THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

DEVELOPMENT OF A CRISPR-CAS9 CONSTRUCT AIMED AT KNOCKING OUT SETD2 IN U2OS CELLS

NICHOLAS SCHULTHEIS
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Yanming Wang
Associate Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Lorraine Santy
Associate Professor of Biochemistry and Molecular Biology
Honors Adviser

Scott Selleck
Department Head for Biochemistry and Molecular Biology

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

SETD2 is a methyltransferase associated with chromatin remodeling that activates areas of the genome for gene expression. SETD2 is the sole trimethyltransferase capable of trimethylating H3K36, and in doing so allows RRM2 to be transcribed, a subunit of ribonucleotide reductase. SETD2 loss is found in a multitude of cancers and has far-reaching effects within them, often allowing them to proliferate and avoid p53-mediated apoptosis. CRISPR is a recently-discovered gene-editing technology that offers higher fidelity and more flexibility with regards to available targets. The purpose of this experiment was to develop a CRISPR-Cas9 product aimed at knocking out SETD2 in U2OS cells. This purpose appeared to have been met, as shown by markedly decreased H3K36 trimethylation. The CRISPR-Cas9 construct appeared to create a cell line of U2OS cells that were heterozygous for the wild-type SETD2 allele, as while the cell line demonstrated lesser H3K36 trimethylation, not all of it was lost.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my personal investigator Dr. Yanming Wang for taking me into his lab and giving me the opportunity to learn how to conduct research. I have learned not only how to perform procedures but also how to think through evidence critically and with an eye for the big picture. I would not have gained these skills had I not joined his lab, and for that I am incredibly grateful.

I would also like to thank my Honors Advisor, Dr. Lorraine Santy. She has helped me through two years of schooling in shaping both my classes and my resume. Her insight has been helpful throughout her time as my Honors Advisor, and I would not be where I am without her.

The entire Wang lab served to create a welcoming and helpful environment. In addition to Dr. Wang, Lai Shi, Jinquan Sun, Alexander Radaoui, Jiayuan Zhou, Lu Yang, and Haonan Xu all helped to contribute to my experience. In particular, Lai Shi taught me many of the particular procedures and answered many of my questions. His experience was vital to the benefit of mine.

Chapter 1

Introduction

Living organisms contain central information that directs the carrying out of biological processes within the organism and is passed from one generation to the next. This information is contained in molecules of deoxyribonucleic acid, otherwise known as DNA, and the functional unit of DNA is known as the "gene". While definitions for what exactly constitutes a gene vary, in general they refer to a continuous, inheritable piece of DNA with some downstream effect on the organism, or more specifically, "the entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide".[1] In effect, a gene is an irreducible piece of information that will affect biological processes.

Because of the gene's role as the basis from which living organisms continue living, there has been much interest in the realm of altering them. Proper manipulation of genes can lead to not only a deeper understanding of their function but also provide methods by which to treat illnesses and improve life. In particular, researchers have looked into the ability of gene-editing techniques to both knock out (KO) and knock in (KI) genes of interest. The KO of a gene will result in an ineffective and non-functional product, effectively removing it from the genome.[2] The KI of a gene allows, through careful manipulation of homologous recombination, for purpose and ability to be added to an area of a genome where there was none before. Specifically, introduction of the desired gene in a vector that facilitates homologous recombination in conjunction with KO of an area of the host's genome can lead to the desired gene being added to the genome of the host.[3]

One method that has been utilized in gene-editing procedures are the Zinc Finger Nucleases (ZFN). ZFN's were artificially synthesized using the DNA binding and recognition domains of eukaryotic transcription factors and the nuclease cleavage domain of Fok I taken from *Flavobacterium okeanokoites*.[4] These portions of the ZFNs come together to form two anti-parallel beta-sheets across from an alpha-helix, wherein the cleavage domain is attached to zinc fingers that recognize the DNA surrounding the desired cleavage site.[5] Because nine base pairs are recognized on each side of the cleavage site, ZFNs have an 18-base-pair specificity to the genome.[6]

ZFNs provide an effective method by which genomes can be altered, but they do suffer from a couple drawbacks. ZFNs find themselves unable to target genes in heterochromatin due to the compact nature of heterochromatin. Furthermore, ZFNs are often imperfectly specific and can often cleave in undesired locations due to the highly-variable nature of choosing the correct zinc fingers.[7] If the Zinc fingers cleave at too many off-target locations the treated cell will often die or undergo apoptosis. Another drawback to the ZFN approach is that it may be inhibited or otherwise affected by other protein domains around DNA surrounding the cleavage site. While ZFNs can work well when properly constructed, there is still an element of luck in its construction. The high rate of failure of ZFNs and its subsequent high cost pushes researchers away from the use of ZFN often.[8]

ZFNs are therefore not the only commonly-utilized gene-editing tool. Transcription activator-like effector nucleases (TALENs) are also available to researchers. TALENs are comprised of a non-specific DNA nuclease fused to domains with great affinity for genomic loci, in much a similar manner to ZFNs.[9] These domains that guide the TALENs come directly from transcription factors. Because the domains specific to DNA do not have to be

constructed from a library of zinc fingers and are instead constructed from an entire domain with high specificity, TALENs are not only more accurate than ZFNs they are also easier to design.[10] Of course, TALENs are not perfect and may still end up cleaving sequences other than the one intended. In particular, this can happen if the sequence the recognition domain recognizes is repetitive as this can lead to much of the genome being affected. TALENs also are sensitive to methylated DNA and may not work in areas of the genome that have been modified in such a way. Furthermore, TALENs can sometimes be difficult to deliver to cells because of its larger size.[2]

The most recent gene-editing technology to be developed uses Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) in conjunction with the Cas9 endonuclease. The entire complex comes evolved from prokaryotic genomes and prokaryotic defense mechanisms. Prokaryotes evolved this mechanism in order to defend against viral infection.[11] RNAs derived from CRISPR sequences, or crRNAs, are incorporated into the endonuclease complex in such a way as to be used to detect possible viral genomic loci and target them for cleavage. The full CRISPR-Cas9 complex involves two crRNAs and trans-acting antisense RNA. In conjunction, these sequences will guide the Cas9 endonuclease to the appropriate gene and in doing so generate a homologous double-stranded DNA break. The two crRNAs will often bond and form a single guiding molecule referred to as gRNA. One aspect of the Cas9 endonucleases are their requirement by which they must recognize a sequence motif referred to as the protospacer adjacent motif (PAM). The PAM is a set of nucleotides approximately 20 nucleotides downstream of the location of cleavage and is unique to each evolved Cas9 version. The Cas9 derived from *Streptococcus pyogenes*, which is used in many CRISPR-Cas9 system today, has a PAM of NGG.[12]

