THE ROLE OF KMT5C ON EGFR INHIBITOR RESISTANCE IN NON-SMALL CELL LUNG CANCER

by

Alejandra Agredo Montealegre

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Biological Sciences

West Lafayette, Indiana

December 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Andrea L. Kasinski, Chair

Department of Biological Sciences

Dr. Scott Briggs

Department of Biochemistry

Dr. Ourania Andrisani

Department of Basic Medical Sciences

Dr. Qing Deng

Department of Biological Sciences

Approved by:

Dr. Janice P. Evans

I dedicate my PhD work to my joyful family,

Marleny Montealegre and Harold Agredo,

My supportive sister Angélica Agredo,

and of course, to my beloved husband Sebastián and my loyal PhD companion Coco

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to the one and only Dr. Andrea Kasinski. Thank you for believing in me when I was just an undergraduate student from Colombia, with little to no laboratory experience. I have learned from you the true meaning of doing science and critical thinking. I really value the time and effort you put into each one of us. You have been key in improving our presentations and writing skills, while always pushing us to think ahead and be critical with our science. I will always admire your passion for science and your mentorship style. Finally, I thank you for your sense of humor, as laughter probably makes us more efficient.

Second, I would like to thank my committee members for all their support and scientific feedback. I thank Dr. Ourania Andrisani for allowing me to rotate in her laboratory and providing unconditional support in every committee meeting. I would also like to thank my committee members Dr. Qing Deng and Dr. Scott Briggs for providing constructive feedback and great ideas for the success of my project. I would also like to thank my previous committee member, Dr. Anderson, for all the great questions and creative ideas. Finally, it has been an honor to engage in discussions about my research with such brilliant scientists.

Third, I would like to especially thank Dr. Arpita Pal for letting me participate in her project since the beginning in 2016. I really must thank you for all you have taught me in and out of the laboratory. I really value our professional and personal relationship, and hope our collaboration continues in the future. Fourth, I would like to thank my amazing laboratory. The Kasinski lab is an actual family, where we all support each other in good and bad times. Especially, I would like to thank Zulaida Soto, Humna Hasan, and Ahmed Abdelaal for being such great friends and supporting me during my entire PhD. I would also like to thank Jihye Son for helping me finish the review experiments and being such a kind and smart mentee. I would like to thank Ikjot Sohal for always being open to answering all my questions with a great attitude. Finally, I would like to thank the entire lab, previous and present members, for their respect for each other and for making going to work every day a pleasant experience. The Kasinski lab is the perfect example of what teamwork really looks like.

Finally, I would like to thank my family and my loved ones. I want to thank my parents for always believing in me, and for always making me realize that my well-being is more important than anything else. This PhD is a testament to your dedication in providing us with the finest education possible. I would also like to thank my sister for always listening to my struggles, knowing that you also went through the same challenges in your PhD helped me tremendously. I would also like to thank my family-in-law, Gladys, Nelson, and Nicolas, for always supporting me from afar. Last but not least, I would like to especially thank Sebastián for never letting me fall and always believing in us. I am really lucky to have shared this life-changing experience alongside you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Lung cancer is the leading cause of cancer-related deaths, and although important therapy advancements have been achieved, ~1.6 million people die from lung cancer annually. Non-small cell lung cancer (NSCLC), which makes up ~85% of lung cancer cases, is mainly treated with radiotherapy, chemotherapies, and targeted agents. Targeted agents are selected based on the mutation spectrum of the tumor. In NSCLC the epidermal growth factor receptor (EGFR) is commonly mutated and, leads to increased proliferation and cell survival. The standard-of-care treatment for patients with activating mutations in EGFR is treatment with tyrosine kinase inhibitors (TKI), such as erlotinib. While tumors initially respond to TKIs, after 1-2 years most patients develop resistance. In ~60% of TKI resistant tumors, resistance is the result of a secondary mutation in EGFR, whereas in the remaining 20%, tumors turn on bypass track-signals to overcome inhibition of the EGFR pathway. However, 15-20% of the cases the mechanisms underlying resistance are unknown. Most studies focus on the gain of function of oncogenes as mediators of resistance; however, little is known about the role that tumor suppressors play in TKI resistance. Hence, we performed a genome-wide CRISPR Cas9 knock-out screen to identify genes that when knocked-out would drive erlotinib resistance, and KMT5C was identified as the top candidate. KMT5C is a histone methyltransferase that trimethylates H4K20 (H4K20me3), enabling the establishment of constitutive and facultative heterochromatin. Data from human samples suggests that the *KMT5C* transcript is globally downregulated in NSCLC and in tumor samples resistant to the third generation TKI osimertinib. Additionally, loss of the modification H4K20me3, influences prognosis of NSCLC, indicating that loss of KMT5C function is a crucial mechanism in carcinogenesis. Here we describe how loss of KMT5C leads to increased transcription of the oncogene MET, due to a loss in H4K20me3-mediated repression of a long non-coding RNA transcription (LINC01510) upstream of MET. This mechanism was found to be partially responsible in driving TKI resistance in EGFR mutant cells. Historically, KMT5C has been associated with generation of constitutive heterochromatin (cHC); however, recent reports, including our own, indicate that KMT5C also regulates transcription in regions outside of cHC. Our preliminary evidence suggests that deposition of H42K0me3 via KMT5C in regions outside of cHC, is less stable than in cHC regions. This novel finding led us to hypothesize that regulation of KMT5C and H42K0me3 at different regions of heterochromatin is a dynamic process.

CHAPTER 1. INTRODUCTION

1.1 Chapter Overview

In accordance with the Frontiers in Genetics policy on author use, the second part of this introduction is modified from the following publication (Agredo A., & Kasinski A. L. 2023) (https://doi.org/10.3389/fgene.2023.1243395) and has been reprinted with permission from Frontiers in Genetics.

1.2 Lung cancer resistance

1.2.1 Lung cancer

Cancer statistics

Cancer is one of the most common diseases worldwide and approximately 609,820 deaths from cancer are expected in the United States (US) in 2023. It is the second most common cause of death in the US, following heart disease. These statistics, however, do not consider the impact that the COVID-19 pandemic had on cancer diagnostics, due to the disruption of health services affecting millions of missed cancer screenings and follow-up exams. Moreover, substantial progress has been achieved against cancer. The overall cancer incidence has reduced since the 20th century, due to reduction of smoking, early detection, and advancement in treatments. Additionally, the population that is most at risk of developing cancer is people that are 65 or older, and lifestyle factors such as smoking, excess body weight, alcohol consumption and unhealthy diets can increase this risk. [1]

Cancer survival is generally described in terms of life expectancy of cancer patients compared to the general population of the same age, race and sex. The 5-year relative survival rate for all cancers has increased since the 1960 and this can be explained by improvement in treatments and early diagnosis. However, these rates do not take into account novel advances in detection technologies and treatment, and external factors such as safe-housing, healthy food, and behavioral differences. [1] Additionally, cancer can be defined by the extend and level of spread of the cancer

at the time of the diagnosis. Staging takes into consideration factors such as size of the primary tumor and spread to other areas of the body indicating metastasis. If cancer cells started growing but do not have the capacity or there hasn't been evidence of penetrating other tissues, it is called in situ. If the cells have managed to penetrate beyond the original tumor tissue the it is considered invasive and can further be categorized as local, regional, or distant[1].

One of the deadliest cancers is lung cancer. It is the second most diagnosed cancer in both men and women, and it leads to more deaths than prostate, breast and colorectal cancers combined. Hence further understanding of the mechanisms involved in lung cancer progression is needed for improvement and design of better therapies.

1.2.2 Non-small Cell Lung Cancer

As previously mentioned, lung cancer is the most common cause of cancer death worldwide with an estimated 1.6 million deaths each year. Lung cancer can be divided into two types, depending on their histological differences. 85% of patients have a histological subtype called Non-small Cell Lung Cancer (NSCLC) and 15% of patients have the subtype Small Cell Lung Cancer (SCLC). NSCLC can further be divided into different subtypes, of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common ones. Moreover, 80% of the cases in the United States and other countries of lung cancer can be attributed to tobacco smoking. However, lung cancer in non-smokers can also occur and it is more common in women and in East Asia. Non-smoker lung cancer is often associated with environmental exposures including pollution, second-hand smoking, exposure of carcinogens and inherited genetic susceptibility. Therefore, strategies to reduce environmental exposure to carcinogens and tobacco prevention programs are essential to prevent and fight lung cancer.

Progress has been achieved for the treatment of NSCLC thanks to the development of targeted therapies and immunotherapies in some patients with advanced NSCLC. However, major challenges still remain, including identification of new molecular targets to broaden the population that would benefit from targeted therapies, understanding mechanisms of resistance to those targeted therapies, and identification of better predictors of response to immunotherapies[2].

1.2.3 Epidermal Growth Factor Receptor

NSCLC can be caused by a variety of different types of mutations. In LUAD the most commonly mutated genes include KRAS (~32%) and epidermal growth factor receptor (EGFR) (~27%), followed by mutations in the tumor suppressors TP53 (~46%), KEAP1(~19%), and NF1(~11%)[2]. EGFR mutations occur in 40%-60% of South-East Asian patients or 10%-20% of Caucasian patients with LUAD. EGFR mutations are more commonly found in women due to the prevalence of this mutation in never smokers or light smokers. Moreover, EGFR is part of a family of receptor tyrosine kinase that include HER2 and HER4. This receptor contains four extracellular domains, a tyrosine kinase domain, a transmembrane domain, and a carboxyl tail. Binding of the ligand EGF leads to EGFR dimerization and trans-phosphorylation of tyrosine residues in the carboxy tail. The latter leads to activation of downstream signaling pathways involved in cell proliferation and tumor growth, survival, invasion, and angiogenesis[2]. Mutations in the ATP-binding pocket of the tyrosine kinase domain may lead to constitutive EGFR activation and ligand independence leading to uncontrolled cell growth. Most of the EGFR mutations occur in exons 18-21 of the tyrosine kinase domain of the receptor. In-frame deletion of exon 19 and L858R substitution point mutation in exon 21 are the most common types of mutations. Other types of EGFR mutations are also found, although they tend to be less common, such as G718X, S768I, L861Q among others[3].

1.2.4 EGFR Tyrosine Kinase Inhibitors

EGFR mutations in the tyrosine kinase domain such as deletion of exon 19 and L858R are both sensitizing mutations to tyrosine kinase inhibitors drugs. In the late 1990s tyrosine kinase inhibitors revolutionized treatment of EGFR mutant lung cancer due to its success in the clinic. First generation tyrosine kinase inhibitors, include gefitinib and erlotinib which were observed to have higher progression-free survival than other cytotoxic therapies used at the time. Both TKIs are quinazoline-based derivatives and act as competitive inhibitors that bind tyrosine kinase pocket of EGFR in a reversible manner. Additionally, they have been shown to be very effective in patients that harbor EGFR mutations, but activity in patients with EGFR wild type is very limited. On the other hand, second generation TKIs such as dacomitinib and afatinib are irreversible inhibitors that also target the HER2 and HER4[3]. Both dacomitinib and afatinib have shown better survival compared to gefitinib, however different patient outcomes seem to be dependent on the

EGFR mutation. These differences can be attributed to distinct conformation changes within the ATP-binding pocket and patterns of auto-phosphorylation induced by each mutation[2].

Despite the success of EGFR targeted therapies, acquired resistance mechanisms of first- and second-generation EGFR TKIs have been extensively characterized. The most common cause of acquired resistance (~50% of resistance cases) is a secondary mutation in EGFR exon 20, with a threonine-to-methionine substitution on codon 790 (T790M). This mutation is responsible for increasing the affinity of the tyrosine kinase domain for ATP, which leads to phosphorylation of EGFR and activation of downstream signaling. The standard care for patients with such mutation is the use of the third generation EGFR-TKI osimertinib Osimertinib irreversibly inhibits the activity of both L858R and exon 19 deletion, and the T790M resistance mutation, but has weaker inhibitor activity against EGFR wild type[4]. Nevertheless, patients also develop resistance to the third generation TKI osimertinib (**Figure 1.1**).

Other causes for first- and second-generation resistance include amplification in HER2 or mutations in MET, BRAF or PI3KCA as well as SCLC transformation. However, 15-20% of mechanisms of resistance to the first generation TKI erlotinib, remain to be elucidated[5]. Most of the studies have focused on identification of oncogenes as drivers of resistance, due to their potential use as therapeutic target. While loss of tumor suppressor genes have also been identified as drivers of resistance, including PTEN, TP53, TET1, and NF1[6–9], further research is needed in better understand the role of tumor suppressors in driving erlotinib resistance.

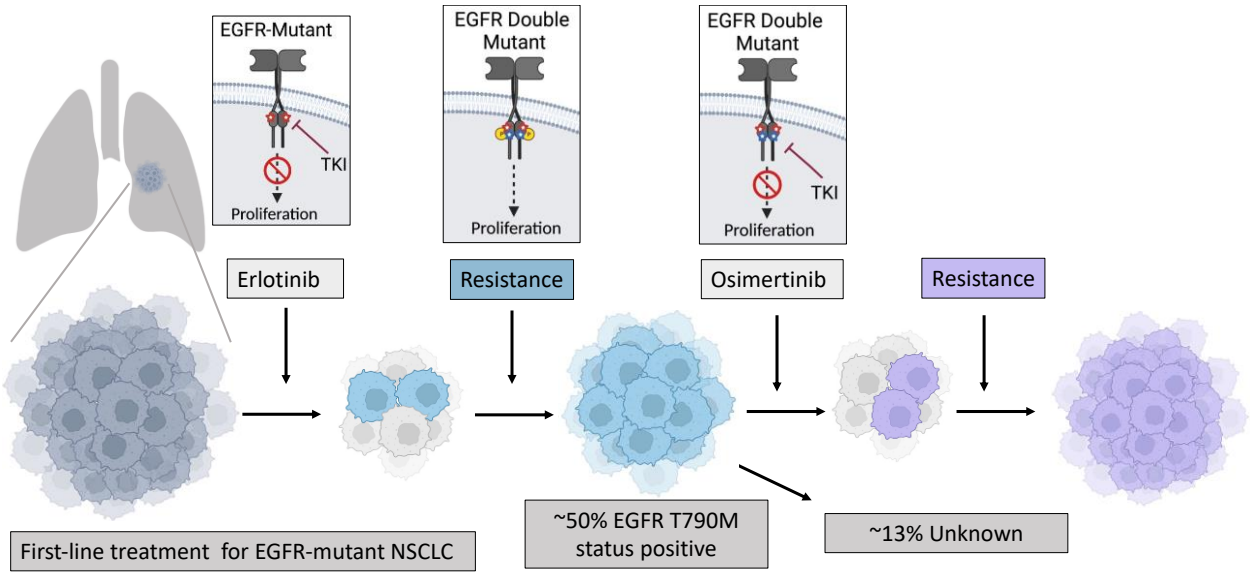

Figure 1.1. Visualization of resistance tumor progression after treatment with EGFR inhibitors erlotinib and osimertinib. After erlotinib treatment, some resistant cells persist (in blue) and continue to grow, forming a resistant tumor. Patients that harbor the secondary mutation T790M in the EGFR receptor are treated with osimertinib. Some cells persist after osimertinib treatment (in purple) which then continue to grow, generating a resistant tumor.

1.2.5 Epigenetics in lung cancer

Mutations and changes in gene expression have been identified as main causes of NSCLC, however, the molecular mechanisms of how changes in gene expression are regulated in cancer cells vs normal cells is still not fully understood. Recent studies have started to better elucidate the role of epigenetics in cancer progression, where chromatin alterations seem to play an important role in tumor formation and progression by regulating gene expression. More specifically, reports indicate that patterns of altered gene expression that do not affect primary DNA sequences can have a detrimental impact in cell homeostasis. In fact, in colon, lung and prostate tumors, pre-malignant cells undergo genetic alterations that might be determined by early epigenetic changes, thereby supporting growth progression of these clones[10]. For instance, changes such as epigenetic silencing have been observed in the early stages of human tumor progression. Early gene-silencing events might be essential for inducing early clonal expansion and might make cells addicted to a certain oncogenic pathway, thereby promoting accumulation of genetic mutations driving tumor progression. More and more studies are emerging as the field of chromatin regulation and gene expression advances. This deeper exploration of the cancer-cell "epigenome" is crucial for gaining

a comprehensive understanding of its role in cancer, paralleling the extensive study of mutations that has been conducted over the years. Examples of epigenetic alterations that have been implicated in cancer progression are loss and gain in DNA methylation, as well as histone modifications such as histone methylation[10].

Histone methylation has started to gain interest due to its reversibility when compared to other more stable epigenetic modifications such as DNA methylation. Histone modifications are mostly affected by histone-modifying enzymes that can either add or remove them. Dysregulation of these enzymes can then result in an unbalanced histone modifications pattern, and correcting their expression should be able to restore cell homeostasis[11]. For example, cancer cells are characterized by dysregulation of histone methyltransferases and histone demethylases, overexpression of deacetylases (HDACs), and reduction in histone acetylation levels[10]. H4K20me3, one of the primary histone modifications linked to cancer progression[12] and drug resistance[13],will be thoroughly examined in this document. Accordingly, we will describe its molecular mechanism in normal cells, its regulation by other molecular players, and review its significance in disease progression.

1.3 Histone 4 Lysine 20 tri-methylation: A key epigenetic regulator in chromatin structure and disease

This section is modified from my publication in Frontiers in Genetics, section Epigenomics and Epigenetics. (https://doi.org/10.3389/fgene.2023.1243395)

1.3.1 Abstract:

Chromatin is a vital and dynamic structure that is carefully regulated to maintain proper cell homeostasis. A great deal of this regulation is dependent on histones proteins which have the ability to be dynamically modified on their tails via various post-translational modifications (PTMs). While multiple histone PTMs are studied and often work in concert to facilitate gene expression, here we focus on the tri-methylation of histone H4 on lysine 20 (H4K20me3) and its function in chromatin structure, cell cycle, DNA repair, and development. The recent studies evaluated in this review have shed light on how H4K20me3 is established and regulated by various interacting partners and how H4K20me3 and the proteins that interact with this PTM are involved

in various diseases. Through analyzing the current literature on H4K20me3 function and regulation, we aim to summarize this knowledge and highlight gaps that remain in the field.

1.3.2 Introduction:

In every eukaryotic cell, genetic information is encoded by nearly identical DNA sequences. Tissues and organs achieve their identity through varied gene expression patterns that are critically regulated[14]. Proper regulation of gene expression is partially dependent on packaging DNA into chromatin, a complex of DNA and proteins[15]. Chromatin is divided into two functional states, initially identified by differential chromosomal staining patterns. Euchromatin corresponds to an open and transcriptionally active conformation of chromatin, while heterochromatin is condensed and transcriptionally inert[16]. The major role for heterochromatin is to protect repetitive regions in the genome from damage and to ensure correct chromosome segregation, thereby preventing genomic instability[17]. Heterochromatin can be further classified into two subtypes, facultative heterochromatin which is present in gene-rich regions regulating the expression of genes under specific cellular contexts, and constitutive heterochromatin which is typically found in gene-poor regions, including repetitive sequences such as satellite repeats and transposable elements[17]. The fundamental unit of chromatin is the nucleosome, which consist of a histone octamer containing two copies each of histone 2A (H2A), histone 2B (H2B), histone 3 (H3) and histone 4 (H4). Each of these histones can be post translationally modified on their non-globular tail domain, leading to various layers of regulation. These posttranslational modifications (PTMs), that include methylation, phosphorylation, acetylation, ubiquitylation, and others, regulate chromatin structure, accessibility, and hence gene expression independently of changes in the DNA sequence, thus their classification as epigenetic marks[14].

The PTMs of lysine 20 on histone 4 (H4K20) are conserved from yeast to human and can be classified into the following four states: un (H4K20)-, mono (H4K20me1)-, di (H4K20me2)- and tri (H4K20me3)-methylated[18]. The different methylation states of H4K20 are established by distinct histone methyltransferases containing SET domains that were first discovered in Drosophila[19,20]. The lysine methyltransferase, KMT5A (SET8 or PR-Set7) catalyzes H4K20me1[21], while KMT5B (SUV420H1) and KMT5C (SUV420H2) catalyze H4K20me2 and H4K20me3, respectively[22,23]. In this review we focus on H4K20me3, as it is of particular interest due to its

association with many physiological processes including heterochromatin structure[24], cell cycle regulation[25], DNA damage[26], development[23], cancer[27], and other cellular processes. We discuss the process involved in establishing H4K20me3 and the role of H4K20me3 in heterochromatin structure, including the mediators and regulators of H4K20me3. We also dissect and analyze the different roles for H4K20me3 in normal cellular homeostasis, and in various diseases.

1.3.3 The role of H4K20me3 in heterochromatin formation and structure

a. H4K20me3 formation

H4K20me3 is abundant at heterochromatin regions that are rich in repetitive sequences, such as satellite repeats and transposable elements, among others (**Figure 1.1A**). This methylation is mediated by multiple events, including catalysis by various methyltransferases and their preceding PMTs. These events can be summarized through the following steps. First, unmethylated H4K20 is converted to H4K20 mono-methylation (H4K20me1) by the methyltransferase SET8 or KMT5A. This methylation plays a critical role in cell cycle regulation and in genomic integrity[26]. Following mono-methylation, KMT5B, also known as the drosophila homologue SUV420H1 (Suppressor of variegation 4-20 homolog 1), catalyzes the formation of H4K20 di-methylation (H4K20me2) which has an important role in DNA damage response, DNA replication, and in cell cycle regulation[28]. Third, KMT5C or SUV420H2 catalyzes H4K20 trimethylation (H4K20me3)[29], which is the primary H4K20 PTM involved in heterochromatin silencing[20,19]. Reports have shown that H4K20me1 can be used as a substrate of KMT5B and KMT5C, directly generating H4K20me2 or H4K20me3 respectively[22,23,30]. Indeed, due to the sequence and structural similarity of KMT5B and KMT5C in their catalytic domain, KMT5B and KMT5C have overlapping functions[22], nevertheless knockout studies indicate that KMT5B loss leads to a 60% reduction of H4K20me2 with no change in H4K20me3, while KMT5C depletion eliminated H4K20me3 without any significant impact on H4K20me2 in primary mouse embryonic fibroblasts (MEFs)[23,31].

While less is known about the events preceding H4K20me3 at facultative heterochromatin, at constitutive heterochromatic regions, another histone PTM has been reported to facilitate H4K20me3 formation. This PTM, H3K9me3, catalyzed by the KMT1A/SUV39H1 methyltransferase, serves as a docking site for Heterochromatin Protein 1 (HP1). HP1 binds to

H3K9me3 though its chromodomain which then leads to the recruitment of KMT5C. Once recruited, KMT5C catalyzes H4K20me3 (**Figure 1.2B**)[32,23]. Previously, it was believed that KMT5C interacted with all HP1 isoforms (HP1a, HP1b, and HP1g) and that each HP1 isoform interacted with different regions of the KMT5C C-terminus[32,33]. However, recent reports have identified a far more complex mechanism. While both HP1a and HP1b are enriched at pericentric heterochromatic (PCH) regions, HP1b appears to have a roll in compacting the chromatin through its direct functional link with H4K20me3 and KMT5C[34]. Hence, the interaction of HP1b with KMT5C appears to be key for H4K20me3 formation at constitutive heterochromatin regions.

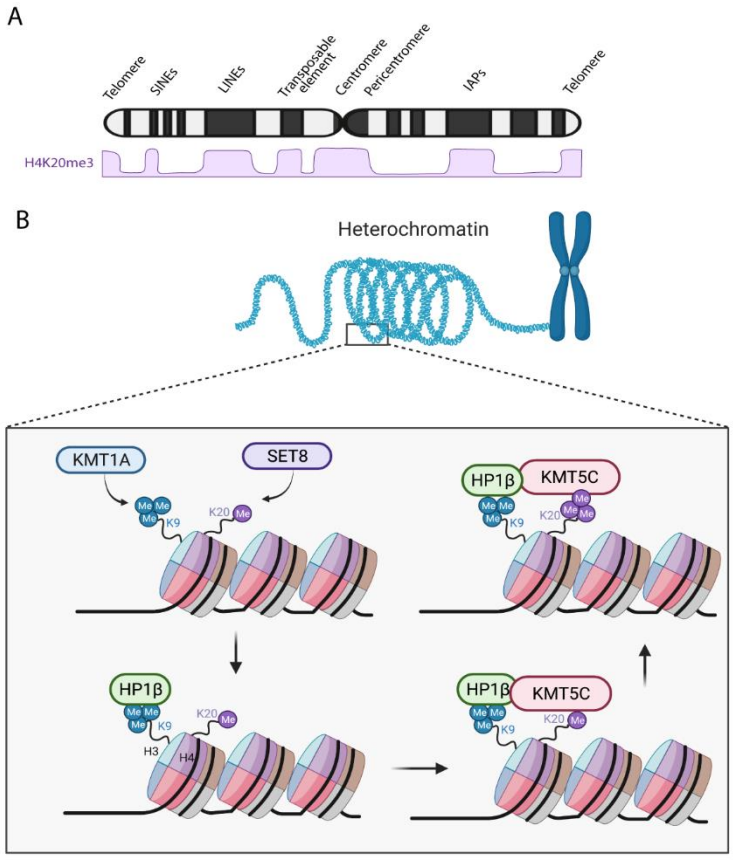

Figure 1.2. H4K20me3 formation at heterochromatin regions and overall abundance in the chromosome. A: H4K20me3 approximated levels at different regions of the chromosome[17]. SINEs: Short interspersed nuclear elements. LINEs: Long interspersed nuclear elements. IAPs: Intracisternal A-particle. **B:** In constitutive heterochromatic regions of the genome KMT1A methyltransferase induces H3K9 trimethylation, which is then recognized by HP1β[34]. HP1β then recruits KMT5C to that region, and KMT5C trimethylates H4K20me1 or H4K20me2 (not shown)[23,32]. HP1β: Heterochromatin protein 1 β isoform.

b. The role of H4K20me3 in chromatin structure:

b.1 The role of H4K20me3 in telomeric heterochromatin

Since H4K20me3 is enriched at telomeric regions leading to their compaction and stabilization, its loss has significant implications on telomeric structure and lenght[24]. Cells deficient for KMT5C methyltransferase, or for both KMT5B/C, have reduced levels of H4K20me3 in both telomeres and subtelomeric regions. This loss is associated with lengthening of telomeres and subtelomeric regions (**Figure 1.3A**). Loss also results in increased sister chromatid recombination globally and at telomeric regions. These changes are specific to loss of H4K20me2/3 as H3K9me3 levels were not affected, lending further support that H3K9me3 is not dependent on H4K20me2/3. And, while telomere length was altered, there was no evidence of defective telomere capping[35]. A similar telomer lengthening defect was observed following depletion of Retinoblastoma1 (RB1), which is known to interact with KMT5C. In this work, loss of RB1 led to reduced H4K20me3 and subsequent telomere lengthening in MEFs.[36] However, whether the increase in telomere length when H4K20me3 is reduced is due to more accessibility to telomerase, or if it is through an alternative lengthening of telomeres mechanism (ALT), which relies on the recombination between telomeric sequences to maintain telomeric length was not determined[35]. In contrast to these findings, PWP1, a chromatin binding protein important for controlling growth downstream of mTOR [37], has been reported to regulate H4K20me3 levels through binding to and stabilizing the shelterin complex in mouse embryonic stem cells, leading to telomere shortening. Reduced expression of PWP1 correlated with reduced levels of H4K20me3 at specific telomeric and subtelomeric regions. PWP1 depletion was also shown to induce telomere shortening and therefore increased DNA damage in telomeric regions (**Figure 1.3B**). Additionally, restoration of telomere length in PWP1 depleted cells was only achieved by overexpressing PWP1 along with KMT5C. In support of this association, PWP1 was found to interact with KMT5C along with the shelterin complex, providing regulation of chromatin length[38]. In addition to proteins such as RB1 and PWP1 that are directly involved in growth, major epigenetic modifiers are also correlated with H4K20me3 levels. In Drosophila, loss of the DNA methyltransferase DNMT2, has been correlated with loss of H4K20me3 at retrotransposons and subtelomeric repeats[39]. An additional study reported that in telomerase-deficient mice with short telomeres, H3K9me3 and H4K20me3 levels were reduced in telomeric and subtelomeric chromatin, accompanied by increase acetylation of

H3 and H4 at these regions[40]. Hence, loss of telomeric repeats appears to lead to loss of heterochromatin features, including H4K20me3. Whether the loss of H4K20me3 in telomeric regions leads to telomere lengthening or shortening is debated and seems to depend on the context. Loss of KMT5C leads to increased telomere length, but loss of PWP1, a protein involved in shelterin complex stabilization, reduces H4K20me3 and decreases telomere length. Contradicting reports expose the complexity of telomere length regulation which could also be due to differences in experimental design and model systems.

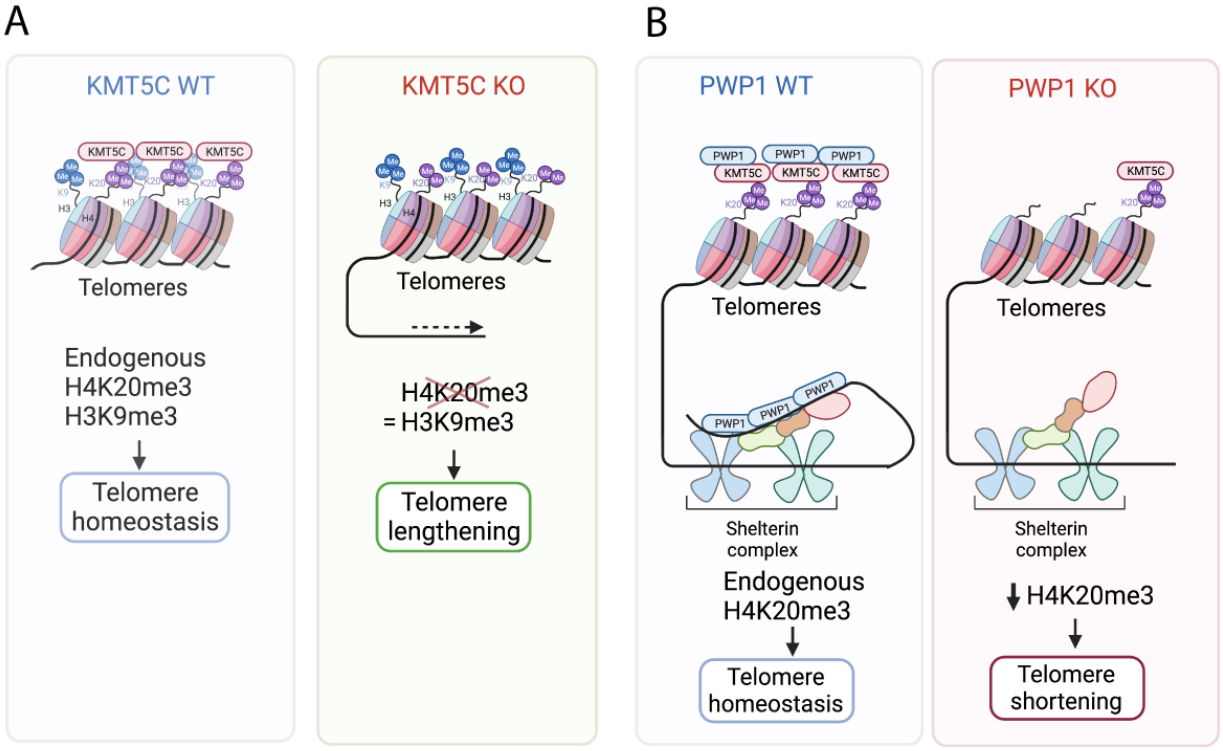

Figure 1.3. The effect of KMT5C alterations on telomere length. A: KMT5C deficiency reduces H4K20me3 levels at telomeres and subtelomeres, thereby increasing telomeric lenght[35]. **B:** Loss of PWP1, a protein involved in shelterin complex stabilization reduces H4K20me3 at telomeres and decreases telomere lenght[38]. PWP1: Periodic tryptophan protein 1.

b.2 A role for H4K20me3 in pericentric heterochromatin and chromocenter structure

The H4K20me3 mark is also highly enriched in pericentric regions and thus changes in H4K20me3 have a direct impact on pericentric chromatin structure and on chromocenter structure. Similar to telomeric regions, pericentric regions are also gene-poor and require proper silencing to maintain cell homeostasis. Silencing is reported to be through both H3K9me3 and H4K20me3[16,22]. Indeed, there is major dysregulation of the 3D chromatin landscape in embryonic stem cells (ES) when

KMT5C is lost. This loss results in A/B compartment switching, perturbed chromatin insulation, and altered chromatin interactions of pericentric heterochromatin, indicative of localized decondensation[41]. While these findings suggest that KMT5C plays a crucial role in maintaining proper chromatin organization in embryonic stem cells, appropriate chromatin structure in pericentric heterochromatin is also critical during mitosis. Because the pericentric regions plays a crucial role in facilitating sister-chromatin cohesion by recruiting cohesin complexes, any dysregulation in chromatin organization can potentially disrupt this process and result in abnormal cell division. For example, in mouse embryonic stem cells, KMT5C interacts with several cohesin subunits, enabling cohesin binding to pericentric heterochromatin which is essential for correct chromosome segregation[16,32]. KMT5C seems to be involved in the initial loading and maintenance of cohesin subunits in G0-phase cells. KMT5B and KMT5C double knock-out cells exhibit reduced chromatin compaction and altered chromocenter organization in interphase, indicating that KMT5C is essential for nuclear architecture. The role of KMT5C in pericentric chromatin structure seems to be dependent on a specific region of the C-terminal domain of the protein, a region where HP1 and cohesin interact. Nonetheless, additional studies are needed to determine the function of the SET enzymatic domain of KMT5C and H4K20me3 methylation in cohesin recruitment[32]. Similarly, KMT5C seems to play a role in the proper structure of chromocenters and clustering of pericentric heterochromatin[42]. Mislocalization of H4K20me3, KMT5C, and HP1 have been observed in cells when the muscle-specific long non-coding RNA ChRO1, is inhibited, which leads to defects in the spatial fusion of chromocenters[43]. In summary, proper regulation of H4K20me3 and chromatin structure in pericentric heterochromatin is critical for maintaining cell homeostasis and normal cellular processes such as mitosis and stem cell function. Dysregulation of these processes, due to loss of KMT5C can lead to chromatin decondensation, altered interactions, and abnormal cell division.

1.3.4 H4K20me3 mediators and regulators

a. H4K20me3 modification mediators (Writers)

The main H4K20me3 writer is KMT5C[23]; although, other methyltransferases have been reported to mediate H4K20me3. SMYD5, a methyltransferase that is a member of the SMYD family of SET and MYND domain-containing proteins, can tri-methylate H4K20 in drosophila, mouse

primary macrophage cells, and mouse embryonic stem cells [44,45,46]. Likewise, SMYD3, a member of the same family can tri-methylate H3K4 [47], H4K5 [48] and H4K20 [49]. These SMYD3-mediated methylation events have been observed in prostate cancer cell lines[49] and in *in vitro* assays using histones from HeLa cells[50]; however, H4K20me3 was not shown to be a substrate of SMYD3 in other cell line tested, including MEFs, HeLa, and MCF7 cells[48],. And, while silencing of SMYD3 was correlated with a decrease in H4K20me3, it appeared to be cell line specific resulting in a need to further validate this correlation with enzymatic assays[49] (**Figure 1.4A**).

b. H4K20me3 erasers

The H4K20me3 mark can be erased by three demethylase enzymes. The first, PHF2, a member of the Jumonji domain family of lysine demethylases, was reported to demethylate H4K20me3 *in vitro* using bacteria purified mononucleosomes, and in macrophages in culture[44]. PHF2 was found to be involved in removing H4K20me3 at promoters responsible for inflammatory responses. Whether a similar response would be observed at other regions of the genome has yet to be evaluated[44]. In agreement with PHF2 acting as an H4K20me3 eraser, its levels were found to be negatively correlated with H4K20me3 levels during mouse embryonic development. Nonetheless, this could simply be due to reduced expression of KMT5B/C in embryonic chromatin[51]. The second predicted H4K20me3 eraser is the lysine demethylase JMJD2A, which also recognizes H3K4me3. The recognition of H3K4me3 and H4K20me3 is somewhat distinct as the crystal structure of JMJD2A identified a specific mutation (D939R) in JMJD2A that impaired its interaction with H3K4me3 but not with H4K20me3[52]. This difference in substrate recognition is useful for technological development and for studying H4K20me3 dynamics; however, the biological and functional significance of JMJD2A in demethylating H4K20me3 remains to be established. The third reported H4K20me3 eraser was identified in a recent screen of ~2,500 nuclear proteins where a human homologue of the yeast protein RAD23 (hHR23A/B) was identified as an eraser of H4K20me1/2/3. Subsequent overexpression of hHR23A or hHR23B in HEK-293T cells reduced the levels of H4K20me1/2/3. hHR23A/B demethylation of H4K20me1 activated transcription of coding genes and demethylation of H4K20me3 activated transcription of repetitive elements, further supporting the role of H4K20me3 in repressing repetitive elements in the genome[53]. Indeed, histone methylation is a very dynamic process mediated by both writers

and erasers. In the case of H4K20me3, additional studies of these, and perhaps other proteins will provide important details on the mechanisms that regulates this dynamic process (**Figure 1.4B**).

c. H4K20me3 readers

Readers are proteins that contain a specialized domain that helps them interact with and interpret modifications such as H4K20me3. For example, elements of the origin of replication complex (ORC) were shown to interact with H4K20me3 in a histone marks proteomics study. More specifically, LRWD1, a leucine-rich and WD40 repeat-containing protein that interacts with an ORC subunit was found to interact with H4K20me3[33]. Consistent with this, H4K20me3 is essential to sustain the licensing and activity of a subset of ORCA/LRWD1-associated origins, which allows proper replication timing, and is critical in the selection of active replication initiation sites in heterochromatin regions in mammalians[54]. More specifically, H4K20me2/3 serve as enhancers for MCM2-7 helicase loading and replication activation at defined origins. The ORC plays a critical role in the initiation of DNA replication and cell cycle progression, highlighting the importance of the H4K20me3 mark in chromatin organization during the cell cycle[25].

Additional readers of H4K20me3 are important for maintaining proper DNA methylation, and hence gene silencing. Recent reports described a direct link between a DNA methyltransferase and H4K20me3, leading to gene repression. DNMT1, a DNA methylase important in mitotic division, reads H4K20me3 through binding of its BAH1 domain. The $DNMT1_{BAH1}$-H4K20me3 interaction triggers a conformational change of DNMT1 into an open conformation. Disruption of the BAH domain of DNMT1 lead to DNA hypomethylation within an H4K20me3-positive LINE-1 but hypermethylation at genomic regions lacking H4K20me3. Hence, the $DNMT1_{BAH1}$-H4K20me3 interaction ensures proper heterochromatin targeting of DNMT1 and DNA methylation at H4K20me3 rich regions such as LINE-1 retrotransposons[55]. Without question the H4K20me3 function at specific genomic regions intimately depends on the protein or reader interacting with the mark (**Figure 1.4C**). It is highly likely that many more readers associating with H4K20me3 across the genome are left to be identified and characterized.

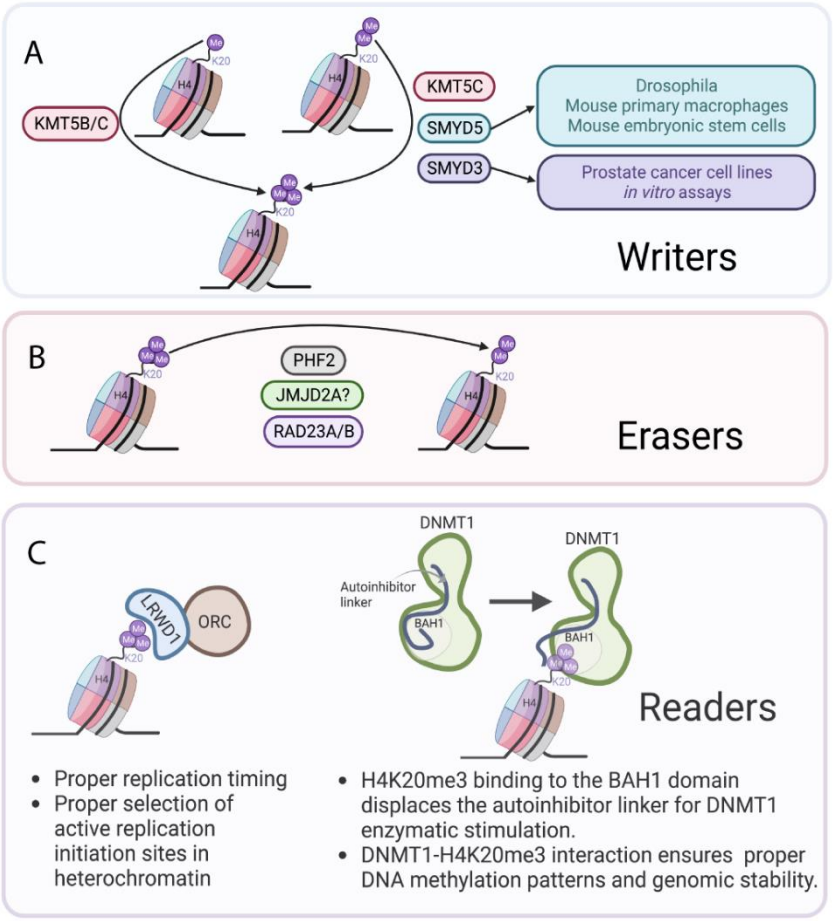

Figure 1.4. **Writers, erasers, and readers of H4K20me3. A:** KMT5C[23] is the main writer of H4K20me3. However, the methyltransferases KMT5B[22], SMYD5[44,45,46], and SMYD3[49,50] have been correlated with H42K0me3 formation depending on the cellular context. **B:** Human RAD23A/B demethylates H4K20me3 in HEK-293T cells[53]. PHF2 demethylates H4K20me3 at promoters of inflammatory response genes[44]. JMJD2A recognizes H4K20me3 but demethylase activity has yet to be demonstrated[52]. **C:** H4K20me3 readers include an element of the origin of replication complex ORC/LRWD1[54], and the DNA demethylase DNMT1[55].

d. H4K20me3 regulation via KMT5C or H4K20me3 interacting partners.

The mechanism of how H4K20me3 is generated at specific genomic regions is not clearly understood, and identification of KMT5C and H4K20me3 interacting partners will help elucidate this phenomenon. For example, KMT5C has been reported to interact with members of the tumor suppressor RB1 family, albeit *in vitro* the interaction of RB1 appears to be greater for the H4K20me2 methyltransferase, KMT5B. Nonetheless, in $RB1^{-/-}$;$RBL1^{-/-}$;$RBL2^{-/-}$ triple knockout mouse embryonic stem cells (MEFs), H4K20me3 levels were reduced in pericentric and telomeric

heterochromatin regions, leading to genomic instability and defects in chromosome segregation (**Figure 1.5A**)[36]. Similarly, KMT5C has also been reported to physically interact with the enzyme Activation-Induced cytidine Deaminase (AID) involved in B cells antibody diversification. Interaction of AID and KMT5C led to increased H4K20me3 in specific regions in the genome of human embryonic kidney (293F) cells (**Figure 1.5B**)[56]. And, while the canonical function of KMT5C is to tri-methylate histones, reports suggest that KMT5C can also methylate non-histone proteins in peptide arrays. *In vitro*, KMT5C catalyzed methylation of Caster Zinc Finger 1 (CASZ1), and OIP5 and CENPU (both centromeric proteins); nevertheless, further research is needed to determine whether these interactions occur *in vivo* or not[57].

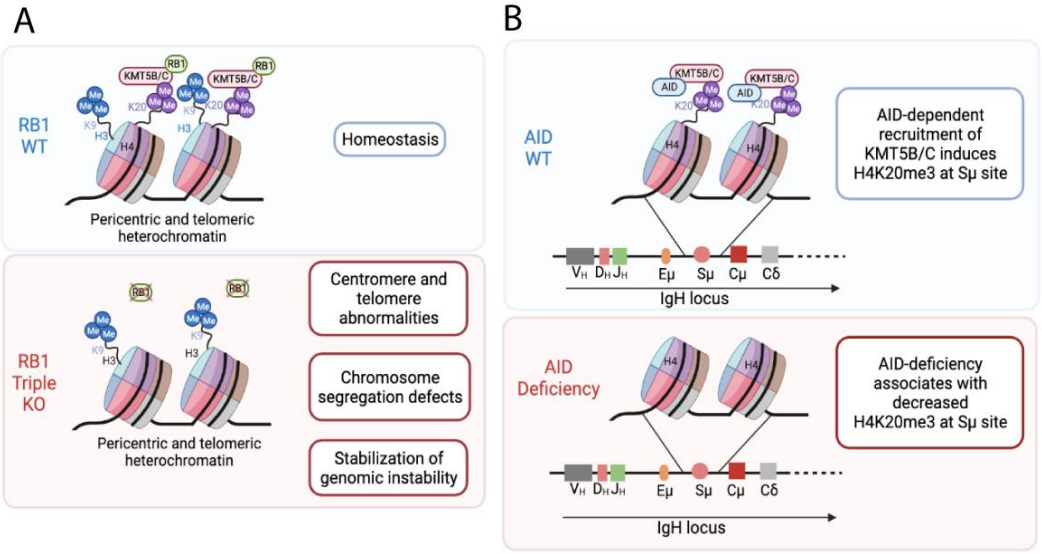

Figure 1.5. KMT5B/C interacting partners and their effect cellular consequence when they are deleted. A: Deletion of the KMT5B/C interacting partner RB1, leads to genomic instability and chromosome segregation defects[36]. **B:** AID deficiency associates with decreased H4K20me3 at Sμ sites[56]. RB1: Retinoblastoma protein 1. AID: Activation-induced cytidine deaminase.

In addition to their interactions with proteins, both KMT5C and H4K20me3 interact with coding and non-coding RNAs, including long non-coding RNAs (lncRNAs). This includes their interaction with the antisense non-coding RNA (asRNA) PAPAS in regulating ribosomal RNA (rRNA) transcription. It was first shown that ectopic expression of PAPAS (Promoter and Pre-rRNA Antisense) was able to trigger H4K20me3 increase at chromatin regions containing ribosomal DNA (rDNA), suggesting that antisense RNA could guide KMT5C to rDNA[58].

Additional studies uncovered a lncRNA-mediated mechanism that facilitates the localization of KMT5C at genomic loci, including rRNA genes and intracisternal A particle (IAP) elements during quiescence or growth arrest[59]. However, this phenomenon seems to be dependent on the environmental context because in hypotonic stress conditions, KMT5C interacts with the E3-ubiquitin ligase Nedd4, leading to KMT5C degradation. In this situation, where KMT5C is lost, rDNA is still silenced, a phenomenon attributed to release of PAPAS, which once released associates with a subunit of a nucleosome remodeling complex, resulting in epigenetic silencing of rDNA[60]. Clearly lncRNAs guiding of KMT5C to rRNA gene regions depends on the cellular context and hence, further research is needed to discover the regulators of KMT5C and H4K20me3 deposition in regions outside of pericentric and telomeric heterochromatin. Additional reports have also identified RNA association with histone post-translational modifications as a possible mechanism of chromatin and gene expression regulation. Using chromatin-associated RNA immunoprecipitation (CARIP) followed by sequencing, Kurup et al. identified mRNAs and noncoding RNAs that associated with H4K20me3 in embryonic stem cells. It appears that H4K20me3 preferentially interacts with longer protein-coding transcripts and ncRNAs, typically those with a greater number of exons. Gene ontology analysis suggests that the transcripts that interact directly with H4K20me3 are involved in RNA processing, DNA repair, cell redox homeostasis, regulation of cell motility/migration, placental development, epithelial cell differentiation, and other processes[61]. However, further research is needed to establish whether the interactions for these RNAs with H4K20me3 are involved in heterochromatin formation and/or stabilization.

1.3.5 H4K20me3 in physiology

a. Cell cycle-dependent tri-methylation of H4K20.

Proper balance of histone methylation is necessary to maintain normal biological functions, including proper cell cycle regulation. Indeed, histone methylation at nearly all methylation sites is dynamic with regard to establishment, reversibility, or maintenance across cell division, conferring genomic integrity, stability or reversibility in response to various stimuli[14]. While H4K20me2 is the most abundant methylation state on histone 4 and represents 80% of H4K20 methylation, historic studies determined that H4K20 and H4K20me1 appear to be the most

dynamic of the H4K20 methylation states throughout the cell cycle[14,62]. It was believed that H4K20me1 and H4K20me2 were more closely linked to DNA replication and DNA damage repair than H4K20me3, which was predominantly associated with silenced heterochromatic regions. However, more recent research has elucidated the essential role of H4K20me3 in multiple stages of the cell cycle.

In a homeostatic cell, the levels of H4K20me3 vary depending on the cell cycle stage (**Figure 1.6**) and on the levels of H4K20me1/2. In resting cells in G1 or G0 H4K20me3 is high in heterochromatic regions, H4K20me2 is present throughout the genome, and H4K20me1 is restricted to specific genes. In early G1, H4K20me1 is reduced as it gets converted to the di- and tri- methylated form (H4K20me2/3) and due to proteolytic degradation of SET8 during G1, new H4K20me1 modifications are greatly reduced[63]. In S phase, during DNA replication when cells are incorporating new histone 4 molecules, very little H4K20me1 is present due to limited SET8 methyltransferase levels. However, at the end of S phase, SET8 expression is stabilized resulting in H4K20me1 establishment on most of the new H4 molecules. During G2/M and in early mitosis H4K20me1 accumulates and is gradually converted to H4K20me2/3[63]. Following mitosis, in early G1, most of the H4K20me1 is again converted to H4K20me2 and H4K20me3 by the KMT5B/C methyltransferases[18].

With the understanding that H4K20 methylation is dynamic during the cell cycle, it is not surprising that deregulation at various stages of the cell cycle can lead to improper replication. For example, in the absence of SET8-driven H4K20 mono-methylation, genome-wide decompaction occurs, leading to excessive loading of the origin recognition complex (ORC), which can result in DNA damage[26]. Additionally, in early stages of replication, the degradation of SET8 is essential to prevent chromatin compaction caused by further H4K20 methylation. When SET8 degradation is impaired, there are massive defects in cell cycle progression and excessive DNA damage. This is attributed to elevated H4K20me3, leading to increased ORC recruitment through binding of ORCA/LRWD1 to H4K20me2/3, which then generates aberrant re-replication[64]. This continued expression of SET8 also leads to increased H4K20me3 in promoters of specific genes, such as histone genes resulting in cell toxicity. This results in DNA damage characterized by activation of P53 and G2 checkpoint pathways, leading to increased apoptosis. SET8 degradation in early

replication is therefore essential for the proper progression of the clockwork-like function of cell cycle, preventing high levels of unregulated H4K20me3 gene repression[65]. In the later stages of the cell cycle, such as during mitosis, H4K20me3 regulation is essential for gene repression. In a sequential pattern, proper regulation of H4K20me1 is essential during mitosis because the mono-methylation enables further gene repression mainly generated by H4K20me3[31]. While not as intensely studied as some of the other histone PTM, it is becoming increasing clearer that H4K20me1/2/3 levels are very dynamic throughout the cell cycle, and their abundance is essential for proper cell replication and homeostasis.

While the process of methylation is dynamic and critical for cell cycle function, so is demethylation of H4K20. For instance, the PHF8 demethylase acts as a cell cycle regulator by demethylating H4K20me1 allowing for G1-S transition[66]. During early development in *Xenopus laevis* embryogenesis, the essential explanation for the kinetics of H4K30me1/2/3 in cycling cells involves nonspecific passive demethylation resulting from cell division dilution, rather than active demethylation. The latter effect also suggests that overall cell-cycle mediated H4K20me dilution through DNA replication is essential for shaping the epigenetic landscape[67].

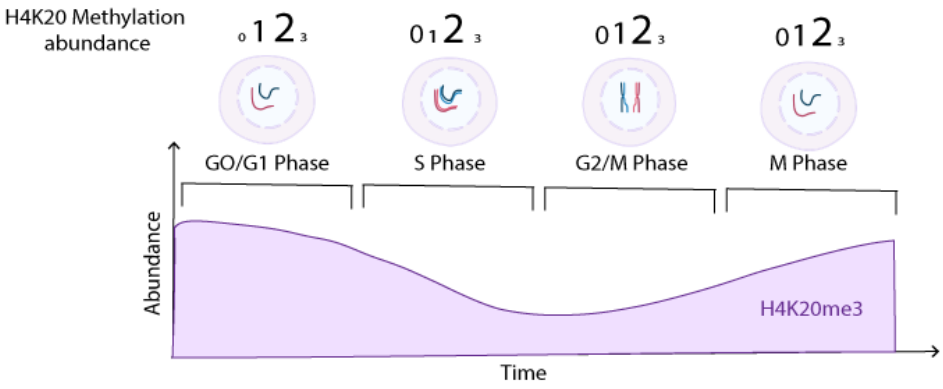

Figure 1.6. H4K20me3 levels in different cell cycle stages and chromosomal locations. Overall H4K20me0/1/2/3 abundance at different cell cycle stages (indicated by number size). While overall H4K20me3 levels are lower than H4K20me1/2 during the cell cycle, they still oscillate as indicated in the lower portion of the diagram.

b. A role for H4K20me3 in DNA damage

The involvement of H4K20 methylation in the DNA-damage response has been established, and recent reports have started unraveling its underlying molecular mechanism. For example, the

known DNA-damage response protein 53BP1 is reported to be recruited to double-strand breaks (DSBs) by direct recognition of H4K20me2, leading to amplification of g-H2AX signaling[68,69]. Although H4K20me1 is also required for 53BP1 binding to DSBs, it is not sufficient and further methylation by KMT5B/C is required for proper 53BP1 nucleation[70]. Recent studies have tried to elucidate the role of KMT5B/C in the DNA repair response by inhibiting the enzymatic activities of these two methyltransferases. Using the KMT5B/C inhibitor A-196, ionizing radiation-induced 53BP1 foci formation was inhibited. Mechanistically, KMT5B/C chemical inhibition repressed non-homologous end joining (NHEJ)-mediated repair but had no effect on homologous-directed repair (HDR)-mediated repair[71]. Henceforth, enzymatic activities of SET8 and both KMT5B/C appear to be essential for 53BP1 recruitment and DSBs DNA repair[70,29].

