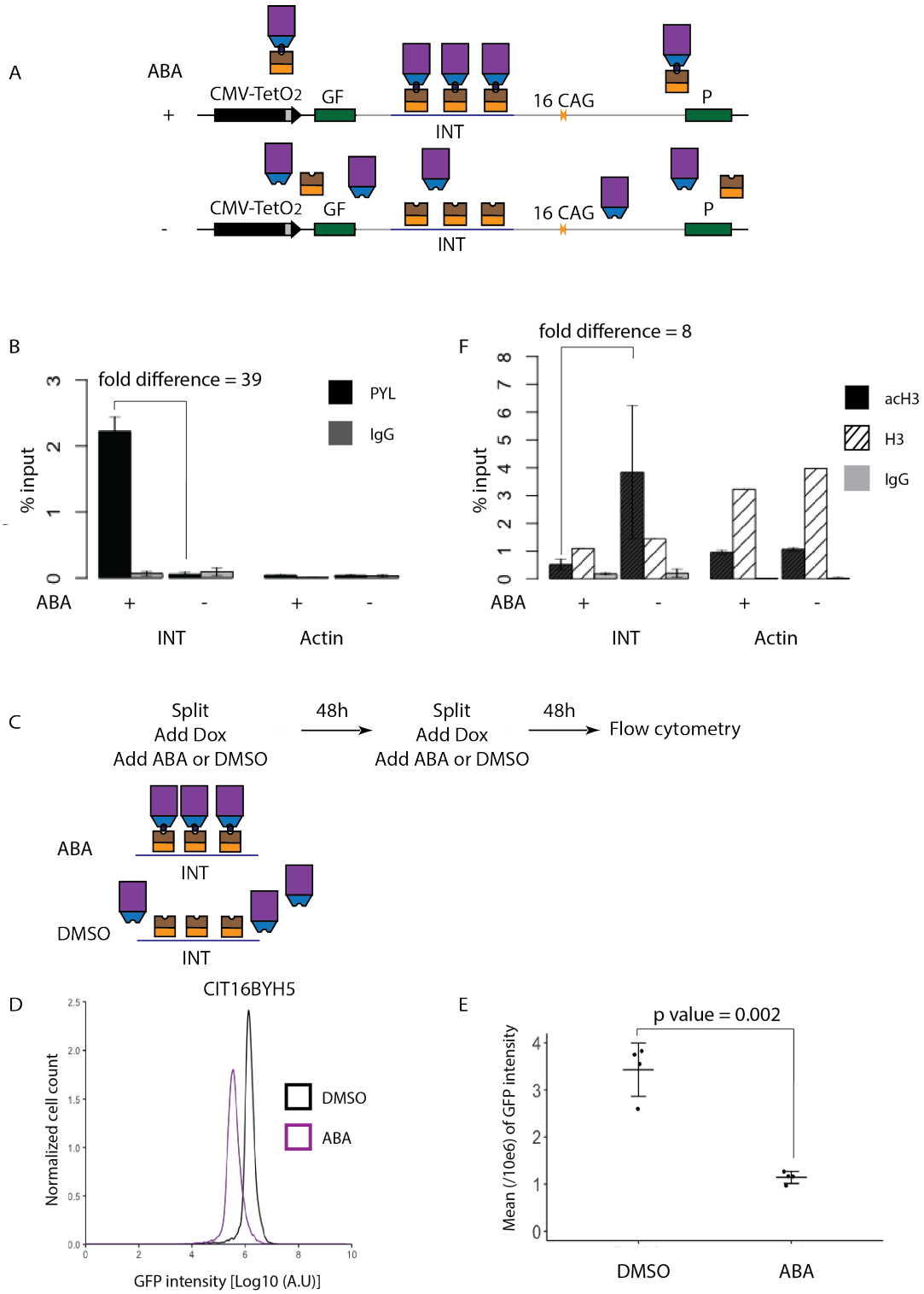


Figure III. 12. PYL-HDAC5 silencing of GFP expression in CIT16BYH5 cells.

A. Cartoon of CIT16BYH5 with and without ABA.
B. ChIP-qPCR experiment at INT and Actin locus in CIT16BYH5 cell line upon ABA addition or not using the antibody against PYL-HDAC5 construct. N=4. Error bars stand for standard error.
C. Experimental design.
D. Typical flow cytometry profile for targeting in CIT16BYH5 cells.
E. Quantification of the data shown in D. N=4 for each condition. Error bars stand for standard deviation.
F. ChIP-qPCR experiment at INT and Actin locus in CIT16BYH5 cell line upon ABA addition or not using the antibody against acH3 (N=2) and and H3 (N=1). Error bars stand for standard error.

Interestingly, the results upon targeting PYL-HDAC5 in CIT59BYH5 cells suggested a repeat-size-dependent effect on GFP expression. I first quantified the shift in GFP expression upon PYL-HDAC5 targeting (Figure III. 13C). I found that HDAC5 targeting in this cell line led to a two-fold decrease in expression, which was significantly less than the threefold seen in CIT16BYH5 cells (Figure III. 12E, $P<0.01$). This was not due to differences in protein levels between the two cell lines (Figure III. 11B) or differences in targeting efficiencies (Figure III. 13E). ChIP efficiency was higher in the cell line with the longer CAG tract (Figure III. 12B, 13E). These data suggest that the efficiency of HDAC5 deacetylation is dependent on the sequence context. To test this directly, I ChIPed acetylated H3 in CIT59BYH5 cells and found that the decrease was only two-fold, i.e., not as dramatic as the sevenfold decrease seen in CIT16BYH5 cells (Figure III. 12F, 13F). Here again, my H3 ChIP indicates the acH3 enrichment difference was not due to H3 default level (Figure III. 13F). In another word, the addition of a 129 bp (59 CAG vs 16 CAG) can significantly induce local chromatin modifications and affect how HDAC5 deacetylates its target.

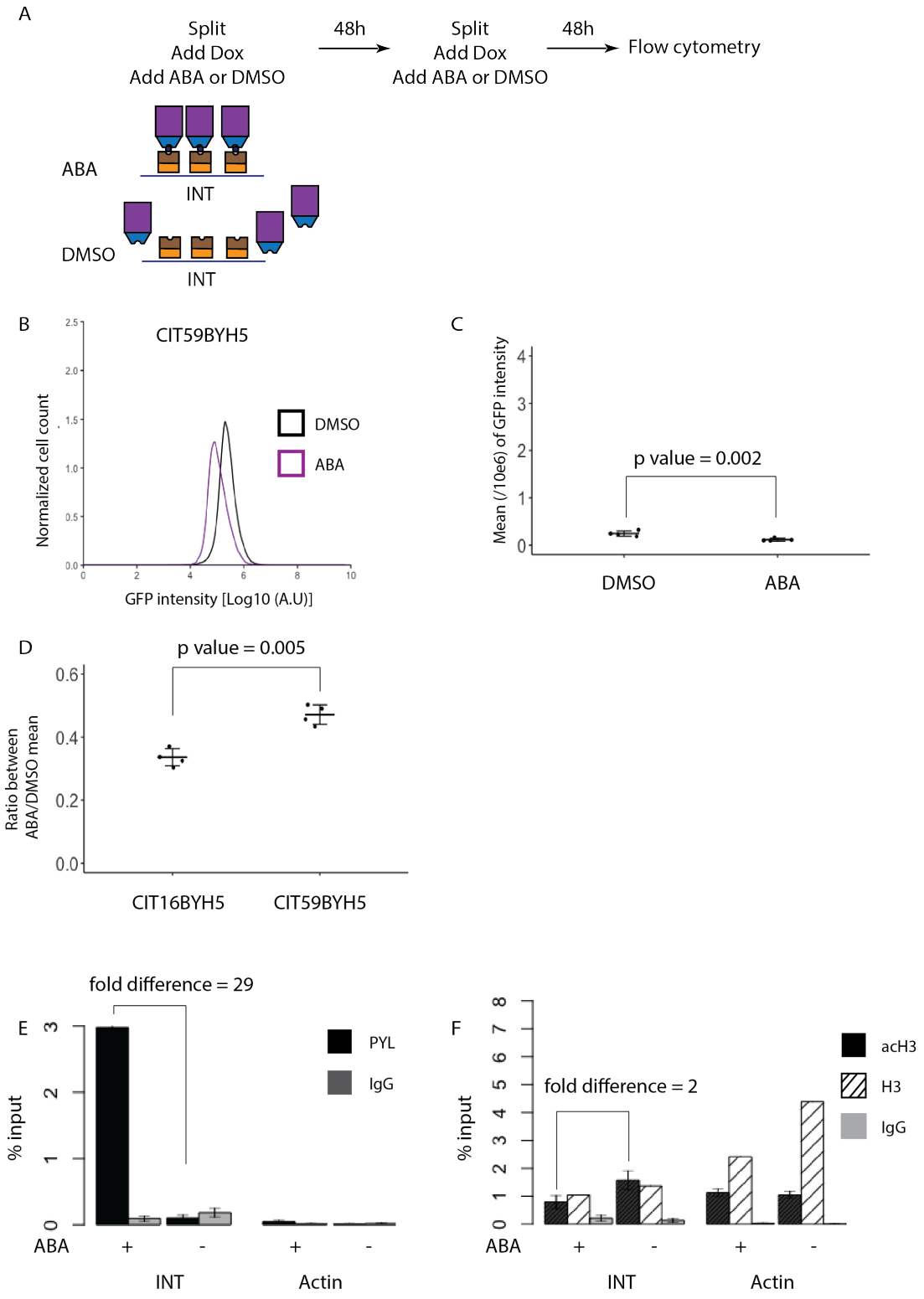


Figure III. 13. PYL-HDAC5 targeting reduces GFP expression.

A. Experimental design.
B. Typical flow cytometry profile for targeting in CIT59BYH5 cells.
C. Quantification of the data shown in B. N=4 for each condition.
D. The ratio of the mean comparison of the targeting effect in CIT16BYH5 and CIT59BYH5 cells.
E. ChIP-qPCR of PYL-HDAC5 in CIT59BYH5. N=4
F. ChIP-qPCR of acH3 (N=2) and H3 (N=1) in CIT59BYH5.

There is still a possibility that the effect we observed is not due to the dimerization of ParB-ABI and PYL-HDAC5 at the INT sequence. To test this possibility, I generated cell lines without an INT sequence but expressing ParB-ABI stably: GFP$(CAG)_0$B and GFP$(CAG)_{101}$B cells (Figure III. 14A). I then transiently transfected PYL-HDAC5 into GFP$(CAG)_0$B cells and GFP$(CAG)_{101}$B cells and measured GFP intensity after five days culture with dox and ABA or DMSO (Figure III. 14B). I found no change in GFP expression (Figure III. 14CD) in contrast to what was seen in CIT40B cells (Figure III. 10D).

Overall, our results suggest that HDAC5 deacetylation activity depends on sequence context.

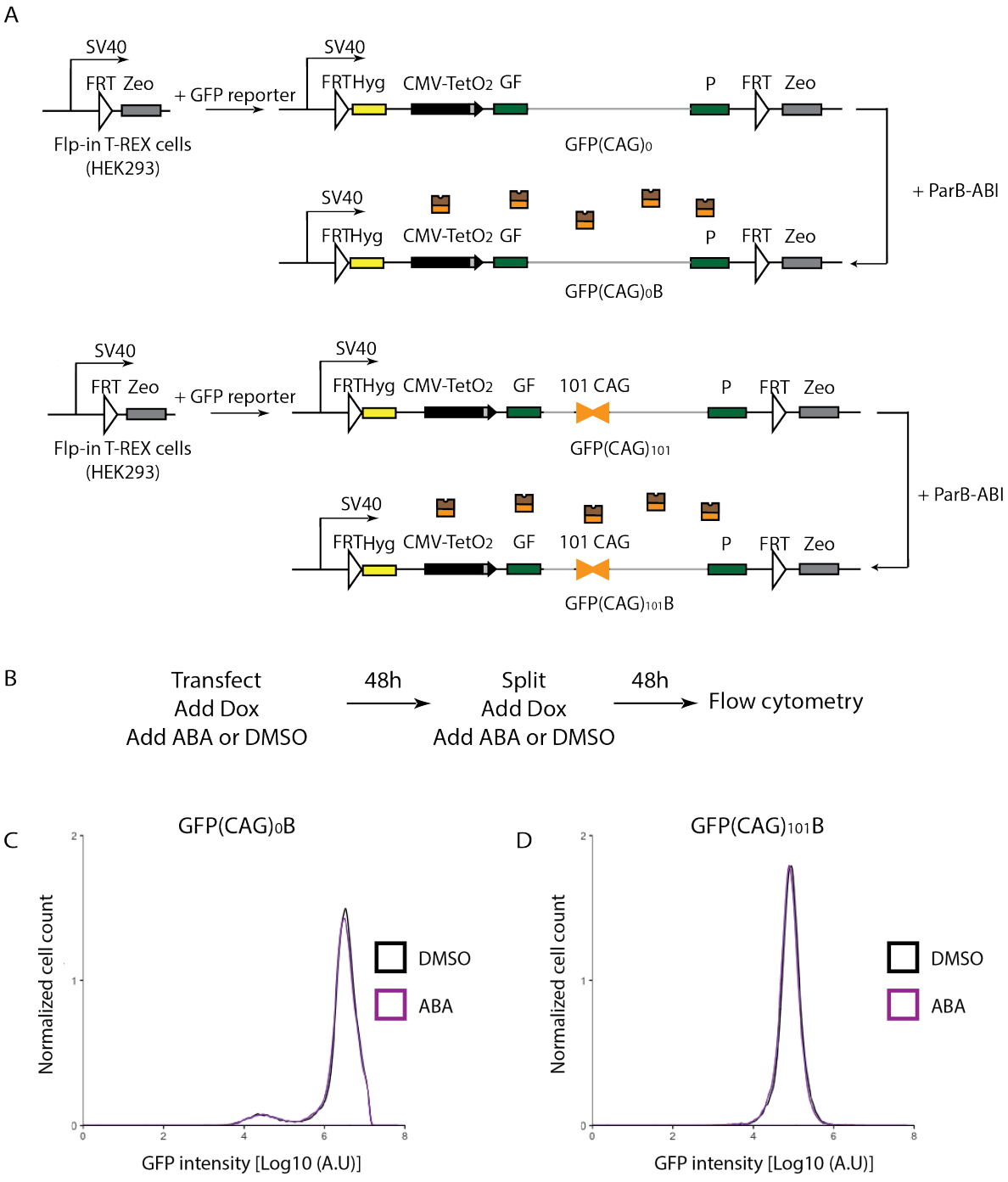


Figure III. 14. The effect of PYL-HDAC5 targeting on GFP expression depends on the presence of the INT sequence.

A. Cell line construction. GFP reporter and ParB-ABI constructs are stably transfected into Flp-in T-REX cells in a sequential manner.

B. Experimental design.

C-D. Representative flow cytometry profiles for $GFP(CAG)_0B$ (C) and $GFP(CAG)_{101}B$ (D) cells.

N-terminal of HDAC5 may be responsible for gene silencing

To gain further insights into the mechanisms of action of HDAC5, we sought to determine which domains are necessary and/or sufficient for silencing. We, therefore, constructed a series of truncation mutants and transfected them into CIT40B cells. We found that the HDAC5 N-terminus is sufficient for decreasing the expression of the GFP reporter, whereas targeting the catalytic domain alone did not have an effect. Indeed, using mutants in the catalytic domain that dramatically improves acetylation levels or entirely abolish them did not have an effect on GFP expression upon targeting. This is not entirely unexpected given that HDAC5, like other types IIa HDACs, is thought to function by recruiting other factors rather than carrying out the deacetylation itself (Fischle et al., 2002; Kao et al., 2000; Lahm et al., 2007). Further truncation of the N-terminus suggests that the coiled-coil domain of HDAC5 responsible for homo- and heterodimerization (Backs et al., 2008), is essential for the function of HDAC5. These experiments show that our targeting system is highly useful in gaining significant insights into the mechanism of gene silencing.

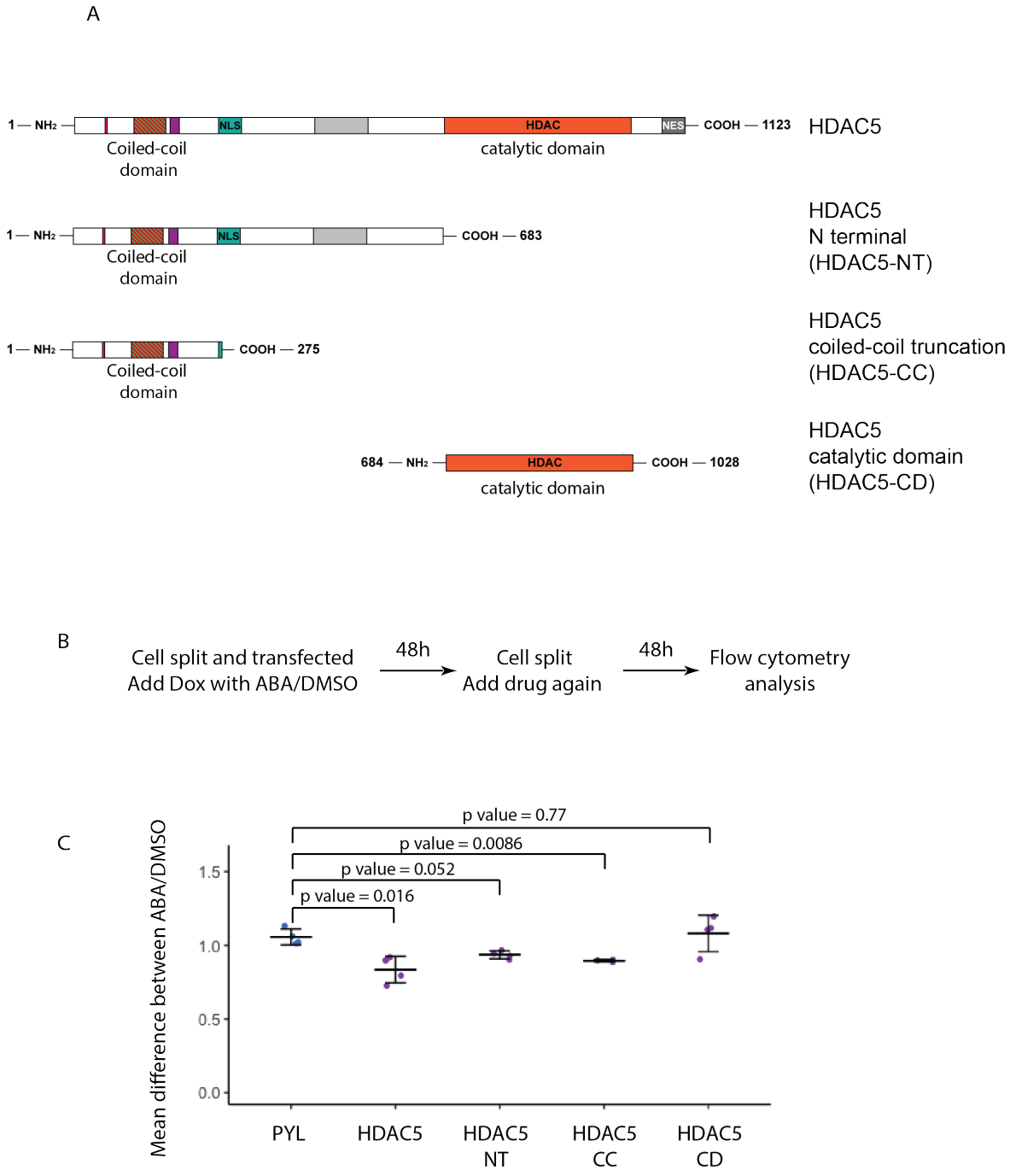


Figure III. 15. Targeting of the N-terminal of HDAC5 is sufficient to shift the expression of GFP. (Adapted from Alicia Borgeaud Master Thesis)

A. HDAC5 mutant constructs. The HDAC5 catalytic domain is shown in orange. All the truncation forms have amino acid sizes indicated on the left and names on the right.
B. Experimental design.
C. Data quantification for GFP intensity differences in ABA or DMSO in CIT40B cells transfected with PYL-HDAC5 truncations. N=4. Error bars indicate standard deviation.

4. HDAC3 targeting promotes local gene expression

HDAC5 is thought to recruit Class I HDACs, including HDAC3. HDAC3 is part of the nuclear co-repressor complexes Nco-R and SMRT and acts as the deacetylase (You et al., 2013). Moreover, both HDAC3 and HDAC5 were shown to promote trinucleotide repeat expansion in human astrocyte cells by Lahue's laboratory (Debacker et al., 2012; Gannon et al., 2012). Thus, I expected that targeting HDAC3 would have a similar effect to HDAC5 targeting.

Transient transfection of PYL-HDAC3 in CITnB cells (Figure III. 16A), unlike PYL-HDAC5, changed GFP expression in neither CIT0B nor CIT40B cells (Figure III. 16BC).

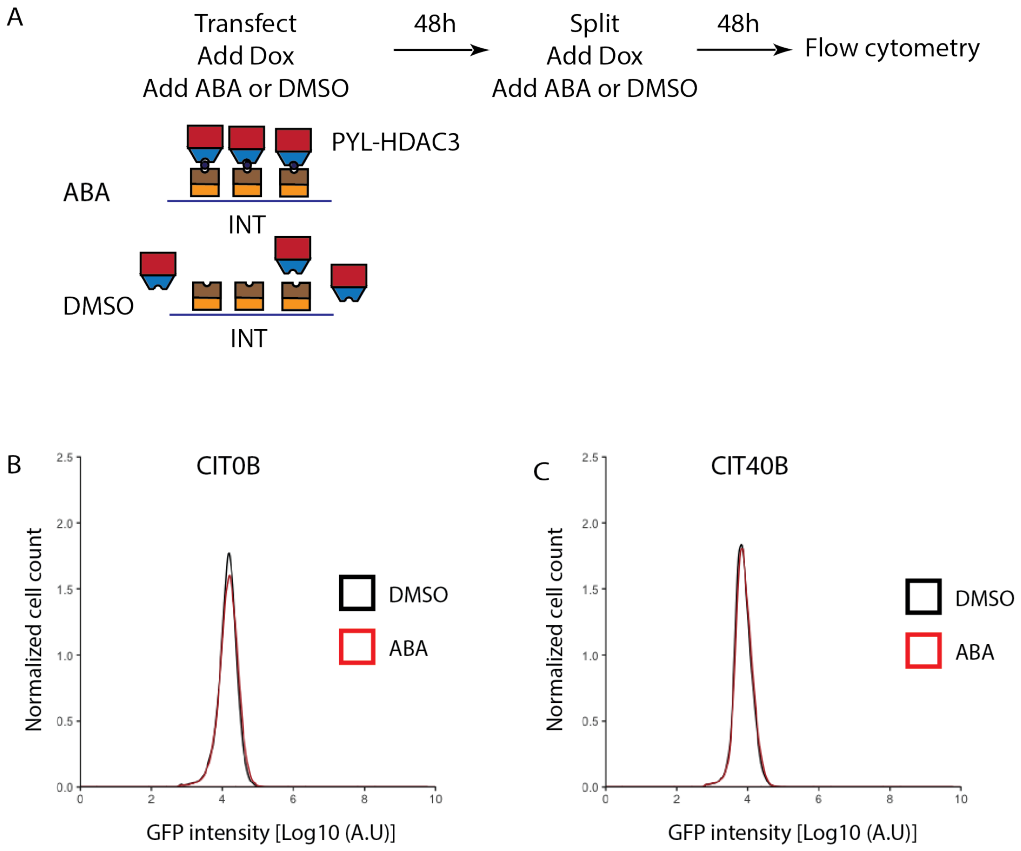
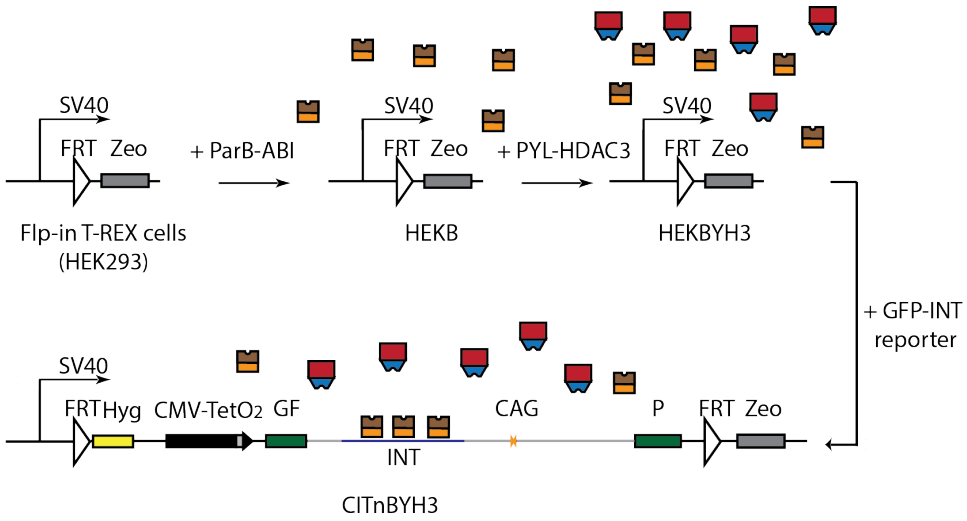


Figure III. 16. The PYL-HDAC3 transient expression did not change GFP intensity upon targeting.

A. Experimental design.
B. Typical flow cytometry profile for targeting of PYL-HDAC3 in CIT0B cells.
C. Typical flow cytometry profile for targeting of PYL-HDAC3 in CIT40B cells.

Since the effect we saw with HDAC5 was dramatically smaller than that seen in stable cell lines, we build CITnBYH3 cells that were isogenic except for the size of the repeat tract in the GFP-INT reporter (Figure III. 17AB, Table 1).

A



B

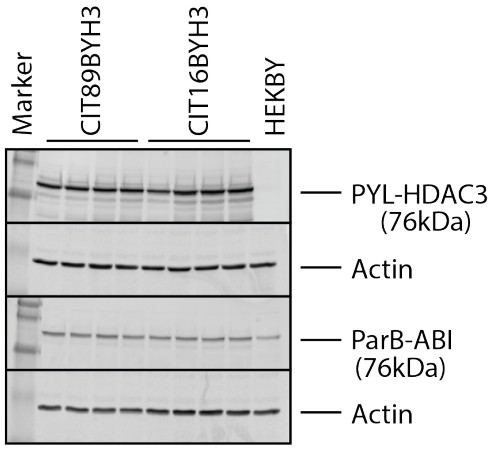


Figure III. 17. CITnBYH3 stable cell line making.

A. Cell line construction. ParB-ABI, PYL-HDAC3 and GFP reporter constructs are stably transfected into Flp-in T-REX cells in a sequential manner.

B. Transgenes expression in CITnBYH3 cells.

I found that PYL-HDAC3 was robustly recruited to INT with a four-fold increase when ABA was added to the media for five days compared to DMSO alone (Figure III. 18A).

Surprisingly, PYL-HDAC3 targeting in CIT16BYH3 cells showed a slight but significant shift towards more GFP expression as measured after five days of targeting using flow cytometry (Figure III. 18 CD). This is surprising, but HDAC3 was shown to bind preferentially to promoters of highly expressed genes, and targeting it with a catalytically null Cas9 had locus-specific effects that are currently unexplained (Kwon et al., 2017; Wang et al., 2009).

To test whether PYL-HDAC3 targeting affected histone H3 acetylation, I ChIPed acetylated H3 in CIT16BYH3 cells. After five day-treatment with either ABA or DMSO, I found that the INT sequence showed significant 2.5 fold enrichment of acetylated H3 upon targeting (Figure III. 18E). H3 ChIP excluded the possibility that acH3 increased because of an increase in H3 (Figure III. 18E). The ACTA1 locus remained unaltered by these treatments, showing that the targeting was specific. The increase in acH3 upon targeting explains well the concomitant increase in GFP expression. Why HDAC3 targeting would increase acH3 is not clear, and it will be further speculated about in the Discussion section.