CRISPR-Cas9 systems do offer significant advantages over the ZFN and TALEN systems. Not only does it have comparable capacity to generate KIs and KOs of genes, but through highly specific utilization of gRNAs the CRISPR-Cas9 system works far more efficiently and with improved accessibility to the genes themselves.[2] This CRISPR technology allows researches who are incapable of engineering proteins from developing a molecule capable of editing genes.[13] Furthermore, the CRISPR-Cas9 complex is not hindered by DNA methylation, which both ZFNs and TALENs can be hindered by.[14] One of the few limitations to the use of CRISPR-Cas9 is the possibility of off-target cleavage, but this does not happen overly frequently.[2]

Relevant to the work performed in this thesis is the nature of epigenetics and the modification of histones. In eukaryotic cells, the DNA within the nucleus is bound to an extensive set of proteins referred to as histones. Histones are small clusters of proteins that can alternatively loosen or condense the DNA bound to them.[15] These histones have four separate proteins, arranged in an octameric fashion. H2A and H2B are two of these proteins, and they form two dimers with each other. H3 and H4 are the other two proteins, and they form a tetramer made out of two H3-H4 dimers.[16] Together, the four dimers make the octameric histone. The bound DNA-histone complexes are referred to as one molecule called chromatin. When the disparate histones are bound closely to each other, the chromatin is in a state called heterochromatin. When a section of DNA is in a heterochromatin state, it has far more limited gene expression, as the biological machinery needed to react with the DNA cannot physically interact with it. When the histones are loosely packed, the sections of DNA they are bound to have much greater gene expression. The aforementioned biological machinery can easily

interact with the DNA, and so through that interaction more RNA is transcribed. DNA that can be transcribed due to the loose nature of the composite histones is referred to as euchromatin.[17]

The manipulation of chromatin allows eukaryotes to add another layer of regulation to their already-extensively regulated genome and gene expression. Through covalent modification of the histones, the extent to which regions of DNA are available for transcription can be modified. In particular, common methods of chromatin modification involve methylation and acetylation of the histones. On each of the proteins that comprise the histone there is a modifiable tail, and it is on this tail that the methylation, acetylation, or citrullination occur.

Citrullination occurs when the imine group of an arginine in the histone tail is hydrolyzed and a keto group is attached. This reduces the positive charge of the histone, and due to the negative charge of DNA the histone becomes less bound to the DNA. Because of weaker binding the DNA then becomes more accessible to RNA polymerases and other transcription factors, thereby making the expression of the DNA more likely.[18] Acetylation is a modification performed on lysines within the tails of histones, wherein an acetyl group is added to the amine. The addition of the acetyl group neutralizes the positive charge of the lysine and affects the interaction between the histone and the DNA in much the same way citrullination does. As the positive charge is lost, the interaction with the DNA weakens and so therefore the chromatin becomes loosely-packed and more-easily transcribed.[19] Methylation is different from the other modifications, however, in that the addition of a methyl group does not radically change the charge of the amino acid to which it is added. Methylation affects the compaction of chromatin, though, in that proteins can detect methylation and subsequently shape the chromatin as biologically directed. Methylation can be carried out on both lysine and arginine, and while lysine can be mono-, di-, or tri-methylated, arginine can only be mono- or di-methylated.[20]

It is through trimethylation that the gene of interest of this project, SETD2, carries out its function. SETD2 trimethylates H3K36, which is the notation used to show that the 36th amino acid, a lysine, on the H3 tail is the affected area. This trimethylation is given the notation H3K36me3 and is generally involved with actively-transcribed genes. One of these actively-transcribed genes is the one for RRM2, which is a subunit of ribonucleotide reductase.[21]

Ribonucleotide reductase converts ribonucleic acids to deoxyribonucleic acids through the reduction of the hydroxide group on the 2' carbon of the ribonucleotide. The hydroxide group is lost and replaced with a hydrogen. The nucleotides are in the tri-phosphate state, and so NTPs are converted to dNTPs. Ribonucleotide reductase is not a singular protein, but is in fact a protein complex comprised of a multitude of protein subunits. One of these subunits is the protein RRM2, which is essential for ribonucleotide reductase function.[22]

One of the main regulatory pathways of RRM2 involves histone modification. Trimethylation of H3K36 recruits transcription initiation factors to RRM2, allowing RRM2 to be transcribed and therefore translated. SETD2 is the only methyltransferase for H3K36me3, and so therefore has a strong regulatory effect on the transcription of RRM2. Due to the effect of SETD2 on RRM2 transcription, loss of SETD2 can lead to incomplete ribonucleotide reductase within a cell, thereby decreasing the amount of available dNTPs. In particular, it has been seen that along with Wee1 inhibition cancer cells deficient in SETD2 suffer dNTP depletion, S-phase arrest, and apoptosis. The inhibition of Wee1 results in the reduction of RRM2 as well because CDKs are activated at times they should not be.[21]

A deficiency in dNTPs leads to stalling in S-phase because during DNA replication the DNA polymerase requires a constant supply of dNTPs for proper function. Without dNTPs, the DNA polymerase will stall and thereby force the cell to undergo apoptosis.[21]

SETD2 was chosen as the gene of interest because of its prevalence in cancers and associated prognoses. Lower expression of SETD2 is associated with worse outcomes in breast cancer and renal cancer, and SETD2 mutations are observed in more than ten percent of cancers in the kidneys, large intestines, endometrium, and ovaries. Furthermore, trimethylation of H3K36 is lost in 54 percent of child cancers of the brain and spinal cord.[21]

SETD2 has also been shown to be necessary for accurate DNA double-strand break repair.[23] Double-stranded DNA breaks threaten genomic stability and are considered the most dire form of DNA damage. Mammalian cells can repair DSBs through one of two pathways: nonhomologous end-joining (NHEJ), which results in errors in the gene sequence, and homologous recombination. Homologous recombination results in an accurate replacement of the broken DNA through the use of homologous sequences from other parts of the DNA. SETD2 is required for the proper recruitment of proteins during the homologous recombination process, and so accurate DNA repair cannot be performed without it.[23]

