University of Central Florida
STARS

Electronic Theses and Dissertations, 2020-

2020

Monitoring Pathological Gene Expression and Studying Endogenous Epigenetic Architecture by CRISPR/Cas9-based Tool Development using alpha-Synuclein as a Model

Levi Adams
University of Central Florida

 Part of the Genetics and Genomics Commons, and the Medical Genetics Commons
Find similar works at: https://stars.library.ucf.edu/etd2020
University of Central Florida Libraries http://library.ucf.edu

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2020- by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation
Adams, Levi, "Monitoring Pathological Gene Expression and Studying Endogenous Epigenetic Architecture by CRISPR/Cas9-based Tool Development using alpha-Synuclein as a Model" (2020). *Electronic Theses and Dissertations, 2020-*. 598.
https://stars.library.ucf.edu/etd2020/598

MONITORING PATHOLOGICAL GENE EXPRESSION AND STUDYING ENDOGENOUS EPIGENETIC ARCHITECTURE BY CRISPR/CAS9-BASED TOOL DEVELOPMENT USING ALPHA-SYNUCLEIN AS A MODEL

by

LEVI ADAMS
B.S. University of North Texas, 2014

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Burnett School of Biomedical Sciences
in the College Medicine
at the University of Central Florida
Orlando, Florida

Summer Term
2020

Major Professor: Yoon-Seong Kim

© 2020 Levi Adams

ABSTRACT

Until recently, complete understanding of the endogenous activity of pathologically relevant genes was out of reach and research was confined to *in situ* work, plasmid-based constructs and artificial model systems. The development and expansion of the CRISPR/Cas9 genome editing technique has enabled us to explore the molecular underpinnings of gene activation using the cell's own endogenous regulatory environment. In this work, we report on the development of a novel tool to monitor the endogenous activity of a causative gene in Parkinson's disease, α-synuclein. We use CRISPR/Cas9 to insert a highly sensitive engineered luciferase at the C-terminal of α-synuclein and assessed its responses to stimuli. Our system responds to epigenetic stimuli, which was unable to be recapitulated by previously available gene activity assays. After development of a sensitive detection tool for epigenetic stimuli, we focused on developed a modular suite of epigenetic writers and erasers by modification of the SunTag protein tagging system and used catalytically dead Cas9 (dCas9) to direct our modular epigenetic toolkit to individual genes. We show that our toolkit of epigenetic effectors successfully writes epigenetic information in a site-specific manner. Using the sensitive α-synuclein reporter we previously developed, we screen the promoter region of this pathologically relevant gene at high resolution and identify the most effective areas for epigenetic intervention in this cell line. These tools allow us to dissect and understand the endogenous regulatory mechanisms of almost any gene targetable by Cas9 in ways that were not previously available may prove to be an effective strategy for persistently altering

pathologic transcriptional activity. This system offers a strong tool for to dissect and understand underlying epigenetic architecture and opens potential new avenues for therapeutic strategies for various disease conditions.

This is dedicated to my parents. My father instilled a love a science from a very early age watching Nature and Nova every week with me. My mother always encouraged me to do my best and never accepted excuses from me. From an early age I learned a strong work ethic and a keen interest in science and medicine from them both and I am proud to be their son.

ACKNOWLEDGEMENTS

I would like to thank my advisor Yoon-Seong Kim for his patience with me and support over the last 5 years. When I suggested a difficult and new scientific direction to explore, he offered encouragement. During more difficult times he continued to back me and guide me to grow as a researcher and I can't thank him enough.

I would also like to thank my committee – Mollie Jewett, Deborah Altomare and Amber Southwell. Every one of them has gone far beyond what is required of a committee member and helped me with experimental design, scientific understanding, and helped guide me through the challenges of a PhD – scientific as well as personal. I would also like to thank Alvaro Estevez – an early committee member and continuous advocate of my success even after he changed institutions.

Thank you to Subhrangshu Guhatakurta. Over the years we worked together he (almost) never lost patience with my questions about experimental design, research, or technique and I'm glad to count him among my friends.

TABLE OF CONTENTS

LIST OF FIGURES xii

LIST OF TABLES xiv

LIST OF ABBREVIATIONS xv

CHAPTER ONE: GENERAL INTRODUCTION 1

1.1. Parkinson's disease and αSynuclein 1

1.1.1. The αSynuclein Protein in PD 2

1.1.2. Increasing our Understanding of αSynuclein Transcriptional Control 5

1.2. The CRISPR/Cas9 Technique 6

1.2.1. Discovery and Development 6

1.2.2. Developing the CRISPR/Cas System into a Genome Engineering Tool 9

1.3 Using CRISPR/Cas9 to Edit Genomes 11

1.4. The NanoLuc Reporter 14

CHAPTER TWO: DESIGNING A NOVEL TOOL FOR MONITORING ENDOGENOUS ALPHA-SYNUCLEIN TRANSCRIPTION BY NANOLUCIFERASE TAG INSERTION AT THE 3'END USING CRISPR/CAS9 GENOME EDITING TECHNIQUE 16

2.1. Abstract 16

2.2. Introduction 17

2.3. Results....18

2.3.1. Generation of a stable cell line endogenously tagged with functional NanoLuc luciferase reporter at the 3'end of SNCA....18

2.3.2. αSyn-NanoLuc luciferase activity reflects SNCA transcriptional regulation....21

2.3.3. Exogenous promoter reporter assays failed to reproduce transcriptional activation of SNCA as seen in endogenous conditions....22

2.4. Discussion....24

2.5. Methods....31

2.5.1 Cell Culture....31

2.5.2. Designing SNCA specific short guide RNA (sgRNA)....31

2.5.3. Generation of HEK-293T cell line stably expressing SNCA-NanoLuc (293T-SNCA-3'NL)....33

2.5.4. Confirmation of stable integration of the NanoLuc reporter at the 3'end of SNCA....33

2.5.5. Western blotting....34

2.5.6. Assay for the NanoLuc luciferase activity....35

2.5.7. SNCA promoter-reporter assay....35

2.5.8. Cell treatment paradigm....36

2.5.9. Bisulfite sequencing....37

2.5.10. Semi quantitative reverse transcriptase PCR (RT-PCR) 37

2.5.11. Statistical analysis 38

2.6. Figures 39

CHAPTER THREE – CRISPR-MEDIATED EPIGENOMIC ENGINEERING 56

3.1. Using dCas9 to Manipulate DNA 56

3.1.1. dCas9-based Tools 57

3.1.2. Using dCas9 to Modify Transcription 57

3.2. CRISPR-mediated epigenome editing 60

3.3. Using dCas9 to Manipulate Histones 61

3.3.1 – Current Approaches to Modify Histone Marks 64

CHAPTER FOUR: PRECISE EPIGENOMIC EDITING BY A NEWLY DEVELOPED MODULAR EPIGENETIC TOOLKIT 65

4.1. Abstract 65

4.2. Introduction 66

4.3. Results 67

4.3.1. Developing the Modular Epigenetic Toolkit 67

4.3.2. The Modular Epigenetic Toolkit Directly Edits Histone Marks in a Targeted Manner 68

4.3.3. Using the Toolkit to Modulate Pathological Gene Activity 69

4.3.4. Screening *SNCA* with NanoLuc 69

4.3.5. High Resolution *SNCA* Screen Reveals Effective Targeting Regions 70

4.4. Discussion 72

4.5. Methods 72

4.5.1. Cloning 72

4.5.2. Cell Culture 73

4.5.3. Chromatin Immunoprecipitation 74

4.5.4. Western Blot 75

4.5.5. Expression of *SNCA* gene 76

4.5.6. Bioinformatics Analysis 76

4.5.7. Statistical Analysis 78

4.6. Figures 79

CHAPTER FIVE: CONCLUSIONS 95

5.1. Limitations of CRISPR 95

5.1.1. Cpf May Provide Alternatives to Cas9 96

5.2. Understanding Endogenous Gene Regulation 96

5.3. Engineered Endogenous Reporter using NanoLuc 98

5.3.1. Future Directions for our NanoLuc Reporter System 98

5.4. Modular Epigenetic Toolkit.....99

5.4.1. Future Directions and Challenges for the Modular Epigenetic Toolkit 100

APPENDIX A: DEFENSE ANNOUNCEMENT AND PUBLICATIONS.....104

APPENDIX B: COPYRIGHT PERMISSIONS 107

LIST OF REFERENCES 110

LIST OF FIGURES

Figure 1 - αSynuclein Structure and Aggregation 4

Figure 2 - The CRISPR/Cas locus 8

Figure 3 – CRISPR/Cas9 Mediated Genome Engineering 12

Figure 4 – Development of 293TSNCA3'NL cells 39

Figure 5 – Sequence alignment of wild type SNCA against 293T-SNCA-3'NL genomic DNA sequence 41

Figure 6 - 293T-SNCA-3'NL cDNA indicates correct insertion of NanoLuc sequence. 43

Figure 7 – Functional expression of the NanoLuc luciferase 45

Figure 8 – Full western blot of 293T-SNCA-3'NL cells showing NanoLuc tagged and WT αSynuclein. 47

Figure 9 – 293TSNCA3'NL cells having the NanoLuc integration can be used to model deregulated SNCA as seen in sporadic PD 49

Figure 10 – Comparable increase of endogenous and NanoLuc-tagged SNCA levels following dopamine treatment 51

Figure 11 – Exogenous overexpression of SNCA driven firefly and Renilla luciferase failed to replicate endogenous SNCA behavior after similar paradigm of drug treatment. 53

Figure 12 – Using dCas9 for Genomic Engineering 59

Figure 13 – H3 Histone Tails Influence Transcription 63

Figure 14 – Toolkit of Epigenetic Editors 79

Figure 15 – Design of Epigenetic Toolkit Writers 81

Figure 16 – Endogenous Epigenetic Structure of Selected Genes in A549 cells 83

Figure 17 – Alterations in endogenous epigenetic structures by SunTag writers.....85

Figure 18 – Screening the α-synuclein gene promoter with NanoLuciferase.....87

Figure 19 – Screening of single gene promoter.....89

Figure 20 - Human SNCA upstream regulatory region with sgRNA binding sites 91

LIST OF TABLES

Table 1 – Alternative Cas9 Orthologs.....10

Table 2 – Primers Referenced in Chapter 2.....55

Table 3 – sgRNA Referenced in Chapter 4.....93

Table 4 – Primers Referenced in Chapter 4.....94

LIST OF ABBREVIATIONS

5-AzadC – 5-azadeoxycytidine

Cas – CRISPR-associated

ChIP – Chromatin Immunoprecipitation

CRISPR – Clustered Regularly Interspaces Short Palindromic Repeats

dCas9 – Catalytically dead Cas9

DNMT – DNA Methyltransferase

DSB – DNA Double Strand Break

ENCODE – Encylcopedia of DNA Elements

FACS – Fluorescent Activated Cell Sorting

H3 – Histone subunit H3

HDAC – Histone deacetylase

HDR – Homology-directed DNA repair pathway

kDa – kiloDalton

LB – Lewy Body

NAC – non-amyloid beta component

NHEJ – Non-homologous End Joining DNA repair pathway

PAM – Protospacer Associated Motif

PD – Parkinson's disease

PTM – Post-translational Modification

qRT-PCR – Quantitative Real Time PCR

RIPA – Radio immunoprecipitation buffer

RT-PCR – Semi-quantitative reverse transcriptase PCR

scFv – Single-chain Variable Fragment of an Antibody

sgRNA – Single guide RNA

SNP – Single Nucleotide Polymorphism

spCas9 – *Streptococcus pyogenes* Cas protein 9

TET – Ten-Eleven Translocase

UTR – Untranslated Region

WT – Wild type

αSyn – αSynuclein

CHAPTER ONE: GENERAL INTRODUCTION

During my graduate studies, I have worked to develop and validate tools to screen and affect pathological transcriptional activity by using the CRISPR/Cas9 system. Using a key Parkinson's disease gene, α-synuclein (αSyn), as a model, we developed an endogenous reporter system using an engineered luciferase to detect transcriptional changes in αSyn levels and allow us to easily screen potential intervention strategies in the cell. We took this system further by developing a suite of targeted epigenetic modulators and using them to screen the promoter region of the α-synuclein gene (*SNCA)* and identify putative genetic targets that may potentially be used to directly affect αSyn levels while increasing our understanding of the endogenous epigenetic architecture of this, and many other, genes. To achieve this, we took advantage of the CRISPR/Cas9 system's ability to make tailored changes to eukaryotic genomes in a site-specific manner.

1.1. Parkinson's disease and αSynuclein

Parkinson's disease (PD) is a progressive neurodegenerative condition affecting about 60,000 people annually in the US[1]. Incidence increases with age and there are approximately one million people in the US with PD. First described in 1817 as 'shaking palsy' by James Parkinson, the primary symptoms of PD are motor related: shaking arms and hands, involuntary movements, and a characteristic rigid and stooped posture. In two centuries since it was initially described, non-motor symptoms have also been identified including constipation,

insomnia, dementia and depression[2,3]. Physiologically PD is characterized by the loss of dopaminergic neurons from the substantia nigra (pars compacta) region of the midbrain. The motor symptoms are a side effect of this, and by the time motor symptoms are apparent in the patient, 60-80% of the dopamine neurons have already been lost[4,5]. Generally, PD is classified in two different ways based on its initial trigger. Inherited genetic abnormalities, often known as familial PD, is based on inherited mutations passed through the germ line. This accounts for less than 10% of total PD cases. The remaining 90% of patients develop PD with no known genetic trigger, referred to as sporadic PD[6].

One of the earliest known pathological hallmarks of PD is the appearance of cytoplasmic proteinaceous inclusion bodies in neuronal perikarya known as Lewy Bodies (LBs)[7]. While LBs have many components such as ubiquitin, in 1997 it was discovered that these inclusion bodies contained primarily aggregates of the protein αSynuclein (αSyn)[8]. Around this time, it was revealed that αSyn was a primary component of the Lewy bodies seen in PD, and that mutations in the αSyn gene (*SNCA*) were revealed to cause genetically heritable, early-onset PD. Later studies confirmed the link between αSynuclein and PD[9-12].

1.1.1. The αSynuclein Protein in PD

αSyn is a 140 kDa protein that exists as a natively unfolded, random coil protein localized primarily in the presynaptic terminals. αSyn can adopt multiple conformations depending on its

localization. When interacting with plasma membranes it can adopt a helical conformation (Figure 1) and the NAC domain (non-AB component of amyloid plaque) can form beta-pleated sheets when two αSyn molecules interact with each other[13]. The exact function of αSyn still remains unclear. It has been shown that αSyn expression is induced during neuronal differentiation and synaptic development and it is believed that αSyn plays a role in modulation of synaptic transmission[14-17]. Knockout or overexpression of αSyn affects neurotransmitter release and it has been shown that αSyn transiently binds vesicles during neuronal firing[18-20].

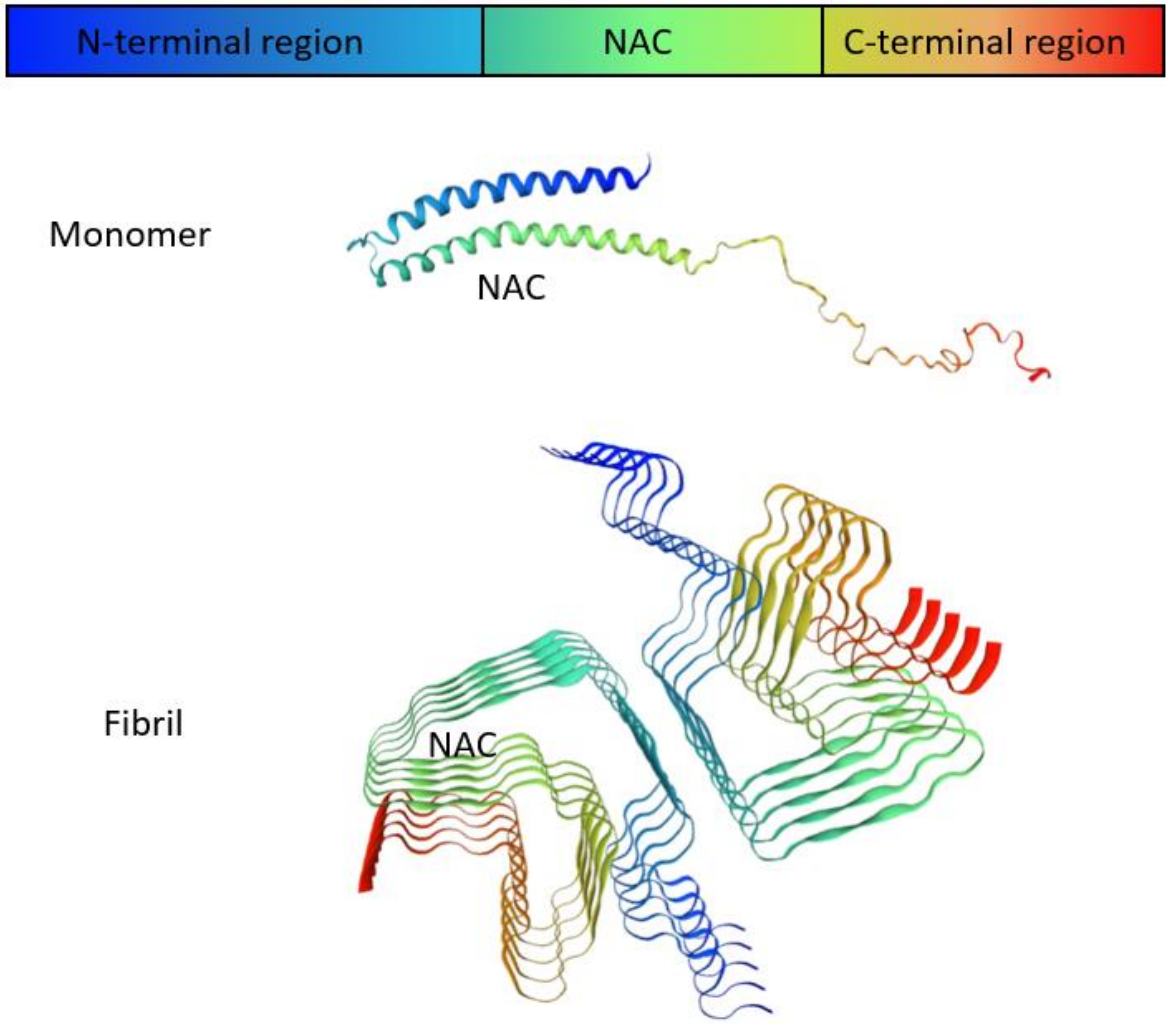


Figure 1 - αSynuclein Structure and Aggregation

Figure 1: Domain Structure of αSynuclein showing the NAC domain of the protein and 3D conformations. The monomer form is shown with α-helical structures. During aggregation, the NAC domain forms β-pleated sheets both intra- and inter-molecularly leading to fibril formation. (3D structures from SWISS model repository publicly available at https://swissmodel.expasy.org/repository/uniprot/P37840)

While the specific functions of αSyn have yet to be fully elucidated, its role in contributing to pathological progression of PD is well established. αSyn is able to form multimeric structures in the cell and these have been shown to drive PD, although there remain some questions as to which specific multimeric forms are toxic. Recently it was shown that αSyn exists in a delicately balanced tetramer in the cell, although the exact details of this remain unclear[21]. αSyn is able to undergo template directed misfolding – where one misfolded protein is able to act as a seed that induces misfolding and aggregation in other normal proteins in a prion-like manner[22]. Misfolded proteins aggregate into oligomers which can condense into ring-like, spherical and string-like structures known as proto-fibrils. Proto-fibirils can further condense into fibrils and eventually develop into the hallmark Lewy bodies observed in PD[23,24] (Figure 1). These aggregated fibril structures are able to propagate through the brain in a characteristic pattern and are thought to be a potential root cause of the neurodegeneration seen in PD. In the time since αSyn was identified as a potential agent in PD, several other genetic and environmental factors have been identified that affect αSyn levels or increase misfolded proteins. These include a wide variety of harmful stimuli such as exposure to heavy metals or pesticides, post-translational modifications, oxidative stress and, more recently, epigenetic dysregulation[13,25-29].

1.1.2. Increasing our Understanding of αSynuclein Transcriptional Control

While much effort has been spent uncovering the factors that influence αSyn aggregation and subsequent neurodegeneration, comparatively little intellectual attention has been focused on

understanding the mechanisms of transcriptional or epigenetic regulation. As early report showed a single copy number variation from gene duplication is sufficient to induce early onset PD[12,30], it can be surmised that αSyn levels play a critical role in the development and pathogenesis of PD and our understanding of this disease could be increased by a clearer picture of the mechanisms driving αSyn in the cell. Transcriptional control includes a wide variety of factors that influence mRNA levels and stability. Post-transcriptional events can degrade or stabilize mRNA, and an enormous array of transcription factors regulate binding and activity of RNA polymerase in response to various stimuli[31]. However, a central event to transcriptional control in eukaryotic cells occurs on the chromatin itself from action of DNA elements, direct modification of the DNA through methylation, and the post-translational histone modifications that control the chromatin structure. The completion of the ENCODE (Encyclopedia of DNA Elements) project has opened the door for a deeper understanding of the complex regulation of the genome by both cis- and trans- acting factors[32].

1.2. The CRISPR/Cas9 Technique

1.2.1. Discovery and Development

In recent years, the technology known as CRISPR has become the central technique used for genetic engineering, although it took decades of work by many researchers to move it from an intriguing feature of prokaryotic genomes into arguably the biggest technological advances in molecular biology since PCR. The basis for the development of this highly specific and

relatively easy-to-use technique began over 30 years ago when, while sequencing the *iap* gene in *E. coli*, a group under Dr. Atsuo Nakata reported the presence of a short repeating palindromic sequence of bases separated by a variable 32-nucleotide spacer sequence[33]. Over a decade later, advances in sequencing technology revealed that 40% of bacteria had these repeat sequences[34], and that they were adjacent to well-conserved coding sequences[35]. These have since been named CRISPR (Clustered regularly Interspaced Short Palindromic Repeats) and Cas (CRISPR-associated) sequences. After the discovery of these regions, it was shown that the spacer regions between the palindromic repeat sequences belong primarily to viral genomes and were extra-chromosomal in origin[36-39]. This led to the idea that the CRISPR/Cas regions function as a bacterial immune system, which was later confirmed after researchers showed that bacteria add new spacer sequences after phage challenge, and that these spacer sequences controlled the targeting specificity of the Cas proteins[40]. This same work identified the adjacent Cas genes as critical to the function of bacterial adaptive immunity and raised the potential idea that the Cas genes function as targetable restriction nucleases (Figure 2).

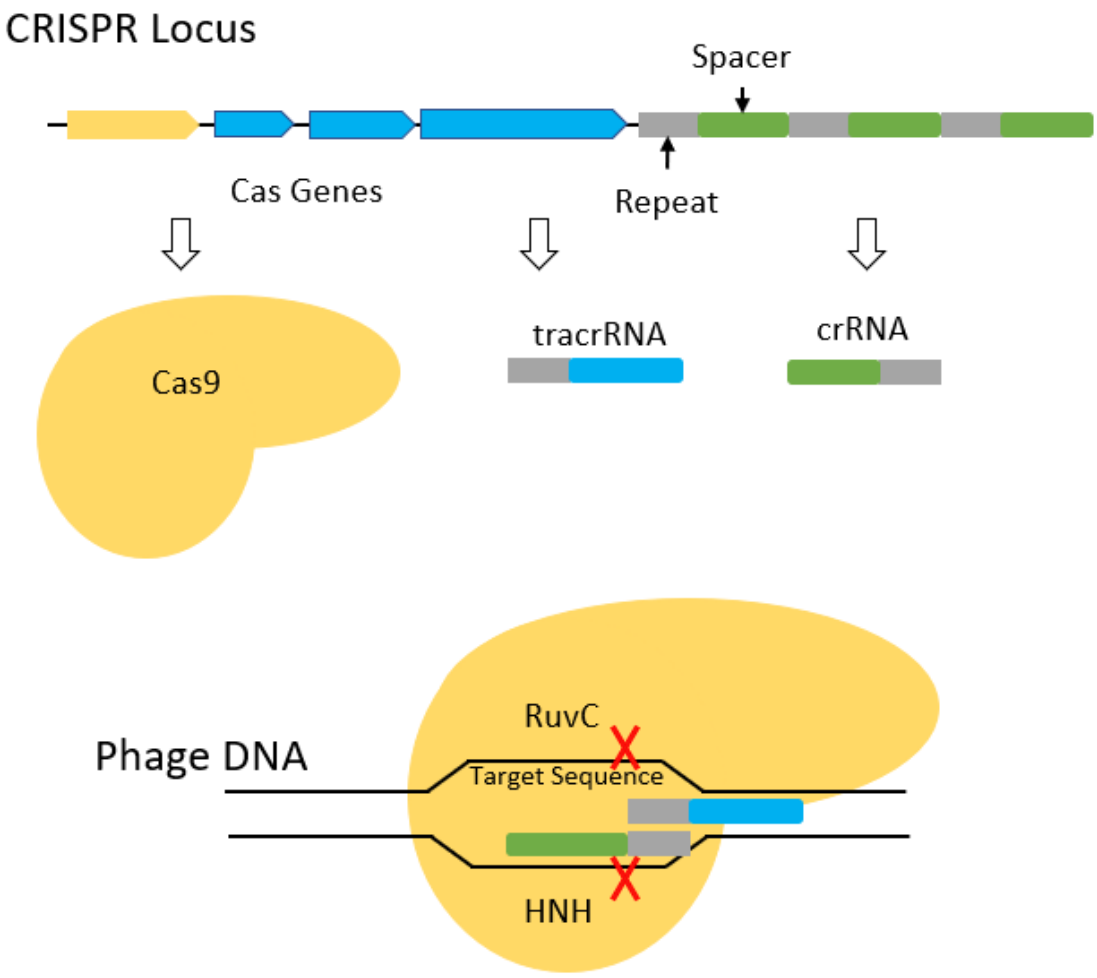


Figure 2 - The CRISPR/Cas locus

Figure 2: In the bacterial CRISPR locus, highly conserved Cas Genes sit adjacent to the CRIPSR Array containing phage-targeting spacer sequences. The Cas genes produce the endonuclease Cas9 and process the trans-activating crRNA (tracrRNA) and the targeting CRISPR RNA (crRNA). When assembled, the RuvC and HNH domains cleave DNA as directed by the tracrRNA-crRNA complex.