While H4K20me3 regulation is crucial for the DNA damage checkpoint, it also plays a critical role in regulating subsequent steps in the cell cycle, including the G0 phase. In the presence of various stimuli, cells can exit G1 and enter into G0. Entry into G0 occurs when cells are aging, are under stress, or are inflicted with DNA damage[72]. In this state, otherwise known as senescence, H4K20me3 levels are elevated, specifically at senescence-associated heterochromatin foci (SAHF). Although increased KMT5C expression upregulates H4K20me3, this does not accelerate senescence in normal human cells, but instead reinforces senescence and slows tumor progression in oncogene-expressing cells. It is thought that elevated H4K20me3 in senescent cells and aged tissues act as a barrier to cancer through enhancing epigenetic and genetic stability. This stability is obtained by suppressing genetic rearrangements that might allow cells to escape from senescence, thereby preventing tumor progression[73]. Additionally, cells can also enter G0 when nutrients and growth factors are scarce, this is known as quiescence[74]. In fact, changes in chromatin compaction and changes in H4K20 methylation are essential for regulating the transition between proliferation and quiescence. In human fibroblasts H4K20me2 and H4K20me3 are increased in quiescent cells. Downregulating KMT5C in these cells results in defects exiting the cell cycle and decreased chromatin compaction. Hence, H4K20me3 is involved in achieving G0 phase and quiescence. However, whether that mechanism is dependent only on chromatin compaction regulation or specific gene repression is still not well understood[75].

c. A role for H4K20me3 in the immune response

During immune responses H4K20me3 functions as a molecular checkpoint by regulating class switch recombination. Specifically, H4K20me3 deposited by the methyltransferase SMYD5 works in conjunction with the NCoR corepressor complex to repress expression of toll-like receptor 4 (TLR4) target genes in macrophages. Upon TLR4 activation, PHF2, a histone lysine demethylase, demethylates H4K20 at these TLR4 promoter regions enabling TLR4 pathway activation[44]. This puts H4K20me3 methylation/demethylation at the forefront of signal-dependent regulation of inflammatory response genes. However, it is unclear what dictates the specificity that SMYD5 has for its substrate in this context. H4K20me3 levels also appear to interplay with DNA deaminases enzymes to regulate class switch recombination. Activation-induced cytidine deaminase (AID) directly interacts with and recruit KMT5B/C enzymes to Ig switch regions, leading to increase H4K20me3 at these sites, stabilizing class switch recombination function. The latter explains why B cells are defective in class switch recombination in KMT5B/C knock-out mice[23]. Similarly, chemical inhibition of KMT5B/C with A-196 can induce class switch recombination, by inhibiting the ability of splenic B cells to switch from IgM to IgG1, IgG3, or IgE[71]. In summary, H4K20me3 contributes to the regulation of immune response genes in specific contexts; however, the mechanisms that dictate KMT5B/C and SMYD5 substrate specificity remain to be established.

d. A role for H4K20me3 in development

H4K20me3 has been reported to play an essential role in development, stem cell differentiation, and aging. With regard to development in general, the essential role for both KMT5B and KMT5C is clear, as loss of both lead to perinatal death[23]. Embryonic lethality also occurs if the H4K20me3 demethylases, mHR23A/B are knocked out, highlighting the importance of both H4K20me3 and demethylation in embryonic development[76]. The specifics with regard to timing suggest that the dynamics are critical later in development. In mice this is attributed to knowledge that preimplantation embryos lack constitutive heterochromatin markers, including H4K20me3 and HP1a, which appear to be late developmental epigenetic markers. Hence, H4K20me3 does not mark the onset of differentiation but marks cells in late fetal development when organs and tissues have formed[77]. In agreement with this, the enzymes that establish H4K20me3, KMT5B/C were

mostly absent in mouse embryos before implantation, and ectopic expression of KMT5C led to development arrest. In agreement, global H4K20me3 sharply decreases after mouse zygote fertilization and starts to increase after implantation[78]. H4K20me3 function after fertilization likely allows the timely and coordinated progression of replication[51].

In embryonic stem (ES) cells, depletion of SMYD5, a methyltransferase reported to mediate H4K20me3, results in chromosomal aberrations and formation of transformed cells during differentiation. In this context, H4K20me3 was shown to be important for regulating endogenous long terminal repeats (LTR) and long interspaced nuclear elements (LINE)-repetitive DNA elements during differentiation[46]. In fact, SMYD5 was demonstrated to repress lineage-specific genes, and thus, contributed to maintenance of ES cell lineage[45]. Somewhat uniquely to ES cells, H4K20me3 is coupled with activating histone modifications, including H3K4me3 and H3K36me3. Association of H4K20me3 with H3K4me3 has been identified in intergenic regions and near transcriptional start sites, whereas H4K20me3/H3K36me3 are located in intergenic regions and within the gene body of actively transcribed genes[79]. Thus, the role of H4K20me3 in gene expression depends on where in the genome it is located and on its interaction with other methylation marks in ES cells. Whether similar patterns are prominent in other cell types, or during diseases such as cancer that often mirror some of the biology of stem cells, has yet to be reported.

And, while the literature is less established, H4K20me3 changes have also been reported in aging, where increased H4K20me3 levels were observed in the kidneys and livers of aged rats[80].

1.3.6 H4K20me3 in disease

a. H4K20me3 in cancer

Controlled cell division is crucial to prevent the development of diseases such as cancer. The regulation of induced pluripotent stem cells (iPS), which are often used as a cancer model, can be influenced by various cellular components, including H4K20me3. Abrogation of KMT5B and KMT5C in induced pluripotent stem cells (iPS), leads to loss of H4K20me3 at heterochromatic regions such as telomeres. The later promotes tumorigenic potential of iPS cells through facilitating telomere elongation during reprograming. When the function of KMT5B/C is lost, iPS

cells *in vitro* are reprogrammed, which is characterized by significant epigenetic changes and gene expression patterns similar to that of embryonic stem cells. One of the major epigenetic changes observed includes loss of histone and DNA methylation, which leads to a more open chromatin conformation in comparison to differentiated cells. This more "relaxed" chromatin, due to loss of H4K20me3, facilitates access of telomerase to the telomere during iPS cell generation, leading to faithful replication of the chromosome ends. And while the telomeres are extended during this process, they are still protected – this is also the case in KMT5B/C double knockout iPS cells[81,82]. These data suggest that H4K20me3 helps to block tumorigenesis in pluripotent stem cells through inhibiting reprogramming and chromatin relaxation[81]. However, whether loss of H4K20me3 only at telomeric regions is responsible for the malignancy phenotype or if loss at other gene-rich regions is also contributing remains to be elucidated.

Changes in H4K20me3 levels are also aberrantly altered in established tumors when compared to normal tissues. While some studies report elevated levels of the methyltransferase and H4K20me3 in cancer, this is typically the exception as overall loss of H4K20me3 appears to be a more common observation[27]. For example, in a panel of tissues and cell lines, H4K20me3 was found to be reduced ~50% of the time in tumor tissues/cells in comparison to corresponding normal tissues. In some tumors, such as in a skin carcinogenesis models, H4K20me3 was progressively lost from early stages to the most malignant stages[27]. In lung cancer, loss of H4K20me3 and reduced levels of the *KMT5C* transcript is observed[83]. In addition to reduced H4K20me3 in tumors, loss also appears to be important for generating resistance to various anti-cancer agents. For example, we recently reported on the loss of KMT5C during the process of acquired resistance to EGFR inhibitors in EGFR mutant non-small cell lung cancer cell lines. Both genetic and chemical inhibition of KMT5C led to resistance, which was partially attributed to loss of H4K20me3 at the locus encoding the long non coding RNA LINC01510, an upstream enhancer of the oncogene MET[13]. These findings highlight the important role of H4K20me3 as a candidate biomarker for early detection and therapeutic approaches of lung cancer[83]. Besides lung cancer, loss of H4K20me3 is also a marker and molecular contributor to colon, bladder, liver, and breast cancers, and osteocarcoma. In colon cancer H4K20me3 helps with stratification of patient groups – high expression of H4K20me3 is associated with good prognosis in early-stage colon cancer[84]. In bladder cancer, global H4K20me3 was found to be lower than in normal urothelium tissue[85], while

progressive loss of H4K20me3 and decreased expression of KMT5C has been reported in rat liver tumors during hepatocarcinogenesis[86]. In breast cancer cell lines, KMT5C and H4K20me3 levels were found to be low in comparison to normal epithelial breast cell lines[87]. In this setting, KMT5C was reported to act as a tumor suppressor through H4K20me3 at the tensin-3 locus leading to silencing of the locus and suppression of breast cancer cell invasiveness[88]. Likewise, another report indicated that suppression of KMT5C also results in increased breast cancer cell proliferation[89]. Similar to the skin carcinogenesis model, expression of KMT5C is lower in breast tumor tissues when compared to adjacent non-cancerous region and higher in early stages of breast cancer when compared to advance stage diseases[89]. In an additional study conducted in breast cancer cell lines, H4K20me3 gene repression of the tumor suppressor TMS1 limited gene reactivation. In this novel study, H4K20me3-mediated gene silencing involved negative regulation of Pol II promoter escape, thereby enforcing Pol II pausing, leading to gene repression[90]. Finally, in osteosarcoma, reduced levels of KMT5C and H4K20me3 were also observed in tumor tissue and malignant cell lines compared to normal counterparts. RNA seq analysis after KMT5C knockdown identified pathways involving mitogen-activated protein kinase, P53, and ErbB to be dysregulated; however, whether these genes are directly regulated by H4K20me3 repression remains to be elucidated[91].

While the mechanisms that contribute to KMT5C downregulation during tumorigenesis are still incomplete, at least one study points to microRNA involvement. In this body of work miR-29a was shown to directly target the *KMT5C* transcript leading to epithelial to mesenchymal transition (EMT), promoting migration and invasion of breast cancer cells, further supporting a tumor suppressive role for KMT5C[92].

Nonetheless, the roles of KMT5C and H4K20me3 in tumorigenesis are clearly context dependent. Expression data from The Cancer Genome Atlas (TCGA) indicates that KMT5C is highly expressed in cancers[93]. And, increased expression of H4K20me3 was reported in esophageal squamous cell carcinoma tumor tissues[94] and in pancreatic cancer, where KMT5C is reported to favor mesenchymal identity, while KMT5C knockdown decreased stemness and increased drug sensitivity[95]. In summary, many reports highlight the loss of H4K20me3 in cancer; however, it appears that H4K20me3 selectively represses various targets depending on the biological context, adding another layer of complexity to this epigenetic regulator and its role in cancer progression.

b. H4K20me3 in other diseases

Besides regulating cancer progression, H4K20me3 is also implicated in other diseases. In fragile X syndrome, expansion of the CGGCCG repeat from the Fragile X Messenger Ribonucleoprotein 1 gene leads to gene silencing and to disease development. It was found that these silenced alleles are elevated for H4K20me3[96]. However, the mechanism of how H4K20me3 is established at that genomic region, or if it is dependent on KMT5C is not well understood.

In obesity, KMT5B and KMT5C regulate metabolism through down-regulating peroxisome proliferator activator receptor gamma (PPAR-g), which regulates lipid storage and glucose metabolism. Abrogation of KMT5B/C and therefore H4K20me3, leads to activation of PPAR-g in brown adipose tissue to increase mitochondria respiration, improve glucose tolerance, and reduce adipose tissue to reduce obesity[97]. While the data from this work suggest that KMT5B/C proteins may be a therapeutic target for the treatment of obesity, other reports suggest a contradictory effect with regards to KMT5C in obesity, more specifically in adipocytes. Knockout of KMT5C in mice, led to loss of H4K20me3 at the transformation related protein 53 (Trp53) promoter, thereby enhancing Trp53 expression. Enhanced Trp53 was found to be responsible for metabolic phenotypes, such as high-fat-diet-induced obesity and glucose intolerance[98]. These contradictory findings suggest an ambiguous role of H4K20me3 in obesity, which may depend on whether KMT5C alone or both KMT5B/C are lost.

While there are a modest number of reports highlighting the dysregulation of H4K20me3 in other diseases, more work is needed to uncover whether H4K20me3 is a direct contributor to these diseases or if it is a consequence of the disease. For example, in Sotos syndrome, which is characterized by overgrowth and increased risk of tumorigenesis, H4K20me3 is reduced. This is likely due to loss-of-function mutations of the SET domain-containing protein NSD1 gene. Epigenetic inactivation of NSD1 correlates with diminished methylation of H4K20me3, however whether this effect is dependent on NSD1 SET domain activity has yet to be determined[99]. In sickle cell disease (SCD), activation of the protein arginine methyltransferase PRMT5 induces repressive marks in the g-globin gene promoter, by assembling a repressor complex that contains KMT5B. Reactivating g-globin gene expression and hence inhibiting H4K20me3 in the g-promoter has been proposed as a potential therapeutic approach for SCD treatment[100]. Based on these studies it is

clear that H4K20me3 and the associated methyltransferases play a major role in normal cellular homeostasis and that dysregulation can lead to deleterious consequence; however, the future depends on gaining further insight into these guardians of the genome and how their misregulation mechanistically contributes to these disease states.

1.3.7 Conclusions and future perspectives

Dysregulation of H4K20me3 is linked to several diseases, including cancer; however, several open questions remain on whether H4K20me3 or KMT5C are good therapeutic targets or not. Due to the potential effects attributed to epigenetic alterations on disease-related genes, oncogenes, and tumor suppressor genes, epigenetic-targeted therapy is becoming a promising strategy for the treatment of cancer and other diseases. First, even though epigenetic modifications are somatically inherited, they can be good therapeutic targets due to their reversibility. Reversibility perhaps makes them more 'druggable' than correcting gene sequences[11]. Additionally, in diseased cells that are addicted to certain epigenetic abnormalities, using an inhibitor of an epigenetic process can be benificial[10] – for example, when tumor cells become dependent on aberrant gene silencing of oncogenes. As has been the case for other therapeutic modalities, the combination of epigenetic therapies with conventional therapies is worth exploring in preclinical and clinical trials. Such attempts have already garnered positive results, such as combinations of histone deacetylase inhibitors and DNA repair inhibitors in the treatment of advanced non-small-cell lung cancer[101]. In addition to therapeutics, epigenetic changes, such as histone methylation, can be used as diagnostics for disease monitoring. For example, studies evaluating H4K20me3 in blood samples determined that H4K20me3 levels can be used as a cancer biomarker. A small clinical study reported that H3K9me3 and H4K20me3 when normalized to nucleosome content could be used as valuable biomarkers to distinguish between cancer patients and healthy patients. In blood samples, H4K20me3 at the centromeric satellites SAT2 was significantly higher in breast cancer and lower in colorectal cancer compared to the respective healthy controls[102]. These blood-based detection methods of H4K20me3 in pericentric heterochromatin-specific circulating nucleosomes represent a potential promising new and non-invasive biomarker for colorectal and breast cancer patients.

While progress has been achieved in utilizing epigenetic therapy to treat various hematological malignancies, further research and clinical trials are necessary to extend the application of these

therapeutics to solid tumors and other diseases, primarily due to numerous challenges[103,104,105]. For example, epigenetic events such as the H4K20me3 mark are present ubiquitously across tissues, and selectivity is one of the main challenges. Would it be possible to target inhibitors to particular regions of chromosomes to prevent potential side effects[11]? This is certainly not trivial as targeting inhibitors to various chromosomal regions has yet to be clinically explored. Another challenge in considering H4K20me3 as a therapeutic target is that a clear distinction between changes in H4K20me3 as a causative or merely correlative effect in each disease is essential for the development of successful epigenetic therapies[11]. Leveraging KMT5C enzymatic activity as a target, restoring (or inhibiting) such a methyltransferase, has the potential to induce side effects due the complexity of KMT5C regulation and due to its ability to interact with multiple proteins and non-coding RNAs. Hence, simply restoring KMT5C might not lead to the intended therapeutic outcome. In diseases and cancers where inhibition of KMT5C is beneficial, the use of the chemical inhibitor A-196 is a possibility. However, based on other methyltransferases encoded from the genome, the possibility that these other histone-modifying enzymes might compensate for KMT5B/C inhibition and thereby conferring drug resistance needs to be considered[11]. It is anticipated that a more thorough understanding on how KMT5C is regulated, how H4K20me3 contributes to gene regulation, discovery of additional readers of H4K20me3, and understanding the role of KMT5C/H4K20me3 in various disease contexts is essential for pushing the therapeutic boundaries (**Table 1.1**).

Table 1.1. Summary table of the advantages and disadvantages in using epigenetics approaches as therapeutics

Advantages	Disadvantages
Epigenetic alteration influence expression of disease-related genes, oncogenes and tumor suppressors	Epigenetic selectivity as many methylation marks are ubiquitously present in cells
Epigenetic modifications are inherited but are also reversible making them more "druggable"	Epigenetic therapy has to be specific to certain regions of chromosome
Advantageous when cancers become dependent on certain epigenetic abnormalities	Difficult distinction between causative and correlative epigenetic alterations
Epigenetic changes such as histone modification can serve as disease biomarkers	Overexpression of epigenetic factors can lead to promiscuity due to lack of proper regulation by other interacting partners
Combination of epigenetics and conventional therapies can be beneficial for cancer treatments	When inhibiting an epigenetic factor, activity compensation by other enzymes should be considered

CHAPTER 2. LOSS OF KMT5C PROMOTES EGFR INHIBITOR RESISTANCE

In accordance with the American Association for Cancer Research (AACR) policy on author use, the following publication (Pal A.S., & Agredo A., *et al.*, 2022) is the accepted version with some additions as indicated in the text of the article published in Cancer Research (https://doi.org/10.1158/0008-5472.CAN-20-0821) and has been reprinted with permission from AACR. Dr. Arpita Pal contributed with the CRISPR Cas9 Screen, validation of KMT5C mutation in EGFR WT cells, and discovery of the LINC01510 regulation by H4K20me3. Alejandra Agredo Montealegre contributed with the initial generation of EKVX Cas9 cell line generation for the CRISPR Cas9 screen, and validation of genetic and chemical loss of KMT5C phenotype in EGFR mutant cells.

2.1 Chapter Overview:

In this chapter, we describe a genome-wide screen to identify tumor suppressor genes that are mediators of erlotinib resistance in NSCLC. KMT5C was the top hit of the screen, and here we elucidate one mechanism in which loss of KMT5C leads to erlotinib resistance in EGFR WT and mutant cells. This chapter is based on the previously mentioned publication from *Cancer Research* with some additional experiments.

2.2 Introduction

Lung cancer is the leading cause of cancer-related mortality, with an estimated 131,880 deaths predicted in 2021 in the United States alone [106]. The majority of lung cancer patients are diagnosed with non-small cell lung cancer (NSCLC), a subtype that represents 85% of lung cancer cases. Since most lung cancer patients are diagnosed at later stages with metastatic disease, surgical resection is not curative, and thus, the most effective treatment strategies are radiotherapy, chemotherapy, and targeted therapy. Targeted therapeutics are selected based on the presence of particular molecular drivers, genes that the cancer cells are essentially addicted to. A few such drivers that are commonly present in NSCLC include KRAS, MEK, MET, HER2, and EGFR, many of which are either mutated or amplified, resulting in constitutive pro-growth signaling [107,108].

Epidermal Growth Factor Receptor (EGFR) is a cell surface receptor required for normal cell growth and proliferation. In 10-35% of NSCLC cases EGFR and its downstream pro-growth signaling pathways are constitutively activated due to mutations in the receptor, the most common of which are an amino acid substitution in exon 21 (L858R) or an in-frame deletion in exon 19. Mutant EGFR can be clinically targeted with a variety of EGFR tyrosine kinase inhibitors (EGFRi), including erlotinib and gefitinib, both of which are first generation EGFRi, afatinib, a second-generation inhibitor, or osimertinib a third generation EGFRi. Osimertinib not only targets the primary EGFR mutation but is also active against a secondary mutation in EGFR, T790M. Erlotinib binds reversibly and specifically to the ATP-binding pocket of EGFR with high efficacy, abrogating downstream growth and survival signaling pathways. While initially beneficial as a first- or second-line treatment, many patients develop resistance within a year post treatment, which is currently a major drawback to its use [5]. The EGFR gene either incurs additional mutations or activates alternative signaling pathways to evade therapy. In the case of erlotinib treated patients that develop resistance, over 60% of patients acquire a secondary mutation, T790M, in their tumor whereas approximately 20% of tumors utilize bypass tracks. Bypass tracks allow the tumor to escape inhibition of the EGFR pathway through the use of alternative mechanisms that sustain their survival. Bypass tracks include signaling through oncogenic proteins such as MET, BRAF, HER2, PIK3CA or histological transformation of cells - NSCLC transformation into small cell lung cancer (SCLC) or through epithelial to mesenchymal transition (EMT) [5,109–111]. In addition to an incomplete understanding of mechanisms that govern these bypass tracks there are also approximately 15-20% of NSCLC tumors that acquire erlotinib resistance by mechanisms that remain unidentified [5].

Although gain-of-function mechanisms that drive resistance have been identified, loss of tumor suppressive genes, such as PTEN, TP53, TET1, and NF1 has also been reported to contribute to resistance [6–9]. Indeed, many tumor suppressive proteins function as gatekeepers of the genome preventing spurious activation of oncogenes. Here to define genes that prevent the development of resistance, a genome-wide loss of function screen was conducted using the CRISPR-Cas9 system. Our data suggest that an epigenetic factor and *bona fide* tumor suppressor, KMT5C can be included among the gatekeepers of the genome. KMT5C catalyzes the trimethylation of histone H4 at

lysine-20 (H4K20) using mono-methylated H4K20 as a substrate, which is required for the establishment of heterochromatin and repression of genes [23,32,57]. Loss of KMT5C has previously been implicated in causation of multiple cancers [27,112], but for the first time we show that loss of KMT5C is a novel mechanism that promotes erlotinib resistance in NSCLC cells. The findings of this study determined that KMT5C mutant cells express high levels of the oncogenic long non-coding RNA, *LINC01510* that transcriptionally upregulates the oncogene MET, mediating erlotinib resistance.

Not included in publication: Moreover, another mechanism that cells use to promote EGFRi resistance likely involves aberrant expression of microRNAs (miRNAs). miRNAs are small non-coding RNAs that mediate gene silencing via translational repression and degradation of targeted mRNAs [113]. Due to the ability of a miRNA to target multiple mRNAs, it is not surprising that misexpression of miRNAs is a common occurrence in cancer. Indeed, various miRNAs have either oncogenic or tumor suppressive functions [114]. In this study we identified the protein coding gene KMT5C that when lost leads to erlotinib resistance[13]. Additionally, in another screen performed in our laboratory, we identified miRNAs that when overexpressed drove resistance to the EGFRi erlotinib[115]. Surprisingly, in our miRNA overexpression screen, miR-4435, which is predicted to target KMT5C, was one of the most significant hits, suggesting that this miRNA is likely an upstream regulator of KMT5C. We determined that miR-4435 is predicted to target KMT5C via canonical mechanism involving the 3′ untranslated region (UTR) and via non-canonical binding to the 5′ UTR. In our preliminary data we determined that overexpression of miR-4435 reduces KMT5C transcript levels by possibly targeting the 5′ UTR. According to these findings, miR-4435 emerges as a potential upstream regulator of KMT5C, thereby inducing its reduction—a novel mechanism contributing to EGFRi resistance.

2.3 Methods

2.3.1 Cell culture:

All cell lines used in the study, except for PC9 (Sigma) were obtained from American Type Culture Collection (ATCC). All lines were routinely confirmed to free of mycoplasma contamination monthly. Cell lines generated during the study were authenticated by ATCC Cell Line

Authentication. All cell lines were grown in RPMI media supplemented with 10% FBS and 1% Penicillin/Streptomycin. ECas9 cells were continuously cultured in media containing 1µg/ml Blasticidin, The EKVX KMT5C mutant clones A, C and E were grown in media containing 100ng/ml Puromycin, inducible-KMT5C Calu6 clones were cultured in 500ng/ml Puromycin containing media, and rescue clones were grown in media containing 100ng/ml Puromycin and 300µg/ml G418 containing media.

2.3.2 Drug Preparation for *in vitro* studies:

Erlotinib (S7786, SelleckChem), afatinib (850140-72-6, Sigma Aldrich), gefitinib (S1025, SelleckChem), almonertinib (S8817, SelleckChem), dacomitinib (S2727, SelleckChem), crizotinib (S1068, SelleckChem), lorlatinib (S7536, SelleckChem), alectinib (S2762, SelleckChem), Sorafenib (S7397, SelleckChem) and osimertinib (S7297, Selleck Chemicals) were dissolved in DMSO to prepare 0.4 M stock solutions, which were aliquoted and stored in -80°C. A 200 µM working dilution of all the drugs was prepared in complete medium and were used to prepare the indicated concentrations for all *in vitro* experiments. A-196 (S7983, Selleck Chemicals) was dissolved in DMSO to prepare 10mg/mL stock solutions, which were aliquoted and stored in -80°C.

2.3.3 Knock-out CRISPR screen:

EKVX cells ($4X10^5$) were plated in 6-well plates and were transfected with 3µg of linearized lentiCas9-Blast (Addgene, 52962) using lipofectamine 2000 (11-668-019, Thermo Fisher Scientific), as per manufacturer's instructions. Forty-eight hours later, cells were selected using 5µg/ml Blasticidin. ECas9 (clone 7) cells stably expressing Cas9 plasmid were clonally selected and characterized. Lentiviral sgRNA library (A and B) were generated and the titer was determined as previously described [116]. The GeCKO V2 library has 6 sgRNAs targeting each protein coding gene and 4 sgRNAs targeting each miRNA. To achieve a 300-fold coverage of the libraries, seventeen 12-well plates were each seeded with $4.5X10^5$ ECas9 cells. Nine plates were transduced with library A, and 8 plates were transduced with library B, both at a multiplicity of infection (MOI) of 0.4 in the presence of polybrene (10µg/ml). Twenty-four hours post transduction, cells were pooled and ~1.31 X 10^7 cells were re-plated in each of seven 15 cm plates containing

complete media supplemented with 2µg/ml blasticidin. Forty-eight hours later cells were plated in six 15 cm plates in media containing 2µg/ml puromycin, to select for library-transduced cells, and 2µg/ml blasticidin. Seventy-two hours later, 2.6 X 10^7 cells were stored for baseline and 2.6 X 10^7 cells were re-plated. The following day, media was replaced with GI75 erlotinib containing media (1.23µM erlotinib) and cells were continuously exposed to GI75 erlotinib for 15 passages. Three biological replicates were performed, and genomic DNA from each baseline and erlotinib treated sample was isolated using the Genomic DNA isolation kit (K1820-01, Thermo Fisher Scientific) following the manufacturer's protocol. For sequencing library preparation, two sequential PCR reactions were conducted for each sample. The first PCR reaction (PCR1) specifically amplified sgRNAs from 1µg of gDNA isolated from each sample. Twenty-five such PCR reactions were conducted, pooled, and gel purified using QIAEX II Gel Extraction Kit (20021, Qiagen). Each PCR1 reaction product (10 ng) was then used for each of 20 PCR2 reactions that were pooled and gel purified. PCR2 fragment sizes and library quality were evaluated on a bioanalyzer (Agilent). Both PCR1 and PCR2 primers are listed in **Table 2.2** (Integrated DNA Technologies). Barcodes included in PCR2 primers were used to identify the samples after deep sequencing. All sequencing was conducted using a NovaSeq 6000 (Illumina). FastQC version 0.11.7 was used to observe sequencing data quality before and after trimming. Cutadapt version 1.13 was used to trim adapters from reads. Reads post-trimming that were shorter than 18nt were discarded. MAGeCK-VISPR v. 0.5.6 was used to perform mapping, allowing no mismatches to ensure accuracy and to reduce bias. Finally, MAGeCK was used to identify over- and under-represented sgRNAs in treated samples relative to baseline, represented as β-scores [117].

2.3.4 Mutant, knockdown, overexpression and rescue experiments:

For EKVX validation studies, KMT5C sgRNA were generated by annealing two oligos (see **Supplementary Table 3**) followed by 5' phosphorylation (T4 Polynucleotide Kinase kit, M0201S, NEB) as described previously [118]. Simultaneously, the CRISPR-Cas9 plasmid, LentiCRISPRv2 (Addgene, 52961) was digested using BsmBI (R0580, NEB), dephosporylated (Antarctic phosphatase, M0289S, NEB), and gel purified using QIAEX II Gel Extraction Kit (20021, Qiagen). The annealed oligos were ligated into the gel purified vector, transformed into Stabl3 bacteria and miniprepped, as outlined previously [118]. Three micrograms of the generated pLV-sgKMT5C plasmid were linearized and forward transfected in 4X10^5 ECas9 (KMT5C wildtype) cells using

lipofectamine 3000 (L3000015, Thermo Fisher Scientific), following the manufacturer's protocol to generate KMT5C mutant clones A, C and E.

For validation studies using PC9 and HCC827 cell lines, $5X10^4$ cells were transfected with Invitrogen TrueCut Cas9 Protein v2 (A36496) along with the Invitrogen TrueGuide Synthetic gRNAs (A35534, Synthego, **Table 2.2**) following the Lipofectamine CRISPRMAX Cas9 transfection protocol (CMAX00001, Thermo Fisher Scientific). Forty-eight hours after transfection, a limiting cell dilution was prepared and 1 cell per well was seeded in a 96-well plate, for clonal isolation and expansion.

For all siRNA-mediated knockdown experiments, 30nM of the respective siRNAs were reverse transfected into $1X10^4$ (for dose curves and proliferation assays) or $4X10^5$ KMT5C mutant clones using Lipofectamine RNAiMAX (13-778-150, Thermo Fisher Scientific) following the manufacturer's protocol. siRNAs used in the study: siMET (Catalog # 4390824, Assay ID # s8700, Thermo Fisher Scientific) and siLINC01510 (Catalog #: 4392420, Assay ID # n506737 Thermo Fisher Scientific).

For generation of DOX inducible overexpression plasmid, the KMT5C sequence was amplified from an ORF expression clone for KMT5C (eGFP tagged) (EX-V0810-M98, GeneCopoeia) introducing a stop codon. The sequence was purified and ligated into the pLVX-Tetone. The oligonucleotides used to perform the sequence exchange are indicated in **Table 2.3**. Following construction of the pLVX-Tetone-KMT5C plasmid, 3µg of the linearized plasmid was transfected into $4X10^5$ Calu6 cells using lipofectamine 3000 to generate the KMT5C-inducible Calu6 clone. Next, to generate the rescue lines from KMT5C mutant clone C, a puromycin resistance gene was cloned into pLVX-Tetone-KMT5C using the primers outlined in **Table 2.3**. Following generation of the pLVX-Tetone-KMT5C-puro plasmid, 3µg of the linearized plasmid was transfected in $4X10^5$ KMT5C mutant cells using lipofectamine 3000 for the generation of inducible-KMT5C rescue clones R1, and R2.

Finally, to test effect of MET or LINC01510 on erlotinib resistance, pT3-EF1a-c-Met (31784, Addgene) or pCMV-Hygro-LINC01510 (Twist Bioscience) were transfected using Lipofectamine 3000 in $4X10^5$ KMT5C wildtype cells.

2.3.5 Genotyping of mutation:

Validation of KMT5C mutations were performed by isolating genomic DNA of each clone (K1820-01, Thermo Fisher Scientific), followed by PCR amplification in the region containing the expected KMT5C mutation using Q5 high fidelity polymerase (M0491L, NEB). PCR products were digested using T7 endonuclease (E3321, NEB) to detect genome editing. PCR products were then purified using QIAquick PCR Purification Kit (28106, Qiagen) and cloned into the TOPO TA cloning vector (K457501, Thermo Fisher Scientific) and six colonies were selected and sequenced for each clone using T7 primer. Primers for amplification and sequencing are outlined in **Table 2.3**.

2.3.6 Bioinformatic analysis of TCGA data:

Cancer Therapeutics Response Portal (CtRPv2) was used to validate the CRISPR-Cas9 knock-out screen [119]. Gene Expression Profiling Interactive Analysis (GEPIA) database [120] (*http://gepia.cancer-pku.cn/*) was used to evaluate *KMT5C, LINC01510,* and *MET* levels in NSCLC patient samples and non-tumorigenic controls. Spearman correlation analysis between *LINC01510* and *MET*, or between *LINC01510* or *MET* and *KMT5C* was also evaluated in LUAD tumor samples using GEPIA. Integrated Genome Viewer (IGV 2.3) was used to view bed files reported by GSE59316 using Human genome 19 (hg19) browser.

2.3.7 Western Blot:

Four-hundred thousand cells were grown in individual wells of a 6 well plate, and lysates were isolated at time points specified in figure legends using RIPA buffer (Sodium chloride (150 mM), Tris-HCl (pH 8.0, 50mM), N P-40 (1 %), Sodium deoxycholate (0.5 %), SDS (0.1 %), ddH2O (up to 100 mL)) containing 1X protease inhibitor cocktail (PIA32955, Thermo Fisher Scientific). Protein quantification was performed using Pierce BCA Protein Assay kit. Equal amounts of protein lysate were resolved through 12% or 4-20% polyacrylamide gels and transferred onto a

polyvinylidene difluoride (PVDF) membrane. Membranes were blocked using LI-COR buffer for 1 hour at room temperature, and incubated overnight in primary antibody at 4°C. The primary antibody was detected using 1:800 IR 800CW secondary antibody. Blots were scanned, and data quantified using the Odyssey LI-COR imaging system and software. Antibodies used: rabbit H4 (61299; Active Motif), mouse H4K20me3 (39672; Active Motif), rabbit H4K20me3 (ab9053, abcam), rabbit MET (D1C2) XP (8198, Cell Signaling), rabbit MKK3 (8535S, Cell Signaling), mouse β-ACTIN (3700, Cell Signaling). Lysates from Supplemental Figure 3B, where prepared using the histones acid extraction protocolas described in Shechter, D., Dormann, H., Allis, C. et al., 2007.

2.3.8 In-Cell Western:

Ten-thousand cells were grown in individual wells of a 96-well plate. Forty-eight hours post plating, cells were fixed using cold 100% methanol for 20 minutes at 4 C. Post fixing, cells were permeabilized using 0.2% TritonX in 1X PBS at room temperature for 30 minutes. Cells were blocked using LI-COR blocking buffer for 1.5 hours followed by overnight incubation with primary antibody at 4°C. The primary antibody was detected using 1:800 IR 800CW secondary antibody (LI-COR). The IR-800 signal was quantified using the Odyssey LI-COR imaging system and software. Antibodies used: 1:400 mouse H4K20me3 (39672, Active Motif), 1:500 rabbit GAPDH (2118, Cell Signaling)

2.3.9 Immunofluorescence:

Two-hundred thousand cells were seeded on collagen coated coverslips that were arranged in individual wells of a 12 or 24-well plate. Forty-eight hours post plating, cells were fixed using cold 100% methanol for 20 minutes in 4°C. Post fixing, cells were permeabilized using 0.2% TritonX in 1X PBS at room temperature for fifteen minutes followed by blocking using LI-COR blocking buffer for 1 hour. For KMT5B/C inhibitor experiments, cells were fixed and permeabilized using cold 100% methanol for 10 minutes at -20ºC followed by blocking using 0.2µ-filtered 1% Bovine Serum Albumin. Following blocking, cells were incubated overnight with 1:50 mouse H4K20me3 (39672, Active Motif) or 1:50 rabbit anti-H4 antibody (13919S, Cell Signaling) at 4°C. After primary antibody incubation, cells were incubated with secondary antibodies and nuclear stain for

2 hours at room temperature. 1:500 anti-mouse Alexa Fluor 647 (A-31571, Thermo Fisher Scientific) and 1:500 anti-rabbit Alexa Fluor 488 (A-11034, Thermo Fisher Scientific) was used to detect H4K20me3 and H4, respectively, and 1:1000 Hoechst dye (H3570, Thermo Fisher Scientific) was used as a nuclear stain. Coverslips were mounted on glass slides using ProLong™ Glass Antifade Mountant (P36982, Thermo Fisher Scientific). Images were acquired using Nikon A1R-MP microscope with a 40X oil objective (Nikon Inc., Melville, NY, U.S.A). The images were acquired and analyzed using the Nikon NIS-Elements imaging software (version 5.20.02) in the ".nd2" format. The acquisition settings were 1K x 1K resolution (pixels) with a scanning frame rate of 1/8 sec. All images were set to the same display lookup table (LUT) settings before exporting the files.

2.3.10 Dual Glo Luciferase assay:

$6X10^4$ EKVX cells were transfected in a 24-well plate using Lipofectamine 2000 (11668019, Thermo Fisher Scientific) following the manufactured protocol. Co-transfection of 200ng pGL3 with the 5' UTR of KMT5C cloned upstream of the Firefly Luciferase gene, along with 2ng of pGL4.75, a vector expressing Renilla Luciferase and with 6nM of miR-4435/pmiRNC (miRNA Scramble negative control) was performed. 48 hours later, Dual Glo Luciferase assay (PR-E2940, Promega) was performed following manufactured protocol.

2.3.11 RNA isolation and Quantitative real time PCR (qRT-PCR):

Four-hundred thousand cells were grown in individual wells of a 6-well plate, and total RNA was isolated after 48 or 96 hours, as indicated, using the miRneasy Kit (217004, Qiagen) according to the manufacturer's instruction. DNase I digestion (79254, Qiagen) was used in each RNA purification reaction to remove genomic DNA. RNA integrity was evaluated on a 1.5% agarose gel, and total RNA quantified using a nanodrop. For quantifying RNA from EGFR wildtype cells, cDNA was then synthesized from 1µg of total RNA using miScript Reverse Transcriptase kit (218161, Qiagen), as indicated by the manufacturer's protocol. Q-RT-PCR was conducted using the miScript SYBR Green PCR Kit (218073, Qiagen) as indicated by the manufacturer's protocol, to quantify target gene mRNA expression. The following primers were obtained:

GAPDH (loading control) (QT00079247, Qiagen), *LINC01510* (LPH09040A, Qiagen), and MET (QT00023408, Qiagen). Primers for *KMT5C* quantification are indicated in **Table 2.3**.

KMT5C transcript from EGFR mutant cell lines was quantified using Taqman assays. cDNA was synthesized from 900ng of total RNA using SuperScript IV VILO Master Mix (11756050, Thermo Fisher Scientific). Q-RT-PCR was conducted using Taqman Fast Advanced Master Mix (4444963, Thermo Fisher Scientific). The following primers were used: KMT5C:(Hs00261961_m1, Thermo Fisher Scientific) and GAPDH (endogenous control) (Hs99999905_m1, Thermo Fisher Scientific).

2.3.12 ChIP-qPCR:

Briefly, a total of $2X10^7$ cells were fixed using 1% of filter-sterilized formadehyde for 10 minutes at room temperature. The formaldehyde was quenched with 2.5M Glycine (55µL per ml of media) for 5 min. Cells were washed with cold PBS and scraped into fresh cold PBS. Cells were pelleted by centrifuging at 1500 rpm for 10 minutes at 4°C. The cell pellet was resuspended in 10 mL of freshly prepared cold cell lysis buffer (5mM PIPES, 85mM KCl, 0.5% NP40), kept on ice for 10 minutes followed by centrifuging at 1000 rm for 10 minutes at 4°C. The lysed cells were resuspended in 1 mL of nuclei lysis buffer (50mM Tris-HCl (pH 8.0),10mM EDTA, 1% SDS) containing 0.1% protease inhibitor cocktail (PIA32955, Thermo Fisher Scientific) and were transferred into 2mL eppendorf tubes, on ice. Cross-linked chromatin from the isolated nuclei was sonicated using a probe sonicator (60% duty cycle) for 10 seconds with a 1minute rest, for 15 cycles to fragment DNA (100-500 bps). Fragmented DNA was immunoprecipitated with antibodies against mouse H4K20me3 (39672, Active Motif), or negative control mouse IgG (5415, Cell Signaling Technology) at 4°C overnight with gentle rotation. The immunoprecipitated DNA was purified using the DNA isolation kit (K1820-01; Thermo Fisher Scientific) following manufacturer's protocol. DNA was used as a template for qRT-PCR as described above. All primer sequences used for qRT-PCR are listed in **Table 2.3**. ChIP data are presented as fold enrichment of DNA immunoprecipitated with H4K20me3 relative to values obtained for DNA immunoprecipitated with IgG control.

2.3.13 Erlotinib dose response:

The protocol followed to evaluate erlotinib dose response was as per the NCI-60 Cell Five-Dose Screen [121]. Briefly, Sulforhodamine B colorimetric assay (SRB assay)[122] was performed by exposing cells to varying concentrations of erlotinib or the highest equivalent volume of DMSO (negative control) containing media for 72 hours. To normalize data, percent of cells was calculated based on first correcting for the number of cells at the start of the assay (time zero = tz), followed by normalization of cell number to respective corrected DMSO values.

2.3.14 Proliferation:

Ten thousand NSCLC cells or transfected cells were seeded in 6 replicates in wells of a 96-well plate, which was placed in a live-imaging system, Incucyte s3 2018A (ESSEN BioScience). Plates were incubated in the system for the specified times. Four images per well were obtained every 2 hours using the 10X objective. Confluence was evaluated using Incucyte s3 2018A software. To normalize data, percent of cells was calculated based on first correcting for the number of cells at the start of the assay (time zero = tz), followed by normalization of cell number to respective corrected DMSO values. Data is represented relative to controls, as described in figure legends.

2.3.15 Clonogenic assay:

Five thousand cells of HCC827 were seeded in 6 well plates and distributed evenly. The next day, media containing 0.1 or 0.01μM of erlotinib or the highest equivalent of DMSO was added. Media containing erlotinib was changed every two days, and the plate was fixed 8 days after seeding using the DIFF-Quick Stain Kit and following the manufacturer's protocol (NC1796273, Polyscience).

2.3.16 Statistical analysis:

All data were analyzed using GraphPad Prism version 9 software (GraphPad Software) and are presented as mean values ± standard deviation (SD). Pearson's correlation was utilized to evaluate linear correlation between KMT5C and/or H4K20me3 and GI50 erlotinib values. Student's t-test or one-way ANOVA were performed, as specified in the figure legends. P-value of < 0.05 was considered significant.

2.4 Results

2.4.1 Identification of mediators of erlotinib resistance

To identify genes that when mutated confer resistance to erlotinib sensitive cells, a genome-wide CRISPR-Cas9 screen was performed. The screen was conducted in EKVX cells, a cell line that was determined to be erlotinib sensitive in data obtained from the Developmental Therapeutics Program, which is maintained by the National Cancer Institute (NCI-60, DTP). EKVX cells were engineered to stably express the Cas9 protein and resulting clones were validated for erlotinib sensitivity, which was similar to the parental EKVX cells **(Figure 2.1).** Cas9-expressing EKVX clone 7 was taken forward to conduct the screen, which is hereafter referred to as ECas9. ECas9 cells were infected with the GeCKO V2 sgRNA lentiviral library targeting 19,052 protein-coding genes and 1,864 miRNA genes (**Figure 2.2A**) [123]. To obtain full coverage of the lentiviral sgRNA library, transduction was performed at 300-fold coverage and was conducted in triplicates to mitigate false positives. One third of the transduced cells were used to determine the representation of the integrated sgRNAs prior to selection in erlotinib (baseline). The remaining cells were grown for 15 passages in the presence of 1.23 μM erlotinib, a concentration that inhibits growth of 75% of ECas9 cells (GI75). Integrated sgRNAs were identified from the resulting population, and from the baseline cells, by PCR amplification and subsequent high-throughput sequencing. Combined analysis of the three replicates using the MAGeCK-VISPR algorithm identified significantly enriched sgRNAs in the population of cells that were cultured in erlotinib (**Table 2.1**, **Table 2.2**, **Figure 2.2B**) [117]. Following the analysis, multiple genes that were previously reported to be 1) downregulated during acquired resistance to chemotherapy treatment (EGFRi or non-EGFRi) [124], 2) highly expressed in erlotinib sensitive cells [125,126], and 3) *bona fide* tumor suppressors [27,88,127–130], were identified among the top hits, validating the sensitivity of the screen and appropriateness of the chosen cell line.

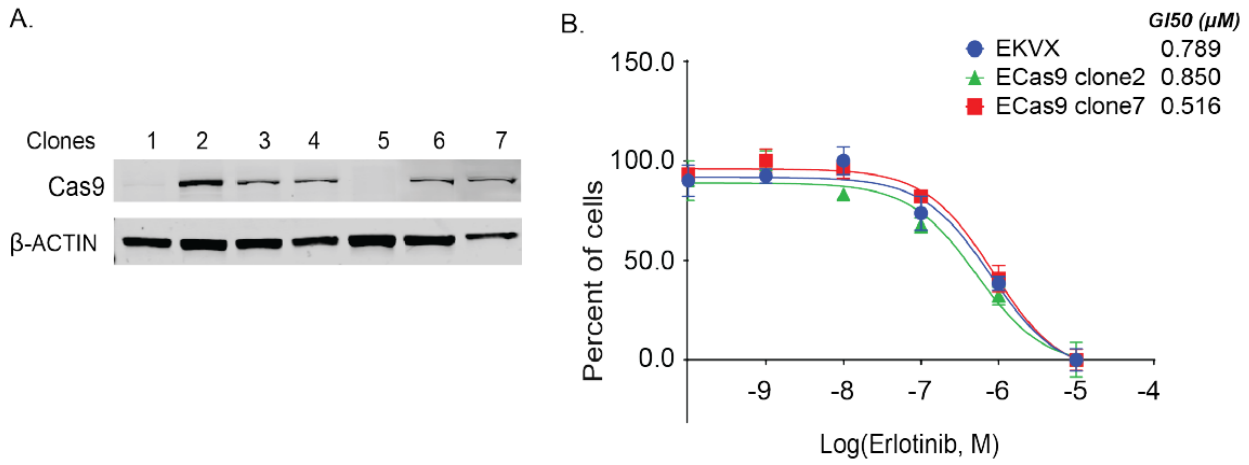

Figure 2.1 Characterization of Cas9 expressing EKVX clones. A) Western Blot analysis of Cas9 levels in EKVX clones stably expressing Cas9. β-ACTIN was used as a loading control. **B)** Parental EKVX cells, ECas9 clone 2, and ECas9 clone 7 were exposed to varying concentrations of erlotinib or the highest equivalent volume of dimethyl sulfoxide (DMSO, negative control) containing media for 72 hours. Erlotinib dose response was evaluated using the SRB assay.

Table 2.1. Candidate genes identified from the CRISPR-Cas9 knock out screen. Thirty-five significant hits identified by MAGeCK-VISPR analysis and β-score, p-value, and false discovery rate (FDR)

Target	β-score	p-value	FDR
KMT5C	97	8.30E-05	0.07
ADSS	91	0.00021	0.07
OPA3	89	0.00028	0.07
LEPREL4	88	0.00032	0.07
GAREM	86	0.00049	0.07
ISG15	83	0.00065	0.07
PROM2	83	0.00065	0.07
hsa-mir-602	77	0.00082	0.07
CCDC130	81	0.00088	0.07
PCSK2	80	0.00091	0.07
FAM120AOS	79	0.001	0.07
CCL23	79	0.0011	0.07
TNFSF12	76	0.0028	0.07
hsa-mir-27b	74	0.0081	0.11
SMN2	25	0.012	0.16
OR6V1	74	0.012	0.16
SYBU	72	0.012	0.17
CASP8	73	0.012	0.17
LDLRAP1	71	0.013	0.17
PFDN2	70	0.013	0.17
CPA3	68	0.013	0.17
PP2D1	68	0.013	0.17
TMEM234	68	0.013	0.17
TMEM147	67	0.013	0.17
hsa-mir-5699	62	0.016	0.21
hsa-mir-512-1	50	0.016	0.21
MLL2	22	0.016	0.21
hsa-mir-648	43	0.016	0.21
AGAP9	22	0.016	0.21
hsa-mir-4669	43	0.016	0.21
RPL41	38	0.016	0.21
hsa-mir-3183	37	0.016	0.21
hsa-mir-1268a	34	0.017	0.22
hsa-mir-147b	34	0.017	0.22
hsa-mir-148a	27	0.018	0.24

Table 2.2. Primer sequences used to conduct the CRISPR-Cas9 screen. Multiple PCR2 primers were used, each with an independent barcode that allows for sorting of sample-specific sgRNAs post sequencing.

PCR	Sample	Primer name	Primer direction	Primer sequence
PCR 1	All samples	1st PCR primer	Forward	TCTTTCCCTACACGACGCTCTTCCGATCTNNNNAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
		1st PCR primer	Reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCC
PCR2	EKVX-Baseline 1	UDA5050	Forward	AATGATACGGCGACCACCGAGATCTACACTGACAATGTCACACTCTTTCCCTACACGAC
		UDA7143	Reverse	CAAGCAGAAGACGGCATACGAGATAGAAGCCAATGTGACTGGAGTTCAGACGTG
	EKVX-Replicate 1	UDA5051	Forward	AATGATACGGCGACCACCGAGATCTACACCGACCTAACGACACTCTTTCCCTACACGAC
		UDA7142	Reverse	CAAGCAGAAGACGGCATACGAGATGACTCACTAAGTGACTGGAGTTCAGACGTG
	EKVX-Baseline 2	UDA5052	Forward	AATGATACGGCGACCACCGAGATCTACACTAGTTCGGTAACACTCTTTCCCTACACGAC
		UDA7141	Reverse	CAAGCAGAAGACGGCATACGAGATAGTCTGTCGGGTGACTGGAGTTCAGACGTG
	EKVX-Replicate 2	UDA5053	Forward	AATGATACGGCGACCACCGAGATCTACACGCCGCACTCTACACTCTTTCCCTACACGAC
		UDA7140	Reverse	CAAGCAGAAGACGGCATACGAGATGTATTCTCTAGTGACTGGAGTTCAGACGTG
	EKVX-Baseline 3	UDA5054	Forward	AATGATACGGCGACCACCGAGATCTACACATTATGTCTCACACTCTTTCCCTACACGAC
		UDA7139	Reverse	CAAGCAGAAGACGGCATACGAGATACGCCTCTCGGTGACTGGAGTTCAGACGTG
	EKVX-Replicate 3	UDA5055	Forward	AATGATACGGCGACCACCGAGATCTACACAGAACCGAGTACACTCTTTCCCTACACGAC
		UDA7138	Reverse	CAAGCAGAAGACGGCATACGAGATTAACCGCCGAGTGACTGGAGTTCAGACGTG

sgRNA sequences used to generate KMT5C mutant cell lines. Designed and purchased from Invitrogen.

sgRNA name	**sgRNA sequence**
Exon3 sgRNA (EGFR WT cell lines)	CGGCCCGCTACTTCCAGAGC
Exon7 sgRNA1 (EGFR Mutant cell lines)	GUGAAUGCCACACCUGUGAG
Exon7 sgRNA2 (EGFR Mutant cell lines)	AAGCAUGUCACCUCGUCCCC

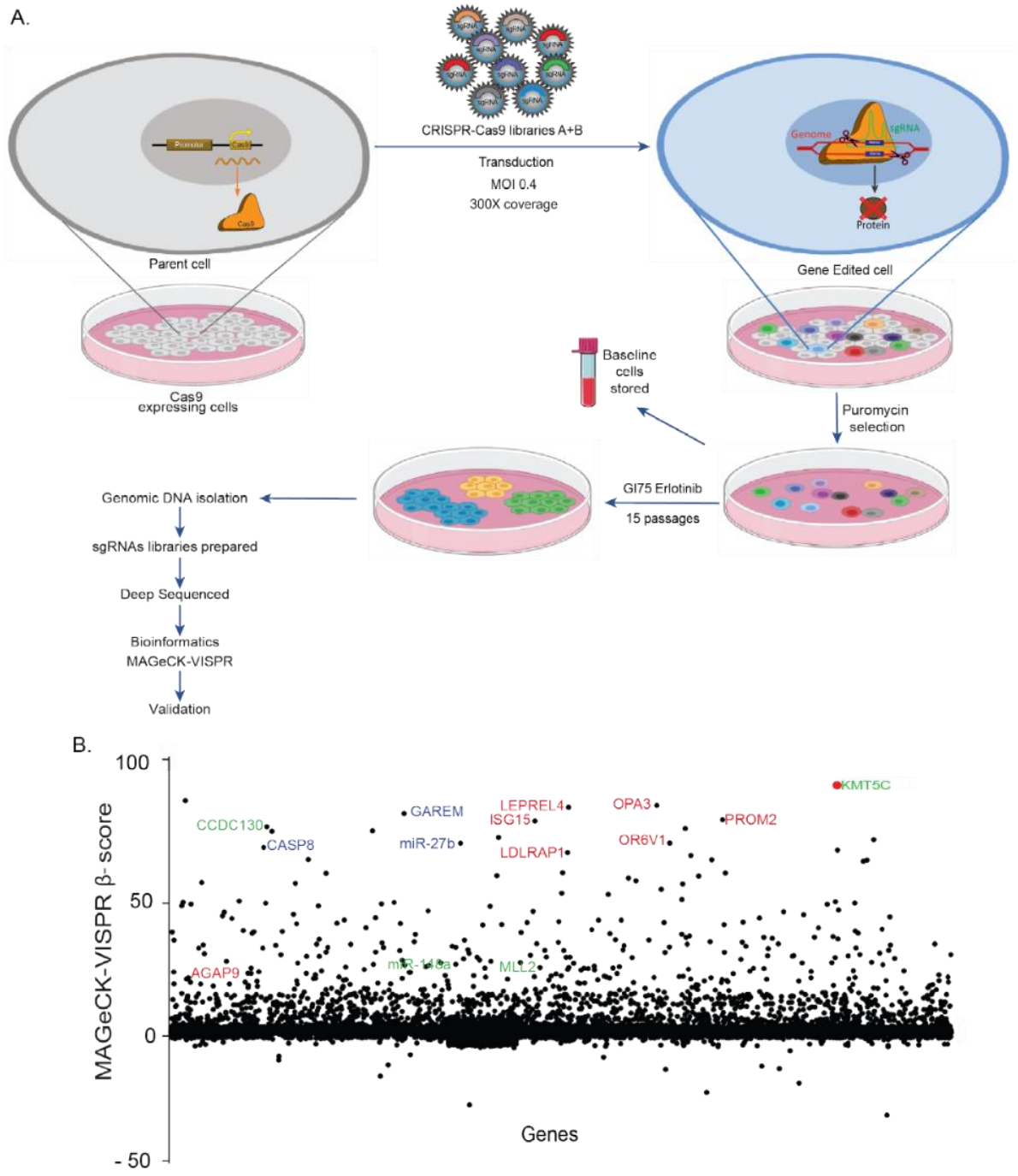

Figure 2.2. A genome-wide CRISPR-Cas9 screen identifies mediators of erlotinib resistance. **A)** Outline of the screen. **B)** Fold enrichment (β-score) analysis of sgRNAs. Blue, genes previously reported to be downregulated in cells after chemotherapeutic treatment; red, genes reported to be high in erlotinib-sensitive cells; green, genes reported as tumor suppressors.