In addition to this, inactivation of SETD2 prevents the p53-mediated apoptotic pathway from completing. Mammalian cells undergo apoptosis when they suffer extensive DNA damage, and the checkpoint of this pathway is controlled by the p53 protein. However, in cells with inactivated SETD2 even extensive DNA damage will not result in the apoptosis of the cell. It has been shown that the checkpoint in the pathway reliant on p53 activity will not proceed properly with inactivated SETD2.[23]

The extensive and wide-reaching effects of SETD2 make it a valuable target for which a CRISPR-Cas9 construct could be made. It can serve not only as an opportunity to study the effects of SETD2 loss on cancerous cell lines that normally possess both wild-type copies of the SETD2 gene, but also as a way to test the efficacy of cutting-edge genome-editing

techniques. The creation of a cell line with SETD2 knocked out had presented difficulties in the past for Dr. Wang's laboratory, and the proper creation of such a cell line would offer the lab another resource. Understanding and overcoming the problems associated with editing the sequence of an oncogene is critical towards the downstream application of this technology, which is the treatment and possible cure of cancers associated with SETD2.

Chapter 2

Materials and Methods

SDS-PAGE[24]

1) Prepare samples in SDS loading buffer (6x: 300 mM Tris-HCl, 20% glycerol, 6% SDS, 4% β-mercaptomethanol, 0.6% bromophenol blue, pH6.8), heat samples at above 80°C for 5 min. Cool down on ice and briefly spin down.
2) Load samples in SDS-PAGE gel, leave one lane for protein marker (Crystalgen, 65-0671), load the lanes at each side with 1x SDS loading buffer.
3) Run the gel in 1x SDS running buffer (10x: 250 mM Tris-HCl, 1.92 M glycine, 1% SDS) at 210V for ~50 min, stop when the dye runs out of the gel.
4) 12% SDS-PAGE resolving gel preparation.

    30% Acrylamide: Bis (37.5:1) 4.4 ml

    1 M Tris-HCl, pH8.8: 4.1 ml

    20% SDS: 0.05 ml

    10% APS: 0.1 ml

    TEMED: 0.005 ml

    ddH2O: 2.35 ml

5) SDS-PAGE stacking gel (5 ml): 0.665 ml 30% Acrylamide: Bis (37.5:1), 0.63 ml 1 M Tris-HCl, pH 6.8, 0.025 ml 20 % SDS, 0.05 ml 10% APS, 0.005 ml TEMED, 3.62 ml ddH2O.

6) Water-saturated isobutanol: Shake equal volumes of water and isobutanol in a glass bottle and allow to separate.

Western blotting[24]

1) After SDS-PAGE, transfer protein to 0.2 μm nitrocellulose membrane (Whatman, BA85) in 1x Transferring buffer (25 mM Tris, 192 mM glycine, 20% methanol) using Semi-dry
Transferring Unit (GE, TE77).

2) Briefly stain the membrane with Ponceau S solution (0.1% Ponceau S in 5% HAc), wash away the background with ddH2O, scan the membrane, and cut to desired portions.

3) Wash away Ponceau S with TBST (10xTBS: 1.37 M NaCl, 27 mM KCl, 250 mM Tris-HCl, pH 7.4. Add 0.1% Tween-20 for TBST).

If use a-Modified Citrulline antibody, wash the membrane with water twice. Incubate the membrane in 0.1% Ovalbumin in TBS for 15 min at room temperature with agitation. Wash membrane twice with water. Prepare the Modification buffer by mixing same volume of Reagent A (0.025% FeCl3 in a solution of sterile, distilled water/98%H2SO4/85%H3PO4 (55%/25%/20%, store at 4°C or room temperature) and Reagent B (0.5% 2,3-butanedione monoxime, 0.25% antipyrine, 0.5 M HAc, store at -20°C, prevent from light). Immediately add the Modification buffer to membrane. Place

the membrane in a lightproof container and incubate at 37°C for at least 3 hr to overnight without agitation. Wash the membrane five times with water, 3 min each.

Proceed to following steps.

4) Block the membrane with the Blotting solution (5% nonfat dry milk in TBST) for 30 min at room temperature.

5) Incubate the membrane in the Blotting solution with the appropriately diluted primary antibody overnight (14-18 hr) at 4°C.

6) Next day take the membrane out, wash with TBST three times, 10 min each.

7) Incubate the membrane in the Blotting solution with the appropriately diluted horseradish peroxidase-conjugated secondary antibody for 2-3 hr at 4°C.

8) Take the membrane out, wash with TBST three times, 10 min each.

9) Incubate the membrane with the Lumi-LightPLUS ECL substrate (Roche, 12015196001) for 5 min.

10) Exposure and develop the film in the dark room.

11) To strip the membrane, incubate the membrane in the Stripping buffer (62.5 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM β-mercaptoethanol, 2% SDS) for 30 min at 50°C. Rinse the membrane with TBST.

Restriction enzyme digestion[24]

1) 50 µl system for plasmid vector digestion:

5 µl 10x buffer

x µl Plasmid vector (~1 µg)

1 µl Restriction enzyme 1

1 µl Restriction enzyme 2

Add ddH2O to 50 µl

2) 60 µl system for cDNA fragment

6 µl 10x buffer

48 µl cDNA PCR product

1 µl Restriction enzyme 1

1 µl Restriction enzyme 2

4 µl ddH2O

3) Digest both at 37°C for 2.5 hr

PCR

1) 15 µl reaction system:

1.5 µl 10x pfu buffer

0.5 µl DNA template (0.1-0.5 µg)

1.2 µl 2.5 mM dNTP mixture

0.65 µl 5 µM Forward primer

0.65 µl 5 µM Reverse primer

0.15 µl pfu

2) Add ddH2O to 15 µl

3) Typical PCR program:

94°C for 2 min

35 cycles of:

94°C for 30 sec

52.5°C for 30 sec

72°C for 1 min

Transformation

1) Plated U2OS in 2 ml of Dulbecco Modified Eagle Medium (DMEM) in each well of a six-well plate and incubate overnight
2) Add 22 µl of plasmid and 10 µl of lipofectamine to 500 µl of DMEM medium. This procedure was done six times, once for each well
3) DMEM medium in wells removed and DMEM with plasmid and lipofectamine added to each well and incubated for two days
4) In each well the DMEM was removed and replaced with 300 µl of DMEM that contained 0.2 µl of puromycin