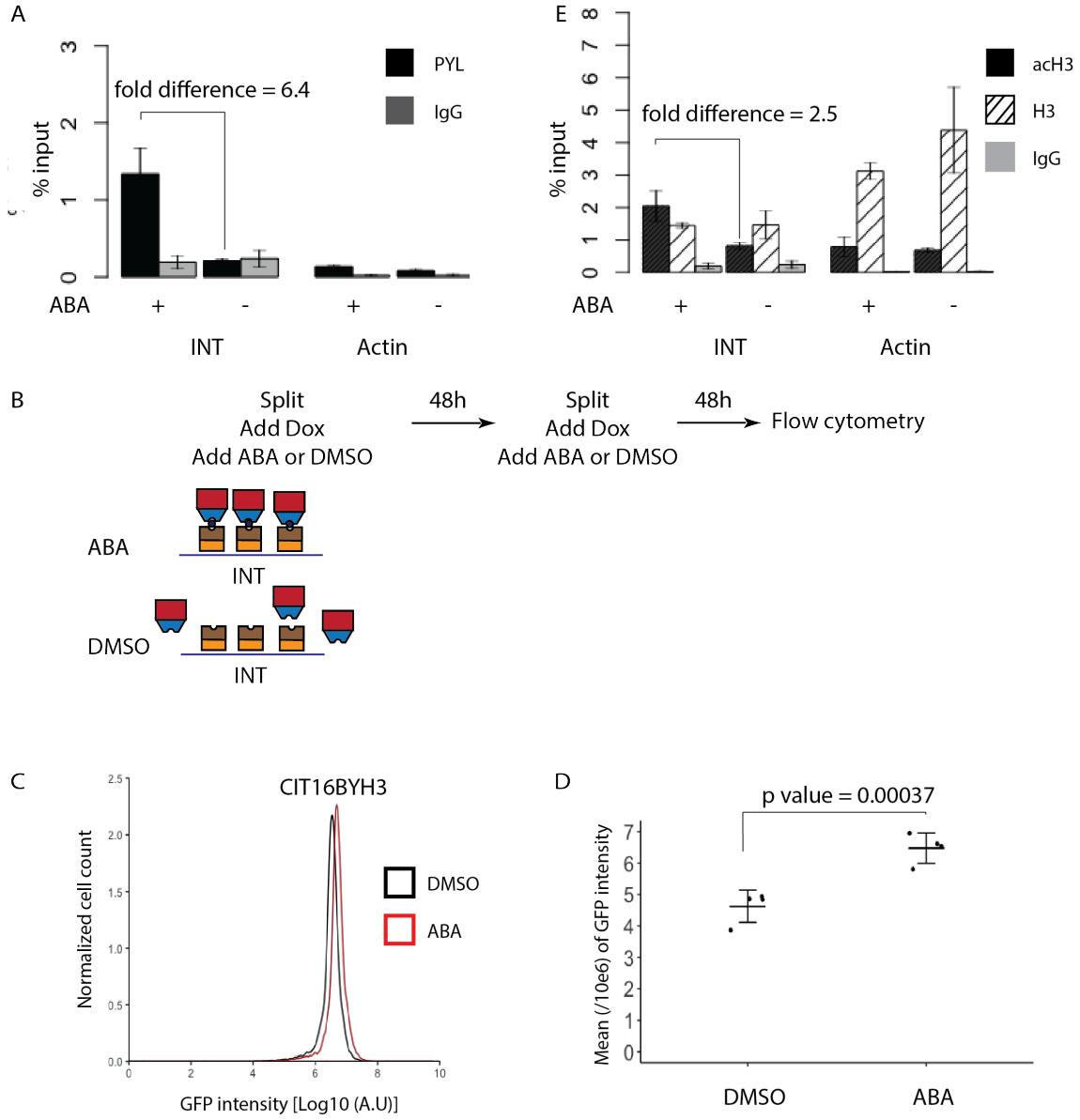


Figure III. 18. PYL-HDAC3 targeting increases GFP expression.

A. ChIP-qPCR experiment at INT and Actin locus in CIT16BYH3 cell line upon ABA addition or not using the antibody against PYL-HDAC3 construct. N=4. Error bars stand for standard error.
B. Experimental design.
C. Typical flow cytometry profile for targeting in CIT16BYH3 cells.
D. Quantification of the data shown in C. N=4 for each condition. Error bars stand for standard deviation.
E. ChIP-qPCR experiment at INT and Actin locus in CIT16BYH3 cell line upon ABA addition or not using the antibody against acH3 and H3. Error bars stand for standard error. N=4

To determine whether HDAC3, like HDAC5, had a repeat-length dependent effect on gene expression, I performed the ChIP and flow cytometry experiments in isogenic CIT89BYH3 cells. The cells were cultured for five days in the presence of dox as well as either ABA or DMSO (Figure III. 19A). I found that PYL-HDAC3 targeting in this context also increased GFP expression significantly (Figure III. 19BC – $P<0.01$). To determine whether HDAC3 acts differently when the repeat tract is in the normal or the expanded range, I performed a statistical analysis on the expression results of CIT16BYH3 and CIT89BYH3 cells. I found that PYL-HDAC3 targeting increased slightly from a 1.2 fold effect in CIT16BYH3 cells to a 1.5 fold in CIT89BYH3 cells, but this increase was not statistically different. (p-value = 0.07) (Figure III. 19D). Thus, unlike HDAC5, which had a repeat-length-dependent effect on gene expression, HDAC3 did not.

HDAC3 targeting and functionality were also tested by ChIP-qPCR of the fusion protein to INT and by the levels of acH3 and H3 ChIP-qPCR in CIT89BYH3. The results were similar to those with the shorter repeat tract: PYL pull down was enriched 3.2 fold when cultured in the presence of ABA compared to the untargeted DMSO (Figure III. 19E). Concomitantly, the levels of acetylation were increased by 2.1 fold, while the H3 levels did not change (Figure III. 18F). It may be useful to test whether the catalytic activity of HDAC3 is required using specific HDAC3 inhibitors.

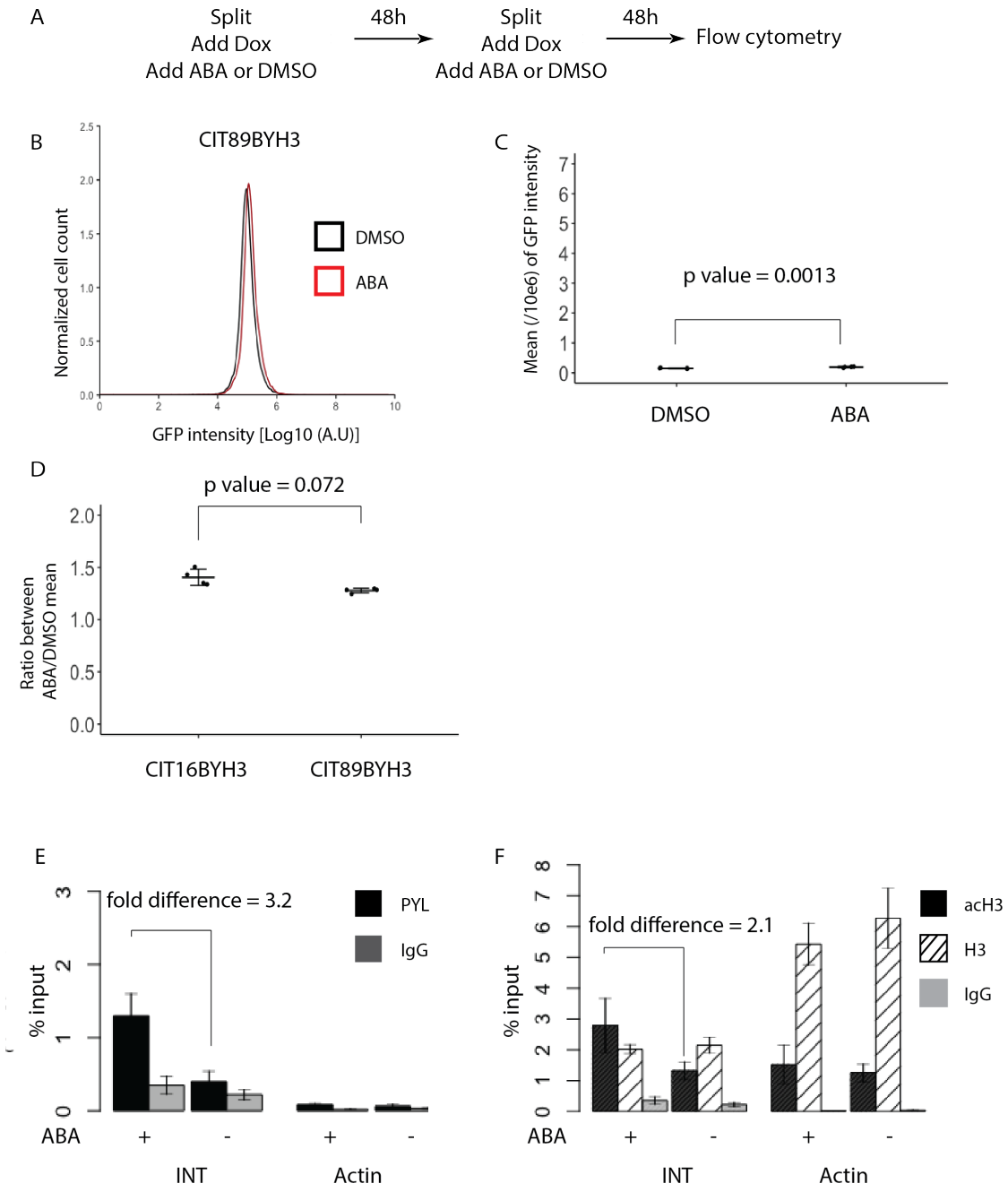


Figure III. 19. PYL-HDAC3 targeting increases GFP expression independently of repeat size.

A. Experimental design.
B. Typical flow cytometry profile for targeting in CIT89BYH5 cells.
C. Quantification of the data shown in C. N=4 for each condition.
D. Comparison of the effect of targeting on GFP expression in CIT16BYH3 and CIT89BYH3 cells.
E. ChIP-qPCR of PYL-HDAC3 in CIT89BYH3. N=4
F. ChIP-qPCR of acH3 and H3 in CIT89BYH3. N=4

As it was the case with PYL-HDAC5 targeting, I found that transient expression of PYL-HDAC3 in GFP$(CAG)_0$B and GFP$(CAG)_{101}$B cells (Figure III. 20A) did not affect gene expression (Figure III. 20BC). I conclude that local targeting of PYL-HDAC3 increases acetylated H3 and GFP expression regardless of repeat size.

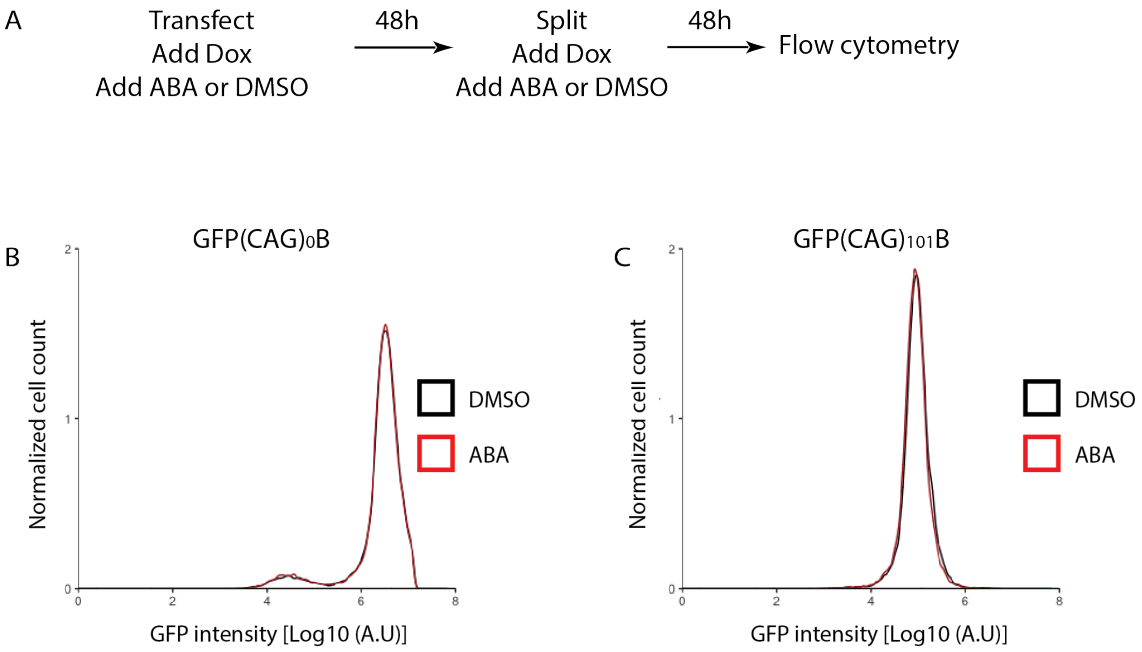


Figure III. 20. The effect of PYL-HDAC5 targeting on GFP expression depends on the presence of the INT sequence.

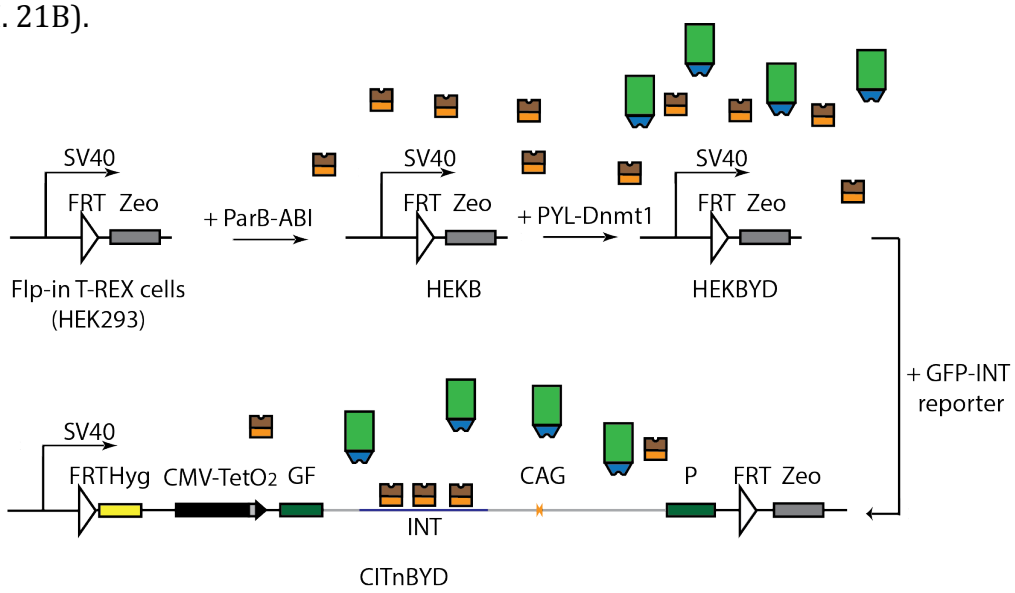
A. Experimental design.

B-C. Representative flow cytometry profiles for GFP$(CAG)_0$B (B) and GFP$(CAG)_{101}$B (C)cells expressing PYL-HDAC3.

5. DNA methyltransferase 1 targeting increases GFP expression in a repeat-size-dependent manner

DNA methyltransferase 1 (DNMT1) has been shown to stabilize an expanded repeat in SCA1 mice and human cells (Dion et al., 2008). The mechanism of action remains unclear. We, therefore, tested whether DNMT1 had a local effect on DNA methylation to affect instability as suggested by Dion et al., or through an indirect effect. I, therefore, built CITnBYD cells that express stably PYL-DNMT1 and contain the GFP-INT reporter with either 16 or 89 CAGs (Figure III. 21A and Table 1). The two isogenic cell lines had, as expected, similar levels of ParB-ABI and PYL-DNMT1 (Figure III. 21B).

A



B

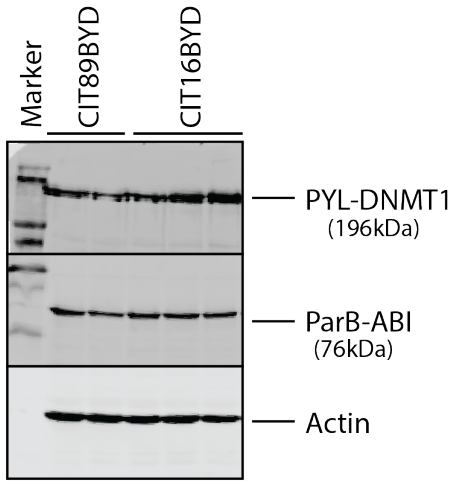


Figure III. 21. CITnBYD stable cell line making.

A. Cell line construction. ParB-ABI, PYL-Dnmt1 and GFP reporter constructs are stably transfected into Flp-in T-REX cells in a sequential manner.
B. Transgene expression in CITnBYD cells.

ChIP-qPCR experiments confirmed PYL-DNMT1 recruitment when adding ABA. The recruitment was significant with a 19 fold increase in the presence of ABA compared to when DMSO was used in CIT16BYD cells and 26 fold in CIT89BYD cells (Figure III. 22A, Figure III. 23A). Recruitment to actin was low and remained unchanged, suggesting that the targeting is both specific and efficient.

I then determined whether PYL-DNMT1 targeting in CIT16BYD and CIT89BYD influenced GFP expression (Figure III. 22CD and 23BC). In both cases, there was a significant decrease in GFP intensity of 2 fold and 1.6 fold ($P = 0.003$ and $P = 0.0005$, respectively). By contrast, transient transfection and targeting of PYL-DNMT1 in $GFP(CAG)_0B$ and $GFP(CAG)_{101}B$ cells did no show changes in GFP expression, suggesting that the silencing observed is due to direct binding of DNMT1 to the INT sequence (Figure III. 24BC).

Interestingly, like in the case of PYL-HDAC5, PYL-DNMT1 targeting led to a significant allele-length dependent silencing of the reporter ($P= 0.04$). We will be able to give a conclusion about DNMT1 influence DNA methylation and/or chromatin modification after bisulfate sequencing, and acH3 ChIP finished. However, the simplest explanation, in this case, is that PYL-DNMT1 targeting was more efficient in the context of a longer repeat tract.

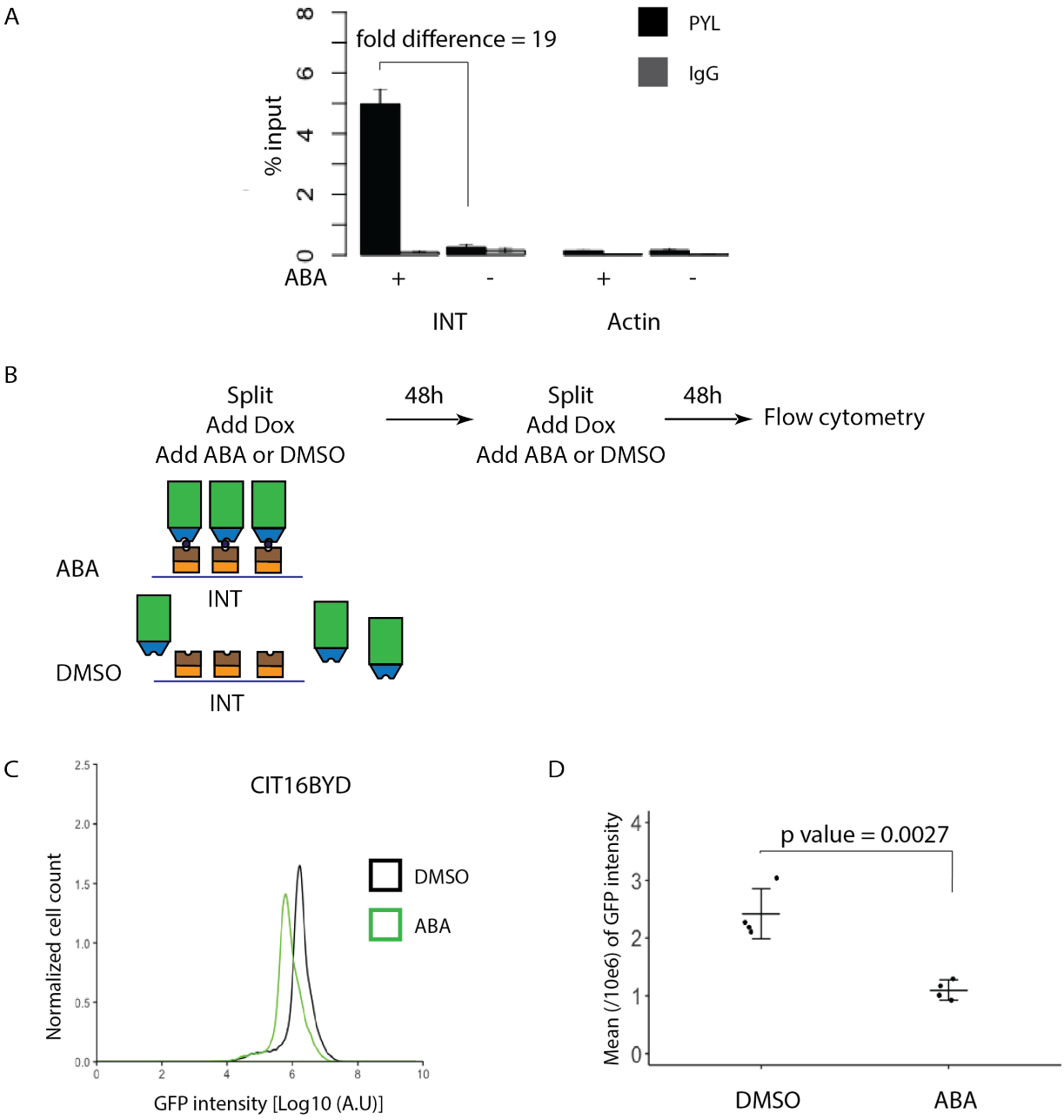


Figure III. 22. PYL-DNMT1 targeting silences GFP expression.

A. ChIP-qPCR experiment at INT and Actin locus in CIT16BYD cell line upon ABA addition or not using the antibody against PYL-Dnmt1 construct. N=4. Error bars stand for standard error.
B. Experimental design.
C. Typical flow cytometry profile for targeting in CIT16BYD cells.
D. Quantification of the data shown in C. N=4 for each condition. Error bars stand for standard deviation.

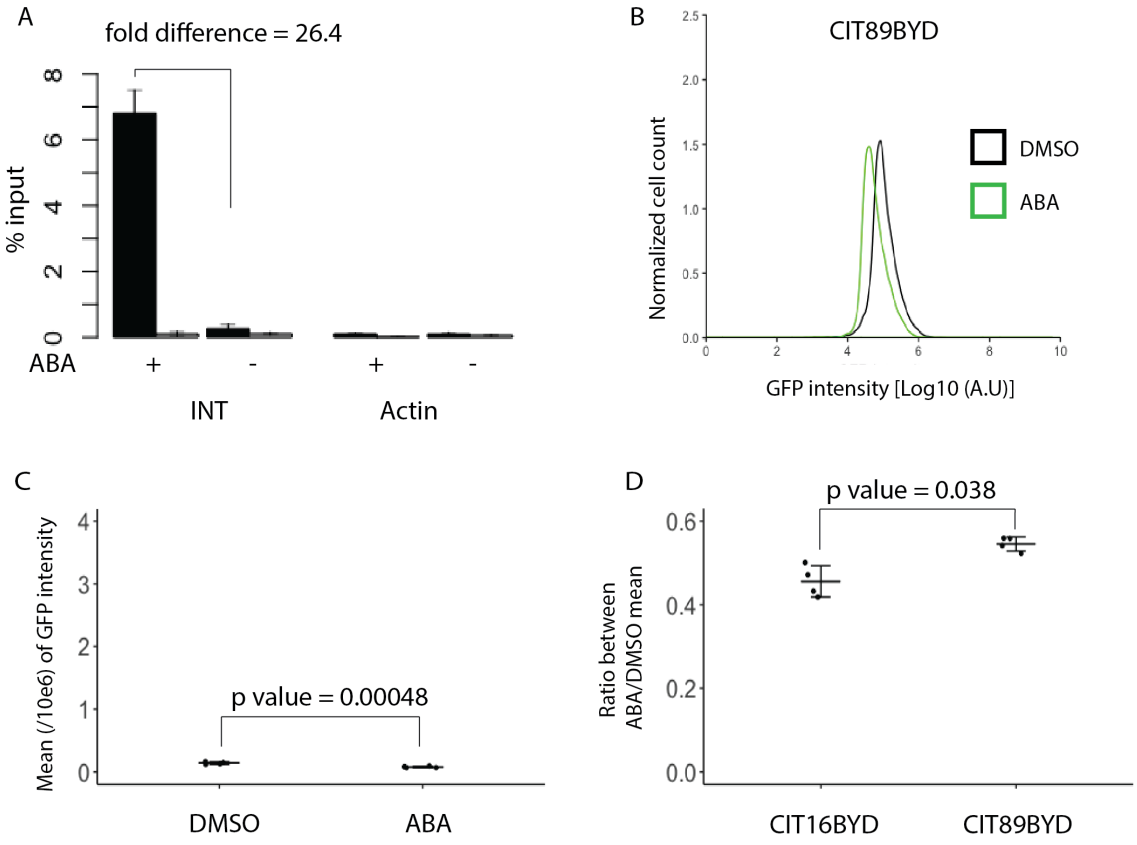


Figure III. 23. The effect of PYL-DNMT1 targeting on GFP expression is repeat-length dependent.

A. ChIP-qPCR of PYL-DNMT1 in CIT89BYD. N=4
B. Typical flow cytometry profile for targeting in CIT89BYD cells.
C. Quantification of the data shown in B. N=4 for each condition. Error bars stand for standard deviation.
D. Comparison of the effect of targeting on GFP expression in CIT16BYD and CIT89BYD cells. Error bars stand for standard deviation.

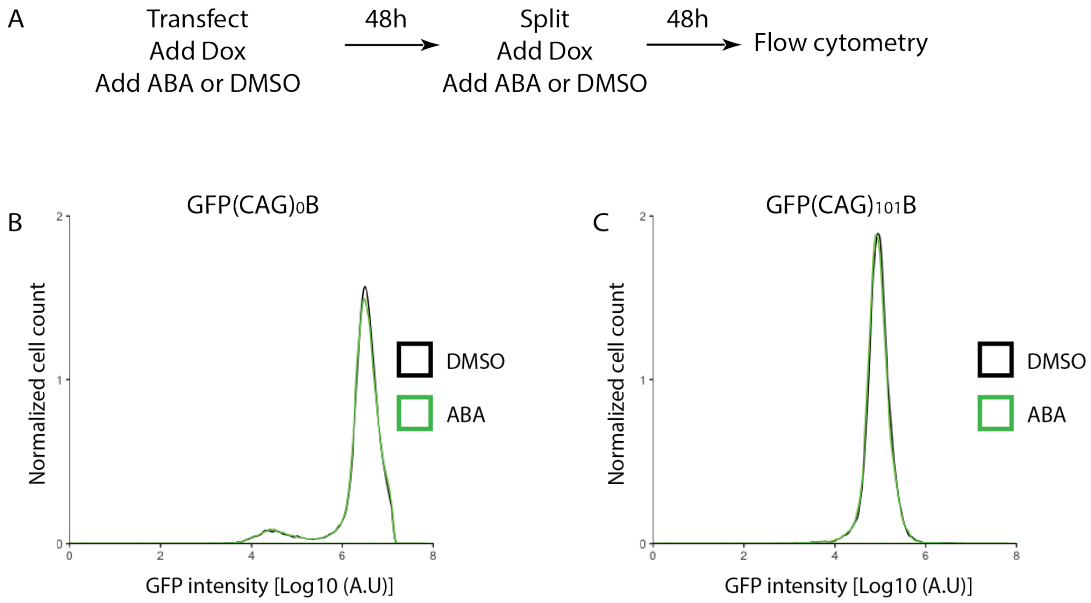


Figure III. 24. The effect of PYL-DNMT1 targeting depends on the presence of the INT sequence.

