Molecular Characterization of Genetic Mutations Altering Lysine Acetyltransferase Complexes

Nicole Robinson

Division of Experimental Medicine

Faculty of Medicine

McGill University

Montreal, Quebec, Canada

August 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Abstract

Epigenetic regulation is a fundamental element for governing eukaryotic gene expression. Eukaryotes have a wide array of proteins that help maintain patterns of epigenetic modifications, which in turn control which genes are expressed in a spatiaotemporal manner. One such epigenetic modification is lysine acetylation, which occurs on a variety of proteins including histones. Due to the proximity of DNA to histones, changes in histone acetylation directly impact gene expression. The addition of acetyl moieties to DNA by lysine acetyltransferases (KATs) allows the decompaction of chromatin, leading to gene transcription, so histone acetylation is considered as an activating mark. Opposing KATs are histone deacetylases (HDACs) which function to remove acetyl moieties from histones, thereby condensing the chromatin and causing transcriptional repression.

The interplay between acetylation and deacetylation is important for cellular functioning. Without appropriate acetylation and deacetylation, there is a dysregulation of gene expression, which can have devastating implications, including development of cancer. With the advent of genetic testing and through the use of mouse models, it is clear that appropriate histone acetylation is critical in embryonic development. Mice lacking KATs or other proteins within KAT complexes often die *in utero* or before weaning, showing substantial defects. Additionally, patients with mutations in genes encoding subunbits of the KAT complexes have been identified through genetic testing of different developmental disorders.

Recently, a patient was identified to possess two different *de novo* mutations in the gene encoding KAT14, one of the two acetyltransferases within the transcriptional adaptor 2A-containing (ATAC) complex. The ATAC complex is known for its roles in cell cycle progression and mammalian development. The patient with KAT14 mutations presents with Diamond Blackfan anemia, a bone marrow disorder that arises during development. These KAT14 mutations alter protein size and stability, while greatly changing the landscape of histone acetylation, affecting both acetylation of histone H3 and H4. Due to the magnitude of alterations in this patients KAT14 it is suspected that the KAT14 variants play a critical role in the mechanism of pathogenesis for this patient.

In addition to the *de novo* mutations found in KAT14, seven different individuals have been identified with WD40 (tryptophan-aspartic acid) repeat domain 5 (WDR5) mutations. WDR5 is a part of the nonspecific lethal (NSL) complex, a KAT complex known for its role in histone H4 acetylation. Patients with WDR5 mutations present with a distinctive global development delay syndrome that mirrors the phenotype of patients with mutations in similar lysine acetyltransferase

complexes. These *de novo* mutations impact WDR5's ability to act as a histone reader, reducing binding to histone H3, as well as alter the subcellular localization of WDR5. Additionally, the mutations impacted the acetylation profile of the NSL complex. Due to similarities in phenotype to other global developmental delay disorders and similar reductions in acetylation found in related KAT complexes, we suspect these seven patients represent a novel global developmental delay syndrome.

These findings about KAT14 and WDR5 shed light on the importance of histone acetylation in development and help classify two new developmental syndromes.

RESUME

La régulation épigénétique est un élément fondamental de l'expression génétique chez les eucaryotes. Les eucaryotes possèdent une grande variété de protéines qui aident à maintenir les motifs épigénétiques, qui à leur tour régulent l'expression des gènes. L'acétylation de la lysine est une des modifications épigénétiques possibles; elle se produit dans diverses protéines, notamment les histones. Étant donné la proximité de l'ADN avec les histones, les changements de l'acétylation des histones ont une incidence directe sur l'expression génétique. L'ajout de résidus d'acétyle à l'ADN par la lysine acétyltransférase (KAT) permet la décompaction de l'ADN chromosomique et, par la suite, la transcription des gènes. Il semble donc que l'acétylation active l'expression génétique. Contrairement aux KAT, les histones désacétylases (HDAC) enlèvent les résidus acétyles, entraînant la condensation de l'ADN et la répression transcriptionnelle.

L'interaction entre l'acétylation et la désacétylation est essentielle au fonctionnement cellulaire. Sans l'acétylation et la désacétylation nécessaires, l'expression génétique devient dérégulée, ce qui peut avoir des conséquences dévastatrices comme le cancer. L'utilisation de tests génétiques et de modèles murins montre clairement que l'acétylation des histones est essentielle au développement embryonnaire. Les souris qui manquent de KAT ou d'autres protéines comprises dans les complexes KAT meurent souvent *in utero* ou avant le sevrage et présentent des anomalies importantes. De même, les patients ayant des mutations dans les complexes KAT présentent différents troubles du développement.