2.4.2 Low expression of KMT5C is associated with erlotinib resistance, and predicts poor prognosis in NSCLC

The top hit from the CRISPR-Cas9 knock-out screen, KMT5C is a histone methyltransferase also referred to as SUV420H2. KMT5C specifically trimethylates histone H4 lysine-20 (H4K20), which is associated with transcriptional repression and is important for establishing constitutive heterochromatic regions [23,32]. Multiple studies have reported on the role of KMT5C as a tumor suppressor, and both KMT5C and H4K20 trimethylation (H4K20me3) are severely downregulated in multiple cancers [27,83,86,88,131]. To determine if KMT5C is also a mediator of erlotinib response, various validation assays were performed. Firstly, using an extensive panel of NSCLC cell lines, cell lines included in the DTP and additional EGFR mutant cell lines, a negative correlation between *KMT5C* transcript and erlotinib response was determined (**Figure 2.3A-D,** Pearson r = -0.83, **Figure 2.4**). Because of the lack of a sensitive and specific KMT5C antibody for immunoblotting, the downstream effector of KMT5C, H4K20me3 was evaluated as a proxy for KMT5C activity (**Figure 2.5A-B)**. Indeed, in the same cell line panel, H4K20me3 levels positively correlate with *KMT5C* transcript (Pearson r = 0.24, **Figure 2.5C**). Additionally, similar to the negative correlation between *KMT5C* transcript and erlotinib response in the NSCLC panel, H4K20me3 also displayed a negative correlation with erlotinib response (Pearson r = -0.50, **Figure 2.5D**). These strong correlations suggest a possible role for KMT5C and H4K20me3 levels in mediating the response of NSCLC cells to erlotinib.

Next, we investigated *KMT5C* transcript levels in NSCLC patient samples using publicly available data provided by The Cancer Genome Atlas [132] and the Genotype-Tissue Expression [133] projects. Patient samples were compared to non-cancerous control tissues using Gene Expression Profiling Interactive Analysis (GEPIA, **Figure 2.3E**) [120]. *KMT5C* transcript levels were generally lower in both lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) samples relative to normal samples suggesting that KMT5C may function as a *bona fide* tumor suppressor.

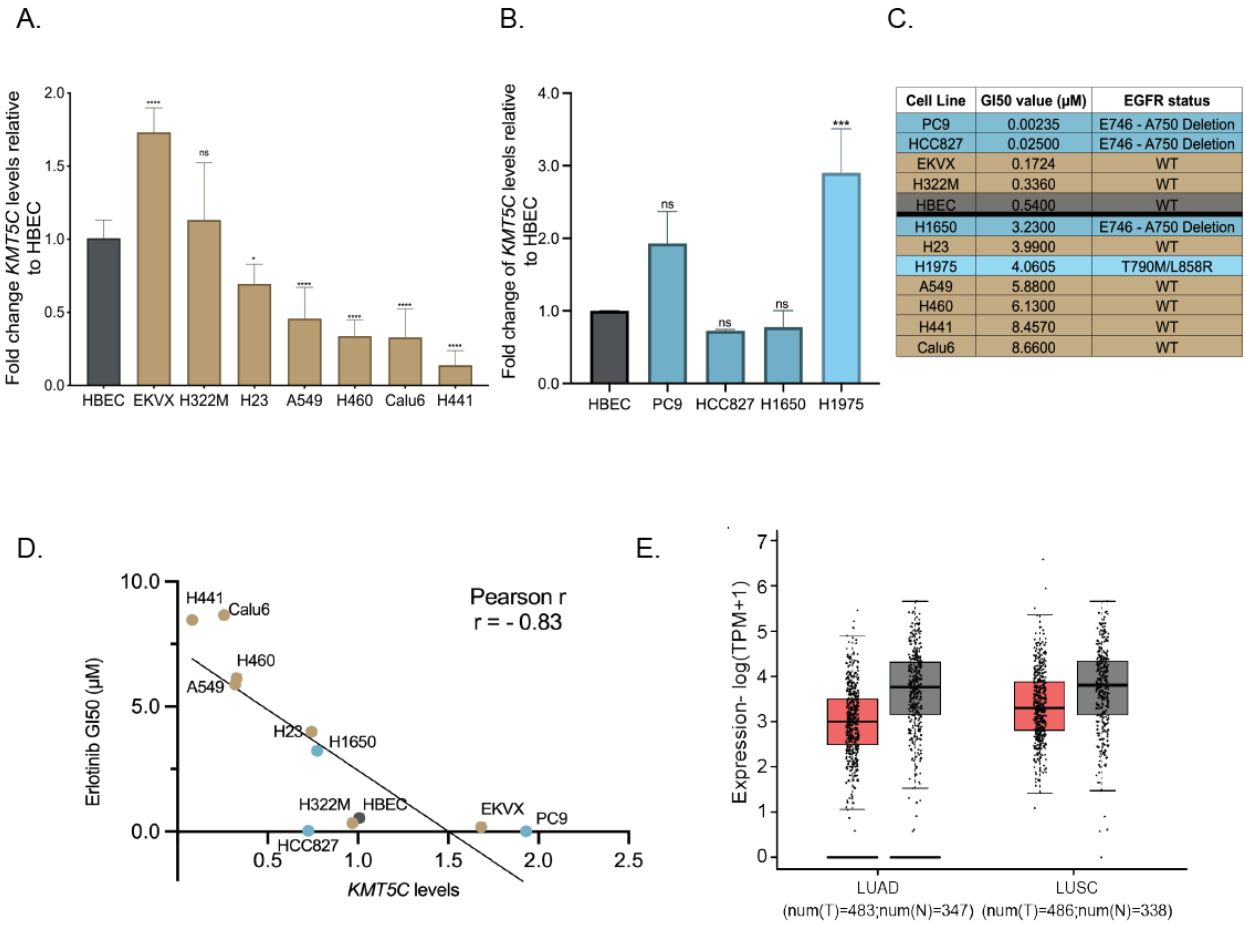

Figure 2.3. Reduced KMT5C transcript correlates with erlotinib resistance in NSCLC cells and poor prognosis in patients with NSCLC. **A** and **B)** Expression of KMT5C in NSCLC cells represented in the DTP (A) or with mutation(s) in EGFR, relative to a nontumorigenic lung epithelial cell line (human bronchial epithelial cells, HBEC; **B)** evaluated by qRT-PCR. Data are normalized to GAPDH and relative to HBEC. One-way ANOVA followed by Dunnett multiple comparison test was used to evaluate statistical significance. Color of bars represents EGFR mutation status: gold, EGFR wt; dark teal, EGFR primary mutation; light teal, EGFR secondary mutation. **C)** Erlotinib dose–response evaluated by exposing cell lines to varying concentrations of erlotinib or the highest equivalent volume of DMSO containing media for 72 hours followed by SRB assay. GI50 concentrations of erlotinib were calculated from respective dose curve. **D)** Correlation analysis between KMT5C transcript from A/B and GI50 erlotinib concentrations from C. **E)** GEPIA analysis for KMT5C transcript levels in normal (gray bars) and tumor samples (pink bars) from LUAD and LUSC data obtained from TCGA and the GTEx databases. TPM, transcripts per million; T, tumor; N, normal. ns, nonsignificant; *, P < 0.05; **,P <0.001; ***,P < 0.0001.

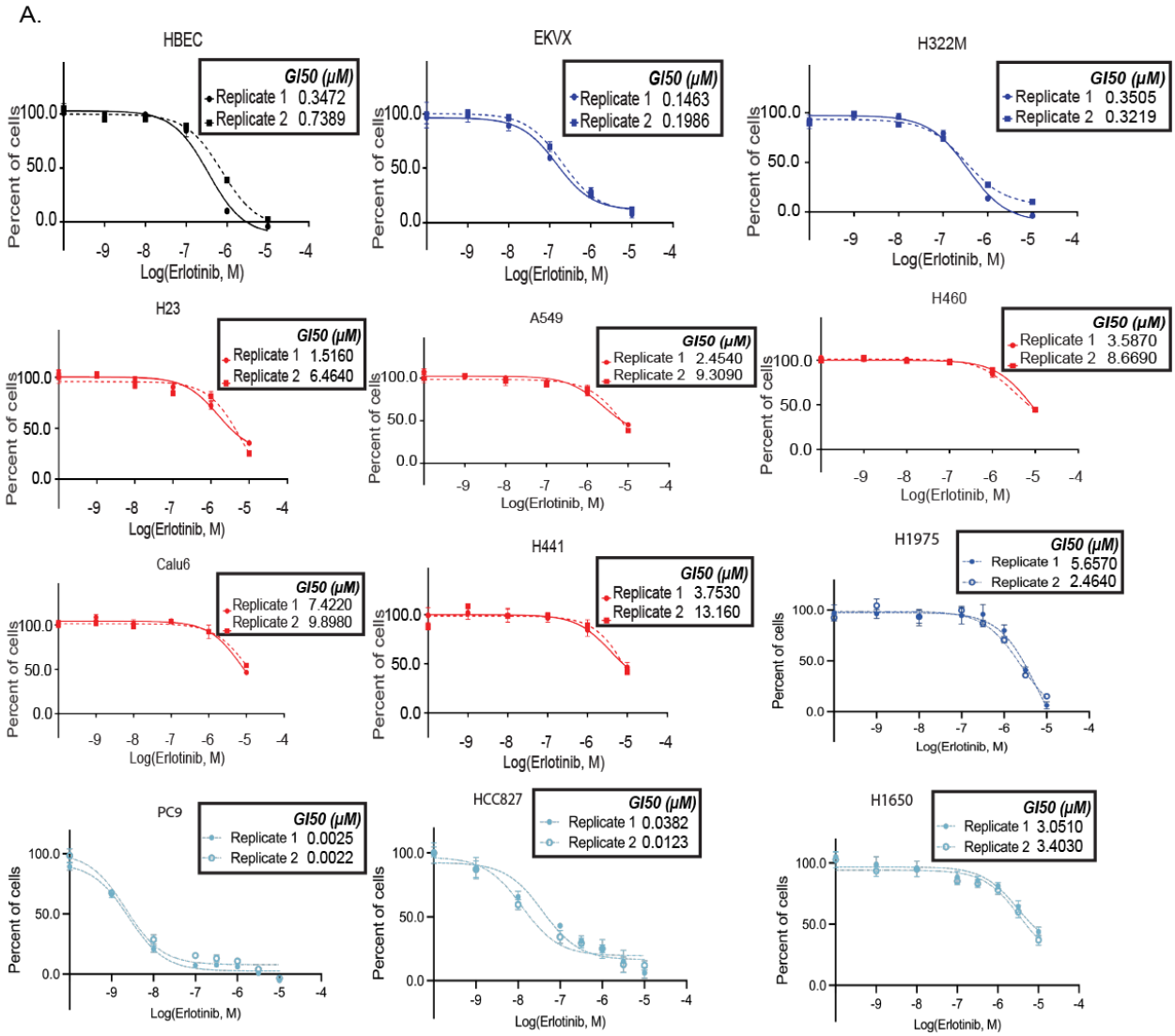

Figure 2.4. Growth inhibition for a panel of NSCLC cell lines following exposure to increasing doses of erlotinib. A panel of NSCLC cell lines were exposed to varying concentrations of erlotinib or the highest equivalent volume of dimethyl sulfoxide (DMSO, negative control) containing media for 72 hours. Erlotinib dose response was evaluated using the SRB assay. Post-normalization, the GI50 concentration of erlotinib was calculated from the respective dose curve for each cell line, two replicates were performed for each cell line. GI50 values in Figure 2C are the average of the replicates indicated here.

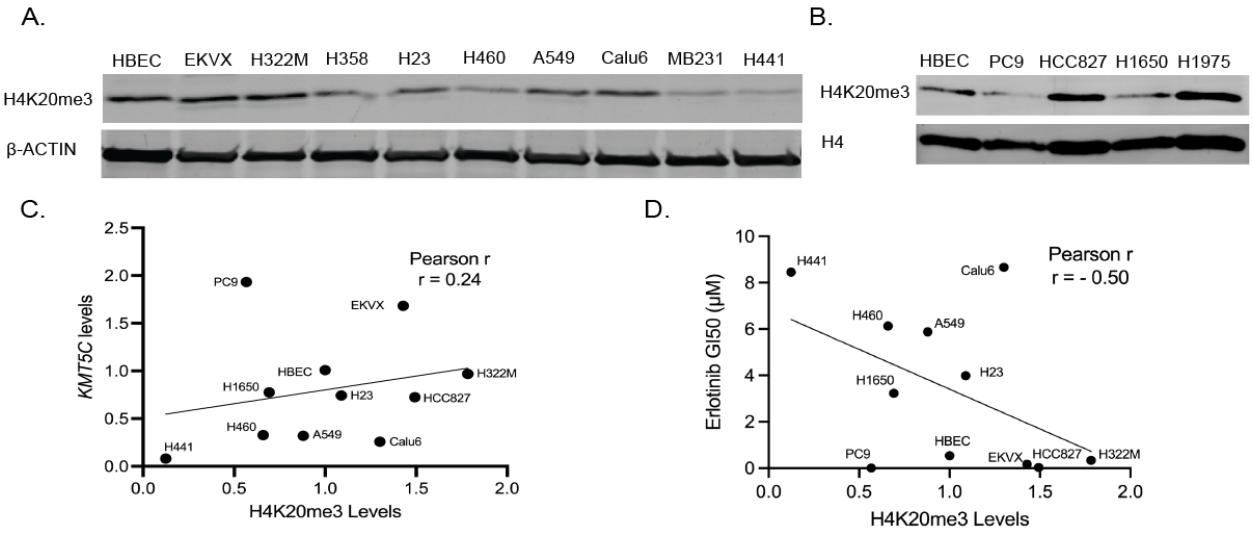

Figure 2.5. Reduced H4K20me3 correlates with erlotinib resistance in NSCLC cells. Representative western blot of H4K20me3 in a panel of NSCLC cells that include **A)** cell lines from the NCI-60 DTP program and **B)** EGFR mutant cell lines. HBEC serves as a control on each blot. β-ACTIN was used as a loading control. MB231, a breast cancer cell line was included as a control cell line reported to have low levels of KMT5C (Shinchi et al., 2015). **C)** Correlation analysis between quantified H4K20me3 levels from panel A/B and KMT5C from Figure 2A/B. **D)** Correlation analysis between quantified H4K20me3 levels from panel A/B and GI50 erlotinib values from Figure 2C. Evaluations in C and D were conducted using the Pearson correlation test.

2.4.3 Loss of KMT5C confers resistance to EGFR inhibitors.

To further validate the findings from the CRISPR-Cas9 screen, ECas9 cells (KMT5C wildtype) were transfected with a sgRNA targeting *KMT5C* to generate three KMT5C mutant lines, clones A, C and E. Genotyping validated that the sgRNA specifically targeted *KMT5C* resulting in various insertions and deletions (**Figure 2.6A**). *KMT5C* transcript levels were reduced in all three clones (**Figure 2.7A**) resulting in downregulation of H4K20me3 (**Figure 2.7B, 2.7C** and **Figure 2.6B**). Comparing the erlotinib response of the mutant clones to that of KMT5C wildtype cells confirmed that loss of KMT5C leads to 5.4 – 11.7 fold increase in erlotinib resistance (**Figure 2.7D**). Additionally, increased proliferation of the mutant clones relative to KMT5C wildtype cells in the presence of erlotinib corroborated the results (**Figure 2.7E**). We also evaluated the response of KMT5C mutant clones to other EGFRi including afatinib, gefitinib, and osimertinib. All the clones were resistant to all three EGFRi (**Figures 2.6C-2.6H**) – resistance to the irreversible inhibitor, afatinib was ~300 to 900-fold higher in the mutant clones relative to wildtype KMT5C cells. Conversely, mutant clones were unaffected in the presence of cisplatin (**Figure 2.8G**) suggesting

that loss of KMT5C is not a global mediator of resistance but may be specific to EGFRi or perhaps other targeted agents.

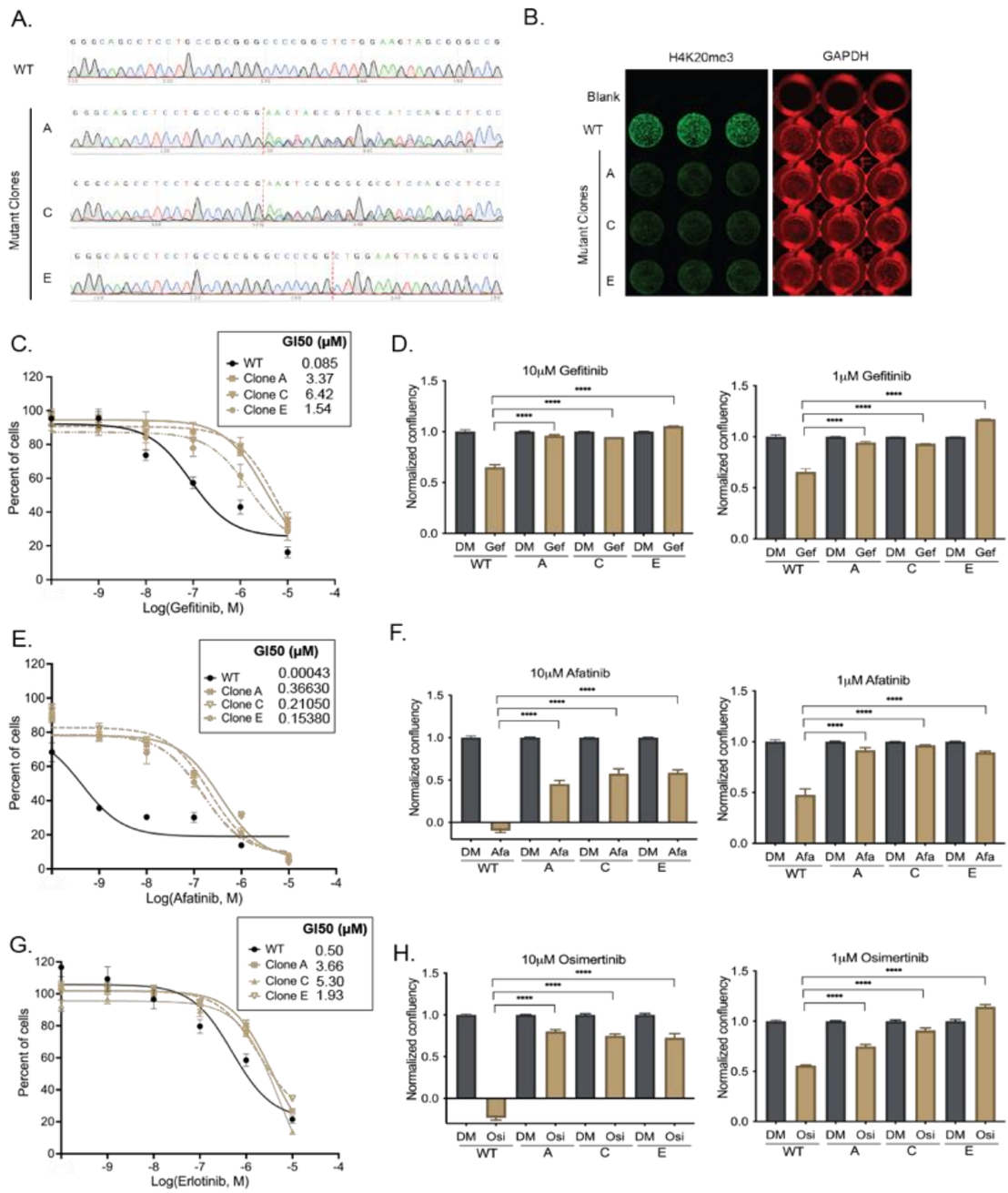

Figure 2.6. KMT5C mutation confers resistance to various EGFRi. A) Genomic DNA of EKVX WT cells or mutant clones A, C, E was isolated, the region targeted by CRISPR-Cas9 sgRNA targeting KMT5C was PCR amplified, purified and sequenced. Representative chromatograms of the wildtype KMT5C (WT) cells, and the specific mutations identified in mutant clones A, C, E. **B)** In-cell western of H4K20me3 levels in EKVX WT cells and mutant clones A, C, E. GAPDH serves as an endogenous control. **C)** Gefitinib, **E)** Afatinib, or **G)** Osimertinib dose response curves. Cells were exposed to the indicated concentration of drug or to the highest equivalent volume of vehicle control containing media for 72 hours. Following normalization, the GI50 concentration of each inhibitor was calculated from the respective dose curve for each cell line. Proliferation of EKVX WT cells or mutant clones A, C, E was evaluated using the Incucyte. Cells were exposed to varying concentrations of **D)** Gefitinib (Gef) **F)** Afatinib (Afa) or **H)** Osimertinib (Osi) or the highest equivalent volume of DMSO (DM) containing media for 72 hours. Data relative to respective normalized DMSO control treatments is represented. Oneway ANOVA followed by Dunnett's Multiple Comparison test was utilized to evaluate statistical significance of normalized confluency of clones A, C, E in the presence of 10 or 1μM of gefitinib, afatinib or osimertinib compared to WT cells.

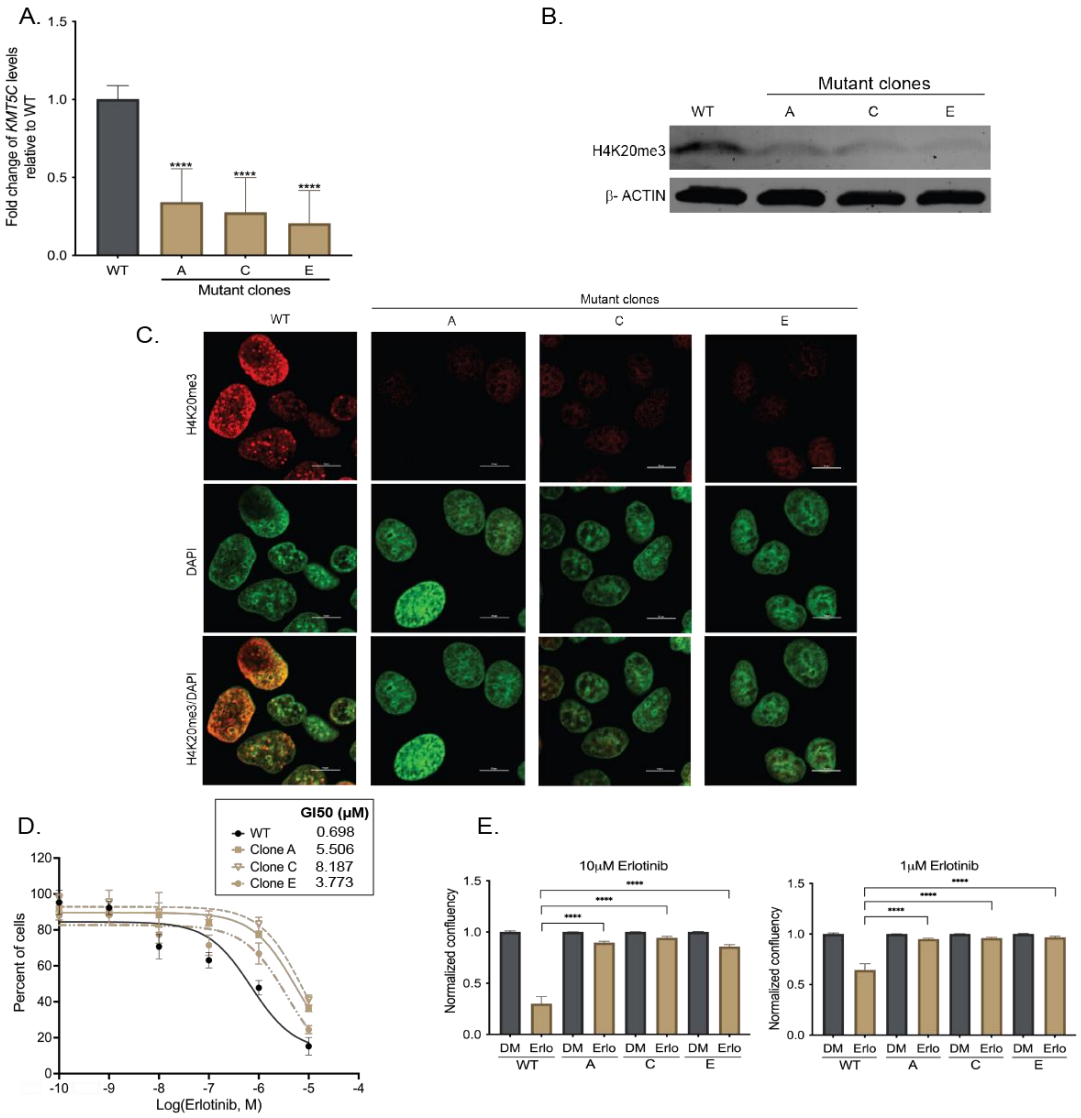

Figure 2.7. Loss of KMT5C confers resistance to erlotinib. A) Expression of KMT5C transcript in EKVX mutant clones A, C, and E. Data were normalized to GAPDH and are represented relative to ECas9 (KMT5C wild type, WT) cells. One-way ANOVA was used to evaluate statistical significance. **B)** Representative Western blot of H4K20me3 in EKVXWT cells and KMT5C mutant clones A, C, and E. b-ACTIN served as a loading control. **C)** Representative immunofluorescent image of H4K20me3 in WT cells and clones A, C, and E. Scale bar, 10 mm. **D)** Erlotinib dose response following exposure to the indicated concentrations of erlotinib or the highest equivalent volume of DMSO for 72 hours. Following normalization, the GI50 concentration of erlotinib was calculated from the respective dose curve. **E)** Live cell imaging of WT or mutant clones (represented as A, C, and E) was conducted to quantify proliferating cells in the presence of erlotinib (Erlo) or vehicle control (DMSO, DM) for 72 hours. Data relative to respective normalized DMSO control treatments are represented. One-way ANOVA followed by Dunnett multiple comparison test was used to evaluate significance. ***, $P < 0.0001$.

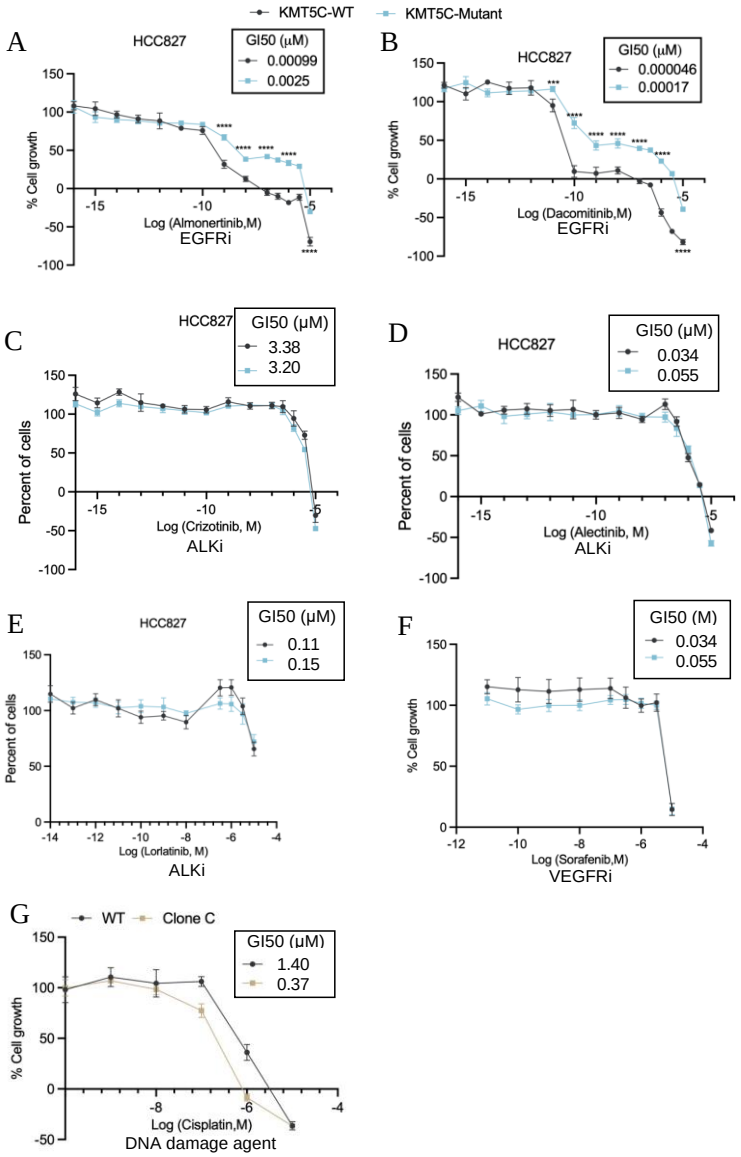

Figure 2.8. KMT5C mutation confers resistance to third generation EGFRi but not to ALKi, VEGFRi and DNA damage agent Cisplatin. A) Almonertinib, **B)** Dacomitinib, **C)** Crizotinib **D)** Alectinib **E)** Lorlatinib **F)** Sorafenib **G)** Cisplatin dose response curves. Cells were exposed to the indicated concentration of drug or to the highest equivalent volume of vehicle control containing media for 72 hours. Following normalization, the GI50 concentration of each inhibitor was calculated from the respective dose curve for each cell line. Oneway ANOVA followed by Dunnett's Multiple Comparison test was utilized to evaluate statistical significance of normalized confluency**.**

The primary CRISPR-Cas9 screen was conducted using EKVX, an EGFR wildtype cell line that was the most sensitive cell line tested in the NCI-60 DTP data. And, although erlotinib has been used in the past to clinically to treat EGFR wildtype tumors, it was imperative to determine if mutant KMT5C could also drive resistance in EGFR mutant cell lines. Thus, four EGFR mutant cell lines were identified, and the two most sensitive, PC9 and HCC827, (**Figure 2.3C**) were validated for erlotinib resistance in the presence of mutant KMT5C. Elevations in EGFR and upregulation of two activating phosphorylation sites at tyrosine 1045 (Y1045) and tyrosine 1068 (Y1068) for the EGFR mutant cells relative to the wildtype cells are presented in **Figure 2.9**. KMT5C Cas9-mediated mutations were generated in both PC9 and HCC827 using divergent sgRNAs located in exon 7, a region of the gene that when translated is essential for the catalytic function of KMT5C (**Figure 2.10A**). Individual clones were isolated and one clone of each cell line with deletions in both alleles (**Figure 2.10B)** and downregulation of H4K20me3 (**Figure 2.10C)** was brought forward. Both cell lines were cultured in the presence of increasing doses of erlotinib along with the respective parental cell line. Similar to the data obtained using EKVX, both PC9 and HCC827 became resistant to erlotinib when KMT5C was mutated (**Figure 2.10E-F**) as evident following dose titration studies. HCC827 was also validated using a colony formation assay where virtually no colonies were visible after culturing in even the lowest dose of erlotinib (0.01uM) (**Figure 2.10D)**. Similar to EKVX, both PC9 and HCC827 also became resistant to osimertinib when KMT5C was mutated (**Figure 2.10G-H**).

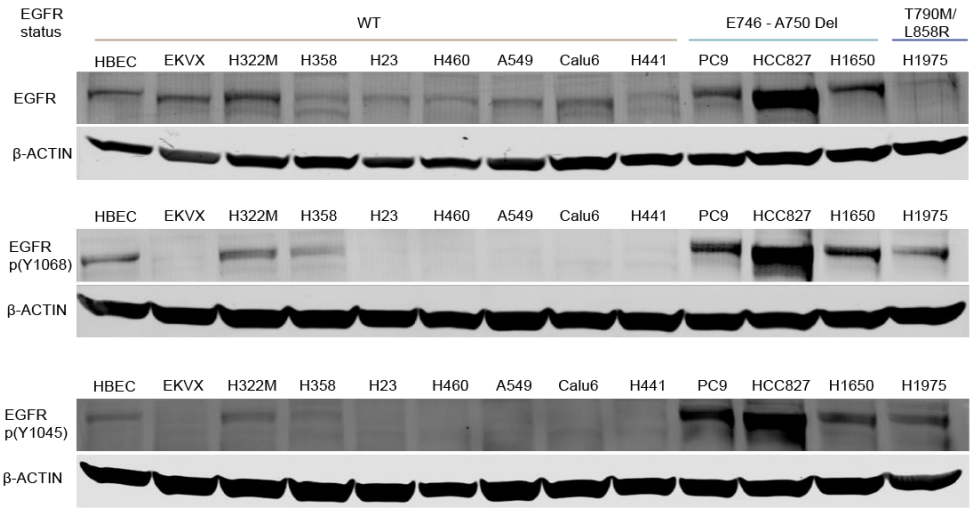

Figure 2.9. EGFR status of cell lines used in this study. Western Blots of EGFR, EGFR p(Y0168) and EGFR p(Y1045) in a panel of EGFR WT and mutant NSCLC cell lines. PC9, HCC827 and H1650 harbor E746-A750 deletion and H1975 harbors the mutation T790M/L858R in EGFR. β-ACTIN was used as a loading control.

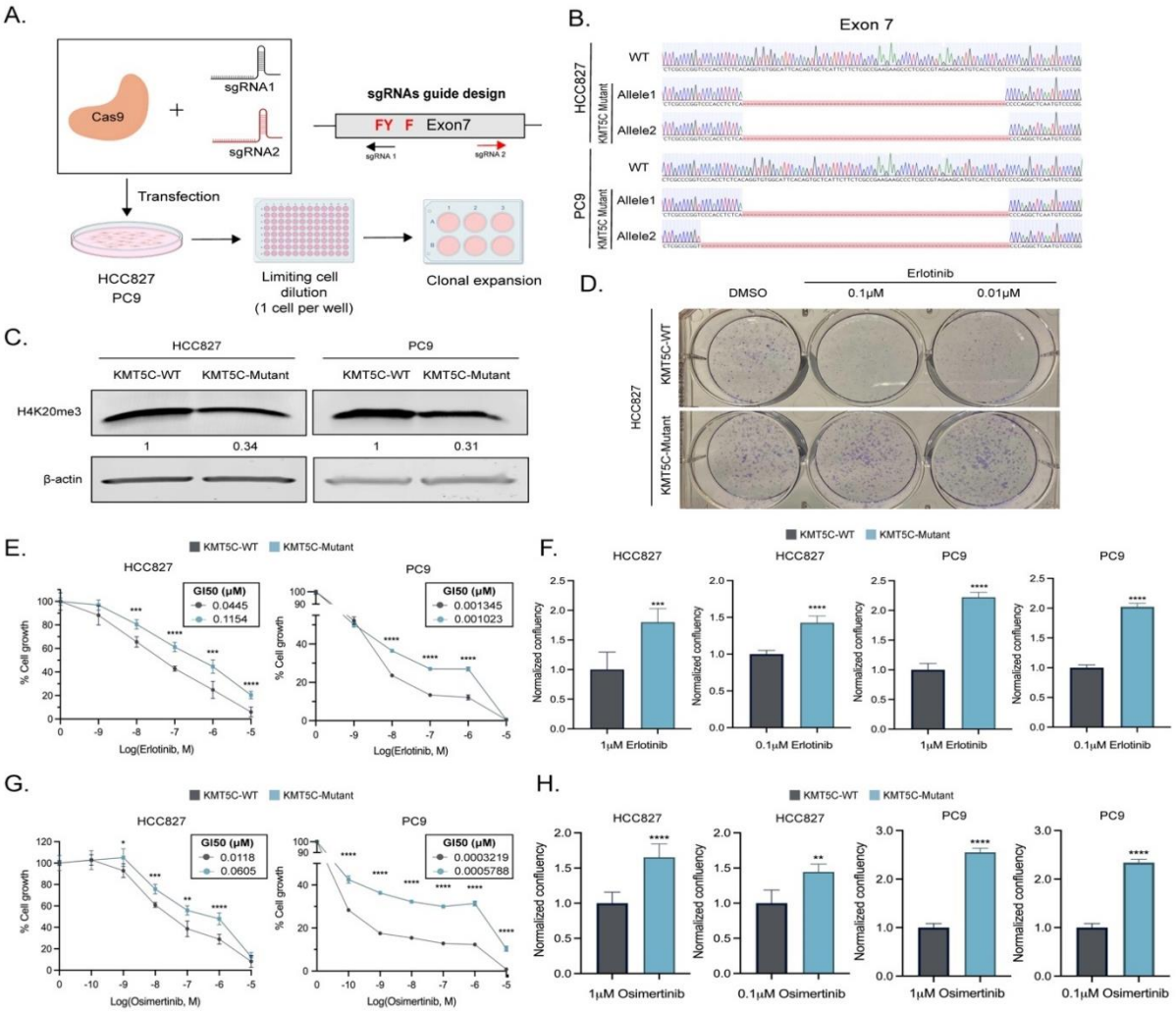

Figure 2.10. Loss of KMT5C confers resistance to erlotinib and osimertinib in EGFR mutant cell lines. A) CRISPR Cas9 strategy to generate KMT5C SET domain mutants. SET domain active site residues are in red. **B)** Alignment of exon 7 sequence in WT and mutant clones using benchling (Sequence Alignment Tool, 2021) retrieved from https://benchling.com. **C)** Representative Western blot analysis of H4K20me3 from WT and mutant HCC827 and PC9 clones. b-ACTIN served as a loading control. **D)** Clonogenic assay in HCC827 KMT5C mutant and WT cells in the presence of 0.1 or 0.01 mmol/L erlotinib containing media for 8 days. E and G, Erlotinib **(E)** or osimertinib **(G)** dose–response curves following exposing the indicated cells to varying concentrations of erlotinib containing media for 72hours. **F and H)** Cell confluency of KMT5C mutant cells was compared with KMT5C WT cells in the presence of 1 or 0.1 mmol/L **(F)** erlotinib or **(H)** osimertinib for 72 hours. Data relative to respective normalized DMSO control treatments are represented.Welch t test was used to evaluate statistical significance. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Data not published: We also evaluated the response of KMT5C mutant clones in the EGFR mutant NSCLC cell line HCC827 to other third generation EGFRi almonertinib and second generation EGFRi dacomitinib. Similar to previous findings, HCC827 became resistant to both EGFRi (**Figure 2.8A-B**) after KMT5C was mutated. Conversely, the HCC827 cell line with KMT5C mutated was unaffected in the presence of ALK inhibitors crizotinib, alectinib and lorlatinib, and VEGFR inhibitor sorafenib **(Figures 2.8C-E)**. These findings and previous findings in EGFR wildtype cell lines suggests that loss of KMT5C may be specific to mediating resistance specifically to EGFRi and not to other targeted agents. '

To complement the genetic studies, the EGFR mutant cell line HCC827 was exposed to A-196, a chemical inhibitor of KMT5B and KMT5C. Treatment with A-196 resulted in a dose- and time-dependent reduction in H4K20me3 (**Figure 2.11A-C, Figure 2.12**) that caused resistance to both erlotinib and osimertinib (**Figure 2.11D**). Collectively, genetic and chemical inhibition of KMT5C provides an advantage to both EGFR wildtype and mutant cells exposed to the EGFRi.
Data not published: Moreover, we evaluated the response of the KMT5C mutated clone derived from EGFR mutant cell line HCC827 to erlotinib resistance, when cells were treated with A-196. Treatment with A-196 in KMT5C wildtype but not KMT5C mutated clone resulted in increased resistance to erlotinib **(Figure 2.13A-C)**. This finding suggests that inhibition of KMT5C and not KMT5B is responsible for erlotinib resistance in HCC827 cells.

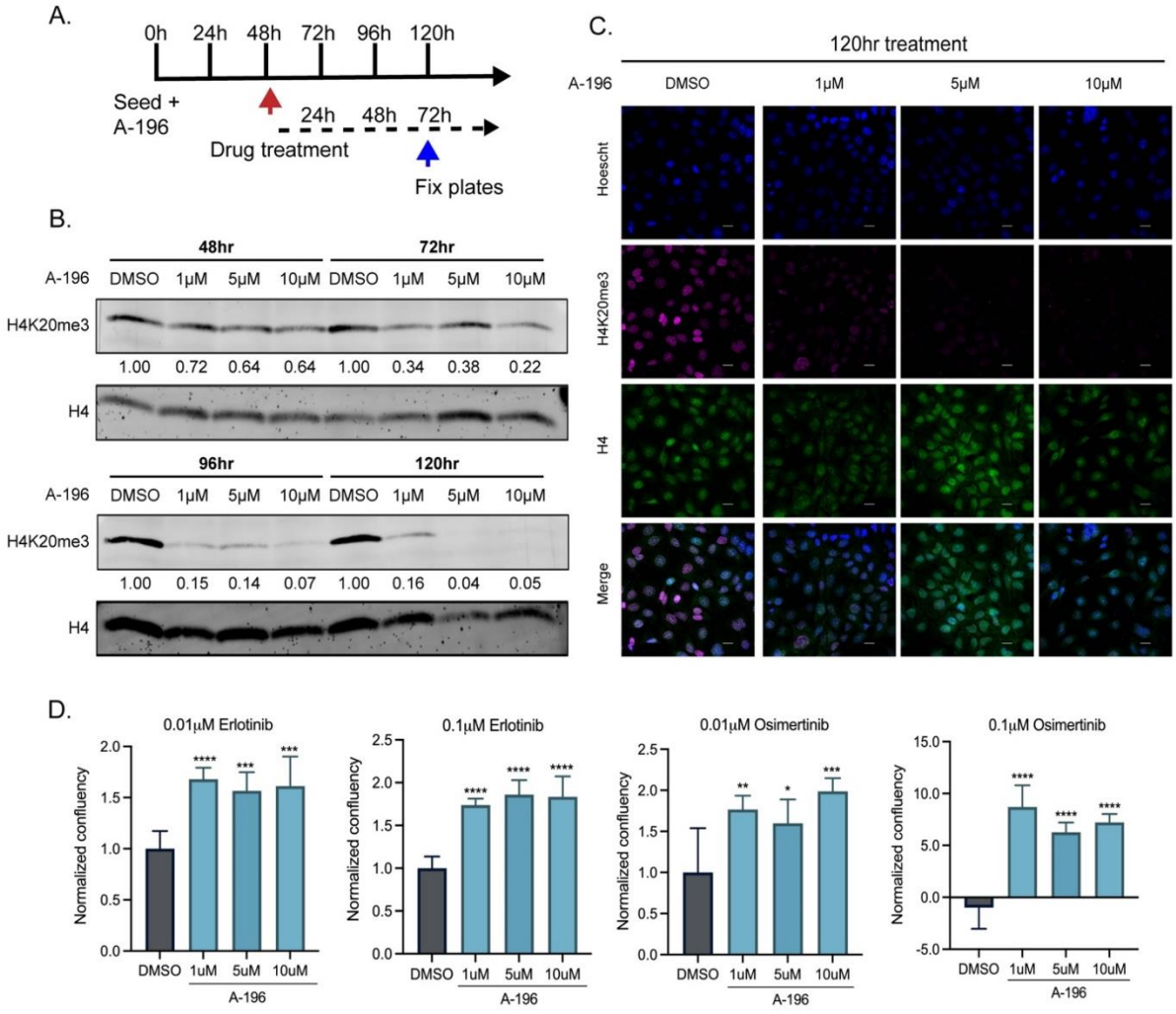

Figure 2.11. Chemical inhibition of KMT5B/C increases erlotinib and osimertinib resistance in HCC827 cells line. A) Experimental timeline. HCC827 cells were treated with the KMT5B/C inhibitor (A-196), 48 hours later erlotinib or osimertinib was added, and cells were fixed 72 hours later for analysis. **B)** Western blot analysis of H4K20me3 in HCC827 cells at different time points, after treatment with A-196. H4 was used as a loading control. **C)** Immunofluorescence of H4K20me3 and H4 in HCC827 cells after treatment with A-196 for 120 hours. **D)** Confluency of HCC827 cells treated with A-196 in the presence of erlotinib/osimertinib for 72 hours. Welch t test was used to evaluate statistical significance. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

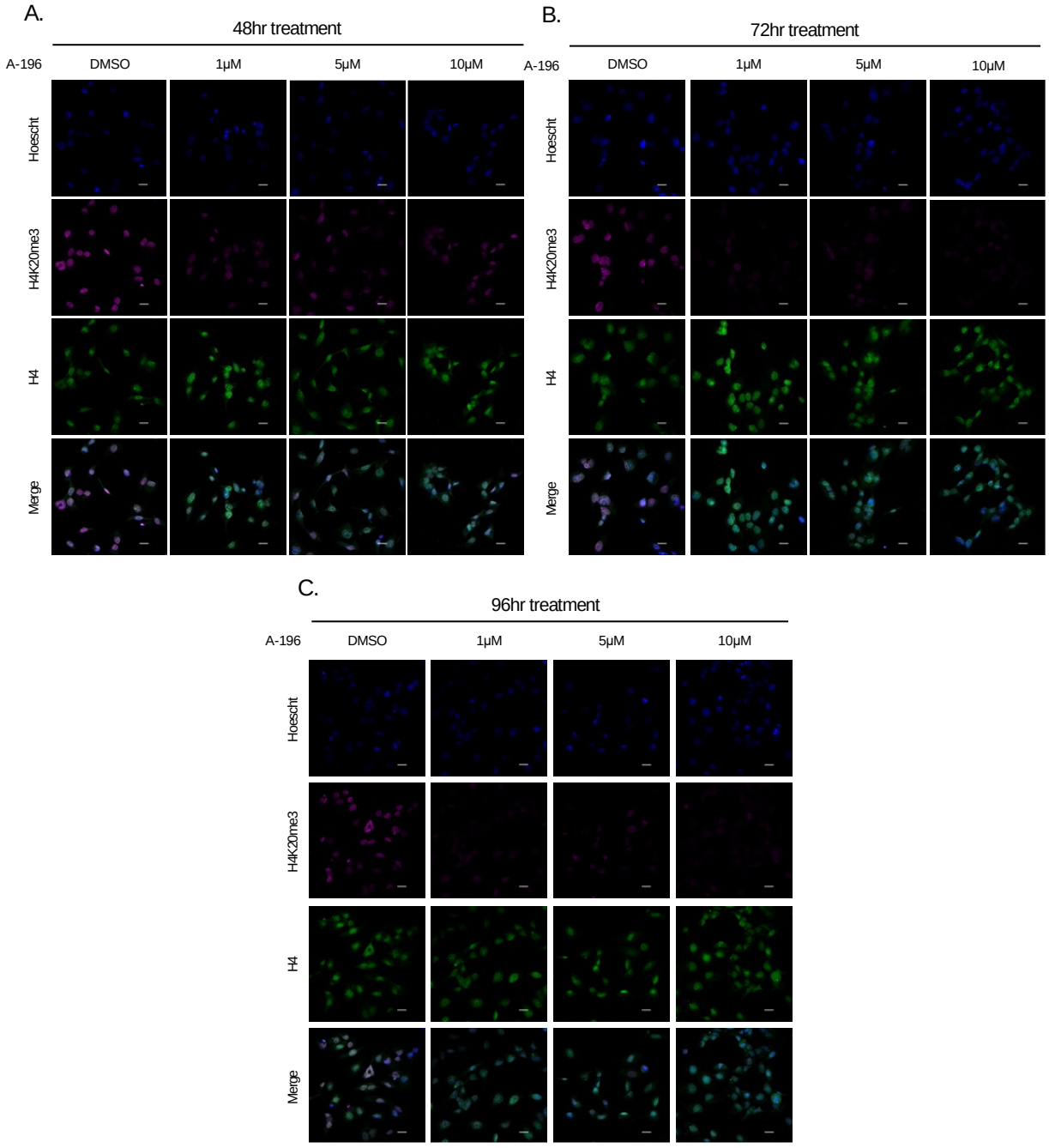

Figure 2.12. Chemical inhibition of KMT5B/C induces global decrease in H4K20me3 with little to no effect on overall H4 levels. Immunofluorescence of H4K20me3 (cyan) and H4 (green) in HCC827 cells after treatment with the indicated doses of A-196 for A) 48h, B) 72h and C) 96h. Hoechst was used as a nuclear stain. Data for 120h timepoint is included in Figure 2.12.

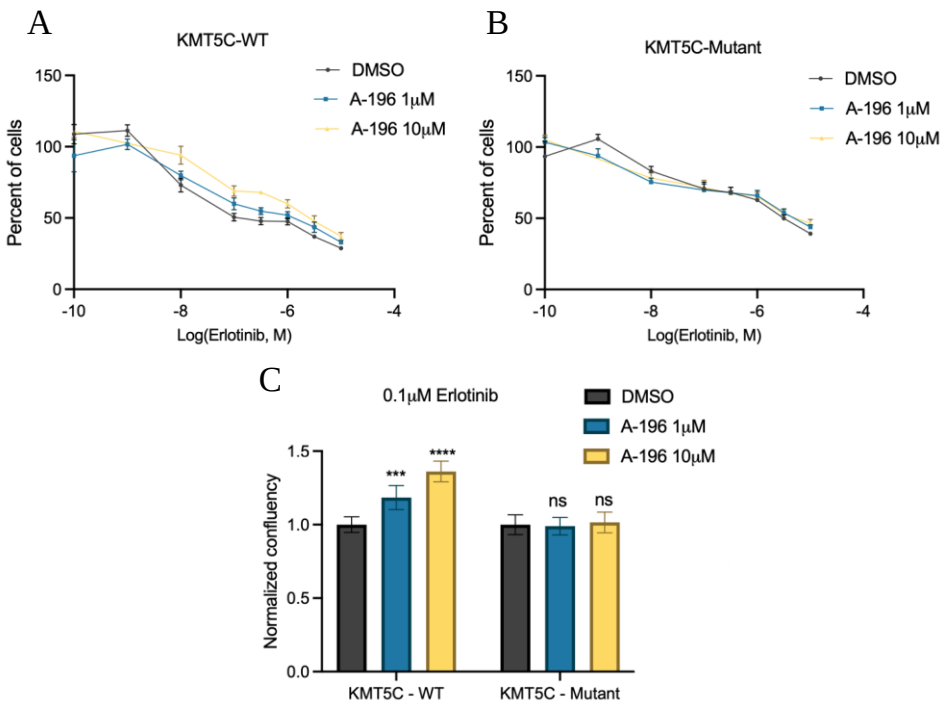

Figure 2.13. Chemical inhibition of KMT5B/C increases erlotinib resistance in KMT5C-WT but not in KMT5C-Mutant cells and has no effect in osimertinib resistant H1975 cell line. Confluency of HCC827 KMT5C-WT **(A)** and HCC827 KMT5C-mutant **(B)** cell lines treated with A-196 in the exposure to increasing doses of erlotinib for 72h. Cells were exposed to varying concentrations of erlotinib or the highest equivalent volume of dimethyl sulfoxide (DMSO, negative control) containing media for 72 hours. Erlotinib/Osimertinib dose response was evaluated using the SRB assay. **C)** Confluency of HCC827 KMT5C-WT or KMT5C-Mutant cells treated with A-196 in the presence of erlotinib for 72 hours. Welch t test was used to evaluate statistical significance., *P < 0.05;**, P < 0.01;***, P < 0.001;****, P < 0.0001.