1.2.2. Developing the CRISPR/Cas System into a Genome Engineering Tool

Once these publications laid the foundational groundwork for understanding the idea that the CRISPR/Cas system could function as a targetable nuclease, research moved forward very quickly. It was shown that the Cas enzymes are guided by short RNA molecules that are coded by the spacer sequences[41,42]. Additionally it was shown that the genomic DNA targeted by the spacer (known as the protospacer) contained short conserved sequences immediately adjacent to the protospacer and that these protospacer-adjacent motifs (PAM) are required for CRISPR activity[43]. Separately it was revealed that in *S. themophlius* Cas5 (later reidentified as Cas9) was the enzyme with nuclease activity[44], that this enzyme is sufficient for DNA cleavage without other Cas proteins and that its function can transfer into *E coli*[45]. Simultaneously, focused work was uncovering the specific functions and processing of the short RNA molecules used to guide the Cas enzymes. The Charpentier group identified two different RNA species (tracRNA and crRNA) responsible for CRISPR activity and identified the biogenesis and processing by RNaseIII and Cas protein Csn1[46]. Once the biochemical underpinnings of the CRISPR system were understood, its potential as a specific editing tool became clear. Several research groups led by Charpentier and Doudna, Siksnys, Church, and Zhang immediately published studies showing Cas9 can be directed by a chimeric single guide RNA strand (sgRNA) that mimics the processed tracrRNA/crRNA to cleave DNA targets in bacterial and eukaryotic cells and developed a range of simple-to-use tools optimized for human cells to allow this technology to be used by researchers around the world[47-51].

The CRISPR systems are classified according to the different Cas proteins associated with the array. Broadly speaking, there are two separate classes of CRISPR systems, although the specifics of each system continue to be defined as researchers identify and classify new subtypes using comparative genomics, structural analysis and biochemical assays[52]. Class 1 CRISPR contains type I and type III CRISPR systems and is generally found in Archea. Class 2 CRISPR contains type II, IV, V, and VI CRISPR arrays[52]. The most commonly used CRISPR is type II CRISPR/Cas9 isolated in *S. pyogenes* (spCas9), primarily because the PAM sequence is the simple and relatively common -NGG motif that can be found very easily in the genome allowing for broad applicability of this system. There are also are a wide variety of spCas9 variants that can be used as alternatives, including Cas13 that can directly target RNA[53]. These orthologs use a variety of PAM target sequences, have different sizes, and have the potential to leave staggered 'sticky' ends after cutting (Table 1).

Table 1 – Alternative Cas9 Orthologs

Cas9 Varieties	PAM	Size(aa)	Cutting Site (From PAM)	Reference
spCas9	NGG	1368	-3	Jinek et al[47], Gasiunas et al[48]
FnCas9	NGG	1629	-3	Hirano et al[54]
St1Cas9	NNAGAA	1121	-3	Gausinas et al[48], Cong et al[50]
St3Cas9	NGGNG	1409	-3	Gausinas et al[48], Cong et al[50]
CjCas9	NNNNACAC	984	-3	Kim et al[55]
AsCpf1	TTTV	1307	-19/24	Yamano et al[56], Kim et al[57]
LbCpf1	TTTV	1228	-19/24	Yamano et al[56], Kim et al[57]
Cas13	--	Multiple	--	Abudayyeh et al[53]

1.3 Using CRISPR/Cas9 to Edit Genomes

Once loaded with a sgRNA sequence, Cas9 will target DNA with a complementary sequence. Once bound, the CRISPR/Cas9 complex will form a three-stranded DNA:RNA nucleic acid structure known as an R-loop[58]. The guideRNA invades the DNA helix and displaces the opposing strand, and then Cas9 introduces a nick on each DNA strand using two separate nickase domains (RuvC and HNH)[58]. This results in a double strand break (DSB) 3-bp upstream of the PAM targeting sequence. In a eukaryotic cell, DSBs are lethal, so the endogenous repair mechanisms will attempt to correct this (Figure 3). In most cells, the non-homologous end joining (NHEJ) mechanisms will activate to quickly seal the break which leads to a random insertion/deletion (indel) of DNA bases immediately surrounding the DSB site. In dividing cells, the alternative homology-directed repair (HDR) may activate if a complete, undamaged template strand is available. During HDR, regions with high homology to the area surrounding the DSB will be used to template the damaged strand to replace any genetic information lost at the break site[59].

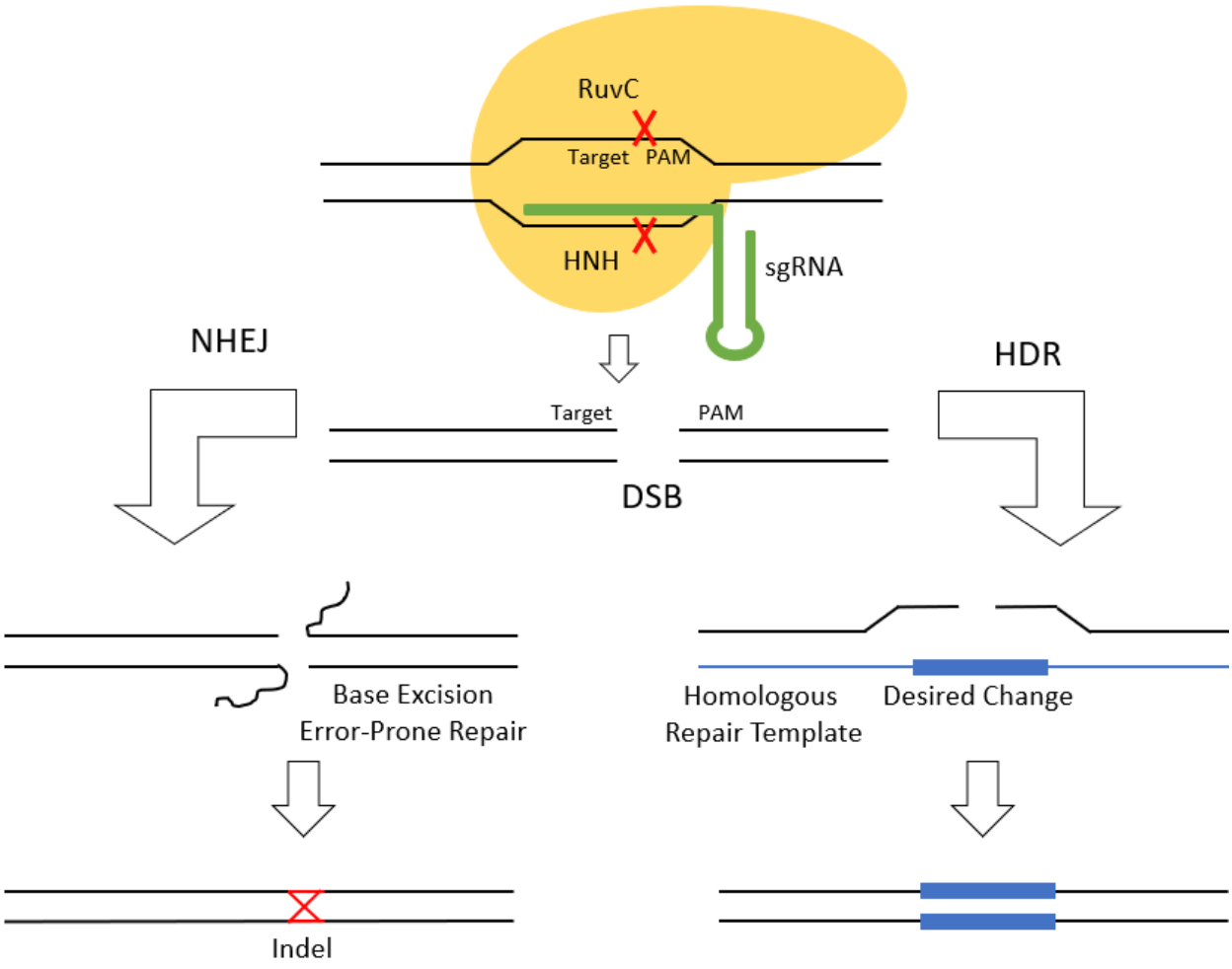


Figure 3 – CRISPR/Cas9 Mediated Genome Engineering

Figure 3: Eukaryotic DNA is targeted using sgRNA to target Cas9 to any PAM-adjacent sequence. Cas9 will create a double strand break which the cell will need to repair. The error prone NHEJ pathway frequently leads to random loss of genetic information and indel formation. Repair by HDR with a donor DNA strand homologous to the original sequence allows replacement of genetic information around the break site with sequences designed by the researcher.

The CRISPR/Cas9 system makes it relatively easy to knockout a protein from cells because NHEJ-mediated indels usually lead to frameshift mutations, disabling target proteins. Additionally, targeting the area around the start codon can be used to prevent translation from occurring. Alternatively, the HDR pathway allows researchers to substitute the original genomic information with a wide variety of new sequences by providing exogenous DNA that has enough homology to the original sequence to be used as a repair template (Figure 3). These two functions are some of the most common uses of the CRISPR/Cas9 system in the laboratory, but several additional tools have been created by re-engineering the CRISPR/Cas9 system and improving its function[60]. One of the main drawbacks to this technology is accidental cleavage at non-target sites that have enough homology to the sgRNA targeting sequence to allow Cas9 to bind – as having a slightly less specific targeting system would be advantageous to bacteria in their defense against rapidly changing phage[61-63]. Several approaches to reduce these potential off-target effects have been used such as rationally designed mutations to increase specificity[64,65]. Additionally, single point mutations to the catalytic cleavage domains can create Cas9 variants that only cleave one DNA strand, requiring that two of these Cas9-nickases be delivered to one area with two separate sgRNAs to nick opposite strands and induce a DSB[66]. Recently a system combining dCas9-nickase with reverse transcriptase and a chimeric sgRNA/RT primer was developed allowing knock-in of edited DNA in a highly specific manner[67]. Finally, development of additional inducible forms of Cas9 requiring small molecule activators, light or ligands allows researchers more control of Cas9 activity and potentially reduces the chance for non-specific effects[68-70].

Taking advantage of the ability of the CRISPR/Cas9 system to allow for targeted insertion of researcher-defined sequences in a precise fashion using HDR, we can now "knock-in" sequences to add features to the genome. As previously discussed, the α-synuclein gene plays a central role in the pathogenesis of Parkinson's disease, but relatively little effort has been spent understanding the endogenous mechanisms that control transcription. Using the CRISPR/Cas9 technique, it is now possible to insert reporter constructs directly into the *SNCA* gene and detect in real time when the gene is being transcribed and translated into protein. During the first part of my dissertation period, we worked to develop a reporter system for αSyn that would allow very sensitive detection of levels in response to various stimuli using only the cell's endogenous regulatory mechanisms. To achieve this, we turned to an engineered luciferase-based reporter: NanoLuc.

1.4. The NanoLuc Reporter

Previous work using firefly-based reporter systems has revealed many key insights into regulation of genes by trans-acting factors such as miRNA and transcription factors. A frequently employed strategy to measure transcriptional regulation involves generation of a plasmid where a target DNA region such as a promoter or 3'UTR is fused to Firefly luciferase[71]. Transfection of this into cells allows researchers to measure changes in luminescence during experimental manipulation to reveal changes in mRNA expression. While Firefly luciferase has been a useful tool in revealing key interactions of various factors with DNA elements in plasmid-

based systems, it is not a practical choice for quantitative genetic reporter systems due to its large size, short signal half-life and poor signal intensity[72,73].

To circumvent these limitations, we used the engineered luciferase, NanoLuc, developed by Promega in 2012 from deep sea shrimp[74]. This rationally designed reporter construct has several advantages over previous Firefly and Renilla luciferase reporters. Importantly, the luminescence activity is approximately 100x brighter than either Firefly or Renilla luciferase. This is an important feature for development of endogenous genetic reporter systems – previous plasmid-based systems rely on powerful viral promoters like CMV and transfections of a high number of copies into cells. Endogenous genes are produced at much lower levels and Firefly luciferase signals may not be detectable. Additionally, the NanoLuc protein is much smaller than the Firefly luciferase (19kDa vs 61kDa)[74], making it more practical to insert into a genome using CRIPSR-mediated HDR. Finally, the novel imidazopyrazinone substrate furimazine produces luminescence with a half-life of nearly 3 hours, making it easier to use in a laboratory setting[74]. For these reasons, we utilized this luciferase reporter to develop an endogenous reporter for α-synuclein.

CHAPTER TWO: DESIGNING A NOVEL TOOL FOR MONITORING ENDOGENOUS ALPHA-SYNUCLEIN TRANSCRIPTION BY NANOLUCIFERASE TAG INSERTION AT THE 3'END USING CRISPR/CAS9 GENOME EDITING TECHNIQUE

Levi Adams#, Sambuddha Basu#, Subrangshu Guhathakurta, Yoon-Seong Kim

#contributed equally

(This chapter has been published as an original research article in Scientific Reports (2017) 47, 45883; doi: 10.1038/srep45883.)

2.1. Abstract

αSynuclein (αSyn) is a major pathologic contributor to Parkinson's disease (PD). Multiplication of αSyn encoding gene (SNCA) is correlated with early onset of the disease underlining the significance of its transcriptional regulation. Thus, monitoring endogenous transcription of SNCA is of utmost importance to understand PD pathology. We developed a stable cell line expressing αSyn endogenously tagged with NanoLuc luciferase reporter using CRISPR/Cas9-mediated genome editing. This allows efficient measurement of transcriptional activity of αSyn in its native epigenetic landscape which is not achievable using exogenous transfection-based luciferase reporter assays. The NanoLuc activity faithfully monitored the transcriptional regulation of SNCA following treatment with different drugs known to regulate αSyn expression; while exogenous promoter-reporter assays failed to reproduce the similar outcomes. To our knowledge, this is the first report showing endogenous monitoring of αSyn

transcription, thus making it an efficient drug screening tool that can be used for therapeutic intervention in PD.

2.2. Introduction

αSynuclein (αSyn) is a key protein involved in the progression and pathogenesis of Parkinson's disease (PD)[11]. Familial PD studies have revealed that multiple copies of the gene encoding αSyn (SNCA) cause severe early onset PD, highlighting the importance of its tight transcriptional control[12,25,30,75]. However, little is known about the transcriptional dysregulation of SNCA. Recently, Gründemann et al. confirmed significant increase in SNCA mRNA levels in individual dopamine neuron from idiopathic PD brains using laser capture microdissection, implying a significant transcriptional de-regulation in pathologic conditions[76]. Moreover, recent progress in research on epigenetic influences on SNCA transcription revealed that hypomethylation of SNCA regulatory region play a significant role towards its higher expression in idiopathic PD[77-79].

Advances in genome mapping and the completion of ENCODE project (Encyclopedia of DNA Elements) highlighted the importance of epigenetic architecture governing transcriptional regulation of a gene[80,81]. In light of these discoveries, complete understanding of SNCA expression in pathologic conditions may require a molecular tool/system that can detect changes in transcription and also account for changes brought about by endogenous epigenetic modulation of the gene. Currently, the most widely used tool for understanding transcriptional activity of a gene is by using luciferase reporter fused to the promotor of a gene of interest[71].

However, the plasmid-based exogenous reporter systems largely ignore the comprehensive aspect of gene expression regulation by complex interaction between different epigenetic factors, transcription factors and various cis elements by artificially limiting investigation on a putative promoter region. To overcome this limitation of exogenous reporter system, we developed a novel tool where a reporter construct is tagged at the 3'end of SNCA endogenously, allowing us to monitor transcriptional activity of the gene keeping its epigenetic architecture unperturbed. The NanoLuc luciferase reporter used in this study is 100-fold brighter and significantly smaller in size than firefly or Renilla luciferase, thus making it an ideal tag for even low expressing genes[73,74]. Recent breakthrough in genome editing techniques like CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats) have made specific genome editing simple and scalable1[61]. Tagging SNCA endogenously with the NanoLuc using CRISPR/Cas9 method allows sensitive and real-time measurement of changes in transcriptional activity under various conditions of stimuli. This strategy can help to shed light on the transcriptional regulation of SNCA, and may serve as a very strong tool for screening of drugs to limit the progression of PD.

2.3. Results

2.3.1. Generation of a stable cell line endogenously tagged with functional NanoLuc luciferase reporter at the 3′end of SNCA

To introduce the NanoLuc reporter at the 3'end of the SNCA, a double-strand DNA break (DSB) was introduced on the reverse strand with the –NGG protospacer adjacent motif sequence

(PAM) directly abutting the stop codon located in the exon 6. This was achieved in human embryonic kidney cell line (HEK293T) by transient transfection with the pSpCas9 (BB)-2A-Puro vector[82]. Along with the CRISPR/Cas9 construct, a donor vector containing the NanoLuc sequence was co-transfected to take advantage of the cell's homology-directed repair (HDR)[83]. This donor construct contained two flanking homology domains each of about 800 base pairs, corresponding to the upstream and downstream of the DSB target site (Figure 4a). The NanoLuc sequence was cloned between these two domains to precisely introduce this reporter construct right before the stop codon of SNCA. Potential positive clones were identified by PCR amplification of the genomic DNA using the NanoLuc forward and reverse primers (Table 2) (Figure 4b; Figure 5). To confirm NanoLuc insertion at the target location in the SNCA gene, a second PCR using "Insertion Confirmation Primers" (Table 2) was performed and later sequenced (Figure 5). The wild-type (WT) allele generated a 280 base pair (bp) PCR amplicon while the NanoLuc-tagged allele generated a 805 bp band, indicating a heterozygous insertion of the reporter construct (Figure 4c). To overcome the PCR amplification bias towards the shorter allele, a separate amplification for NanoLuc-tagged allele was performed using a forward primer (NanoLuc Internal Forward Primer) on the NanoLuc insert in combination with the same reverse primer (cDNA sequencing Reverse Primer) on the 3'UTR, generated a comparable amplification of 356 bp product for NanoLuc insert (Figure 4c, lane 2). The PCR using NanoLuc internal forward primer failed to amplify any band in wild type HEK293T cells (Figure 4c, lane 4). A second potential positive clone was found to have an incomplete insertion of the NanoLuc reporter tag (colony 14, Figure 4b), and thus not used any further. To confirm

the expression of NanoLuc-tagged SNCA allele tagged with the NanoLuc in the cell line, hereafter referred to as 293T-SNCA-3'NL, RT-PCR was conducted using primers encompassing the entire coding region of the gene and the 3' UTR. The amplicon was then sequence verified to confirm the presence of NanoLuc insertion (Figure 4d, Figure 6).

Following confirmation of the NanoLuc insertion in the SNCA genomic region and the presence of mature mRNA, we sought to confirm protein expression and functional activity of the NanoLuc. Western blot analysis of cell lysates with a polyclonal anti αSyn antibody confirmed the presence of both wild type αSyn (~15 KDa) and a NanoLuc-tagged protein (~34 KDa) matching αSyn fused with the NanoLuc (19.1 KDa) (Figure 7a, Figure 8). We performed luciferase activity assay on 293T-SNCA-3'NL cells by measuring luminescence after addition of substrate furimazine. 293T-SNCA-3'NL cells produced a considerably high signal distinguishable from cells without the NanoLuc incorporation or 293T-SNCA-3'NL cells without furimazine (Figure 7b). Titration of cell counts from 2,500 to 50,000 produced a linear increase in luminescence activity ($R2 = 0.95$), indicating that the luminescence of the SNCA-tagged NanoLuc reporter is internally consistent (Figure 7c).

Taken together, these results show that SNCA was successfully tagged with the NanoLuc construct at the 3'end, and that expression of the NanoLuc-tagged allele leads to generation of a fusion protein.

2.3.2. αSyn-NanoLuc luciferase activity reflects SNCA transcriptional regulation

To validate whether this system would be able to monitor changes in endogenous SNCA transcription, 293T-SNCA-3'NL cells were treated with known epigenetic modulators like DNA methyltransferase 1(DNMT1) inhibitor (5-AzadC), histone deacetylase (HDAC) inhibitors (sodium butyrate) and also dopamine which may have a toxic effect beyond a certain threshold concentration[84]. SNCA harbors CpG islands at the regulatory regions encompassing the promoter and intron[79,84]. The CpG island in the intron1of SNCA in HEK293T remains completely methylated which upon demethylation can increase gene expression[79]. We treated the 293T-SNCA-3'NL cells with 5-AzadC for 72 hours to allow more than one round of cell division. A significant increase in the NanoLuc activity was observed in cells treated with 5-AzadC as compared to vehicle treated ones (3.68 times increase; $p \leq 0.0001$) (Figure 9a). SNCA transcript level from sister cultures correlated well with the observed increase in the NanoLuc activity (Figure 6a). Changes in methylation of the SNCA-intron1 CpG island were detected using bisulfite sequencing as done by Jowaed et al[78]. Ten clones from each sample were analyzed and a significant reduction in mean intron 1 methylation by 31.7% ($p < 0.05$) was observed (Figure 9b). The reduction of cytosine methylation in the intron1 positively correlated with increase in SNCA transcript, as we saw with the increased NanoLuc activity.

It has already been reported that dopamine at 100 µM concentration can enhance SNCA transcription in HEK293T cells without inducing subsequent toxicity[79,84]. Likewise, we treated

293T-SNCA-3'NL cells with 100 µM dopamine for 48 hours and a significant 1.31 times increase in the NanoLuc activity was observed as compared to the controls ($p < 0.0001$) (Figure 9c). Again this increase in the NanoLuc luciferase activity complied with an increasing trend in αSyn/NanoLuc mRNA and protein expression after dopamine treatment as seen by RT-PCR and western blot analyses (Figure 9c, Figure 10). Hyper-acetylation of histone is expected to unwind underlying DNA, which in turn favors transcription[66,85]. To test whether HDAC inhibition in 293T-SNCA-3'NL cells would faithfully monitor transcription in response to histone hyper-acetylation, cells were treated with sodium butyrate (class I and IIa inhibitor of HDAC) at concentrations 2.5 mM and 5.0 mM for 24 hours[86]. This treatment paradigm significantly increased the NanoLuc activity by 1.5 and 2.35 times respectively as compared to the controls ($p < 0.001$) (Figure 9d). αSyn/NanoLuc transcript levels also corroborated well with the activity measurement and showed a dose-dependent increase (Figure 9d).

2.3.3. Exogenous promoter reporter assays failed to reproduce transcriptional activation of SNCA as seen in endogenous conditions

Firefly luciferase-based promoter assay is considered a gold standard for assessing promoter activity of a target gene, which in turn reflects the transcriptional activity of the gene[87]. We compared the reporter activity of SNCA transcription between transient transfection-based luciferase system and the endogenous NanoLuc system that we designed. HEK293T cells were co-transfected with pGL3 basic plasmid containing SNCA promoter-intron1 region cloned

upstream of luciferase coding sequence and CMV-Renilla (transfection control). Twenty-four hours later, transfected cells were treated exactly with the same modulators of SNCA expression as described in the previous result. We observed a significant decrease in normalized reporter activity upon 5-AzadC treatment post 72 hrs ($p < 0.0001$) Figure 11a) contrary to the increased NanoLuc activity and transcript expression (Figure 6a). The CMV promoter-driven Renilla luciferase activity which was used as an internal control significantly varied upon 5-AzadC treatment, while the SNCA promoter-intron1 driven firefly luciferase activity remained largely unaffected, thereby leading to reduction in the normalized reporter activity (firefly/Renilla) (Figure 11a). Next, to compare the effect of dopamine on the exogenous luciferase activity driven by SNCA promoter-intron1 and the endogenous SNCA behavior, the transfected cells were treated with 100 µM dopamine for 48 hrs, exactly following paradigm followed for the 293T-SNCA-3'NL cells. Interestingly, we did not observe any significant change in the normalized luciferase activity ($p = 0.22$) (Figure 11b). In this treatment, no significant change was observed for either firefly luciferase ($p = 0.15$) or Renilla luciferase activity ($p = 0.16$) (Figure 11b), although an increasing trend could be seen for both. This data again failed to demonstrate the endogenous state of regulation upon dopamine treatment (Figure 11c).

We also investigated the effect of HDAC inhibition (for 24 hrs) on the SNCA promoter-intron1 driven luciferase activity. Similar to 5-AzadC treatment, normalized firefly activity showed a significant decrease from control ($p < 0.0001$) (Figure 11c). This time we saw a significant

increase in the firefly luciferase activity ($p < 0.05$), along with a significant increase in the Renilla activity ($p < 0.001$) thereby causing an artefactual reduction in SNCA promoter activity after normalization ($p < 0.0001$). This observation is again opposite to what we found with the NanoLuc-based endogenous system after addition of HDAC inhibitor (Figure 11d). Together, these observations showed that the exogenous luciferase-based reporter system largely failed to demonstrate original state of endogenous transcriptional activity of SNCA.

2.4. Discussion

In the present study, we have successfully incorporated the NanoLuc reporter construct at the 3'end of SNCA in HEK293T cells using CRISPR/Cas9 technology and monitored changes in expression induced by two epigenetic modulators and dopamine which are known to deregulate the gene's expression. The newly emerging endogenous reporter system represents a significant paradigm shift in the study of gene regulation and may provide new and exciting opportunities for both basic and translational research. Such strategies allow to insert a reporter directly into the targeted genome, enabling us to investigate endogenous gene regulation while keeping the epigenetic structure intact.

This particular feature is extremely relevant in studying SNCA expression, as this gene has been shown to get extensively regulated by its epigenetic structure[77-79]. Moreover, αSyn is a molecule whose level of expression is directly correlated with the severity of the PD

pathogenesis, thus making this tool very helpful for developing potential therapeutic options[12,30,75,76]. Recent studies have demonstrated that epigenetic regulation by 5-AzadC increases SNCA expression in SK-SN-SH and SH-SY5Y cells[78,88]. This cytidine analogue passively demethylates by irreversibly trapping DNA methyltransferase I (DNMTI), which positively correlates with higher expression of various genes[89]. SNCA harbors CpG islands at the regulatory regions encompassing the promoter and intron[79,84]. It has been shown that hypomethylation of the intron1 CpG island but not the promoter CpG island is strongly associated with PD[78]. So, we sought to check the methylation status of the 23 CpG sites[78] in the SNCA-intron1 by bisulfite sequencing. As expected, a decrease in methylation of intron1 CpG correlated with an increase in transcript levels as measured by the NanoLuc activity. Conventionally, to study the effect of methylation on the promoter's transcriptional efficiency, that promoter-reporter construct is fully methylated in vitro and the reporter activity is then compared with the unmethylated one[78,90]. Consistent with our observation for endogenous αSyn with NanoLuc activity, Jowaed et al. showed that in-vitro methylation of regulatory promoter-reporter construct of αSyn reduces transcription of αSyn. Studying gene expression this way could lead to different outcome than actual transcriptional state of the target gene as the regulatory region of that gene might exist as partially methylated condition endogenously, which cannot be replicated using the exogenous system. Moreover, comprehensive regulation of the target gene promoter with inputs coming from other local epigenetic modifications such as histone post translational modifications (PTMs)[91] and trans factors which are usually present

endogenously, might not be able to regulate the exogenously introduced promoter of the same gene.