A. Experimental design.

B-C. Typical flow cytometry profile for targeting in PYL-DNMT1 in $GFP(CAG)_0B$ (B) and $GFP(CAG)_{101}B$ (C)cells.

6. GFP expression changes are not due to CAG repeat instability

The GFP assay was designed to look for repeat instability (Cinesi et al., 2016; Santillan et al., 2014). Moreover, we constructed CIT to test the hypothesis that local recruitment of chromatin modifiers enzymes is sufficient to drive CAG repeat instability. However, given that the reporter requires transcription and that there is an alternative CAG exon, our system can, in principle, monitor three processes simultaneously: instability, transcription, and splicing. I, therefore, tested the hypothesis that targeting of HDAC5, HDAC3, or DNMT1 may be due to CAG repeat expansion or contraction. To do so, Oscar Rodriguez Lima and I collaborated to perform small pool PCR (SP-PCR) to detect CAG repeats length change (Method). SP-PCR is a gold-standard method to detect CAG repeats variability. Briefly, genomic DNA is diluted down to only a few genomes per PCR, and multiple reactions are set up for a single sample. This minimizes the bias towards amplifying shorter repeats. Because the yield is weak because of a limited number of templates, Southern blotting is necessary. To maximize the chances of detecting changes in CAG repeat instability, I prepared the genomic DNA samples after four weeks of continuous culturing in the presence of ABA or DMSO and compared the flow cytometry results (Figure III. 25A).

We first tested HDAC5 and DNMT1 given that they had the most exciting effects on GFP expression. We saw no significant changes in CAG repeat instability between the ABA and DMSO treated samples (Figure III. 25BC). This result indicates the silencing events when targeting HDAC5 or DNMT1 in the CIT system may not due to CAG instability. It remains possible that the frequencies of instability are too low for SP-PCR to detect and that waiting longer would reveal then.

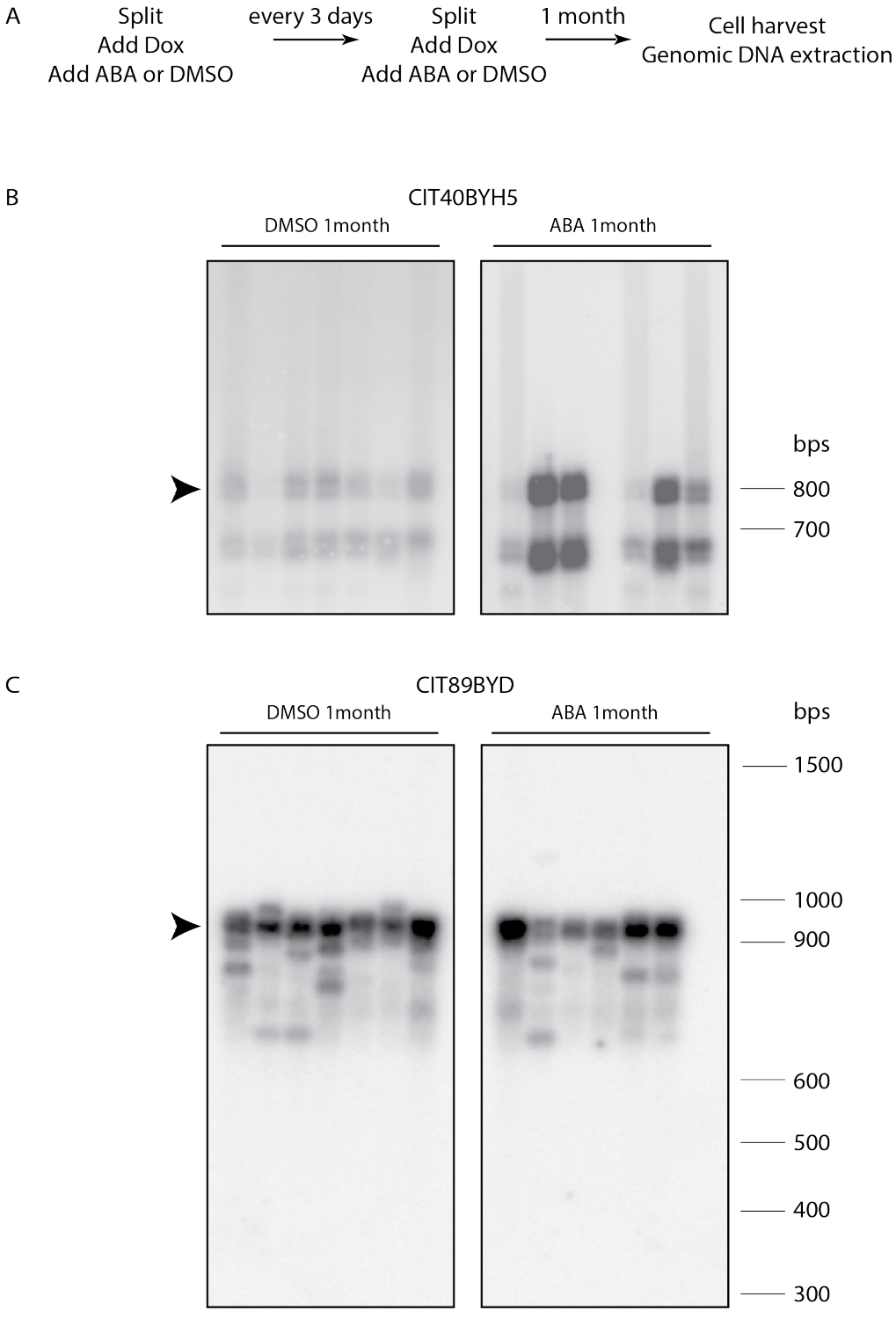


Figure III. 25. GFP intensity change may not due to CAG repeat instability in one-month treatment.

A. Experimental design. Compare to the typical flow cytometry experiment timeline, samples for small pool PCR analysis are treated for 1 month.
B. Southern blot for CIT40BYH5. Arrowhead indicates the original repeat number fragment.
C. Southern blot for CIT89BYD. Arrowhead indicates the original repeat number fragment.

7. An alternative targeting strategy to test HDAC5 direct effect on gene transcription and/or silencing

Epigenome editing is increasingly done via dCas9 fusions. We, therefore, tested whether we could reproduce the results found here with dCas9 fusion. To build a dCas9-dependent targeting of HDAC5, I used CITn cells (Figure III. 26A). The experimental design was to have dCas9 fused to a protein of interest, guided by single guide RNA to the GFP locus to achieve targeting. There were significant challenges to overcome to achieve this. First, the molecular weight of dCas9 is limiting, and a dCAS9-HDAC5 fusion protein would be 280kDa, which is not practical to work with. Besides, since ParB oligomerizes, there is a large number of molecules of ParB that get recruited, which is not necessarily the case for dCas9.

We first designed seven different sgRNAs targeted to the GFP locus as showed in Figure III. 26B with one control sgRNA being used. To test how well they worked in cells, I transfected wild-type Cas9 with seven sgRNAs separately into the CIT40 cells for three days. Cells were then harvested, and genomic DNA was extracted for PCR amplification. If a sgRNA successfully targeted Cas9 to its target site, cleavage will produce indels. The PCR product of the region after transfection is then mixed in equal molar ratios with the same PCR but from the untransfected cells. After denaturation and slow annealing, fragments will form mismatches due to on-targeting cleavage and indel formation. The T7 endonuclease I is then used to digest mismatched fragment, given multiple bands on an agarose gel. Figure III. 26C shows that at least sgVIN-006 had a second band after T7endoI treatment. I also included sgRNAs targeting the repeat tract itself, sgRNA sgVIN-050 and sgVIN-052, because they have been shown to bring the Cas9 nickase to the repeat tract (Cinesi et al., 2016). I also included these two sgRNAs even though they did not show an evident T7endoI endonuclease cleavage (Figure III. 26C).

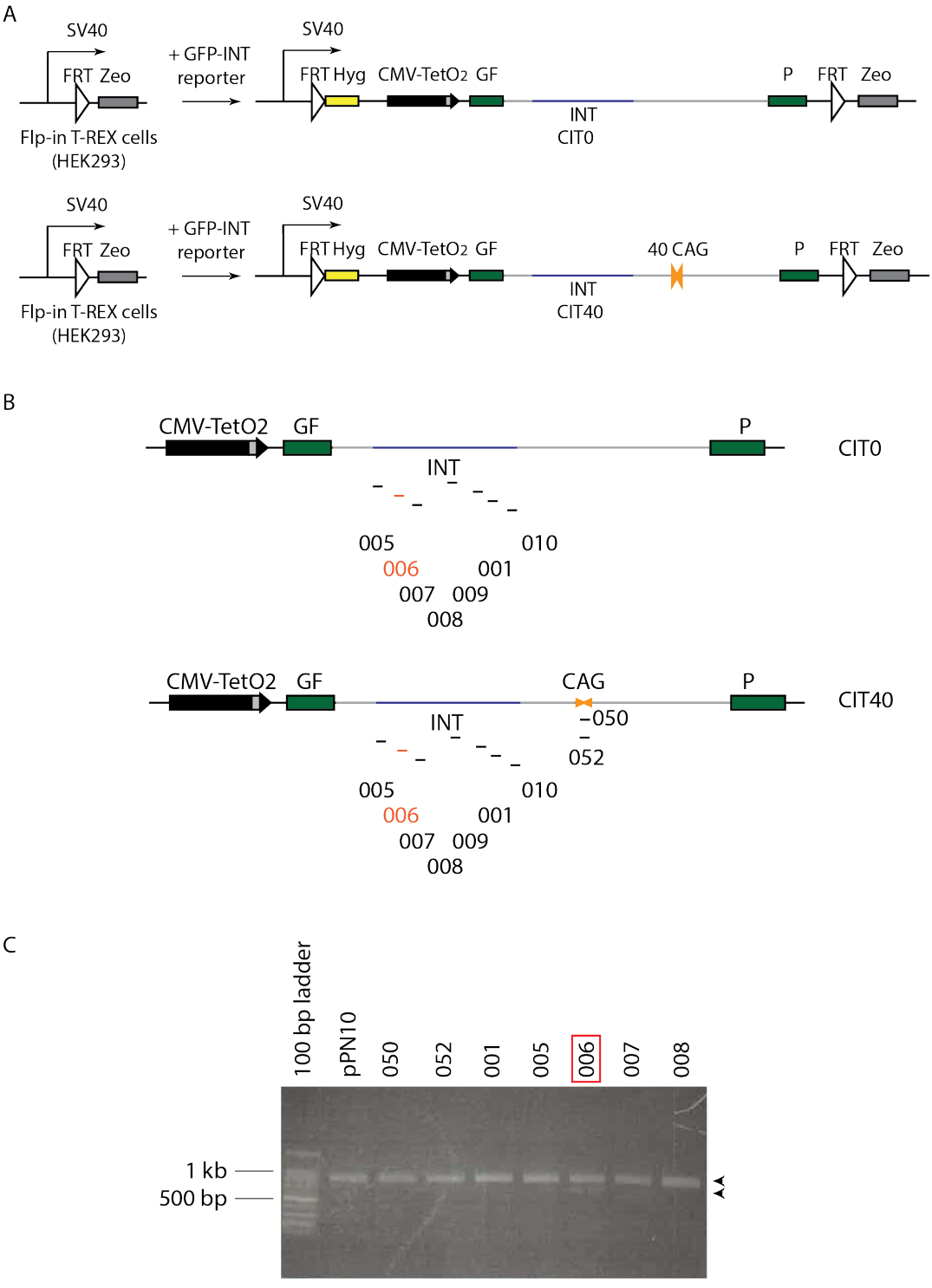


Figure III. 26. dCas9 targeting using sgRNAs in CIT0 and CIT40 cells.

A. Cell lines construction. GFP reporter is stably integrated into Flp-in T-REX cells.
B. Short lines indicate sgRNA localization within the GFP transgene along with their identifying numbers.
C. T7endI test for editing efficiency.

I used the dCas9-KRAB fusion construct first to see whether targeting of a silencing factor to the INT sequence, which is several kb away from the promoter, could silence it. The control is an empty vector construct called pPN10 that does not encode a targeting sequence. sgVIN-006 was used to target the INT sequence. I found that dCas9-KRAB targeting to INT had no significant effect on the expression of the GFP compared to dCas9 expression without a sgRNA (Figure III. 27B). The same was true in CIT40 cells (Figure III. 27C). As a further control, I targeted a blue fluorescent protein (BFP) fused to dCas9. As expected, no GFP intensity change was seen for this construct (Figure III. 27DE). dCas9 to bring transcription silencing factor KRAB to INT sequence could induce minor GFP decrease in CIT cells with 0 CAG and 40 CAGs.

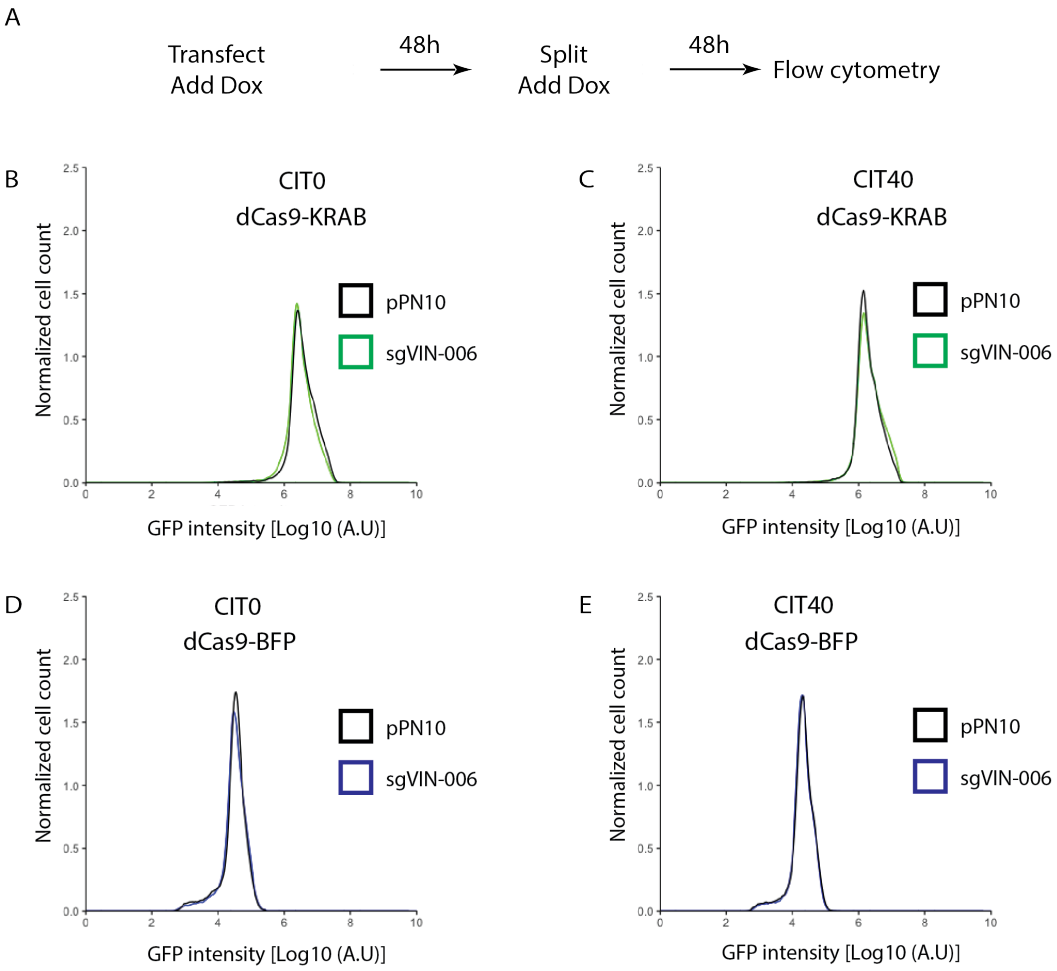


Figure III. 27. dCas9-KRAB and dCas9-BFP targeting in CIT0 and CIT40 cells.

A. Experimental design.
B. Typical flow cytometry profile for targeting in dCas9-KRAB in CIT0 cells.
C. Typical flow cytometry profile for targeting in dCas9-KRAB in CIT40 cells.
D. Typical flow cytometry profile for targeting in dCas9-BFP in CIT0 cells.
E. Typical flow cytometry profile for targeting in dCas9-BFP in CIT40 cells.

Considering the size of the dCas9-HDAC5 fusion protein, we opted to use only the N-terminal based on our results with transient transfection (Figure III. 15).

The dCas9-HDAC5 truncation constructs with sgRNAs were transiently transfected into CIT0 and CIT40 cells. I found that the dCas9-HDAC5 N terminal domain (dCas9-HDAC5NT) fusion protein did not affect GFP intensity shift whether the sgRNA against INT was included or not (Figure III. 28B). Furthermore, the dCas9 fused to an HDAC5 catalytic domain (dCas9-HDAC5 CD) seemed to narrow the GFP intensity peak in sgVIN-006-transfected compare to pPN10-transfected cells. (Figure III. 28C). The same experiment with dCas9-HDAC5NT in CIT40 cells seemed to increase GFP expression slightly when the sgRNA against the INT sequence was included, whereas dCas9-HDAC5 CD did not have any effect (Figure III. 28DE). It is possible that this lack of significant effects on GFP expression of the dCas9-construct reflects the requirement to recruit silencers closer to the promoter (Gilbert et al., 2014) or that there is not enough HDAC5 recruited to the INT sequence when dCas9 is used compared to ParB. Indeed, many silencing proteins appear as foci in the nucleus that correspond to several hundred molecules and may work via liquid phase separation (Larson et al., 2017; Strom et al., 2017). This interpretation has significant implications when it comes to designing epigenome editing strategies with dCas9.

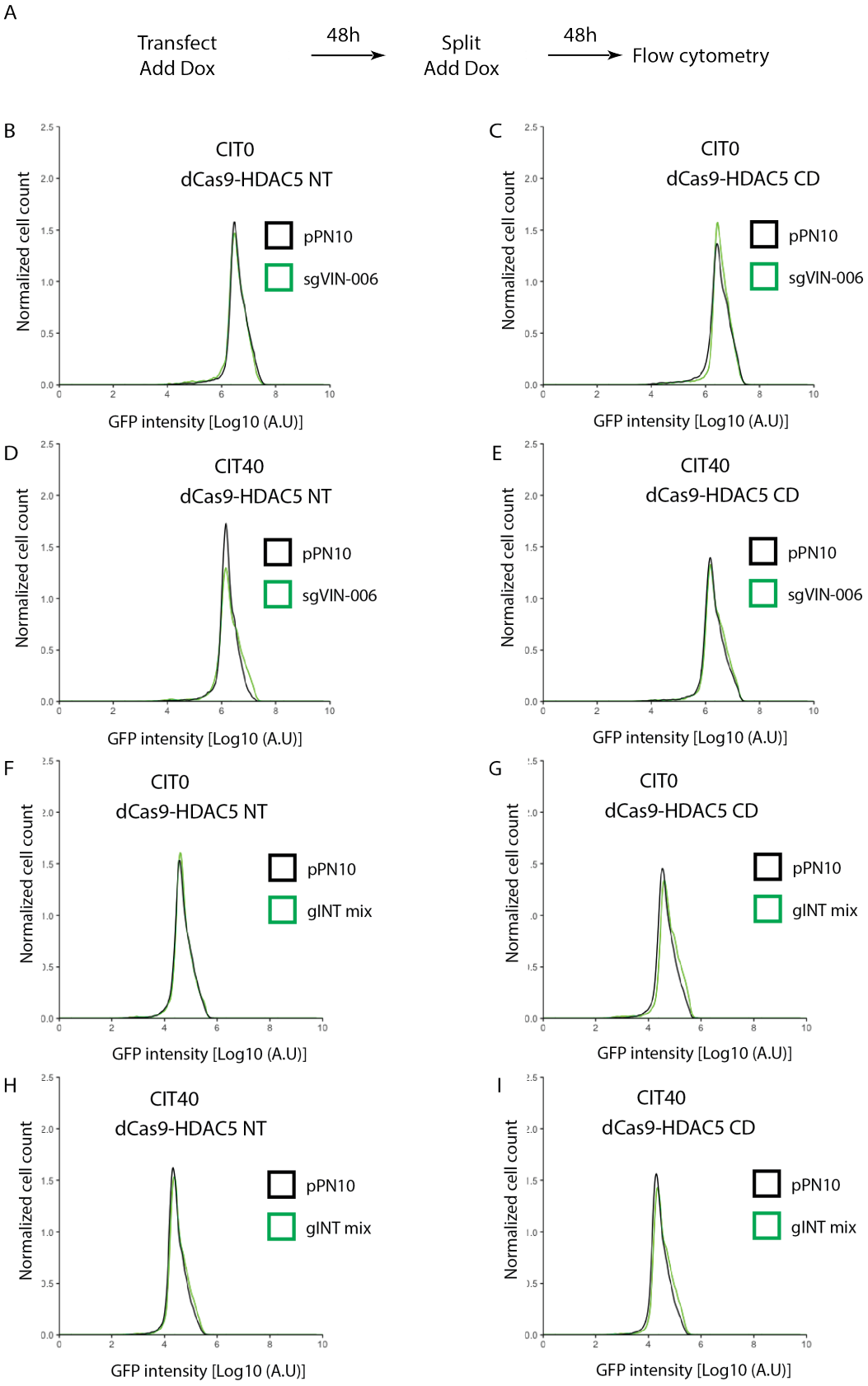


Figure III. 28. dCas9-HDAC5 truncation targeting in CIT0 and CIT40 cells.

A. Experimental design.

B–I. Typical flow cytometry profiles for targeting of dCas9-HDAC5 truncations.

Chapter IV

Discussion

Here we present the chromatin inducible targeting (CIT) system and its application to understanding how chromatin modifying enzymes affect gene regulation. CIT uses cell lines with a stable transgene integration that is carefully designed to make transgenes comparable to each other. We designed it to make up for what loss-of-function approaches could not tackle: whether the effect of any given chromatin modifying enzyme is direct or is due to changes in gene expression *in trans.* We used it to investigate the effect of HDAC5, HDAC3 and DNMT1 targeting on either silencing or promoting gene expression. CIT can be integrated into any genome locus, not restricted to usage we showed here. Furthermore, it may have an extensive usage for screening.

1. The advantages of CIT

1.1 Protein-DNA targeting system - ParB-INT

Targeting systems have been used widely for tethering specific protein to specific locus and induces diverse cellular functions (Waryah et al., 2018). What we have here is a targeting strategy with a bacteria derived ParB-INT system fused to plant hormone ABA-induced proximity system. The feature of the ParB-INT system is its efficient targeting rate (up to 44 fold), insensitive to the DNA contexts tested here and can attract multiple molecules to the target DNA sequence.

A characteristic of significant importance for targeting systems is the tethering efficiency. Indeed, one concern is that large fusion proteins may not be recruited efficiently. Here, however, we found that four different POIs ranging from 26 kDa to 196 kDa, achieved robust targeting efficiency upon ABA addition, up to 44-fold increase compared to DMSO alone, in PYL as well as PYL-HDAC5, PYL-HDAC3 or PYL-DNMT1, based on the validation from ChIP-qPCR. We conclude that the targeting efficiency within the system is robust regardless of the size of the proteins targeted, at least within the range that we tested.

Chromatin context could also affect targeting efficiently. This is the case within other existing targeting systems. TALEs can be impaired by condensed chromatin

(Bultmann et al., 2012; Chen et al., 2013) and dCas9 binds poorly to hypermethylated CpG islands (Cano-Rodriguez et al., 2016; Qi et al., 2013). Targeting efficiency for ParB-INT, by contrast, was unaffected by the presence of an expanded CAG tract. Expanded TNRs correlate with altered chromatin structure and make it more heterochromatic (Dion and Wilson, 2009; Nageshwaran and Festenstein, 2015), suggesting that targeting of ParB-INT is independent of changes brought by an expanded CAG tract. In CIT59BYH5 DMSO condition, acetylated histone 3 levels dropped to below 50% compared to CIT16BYH5 in DMSO. With active transcription marks significantly decreased, HDAC5 still shows 30-fold PYL pull down enrichment upon ABA addition in 59 CAG samples and 40-fold when CIT16BYH5 cells were used. Until now, we observed no impairment of targeting by expanded trinucleotide repeats at the GFP reporter.

Last but not least, is the feature of the ParB-INT targeting that multiple ParB molecules are recruited to the INT sequence. This may explain why CIT and dCas9 system give different results with the same POI domain treatment. Clearly, in contrast to CIT, dCas9-HDAC5 N terminal in our system did not show direct silencing effect. We hypothesize that this might be due to local protein concentration. Since ParB has been reported to accumulate 100-200 proteins targeting to INT *in vitro* (Saad et al., 2014), this may provide a dramatic difference with having one dCas9 bringing one HDAC5 molecule to the same locus. Of course, there are other possibilities like the steric hindrance or that HDAC5NT fusion to dCas9 prevents its efficient targeting. To test this, I may perform a dCas9-HDAC5NT ChIP-qPCR to test the targeting efficiency. Unlike the dCas9 targeting approach, CIT system multiple molecules binding enlarges POI signal, and it may help to fish out some weaker effect using this system.

1.2 Chemical-induced proximity system - ABA system

It is important to note that different levels of the PYL fusions between cell lines with different repeat length would have prevented us from making the conclusions found herein. Indeed, we took care to design CIT such that PYL fusions are integrated into the genome before adding the GFP reporter with differing repeat sizes. The FLP-In system further allows us to integrate the reporter at a specific single locus. Thus, we can make the comparisons directly, yet not between the lines with different PYL fusions.

We chose the ABI-PYL system among several CIP systems (Liang et al., 2011). This was because it did not show evident toxicity to human and animals proved by our regular vegetable consumption (Stanton et al., 2018). Compared to ABA, the commonly used FKBP-rapamycin system contains endogenous proteins receptors (Kang et al., 2008). Overexpression of FKBP may disturb *in vivo* signal transduction and make it less ideal for our purpose. On the other hand, the trivial addition and removal of ABA to induce or reverse proximity provide a convenient way to control the process. The light-induced dimerization systems also show some advantages in spatiotemporal protein-protein interaction studies (Kennedy et al., 2010; Levskaya et al., 2009; Yazawa et al., 2009). Compared to expose cell culture under blue light, which requires LEDs and black box, ABA system would be more efficient to be applied in our application and more efficiently to be implemented by merely adding ABA to cell medium. In conclusion, including ABA-inducible proximity system in the CIT system gives well-controlled POI local effect and is trivial to use in daily research.

1.3 Convenient reporter assay to measure transcription and splicing alternation - GFP reporter

A widely used detection for transcription and/or splicing change is quantitative PCR. Instead of using qPCR as the readout on a daily basis, I took advantage of our lab common tool, a GFP minigene reporter, as a fast and sensitive readout of gene expression. It can detect repeat instability (Cinesi et al., 2016; Santillan et al., 2014), changes in alternative splicing and transcription, making it ideal for a wide array of purposes.

Five days of treatment with Cas9 nickase and chromatin modifying enzymes leads to significant GFP intensity shift shown in previous publications from our lab and others (Cinesi et al., 2016). This indicates that a short time frame treatment with GFP reporter is enough to reveal GFP intensity difference after local genome or epigenome editing. It can also detect GFP expression up-regulation and down-regulation result at the same time. This would be very informative for researches on gene regulation and/or repeat instability.

2. Testing chromatin modifying enzymes effect using CIT

The first group of proteins we tested with CIT were HDAC5, HDAC3 and DNMT1. HDAC5 and HDAC3 affect local chromatin modification and further impact on gene regulation and trinucleotide repeat instability(Debacker et al., 2012; Gannon et al., 2012; Suelves et al., 2017). DNMT1 is responsible for maintaining DNA methylation and shows its role also in trinucleotide repeat instability (Dion et al., 2008; Robert et al., 2003).