Protein Isolation

1) DMEM was removed from cell culture
2) 5 ml of Phosphate-buffered Saline (PBS) was added to the colony of U2OS cells
3) The PBS was removed
4) 2 ml of Trypsin was added to the colony and incubated for 3 minutes
5) 10 ml of DMEM was added to the colony with Trypsin
6) Cell concentrations were counted

7) The colony was spun down in a centrifuge
8) The medium was removed from the pellet
9) 1 ml PBS was used to wash the pellet
10) The solution was spun down in a centrifuge
11) The PBS was removed
12) 300 µl of Digestion Buffer was added to the pellet
13) 3 µl of Protein Kinase A (PKA) was added to the pellet and the pellet was resuspended
14) 300 µl of phenol/chloroform/isoamyl alcohol was added
15) Centrifuged for 10 minutes at maximum speed
16) The aqueous top layer was transferred to a new tube
17) 2.8 µl of 3M NaAc and 560 µl of 100% ethanol was added
18) Centrifuged for 2 minutes at maximum speed
19) The supernatant was removed
20) The pellet was rinsed with 70% ethanol
21) The ethanol was decanted and the pellet was allowed to air dry
22) The pellet was solubilized in 50 µl of ddH_2O

Chapter 3

Results

CRISPR Target and Primer Design

The area of the SETD2 gene targeted for Cas9 repression was Exon 1 along the nucleotides 45411 to 46067. This section of DNA is important for SETD2 function as it is a portion of an exon, and its cleavage would result in the reduction of SETD2 activity. This section of DNA also carries the benefit of containing no sites which could be targeted by the EcoR1 or BamH1 restriction enzymes. The EcoR1 and BamH1 restriction enzymes were utilized to add the desired length of nucleotides to the CRISPR-Cas9 plasmid pSpCas9(BB)-2A-Puro (PX459).

During creation of the primers, the restriction site for BamH1 was added to the 5' end of the section of the forward primer complementary to the SETD2 gene. In a similar fashion, the restriction site for EcoR1 was added to the 5' end of the section of the reverse primer complementary to the SETD2 gene. Two nucleotides, a cytosine and a guanine, were added to the 5' end of the forward primer in order to create a stable primer end. One guanine was added to the 5' end of the reverse primer for the same reason. The primers as ordered from Integrated DNA Technologies are presented in **Figure 1**.

Forward Primer:

5′-CG|G GAT CC|C CTT TAA AAA GTG AGG ATC TAG-3′

BamH1

Reverse Primer:

5′-G|GA ATC C|AT TTA TCA GAC TTG GGT ATA GG-3′

EcoR1

Figure 1. Primers Used for SETD2 PCR

Primers were designed and used to copy and multiply the PCR gene for insertion into the CRISPR Plasmid. The nucleotides between the vertical lines represent the areas by corresponding to restriction enzyme sites.

PCR of SETD2

The primers were used to replicate the portion of the SETD2 gene of interest in a Polymerase Chain Reaction (PCR). The portion of the gene of interest was 667 base pairs long, and the additional nucleotides used for restriction enzyme sites and primer stability brought the entire length of the PCR product to 682 nucleotides. In order to determine the efficacy of the PCR the product was run on an agarose gel next to a 100-bp ladder. The two bright bands represent DNA of 1000 and 500 base pairs wherein the 1000-bp band is above the 500-bp band. The presence of a bright band in the loaded sample lane in **Figure 2** demonstrates the successful creation of a PCR product, and the demonstrated band size is consistent with the intended PCR product.

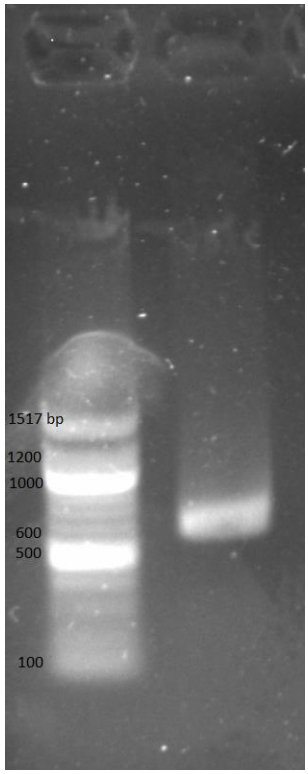

Figure 2. Agarose Gel of PCR Product

The product of the SETD2 PCR was run on a gel adjacent to a 100-base pair ladder.

Plasmid Ligation Verification

The SETD2 PCR product was incubated and digested with EcoR1 and BamH1 in NEBuffer 3.1. The pSpCas9(BB)-2A-Puro (PX459) plasmid was digested with the same enzymes and the two were ligated together. The ligated plasmid was treated with Bbs1 and Not1, other restriction enzymes, and the results were run on an agarose gel. The plasmid has a Not1 site unassociated with the location of insertion, but the Bbs1 site is within the portion of the plasmid removed by digestion by EcoR1 and BamH1. Ineffective digestion by EcoR1 and BamH1 would lead to the plasmid being cut in two places and two bands would appear on the agarose gel. However, the presence of one band in the gel would indicate that only the Not1 site was cut and therefore the digestion was effective. Only one band is present in the experiment, indicating the successful removal of the Bbs1 restriction enzyme site, as seen in **Figure 3**.

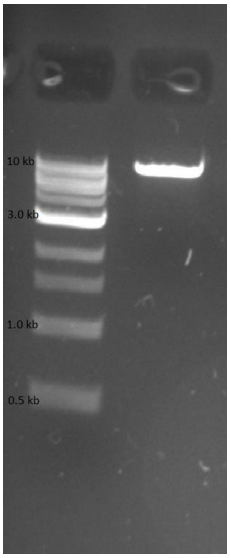

Figure 3. Agarose Gel of Ligation Product

Product of the plasmid and PCR product ligation run adjacent to a 1-kb DNA ladder

An illustration of the pSpCas9(BB)-2A-Puro (PX459) plasmid can be found in **Figure 4**.