Récemment, un patient ayant deux mutations *de novo* dans le gène codant KAT14, l'une des deux acétyltransférases dans le complexe de l'adaptateur transcriptionnel contenant 2A (ATAC). Le complexe ATAC est connu pour son rôle dans la progression du cycle cellulaire et le développement mammalien. Le patient ayant les mutations KAT14 est atteint de l'anémie de Diamond-Blackfan, trouble de la moelle osseuse qui se manifeste pendant le développement. Ces mutations KAT14 altèrent la taille et la stabilité de la protéine ce qui a un substantielle effet sur l'acétylation des histones H3 et H4. L'importance des altérations du KAT14 chez le patient laisse supposer que les variants KAT14 jouent un rôle critique dans le mécanisme pathogénique de ce patient.

En plus des mutations *de novo* dans KAT14, des mutations des protéines à domaine de répétition WD40 (acide tryptophane-aspartique) 5 (WDR5) ont été détectées chez sept autres individus. La protéine WDR5 fait partie du complexe NSL (nonspecific lethal), soit un complexe KAT connu pour son rôle dans l'acétylation de l'histone H4. Les patients ayant des mutations de WDR5

présentent un syndrome de retard de développement global particulier qui ressemble au phénotype des patients ayant des mutations dans des complexes lysines acétyltransférases similaires. Nous avons trouvé que les mutations *de novo* ont une incidence sur la capacité de WDR5 d'agir comme lecteur d'histone, ce qui réduit la liaison à l'histone H3 et change la localisation subcellulaire de WDR5. De plus, les mutations changent le profil d'acétylation du complexe NSL. Les similarités entre ce phénotype et d'autres syndromes de retard de développement global ainsi que la réduction de l'acétylation dans les complexes KAT nous portent à croire que les sept patients présentent un nouveau syndrome de retard de développement global.

Nos constatations quant au rôle de KAT14 et WDR5 démontrent l'importance de l'acétylation des histones dans le développement et aident à classer deux nouveaux syndromes développementaux.

ACKNOWLEDGEMENTS

The journey to complete my master's has been one of the most fulfilling experiences of my life. Without the support from other, this journey would not have been possible.

A heartfelt thank you Dr. Xiang-Jiao Yang who made this journey possible. Thank you for the unending support, care and guidance throughout this project. Your passion for science and learning drove me to learn more, aim higher and think bigger. Thank you for pushing me to think more critically, adapt to challenges and never give up. I have grown so much not only as a researcher, but as a person through these past three years, and a lot of that is due to your presence. I can't thank you enough.

I would like to thank the members of my thesis commity, Dr. Andrew Mouland, Dr. William Pastor and Dr. Jason Tanny for insightful discussions and meaningful suggestions to help my project progress.

I must also thank lab members, Dr. Lin Li, Dr. Kezhi Yan, Dr. Jianliang Jin and Harvey Li for their commradance, help and joy they brought to the lab. A special thanks goes Dr. Kezhi Yan and Dr. Jianliang Jin who spent time training me, answering countless questions and helping me develop into a better scientist.

A thank you goes out to the scholars who collaborated on this thesis. Their insights and discovery of the patient muations were integral to my work. Dr. Pankaj B. Agrawal who identified the *KAT14* patient mutations and to the researchers who identified and initially characterized patients with *WDR5* mutations, J. Verseput, Lot Snijders Block, and Doctors Constance Stumpel, Karit Reinson, Katrin Õunap, David Burton, Micheil Innes, Johanna va Hagen, Pascal Joset, Matthew Deardorff, Cara Skraban, Hanka Venselaar, Tjitske Kleefstra and P. Campeau.

I will forever be grateful to the friends I have made over the course of my masters, Sara Matthews, Megan Hammell and Sheri McDowell. I am constantly in awe of their intelligence, strength, humor and kindness. Their friendship and never ending support meant the world to me through this journey. I am especially grateful for Sara Matthews and Johanne Guibault for helping with the French translation of my abstract.

Finally, to my parents, Jocelyn and David Robinson. I couldn't have completed this journey without you. Your unconditional love is why I am the person I am today. Thank you for listening to my near constant rambling about my research, even if you didn't always understand. Thank you for being a calming presence when I'm nervous, and for believing in me when I lost belief in myself.