During genetic typing, we saw that the intensity of the NanoLuc-tagged allele after PCR is relatively weaker when compared to the wild-type PCR product for both the genomic DNA and cDNA (Figure 4c, Figure 6). This may be a result of PCR bias towards the amplification of the shorter allele over the longer ones involving reaction in a single tube[92]. This problem of preferential amplification was overcome by using separate primer set to amplify only the NanoLuc-tagged allele (not encompassing the wild-type allele) which could amplify the NanoLuc-tagged allele very efficiently and comparably to the wild-type allele (Figure 4c, Figure 6). In addition to using separate set of primers, we used equal number of cycles (30 cycles) to ensure that the PCR products are not saturated, which is indicated by the increase in expression after dopamine treatment. However, the western blot shown in the Figure 5a indicated that the NanoLuc-tagged αSyn protein (band shown at ~34 KDa) is higher in intensity than the wild type protein (band at ~15 KDa) when probed with αSyn specific antibody. This unequal distribution of protein bands between wild type (low) and NanoLuc-tagged allele (high) may indicate that either that the NanoLuc luciferase has been targeted to αSyn in more than one allele of the locus, or it is also possible that fusion of a highly stable NanoLuc luciferase protein with αSyn may affect the stability of the target protein positively. HEK293T cells are not typically diploid, and are instead complex hypo-triploid in nature, containing less than three times the number of

chromosomes of a normal diploid human cell[93]. Our results show that the NanoLuc reporter construct is precisely targeted to the end of SNCA and faithfully reports the transcriptional changes of αSyn.

The advantage of using an endogenous reporter system to study regulation of transcription by epigenetic environment of the gene was highlighted by the data obtained from conventional reporter assays after treatment with different drugs such as 5-AzadC, dopamine and sodium butyrate. As 5-AzadC inhibits DNA methylation in dividing cells therefore we previously observed that the drug reduced endogenous SNCA-intron1 methylation and increased transcription or NanoLuc activity in our endogenous reporter system. However, we did not observe such increase in exogenous SNCA-promoter/intron1 firefly reporter activity upon drug treatment, may be due to lack of DNA methylation in the exogenous plasmid (Figure 11a first panel). Surprisingly, we observed a significant increase in Renilla luciferase activity upon 5-AzadC treatment (Figure a second panel) and a significant decrease in normalized reporter activity (Figure 11a third panel). However, the observed decrease in reporter activity can be attributed to CMV-Renilla luciferase activity which, although equally transfected, significantly varied upon 5-AzadC treatment. The mechanism of this apparent variation in Renilla luciferase activity is yet unknown. However, it is known that Renilla luciferase activity can vary significantly upon treatment paradigm depending on the promoter driving its activity. This could lead to erroneous interpretation of the observed data after normalization to Renilla as is

a standard protocol for these type of assays[94]. Therefore it was suggested to use different or multiple endogenous controls in case of employing exogenous reporter systems[94].

It was also shown that dopamine can enhance αSyn transcription in HEK293 cells[79]. Similarly, we also observed a significant increase in the NanoLuc activity upon dopamine treatment to 293T-SNCA-3'NL cells (Figure 9b) unlike exogenous promoter-reporter assay (Figure 11b), which further fortifies the reliability of our endogenous tagging system. This apparent discrepancy between the outcome coming from endogenous and exogenous systems may be attributed to either lack of appropriate crosstalk between cis/trans elements in exogenous reporter system or due to lack of methylation structure in the SNCA promoter-intron1 construct in the plasmid-based luciferase system. Since, the aim of our study was to design a tool that can faithfully monitor changes in endogenous αSyn transcription at physiological level, the question that we seek to address was whether this NanoLuc-tagged αSyn reflected accurately the change in transcription of wild-type αSyn (although present in higher copy) after treatment with dopamine. The comparable increasing trend of mRNA and protein of both the wild-type and the tagged αSyn after treatment with dopamine (Figure 10) indicated that the engineered cell line could efficiently replicate transcriptional changes of endogenous wild-type αSyn.

Next we investigated how HDAC inhibitor, sodium butyrate can regulate transcription of SNCA. As the dynamics between histone acetylation or chromatin relaxation and histone

deacetylation or chromatin condensation play an important role in regulation of gene transcription[95,96], it can be envisaged that the application of HDAC inhibitor would result in hyper-acetylated condition in the gene promoter, which in turn might favor induction in transcription. As expected, we observed a significant increase in the NanoLuc activity in response to sodium butyrate ($p < 0.001$) (Figure 9d). However, a significant opposite observation was noticed using exogenous reporter system (Figure 11c). This prominent decrease in the normalized reporter activity can again be attributed to the significant increase of Renilla luciferase activity following sodium butyrate treatment. Interestingly, it has been shown that this HDAC inhibitor can change the overall chromatin structure of the cell and functions through the butyrate response elements on the gene regulatory regions that encompass Sp1/Sp3 binding sites[97]. Since, SNCA regulatory region contains multiple Sp1/Sp3 binding sites, it is reasonable to assume that recruitment of acetylated Sp1/Sp3 can enhance SNCA expression. Analysis of CMV promoter sequence revealed that it harbors around 5 cognate sequences for Sp1/Sp3 binding, suggesting a possible mechanism for sodium butyrate mediated significant increase in the Renilla luciferase activity. Since, Renilla luciferase plasmid is usually used as endogenous control for this kind of conventional luciferase assay, it is very important to interpret the reporter assay results with caution using this type of transfection controls[94].

Transcriptional upregulation of SNCA has long been a concern in PD pathogenesis. Most intellectual effort has been focused on aggregation behavior, resulting in a lacuna of information on transcriptional regulation of this gene despite its immense importance. In this study we developed a novel screening tool that can efficiently monitor SNCA transcript levels under different treatment conditions known to up-regulate transcription. This tool can provide a new diagnostic platform for drug development and testing of compounds believed to regulate SNCA in the cell in an inexpensive and precise way. As the epigenetic environment for this gene regulation is kept unchanged, the effects of treatments would more closely mimic the state seen in the cells than has previously been available. It is also worth mentioning that the stability and sensitivity of this NanoLuc luciferase reporter makes it suitable to monitor very low expressing genes which is unlikely to be achievable by any other conventional reporter systems. At this juncture it is also important to mention that our study highlights the importance of using endogenous reporter system over exogenous reporter system particularly for studying comprehensive epigenetic regulation of gene transcription. This SNCA-NL reporter system is very useful to study the effects of drugs or agents that are known to modulate gene's chromatin structure or any other epigenetic environment. Therefore previous studies which used exogenous firefly/Renilla-based reporter system to demonstrate the important cis elements of the SNCA promoter or novel transcription factors and pathways for the gene are still useful[98-101].

2.5. Methods

2.5.1 Cell Culture

HEK293T LVX cells were maintained in DMEM (Dulbecco's Modified Eagle Medium) supplemented with heat-inactivated 10% fetal bovine serum and 1% Penicillin/Streptomycin (Gibco, 10000 U/mL). Cells were maintained at 37 °C in humidified incubators with 5% CO2 and passaged following trypsinaization with 0.25% Trypsin/0.53 mM EDTA (ThermoFisher Scientific).

2.5.2. Designing SNCA specific short guide RNA (sgRNA)

The vector for cloning of the sgRNA pSpCas9 (BB)-2A-Puro (PX459) was a gift from Feng Zhang Lab (Addgene plasmid # 48139)16. The sgRNA targeting the stop codon area of the SNCA gene was ligated into the pSpCas9(BB)-2A-Puro vector as previously described with minor modifications (Table 1). Briefly, sgRNA oligos corresponding to 20 base pairs immediately upstream of –AGG PAM sequence located on the reverse strand at the SNCA stop codon were ligated into pSpCas9(BB)-2A-Puro vector using Fast-Digest BbsI (ThermoFisher Scientific) and T7 ligase (New England Biolabs) by cycling between 37 °C and 21 °C 10 times for 6 minutes. Ligation reaction was treated with Plasmid-Safe ATP-Dependent DNase (Epicentre) and transformed into CaCl2 competent cells DH5α E. coli by heat shock (42°, 50 sec). Plasmid was purified using Gene-Jet plasmid Miniprep Kit (ThermoFisher Scientific) according to manufacturer's protocol and successful insertion of the sgRNA was confirmed by sequencing.

Cloning of NanoLuc-homology donor vector

Primers used for amplification, addition of restriction sites and insert confirmation are listed in Table 2. Approximately 800 bp-length homology domains were amplified from HEK293T genomic DNA using Q5-Polymerase (New England Biolabs). Upstream homology arm (Chromosome 4, 89,727,231–89,728,021, reverse strand GRCh38:CM000666) was modified to include a 5′ NotI restriction site and a 3′ SacI restriction site (Figure 4). Downstream homology arm (Chromosome 4, 89,726,457–89,727,235, reverse strand GRCh38:CM000666) was modified to include a 5′ HindIII cut site and 3′ NotI/AatII cut site. The NanoLuc luciferase sequence was amplified from pNL1.1 vector (a gift from Promega Corporation). To allow for sequential plasmid ligation, NanoLuc coding sequence was modified to include terminal 5′ SacI and 3′ HindIII/AatII restriction enzyme sites. Upstream homology arm and the NanoLuc were ligated together and cloned into NotI and AatII digested pGEM -T Easy vector (Promega Corporation; cat no. A137A) and transformed into competent DH5α cells. Plasmid was purified using Gene-Jet Plasmid Miniprep Kit according to manufacturer's protocol (ThermoFisher Scientific). Downstream homology arm was then cloned into AatII-digested vector. Completed homology sequence was digested with NotI to release sequence from pGEM -T Easy vector and delete AatII sequence, then ligated into NotI digested pAAV-IRES-hrGFP backbone (2,846 bp) and transformed into chemically competent DH5α cells. Presence of insert was confirmed by sequencing.

2.5.3. Generation of HEK-293T cell line stably expressing SNCA-NanoLuc (293T-SNCA-3′NL)

Vector constructs were transfected into HEK293T LVX (CloneTech) cells using X-Fect Polymer (CloneTech) according to manufacturer's protocol. Briefly, HEK293T LVX cells were seeded in a 6-well plate with 1 × 106 and allowed to grow to 80% confluency. X-Fect Polymer was mixed with 1.25 µg pSpCas9 (BB)-2A-Puro with SNCA sgRNA and 1.25 µg NanoLuc-homology arm donor vector and incubated with cells for 4 hours, followed by a change with fresh media. After 48 hours, cells were subjected to 5 µg/mL puromycin (ThermoFisher Scientific, cat no. AC227420100) selection for 48 hours. Surviving puromycin resistant cells were diluted to a single cell level and plated in 96 well plate then allowed to propagate for approximately two weeks. 15 pure colonies were recovered and grown to 50% confluency before passaging them to 1:2 in a 24 well plate.

2.5.4. Confirmation of stable integration of the NanoLuc reporter at the 3′end of SNCA

Genomic DNA of the puromycin resistant colonies was extracted by 16 hours incubation in lysis buffer (10 mM Tris-HCl pH 7.6, 0.5 mM EDTA, 0.67% SDS, 132 µg/ml Proteinase-K) at 55 °C, followed by precipitation in 100% ethanol with 150 mM NaCl. Presence of the NanoLuc insert was confirmed by PCR using NanoLuc forward and reverse cloning primers and identity of insert was confirmed by gene specific confirmation primers in intron 5′ and the 3′ UTR of SNCA using PCR master mix (GenDepot, P0311-200). The details of the "Insert Confirmation primers" are listed in Table 1. Individual bands were excised and sequenced to confirm the sequence to

ensure introduction of any mutations and location of insert in the genomic DNA and cDNA (Macrogen USA).

2.5.5. Western blotting

For Western blotting, protein was extracted using RIPA (Radio Immuno Precipitation buffer) buffer (PBS, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% PMSF, 100 ng/ml protease inhibitor, dH2O) and 40 µg protein were ran on 10% denaturing SDS gel and transferred to PVDF membrane (Millipore cat no. IPFL00010). Following the transfer process, the membrane was fixed in 0.4% PFA in PBS as suggested by Lee et al.36. The membrane was then blocked with Odyssey Blocking Buffer (LI-COR cat no. 927-50000) mixed 1:1 with TBS. Specific protein bands were detected using rabbit anti-αSyn antibodies (Santa Cruz Biotechnology, cat no. SC-7011R, 1:1000 dilution) overnight and followed by goat anti-rabbit secondary antibodies for fluorescent detection at 680 nm (LI-COR, cat no. 926-68020). Western blotting for the experiments related to 48 hour dopamine treatment, proteins were extracted as before and transferred to PVDF membrane. Membrane was blocked with 5% milk in TBS-T. Specific protein bands were detected using mouse anti-αSyn antibodies (BD Transduction Laboratories, cat no. 610786) at 1:250 dilution, followed by goat anti-mouse HRP secondary antibodies for chemiluminescent detection (Jackson Immuno Research laboratories Inc. cat no. 115-035-146).

2.5.6. Assay for the NanoLuc luciferase activity

To measure the NanoLuc luciferase activity, 200,000 cells were seeded per well of a 24 well plate. After 24 hours, appropriate treatments were done and incubated for designated periods of time in duplicate. Cells were detached by trypsinization and two wells were combined in a 1.5 mL microcentrifuge tube and pelleted at 2,000 × g for 3 minutes. Cell pellets were resuspended in 500 µL colorless DMEM and counted using an automated cell counter (BIORAD, USA) three times and then luciferase activity was monitored in technical triplicate using the Nano-Glo Luciferase Assay System according to manufacturer's protocol with minor modifications (Promega corporation, cat no. N1110). Briefly, 5,000 cells in 30 µL colorless DMEM were assayed using 30 µL of assay buffer mixed 1:100 with NanoGlo substrate. Plates were incubated for 5 minutes and then luminescence was recorded in triplicate on a multi-plate reader (EnVision, PerkinElmer). This experiment was performed three times independently.

2.5.7. SNCA promoter-reporter assay

To generate the SNCA promoter luciferase construct, SNCA promoter-intron1 region (−2,200 to +118 bp; with respect to ATG) was amplified from HEK293T genomic DNA and cloned into XhoI/HindIII sites of promoter less pGl3 basic vector (Promega, USA, cat no. E1751). The presence of insert was validated by sequencing. For the promoter-reporter assays, 200,000 HEK293T LVX cells were co-transfected with SNCA-pGL3 and CMV-pRL (as a transfection control) vectors in a 24 well plate format. Each well were co-transfected with 500 ng of SNCA-

pGL3 and 10 ng of CMV-pRL plasmids using X-Fect polymer as described above and incubated 24 hours before proceeding to any chemical/drug treatments. Following the drug treatments for appropriate time HEK293T LVX cells were collected for lysis. Briefly, cell pellet was lysed in 100 µL of Lysis Buffer (25 mM Tris-Phosphate Buffer pH 8.0, 4 mM EGTA, 2 mM DTT, 20% Glycerol & 1% Triton X-100) for 10 minutes at room temperature with occasional shaking, then centrifuged at 16,000 × g for 5 minutes and the lysate was collected in a separate tube. 10 µL of this lysate was assayed with 140 µL of luciferase assay buffer (25 mM Tris-Phosphate Buffer pH 8.0, 4 mM EGTA, 20 mM MgSO4, 1 mM DTT, and 2 mM ATP). Firefly luciferase activity (0.75 mM luciferin, 10 mM DTT) and Renilla luciferase activity (1.5 µM Coelenterazine, 100 mM NaCl, 25 mM Tris pH 7.5) were measured in triplicate using a multi-plate reader (EnVision, PerkinElmer). Relative luciferase activity was measured by normalizing the firefly luciferase activity to Renilla luciferase activity (F/R).

2.5.8. Cell treatment paradigm

All the treatments were performed in duplicate wells. Compounds used in the study are as follows: Sodium butyrate (Alfa Aeser; cat no. A11079), (5-AzadC) (Sigma, cat no. A3656) and Dopamine hydrochloride (Sigma, cat no. H8502). Sodium butyrate stock was prepared in distilled water at a concentration of 500 mM and cells were treated with sodium butyrate for 24 hours at 2.5 mM and 5.0 mM. 5-AzadC stock of 5 mM was prepared in 50% acetic acid and immediately aliquoted and frozen. Cells were treated with 10 µM 5-AzadC for 72 hours,

refreshed every 8–12 hours8. Dopamine stock was prepared in distilled water at a concentration of 10 mM. Cells were treated with 100 µM dopamine for 48 hours, refreshed every 24 hours9. Control wells were treated with respective vehicles for indicated times and no visible cellular toxicity was noticed under aforementioned treatment paradigms.

2.5.9. Bisulfite sequencing

Genomic DNA from 293T-SNCA-3'NL cells following 5-AzadC treatment was extracted as mentioned above. Approximately, 500 ng high quality DNA (260/280 > 1.8) was used for bisulfite conversion using EZ DNA Methylation-Direct Kit (Zymo Research cat no. D5020) then used as template for PCR amplification of SNCA intron1 region using specific primers designed for bisulfite modified DNA as described in article by Jowaed et al.8. EpiMark Hot Start Taq DNA Polymerase (NEB Inc; cat No. M0490S) was used for PCR and the amplicon was cloned into pGEM -T Easy vector (Promega Corporation). 10 positive colonies were selected for sequencing analysis using T7 universal primer. The sequenced products were analyzed for degree of methylation using BISMA and QUMA software[102,103].

2.5.10. Semi quantitative reverse transcriptase PCR (RT-PCR)

RNA was extracted using TRIzol Reagent according to manufacturer protocols (Life Technologies Inc.; cat no. 15596-026). Complementary DNA (cDNA) was generated by conversion of 1 µg total RNA using amfiRivert cDNA Synthesis Platinum Master Mix according to manufacturer's

protocol (GenDEPOT; cat no. R5600-50). The cDNA was diluted 1:1 with nuclease free water before PCR. To check the expression of SNCA, amplification was done using primers against the NanoLuc sequence and normalized by the amplification of an endogenous control gene, β-actin. The scheme of the differential amplification of NanoLuc-tagged allele and wild type SNCA alleles are shown in Figure 11. Details of all the primers are listed in the Table 1.

2.5.11. Statistical analysis

Statistical analysis was performed using GraphPad Prism 5 (GraphPad Software Inc.). Data are presented as mean ± SEM. To get statistically meaningful data, all experiments were performed in technical triplicates. The times change in the NanoLuc luciferase activities in treated groups were calculated by normalizing the value with respective control for each experiment. All the experiments were independently repeated three separate times using separate batches of cells and the means of each experimental set were analyzed. In case of exogenous luciferase reporter assays, minimum of four independent repeats were performed. Statistical significance was determined by comparing means of different groups and conditions using unpaired 2-tailed Student's t test, and one-way ANOVA. Multiple corrections were made using post-hoc Tukey test, Bonferroni test and Scheffe's test whenever it was required. Significance was assessed at 95% level.

2.6. Figures

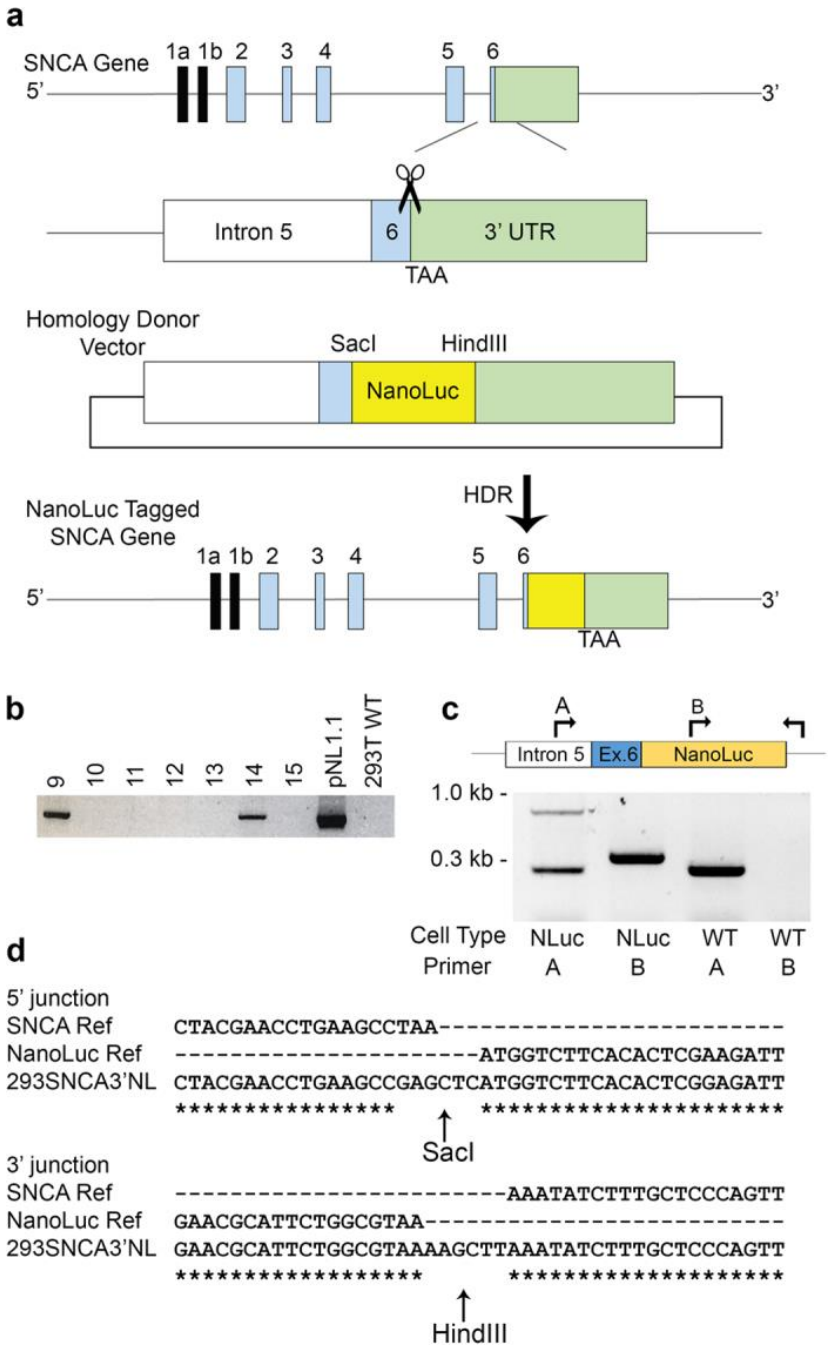


Figure 4 – Development of 293TSNCA3'NL cells

Figure 4: a Schematic representation of cloning strategy. *SNCA* gene map showing exons (1a and 1b non-coding, 2–6 coding) and the 3'UTR. Transfection of sgRNA targeting the 3'end of exon 6 induces a DSB near the stop codon (TAA). Donor vector design contains 5' homology arm of 790 bp encompassing part of intron 5 and exon 6 upstream from the stop codon and the NanoLuc-3' homology arm of 800 bp downstream of the stop codon containing part of the 3'UTR. Co-transfection of donor vector with the CRISPR/Cas9 construct precisely incorporated the NanoLuc right before the stop codon by HDR of the *SNCA* gene. **b** Following puromycin selection and single cell dilution, genomic DNA from all surviving isogenic colonies were screened for the NanoLuc insert with pNL1.1 NanoLuc vector and HEK293T LVX cells as controls. From 15 colonies recovered, two were positive for the NanoLuc insertion. **c** Gene specific PCR with primers in the intron 5 (A) and the 3'UTR of *SNCA* showed colony 9 had a heterozygous insertion in 293T-SNCA-3'NL cells (Lane 1); PCR with forward primer on the NanoLuc (B or NanoLuc Internal Forward Primer) and the same 3'UTR reverse primer (cDNA sequencing Reverse Primer) showed comparable amplification of the NanoLuc tagged allele (Lane 2); PCR of the wild-type αSyn and NanoLuc from the HEK293T LVX as controls (Insertion Confirmation Forward Primer and cDNA sequencing Reverse Primer) (Lanes 3 and 4). **d** Excerpt of Sanger sequencing results showing insertion of the NanoLuc sequence with restriction sites precisely before the stop codon of *SNCA* and with correct continuation of the 3'UTR after the NanoLuc sequence.