Making of some additional control cell lines is still ongoing due to unseen technical delay. It would be better to include PYL control cell lines (CIT16BY and CIT89BY) with the same CAG repeat length as being used in HDACs and DNMT1 work. Hopefully, this part of the experiment will be finished before my defense. Another ongoing work is HDAC5, HDAC3 and DNMT1 bisulfite sequencing for local gene DNA methylation test. We would be able to tell the change of DNA methylation upon POI targeting soon.

2.1 The role of HDAC5 N-terminus in local silencing

Targeting PYL-HDAC5 has a silencing effect in CIT. As a class IIa HDAC, HDAC5 possesses a poor deacetylase activity (Fischle et al., 2002; Lahm et al., 2007). While in the CIT system with 16 CAG and 59 CAG GFP reporter, HDAC5 targeting induces clear deacetylation of histone H3 without changing the levels of H3 locally. Whether the gene silencing is due to HDAC5 deacetylase activity or HDAC5 binding to other

endogenous HDACs and/or silencing complex could be an interesting topic to discuss.

We first tested whether the catalytic activity is responsible for HDAC5 targeting dependent silencing. One advantage of CIT is that we can target truncations and mutants to understand the mechanism of action. For instance, we found that HDAC5 mediated silencing was independent of its catalytic activity. However, HDAC5 truncations revealed that the N terminal appears responsible for silencing activity. How the HDAC5 N terminal induces gene silencing is not clear. One possibility is that by recruiting other HDACs that would deacetylate the locus. This could be done via the coiled-coil domain in the N-terminal, which forms heterodimers and homodimers with HDAC4 and HDAC5, respectively (Backs et al., 2008). To test this directly, I would perform immunoprecipitation on HDAC5-coiled coil truncation targeting samples and look for endogenous HDAC4 and HDAC5 presence by western blot.

Current models of silencing complex suggest that HDAC5 recruits HDAC3 as part of the N-CoR/SMRT complex to deacetylate histones and to promote gene silencing (Fischle et al., 2002; Kao et al., 2000; You et al., 2013). In support of this model, the C-terminal of HDAC5 interacts with the SMRT component SMRT repression domain 3 (SRD3) (Kim et al., 2015). Given that HDAC3 and HDAC5 targeting leads to different outcomes in CIT, I propose that HDAC5-dependent gene silencing, at least in some contexts, does not function via SMRT recruitment. To test this directly, pulling down PYL-HDAC5 in CITnBYH5 cells and looking for SRD3 may give the answer. These results would challenge the current model on how HDAC5 promotes gene silencing.

One significant finding is that HDAC5 promotes gene silencing in a repeat-length dependent manner. In this case, we could rule out that the effect was due to different targeting efficiencies. A simpler model, therefore, could be that 59 CAG repeats are sufficient to induce the accumulation of heterochromatic marks, like H3K9me2/3 and the loss of acetylated H3. As a result, HDAC5 would have less

substrate to work with and therefore exhibit a less pronounced silencing effect than that in the context of 16 CAG repeats.

Overall, HDAC5 in CIT uncovered its direct role in local gene silencing. This silencing acts differently depending on sequence contexts. Even though I did not observe any CAG repeat expansion or contraction, my data suggests an SMRT-independent role of HDAC5 in gene silencing that will be interesting to understand further.

2.2 PYL-HDAC3 targeting increases local histone acetylation and gene expression.

We chose to study HDAC3 along with HDAC5, because of their roles to promote CAG repeat expansion in human cells (Debacker et al., 2012). Moreover, HDAC3 was recently shown to influence Huntington's disease pathology which marks its potential for clinical application (Jia et al., 2015, 2016; Suelves et al., 2017). Therefore, we asked whether HDAC3 has a direct role on CAG repeats and/or gene regulation in the CIT system. Stable CITnBYH3 cells showed that PYL-HDAC3 targeting promoted GFP expression regardless of the size of the repeat tract, because HDAC3, as a histone deacetylase, is thought to promote silencing. For the first time, we show that HDAC3 targeting increases acH3 levels locally upon targeting. This is unlikely due to an inactive PYL-HDAC3 fusion. With its overexpression via transient transfection, non-targeting PYL-HDAC3 sample decreased GFP intensity compare to not transfected control. This suggests a genome-wide effect in gene silencing as expected (data not shown) and this is the evidence that PYL-HDAC3 fusion may be functional.

How does HDAC3 targeting increases gene expression? One ChIP-seq study shows that HDAC3 co-localizes with histone acetyltransferases (HATs) at highly transcribed regions instead of at the expected heterochromatin (Wang et al., 2009). HDACs are associated with active gene and are positively correlated with transcription. Whereby, they proposed a model that HDAC3 and HATs work together to influence gene transcription dynamics (Wang et al., 2009). In this system, HDAC3 associates at genes also enriched by several HATs, suggesting that they may keep a

balance between activation and repression at any given locus. In addition, Rpd3 and Hos2 in yeast associate with highly transcribed genes (Kurdistani et al., 2002; Wang et al., 2002). The observation further supports this that dCas9-HDAC3 targeting in MC3T3-e1 pre-osteoblasts induced MecP2 mRNA up-regulation (Kwon et al., 2017). Another overlapping model (Greer et al., 2015) is that HDACs promote gene transcription by inhibiting negative elongation factor (NELF) and by limiting acetylation in gene body and intergenic regions, less acetylation within gene bodies preventing cryptic transcription and facilitating transcription elongation. This model is, however, not supported by ChIP-seq data showing that HDAC3 associates mostly with the promoter region of highly transcribed genes (Wang et al., 2009).

Our data are consistent with the idea that HDAC3 targeting promotes gene expression. However, it would not do so by deacetylating histones in gene body since its targeting increases acetylation. This raises the question of the necessity of the deacetylase activity of HDAC3. The next step is to test whether the HDAC3 specific inhibitor RGFP966 (Malvaez et al., 2013) can minimize the increase in gene expression upon targeting. Our speculative model is that HDAC3 targets non-histone proteins (Glozak et al., 2005; Spange et al., 2009) which would act as inhibitors for gene expression. Those proteins, when being acetylated, tend to dissociate from chromatin, allowing transcription to go through and enabling HATs to access the locus better. Alternatively, deacetylation of those proteins could activate some transcription factor or some remodeling enzyme that promotes transcription elongation and/or initiation. If this model were correct, the deacetylase activity of HDAC3 would be required for overexpression.

IP-Mass spec using HDAC3 as bait has been reported with some unusual potential factors for us to work on (Armour et al., 2017). Besides the known interactors such as NCoR, Armour and colleagues also found proteins essential for development as well as several transcription factors. These candidates could be knockout, and their effect on HDAC3-dependent gene activation will be determined.

One possibility is that HDAC3 brings in HATs that target histones for acetylation. This model predicts that HDAC3 has a structural role and does not

require its deacetylase activity. If this model is correct, we may be able to detect: 1. accumulation of HATs around the GFP locus upon PYL-HDAC3 targeting; 2. the recruitment of a HAT should be indispensable to have the same effect as HDAC3 and should no longer depend on HDAC3; 3. HDAC3-targeting in HAT knockouts would be predicted not to affect gene expression.

2.3 DNMT1 has a direct effect on gene silencing

Similar to HDAC5, PYL-DNMT1 stable integration into the CIT system gives direct evidence that DNMT1 targeting to gene body can induce further silencing of a strongly expressed gene in human cells. Unlike promoter regions with low CpG methylation, gene body usually is highly methylated and correlates positively with the transcriptional output (Jones, 1999, 2012). Evidence comes from human active X chromosome and inactive X chromosome DNA methylation profiles, in which Xa has at least two times more allele-specific methylation than Xi (Hellman and Chess, 2007). The methylation pattern concentrates on gene-bodies in Xa. A shotgun genomic bisulfite sequencing in plants and animals also shows the same hypermethylation in the gene body (Feng et al., 2010). The role of gene-body methylation is highly debated, however, mainly because of the correlational nature of the experiments. Nevertheless, it is reported that DNMT inhibition in HCT116 cells reduces DNA methylation in the gene body, which leads to alleviated overexpression of tumor-suppressor genes and helps reactivation of oncogenes like c-Myc. (Yang et al., 2014). This complicated relationship suggests that gene-body methylation regulates gene expression, but the mechanisms remain unknown.

CIT would be ideal to tackle the mechanism of gene body methylation-dependent gene expression. For instance, if bisulfite sequencing reveals that DNMT1 targeting leads to an increase in DNA methylation, it would be interesting to see how DNMT1 effect get around classic understanding about gene body methylation. If local DNA methylation increased after DNMT1 targeting, what can be done is to make a DNMT1 catalytic dead mutant and test whether the mutant construct is still able to induce DNA hypermethylation or not. This may indicate that DNMT1 targeting dependent silencing is due to its catalytic activity or other factors like DNMT3a/b recruited by DNMT1 (Kim et al., 2002) to perform methylation.

One critical observation was that targeting PYL-DNMT1 to INT led to a repeat-dependent silencing of the GFP reporter. The first potential explanation may be the large size of the fusion with a molecular weight of 196 kDa. Long CAG repeats may induce heterochromatic markers accumulation and lower DNA accessibility (Cho et al., 2005). This may influence the efficiency of a big fusion protein binding to the GFP intron like DNMT1. We also have stable cell lines with a relatively big size (147kDa) protein expression, FANCI, and targeting FANCI does not reproduce the silencing effect observed in HDAC5 and DNMT1. This may indicate the size of the fusion protein is not the main reason for repeat dependent silencing.

Secondly, this short repeat specific silencing may merely due to the GFP intensity detection limitation. As the GFP reporter with 16 CAG repeats has higher baseline intensity after transcription than the 89 ones, the silencing effect difference of DNMT1 targeting compares to non-targeting can be less evident in long repeat cell lines. To test this hypothesis, GFP reporter mRNA qPCR may give a better resolution. If the suspect is correct, we may observe similar fold difference between CIT16BYD and CIT89BYD in ABA versus DMSO dataset. Polymerase II ChIP can also provide insight about local transcription.

Local DNA methylation may also affect splicing. It has been shown that exons tend to have more CpG methylation than introns in hESC (Chodavarapu et al., 2010; Laurent et al., 2010; Shukla et al., 2011). Sharp methylation level differences on exon-intron boundaries indicate the potential correlation between DNA methylation and splicing. The higher the methylation is, the more likely the gene-body fragment gets involved (Laurent et al., 2010; Shukla et al., 2011). So if DNMT1 targeting is inducing DNA methylation increase locally, it may result in the inclusion of hypermethylated INT-CAG intron.

Our GFP reporter is a splicing driven assay. The longer the CAG repeats are in the intron, the more likely GFP mRNA will be included the CAG repeat as part of an alternative exon, resulting in lower GFP intensity in CIT89BYD cells than in CIT16BYD cells. If the hypothesis is correct, that DNMT1 targeting ends up with

more intron inclusion into mRNA, GFP expression should show further silencing compare to non-targeted cell samples.

Furthermore, the observed allele specificity may also due to DNA methylation level. With more heterochromatic structure property for expanded CAG/CTG repeats, DNA methylation starting rate in CIT89BYD cells may be higher than that in CIT16BYD cells. DNMT1 targeting in expanded CAG repeat cell lines would have fewer substrates available to catalyze. It could explain why the reporter with 16 CAG repeats is more sensitive to DNMT1 targeting than 89 repeats.

CpG methylation and heterochromatic marks increase upon TNR expansion, regardless of the type of repeat (Dion and Wilson, 2009). It has been widely accepted that DNA methylation interplays with histone modifications like methylation and acetylation (Fuks et al., 2000; Robert et al., 2003). If the hypothesis proves to be true, that in extended CAG repeat-containing GFP locus has a different level of DNA methylation, it can be interesting to determine whether heterochromatic marks like H3K9me2/H3K9me3 appear upon PYL-DNMT1 targeting versus non-targeting. This observation can provide insights into whether DNA methylation is indispensable to recruit histone modifiers, or whether histone marks affect DNA methylation.

DNMT1 has been shown to cooperate with other DNA methyltransferases (DNMTs), histone methyltransferases (HMTs), histone deacetylases (HDACs) (Fuks, 2005; Fuks et al., 2000; Robert et al., 2003). Considering that DNMT1 directly affects gene silencing with low repeat specificity, I would imagine its working model by which DNMT3a/b and HDAC1 and 2 involved. DNMT3a/b are responsible for establishing CpG methylation and DNMT1 maintains DNA methylation after replication (Robert et al., 2003). If our bisulfate sequencing data shows that local DNA methylation increases upon PYL-DNMT1 targeting, it may indicate that DNMT3a/b are attracted by DNMT1 to the gene and catalyze CpG methylation, which further results in gene silencing. DNMT1 catalytic dead mutant targeting can also provide the correlation between DNMT1 catalytic activity and gene silencing. HDAC1/2 are also DNMT1 working partners for removing acetylation marks from histone proteins and cause heterochromatic structure formation and gene down-

regulation (Fuks et al., 2000). We can perform DNMT1 targeting combined with HDAC inhibitor-like TSA to see whether silencing effect disappears or not. If the answer is yes, it indicates that PYL-DNMT1 targeting induces silencing in an HDAC-dependent manner. Above all, DNMT1 targeting dependent silencing works in a DNA context-sensitive manner. This provides insight into the DNMT1 function in gene regulation.

3. Further applications of CIT

3.1 CIT can be used in other sequence contexts

Other apparent applications of CIT include replacing CAG repeat expansion with any sequence of interest, for instance, Friedreich's Ataxia's GAA repeats or the CGG repeats seen in Fragile X syndrome or even random sequence from the genome. If targeting POI induces GFP intensity change comparable to the DMSO group, it suggests that POI has a direct effect on the targeting region. In conclusion, CIT can be transplanted to any locus of interest, promising more extensive applications of this inducible chromatin targeting system to test protein working mechanism in any loci.

3.2 Screening of expanded allele silencers using CIT

Here we use a GFP reporter with different lengths of CAG repeats to uncover novel roles of chromatin modifying enzymes in gene regulation at a disease-relevant locus. Patients with CAG repeat expansions are generally heterozygous and harbor a normal allele that is important for function (Orr and Zoghbi, 2007; Ross and Tabrizi, 2011). This system will be particularly useful to find factors that could silence the expanded allele specifically.

The PYL construct designed here is amenable to large-scale studies since we can use a simple Gateway cloning reaction to fuse any protein of interest to PYL using currently available cDNA libraries (Katzen, 2007). Here we can take advantage of a ready-to-use database of epigenetic modifiers (Nanda et al., 2016) and clone them into PYL construct for transfection. After transfected into CITnB cells, positive transfected ones will be selected by proper selection marker. Then these samples

will pass through flow cytometry and western blot to further confirm their effect on the local gene expression. This large-scale chromatin modifiers screening will provide a further understanding of chromatin modifiers and disease-causing TNR expansion interplay.

4. Conclusion

In conclusion, the whole project goal was to build an inducible chromatin targeting system. I achieved this successfully and tested the effect of three chromatin modifying enzymes on gene expression and partially repeat instability (HDAC5 and DNMT1). CIT system can be further transplanted for mechanistic studies. My results challenge current models of how these proteins affected gene expression and clarified their indirect role in repeat instability. Moreover, the system can be used for further mechanistic studies and to uncover novel therapeutic targets for devastating expanded repeat disorders.

Chapter V

References

Abel, T., and Zukin, R.S. (2008). Epigenetic targets of HDAC inhibition in neurodegenerative and psychiatric disorders. Curr Opin Pharmacol *8*, 57–64.

Aeschbach, L., and Dion, V. (2017). Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications. Scientific Reports *7*, 18026.

van Agtmaal, E.L., André, L.M., Willemse, M., Cumming, S.A., van Kessel, I.D.G., van den Broek, W.J.A.A., Gourdon, G., Furling, D., Mouly, V., Monckton, D.G., et al. (2017). CRISPR/Cas9-Induced (CTG·CAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing. Molecular Therapy *25*, 24–43.

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell *167*, 219-232.e14.

An, M.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., Melov, S., and Ellerby, L.M. (2012). Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. Cell Stem Cell *11*, 253–263.

Andrew, S.E., Goldberg, Y.P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B., Kalchman, M.A., et al. (1993). The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nature Genetics *4*, 398.

Armour, S.M., Remsberg, J.R., Damle, M., Sidoli, S., Ho, W.Y., Li, Z., Garcia, B.A., and Lazar, M.A. (2017). An HDAC3-PROX1 corepressor module acts on HNF4α to control hepatic triglycerides. Nature Communications *8*, 549.

Axford, M.M., Wang, Y.-H., Nakamori, M., Zannis-Hadjopoulos, M., Thornton, C.A., and Pearson, C.E. (2013). Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. PLoS Genet. *9*, e1003866.

Backs, J., Backs, T., Bezprozvannaya, S., McKinsey, T.A., and Olson, E.N. (2008). Histone Deacetylase 5 Acquires Calcium/Calmodulin-Dependent Kinase II Responsiveness by Oligomerization with Histone Deacetylase 4. Mol. Cell. Biol. *28*, 3437–3445.

Banerjee, D., Mandal, S.M., Das, A., Hegde, M.L., Das, S., Bhakat, K.K., Boldogh, I., Sarkar, P.S., Mitra, S., and Hazra, T.K. (2011). Preferential Repair of Oxidized Base Damage in the Transcribed Genes of Mammalian Cells. J. Biol. Chem. *286*, 6006–6016.

Barbé, L., Lanni, S., López-Castel, A., Franck, S., Spits, C., Keymolen, K., Seneca, S., Tomé, S., Miron, I., Letourneau, J., et al. (2017). CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy. Am. J. Hum. Genet. *100*, 488–505.

Barlow, A.D., Xie, J., Moore, C.E., Campbell, S.C., Shaw, J. a. M., Nicholson, M.L., and Herbert, T.P. (2012). Rapamycin toxicity in MIN6 cells and rat and human islets is mediated by the inhibition of mTOR complex 2 (mTORC2). Diabetologia *55*, 1355–1365.

Barlow, A.D., Nicholson, M.L., and Herbert, T.P. (2013). Evidence for rapamycin toxicity in pancreatic β-cells and a review of the underlying molecular mechanisms. Diabetes *62*, 2674–2682.

Baylis, F., and McLeod, M. (2017). First-in-human Phase 1 CRISPR Gene Editing Cancer Trials:Are We Ready? Curr Gene Ther *17*, 309–319.

Beerli, R.R., Dreier, B., and Barbas, C.F. (2000). Positive and negative regulation of endogenous genes by designed transcription factors. Proc. Natl. Acad. Sci. U.S.A. *97*, 1495–1500.

Belmont, A.S. (2001). Visualizing chromosome dynamics with GFP. Trends in Cell Biology *11*, 250–257.

Belshawl, P.J., Spencer, D.M., Crabtree, G.R., and Schreiber, S.L. (1996). Controlling programmed cell death with a cyclophilincyclosporin-based chemical inducer of dimerization. Chemistry & Biology *3*, 731–738.

Beltran, A., Parikh, S., Liu, Y., Cuevas, B.D., Johnson, G.L., Futscher, B.W., and Blancafort, P. (2007). Re-activation of a dormant tumor suppressor gene maspin by designed transcription factors. Oncogene *26*, 2791–2798.

Beltran, A.S., Sun, X., Lizardi, P.M., and Blancafort, P. (2008). Reprogramming epigenetic silencing: artificial transcription factors synergize with chromatin remodeling drugs to reactivate the tumor suppressor mammary serine protease inhibitor. Mol Cancer Ther *7*, 1080–1090.

Biacsi, R., Kumari, D., and Usdin, K. (2008). SIRT1 Inhibition Alleviates Gene Silencing in Fragile X Mental Retardation Syndrome. PLoS Genet *4*.

Blackwood, J.K., Okely, E.A., Zahra, R., Eykelenboom, J.K., and Leach, D.R.F. (2010). DNA tandem repeat instability in the Escherichia coli chromosome is stimulated by mismatch repair at an adjacent CAG·CTG trinucleotide repeat. Proc. Natl. Acad. Sci. U.S.A. *107*, 22582–22586.

Blancafort, P., Steinberg, S.V., Paquin, B., Klinck, R., Scott, J.K., and Cedergren, R. (1999). The recognition of a noncanonical RNA base pair by a zinc finger protein. Chem. Biol. *6*, 585–597.

Blencowe, B.J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends in Biochemical Sciences *25*, 106–110.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science *326*, 1509–1512.

Boga (Pekmezekmek, A., Binokay, S., and Sertdemir, Y. (2009). The toxicity and teratogenicity of gibberellic acid (GA 3) based on the frog embryo teratogenesis assay-Xenopus (FETAX). Turkish Journal of Biology *33*.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536–2539.

Braun, S.M.G., Kirkland, J.G., Chory, E.J., Husmann, D., Calarco, J.P., and Crabtree, G.R. (2017). Rapid and reversible epigenome editing by endogenous chromatin regulators. Nature Communications *8*, 560.

Broek, V.D., A, W.J.A., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., te Riele, H., Groenen, P.J.T.A., and Wieringa, B. (2002). Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch–repair proteins. Hum Mol Genet *11*, 191–198.

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.-P., Hudson, T., et al. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell *68*, 799–808.

Bultmann, S., Morbitzer, R., Schmidt, C.S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T., and Leonhardt, H. (2012). Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. Nucleic Acids Res *40*, 5368–5377.

Bustos, A.D., Cuadrado, A., and Jouve, N. (2016). Sequencing of long stretches of repetitive DNA. Scientific Reports *6*, 36665.

Butler, R., and Bates, G.P. (2006). Histone deacetylase inhibitors as therapeutics for polyglutamine disorders. Nature Reviews Neuroscience *7*, 784.

Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996). Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. Science *271*, 1423–1427.

Cano-Rodriguez, D., Gjaltema, R.A.F., Jilderda, L.J., Jellema, P., Dokter-Fokkens, J., Ruiters, M.H.J., and Rots, M.G. (2016). Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat Commun *7*, 12284.

Carpenter, A.E., Memedula, S., Plutz, M.J., and Belmont, A.S. (2005). Common effects of acidic activators on large-scale chromatin structure and transcription. Mol. Cell. Biol. *25*, 958–968.

Castel, A.L., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol *11*, 165–170.

Celik, I., Tuluce, Y., and Isik, I. (2007). Evalution of toxicity of abcisic acid and gibberellic acid in rats: 50 days drinking water study. Journal of Enzyme Inhibition and Medicinal Chemistry *22*, 219–226.

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.-W., and Xi, J.J. (2013). Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. Cell Research *23*, 465–472.

Chen, S., Oikonomou, G., Chiu, C.N., Niles, B.J., Liu, J., Lee, D.A., Antoshechkin, I., and Prober, D.A. (2013). A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. Nucleic Acids Res *41*, 2769–2778.

Cho, D.H., Thienes, C.P., Mahoney, S.E., Analau, E., Filippova, G.N., and Tapscott, S.J. (2005). Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. Mol. Cell *20*, 483–489.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nature Biotechnology *31*, 230–232.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., et al. (2010). Relationship between nucleosome positioning and DNA methylation. Nature *466*, 388–392.

Chutake, Y.K., Lam, C.C., Costello, W.N., Anderson, M.P., and Bidichandani, S.I. (2016). Reversal of epigenetic promoter silencing in Friedreich ataxia by a class I histone deacetylase inhibitor. Nucleic Acids Res *44*, 5095–5104.

Cinesi, C., Aeschbach, L., Yang, B., and Dion, V. (2016). Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. Nature Communications *7*, 13272.

Claassen, D.A., and Lahue, R.S. (2007). Expansions of CAG·CTG repeats in immortalized human astrocytes. Hum. Mol. Genet. *16*, 3088–3096.

Cleary, J.D., and Pearson, C.E. (2003). The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. Cytogenet. Genome Res. *100*, 25–55.

Coffee, B., Zhang, F., Warren, S.T., and Reines, D. (1999). Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. Nat Genet *22*, 98–101.

Coffee, B., Zhang, F., Ceman, S., Warren, S.T., and Reines, D. (2002). Histone Modifications Depict an Aberrantly Heterochromatinized FMR1 Gene in Fragile X Syndrome. The American Journal of Human Genetics *71*, 923–932.

Cong, L., Zhou, R., Kuo, Y., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nature Communications *3*, 968.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Dabrowska, M., Juzwa, W., Krzyzosiak, W.J., and Olejniczak, M. (2018). Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases. Front Neurosci *12*.

Dandelot, E., and Tomé, S. (2017). Genetic Modifiers of CAG.CTG Repeat Instability in Huntington's Disease Mouse Models.

Debacker, K., Frizzell, A., Gleeson, O., Kirkham-McCarthy, L., Mertz, T., and Lahue, R.S. (2012). Histone Deacetylase Complexes Promote Trinucleotide Repeat Expansions. PLOS Biol *10*, e1001257.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y., and Yan, N. (2012). Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. Science *335*, 720–723.

Di Stasi, A., Tey, S.-K., Dotti, G., Fujita, Y., Kennedy-Nasser, A., Martinez, C., Straathof, K., Liu, E., Durett, A.G., Grilley, B., et al. (2011). Inducible Apoptosis as a Safety Switch for Adoptive Cell Therapy. New England Journal of Medicine *365*, 1673–1683.

Didonna, A., and Opal, P. (2015). The promise and perils of HDAC inhibitors in neurodegeneration. Ann Clin Transl Neurol *2*, 79–101.

Dion, V., and Wilson, J.H. (2009). Instability and chromatin structure of expanded trinucleotide repeats. Trends in Genetics *25*, 288–297.

Dion, V., Lin, Y., Hubert, L., Waterland, R.A., and Wilson, J.H. (2008). Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. Hum. Mol. Genet. *17*, 1306–1317.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nature Biotechnology *32*, 1262–1267.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology *34*, 184–191.

Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E.T., Friedberg, E.C., Kucherlapati, R., Edelmann, W., Lunetta, K.L., MacDonald, M.E., et al. (2009). Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol. Dis. *33*, 37–47.

Dubarry, N., Pasta, F., and Lane, D. (2006). ParABS Systems of the Four Replicons of Burkholderia cenocepacia: New Chromosome Centromeres Confer Partition Specificity. J. Bacteriol. *188*, 1489–1496.

Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M., et al. (1993). Trinucleotide repeat length instability and age of onset in Huntington's disease. Nature Genetics *4*, 387.

Echeverria, G.V., and Cooper, T.A. (2012). RNA-binding proteins in microsatellite expansion disorders: mediators of RNA toxicity. Brain Res. *1462*, 100–111.

Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy.

Eisenstein, M. (2012). Sangamo's lead zinc-finger therapy flops in diabetic neuropathy.

Elrick, L.L., Humphrey, M.B., Cooper, T.A., and Berget, S.M. (1998). A Short Sequence within Two Purine-Rich Enhancers Determines 5′ Splice Site Specificity. Mol. Cell. Biol. *18*, 343–352.

Elrod-Erickson, M., Rould, M.A., Nekludova, L., and Pabo, C.O. (1996). Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. Structure *4*, 1171–1180.

Ester, K., Ćurković-Perica, M., and Kralj, M. (2009). The Phytohormone Auxin Induces G1 Cell-Cycle Arrest of Human Tumor Cells. Planta Med *75*, 1423–1426.

Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature *366*, 483–487.

Falahi, F., Huisman, C., Kazemier, H.G., Vlies, P. van der, Kok, K., Hospers, G.A.P., and Rots, M.G. (2013). Towards Sustained Silencing of HER2/neu in Cancer By Epigenetic Editing. Mol Cancer Res *11*, 1029–1039.

Farrell, B.T., and Lahue, R.S. (2006). CAG*CTG repeat instability in cultured human astrocytes. Nucleic Acids Res. *34*, 4495–4505.

Feinberg, A.P. (2007). Phenotypic plasticity and the epigenetics of human disease.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. PNAS *107*, 8689–8694.

Filla, A., De Michele, G., Cavalcanti, F., Pianese, L., Monticelli, A., Campanella, G., and Cocozza, S. (1996). The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. Am J Hum Genet *59*, 554–560.