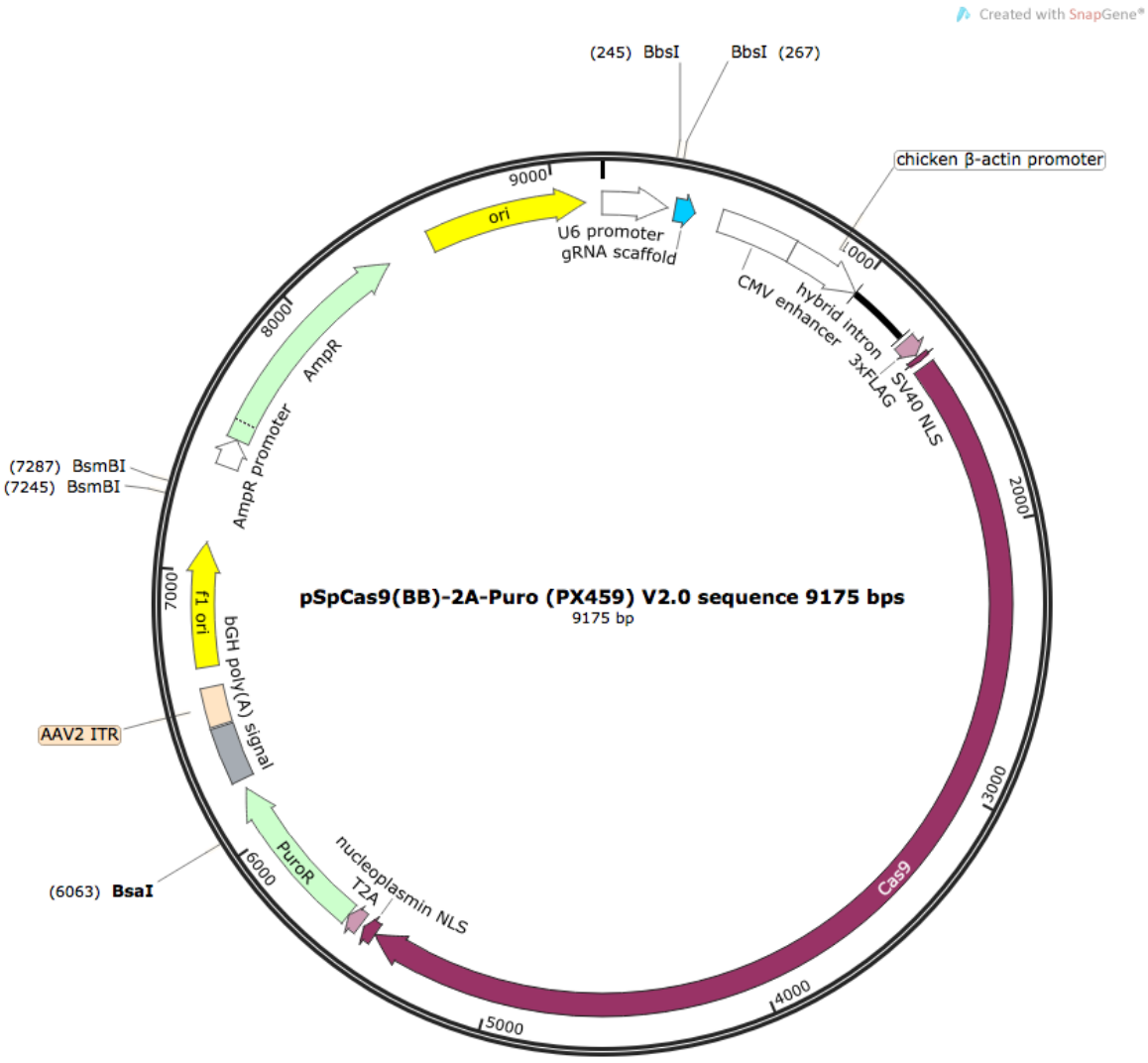

Figure 4. pSpCas9(BB)-2A-Puro (PX459) Plasmid

Supplied by Addgene[25], this graphic demonstrates the general organization of the pSpCas9(BB)-2A-Puro (PX459) plasmid, into which the SETD2 PCR product was inserted.

CRISPR Sequencing

Before transformation of the pSpCas9(BB)-2A-Puro (PX459) plasmid into U2OS cells, the completed plasmid was transformed into *E. coli* cells for the purposes of cloning. The plasmid carried with it ampicillin resistance, which it conferred to the *E. coli* cells. The *E. coli* cells were grown in LB medium with added ampicillin, and surviving colonies were subsequently collected and grown. Plasmid isolation was performed on the surviving colonies, and the collected DNA was sent to the Genomics Core Facility at the Huck Institutes of the Life Sciences of Penn State for confirmation on the presence of the plasmid. **Figure 5** contains the resultant sequence of interest:

Sequencing Result:

NNNNNNNNNCGATACNGGCTGTTAGAGAGANAATTGGAATTAATTTG
ACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAAT
AATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTAT
CATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATAT
CTTGTGGAAAGGACGAAACACCGACTCTGATCGTCGCTACCATGTTTTA
GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA
AAAGTGGCACCGAGTCGGTGCTTTTTTGTTTTAGAGCTAGAAATAGCAA
GTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGAC
AAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCC
TGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCC
AATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTA
TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCCCAGTACA
TGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
GCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCA
TCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTAT
TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGN

Figure 5. Sequencing Result of the Isolated Plasmid

The pSpCas9(BB)-2A-Puro (PX459) plasmid was cloned and isolated from E. coli, and the resulting sequence was recorded for confirmation.

This sequence was consistent with that of the plasmid, thereby indicating that the cloning of the plasmid was a success and could thereby be used to transform into mammalian cells.

Observations on the Growth of CRISPR-Transfected U2OS Cells

The pSpCas9(BB)-2A-Puro (PX459) plasmid also conferred puromycin resistance, which acted as a screen for the U2OS cells it was transformed into. The transfected cells were cultured in DMEM with added puromycin. The cell concentration was determined and the population was subsequently diluted to allow the cells to be aliquoted into three 96-well plates at a concentration of 0.5 cells per well.

A colony of cells was found and its growth was observed. As this colony grew, the rate by which it became confluent throughout the well was recorded, as it appeared to be multiplying at a reduced rate. The spread of the cells is displayed in **Figure 6**.