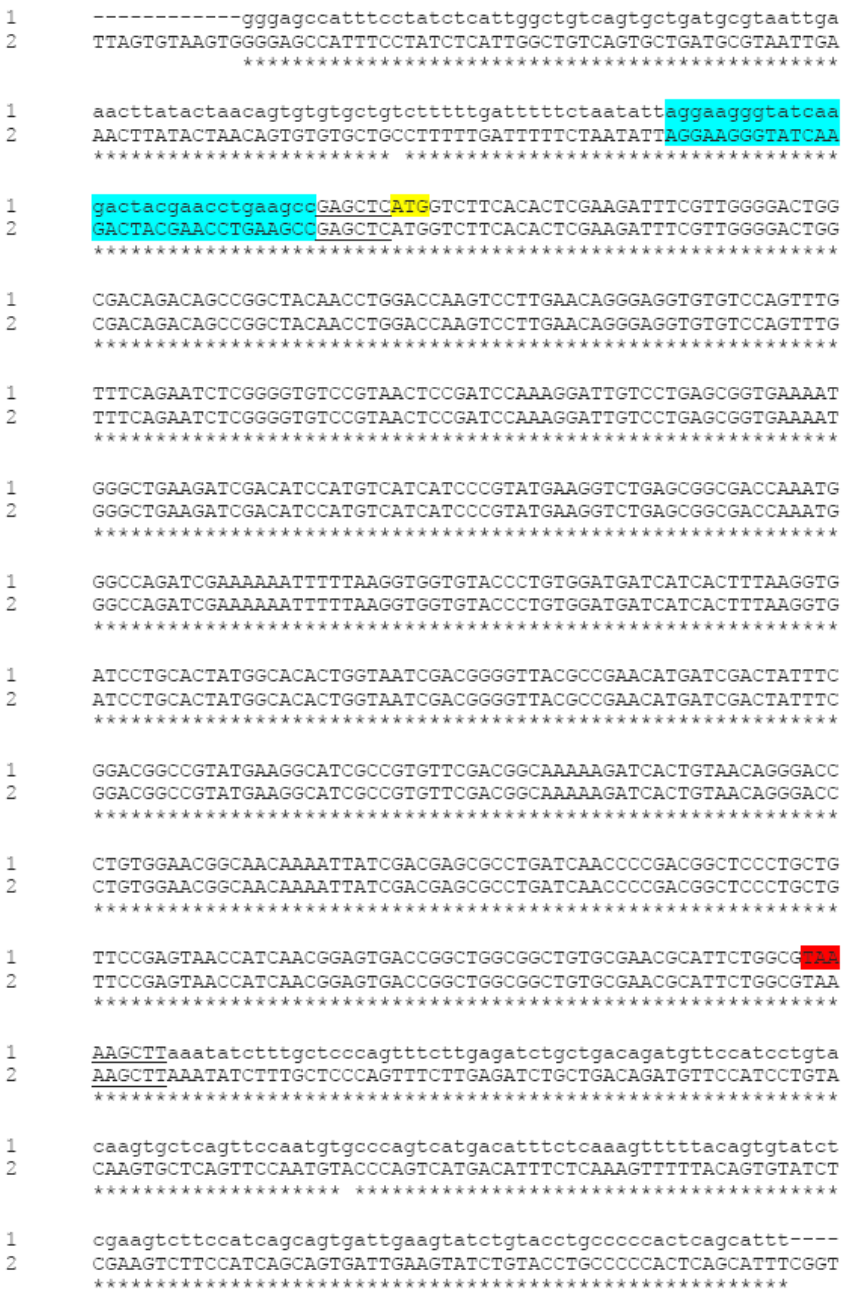
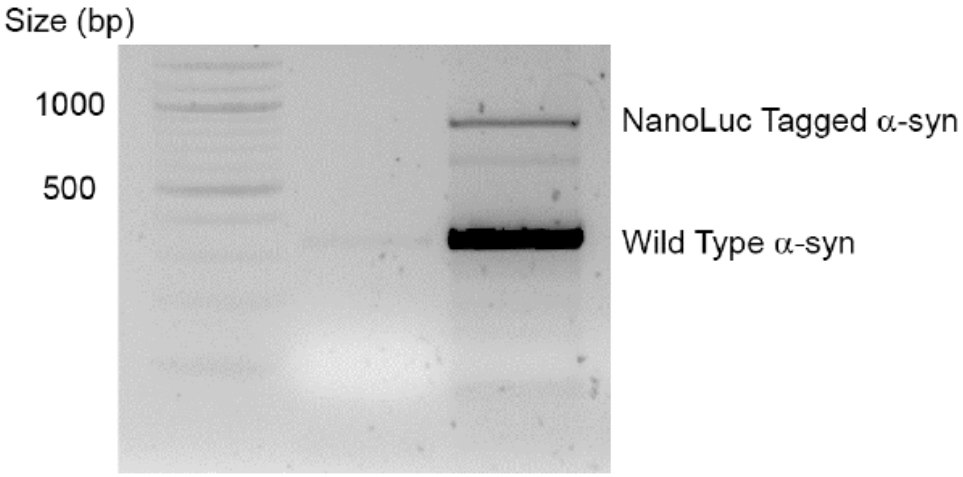


Figure 5 – Sequence alignment of wild type SNCA against 293T-SNCA-3'NL genomic DNA sequence

Figure 5: Part of the sequence alignment is shown depicting the 3' end of the *SNCA* of 293T-SNCA-3'NL. The wild type (WT) *SNCA* genomic DNA sequence (reference sequence: GRCh38.p7) was aligned with the sequencing data from 293T-SNCA-3'NL to check for the proper incorporation of the NanoLuc construct at the end of the coding sequence of *SNCA*. Sequence 1 is expected *SNCA* genomic DNA, sequence 2 is 293T-SNCA-3'NL sequence. The last exon of *SNCA* is highlighted in blue. The start and stop codons of NanoLuc are highlighted by yellow and red colors respectively. The aligned bases between both the sequences are marked by star (*). The in-frame incorporation of the NanoLuc can be seen in between the two restriction enzyme sites, SacI and HindIII (underlined) in 293T-SNCA-3'NL cells (2). Sequence alignment was done using Clustal Omega program.

a



b

```
1       TGGTGTAA-AGGAATTCATTAGCCATGGATGTATTCATGAAAGGACTTTCAAAGGCCAAG
2       TGGTGTAAAGGAATTACATTAGCCATGGATGTATTCATGAAAGGACTTTCAAAGGCCAAG
        ********  * * * ********************************************

1       GAGGGAGTTGTGGCTGCTGCTGAGAAAACCAAACAGGGTGTGGCAGAAGCAGCAGGAAAG
2       GAGGGAGTTGTGGCTGCTGCTGAGAAAACCAAACAGGGTGTGGCAGAAGCAGCAGGAAAG
        ************************************************************

1       ACAAAAGAGGGTGTTCTCTATGTAGGCTCCAAAACCAAGGAGGGAGTGGTGCATGGTGTG
2       ACAAAAGAGGGTGTTCTCTATGTAGGCTCCAAAACCAAGGAGGGAGTGGTGCATGGTGTG
        ************************************************************

1       GCAACAGTGGCTGAGAAGACCAAAGAGCAAGTGACAAATGTTGGAGGAGCAGTGGTGACG
2       GCAACAGTGGCTGAGAAGACCAAAGAGCAAGTGACAAATGTTGGAGGAGCAGTGGTGACG
        ************************************************************

1       GGTGTGACAGCAGTAGCCCAGAAGACAGTGGAGGGAGCAGGGAGCATTGCAGCAGCCACT
2       GGTGTGACAGCAGTAGCCCAGAAGACAGTGGAGGGAGCAGGGAGCATTGCAGCAGCCACT
        ************************************************************

1       GGCTTTGTCAAAAAGGACCAGTTGGGCAAGAATGAAGAAGGAGCCCCACAGGAAGGAATT
2       GGCTTTGTCAAAAAGGACCAGTTGGGCAAGAATGAAGAAGGAGCCCCACAGGAAGGAATT
        ************************************************************

1       CTGGAAGATATGCCTGTGGATCCTGACAATGAGGCTTATGAAATGCCTTCTGAGGAAGGG
2       CTGGAAGATATGCCTGTGGATCCTGACAATGAGGCTTATGAAATGCCTTCTGAGGAAGGG
        ************************************************************

1       TATCAAGACTACGAACCTGAAGCCTAA----------------------------
2       TATCAAGACTACGAACCTGAAGCCGAGCTCATGGTCTTCACACTCGAAGATTTCGTT
```

Figure 6 - 293T-SNCA-3'NL cDNA indicates correct insertion of NanoLuc sequence.

Figure 6: a PCR of SNCA from 293T-SNCA-3'NL cDNA using "cDNA sequencing primers" (Table 1) produces two bands, one matching expected wild-type size and one matching expected size for wild-type with NanoLuc insertion (NanoLuc Tagged αSyn). **b** Part of the sequence alignment is shown depicting the coding region of the gene and incorporation of the NanoLuc sequence in 293T-SNCA-3'NL cell line. The WT SNCA mRNA sequence (reference sequence: NM_007308.2, transcript variant 4) was aligned with the sequencing data from 293T-SNCA3'NL to check for the proper incorporation of the NanoLuc construct at the end of the coding sequence of SNCA. Sequence (1) is WT SNCA mRNA transcript, sequence (2) is 293T-SNCA3'NL sequence. The start and stop codons in the WT SNCA mRNA sequence are highlighted by green and red colors respectively. The aligned bases between both the sequences are marked by star (*). The in-frame incorporation of the NanoLuc can be seen in between the two restriction enzyme sites, SacI and HindIII (underlined) in 293T-SNCA-3'NL cells (2). The NanoLuc insert is unique to 293T-SNCA-3'NL cell line and not found in WT SNCA mRNA sequence. The new stop codon after the NanoLuc sequence in 293T-SNCA-3'NL cell is marked by yellow. "N" denotes unread base in the sequence. Sequence alignment was done using Clustal Omega program.

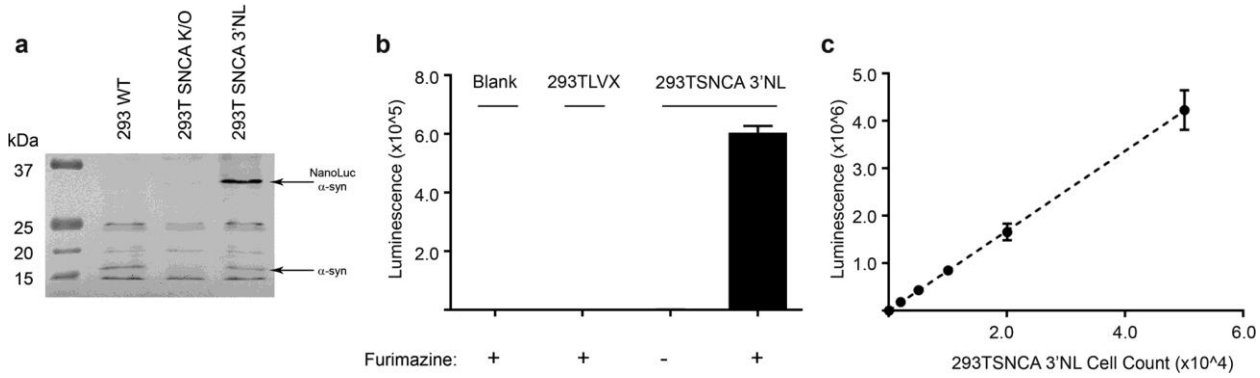


Figure 7 – Functional expression of the NanoLuc luciferase.

Figure 7: **a** Lysates from HEK293T LVX, HEK293T *SNCA* K/O, and 293T-*SNCA*-3'NL cells were Western blotted with anti-αSyn antibody. As expected, a wild-type band appeared in HEK293T LVX and 293T-*SNCA*-3'NL cells, but in the 293T-*SNCA*-3'NL cells an additional band was identified at approximately 34 kDa corresponding to αSyn fused with the NanoLuc that was absent in non-NanoLuc tagged cells. **b** Luminescence activity of HEK293T LVX cells compared with 293-*SNCA*-3'NL cells. Only 293T-*SNCA*-3'NL cells generate significant NanoLuc activity in the presence of furimazine as the substrate. **c** Luminescence of 293T-*SNCA*-3'NL cells follows a linear trend with increasing cell numbers ($R^2 = 0.95$). All the experiments have been performed in triplicate.

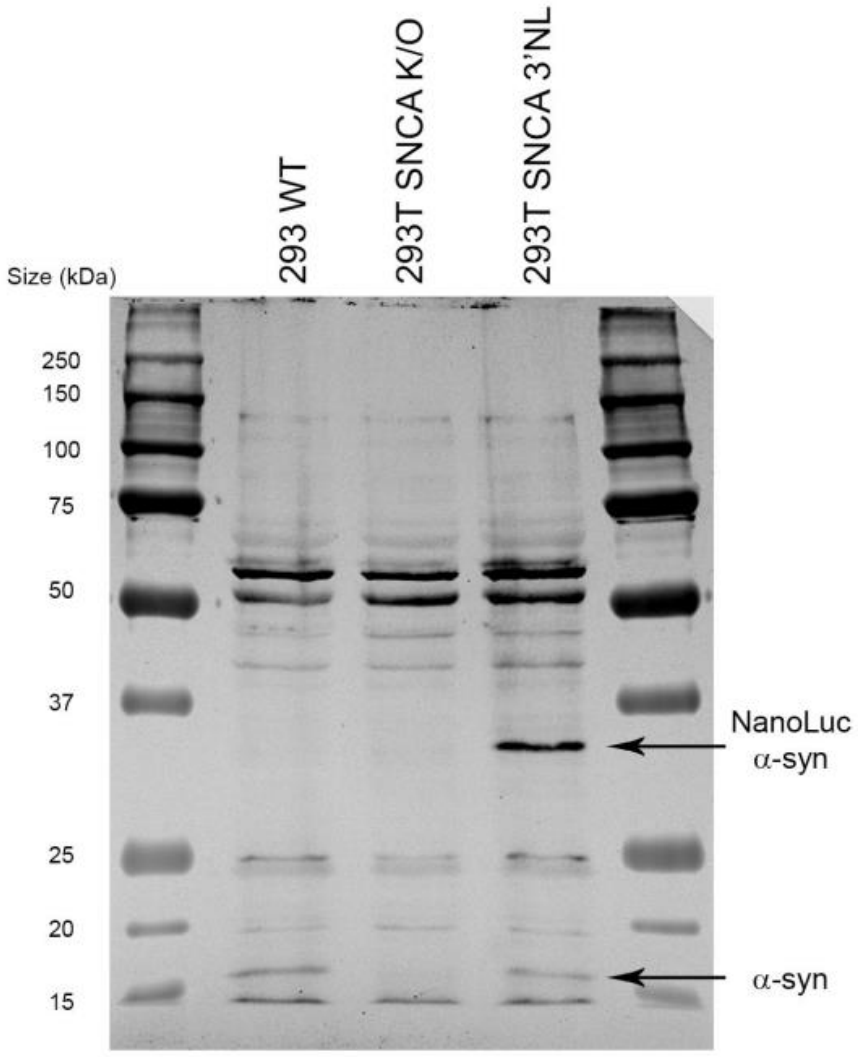


Figure 8 – Full western blot of 293T-SNCA-3'NL cells showing NanoLuc tagged and WT αSynuclein.

Figure 8: Full western image from Figure 5a. αSYN was detected using polyclonal rabbit antibody (Santa Cruz, SC-7011-R). Wild-type αSYN is shown at approximately 15 kDa and is present in both wild-type 293T cells and 293T-SNCA-3'NL cells, but absent from 293T-SNCA-knockout cells. NanoLuc-fused αSYN is shown at approximately 34 kDa, and only appears in 293T-SNCA-3'NL cells. The unmarked bands present in the blot are non-specific bands that commonly appear with this particular polyclonal antibody.

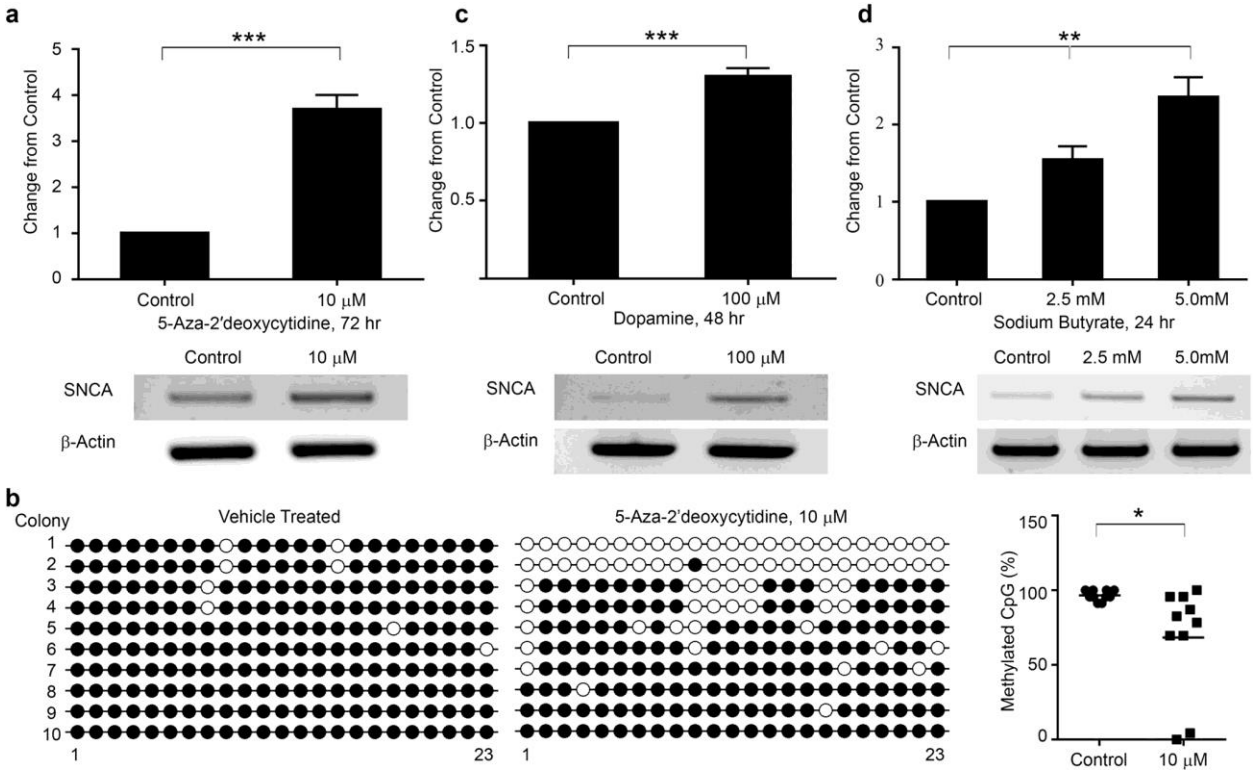


Figure 9 – 293TSNCA3'NL cells having the NanoLuc integration can be used to model deregulated SNCA as seen in sporadic PD

Figure 9: a 293T-*SNCA*-3'NL cells treated with 10 µM 5-AzadC for 72 hours which induced a significant increase in the NanoLuc activity as compared to the control. The NanoLuc activity was corroborated by increase in αSyn transcript as shown in the RT-PCR. **b** Methylation status of 23 CpG sites on the *SNCA* intron1 was determined by bisulfite sequencing. Amplified PCR products were cloned into pGEM-T Easy vector and 10 clones were sequenced. (Left) Comparison of vehicle and 5-AzadC treatment (10 µM, 72 hours) showing unmethylated (open circles) and methylated (closed circles) cytosines for all 10 clones (y-axis) at each of the 23 CpGs in intron1 (x-axis). (Right) Scatter plot showing overall decrease in methylation by 31.7% compared to the control. **c** Similarly, 293T-*SNCA*-3'NL cells treated with dopamine at 100 µM concentration for 48 hours, increased NanoLuc activity significantly. Increase in the NanoLuc activity was confirmed by RT-PCR after dopamine treatment. **d** Following HDAC inhibitor (sodium butyrate) treatment at concentrations of 2.5 mM and 5.0 mM for 24 hours, the 293T-*SNCA*-3'NL cells showed a significant dose dependent increase in the NanoLuc activity. This dose-dependent increase in the NanoLuc activity was also confirmed by RT-PCR following same treatment paradigm. β-actin amplification was used as an internal control for all the PCRs. Error bars show the mean from three technical repeats. p values are given for t-test (5-AzadC, dopamine), one-way ANOVA (Sodium butyrate) where *represents $p < 0.05$, **represents $p < 0.01$, ***represents $p < 0.0001$.

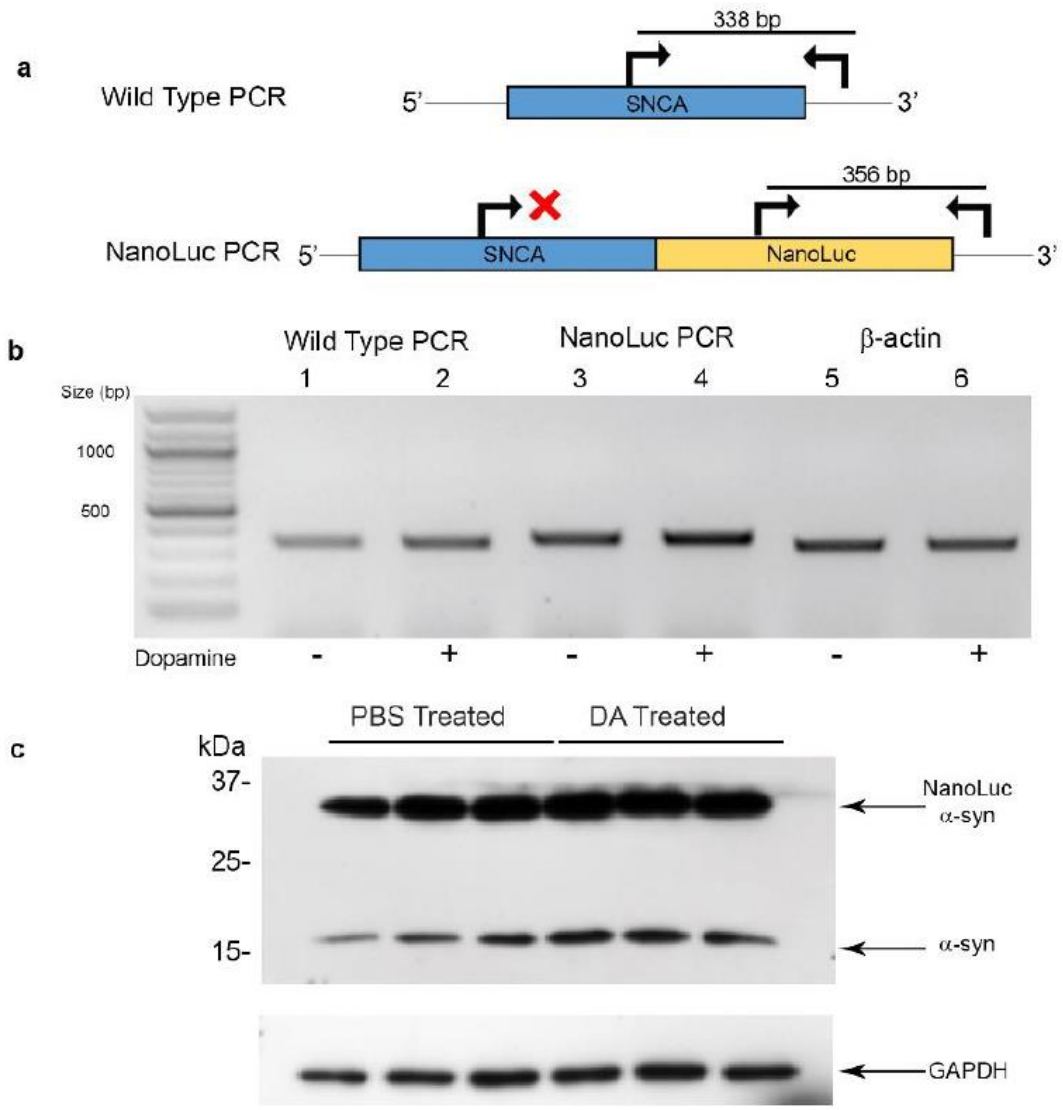


Figure 10 – Comparable increase of endogenous and NanoLuc-tagged SNCA levels following dopamine treatment

Figure 10: **a** Schematic representation of PCR strategy to avoid preferential amplification issues from heterozygous allele sizes. For WT amplification, primers used were "*SNCA* Exon 4 Forward" and "Insert Confirmation Reverse" (Table 1) to give an amplification product size of 338 bp. For NanoLuc specific amplification, primer pair used were "NanoLuc Internal Forward" and "Insert confirmation reverse" to get a product size of 356 bp. Elongation time was restricted for NanoLuc specific amplification to prevent competitive wild-type amplification. **b, c** Qualitative image of RT-PCR and western blot analyses of 293T-*SNCA*-3'NL cDNA and protein under vehicle and dopamine (100 uM) treated condition show comparable increasing trend for both wild-type and NanoLuc tagged αSym.

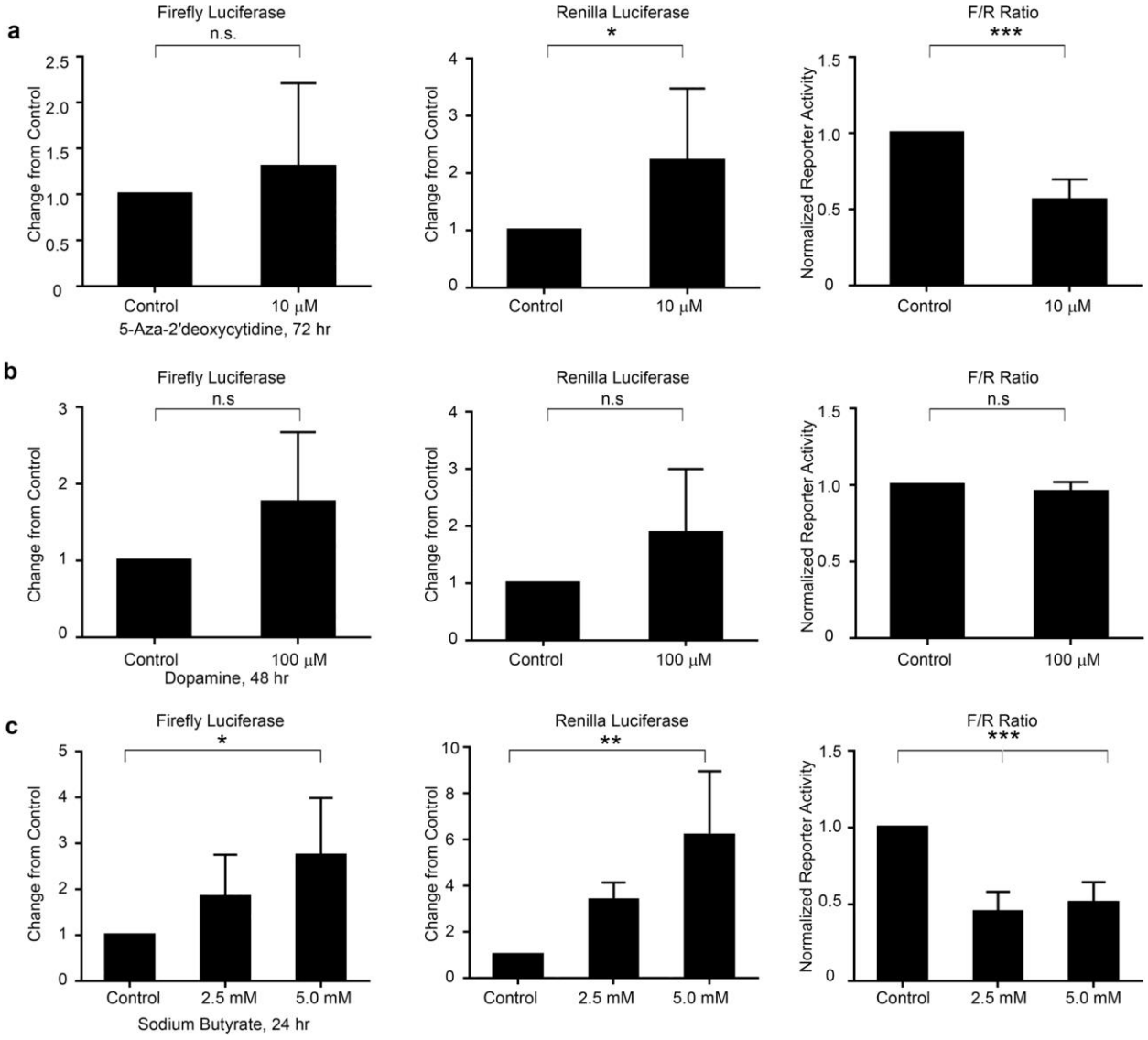


Figure 11 – Exogenous overexpression of SNCA driven firefly and Renilla luciferase failed to replicate endogenous SNCA behavior after similar paradigm of drug treatment.