The research presented here was funded by grants from CHIR and NSERC (to Dr. Xiang-Jiao Yang)

CONTRIBUTIONS OF AUTHORS

Chapter One and Two

N.R. completed the literature review, designed the figures and wrote the chapters.

Chapter Three

N.R. performed all experiments within these chapter, bar several DNA constructs that were given by others, listed below. N.R. wrote the manuscript and prepared all figures seen. N.R. in collaboration Dr. Yang designed experiments, edited drafts and finalized the chapter.

The KAT14 patient was clinically evaluated by Dr. Pankaj B. Agrawal of Harvard Medical School, Ma…

The seven patients identified with *de novo* mutations were found through the Max Plank Institute for psycholinguistics or gene matching sites. Initial patient and mutant characterization, including pathogenic predictions, using CADD and SIFT scores were completed by J. Versput and colleagues.[1] This project involves collaborations with several other clinical scientsits including Dr. P. Campeau at U of Montreal.

The following cDNA constructs were prepared by former members of Dr. Yang's lab: FLAG-GCN5, FLAG-KAT8, FLAG-PCAF, GFP-H3, HA-YEATS2 and lab member Mohammad Ghorbani prepared HA-KANSL1, HA-KANSL2 and HA-WDR5

Discussion and Conclusion

N.R. wrote and edited the chapter.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Acetyl-CoA: Acetyl-coenzyme A
AML: Acute myeloid leukaemia
ATAC: Ada2a Containing Complex
ATP: Adenosine-5'-triphosphate
BAF53a: BRG1-Associated factor 53
BRD8: Bromodomain containing 8
BRPF: Bromodomain-containing protein
BSA: Bovine serum albumin
CDK2: Cyclin dependent kinase 2
CHD: Chromodomain-helicase-DNA-binding protein
Chromo: Chromatin binding
CNS: Central Nervous System
Co-IP: Co-immunoprecipitation
DAPI: 4',6-diamidino-2-phenylindole
DMAP: DNA methyltransferase 1 associated protein 1
DMEM: Dulbecco's modified eagle medium
DNA: Deoxyribonucleic acid
DNMT: DNA methyltransferase
DR1: Down-Regulator of Transcription 1
E2F: E2F transcription factor
EPC1: Enhancer of polycomb homolog 1
ER: Endoplasmic reticulum
eRNA: Enhancer RNA
ESC: Embryonic Stem Cells
EZH2: Enhancer of Zeste Homolog 2
FBS: Fetal bovine serum
FOS: FBJ Murine Osteosarcoma viral oncogene homolog
GAS41: Glioma-amplified sequence 41 (YEATS4)
GCN5: General Control of Amino Acid Synthesis 5
GPS: Genitopatellar Syndrome
HA: Human influenza hemagglutinin
HAT: Histone acetyltransferase
HBO1: Histone acetyltransferase binding to ORC1
HCF1: Host Cell Factor 1
HDAC: Histone Deacetylase
HEK 293: Human embryonic kidney 293 cell line
HeLa: Human epithelial carcinoma cell line from fatal cervical carcinoma
HIV: Human immunodeficiency virus
HSC: Hematopoietic Stem Cells
IF: Immunofluorescence
ING: Inhibitor of growth family member
ISWI: Imitation SWI
JADE: Jade family of PHD finger
JNK: Jun-N-terminal kinase
JUN: Jun proto-oncogene, AP-1 transcription factor subunit
JRA: Jun-related antigen
KANSL: KAT8 Regulatory NSL Complex Subunit