2.4.4 <u>Data not published:</u> KMT5C reduction could be partially explained by miR-4435 downregulation.

<u>Data not published:</u> As previously mentioned, multiple studies have reported reduction of KMT5C transcript in tumor samples vs healthy adjacent tissues[27,83,86,88,131]. In a recent publication in NSCLC, the authors performed RNA sequencing in tumors of patients before and after being treated with the EGFRi osimertinib[4]. Further analysis of this data, lead to the finding that KMT5C transcript levels was indeed downregulated in tumors after patients were treated with osimertinib **(Figure 2.14).** This further confirms the fact that KMT5C might be acting as a tumor suppressor and is involved in resistance. However, what leads to downregulation of KMT5C transcript in the first place is not known. Previous data from our laboratory elucidated that miRNAs can also

increase resistance to erlotinib. More specifically, miR-4435 was found to be among the top hits of a miRNA overexpression screen[115], and as observed in **Figure 2.15A**, miR-4435 overexpression in EKVX sensitive cells can drive erlotinib resistance. Additionally, according to miRNA target prediction software, miR-4435 is predicted to target the KMT5C transcript via non-canonical mechanism in the 5′ UTR (RNA22 & miRSearch) and via canonical mechanism at the 3′ UTR (miRWalk). Hence, we overexpressed miR-4435 in EKVX sensitive cell line and KMT5C transcript was shown to be downregulated **(Figure 2.15B).** We then evaluated miR-4435 ability to target the 5' UTR **(Figure 2.15C-D)** or 3' UTR (data not shown) in EKVX cell lines using luciferase reporter assays. miR-4435 lead to reduction in Firefly luciferase signal when testing the 5' UTR of KMT5C, however miR-4435 was not found to target the 3' UTR of KMT5C (data not shown). Because miR-4435 overexpression modestly drove erlotinib resistance, and miR-4435 potentially targets the 5' UTR of KMT5C **(Figure 2.15E)**, we can hypothesize that miR-4435 regulation may represent only one of many mechanisms regulating KMT5C.

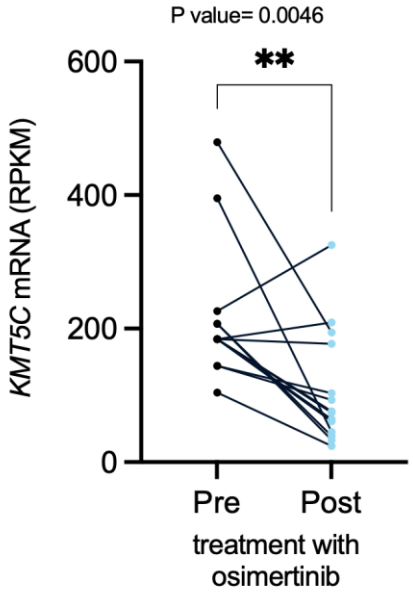

Figure 2.14. KMT5C transcript levels are downregulated in tumors post-treatment with osimertinib in NSCLC. KMT5C mRNA reads per million in resistant tumors derived from patients pre and post treatment with Osimertinib. Two-tailed paired t-test. **, $P < 0.005$**.**

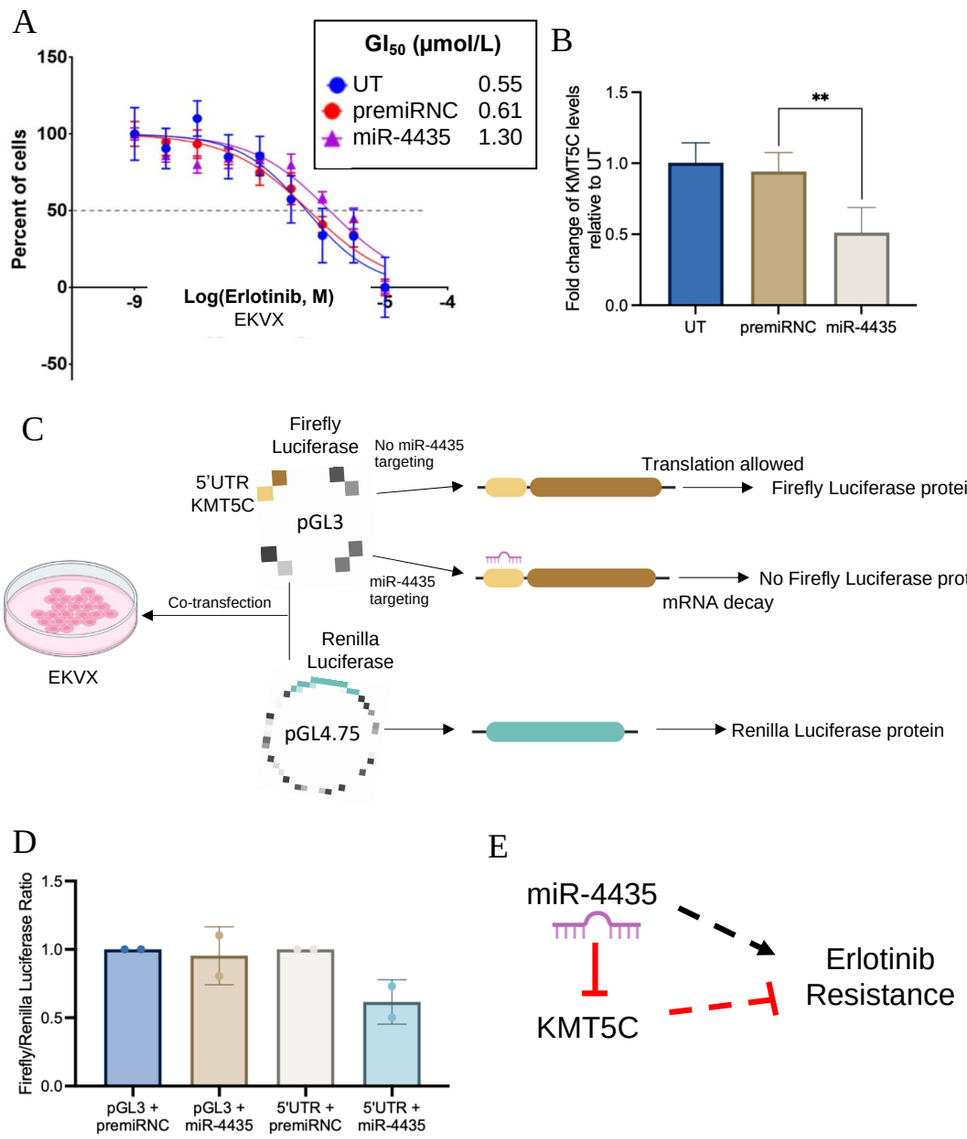

Figure 2.15. miR-4435 is predicted to target the 5' UTR of KMT5C thereby contributing to erlotinib resistance A) EKVX parental cells were reverse transfected with 100nM premiR negative control (premiRNC) or miR-4435 or were untransfected (UT). Erlotinib dose response via SRB assay was evaluated by exposing cells to varying concentrations of erlotinib or the highest equivalent volume of DMSO (negative control) containing media for 72 hours. For percent of cells calculation, number of cells at the time of addition of erlotinib or DMSO was first corrected for, followed by normalization of cell number to respective corrected DMSO values. **B)** Expression of KMT5C in EKVX cells transfected with pmiRNC or miR-4435. Data is normalized to GAPDH and relative to untransfected (UT). **C)** Experimental procedure diagram of Dual Glo Luciferase assay: co-transfection of pGL3 with the 5' UTR of KMT5C cloned upstream of the Firefly Luciferase gene and pGL4.75 containing Renilla Luciferase gene to evaluate transfection efficiency in EKVX cells. **D)** Dual-Glo Luciferase assay results after co-transfection of vectors pGL3 parent/pGL3+5' UTR KMT5C with pGL4.75 and premiRNC/miR-4435. Firefly Luciferase signal was normalized to Renilla Luciferase signal relative to negative control pGL3 with premiRNC (n=2). **E)** Hypothesis model: miR-4435 downregulated KMT5C via non-canonical 5' UTR targeting thereby promoting erlotinib resistance.

2.4.5 Ectopic expression of KMT5C partially sensitizes EGFRi resistant cells.

Since loss of KMT5C led to erlotinib resistance, we evaluated if the converse holds true by overexpressing KMT5C. A doxycycline (DOX) inducible *KMT5C* plasmid was stably expressed in Calu6 cells, which have low levels of *KMT5C* (**Figure 2.3A**) and are resistant to erlotinib (**Figure 2.3C**). Culturing the two clonally-derived lines in the presence of DOX resulted in a 4 to 8-fold increase of *KMT5C* relative to cells grown in PBS containing media (**Figure 2.16A**). H4K20me3 levels were also significantly increased following DOX induction in both clones, but not in Calu6 parental cells (**Figure 2.16B**). Exposure of clones to increasing concentrations of erlotinib resulted in ~2-fold increase in GI50 values for clones cultured in DOX (**Figure 2.16C**). Live-cell proliferation analysis of clone 2 in the presence of three different concentrations of erlotinib validated these findings (**Figure 2.16D**). With respect to gefitinib, afatinib and osimertinib, KMT5C overexpressing clones were sensitized (**Figure 2.16D-F**), most notably at higher concentrations of each drug.

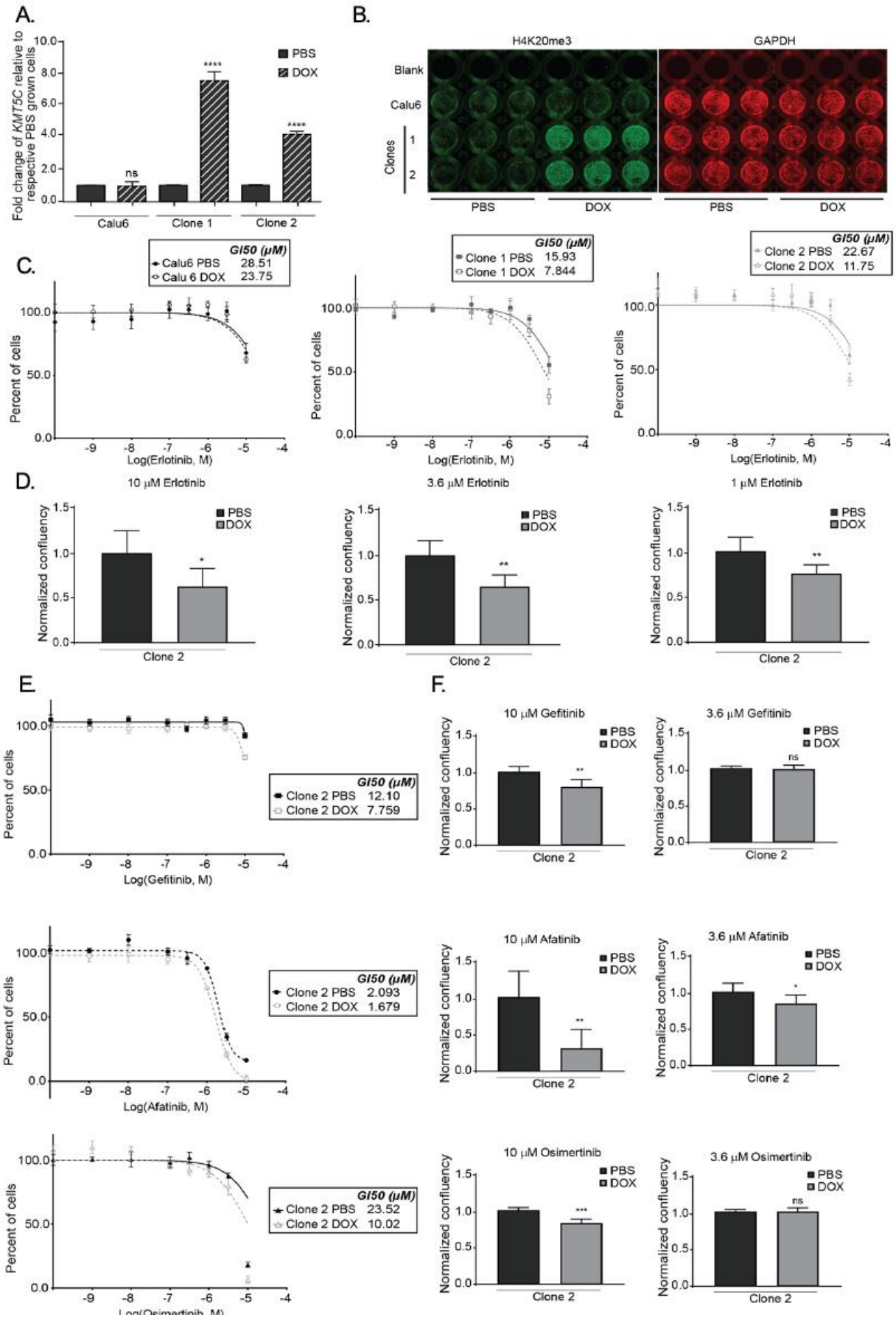

Figure 2.16. Ectopic expression of KMT5C partially sensitizes EGFRi resistant cells to EGFRi. A) KMT5C transcript levels evaluated by qRT-PCR in Calu6 cells and Calu6 clones 1, 2 stably expressing DOX-inducible KMT5C. One-way ANOVA followed by Dunnett's Multiple Comparison test was used to evaluate statistical significance of KMT5C transcript levels relative to respective PBS treated cells. **B)** H4K20me3 levels evaluated by in-cell western. DOX (or PBS control) treatment was for two weeks. GAPDH serves as an endogenous control. **C)** Erlotinib dose response measured by SRB was evaluated after a two-week exposure to PBS or DOX containing media. Cells were then exposed to varying concentrations of erlotinib or the highest equivalent volume of DMSO containing media for 72 hours following normalization, the GI50 concentration of erlotinib was calculated from the respective dose curve for each cell line. **D)** Proliferation of clone 2 was evaluated using the Incucyte. Cells grown in PBS or DOX containing media for two weeks were exposed to varying concentrations of erlotinib or the highest equivalent volume of DMSO containing media for 72 hours. Normalized data relative to respective normalized PBS treated samples is represented. Unpaired t-test was used to evaluate the statistical significance for each pair. **E)** Dose response measured by SRB was evaluated after a two-week exposure to PBS or DOX containing media for Calu6 or clones 1, 2 for gefitinib, afatininb or osimertinib. EGFRi treatments lasted for 72 hours. Following normalization, the GI50 concentration of each EGFRi was calculated from the respective dose curve for each cell line. **F)** Proliferation of clone 2 was evaluated using the Incucyte. Cells grown in PBS or DOX containing media for two weeks, were exposed to varying concentrations of gefitinib, afatinib, osimertinib, or the highest equivalent volume of DMSO containing media for 72 hours. Unpaired t-test was used to evaluate statistical significance of normalized confluency of DOX-cultured clone 2 cells in the presence of either 10 or 3.6 μM of gefitinib, afatinib, or osimertinib compared to respective normalized confluency of PBS-treated cells.

2.4.6 KMT5C negatively regulates the oncogenic long non-coding RNA, LINC01510, and the oncogene, MET

Because KMT5C functions as a tumor suppressor, and is associated with repression of oncogenes [88,134], GEPIA analysis was used to determine if any of the common bypass tracks involved in erlotinib resistance were negatively correlated with *KMT5C*.

Data not published: A significant negative correlation was identified between MKK3, MET and KMT5C in LUAD (Spearman r = -0.34, p-value = 0.00), (Spearman r = -0.44, p-value = $1.0e^{-37}$**, Figure 2.18A)** MKK3, is a gene involved in pro-growth Map Kinase signaling, and is reported to be upregulated in gefitinib resistant cell lines[135] and to drive resistance to DNA damaging agents in NSCLC cell lines[136,137]. Hence, we evaluated the levels of MKK3 in the HCC827 EGFR mutant cell line when KMT5C was mutated. Indeed, after loss of KMT5C, MKK3 levels were increased **(Figure 2.17B)**. Similarly, MKK3 levels were sustained after HCC827 cells with KMT5C mutation were treated with osimertinib, contrary to KMT5C wildtype cells **(Figure 2.17C).** This suggests that loss of KMT5C might contribute to sustained expression of oncogenes such as MET and MKK3 in the presence of EGFRi such as osimertinib. Whether this result is dependent on transcriptional regulation of MKK3 by decreased H4K20me3 levels at this locus when KMT5C is lost, will need to be further investigated.

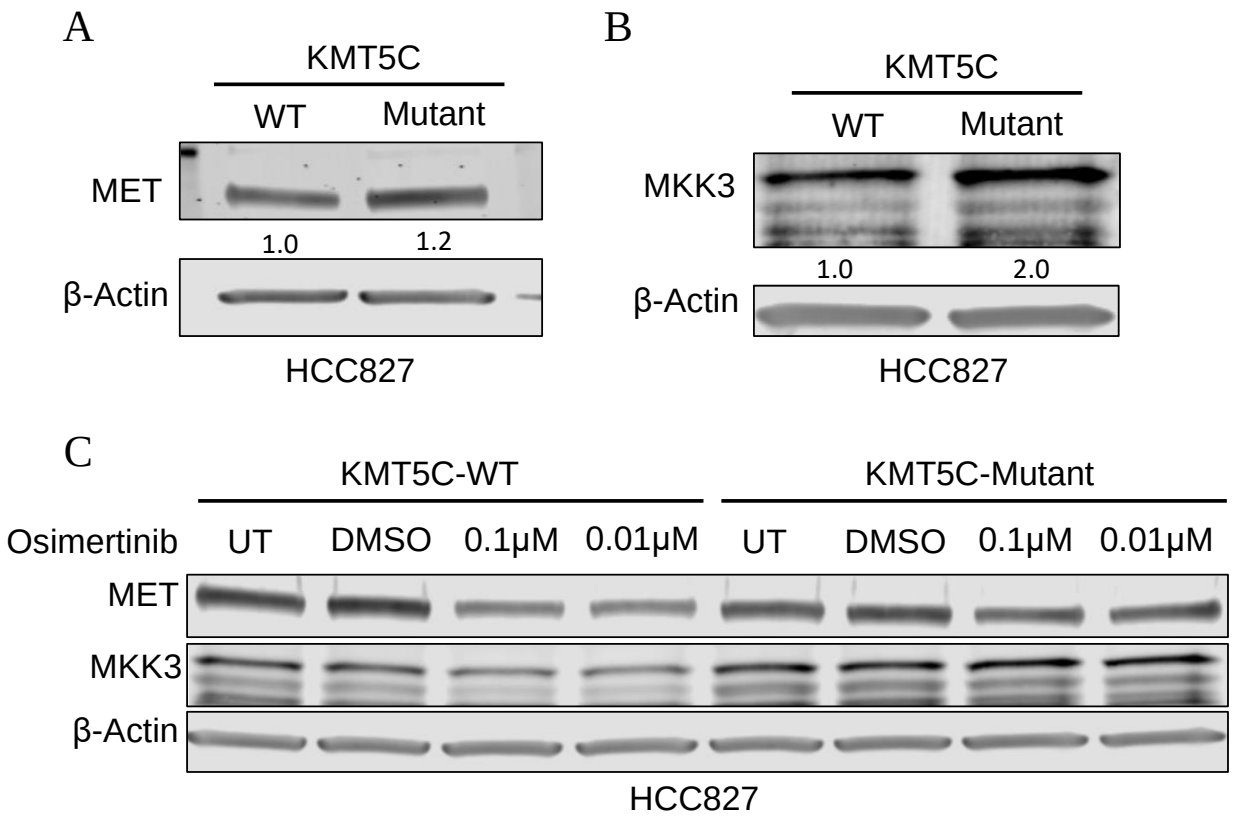

Figure 2.17. Loss of KMT5C induces MET and MKK3 overexpression in EGFR-mutant cell line HCC827. A) B) Representative Western blot analysis of MET and MKK3 in HCC827 cell lines with either KMT5C-WT or KMT5C-Mutant. β-Actin was used as a loading control. **C)** Western Blot analysis of HCC827 KMT5C-WT or HCC827 KMT5C-Mutant cells in the presence of osimertinib at the indicated concentrations for 72h. β-Actin was used as a loading control. UT: Untreated.

Moreover, MET amplification is one of the more common bypass mechanisms that cells use to overcome inhibition of EGFR signaling by erlotinib [5,138]. As expected, *MET* was higher in LUAD relative to normal tissues (**Figure 2.18B**). To determine if the negative correlation between *MET* and *KMT5C* held true in the NSCLC cell lines, KMT5C mutant cells were evaluated for MET.

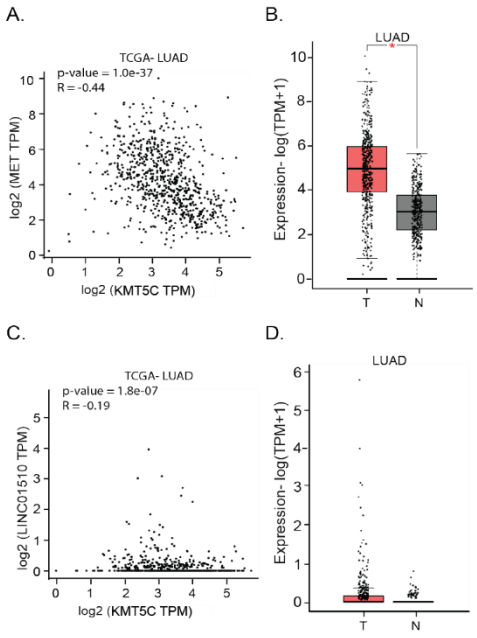

Figure 2.18. LINC01510 correlates poorly with LUAD prognosis. Correlation analysis between **A)** MET and KMT5C and **C)** LINC01510 and KMT5C transcript levels in TCGALUAD dataset, evaluated using GEPIA. GEPIA analysis for **B)** MET and **D)** LINC01510 transcript levels in normal (N, n= 347) and tumor samples (T, n = 483) form LUAD data obtained from TCGA and GTEx databases. The majority of the samples in the normal subgroup had undetectable levels of LINC01510. TPM= Transcripts per million.

Data not published: Indeed, following loss of KMT5C, MET protein was increased relative to levels in KMT5C wildtype cells in EGFR WT (**Figure 2.19Ai**) and EGFR Mutant cells **(Figure 2.17A)**. Moreover, similar to MKK3, MET levels are sustained after EGFR Mutant cells with KMT5C mutations are treated with osimertinib **(Figure 2.17C).**

Additionally, the *MET* transcript was elevated in the mutant cells, suggesting that loss of KMT5C enhanced MET via a transcriptional mechanism (**Figure 2.19Bi**). Conversely, induction of KMT5C in dox-inducible clones resulted in reductions in both MET RNA and protein (**Figure 2.19Aii, Bii**).

Previous studies determined that MET can be induced through both genomic amplification and transcriptional upregulation [138–140]. Although there are multiple mechanisms that are involved in regulating transcription from the *MET* locus, a recent study identified a long non-coding RNA (lncRNA) that functions as a positive regulator of *MET* [141]. A short variant of the long non-coding RNA, LINC01510, referred to as COMET (Correlated-to-MET) was also identified to positively

regulate *MET* transcription in papillary thyroid carcinomas [142]. Similar to *MET*, high *LINC01510* correlates with poor prognosis in various cancers, including NSCLC [141,143,144]. Further evaluation of *LINC01510* transcript levels in NSCLC via GEPIA analysis in LUAD (TCGA and GTEx) indicated that *LINC01510* was higher in a subset of tumors relative to normal tissues (**Figure 2.18D**).

Since *LINC01510* and *MET* levels are reported to positively correlate in colorectal cancer [141], their correlation was evaluated in NSCLC. Correlation analyses using TCGA LUAD and LUSC datasets via GEPIA highlight a positive correlation in both LUAD (Spearman r = 0.38, p-value = $1.6e^{-27}$) and LUSC (Spearman r = 0.25, p-value = $1.1e^{-12}$) (**Figure 2.19C**). Based on the reported and evaluated positive correlation between *MET* and *LINC01510*, and the negative correlation between KMT5C and MET, we hypothesized that *KMT5C* would also negatively correlate with *LINC01510*. The correlation analysis between *KMT5C* and *LINC01510* suggests a significant, modest negative correlation in LUAD tissues (Spearman r = -0.19, p-value = $1.8e^{-7}$, **Figure 2.18C).** As also hypothesized, in KMT5C mutant clones *LINC01510* was significantly upregulated between 8 and 10-fold (**Figure 2.19Di**). Conversely, in the KMT5C inducible clones, *LINC01510* was significantly lower when cells were cultured in the presence of DOX (**Figure 2.19Dii**).

KMT5C mediates its repressive effects via the H4K20me3 modification [88], hence we hypothesized that *MET* and/or *LINC01510*, are likely negatively regulated by KMT5C via H4K20me3 mediated repression. To this end, we analyzed the reported ChIP-seq profile of H4K20me3 obtained from a human lung fibroblast cell line, IMR90 (GSE59316) [134]. Surprisingly the H4K20me3 modification in this dataset was not present within or near the *MET* locus but instead was localized in the gene body of *LINC01510*, i.e. ~55kb upstream of the start site of its neighboring gene, *MET* (**Figure 2.19E**). To identify the region of the chromosome associated with the H4K20me3 modification in the erlotinib sensitive cells, chromatin immunoprecipitation followed by q-RT-PCR (ChIP-qPCR) was conducted. Sensitivity of the assay was first established using primers designed to pulldown the *FOXA1* locus, a target previously reported to be regulated by KMT5C [95]. Two primer sets were tested, one based on the original publication [95], and another that overlaps with the predicted H4K20me3 mark (**Figure 2.20A and Table 2.3**). As expected, pulldown of the FOXA1 region using an antibody to H4K20me3 was depended on the presence of KMT5C. A significant reduction

in pulldown was observed in the KMT5C mutant clones (**Figure 2.20B**) and an increase in pulldown was evident when KMT5C was induced (**Figure 2.20C**).

Following the results obtained from ChIP-qPCR for *FOXA1*, ChIP-qPCR analysis at the *LINC01510* and *MET* loci was conducted using primers that overlapped with the predicted H4K20me3 site and with primers both up and downstream of the predicted site (**Figure 2.19E**, **Table 2.3**). Similar to the *FOXA1* locus, pulldown varied depending on the status of KMT5C. The most abundant reduction in pulldown in the KMT5C mutant occurred just upstream of the *LINC01510* locus with no obvious difference at the *MET* locus (**Figure 2.19F** compare upstream primer 1 (U1) to MET primers). In concordance, induction of KMT5C followed by ChIP-qPCR resulted in enrichment of the H4K20me3 mark in regions surrounding the lncRNA, with the LINC01510 mark having the most significant increase, with only a marginal increase at the *MET* locus (**Figure 2.19G**, compare LINC01510 primers to MET primers). Importantly the observed enrichment in ChIP-qRT-PCR for the LINC01510 mark is similar to the enrichment observed at the *FOXA1* locus. These results further support the hypothesis that KMT5C regulates *LINC01510* expression via the H4K20me3 modification present within its gene body.

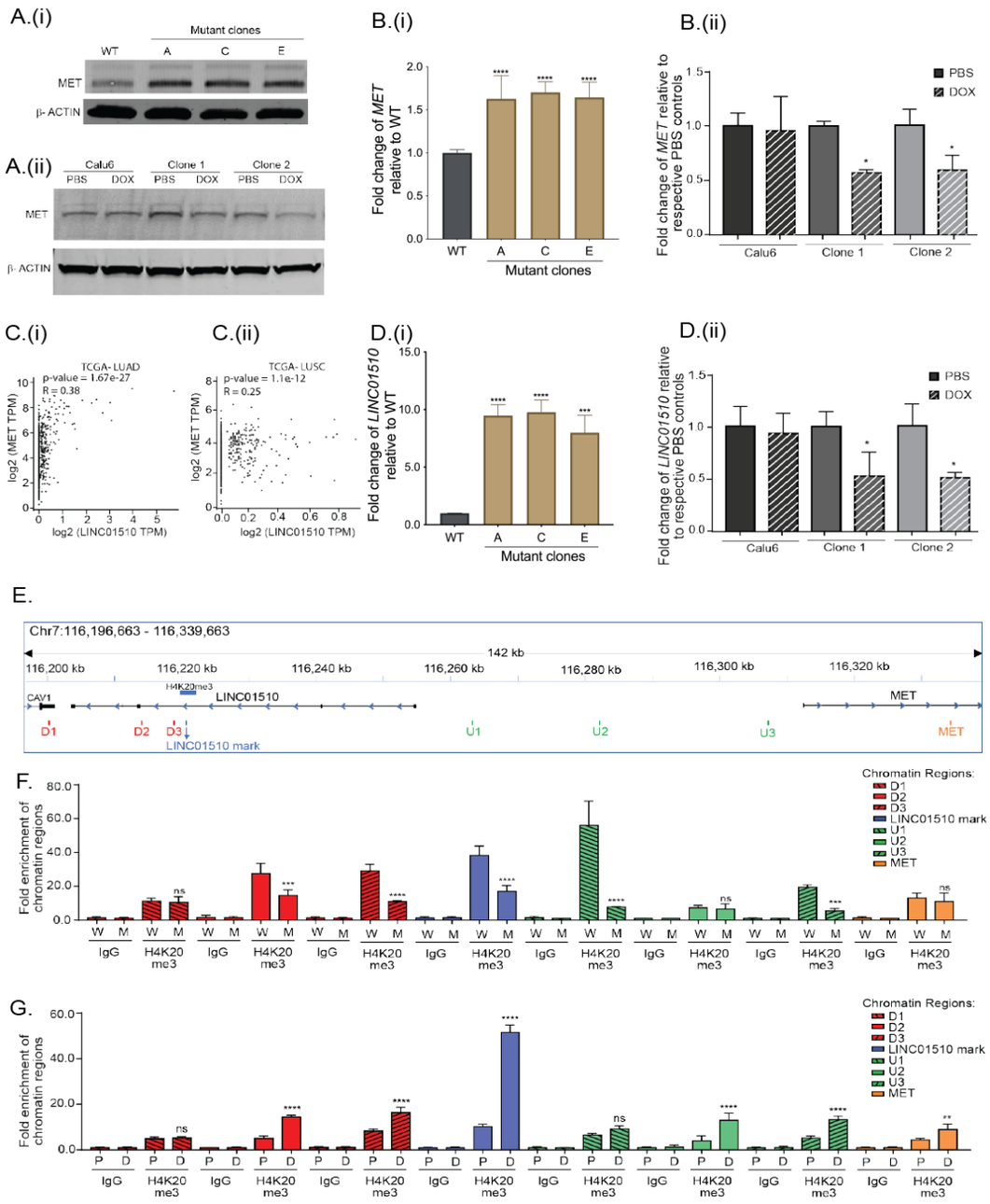

Figure 2.19. KMT5C represses LINC01510 and MET via H4K20me3. **A)** Representative Western blot analysis of MET in (i) EKVX KMT5C WT cells and mutant clones, and (ii) Calu6 cells and clones stably expressing a DOX-inducible KMT5C vector. **B)** qRT-PCR data for MET in (i) WT cells and KMT5C mutant clones, or (ii) Calu6 cells and clones stably expressing a DOX-inducible KMT5C vector. **C)** Correlation analysis between LINC01510 and MET transcripts obtained from (i) LUAD and (ii) LUSC datasets, evaluated using GEPIA. **D)** Expression of LINC01510 in (i) KMT5C mutant lines, or in (ii) KMT5C-inducible clones. **E)** Diagram of the genomic region representing the predicted H4K20me3 modification on the LINC01510 gene body, upstream of MET, as identified from GSE59316. ChIP-qPCR primers designed on and around the H4K20me3 mark are indicated as LINC01510 mark, regions downstream (D1, D2, D3) and upstream (U1, U2, U3) of the H4K20me3 mark, and on MET. **F and G)** ChIP was performed on chromatin isolated from WT (W) or KMT5C mutant clone C (M; F), DOX-inducible KMT5C cells following growth in DOX (D, induced) or PBS (P, uninduced; G). qPCR using the immunoprecipitated chromatin was conducted using primers depicted in E. Data are represented as fold enrichment of the chromatin region pulled down by H4K20me3 primary antibody relative to IgG. Statistical significance is represented for fold enrichment of chromatin regions in KMT5C mutant clone C relative to WT, or DOX relative to PBS. For panels showing statistical significance, one-way ANOVA followed by Dunnett multiple comparison test was used. ns, nonsignificant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. TPM, transcripts per million.

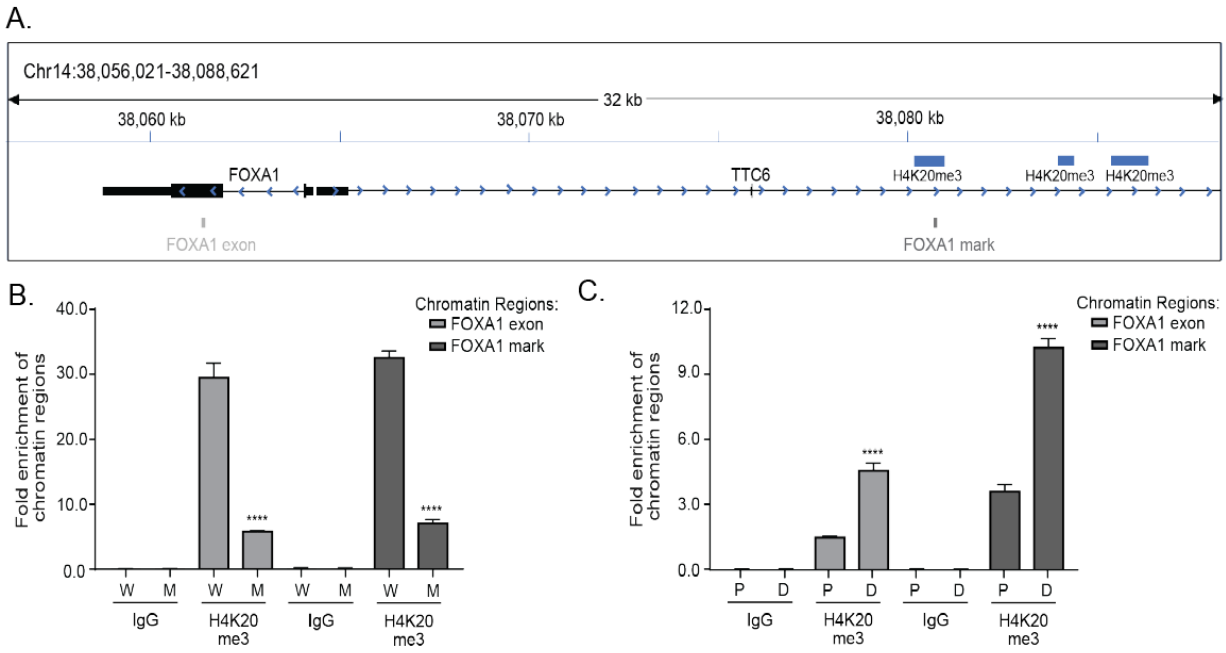

Figure 2.20. **H4K20me3 is enriched at the FOXA1 locus in an KMT5C dependent manner. A)** ChIP-qPCR primers designed to evaluate enrichment of H4K20me3 at the FOXA1 exonic region (FOXA1 exon), and at the predicted H4K20me3 modification upstream of the FOXA1 promoter region (FOXA1 mark). ChIP was performed using either IgG or H4K20me3 primary antibodies on chromatin isolated from **B)** WT or KMT5C mutant clone C or **C)** inducible KMT5C cells (in the presence of DOX or PBS). qPCR using the immunoprecipitated chromatin was conducted using primers shown in A (Table 3). Data are represented as fold enrichment of the chromatin region pulled-down by the H4K20me3 primary antibody relative to IgG and was evaluated for significance using one-way ANOVA. W = WT cells, M = KMT5C mutant clone C cells, P = Calu6 clones grown in PBS containing media, D = Calu6 clones grown in DOX containing media.

Table 2.3. Primers utilized in the study. Designed and purchased from Integrated DNA Technologies.

Primer use		**Primer direction**	**Primer sequence**
pLV-sgKMT5C		Forward	CACCGCGGCCCGCTACTTCCAGAGC
		Reverse	AAACGCTCTGGAAGTAGCGGGCCGC
pLVX-Tetone-KMT5C		Forward	TCGTAAAGAATTCACCATGGGGCCCGACAGAGTGACAGCA
		Reverse	GAGATCTGGATCCTCAGTACAGCTCTTCACCGCCGAC
pLVX-Tetone-KMT5C-puro		Forward	CCGCTACGCGTTCAGAAGAACT
		Reverse	AGCGGCGTACGATGATTGAACA
KMT5C genomic locus amplification		Forward	GAGCAGATGGGAGGTGCGGCGACAGT
		Reverse	GAGCTCAGAAGAAAGGAGACAGAT
KMT5C Exon7 locus amplification for T7 endonuclease assay		Forward	CTCAGCTGTTGCCCCATTCCAG
		Reverse	CTTGGTCTCACGCAGCTGGTA
KMT5C genomic locus sequencing		Forward	CCTCTCCTTAGCCTGGTCCT
		Reverse	CAAGGGCTAGGAAGTCAGGG
KMT5C quantification		Forward	TCGGTTTCCGCACCCATAAG
		Reverse	CGGAGGTAGCGATAGACGTG
ChIP-QPCR	FOXA1 mark	Forward	AAGGAGAGGTGCGTTGTTTG
		Reverse	CATTCTCCCACGAAAGGCAG
	FOXA1 exon	Forward	AAGACTCCAGCCTCCTCAAC
		Reverse	CGGGTGGTTGAAGGAGTAGT
	Linc01510 mark	Forward	GCTTCTTGTCCCTCCCAGAT
		Reverse	GCAGAAGTGAGAGGAAGGGT
	Up 1	Forward	CACACTGGAGTTCTTGCCAC
		Reverse	TATGCACTCCTTCACTGGGG
	Up 2	Forward	GCAGTCCAGCTAAGCAATCC
		Reverse	GACATCTTGGGAAGGGGACA
	Up 3	Forward	CCTCTTCACATCCCACAGGT
		Reverse	CTCTGCTGGCTTGATCATTG
	MET	Forward	GATCAAGGAAATGGGGCGTT
		Reverse	GGGACTAGGGCCTATTGTCA
	Down 1	Forward	CCCTGCCTCTCATCAACTGA
		Reverse	GTTGAGCCACTAAACCACCC
	Down 2	Forward	TGCCTGGTCTCCTGTTAACA
		Reverse	ATCTGTCTTCTCCCTGTGCC
	Down 3	Forward	AGTCCAAGATCAAGGCACCA
		Reverse	AGGCCTTTCTTGTACCCCTT

2.4.7 Loss of LINC01510 or MET partially re-sensitizes KMT5C mutant cells to erlotinib, conversely overexpression promotes erlotinib resistance in KMT5C wildtype cells

From Figure 2.19, it can be inferred that KMT5C negatively regulates both *LINC01510* and *MET* transcript levels and MET protein levels. It was also determined that KMT5C negatively regulates *LINC01510* transcription, likely via H4K20me3 within its gene body. Therefore, we evaluated if KMT5C negatively regulates *MET* indirectly via repression of *LINC01510*. *LINC01510* or *MET* were knocked down in an KMT5C mutant clone, which expresses high levels of *LINC01510* and *MET* (**Figures 2.21Ai, 2.21Bi, 2.21Di**). By western blot and qRT-PCR analyses, it was confirmed that siRNAs targeting either *MET* or *LINC01510* downregulate MET at both the transcript and protein level (**Figure 2.21A, B**). To determine if loss of KMT5C partially mediates erlotinib resistance via upregulation of *LINC01510* and *MET*, *LINC01510* or *MET* were downregulated and erlotinib dose response and proliferation analyses were conducted. Both results validate that erlotinib resistant KMT5C mutant cells can be partially re-sensitized to erlotinib post knockdown of either *LINC01510* or *MET* (**Figure 2.21C, D**).

Data presented in Figure 2.21A/B suggests that knockdown of *LINC01510* reduces MET at the transcript level, therefore, we further evaluated if overexpression of *LINC01510* in KMT5C wildtype cells can positively regulate MET. Following transfection of a *LINC01510* or *MET* overexpressing plasmid, a modest, yet significant increase in MET was observed (**Figure 2.21E, F**). Additionally, as hypothesized, *LINC01510* or *MET* overexpression also led to acquired resistance in KMT5C wildtype cells, evaluated by both dose curve and proliferation analyses (**Figure 2.21G, H**).

Overall, the findings of this study, depicted in the model in **Figure 2.21I** suggests that wildtype KMT5C in NSCLC cells negatively regulates *LINC01510* via the downstream modification, H4K20me3. In cells with high KMT5C, repression of *LINC01510* inhibits full expression of MET. However, upon loss of KMT5C in mutant cells, *LINC01510* becomes de-repressed due to reductions in the H4K20me3 modification, resulting in increased expression of *LINC01510*. Simultaneously, *LINC01510* positively regulates the transcription of *MET*. Therefore, increased levels of *LINC01510* and MET function as mediators of erlotinib resistance in KMT5C mutant cells.

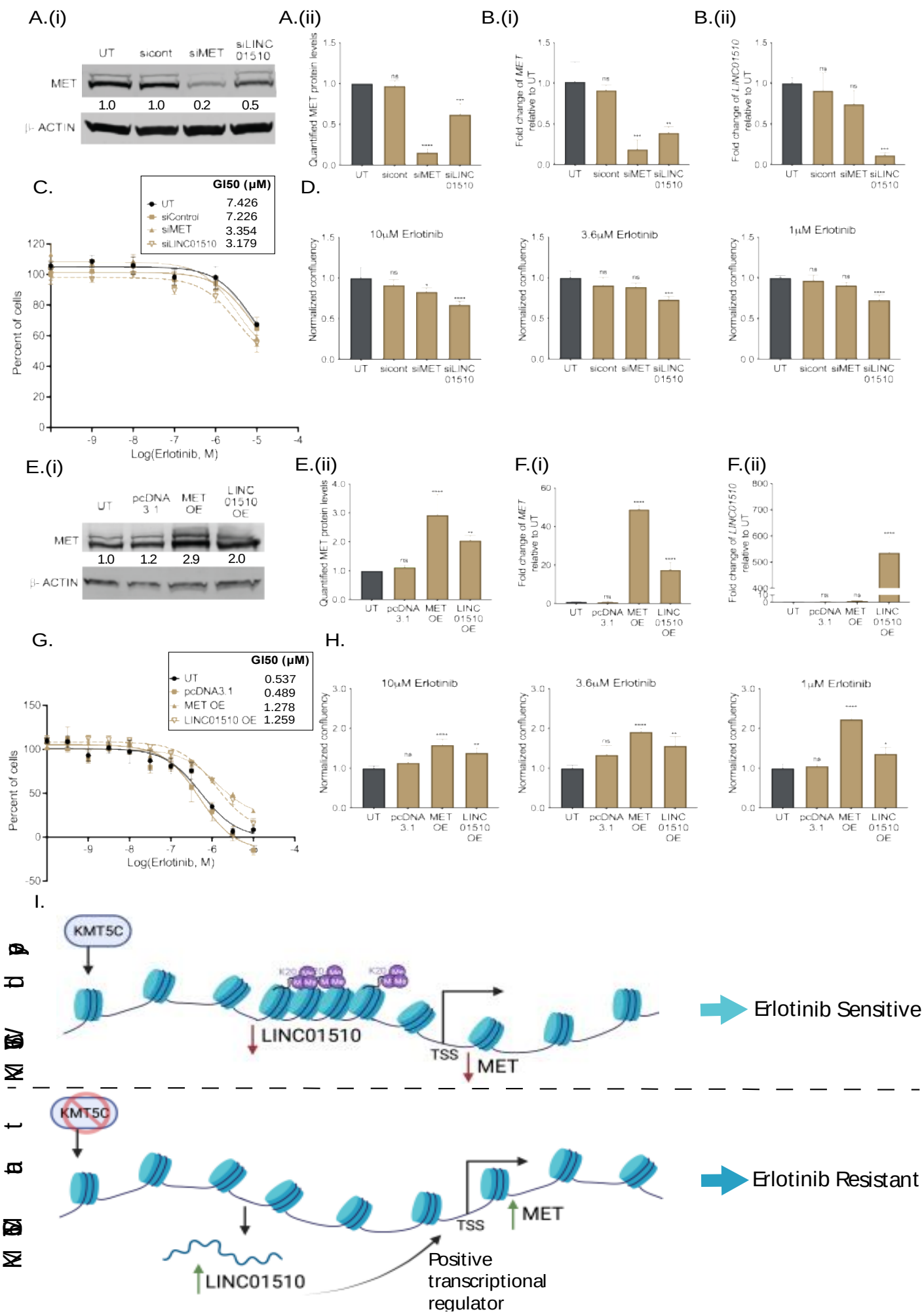

Figure 2.21. **Modulation of LINC01510 or MET is partially responsible for the erlotinib response. A)** (i) Representative Western blot analysis of MET in KMT5C mutant cells that were either untransfected (UT) or reverse transfected with siRNA control (sicont), siRNA to MET (siMET), or siRNA to LINC01510 (siLINC01510) for 96 hours. b-ACTIN served as a loading control. Densitometry values normalized to b-ACTIN and relative to untransfected are indicated. (ii) Quantification of protein levels from three biological replicates as done in Ai. **B)** Expression of (i) MET and (ii) LINC01510 in KMT5C mutant cells that were either untransfected or reverse transfected with sicont, siMET, or siLINC01510 for 96 hours. Data were normalized to GAPDH and are graphed relative to data from untransfected cells. **C)** Erlotinib dose response of KMT5C mutant cells following transfection with the indicted siRNAs. Twenty-four hours after transfection, cells were exposed to varying concentrations of erlotinib or DMSO for 72 hours. Post-normalization, the GI50 concentration of erlotinib was calculated. **D)** Proliferation of KMT5C mutant cells following transfection with the indicated siRNAs. Twenty-four hours after transfection, cells were exposed to erlotinib for 72 hours. Normalized data are represented relative to untransfection. One-way ANOVA followed by Dunnett multiple comparison test was used to evaluate significance. **E)** (i) Representative Western blot analysis of MET in KMT5C WT cells that were untransfected, or transfected with pcDNA3.1 control plasmid or plasmids to overexpress to MET (MET OE) or LINC01510 (LINC01510 OE) for 96 hours. b-ACTIN was used as a loading control. Densitometry values for the representative blots are shown. (ii) Quantification of MET from three biological replicates as in Ei. **F)** Expression of (i) MET and (ii) LINC01510 in KMT5CWT cells that were either untransfected or transfected with the indicated vectors. Data are normalized to GAPDH. **G)** Erlotinib dose response via SRB assay was evaluated in WT cells that were either untransfected or that were transfected with the indicated vectors, as described in C. **H)** Proliferation of WT cells transfected as in G was evaluated as described in D. **I)** Model depicting loss of KMT5C in NSCLC results in development of erlotinib resistance via LINC01510-mediated upregulation of MET. ns, nonsignificant; *, P < 0.05; **, P<0.01; ***, P < 0.001; ****, P < 0.0001.

2.5 Discussion and future directions

Changes to the epigenome are common occurrences that influence all aspects of cancer, including chemoresistance [145]. However, only a limited subset of epigenetic factors have been determined to have a role in resistance to therapeutic drugs in cancer [146]. The aim of this study was to identify unknown mechanisms by which acquired erlotinib resistance manifests in NSCLC in an unbiased way, and loss of KMT5C, an epigenetic factor was the top hit. KMT5C is a histone methyltransferase responsible for maintaining constitutive heterochromatic regions of the genome and for repressing specific genes, via the repressive mark H4K20me3. Both KMT5C and H4K20me3 are significant for maintaining cells in their differentiated states, loss of which is consequentially reported to cause enhanced survival due to elongation of telomeres [81], and spontaneous carcinogenesis [27,112].

Catalysis of H4K20me3 modification of the genome is a sequential process. SUV39H2, another histone methyltransferase first catalyzes the H3K9me3 modification, that recruits HP1 which physically associates with KMT5C to mediate H4K20me3 [32,34]. Although the findings of this study, for the first time identify a role for KMT5C in mediating drug resistance, loss of a key upstream regulator of KMT5C activity, SUV39H1/2 has previously been reported to be associated with resistance [147,148]. SUV39H null mice displayed chromosomal instabilities and increased tumorigenicity [32,34,149,150]. Apart from SUV39H1/2, it is also possible that other upstream regulators of KMT5C such as HP1 may have an unidentified role in mediating resistance to drugs, such as EGFRi [32,34]. Indeed, the first identified demethylase for H4K20me3, mineral dust-induced gene (Mdig) was determined to be overexpressed in breast and lung cancer cells antagonizing the effects of the H4K20me3 modification which led to induction of oncogenes [151]. Analogous to Mdig overexpression in cancer cells, leading to reduction of H4K20me3, we found that loss of KMT5C also leads to depletion of H4K20me3 mark that in turn enhances expression of oncogenes such as MKK3, *LINC01510* and MET.

It has been long appreciated that genomic instability generates tumor heterogeneity and in the presence of a drug gives rise to resistant cells [145,152], also a reported mechanism of EGFRi resistance [153,154]. In the current study, complete loss of KMT5C function may have led to spontaneous genetic aberrations leading to rapid establishment of resistant population of cells in

the presence of erlotinib and other EGFRi. Indeed, previous reports determined that loss of KMT5B/C impairs the DDR mechanism, inadvertently leading to accumulation of damaged DNA and increased tumorigenicity [18,23,32,71,149,155,156]. Therefore, it is possible that in the KMT5C mutant cells, the chromatin may have suffered *massive* loss of H4K20me3, which disrupted the heterochromatic shield protecting the DNA from damage. On the contrary, in Calu6 cells, which still have modest amounts of H4K20me3 (**Figure 2.5**) the regions of the chromatin lacking H4K20me3 could be localized at oncogenes leading to their upregulation, while the constitutive heterochromatic regions remained marked and compact, preventing genomic instability. Indeed, increased H4K20me3 in Calu6 cells due to DOX-induction of KMT5C resulted in reductions in MET **(Figure 2.19)** and promoted sensitivity to EGFRi (**Figure 2.16**) suggesting that even modest changes in H4K20me3, or other unidentified mechanisms of KMT5C can alter the response of cells to EGFRi. Additional studies addressing the dynamics of KMT5C and H4K20me3 and their role in maintaining genomic stability will need to be conducted to support these observations.

While this study defines a role for MET and *LINC01510* upregulation that is mediated by loss of KMT5C in EGFRi resistance, there are likely to be several other oncogenes regulated by KMT5C that contribute to this phenotype. Using the NCI Cell Miner Database that has sequencing data for the NCI-60 cell lines [157], multiple genes previously determined to be involved in NSCLC or in EGFRi resistance were found to be negatively correlated with *KMT5C*. Some of the top genes negatively correlated with *KMT5C* include Annexin A5 (negative correlation , nc = -0.616), Vimentin (nc = -0.636), *CD44* (nc = -0.637), *AKT3* (nc = -0.612), *PRKD1* (nc = -0.632) a member of the PKC family, *NOTCH* (nc = -0.565), *JUN* (nc = -0.359), MKK3 (nc=-0.343) and *ERK* (nc = -0.343) all with p-values <0.01. In this analysis the negative correlation between *MET* and *KMT5C* was -0.337, p-value <0.01. Similar to MET, many of these genes are predicted to contain a H4K20me3 modification as determined using H4K20me3 ChIP from IMR90 (GSE59316) [134] including *AKT, NOTCH, CD44, MKK3, ERK* and others. It is possible that aberrant KMT5C may alter a cohort of genes that could ultimately synergize to promote resistance, similar to the effects observed following aberrant microRNA expression [158,159]. Whether or not these additional candidates are also KMT5C targets and what their contribution is to resistance remains an active area of investigation.

Not included in publication: It has been reported that miRNAs can induce resistance to EGFRi like erlotinib in cancer[115]. One possible mechanism that can explain this phenomenon is the ability of miRNAs to downregulate multiple target genes such as tumor suppressors by destabilizing the mRNA, and thereby preventing translation. In this study we reveal that miR-4435 induces erlotinib resistance by potentially downregulating KMT5C via non-canonical 5' UTR targeting. However, multiple other mechanisms other than KMT5C downregulation can regulate H4K20me3 in oncogenes. For example, the DNA methylase enzyme DNMT1 has shown to interact with H4K20me3 thereby ensuring DNA methylation at H4K20me3 rich sites. However, whether DNMT1 interaction with H4K20me3 regulates the level of trimethylation at specific loci remains to be investigated[55]. Moreover, demethylase enzymes can potentially have an effect in downregulating overall H4K20me3 levels in cells. For instance, the human homologue of RAD23 protein from yeast (hHR23A/B) has shown to be an eraser of H4K20me1/2/3, as well as PHF2, a member of Jumonji domain family of lysine demethylases[44]. Hence, downregulation of KMT5C or H4K20me3 cannot solely be attributed to miR-4435 targeting the 5' UTR of KTM5C transcript, and further regulation mechanisms must be determined.