Figure 11: **a** HEK293T LVX cells were co-transfected with *SNCA*-pGL3 (*firefly* luciferase) and CMV-pRL (*Renilla* luciferase) constructs. Twenty-four hours after transfection, cells were subjected to 10 µM 5-AzadC treatment for 72 hours. No significant change in *firefly* luciferase activity was seen, however the *Renilla* luciferase activity showed a significant increase. Thus overall normalized luciferase assay showed a significant decrease in the activity (F/R) after treatment with 5-AzadC. **b** Similarly, HEK293T LVX cells were also co-transfected with *SNCA*-pGL3 and CMV-pRL plasmids, treated dopamine at concentration of 100 µM for 48 hours. No significant change was observed either in *firefly* luciferase activity or *Renilla* luciferase activity. Thereby, the normalized luciferase activity (F/R) also did not show any significant change after treatment with dopamine. **c** Treatment with sodium butyrate at concentrations of 2.5 mM and 5.0 mM respectively for 24 hours, gave comparable results (like 5-AzadC) in normalized luciferase activity (F/R) which showed a significant decrease. This decrease in the normalized luciferase was attributed to the significant increase in *firefly* luciferase activity and *Renilla* luciferase activity. Error bars show the mean from three technical repeats. p values given for t-test (5-AzadC, dopamine), one-way ANOVA (Sodium butyrate) for three independent experiments where *represents $p < 0.05$, **represents $p < 0.01$, ***represents $p < 0.0001$.

Table 2 – Primers Referenced in Chapter 2

Primer Name	Sequence 5' - 3'
SNCA sgRNA Top Strand	CACCGTGGGAGCAAAGATATTTCTT
SNCA sgRNA Bottom Strand	AAACAAGAAATATCTTTGCTCCCAC
Upstream Homology Forward Primer	TATGGCGGCCGCTTAGGAACAAGGAAAAT
Upstream Homology Reverse Primer	AGTGAGCTCGGCTTCAGGTTCGTAGTC
NanoLuc Forward Primer	AAAGAGCTCATGGTCTTCACACTCGAA
NanoLuc Reverse Primer	CATGACGTCAAGCTTTTACGCCAGAATGCGTG
Downstream Homology Forward Primer	AAGCTTAAATATCTTTGCTCCCAGTTTCTTGA
Downstream Homology Reverse Primer	GTTGACGTCGCGGCCGCATACCAAAACA
Insertion Confirmation Forward Primer	CTGCAGAATATTTGCAAAAACATTGATTG
Insertion Confirmation Reverse Primer	TAAAAACTTTGAGAAATGTCATGACTGGG
cDNA sequencing Forward Primer	GGAGTGGCCATTCGACGACAGTG
cDNA sequencing Reverse Primer	TAAAAACTTTGAGAAATGTCATGACTGGG
SNCA Exon 4 Forward Primer	CAAATGTTGGAGGAGCAGTGGTGA
NanoLuc Internal Forward Primer	AAGGTGATCCTGCACTATGGCA

CHAPTER THREE – CRISPR-MEDIATED EPIGENOMIC ENGINEERING

We used the CRIPSR/Cas9 system to insert an endogenous reporter into the *SNCA* gene to generate a reporter by taking advantage of its double strand break capability to induce homology-directed repair. While the direct editing of DNA by Cas9-mediated cleavage is a major use of this technology, the highly specific targeting capabilities of the sgRNA allows a broad range of other chromatin and genomic manipulation. Here we will discuss the different ways in which the Cas9 platform can be used to affect DNA without double-stranded breaks including manipulation of gene expression, DNA labeling and conformational control, and epigenomic engineering. Application of these concepts allowed us to develop and submit a publication describing our modular epigenetic toolkit that allows us to directly write epigenetic information in a site-specific manner (See Chapter 5).

3.1. Using dCas9 to Manipulate DNA

After being directed to the DNA strand by a targeting sgRNA, Cas9 melts the DNA duplex and generates a double stranded break by nicking each DNA strand[58]. Introduction of a point mutation in both the RuvC and HNH domains generates catalytically dead Cas9 (dCas9) that can complex with guide RNA and bind target genomic targets without introducing a DSB at the site[104]. Fusion of different enzymes to dCas9 allows many different types of targeted chromatin editing to occur (Figure 12).

3.1.1. dCas9-based Tools

Using dCas9 as a platform for deaminase enzymes such as APOBEC1 allows direct conversion of C to T or A to G in a site-specific manner, allowing correction of single-nucleotide polymorphism (SNP) mutations or creating an early stop codon[105-107]. The organization of chromatin can be visualized in live cells by fluorescently labeling dCas9 or by using an sgRNA aptamer scaffold to bind multiple MS2 fluorophores[108,109]. Interestingly, tethering dimerizable proteins to dCas9 scaffolds allows researchers to force formation of chromatin loops to increase or prevent enhancer activity on local promoters[110]. Finally, fusion of dCas9 with epigenetic modifying enzymes allows researchers to directly write and edit epigenetic information on the chromatin and influence gene expression[60].

3.1.2. Using dCas9 to Modify Transcription

A wide variety of CRISPR-based tools have been generated to affect gene expression and the epigenome (Figure 12). Most simply, genes can be directly downregulated by targeting dCas9 to the polymerase binding site where it can sterically hinder RNA polymerases, an approach known as CRISPRi[104]. This technique was taken further with the fusion of the transcriptional repressor KRAB-box associated protein 1 and heterochromatin protein 1 to further reduce targeted gene expression[111,112]. Conversely, multiple studies demonstrate that fusion of dCas9 with four copies of the 16-amino acid Herpes simplex transactivation domain (VP64) can robustly induce transcription[62,113,114]. Further work has expanded this system by combining VP64 fusion with other chimeric constructs such as the NF-kB transactivation domain P65 or adapting the sgRNA scaffold to form hairpin apatmers that can recruit activation domains[115-117].

Also, Tanenbaum et al developed the SunTag system consisting of a repeating polypeptide array of GCN4 moieties that allow multiple effector molecules to be co-localized with a targeting protein[118]. Effector molecules are fused to a scFv antibody fragment targeting the GCN4 moiety, recruiting the effectors to the site of the base protein. By fusing the GCN4 tail to dCas9 and attaching VP64 to the scFv fragments, they demonstrated robust transcriptional activation.

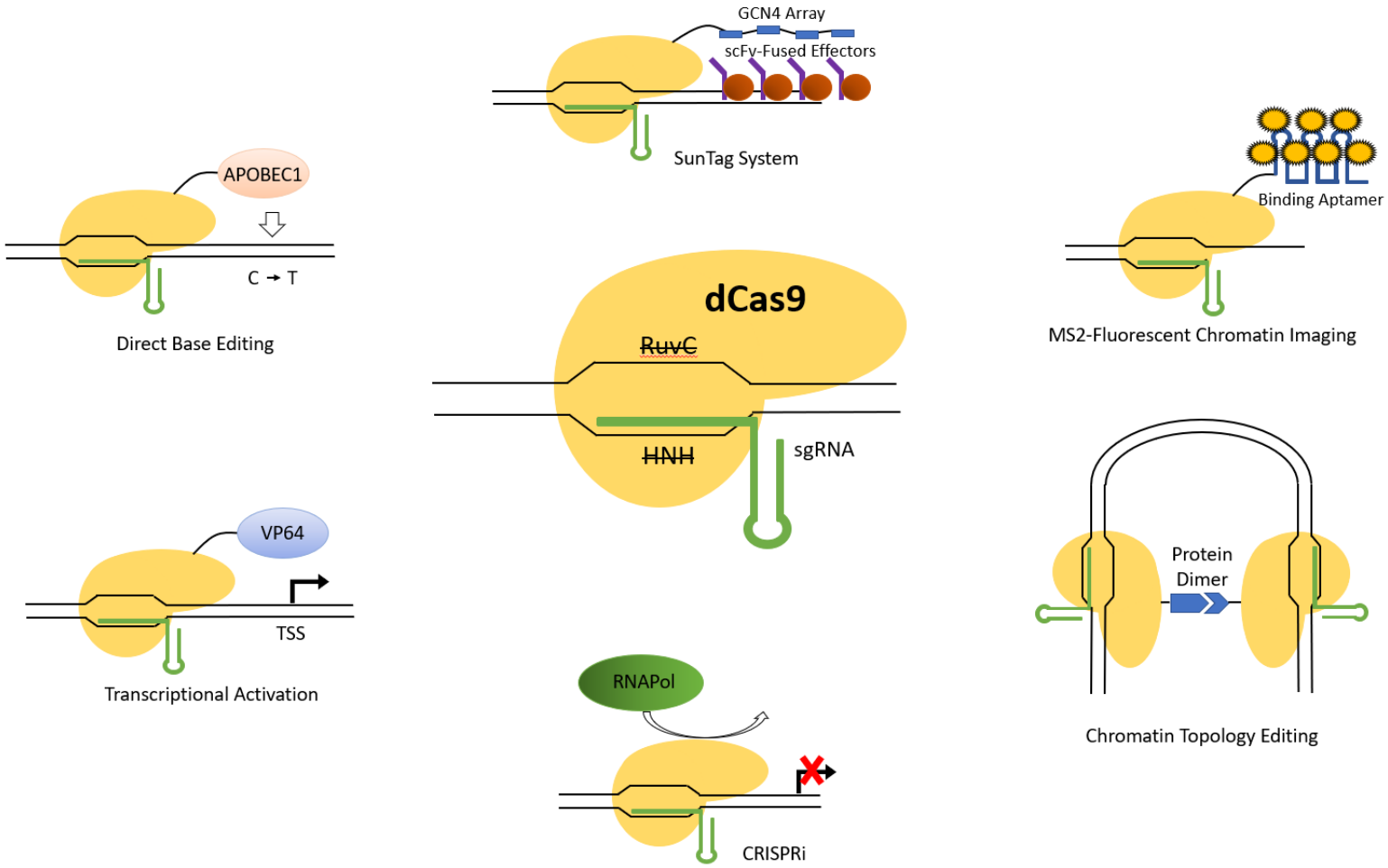


Figure 12 – Using dCas9 for Genomic Engineering

Figure 12: Catalytically dead Cas9 (Center) has been fused to a wide variety of effector molecules, enabling a broad spectrum of genomic effects, such as base editing, chromatin imaging, altering chromatin topology, and affecting transcription.

3.2. CRISPR-mediated epigenome editing

While the previously described work has been able to directly alter gene expression, increasing attention has been focused on the role of epigenetics in gene expression and using the CRIPSR/Cas9 system to understand and edit epigenetic architecture. Broadly speaking, epigenetics can be defined as any heritable change in gene expression that is not directly related to the actual DNA sequence. A large variety of epigenetic mechanisms exist, such as non-coding RNA, methylation of cytosine bases in DNA, and post-translational modification of histones[60]. Among these, DNA methylation is the most widely studied mechanism of epigenetic control. There are two major enzymes that catalyze *de novo* DNA methylation, DNMT3A (DNA Methyltransferase 3) and DNMT3B[119]. Generally, DNA methylation at promoter regions or other regulatory regions is associated with transcriptional repression. Aberrant DNA methylation has been associated with many types of cancer and small molecule methylation inhibitors such as 5-Azacytidine are FDA approved, although this is a global effect and targets the entire genome in a non-specific manner[120]. To circumvent this, the catalytic domain of DNMT3A was fused to dCas9 to allow for target-specific DNA methylation by several independent research groups[121-123]. Similarly, DNA methylation was reduced in a site-specific manner by fusion of the TET (Ten-Eleven Translocase) enzymes to dCas9[124-126].

Once it was established that DNA methylation could be influenced in a site-specific manner, researchers also began to target histone post-translational modifications as a strategy to alter

epigenomic activity. In the eukaryotic cell, DNA is wrapped around these proteins. Each histone is a dimer composed to two hetero-tetramers with 4 subunits: Histone protein H2A, H2B, H3, and H4[127] (Figure 13). The histone functions as a spool which is able to organize, compact and regulate DNA accessibility. One of the major regulatory mechanisms of DNA control is the post-translational modifications of the long protein tails that extend from each histone core protein. While each protein tail has different regulatory functions, we will focus on two major regulatory lysine residues on histone protein H3.

3.3. Using dCas9 to Manipulate Histones

Lysine 4 on histone H3 (H3K4) is a critical site for marking active regions of transcription. H3K4 can be appended with mono-, di-, or tri-methylation (me1, me2, me3)[127]. Active distal regulatory elements are generally marked by H3K4me1 or H3K4me2[60]. H3K4me3 generally indicates an active promoter region or one that is 'poised' – meaning it contains both permissive and inhibitory marks simultaneously and is frequently present at high levels around transcriptional start sites[128]. H3K27 is also a central residue for epigenetic information. When acetylated, histones adopt an open conformation, exposing the DNA for recruitment of transcriptional machinery leading to increased transcription[128,129]. Increases in H3K27ac are found at both active distal enhancer regions as well as promoters. Conversely, when H3K27 is tri-methylated, histones adopt a closed conformation and the DNA has reduced accessibility to transcriptional activation[128]. (Figure 13). It should be noted that none of these marks are sole

determinants of transcriptional activity and the specific combination of DNA methylation, histone marks, and cellular context determines the actual transcriptional activity[60]. Despite this complexity, the H3K4 and H3K27 residues remain attractive targets for CRISPR-based editing strategies.

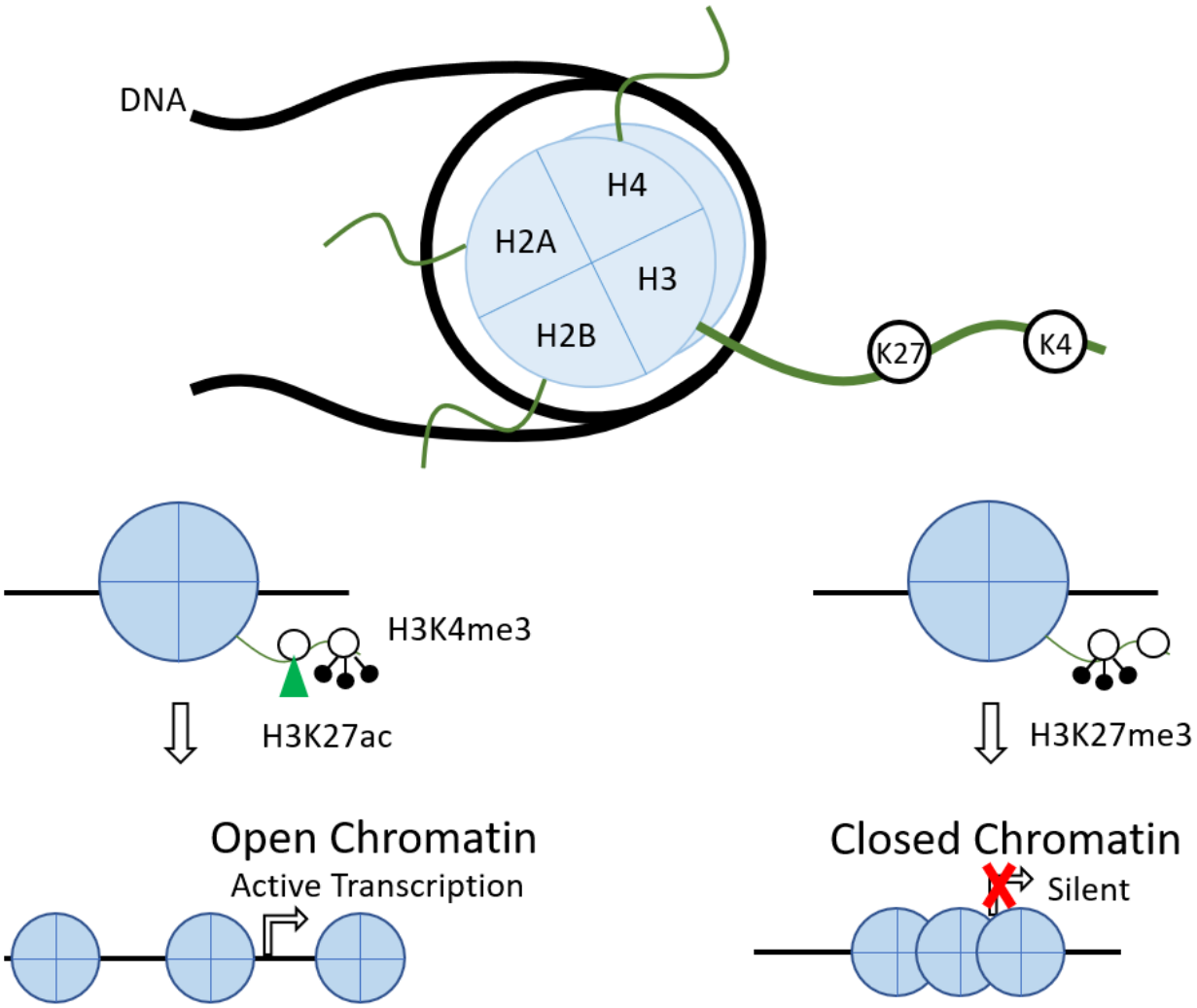


Figure 13 – H3 Histone Tails Influence Transcription

Figure 13 – Histones consist of two tetramers of histone proteins (H2A, H2B, H3, H4) that DNA is wrapped around. Modifications on the tail of H3 at K27 and K4 can modify the chromatin structure to allow or block transcription factors and RNA polymerase to initiate transcription.

3.3.1 – Current Approaches to Modify Histone Marks

While the role of histone marks is now well established, relatively little intellectual effort has been focused on careful manipulation of histone marks in a site-specific manner. Fusion of the histone H3K4me2 demethylase LSD1 to Cas9 has been reported and shown to reduce target gene expression by removing activating K4 methylation[130]. Conversely, depositing the transcriptionally activating mark H3K27ac by fusion of dCas9 with P300 has been shown to increase target expression[131]. H3K27ac was also targeted for removal by fusion of histone deacetylase 3 (HDAC3) to dCas9, which reduced both acetylation levels as well as expression[132]. Finally, dCas9 fusion to the methyltransferase PRDM9 has been shown to modulate local H3K4me3 levels and allow re-expression of silenced genes in several cell types[133]. While these works are important first steps in expanding our ability to manipulate and understand local epigenetic effects, much work remains to be done. Identifying a causal link between gene expression and epigenetic architecture remains a central focus in chromatin biology and using the Cas9 platform to deeply explore and understand the specific effects of locus specific mark manipulation will be a crucial step toward realizing this goal. To assist in answering these central questions, during my dissertation period our group has worked to develop a suite of modular epigenetic modulators to aid increase our understanding of focal epigenetic effects. (See Chapter 4)

CHAPTER FOUR: PRECISE EPIGENOMIC EDITING BY A NEWLY DEVELOPED MODULAR EPIGENETIC TOOLKIT

Subhrangshu Guhathakurta[#]; Levi Adams[#]; Anishaa Sivakumar; Mingyu Cha; Mariana Bernardo Fiadeiro ; Haiyan Nancy Hu; Yoon-Seong Kim*

[#] contributed equally

4.1. Abstract

Epigenetic regulation of gene expression is one of the master regulators in the body to control gene-transcription. Therefore, any disease-associated gene whose transcriptional deregulation is responsible for the disease pathology could be rectified by locus-specific modulation of that gene's epigenetic niche. In this study, we created a novel CRISPR/dCas9-based epigenetic toolkit, which can be used to modulate gene-specific epigenetic environment to ameliorate its expression associated with any disease. We have shown usefulness of this system using several key enzymes that regulate histone post-translational modifications associated with different states of gene expression which could be deployed to alter any gene's expression.

4.2. Introduction

Utilization of the CRISPR/Cas9 system is now well established for precise genome editing[50,112,115,131,134-141]. Taking advantage of the genomic locus-specific targeting abilities with catalytically dead Cas9[142] (dCas9), this system has also been employed to modulate focal epigenetic architecture and gene expression[60]. Early transcription modulation systems using dCas9 fused to synthetic activators such as VP64, SAM, and VPR have exhibited to alter a specific gene expression. But the effect is transient[143] and, moreover, they ignore the endogenous epigenetic environment. However, each gene's regulatory regions such as enhancer, promoter or gene body have unique combinations of different histone posttranslational modifications (PTM) that predominantly determine their status of expression. Therefore, to understand the precise epigenetic regulation of a gene, we need to develop a system which can systematically explore the effect of every such histone PTM of the gene on their expression with a high resolution. To address this gap, we developed a modular epigenetic toolkit by modifying previously developed SunTag framework[118]. dCas9 tagged with a repeating polypeptide array of GCN4 moieties recruits multiple anti-GCN4 scFv-fused endogenous epigenetic effectors JMJD3, EZH2, PRDM9, p300 and JARID1A. Here, we reported that this switchable modular system efficiently works in precisely modifying each histone PTM of any gene with a couple of hundred-base resolution. This CRISPR/dCas9-based modular epigenetic toolkit would pave a new avenue to study the epigenetic architectures of an individual gene with great precision.

4.3. Results

We envisioned a system where highly specific changes to individual genes can be achieved using a suite of epigenetic modifiers that adjusts the endogenous structure while minimally disturbing other loci. To this end, we took advantage of the previously developed SunTag system[118].

4.3.1. Developing the Modular Epigenetic Toolkit

To recruit multiple functional domains of epigenetic writers, we used dCas9 with 5x GCN4 moieties (maintaining 22 amino acid separation as previously reported[124]). This allows us to recruit multiple anti-GCN4 antibody-fused epigenetic effectors to the sgRNA-directed loci and directly modulate the local histone PMTs (Figure 14a). We generated five plasmids constructs with scFv and sfGFP fused to different key epigenetic effectors (Figure 14b, Figure 17 S1a,b;). Histone acetyltransferase p300 catalyzes the acetylation at the 27th lysine residueon histone H3 (H3K27ac)[144-146]. JMJD3 (Jumonji domain-containing protein D3) and EZH2 (Enhancer of zeste homolog 2) decreases and increases H3K27 tri-methylation (H3K27me3), respectively [147-151]. We also generated constructs to decrease or increase H3K4 tri-methylation (H3K4me3) using JARID1A (Jumonji, AT-rich interactive domain 1) and PRDM9 (PR/SET Domain 9) [152,153], respectively. Previously, it was shown that only using the catalytic core of an enzyme with dCas9 is more effective[131], so we subcloned the catalytic core of each epigenetic writer into the SunTag-scFv system (Figure 17 S1c) to develop our modular epigenetic toolkit.

4.3.2. The Modular Epigenetic Toolkit Directly Edits Histone Marks in a Targeted Manner

To assess the efficiency of the modular epigenetic toolkit, we analyzed relevant epigenetic marks in specific genes in the well-characterized A549 cell line [32,81]. We data mined available epigenetic information around ± 2.5 kb of transcriptional start sites, sorted genes for epigenetic marks, and selected genes whose promoter region epigenetic signatures correlate well with their expression profile (Figure 18 S2a). We selected four genes with distinctive patterns of histone modifications that would demonstrate the toolkit's flexibility—RPLP0, FTL, DLX5 and NEUROG2 (Figure 18 S2a,b), and confirmed the presence of these marks in our cells (Figure 18 S2c). For each gene, we chose a sgRNA at the region of the gene enriched for the relevant epigenetic target (Table 3). We generated A549 cell lines stably expressing dCas9-5xGCN4 and the desired sgRNA and confirmed the presence of dCas9-5xGCN4 at the sgRNA target site. (Figure 19 S3a). The appropriate dCas9-5XGCN4/sgRNA cell lines were transfected with the scFv-fused epigenetic writers (p300, JMJD3, EZH3, JARID1A or PRDM9) or empty scFv, FACS sorted followed by ChIP analysis, indicating that our epigenetic modulators effectively alter the epigenetic landscape of the targeted genes (Figure 14c, Figure 19 S3b). PRDM9 increased DLX5 H3K4me3 enrichment by 6.9 folds, conversely, JARID1A decreased RPLP0 H3K4me3 enrichment by 11 folds. p300 led to 5.8-fold increase in DLX5 H3K27ac. EZH2 increased H3K27me3 in FTL regulatory region from below detection limit by, while JMJD3 reduced Neurog2 H3K27me3 by 5.3 folds (Figure 14c). This data shows that our modular epigenetic toolkit system target-specifically modulates endogenous histone marks. By directly targeting the native epigenetic

architecture, we may be able to reveal new insights into endogenous gene regulation and inform future decisions about where to target epigenetic writers to achieve maximum effect.

4.3.3. Using the Toolkit to Modulate Pathological Gene Activity

To test this idea, we screened a large gene promoter region with JARID1A and EZH2 and compared it to the established HEK293T epigenetic landscape. The promoter region of *SNCA*, a gene encoding αSynuclein, spans over 3200 base pairs; to thoroughly investigate this area, we selected 10 sgRNA targets spaced every 200-400 base pairs (Figure 15, Figure 20, Table 3). Previous HEK293T data shows *SNCA* promoter enrichment for H3K4me3 as well as H3K27ac and low H3K27me3 levels (Figure 15b) [32,81]. We hypothesized that targeting our modular epigenetic toolkit to the sites of histone mark enrichment would allow us to modulate *SNCA* expression and provide insight into the functional importance of epigenetic architecture with a hundred-base resolution.