KAT: Lysine acetyltransferase
KDM: Lysine demethylase
KMT: Lysine methyltransferase
lncRNA: Long non-coding RNA
LUZP1: Leucine zipper motif-containing protein 1
MAPK: Mitogen activated protein kinase
MBD: Methyl-CpG binding domain protein
MBIP: MAP3K12 Binding Inhibitory Protein 1
MCM: Mini-Chromosome-Maintenance
MCRS1: Microspherule Protein 1
MDM-2 Mouse double minute 2, proto-oncogene, E3 ubiquitin protein ligase
mEAF6: MYST/Esa1 associated factor 6
MECO: Meta-coactivator complex
MECP2: Methyl-CpG binding protein
MED: Mediator coactivator complex
miRNA: MicroRNA
MLE: Maleless
MLL: Mixed-lineage leukemia
MOF: Males Absent on the First
MOZ: Monocytic Leukemia Zinc Finger Protein
MORF: MOZ-Related Factor
MRG15: MORF-Related gene on chromosome 15
MRGBP: MRG Domain Binding Protein
mRNA: Messenger RNA
MSL: Male-Specific Lethal
MTA: Metastasis-associated proteins 1/2/3
MYST: Moz, YBF/SAS3, SAS2, TIP60
NC2-α: Negative cofactor 2 α
NC2-β: Negative Cofactor 2-Beta
NCoR-SMRT: nuclear receptor corepressor- silencing mediator or retinoic acid and thyroid hormone receptor
ncRNA: Non-coding RNA
NSL: Nonspecific Lethal
NuA4: Nucleosome acetyltransferase of H4
NuRD: Nucleosome remodeling and deacetylase
NURF: Nucleosome remodeling factor
OGT: O-Linked N- Acetylglucosamine (GlcNAc) Transferase
ORC1: Origin recognition complex protein 1
P/S: Penicillin-streptomycin
PBS: Phosphate buffered saline
PCAF: P300/CBP-associated factor
PCR: Polymerase chain reaction
PHD: Plant homeodomain
PHF20: PHD Finger Protein 20
PIC: Preinition complex
PMSF: Phenylmethanesulfonylfluoride
PKC: Protein kinase C
PTM: Post-Translational Modifications

PWWP: Pro-Trp-Trp-Pro
REST: Restrictive element 1-silencing transcription factor
RNA: Ribonucleic acid
RNA Pol II: RNA polymerase II
RP: Ribosomal protein
RUVBL: RuvB Like AAA ATPase
SAGA: Spt-Ada-GCN5 acetyltransferase
SANT: Switching-defective protein 3, Ada2, nuclear receptor co-repressor & transcription factor IIIB
Sas: Something about silencing
SBBYS: Say-Barber-Biesecker-Young-Simpson variant of Ohdo Syndrome
SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SET: SET nuclear proto-oncogene
SETD5: SET Domain containing 5
SGF29: SAGA Complex Associated Factor 29
SIRT2: Sirtuin 2
sncRNA: Small non-coding RNA
SNP: Single nucleotide polymorphism
SWI-SNF: SWItch/sucrose non-fermentable
TADA: Transcriptional Adaptor
TBP: TATA-binding protein
TBS: Tris-buffered saline
TGF-β: Transforming growth factor β
TIP60: HIV Tat interacting protein, 60 kDa
TPA: 12-O-tetradecanoylphorbol-13-acetate
TTRAP: Transformation/transcription domain-associated protein
UBP8: Ubiquitin carboxyl-terminal hydrolase 8
WBM: WDR5 binding motif
WDR5: WD Repeat Domain 5
WES: Whole Exome Sequencing
WIN: WDR5 interacting site
XPC: Xeroderma Pigmentosum, Complementation Group C
YEATS: Yaf9, ENL, AF9, Taf14, Sas5
ZZZ3: Zinc Finger ZZ-Type Containing

CHAPTER 1: LITERATURE REVIEW

1.1 EPIGENETIC REGULATION AND HUMAN DISORDERS

The foundation of eukaryotic life's diversity comes through its genetic blueprint, DNA (deoxyribonucleic acid). In order to contain all of genetic information, DNA is packaged into a structure called chromatin, which are tightly packaged nucleosome proteins of core octamers histones surrounded by DNA.[2] Through the compaction of chromatin DNA, differential expression of genetic information occurs, generating different cell and tissues types. Appropriate gene expression is integral in functioning, therefore changes in gene expression can have detrimental consequences. Thus, understanding DNA sequences and gene expression are essential to deciphering the human body in conditions of health and illness.

Gene expression is regulated through post-translational modifications (PTMs) of histones or other proteins. Epigenetics refers to the study of these reversible PTMs. Specifically epigenetic traits are "stably heritable phenotypes resulting from changes in chromosome without alterations in the DNA sequence".[3] In general, there are three types of signals that establish epigenetic traits. Firstly, epigenators are environmental factors that trigger intracellular pathways.[3] The second signal is epigenetic initiators, which respond to the epigenator in order to define the location of the epigenetic chromatin.[3] Finally, epigenetic maintainers maintain the chromatin environment.[3] Maintenance occurs through different phases of the cell cycle and through the following generations.[3]

As a whole, epigenetics is the study of regulatory mechanisms that control gene expression. Changes in gene expression occur through responses to stimuli and can be altered without impacting the DNA sequence itself.