In conclusion, the results of this study describe that loss of KMT5C confers EGFRi resistance in NSCLC cells via a novel mechanism. Loss of KMT5C abrogates the H4K20me3 modification at an oncogenic long non-coding RNA, *LINC01510,* resulting in enhanced transcription of *LINC01510*. *LINC01510* in turn functions as a positive transcriptional regulator of the oncogene *MET* consequently resulting in *MET* upregulation, a predominant mechanism of acquired resistance to erlotinib. Therefore, this study establishes a mechanism of erlotinib resistance mediated by loss of KMT5C, which in part is due to indirect overexpression of *MET*.

CHAPTER 3. IDENTIFICATION OF ONCOGENES THAT DRIVE OSIMERTINIB RESISTANCE VIA DYNAMIC H4K20ME3 REGULATION

3.1 Chapter overview:

In this chapter, we evaluate whether our previous finding that loss of KMT5C induces erlotinib resistance in NSCLC is also true *in vivo,* using xenograft models. Additionally, previous data in our laboratory suggested that KMT5C deposition at constitutive heterochromatin was more stable than at facultative heterochromatin regions such as LINC01510. To test this hypothesis, we optimized techniques such as FISH combined with IF and CUT & RUN, and we generated tools like inducible knockdown cell lines. We further evaluated gene regulation at different time points when KMT5C is modulated for future studies.

3.2 Abstract

Chromatin, a pivotal and dynamic organizational framework, is controlled to sustain cellular equilibrium. Histone proteins, subject to diverse post-translational modifications (PTMs) on their tails, contribute to this regulation. One prominent PTM, histone H4 lysine 20 tri-methylation (H4K20me3), emerges as a critical modulator of chromatin structure, cell cycle progression, DNA repair, and developmental processes. The histone methyltransferase KMT5C catalyzes this modification, establishing both constitutive and facultative heterochromatin, while also demonstrating novel roles beyond traditional domains. Notably, our investigations uncover a novel interplay between KMT5C and the oncogenic long non-coding RNA LINC01510. Elevated LINC01510 levels, triggered by KMT5C deficiency, drive the activation of the MET oncogene, consequently driving resistance to inhibitors targeting the epidermal growth factor receptor (EGFR) in Non-Small Cell Lung Cancer (NSCLC) lines. Expanding from in vitro findings, our *in vivo* xenograft models validate the growth advantage conferred by KMT5C depletion in the presence of EGFR inhibitor Erlotinib. Moreover, while historically confined to constitutive heterochromatin (cHC) regulation, KMT5C's influence extends beyond this role, as corroborated by our own investigation. Intriguingly, our initial evidence indicates variable stability in H4K20me3 deposition by KMT5C, distinguishing regions within cHC from others. This observation drives

the hypothesis of a context-dependent, dynamic interplay between KMT5C and H4K20me3. Our exploration supports this notion, revealing dynamic H4K20me3 regulation within LINC01510-associated domains, divergent from the global H4K20me3 landscape. Furthermore, similar dynamism characterizes the regulation of FOXA1, a predicted target of H4K20me3, upon KMT5C downregulation. Prospective endeavors will aim to identify of additional dynamic targets, real-time visualization of H4K20me3 dynamics in live cells using imaging techniques, and an assessment of its relevance in EGFR inhibitor resistance. In summary, this study elucidates the multifaceted roles of KMT5C and its product, H4K20me3, unveiling a transition from conventional cHC-centric perspectives to a dynamic, adaptable regulatory framework.

3.3 Introduction

Genetic information is encoded by nearly identical DNA sequences and proper regulation of gene expression is partially dependent on packaging DNA into chromatin, a complex of DNA and proteins[15]. Chromatin is divided into euchromatin, which corresponds to an open and transcriptionally active conformation of chromatin, and heterochromatin which is condensed and transcriptionally inert[16]. The major role for heterochromatin is to prevent genomic instability by protecting repetitive regions in the genome from damage and by ensuring correct chromosome segregation [17]. Heterochromatin can be further classified into two subtypes, facultative heterochromatin which is present in gene-rich regions regulating the expression of genes under specific cellular contexts, and constitutive heterochromatin which is typically found in gene-poor regions, including repetitive sequences such as satellite repeats and transposable elements[17]. Chromatin is organized in nucleosomes and each nucleosome consists of a histone octamer. Each of these histones can be post translationally modified on their tail domain, leading to various layers of regulation. One major posttranslational modification (PTM) includes the tri-methylation of histone 4 in lysine 20 (H4K20me3), which is known to be enriched in heterochromatin rich regions.

In our recent study, we identified that this methylation mark H4K20me3 has a role in resistance to drugs used for treatment of Non-Small Cell Lung Cancer (NSCLC). We performed a genome-wide loss of function screen using the CRISPR-Cas9 system to identify potential tumor suppressors that when lost would drive Erlotinib resistance, an Epidermal Growth Factor Receptor inhibitor (EGFRi) used in NSCLC treatment. We discovered that KMT5C loss, drives erlotinib

resistance in this context. KMT5C enables the establishment of constitutive and facultative heterochromatin by generating the H4K20me3 mark[148]. The process by which KMT5C is reduced in tumors is unknown, yet data from human samples suggest that KMT5C is globally downregulated in NSCLC[1]. Loss of KMT5C is a poor prognostic marker in breast cancer and is associated with increased invasiveness[166], migration and epithelial-mesenchymal transition[92]. Additionally, loss of H4K20me3, the modification made by KMT5C, in preneoplasia influences prognosis of NSCLC[83], indicating that loss of KMT5C function is a crucial mechanism in carcinogenesis. Our work was the first to establish its role in EGFRi resistance the 1st generation inhibitor erlotinib *in vitro* however, the role of KMT5C in generating resistance *in vivo* needed to be established. Here we show that loss of KMT5C partially gives a growth advantage to tumor cells when mice are treated with erlotinib vs with negative vehicle control using xenograft models. Additionally, in our effort to elucidate KMT5C role in resistance in various different cell lines, we discovered a novel role for KMT5C and H4K20me3 in their transcriptional regulation of targets.

KMT5C expression has historically been associated with silencing of pericentric regions of the chromatin, telomeres, and repetitive DNA elements generating a constitutive heterochromatic (cHC) state that is not easily reversed[90]. Canonical formation of H4K20me3 in cHC initiates with the interaction between Heterochromatin protein 1 (HP1) and the H3K9me3 silencing mark, facilitating the subsequent recruitment of KMT5C to that specific region through HP1 binding. However, more recently KMT5C was found to regulate gene-rich regions of the genome[73], including our work which identified that KMT5C silences the oncogenic long non-coding RNA LINC01510 leading to reduced transcription of *MET*. The extent to which KMT5C's regulation of gene-rich regions relies on canonical H3K9me3 and HP1 binding is yet to be fully understood within the field. Moreover, our preliminary evidence suggests that the modification made by KMT5C, (H4K20me3) is more easily lost and less stable at these gene-rich/facultative heterochromatic (fHC) regions relative to cHC regions, leading to dynamic gene expression. KMT5C-mediated silencing of genes is likely in line with a fHC state where the silencing can be easily reversed. However, nothing is known about the dynamics of the KMT5C methyltransferase in regulating fHC in NSCLC. It is viable that KMT5C-mediated genomic silencing in cHC regions is stable, likely through sequestration of KMT5C at these genomic regions[148], but KMT5C-mediated silencing of fHC regions is more dynamic and is dependent on the abundance of available

KMT5C. We consider that at these more dynamic locations KMT5C is not sequestered, but instead it enzymatically trimethylates H4K20 and then releases the substrate, and only when KMT5C is plentiful, these marks are maintained. When KMT5C levels drop, the majority of it is sequestered and locked-in at the cHC due to binding to histone protein 1 (HP1), leading to an insufficient amount of soluble KMT5C available to maintain the fHC, leading to gene expression. In this study we further evaluated KMT5C-mediated gene regulation at some fHC regions and tested gene expression regulation at different time points after KMT5C knockdown. However, future work is needed to further confirm this hypothesis and further identification of authentic genes that are genuinely regulated by KMT5C within those gene-rich regions is crucial. This emphasis will enable a more comprehensive evaluation of the hypothesis regarding the dynamic regulation of KMT5C.

3.4 Methods

3.4.1 Cell Culture:

All cell lines used in the study, except for PC9 (Sigma) were obtained from American Type Culture Collection (ATCC). All lines were routinely confirmed to free of mycoplasma contamination monthly. Cell lines generated during the study were authenticated by ATCC Cell Line Authentication. All cell lines were grown in RPMI media supplemented with 10% FBS and 1% Penicillin/Streptomycin. ECas9 cells were continuously cultured in media containing 1µg/ml Blasticidin, The EKVX KMT5C mutant clone C was grown in media containing 100ng/ml Puromycin. The inducible-KMT5C overexpressing Calu6 clone cell line was cultured in 500ng/ml Puromycin containing media in the presence of PBS or Doxycycline (DOX).

3.4.2 KMT5B/C chemical inhibitor preparation:

A-196 (S7983, Selleck Chemicals) was dissolved in DMSO to prepare 10mg/mL stock solutions, which were aliquoted and stored in -80°C.

3.4.3 Generation of inducible KMT5C knockdown cell lines:

Three vectors containing different shRNAs against KMT5C were purchased from Dharmacon using the backbone SMARTvector inducible hEF1α Turbo GFP (V3IHSHEG_8615413, V3IHSHEG_6431737 (sh37), V3IHSHEG_5432959). The second one showed better knockdown efficiency in transient transfection assays and was used for stable cell line generation using transduction. HEK-293T cells were transfected with 20μg of sh37 vector using Lipofectamine 2000 (11668019, Thermo Fisher Scientific) along with 13.3 μg of psPAX2 packaging plasmid, and 6.68μg of pMD2.G envelop plasmid. Media was changed after 5h and 30mL of virus containing media was filtered through a 0.2μm pore size filter and centrifuged 20 000rpm for 3h at 4C. Pellet was resuspended in PBS and used in the transduction of HCC827 cell line. Stable singles clones were selected under the pressure of 1μg of Puromycin using clonal rings.

3.4.4 *In vivo* tumor models:

Animal studies were performed at Hansen Life Sciences Animal Facility at Purdue University. In brief, EKVX with Cas9/ Clone C (5×10^6 with Matrigel) and KMT5C-WT/KMT5C-mutant HCC827 cell lines (5×10^6 with Matrigel) were injected subcutaneously into opposing flanks of female 8 weeks old NU/NU Nude immunodeficient animals (Jackson Laboratories) Xenograft animals whose tumors have reached 100 mm3 were randomly assigned to receive either vehicle (1% polysorbate 80) (n=4), osimertinib (5mg/kg) (n=4) or no treatment (UT) (n=3). Agents were administered by oral gavage once daily at a weekly schedule of 5 days on and 2 days off. Using Vernier caliper, tumor volume was measured four days after injection and every three days after using the following formula: tumor volume (mm3) = width x (length2) x 2-1.

3.4.5 ChIP-qPCR:

Briefly, a total of $2X10^7$ cells were fixed using 1% of filter-sterilized formadehyde for 10 minutes at room temperature. The formaldehyde was quenched with 2.5M Glycine (55μL per ml of media) for 5 min. Cells were washed with cold PBS and scraped into fresh cold PBS. Cells were pelleted by centrifuging at 1500 rpm for 10 minutes at 4°C. The cell pellet was resuspended in 10 mL of freshly prepared cold cell lysis buffer (5mM PIPES, 85mM KCl, 0.5% NP40), kept on ice for 10 minutes followed by centrifuging at 1000 rm for 10 minutes at 4°C. The lysed cells were

resuspended in 1 mL of nuclei lysis buffer (50mM Tris-HCl (pH 8.0),10mM EDTA, 1% SDS) containing 0.1% protease inhibitor cocktail (PIA32955, Thermo Fisher Scientific) and were transferred into 2mL eppendorf tubes, on ice. Cross-linked chromatin from the isolated nuclei was sonicated using a probe sonicator (60% duty cycle) for 10 seconds with a 1minute rest, for 15 cycles to fragment DNA (100-500 bps). Fragmented DNA was immunoprecipitated with antibodies against mouse H4K20me3 (39672, Active Motif), or negative control mouse IgG (5415, Cell Signaling Technology) at 4°C overnight with gentle rotation. The immunoprecipitated DNA was purified using the DNA isolation kit (K1820-01; Thermo Fisher Scientific) following manufacturer's protocol. DNA was used as a template for qRT-PCR as described above. All primer sequences used for qRT-PCR are listed in Supplementary Table 3 from the Cancer Research Manuscript. ChIP data are presented as fold enrichment of DNA immunoprecipitated with H4K20me3 relative to values obtained for DNA immunoprecipitated with IgG control.

3.4.6 In-Cell Western:

Ten-thousand cells were grown in individual wells of a 96-well plate. Forty-eight hours post plating, cells were fixed using cold 100% methanol for 20 minutes at 4 C. Post fixing, cells were permeabilized using 0.2% TritonX in 1X PBS at room temperature for 30 minutes. Cells were blocked using LI-COR blocking buffer for 1.5 hours followed by overnight incubation with primary antibody at 4°C. The primary antibody was detected using 1:800 IR 800CW secondary antibody (LI-COR). The IR-800 signal was quantified using the Odyssey LI-COR imaging system and software. Antibodies used: 1:400 mouse H4K20me3 (39672, Active Motif), 1:500 rabbit GAPDH (2118, Cell Signaling).

3.4.7 Western Blot:

$3X10^5$ cells were grown in individual wells of a 6 well plate, and lysates were isolated at time points specified in figure legends using RIPA buffer (Sodium chloride (150 mM), Tris-HCl (pH 8.0, 50mM), N P-40 (1 %), Sodium deoxycholate (0.5 %), SDS (0.1 %), ddH2O (up to 100 mL)) containing 1X protease inhibitor cocktail (PIA32955, Thermo Fisher Scientific). Protein quantification was performed using Pierce BCA Protein Assay kit. Equal amounts of protein lysate were resolved through 4-20% polyacrylamide gels and transferred onto a polyvinylidene difluoride

(PVDF) membrane. Membranes were blocked using LI-COR buffer for 1 hour at room temperature, and incubated overnight in primary antibody at 4°C. The primary antibody was detected using 1:1000 IR 800CW or 1:1000 IR 680CW secondary antibody. Blots were scanned, and data quantified using the Odyssey LI-COR imaging system and software. Antibodies used: rabbit H4 (61299; Active Motif), mouse H4K20me3 (39672; Active Motif).

3.4.8 RNA isolation and Quantitative real time PCR (qRT-PCR):

$1X10^5$ cells were grown in individual wells of a 12-well plate, and total RNA was isolated after 24 hours, using the miRneasy Kit (217004, Qiagen) according to the manufacturer's instruction. DNase I digestion (79254, Qiagen) was used in each RNA purification reaction to remove genomic DNA. RNA integrity was evaluated on a 1.5% agarose gel, and total RNA quantified using a nanodrop. For quantifying transcript levels from inducible Calu6 KMT5C overexpressing cell line, cDNA was then synthesized from 1µg of total RNA using MiScript Reverse Transcriptase kit (218161, Qiagen), as indicated by the manufacturer's protocol. Q-RT-PCR was conducted using the miScript SYBR Green PCR Kit (218073, Qiagen) as indicated by the manufacturer's protocol, to quantify target gene mRNA expression. The following primers were obtained: *GAPDH* (loading control) (QT00079247, Qiagen), *LINC01510* (LPH09040A, Qiagen). Primers for *KMT5C* quantification are indicated in Supplementary Table 3 from Cancer Research Manuscript.

KMT5C transcript from inducible KMT5C knockdown cell line derived from HCC827, was quantified using Taqman assays. cDNA was synthesized from 1µg of total RNA using SuperScript IV VILO Master Mix (11756050, Thermo Fisher Scientific). Q-RT-PCR was conducted using Taqman Fast Advanced Master Mix (4444963, Thermo Fisher Scientific). The following primers were used: KMT5C:(Hs00261961_m1, Thermo Fisher Scientific), GAPDH (endogenous control) (Hs99999905_m1, Thermo Fisher Scientific) and FOXA1 (Hs04187555_m1, Thermo Fisher Scientific).

3.4.9 CUT&RUN and quantitative real time PCR (qRT-PCR):

CUT&RUN experiments were conducted following the guidelines provided by the manufacturer (EpiCypher CUTANA™ pAG-MNase for ChIC/CUT&RUN, Cat# 15-1116). Initially, cells were washed with CUT&RUN wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1× Roche Complete Protease Inhibitor) and approximately one million cells were bound to activated ConA beads (Bangs Laboratories, cat# BP531). Subsequently, anti-H4K20me3, SNAP-Certified for CUT&RUN (13-0054, Epicypher) was added, and cell permeabilization was carried out using the digitonin buffer (CUT&RUN wash buffer plus 0.01% digitonin). Following washing with the digitonin buffer, samples were incubated with pAG-MNase, followed by additional washes with the digitonin buffer. After the final wash, pAG-MNase activation was induced for DNA digestion by suspending cell samples in the pAG-MNase digestion buffer (digitonin buffer plus 2 mM CaCl2) and incubating on a nutator at 4 °C for 2 h. The solubilized chromatin was released using the stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml RNase A, 50 µg/ml glycogen) and collected using a PCR cleanup kit from the CUT&RUN kit. For purification of input DNA, the traditional DNA Salt precipitation technique was employed.

Primers for CUT&RUN followed by quantitative real time PCR were designed using the Primer Blast website. Verification of best primer set was performed by regular PCR and gel electrophoresis. The primers that produced only one band were selected for CUT&RUN qPCR.

Table 3.1. Primers utilized in the CUT&RUN studies. Designed and purchased from Integrated DNA Technologies.

Primer Name	**Sequence**
LINC15010_1_F:	AGTGCCTGCACAGTTTCTGA
LINC15010_1_R:	GTCCCCTTCCCTTCAGCTTC
GAPDH_2_F:	CCAGTTGAACCAGGCGGC
GAPDH_2_R:	CGCCCGTAAAACCGCTAGTA
FOXA1_F:	GAGAGGTGCGTTGTTTGGG
FOXA1_R:	CCACAGCGGATTAGCGAGG
ZNF_3_F:	TGTAATGACAAAACAGTCCTTTCAA
ZNF_3_R:	GACGCTATGTTTGCCAGGTG
OR_2_F:	TTGGCCTCCTACAAAGTCATTCT
OR_2_R:	GTAGAGAGAGCTTTGCGTCG

3.4.10 Immunofluorescence (IF) and Fluorescence In Situ Hybridization (FISH):

Protocol was modified from a published protocol[169]. In brief, cells were plated in a 24-well plate with collagen coated coverslips. The adherent cells were then fixed with cold 100% methanol for 10min at -20C. Cells were then blocked and RNAse treated with the ABDIL (Buffer composition described in original protocol) and RNase A blocking buffer for 1h. Primary mouse H4K20me3 antibody (39672; Active Motif) in the concentration 1:50 was made in the ABDIL blocking buffer and added to the cells. Coverslips were incubated in a humidified chamber overnight at 4C. Coverslips were then washed three times in PBST for 10 min at room temperature. Secondary antibody 1:100 (Donkey anti-mouse alexa 647 (650nm), LICOR) was added to the coverslips and incubated for 1h in a nutator at room temperature (RT) protected from light. Coverslips were washed again three times for 10min with PBST. Coverslips were then fixed with 2% paraformaldehyde for 20 or 30 min at RT and washed with deionized water. Ethanol series of dehydration was performed starting with 70% ethanol, followed by 90% ethanol and finishing with 100% ethanol. The PNA telomeric probe (2.5μg/mL Cy3-conjugated C-strand telomere PNA probe 550nm, PAN BIO) was added to the coverslips, denatured for 5min at 80C and hybridized overnight at RT in a dark humidified chamber. Coverslips were washed with PNA wash A and B (Buffer recipe in original protocol) and DAPI (1:2000) was added. After washing with deionized water, coverslips were again dehydrated though a graded ethanol series and air dry before mounting and imaging in the confocal microscope.

3.5 Results

3.5.1 Loss of KMT5C in EGFR-WT and EGFR-mutant effect in erlotinib resistance

Previous findings in our laboratory determined that loss of the tumor suppressor KMT5C via upregulation of LINC01510 transcription thereby enhancing MET expression was a mechanism that induced resistance to EGFRi in NSCLC cell lines. However, it was imperative to determine this effect in *in vivo* models. Hence, we injected EGFR wildtype EKVX cells with (Clone C) or without KMT5C mutation (WT) into each flank of an NRG mouse. When tumors reached 200mm^3 in size, the EGFR inhibitor erlotinib was administered to the mice via oral gavage **(Figure 3.1A)**. As observed in **Figure 3.1B-D**, the average tumor volume in KMT5C WT derived xenograft (WT) was lower than in Clone C when mice were in the untreated group, indicating that KMT5C

mutation in EKVX cell line might provide a slight tumor proliferation. Similar results were observed for mice treated with DMSO vehicle control advantage **(Figure 3.1B-D)**. However, EKVX with wildtype KMT5C tumor growth did not reach 200mm^3 which is uncommon for this cell line. One possible explanation is that EKVX with KMT5C wildtype was derived from single clonal selection and expressed Cas9 (ECas9), suggesting that possibly this clone does not intrinsically grow well *in vivo*. Nevertheless, upon erlotinib treatment, tumor growth from EKVX with KMT5C-WT cells (WT) regressed significantly when compared to EKVX with KMT5C mutated (Clone C) derived xenograft. This significance was observed 28 and 32 days after treatment with erlotinib. Evaluation of this phenotype for more days after erlotinib treatment was not possible due to extensive growth of Clone C derived xenograft, and mice in this group had to be euthanized. This finding suggests that KMT5C potentially plays a role in sustaining tumor growth after erlotinib administration. To gain a more comprehensive understanding of the underlying mechanisms through which KMT5C influences resistance to erlotinib and other EGFR inhibitors, it is imperative to investigate further. For instance, an assessment of tumor progression in the presence or absence of erlotinib could be executed utilizing xenograft models established from a KMT5C doxycycline-inducible knockdown cell line. By exercising precise control over KMT5C downregulation through doxycycline induction, an enhanced experimental framework can be established. This strategy allows for a categorization of the experimental cohort, with a subset of mice receiving doxycycline treatment and others serving as a vehicle control group. Such a discriminative approach enables a more robust comparison of tumor growth, effectively mitigating the intrinsic impact of inherent cell growth kinetics.

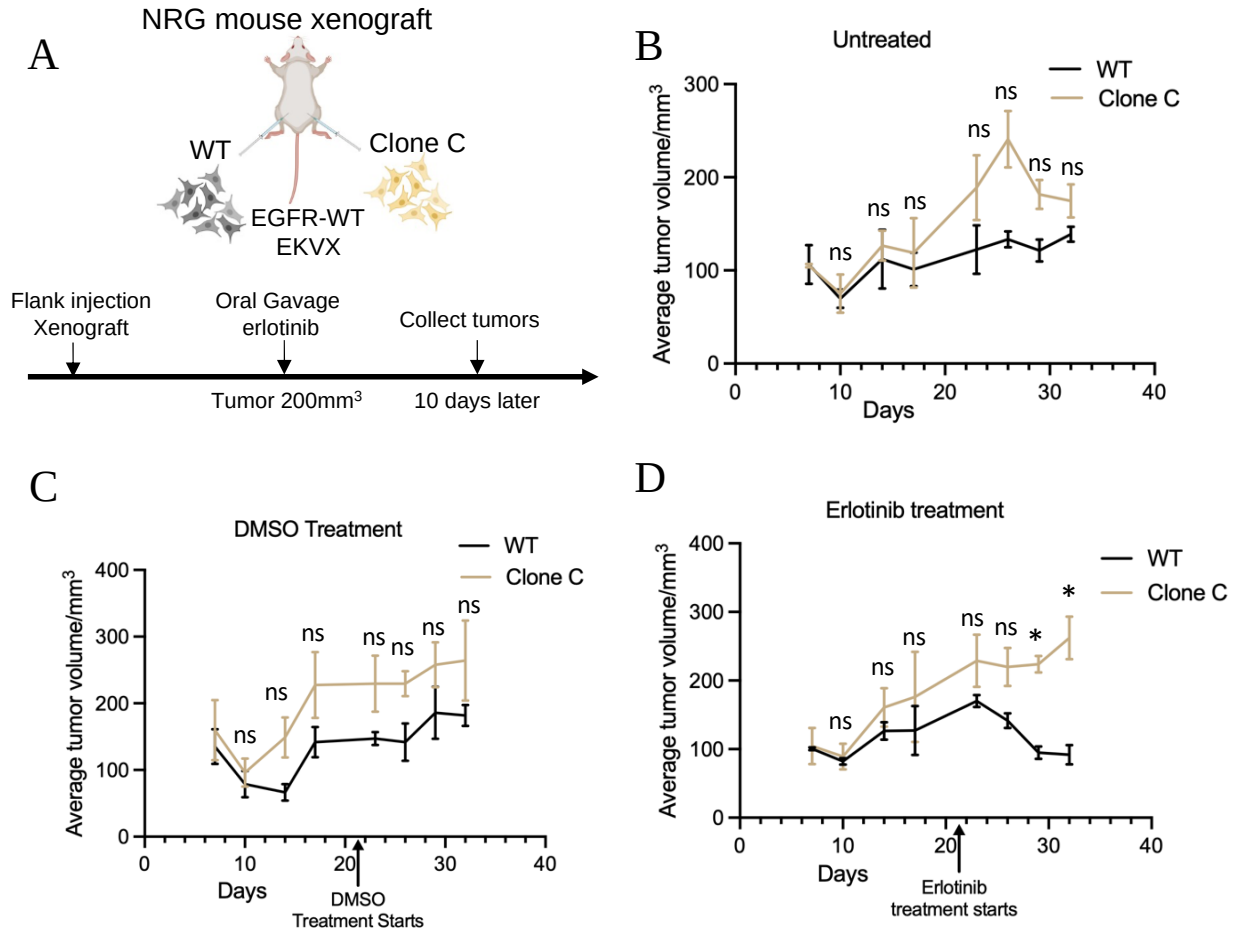

Figure 3.1. In vivo effect of erlotinib in EGFR-WT with KMT5C-WT or KMT5C-Mutant derived xenograft model. A) Experiment schematic and timeline of WT and Clone C (EGFR-WT) derived xenografts in NRG mice. Drug administrated via Oral Gavage three times a week. **B)** Tumor growth curves of WT or Clone C derived xenograft in the untreated group. n=3 **C)** Tumor growth curves of WT or Clone C derived xenograft in the DMSO vehicle treated group. n=3 **D)** Tumor growth curves of WT or Clone C derived xenograft in the erlotinib treated group. n=3. Tumor volume for all figures was determined using the equation Tumor Volume (mm^3) = Length * $Width^2/2$. *, $P < 0.05$.

3.5.2 H4K20me3 regulation of LINC01510 is dynamic when KMT5C overexpression induction is removed.

Furthermore, the evaluation of KMT5C's role in erlotinib resistance in vivo, more specifically that loss of KMT5C provides a growth advantage to tumors when mice are treated with erlotinib, confirmed the importance of understanding KMT5C and H4K20me3 regulation in cancer. Therefore, we generated multiple stable cell lines with either loss, overexpression, or rescue of KMT5C. We then evaluated how changes in KMT5C levels of expression affected the

transcription of specific oncogenes and oncogenic non-coding RNAs such as MET and LINC01510 respectively. More specifically, for our overexpression model, we used the Calu-6 NSCLC cell line previously reported to have low levels of endogenous KMT5C, to induce KMT5C expression after doxycycline (DOX) was added to the media **(Figure 3.2A)**. These cells were then cultivated in three different conditions, 1) cells grown in negative control PBS, 2) cells grown in doxycycline constantly and 3) cells grown in DOX and the without DOX for only 24h **(Figure 3.2B)**. Evaluation of overall H4K20me3 in these cells grown in the three different conditions was performed, and no change was observed between cells grown in conditions of constant DOX and DOX withdrawal for 24h **(Figure 3.2C)**. Similarly, when evaluating the overall H4K20me3 in these three different conditions by using In Cell Western, we observed comparable results **(Figure 3.2D)**. However, the levels of LINC01510 transcript were different from DOX treatment to DOX withdrawal treatment **(Figure 3.2E).** Furthermore, H4K20me3 regulation of LINC01510 genomic region was validated by ChIP qPCR as shown in **Figure 3.2F**. Overall, these results indicate that when KMT5C induction is removed for 24h, overall levels of H4K20me3 in the cells do not change but H4K20me3 mark at specific regions such as LINC01510 do seem to be dynamic and changing. This suggests that H42K0me3 mark is dynamic at genomic regions, but it is likely sequestered at other regions like repetitive-rich regions, where H4K20me3 is normally very abundant.

Further work is needed to determine the regulation of other genes that are regulated by H4K20me3 in this system and to identify which genes are dynamically regulated and which ones are not. To further identify these candidate genes, we are currently working with Dr. Paula Vertino in Rochester University, where CUT&RUN sequencing studies will be performed in our cell lines for better identification of genes that are regulated by H4K20me3. We will then have a better picture of the genes that are regulated by H4K20me3 in our cells, and we will evaluate their transcriptional regulation vs other repetitive-rich H4K20me3 regions when KMT5C is modulated.

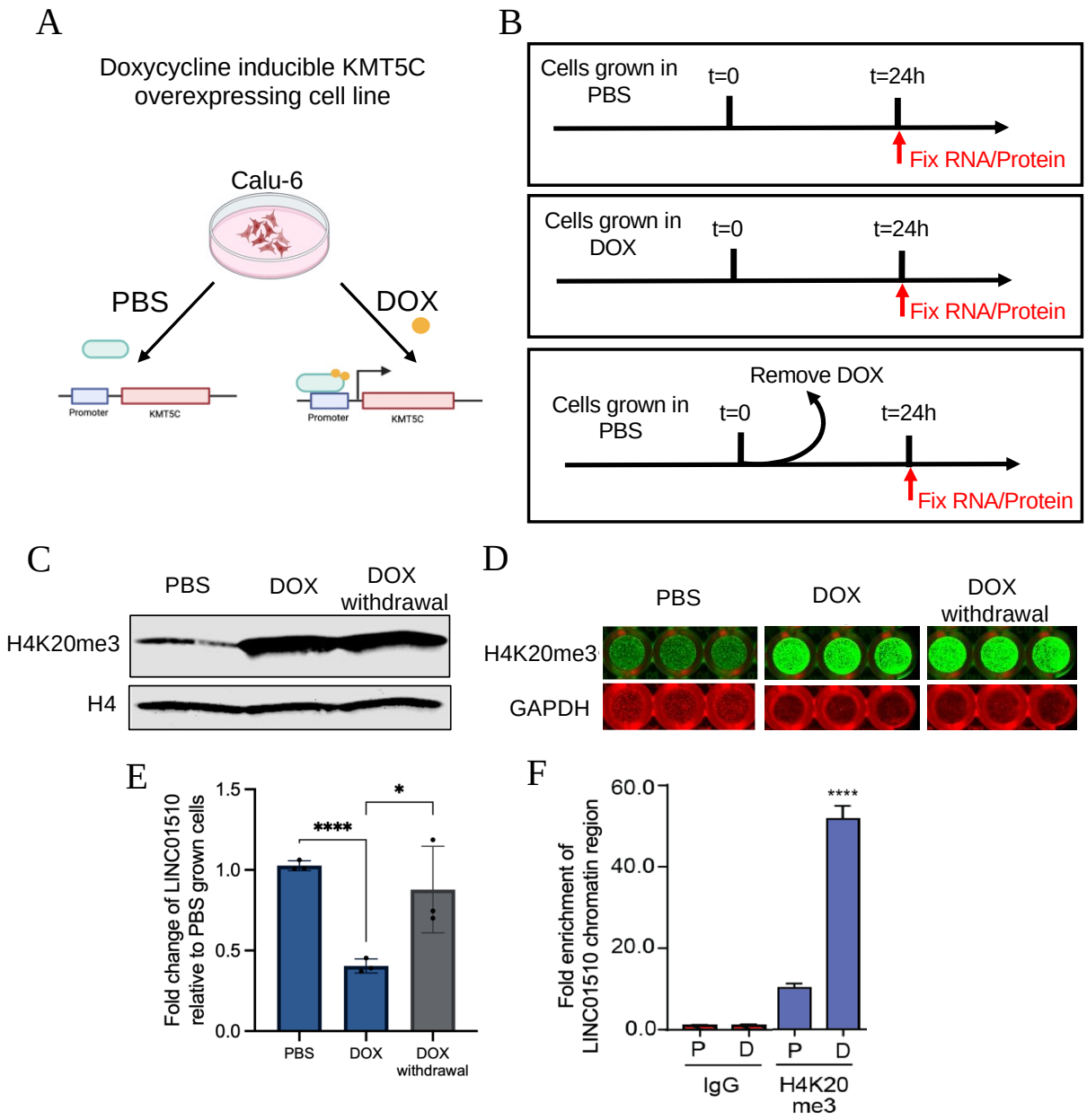

Figure 3.2. H4K20me3 overall levels are stable but regulation in target genes such as LINC01510 is dynamic. **A)** Schematic of doxycycline inducible KMT5C overexpression system in Calu-6 cell line. **B)** Diagram of three different experimental conditions used in C), D) and E). Cells were grown in PBS (1), DOX for over a month (2) or in DOX for over a month and then in PBS for 24h (DOX withdrawal) (3). C) Representative Western blot analysis of H4K20me3 in Calu-6 inducible cell line in the indicated treatment B). H4 was used as a loading control. **D)** H4K20me3 levels in Calu-6 doxycycline (DOX) inducible clone using In Cell Western, cells were grown in PBS, DOX or in DOX withdrawal for 24h (3). **E)** LINC01510 transcript levels in Calu-6 clones grown in PBS (1), DOX (2), or DOX withdrawal for 24h (3) measured by RT-qPCR Taqman assays. **F)** ChIP followed by RT-qPCR. Fold enrichment of LINC01510 chromatin region pulled-down by H4K20me3 relative to IgG. P:PBS, D:DOX. One-way ANOVA followed by Dunnett's Multiple Comparison test. ****P<0.0001

3.5.3 Evaluation of H4K20me3 dynamic regulations of targets when KMT5C is knocked down.

Generation of KMT5C inducible knockdown cell line.

To further confirm the previous finding of H4K20me3 dynamic regulation of gene-rich regions vs other genomic regions, we generated a KMT5C inducible knockdown cell line to test whether this effect was also true with endogenous KMT5C expression. The inducible knockdown cell line had an inducible promoter dependent on doxycycline treatment and a Green Fluorescent Protein (GFP) reporter upstream of the shRNA against KMT5C sequence **(Figure 3.3A)**. After generation of stable cell lines, three different clones were selected and *KMT5C* transcript levels were evaluated after DOX treatment for 48h **(Figure 3.3B).** The knockdown efficiency was higher for the clone 1 and clone 8. Clone 8 was selected for further experiments because of knockdown efficiency and high GFP signal **(Figure 3.3C)**. The best knockdown efficiency in clone 8 was found to be at 72h **(Figure 3.3D)** after DOX treatment, indicating that KMT5C knockdown seems to be time dependent. Moreover, FOXA1, a gene previously reported to be regulated by H4K20me3[95], was evaluated and indeed upregulation of FOXA1 transcript was found when clone 8 was treated with DOX at either 24, 48 or 72h **(Figure 3.3E)**. This suggests that upon DOX induction, knockdown of *KMT5C* leads to upregulation of genes that are regulated by H4K20me3 such as FOXA1.

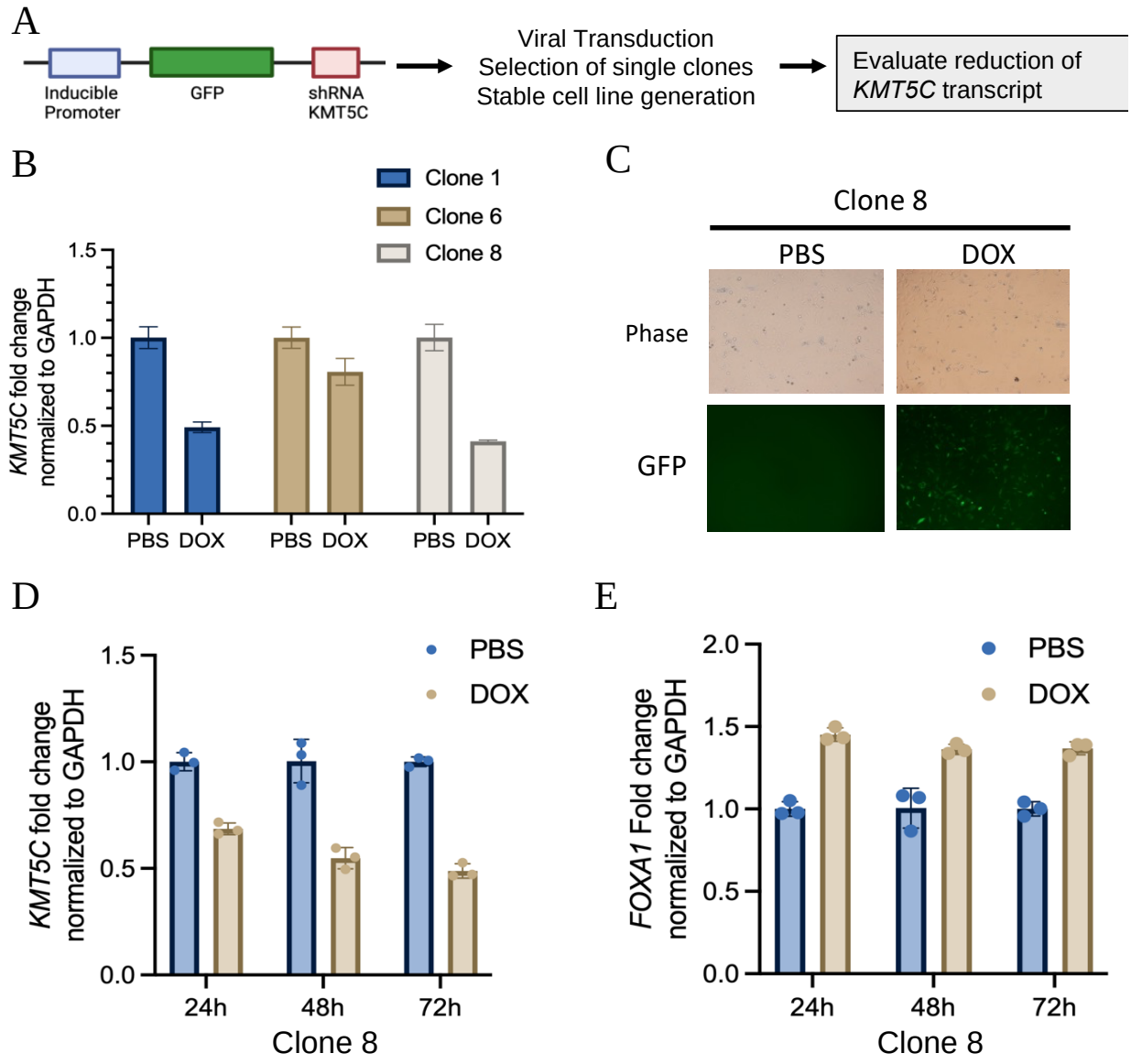

Figure 3.3. Knockdown of KMT5C upon doxycycline (DOX) regulation leads to FOXA1 increased transcript. A) Schematic of KMT5C inducible knockdown cell line generation. GFP upstream of shRNA-KMT5C serves as a transduction control. Inducible promoter is regulated by presence of doxycycline (Tet-on system). **B)** qRT-PCR (Taqman) data for KMT5C in three different single clones with DOX-inducible shRNA KMT5C vector. **C)** Fluorescent microscopy of inducible single clone 8 in presence of PBS negative control or DOX. **D)** qRT-PCR (Taqman) data for KMT5C in single clone 8, 24h, 48h and 72h after DOX knockdown induction. **E)** qRT-PCR (Taqman) data for FOXA1 in single clone 8, 24h, 48h and 72h after DOX knockdown induction.

Time point determination for observing H4K20me3 dynamic regulation when KMT5C is knocked down.

Following KMT5C inducible knockdown cell line generation, we wanted to determine the time point after KMT5C knockdown induction where we start to see a dynamic regulation of genes regulated by H4K20me3, such as FOXA1. For this, we first evaluated the knockdown efficiency of clone 8 at different concentrations of DOX and at different time points **(Figure 3.4A).** KMT5C downregulation of transcript was found to be better 48h after DOX treatment, however there was no significant difference among doses. This suggests that KMT5C knockdown in this cell line reaches its highest efficiency after 48h of induction, but it is not dose dependent **(Figure 3.4A).** Moreover, KMT5C downregulation 24h after DOX treatment was enough to derepress transcription of FOXA1 gene **(Figure 3.4B).** FOXA1 de-repression was not significantly different between time points and between doses of DOX. This is concomitant with the KMT5C knockdown result, where knockdown efficiency was similar between time points and doses. To further validate these finding and evaluate H4K20me3 levels after KMT5C knockdown, we performed a time course experiment with a low dose of DOX. Low dose of DOX (100ng/mL) was found to be enough to reach the highest KMT5C knockdown efficiency in this clone (Data not shown). Similarly, we added a treatment with the KMT5B/C inhibitor A-196 to compare H4K20me3 reduction levels when KMT5B/C is inhibited vs when KMT5C is knocked-down by the shRNA **(Figure 3.4C).** Quantification of the western blot showed that when KMT5C is knocked-down H4K20me3 levels reduce around 40% 96h after DOX treatment, however when cells are treated with KMT5B/C inhibitor, H4K20me3 levels are completely abolished **(Figure 3.4D).** This indicates that KMT5C knockdown is not as efficient as enzymatic inhibition by the small molecule. This also suggests that KMT5B activity could be compensating for KMT5C downregulation or that other methyltransferase is compensating for KMT5C knockdown.

To better test our hypothesis we need a precise modulation of KMT5C levels therefore, future work is needed where we will proceed to analyze dynamics when the small molecule KMT5B/C inhibitor A-196 is used, instead using the knockdown model. By using the small molecule, we can better control the H4K20me3 levels in the cells thereby, allowing better understanding of H4K20me3 transcriptional regulation of genes in a time dependent manner. We are currently working on determining the best time point for evaluation of H4K20me3 transcriptional regulation

of genes using the small molecule. We will test different A-196 concentrations by performing a titration of this inhibitor at different time points. This will us to identify a concentration in which H4K20me3 levels are reduced but not completely abolished and another concentration in which H4K20me3 levels are abrogated. We will then evaluate the regulation of H4K20me3 target genes, when H4K20me3 levels are partially reduced, when H4K20me3 levels are completely gone and when H4K20me3 is intact (No A-196 treatment). By evaluating these three time points, we will be in a better state to identify those genes that are dynamically regulated by H4K20me3 and those genes or regions where H4K20me3 is not dynamic but is very stable. This will further confirm our hypothesis model of H4K20me3 dynamic regulation of genes vs other regions where H4K20me3 is believed to be less dynamic due to sequestration of KMT5C by HP1ß binding.

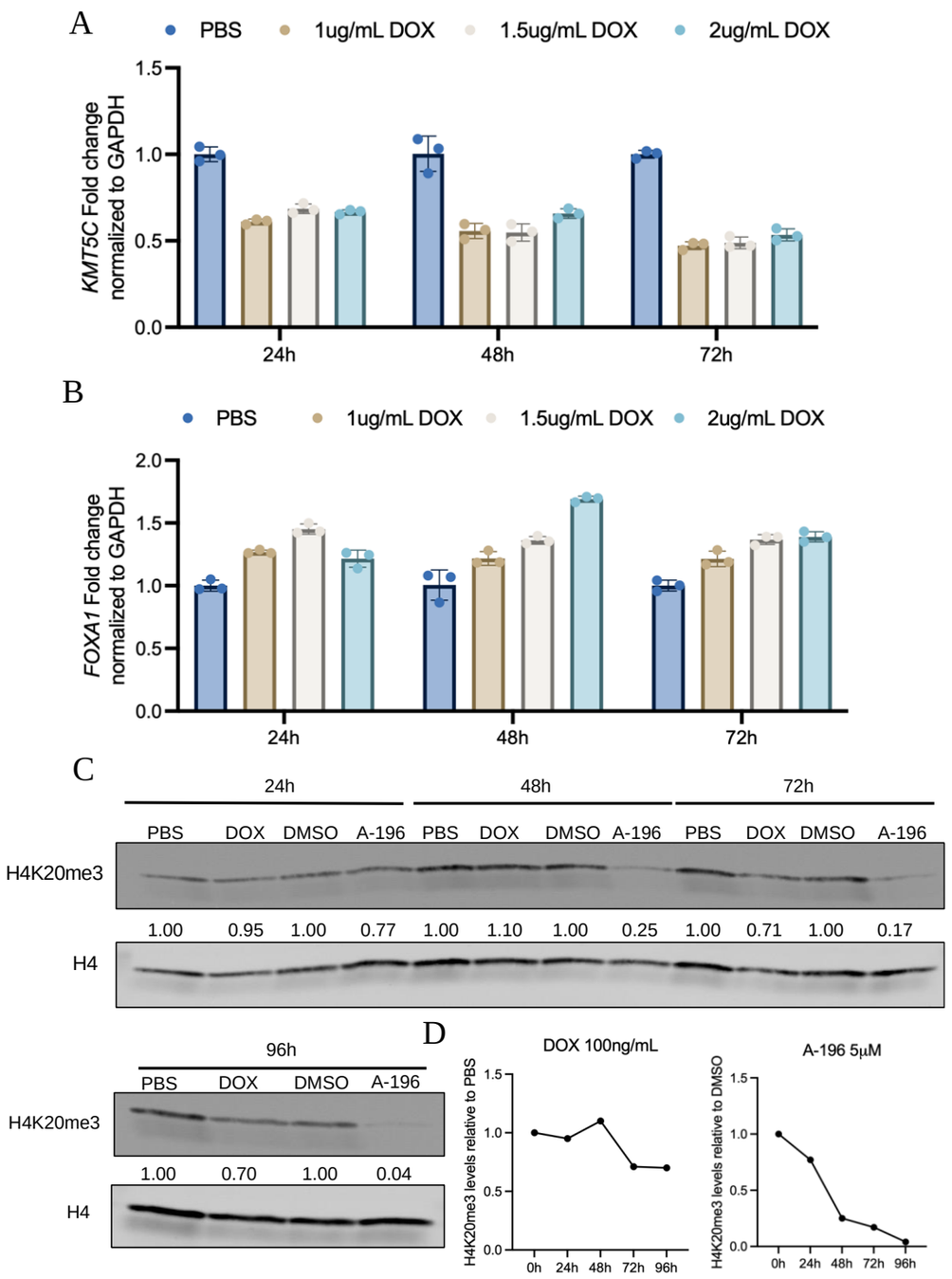

Figure 3.4. KMT5C downregulation leads to increase FOXA1 transcript levels in a dose dependent manner. A) qRT-PCR (Taqman) data for KMT5C in single clone 8 upon treatment with different concentrations of doxycycline. **B)** qRT-PCR (Taqman) data for FOXA1 in single clone 8 upon treatment with different concentrations of doxycycline. **C)** Western blot analysis of H4K20me3 in single clone 8 after treatment with DOX (100ng/mL) or A-196 (5μM) for 24h, 48h, 72h and 96h. H4 was used as a loading control. DMSO is used as a vehicle control of A-196. PBS is the negative control of DOX. **D)** Summary of H4K20me3 quantification results from western blot analysis when single clone 8 was treated with DOX (100ng/mL) or A-196 (5μM).

3.5.4 H4K20me3 regulation of telomeres vs facultative heterochromatin in live cells.

In addition to evaluating modulation of *LINC01510* and *FOXA1* transcript levels after KMT5C is inhibited by the small molecule, we wanted to test overall changes of H4K20me3 when A-196 is added into the cells by using high-resolution microscopy. This will give us a better picture of how H4K20me3 association with dense heterochromatic regions, such as telomere and with other gene-rich regions, varies when KMT5C is inhibited. More specifically, we want to quantify the co-localization of H4K20me3 with genomic regions where H4K20me3 regulation is dynamic, for example LINC01510 and FOXA1 regions, and compare that to H4K20me3 co-localization with repetitive rich regions such as telomeres at different time points after A-196 treatment. This will indicate whether H4K20me3 localization at telomeres is more stable or not when compared to H4K20me3 localization at other genomic regions thereby, testing our hypothesis of H4K20me3 dynamic regulation of gene-rich regions vs repetitive-rich regions. For this, we optimized a previously described protocol[169] where we labeled telomere repetitive sequences using a Fluorescent In-situ Hybridization probe along with labeling of H4K20me3 using a Immunofluorescence. Combination of these two methodologies allows us to track both components and compare H4K20me3 co-localization with telomeres vs co-localization with other gene-rich regions.

In the following figure we tested and optimized this technique using different experimental conditions. In **Figure 3.5**, we observe successful telomere labeling in the top right panel along with DAPI staining. Similarly, we performed different conditions of fixation after H4K20me3 immunofluorescence and then coverslips were hybridized with the telomere probe specific for the CCCTAA repetitive motif. H4K20me3 foci dense areas in this experiment do not necessarily co-localize with telomere foci **(Figure 3.5)**. This suggests that for better co-localization analysis we would need super high-resolution microscopy or perform a z stack acquisition. For this, we will partner with our collaborators from the Yuan lab in the Chemical Engineering Department at Purdue University. Future work will focus on testing other probes against gene-rich regions such as LINC01510, using this methodology and testing their co-localization with H4K20me3.

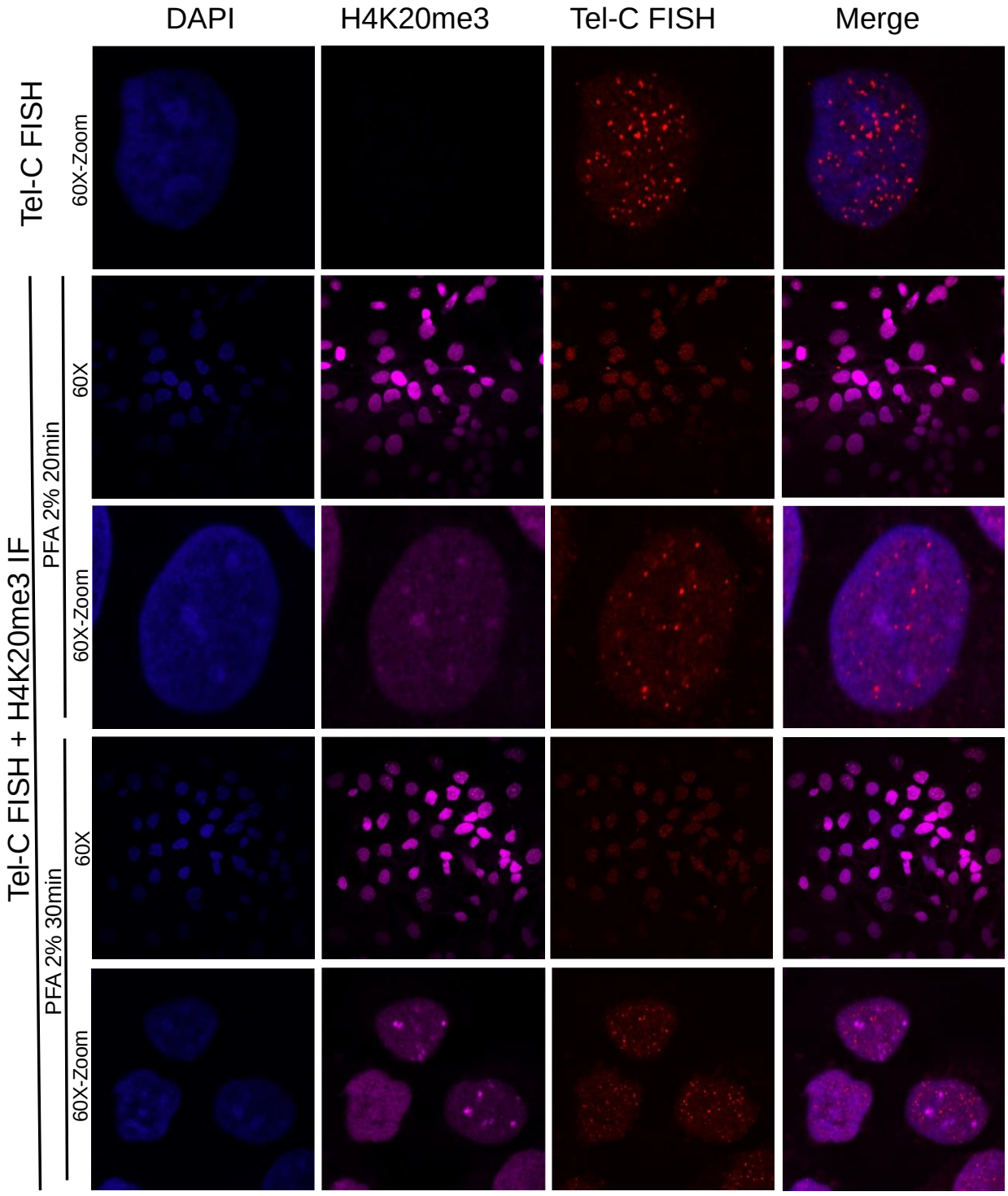

Figure 3.5. H4K20me3 and Telomere labeling optimization. Top panel: Telomere hybridization using Fluorescence In-Situ Hybridization (FISH) of the DNA Tel-C probe (CCCTAA). Zoomed image was acquired using 60X lens in a Nikon confocal microscope. Four other panels: Immunofluorescence of H4K20me3 combined with FISH to label telomeres in HCC827 after samples were fixed with different incubation times of Paraformaldehyde (PFA).