4.3.4. Screening *SNCA* with NanoLuc

For initial screening of the *SNCA* promoter we used the NanoLuc cell line we previously developed (see Chapter 2). The high sensitivity of this cell line and its ease of use made it a good fit for this type of screening. To screen this, we selected 8 sgRNA spanning the promoter region: one at each previously identified H3K4me3 and H3K27ac peak, and several more distal sites to get a complete picture of the epigenetic effects and range (Figure 15a,b). We

transfected the dCas9-5xGCN4 construct, JARID1A-scFv, and sgRNA into NanoLuc cells in a 96 well plate. After 72 hours we performed a NanoLuc assay to determine what effects modulation by JARID1A would have on the NanoLuc cell line.

As expected, removal of H3K4me3 by JARID1A significantly reduced αSynuclein expression compared to control but surprisingly the effective sgRNA targets did not directly align with the H3K4me3 enrichment peaks previously reported (Figure 15 b,c). Instead, maximal impact on expression occurred when sgRNA was targeted approximately 400-800bp upstream and downstream from reported enrichment peaks and targeting the area immediately around the peak enrichment had no effect. EZH2 was also able to significantly reduce expression of *SNCA* when targeted to the 5' end of the promoter but had no effect in other regions (Figure 15d). We also noted that the actual reduction in NanoLuciferase activity was relatively weak, although the mixed nature of the cell culture could be masking the real effect.

4.3.5. High Resolution *SNCA* Screen Reveals Effective Targeting Regions

To further explore this data, we decided to further investigate this promoter using a FACS sorter to select only cells that were successfully transfected. We transfected sgRNA with JARID1A, EZH2 or control constructs into HEK293 cells that stably express dCas9-5xGCN4. After 72 hours, we FACS-sorted cells and compared expression of αSynuclein by qRT-PCR in control (empty-scFv) to JARID1A-scFv. We also added additional sgRNA targets near the areas that

appeared most affected in the NanoLuc screen to increase resolution around those target areas.

We found similar trends using the FACS/qPCR method as we found using the NanoLuc screen, although some key differences (Figure 16). As expected, sorting out only cells that had been transfected increased our ability to detect reductions in the level of α-synuclein. We also found some intriguing differences between JARID1A and EZH2 effective areas. The action of JARID1A but not EZH2 at sgRNAb and sgRNAc underscores the highly selective nature of these modifications in gene expression (Figure 16 c,d). Notably, the same effectors exhibited dramatically different effects when the sgRNAs were only 200bp apart (sgRNA h,i,j), highlighting the ability of our system to enact change at high resolution determined only by the availability of specific sgRNA sites. Each cell type has a specific epigenetic landscape for any given gene, for example, *SNCA*'s epigenetic landscape is little different between neuronal and non-neuronal cells, and this may influence the extent of effect by modular epigenetic toolkit as well as selection of optimal targeting sgRNA.

4.4. Discussion

This system has the potential to make highly specific changes to the epigenetic architecture of any gene targetable by dCas9, and focused targeting of specific endogenous epigenetic modifiers may prove to be an effective strategy for persistently altering pathologic transcriptional activity. Recently there has been increased focus on using specific gene targeting as a therapeutic avenue for genetic disease. This system offers a strong tool for to dissect and understand underlying epigenetic architecture and opens potential new avenues for therapeutic strategies for various disease conditions.

4.5. Methods

4.5.1. Cloning

The template plasmids were purchased from Addgene. The pCAG-dCas9-5xPlat2AflD plasmid (Addgene #82560 a gift from Izuho Hatada; [124]). The lentiGuide-puro (Addgene #52963; a gift from Feng Zhang lab[154]) was used for subcloning all the guide RNAs used in the study. For sub-cloning catalytic domains of different epigenetic writers to ScFv expressing plasmid, pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (Addgene #60910 a gift from Ron Vale[155]). Details for catalytically active domains for all enzymes used in this study are listed in Supplementary Figure 1b. Enzyme sequences were amplified by PCR with *RsrII* sites and 5' and 3' linker sequences, both inserts were cloned into pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS using *RsrII* (NEB) to remove the VP64 fragment. Empty constructs were generated by plasmid digestion with *RsrII* and ligation. To allow for lentivirus production, each vector was subcloned into pLVX-dsRed

(Clontech) using *Xma/NotI* restriction sites to remove DsRed, PGK promoter, and Puromycin^r from the parent vector. We also transferred the dCas9-5xGCN4 construct into the pLVX background in a similar manner. All the constructs were sequence verified by multiple internal primers to make sure inserts are free from any unintended mutations. We confirmed expression by transiently transfecting HEK293 cells with each vector and immunoblotting against GFP (ThermoScientific MS-1315-P0). pcDNA-dCas9-p300 Core was a gift from Charles Gersbach[131]. pCMV-HA-JMJD3 was a gift from Kristian Helin[156] (Addgene plasmid # 24167). pGEX-EZH2 was a gift from Mien-Chie Hung[157] (Addgene plasmid # 28060). pcDNA3/HA-FLAG-RBP2 was a gift from William Kaelin[158] (Addgene plasmid # 14800). 4IJD was a gift from Cheryl Arrowsmith (Addgene plasmid # 51328).

4.5.2. Cell Culture

A549 cells were cultured in DMEM medium (GenDEPOT) supplemented with 10% fetal bovine serum (FBS; Gemini) and maintained in a humidified atmosphere at 37°C with 5% CO2. Cells were seeded at a density of 4 x 105/well of 6-well plates for all the experiments and maintained in 6 or 10 cm dishes (Corning). HEK293 cells were cultured in DMEM medium (GenDEPOT) supplemented with 10% fetal bovine serum (FBS; Gemini) and maintained in a humidified atmosphere at 37°C with 5% CO2. To establish stable cell lines (A549 and HEK293) expressing dCas9-5xGCN4, we transduced cells with lentiviral particles generated with pLVX-dCas9-5xGCN4 plasmid. 48 hours after, we selected transduced cells with Blasticidin (5ug/mL) for 4 days. For

A549 cells, after stable expression of dCas9-5xGCN4 was confirmed, we transduced gene-specific sgRNA into cells using lentiviral particles generated from Lentiguide-Puro vectors with specific sgRNA and selected for transduced cells using puromycin (2 ug/mL) for 48 hours. For epigenetic expression inA549 cells, plasmids with epigenetic vectors were transiently transfected using Helix-IN reagent (OZ Biosciences) according to manufacturer protocols. In 293 cells, plasmids with sgRNA and epigenetic effectors were transfected using Xfect (Takara) according to manufacturer protocols.

4.5.3. Chromatin Immunoprecipitation

ChIP was performed following the protocol for EZ ChIP™ Chromatin Immunoprecipitation kit with the following modifications. 72 hours after transfection with epigenetic writers, A549-dCas95xGCN4-sgRNA cells were fixed with 1% formaldehyde for 5 min followed by quenching in glyine (125mM). After glycine incubation, fixed cells washed with PBS + Protease Inhibitor Cocktail (ThermoFisher), filtered and FACS sorted for GFP expression directly into lysis buffer (buffer (1% SDS; 10-mM EDTA pH 8.0; 50-mM Tris-Cl pH 8.0). Lysed cells were sonicated to shear DNA (5x20 sec, 20Hz) and precleared with protein A agarose/salmon sperm slurry (Millipore 16-57, 40uL/mL) overnight at 4°C with rotation. Samples were incubated with appropriate antibody for 8 hours, then 20uL protein A agarose/salmon sperm slurry was added and incubated overnight at 4 °C. Samples were washed 1 time in low salt wash buffer (0.1% SDS; 1.0% Triton X-100; 2-mM EDTA pH 8.0; 20-mM Tris-C, pH 8.1; 150-mM NaCl), 1 time in

high salt buffer (0.1% SDS; 1.0% Triton X-100; 2-mM EDTA pH 8.0; 20-mM Tris-Cl pH 8.0; 500-mM NaCl), 1 time in Lithium Chloride (LiCl) wash buffer (250-mM LiCl; 1.0% IGEPAL; 1-mM EDTA pH 8.0; 10-mM Tris-Cl pH 8.1; 1.0% Deoxycholic acid), and finally washed twice in Tris-EDTA buffer, pH 8.0 (10-mM Tris-Cl; 1-mM EDTA) before being eluted with 0.1-mM NaHCO3; 1.0% SDS buffer (2x, 15 min). Samples were reverse crosslinked (200mM NaCl overnight at 65°C followed by 10ug RNase A, 10-mM EDTA pH 8.0; 40-mM Tris-Cl,pH 8.0; 50 µg Proteinase K at 45°C for 2 hours). Fragmented DNA was isolated by phenol/chloroform/isoamyl alcohol extraction. For chromatin immunoprecipitation experiments, the following primary antibodies were used: H3K4me3 (1 mg/mL), ab8580, Abcam; H3K27ac (1 mg/mL), ab4729, Abcam; H3K27me3 (1 mg/mL), 39155, Active Motif; Normal mouse IgG (1 mg/mL), 12-371, EMD Millipore Corp. Finally, samples were amplified using PCR (See Additional File 5 for primers) and visualized using a 1.5% agarose gel with Ethidium Bromide.

4.5.4. Western Blot

293 cells were transfected with appropriate scFv-fusion vectors for 72 hours in a 6-well plate. After washing, cells were collected in RIPA buffer supplemented with Protease Inhibitor Cocktail (ThermoFisher 1860932) and incubated for 15 minutes on ice before centrifugation. Approximately 10ug supernatant was boiled in 4X loading dye (50-mM Tris-HCl, pH 6.8; 2% SDS; 10% glycerol; 1% β-mercaptoethanol; 12.5-mM EDTA; 0.02% Bromophenol blue) and loaded into a 10% acrylamide gel for electrophoresis (100V, 1 hr). After transfer to PVDF membrane

(100V, 1 hr), blots were blocked (5% milk) and immunoblotted using anti-GFP antibody (Millipore AB3080), followed by anti-Rabbit IgG secondary (Jackson ImmunoResearch 111-035-045) and then visualized with ECL detection reagents (GE Amersham ECL Prime).

4.5.5. Expression of *SNCA* gene

In HEK293-dCas95xGCN4 cells, we transfected specified RNA with relevant epigenetic writer plasmids. 72 hours after transfection, cells were FACS sorted for GFP expression directly into TriZOL and RNA extracted according to manufacturer protocols. cDNA was generated with 300ng total RNA (AmfiRivert, GenDepot) and used to check expression of b-actin and *SNCA* with qPCR. We performed qPCR for biological repeat in triplicate using PowerUP SYBR Green Master Mix (Applied Biosystems) according to manufacturer protocols on a QuantStudio 7 Flex instrument (Applied Biosystems). To determine the difference between empty control and epigenetic modulator, we calculated fold change (2^(- ΔΔCt)) of treated cells.

4.5.6. Bioinformatics Analysis

Datasets corresponding to histone modifications (H3K4me3, H3K9ac, H3K27ac, H3K27me3) and DNA methylation in A549 cell lines were collected from the ENCODE project (access number: ENCSR000ASH, ENCSR000ASV, ENCSR000AUI, ENCSR000AUK, and ENCSR000ABU) (Davis et al 2018; https://www.encodeproject.org/)[32]. The RNA-Seq data in A549 were downloaded from the UCSC Genome Browser (accession number: wgEncodeEH000314). The CpG island

annotation data corresponding to the hg19 genome assembly was also obtained from the UCSC Genome Browser.

To define bonafide histone marker enrichment in the promoter region of a specific gene, we used RNA-Seq gene expression data to set up a threshold for each histone marker as follows. We first identified 1000 most and 1000 least expressed genes based on RPKM (Reads per Kilobase Million) values calculated from the RNA-Seq measurements. For each of these 2000 genes, we then extracted the [-1000,500] region surrounding its TSS. We divided this 1,500 bp promoter region into 30 bins with 50 bp length each. We then generated the averaged histone marker signal values across the 30 bins for the1000 most and least expressed genes respectively. We then defined the threshold for enrichment calling of each histone maker as a number that can maximally distinguish the averaged histone marker profiles corresponding to the most and least expressed genes. DNA methylation enrichment was determined if the DNA methylation signal is present in a gene's upstream 1k region. Similarly, CpG island presence was also based on its annotation in a gene's upstream 1k region.

For HEK293 cells H3K4me3 and H3K27ac, we downloaded the call sets from the ENCODE portal with the following identifiers: ENCSR000DTU, ENCSR000FCH.

For H3K27me3 data for HEK293 cells, we downloaded data deposited with ChIP-Atlas (Oki et al 2018)[159] with the following identifier: DRX013192.

4.5.7. Statistical Analysis

All ChIP samples were normalized to input and all qPCR samples were normalized to β-Actin. ChIP sample statistical analysis was performed with a parametric t-test comparing empty and enzyme sample sets. Statistical analysis of the qPCR data fold change from 1 was performed via the bootstrap resampling method using the open web portal http://pdo.iconcologia.net/stats/br/index.html. [160,161]

4.6. Figures

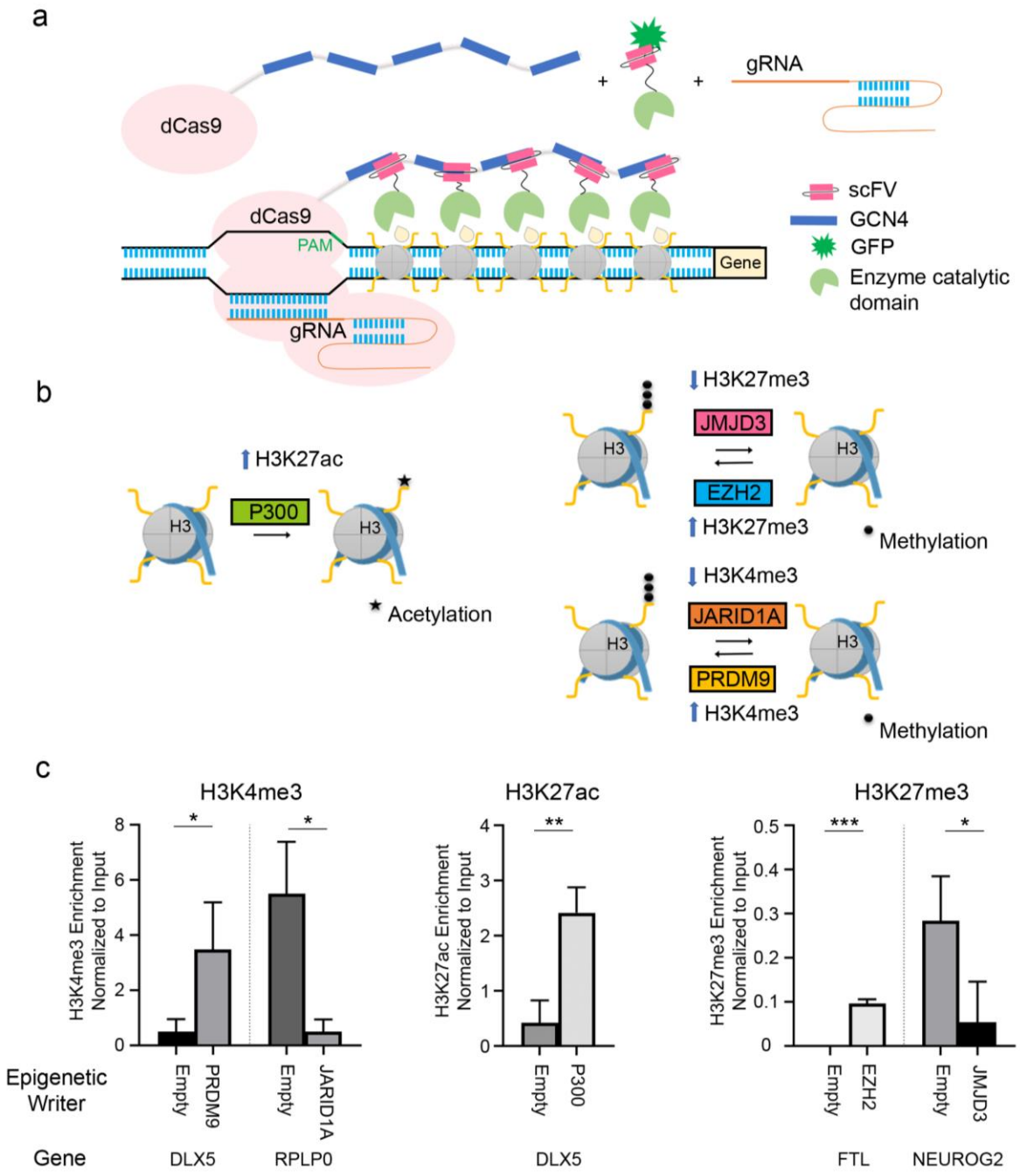
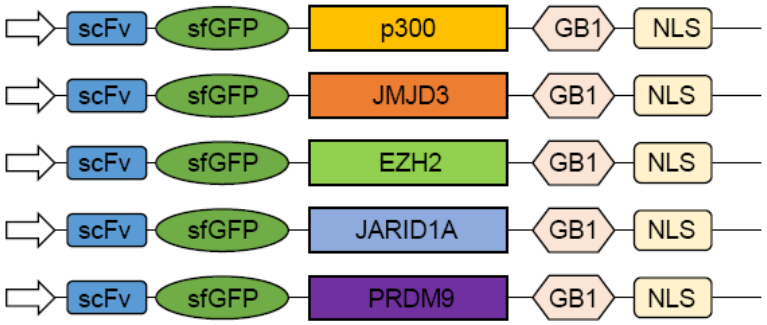


Figure 14 – Toolkit of Epigenetic Editors

Figure 14: a Schematic of epigenetic modulator design based on SunTag framework. 5xGCN4 sequences bound to dCas9 recruit epigenetic effectors fused with anti-GCN4 scFv antibody fragments to the sites directed by sgRNA. **b** Epigenetic modulators used in this study and their effects on histone marks. **c** Epigenetic editors alter histone marks in a site-specific manner. Each graph is normalized to input and represents biological triplicates. PRDM9 increases H3K4me3 enrichment 6.9 fold ($p=0.04$) and JARID1A decreases it by 11 fold ($p=0.01$). H3K27ac is increased 5.8 fold by p300 ($p=0.005$). H3K27me3 is increased from below detection limits by EZH2 ($p<0.001$) while JMJD3 decreases H3K27me3 by 5.3 fold ($p=0.04$) * = $p < 0.05$, **=$p <0.01$, ***=$p<0.001$.

S1a



S1b

Plasmid name	Addgene ID	Target Histone PTM	Cloned	Predicted Final Protein Size
pcDNA-dCas9-p300 core	61357	H3K27ac	HAT domain, 617 aa (1851 bp)	148 kDa
pCMV-HA-JMJD3	24167	H3K27me3	Demethylase domain, 647 aa (1941 bp)	152 kDa
pGEX-EZH2	28060	H3K27me3	SET domain, 122 aa (366 bp)	90 kDa
pcDNA3/HA-FLAG-RBP2 (JARID1A)	14800	H3K4me3	JmjN to JmjC domain, 797aa (2,391 bp)	169 kDa
PRDM9(4IJD)	51328	H3K4me3	SET domain, 121 aa (363 bp)	90 kDa

S1c

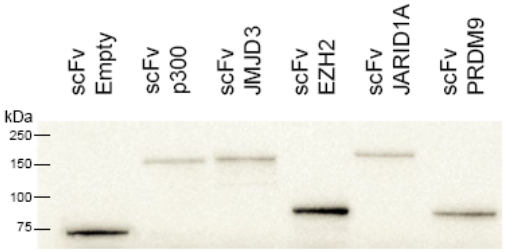


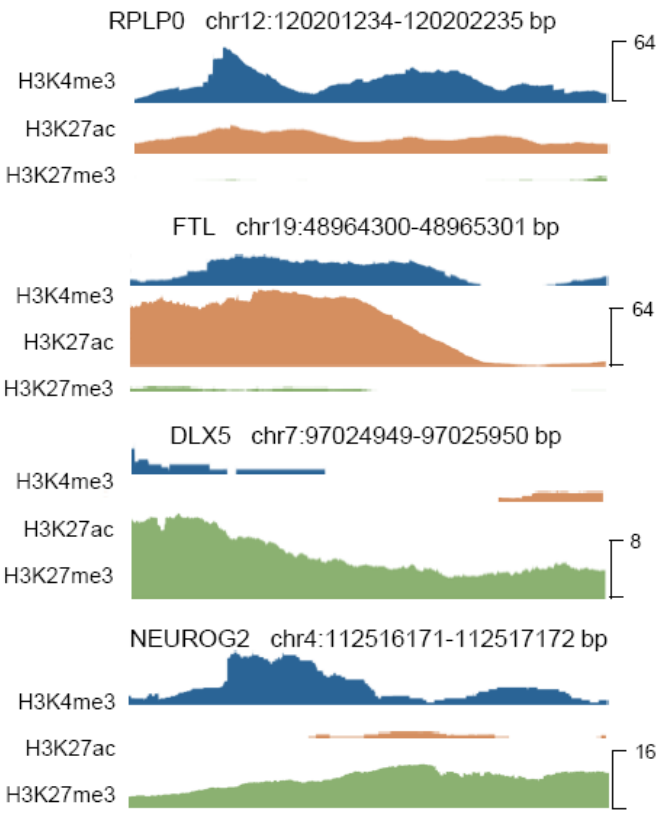
Figure 15 – Design of Epigenetic Toolkit Writers

Figure 15: **a** Schematic design of expression constructs used in this study. Expression is driven by CMV promoter and contain an anti-GCN4 scFvantibody fragment, sfGFP, enzyme catalytic core, the GB1 solubility sequence and a nuclear localization sequence. **b** Details about specific sub-cloning for each epigenetic writer including original vector name and AddgeneID, specific histone post-translational mark targeted, which specific domains were subcloned and the approximate expected size of the entire expression cassette. **c** Vectors were transfected into 293 cells and extracted protein was immunoblotted (anti-GFP antibody) for the indicated construct. Empty vector has a predicted molecular weight of 74 kDa, and each construct expresses the expected size protein band.

S2a

	H3K4me3	H3K27ac	H3K27me3
RPLP0	24.43	3.12	0.00
FTL	18.76	44.98	0.00
DLX5	0.00	0.00	11.38
NEUROG2	16.43	0.00	10.25

S2b



S2c

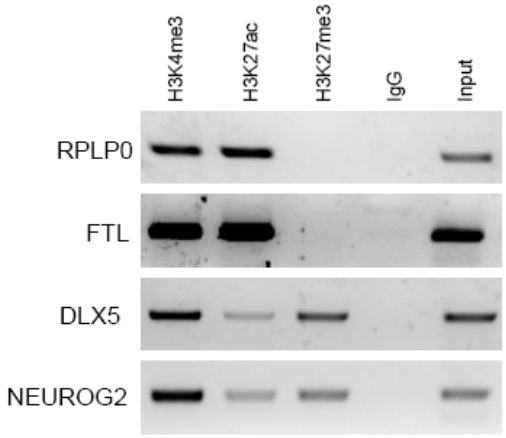


Figure 16 – Endogenous Epigenetic Structure of Selected Genes in A549 cells .

Figure 16: a Enrichment scores (Fold change from input) for histone post translational modifications in final selected genes from A549 cells. **b** ENCODE data showing peak enrichment for specified histone post translational modifications in the promoter region of the specified gene confirms our initial analysis. **c** ChIP data showing our A549 cells have similar histone mark enrichment as reported in ENCODE database with only minor differences.

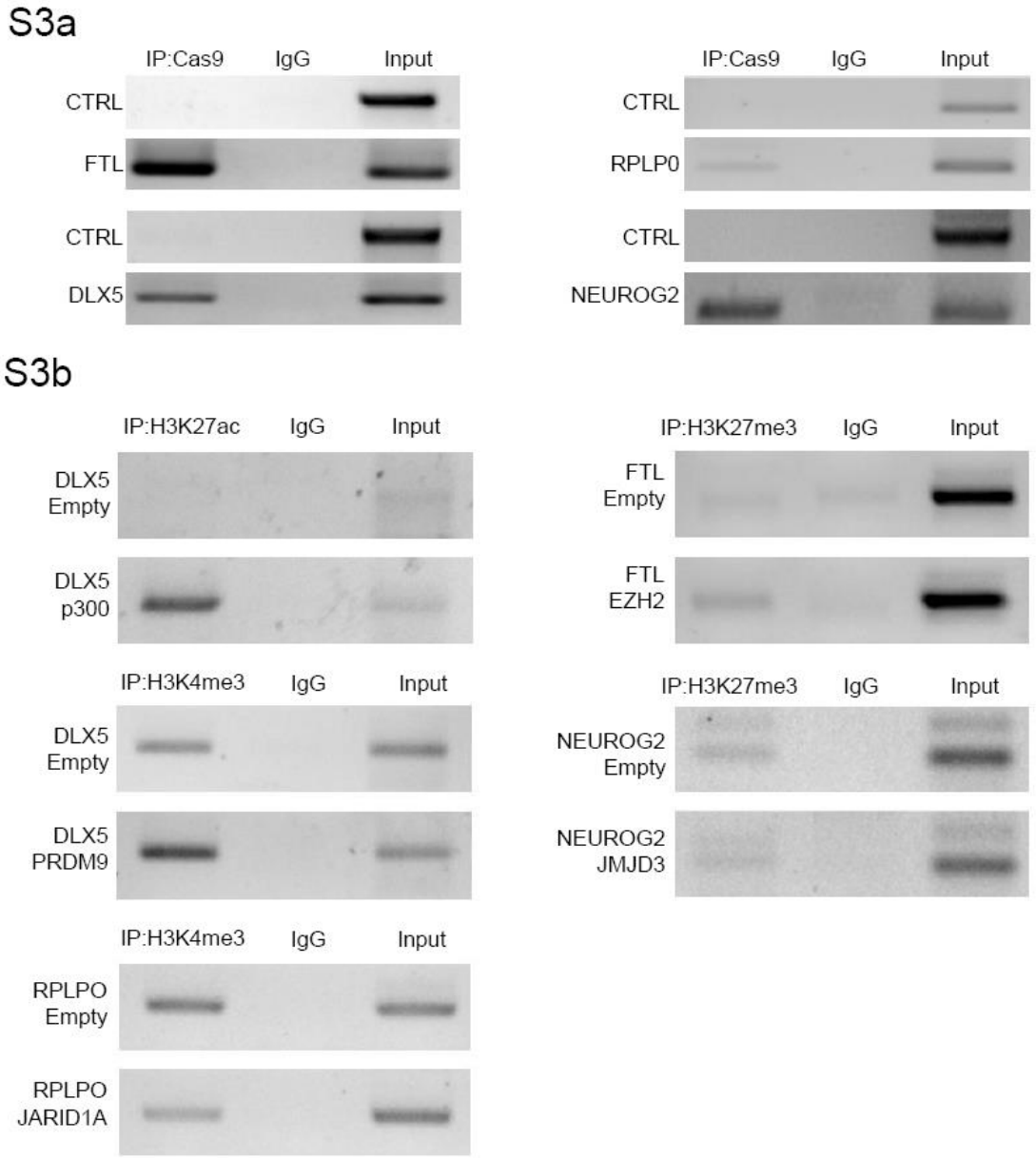


Figure 17 – Alterations in endogenous epigenetic structures by SunTag writers.