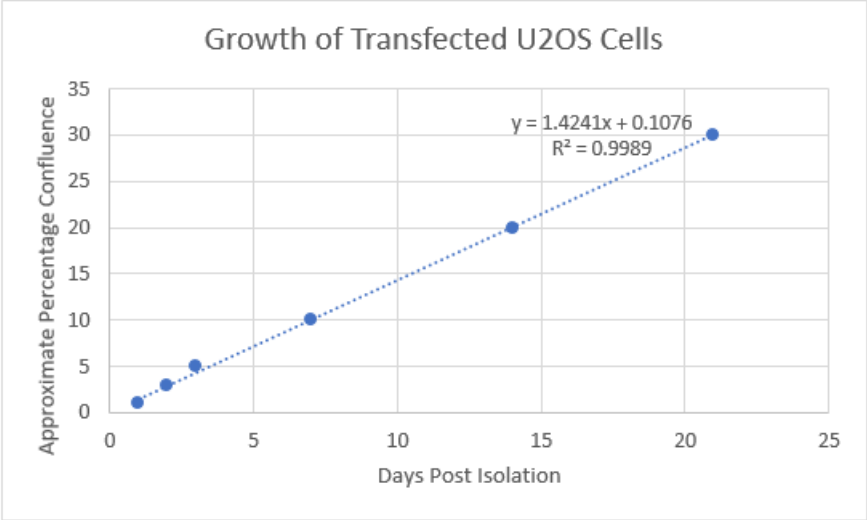

Figure 6. Growth of Transfected U2OS Cells

The approximate percentage confluence of transfected U2OS cells was recorded and graphed.

Preliminary U2OS Growth Test

Due to the curiosity of the lowered growth rate of transfected U2OS cells, a small test was performed regarding the growth of wild-type U2OS cells. A culture of U2OS cells was diluted and seeded in a 96-well plate such that the concentration was 0.5 cells per well. Three wells formed a colony, and after a 24-hour incubation period they started to multiply. Over the

first 24 hours there was no change in the number of cells in any of the wells, but in the subsequent 18 hours one well increased from five cells to eight, the second increased from three cells to six, and the third increased from five cells to fifteen.

Western Blot

The colony of U2OS cells that were transfected with the SETD2 CRISPR plasmid were isolated and allowed to grow. Protein isolation was performed on this colony and a western blot was run on wild-type U2OS cells and those that had been transfected with the SETD2 CRISPR. The western blot indicates the two populations have similar levels of α-Tubulin and the H3 protein, but the cells transfected with SETD2 CRISPR have markedly lower levels of H3K36 trimethylation, as seen in **Figure 7**.

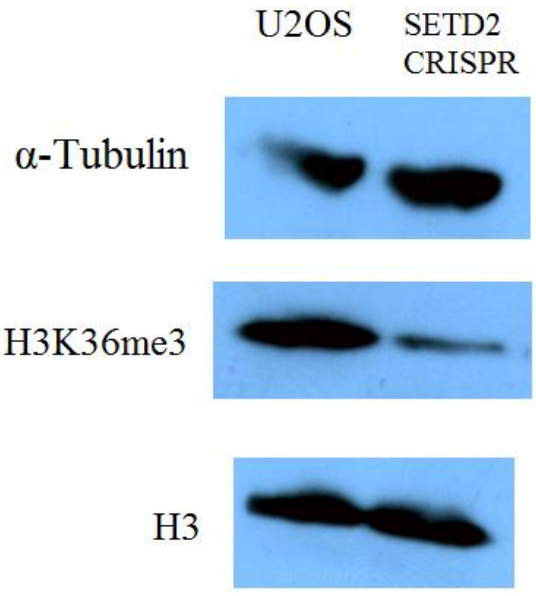

Figure 7. Western Blot of H3K36me3

Protein levels of α-Tubulin, H3K36me3, and H3 were measured and visualized through a Western Blot for WT U2OS and U2OS cells transfected with the SETD2 CRISPR plasmid.

Chapter 4

Discussion

The marked decrease in H3K36me3 between non-transfected and transfected U2OS cells demonstrates success of the CRISPR-Cas9 construct. As SETD2 is the sole methyltransferase of H3K36me3, loss of SETD2 is the singular gene that could be removed and result in decreased trimethylation of H3K36 without dramatic effect on all biological function. This is demonstrated by the consistent levels of H3 and α-Tubulin expression in both non-transfected and transfected U2OS cells. The decreased but present band of H3K36me3 in the transfected U2OS cells is an indication that some amount of SETD2 function is still being performed. This could possibly be caused by the CRISPR-Cas9 construct performing a heterozygous knockout of the gene as opposed to a complete homozygous knockout of both SETD2 alleles.

A heterozygous knockout of SETD2 alleles appears to be the most likely scenario to have occurred, which implicates the added CRISPR construct as being only partially successful in this particular cell line. A homozygous knockout of SETD2 should result in complete or near-complete loss of H3K36me3, which is not supported by the western blot of the transfected U2OS cell line. Complete failure of the CRISPR construct, however, should leave H3K36me3 levels unchanged, which is also not supported by the western blot. It is possible that homozygous knockouts place more stress on the cell than that caused by heterozygous knockouts, making heterozygous knockouts more likely to survive and form colonies. The difficulty associated in obtaining a viable colony post-transfection might attest to this. Most transfected and effective CRISPR constructs should theoretically result in the loss of both copies

of SETD2 as the construct searches the entire genome, but if a CRISPR construct were to only knock out one copy of SETD2 it may make the colony more viable. Perhaps the particular laboratory environment used in this experiment caused the stress to make homozygous knockouts nonviable. Another possibility is that homozygous knockouts of SETD2 in U2OS cells are generally nonviable, but this would contradict *Pfister et. Al.* Further experimentation in a range of environments would be required to determine the viability of U2OS cells post-transfection of a SETD2 CRISPR Plasmid.

While isolating a colony of U2OS cells that had undergone successful transformation with the SETD2 CRISPR plasmid, the growth rate had appeared curiously slow, which predicated an examination of the literature surrounding wild-type U2OS growth rates. Wild-type U2OS cells have a doubling time of approximately 23.7 hours in regular 96-well plates.[26] In cultures of monolayer cells the yield is approximately 50000 cells/cm^2 to 100000 cells/cm^2.[27] The bottom of wells in a 96-well plate have a diameter of 6.35 millimeters.[28] From these numbers one can determine a predicted rate by which the confluency of wild-type U2OS cells should increase.