3.5.5 H4K20me3 levels are reduced when KMT5C is mutated or inhibited when using CUT&RUN methodology.

To further verify whether H4K20me3 regulation is dynamic, it is essential to look at transcription changes when KMT5C is modulated but also to evaluate H4K20me3 mark at different regions of the chromosome using a chromatin profiling technique. For this, we performed CUT&RUN followed by qPCR at regions where we predict more dynamic regulation by H4K20me3 and regions where H4K20me3 is highly enriched due to high abundance of repetitive sequences. CUT&RUN or Cleavage Under Target and Release Using Nuclease is a chromatin profiling strategy similar to Chromatin Immunoprecipitation (ChIP). However, unlike ChIP, it is performed in situ which allows for quantitative high-resolution chromatin mapping and probing of the local chromatin environment[170]. We first confirmed the specificity of the antibody by performing CUT&RUN of H4K20me3 and H3K4me3 positive control in HCC827 cells. Indeed, H3K4me3 was highly enriched in GAPDH genomic region as seen in **Figure 3.6A**. Similarly, to confirm that the assay is sensitive, we evaluated H4K20me3 mark in the previously identified LINC01510 gene in HCC827 cell line with KMT5C mutated or wildtype. H4K20me3 was reduced in cells where KMT5C was mutated confirming our previous results[13] when using ChIP-qPCR (Figure 3.6B). Furthermore, we proceeded to evaluate H4K20me3 enrichment in the FOXA1 gene in HCC827 cells treated with DMSO negative control or KMT5B/C A-196 inhibitor for 72h. H4K20me3 was highly enriched in FOXA1 gene in DMSO treated compared to A-196 treated, indicating that chemical inhibition of KMT5B/C for 72h effectively abolishes H4K20me3 mark at FOXA1 gene **(Figure 3.6C).** Likewise, LINC01510 region also loses H4K20me3 when the cells are treated with A-196 inhibitor for 72h **(Figure 3.6D).** H4K20me3 dramatic reduction after the chemical inhibitor treatment was expected, as we hypothesis that H4K20me3 is not strongly sequestered at these regions.

Moreover, we wanted to evaluate other genomic regions where both H4K20me3 and H3K9me3 are highly enriched. For this we used publicly available ChIP seq data for H4K20me3[73] and H3K9me3[171] in human normal lung fibroblasts IMR-90 cell line and searched for highly silenced heterochromatic regions such as centromeres and peri-centromeres. Two genomic regions highly silenced and highly repetitive were evaluated: ZNF510 (ZNF_3) and a region downstream of the OR4Q3 (OR_2) gene. Both regions lost H4K20me3 methylation after treatment with 5μM of

KMT5B/C inhibitor for 72h **(Figure 3.6E, F)**. We expected less dramatic reduction of H4K20me3 at these regions, however this could be because the inhibitor dose and exposure time might be leading to a dramatic reduction of H4K20me3 overall, obscuring dynamic changes of H4K20me3 in the chromosome. Hence, we will vary A-196 concentration to get a less stringent inhibition of KMT5C, to better evaluate dynamic H4K20me3 silencing in different regions. A more precise modulation or inhibition of KMT5C, will allows us to better compare H4K20me3 regulation at dynamic regions vs more stable regions.

Additionally, upon analysis of Dr. Paula Vertino's CUT&RUN seq data, we will have a list of genuine candidate genes that are regulated by H4K20me3 in our cell lines, thereby ensuring that we are indeed evaluating true H4K20me3 targets. Once we get the candidate list, we will validate that H4K20me3 is enriched at those candidate genes by performing CUT&RUN followed by qPCR. We will then compare the regulation of some of those genes using CUT&RUN qPCR at different time points after KMT5C inhibition by the small molecule A-196 and using regular RT-qPCR to evaluate transcriptional regulation of H4K20me3 as well. This experiment will elucidate the mechanism of H4K20me3 regulation in different genomic regions upon KMT5C inhibition thereby, testing our hypothesis.

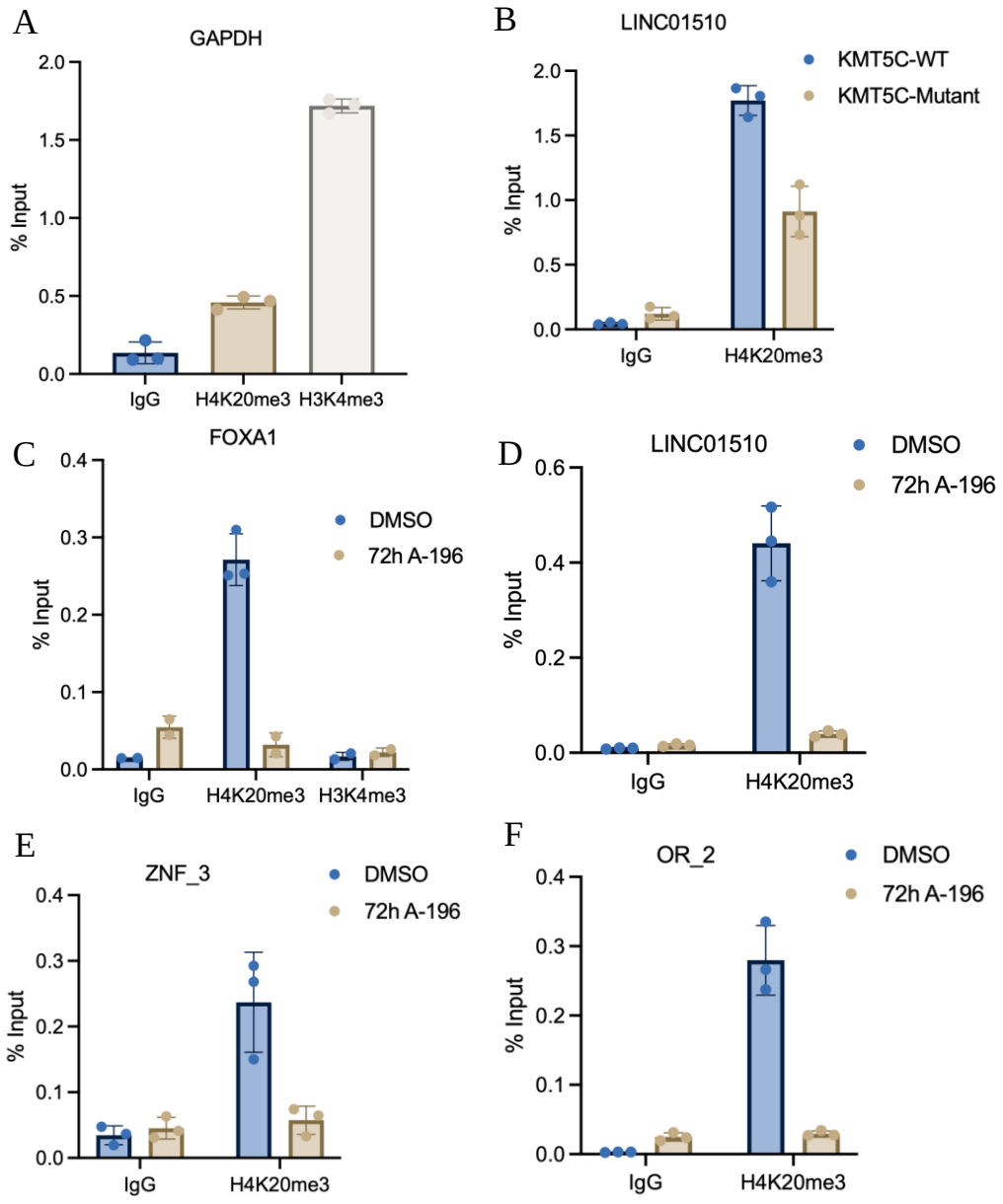

Figure 3.6. KMT5B/C inhibition for 72h leads to decreased H4K20me3 in LINC01510 , Zinc Finger Genes (ZNF) and Olfactory genes (OR). A) CUT&RUN qPCR of H4K20me3 and H3K4me3 in GAPDH gene in HCC827 cells. IgG was used as a negative control for all CUT&RUN experiments. **B)** CUT&RUN qPCR of H4K20me3 and H3K4me3 in FOXA1 gene in HCC827 cells treated with DMSO or A-196 5µM for 72h. DMSO is used as a vehicle control. **C)** CUT&RUN qPCR of H4K20me3 in LINC01510 gene in HCC827 KMT5C-WT/KMT5C-Mutant cells. **D) E) F)** CUT&RUN qPCR of H4K20me3 in LINC01510/ZNF/OR genes in HCC827 cells treated with DMSO or A-196 5µM for 72h.

3.6 Discussion and future directions

NSCLC therapeutics focused on targeting EGFR-mutant patients, has made a lot of progress in the recent years thanks to the development of novel third-generation EGFR inhibitors, such as osimertinib. However, acquired resistance remains a challenge and in most instances it is inevitable[162]. Most of the described mechanisms responsible for resistance have been secondary mutations of the EGFR receptor or alternative growth pathways that get overactivated. Additionally, some nongenetic mechanisms, such as epigenetic alterations have also been reported to contribute to resistance. More specifically, reports have identified novel roles of epigenetic regulatory proteins in the development of EGFRi resistance[172,165,173]. A recent report identified that loss of CBX5 also known as HP1, a methyl-lysine-binding protein that localizes to heterochromatin and has a function in gene silencing, conferred resistance to multiple EGFRi[174]. Similarly, our report indicated that loss of KMT5C, a methyltransferase responsible for H4K20me3 repressive mark, can also drive EGFRi resistance. Most of the studies where epigenetic factors have an effect on resistance, more specifically to EGFRi, have been performed in cell culture[174,13]. However, to translate this basic biology finding to the clinic, it is imperative to determine the effect *in vivo*. We reported here that loss of KMT5C in EKVX (EGFR-WT) derived xenografts, has a partial effect on tumor resistance. However, the effect was not as significant due to reduced tumor growth of the control cell line EKVX. We also tested another cell line HCC827 (EGFR-Mutant) were KMT5C was lost in a similar experimental setup (data not shown). However, KMT5C knock-out or mutated cells failed to form tumors *in vivo* in this context thereby, impeding the evaluation of the phenotype. One of the possible reasons for this finding is that loss of KMT5C cells may have led to a loss of genomic stability and increase genetic aberrations. The latter is concomitant with previous reports where the loss of KMT5B/C was shown to impair the DDR mechanism, leading to accumulation of DNA damage[18,23,32,71,149,155,156]. Hence, it is possible that disruption in genomic stability in cells where KMT5C was lost, might impair tumor growth *in vivo* thereby complicating the testing of our hypothesis. Additionally, it is important to note that we evaluated tumor growth of two completely different cell lines due to single cell stable cell line generation. Because of this, it is hard to make a conclusion based on tumor growth comparison, as different cell lines tend to have intrinsically different proliferation patterns *in vitro* and *in vivo*. Similarly, there is a slight possibility that tumors in the same animal could lead to some sort of competition for metabolic resources to facilitate tumor growth. Therefore, to better account for

these issues, it is important to evaluate tumor growth from a cell line that contains a doxycycline (DOX) inducible promoter and where mice can be divided in two groups; one grouped treated with the negative control PBS and the other group treated with DOX. Conclusions from such experiment will lead to a better conclusion on whether KMT5C loss drives EGFRi resistance *in vivo*.

Moreover, cells need a balance between stability and reversibility in terms of gene expression regulation. Epigenetic regulation is one the integral components of this dynamic control when cells are responding to different stimuli[175,14]. Specifically, appropriate maintaining of chromatin modifications, such as H4K20me3 are essential for cell homeostasis, as its dysregulation has been linked to multiple diseases[69,176,89]. Likewise, KMT5C has been linked to heterochromatin rich regions of the chromosomes, thereby regulating gene transcription and gene expression[23]. Similarly, there has been evidence that the methylation mark H4K20me3 can be removed by a variety of demethylases such as HsRAD23[76], Mdig[151], JMJD2[52] and PHF2[44]. The latter indicates that H4K20me3 mark is most likely dynamic in response to stimuli as it is highly regulated by the cell. However, there are very few reports on how H4K20me3 is regulated in the cells and how it is directed to specific regions of the chromosome. One report indicates the mechanism of how H4K20me3 is directed to specific regions of the chromosome by binding to the long non-coding RNA PAPAS[60]. However, whether this regulation is true in other biological contexts and whether this mechanism is dynamic has yet to be determined. Here we report that H4K20me3 levels in different genomic regions might be dynamically dependent on KMT5C overall abundancy in the cells. More specifically, we found that when induction of KMT5C overexpression was removed for 24 hours, overall H4K20me3 levels did not change but H4K20me3 levels at other genomic regions such as LINC01510 did change. Similarly, in our KMT5C knockdown system, we found that *FOXA1* (gene known to be methylated by KMT5C[95]) transcript levels increase after 24h of KMT5C downregulation, however overall levels of H4K20me3 are relatively sustained. It is possible that this differential regulation of genomic regions by H4K20me3 is an adaptation of cells to maintain genomic stability by keeping heterochromatic regions rich in repetitive sequences highly methylated and silenced, whereas other more transcriptionally open regions lose H4K20me3 mark more easily. Moreover, H4K20me3 evaluation of more silenced and repetitive regions such as the Zing Finger gene ZNF510 and the olfactory gene OR4Q3 using CUT&RUN

showed a dramatic decrease after treatment with KMT5B/C inhibitor. Treatment with KMT5B/C can give us an answer to H4K20me3 dynamics if we change it slightly over time, however time point and dose of the inhibitor treatment will need to be further optimize for better evaluation of H4K20me3 dynamics regulation of genomic regions. Future work will aim at determining the best time point and dose to evaluate H4K20me3 dynamics and at evaluating overall H4K20me3 enrichment in the genome by CUT&RUN sequencing and RNA sequencing. By employing CUT&RUN sequencing to unravel the genuine targets of KMT5C regulation through H4K20me3 methylation, we can significantly enhance our comprehension of the dynamic regulation of H4K20me3 in fHC versus cHC. This approach will provide us with valuable insights into the differential regulation of these distinct chromatin states.

CHAPTER 4. IN-CELL WESTERN PROTOCOL FOR HIGH THROUGHPUT SCREENING OF SINGLE CLONES

In accordance with the Bio-protocol policy on author use, the following publication (Pal A.S., & Agredo A., Kasinski A.L., 2022) is the accepted version of the article published in Bio-protocol (https://en.bio-protocol.org/en/bpdetail?id=4489&type=0) and has been reprinted with permission from Bio-protocol.

4.1 Chapter Overview:

In this chapter, we describe a protocol for screening single clones using a technique called In-Cell Western, which is found to be useful for high throughput experiments clones screening, when compared to regular Western Blot.

4.2 Introduction

Knockout or overexpression screens are efficient methods for identifying the involvement of novel genes that contribute to phenotypes such as drug resistance. Perturbation of gene function is enabled through either loss-of-function studies, using biological tools such as the CRISPR-Cas9 system (Clustered Regularly Interspaced Short Palindromic Repeats)[177,178,179], or by gain-of-function studies through overexpression of human ORFs (Open Reading Frame)[180]. Several genes can potentially be identified by such screening methodologies; however, validation is a key step. To validate and dissect the cellular function of the gene(s) of interest, gene modulation has become a prevalent technique in the field. To this end, the candidate gene is either individually knocked out or overexpressed, single clones are isolated, and finally the phenotype observed via the screen is reevaluated. Nevertheless, to confirm that the gene of interest is accurately knocked out or overexpressed in single clones, protein quantification is a usual process.

Differential protein expression of individual clones is conventionally determined by western blot. This and other techniques such as immunofluorescence and immunohistochemistry are indispensable methods for protein analysis; however, these protocols require increased amounts of resources (antibodies, reagents) and often need one or more days to complete. Therefore, a semi-

high-throughput screen that allows for rapid identification of differential protein expression post-clonal selection that reduces expenses, labor, and time should be considered. The in-cell western (ICW) is a powerful, simple, and reproducible technique that is underutilized in the field. It is a cost-effective method to quantify intracellular signaling in intact cells. The ICW protocol involves fixation and immunostaining of cells and combines the specificity of a western blot with the reproducibility and throughput of an enzyme-linked immunosorbent assay (ELISA)[181]. From previous reports, semi-high-throughput cell-based applications of ICW include: 1) identification of efficient siRNAs from libraries and 2) identification of small molecule inhibitors targeting a particular signaling pathway[182]. Additionally, ICW has been successfully utilized for screening genotoxic drugs by quantifying the expression of γH2AX, a well-known DNA damage and repair marker[183]. Another example of the versatility of ICW for throughput experiments is during the screening of chemical libraries for compounds that modulate the intensity and duration of growth factor-induced MAPK activity, an important regulator in cancer progression[184].

Here, we describe a semi-high-throughput screening mechanism using the ICW protocol for validation of single knockout or overexpression clones for a protein of interest, initially identified using the CRISPR Cas9 screening system[13]. ICW has proved to be an efficient technique for clonal selection of cells because it allows rapid analysis of numerous samples, conserving the accuracy of the quantifiable output. To the best of our knowledge, the ICW protocol described below is the first reported use of ICW for the selection of multiple single clones simultaneously. The pros and cons of using ICW versus western blot for efficient clonal selection of cells are enlisted below **(Table 4.1).**

Table 4.1. Pros and cons of using in-cell western over using the more conventional technique of western blot for high-throughput selection of single clones.

Pros	Cons
• Fewer cells needed • Reliable quantification if unequal cells are plated • Less volume of reagents and buffers needed • Primary antibodies can be reused • Plates can be stored at 4 °C in PBS for processing at a later time • Experimental replicates are easy to obtain • Easier to visualize radical changes in protein expression in multiple samples • Fixation preserves post-translational modifications • Error-prone steps such as cell lysis, gel electrophoresis, and membrane transfer are eliminated • Less time-consuming than western blot (many samples in parallel) • Performed in situ, relevant to cellular context	• No molecular weight–based separation step; thus, antibody specificity is critical. • Optimization for each antibody may be necessary • More concentrated primary antibodies are required • Slight changes in protein expression may not be detected • A dual-fluorescent imaging system, such as the Odyssey LI-COR Imaging System is required

The protocol described here is validated for one protein; therefore, optimization of various parameters may be necessary to achieve study-specific goals, described in **Table 4.2**.

Table 4.2. Parameters at specific steps of ICW that can be optimized to achieve high-throughput selection of single clones for study-specific goals.

Steps in protocol	Parameters to be optimized
Antibody Optimization	• Specificity is typically better if antibody is ChIP grade and reported in immunofluorescence studies • Incubation time overnight or >2.5h • Concentration of antibody
Cells	• Cell number • Cell type
Choice of plate	• Type of plate • Focus for scanning (3.0 mm to 4.0 mm)
Fixation	• Type of fixative: methanol, formaldehyde, ethanol, acetone • Use the recommended fixative from primary antibody manufacturer • Incubation time • Incubation temperature
Permeabilization	• Type of permeabilizer: methanol, Triton X-100, Saponin • Incubation time

4.3 Materials and Reagents

1. 96-well clear flat-bottom polystyrene tissue-culture plates (Corning, catalog number: 3596)
2. 15 mL Falcon tubes (Corning, Falcon®, catalog number: 352097)
3. Reagent reservoir nonsterile (VWR, catalog number: 89094-684)
4. 100% methanol (Thermo Fisher, catalog number: A412-20), storage: room temperature
5. Distilled water, storage: room temperature
6. NaCl (Sigma Aldrich, catalog number: S3014-1KG), storage: room temperature
7. KCl (Sigma Aldrich, catalog number: P9333-1KG), storage: room temperature
8. KH2PO4 (Sigma Aldrich, catalog number: P9791-500G), storage: room temperature

9. Na2HPO4·2H2O (Sigma Aldrich, catalog number: 71643-1KG), storage: room temperature
10. Triton X (Sigma Aldrich, catalog number: X100), storage: room temperature
11. Tween-20 (Sigma Aldrich, catalog number: P9416), storage: room temperature
12. Primary Antibodies (varies), storage: either 4 °C or -20 °C, depending on the antibody
13. LI-COR Secondary Antibodies (varies), storage: 4 °C
14. 1× PBS (see Recipes)
15. Permeabilizing buffer (0.2% Triton X in 1× PBS) (see Recipes)
16. 1× Phosphate Buffered Saline Tween-20 (PBST) (see Recipes)

4.4 Equipment

1. LI-COR Blocking Buffer (LI-COR, Odyssey, catalog number: 927-40003)
2. Multichannel pipettes (Mettler-Toledo International, catalog number: L12-200XLS)
3. FinnpipetteTM Novus multichannel pipette (Thermo Fisher Scientific, Thermo ScientificTM, catalog number: 46300800)

4.5 Software

1. Image Studio Lite (LI-COR Biosciences, https://www.licor.com/bio/image-studio/)

4.6 Procedure

1. Plate 2,000 to 10,000 cells in single wells or duplicates in a 96-well plate, in appropriate culture medium, for 24 to 48 h prior to performing ICW. *Note: Include appropriate positive control cells and blank wells containing only medium for every assay. Cell number may need to be optimized for the cell type.*
2. Chill 100% methanol at -20 °C for 15 min.
3. Remove media by flicking plate, then using a multichannel pipette to remove any residual media. *Note: It was determined that flicking the plate on a stack of paper towels (turning*

the plate upside down rapidly and tapping gently) to decant the media was preferred over mechanical removal. Fewer cells were disrupted. This is especially true for less adherent cells.

4. Add 150 µL of methanol to each of the 96 wells and incubate for 20 min at 4 °C (without shaking). *Notes:*
    a. *Methanol should be added very gently down the walls of each of the wells.*
    b. *Be cautious that wells do not dry out throughout the process. A manual multichannel pipette will suffice for few wells in a 96-well plate. However, an electronic multichannel pipette is advisable for a high-throughput experiment.*
5. Flick plate to remove methanol, then using the multichannel pipette remove any residual methanol.
6. Permeabilize cells using 150 µL of permeabilizing buffer with gentle shaking at room temperature for 30 min.
7. Flick plate to remove permeabilizing buffer, then using the multichannel pipette, remove any residual permeabilizing buffer.
8. Block for 1.5 h using 50 µL LI-COR blocking buffer with gentle shaking at room temperature. *Note: At this point, the plate can be stored at 4 °C overnight. Plates stored for longer than 4 days need to be checked for bacterial growth.*
9. Flick plate to remove blocking buffer, then using the multichannel pipette, remove any residual buffer.
10. Add 50 µL of 1:50 to 1:500 primary antibody and incubate overnight at 4 °C with gentle shaking. *Note: It is advised to use antibodies that are validated for ChIP or immunofluorescence. Each antibody will need to be optimized for concentration; typically, 1:50–1:500 dilutions are used due to differences in antibody affinity.*
11. Carefully remove antibody. This can be accomplished by either flicking the plate or using the multichannel pipette if the antibody will be saved. For the latter, care should be taken to avoid disrupting the cells on the bottom of the well with the pipette tips. To avoid disrupting the cell monolayer, angle the plate at ~45° and position the pipette tips at the

interface of the wall of the well and the bottom of the well. Slowly pipette the antibody solution into the pipette. *Note: Some primary antibodies can be saved and reused if desired. However, the number of reuses has to be determined for each antibody separately.*

12. Add 150 µL of PBST to each well using the multichannel pipette and place plate on a shaker for 5 min at room temperature.
13. Remove PBST by flicking plate.
14. Repeat steps 12 and 13 for a total of five times.
15. Add 50 µL of 1:800 secondary antibody, cover plate with aluminum foil to protect antibody from light and incubate 1 h at room temperature with gentle shaking. *Note: Due to the sensitivity of the fluorescent antibodies to light, be sure to keep plate covered with aluminum foil after this step.*
16. Carefully remove antibody. See step 11.
17. Add 150 µL of PBST to each well using the multichannel pipette and place plate on a shaker for 5 min at room temperature.
18. Carefully remove the wash buffer by flicking plate.
19. Repeat steps 17 and 18 for a total of five times.
20. Perform one wash with PBS for 5 min while shaking the plate.
21. Flick plate to remove PBS, blot the plate, clean the bottom of the plate, and scan. *Note: At this point, since the plate has yet to be blotted for the endogenous control, it is necessary to ensure a layer of PBS is maintained in the plate to prevent it from drying out.*
22. In order to scan the plate on the LI-COR Biosciences software, select the "plate" setting and focus at "4.0 mm" along with resolution and quality according to preference. *Note: Focus set to 4.0 mm usually works best. However, this setting may need to be adjusted based on the manufacturer of the plate used. See Figures 4.1 and 4.2 below for location of plate and software settings.*
23. Reblot the plate with the endogenous control antibody at a dilution of 1:500 to 1:1,000 either for 2.5 h at room temperature or overnight at 4 °C.

24. Carefully remove antibody. See step 11.
25. Wash each well by adding 150 μL of PBST using the multichannel pipette and incubate the plate on a shaker for 5 min at room temperature.
26. Repeat step 25 for a total of five times.
27. Add 50 μL of 1:800 secondary antibody to each well, cover the plate with aluminum foil, and incubate 1 h at room temperature with gentle shaking.
28. Carefully remove antibody.
29. Wash each well by adding 150 μL of PBST using the multichannel pipette and incubate the plate on a shaker for 5 min at room temperature.
30. Repeat step 29 for a total of five times.
31. Perform one wash with PBS for 5 min while shaking the plate.
32. Flick plate to remove PBS and blot the plate dry before the final scan.
33. Scan the plate using the suggested starting Odyssey scan parameters:
    - Resolution: 84 μM
    - Quality: Medium
    - Focus offset: 4.0 mm
    - Intensity: Adjust as necessary such that signal is evident for positive samples but not for negative controls (i.e., empty wells or cells not incubated with the primary antibody). *Note: See Figure 4.2 from step 22 for the setting parameters*.
34. The specific plate format selected under the "analysis settings" tab of the software will create a grid that can be adjusted to match the wells on the scanned image for further analysis (see Figure 4.3 below). *Note: It is important to place the plate parallel to the scale markings on the scanner in order to align the plate template onto the scanned image*.

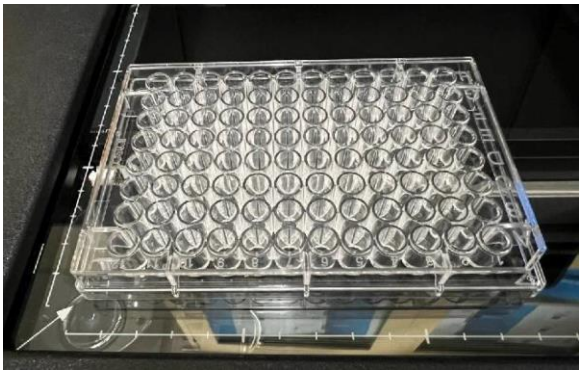

Figure 4.1. **Plate location on LI-COR, in this case positioned in the bottom left corner.**

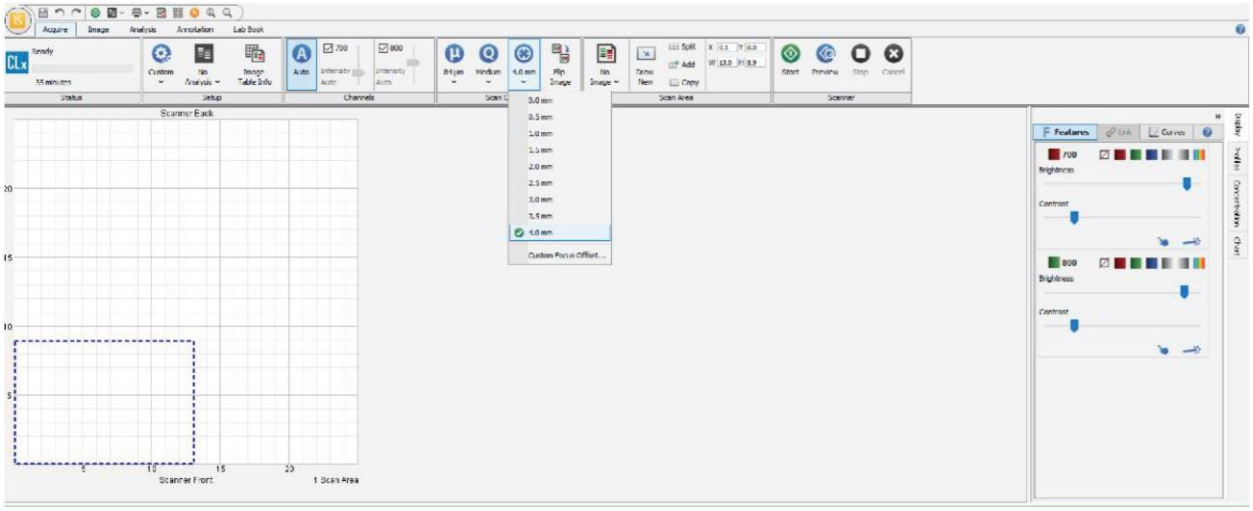

Figure 4.2. Plate settings on LI-COR, including selected region for scanning based on the location of the plate.

Examples of using the ICW protocol to identify clones knocked out for a particular gene are shown in **Figures 4.4–4.6**. In this case, Cas9 targeting KMT5C was transfected into cells, and individual clones were isolated and propagated. Using the ICW protocol, expanded clones were then evaluated for the downstream histone modification generated of KMT5C, histone 4 lysine 20 trimethylation (H4K20me3). Clones with variability in KMT5C activity were identified using the ICW protocol **(Figure 4.4)** and were confirmed via western blotting **(Figure 4.5)**. In a similar way, clones overexpressing a doxycycline-inducible KMT5C were screened using the ICW protocol **(Figure 4.6).**

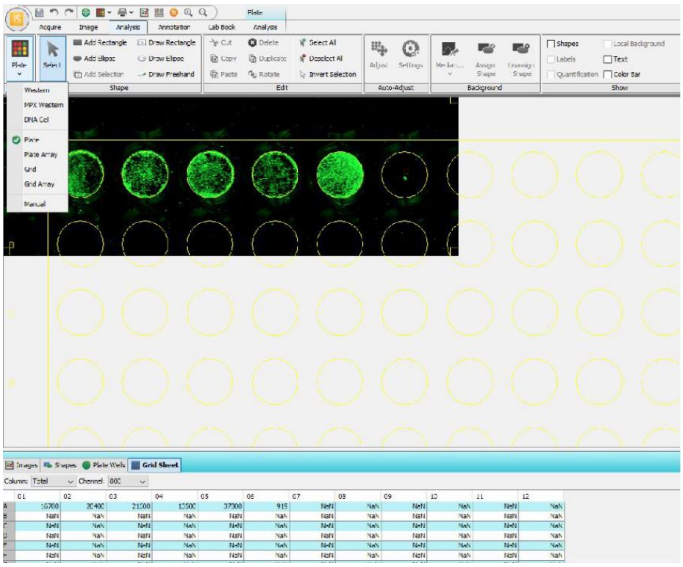

Figure 4.3. Image depicting the analysis setting options on the LI-COR software, including quantification of signal intensity in each identified well in the table at the bottom.

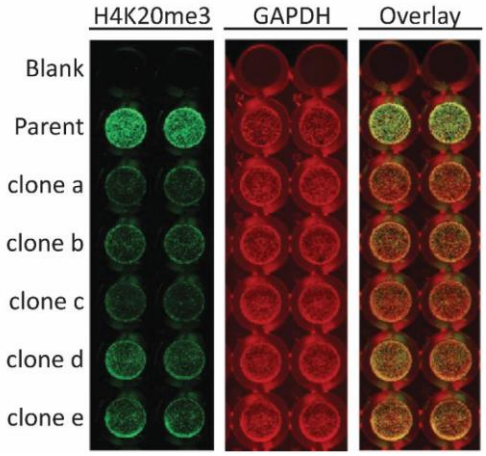

Figure 4.4. Selection of single clones knocked out for KMT5C by quantifying a downstream effector, H4K20me3 mark via ICW.

Parent cell or KO clones (clones a-e) were plated in duplicates in a 96-well plate (10,000 cells/well). 48 h post-plating, cells were blocked using LI-COR blocking buffer, permeabilized using permeabilizing buffer, incubated with 1:400 H4K20me3 (Active Motif) primary antibody overnight on a shaker at 4 °C, and detected using anti-mouse LI-COR secondary antibodies. Then, the plate was scanned and re-blotted overnight on a shaker at 4 °C using a 1:500 concentration of GAPDH (Cell Signaling) primary antibody and detected using anti-rabbit LI-COR secondary antibodies.

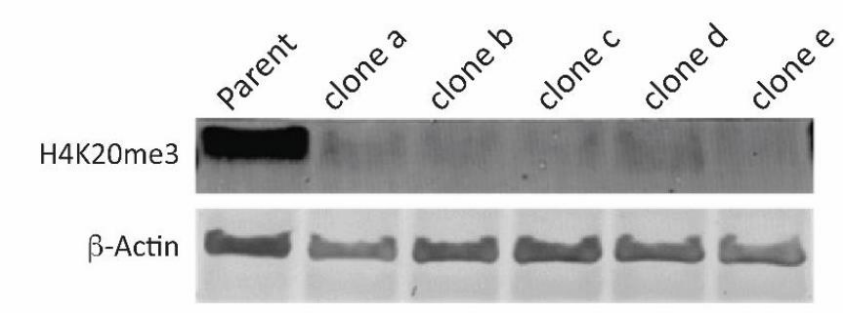

Figure 4.5. Clones identified through ICW, validated via western blot.

Parent cells or KO clones (Clones a-e) were plated in a 6-well plate at 4×10^5 cells/well. 48 h post-plating, lysates were isolated, quantified, and separated using polyacrylamide gel electrophoresis. Post-transfer onto a PVDF membrane, the membrane was blocked using LI-COR blocking buffer and incubated overnight in 1:500 H4K20me3 antibody or in 1:10,000 β-ACTIN (Cell Signaling) primary antibody overnight on a shaker at 4 °C and detected using anti-mouse or anti-rabbit LI-COR secondary antibodies, respectively.

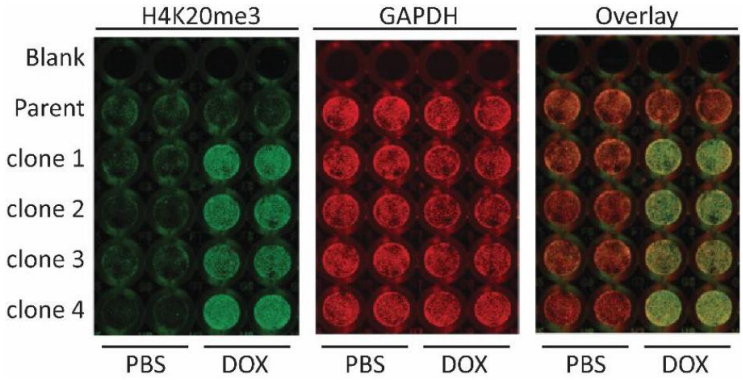

Figure 4.6. Single clones selected post-doxycycline mediated induction of KMT5C via H4K20me3 quantification by ICW.

Either parental cells or doxycycline inducible KMT5C-single clones (clones 1–4) were plated at 10,000 cells/well in replicates of four in a 96-well plate. Doxycycline (2 μg/mL) or equivalent volume of PBS was added to two wells for each cell line at the time of plating. 48 h post-treatment, cells were blocked using LI-COR buffer, permeabilized using permeabilizing buffer, incubated with 1:400 H4K20me3 (Active Motif) primary antibody overnight on a shaker at 4 °C, and detected using anti-mouse LI-COR secondary antibodies. Then, the plate was scanned and

reblotted overnight on a shaker at 4 °C in 1:500 GAPDH (Cell Signaling) primary antibody and detected using anti-rabbit LI-COR secondary antibodies.

4.7 Notes

1. During screening of single clones, cells can be plated in single wells, without counting. However, calculating relative signal of protein of interest to that of the positive control, post-normalizing to signal of the endogenous control is necessary.
2. Scanning overnight dried plates (in the dark) after blotting for both the protein of interest and the endogenous control can yield more uniform and sharper signals.
3. We recommend validating the candidate single clones identified through ICW via other protein quantification techniques such as western blot or immunofluorescence.

4.8 Recipes

1. 1× PBS
   800 g NaCl
   20 g KCl
   144 g Na2HPO4·2H2O
   24 g KH2PO4
   8 L of distilled water

2. Permeabilizing buffer (0.2% Triton X in 1× PBS)
   50 mL 1× PBS
   100 μL Triton X

3. 3. 1× PBST
   1 L of 1× PBS
   1 mL Tween-20

CHAPTER 5. DISCUSSION AND FUTURE DIRECTIONS.

5.1 Chapter overview:

In this chapter, we discuss alternative mechanisms that contribute to erlotinib resistance when KMT5C is loss in cancer. Furthermore, we explore the potential involvement of the constitutive heterochromatin formation pathway in driving resistance to drugs like tyrosine kinase inhibitors. Future directions insights of this project are added to this discussion as well.

5.2 The contribution of heterochromatin formation pathway to erlotinib resistance.

In our endeavor to identify tumor suppressor genes associated with erlotinib resistance, we conducted a CRISPR-Cas9 screen, which revealed KMT5C as the most significant finding. KMT5C is a methyltransferase responsible for tri-methylating histone 4 lysine 20 (H4K20me3), a methylation mark frequently observed in heterochromatic regions where chromatin is densely compacted, thus impeding active transcription. Another commonly found methylation mark in heterochromatic regions is H3K9me3, which has been linked to H4K20me3 through HP1 binding. Given the co-localization of H3K9me3 and H4K20me3 in genomic regions, we explored whether the H3K9me3 pathway was also found amongst the significant hits from the CRISPR-Cas9 screen. We revisited the top hits database and examined whether the methyltransferase responsible for H3K9me3 (SUV39H1) or the HP1 protein were among the significant hits, thus supporting our hypothesis.

For reference, our top hit, KMT5C, attained a beta score of 96.85, indicating a high enrichment of sgRNAs targeting KMT5C in erlotinib-resistant cells compared to the baseline. However, SUV39H1 exhibited a considerably low beta score of 2.29 and a non-significant p-value of 0.2. Similarly, HP1 (also known as CBX5) achieved a beta score of 2.86 and a non-significant p-value of 0.17. Additionally, KMT5B (also known as SUV420H1) sgRNAs demonstrated enrichment in erlotinib-resistant cells compared to the baseline, with a beta score of 90, although the p-value was not significant[13]. These findings could be attributed to the inherent sensitivity of the screening method, and further confirmation in other cell lines would help elucidate the role of KMT5C in erlotinib resistance. It is worth noting that a recent report indicated the involvement of HP1 in

EGFRi resistance in NSCLC, further supporting the relevance of our findings. In their study they found that HP1 (also known as CBX5) loss in EGFR-mutant in LUAD cells, leads to increased expression of the transcription factor E2F1, which then stimulates expression of the antiapoptotic gene BIRC5 (*survivin*). Restoring HP1 expression with a bromo- and extra-terminal (BET)-domain inhibitor (BETi) sensitized them to EGFRi/BETi combination therapy. Similarly, chemical inhibition of the gene BIRC5 was shown to suppress growth of EGFRi-resistant LUAD cells[174]. This finding suggests that reinstating HP1's control over gene transcription could be advantageous for EGFRi resistance therapy. However, it remains to be investigated whether this effect is dependent on H3K9me3, H4K20me3, or both. To explore the potential of KMT5C as a therapeutic target, a screen utilizing small molecule inhibitors targeting various epigenetic regulators from the Structural Genome Consortium[174] could be conducted. This would enable the identification of a suitable candidate capable of restoring KMT5C expression. Similarly, the combination of EGFRi with MET inhibitors holds promise and may potentially enhance the sensitivity of EGFR mutant cells. However, if MET inhibitors fail to sensitize KMT5C mutant cells, it suggests that KMT5C mutation drives erlotinib resistance partially through MET overexpression, but likely involves the regulation of other targets or changes in chromatin structure.

5.3 Alternative mechanisms of resistance in KMT5C mutated cells

5.3.1 Candidate genes regulated by H4K20me3 and their role in resistance.

The resistance phenotype observed in EGFR mutant cells following the loss of KMT5C could potentially be explained by the upregulation of MET. However, it is likely that other genes are also involved and contribute to the resistance mechanism. To identify these genes, it would be beneficial to perform Chromatin Immunoprecipitation (ChIP) or CUT&RUN sequencing in conjunction with RNA sequencing studies. By doing so, we can pinpoint genes that exhibit a loss of H4K20me3 at their genomic loci and show increased transcript levels in KMT5C-mutant cells compared to KMT5C-WT cells, thus establishing them as potential candidates. This approach would enable the identification of genuine oncogenic targets influenced by H4K20me3 and KMT5C in this specific biological context.

MKK3 as a potential candidate target involved in TKI resistance.

Furthermore, utilizing the NCI Cell Miner Database, we identified several candidate genes previously implicated in NSCLC and EGFRi resistance that displayed a negative correlation with KMT5C. Among these genes, MKK3 or MAP2K3 emerged as one of the top candidates negatively associated with KMT5C. Available ChIP data from IMR90 cells[73] suggest that MKK3 is regulated by H4K20me3, and notably, we observed increased protein expression of MKK3 in KMT5C-mutant cells compared to KMT5C-WT cells. To confirm that MKK3 is indeed regulated by H4K20me3 and KMT5C expression, additional experiments would involve measuring MKK3 transcript levels and conducting ChIP qPCR for the specific genomic region of interest in both KMT5C-WT and KMT5C-mutant cells, similar to our studies with MET.

MKK3, a member of the mitogen-activated protein kinase kinase (MAPKK) family, plays a crucial role as a cellular signal transducer in response to extracellular stress. It functions by phosphorylating and activating the MAPK p38, which is involved in regulating cell cycle checkpoints, cell differentiation, and cell survival. In the context of non-small cell lung cancer (NSCLC), it has been observed that cells resistant to the second-generation TKI gefitinib induce tetraploidization through the activation of YAP-MKK3/6-p38 MAPK-STAT3 signaling[135]. This finding suggests that MKK3 might contribute to resistance by engaging the MAPK-STAT3 signaling pathway. Therefore, in our KMT5C-mutant cells, it is plausible that the overexpression of MKK3 leads to enhanced MAPK-STAT3 signaling, thereby contributing to TKI resistance. However, our current studies only demonstrate an increase in MKK3 protein levels, and further investigation into the levels of phosphorylated MKK3 in KMT5C-mutant cells is necessary to validate this hypothesis.

Notably, the induction of tetraploidy in cells has been associated with uncontrolled cancer cell growth and has been identified in 37% of early-stage cancers. Furthermore, tetraploid cells have been linked to multidrug resistance and poor cancer prognosis. Importantly, tetraploidization is known to contribute to genomic instability, a recognized hallmark of cancer. In line with this, our preliminary data from laboratory experiments indicate an increase in genomic instability in KMT5C-mutant cells, as evidenced by an elevated presence of DNA breaks detected using a

Comet Assay. It is conceivable, therefore, that KMT5C-mutant cells exhibit genomic instability due to the heightened MKK3 signaling, which can drive tetraploidization.

Stem cell markers as potential targets involved in TKI resistance.

As mentioned previously, EGFR inhibitor resistance can occur through various well-described mechanisms, such as cell transformation or epithelial to mesenchymal (EMT) transition. For instance, a study conducted on NSCLC cell lines demonstrated that gefitinib-resistant cell lines exhibit stem cell-like properties compared to their parental cells. Additionally, these TKI-resistant cell lines also displayed EMT features[185]. However, the exact development of stem cell-like properties in resistance cells remains poorly understood. Recent reports have shed light on the role of H4K20me3 and KMT5C regulation in embryonic stem cells, indicating that a genome-wide loss of H4K20me3 leads to dysregulated gene expression and delayed differentiation in embryonic stem cells. Transcriptome data analysis of KMT5C-WT, KMT5C-knockout (KO), and KMT5B/C-KO cells revealed enriched expression of undifferentiated ES cell genes in KMT5C-KO compared to KMT5C-WT cells, suggesting that KMT5C plays a vital role in regulating the transcriptional landscape during differentiation[41]. Since stem cell-like properties are likely involved in resistance, we hypothesized that loss of KMT5C, known to delay differentiation, could potentially affect the transcriptional regulation of differentiation genes, leading to resistance. To test this hypothesis, we assessed the expression of EMT markers (such as Snail and Vimentin) and stem cell markers (such as CD44 and OCT-4). Preliminary data (not shown here) demonstrated increased protein levels of Vimentin in KMT5C-mutant cells compared to KMT5C-WT cells. However, other stem and EMT markers did not exhibit increased expression in KMT5C-mutant cells. Transcriptional upregulation was not observed for Vimentin suggesting that increased protein expression of Vimentin might not be regulated by H4K20me3, but rather could be an indirect effect of clonal selection. Further examination of other genes and transcriptome analysis in our cell lines is necessary to gain a better understanding of the role of EMT or stem cell-like pathways in regulating and contributing to EGFR inhibitor resistance when KMT5C is mutated.

5.3.2 DNA repair pathway and H4K20me3

Early studies in mice have provided valuable insights into the significance of H4K20me2 and H4K20me3 in DNA repair. The knockout of both KMT5B and KMT5C in mice leads to a genome-wide shift to an H4K20me1 state, resulting in less efficient DNA double-strand break (DSB) repair and chromosomal aberrations. This impact on chromatin structure is so critical that KMT5B and KMT5C double knockout mice do not survive beyond the perinatal period[23]. Additionally, structural studies have confirmed the association between the DNA damage response factor 53BP1 and H4K20me2, and to a lesser extent, H4K20me1. However, structural predictions suggest that 53BP1 interacts with H4K20me2 but not with H4K20me3. DNA repair proteins directly recognize H4K20me2, enabling their targeting to DSBs[68]. Furthermore, recent studies have shed further light on this process, revealing that KMT5A, the H4K20 mono-methyltransferase responsible for H4K20me1, facilitates the recruitment and catalysis of KMT5B methyltransferase, which generates H4K20me2 essential for 53BP1 binding and DSB repair[70]. Moreover, H4K20me1 has been identified as a determinant of Non-Homologous End Joining (NHEJ)-directed repair in the DSB repair pathway[70]. The significance of KMT5B and KMT5C in NHEJ repair was further validated using the KMT5B/C inhibitor A-196, which significantly inhibited NHEJ repair while having no impact on HDR-mediated repair[71]. These findings suggest that KMT5B/C plays a crucial role in 53BP1 recruitment and efficient DNA repair.

Furthermore, recent studies have begun unraveling the importance of DNA repair pathways in resistance to EGFR inhibitors, including osimertinib. Impaired DNA damage repair pathways were observed in osimertinib-resistant cell lines, as evidenced by increased levels of the DNA damage marker γH2AX and a greater intensity of the comet tail after treatment with the DNA damage agent cisplatin[186]. Therefore, it is plausible that the loss of KMT5C reduces DNA repair pathways, such as NHEJ-mediated DNA repair, potentially contributing to the observed resistance phenotype. However, the precise role of H4K20me3 alone in DNA repair pathways is less clear compared to H4K20me2. Consequently, there is a need for further evaluation of the DNA repair response in KMT5C-mutant cell lines to determine the essentiality of this pathway in the resistance phenotype.

5.3.3 Chromatin structure's role in resistance.

Telomeric length implication when KMT5C is lost.

H4K20me3 has been implicated in the regulation of telomere length during reprogramming from somatic cells to induced pluripotent stem (iPS) cells[82]. In the context of mouse embryonic fibroblasts, iPS cells have been found to exhibit telomere elongation during reprogramming[81]. Additionally, the telomeric chromatin undergoes dynamic changes depending on the differentiation stage of the cells. Notably, a decrease in H4K20me3 density has been observed at telomeric repeats in iPS cells compared to parental cells. Furthermore, simultaneous inhibition of KMT5B and KMT5C has been shown to promote increased cell growth rate and elongated telomeres[81,82]. This is concomitant with the fact that the removal of multilayered marks of epigenetic silencing, including histone and DNA methylation, is crucial for the de-differentiation process[81]. It is possible that cells that lose KMT5C benefit from dynamic changes in telomere length, thereby promoting their survival by transitioning to a more de-differentiated state. However, the specific role of KMT5C loss in affecting telomeric length in our model requires further examination.

Furthermore, studies on advanced non-small cell lung cancer (NSCLC) patient samples have reported shorter telomeres in comparison to early-stage NSCLC[187]. These tumor samples also exhibit high levels of telomerase and abnormal expression of shelterin genes, which correlates with shorter survival. This finding contradicts the notion that telomere length is associated with a more malignant phenotype. Therefore, additional research is necessary to determine whether telomeric length and the role of H4K20me3 regulation in this context have an impact on survival and resistance to EGFR inhibitors.

Chromatin structure and role of other methylation marks

Nevertheless, it is crucial to consider other significant factors that contribute to tumorigenesis, such as the overall chromatin structure beyond just telomeres, as well as the interaction between H4K20me3 and other histone methylations. H4K20me3 plays a vital role in the regulation of repetitive element silencing, and any disruption in this process can have a profound impact on chromatin interactions and overall structure[73]. Moreover, epigenetic changes have been linked to

drug resistance, including resistance to chemotherapy[188]. Therefore, it is possible that alterations in KMT5C-mediated silencing of repeated elements directly affect the overall condensation of chromatin structure, providing cells with greater plasticity to adapt to adverse conditions like drug treatments.

Furthermore, studies conducted on embryonic stem (ES) cells have revealed a novel interaction between H4K20me3 and other methylation marks, such as H3K4me3 and H3K36me3, at repetitive LINE and LTR elements[79]. Both of these marks are associated with transcriptional activation, and indeed, the presence of H4K20me3 alongside H3K4me3 and H4K20me3 alongside H3K36me3 is associated with active transcription in ES cells[79]. Therefore, it is plausible that altering endogenous levels of H4K20me3 could disrupt the interaction with activating histone methylation marks, thereby disturbing the homeostasis either by increasing transcription at those specific regions or by not affecting them. Further research is required to comprehend the role of this silencing methylation mark in regions of the chromosome where active transcription occurs. Similarly, the impact of changes in these regions on resistance to EGFR inhibitors remains to be elucidated.

5.3.4 hHR23 potential role in resistance.

In the introduction, the advantages and disadvantages of targeting epigenetic factors as cancer therapies were discussed. One significant drawback is the potential for adverse effects due to the lack of specificity when modulating these key epigenetic players. However, regulating the upstream regulators of these epigenetic factors holds potential therapeutic benefits. For instance, the yeast RAD23B homologue (hHR23B) serves as an upstream regulator of H4K20me3 and has been demonstrated to demethylate H4K20me2/3 in human-derived cells[53]. Given that cancer often leads to a loss of H4K20me3, restoring endogenous H4K20me3 levels could be advantageous for cancer therapeutics. Therefore, inhibiting hHR23B with a small molecule inhibitor may prove beneficial for patients with reduced levels of *KMT5C*. Moreover, reports indicate that upregulation of hHR23B expression enhances DNA repair activity and cell survival in the NSCLC A549 cell line following treatment with the DNA damage drug cisplatin[189]. Consequently, inhibiting hHR23B could potentially increase sensitivity to cisplatin in this specific cellular context. Thus, inhibiting hHR23B might offer benefits by indirectly upregulating H4K20me3 levels while decreasing DNA repair, rendering cancer cells more susceptible to drug therapies.

5.3.5 Contribution of KMT5C interacting protein RB1 in resistance.

The loss of the methyltransferase enzyme KMT5C, responsible for H4K20me3 methylation, can potentially impact the function and regulation of proteins that interact with KMT5C. One such protein is Retinoblastoma Protein 1 (RB1), which interacts with both KMT5B and KMT5C, playing a role in stabilizing H4K20me3 at constitutive heterochromatin[36]. Consequently, the RB1 family is known to control H4K20me3 and maintain overall chromatin structure, aligning with its role as a tumor suppressor. Additionally, studies involving mouse embryonic fibroblasts deficient in RB1, RBL1, and RBL2 demonstrated increased genomic instability and decreased H4K20me3 levels, highlighting the impact of RB1 protein function on H4K20me3 regulation[36].

Furthermore, the RB family of proteins plays a significant role in lung cancer and resistance to tyrosine kinase inhibitors (TKIs). Analysis of tumor samples from EGFR mutant patients who transitioned from non-small cell lung cancer (NSCLC) to small cell lung cancer (SCLC), a phenotype associated with TKI resistance, revealed RB loss in 100% of SCLC-transformed cases but rarely in those that remained NSCLC[110]. This indicates a global loss of the RB tumor suppressor protein in resistant cancers that adopt SCLC characteristics. Another study focusing on SCLC-transformed tumor samples resistant to EGFR inhibitors identified mutations in RB1 and TP53, both well-known tumor suppressor genes[190]. Consequently, it is plausible that the loss of KMT5C leads to downregulation of H4K20me3, impacting RB1 function and contributing to an increased transformed phenotype associated with TKI resistance. However, the precise mechanism through which RB1 mutations drive transformation and resistance in NSCLC remains unclear.