Figure 17: **a** ChIP data for A549-dCas95xGCN4 cell lines show that dCas9 does not occupy the promoter regions of target genes. Stable transduction of A549-dCas95xGCN4 cells with gene-specific sgRNA results in recruitment of dCas9 to the targeted promoter. **b** Original gel images for results reported in Figure 14c. Please note upper band in NEUROG2 is non-specific artifact generated during PCR.

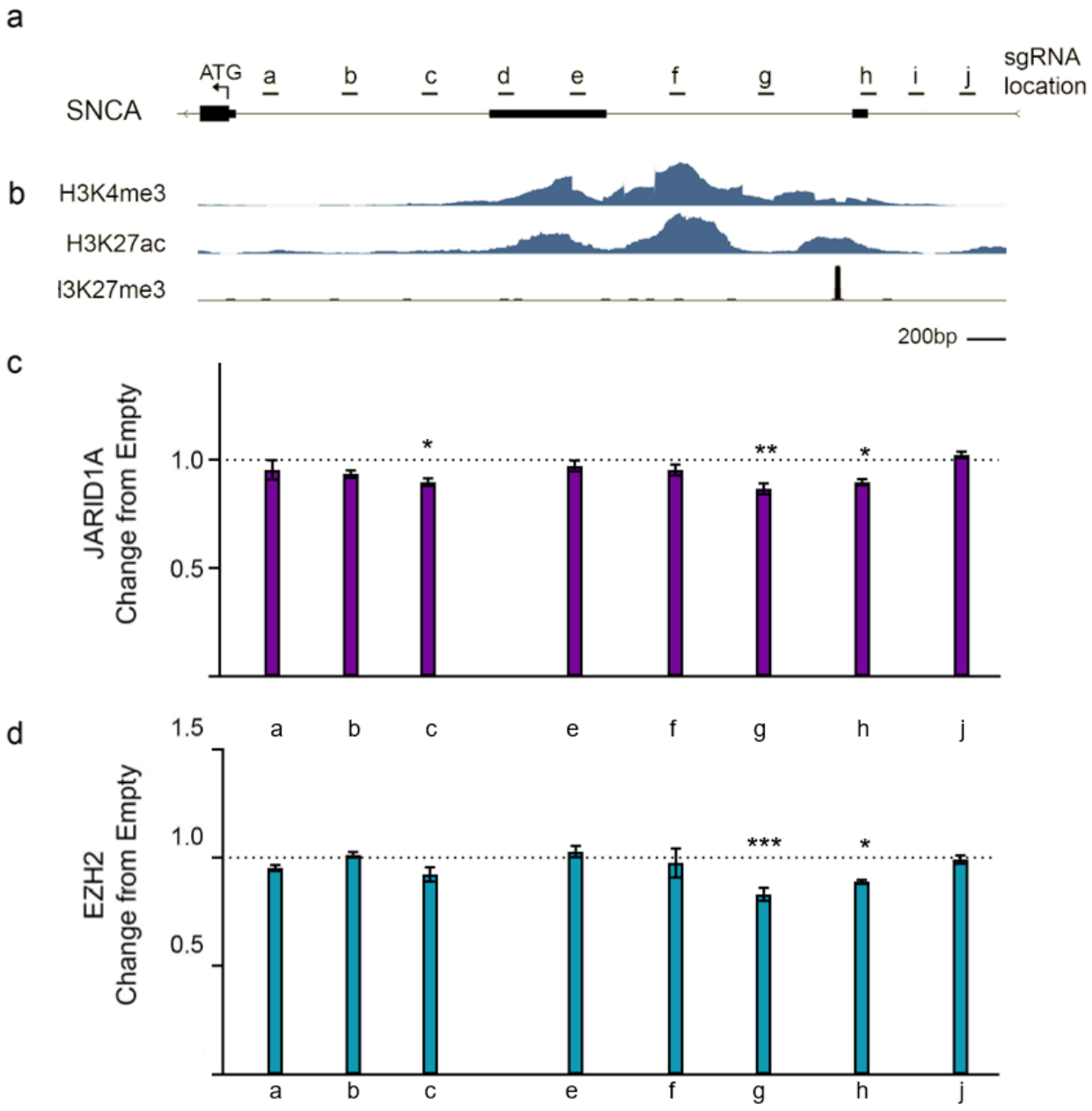


Figure 18 – Screening the α-synuclein gene promoter with NanoLuciferase

Figure 18 - **a** Scale diagram of promoter region of SNCA gene showing two alternative non-coding Exon 1 and Exon 2 with a start codon (ATG) as well as locations of sgRNA used. **b** Histone mark enrichment from ENCODE database (H3K4me3, H3K27ac) or ChIP-atlas (H3K27me3) for HEK293 cells. (Peak locations are scaled to diagram shown in a). **c** NanoLuciferase assay was carried out to determine the change in αSynuclein expression levels relative to empty scFv control (Set to 1) for JARID1A targeted to specified loci by sgRNA. (Locations are scaled to diagram shown in a). **d** Same at Figure 15c, but for EZH2 enzyme. Each point is normalized to empty control and represents biological triplicates. * = $p < 0.05$, **=$p < 0.01$, ***=$p<0.005$.

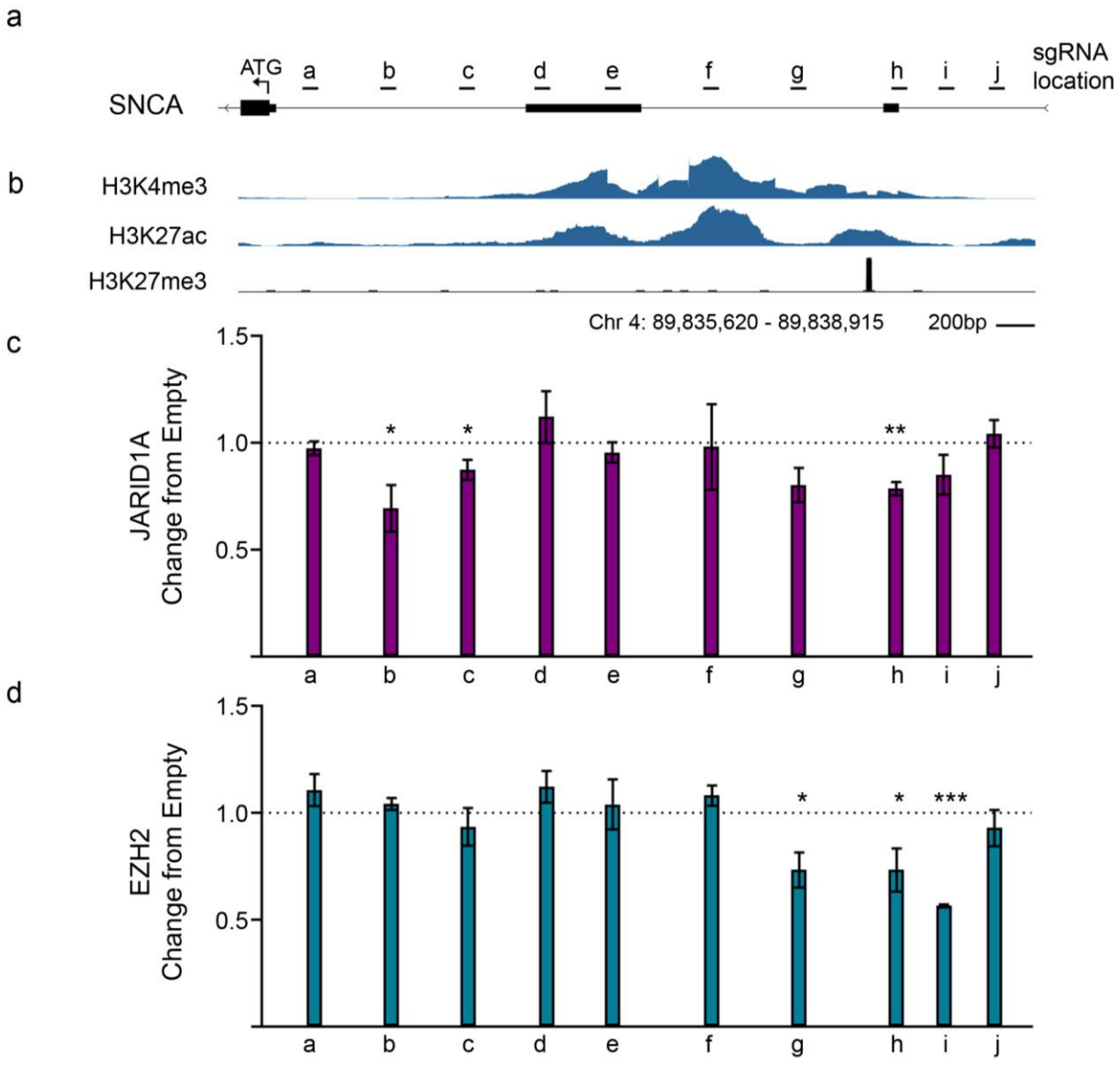


Figure 19 – High-resolution Screening the α-synuclein Gene Promoter

Figure 19: a Scale diagram of promoter region of SNCA gene showing two alternative non-coding Exon 1 and Exon 2 with a start codon (ATG) as well as locations of sgRNA used. **b** Histone mark enrichment from ENCODE database (H3K4me3, H3K27ac) or ChIP-atlas (H3K27me3) for HEK293 cells. (Peak locations are scaled to diagram shown in a. **c** qPCR assay was carried out to determine the change in αSynuclein expression levels relative to empty scFv control (Set to 1) for JARID1A targeted to specified loci by sgRNA. (Locations are scaled to diagram shown in a. **d** Same at Figure 16c, but for EZH2 enzyme. Each point is normalized to empty control and represents biological triplicates. * = $p < 0.05$, **=$p < 0.01$, ***=$p<0.005$.

Figure 20 - Human SNCA upstream regulatory region with sgRNA binding sites

Figure 20: sgRNA binding sites on SNCA promoter. Genomic sequence of promoter region of SNCA with sgRNA locations highlighted in blue. (GRCh38.p13 chr4: 89,835,620 - 89,838,915)

Table 3 – sgRNA Referenced in Chapter 4

sgRNA Name	Sequence 5' - 3'
RPLPO	ACTTAAAGGCGGCTTTCCGT
FTL	TATCTCGCAGCGCAAACCTC
DLX5	CAAAAACACACACAAGCGCG
NEUROG2	CGCAGGCCTCCCGGAGTCCA
SCNAa	AAAGCAGACATTTTTAGCTC
SCNAb	AACAGCAGGCCCAAGTGTGA
SCNAc	GCTTTTCCCCGGGAAACGCG
SCNAd	ATTCCCAAATAATATTTAAT
SCNAe	CACTTCCGCGTCGCGGCGCT
SCNAf	GCGACTCTGACGAGGGGTAG
SCNAg	TGGGAAAATCAGCGTCTGGC
SCNAh	AAGCAAAGGCTTTCTGCTAG
SCNAi	ACTTTAAAACCACAAGGAAC
SCNAj	CAAGTCCAACCTTCTTGCTC

Table 4 – Primers Referenced in Chapter 4

Primer Name	Sequence 5' - 3'
RPLPO-1 F	CGAGGCAGCGCCTTCCTT
RPLPO-1 R	CCCGCGCGTGCCTTTTAT
FTL-1 F	TCCAGAGATCTCCAGGGGTC
FTL-1 R	GGGTGCCTCGGGAAAGTAAG
DLX5-1 F	TGAGGCTTCTGATTGGAACACA
DLX5-1 R	AGTAACACCCTAACTCGTCCAAC
Neurog2 F	TAATGAGCTGCTGAAAGGGAGC
Neurog2 R	CCGCCGCTGTCCATTGT
SNCA Expression F	CCAGAAGACAGTGGAGGGAGCAGG
SNCA Expression R	GCCTCATTGTCAGGATCCACAGGC
β-Actin F	GGAGTCCTGTGGCATCCACG
β-Actin R	CTAGAAGCATTTGCGGTGGA

CHAPTER FIVE: CONCLUSIONS

The CRISPR/Cas systems have been an invaluable tool in recent years to empower researchers to make precise edits to the genome. The relative ease in using the techniques to quickly and effectively make targeted changes has had great impact in laboratories around the world. This technology has even been widely covered in the media, movies and on social networks, generating interest from the general public in a way few scientific techniques even have before. We have used this system in two different ways to elucidate epigenetic effects in ways that were not available even a few years ago. As previously discussed, the CRISPR/Cas toolbox has many more potential uses than we showed here, and it remains to be seen how far these techniques can take us. For all its power though, CRISPR/Cas systems do have a number of technical limitations and ethical issues that remain to be solved.

5.1. Limitations of CRISPR

The commonly used spCas9 is quite a large protein and this size places some limits on its utility. In vector constructs containing spCas9, the size can make transfection into sensitive or difficult cells such as primary neurons a significant challenge. This size also presents a barrier to therapeutic delivery of Cas9 into living organisms. The AAV system is widely used to ferry genes of interest *in vivo,* but they have a limited genomic capacity (~4.4kb) and the spCas9 ORF is 4.2kb (which does not include promoters, sgRNA, or any modifications). Aside from the size, the PAM sequence of -NGG presents some difficulty of its own. The relatively common -NGG sequence has a high chance of being found, and so this increases the potential for off target

effects. Conversely, not all desirable target sites are in C/G rich areas of the genome – in particular the 3' end of the *SNCA* gene discussed here is very A/T rich and left very few potential target sites for us to work with.

5.1.1. Cpf May Provide Alternatives to Cas9

To circumvent these problems, researchers are now turning their attention to alternative Cas9 options. Cpf1 orthologs (Table 1) present an exciting new opportunity. The smaller size of these proteins makes them a better fit for *in vivo* delivery and easier to work with. They also make use of an alternative PAM site (TTTV) that can allow targeting to A/T rich areas. They also may reduce another major hurdle preventing widescale use of CRISPR/Cas9 as a therapeutic system in humans – immunogenicity. The commonly used Cas9 systems are from *S. pyogenes* and *S. aureus*, which are human pathogens. One study indicated that more than half of humans have pre-existing humoral immune responses to Cas9 proteins from these species[162], so a shift to alternative Cas9 proteins may be needed. I suspect that engineered Cas9 alternatives will be the next major breakthrough with the CRISPR/Cas system as research groups work to develop artificial CRISPR-guided endonucleases that have a small size, flexible PAM sequence, and low immunogenicity and I am excited to see what the future brings on this front.

5.2. Understanding Endogenous Gene Regulation

Understanding the endogenous regulatory mechanisms of genetic control is an essential piece of the puzzle in fully realizing our ability to alter pathogenic genetic activity. Recently there has

been focus on precision medicine, and the ability to affect individual genes in specific groups of cells is a critical part of that. Using a highly specific set of tools like those described here allows a degree of control that hasn't been previously available. While every cell in an individual has the same genetic sequence each different cell type has a drastically different epigenetic code, making it a very real possibility for a tailored treatment that only has effects on the desired subpopulation. Using the α-synuclein example presented here, it could be feasible to target specific alterations to the epigenetic code that would only be effective in the exact context of a dopamine neuron – allowing us to reduce expression in those cells while not affecting the rest of the body. While there is certainly a lot of work remaining before we have that type of solution available, increasing our understanding of individual epigenetic effects may help move us towards that goal.

In this work, we have presented a tool that allows us to quickly and easily measure real-time changes to α-synuclein levels in a highly sensitive manner. The advantage of having an endogenous reporter system is that we are detecting the cells stimulus response using only its native regulatory mechanisms. Previous systems involve plasmids, gene knockout or knockdown, or overexpression constructs to model disease conditions but fail to take the endogenous environment into account. Once we validated that reporter system, we developed a modular epigenetic toolkit that would allow us to make very precise changes to the endogenous epigenetic environment. We showed the ability of our modular epigenetic toolkit

to affect histone marks in a target-specific manner and identify potential targeting regions for modulating expression using our reporter system that we confirmed with more rigorous testing.

5.3. Engineered Endogenous Reporter using NanoLuc

The endogenous tagging of the α-synuclein gene provides a new way to monitor cell activity. This creates a great opportunity for quickly screening treatment options in a high-throughput manner. NanoLuciferase is a very bright reporter, making it very sensitive at detecting even subtle changes in α-synuclein levels. This, as well as its relatively small size (>19kDa) makes it a practical option for understanding endogenous effects. Having a simple readout of luminescence makes it easy to compare activity of large small-molecule compound libraries to identify novel candidates that can target α-synuclein and reduce its levels in the cell. Similarly, the sensitive reporter allows more focused and detailed changes to be examined – such as writing or erasing histone marks.

5.3.1. Future Directions for our NanoLuc Reporter System

While the NanoLuc reporter cell line has some potential value, to fully realize its potential for Parkinson's disease research it will need to be moved into a more relevant cell type. While the HEK293 cell line is easy to grow and convenient for preliminary data collection, it lacks the specific regulatory and epigenetic environment of the cells directly affected in PD: dopamine producing neurons. With this limitation in mind, we have already begun work to generate

dopaminergic neuronal cell lines with the NanoLuc tag inserted in the *SNCA* locus. Once these cells have been developed, we will be able to quickly advance our understanding of the dynamics of α-synuclein in the cell. The HEK293 cell line may make a great option for the first round of data collection for drug screening, novel therapeutics and understanding basic epigenetic dynamics, due to the relative ease of their culture and speed of growth. Once that first round has been completed, the more sensitive but delicate neuronal cell lines can be used to carefully dissect the candidates already identified to find potential therapeutic options. Having both a fast screening tool and a sensitive, relevant system empowers us to move quickly through therapeutic options and may hopefully enable identification of novel treatments.

5.4. Modular Epigenetic Toolkit

The modular epigenetic toolkit is an excellent tool to increase our understanding of the dynamics surrounding epigenetic regulation in eukaryotic cells. While in the work described here, we confined our experiments to a small group of genes and conducted a limited screen of the α-synuclein promoter, the ability to focally edit histone marks and epigenetic information has much broader application once a few technical details have been ironed out. While we were able to show that we can focally edit marks in a region using ChIP, we still need a clearer picture of the specific chromatin location where the editing takes place. It was interesting to note that targeting the epigenetic writer to a putative histone mark peak was not effective, but that maximal impact happens when the complex is targeted 200-400 base pairs away. Previous work has identified that the Cas9 molecule occupies a 50-100 base pair 'footprint' on the

DNA[163], so with a GCN4 tail it seems reasonable that the complex may be more effective at targeting nearby genomic regions.

5.4.1. Future Directions and Challenges for the Modular Epigenetic Toolkit

In future work, we will need to use a full ChIP-seq to determine with high precision where the histones are actually being edited in relation to the sgRNA target site. Once we have the data comparing control histone peaks to treated histone peaks, we will begin to understand how to control the epigenome in a way not previously available. After showing exactly where a given mark is being appended, we can connect individual epigenetic marks to expression data and then begin to influence transcription with a high degree of control. No single epigenetic mark acts in a vacuum, so to fully realize the potential of this system we will need to combine our editing capabilities with computational tools and algorithms to predict the maximum impact of any given edit in reference to the endogenous epigenetic architecture present in the cell.

However, this complexity also opens up the exciting new possibility of highly specific genetic therapies. As every cell has its own complex regulatory environment, focal epigenetic changes may only be applicable to one specific cell type. Consider the α-synuclein gene we used to demonstrate the techniques outlined here. In Parkinson's disease, it is elevated to pathological effect and we showed targeting H3K27me3 to very specific regions in may reduce its level and alleviate pathological effects. As previously discussed, the specific patterns of DNA

methylation, chromatin state, and other histone marks will be different in – for instance – a liver cell, greatly reducing the chances of unintended effects occurring in that cell. One of the main issues facing genetic therapy today is the non-specific nature of genetic effects and the difficulty getting very specific delivery to target tissues *in vivo.* This type of epigenetic approach may be a way to help bridge that gap, although there remains a lot of work understanding the complete epigenome before that could be possible. Although in these publications we focused on α-synuclein, this type of directed epigenetic manipulation has broad potential reach outside Parkinson's disease and neuroscience. Accumulating data has been highlighting the role of epigenetics in many areas of human health in areas such as aging and cancer research.

As medicine continues to improve, aging and age-related diseases have become focal points for research. Aging is a massively complex process with many contributing factors such as nutrition, DNA damage accumulation, oxidative stress, and cellular senescence, but recently the role of epigenetics is increasingly being recognized as an important factor. During aging there are many changes to the chromatin structure such as general loss of histones, changes in activating/repressive histone marks, and changes in DNA methylation and heterochromatin[164]. Studies with senescent cells have shown major reorganization of heterochromatin related to changes in EZH2 expression levels with concomitant shifts in H3K27me3 that play direct roles in replicative lifespan and post-mitotic lifespan[165]. In agreement with this, patients with the rare premature aging laminopathy Hutchinson-Gilford progeria syndrome (HGPS) show H3K27me3

loss is observed along with heterochromatin reductions[166]. Additionally, regions of activating H3K4me3 are observed, indicating major chromatin reorganization takes place in these cells[167]. Taken together, these observations highlight the fact that epigenetics have a significant role in the aging process and using a targeted epigenetic writing system we can begin to understand exactly what effect individual changes can have on a cells lifespan and hopefully being to inform decisions about targeted therapies in the future. While it is unlikely that making targeted changes to individual histone marks may alone drastically affect the aging process, epigenetics is an important piece of the puzzle.

In the field of cancer research, epigenetics has long been a research focus, particularly DNA methylation. More recently, whole genome sequencing of tumors has identified dysregulation of chromatin modifiers as drivers in many types of cancers[168], and epigenetic activation or silencing of central cancer genes such as p53[169]. Intriguingly, the same functions that protect cells from age-related senescence also can drive unrestricted tumor growth when they are not properly managed. Overexpression of the H3K27 methyltransferase EZH2 has revealed it to be a putative oncogene in breast, prostate, and bladder cancers[170]. Aberrant H3K27me3 has also been identified in medulloblastoma[171]. Similarly, histone acetyltransferases such as p300 can recruit transcriptional activators to oncogenic genes to drive tumor growth and specific inhibitors such as JQ1 showed potential as anti-cancer therapeutics in early studies[172]. The role of epigenetic modulation is clear in these malignancies, and it will be very interesting to see

what discoveries the future will bring on this front. The CRISPR-based epigenetic platform we developed here could easily be used in a large-scale screening with sgRNA libraries to quickly identify new gene targeting strategies to slow cell growth or make cancer cells for responsive to drugs. What is particularly interesting is the idea that previous work has shown histone marks to be stable over cell divisions, opening up the possibility that identifying epigenetic targets and developing a clear understanding of the genetic mechanics could lead intervention strategies could have lasting impact on cell differentiation or cancer proliferation.

Developing a deep understanding of endogenous epigenetic regulation remains a key goal of chromatin biology. We are still a long way from being able to directly edit the epigenome for therapeutic benefit but development of tools such as the ones described here may help to move us towards that end. Despite the work still remaining to fully optimize the CRISPR/Cas systems for direct human use and the challenges of untangling the complex epigenome I am hopeful that these techniques will soon be available to help treat and prevent genetic diseases and form the basis for new interventions in human health.

APPENDIX A: DEFENSE ANNOUNCEMENT AND PUBLICATIONS

Announcing the Final Examination of Mr. Levi Adams for the Degree of Doctor of Philosophy in Biomedical Sciences

Date: Friday, May 15, 2020
Time: 1 PM
Room: Zoom meeting
Dissertation Title: Monitoring Pathological Gene Expression and Studying Endogenous Epigenetic Architecture by CRISPR/Cas9-Based Tool Development using α-synuclein as a Model

Abstract

Until recently, complete understanding of the endogenous activity of pathologically relevant genes was out of reach and research was confined to in situ work, plasmid-based constructs and artificial model systems. The development and expansion of the CRISPR/Cas9 genome editing technique has enabled us to explore the molecular underpinnings of gene activation using the cell's own endogenous regulatory environment. In this work, we report on the development of a novel tool to monitor the endogenous activity of a causative gene in Parkinson's disease, α-synuclein. We use CRISPR/Cas9 to insert a highly sensitive engineered luciferase at the C-terminal of α-synuclein and assessed its responses to stimuli. Our system responds to epigenetic stimuli, which was unable to be recapitulated by previously available gene activity assays. After development of a sensitive detection tool for epigenetic stimuli, we focused on developed a modular suite of epigenetic writers and erasers by modification of the SunTag protein tagging system and used catalytically dead Cas9 (dCas9) to direct our modular epigenetic toolkit to individual genes. We show that our toolkit of epigenetic effectors successfully writes epigenetic information in a site-specific manner. Using the sensitive α-synuclein reporter we previously developed, we screen the promoter region of this pathologically relevant gene at high resolution and identify the most effective areas for epigenetic intervention in this cell line. These tools allow us to dissect and understand the endogenous regulatory mechanisms of almost any gene targetable by Cas9 in ways that were not previously available may prove to be an effective strategy for persistently altering pathologic transcriptional activity. This system offers a strong tool for to dissect and understand underlying epigenetic architecture and opens potential new avenues for therapeutic strategies for various disease conditions.

Dissertation Committee:

Dr. Yoon-Seong Kim (Chair)
Dr. Mollie Jewett
Dr. Deborah Altomare
Dr. Amber Southwell

Publications

Adams L, Franco MC, Estevez AG. Reactive nitrogen species in cellular signaling. Exp Biol Med (Maywood). 2015 Jun;240(6):711-7. doi: 10.1177/1535370215581314. Epub 2015 Apr 16.

Basu S#, **Adams L#**, Guhathakurta S, Kim YS. A novel tool for monitoring endogenous alpha-synuclein transcription by NanoLuciferase tag insertion at the 3'end using CRISPR-Cas9 genome editing technique. Sci Rep. 2017 Apr 4;8:45883. doi: 10.1038/srep45883.