1.1.1 EPIGENETIC REGULATION

Epigenetic regulation processes consist of four different types of modifications: 1) DNA methylation, 2) non-coding RNA associations , 3) chromatin remodelling, and 4) histone modifications.[4] These four regulation mechanisms affect gene expression and utilize PTMs to exert their functions. PTMs are reactions where biochemical moieties are added onto specific amino acids of proteins through enzymatic catalysis.

Many PTMs exist, including methylation, phosphorylation and acetylation. PTMs are added to proteins by transferases and removed by enzymes such as deacetylases, or phosphatases.[5] Each PTM is unique and able to act in different cellular functions. PTM's are able to work alone, or

alongside one another, to coordinate the regulation of cellular and systemic processes. The presence and absence of different PTMs allows varying cell types to arise.

Proteins involved in controlling or interacting with PTMs are called epigenetic regulators, due to their ability to control transcription. Regulators are critical for many cellular processes as well as genomic integrity. Three common types of epigenetic regulators are writers, erasers and readers. Writers are epigenetic regulators that can enzymatically add PTM moieties on histones, proteins or the DNA.[6] Examples of writer proteins are acetyltransferases or methyltransferases. Erasers oppose the function of readers by enzymatically removing the moieties from their respective locations.[6] Different erasers include deacetylases and demethylases.[4] The interplay between the addition and removal of modifications indicates that gene expression is dynamic instead of permeant.[4] Readers are proteins that recognize, and or bind to cellular targets that are modified or unmodified.[4] Reader domains can be found in many proteins and help control the specificity of where reader and eraser proteins exert their effects.[6] Common reader domains include the tudor domain, which recognizes H3K4me2/3, as well as WD40, and bromodomains, which both recognize modified or unmodified histone H3.[7]

With an ever-expanding range of PTMs being identified, the understanding of patterns and networks of PTMs is complex and intertwined. The most common PTMs and the epigenetic regulators that are responsible for the modifications are described in detail in the following sections.

1.1.1.1 DNA Methylation

Methylation and demethylation are one set of unique modifications. Methylation based modifications occur on proteins, including histones and on the DNA itself.[8] On DNA, methylation occurs on cytosine residues, particularly CpG dinucleotides, where the mark is associated with transcriptional repression. DNA methyltransferases (DNMT) are responsible for catalyzing the methylation reaction. DNA methylation patterns can be propagated, for example DNMT1 is able to reproduce CpG methylation patterns on the DNA of daughter progeny to match the parental DNA after cell division.[8] Although many DNA methylation sites act to directly repress gene expression, some can act indirectly, allowing the recruitment of other enzymes or complexes to further compact the chromatin.[8]

1.1.1.2 Non-coding RNA Associations

Like DNA, non-coding RNA (ncRNA) can also undergo modification. The vast majority of the human genome is not translated into proteins, however these non-coding regions have regulatory

roles in gene expression patterns, among other cellular functions.[9,10] Ribosomal and transfer RNAs (rRNA and tRNA) are two types of ncRNA and are involved in the synthesis of proteins from messenger RNA (mRNA).[9] ncRNA can be divided into two major groups, small or large ncRNA (sncRNA, lncRNA), both of which are able to regulate and recruit DNMTs or histone modifying groups to exert transcriptional effects.[9] Examples of sncRNA include microRNA (miRNA) and small interfering RNA (siRNA). A single stranded variant of ncRNA, miRNA, is important in regulating gene silencing after transcription, through supressing translation and enhancing endonucleic cleavage.[11,12] Small interfering RNA (siRNA) are similar to miRNA and lead to transcriptional silencing mediated through the degradation of mRNA.[12] Conversely, enhancer RNA's (eRNA), a type of lcRNA, correlates to the presence of active gene transcription.[13] lncRNA are thought to have specific functions in the nucleus, helping bring regulatory complexes to specific areas of the genome either activating or repressing gene expression.[13] In all, ncRNAs are involved in both gene silencing and activation, making these RNAs invaluable to gene expression and an important component of epigenetic regulation.[14]