Fischle, W., Dequiedt, F., Hendzel, M.J., Guenther, M.G., Lazar, M.A., Voelter, W., and Verdin, E. (2002). Enzymatic Activity Associated with Class II HDACs Is Dependent on a Multiprotein Complex Containing HDAC3 and SMRT/N-CoR. Molecular Cell *9*, 45–57.

Foiry, L., Dong, L., Savouret, C., Hubert, L., Riele, H. te, Junien, C., and Gourdon, G. (2006). Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. Hum Genet *119*, 520–526.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Curr. Opin. Genet. Dev. *15*, 490–495.

Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nature Genetics *24*, 88–91.

Gacy, A.M., Goellner, G., Juranić, N., Macura, S., and McMurray, C.T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell *81*, 533–540.

Gannon, A.-M.M., Frizzell, A., Healy, E., and Lahue, R.S. (2012). MutSβ and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells. Nucl. Acids Res. *40*, 10324–10333.

Gao, Y., Xiong, X., Wong, S., Charles, E.J., Lim, W.A., and Qi, L.S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. Nature Methods *13*, 1043–1049.

Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2015). Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell *162*, 516–526.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Glozak, M.A., Sengupta, N., Zhang, X., and Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. Gene *363*, 15–23.

Goldberg, Y.P., McMurray, C.T., Zeisler, J., Almqvist, E., Sillence, D., Richards, F., Gacy, A.M., Buchanan, J., Telenius, H., and Hayden, M.R. (1995). Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. Hum. Mol. Genet. *4*, 1911–1918.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395–398.

Gorbunova, V., Seluanov, A., Dion, V., Sandor, Z., Meservy, J.L., and Wilson, J.H. (2003). Selectable System for Monitoring the Instability of CTG/CAG Triplet Repeats in Mammalian Cells. Mol. Cell. Biol. *23*, 4485–4493.

Gori, J.L., Hsu, P.D., Maeder, M.L., Shen, S., Welstead, G.G., and Bumcrot, D. (2015). Delivery and Specificity of CRISPR/Cas9 Genome Editing Technologies for Human Gene Therapy. Human Gene Therapy *26*, 443–451.

Graef, I.A., Holsinger, L.J., Diver, S., Schreiber, S.L., and Crabtree, G.R. (1997). Proximity and orientation underlie signaling by the non-receptor tyrosine kinase ZAP70. EMBO J *16*, 5618–5628.

Greene, E., Mahishi, L., Entezam, A., Kumari, D., and Usdin, K. (2007). Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. Nucleic Acids Res *35*, 3383–3390.

Greer, C.B., Tanaka, Y., Kim, Y.J., Xie, P., Zhang, M.Q., Park, I.-H., and Kim, T.H. (2015). Histone Deacetylases Positively Regulate Transcription through the Elongation Machinery. Cell Rep *13*, 1444–1455.

Gregory, D.J., Zhang, Y., Kobzik, L., and Fedulov, A.V. (2013). Specific transcriptional enhancement of inducible nitric oxide synthase by targeted promoter demethylation. Epigenetics *8*, 1205–1212.

Grimmer, M.R., Stolzenburg, S., Ford, E., Lister, R., Blancafort, P., and Farnham, P.J. (2014). Analysis of an artificial zinc finger epigenetic modulator: widespread binding but limited regulation. Nucleic Acids Res *42*, 10856–10868.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. PLOS Genetics *6*, e1000869.

Groote, D., L, M., Verschure, P.J., and Rots, M.G. (2012). Epigenetic Editing : targeted rewriting of epigenetic marks to modulate expression of selected target genes. Nucleic Acids Res *40*, 10596–10613.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., and Nasmyth, K. (2006). Evidence that Loading of Cohesin Onto Chromosomes Involves Opening of Its SMC Hinge. Cell *127*, 523–537.

Guilinger, J.P., Pattanayak, V., Reyon, D., Tsai, S.Q., Sander, J.D., Joung, J.K., and Liu, D.R. (2014). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. Nature Methods *11*, 429–435.

Hanawalt, P.C. (2002). Subpathways of nucleotide excision repair and their regulation. Oncogene *21*, 8949.

Harley, H.G., Rundle, S.A., MacMillan, J.C., Myring, J., Brook, J.D., Crow, S., Reardon, W., Fenton, I., Shaw, D.J., and Harper, P.S. (1993). Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. Am. J. Hum. Genet. *52*, 1164–1174.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and Memory of Heterochromatin in Living Cells. Cell *149*, 1447–1460.

Hellman, A., and Chess, A. (2007). Gene Body-Specific Methylation on the Active X Chromosome. Science *315*, 1141–1143.

Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L., and Gottesfeld, J.M. (2006a). Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. Nat. Chem. Biol. *2*, 551–558.

Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L., and Gottesfeld, J.M. (2006b). Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. Nature Chemical Biology *2*, 551.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nature Biotechnology *33*, 510–517.

Ho, S.N., Biggar, S.R., Spencer, D.M., Schreiber, S.L., and Crabtree, G.R. (1996). Dimeric ligands define a role for transcriptional activation domains in reinitiation. Nature *382*, 822–826.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B., et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nature Biotechnology *27*, 851–857.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nature Biotechnology *29*, 731–734.

Hockly, E., Richon, V.M., Woodman, B., Smith, D.L., Zhou, X., Rosa, E., Sathasivam, K., Ghazi-Noori, S., Mahal, A., Lowden, P.A.S., et al. (2003). Suberoylanilide hydroxamic acid, a histone deacetylase inhibitor, ameliorates motor deficits in a mouse model of Huntington's disease. PNAS *100*, 2041–2046.

Holkers, M., Maggio, I., Liu, J., Janssen, J.M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T., and Gonçalves, M.A.F.V. (2013). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucleic Acids Res *41*, e63–e63.

Holmans, P.A., Massey, T.H., and Jones, L. (2017). Genetic modifiers of Mendelian disease: Huntington's disease and the trinucleotide repeat disorders. Hum Mol Genet *26*, R83–R90.

Holsinger, L.J., Spencer, D.M., Austin, D.J., Schreiber, S.L., and Crabtree, G.R. (1995). Signal transduction in T lymphocytes using a conditional allele of Sos. Proc Natl Acad Sci U S A *92*, 9810–9814.

Hong, D.-H., and Li, T. (2002). Complex Expression Pattern of RPGR Reveals a Role for Purine-Rich Exonic Splicing Enhancers. Invest. Ophthalmol. Vis. Sci. *43*, 3373–3382.

Hornstra, I.K., Nelson, D.L., Warren, S.T., and Yang, T.P. (1993). High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. Hum. Mol. Genet. *2*, 1659–1665.

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., and Schmid, B. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. Development *140*, 4982–4987.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology *31*, 827–832.

Huang, Y.-H., Su, J., Lei, Y., Brunetti, L., Gundry, M.C., Zhang, X., Jeong, M., Li, W., and Goodell, M.A. (2017). DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. Genome Biology *18*, 176.

Hubert, L., Lin, Y., Dion, V., and Wilson, J.H. (2011). Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. Hum. Mol. Genet. *20*, 4822–4830.

Huisman, C., Wijst, M.G. van der, Falahi, F., Overkamp, J., Karsten, G., Terpstra, M.M., Kok, K., Zee, A.G. van der, Schuuring, E., Wisman, G.B.A., et al. (2015). Prolonged re-expression of the hypermethylated gene EPB41L3 using artificial transcription factors and epigenetic drugs. Epigenetics *10*, 384–396.

Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O., and Joung, J.K. (2003). Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. PNAS *100*, 12271–12276.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature Biotechnology *31*, 227–229.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429–5433.

Jacobs, J.Z., Ciccaglione, K.M., Tournier, V., and Zaratiegui, M. (2014). Implementation of the CRISPR-Cas9 system in fission yeast. Nature Communications *5*, 5344.

Jacome, A., and Fernandez-Capetillo, O. (2011). Lac operator repeats generate a traceable fragile site in mammalian cells. EMBO Reports *12*, 1032–1038.

Jaworski, A., Rosche, W.A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R.P., Sinden, R.R., and Wells, R.D. (1995). Mismatch repair in Escherichia coli enhances instability of (CTG)n triplet repeats from human hereditary diseases. PNAS *92*, 11019–11023.

Je, H.-S., Lu, Y., Yang, F., Nagappan, G., Zhou, J., Jiang, Z., Nakazawa, K., and Lu, B. (2009). Chemically Inducible Inactivation of Protein Synthesis in Genetically Targeted Neurons. J Neurosci *29*, 6761–6766.

Jia, H., Morris, C.D., Williams, R.M., Loring, J.F., and Thomas, E.A. (2015). HDAC inhibition imparts beneficial transgenerational effects in Huntington's disease mice via altered DNA and histone methylation. PNAS *112*, E56–E64.

Jia, H., Wang, Y., Morris, C.D., Jacques, V., Gottesfeld, J.M., Rusche, J.R., and Thomas, E.A. (2016). The Effects of Pharmacological Inhibition of Histone Deacetylase 3 (HDAC3) in Huntington's Disease Mice. PLOS ONE *11*, e0152498.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821.

Jones, P.A. (1999). The DNA methylation paradox. Trends Genet. *15*, 34–37.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics *13*, 484–492.

Joung, J.K., and Sander, J.D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nature Reviews Molecular Cell Biology *14*, 49–55.

Juillerat, A., Dubois, G., Valton, J., Thomas, S., Stella, S., Maréchal, A., Langevin, S., Benomari, N., Bertonati, C., Silva, G.H., et al. (2014). Comprehensive analysis of the specificity of transcription activator-like effector nucleases. Nucleic Acids Res *42*, 5390–5402.

Kang, C.B., Hong, Y., Dhe-Paganon, S., and Yoon, H.S. (2008). FKBP Family Proteins: Immunophilins with Versatile Biological Functions. NSG *16*, 318–325.

Kao, H.-Y., Downes, M., Ordentlich, P., and Evans, R.M. (2000). Isolation of a novel histone deacetylase reveals that class I and class II deacetylases promote SMRT-mediated repression. Genes Dev. *14*, 55–66.

Katzen, F. (2007). Gateway(®) recombinational cloning: a biological operating system. Expert Opin Drug Discov *2*, 571–589.

Kennedy, M.J., Hughes, R.M., Peteya, L.A., Schwartz, J.W., Ehlers, M.D., and Tucker, C.L. (2010). Rapid blue light induction of protein interactions in living cells. Nat Methods *7*, 973–975.

Khare, D., Ziegelin, G., Lanka, E., and Heinemann, U. (2004). Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. Nature Structural & Molecular Biology *11*, 656–663.

Kick, L., Kirchner, M., and Schneider, S. (2017). CRISPR-Cas9: From a bacterial immune system to genome-edited human cells in clinical trials. Bioengineered *8*, 280–286.

Kim, G.-D., Ni, J., Kelesoglu, N., Roberts, R.J., and Pradhan, S. (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. The EMBO Journal *21*, 4183–4195.

Kim, G.S., Jung, H.-E., Kim, J.-S., and Lee, Y.C. (2015). Mutagenesis Study Reveals the Rim of Catalytic Entry Site of HDAC4 and -5 as the Major Binding Surface of SMRT Corepressor. PLOS ONE *10*, e0132680.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P.W., Li, Z., Peterson, R.T., Yeh, J.-R.J., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Knight, S.J.L., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R., Barnicoat, A., Pembrey, M.E., et al. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell *74*, 127–134.

Kohler, J.J., and Bertozzi, C.R. (2003). Regulating Cell Surface Glycosylation by Small Molecule Control of Enzyme Localization. Chemistry & Biology *10*, 1303–1311.

Kohwi, Y., Wang, H., and Kohwi-Shigematsu, T. (1993). A single trinucleotide, 5'AGC3'/5'GCT3', of the triplet-repeat disease genes confers metal ion-induced non-B DNA structure. Nucleic Acids Res. *21*, 5651–5655.

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral–pallidoluysian atrophy (DRPLA). Nature Genetics *6*, 9.

Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Le Cong, Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472–476.

Kovtun, I.V., Thornhill, A.R., and McMurray, C.T. (2004). Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. Hum. Mol. Genet. *13*, 3057–3068.

Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H., and McMurray, C.T. (2007). OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature *447*, 447–452.

Kumari, D., Biacsi, R.E., and Usdin, K. (2011). Repeat Expansion Affects Both Transcription Initiation and Elongation in Friedreich Ataxia Cells. J. Biol. Chem. *286*, 4209–4215.

Kunkel, T.A., and Erie, D.A. (2005). Dna Mismatch Repair. Annual Review of Biochemistry *74*, 681–710.

Kurdistani, S.K., Robyr, D., Tavazoie, S., and Grunstein, M. (2002). Genome-wide binding map of the histone deacetylase Rpd3 in yeast. Nature Genetics *31*, 248–254.

Kwon, D.Y., Zhao, Y.-T., Lamonica, J.M., and Zhou, Z. (2017). Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. Nature Communications *8*, 15315.

Lahm, A., Paolini, C., Pallaoro, M., Nardi, M.C., Jones, P., Neddermann, P., Sambucini, S., Bottomley, M.J., Surdo, P.L., Carfí, A., et al. (2007). Unraveling the hidden catalytic activity of vertebrate class IIa histone deacetylases. PNAS *104*, 17335–17340.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Sung, K.W.K., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. Genome Res. *20*, 320–331.

Lee, J.-M., Ramos, E.M., Lee, J.-H., Gillis, T., Mysore, J.S., Hayden, M.R., Warby, S.C., Morrison, P., Nance, M., Ross, C.A., et al. (2012). CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. Neurology *78*, 690–695.

Lee, M.G., Wynder, C., Bochar, D.A., Hakimi, M.-A., Cooch, N., and Shiekhattar, R. (2006). Functional Interplay between Histone Demethylase and Deacetylase Enzymes. Mol. Cell. Biol. *26*, 6395–6402.

Levskaya, A., Weiner, O.D., Lim, W.A., and Voigt, C.A. (2009). Spatiotemporal control of cell signalling using a light-switchable protein interaction. Nature *461*, 997–1001.

Li, F., Papworth, M., Minczuk, M., Rohde, C., Zhang, Y., Ragozin, S., and Jeltsch, A. (2007). Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. Nucleic Acids Res. *35*, 100–112.

Li, Y., Polak, U., Bhalla, A.D., Rozwadowska, N., Butler, J.S., Lynch, D.R., Dent, S.Y.R., and Napierala, M. (2015). Excision of Expanded GAA Repeats Alleviates the Molecular Phenotype of Friedreich's Ataxia. Mol. Ther. *23*, 1055–1065.

Liang, F.-S., Ho, W.Q., and Crabtree, G.R. (2011). Engineering the ABA Plant Stress Pathway for Regulation of Induced Proximity. Sci. Signal. *4*, rs2–rs2.

Lin, Y., and Wilson, J.H. (2007). Transcription-Induced CAG Repeat Contraction in Human Cells Is Mediated in Part by Transcription-Coupled Nucleotide Excision Repair. Mol. Cell. Biol. *27*, 6209–6217.

Lin, A., Giuliano, C.J., Sayles, N.M., and Sheltzer, J.M. CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. ELife *6*.

Lin, Y., Dion, V., and Wilson, J.H. (2005). A novel selectable system for detecting expansion of CAG.CTG repeats in mammalian cells. Mutat. Res. *572*, 123–131.

Lin, Y., Dion, V., and Wilson, J.H. (2006). Transcription promotes contraction of CAG repeat tracts in human cells. Nat Struct Mol Biol *13*, 179–180.

Lin, Y., Dent, S.Y.R., Wilson, J.H., Wells, R.D., and Napierala, M. (2010). R loops stimulate genetic instability of CTG·CAG repeats. PNAS *107*, 692–697.

Liu, G., Chen, X., Bissler, J.J., Sinden, R.R., and Leffak, M. (2010). Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells. Nat. Chem. Biol. *6*, 652–659.

Liu, K.J., Arron, J.R., Stankunas, K., Crabtree, G.R., and Longaker, M.T. (2007). Chemical rescue of cleft palate and midline defects in conditional GSK-3beta mice. Nature *446*, 79–82.

Liu, P.Q., Rebar, E.J., Zhang, L., Liu, Q., Jamieson, A.C., Liang, Y., Qi, H., Li, P.X., Chen, B., Mendel, M.C., et al. (2001). Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. J. Biol. Chem. *276*, 11323–11334.

Liu, Q., Zhang, P., Wang, D., Gu, W., and Wang, K. (2017). Interrogating the "unsequenceable" genomic trinucleotide repeat disorders by long-read sequencing. Genome Medicine *9*, 65.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247.e17.

Liu, X.S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., Hnisz, D., Li, C.H., Yuan, B., Xu, C., et al. (2018). Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. Cell *172*, 979-992.e6.

Liu, X.S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., Hnisz, D., Li, C.H., Yuan, B., Xu, C., et al. Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. Cell.

Lo, C.-L., Choudhury, S.R., Irudayaraj, J., and Zhou, F.C. (2017). Epigenetic Editing of *Ascl1* Gene in Neural Stem Cells by Optogenetics. Scientific Reports *7*, 42047.

Long, C., McAnally, J.R., Shelton, J.M., Mireault, A.A., Bassel-Duby, R., and Olson, E.N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9–mediated editing of germline DNA. Science *345*, 1184–1188.

Loomis, E.W., Eid, J.S., Peluso, P., Yin, J., Hickey, L., Rank, D., McCalmon, S., Hagerman, R.J., Tassone, F., and Hagerman, P.J. (2013). Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene. Genome Res *23*, 121–128.

Loomis, E.W., Sanz, L.A., Chédin, F., and Hagerman, P.J. (2014). Transcription-Associated R-Loop Formation across the Human FMR1 CGG-Repeat Region. PLOS Genetics *10*, e1004294.

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nature Reviews Genetics *19*, 81–92.

MacDonald, M.E., Ambrose, C.M., Duyao, M.P., Myers, R.H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S.A., James, M., Groot, N., et al. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell *72*, 971–983.

Maeder, M.L., Linder, S.J., Reyon, D., Angstman, J.F., Fu, Y., Sander, J.D., and Joung, J.K. (2013a). Robust, synergistic regulation of human gene expression using TALE activators. Nature Methods *10*, 243–245.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013b). CRISPR RNA–guided activation of endogenous human genes. Nature Methods *10*, 977.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013a). RNA-Guided Human Genome Engineering via Cas9. Science *339*, 823–826.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013b). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology *31*, 833.

Malvaez, M., McQuown, S.C., Rogge, G.A., Astarabadi, M., Jacques, V., Carreiro, S., Rusche, J.R., and Wood, M.A. (2013). HDAC3-selective inhibitor enhances extinction of cocaine-seeking behavior in a persistent manner. PNAS *110*, 2647–2652.

Manley, K., Shirley, T.L., Flaherty, L., and Messer, A. (1999). *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. Nature Genetics *23*, 471.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nature Reviews Genetics *11*, 285–296.

Mariamé, B., Kappler-Gratias, S., Kappler, M., Balor, S., Gallardo, F., and Bystricky, K. (2018). Real-time visualization and quantification of human Cytomegalovirus replication in living cells using the ANCHOR DNA labeling technology. BioRxiv 300111.

McCampbell, A., Taye, A.A., Whitty, L., Penney, E., Steffan, J.S., and Fischbeck, K.H. (2001). Histone deacetylase inhibitors reduce polyglutamine toxicity. PNAS *98*, 15179–15184.

McFarland, K.N., Liu, J., Landrian, I., Godiska, R., Shanker, S., Yu, F., Farmerie, W.G., and Ashizawa, T. (2015). SMRT Sequencing of Long Tandem Nucleotide Repeats in SCA10 Reveals Unique Insight of Repeat Expansion Structure. PLoS ONE *10*, e0135906.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet *11*, 786–799.

Meckler, J.F., Bhakta, M.S., Kim, M.-S., Ovadia, R., Habrian, C.H., Zykovich, A., Yu, A., Lockwood, S.H., Morbitzer, R., Elsäesser, J., et al. (2013). Quantitative analysis of TALE–DNA interactions suggests polarity effects. Nucleic Acids Res *41*, 4118–4128.

Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. Nature Biotechnology *31*, 1133–1136.

Merienne, N., Vachey, G., de Longprez, L., Meunier, C., Zimmer, V., Perriard, G., Canales, M., Mathias, A., Herrgott, L., Beltraminelli, T., et al. (2017). The Self-Inactivating KamiCas9 System for the Editing of CNS Disease Genes. Cell Reports *20*, 2980–2991.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. *4*, 1609–1614.

Mirkin, S.M. (2007). Expandable DNA repeats and human disease. Nature *447*, 932–940.

Mittelman, D., Moye, C., Morton, J., Sykoudis, K., Lin, Y., Carroll, D., and Wilson, J.H. (2009). Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. Proc. Natl. Acad. Sci. U.S.A. *106*, 9607–9612.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.

Monckton, D.G., Wong, L.J., Ashizawa, T., and Caskey, C.T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. Hum. Mol. Genet. *4*, 1–8.

Moore, M., Klug, A., and Choo, Y. (2001). Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. Proc. Natl. Acad. Sci. U.S.A. *98*, 1437–1441.

Mootz, H.D., and Muir, T.W. (2002). Protein Splicing Triggered by a Small Molecule. J. Am. Chem. Soc. *124*, 9044–9045.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., et al. (2016). Targeted DNA demethylation *in vivo* using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. Nature Biotechnology *34*, 1060–1065.

Nadel, Y., Weisman-Shomer, P., and Fry, M. (1995). The fragile X syndrome single strand d(CGG)n nucleotide repeats readily fold back to form unimolecular hairpin structures. J. Biol. Chem. *270*, 28970–28977.

Nageshwaran, S., and Festenstein, R. (2015). Epigenetics and Triplet-Repeat Neurological Diseases. Front. Neurol. 262.

Nanda, J.S., Kumar, R., and Raghava, G.P.S. (2016). dbEM: A database of epigenetic modifiers curated from cancerous and normal genomes. Scientific Reports *6*, 19340.

Niu, J., Zhang, B., and Chen, H. (2014). Applications of TALENs and CRISPR/Cas9 in Human Cells and Their Potentials for Gene Therapy. Mol Biotechnol *56*, 681–688.

Nunna, S., Reinhardt, R., Ragozin, S., and Jeltsch, A. (2014). Targeted Methylation of the Epithelial Cell Adhesion Molecule (EpCAM) Promoter to Silence Its Expression in Ovarian Cancer Cells. PLOS ONE *9*, e87703.

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boué, J., Bertheas, M.F., and Mandel, J.L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science *252*, 1097–1102.

Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide Repeat Disorders. Annual Review of Neuroscience *30*, 575–621.

Orr, H.T., Chung, M., Banfi, S., Jr, T.J.K., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P.W., and Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genetics *4*, 221.

Otten, A.D., and Tapscott, S.J. (1995). Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. PNAS *92*, 5465–5469.

Ouellet, D.L., Cherif, K., Rousseau, J., and Tremblay, J.P. (2017). Deletion of the GAA repeats from the human frataxin gene using the CRISPR-Cas9 system in YG8R-derived cells and mouse models of Friedreich ataxia. Gene Therapy *24*, 265–274.

Park, C.-Y., Halevy, T., Lee, D.R., Sung, J.J., Lee, J.S., Yanuka, O., Benvenisty, N., and Kim, D.-W. (2015). Reversion of FMR1 Methylation and Silencing by Editing the Triplet Repeats in Fragile X iPSC-Derived Neurons. Cell Reports *13*, 234–241.

Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809–817.

Pearson, C.E., Sinden, R.R., Wang, Y.-H., and Griffith, J.D. (1998). Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)n·(CAG)n repeats from the myotonic dystrophy locus. Nucl. Acids Res. *26*, 816–823.

Pearson, C.E., Tam, M., Wang, Y.-H., Montgomery, S.E., Dar, A.C., Cleary, J.D., and Nichol, K. (2002). Slipped-strand DNAs formed by long (CAG)*(CTG) repeats: slipped-out repeats and slip-out junctions. Nucleic Acids Res. *30*, 4534–4547.

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A. (2013a). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nature Methods *10*, 239–242.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013b). RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nature Methods *10*, 973.

Pinto, B.S., Saxena, T., Oliveira, R., Méndez-Gómez, H.R., Cleary, J.D., Denes, L.T., McConnell, O., Arboleda, J., Xia, G., Swanson, M.S., et al. (2017). Impeding Transcription of Expanded Microsatellite Repeats by Deactivated Cas9. Molecular Cell *68*, 479-490.e5.

Pinto, R.M., Dragileva, E., Kirby, A., Lloret, A., Lopez, E., Claire, J.S., Panigrahi, G.B., Hou, C., Holloway, K., Gillis, T., et al. (2013). Mismatch Repair Genes Mlh1 and Mlh3

Modify CAG Instability in Huntington's Disease Mice: Genome-Wide and Candidate Approaches. PLOS Genet *9*, e1003930.

Polstein, L.R., Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Bledsoe, P., Song, L., Safi, A., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. Genome Res. *25*, 1158–1169.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nature Biotechnology *28*, 1057–1068.

Punga, T., and Bühler, M. (2010). Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. EMBO Mol Med *2*, 120–129.

Qasim, W., Amrolia, P.J., Samarasinghe, S., Ghorashian, S., Zhan, H., Stafford, S., Butler, K., Ahsan, G., Gilmour, K., Adams, S., et al. (2015). First Clinical Application of Talen Engineered Universal CAR19 T Cells in B-ALL. Blood *126*, 2046–2046.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Rafalska-Metcalf, I.U., and Janicki, S.M. (2007). Show and tell: visualizing gene expression in living cells. J Cell Sci *120*, 2301–2307.

Rando, O.J., and Chang, H.Y. (2009). Genome-Wide Views of Chromatin Structure. Annual Review of Biochemistry *78*, 245–271.

Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.-H., and Pearson, C.E. (2011). Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic Acids Res. *39*, 1749–1762.

Reddy, K., Schmidt, M.H.M., Geist, J.M., Thakkar, N.P., Panigrahi, G.B., Wang, Y.-H., and Pearson, C.E. (2014). Processing of double-R-loops in (CAG)·(CTG) and C9orf72 (GGGGCC)·(GGCCCC) repeats causes instability. Nucleic Acids Res. *42*, 10473–10487.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nature Biotechnology *30*, 460–465.

Rinaldi, F.C., Doyle, L.A., Stoddard, B.L., and Bogdanove, A.J. (2017). The effect of increasing numbers of repeats on TAL effector DNA binding specificity. Nucleic Acids Res *45*, 6960–6970.

Robert, M.-F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I.C., Barsalou, A., and MacLeod, A.R. (2003). DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. Nature Genetics *33*, 61–65.

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

Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. Nature Reviews Genetics *1*, 11.

Rogers, J.M., Barrera, L.A., Reyon, D., Sander, J.D., Kellis, M., Joung, J.K., and Bulyk, M.L. (2015). Context influences on TALE–DNA binding revealed by quantitative profiling. Nature Communications *6*, 7440.

Ross, C.A., and Tabrizi, S.J. (2011). Huntington's disease: from molecular pathogenesis to clinical treatment. The Lancet Neurology *10*, 83–98.

Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D., and Bystricky, K. (2014). DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. PLoS Genet. *10*, e1004187.

Santillan, B.A., Moye, C., Mittelman, D., and Wilson, J.H. (2014). GFP-based fluorescence assay for CAG repeat instability in cultured human cells. PLoS ONE *9*, e113952.

Saveliev, A., Everett, C., Sharpe, T., Webster, Z., and Festenstein, R. (2003). DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. Nature *422*, 909–913.

Schmidt, M.H.M., and Pearson, C.E. (2016). Disease-associated repeat instability and mismatch repair. DNA Repair *38*, 117–126.

Schweitzer, J.K., and Livingston, D.M. (1997). Destabilization of CAG Trinucleotide Repeat Tracts by Mismatch Repair Mutations in Yeast. Hum. Mol. Genet. *6*, 349–355.

Segal, D.J., Dreier, B., Beerli, R.R., and Barbas, C.F. (1999). Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. PNAS *96*, 2758–2763.