Because of the known diameter of the well, the approximate area of the well can be calculated.

$$Radius = 6.35\ mm \times \frac{1\ cm}{10\ mm} \times \frac{1}{2} = 0.3175\ cm$$

$$Area = \pi r^2 = \pi(0.3175\ cm)^2 \cong 0.317\ cm^2$$

From the known area of the bottom of the well can be calculated the approximate minimum number of cells needed to achieve complete confluency.

$$0.5 \times 10^5 \frac{cells}{cm^2} \times 0.317\ cm^2 \cong 15800\ cells$$

Assuming a uniform doubling time and a starting colony of one cell allows the approximate number of doubling times required to reach complete confluency to be calculated.

$$2^{Number\ of\ Doublings} = 15800\ cells$$

$$Number\ of\ Doublings = \log_2 15800\ cells = \frac{\log 15800\ cells}{\log 2} \cong 14.0\ doublings$$

The doubling time of U2OS cells is 23.7 hours, and because of this, one may estimate it to be a full day, although the exact time for 14.0 doublings is as follows:

$$23.7 \frac{hours}{doubling} \times 14.0\ doublings = 331\ hours \times \frac{1\ day}{24\ hours} \cong 13.8\ days$$

Therefore, the expected time for wild-type U2OS cells to reach complete confluency is approximately 13.8 days. This can be compared to the predicted rate of confluency post-isolation based on the linear equation used to fit the data for U2OS cells transfected with SETD2 CRISPR plasmid.

$$Approximate\ Percent\ Confluency = 1.4241 \times Days\ Post\ Isolation + 0.1076$$

$$100 = 1.4241 \times Days\ Post\ Isolation + 0.1076$$

$$Days\ Post\ Isolation = \frac{(100 - 0.1076)}{1.4241} = 70.14$$

Based off of the rate of confluency increase in the transfected U2OS cell colony, it would require approximately 70 days to reach full confluency in a well with a diameter of 6.35 millimeters. This would appear to indicate a rate of growth far lower than that of wild-type

U2OS cells, which would be consistent with a lack of H3K36 trimethylation causing lowered transcription levels of RRM2 and therefore decreased replicative ability.

The preliminary test of wild-type U2OS cells appears to support the possibility that U2OS cells transfected with the SETD2 CRISPR plasmid will replicate much slower. Furthermore, the difficulty involved in finding even one viable colony of transfected cells would also support this possibility. Further experimentation is required for an accurate assessment, however, as only one colony of transfected U2OS cells was compared to a standard doubling rate for wild-type U2OS cells. The presence of puromycin may also have affected the growth rate of the transfected U2OS cells despite the resistance conferred by the CRISPR plasmid. Experiments comparing the growth rate of wild-type U2OS cells with transfected U2OS cells could include scenarios wherein both are transfected with puromycin resistance but only one population is also transfected with SETD2 knockout CRISPR plasmids. Growth rates in medium with no puromycin could also be examined.

The test for the success of the restriction enzyme digestion and ligation with the SETD2 PCR product carries a particular amount of interest because it demonstrates not only the success of the restriction enzyme digest, as mentioned before, but also the success of the annealing of the SETD2 PCR product into the plasmid. This can be seen because of the lack of a band near the area of the DNA ladder corresponding to DNA approximately 700 base pairs long. If the PCR product had not properly annealed to the plasmid then a noticeable band in that location should be present, as the PCR product is approximately 682 base pairs long.

The size of the one present band also reinforces the evidence that the annealing between the plasmid and the PCR product was successful. As the plasmid has 9175 base pairs and the PCR product as 682 base pairs, the two annealed together should have approximately 9875 base

pairs, which is consistent with the band size seen in the gel. It is close in line with the 10000-base pair band in the DNA ladder, as one would expect from a band of DNA of size 9875 base pairs.

The next step by which one could check the success of the CRISPR construct would be to sequence the SETD2 gene out of the U2OS cell line previously transfected with the Cas9 plasmid. One would first isolate the genomic DNA of the cells, then PCR the SETD2 gene and isolate the PCR product. The PCR product could then be inserted into a plasmid and cloned in *E. coli* culture before being sent for sequencing.

A U2OS cell line heterozygous for SETD2 allows for further investigations and uses to be performed. Because of the multitude of cancers associated with SETD2 loss, this allows for research to be performed on this U2OS cell line that can be applied towards other cancers with similar genotypes. For example, previous work has shown that the drug Wee1 targets SETD2-deficient cancers with high efficacy. Similar drugs could be tested or examined in this way, allowing for rather rapid assessments of the utility of new treatments.

In a similar fashion, SETD2-deficient U2OS cells can be utilized to examine their influence on living organisms. Insertion into immunodeficient mice and subsequent dissection and analyzation can provide insight into the mechanisms by which cancers grow and spread, especially those cancers missing copies of SETD2, of which there are many.

The creation of this construct and the subsequent slow growth of the transfected U2OS cells has raised the question about the speed by which cancer cells lacking wild-type SETD2 copies multiply and grow. SETD2, while it is associated with DNA repair, is also associated with the transcription of RRM2. Proper transcription of RRM2 is needed for DNA replication, so the viability assay and this train of thought both indicate that loss of SETD2 will

cause slower cancer growth rates. Future research should look into the viability and growth rates of different cancers based on the presence or loss of SETD2.

Difficulties Faced and Overcome

Pfister et. Al. discuss in their paper that cancer cells with missing SETD2 copies are viable and capable of thriving. Only under Wee1 inhibition do they theoretically lose their ability to reproduce as RRM2 production halts. However, during conductance of this research there were difficulties in developing a cell line of U2OS cells with knockout of SETD2. During the screening and incubation progress the transformation and incubation procedure had to be repeated multiple times on new sets of U2OS cells. The addition of the CRISPR plasmid consistently resulted in non-replicating cell populations, which would appear to be incongruent with what was reported in *Pfister et. Al.*

Eventually one colony of cells was obtained, which was incubated and for which the above western blot was performed. Based on the levels of H3K36 trimethylation the conclusion appears to be valid that the cell line generated was a heterozygous knockout of SETD2, and even this cell line demonstrated a reduced ability to multiply and proliferate.