1.1.1.3 Chromatin Remodeling

Another form of epigenetic regulation is chromatin remodeling. Chromatin remodeling is vital in regulating gene expression, acting to package chromatin into specific regions.[15] Despite chromatin regions being tightly packed, the DNA can still be accessed and this process is regulated by chromatin remodelers.[15] In chromatin remodeling, remodeling complexes hydrolyze ATP in order to generate energy.[16] This energy can then be used to alter the structure of chromatin, which modifies where nucleosomes are.[16] Remodeling complexes are able to effect various DNA elements, such as enhancers, origins of replication or promoters, mediating how these elements are packaged in the cell, and determining gene expression patterns.[15] There are four major chromatin remodeling families, SWItch/Sucrose Non-Fermentable (SWI/SNF), Imitation SWI (ISWI), Chromodomain-helicase-DNA-binding protein (CHD), and Chromatin-remodeling ATPase INO80 (INO80). These families of remodelers share certain properties, such as recognition of histone modifications, nucleosomal and DNA affinity, ATPase regulatory domains, interaction domains and the ability break the binding between histones and DNA through DNA dependent ATPase domains.[15]

Remodelers can alter gene expression by mediating the proper location and addition of nucleosomes for the chromatin to bind to after DNA replication.[15] Remodelers can also be used to access hidden *cis* DNA elements, through nucleosomal sliding.[15] Additionally, remodelers are

important in transcription by regulating nucleosomes, either removing them to allow polymerase functioning to continue un-impeded or properly replacing the remodeler once this process is finished.[15]

1.1.1.4 Histone Modifications

Another modification that affects nucleosomes, or more specifically their octamers are histone modifications. In the nucleus, PTMs can occur on histones, where moieties are added or removed from histone tails. Modifications of the tails can serve to transcriptionally activate or repress regions of DNA.[7] The changes between active and inactive states are reversible, leading to dynamic control of gene expression.[7] PTMs are identical to histone modifications and occur outside of the nucleus on cytoplasmic proteins.[16] By modifying proteins, signalling patterns are stimulated or adjusted, allowing for further alterations in gene expression. Here after, four common types of histone modifications will be examined in further detail.

1.1.1.4.1 Methylation and Demethylation

Methylation and demethylation are a unique set of modifications. Methylation based modifications can occur both on proteins and on the DNA itself, as previously mentioned.[8]

A common place for modifications is on histone tails, and this is no exception for methylation. Protein methylation occurs due to the activity of lysine methyltransferases (KMT), which are able to either mono, di or tri methylate lysine residues on target proteins or within the histone tails. An example of this is KMT2D or MLL2, a member of the SET1/MLL (SET nuclear proto-oncogene 1/ Mixed-lineage leukemia) family of histone methyltransferases, which specifically targets the methylation of histone H3K4. Unlike acetylation, methylation will not affect the charge of the histone tail, instead lysine methylation serves as a recognition site for reader proteins, which will bind to these marked histones.[17] Histone methylation can be both activating or repressing depending on which residue is affected or the amount of methyl groups added, for example H3K4 methylation indicates active transcription, whereas H3K9 methylation silences chromatin regions.[18] Opposing the action of KMT, are lysine demethylases (KDM), which act to remove methylation marks from histones.[18]

1.1.1.4.2 Phosphorylation

Phosphorylation is another common catalytic process that occurs on proteins. Kinases are enzymes that catalyze phosphorylation and dephosphorylation.[4] These enzymes can act as both activators or inhibitors of their targets.[4] Kinases can directly phosphorylate targets or indirectly phosphorylate through binding to chromatin or other proteins to exert effects downstream.[4]

Phosphorylation interacts with other types of modifications. An example of this can be found through the phosphorylation of the methyltransferase EZH2 (Enhancer of Zeste Homolog 2). This causes a reduction in EZ2H's ability to bind in complexes and subsequently a loss of histone H3K27me3 causing increases in gene expression.[19] The addition of phospho moieties can act to either enhance or suppress gene expression, and is dependent on the marker or protein being phosphorylated.[4] Phosphorylation can allow increases in binding, changes in specificity of the binding partners, either an activation or termination of protein function and enhance or supress other modifications.[4] In all phosphorylation is a wide-spread mark with many functions.