Shah, S.A., Erdmann, S., Mojica, F.J.M., and Garrett, R.A. (2013). Protospacer recognition motifs. RNA Biology *10*, 891–899.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of Site-Specific Histone Acetylation and Deacetylation. Annual Review of Biochemistry *76*, 75–100.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. Cell *119*, 941–953.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88.

Snowden, A.W., Gregory, P.D., Case, C.C., and Pabo, C.O. (2002). Gene-Specific Targeting of H3K9 Methylation Is Sufficient for Initiating Repression In Vivo. Current Biology *12*, 2159–2166.

Soragni, E., and Gottesfeld, J.M. (2016). Translating HDAC inhibitors in Friedreich's ataxia. Expert Opinion on Orphan Drugs *4*, 961–970.

Soragni, E., Miao, W., Iudicello, M., Jacoby, D., De Mercanti, S., Clerico, M., Longo, F., Piga, A., Ku, S., Campau, E., et al. (2014). Epigenetic therapy for Friedreich ataxia. Ann. Neurol. *76*, 489–508.

Spange, S., Wagner, T., Heinzel, T., and Krämer, O.H. (2009). Acetylation of non-histone proteins modulates cellular signalling at multiple levels. The International Journal of Biochemistry & Cell Biology *41*, 185–198.

Spencer, D.M., Wandless, T.J., Schreiber, S.L., and Crabtree, G.R. (1993). Controlling signal transduction with synthetic ligands. Science *262*, 1019–1024.

Stankunas, K., Bayle, J.H., Gestwicki, J.E., Lin, Y.-M., Wandless, T.J., and Crabtree, G.R. (2003). Conditional protein alleles using knockin mice and a chemical inducer of dimerization. Mol. Cell *12*, 1615–1624.

Stanton, B.Z., Chory, E.J., and Crabtree, G.R. (2018). Chemically induced proximity in biology and medicine. Science *359*, eaao5902.

Steffan, J.S., Bodai, L., Pallos, J., Poelman, M., McCampbell, A., Apostol, B.L., Kazantsev, A., Schmidt, E., Zhu, Y.-Z., Greenwald, M., et al. (2001). Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in *Drosophila*. Nature *413*, 739.

Steinbach, P., Gläser, D., Vogel, W., Wolf, M., and Schwemmle, S. (1998). The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. Am J Hum Genet *62*, 278–285.

Stine, O.C., Pleasant, N., Franz, M.L., Abbott, M.H., Folstein, S.E., and Ross, C.A. (1993). Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. Hum Mol Genet *2*, 1547–1549.

Stöger, R., Kajimura, T.M., Brown, W.T., and Laird, C.D. (1997). Epigenetic Variation Illustrated by DNA Methylation Patterns of the Fragile-X Gene FMR1. Hum Mol Genet *6*, 1791–1801.

Stolzenburg, S., Rots, M.G., Beltran, A.S., Rivenbark, A.G., Yuan, X., Qian, H., Strahl, B.D., and Blancafort, P. (2012). Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer. Nucleic Acids Res *40*, 6725–6740.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245.

Suelves, N., Kirkham-McCarthy, L., Lahue, R.S., and Ginés, S. (2017). A selective inhibitor of histone deacetylase 3 prevents cognitive deficits and suppresses striatal CAG repeat expansions in Huntington's disease mice. Scientific Reports *7*, 6082.

Sutcliffe, J.S., Nelson, D.L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D., and Warren, S.T. (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum. Mol. Genet. *1*, 397–400.

Tacke, R., and Manley, J.L. (1999). Determinants of SR protein specificity. Current Opinion in Cell Biology *11*, 358–362.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell *159*, 635–646.

Tomé, S., Holt, I., Edelmann, W., Morris, G.E., Munnich, A., Pearson, C.E., and Gourdon, G. (2009). MSH2 ATPase Domain Mutation Affects CTG•CAG Repeat Instability in Transgenic Mice. PLOS Genetics *5*, e1000482.

Tomé, S., Manley, K., Simard, J.P., Clark, G.W., Slean, M.M., Swami, M., Shelbourne, P.F., Tillier, E.R.M., Monckton, D.G., Messer, A., et al. (2013). MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet. *9*, e1003280.

Tsai, Y.-C., Greenberg, D., Powell, J., Hoijer, I., Ameur, A., Strahl, M., Ellis, E., Jonasson, I., Pinto, R.M., Wheeler, V., et al. (2017). Amplification-free, CRISPR-Cas9 Targeted Enrichment and SMRT Sequencing of Repeat-Expansion Disease Causative Genomic Regions. BioRxiv 203919.

Usdin, K., and Woodford, K.J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. Nucleic Acids Res. *23*, 4202–4209.

Usdin, K., House, N.C.M., and Freudenreich, C.H. (2015). Repeat instability during DNA repair: Insights from model systems. Critical Reviews in Biochemistry and Molecular Biology *50*, 142–167.

Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Maréchal, A., Macmaster, R., Melliand, K., Juillerat, A., and Duchateau, P. (2012). Overcoming Transcription Activator-like Effector (TALE) DNA Binding Domain Sensitivity to Cytosine Methylation. J. Biol. Chem. *287*, 38427–38432.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nature Reviews Molecular Cell Biology *16*, 178–189.

Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905–914.

Verschure, P.J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A.E., Belmont, A.S., and van Driel, R. (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. Mol. Cell. Biol. *25*, 4552–4564.

Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., Klasić, M., and Zoldoš, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res *44*, 5615–5628.

Wang, A., Kurdistani, S.K., and Grunstein, M. (2002). Requirement of Hos2 Histone Deacetylase for Gene Activity in Yeast. Science *298*, 1412–1414.

Wang, L., Lin, J., Zhang, T., Xu, K., Ren, C., and Zhang, Z. (2013). Simultaneous screening and validation of effective zinc finger nucleases in yeast. PLoS ONE *8*, e64687.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019–1031.

Warby, S.C., Montpetit, A., Hayden, A.R., Carroll, J.B., Butland, S.L., Visscher, H., Collins, J.A., Semaka, A., Hudson, T.J., and Hayden, M.R. (2009). CAG Expansion in the Huntington Disease Gene Is Associated with a Specific and Targetable Predisposing Haplogroup. Am J Hum Genet *84*, 351–366.

Waryah, C.B., Moses, C., Arooj, M., and Blancafort, P. (2018). Zinc Fingers, TALEs, and CRISPR Systems: A Comparison of Tools for Epigenome Editing. Methods Mol. Biol. *1767*, 19–63.

Xu, G.L., and Bestor, T.H. (1997). Cytosine methylation targetted to pre-determined sequences. Nat. Genet. *17*, 376–378.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discovery *2*, 16009.

Yandim, C., Natisvili, T., and Festenstein, R. (2013). Gene regulation and epigenetics in Friedreich's ataxia. J. Neurochem. *126 Suppl 1*, 21–42.

Yang, S., Chang, R., Yang, H., Zhao, T., Hong, Y., Kong, H.E., Sun, X., Qin, Z., Jin, P., Li, S., et al. CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. J Clin Invest *127*, 2719–2724.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell *26*, 577–590.

Yazawa, M., Sadaghiani, A.M., Hsueh, B., and Dolmetsch, R.E. (2009). Induction of protein-protein interactions in live cells using light. Nature Biotechnology *27*, 941–945.

Yeakley, J.M., Morfin, J.P., Rosenfeld, M.G., and Fu, X.D. (1996). A complex of nuclear proteins mediates SR protein binding to a purine-rich splicing enhancer. PNAS *93*, 7582–7587.

You, S.-H., Lim, H.-W., Sun, Z., Broache, M., Won, K.-J., and Lazar, M.A. (2013). Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 *in vivo*. Nature Structural & Molecular Biology *20*, 182–187.

Zetsche, B., Volz, S.E., and Zhang, F. (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. Nat. Biotechnol. *33*, 139–142.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nature Biotechnology *29*, 149–153.

Zhang, F., Bodycombe, N.E., Haskell, K.M., Sun, Y.L., Wang, E.T., Morris, C.A., Jones, L.H., Wood, L.D., and Pletcher, M.T. (2017). A flow cytometry-based screen identifies MBNL1 modulators that rescue splicing defects in myotonic dystrophy type I. Hum Mol Genet *26*, 3056–3068.

Zhang, L., Spratt, S.K., Liu, Q., Johnstone, B., Qi, H., Raschke, E.E., Jamieson, A.C., Rebar, E.J., Wolffe, A.P., and Case, C.C. (2000). Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. J. Biol. Chem. *275*, 33850–33860.

Zhao, X.-N., and Usdin, K. (2015). The Repeat Expansion Diseases: The dark side of DNA repair. DNA Repair *32*, 96–105.

Zharkov, D.O. (2008). Base excision DNA repair. Cell. Mol. Life Sci. *65*, 1544–1565.

Zhou, X., Naik, S., Dakhova, O., Dotti, G., Heslop, H.E., and Brenner, M.K. (2016). Serial Activation of the Inducible Caspase 9 Safety Switch After Human Stem Cell Transplantation. Mol Ther *24*, 823–831.

Appendix A: Articles

nature COMMUNICATIONS

ARTICLE

Received 15 Jun 2016 | Accepted 12 Sep 2016 | Published 9 Nov 2016

DOI: 10.1038/ncomms13272 OPEN

Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase

Cinzia Cinesi[1], Lorène Aeschbach[1], Bin Yang[1] & Vincent Dion[1]

CAG/CTG repeat expansions cause over 13 neurological diseases that remain without a cure. Because longer tracts cause more severe phenotypes, contracting them may provide a therapeutic avenue. No currently known agent can specifically generate contractions. Using a GFP-based chromosomal reporter that monitors expansions and contractions in the same cell population, here we find that inducing double-strand breaks within the repeat tract causes instability in both directions. In contrast, the CRISPR-Cas9 D10A nickase induces mainly contractions independently of single-strand break repair. Nickase-induced contractions depend on the DNA damage response kinase ATM, whereas ATR inhibition increases both expansions and contractions in a MSH2- and XPA-dependent manner. We propose that DNA gaps lead to contractions and that the type of DNA damage present within the repeat tract dictates the levels and the direction of CAG repeat instability. Our study paves the way towards deliberate induction of CAG/CTG repeat contractions *in vivo*.

[1] Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. Correspondence and requests for materials should be addressed to V.D. (email: Vincent.dion@unil.ch).

Uncovering the Interplay Between Epigenome Editing Efficiency and Sequence Context Using a Novel Inducible Targeting System

Bin Yang[1], Alicia Borgeaud[1,2], Lorène Aeschbach[1], and Vincent Dion[1*]

1: University of Lausanne, Faculty of Biology and Medicine, Center for Integrative Genomics, Bâtiment Génopode, 1015-Lausanne, Switzerland.
2: current address: MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK
*: Corresponding author: Vincent.dion@unil.ch

Keywords: Epigenome editing, synthetic biology, chromatin, gene expression, histone deacetylases, expanded CAG/CTG repeats.

Abstract

Epigenome editing is an attractive way to manipulate gene expression. However, editing efficiencies depend on the DNA sequence context in a manner that remains poorly understood. Here we developed a novel system in which any protein can be recruited at will to a GFP reporter. We named it ParB/ANCHOR-mediated Inducible Targeting (PInT). Using PInT, we tested how CAG/CTG repeat size affects the ability of histone deacetylases to modulate gene expression. We found that repeat expansion reduces the effectiveness of silencing brought about by HDAC5 targeting. This repeat-length specificity was abolished when we inhibited HDAC3 activity. Our data guide the use of these histone deacetylases in manipulating chromatin. PInT can be adapted to study the effect of virtually any sequence on epigenome editing.

Key points:

- PInT: a novel assay to test the effect of DNA sequence on epigenome editing.
- HDAC5-mediated silencing is more efficient at short compared to expanded CAG/CTG repeats.
- HDAC3 activity is responsible for the allele-size effect of HDAC5-mediated gene silencing.

Introduction:

Chromatin structure impinges on every DNA-based transaction, from replication and DNA repair to transcription. Thus, it is not surprising that epigenome editing is being harnessed both to understand basic molecular mechanisms and to treat disease (1). Epigenome editing is now most commonly carried out via the use of the domain of a chromatin modifier, or EpiEffector, fused to a catalytically dead Cas9 (dCas9). The fusion protein is targeted to a locus of choice by way of a customizable single guide RNA (sgRNA) (2-10). Examples of dCas9-mediated epigenome editing include altering chromatin states by either targeting Krüppel-associated box (KRAB) (6) or the histone acetyltransferase domain of p300 (2), thereby reducing or promoting enhancer function, respectively. Moreover, epigenome editing using Cas9-based approaches have been used to modify disease phenotypes in cells and *in vivo* (11,12).

It is currently not possible to predict whether targeting a specific locus with a particular dCas9-EpiEffector fusion will result in efficient chromatin modification and alteration of gene expression. Several reasons have been proposed to account for this, ranging from the sequence of the sgRNA and the distance of its target from the transcriptional start site (3-5), the chromatin structure already present at the target locus (13-18)(BioRxiv: https://doi.org/10.1101/228601) , and/or the exact EpiEffector used (2,4,10,16,18). Indeed, the same EpiEffector targeted at different loci can have very different effects (10,16), highlighting that DNA context affects EpiEffectors in ways that are not understood.

Some DNA sequences can have profound effects on nucleosome positioning and chromatin structure (19). A prime example of this is the expansion of CAG/CTG repeats, which causes 14 different neurological and neuromuscular diseases (20,21). In healthy individuals, these sequences have less than 35 repeats at any one disease locus. However, they can expand and reach up to thousands of triplets. Once expanded, CAG/CTG repeats lead to changes in gene expression in their vicinity and to a heterochromatic-like state (22-25). How these repetitive sequences might affect epigenome editing is unknown.

Here, we developed a method to understand how DNA sequence context can influence epigenome editing efficiency. We named the system ParB/ANCHOR-mediated induced targeting (PInT). With PInT, any protein of interest can be targeted near a sequence of choice. ParB, a bacterial protein, forms oligomers once it nucleates at its non-repetitive binding site, INT (26). Fusing a protein of interest to ParB leads to the recruitment of many of the desired molecules to the INT locus. The targeting is inducible as we coupled ParB/ANCHOR to a chemically induced proximity (CIP) system derived from plants (27). The target sequence is embedded in a GFP mini gene (28) such that the effect of targeting on gene expression is easily monitored. Using PInT, we uncovered an unexpected effect of expanded CAG/CTG repeats on the effectiveness of histone deacetylase 5 (HDAC5) to modulate gene expression and found that this was due to the catalytic activity of HDAC3.

Materials and Methods:

Cell culture conditions and cell line construction

The majority of the cell lines used, including all the parental lines, were genotyped by Microsynth, AG (Switzerland) and found to be HEK293.2sus. They were free of mycoplasma as assayed by the Mycoplasma check service of GATC Biotech. The cells were maintained in DMEM containing 10% FBS, penicillin, and streptomycin, as well as the appropriate selection markers at the following concentrations: 15 µg ml^{-1} blasticidin, 1µg ml^{-1} puromycin, 150µg ml^{-1} hygromycin, 400 µg ml^{-1} G418, and/or 400 µg ml^{-1} zeocin. The incubators were set at 37 °C with 5% CO_2. Whereas FBS was used to maintain the cells, dialyzed calf serum was used at the same concentration for all the experiments presented here. The ABA concentration used was 500 µM, unless otherwise indicated. Doxycycline (dox) was used at a concentration of 2 µg ml^{-1} in all experiments.

A schematic of cell line construction and pedigree is found in Figure S1, and the lines are listed in Table S1. This table includes the plasmids made for cell line construction. The plasmids used for transient transfections are found in Table S2. For each cell line, single clones were picked and tested for expression of ParB-ABI and PYL-fusions by western blotting using the protocol described before (29). Briefly, whole cell extracts were obtained, and their protein content was quantified using the Pierce BCA Protein Assay Kit (ThermoScientific). Proteins were then run onto Tris-glycine 10% SDS PAGE gels before being transferred onto nitrocellulose membrane (Axonlab). The membranes were blocked using the Blocking Buffer for Fluorescent Western Blotting (Rockland), and primary antibodies were added overnight. Membranes were then washed followed by the addition of the secondary antibody (diluted 1 to 2000). The fluorescent signal was detected using an Odyssey Imaging System (Li-CoR). All antibodies used are found in Table S3. To assess repeat sizes, we amplified the repeat tracts using oVIN-0459 and oVIN-0460 with the UNG and dUTP-containing PCR as described(30) and then Sanger-sequenced by Microsynth AG (Switzerland). The sequences of all the primers used in this study are found in Table S4.

The ParB-INT sequence system used here is the c2 version described previously (26), except that the ParB protein was codon-optimized for expression in human cells. It is also called ANCHOR1 and is distributed by NeoVirTech. ParB-ABI (pBY-008), PYL (pAB-

NEO-PYL), PYL-HDAC5 (pAB(EXPR-PYL-HDAC5-NEO)) and PYL-HDAC3 (pAB(EXPR-PYL-HDAC3-NEO)) constructs were randomly inserted and single clones were then isolated (Table S1). GFP-reporter cassettes were inserted using Flp-mediated recombination according to the manufacturer's instruction (Thermo Scientific). Single colonies were picked and screened for zeocin sensitivity to ensure that the insertion site was correct.

Targeting assays

For targeting assays involving transient transfections, cells were plated onto poly-D-lysine-coated 12-well plates at a density of $6x10^5$ cells per well and transfected using 1 µg of DNA per well and Lipofectamine 2000 or Lipofectamine 3000 (Thermofisher Scientific). 6 hours after transfection, the medium was replaced with one containing dox and ABA or DMSO. 48h after the transfection, the cells were split, and fresh medium with dox and ABA or DMSO was replenished. On the fifth day, samples were detached from the plate with PBS + 1 mM EDTA for flow cytometry analysis.

In the case of the stable cell lines, cells were seeded at a density of $4x10^5$ per well in 12-well plates. The media included dox and ABA or DMSO. The medium was changed 48 hours later and left to grow for another 48 hours. The cells were then resuspended in 500µl PBS + 1 mM EDTA for flow cytometry analysis.

Flow cytometry and analysis

We used an Accuri C6 flow cytometer from BD and measured the fluorescence in at least 12 500 cells for each treatment. The raw data was exported as FCS files and analyzed using FlowJo version 10.0.8r1.

Chromatin immunoprecipitation

For chromatin immunoprecipitation, the cells were treated as for the targeting experiments except that we used 10 cm dishes and $4x10^6$ cells. After 96 hours of incubation, paraformaldehyde was added to the medium to a final concentration of 1% and the cells were incubated for 10 minutes at room temperature. The samples were then quenched with 0.125 M PBS-glycine for 5 minutes at room temperature. Samples were then centrifuged, the supernatant was discarded, and the cell pellets were washed with ice-cold PBS twice. The samples were split into 10^7 cell aliquots and either used immediately or stored -75 °C for later use. Sonication was done using a Bioruptor for 25 to 30 min. DNA shearing was visualized by agarose gel electrophoresis after crosslink reversal and RNase treatment. 20% of sonicated supernatant was used per IP, with 3 µg

anti-FLAG (M2, Sigma), anti-PAN acetylated H3 (Merck), or anti-IgG (3E8, Santa Cruz Biotechnology) on Protein G Sepharose 4 Fast Flow beads (GE healthcare). The samples were incubated at 4°C overnight and then washed with progressively more stringent conditions. After the IP, the samples were de-crosslinked and purified using a QIAquick PCR purification kit (Qiagen) and analyzed using a qPCR.

Quantitative PCR

Quantitative PCR was performed with the FastStart Universal SYBR Green Master Mix (Roche) using a 7900HT Fast Real-Time PCR System in a 384-Well Block Module (Applied Biosystems™). Primers used to detect enrichment at the INT sequence and at *ACTA1* gene are listed in Table S4. Ct values were analyzed using the SDS Software v2.4. The percentage of input reported was obtained by dividing the amount of precipitated DNA for the locus of interest by the amount in the input samples multiplied by 100%.

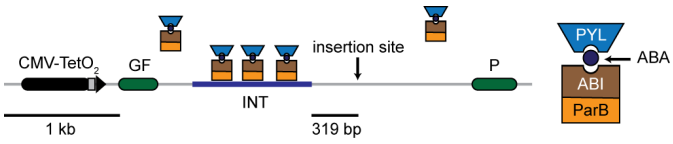
Statistics

We determined statistical significance in the targeting experiments using a two-tailed paired Student's t-test because the samples treated with DMSO and ABA were from the same original population and treated side-by-side. For the ChIP samples, the test used was a two-tailed Student's t-test. All the statistical analyses were done using R studio version 3.4.0. We concluded that there was a significant difference when $P < 0.05$.

Results:

ParB/ANCHOR-mediated induced targeting (PInT)

We designed PInT (Fig. 1) to be modular and highly controllable. It contains a GFP mini gene that harbours two GFP exons flanking the intron of the mouse *Pem1* gene (28,29). A doxycycline-inducible TetOn promoter drives the expression of the reporter. This cassette is always inserted at the same genomic location as a single copy integrant in T-Rex Flp-In HEK293 cells. Inside the intron, we inserted a 1029 bp non-repetitive sequence, INT, that contains four binding sites for dimers of the *Burkholderia cenocepacia* ParB protein. Once bound to INT, ParB oligomerizes in a sequence-independent manner, recruiting up to 200 ParB molecules (31). This ParB/ANCHOR system was first used in live yeast cells to visualize double-strand break repair (26). More recently, it has been used to monitor the mobility of a genomic locus upon activation of transcription (32) and to visualize viral replication (33) in live mammalian cells. We made the system inducible by fusing ParB to a plant protein called ABSCISIC ACID INSENSITIVE (ABI), which dimerizes with PYRABACTIN RESISTANCE1-LIKE (PYL) upon addition of abscisic acid (ABA) to the culture medium (27). ABA is a plant hormone that is not toxic to human cells, making this CIP system especially convenient. Within 319bp of the INT sequence, there is a cloning site that can be used to insert any DNA motif. Thus, fusing any protein of interest to PYL allows for full temporal control over the recruitment of a protein of interest near a DNA sequence of choice.



Yang et al - Fig. 1

***Fig. 1:* Schematic of PInT**. *The GFP reporter is driven by an inducible Tet-On promoter. It contains an intron harbouring an INT sequence, which mediates the recruitment and oligomerization of ParB. We fused ParB to ABI, a plant protein that binds PYL only in the presence of abscisic acid (ABA). Fusing PYL to any protein of interest leads to its inducible recruitment 319bp away from a cloning site that can be used to insert a sequence of choice. The PYL construct contains a 3xFLAG tag whereas the ParB-ABI fusion includes 3xHA. They both contain SV40 nuclear localization signals.*

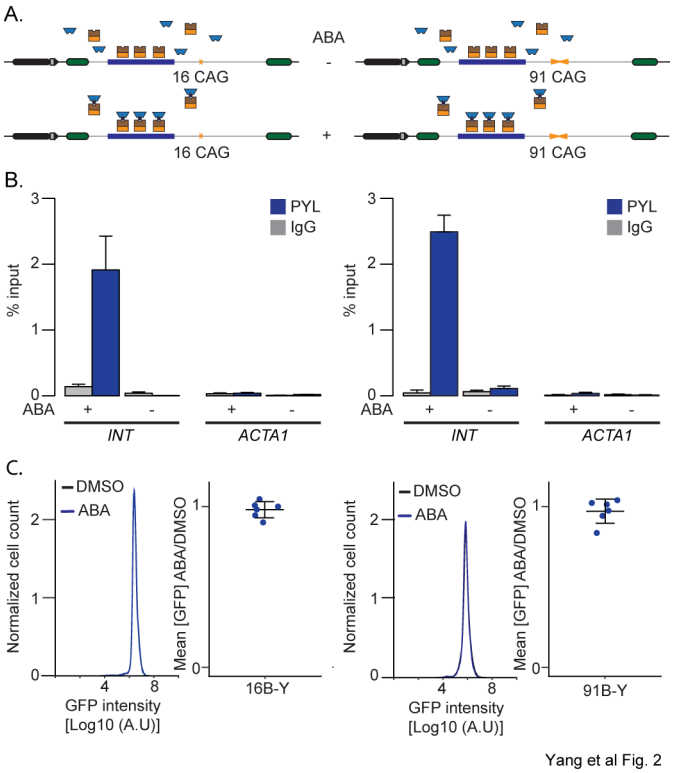
It was important to determine whether the components of PInT affect the expression of the GFP reporter. We first tested whether ABA changed GFP expression in

GFP$(CAG)_0$ cells (28). These cells carry the GFP mini gene without the INT sequence or any additional sequences in the intron (see Table S1 and Fig. S1 for details on cell line construction). We found that treatment with up to at least 500 μM of ABA, which induces the dimerization between PYL and ABI, did not affect GFP expression (Fig. S2AB). We also transiently transfected GFP$(CAG)_0$ cells with plasmids expressing the ParB-ABI fusion. This had no detectable effect on the behaviour of the reporter (Fig. S2C). We next inserted the INT sequence inside the *Pem1* intron and integrated this construct using site-directed recombination, generating GFP-INT cells. These cells contain INT but no additional sequence within the intron. They do not express ParB-ABI. We found that the insertion of the INT sequence had little, if any, discernable effect on the GFP expression (Fig. S2D). we conclude that by themselves, the individual components of PInT do not significantly interfere with GFP expression.

We then stably integrated both the GFP-INT reporter and the ParB-ABI fusion to generate GFP-INT-B cells. We found a decrease in GFP expression that correlated with high levels of ParB-ABI (Fig. S2EFG), suggesting that the binding of ParB-ABI has a predictable effect on the GFP reporter. To avoid any complication, we integrated ParB-ABI early in the construction pipeline such that all the cell lines presented here contain the same amount of ParB-ABI (Fig. S1).

Next, we determined the efficiency of targeting PYL to the INT sequence and the consequences on GFP expression. We used nB-Y cells, which contain the GFP mini gene with the INT sequence, stably express both ParB-ABI (B) and PYL (Y), and contain n CAG repeats, in this case either 16, which is in the normal range, or an expanded repeat of 91 triplets (Fig. 2A, Fig. S3A). We found, using chromatin immunoprecipitation followed by qPCR (ChIP-qPCR), that only 0.1% of the input DNA could be precipitated when we treated the cells with the vehicle, DMSO, alone. By contrast, the addition of ABA to the cell media increased the association of PYL to the INT locus significantly, reaching 1.9 and 2.5% of the input chromatin pulled down in 16B-Y or 91B-Y cells, respectively (Fig. 2B). These results demonstrate the inducible nature of the system and show that the presence of the expansion does not interfere with the targeting efficiency. Importantly, PYL targeting had no effect on GFP expression as measured by flow cytometry (Fig. 2C). We conclude that PInT

works as an inducible targeting system and that PYL targeting is efficient and does not further affect gene expression.



Yang et al Fig. 2

Fig. 2: Inducible targeting of PYL at the GFP reporter*. A) Schematic representation of 16B-Y (left) and 91B-Y (right) cell lines. B) ChIP-qPCR using antibodies against FLAG to pull down PYL at INT and ACTA1 in 16B-Y cells (left, N=4) and 91B-Y cells (right, N=4). The error bars represent the standard error. C) Representative flow cytometry profiles as well as quantification of the GFP expression in 16B-Y (left, N=6) and 91B-Y (right, N=6) cells. The error bars are the standard deviation around the mean.*

HDAC5 silencing depends on CAG/CTG repeat size

We next sought to test whether we could manipulate GFP expression using HDAC5. This class IIa deacetylase impacts gene silencing and heterochromatin maintenance (34,35) as well as cell proliferation (35,36). The PYL-HDAC5 fusion was functional since GFP-INT cells transiently expressing this fusion had slightly lower GFP expression than those expressing PYL alone (Fig. S4A). We created isogenic nB-Y-HDAC5 cells that express stably a PYL-HDAC5 fusion and have 16 or 59 CAG repeats within the GFP reporter (Fig. 3A). We found that adding ABA to these cells led to an increase in pull-down efficiency of PYL-HDAC5 at the INT locus from 0.06% to 2.2% in 16B-Y-HDAC5 cells and from 0.1% to 3% of input in the presence of 59 repeats (Fig. 3B). This was accompanied by a significant 2-fold decrease in GFP expression in 59B-Y-HDAC5 cells, whereas the decrease was of 3 folds in 16B-Y-HDAC5 cells (Fig. 3C, P=0.001 and P= 0.0015 using a paired Student's t-test comparing conditions with ABA to those with DMSO only in 16B-Y-HDAC5 and 59B-Y-HDAC5 cells, respectively). Remarkably, the decrease in expression was significantly lower in the context of an expanded repeat (Fig. 3C, P=0.0001 comparing the decrease in expression upon ABA addition between the 16B-Y-HDAC5 and 59B-Y-HDAC5 using a Student's t-test).