Chapter 5

Conclusion

There is significant evidence to indicate that through this procedure a proper CRISPR-Cas9 construct has been created capable of knocking out SETD2 in U2OS cells. In the construction of this CRISPR, a cell line of U2OS cells has been generated that is likely heterozygous for Wild-Type SETD2 alleles. This is seen through lowered, but not eliminated, levels of H3K36me3 present in the cell. Both the cell line and the construct itself may be used to further the knowledge of not only SETD2-associated cancers but also for the current leading edge of gene-editing technology.

BIBLIOGRAPHY

[1] Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000. Section 9.1, Molecular Definition of a Gene. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21640/

[2] Gilles AF, Averof M. Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. EvoDevo. 2014;5:43

[3] Maresca M, Lin VG, Guo N, et al. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through non-homologous end joining. Genome Res. 2013;23:539–546.

[4] Bitinaite J, Wah D, Aggarwal AK, et al. FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci USA. 1998;95:10570–10575.

[5] De Souza N. Primer: genome editing with engineered nucleases. Nat Methods. 2011;9:27.

[6] Carroll D. Genome engineering with zinc-finger nucleases. Genetics. 2011;188:773–782.

[7] Sander JD, Dahlborg EJ, Goodwin MJ, et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods. 2011;8:67–69

[8] Maeder ML, Thibodeau-Beganny S, Osiak A, et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell. 2008;31:294–301.

[9] Li T, Liu B, Spalding MH, et al. High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol. 2012;30:390–392.

[10] Ain QU, Chung JY, Kim Y-H. Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. J Control Release. 2014;205:120–127.

[11] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709–1712. http://dx.doi. org/10.1021/sb500255k.

[12] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual RNA guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–821

[13] Doench JG, Hartenian E, Graham DB, et al. Rational design of highly active sgRNAs for CRISPR-Cas9mediated gene inactivation. Nat Biotechnol. 2014;32:1262–1267.

[14] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014;157:1262–1278.

[15] Bhasin M, Reinherz EL, Reche PA (2006). "Recognition and classification of histones using support vector machine". Journal of Computational Biology. 13 (1): 102–12. doi:10.1089/cmb.2006.13.102. PMID 16472024.

[16] Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (Sep 1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution". Nature. 389 (6648): 251–60. doi:10.1038/38444. PMID 9305837. PDB: 1AOI

[17] Strahl BD, Allis CD (Jan 2000). "The language of covalent histone modifications". Nature. 403 (6765): 41–5. doi:10.1038/47412. PMID 10638745.

[18] Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M, Schneider R, Gregory PD, Tempst P, Bannister AJ, Kouzarides T (Sep 2004). "Histone deimination antagonizes arginine methylation". Cell. 118 (5): 545–53. doi:10.1016/j.cell.2004.08.020. PMID 15339660.

[19] Koch CM, Andrews RM, Flicek P, Dillon SC, Karaöz U, Clelland GK, Wilcox S, Beare DM, Fowler JC, Couttet P, James KD, Lefebvre GC, Bruce AW, Dovey OM, Ellis PD, Dhami P, Langford CF, Weng Z, Birney E, Carter NP, Vetrie D, Dunham I (Jun 2007). "The landscape of histone modifications across 1% of the human genome in five human cell lines". Genome Research. 17 (6): 691–707. doi:10.1101/gr.5704207

[20] Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (May 2007). "High-resolution profiling of histone methylations in the human genome". Cell. 129 (4): 823–37. doi:10.1016/j.cell.2007.05.009. PMID 17512414.

[21] Pfister, S. X., Markkanen, E., Jiang, Y., Sarkar, S., Woodcock, M., Orlando, G., … Humphrey, T. C. (2015). Inhibiting WEE1 Selectively Kills Histone H3K36me3-Deficient Cancers by dNTP Starvation. Cancer Cell, 28(5), 557–568. http://doi.org/10.1016/j.ccell.2015.09.015

[22] Elledge SJ, Zhou Z, Allen JB (March 1992). "Ribonucleotide reductase: regulation, regulation, regulation". Trends Biochem. Sci. 17 (3): 119–23. doi:10.1016/0968-0004(92)90249-9. PMID 1412696.

[23] Carvalho, S., Vítor, A. C., Sridhara, S. C., Martins, F. B., Raposo, A. C., Desterro, J. M., … de Almeida, S. F. (2014). SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. *eLife*, *3*, e02482. http://doi.org/10.7554/eLife.02482

[24] Li, Pingxin. (2010). *Analysis of Peptidylarginine Demininase 5 and Histone Citrullination in Transcriptional Regulation and Innate Immunity* (Doctoral Dissertation). Retrieved from the PennState Electronic Theses and Dissertations Database.

[25] Genome engineering using the CRISPR-Cas9 system. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. *Nat Protoc. 2013 Nov;8(11):2281-308. doi: 10.1038/nprot.2013.143. Epub 2013 Oct 24.* 10.1038/nprot.2013.143

[26] Solly K., Wang X., Xu X., Strulovici B. & Zheng W. Application of real-time cell electronic sensing (RT-CES) technology to cell-based assays. Assay and drug development technologies 2, 363–372, doi: 10.1089/adt.2004.2.363 (2004).

[27] ATCC. (2016). *ATCC® Animal Cell Culture Guide*. Retrieved from https://www.atcc.org/~/media/PDFs/Culture%20Guides/AnimCellCulture_Guide.ashx

[28] Corning. (2015). *96 Well Microplates*. Retrieved from http://csmedia2.corning.com/LifeSciences/Media/equipment_compatibility/MD_Microplate_Dimension_Sheets.htm

ACADEMIC VITA

Academic Vita of Nicholas Schultheis
schultheisnicholas@gmail.com

Education
Major(s) and Minor(s): Biochemistry and Molecular Biology

Thesis Title: DEVELOPMENT OF A CRISPR-CAS9 CONSTRUCT AIMED AT KNOCKING OUT SETD2 IN U2OS CELLS
Thesis Supervisor: Yanming Wang

Work Experience
2014-2017
Undergraduate Assistant
Helped with research and analyzed data sets to determine patterns and scale of replication, and extracted tumors from oncogenic mice and determined gene expression within them
Pennsylvania State University at University Park
Yanming Wang

Awards: Provost Award