Multiple mechanisms could potentially explain why the loss of the methyltransferase KMT5C leads to resistance to erlotinib and EGFR inhibitors. While one possible explanation involves the transcriptional regulation of oncogenes through H4K20me3, particularly involving the gene MET, we do not consider this to be the primary mechanism underlying acquired resistance. It is plausible that other genes might also be involved, and the overall loss of KMT5C could drive a broader regulation of multiple oncogenes. Additionally, the impact of KMT5C loss on chromatin structure, its interaction with protein partners, and DNA damage may play a significant role. Further investigations utilizing advanced sequencing techniques such as Hi-C seq (chromatin loops and interactions), CUT & RUN (histone post-translational modifications), ChIA-PET (chromatin

interaction analysis by paired-end tag sequencing), and ATAC-seq (open chromatin) could provide valuable insights into the observed phenotype in cells with loss of KMT5C, allowing for a better understanding of the underlying mechanisms.

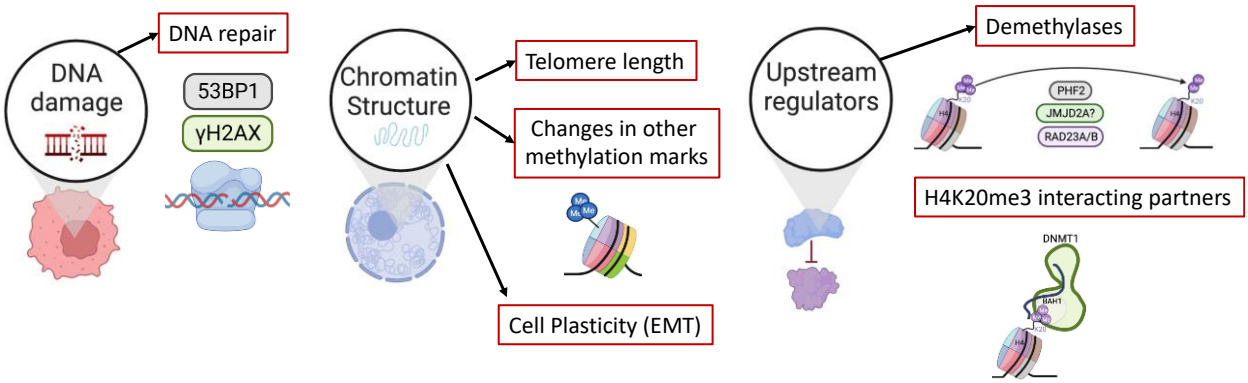

Figure 5.1. Summary of alternative ways KMT5C and H4K20me3 dysregulation can lead to resistance to EGFR inhibitors.

REFERENCES

1. America Cancer Society, A., & Society AC. American Cancer Society: Cancer Facts & Figures. Atlanta: American Cancer Society, 1(1),.

2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature*. 2018;553(7689):446-454. doi:10.1038/nature25183

3. Hsu WH, Yang JCH, Mok TS, Loong HH. Overview of current systemic management of EGFR-mutant NSCLC. *Ann Oncol*. 2018;29(Supplement 1):i3-i9. doi:10.1093/annonc/mdx702

4. Roper N, Brown AL, Wei JS, et al. Clonal Evolution and Heterogeneity of Osimertinib Acquired Resistance Mechanisms in EGFR Mutant Lung Cancer. *Cell Reports Med*. 2020;1(1):100007. doi:10.1016/j.xcrm.2020.100007

5. Liao BC, Griesing S, Yang JCH. Second-line treatment of EGFR T790M-negative non-small cell lung cancer patients. *Ther Adv Med Oncol*. 2019;11:1-16. doi:10.1177/1758835919890286

6. Sos ML, Koker M, Weir BA, et al. PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of akt and EGFR. *Cancer Res*. 2009;69(8):3256-3261. doi:10.1158/0008-5472.CAN-08-4055

7. Huang S, Benavente S, Armstrong EA, Li C, Wheeler DL, Harari PM. P53 modulates acquired resistance to EGFR inhibitors and radiation. *Cancer Res*. 2011;71(22):7071-7079. doi:10.1158/0008-5472.CAN-11-0128

8. Forloni M, Gupta R, Nagarajan A, Politi K, Dogra SK. Oncogenic EGFR Represses the TET1 DNA Demethylase to Induce Silencing of Tumor Suppressors in Cancer Cells. *Cell Rep*. 2016;16:457-471. doi:10.1016/j.celrep.2016.05.087

9. de Bruin EC, Cowell C, Warne PH, et al. Reduced NF1 expression confers resistance to EGFR inhibition in lung cancer. *Cancer Discov*. 2014;4(5):606-619. doi:10.1158/2159-8290.CD-13-0741

10. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - A mechanism for early oncogenic pathway addiction? *Nat Rev Cancer*. 2006;6(2):107-116. doi:10.1038/nrc1799

11. Kelly TK, De Carvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. *Nat Biotechnol*. 2010;28(10):1069-1078. doi:10.1038/nbt.1678

12. Jones PA, Baylin SB. The Epigenomics of Cancer. *Cell*. 2007;128(4):683-692. doi:10.1016/j.cell.2007.01.029

13. Pal AS, Agredo A, Lanman NA, et al. Loss of KMT5C Promotes EGFR Inhibitor Resistance in NSCLC via LINC01510-Mediated Upregulation of MET. *Cancer Res*. 2022:OF1-OF14. doi:10.1158/0008-5472.can-20-0821

14. Greer EL, Shi Y. Histone methylation: A dynamic mark in health, disease and inheritance. *Nat Rev Genet*. 2012;13(5):343-357. doi:10.1038/nrg3173

15. Millán-Zambrano G, Burton A, Bannister AJ, Schneider R. Histone post-translational modifications — cause and consequence of genome function. *Nat Rev Genet*. 2022;0123456789. doi:10.1038/s41576-022-00468-7

16. Saksouk N, Simboeck E, Déjardin J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics and Chromatin*. 2015;8(1):1-17. doi:10.1186/1756-8935-8-3

17. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol*. 2018;19(4):229-244. doi:10.1038/nrm.2017.119

18. Jørgensen S, Schotta G, Sørensen CS. Histone H4 Lysine 20 methylation: Key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res*. 2013;41(5):2797-2806. doi:10.1093/nar/gkt012

19. Fodor BD, Shukeir N, Reuter G, Jenuwein T. Mammalian Su (var) Genes in Chromatin Control . *Annu Rev Cell Dev Biol*. 2010;26(1):471-501. doi:10.1146/annurev.cellbio.042308.113225

20. Grigliatti T. Position-effect variegation—an assay for nonhistone chromosomal proteins and chromatin assembly and modifying factors. *Methods Cell Biol*. 1991;35:587-627.

21. Nishioka K, Rice JC, Sarma K, et al. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent Chromatin. *Mol Cell*. 2002;9(6):1201-1213. doi:10.1016/S1097-2765(02)00548-8

22. Schotta G, Lachner M, Sarma K, et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. 2004;5:1251-1262. doi:10.1101/gad.300704.chromatin

23. Schotta G, Sengupta R, Kubicek S, et al. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev*. 2008;22(15):2048-2061. doi:10.1101/gad.476008

24. Schoeftner S, Blasco MA. A higher order of telomere regulation: Telomere heterochromatin and telomeric RNAs. *EMBO J*. 2009;28(16):2323-2336. doi:10.1038/emboj.2009.197

25. Shen Z, Sathyan KM, Geng Y, et al. A WD-repeat protein stabilizes ORC binding to chromatin. *Mol Cell*. 2010;40(1):99-111. doi:10.1016/j.molcel.2010.09.021

26. Shoaib M, Walter D, Gillespie PJ, et al. Histone H4K20 methylation mediated chromatin compaction threshold ensures genome integrity by limiting DNA replication licensing. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-06066-8

27. Fraga MF, Ballestar E, Villar-Garea A, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet*. 2005;37(4):391-400. doi:10.1038/ng1531

28. Corvalan AZ, Coller HA. Methylation of histone 4's lysine 20: A critical analysis of the state of the field. *Physiol Genomics*. 2021;53(1):22-32. doi:10.1152/physiolgenomics.00128.2020

29. Yang H, Pesavento JJ, Starnes TW, et al. Preferential dimethylation of histone H4 lysine 20 by Suv4-20. *J Biol Chem*. 2008;283(18):12085-12092. doi:10.1074/jbc.M707974200

30. Wu H, Siarheyeva A, Zeng H, et al. Crystal structures of the human histone H4K20 methyltransferases SUV420H1 and SUV420H2. *FEBS Lett*. 2013;587(23):3859-3868. doi:10.1016/j.febslet.2013.10.020

31. Karachentsev D, Sarma K, Reinberg D, Steward R. PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. *Genes Dev*. 2005;19(4):431-435. doi:10.1101/gad.1263005

32. Hahn M, Dambacher S, Dulev S, et al. Suv4-20h2 mediates chromatin compaction and is important for cohesion recruitment to heterochromatin. *Genes Dev*. 2013;27(8):859-872. doi:10.1101/gad.210377.112

33. Vermeulen M, Eberl HC, Matarese F, et al. Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell*. 2010;142(6):967-980. doi:10.1016/j.cell.2010.08.020

34. Bosch-Presegué L, Raurell-Vila H, Thackray JK, et al. Mammalian HP1 Isoforms Have Specific Roles in Heterochromatin Structure and Organization. *Cell Rep*. 2017;21(8):2048-2057. doi:10.1016/j.celrep.2017.10.092

35. Benetti R, Gonzalo S, Jaco I, et al. Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J Cell Biol*. 2007;178(6):925-936. doi:10.1083/jcb.200703081

36. Gonzalo S, García-Cao M, Fraga MF, et al. Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. *Nat Cell Biol*. 2005;7(4):420-428. doi:10.1038/ncb1235

37. Liu Y, Matos RC, Heino TI, Hietakangas V. PWP1 promotes nutrient-responsive expression of 5S ribosomal RNA. *Biol Open*. 2018;7(11):5-8. doi:10.1242/bio.037911

38. Yu Y, Jia W, Lyu Y, et al. Pwp1 regulates telomere length by stabilizing shelterin complex and maintaining histone H4K20 trimethylation. *Cell Discov*. 2019;5(1). doi:10.1038/s41421-019-0116-8

39. Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G. Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet*. 2009;41(6):696-702. doi:10.1038/ng.360

40. Benetti R, García-Cao M, Blasco MA. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat Genet*. 2007;39(2):243-250. doi:10.1038/ng1952

41. Kurup JT, Han Z, Jin W, Kidder BL. H4K20me3 methyltransferase SUV420H2 shapes the chromatin landscape of pluripotent embryonic stem cells. *Dev*. 2020;147(23). doi:10.1242/dev.188516

42. Hagihara Y, Asada S, Maeda T, Nakano T, Yamaguchi S. Tet1 regulates epigenetic remodeling of the pericentromeric heterochromatin and chromocenter organization in DNA hypomethylated cells. *PLoS Genet*. 2021;17(6):1-20. doi:10.1371/journal.pgen.1009646

43. Park J, Lee H, Han N, et al. Long non-coding RNA ChRO1 facilitates ATRX/DAXX-dependent H3.3 deposition for transcription-associated heterochromatin reorganization. *Nucleic Acids Res*. 2018;46(22):11759-11775. doi:10.1093/nar/gky923

44. Stender JD, Pascual G, Liu W, et al. Control of Proinflammatory Gene Programs by Regulated Trimethylation and Demethylation of Histone H4K20. *Mol Cell*. 2012;48(1):28-38. doi:10.1016/j.molcel.2012.07.020

45. Kidder BL, Hu G, Cui K, Zhao K. SMYD5 regulates H4K20me3-marked heterochromatin to safeguard ES cell self-renewal and prevent spurious differentiation. *Epigenetics and Chromatin*. 2017;10(1):1-20. doi:10.1186/s13072-017-0115-7

46. Kidder BL, He R, Wangsa D, et al. SMYD5 controls heterochromatin and chromosome integrity during embryonic stem cell differentiation. *Cancer Res*. 2017;77(23):6729-6745. doi:10.1158/0008-5472.CAN-17-0828

47. Chen YJ, Tsai CH, Wang PY, Teng SC. SMYD3 Promotes Homologous Recombination via Regulation of H3K4-mediated Gene Expression. *Sci Rep*. 2017;7(1):1-12. doi:10.1038/s41598-017-03385-6

48. van Aller GS, Reynoird N, Barbash O, et al. Smyd3 regulates cancer cell phenotypes and catalyzes histone H4 lysine 5 methylation. *Epigenetics*. 2012;7(4):340-343. doi:10.4161/epi.19506

49. Vieira FQ, Costa-Pinheiro P, Almeida-Rios D, et al. SMYD3 contributes to a more aggressive phenotype of prostate cancer and targets Cyclin D2 through H4K20me3. *Oncotarget*. 2015;6(15):13644-13657. doi:10.18632/oncotarget.3767

50. Foreman KW, Brown M, Park F, et al. Structural and functional profiling of the human histone methyltransferase SMYD3. *PLoS One*. 2011;6(7). doi:10.1371/journal.pone.0022290

51. Eid A, Rodriguez-Terrones D, Burton A, Torres-Padilla ME. SUV4-20 activity in the preimplantation mouse embryo controls timely replication. *Genes Dev*. 2016;30(22):2513-2526. doi:10.1101/gad.288969.116

52. Lee J, Thompson JR, Botuyan MV, Mer G. Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat Struct Mol Biol*. 2008;15(1):109-111. doi:10.1038/nsmb1326

53. Cao X, Chen Y, Wu B, et al. Histone H4K20 Demethylation by Two hHR23 Proteins. *Cell Rep*. 2020;30(12):4152-4164.e6. doi:10.1016/j.celrep.2020.03.001

54. Brustel J, Kirstein N, Izard F, et al. Histone H4K20 tri-methylation at late-firing origins ensures timely heterochromatin replication. *EMBO J*. 2017;36(18):2726-2741. doi:10.15252/embj.201796541

55. Ren W, Fan H, Grimm SA, et al. DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. *Nat Commun*. 2021;12(1). doi:10.1038/s41467-021-22665-4

56. Rodríguez-Cortez VC, Martínez-Redondo P, Català-Moll F, et al. Activation-induced cytidine deaminase targets SUV4-20-mediated histone H4K20 trimethylation to class-switch recombination sites. *Sci Rep*. 2017;7(1):1-13. doi:10.1038/s41598-017-07380-9

57. Weirich S, Kudithipudi S, Jeltsch A. Specificity of the SUV4-20H1 and SUV4-20H2 protein lysine methyltransferases and methylation of novel substrates. *J Mol Biol*. 2016;428(11):2344-2358. doi:10.1016/j.jmb.2016.04.015

58. Bierhoff H, Schmitz K, Maass F, Ye J, Grummt I. Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. *Cold Spring Harb Symp Quant Biol*. 2010;75:357-364. doi:10.1101/sqb.2010.75.060

59. Bierhoff H, Dammert MA, Brocks D, Dambacher S, Schotta G, Grummt I. Quiescence-Induced LncRNAs Trigger H4K20 Trimethylation and Transcriptional Silencing. *Mol Cell*. 2014;54(4):675-682. doi:10.1016/j.molcel.2014.03.032

60. Zhao Z, Dammert MA, Grummt I, Bierhoff H. LncRNA-Induced Nucleosome Repositioning Reinforces Transcriptional Repression of rRNA Genes upon Hypotonic Stress. *Cell Rep*. 2016;14(8):1876-1882. doi:10.1016/j.celrep.2016.01.073

61. Kurup JT, Kidder BL. Identification of H4K20me3- and H3K4me3-associated RNAs using CARIP-Seq expands the transcriptional and epigenetic networks of embryonic stem cells. *J Biol Chem*. 2018;293(39):15120-15135. doi:10.1074/jbc.RA118.004974

62. Pesavento JJ, Yang H, Kelleher NL, Mizzen CA. Certain and Progressive Methylation of Histone H4 at Lysine 20 during the Cell Cycle. *Mol Cell Biol*. 2008;28(1):468-486. doi:10.1128/mcb.01517-07

63. Kohi SM, Feng T, Tian Y, Zhu W-G. Biological function and regulation of histone 4 lysine 20 methylation in DNA damage response. *Genome Instab Dis*. 2022;3(1):33-46. doi:10.1007/s42764-022-00063-4

64. Beck DB, Burton A, Oda H, Ziegler-Birling C, Torres-Padilla ME, Reinberg D. The role of PR-Set7 in replication licensing depends on Suv4-20h. *Genes Dev*. 2012;26(23):2580-2589. doi:10.1101/gad.195636.112

65. Abbas T, Shibata E, Park J, Jha S, Karnani N, Dutta A. CRL4Cdt2 regulates cell proliferation and histone gene expression by targeting PR-Set7/Set8 for degradation. *Mol Cell*. 2010;40(1):9-21. doi:10.1016/j.molcel.2010.09.014

66. Liu W, Tanasa B, Tyurina O V., et al. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature*. 2010;466(7305):508-512. doi:10.1038/nature09272

67. Schuh L, Loos C, Pokrovsky D, Imhof A, Rupp RAW, Marr C. H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos. *Cell Syst*. 2020;11(6):653-662.e8. doi:10.1016/j.cels.2020.11.003

68. Botuyan MV, Lee J, Ward IM, et al. Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell*. 2006;127(7):1361-1373. doi:10.1016/j.cell.2006.10.043

69. Hsiao KY, Mizzen CA. Histone H4 deacetylation facilitates 53BP1 DNA damage signaling and double-strand break repair. *J Mol Cell Biol*. 2013;5(3):157-165. doi:10.1093/jmcb/mjs066

70. Tuzon CT, Spektor T, Kong X, et al. Concerted Activities of Distinct H4K20 Methyltransferases at DNA Double-Strand Breaks Regulate 53BP1 Nucleation and NHEJ-Directed Repair. *Cell Rep*. 2014;8(2):430-438. doi:10.1016/j.celrep.2014.06.013

71. Bromberg KD, Mitchell TRH, Upadhyay AK, et al. The SUV4-20 inhibitor A-196 verifies a role for epigenetics in genomic integrity. *Nat Chem Biol*. 2017;13(3):317-324. doi:10.1038/nchembio.2282

72. Di Micco R, Krizhanovsky V, Baker D, d'Adda di Fagagna F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat Rev Mol Cell Biol*. 2021;22(2):75-95. doi:10.1038/s41580-020-00314-w

73. Nelson DM, Jaber-Hijazi F, Cole JJ, et al. Mapping H4K20me3 onto the chromatin landscape of senescent cells indicates a function in control of cell senescence and tumor suppression through preservation of genetic and epigenetic stability. *Genome Biol*. 2016;17(1):1-20. doi:10.1186/s13059-016-1017-x

74. Marescal O, Cheeseman IM. Cellular Mechanisms and Regulation of Quiescence. *Dev Cell*. 2020;55(3):259-271. doi:10.1016/j.devcel.2020.09.029

75. Evertts AG, Manning AL, Wang X, Dyson NJ, Garcia BA, Coller HA. H4K20 methylation regulates quiescence and chromatin compaction. *Mol Biol Cell*. 2013;24(19):3025-3037. doi:10.1091/mbc.E12-07-0529

76. Ng JMY, Vermeulen W, Van der Horst GTJ, et al. A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. *Genes Dev*. 2003;17(13):1630-1645. doi:10.1101/gad.260003

77. Wongtawan T, Taylor JE, Lawson KA, Wilmut I, Pennings S. Histone H4K20me3 and HP1α are late heterochromatin markers in development, but present in undifferentiated embryonic stem cells. *J Cell Sci*. 2011;124(11):1878-1890. doi:10.1242/jcs.080721

78. Burton A, Brochard V, Galan C, et al. Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat Cell Biol*. 2020;22(7):767-778. doi:10.1038/s41556-020-0536-6

79. Xu J, Kidder BL. H4K20me3 co-localizes with activating histone modifications at transcriptionally dynamic regions in embryonic stem cells. *BMC Genomics*. 2018;19(1):1-15. doi:10.1186/s12864-018-4886-4

80. Sarg B, Koutzamani E, Helliger W, Rundquist I, Lindner HH. Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. *J Biol Chem*. 2002;277(42):39195-39201. doi:10.1074/jbc.M205166200

81. Marión RM, Schotta G, Ortega S, Blasco MA. Suv4-20h abrogation enhances telomere elongation during reprogramming and confers a higher tumorigenic potential to iPs cells. *PLoS One*. 2011;6(10):2-11. doi:10.1371/journal.pone.0025680

82. Marion RM, Strati K, Li H, et al. Telomeres Acquire Embryonic Stem Cell Characteristics in Induced Pluripotent Stem Cells. *Cell Stem Cell*. 2009;4(2):141-154. doi:10.1016/j.stem.2008.12.010

83. Van Den Broeck A, Brambilla E, Moro-Sibilot D, et al. Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. *Clin Cancer Res*. 2008;14(22):7237-7245. doi:10.1158/1078-0432.CCR-08-0869

84. Benard A, Goossens-Beumer IJ, van Hoesel AQ, et al. Histone trimethylation at H3K4, H3K9 and H4K20 correlates with patient survival and tumor recurrence in early-stage colon cancer. *BMC Cancer*. 2014;14(1):1-9. doi:10.1186/1471-2407-14-531

85. Schneider AC, Heukamp LC, Rogenhofer S, et al. Global histone H4K20 trimethylation predicts cancer-specific survival in patients with muscle-invasive bladder cancer. *BJU Int*. 2011;108(8 B):290-296. doi:10.1111/j.1464-410X.2011.10203.x

86. Pogribny IP, Ross SA, Tryndyak VP, Pogribna M, Poirier LA, Karpinets T V. Histone H3 lysine 9 and H4 lysine 20 trimethylation and the expression of Suv4-20h2 and Suv-39h1 histone methyltransferases in hepatocarcinogenesis induced by methyl deficiency in rats. *Carcinogenesis*. 2006;27(6):1180-1186. doi:10.1093/carcin/bgi364

87. Tryndyak VP, Kovalchuk O, Pogribny IP. Loss of DNA methylation and histone H4 lysine 20 trimethylation in human breast cancer cells is associated with aberrant expression of DNA methyltransferase 1, Suv4-20h2 histone methyltransferase and methyl-binding proteins. *Cancer Biol Ther*. 2006;5(1):65-70. doi:10.4161/cbt.5.1.2288

88. Shinchi Y, Hieda M, Nishioka Y, et al. SUV420H2 suppresses breast cancer cell invasion through down regulation of the SH2 domain-containing focal adhesion protein tensin-3. *Exp Cell Res*. 2015;334(1):90-99. doi:10.1016/j.yexcr.2015.03.010

89. Isin H, Özgür E, Talu CK, Trabulus DC, Karaçetin D, Gezer U. Impact of histone methyltransferase suv420h2 in breast cancer. *Biomed Reports*. 2020;13(4):1-7. doi:10.3892/br.2020.1336

90. Kapoor-Vazirani P, Kagey JD, Vertino PM. SUV420H2-Mediated H4K20 Trimethylation Enforces RNA Polymerase II Promoter-Proximal Pausing by Blocking hMOF-Dependent H4K16 Acetylation. *Mol Cell Biol*. 2011;31(8):1594-1609. doi:10.1128/mcb.00524-10

91. Piao L, Yuan X, Wang L, et al. Loss of histone H4 lysine 20 trimethylation in osteosarcoma is associated with aberrant expression ofhistone methyltransferase SUV420H2. *Oncol Lett*. 2020;20(4):1-8. doi:10.3892/ol.2020.11887

92. Wu Y, Shi W, Tang T, et al. miR-29a contributes to breast cancer cells epithelial–mesenchymal transition, migration, and invasion via down-regulating histone H4K20 trimethylation through directly targeting SUV420H2. *Cell Death Dis*. 2019;10(3). doi:10.1038/s41419-019-1437-0

93. Herlihy CP, Hahn S, Hermance NM, Crowley EA, Manning AL. Suv420 enrichment at the centromere limits Aurora B localization and function. *J Cell Sci*. 2021;134(15). doi:10.1242/jcs.249763

94. Zhou M, Li Y, Lin S, et al. H3K9me3, H3K36me3, and H4K20me3 Expression Correlates with Patient Outcome in Esophageal Squamous Cell Carcinoma as Epigenetic Markers. *Dig Dis Sci*. 2019;64(8):2147-2157. doi:10.1007/s10620-019-05529-2

95. Viotti M, Wilson C, McCleland M, et al. SUV420H2 is an epigenetic regulator of epithelial/mesenchymal states in pancreatic cancer. *J Cell Biol*. 2018;217(2):763-777. doi:10.1083/jcb.201705031

96. Kumari D, Usdin K. The distribution of repressive histone modifications on silenced FMR1 alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum Mol Genet*. 2010;19(23):4634-4642. doi:10.1093/hmg/ddq394

97. Pedrotti S, Caccia R, Neguembor MV, et al. The Suv420h histone methyltransferases regulate PPAR-γ and energy expenditure in response to environmental stimuli. *Sci Adv*. 2019;5(4). doi:10.1126/sciadv.aav1472

98. Zhao Q, Zhang Z, Rong W, et al. KMT5c modulates adipocyte thermogenesis by regulating Trp53 expression. *Proc Natl Acad Sci U S A*. 2020;117(36):22413-22422. doi:10.1073/pnas.1922548117

99. Berdasco M, Ropero S, Setien F, et al. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc Natl Acad Sci U S A*. 2009;106(51):21830-21835. doi:10.1073/pnas.0906831106

100. Rank G, Cerruti L, Simpson RJ, Moritz RL, Jane SM, Zhao Q. Identification of a PRMT5-dependent repressor complex linked to silencing of human fetal globin gene expression. *Blood*. 2010;116(9):1585-1592. doi:10.1182/blood-2009-10-251116

101. Ramalingam SS, Maitland ML, Frankel P, et al. Carboplatin and paclitaxel in combination with either vorinostat or placebo for first-line therapy of advanced non-small-cell lung cancer. *J Clin Oncol*. 2010;28(1):56-62. doi:10.1200/JCO.2009.24.9094

102. Leszinski G, Gezer U, Siegele B, Stoetzer O, Holdenrieder S. Relevance of histone marks H3K9me3 and H4K20me3 in cancer. *Anticancer Res*. 2012;32(5):2199-2206.

103. Cheng Y, He C, Wang M, et al. Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials. *Signal Transduct Target Ther*. 2019;4(1). doi:10.1038/s41392-019-0095-0

104. Balch C, Fang F, Matei DE, Huang THM, Nephew KP. Minireview: Epigenetic changes in ovarian cancer. *Endocrinology*. 2009;150(9):4003-4011. doi:10.1210/en.2009-0404

105. Lin J, Gilbert J, Rudek MA, et al. A phase I dose-finding study of 5-azacytidine in combination with sodium phenylbutyrate in patients with refractory solid tumors. *Clin Cancer Res*. 2009;15(19):6241-6249. doi:10.1158/1078-0432.CCR-09-0567

106. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7-30. doi:10.3322/caac.21590

107. Luo B, Wing H, Subramanian A, et al. Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci*. 2008;105(51):20380–20385. doi:10.1080/02331888.2012.748768

108. Onitsuka T, Uramoto H, Ono K, et al. Comprehensive molecular analyses of lung adenocarcinoma with regard to the epidermal growth factor Receptor, K-ras, MET, and hepatocyte growth factor status. *J Thorac Oncol*. 2010;5(5):591-596. doi:10.1097/JTO.0b013e3181d0a4db

109. Yun CH, Mengwasser KE, Toms A V., et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A*. 2008;105(6):2070-2075. doi:10.1073/pnas.0709662105

110. Niederst MJ, Sequist L V., Poirier JT, et al. RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. *Nat Commun*. 2015;6:6377. doi:10.1038/ncomms7377

111. Sequist L V., Waltman BA, Dias-Santagata D, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med*. 2011;3(75):75ra26. doi:10.1126/scitranslmed.3002003

112. Pogribny IP, Tryndyak VP, Muskhelishvili L, Rusyn I, Ross SA. Methyl deficiency, alterations in global histone modifications, and carcinogenesis. *J Nutr*. 2007;137(1 Suppl):216S-222S. doi:10.1093/jn/137.1.216S

113. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol*. 2018;(Mid). doi:10.1038/s41580-018-0045-7

114. Baylin SB, Jones PA. Epigenetic determinants of cancer. *Cold Spring Harb Perspect Biol*. 2016;8(9):1-35. doi:10.1101/cshperspect.a019505

115. Pal AS, Bains M, Agredo A, Kasinski AL. Identification of microRNAs that promote erlotinib resistance in non-small cell lung cancer. *Biochem Pharmacol*. 2021;189:114154. doi:10.1016/J.BCP.2020.114154

116. Golden RJ, Chen B, Li T, et al. An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature*. 2017;542(7640):197-202. doi:10.1038/nature21025

117. Liu XS, Zhang F, Irizarry RA, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. 2014;15(12):1-12. doi:10.1186/s13059-014-0554-4

118. LentiGuide-Puro and LentiCRISPRv2. http://www.genome-engineering.org/gecko/. doi:10.1126/science.1247005

119. Rees MG, Seashore-Ludlow B, Cheah JH, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol*. 2016;12(2):109-116. doi:10.1038/nchembio.1986

120. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98-W102. doi:10.1093/nar/gkx247

121. NCI-60 DTP. NCI-60 Screening Methodology | NCI-60 Human Tumor Cell Lines Screen | Discovery & Development Services | Developmental Therapeutics Program (DTP) https://dtp.cancer.gov/discovery_development/nci-60/methodology.htm.

122. Orellana E, Kasinski A. Sulforhodamine B (SRB) Assay in Cell Culture to Investigate Cell Proliferation. *BIO-PROTOCOL*. 2016;6(21):e1984. doi:10.21769/bioprotoc.1984

123. Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet*. 2015;16(5):299-311. doi:10.1038/nrg3899

124. Chen S, Wang Q, Zhou XM, Zhu JP, Li T, Huang M. MicroRNA-27b reverses docetaxel resistance of non-small cell lung carcinoma cells via targeting epithelial growth factor receptor. *Mol Med Rep*. 2016;14(1):949-954. doi:10.3892/mmr.2016.5332

125. Orzáez M, Guevara T, Sancho M, Pérez-Payá E. Intrinsic caspase-8 activation mediates sensitization of erlotinib-resistant tumor cells to erlotinib/cell-cycle inhibitors combination treatment. *Cell Death Dis*. 2012;3(10):1-9. doi:10.1038/cddis.2012.155

126. CtRP v2. Cancer Therapeutics Response Portal https://portals.broadinstitute.org/ctrp.v2.1/.

127. Aprelikova O, Palla J, Hibler B, et al. Silencing of miR-148a in cancer-associated fibroblasts results in WNT10B-mediated stimulation of tumor cell motility. *Oncogene*. 2013;32(27):3246-3253. doi:10.1038/onc.2012.351

128. Kantidakis T, Saponaro M, Mitter R, et al. Mutation of cancer driver MLL2 results in transcription stress and genome instability. *Genes Dev*. 2016;30(4):408-420. doi:10.1101/gad.275453.115

129. Chen Y, Saif Zaman M, Deng G, et al. MicroRNAs 221/222 and genistein-mediated regulation of ARHI tumor suppressor gene in prostate cancer. *Cancer Prev Res*. 2011;4(1):76-86. doi:10.1158/1940-6207.CAPR-10-0167

130. Ryu SW, Yoon J, Yim N, Choi K, Choi C. Downregulation of OPA3 Is Responsible for Transforming Growth Factor-β-Induced Mitochondrial Elongation and F-Actin Rearrangement in Retinal Pigment Epithelial ARPE-19 Cells. *PLoS One*. 2013;8(5):1-9. doi:10.1371/journal.pone.0063495

131. Chekhun VF, Tryndyak VP, Kovalchuk O, Pogribny IP, Lukyanova NY. Epigenetic profiling of multidrug-resistant human MCF-7 breast adenocarcinoma cells reveals novel hyper- and hypomethylated targets. *Mol Cancer Ther*. 2007;6(3):1089-1098. doi:10.1158/1535-7163.mct-06-0663

132. TCGA. The Cancer Genome Atlas Program https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

133. GTEx. GTEx Portal https://gtexportal.org/home/.

134. Nelson DM, Jaber-Hijazi F, Cole JJ, et al. Mapping H4K20me3 onto the chromatin landscape of senescent cells indicates a function in control of cell senescence and tumor suppression through preservation of genetic and epigenetic stability. *Genome Biol*. 2016;17(1):158. doi:10.1186/s13059-016-1017-x

135. Yeung YT, Yin S, Lu B, et al. Losmapimod Overcomes Gefitinib Resistance in Non-small Cell Lung Cancer by Preventing Tetraploidization. *EBioMedicine*. 2018;28:51-61. doi:10.1016/j.ebiom.2018.01.017

136. Tsai MS, Weng SH, Chen HJ, et al. Inhibition of p38 MAPK-dependent excision repair cross-complementing 1 expression decreases the DNA repair capacity to sensitize lung cancer cells to etoposide. *Mol Cancer Ther*. 2012;11(3):561-571. doi:10.1158/1535-7163.MCT-11-0684

137. Tseng SC, Huang YC, Chen HJ, et al. Metformin-mediated downregulation of p38 mitogen-activated protein kinase-dependent excision repair cross-complementing 1 decreases DNA repair capacity and sensitizes human lung cancer cells to paclitaxel. *Biochem Pharmacol*. 2013;85(4):583-594. doi:10.1016/j.bcp.2012.12.001

138. Jakobsen KR, Demuth C, Madsen AT, et al. MET amplification and epithelial-to-mesenchymal transition exist as parallel resistance mechanisms in erlotinib-resistant, EGFR-mutated, NSCLC HCC827 cells. *Oncogenesis*. 2017;6(4):e307. doi:10.1038/oncsis.2017.17

139. Pennacchietti S, Michieli P, Galluzzo M, Mazzone M, Giordano S, Comoglio PM. Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. *Cancer Cell*. 2003;3(4):347-361. doi:10.1016/S1535-6108(03)00085-0

140. Seol DW, Chen Q, Zarnegar R. Transcriptional activation of the Hepatocyte Growth Factor receptor (c-met) gene by its ligand (Hepatocyte Growth Factor) is mediated through AP-1. *Oncogene*. 2000;19(9):1132-1137. doi:10.1038/sj.onc.1203404

141. Li J, Lin C, Zhang T, et al. Long noncoding RNA LINC01510 promotes the growth of colorectal cancer cells by modulating MET expression. *Cancer Cell Int*. 2018;18(1):1-12. doi:10.1186/s12935-018-0503-5

142. Esposito R, Esposito D, Pallante P, Fusco A, Ciccodicola A, Costa V. Oncogenic properties of the antisense lncRNA COMET in BRAF- and RET-driven papillary thyroid carcinomas. *Cancer Res*. 2019;79(9):2124-2135. doi:10.1158/0008-5472.CAN-18-2520

143. Li J, Wei L. Increased expression of LINC01510 predicts poor prognosis and promotes malignant progression in human non-small cell lung cancer. *Biomed Pharmacother*. 2019;109(7):519-529. doi:10.1016/j.biopha.2018.10.136

144. Li Q, Wang X jun, Jin J hong. SOX2-induced upregulation of lncRNA LINC01510 promotes papillary thyroid carcinoma progression by modulating miR-335/SHH and activating Hedgehog pathway. *Biochem Biophys Res Commun*. 2019;520(2):277-283. doi:10.1016/j.bbrc.2019.09.138

145. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science (80-)*. 2017;357:eaal2380. doi:10.1126/science.aal2380

146. Wilting RH, Dannenberg JH. Epigenetic mechanisms in tumorigenesis, tumor cell heterogeneity and drug resistance. *Drug Resist Updat*. 2012;15(1-2):21-38. doi:10.1016/j.drup.2012.01.008

147. Braig M, Lee S, Loddenkemper C, et al. Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature*. 2005;436(7051):660-665. doi:10.1038/nature03841

148. Souza PP, Völkel P, Trinel D, et al. The histone methyltransferase SUV420H2 and heterochromatin proteins HP1 interact but show different dynamic behaviours. *BMC Cell Biol*. 2009;10:1-16. doi:10.1186/1471-2121-10-41

149. Kovaríková AS, Legartová S, Krejcí J, Bártová E. H3K9me3 and H4K20me3 represent the epigenetic landscape for 53BP1 binding to DNA lesions. *Aging (Albany NY)*. 2018;10(10):2585-2605. doi:10.18632/aging.101572

150. Peters AHFM, O'Carroll D, Scherthan H, et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*. 2001;107(3):323-337. doi:10.1016/S0092-8674(01)00542-6

151. Zhang Q, Thakur C, Fu Y, et al. Mdig promotes oncogenic gene expression through antagonizing repressive histone methylation markers. *Theranostics*. 2020;10(2):602-614. doi:10.7150/thno.36220

152. Gillies RJ, Verduzco D, Gatenby RA. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat Rev Cancer*. 2012;12(7):487-493. doi:10.1038/nrc3298

153. Serizawa M, Takahashi T, Yamamoto N, Koh Y. Genomic aberrations associated with erlotinib resistance in non-small cell lung cancer cells. *Anticancer Res*. 2013;33(12):5223-5234.

154. Nahar R, Zhai W, Zhang T, et al. Elucidating the genomic architecture of Asian EGFR-mutant lung adenocarcinoma through multi-region exome sequencing. *Nat Commun*. 2018;9(1):216. doi:10.1038/s41467-017-02584-z

155. Sanders SL, Portoso M, Mata J, Bähler J, Allshire RC, Kouzarides T. Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell*. 2004;119(5):603-614. doi:10.1016/j.cell.2004.11.009

156. Celeste A, Difilippantonio S, Difilippantonio MJ, et al. H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. *Cell*. 2003;114(3):371-383. doi:10.1016/S0092-8674(03)00567-1

157. Shankavaram UT, Varma S, Kane D, et al. CellMiner: A relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics*. 2009;10(1):277. doi:10.1186/1471-2164-10-277

158. Adams BD, Kasinski AL, Slack FJ. Aberrant regulation and function of microRNAs in cancer. *Curr Biol*. 2014;24(16):R762. doi:10.1016/j.cub.2014.06.043

159. Orellana EA, Kasinski AL. Micrornas in cancer: A historical perspective on the path from discovery to therapy. *Cancers (Basel)*. 2015;7(3):1388-1405. doi:10.3390/cancers7030842

160. American Cancer Society. Cancer Statistics Center. http://cancerstatisticscenter.cancer.org.

161. Chansky K, Detterbeck FC, Nicholson AG, et al. The IASLC Lung Cancer Staging Project: External Validation of the Revision of the TNM Stage Groupings in the Eighth Edition of the TNM Classification of Lung Cancer. *J Thorac Oncol*. 2017;12(7):1109-1121. doi:10.1016/j.jtho.2017.04.011

162. Leonetti A, Sharma S, Minari R, Perego P, Giovannetti E, Tiseo M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br J Cancer*. 2019;121(9):725-737. doi:10.1038/s41416-019-0573-8

163. Demuth C, Madsen AT, Weber B, Wu L, Meldgaard P, Sorensen BS. The T790M resistance mutation in EGFR is only found in cfDNA from erlotinib-treated NSCLC patients that harbored an activating EGFR mutation before treatment. *BMC Cancer*. 2018;18(1):1-5. doi:10.1186/s12885-018-4108-0

164. de Bruin EC, Cowell C, Warne PH, et al. Reduced NF1 expression confers resistance to EGFR inhibition in lung cancer. *Cancer Discov*. 2014;4(5):606-619. doi:10.1158/2159-8290.CD-13-0741

165. Forloni M, Gupta R, Nagarajan A, et al. Oncogenic EGFR Represses the TET1 DNA Demethylase to Induce Silencing of Tumor Suppressors in Cancer Cells. *Cell Rep*. 2016;16(2):457-471. doi:10.1016/j.celrep.2016.05.087

166. Yokoyama Y, Matsumoto A, Hieda M, et al. Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. *Breast Cancer Res*. 2014;16(3). doi:10.1186/bcr3681

167. Zhao QY, Lei PJ, Zhang X, et al. Global histone modification profiling reveals the epigenomic dynamics during malignant transformation in a Four-Stage breast cancer model. *Clin Epigenetics*. 2016;8(1):1-15. doi:10.1186/s13148-016-0201-x

168. Pal, A.S, Agredo, A, Lanman, N.A, Son, J, Sohal, I.S, Bains, M, Li, C, Clingerman, J, K, Gates, Kasinski A. Loss of KMT5C promotes EGFR inhibitor resistance in NSCLC through upregulation of MET via LINC01510. *Press Cancer Res*. 2022.

169. Cesare AJ, Heaphy CM, O'Sullivan RJ. *Visualization of Telomere Integrity and Function in Vitro and in Vivo Using Immunofluorescence Techniques*. Vol 2015.; 2015. doi:10.1002/0471142956.cy1240s73

170. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*. 2017;6:1-35. doi:10.7554/eLife.21856

171. Wong KM, King DA, Schwartz EK, Herrera RE, Morrison AJ. Retinoblastoma protein regulates carcinogen susceptibility at heterochromatic cancer driver loci. *Life Sci Alliance*. 2022;5(4):1-15. doi:10.26508/LSA.202101134

172. Sharma S V., Lee DY, Li B, et al. A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell*. 2010;141(1):69-80. doi:10.1016/j.cell.2010.02.027

173. Liao S, Davoli T, Leng Y, Li MZ, Xu Q, Elledge SJ. A genetic interaction analysis identifies cancer drivers that modify EGFR dependency. *Genes Dev*. 2017;31(2):184-196. doi:10.1101/gad.291948.116

174. Suresh Bugide, Yvonne J. K. Edwards , Romi Gupta, Michael R. Greenc NW. CBX5 loss drives EGFR inhibitor resistance and results in therapeutically actionable vulnerabilities in lung cancer. *Proc Natl Acad Sci*. 2017;120:2017. doi:10.1073/pnas

175. Barth TK, Imhof A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci*. 2010;35(11):618-626. doi:10.1016/j.tibs.2010.05.006

176. Wickramasekara RN, Stessman HAF. Histone 4 lysine 20 methylation: A case for neurodevelopmental disease. *Biology (Basel)*. 2019;8(1):1-18. doi:10.3390/biology8010011

177. Humphrey SE, Kasinski AL. RNA-guided CRISPR-Cas technologies for genome-scale investigation of disease processes. *J Hematol Oncol*. 2015;8(1):1-9. doi:10.1186/s13045-015-0127-3

178. Szlachta K, Kuscu C, Tufan T, et al. CRISPR knockout screening identifies combinatorial drug targets in pancreatic cancer and models cellular drug response. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-06676-2

179. Li C, Kasinski A. In vivo Cancer Based Functional Genomics. *Trends in Cancer*. 2020;(6(12)):1002-1017. doi:https://doi.org/10.1016/j.trecan.2020.07.004

180. Arnoldo A, Kittanakom S, Heisler LE, et al. A genome scale overexpression screen to reveal drug activity in human cells. *Genome Med*. 2014;6(4):1-16. doi:10.1186/gm549

181. Boveia V, Schutz-Geschwender A. Quantitative Analysis of Signal Transduction with In-Cell Western Immunofluorescence Assays. In: Kurien, B., Scofield, R. (eds) Detection of Blotted Proteins. *Methods Mol Biol*. 2015;1314. doi:https://doi.org/10.1007/978-1-4939-2718-0_13

182. Hoffman GR, Moerke NJ, Hsia M, Shamu CE, Blenis J. A high-throughput, cell-based screening method for siRNA and small molecule inhibitors of mTORC1 signaling using the in cell western technique. *Assay Drug Dev Technol*. 2010;8(2):186-199. doi:10.1089/adt.2009.0213

183. Khoury L, Zalko D, Audebert M. Validation of High-Throughput Genotoxicity Assay Screening Using cH2AX In-CellWestern Assay on HepG2 Cells. *Environ Mol Mutagen*. 2013;(54):737-746. doi:10.1002/em.21817

184. Schnaiter S, Fürst B, Neu J, et al. Screening for MAPK modulators using an in-cell western assay. *Methods Mol Biol*. 2014;1120:121-129. doi:10.1007/978-1-62703-791-4_8

185. Shien K, Toyooka S, Yamamoto H, et al. Acquired resistance to EGFR inhibitors is associated with a manifestation of stem cell-like properties in cancer cells. *Cancer Res*. 2013;73(10):3051-3061. doi:10.1158/0008-5472.CAN-12-4136

186. Liang X mei, Qin Q, Liu B ning, et al. Targeting DNA-PK overcomes acquired resistance to third-generation EGFR-TKI osimertinib in non-small-cell lung cancer. *Acta Pharmacol Sin*. 2021;42(4):648-654. doi:10.1038/s41401-020-00577-1

187. Faugeras E, Véronèse L, Jeannin G, et al. Telomere Status of Advanced Non-Small-Cell Lung Cancer Offers a Novel Promising Prognostic and Predictive Biomarker. *Cancers (Basel)*. 2023;15(1). doi:10.3390/cancers15010290

188. Ansari J, Shackelford RE, El-Osta1 H. Epigenetics in non-small cell lung cancer: From basics to therapeutics. *Transl Lung Cancer Res*. 2016;5(2):155-171. doi:10.21037/tlcr.2016.02.02

189. Shen YH, Chen BR, Cherng SH, et al. Cisplatin transiently up-regulates hHR23 expression through enhanced translational efficiency in A549 adenocarcinoma cells. *Toxicol Lett*. 2011;205(3):341-350. doi:10.1016/j.toxlet.2011.06.028

190. Lu Y, Li Z, Zhu H, Zhao J, Xu M, Gu W. Case report: EGFR-mutant lung adenocarcinoma with the TP53 and RB1 mutations showed resistance to TKI therapy. *Ann Palliat Med*. 2021;10(12):12886-12893. doi:10.21037/apm-21-2016

191. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet*. 2015;16(7):421-433. doi:10.1038/nrg3965

192. Grosswendt S, Rajewsky N. *Essentials of MiRNA-Dependent Control of MRNA Translation and Decay, MiRNA Targeting Principles, and Methods for Target Identification*. Elsevier Inc.; 2017. doi:10.1016/B978-0-12-804402-5.00002-9

193. Helwak A, Tollervey D. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat Protoc*. 2014;9(3):711-728. doi:10.1038/nprot.2014.043

194. Grosswendt S, Filipchyk A, Manzano M, et al. Unambiguous Identification of miRNA: Target site interactions by different types of ligation reactions. *Mol Cell*. 2014;54(6):1042-1054. doi:10.1016/j.molcel.2014.03.049

195. Moore MJ, Scheel TKH, Luna JM, et al. MiRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat Commun*. 2015;6(May). doi:10.1038/ncomms9864

196. Salzman DW, Nakamura K, Nallur S, et al. MiR-34 activity is modulated through 5′-end phosphorylation in response to DNA damage. *Nat Commun*. 2016;7(May 2015):1-9. doi:10.1038/ncomms10954

197. Li KW, Schwer B, Shuman S. Structure-guided mutational analysis of T4 RNA ligase 1. *Rna*. 2006;12(12):2126-2134. doi:10.1261/rna.271706

198. El Omari K, Ren J, Bird LE, et al. Molecular architecture and ligand recognition determinants for T4 RNA ligase. *J Biol Chem*. 2006;281(3):1573-1579. doi:10.1074/jbc.M509658200

199. Belasco JG. All things must pass: Contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat Rev Mol Cell Biol*. 2010;11(7):467-478. doi:10.1038/nrm2917

200. Iwakawa H, Tomari Y. The {Functions} of {MicroRNAs}: {mRNA} {Decay} and {Translational} {Repression}. *Trends Cell Biol*. 2015;25(11):651-665. doi:10.1016/j.tcb.2015.07.011

201. Rna ST, Hutvágner G, Mclachlan J, et al. A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let- Published by : American Association for the Advancement of Science Linked references are available on JSTOR for this article : A Cellular Function for the RNA-Interfer. 2016;293(5531):834-838.

202. Han K, Tjaden B, Lory S. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. *Nat Microbiol*. 2016;2(December 2016). doi:10.1038/nmicrobiol.2016.239

203. Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*. 2013;153(3):654-665. doi:10.1016/j.cell.2013.03.043

204. Elefant N, Altuvia Y, Margalit H. A wide repertoire of miRNA binding sites: Prediction and functional implications. *Bioinformatics*. 2011;27(22):3093-3101. doi:10.1093/bioinformatics/btr534

205. Romani AMP. *Magnesium in the Central Nervous System*.; 2011.

206. Rubin H. The membrane, magnesium, mitosis (MMM) model of cell proliferation control. *Magnes Res*. 2005;18 (4):268–274.

APPENDIX: IN VITRO LIGATION AND SEQUENCING OF HYBRIDS USING T4 RNA LIGASE 1

Introduction

microRNAs (miRNAs) are short non-coding RNAs (ncRNAs) of around 22 nucleotides that mediate gene silencing by guiding Argonaute (AGO) proteins to their target sites in the 3' untranslated region (UTR) of mRNAs [191]. AGO proteins use a single stranded small nucleic acid as guides to complementary sequences in RNA or DNA targeted for silencing. The miRNA loaded into AGO, forms a complex named miRNA-induced silencing complex (miRISC) which promotes translational repression and degradation of targeted mRNAs. Deregulation of miRNA function is associated with a variety of diseases, but it is predominantly seen in cancer [113]. miRNAs can act as an oncogene (oncomiR) or as a tumor suppressor depending on the identity of the miRNA target.

Moreover, numerous bioinformatic miRNA target prediction algorithms search for canonical seed matches in 3' UTR as well as using conservation, target site accessibility and binding-free energies to define potential interactions. However, this analysis is very limited due to both high false positive and false negative rates. In addition, imperfect seed complementarities constitute a large fraction of miRNA binding. These algorithms are unable to account for context specific targets, e.g different expression of transcripts in different cell lines. Furthermore, a longstanding riddle of miRNA biology is the existence and conservation of miRNA families as they share the same seed sequence, but, have different sequences at the 3' end. The unknown lies within understanding what the targeting specificity between family members is and if there is any target preference [192]. Uncovering these target preferences will explain the biological need for conservation of miRNA families.

Currently, there are different methods to identify endogenous miRNA targets by generating miRNA target-chimeras. CLASH [193], iPAR-CLIP [194] and CLEAR-CLIP [195] use an ectopic RNA ligase to form miRNA-mRNA chimeras in cell lysates in their protocol. Nevertheless, one of the biggest challenges of these protocols is the low efficiency of ligation reaction, for instance CLASH protocol has chimeric reads lower than 2% [193]. These methodologies identify endogenous miRNA

targets using immunoprecipitation of Argonaute proteins crosslinked to transcripts. Although miRNA loading into Argonaute has been defined as the main mechanism of targeting, it might not always be the case. Perhaps miRNAs can bind first to their target and then Argonaute is recruited for silencing. If this is the case, there might be other protein complex that can mediate miRNA targeting. Moreover, miR-34 family are direct targets of p53 and their upregulation induces apoptosis and cell-cycle arrest. It is a well-studied miRNA family, since it has been shown to target many oncogenes such as Bcl-2, c-MYC, E2F3 among others. Despite knowing their major targets, it is to be determined if all miR-34 family members have a functional redundancy or if there are target specificities not described [196].

In an aim to expand our current miRNA targeting knowledge and to answer these questions we propose to develop a technique to increase miRNA targets-chimeras formation in cells. We will use the following approach: overexpression of exogenous T4 RNA Ligase 1 (T4Rnl1) in mammalian cells to enrich miRNA target-chimeras formation.

Bacteriophage T4Rnl1 is the founding member of the RNA ligases and it is a representative of the nucleotidyltransferase superfamily, which includes RNA ligases, DNA ligases and RNA capping enzymes [197]. The biological role of T4Rnl1 is the countering of a host defense mechanism invoked following bacteriophage infection of the bacterial host. This phage protein is dedicated to the repair of programmed tRNA breaks *in vivo*[198]. T4Rnl1 catalyzes the formation of phosphodiester bonds between the 5' phosphate and the 3'-hydroxyl termini of single-stranded nucleic acids[198]. In addition, it has been shown that disruption of the amino acids K99 or E159 impairs T4Rnl1 activity [197]. In animals, miRNA-mediated mRNA decay, the dominant effect of mammalian miRNAs, consists initially in a deadenylation process that leaves a free hydroxyl group at the 3'end of the transcript. Next, the deadenylated intermediate is susceptible to decapping leaving a 5' monophosphorylated decay intermediate [199,200]. mRNA decay pathway can therefore leave the correct ends for T4 RNA ligase 1 reaction. Additionally, mature miRNAs can serve as a substrate for ligation without further modification, since they require a 5' monophosphate to be loaded into Argonaute[201]. Hence, we reasoned that expression of the T4Rnl1 gene would lead to ligation between the 5' monophosphate of the miRNA, the decaying mRNA or cleaved RNA and the 3' hydroxyl end of miRNAs or mRNAs.

In this section, we present a protocol for generating miRNA-mRNA hybrids in cells, which will enhance the identification of miRNA targets. Through our research, we discovered that the expression of T4Rnl1 enzyme from bacteria remains active within cancer cells after transfection. However, there are still challenges that need to be addressed. Specifically, a deeper understanding and optimization of increasing the concentration of free Mg2+ in the cytoplasm of cells is vital for the advancement of this innovative technique. The successful development of this approach will not only shed light on new aspects of miRNA biology but also uncover novel patterns of target binding, ultimately leading to the identification of previously unknown targets.