Targeting efficiency of PYL-HDAC5 does not account for the repeat size-dependent effect since it was slightly higher in 59B-Y-HDAC5 than in 16B-Y-HDAC5 cells (Fig. 3B). To determine whether the effect is due to targeting at the INT locus, we transiently expressed PYL-HDAC5 in GFP$(CAG)_0$B cells, which have no INT in their GFP reporter but express ParB-ABI. Adding ABA to these cells did not affect GFP expression (Fig. S4BC), suggesting that the presence of the INT sequence is essential. Moreover, PYL-HDAC5 targeting reduced the levels of acetylated histone H3 (acH3) (P=0.0001 and P=0.024 comparing DMSO treated and ABA-treated 16B-Y-HDAC5 and 59B-Y-HDAC5, respectively, using a Student's t-test), as measured by ChIP-qPCR. The decrease in acH3 upon targeting was greater in 16B-Y-HDAC5 than in 16B-Y cells (P=0.006 using a Student's t-test), consistent with a role for HDAC5 in silencing gene expression. Interestingly, the acH3 levels at the INT sequence were similar between 16B-Y and 91B-Y and between 16B-Y-HDAC5 and 59B-Y-HDAC5 (Fig. 3DE, P=0.95, and P=0.25, respectively using a Student's t-test), suggesting that the acH3 levels are unaffected by the expansion. We conclude that PYL-HDAC5 targeting silences better the lines with the shorter repeats.

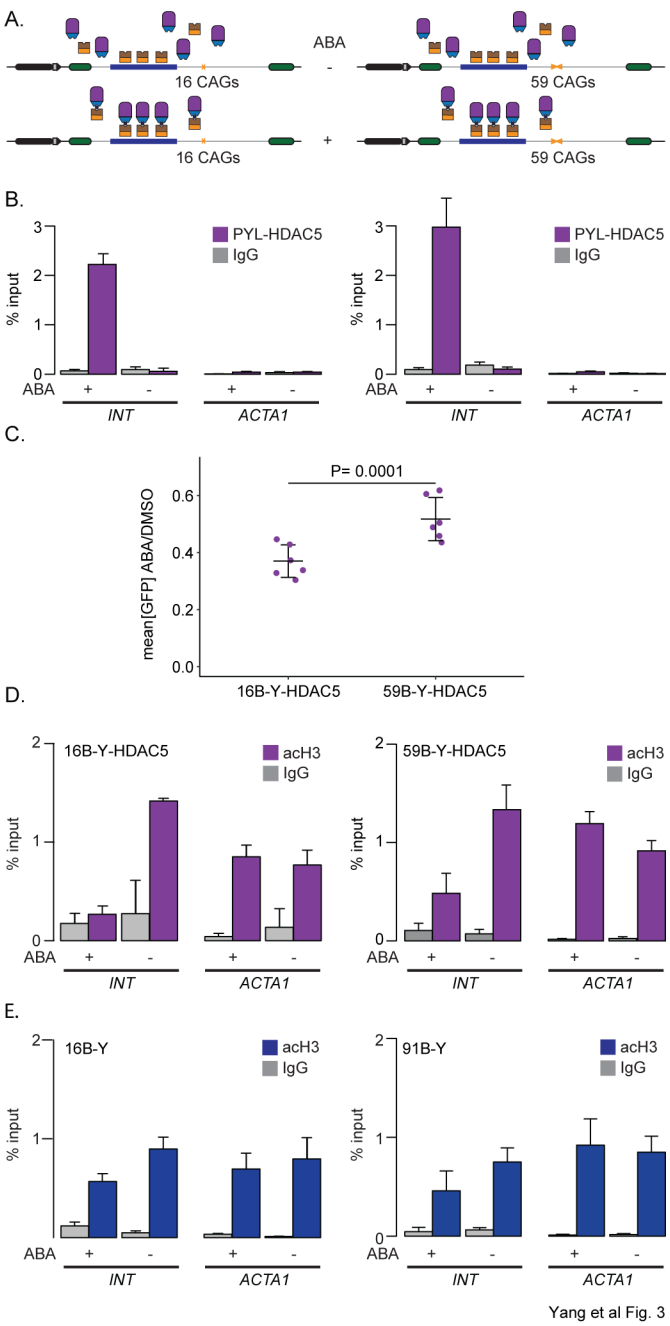
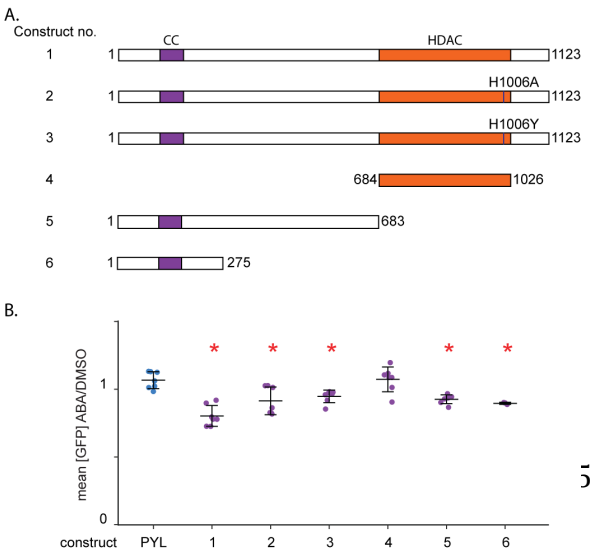


Fig. 3: PYL-HDAC5 targeting induces silencing in a repeat-length dependent manner*. A) Schematic representation of 16B-Y-HDAC5 (left) and 59B-Y-HDAC5 (right) cells. B) ChIP-qPCR*

using antibodies against FLAG to pull down PYL-HDAC5 at INT and ACTA1 in 16B-Y-HDAC5 cells (left, N=4) and 59B-Y-HDAC5 cells (right, N=4). The error bars represent the standard error. C) Quantification of GFP expression upon incubation with ABA or DMSO in 16B-Y-HDAC5 (left, N=6) and 59B-Y-HDAC5 (right, N=6) cells. The error bars show the standard deviation around the mean. D) ChIP-qPCR data using a pan-acetylated H3 antibody to pull down the INT and ACTA1 loci in 16B-Y-HDAC5 (left, N=4) and 59B-Y-HDAC5 (right, N=4) cells. The error bars represent the standard error. E) ChIP-qPCR data using a pan-acetylated H3 antibody to pull down the INT and ACTA1 loci in 16B-Y (left, N=4) and 91B-Y (right, N=4) cells. The error bars represent the standard error.

The N-terminal domain of HDAC5 mediates silencing

Class I HDACs derive their catalytic activity *in vitro* from a conserved tyrosine residue that helps coordinate a zinc ion essential for catalysis (37). By contrast, class IIa enzymes, like HDAC5, have a histidine instead of tyrosine at the analogous site, which considerably lowers HDAC activity (37). In fact, the H1006Y mutant had more than 30-fold increase in its HDAC activity compared to the wild type enzyme (37). To determine whether the catalytic activity of HDAC5 potentiates the decrease in GFP expression upon targeting, we compared the silencing activity of wild-type PYL-HDAC5, the H1006A loss-of-function mutant, and the H1006Y gain-of-function mutant by transient transfection in 40B cells, which contain the GFP-INT reporter with 40 CAGs and express ParB-ABI (Fig. 4A). The effect on silencing seen upon targeting of the wild-type PYL-HDAC5 fusion was lower when delivered by transient transfection compared to the stable cell lines. Nevertheless, under these conditions, targeting PYL-HDAC5-H1006A or PYL-HDAC5-H1006Y could both silence the transgene compared to targeting PYL alone (Fig 4B; P= 0.01 and 0.0008, respectively, using a Student's t-test), suggesting that tampering with the catalytic activity of HDAC5 does not influence silencing of our GFP reporter. Moreover, targeting PYL fused to the catalytic domain of HDAC5 did not shift GFP expression (Fig. 4B). Indeed, we find that the silencing activity was contained within the N-terminal part of HDAC5, which characterizes class IIa enzymes. Further truncations (Fig. 4AB) are consistent with a model by which the coiled-coil domain of HDAC5, which is responsible for homo and



heterodimerization of class IIa enzymes *in vitro (38)*, contains the silencing activity.

***Fig. 4:* The silencing activity of HDAC5 is contained in its N-terminal domain**. *A) Mutants and truncations of HDAC5 fused to PYL. The coiled-coil (CC) domain is indicated in purple, the deacetylase domain (HDAC) in orange. B) Ratio of the mean GFP intensities in 40B cells between ABA and DMSO only treatment of 40B cells transiently transfected with plasmids containing the constructs shown in A. Construct 1: N=7, P=0.0002 versus PYL; construct 2: N=7, P=0.01 versus PYL; construct 3: N=7, P=0.0008 versus PYL; construct 4: N=7, P=0.88 versus PYL; construct 5: N=7, P=0.0012 versus PYL; construct 6: N=3, P=0.0003 versus PYL. In all cases, we used a Student's t-test to calculate the P-values. The error bars show the standard deviation around the mean. *: P≤0.01 compared to PYL targeting.*

PYL-HDAC3 targeting enhances GFP expression independently of its catalytic activity

HDAC5 is thought to mediate histone deacetylation by recruiting other HDACs, including HDAC3 (39). Therefore, we hypothesized that PYL-HDAC3 targeting should have the same effect on GFP expression as PYL-HDAC5 targeting. To address this directly, we made a PYL-HDAC3 fusion and overexpressed it in 40B cells without targeting (Fig. S4D). We found that there was a slight decrease in GFP expression, suggesting that the construct could silence gene expression. Next, we generated nB-Y-HDAC3 cells and compared GFP intensities with and without ABA. Contrary to our initial hypothesis, we found that targeting PYL-HDAC3 in both 16B-Y-HDAC3 and 89B-Y-HDAC3 increased GFP expression by 1.5 fold (Fig. S5AB, P=0.0004 and P=0.001 using paired Student's t-tests comparing ABA and DMSO treatments in 16B-Y-HDAC3 and 89B-Y-HDAC3, respectively). The effect appeared direct since adding ABA to GFP$(CAG)_0$B cells transiently expressing PYL-HDAC3 did not affect GFP expression (Fig. S4E). The increase in GFP expression in nB-Y-HDAC3 cells was accompanied by an efficient targeting of the PYL-HDAC3 fusion (Fig. S5C) and an increase in acH3 levels (Fig. S5D). However, treatment with the HDAC3-specific small molecule inhibitor RGFP966 (40) did not affect the increase in GFP expression in neither 16B-Y-HDAC3 nor 89B-Y-HDAC3 cells (Fig. S5E). We conclude that targeting PYL-HDAC3 increases GFP expression independently of its HDAC activity, consistent with the observation that HDAC3 has an essential role during development that does not involve its HDAC activity (41).

HDAC3 activity is required for the repeat size-specificity upon HDAC5-mediated silencing

Although HDAC3 targeting did not have the expected effect on GFP expression, evidence shows that its catalytic activity is implicated in HDAC5-mediated silencing (39). To determine the potential catalytic role of HDAC3 in this context, we targeted PYL (Fig. 5A) or PYL-HDAC5 (Fig. 5B) to our GFP reporter in nB-Y and nB-Y-HDAC5 cells while cultivating the cells in the presence of RGFP966. We find that although this treatment had no effect on PYL targeting (Fig. 5A), it abolished the allele-length specificity of PYL-HDAC5 targeting, leading to a silencing efficiency of 2.4 and 2.5 folds for 16B-Y-HDAC5 and 59B-Y-HDAC5, respectively (Fig. 5B, P= 0.77 using a Student's t-test). This is in contrast to the RGFP966-free conditions where targeting PYL-HDAC5 silenced better the normal-sized allele (Fig. 3). These results suggest that HDAC3 mediates the CAG repeat size-dependency upon PYL-HDAC5 targeting.

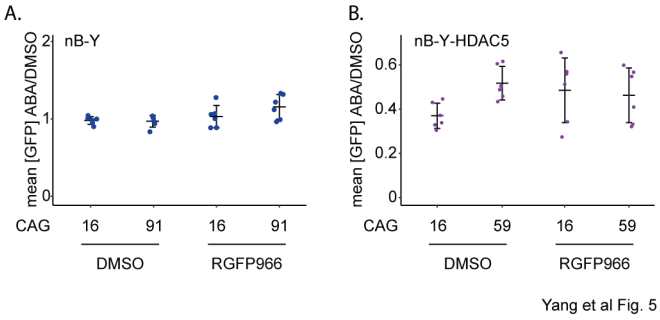


Fig. 5: HDAC3 activity is required for the allele-specificity upon HDAC5-mediated silencing*. A) Quantification of GFP intensity upon targeting in the presence of RGFP966 or the vehicle, DMSO, in 16B-Y (N=6) and 91B-Y cells (N=6). Note that the data for the DMSO-treated cells is the same as in Fig. 2C. The addition of DMSO did not affect GFP expression, and we therefore pooled the data. B) Quantification of GFP intensity upon targeting in the presence of RGFP966 or the vehicle, DMSO, in 16B-Y-HDAC5 (N=6) and 59B-Y-HDAC5 cells (N=6). Note that the data for the DMSO-treated cells is the same as in Fig. 3C. The addition of DMSO did not affect GFP expression, and we therefore pooled the data. The error bars show the standard deviation around the mean.*

Discussion:

We presented here a novel assay to investigate the effect of a DNA sequence of interest on the efficiency of a chosen EpiEffector in altering gene expression. As an example of how the DNA context may affect the activity of an EpiEffector, we showed that expanded CAG/CTG repeats decrease the silencing efficiency of HDAC5. Moreover, we determined that this allele-length specificity depends on HDAC3 activity, highlighting the potential of PInT in uncovering unique mechanistic insights. These data provide evidence that local DNA sequence context is an important determinant of epigenome editing, independently of the efficiency or mode of targeting.

PInT could be used for many different applications. First, the intron can host sequences beyond CAG/CTG repeats. Indeed, the GFP mini gene we used here, without the targeting components, was recently used to monitor the effect of a RNA polymerase III gene on RNA polymerase II-mediated transcription (42). Second, it is often difficult to differentiate between a chromatin modifier changing gene expression because of a local effect on chromatin structure or indirectly through changes in the transcriptome. PInT allows making that distinction thanks to its inducible nature. Indeed, we found that overexpressing PYL-HDAC5 had a small effect on gene expression at the GFP reporter and that targeting it further decreased expression. We could conclude that PYL-HDAC5 can act locally to silence the transgene. This is useful in dissecting the mechanisms of action of EpiEffectors. Third, we demonstrated, using mutants and truncations of HDAC5, that we can quickly screen for protein domains and mutants that are most effective in modulating gene expression. Thus, PInT could be used to design peptides with sufficient activity to be useful in downstream epigenomic editing applications, for example when using dCas9 fusions *in vivo*. A current limitation of the *S. pyogenes* Cas9 for *in vivo* applications is its large size, which is at the limit of what adeno-associated viral vectors can accommodate (43). Even with the smaller orthologues, fitting a dCas9 fusion inside a gene delivery vector is a challenge. Therefore, being able to trim an EpiEffector down to a small domain may help optimizing downstream applications and translation.

The observation that HDAC5 targeting has a differential effect on gene expression depending on the size of the repeat tract is surprising. Our data suggest that the deacetylase activity of HDAC3 is required for this effect. Importantly, we cannot currently rule out that RGFP966 may inhibit other HDACs that would be responsible for this effect. Nevertheless, this small molecule is highly selective for HDAC3 (40), making this HDAC the most likely candidate for driving allele-specific silencing. HDAC3 could be setting up an asymmetry between the two size alleles in several ways. For instance, it could deacetylate histones (those residues not recognized by the pan-acetylated histone H3 antibody that we used) or non-histone proteins in the vicinity of the expanded CAG/CTG repeat prior to HDAC5 targeting. More work is required to understand further the mechanism of the repeat length-specific silencing.

Several studies have suggested that the ectopic insertion of an expanded CAG/CTG repeat in mice could induce changes in chromatin structure in the abutting sequences. An early example was the random insertion of arrays of transgenes, each carrying 192 CAGs (44), which led to the silencing of the transgenes independently of the site of genomic integration. In addition, inserting a 40 kb human genomic region containing the *DMPK* gene along with an expansion of 600 CTGs (45), or a 13.5Kb region containing the human SCA7 gene with 92 CAGS (46) all led to changes in chromatin marks near the expansion. It has been unclear, however, whether the presence of endogenous sequence elements, like CpG islands (47) and CTCF binding sites (25,48), is necessary for this effect. Our data show that 91 CAGs, without the flanking sequences normally present at the *DMPK* gene from whence this repeat was cloned (49), does not lead to significant changes in the levels of acetylated histone H3 in its vicinity. These data suggest that the flanking sequence elements play important roles in the induction and/or maintenance of heterochromatic marks surrounding expanded CAG/CTG repeats.

Recently, a number of studies have proposed that silencing the expanded repeat allele without affecting the expression of the normally sized allele may lead to a novel therapeutic approach for expanded CAG/CTG repeats (50-52). However, only one factor, which is essential for mouse development (52), has been identified so far. We speculate that PInT may be adapted to screen for allele length-specific silencers,

which could help uncover novel therapeutic options for expanded CAG/CTG repeat disorders.

Data availability statement:

The datasets, cell lines, and plasmids generated and analysed during the current study are available from the corresponding author. Note that to obtain some of the plasmids you will also need the permission of NeoVirTech, which owns the rights to the ANCHOR technology.

Funding:
This work was funded by SNSF professorship grants #144789 and #172936 to V.D.

Acknowledgements:

We thank John H. Wilson and Kerstin Bystricky for sharing reagents, Oscar Rodriguez Lima for cloning pBY-050, as well as Fisun Hamaratoglu, Helder C. Ferreira, Ana C. Marques, Johanna E. Martin, and Nastassia Gobet for critical reading of the manuscript.

Author contributions:

BY performed the experiments except for those presented in Fig. 4, which were done by ACB. ACB and LA helped BY in generating the cell lines. BY, ACB, and VD designed the experiments. BY and VD wrote the paper and prepared the figures.

References

1. Kungulovski, G. and Jeltsch, A. (2016) Epigenome Editing: State of the Art, Concepts, and Perspectives. *Trends Genet*, **32**, 101-113.
2. Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E. and Gersbach, C.A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*, **33**, 510-517.
3. Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C. *et al.* (2014) Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, **159**, 647-661.
4. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442-451.
5. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173-1183.
6. Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E. and Gersbach, C.A. (2015) Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*, **12**, 1143-1149.
7. Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A. and Jaenisch, R. (2016) Editing DNA Methylation in the Mammalian Genome. *Cell*, **167**, 233-247 e217.
8. Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., Klasic, M. and Zoldos, V. (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res*, **44**, 5615-5628.
9. Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M. and Maehr, R. (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods*, **12**, 401-403.
10. O'Geen, H., Ren, C., Nicolet, C.M., Perez, A.A., Halmai, J., Le, V.M., Mackay, J.P., Farnham, P.J. and Segal, D.J. (2017) dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression. *Nucleic Acids Res*, **45**, 9901-9916.
11. Liao, H.K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M.Z., Sui, Y., Yamauchi, T., Sakurai, M., O'Keefe, D.D., Nunez-Delicado, E. *et al.* (2017) In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell*, **171**, 1495-1507 e1415.
12. Liu, X.S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., Hnisz, D., Li, C.H., Yuan, B., Xu, C. *et al.* (2018) Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. *Cell*, **172**, 979-992 e976.
13. Isaac, R.S., Jiang, F., Doudna, J.A., Lim, W.A., Narlikar, G.J. and Almeida, R. (2016) Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife*, **5**.
14. Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R. and Weissman, J.S. (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife*, **5**.
15. Jensen, K.T., Floe, L., Petersen, T.S., Huang, J., Xu, F., Bolund, L., Luo, Y. and Lin, L. (2017) Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett*, **591**, 1892-1901.
16. Kwon, D.Y., Zhao, Y.T., Lamonica, J.M. and Zhou, Z. (2017) Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat Commun*, **8**, 15315.
17. Lee, C.M., Davis, T.H. and Bao, G. (2018) Examination of CRISPR/Cas9 design tools and the effect of target site accessibility on Cas9 activity. *Exp Physiol*, **103**, 456-460.

18. Cano-Rodriguez, D., Gjaltema, R.A., Jilderda, L.J., Jellema, P., Dokter-Fokkens, J., Ruiters, M.H. and Rots, M.G. (2016) Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nat Commun*, **7**, 12284.
19. Struhl, K. and Segal, E. (2013) Determinants of nucleosome positioning. *Nat Struct Mol Biol*, **20**, 267-273.
20. Orr, H.T. and Zoghbi, H.Y. (2007) Trinucleotide repeat disorders. *Annu Rev Neurosci*, **30**, 575-621.
21. Lopez Castel, A., Cleary, J.D. and Pearson, C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol*, **11**, 165-170.
22. Dion, V. and Wilson, J.H. (2009) Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet*, **25**, 288-297.
23. Lopez Castel, A., Nakamori, M., Tome, S., Chitayat, D., Gourdon, G., Thornton, C.A. and Pearson, C.E. (2011) Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum Mol Genet*, **20**, 1-15.
24. Barbe, L., Lanni, S., Lopez-Castel, A., Franck, S., Spits, C., Keymolen, K., Seneca, S., Tome, S., Miron, I., Letourneau, J. *et al.* (2017) CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy. *Am J Hum Genet*, **100**, 488-505.
25. Cho, D.H., Thienes, C.P., Mahoney, S.E., Analau, E., Filippova, G.N. and Tapscott, S.J. (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol Cell*, **20**, 483-489.
26. Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D. and Bystricky, K. (2014) DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. *PLoS Genet*, **10**, e1004187.
27. Liang, F.S., Ho, W.Q. and Crabtree, G.R. (2011) Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci Signal*, **4**, rs2.
28. Santillan, B.A., Moye, C., Mittelman, D. and Wilson, J.H. (2014) GFP-based fluorescence assay for CAG repeat instability in cultured human cells. *PLoS One*, **9**, e113952.
29. Cinesi, C., Aeschbach, L., Yang, B. and Dion, V. (2016) Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat Commun*, **7**, 13272.
30. Aeschbach, L. and Dion, V. (2017) Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications. *Scientific Reports*, **7**, 18026.
31. Khare, D., Ziegelin, G., Lanka, E. and Heinemann, U. (2004) Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nat Struct Mol Biol*, **11**, 656-663.
32. Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H.A., Sellou, H., Politi, A.Z., Ellenberg, J., Gallardo, F. and Bystricky, K. (2017) Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophys J*, **113**, 1383-1394.
33. Mariame, B., Kappler-Gratias, S., Kappler, M., Balor, S., Gallardo, F. and Bystricky, K. (2018) Real-time visualization and quantification of human Cytomegalovirus replication in living cells using the ANCHOR DNA labeling technology. *J Virol*.
34. Lemercier, C., Verdel, A., Galloo, B., Curtet, S., Brocard, M.P. and Khochbin, S. (2000) mHDA1/HDAC5 histone deacetylase interacts with and represses MEF2A transcriptional activity. *J Biol Chem*, **275**, 15594-15599.
35. Peixoto, P., Castronovo, V., Matheus, N., Polese, C., Peulen, O., Gonzalez, A., Boxus, M., Verdin, E., Thiry, M., Dequiedt, F. *et al.* (2012) HDAC5 is required for maintenance of pericentric heterochromatin, and controls cell-cycle progression and survival of human cancer cells. *Cell Death Differ*, **19**, 1239-1252.
36. Huang, Y., Tan, M., Gosink, M., Wang, K.K. and Sun, Y. (2002) Histone deacetylase 5 is not a p53 target gene, but its overexpression inhibits tumor cell growth and induces apoptosis. *Cancer Res*, **62**, 2913-2922.

37. Lahm, A., Paolini, C., Pallaoro, M., Nardi, M.C., Jones, P., Neddermann, P., Sambucini, S., Bottomley, M.J., Lo Surdo, P., Carfi, A. *et al.* (2007) Unraveling the hidden catalytic activity of vertebrate class IIa histone deacetylases. *Proc Natl Acad Sci U S A*, **104**, 17335-17340.
38. Backs, J., Backs, T., Bezprozvannaya, S., McKinsey, T.A. and Olson, E.N. (2008) Histone deacetylase 5 acquires calcium/calmodulin-dependent kinase II responsiveness by oligomerization with histone deacetylase 4. *Mol Cell Biol*, **28**, 3437-3445.
39. Fischle, W., Dequiedt, F., Hendzel, M.J., Guenther, M.G., Lazar, M.A., Voelter, W. and Verdin, E. (2002) Enzymatic activity associated with class II HDACs is dependent on a multiprotein complex containing HDAC3 and SMRT/N-CoR. *Mol Cell*, **9**, 45-57.
40. Malvaez, M., McQuown, S.C., Rogge, G.A., Astarabadi, M., Jacques, V., Carreiro, S., Rusche, J.R. and Wood, M.A. (2013) HDAC3-selective inhibitor enhances extinction of cocaine-seeking behavior in a persistent manner. *Proc Natl Acad Sci U S A*, **110**, 2647-2652.
41. You, S.H., Lim, H.W., Sun, Z., Broache, M., Won, K.J. and Lazar, M.A. (2013) Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. *Nat Struct Mol Biol*, **20**, 182-187.
42. Yeganeh, M., Praz, V., Cousin, P. and Hernandez, N. (2017) Transcriptional interference by RNA polymerase III affects expression of the Polr3e gene. *Genes Dev*, **31**, 413-421.
43. Epstein, B.E. and Schaffer, D.V. (2017) Combining Engineered Nucleases with Adeno-associated Viral Vectors for Therapeutic Gene Editing. *Adv Exp Med Biol*, **1016**, 29-42.
44. Saveliev, A., Everett, C., Sharpe, T., Webster, Z. and Festenstein, R. (2003) DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature*, **422**, 909-913.
45. Brouwer, J.R., Huguet, A., Nicole, A., Munnich, A. and Gourdon, G. (2013) Transcriptionally Repressive Chromatin Remodelling and CpG Methylation in the Presence of Expanded CTG-Repeats at the DM1 Locus. *J Nucleic Acids*, **2013**, 567435.
46. Libby, R.T., Hagerman, K.A., Pineda, V.V., Lau, R., Cho, D.H., Baccam, S.L., Axford, M.M., Cleary, J.D., Moore, J.M., Sopher, B.L. *et al.* (2008) CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genet*, **4**, e1000257.
47. Gourdon, G., Dessen, P., Lia, A.S., Junien, C. and Hofmann-Radvanyi, H. (1997) Intriguing association between disease associated unstable trinucleotide repeat and CpG island. *Ann Genet*, **40**, 73-77.
48. Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V. and Tapscott, S.J. (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat Genet*, **28**, 335-343.
49. Gorbunova, V., Seluanov, A., Dion, V., Sandor, Z., Meservy, J.L. and Wilson, J.H. (2003) Selectable system for monitoring the instability of CTG/CAG triplet repeats in mammalian cells. *Mol Cell Biol*, **23**, 4485-4493.
50. Liu, C.R. and Cheng, T.H. (2015) Allele-selective suppression of mutant genes in polyglutamine diseases. *J Neurogenet*, **29**, 41-49.
51. Liu, C.R., Chang, C.R., Chern, Y., Wang, T.H., Hsieh, W.C., Shen, W.C., Chang, C.Y., Chu, I.C., Deng, N., Cohen, S.N. *et al.* (2012) Spt4 is selectively required for transcription of extended trinucleotide repeats. *Cell*, **148**, 690-701.
52. Cheng, H.M., Chern, Y., Chen, I.H., Liu, C.R., Li, S.H., Chun, S.J., Rigo, F., Bennett, C.F., Deng, N., Feng, Y. *et al.* (2015) Effects on murine behavior and lifespan of selectively decreasing expression of mutant huntingtin allele by supt4h knockdown. *PLoS Genet*, **11**, e1005043.