(Submitted to Genome Medicine) Subhrangshu Guhathakurta#; **Levi Adams#**; Anishaa Shivakumar; Mingyu Cha; Mariana Bernardo Fiadeiro ; Haiyan Nancy Hu; Yoon-Seong Kim. Precise epigenomic editing by a newly developed modular epigenetic toolkit.

(Under Revision) Guhathakurta S, Kim JI, Basu S, Adler E, Je G, Bernardo Fiadeiro M, **Adams L**, Kim YS. Targeted attenuation of elevated histone marks at SNCA alleviates α-synuclein in Parkinson's disease. EMBO Molecular Medicine.

Approved for distribution by Dr. Yoon-Seong Kim, Committee Chair, on May 1, 2020.

Any member of the public who wishes to virtually participate as part of the audience can request the meeting information from the Program Office at BSBSGradInfo@ucf.edu at least 1 week in advance.

APPENDIX B: COPYRIGHT PERMISSIONS

From https://www.nature.com/nature-research/reprints-and-permissions/permissions-requests. Accessed 5/1/2020.

Permissions requests

Permission requests from authors

The author of articles published by Springer Nature do not usually need to seek permission for re-use of their material as long as the journal is credited with initial publication.

Ownership of copyright in in original research articles remains with the Author, and provided that, when reproducing the contribution or extracts from it or from the Supplementary Information, the Author acknowledges first and reference publication in the Journal, the Author retains the following non-exclusive rights:

To reproduce the contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).

The author and any academic institution where they work at the time may reproduce the contribution for the purpose of course teaching.

To reuse figures or tables created by the Author and contained in the Contribution in oral presentations and other works created by them.

To post a copy of the contribution as accepted for publication after peer review (in locked Word processing file, of a PDF version thereof) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the contribution on the publisher's website.

The above use of the term 'Contribution' refers to the author's own version, not the final version as published in the Journal.

Self-archiving

Authors retain the right to self-archive the final accepted version of their manuscript. Please see our self-archiving policy for full
details: http://www.nature.com/authors/policies/license.html

Author reuse

Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution.

Authors must properly cite the published article in their thesis according to current citation standards.

LIST OF REFERENCES

1 Beitz, J. M. Parkinson's disease: a review. *Front Biosci (Schol Ed)* **6**, 65-74, doi:10.2741/s415 (2014).

2 Sherer, T. B., Chowdhury, S., Peabody, K. & Brooks, D. W. Overcoming obstacles in Parkinson's disease. *Mov Disord* **27**, 1606-1611, doi:10.1002/mds.25260 (2012).

3 Steele, J. C., Richardson, J. C. & Olszewski, J. Progressive Supranuclear Palsy. A Heterogeneous Degeneration Involving the Brain Stem, Basal Ganglia and Cerebellum with Vertical Gaze and Pseudobulbar Palsy, Nuchal Dystonia and Dementia. *Arch Neurol* **10**, 333-359, doi:10.1001/archneur.1964.00460160003001 (1964).

4 Dauer, W. & Przedborski, S. Parkinson's disease: mechanisms and models. *Neuron* **39**, 889-909, doi:10.1016/s0896-6273(03)00568-3 (2003).

5 Morrish, P. K., Rakshi, J. S., Bailey, D. L., Sawle, G. V. & Brooks, D. J. Measuring the rate of progression and estimating the preclinical period of Parkinson's disease with [18F]dopa PET. *J Neurol Neurosurg Psychiatry* **64**, 314-319, doi:10.1136/jnnp.64.3.314 (1998).

6 Gasser, T. Molecular pathogenesis of Parkinson disease: insights from genetic studies. *Expert Rev Mol Med* **11**, e22, doi:10.1017/S1462399409001148 (2009).

7 Engelhardt, E. & Gomes, M. D. M. Lewy and his inclusion bodies: Discovery and rejection. *Dement Neuropsychol* **11**, 198-201, doi:10.1590/1980-57642016dn11-020012 (2017).

8 Spillantini, M. G. *et al.* Alpha-synuclein in Lewy bodies. *Nature* **388**, 839-840, doi:10.1038/42166 (1997).

9 Golbe, L. I., Di Iorio, G., Bonavita, V., Miller, D. C. & Duvoisin, R. C. A large kindred with autosomal dominant Parkinson's disease. *Ann Neurol* **27**, 276-282, doi:10.1002/ana.410270309 (1990).

10 Polymeropoulos, M. H. *et al.* Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science* **274**, 1197-1199, doi:10.1126/science.274.5290.1197 (1996).

11 Stefanis, L. alpha-Synuclein in Parkinson's disease. *Cold Spring Harb Perspect Med* **2**, a009399, doi:10.1101/cshperspect.a009399 (2012).

12 Chartier-Harlin, M. C. *et al.* Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167-1169, doi:10.1016/S0140-6736(04)17103-1 (2004).

13 Guhathakurta, S., Bok, E., Evangelista, B. A. & Kim, Y. S. Deregulation of alpha-synuclein in Parkinson's disease: Insight from epigenetic structure and transcriptional regulation of SNCA. *Prog Neurobiol* **154**, 21-36, doi:10.1016/j.pneurobio.2017.04.004 (2017).

14 Withers, G. S., George, J. M., Banker, G. A. & Clayton, D. F. Delayed localization of synelfin (synuclein, NACP) to presynaptic terminals in cultured rat hippocampal neurons. *Brain Res Dev Brain Res* **99**, 87-94, doi:10.1016/s0165-3806(96)00210-6 (1997).

15 Kholodilov, N. G. *et al.* Increased expression of rat synuclein in the substantia nigra pars compacta identified by mRNA differential display in a model of developmental target injury. *J Neurochem* **73**, 2586-2599, doi:10.1046/j.1471-4159.1999.0732586.x (1999).

16 Murphy, D. D., Rueter, S. M., Trojanowski, J. Q. & Lee, V. M. Synucleins are developmentally expressed, and alpha-synuclein regulates the size of the presynaptic vesicular pool in primary hippocampal neurons. *J Neurosci* **20**, 3214-3220 (2000).

17 Rideout, H. J., Dietrich, P., Savalle, M., Dauer, W. T. & Stefanis, L. Regulation of alpha-synuclein by bFGF in cultured ventral midbrain dopaminergic neurons. *J Neurochem* **84**, 803-813, doi:10.1046/j.1471-4159.2003.01574.x (2003).

18 Abeliovich, A. *et al.* Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system. *Neuron* **25**, 239-252, doi:10.1016/s0896-6273(00)80886-7 (2000).

19 Cabin, D. E. *et al.* Synaptic vesicle depletion correlates with attenuated synaptic responses to prolonged repetitive stimulation in mice lacking alpha-synuclein. *J Neurosci* **22**, 8797-8807 (2002).

20 Fortin, D. L. *et al.* Neural activity controls the synaptic accumulation of alpha-synuclein. *J Neurosci* **25**, 10913-10921, doi:10.1523/JNEUROSCI.2922-05.2005 (2005).

21 Wang, W. *et al.* A soluble alpha-synuclein construct forms a dynamic tetramer. *Proc Natl Acad Sci U S A* **108**, 17797-17802, doi:10.1073/pnas.1113260108 (2011).

22 Steiner, J. A., Quansah, E. & Brundin, P. The concept of alpha-synuclein as a prion-like protein: ten years after. *Cell Tissue Res* **373**, 161-173, doi:10.1007/s00441-018-2814-1 (2018).

23 Khurana, R. *et al.* A model for amyloid fibril formation in immunoglobulin light chains based on comparison of amyloidogenic and benign proteins and specific antibody binding. *Amyloid* **10**, 97-109, doi:10.3109/13506120309041731 (2003).

24 Khurana, R. *et al.* A general model for amyloid fibril assembly based on morphological studies using atomic force microscopy. *Biophys J* **85**, 1135-1144, doi:10.1016/S0006-3495(03)74550-0 (2003).

25 Basu, S., Je, G. & Kim, Y. S. Transcriptional mutagenesis by 8-oxodG in alpha-synuclein aggregation and the pathogenesis of Parkinson's disease. *Exp Mol Med* **47**, e179, doi:10.1038/emm.2015.54 (2015).

26 Cristovao, A. C. *et al.* NADPH oxidase 1 mediates alpha-synucleinopathy in Parkinson's disease. *J Neurosci* **32**, 14465-14477, doi:10.1523/JNEUROSCI.2246-12.2012 (2012).

27 Mouradian, M. M. MicroRNAs in Parkinson's disease. *Neurobiol Dis* **46**, 279-284, doi:10.1016/j.nbd.2011.12.046 (2012).

28 Tanner, C. M. *et al.* Rotenone, paraquat, and Parkinson's disease. *Environ Health Perspect* **119**, 866-872, doi:10.1289/ehp.1002839 (2011).

29 Venda, L. L., Cragg, S. J., Buchman, V. L. & Wade-Martins, R. alpha-Synuclein and dopamine at the crossroads of Parkinson's disease. *Trends Neurosci* **33**, 559-568, doi:10.1016/j.tins.2010.09.004 (2010).

30 Singleton, A. B. *et al.* alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841, doi:10.1126/science.1090278 (2003).

31 Mata, J., Marguerat, S. & Bahler, J. Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* **30**, 506-514, doi:10.1016/j.tibs.2005.07.005 (2005).

32 Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794-D801, doi:10.1093/nar/gkx1081 (2018).

33 Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. 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, doi:10.1128/jb.169.12.5429-5433.1987 (1987).

34 Mojica, F. J., Diez-Villasenor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36**, 244-246, doi:10.1046/j.1365-2958.2000.01838.x (2000).

35 Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565-1575, doi:10.1046/j.1365-2958.2002.02839.x (2002).

36 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-182, doi:10.1007/s00239-004-0046-3 (2005).

37 Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561, doi:10.1099/mic.0.28048-0 (2005).

38 Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653-663, doi:10.1099/mic.0.27437-0 (2005).

39 Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**, 7, doi:10.1186/1745-6150-1-7 (2006).

40 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

41 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:10.1126/science.1159689 (2008).

42 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845, doi:10.1126/science.1165771 (2008).

43 Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**, 1390-1400, doi:10.1128/JB.01412-07 (2008).

44 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71, doi:10.1038/nature09523 (2010).

45 Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* **39**, 9275-9282, doi:10.1093/nar/gkr606 (2011).

46 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607, doi:10.1038/nature09886 (2011).

47 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

48 Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586, doi:10.1073/pnas.1208507109 (2012).

49 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471, doi:10.7554/eLife.00471 (2013).

50 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

51 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

52 Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67-78, doi:10.1016/j.mib.2017.05.008 (2017).

53 Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280-284, doi:10.1038/nature24049 (2017).

54 Hirano, H. *et al.* Structure and Engineering of Francisella novicida Cas9. *Cell* **164**, 950-961, doi:10.1016/j.cell.2016.01.039 (2016).

55 Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**, 14500, doi:10.1038/ncomms14500 (2017).

56 Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949-962, doi:10.1016/j.cell.2016.04.003 (2016).

57 Kim, D. *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* **34**, 863-868, doi:10.1038/nbt.3609 (2016).

58 Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871, doi:10.1126/science.aad8282 (2016).

59 Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* **5**, a012740, doi:10.1101/cshperspect.a012740 (2013).

60 Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1911, doi:10.1038/s41467-018-04252-2 (2018).

61 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013).

62 Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838, doi:10.1038/nbt.2675 (2013).

63 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839-843, doi:10.1038/nbt.2673 (2013).

64 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).

65 Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88, doi:10.1126/science.aad5227 (2016).

66 Chiang, T. W., le Sage, C., Larrieu, D., Demir, M. & Jackson, S. P. CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. *Sci Rep* **6**, 24356, doi:10.1038/srep24356 (2016).

67 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157, doi:10.1038/s41586-019-1711-4 (2019).

68 Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat Biotechnol* **33**, 390-394, doi:10.1038/nbt.3155 (2015).

69 Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* **33**, 755-760, doi:10.1038/nbt.3245 (2015).

70 Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat Biotechnol* **34**, 646-651, doi:10.1038/nbt.3528 (2016).

71 Solberg, N. & Krauss, S. Luciferase assay to study the activity of a cloned promoter DNA fragment. *Methods Mol Biol* **977**, 65-78, doi:10.1007/978-1-62703-284-1_6 (2013).

72 He, S. X., Song, G., Shi, J. P., Guo, Y. Q. & Guo, Z. Y. Nanoluciferase as a novel quantitative protein fusion tag: Application for overexpression and bioluminescent receptor-binding assays of human leukemia inhibitory factor. *Biochimie* **106**, 140-148, doi:10.1016/j.biochi.2014.08.012 (2014).

73 Norisada, J., Hirata, Y., Amaya, F., Kiuchi, K. & Oh-hashi, K. A sensitive assay for the biosynthesis and secretion of MANF using NanoLuc activity. *Biochem Biophys Res Commun* **449**, 483-489, doi:10.1016/j.bbrc.2014.05.031 (2014).

74 Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem Biol* **7**, 1848-1857, doi:10.1021/cb3002478 (2012).

75 Ibanez, P. *et al.* Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet* **364**, 1169-1171, doi:10.1016/S0140-6736(04)17104-3 (2004).

76 Grundemann, J., Schlaudraff, F., Haeckel, O. & Liss, B. Elevated alpha-synuclein mRNA levels in individual UV-laser-microdissected dopaminergic substantia nigra neurons in idiopathic Parkinson's disease. *Nucleic Acids Res* **36**, e38, doi:10.1093/nar/gkn084 (2008).

77 de Boni, L. *et al.* Next-generation sequencing reveals regional differences of the alpha-synuclein methylation state independent of Lewy body disease. *Neuromolecular Med* **13**, 310-320, doi:10.1007/s12017-011-8163-9 (2011).

78 Jowaed, A., Schmitt, I., Kaut, O. & Wullner, U. Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains. *J Neurosci* **30**, 6355-6359, doi:10.1523/JNEUROSCI.6119-09.2010 (2010).

79 Matsumoto, L. *et al.* CpG demethylation enhances alpha-synuclein expression and affects the pathogenesis of Parkinson's disease. *PLoS One* **5**, e15522, doi:10.1371/journal.pone.0015522 (2010).

80 Consortium, E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-640, doi:10.1126/science.1105136 (2004).

81 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

82 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

83 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766, doi:10.7554/eLife.04766 (2014).

84 Michel, P. P. & Hefti, F. Toxicity of 6-hydroxydopamine and dopamine for dopaminergic neurons in culture. *J Neurosci Res* **26**, 428-435, doi:10.1002/jnr.490260405 (1990).

85 Sharma, S. & Taliyan, R. Targeting histone deacetylases: a novel approach in Parkinson's disease. *Parkinsons Dis* **2015**, 303294, doi:10.1155/2015/303294 (2015).

86 Chuang, D. M., Leng, Y., Marinova, Z., Kim, H. J. & Chiu, C. T. Multiple roles of HDAC inhibition in neurodegenerative conditions. *Trends Neurosci* **32**, 591-601, doi:10.1016/j.tins.2009.06.002 (2009).

87 Williams, T. M., Burlein, J. E., Ogden, S., Kricka, L. J. & Kant, J. A. Advantages of firefly luciferase as a reporter gene: application to the interleukin-2 gene promoter. *Anal Biochem* **176**, 28-32, doi:10.1016/0003-2697(89)90267-4 (1989).

88 Wang, Y. *et al.* A DNA methyltransferase inhibitor, 5-aza-2'-deoxycytidine, exacerbates neurotoxicity and upregulates Parkinson's disease-related genes in dopaminergic neurons. *CNS Neurosci Ther* **19**, 183-190, doi:10.1111/cns.12059 (2013).

89 Christman, J. K. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483-5495, doi:10.1038/sj.onc.1205699 (2002).

90 Elliott, E., Ezra-Nevo, G., Regev, L., Neufeld-Cohen, A. & Chen, A. Resilience to social stress coincides with functional DNA methylation of the Crf gene in adult mice. *Nat Neurosci* **13**, 1351-1353, doi:10.1038/nn.2642 (2010).

91 Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295-304, doi:10.1038/nrg2540 (2009).

92 Walsh, P. S., Erlich, H. A. & Higuchi, R. Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods Appl* **1**, 241-250, doi:10.1101/gr.1.4.241 (1992).

93 Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun* **5**, 4767, doi:10.1038/ncomms5767 (2014).

94 Shifera, A. S. & Hardin, J. A. Factors modulating expression of Renilla luciferase from control plasmids used in luciferase reporter gene assays. *Anal Biochem* **396**, 167-172, doi:10.1016/j.ab.2009.09.043 (2010).

95 Kim, H. J. & Bae, S. C. Histone deacetylase inhibitors: molecular mechanisms of action and clinical trials as anti-cancer drugs. *Am J Transl Res* **3**, 166-179 (2011).

96 Xu, W. S., Parmigiani, R. B. & Marks, P. A. Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene* **26**, 5541-5552, doi:10.1038/sj.onc.1210620 (2007).

97 Davie, J. R. Inhibition of histone deacetylase activity by butyrate. *J Nutr* **133**, 2485S-2493S, doi:10.1093/jn/133.7.2485S (2003).

98 Chiba-Falek, O. & Nussbaum, R. L. Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system. *Hum Mol Genet* **10**, 3101-3109, doi:10.1093/hmg/10.26.3101 (2001).

99 Clough, R. L., Dermentzaki, G., Haritou, M., Petsakou, A. & Stefanis, L. Regulation of alpha-synuclein expression in cultured cortical neurons. *J Neurochem* **117**, 275-285, doi:10.1111/j.1471-4159.2011.07199.x (2011).

100 Clough, R. L., Dermentzaki, G. & Stefanis, L. Functional dissection of the alpha-synuclein promoter: transcriptional regulation by ZSCAN21 and ZNF219. *J Neurochem* **110**, 1479-1490, doi:10.1111/j.1471-4159.2009.06250.x (2009).

101 Clough, R. L. & Stefanis, L. A novel pathway for transcriptional regulation of alpha-synuclein. *FASEB J* **21**, 596-607, doi:10.1096/fj.06-7111com (2007).

102 Kumaki, Y., Oda, M. & Okano, M. QUMA: quantification tool for methylation analysis. *Nucleic Acids Res* **36**, W170-175, doi:10.1093/nar/gkn294 (2008).

103 Rohde, C., Zhang, Y., Reinhardt, R. & Jeltsch, A. BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics* **11**, 230, doi:10.1186/1471-2105-11-230 (2010).

104 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).

105 Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

106 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016).

107 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, doi:10.1126/science.aaf8729 (2016).

108 Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491, doi:10.1016/j.cell.2013.12.001 (2013).

109 Qin, P. *et al.* Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat Commun* **8**, 14725, doi:10.1038/ncomms14725 (2017).

110 Morgan, S. L. *et al.* Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat Commun* **8**, 15993, doi:10.1038/ncomms15993 (2017).

111 Friedman, J. R. *et al.* KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev* **10**, 2067-2078, doi:10.1101/gad.10.16.2067 (1996).

112 Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149, doi:10.1038/nmeth.3630 (2015).

113 Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10**, 977-979, doi:10.1038/nmeth.2598 (2013).

114 Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* **10**, 973-976, doi:10.1038/nmeth.2600 (2013).

115 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326-328, doi:10.1038/nmeth.3312 (2015).

116 Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350, doi:10.1016/j.cell.2014.11.052 (2015).

117 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).

118 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**, 635-646, doi:10.1016/j.cell.2014.09.039 (2014).

119 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257, doi:10.1016/s0092-8674(00)81656-6 (1999).

120 Kaminskas, E., Farrell, A. T., Wang, Y. C., Sridhara, R. & Pazdur, R. FDA drug approval summary: azacitidine (5-azacytidine, Vidaza) for injectable suspension. *Oncologist* **10**, 176-182, doi:10.1634/theoncologist.10-3-176 (2005).

121 Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247 e217, doi:10.1016/j.cell.2016.08.056 (2016).

122 Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* **44**, 5615-5628, doi:10.1093/nar/gkw159 (2016).

123 McDonald, J. I. *et al.* Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biol Open* **5**, 866-874, doi:10.1242/bio.019067 (2016).

124 Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol* **34**, 1060-1065, doi:10.1038/nbt.3658 (2016).

125 Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479, doi:10.1038/nature12750 (2013).

126 Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7**, 46545-46556, doi:10.18632/oncotarget.10234 (2016).

127 Lehninger, A. L., Nelson, D. L. & Cox, M. M. *Lehninger principles of biochemistry*. 6th edn, (W.H. Freeman, 2013).

128 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).

129 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).

130 Kearns, N. A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **12**, 401-403, doi:10.1038/nmeth.3325 (2015).

131 Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-517, doi:10.1038/nbt.3199 (2015).

132 Kwon, D. Y., Zhao, Y. T., Lamonica, J. M. & Zhou, Z. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat Commun* **8**, 15315, doi:10.1038/ncomms15315 (2017).

133 Cano-Rodriguez, D. *et al.* Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nat Commun* **7**, 12284, doi:10.1038/ncomms12284 (2016).

134 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

135 Holtzman, L. & Gersbach, C. A. Editing the Epigenome: Reshaping the Genomic Landscape. *Annu Rev Genomics Hum Genet* **19**, 43-71, doi:10.1146/annurev-genom-083117-021632 (2018).

136 Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863, doi:10.1038/nprot.2017.016 (2017).

137 Klann, T. S. *et al.* CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* **35**, 561-568, doi:10.1038/nbt.3853 (2017).

138 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20-36, doi:10.1016/j.cell.2016.10.044 (2017).

139 Park, M., Keung, A. J. & Khalil, A. S. The epigenome: the next substrate for engineering. *Genome Biol* **17**, 183, doi:10.1186/s13059-016-1046-5 (2016).

140 Pulecio, J., Verma, N., Mejia-Ramirez, E., Huangfu, D. & Raya, A. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* **21**, 431-447, doi:10.1016/j.stem.2017.09.006 (2017).

141 Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods* **13**, 127-137, doi:10.1038/nmeth.3733 (2016).

142 Brocken, D. J. W., Tark-Dame, M. & Dame, R. T. dCas9: A Versatile Tool for Epigenome Editing. *Curr Issues Mol Biol* **26**, 15-32, doi:10.21775/cimb.026.015 (2018).

143 O'Geen, H. *et al.* dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression. *Nucleic Acids Res* **45**, 9901-9916, doi:10.1093/nar/gkx578 (2017).

144 Eckner, R. *et al.* Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes Dev* **8**, 869-884, doi:10.1101/gad.8.8.869 (1994).

145 Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953-959, doi:10.1016/s0092-8674(00)82001-2 (1996).

146 Clayton, A. L., Hazzalin, C. A. & Mahadevan, L. C. Enhanced histone acetylation and transcription: a dynamic perspective. *Mol Cell* **23**, 289-296, doi:10.1016/j.molcel.2006.06.017 (2006).

147 Koyanagi, M. *et al.* EZH2 and histone 3 trimethyl lysine 27 associated with Il4 and Il13 gene silencing in Th1 cells. *J Biol Chem* **280**, 31470-31477, doi:10.1074/jbc.M504766200 (2005).

148 Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-874, doi:10.1038/nature04431 (2006).

149 Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043, doi:10.1126/science.1076997 (2002).

150 Zhang, X., Liu, L., Yuan, X., Wei, Y. & Wei, X. JMJD3 in the regulation of human diseases. *Protein Cell* **10**, 864-882, doi:10.1007/s13238-019-0653-9 (2019).

151 Burchfield, J. S., Li, Q., Wang, H. Y. & Wang, R. F. JMJD3 as an epigenetic regulator in development and disease. *Int J Biochem Cell Biol* **67**, 148-157, doi:10.1016/j.biocel.2015.07.006 (2015).

152 DiTacchio, L. *et al.* Histone lysine demethylase JARID1a activates CLOCK-BMAL1 and influences the circadian clock. *Science* **333**, 1881-1885, doi:10.1126/science.1206022 (2011).

153 Paigen, K. & Petkov, P. M. PRDM9 and Its Role in Genetic Recombination. *Trends Genet* **34**, 291-300, doi:10.1016/j.tig.2017.12.017 (2018).

154 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).

155 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646, doi:10.1016/j.cell.2014.09.039 (2014).

156 Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731-734, doi:10.1038/nature06145 (2007).

157 Wei, Y. *et al.* CDK1-dependent phosphorylation of EZH2 suppresses methylation of H3K27 and promotes osteogenic differentiation of human mesenchymal stem cells. *Nat Cell Biol* **13**, 87-94, doi:10.1038/ncb2139 (2011).

158 Klose, R. J. *et al.* The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128**, 889-900, doi:10.1016/j.cell.2007.02.013 (2007).

159 Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* **19**, doi:10.15252/embr.201846255 (2018).

160 Cleries, R. *et al.* BootstRatio: A web-based statistical analysis of fold-change in qPCR and RT-qPCR data using resampling methods. *Comput Biol Med* **42**, 438-445, doi:10.1016/j.compbiomed.2011.12.012 (2012).

161 *Disease Registry Statistical Tools - BootStratio*, <http://pdo.iconcologia.net/stats/br/index.html> (

162 Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat Med* **25**, 249-254, doi:10.1038/s41591-018-0326-x (2019).

163 Josephs, E. A. *et al.* Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* **44**, 2474, doi:10.1093/nar/gkv1293 (2016).

164 Sen, P., Shah, P. P., Nativio, R. & Berger, S. L. Epigenetic Mechanisms of Longevity and Aging. *Cell* **166**, 822-839, doi:10.1016/j.cell.2016.07.050 (2016).

165 Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* **21**, 525-530, doi:10.1101/gad.415507 (2007).

166 Shumaker, D. K. *et al.* Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proc Natl Acad Sci U S A* **103**, 8703-8708, doi:10.1073/pnas.0602569103 (2006).

167 Shah, P. P. *et al.* Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev* **27**, 1787-1799, doi:10.1101/gad.223834.113 (2013).

168 Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17-37, doi:10.1016/j.cell.2013.03.002 (2013).

169 Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153**, 38-55, doi:10.1016/j.cell.2013.03.008 (2013).

170 Bracken, A. P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* **22**, 5323-5335, doi:10.1093/emboj/cdg542 (2003).

171 Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43-48, doi:10.1038/nature11213 (2012).

172 Qi, J. Bromodomain and extraterminal domain inhibitors (BETi) for cancer therapy: chemical modulation of chromatin structure. *Cold Spring Harb Perspect Biol* **6**, a018663, doi:10.1101/cshperspect.a018663 (2014).