Methodology

Cell culture:

A549 cell line used in this study was obtained from American Type Culture Collection (ATCC). Cell line weas routinely confirmed to free of mycoplasma contamination monthly. A549 cell line was grown in RPMI media supplemented with 10% FBS and 1% Penicillin/Streptomycin (complete media).

Transfection:

For all the T4Rnl1 and mCherry transfection experiments, 2μg of the respective vector were transfected into $4X10^5$ using Lipofectamine 2000 (11668019, Thermo Fisher Scientific) following the manufacturer's protocol. Cells were seeded and 24 hours later transfection was performed. Four hours after transfection, Lipofectamine containing media was changed to complete media.

Site directed mutagenesis:

Site directed mutagenesis was performed following the manufacturer instructions (210518, Agilent).

Primers used for K99N mutation:

F: 5' CTACATCCTGACCAACGAGGACGGATCCCTC 3'

R: 5' GAGGGATCCGTCCTCGTTGGTCAGGATGTAG 3'

Primers for E159A mutation:

Forward: 5' CTTCACCGCCAACTTCGCCTTTGTGGCTCCCACCAAC 3'

Reverse: 5' GTTGGTGGGAGCCACAAAGGCGAAGTTGGCGGTGAAG 3'

In vitro enzymatic ligation assay:

Lysates after transfection of the respective T4Rnl1 vectors were isolated using a mild lysis buffer (50nM Tris HCL pH 8.5, 150mM NaCl, 1% NP-40) in combination with fresh 1X protease inhibitor cocktail tablets (PIA32955, Thermo Fisher Scientific). The 6 well plate with cells was placed on ice and media was removed. 1mL of cold 1X PBS was added and then removed for a total of 1 wash. Then, 50µL of Lysis Buffer with protease inhibitor was added to the 6 well plate. The plate was shacked for 10min on ice and cells were scraped out into a microcentrifuge tube. Lysates in the tubes were centrifuged for 10min at 10 000xg at 4C, aliquoted and stored at -20C. Before the enzymatic assay, 5µL lysates were thawed on ice and mixed with 1X of T4 Commercial Reaction Buffer (M0204S, NEB), 1X of ATP (Stock 10mM), 5µM of 5' phosphate miR34a 5p (Stock 200µM), 2U/µL of RNase Inhibitor (AM2696, ThermoFisher), 10U of commercial T4Rnl1 enzyme (M0204S, NEB) for positive control reactions, and volume was adjusted with nuclease free water. After combining these elements, microcentrifuge tubes were incubated at 37C for 1h or indicated time. Reactions were then placed on ice before proceeding with electrophoresis.

12% Polyacrylamide Urea Gel Electrophoresis:

Three mL of RNAse free water were mixed with 1.5 mL of 10X TBE RNAse free. 7.2g of urea was dissolved into the mix. Mix was heated up using the microwave (10s). Before the urea started to come out solution 4.625 mL of acrylamide stock was added and mixed well. 75µL of 10% APS and 7.5µL of TEMED was then added. The gel was then mixed quickly and poured into gel cassette

(12 + 2 wells). The gel was let to polymerize for at least 1 hour and the gel was pre-run gel for 30 minutes at 80V using 1X TBE buffer. Once polymerized, samples were combined with RNA loading buffer and heated at 70C for 10min. Samples were incubated for 1min on ice and then loaded into the gel. The gel was left to ran for around 3h and then the gel was incubated in 30mL of 0.5 X TBE buffer RNAse free with 10μL of Gel red for 10min while shacking. The gel was then imaged using a UV transilluminator.

Northern Blot:

Samples were run using the 12% Polyacrylamide previously described gel. The entire sample of T4Rnl1 lysates reactions were loaded to the gel. After running, transfer was performed using a semidry blotter (Amershan TE70) using 0.5X TBE buffer, four 3M filter paper and a nylon membrane. The transfer was left to run for 1h at constant amperage less than 75mA. Membrane was then crosslinked in a UV crosslinker in the optimal crosslinking button twice. After this, the membrane was prehybridized in the ULTRAhyb-Oligo Hybridization buffer for 30min at 42C. Later on, the antisense of miR34a 5p NIR probe was mixed with 50mL of ULTRAhyb-Oligo Hybridization buffer and was left to incubate overnight at 42C with gentle shacking. Washes with 2X SSC and 0.5% SDS were performed and then membrane was imaged in the LICOR visualizer.

Results

Methodology description for miRNA target identification using in vitro ligation by T4Rnl1 enzyme.

To efficiently identify miRNA targets and investigate complex miRNA interactions, we proposed a protocol involving the expression of T4Rnl1 in cells, which would facilitate the ligation of miRNAs and mRNAs intracellularly, ultimately resulting in the formation of chimeric RNA molecules for subsequent sequencing. The initial step of this protocol entails the expression of codon-optimized T4Rnl1 wild type in cells using lipofectamine transfection. Then, the cells will be transfected with the miRNA of interest. During the deadenylation process, the 3'OH end of the mRNA (shown in blue) becomes available to interact with the 5' phosphate of the miRNA. Likewise, after the decapping process, the 3'OH end of the miRNA can be ligated to the 5' phosphate of the mRNA. After 48 hours of transfection, total RNA will be isolated. To specifically

select the mRNA chimeras of interest from the total RNA pool, we will use a miRNA probe containing biotin at either its 5' phosphate end or its 3'OH end. Subsequently, a streptavidin magnetic bead will be used to enrich for the chimeric RNA molecules. The RNA obtained from the pulldown will be subjected to purification, followed by the construction of a cDNA library. The generated cDNA will then undergo sequencing, and the identification of miRNA-mRNA chimeras will be analyzed (Figure A1A).

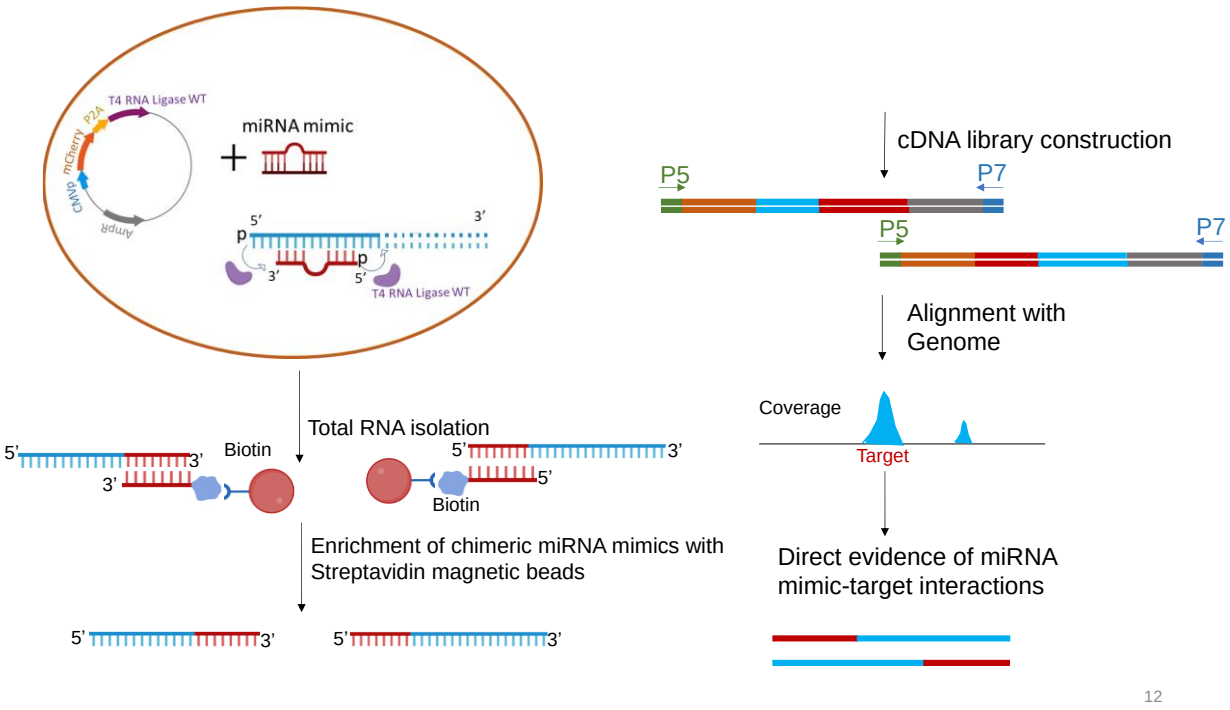

Figure A1. **Proposed methodology for miRNA-target identification using *in vitro* ligation by T4 Rnl1 enzyme.** miRNA of interest is co-transfected with T4 RNA Ligase into cells. After miRNA-target interaction T4 RNA Ligase ligates the 3'OH of the mRNA with the 5'P of the miRNA or ligate the 3'OH of the miRNA with the 5'P of the mRNA. RNA is isolated and using a miRNA Biotin labeled miRNA-mRNA chimeras are pulldown using Streptavidin magnetic beads. Then cDNA library is constructed using known adaptor sequences such as P5 and P7 and then bioinformatic analysis of chimeras is performed thereby identifying miRNA-mRNA interaction sequence.

Description of T4Rnl1 vector used for the in vitro ligation.

To optimize the expression of the bacterial T4Rnl1 protein in human cancer cells, the T4Rnl1 Open Reading Frame (ORF) sequence was codon-optimized by Dr. Orellana. Furthermore, a reporter gene was incorporated upstream of the T4Rnl1 gene to assess proper vector expression

and transfection efficiency. The vector used as the backbone for this construct is PCDNA3.0, which features a CMV promoter and a Kanamycin resistance gene marker. To serve as a negative control for nonspecific ligation activity, two mutations were introduced: a single mutant K99N and a double mutant K99N E159A. These specific mutations are known to impair T4Rnl1 activity[197]. Site-directed mutagenesis was performed using Quick Change, and the primer design for this mutagenesis is illustrated in Figure A2B. The correctness of the introduced mutations was confirmed by plasmid sequencing and restriction digestion.

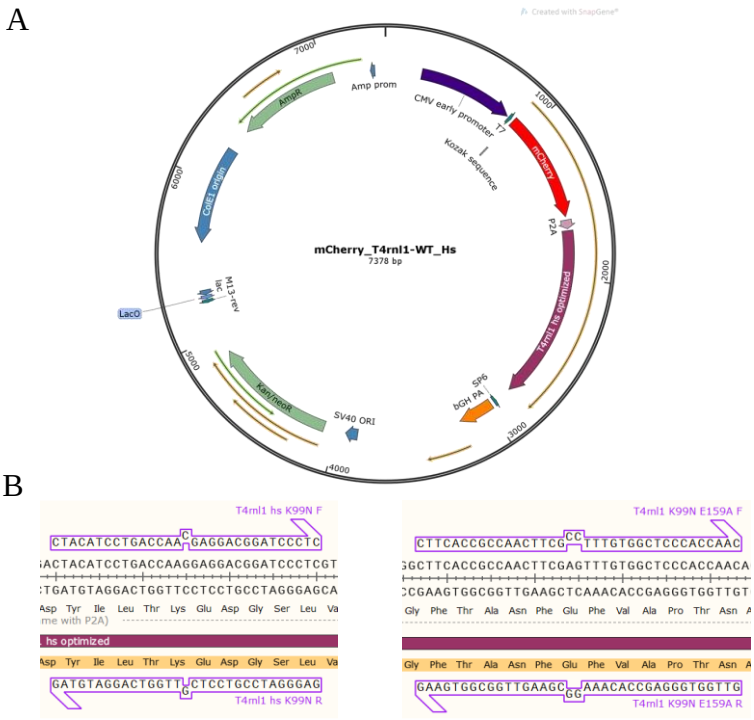

Figure A2. **T4Rnl1-WT vector used for miRNA-target identification technique.** A) Map of *Homo Sapiens* codon-optimized pCDNA3-T4rnl1 plasmid. mCherry gene serves as a reporter for transfection efficiency. Kozak Sequence is a consensus sequence for translation initiation. P2A: 2A self-cleaving peptide leads to ribosomal skipping. B) Primers used in site directed mutagenesis for generation of mutated vectors. Mutations were designed to disrupt enzymatic activity of T4 Rnl1. Plasmid graphics from SnapGene viewer.

T4Rnl1 transcript is highly expressed 18h after transfection.

To confirm the expression of the T4Rnl1 sequence derived from bacteria in human lung cancer cells (A549), we transfected the T4Rnl1 WT, K99N, and double mutant K99N E159A vectors and assessed their respective expression levels. Transcript levels were found to be high 18 hours after

transfection for the T4Rnl1 wild-type vector. In contrast, the expression of the single and double mutants was not as high as that of the wild-type vector, but their expression levels appeared to be higher at 24 hours (Figure A3A). These results indicate that T4Rnl1 can be efficiently transcribed in mammalian cells, with higher expression observed for the T4Rnl1 wild-type sequence. Similarly, the expression of the reporter gene mCherry was evaluated 24 hours after transfection. The mCherry vector control exhibited higher transfection efficiency compared to the T4Rnl1 WT, K99N, and K99N E159A vectors (Figure A3B). However, mCherry expression was similar between the T4Rnl1 wild-type and mutated vectors. Therefore, it appears that the overall gene expression of T4Rnl1 WT and the mutated versions is not significantly different after 24 hours.

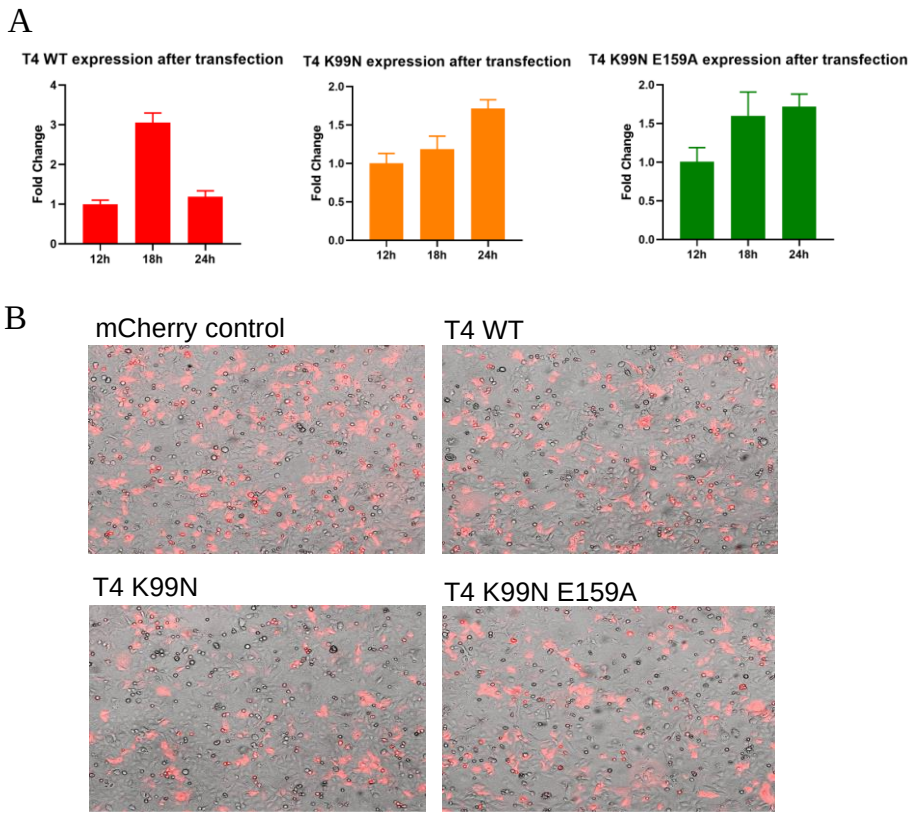

Figure A3. **T4 Rnl1(T4) wildtype or mutated expression levels in transfected A549 cells**. A) qRT-PCR (SYBR Green) data for T4 transcript in A549 cells after 12h, 18h or 24h transfection with T4-WT/K99N/K99N-E159A vectors. Data normalized to GAPDH, relative to 12h time point. B) Fluorescence microscopy 24h after transfection with the indicated vectors. mCherry serves as a transfection efficiency proxy.

T4Rnl1 wild-type can circularize miR34a-5p in vitro when buffer is added to reaction.

To assess the function of the T4Rnl1 protein expressed in mammalian cells and its ability to catalyze ligation, we conducted an *in vitro* test using a miRNA molecule that had the necessary 5' and 3' ends for ligation. Specifically, we utilized miR34a, which was available in the laboratory with a 5' phosphate group modification and a 3' OH group. Lysates from the T4Rnl1-transfected cells were combined with the commercially available T4 Buffer and the 5' phosphate miR34a 5p strand. The mixture was then incubated for 1 hour at 37°C. In the resulting gel electrophoresis analysis, the top band represented the linear form of the miRNA without ligation, while the bottom band indicated the ligated circular product. We observed a single top band for the negative control (no lysate) and for the T4Rnl1 single and double mutant samples. However, in the lysates from cells transfected with the T4Rnl1 wild-type vector, a lower molecular weight band was observed, indicating successful circularization of the miRNA. Furthermore, an even higher molecular weight band was visible in the T4Rnl1 wild-type reaction, suggesting the presence of concatemer ligation products. In each reaction, a positive control containing the commercially available T4Rnl1 enzyme was included (Figure A4A). These findings demonstrate that the T4Rnl1 wild-type gene was expressed in mammalian cells and exhibited ligation activity in the presence of the T4 Buffer components. '

To gain a better understanding of the kinetics of T4 ligation efficiency within cells, we first determined the kinetics of T4 ligation *in vitro*. We performed the same ligation assay in vitro but allowed the incubation reaction to proceed for different durations. In the case of T4Rnl1 WT lysates, we observed a continuous ligation reaction, and at 20 min of incubation, the reaction appeared to reach saturation, indicating that the ligation process was effectively occurring. On the other hand, the negative control T4Rnl1 double-mutant reactions did not circularize the miRNA as expected, confirming the specificity of the ligation reaction (Figure A4B). This finding suggests that in the *in vitro* setting, T4Rnl1 ligation reactions reach their peak after 20 min of incubation at 37°C.

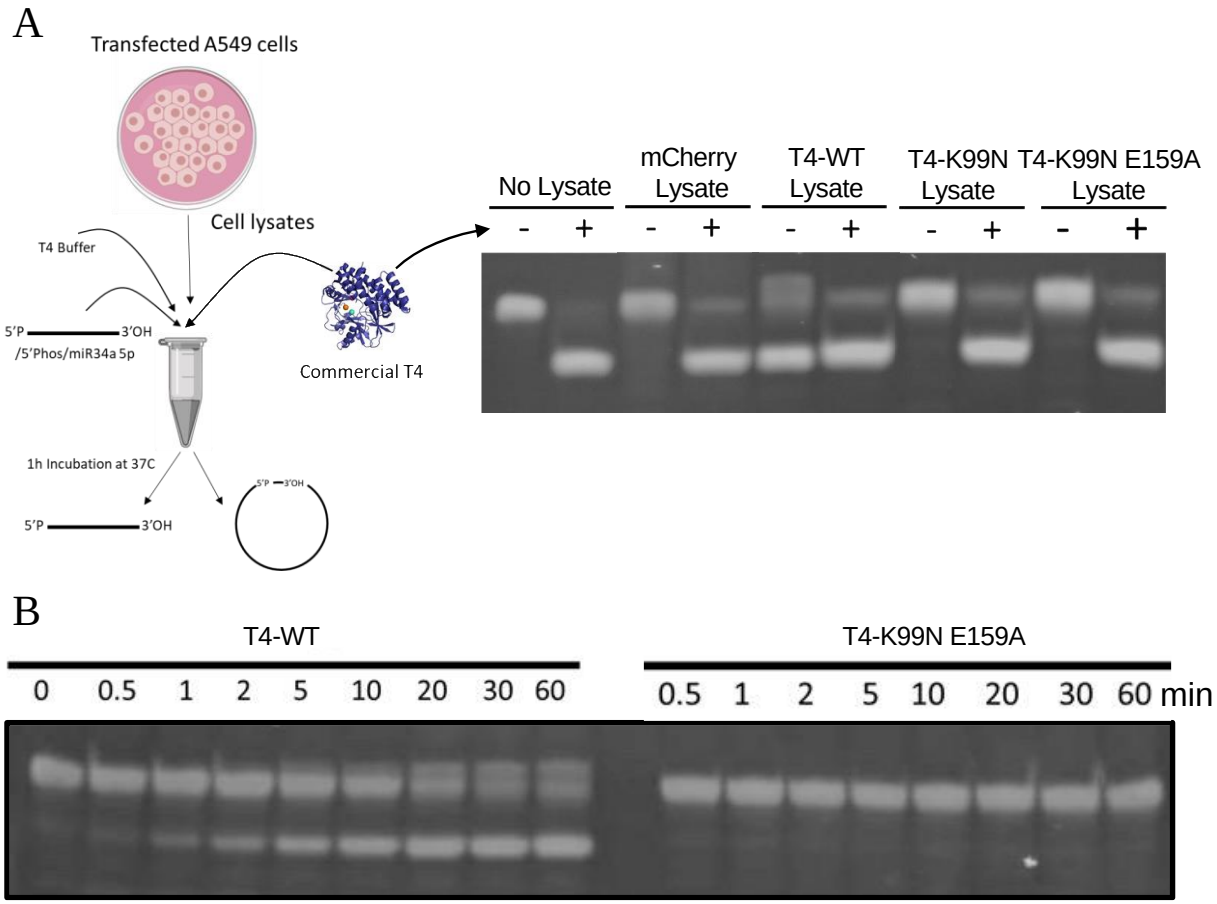

Figure A4. T4-WT enzyme from cell lysates can ligate a miRNA with 5'Phosphate (Phos) and 3'OH *in vitro* (in tubes). **A)** Schematic of T4 ligase in vitro assay. Cells were transfected with different vectors and T4 activity after collecting cell lysates was evaluated. The high molecular band observed in Polyacrylamide Gel corresponds to linear product and lower molecular band corresponds to circular ligated product. **B)** T4 ligase in vitro assay after T4-WT/T4-K99N-E159A transfection and incubation of reaction at 37ºC for different durations.

T4Rnl1 wild-type ligation activity is not detectable in cells.

To validate the activity of T4Rnl1 in cells and determine if ligation of miRNA with its corresponding mRNA targets occurs within the cellular context, we conducted co-transfection experiments. We co-transfected miR34a with a 5' phosphate end modification into cells along with the T4Rnl1 wild-type vector. Our hypothesis was that if T4Rnl1 is capable of ligating miR34a with its mRNA targets, the resulting ligation products would lead to a smear pattern in a Northern Blot analysis. In Figure A5A, we detected miR34a expression in the Northern Blot for all samples except the untransfected control and the mCherry vector with a miRNA scramble negative control.

However, contrary to our expectations, no smear pattern indicative of ligation products was observed in any of the samples, including the T4Rnl1 combined with miR34a 5' phosphate miRNA (T4 WT + miR34a 5p) sample. This suggests that although T4Rnl1 can be expressed in cells, the ligation of miRNAs with their respective targets might not be occurring in live cells. It is possible that the single-stranded miRNA we transfected into the cells is susceptible to degradation by nucleases, compromising the end modification necessary for ligation.

To account for this, we performed a similar experiment using the miR34a mimic, which is a double-stranded RNA that is very stable and has been shown to bind and downregulate its mRNA targets. Interestingly, the results obtained were similar, with only a single band observed for the T4 WT with mimic miR34a and other control samples (Figure A5B). This suggests that either ligation of miR34a to its targets after T4Rnl1 wild-type co-transfection is not occurring in cells, or that the ligation products are not detectable via Northern Blot analysis. It should be noted that additional top bands observed in both experiments may be due to non-specific binding of the miR34a antisense probe. Further investigations are required to elucidate the exact mechanisms underlying the lack of observed ligation products and to explore alternative methods for their detection.

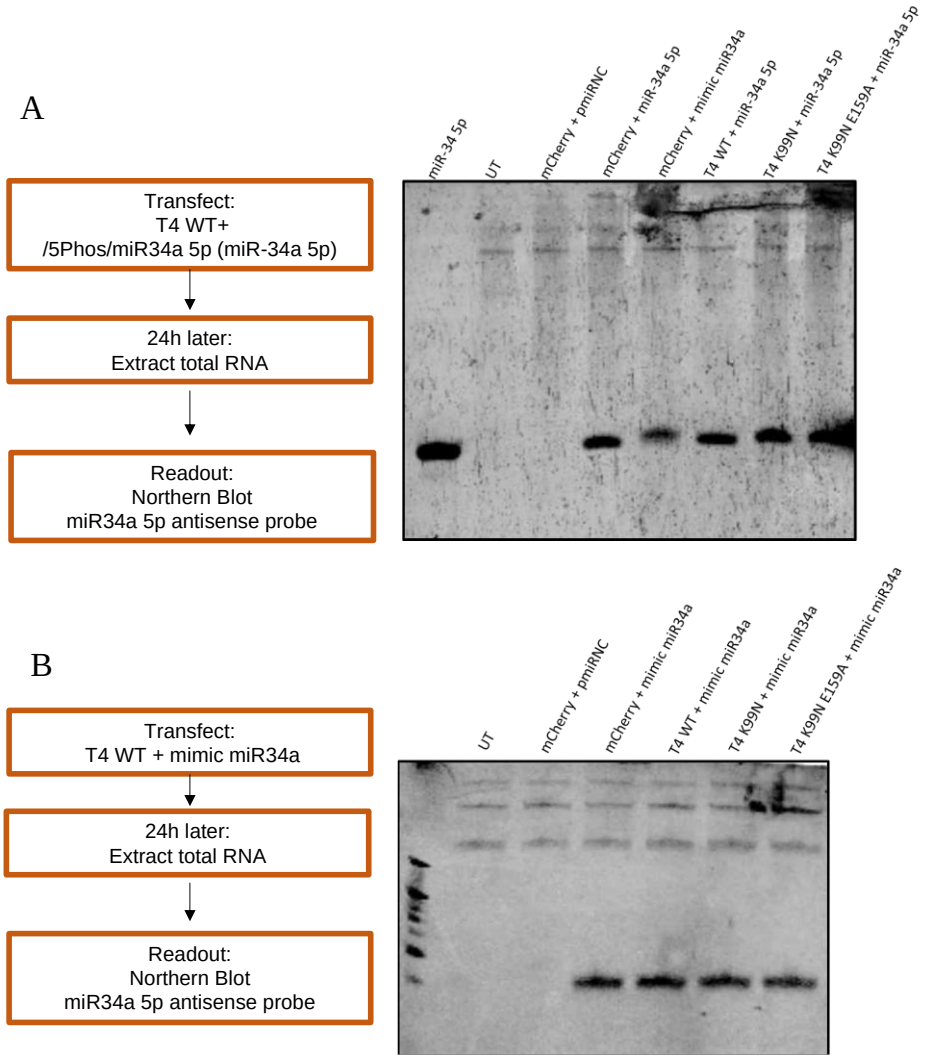

Figure 5. **T4 activity in cells is not evidenced by Northern Blot.** A) Schematic of experimental procedure and Northern Blot result after A549 cells were transfected with different vectors as indicated. Cells were transfected with 5'Phos miR34a 3'OH (miR-34a 5p) or commercial mimic miR34a. **B)** Schematic of experimental procedure and Northern Blot result after A549 cells were transfected with different vectors as indicated. Cells were transfected with commercial mimic miR34a or premiRNC (pmiRNC) negative control. mCherry is the vector control.

MgCl2 ion is essential for T4Rnl1 ligase activity in vitro.

In our previous experiments, we observed that T4Rnl1 was expressed in cells and demonstrated its enzymatic activity *in vitro* when the T4 commercial buffer was included. However, when the T4 buffer was not added to the cells, the T4Rnl1 ligase activity was not detected in the *in vitro* assay. This suggests that a component of the T4 buffer is crucial for the enzymatic reaction of

T4Rnl1 and that this component may not be readily available for the T4Rnl1 enzyme to utilize within cells. Consequently, we conducted further investigations to identify which specific component of the T4 commercial buffer is essential for T4Rnl1 enzymatic activity. In Figure A6A, we evaluated the predicted involvement of different ions in the enzymatic pocket of the T4Rnl1 protein in an in vitro enzymatic assay. However, we focused on the main ingredients of the T4 commercial buffer. In Figure A5B, we observed that adding MgCl2 alone or in combination with ATP enabled the ligation or circularization of the miRNA to occur. Similarly, adding Tris-HCl to the reaction also allowed for ligation, but less effectively than with MgCl2. These findings indicate that Mg2+ ion is the essential ion for T4Rnl1 enzymatic activity. Although Mg2+ is abundant in cells, it is predominantly associated with ATP. Therefore, we hypothesized that the free Mg2+ available in the cytoplasm of cells might not be abundant enough for the T4Rnl1 enzyme to utilize. To test this hypothesis, we added MgCl2 to the cell culture media at different concentrations, aiming to increase the ligation efficiency of T4Rnl1. However, our experiments yielded similar results, with no smear pattern observed via Northern Blot when MgCl2 was added to the media before or after transfection, even at different concentrations (Figure A6B). Excessive levels of MgCl2 in cells can lead to toxicity, and it is also possible that MgCl2 is not being effectively internalized by the cells. Therefore, future experiments should focus on resolving the question of whether we can increase the available Mg2+ within the cell.

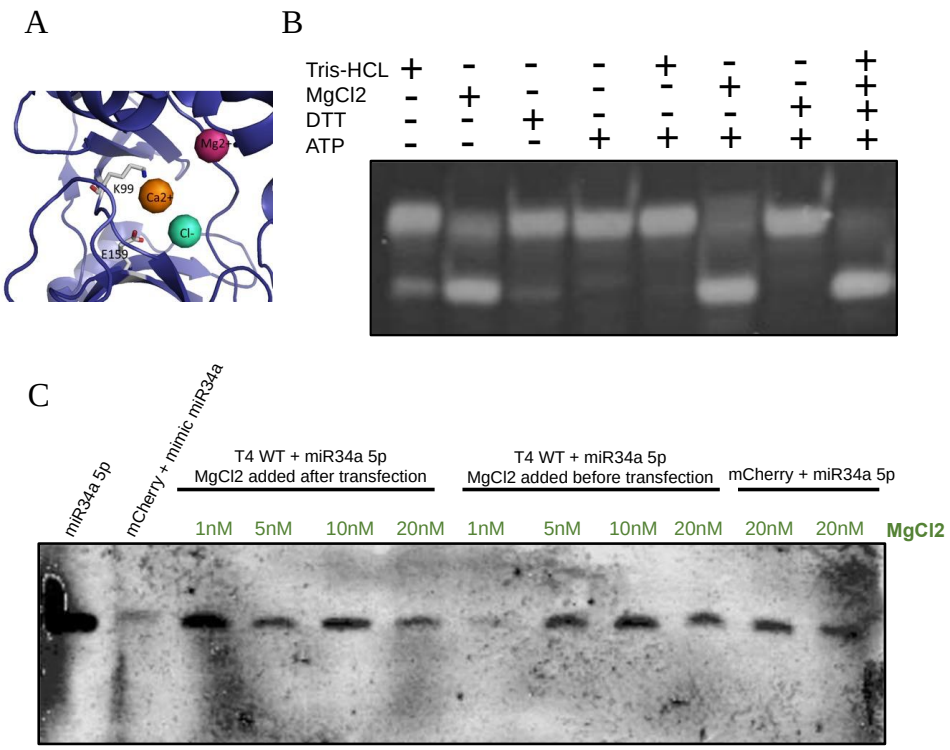

Figure A6. $MgCl_2$ ion is essential for T4 ligase activity in the T4 *in vitro* assay. **A)** Zoom of T4 active site protein structure and the co-factors predicted to be involved in activity. Structure was obtained using Pymol. **B)** T4 in vitro assay using different component of the T4 commercial buffer. The high molecular band observed in Polyacrylamide Gel corresponds to linear product and lower molecular band corresponds to circular ligated product. **C)** Northern Blot of A549 cells transfected with different vectors and miRNAs and treated with different concentrations of $MgCl_2$.

Discussion and future directions

Dysregulation of miRNAs has been implicated in various diseases, highlighting the significance of gaining a deeper understanding of their mechanism of action[158]. One of the main challenges of miRNA study is that predicting miRNA targets by algorithms is imprecise due to non-contiguous base-pairing regions with frequent internal secondary structures [202]. The rule over the last decade is that miRNA-targeting interaction is mediated by the "seed" region and that pairing outside the seed region stabilizes the interaction. However, there is substantial evidence for exceptions. For instance, it has been found that interactions often contain bulged nucleotides, multiple mismatches and wobbles targeting [203]. Therefore, there is a need to better understand the dynamics of miRNA. Moreover, it has been debated in the field if miRNAs are always first loaded into Argonaute and consequently directed to its target. If this is not the case, then miRNA-target interactions could

also be regulated by other protein complex. Furthermore, it has been shown that even members of the same miRNA family can manifest distinct base pairing patterns, showing that many miRNA families do not always have a redundant function [204]. Therefore, there is a need to develop a technique that will allow us to efficiently and unbiasedly identify miRNA targets and complex miRNA interactions.

Current approaches are focused on unraveling these interactions in greater detail, but there is a need to enhance their accuracy and efficiency. Techniques like CLASH have elucidated the non-canonical ways in which miRNAs interact with their mRNA targets. Despite the insights provided by CLASH, the estimated efficiency of miRNA-mRNA ligation is only 2%[193]. In this study, we propose an alternative approach where the ligation reaction takes place within the cells, potentially resulting in improved ligation efficiency. However, several challenges need to be addressed before this approach can be used effectively for miRNA target identification.

In our investigation, we successfully expressed the bacterial T4Rnl1 ligase enzyme in mammalian cells, including cancer cells, and confirmed its functionality through *in vitro* enzymatic activity assays. By utilizing mild lysis buffers, we were able to retain the enzymatic function of T4Rnl1 during enzyme extraction. However, in Figure A2A, the transfection efficiency of T4Rnl1 was found to be less than 100%. Hence, for future assays, it is crucial to optimize the transfection ratios to improve the detection and enhance the efficiency of miRNA-mRNA ligated hybrid products. Despite observing increased transcript levels 18 hours after transfection in T4Rnl1 WT compared to single and double mutants, protein expression, as indicated by similar mCherry expression across all samples, did not exhibit a corresponding increase. This discrepancy could potentially be attributed to post-transcriptional regulation of T4rnl1 mRNA, which may affect efficient translation.

Additionally, detection of miRNA-mRNA hybrids was difficult to detect via Northern Blot, where sensitivity might not be enough to quantify the hybrids. We also tested a slightly more sensitive Northern Blot using isotope radiation (data not shown); however, detection of hybrids was not observed either. An alternative approach to better evaluate T4Rnl1 enzymatic activity in cells could be to co-transfect the T4Rnl1 along with an RNA oligo of around 60 nucleotides that

contains the 5' phosphate end and the 3' OH of known sequence, and design primers that will allow us to detect the chimeras via PCR. The primers will be designed so that a PCR product can only be detected if there is ligation of the RNA transfected and no product if there is no ligation reaction. Multiple size bands should be identified for the samples with T4Rnl1 activity and no band for the no ligase negative control.

In our investigation of the essential co-factors required for T4Rnl1 enzymatic activity, we determined that MgCl2 is necessary and sufficient for ligation. Mammalian cells contain a substantial amount of total and free magnesium (Mg2+) ion, with an average concentration ranging from 17 to 20 mM[205]. Considering that we have demonstrated successful ligation in vitro with as little as 6 mM of Mg2+ (data not shown), this high concentration within cells should be more than adequate. However, it is important to note that the majority of Mg2+ in cells is localized within the mitochondria, nucleus, and endoplasmic reticulum. Mg2+ plays a crucial role by binding and interacting with phospholipids, proteins, nucleic acids, chromatin, and nucleotides. Moreover, only a fraction of the total magnesium content exists as free ions in the lumen of these organelles. Notably, there is currently no available determination of free Mg2+ concentration in the nucleus and cytoplasm. However, the concentration of Mg2+ complexed with ATP and other phospho-nucleotides in the cytoplasm is high due to Mg2+ high binding affinity with ATP[205]. This observation may explain why Mg2+ ions are predominantly associated with ATP and are not readily available for utilization by the T4Rnl1 enzyme. Future efforts will aim to find strategies to increase the concentration of free cytoplasmic Mg2+ for efficient T4Rnl1 ligation within cells.

Considering that Mg2+ is essential for cell growth, its concentration is higher in dividing cells compared to quiescent cells[206]. In fact, intracellular Mg2+ concentration increases upon exposure to growth factors during the G1 and S phases[206]. Therefore, a potential approach could involve synchronizing the cell cycle by serum starvation followed by the transfection of T4Rnl1 and the addition of growth factors. This would potentially lead to an increase in intracellular Mg2+ concentration, which could benefit the T4Rnl1 ligation reaction. However, it is important to note that optimizing the time points is crucial to avoid extensive interaction of all free Mg2+ ions with ATP due to the increased concentration of ATP in the cytoplasm. Additionally, by synchronizing

cells to a specific cell cycle stage, there is a possibility of missing out on other true targets of the miRNA of interest during the other cell cycle phases.

Lastly, it is hypothesized that both the miRNA and the mRNA should possess the appropriate end modifications for the ligation reaction to occur successfully. However, there is a possibility that long mRNA transcripts may be overlooked if the RNA ends are not in close proximity for the T4Rnl1 enzyme to facilitate ligation. Another challenge associated with this technique is the potential for unspecific ligation of other intracellular RNAs that contain a 5' phosphate and 3' OH when T4Rnl1 ligase is overexpressed within cells. This could significantly contribute to background noise and the formation of other RNA chimeras during the sequencing step. Moreover, the ligation of miRNAs with mRNAs and other RNAs may have adverse effects on cell viability, further underscoring the importance of evaluating the impact on cellular health. Therefore, once the cytoplasmic Mg^{2+} concentration is increased and T4Rnl1 activity is confirmed, it is crucial to conduct viability assays and optimize the ligation activity duration to avoid the detection of chimeric products resulting from cell toxicity rather than genuine miRNA targets.

VITA

Alejandra Agredo

Graduate Research Assistant. Dr. Kasinski Laboratory; Department of Biological Sciences

Hansen Life Sciences Research Building; West Lafayette, Indiana 47906

Education

Purdue University | *Ph.D. candidate in Biological Sciences (PULSe)*	**West Lafayette, IN**
Advisor: Dr. Andrea Kasinski - miRNAs as cancer therapeutics	2017 - August 2023
National University of Colombia | *Bachelor of Science in Biology*	**Bogotá, Colombia**
Advisor: Dr. Luis Fernando Cadavid	August 2011- April
- Evolutionary Immunology and Immunogenetics (2014-2017)	2017
Advisor: MSc. Martha Lucía Bueno	
- Cytogenetics (2011-2014)	

Awards and Honors

Purdue University | College of Science	**West Lafayette, IN**
College of Science Graduate Student Travel Award	April 2023
- Awarded a travel grant by the College of Science to present a poster at the American Association for Cancer Research 2023 in Orlando, FL, USA	
Purdue University | Purdue Institute for Cancer Research	**West Lafayette, IN**
SIRG Graduate Research Assistantship	May 2022 - May
- Designed and composed a written grant proposal and won one year of funding designated to an outstanding graduate student undertaking cancer-related research.	2023
Purdue University | Colombian Student Association at Purdue (CSAP)	**West Lafayette, IN**
Travel Grant Award	Spring 2022

- Awarded travel grant by the Colombian Student Association to present a poster at the American Association for Cancer Research 2022, New Orleans, LA, USA.

Purdue University | Ross-Lynn Research Scholar Fund — **West Lafayette, IN**
Graduate Research Assistantship — August 2020 – May 2021
- Formulated and composed a written grant proposal in collaboration with advisor and won one year of Ph.D. stipend.

National University of Colombia — **Bogotá, Colombia**
- Awarded automatic Master admission by placing among top 5 best GPA of 50 graduates. — February 2017

Administrative Department of Science, Technology, and Innovation — **Bogotá, Colombia**
Nexo Global Research Internship — January 2016 – July 2016
- Awarded fellowship for full economic support to perform an international research internship for six months.

Publications

- Pal, A. S.*, **Agredo, A.***, Lanman, N. A., Son, J., Sohal, I. S., Bains, M., Li, C., Clingerman, J., Gates, K., & Kasinski, A. L. (2022). Loss of KMT5C Promotes EGFR Inhibitor Resistance in NSCLC via LINC01510-Mediated Upregulation of MET. ***Equal contribution.** ***Cancer research,*** 82(8), 1534–1547. https://doi.org/10.1158/0008-5472.CAN-20-0821
- Pal, A. S.*, **Agredo, A.** M.*, & Kasinski, A. L. (2022). In-Cell Western Protocol for Semi-High-Throughput Screening of Single Clones. ***Equal contribution**. ***Bio-protocol,*** 12(16), e4489. https://doi.org/10.21769/BioProtoc.4489
- **Agredo A.,** Kasinski A. L., Histone 4 Lysine 20 trimethylation: A key epigenetic regulator in chromatin structure and disease. (Review, submitted to Frontiers in Genetics, Epigenomics and Epigenetics)

- Pal, A. S., Bains, M., **Agredo, A**., Kasinski, A.L. (2020) Identification of microRNAs that promote erlotinib resistance in non-small cell lung cancer. *Biochemical Pharmacology*. Volume 189, 114154. https://doi.org/10.1016/j.bcp.2020.114154

Research Experience

Purdue University **West Lafayette, IN**

Graduate Research Assistant in Biological Sciences Department 2017–December 2022

- Discovered a protein involved in driving lung cancer resistance to targeted therapies by performing a CRISPR-Cas9 screen.

- Validated CRISPR-Cas9 screen results by generating stable cell lines with gene knockout, gene overexpression and inducible expression systems.

- Conducted drug response curves to multiple drugs used in the clinic for lung cancer treatment in cancer cell lines.

- Evaluated protein function in driving resistance to drugs *in vivo*, by performing cell-line derived mouse xenografts and drug administration via oral gavage.

- Acquired technical and theoretical skills in the study of histone methylation such as ChIP-qPCR and CUT&RUN-qPCR.

- Tested the use of a new technique for identification of miRNA gene targets using T4 enzyme RNA ligation in cells.

- Incorporated the use of digital tools for time management and organization that helped achieve publication deadlines.

Undergraduate Research Assistant in Biological Sciences Department January 2016 – July 2016

Mentor: Dr. Andrea L. Kasinski

- Generated stable Cas9 expressing cell line and transduced a library of sgRNAs to perform a **CRISPR-Cas9 screen** in cancer cells.

- Confirmed Cas9 expression in mammalian cancer cells using western-blotting, conducted bacterial cloning, mini-prep, maxi-prep and DNA transfections.

National University of Colombia **Bogota D.C., Colombia**

Undergraduate Research Assistant in Biology Department January 2015 – April 2017

Mentor: Dr. Luis F. Cadavid

- Cloned genes involved in the immune response of the cnidarian *H. symbiolongicarpus* in vectors for overexpression.
- Maintained living colonies of *H.symbiolongicarpus* as well as colony propagation.

National University of Colombia **Bogota D.C., Colombia**

Undergraduate Research Assistant in Biology Department August 2011 – December 2014

Mentor: Msc Marta L. Bueno

- Worked in the Chromosomic Identification Group elucidating the karyotype of the cnidarian *H. symbiolongicarpus*, Resulting in one conference presentation.
- Performed karyological analyses of human samples using peripheral blood primary cell cultures. For karyological analyses: carried out chromosome banding staining techniques to identify the karyotype of several samples, performed Fluorescent in situ hybridization (FISH) techniques, and assisted with solution preparation.

Communication Experience

Purdue University **West Lafayette, IN**

Graduate Research Assistant in Biological Sciences Department 2017–December 2022

- Wrote two successful research grants that were funded to support one year salary each.
- Clarified cell biology concepts to sophomore undergraduate students by giving one-on-one consultation and graded exams.
- Collaborated with Materials Engineering laboratory by providing my technical and critical skills in molecular biology.
- Presented poster in the largest cancer research conference in the United States (AACR, 2022) and in multiple other internal Purdue University symposiums.

Teaching and Mentorship Experiences

Purdue University **West Lafayette, IN**

Teaching Assistant Experience

BIO110, Fundamentals of Biology I, Laboratory and Recitations Fall 2019

- Instructed principles of basic biology to Biology Non-Majors in a classroom setting.
- Prepared and organized weekly laboratory sessions and provided guidance to complete experiments.

BIO231, Cell Structure and Function, Grading, office hours and tutoring. Fall 2021

- Taught and clarified cell biology concepts to sophomore undergraduate students for homework completion and exam preparation by giving one-on-one consultations.
- Graded open question exams for 600 undergraduate students and proctored exams.

Undergraduate Mentoring Experience

- Mentored undergraduate students in a laboratory setting to enrich their research experience and provided guidance to complete individual projects.

Mentees: Manvir Bains (September 2019- Spring 2021), Max Hellrun (Spring 2022-Current)

Graduate Mentoring Experience

- Advised and guided first-year graduate students in their laboratory rotation to ensure accurate experimental skills and encouraged critical thinking.

Mentees: Jenna Clingerman (Spring 2019), Jihye Son (Fall 2020), Basilia Liu (Spring 2022)

Conferences presentations

- Poster: **Agredo A**., Pal A.S., Son J., Lanman N., Kasinski A.L., Loss of the methyltransferase KMT5C drives resistance to tyrosine kinase inhibitors via H4K20me3 regulation in non-small cell lung cancer. American Association of Cancer Research Annual Meeting: **AACR,** Orlando, USA (2023)

- Poster: **Agredo A**., Pal A.S., Son J., Lanman N., Kasinski A.L., Loss of the methyltransferase KMT5C drives resistance to tyrosine kinase inhibitors via H4K20me3 regulation in non-small cell lung cancer. Purdue Institute for Cancer Research Day (PICR). Purdue University (2023)

- Speaker: **Agredo A**., Pal A.S., Son J., Lanman N., Kasinski A.L., Loss of KMT5C drives EGFR inhibitor resistance in EGFR mutant Non-Small Cell Lung Cancer. Cell and Molecular Biology internal seminar series, Purdue University (2023)

- Speaker: **Agredo A**., Pal A.S., Son J., Lanman N., Kasinski A.L., Loss of KMT5C drives EGFR inhibitor resistance in EGFR mutant Non-Small Cell Lung Cancer. Chromatin Club internal seminar series, Purdue University (2022)

- Poster: **Agredo A**., Pal A.S., Son J., Lanman N., Kasinski A.L., Loss of KMT5C drives EGFR inhibitor resistance in EGFR mutant Non-Small Cell Lung Cancer. Biology Department Retreat, Purdue University (2022)

- Poster: **Agredo A**., Pal A.S., Son J., Lanman N., Kasinski A.L., Loss of KMT5C (SUV420H2) drives EGFR inhibitor resistance in EGFR mutant NSCLC cell lines. American Association of Cancer Research Annual Meeting: **AACR,** New Orleans, USA (2022)

- Poster: **Agredo A.,** Son J., Pal A.S., Kasinski A. L., Loss of SUV420H2 drives EGFR inhibitor resistance in EGFR mutant NSCLC cell lines. Purdue Institute for Cancer Research Day (PICR). Purdue University (2021)

- Poster: **Agredo A**., Pal A. S., Kasinski A. L., miR-4435 mediated downregulation of SUV420H2 drives Erlotinib resistance in non-small cell lung cancer. Biology Retreat, Purdue University. (2019)

- Poster: **Agredo A.,** Kasinski A. Direct identification of Argonaute dependent and independent miRNA targeting. Research Symposium Purdue University. Biology Graduate Student Recruitment Event. (2019)

- Poster: Pal, A. S., **Agredo, A**, Kasinski, A. Aberrantly expressed microRNAs drive the development of acquired erlotinib-resistance in non-small cell lung cancer. Biology Retreat, Purdue University. (2017)

- Poster: **Agredo A.;** Xu N., Robles E. A genetic model to study dendritic spines in the larval zebrafish brain. Purdue University Interdisciplinary Life Sciences Program (PULSe) first-year poster competition, Purdue University. (2018)

- Poster: **Agredo A.,** Pal A. S., Kasinski A. Identification of miRNAs that induce erlotinib resistance in lung cancer cell lines using CRISPR Cas9 system. Nexo Global Poster Presentation. Bogotá, Colombia. (2017)

- Speaker: **Agredo A**., Pal A. S., Kasinski A. Identification of miRNAs that induce erlotinib resistance in lung cancer cell lines using CRISPR Cas9 system. Nexo Global Symposium. Purdue University. (2016)

Affiliations and Leadership

Purdue University **West Lafayette, IN**

Member of American Association for Cancer Research (AACR) 2020 - Current

- Communicated my research publication findings by presenting a poster at the AACR conference for two consecutive years.

Purdue University **West Lafayette, IN**

Member of the Association for Women In Science at Purdue (WISP) 2017- 2023

- Attended multiple workshops to develop social skills and to build a sense of belonging among the Women in Science community.

Purdue University **West Lafayette, IN**

Volunteer for the "Big Sibling Program" from the Colombian Student Association at Purdue Fall 2021

- Advised an undergraduate visitor scholar and supported him with graduate school applications.

Purdue University **West Lafayette, IN**

Treasurer at the Colombian Student Association at Purdue 2019 – 2020

- Managed annual budget of $7000 to run club activities, virtual conferences, social events, and professional development workshops.
- Collaborated with a group of six members to gather funds for the association by promoting social media presence, which resulted in $5000 recollection.

Purdue University **West Lafayette, IN**

Volunteer at Science in School (PULSe) 2017

- Demonstrated and lead hands on science experiments to encourage scientific discovery to school-aged students.

Professional development

Purdue University | *Intensive Writing Workshop* **West Lafayette, IN**

- Accomplished writing goals by participating in a three-day intensive writing workshop to complete a review paper. Spring 2022

Purdue University | *Biocomputing in Life Sciences* **West Lafayette, IN**

- Acquired basic knowledge in Python to complete a project using Pandas Dataframes for analysis of publicly available ChIP-seq and RNA-seq data. Spring 2021

Purdue University | *Developing Your Teaching Toolkit: Professionalism* **West Lafayette, IN**

Fall 2019

- Participated in a two-hour professionalism workshop to improve mentoring skills in a classroom setting.

Purdue University | *Developing Your Teaching Toolkit: Engaging ALL Learners* **West Lafayette, IN** Fall 2019

- Participated in a two-hour teaching workshop to improve students' engagement in lecture and laboratory sections.

Skills

Molecular Biology assays: CRISPR/Cas 9/sgRNA | RNA purification | DNA/Protein extraction | cDNA synthesis | Taqman/SYBR Green qPCR | PCR | Lentiviral production/transduction | Transfection | Primer Design | Amplicon Production | Sanger sequencing | Western blots | ELISA | Flow cytometry | Confocal Microscopy | Co-Immunoprecipitation | Molecular cloning | Quick Change/Miniprep | Immunofluorescence | CUT & RUN qRT-PCR | Northern Blot | Polyacrylamide Gel Electrophoresis | miRNA qRT-PCR | Comet assays | FISH

Cell-based assays: Cell culture cancer cell lines propagation | Proliferation assays | Migration assay (Scratch wound) | Luciferase assays | In-cell westerns | Drug response curves | Incucyte proliferation assays (live cell imaging)

Animal-based assays: Rodent restraint | Xenograft | Tumor implantation | Oral gavage | Isoflurane gas anesthesia | Euthanasia

Computer programs: GraphPad PRISM | UCSC Genome Browser | TCGA | IGV | JMP Graphing Software | Microsoft Office | OneNote | Adobe Illustrator

Language Skills

- Advanced English - Nexo Global English Language Program, English for academic purposes - Purdue University
- Advanced Spanish – Native Language
- Advanced French - Lycée Français Louis Pasteur, Bogotá, Colombia (High School)

PUBLICATIONS

- **Agredo, A.,** & Kasinski, A. L. (2023). Histone 4 lysine 20 tri-methylation: A key epigenetic regulator in chromatin structure and disease. *Frontiers in Genetics*, 14, 1243395. https://doi.org/10.3389/fgene.2023.1243395

- Pal, A. S.*, **Agredo, A.***, Lanman, N. A., Son, J., Sohal, I. S., Bains, M., Li, C., Clingerman, J., Gates, K., & Kasinski, A. L. (2022). Loss of KMT5C Promotes EGFR Inhibitor Resistance in NSCLC via LINC01510-Mediated Upregulation of MET. ***Equal contribution.** ***Cancer Research,*** 82(8), 1534–1547. https://doi.org/10.1158/0008-5472.CAN-20-0821

- Pal, A. S.*, **Agredo, A.** M.*, & Kasinski, A. L. (2022). In-Cell Western Protocol for Semi-High-Throughput Screening of Single Clones. ***Equal contribution**. ***Bio-protocol,*** 12(16), e4489. https://doi.org/10.21769/BioProtoc.4489

- Pal, A. S., Bains, M., **Agredo, A**., Kasinski, A.L. (2020) Identification of microRNAs that promote erlotinib resistance in non-small cell lung cancer. *Biochemical Pharmacology*. Volume 189, 114154. https://doi.org/10.1016/j.bcp.2020.114154