Appendix B: R script

Install the necessary R package

```
install.packages("ggplot2")
install.packages("limma")
install.packages("tidyr")
install.packages("dplyr")
install.packages("Hmisc")
require(ggplot2)
require(tidyr)
require(dplyr)
require(limma)
require(Hmisc)
```

Flow cytometry histogram

```
# load data
library(readxl)
HEKBYH516 <- read_excel("~/Desktop/Capri/FC/FC example/HEKBYH516.xlsx")
View(HEKBYH516)

# reorganize dataset
b <- gather(HEKBYH516,
        "Sample",
        "Value",
        control, sample)
print(b)

#try to plot with default setting
c <- ggplot(data=b,
        aes(x=log10(Value),

            colour=Sample
        )) +
 geom_density()

# plot the graph with desired parameters
d <- c +
 scale_color_manual(values=c("control"="black", "sample"="darkorchid3"),
                labels=c("DMSO", "ABA"))+
 ##remove margin between 0 and data
 scale_y_continuous(expand = c(0,0),
                breaks = seq(0, 2.5, 0.5),
                limits=c(0, 2.5))+
 scale_x_continuous(expand = c(0,0),
                breaks = seq(0, 10, 2),
                limits=c(0, 10))+
 xlab("GFP intensity")+
 ylab("Density")+
```

```
  theme(panel.grid.major = element_blank(),
        panel.background = element_blank(),
        axis.line.y  = element_line(colour = "black"),
        axis.line.x  = element_line(colour = "black")

  )
print(d)
```

Flow cytometry data analysis in same cell line

```
# Load data
HEBYH5161 <- read_excel("~/Desktop/Capri/HDAC5
FC/H516/HEKBYH5161.xlsx")

# h516 Sample vs control #
h516 <- gather(HEBYH5161,
           "Sample",
           "Value",
           control1,control2,control3,control4,sample1,sample2,sample3,sample4)
# Raw data calculation
sampleh516m<-c(mean(h516$Value[h516$Sample=="sample1"],na.rm = T),
mean(h516$Value[h516$Sample=="sample2"],na.rm = T),
           mean(h516$Value[h516$Sample=="sample3"],na.rm =
T),mean(h516$Value[h516$Sample=="sample4"],na.rm = T))

controlh516m<-c(mean(h516$Value[h516$Sample=="control1"],na.rm = T),
mean(h516$Value[h516$Sample=="control2"],na.rm = T),
            mean(h516$Value[h516$Sample=="control3"],na.rm = T),
mean(h516$Value[h516$Sample=="control4"],na.rm = T))

#calculate p value
h516ttest <- t.test(sampleh516m,controlh516m, paired= TRUE)
h516pval <- formatC(h516ttest$p.value, format = "e", digits = 1)

# ggPlot
# prepare dataframe
h516data <- data.frame(mfi=c(sampleh516m/1e6, controlh516m/1e6),
                 sampleid=rep(c("ABA", "DMSO"), c(length(sampleh516m),
length(controlh516logm))))
print(h516data)

#reorder the group order on the boxplot
h516data$sampleid <- factor(h516data$sampleid,
                      levels = c("DMSO","ABA"), ordered = TRUE)

# Plot
ggplot(data=h516data, aes(x=sampleid, y=mfi)
) +
```

```
# add points
 geom_jitter(data=h516data, size=1, color="black",width = 0.1
 ) +

# Changing axis ticks, y axis name with numbers on y axis
 scale_y_continuous(name = " ",
                    breaks = seq(0, 4, 1),
                    limits=c(0, 4))+

# add mean (if you like it dashed, add 'linetype = "dashed"' within brackets)
 stat_summary(fun.y = mean, geom = "errorbar", aes(ymax = ..y.., ymin = ..y..), width
= 0.2, color="black") +

 # add error bars corresponding to mean+/-sd
 stat_summary(fun.data=mean_sdl, fun.args = list(mult=1), geom="errorbar",
width=0.1, color="black") +

# Add x axis label
 xlab("") +

 # Add title
 ggtitle(" ") +

 # Choose font size
 theme(axis.title.y = element_text(size=20),
      plot.title = element_text(size=20, hjust = 0.5),
      text = element_text(size=20)) +

# Diplay only x and y axes without any backgroud
 theme(axis.line = element_line(colour = "black"),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      panel.border = element_blank(),
      panel.background = element_blank())
```

Flow cytometry data analysis between different repeat length lines

```
library(readxl)

# Load data
HEBYH5161 <- read_excel("~/Desktop/Capri/HDAC5
FC/H516/HEKBYH5161.xlsx")
HEKBYH559 <- read_excel("~/Desktop/Capri/HDAC5 FC/H559/HEKBYH559.xlsx")

# h516 Sample vs control #
h516 <- gather(HEBYH5161,
           "Sample",
           "Value",
           control1,control2,control3,control4,sample1,sample2,sample3,sample4)
# Raw data calculation #
```

```
sampleh516m<-c(mean(h516$Value[h516$Sample=="sample1"],na.rm = T),
mean(h516$Value[h516$Sample=="sample2"],na.rm = T),
          mean(h516$Value[h516$Sample=="sample3"],na.rm =
T),mean(h516$Value[h516$Sample=="sample4"],na.rm = T))
controlh516m<-c(mean(h516$Value[h516$Sample=="control1"],na.rm = T),
mean(h516$Value[h516$Sample=="control2"],na.rm = T),
           mean(h516$Value[h516$Sample=="control3"],na.rm = T),
mean(h516$Value[h516$Sample=="control4"],na.rm = T))

#calculate ratio between aba/dmso #
h516ratio<- data.frame(sampleh516m/controlh516m,fix.empty.names = FALSE)

# h559 Sample vs control #
h559 <- gather(HEKBYH559,
          "Sample",
          "Value",
          control1,control2,control3,control4,sample1,sample2,sample3,sample4)
# Raw data calculation #
sampleh559m<-c(mean(h559$Value[h559$Sample=="sample1"],na.rm = T),
mean(h559$Value[h559$Sample=="sample2"],na.rm = T),
          mean(h559$Value[h559$Sample=="sample3"],na.rm =
T),mean(h559$Value[h559$Sample=="sample4"],na.rm = T))

controlh559m<-c(mean(h559$Value[h559$Sample=="control1"],na.rm = T),
mean(h559$Value[h559$Sample=="control2"],na.rm = T),
           mean(h559$Value[h559$Sample=="control3"],na.rm = T),
mean(h559$Value[h559$Sample=="control4"],na.rm = T))

#calculate ratio between aba/dmso #
h559ratio<- data.frame(sampleh559m/controlh559m, fix.empty.names = FALSE)

# ggPlot #
# prepare dataframe #
h516ratio$cells <- rep("H5CAG16", 4)
h559ratio$cells <- rep("H5CAG59", 4)
colnames(h516ratio) <- c("ratio", "cells")
colnames(h559ratio) <- c("ratio", "cells")
ratios <- rbind(h516ratio, h559ratio)
#calculate p value #
h5ttest <- t.test(h516ratio$ratio,h559ratio$ratio, paired= TRUE)
h5pval <- formatC(h5ttest$p.value, format = "e", digits = 1)

#draw plot#
ggplot(data=ratios, aes(x=cells, y=ratio)
) +
 # add points #
 geom_jitter(data=ratios, size=1, color="black", width = 0.1
 ) +
 # Changing axis ticks, y axis name with numbers on y axis #
```

```
 scale_y_continuous(name = " ",
                    breaks = seq(0, 0.6, 0.2),
                    limits=c(0, 0.6))+

 # add mean (if you like it dashed, add 'linetype = "dashed"' within brackets) #
 stat_summary(fun.y = mean, geom = "errorbar", aes(ymax = ..y.., ymin = ..y..), width
= 0.2, color="black") +

 # add error bars corresponding to mean+/-sd #
 stat_summary(fun.data=mean_sdl, fun.args = list(mult=1), geom="errorbar",
width=0.1, color="black") +

 # Add x axis label #
 xlab("") +

 # Add title
 ggtitle(" ") +

 # Choose font size
 theme(axis.title.y = element_text(size=20),
       plot.title = element_text(size=20, hjust = 0.5),
       text = element_text(size=20)) +

 # Diplay only x and y axes without any backgroud
 theme(axis.line = element_line(colour = "black"),
       panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       panel.border = element_blank(),
       panel.background = element_blank())
```

Flow cytometry data analysis for HDAC5 truncations

```
# Load data
aliciapyl <- read_excel("~/Desktop/PhDthesis/Figure III./Alicia
truncation/FC/aliciapyl.xlsx")
aliciahdac5 <- read_excel("~/Desktop/PhDthesis/Figure III./Alicia
truncation/FC/aliciahdac5.xlsx")
aliciaNT <- read_excel("~/Desktop/PhDthesis/Figure III./Alicia
truncation/FC/aliciaNT.xlsx")
aliciaCC <- read_excel("~/Desktop/PhDthesis/Figure III./Alicia
truncation/FC/aliciaCC.xlsx")
aliciaCD <- read_excel("~/Desktop/PhDthesis/Figure III./Alicia
truncation/FC/aliciaCD.xlsx")

# PYL Sample vs control #
citb40pyl <- gather(aliciapyl,
                    "Sample",
                    "Value",
     control1,control2,control3,control4,sample1,sample2,sample3,sample4)
```

```
# Raw data calculation
samplecitb40pylm<-c(mean(citb40pyl$Value[citb40pyl$Sample=="sample1"],na.rm
= T), mean(citb40pyl$Value[citb40pyl$Sample=="sample2"],na.rm = T),
            mean(citb40pyl$Value[citb40pyl$Sample=="sample3"],na.rm =
T),mean(citb40pyl$Value[citb40pyl$Sample=="sample4"],na.rm = T))

controlcitb40pylm<-c(mean(citb40pyl$Value[citb40pyl$Sample=="control1"],na.rm
= T), mean(citb40pyl$Value[citb40pyl$Sample=="control2"],na.rm = T),
            mean(citb40pyl$Value[citb40pyl$Sample=="control3"],na.rm = T),
mean(citb40pyl$Value[citb40pyl$Sample=="control4"],na.rm = T))

#calculate p value
citb40pylttest <- t.test(samplecitb40pylm,controlcitb40pylm, paired= TRUE)
citb40pylpval <- formatC(citb40pylttest$p.value, format = "e", digits = 1)

#calculate ratio between aba/dmso
citb40pylratio<- data.frame(samplecitb40pylm/controlcitb40pylm,fix.empty.names
= FALSE)

# wildtype HDAC5 Sample vs control #
citb40hdac5 <- gather(aliciahdac5,
            "Sample",
            "Value",
control1,control2,control3,control4,sample1,sample2,sample3,sample4)
# Raw data calculation
samplecitb40hdac5m<-
c(mean(citb40hdac5$Value[citb40hdac5$Sample=="sample1"],na.rm = T),
mean(citb40hdac5$Value[citb40hdac5$Sample=="sample2"],na.rm = T),
            mean(citb40hdac5$Value[citb40hdac5$Sample=="sample3"],na.rm =
T),mean(citb40hdac5$Value[citb40hdac5$Sample=="sample4"],na.rm = T))

controlcitb40hdac5m<-
c(mean(citb40hdac5$Value[citb40hdac5$Sample=="control1"],na.rm = T),
mean(citb40hdac5$Value[citb40hdac5$Sample=="control2"],na.rm = T),
            mean(citb40hdac5$Value[citb40hdac5$Sample=="control3"],na.rm = T),
mean(citb40hdac5$Value[citb40hdac5$Sample=="control4"],na.rm = T))
#calculate p value
citb40hdac5ttest <- t.test(samplecitb40hdac5m,controlcitb40hdac5m, paired=
TRUE)
citb40hdac5pval <- formatC(citb40hdac5ttest$p.value, format = "e", digits = 1)

#calculate ratio between aba/dmso
citb40hdac5ratio<-
data.frame(samplecitb40hdac5m/controlcitb40hdac5m,fix.empty.names = FALSE)

# N terminal Sample vs control #
citb40nt <- gather(aliciaNT,
            "Sample",
            "Value",
```

```
control1,control2,control3,control4,sample1,sample2,sample3,sample4)

# Raw data calculation
samplecitb40ntm<-c(mean(citb40nt$Value[citb40nt$Sample=="sample1"],na.rm =
T), mean(citb40nt$Value[citb40nt$Sample=="sample2"],na.rm = T),
            mean(citb40nt$Value[citb40nt$Sample=="sample3"],na.rm =
T),mean(citb40nt$Value[citb40nt$Sample=="sample4"],na.rm = T))

controlcitb40ntm<-c(mean(citb40nt$Value[citb40nt$Sample=="control1"],na.rm =
T), mean(citb40nt$Value[citb40nt$Sample=="control2"],na.rm = T),
             mean(citb40nt$Value[citb40nt$Sample=="control3"],na.rm = T),
mean(citb40nt$Value[citb40nt$Sample=="control4"],na.rm = T))

#calculate p value
citb40ntttest <- t.test(samplecitb40ntm,controlcitb40ntm, paired= TRUE)
citb40ntpval <- formatC(citb40ntttest$p.value, format = "e", digits = 1)

#calculate ratio between aba/dmso
citb40ntratio<- data.frame(samplecitb40ntm/controlcitb40ntm,fix.empty.names =
FALSE)

# Coiled-coil HDAC5 Sample vs control #
citb40cc <- gather(aliciaCC,
            "Sample",
            "Value",
            control1,control2,control3,sample1,sample2,sample3)

# Raw data calculation
samplecitb40ccm<-c(mean(citb40cc$Value[citb40cc$Sample=="sample1"],na.rm =
T), mean(citb40cc$Value[citb40cc$Sample=="sample2"],na.rm = T),
            mean(citb40cc$Value[citb40cc$Sample=="sample3"],na.rm = T))

controlcitb40ccm<-c(mean(citb40cc$Value[citb40cc$Sample=="control1"],na.rm =
T), mean(citb40cc$Value[citb40cc$Sample=="control2"],na.rm = T),
             mean(citb40cc$Value[citb40cc$Sample=="control3"],na.rm = T))

#calculate p value
citb40ccttest <- t.test(samplecitb40ccm,controlcitb40ccm, paired= TRUE)
citb40ccpval <- formatC(citb40ccttest$p.value, format = "e", digits = 1)

#calculate ratio between aba/dmso
citb40ccratio<- data.frame(samplecitb40ccm/controlcitb40ccm,fix.empty.names =
FALSE)

# catalytic domain Sample vs control #
citb40cd <- gather(aliciaCD,
            "Sample",
            "Value",
   control1,control2,control3,control4,sample1,sample2,sample3,sample4)
```

```
# Raw data calculation
samplecitb40cdm<-c(mean(citb40cd$Value[citb40cd$Sample=="sample1"],na.rm =
T), mean(citb40cd$Value[citb40cd$Sample=="sample2"],na.rm = T),
            mean(citb40cd$Value[citb40cd$Sample=="sample3"],na.rm =
T),mean(citb40cd$Value[citb40cd$Sample=="sample4"],na.rm = T))

controlcitb40cdm<-c(mean(citb40cd$Value[citb40cd$Sample=="control1"],na.rm =
T), mean(citb40cd$Value[citb40cd$Sample=="control2"],na.rm = T),
             mean(citb40cd$Value[citb40cd$Sample=="control3"],na.rm = T),
mean(citb40cd$Value[citb40cd$Sample=="control4"],na.rm = T))

#calculate p value
citb40cdttest <- t.test(samplecitb40cdm,controlcitb40cdm, paired= TRUE)
citb40cdpval <- formatC(citb40cdttest$p.value, format = "e", digits = 1)

#calculate ratio between aba/dmso
citb40cdratio<- data.frame(samplecitb40cdm/controlcitb40cdm,fix.empty.names =
FALSE)

# prepare dataframe
citb40pylratio$cells <- rep("PYL", 4)
colnames(citb40pylratio) <- c("ratio", "cells")

citb40hdac5ratio$cells <- rep("HDAC5", 4)
colnames(citb40hdac5ratio) <- c("ratio", "cells")

citb40ntratio$cells <- rep("N terminal", 4)
colnames(citb40ntratio) <- c("ratio", "cells")

citb40ccratio$cells <- rep("Coiled-Coil", 3)
colnames(citb40ccratio) <- c("ratio", "cells")

citb40cdratio$cells <- rep("catalytic domain", 4)
colnames(citb40cdratio) <- c("ratio", "cells")

ratioABall <- rbind(citb40pylratio,citb40hdac5ratio,
          citb40ntratio,citb40ccratio, citb40cdratio)

# P value calculation between POI verses PYL
HDAC5ttest <- t.test(citb40pylratio$ratio,citb40hdac5ratio$ratio, paired= TRUE)
HDAC5pval <- formatC(HDAC5ttest$p.value, format = "e", digits = 1)

NTttest <- t.test(citb40pylratio$ratio,citb40ntratio$ratio, paired= TRUE)
NTpval <- formatC(NTttest$p.value, format = "e", digits = 1)

CCttest <- t.test(citb40pylratio$ratio,citb40ccratio$ratio, paired= F)
CCpval <- formatC(CCttest$p.value, format = "e", digits = 1)
```

```
CDttest <- t.test(citb40pylratio$ratio,citb40cdratio$ratio, paired= TRUE)
CDpval <- formatC(CDttest$p.value, format = "e", digits = 1)

#try to order data manually
ratioABall_table <- table(ratioABall$cells)
ratioABall_levels <- names(ratioABall_table)[order(ratioABall_table)]

#set the presentation order
ratioABall$cell2 <- factor(ratioABall$cells, levels = c("PYL","HDAC5","N
terminal","Coiled-Coil",
                                    "catalytic domain"))

#prepare to set color for each sample
ratioABall$cell2 <- as.factor(ratioABall$cell2)
colorprotein =
c("dodgerblue2","darkorchid3","darkorchid3","darkorchid3","darkorchid3")

#draw an overall ggplot
ggplot(data=ratioABall, aes(x=cell2, y=ratio)
) +

 # add points
 geom_jitter(data=ratioABall, size=1, aes(color= cell2),
          width = 0.1
 ) +

#set different color for different POI
 scale_color_manual(values = colorprotein)+

 # Changing axis ticks, y axis name with numbers on y axis
 scale_y_continuous(name = " ",
                breaks = seq(0, 1.6, 0.5),
                limits=c(0, 1.6))+

# add mean (if you like it dashed, add 'linetype = "dashed"' within brackets)
 stat_summary(fun.y = mean, geom = "errorbar", aes(ymax = ..y.., ymin = ..y..), width
= 0.4, color="black") +

 # add error bars corresponding to mean+/-sd
 stat_summary(fun.data=mean_sdl, fun.args = list(mult=1), geom="errorbar",
width=0.2, color="black") +

# Add x axis label
 xlab("") +

 # Add title
 ggtitle(" ") +

 # Choose font size
```

```
 theme(axis.title.y = element_text(size=12),
      plot.title = element_text(size=20, hjust = 0.5),
      text = element_text(size=12)) +

 # Diplay only x and y axes without any backgroud
 theme(axis.line = element_line(colour = "black"),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      panel.border = element_blank(),
      panel.background = element_blank(),
      axis.text.x = element_text(angle = 90),
      legend.position='none')
```

ChIP barplot

```
# load data
library(readxl)
chiph516 <- read_excel("~/Desktop/Capri/H516
ChIP/HEKBYH5CAG161CHIP.xlsx")
View(chiph516)

#start plot the data
par(mar = c(5, 6, 4, 5) + 0.1)
chiph516mean<- barplot(chiph516$mean,
              space =c(0,0,0.2,0,1,0,0.2,0),
              col=c("black", "grey"),
              ylim=c(0,3),
              border = "black",
              ylab="% of input",
              legend.text = TRUE,
              yaxt="n", ann=FALSE
)
segments(chiph516mean, chiph516$mean-chiph516$sderror,
       chiph516mean, chiph516$mean+chiph516$sderror)
arrows(chiph516mean, chiph516$mean-chiph516$sderror,
     chiph516mean, chiph516$mean+chiph516$sderror, lwd = 0.5, angle = 90,
     code = 3, length = 0.05)
axis(2, at=0:3, labels=c(0:3))
```

Appendix C: ChIP Protocal

ChIP INT:

ChIP material: Trypsin cells and resuspend cells in 10ml DMEM media in 15ml tubes.
Take 20μl cells in the mean time for cell counting.

Fix
Add formaldehyde (find in common chemical room, drawer under the hood) to a final concentration of 1% (for 37% stock, add 270μl per 10ml media), RT 10min on shaker or wheel. Time here is essential for achieving a perfect sonication pattern.
Add Glycine to a final concentration of 0.125M (for 2M stock, add 757μl per 10ml) to stop the crosslink, RT 5min on wheel.
1000rpm 5min, discard supernatant to special liquid trash under the hood.
Add ice cold PBS to achieve 10 million cells per ml, then separate cells to 1.5ml Eppendorf tubes.
Wash with ice cold PBS twice, 3min 1000rpm at 4 degree.
Freeze the pellet.

Lysis (NO Chromatin extraction")
- Resuspend cell pellet in LB3 buffer (Complete EDTA-free): 1 ml/10 million cells in the red cap 15ml TPX tubes (normal 15ml TPP tube need much longer time to sonicate)
- Keep on ice

Sonication

Change the ddH2O water in Biorupter everytime before using it
- Biorupter sonication
- Sonication conditions: normally 15min High (CITB cells), 25min High (CITBY40 and CITBY40H5 cells)
- Spin 15 min. @ full speed 4°C
- **Collect each SN and** (Keep 20 μl to check sonication pattern, keep 20 μl as qPCR input, keep 15 μl for protein analysis)
- store all the samples except sonication tubes @ -80°C

Check sonication
Add 80 μl of crosslink reverse buffer to 20 μl samples.

Put all the sample tubes in the grey metal container next to the water bath.

Incubate at least 6H @ 65°C

Purify the samples with Qiagen QIAquick PCR Purification Kit, elute into 20 μl and load on a 1% gel and run for 1h in 80V.

Optimal sonication pattern will be a fat DNA band around 300 bp to 500 bp.

Extract Dilution + Pre-clearing

Dilute cell extracts 1:5 with IP dilution buffer per sample

200 µl extract (2 million cells) + 800 µl IP dilution buffer

Block beads

Well resuspend beads and take the needed amount out into 1.5ml Eppendorf. Centrifuge 800 rpm 2min 4 °C to remove the storage liquid (20% ethanol) and wash 2 times with PBS and once with IP dilution buffer
After final wash, add final concentration of 1mg/ml BSA and 0.3mg/ml salmon sperm DNA (pre-denatured 95°C 5min and put on ice immediately) in IP dilution buffer, 4 degree on wheel for 1h

Preclearing + IP

- Preclear each extract for 30 min to 1 h @ 4°C on wheel (20 µl protein G beads (blocked, 50% slurry in IP dilution buffer) for 1ml diluted sample
- Spin 2 min @ 800 rpm, collect SN (1ml)
- For each IP:
    - 1ml for each (2 million cells) for IP

- Add following beads and antibodies:

	HA- IP
Ab9110 αbody Abcam 1µg/µl	3 µl
G beads Sepharose 4 Fast Flow	50 µl
	FLAG- IP
Flag M2 1µg/µl	3 µl
G beads Sepharose 4 Fast Flow	50 µl

G Beads volumes: slurry 50 % in IP buffer, (blocked: BSA 1 mg/ml, 0.3mg/ml salmon sperm DNA)

- Incubate overnight on wheel @ 4°C

Washes

- Spin IPs for 2 min. @ 800 rpm @ 4°C
- Remove SN (keep if needed in WB) and wash as following (**every wash for 5 min. on the wheel, 4°C**)
- Wash with following buffers once: wash buffer 1, wash buffer 2 and wash buffer 3
- Finally wash all samples twice in wash buffer 4

- At the end of the last wash, keep ¼ of sample (250 µl) for protein analysis, spin as above and resuspend into 2× SDS sample buffer (20 µl), use ¾ of sample (750 µl) for DNA analysis as usual.

Crosslink reversal

- (Thaw input DNA and sonication check samples)
- Treat with Crosslink reversal buffer (final volume: 100 µl, beads do not count as volume:
    - Beads (IPs): 100 µl
    - Input: 80 µl
    - Sonication: 80 µl
- Incubate overnight (or at least 6 h) @ 65°C

DNA extraction

- Take samples from 65°C and put them on ice, proceed with DNA extraction
- Qiagen purification kit:
    - Add 5 × volume (500 µl) binding buffer PB
    - Mix and transfer in a Qiagen column and onto a vacuum manifold (or spin, 30-60 sec, ≥10'000 g)
    - Wash with 750 µl wash buffer PE (spin, 30-60 sec, ≥10'000 g)
    - Spin 1 min full speed to remove EtOH
    - In clean Eppendorf tubes elute DNA with 20 µl dH_2O for sonication pattern check and 50 µl dH_2O for qPCR
    - Spin 1 min full speed

Buffer

LB3 (Julien Marquis, NIHS)

- 10 mM Tris-HCl pH 8.0
- 200 mM NaCl
- 1 mM EDTA
- 0.5 mM EGTA
- 0.1% Na-Deoxycholate (NaDOC)
- 0.25% Sodium lauroyl sarcosinate (NLS)-Sigma 61743-25G
- Protease Inhibitor

For 50 ml:

- 0.5 ml 1 M Tris-HCl (pH 8.0)
- 2 ml 5 M NaCl
- 0.1 ml 0.5 M EDTA (pH 8.0)
- 50 µl 0.5 M EGTA (pH 8.0)
- 0.5 ml 10% Na-Deoxycholate
- 2.5 ml 5% NLS
- 1 x Complete protease inhibitor Add fresh
- dH_2O add to 50 ml (44.4 ml)

IP dilution buffer

For 40 ml:

- 1.25% Triton-X	5 ml (10%)
- 1 mM EDTA pH 8.0	80 µl (0.5M stock)
- 0.5 mM EGTA pH 8.0	40 µl (0.5M stock)
- 16.25 mM Tris-HCl pH 8.0	650 µl (1M stock)
- 137.5 mM NaCl	1.1 ml (5M stock)
- dH_2O	32.73 ml

1 x Complete protease inhibitor **Add fresh**

Wash 1

For 100 ml:

-0.1% SDS	1ml (10% stock)
-1% Trition	1ml
-20mM Tris-HCl pH 8.0	2ml (1M stock)
-2mM EDTA pH 8.0	400µl (0.5M stock)
-300 mM NaCl	6ml (5M stock)

Filter

Wash 2

For 100 ml:

-0.1% SDS	1ml (10% stock)
-1% Trition	1ml
-20mM Tris-HCl pH 8.0	2ml (1M stock)
-2mM EDTA pH 8.0	400µl (0.5M stock)
-500 mM NaCl	10ml (5M stock)

Filter

Wash 3

For 100 ml:

-1% NP-40	1ml
-1% Na-deoxycholate	10ml (10% stock)
-10mM Tris-HCl pH 8.0	1ml (1M stock)
-1mM EDTA pH 8.0	200µl (0.5M stock)
-250mM LiCl	26ml (4.8M stock)

Filter

Wash 4

For 100 ml:

-10mM Tris-HCl pH 8.0	1ml (1M stock)
-1mM EDTA pH 8.0	200µl (0.5M stock)

Filter

Crosslink Reversal Buffer

For 500ml:

-1% SDS	50ml (10% stock)
-0.1M NaHCO3	50ml (1M stock)
-0.5mM EDTA pH 8.0	0.5ml (0.5M stock)
-20mM Tris-HCl pH 8.0	10ml (1M stock)
-RNase: final concentration 100µg/ml	Add fresh