1.1.1.4.3 Lysine Acetylation and Deacetylation

One type of lysine modification is acetylation. Acetylation involves the transfer of an acetyl group from acetyl Co-A to the ε-amino group of histone tail lysine residue or other non-histone protein targets.[20] This reaction is catalyzed by lysine acetyltransferases (KAT) and can be reversed by histone deacetylases (HDAC). Histone acetylation alters gene expression through its ability to neutralize the positive charge of histones, allowing for a weakened association between histones and negatively charged DNA.[20] In this relaxed state chromatin is transcriptionally active, with the presence of acetylated histones indicative of active transcriptional.[20,21] For this reason, a common location for acetylation modifications to occur is on gene promoters or enhancers.[21] Conversely, when HDACS catalyze deacetylation there is transcriptional repression and a tightening of the association between the DNA and the nucleosomes.[20]

Historically, KATs were referred to as HATs (histone acetyltransferases), but the name has been updated to reflect the enzymes ability to acetylate both histone and non-histone targets. However, the term HDAC is still used, since a common consensus in nomenclature was agreed upon.[22] More than a dozen KATs have been identified, with the majority of KAT's belonging to one of three groups, GCN5-related-N-acetyltransferase (General control of amino acid synthesis 5) (GNAT) family, p300/CREB binding protein (CBP) family and the Monocytic Leukemia Zinc Finger Protein (MOZ), Ybf2/Something about silencing 3 (Ybf2/Sas3), Sas2, HIV-1 Tat Interactive Protein, 60kDa(TIP60) or as its commonly known, the MYST family.[23–27] To date 18 different HDAC's have been identified.[28] These 18 enzymes can be categorized into four classes: Rpd3-like proteins are in class one (HDAC1-3/8), Had1-like or class two (HDAC4-7/9/10), class three consists of Sir2-like proteins (SIRT1-7) and finally there is class four, featuring only HDAC11.[28] All HDAC's in classes one, two and four have conserved active sites and rely on zinc dependent

mechanisms.[28] Whereas SIRT enzymes utilize NAD^+ dependent mechanisms to perform deacetyltransferase activities.[28]

The regulation of acetylation modifications on histones plays an integral role in many cellular functions. Acetylation is an activating modification, and thus can stimulate transcription. Depending on the acetyltransferase different histone lysine substrates will be targeted, allowing for specificity of gene transcription. Acetylation has been implicated in DNA repair mechanisms, such as nucleosome acetyltransferase of 4 (NuA4) complex, a KAT5 containing complex, which is necessary in the formation of DNA double stranded breaks.[29] Furthermore, p300 is involved with proliferating cell nuclear antigen (PCNA), which is found in DNA replication and repair processes.[30]

1.1.1.4.4 ACYLATION

Acylation modifications are similar to the previously mentioned acetylation modifications. There are eight identified types of short chain acylation modifications, which occur on the ε-amino group of histone tail lysine residues, including: butyrylation (bu), hydroxyisobutyrylation (hib) propionylation (pr), succinylation (succ), crotonylation (cr), β-hyrdroxybutyrylation (bhb), glutarylation (glu), and malonylation (ma).[31] Within the eight types of acylations, three groups arise based on their chemical properties.[31] β-hyrdroxybutyrylation and hydroxyisobutyrylation are both polar acyl groups, and when bound to lysine their hydroxyl group allows them to form hydrogen bonds with proteins.[31] Succinyl, malonyl, and glutaryl all contain acidic acyl groups with negative changes which influence lysine residues to adopt a negative charge.[31] Finally, propionylation, crotonylation and butyrylation are hydrophobic, and cause lysine residues to have increased hydrophobicity.[31]

Many acylation modifications are catalyzed by KATs. KATs are able to use both acetyl-CoA and acyl-CoA as substrates to modify gene expression.[31] One example is KAT8, a member of the MYST family, which propionylates histone H4K16.[32] For this reason, KATs that catalyze acylation reactions may be called acyltransferases.[31] Like acetylation, these acylation's can be removed by one of the four classes of HDACs.[32] Patterns of acylation's, histone modifications and the larger networks of PTMs, show the complexity of epigenetic patterns and how important they are to gene expression.

1.1.2 EPIGENETIC REGULATION AND DISEASE

Maintenance of epigenetic patterns is critical for cellular and physiological functions in humans. When dysregulation of epigenetic networks occurs, there can be severe and detrimental

consequences. Changes in gene expression caused by alterations in epigenetic regulators are a common cause of disease. The following sections provide examples of disorders arising from defects in epigenetic regulators that alter epigenetic patterns.

1.1.2.1 CANCER

One disorder well known for both aberrant genetic and epigenetic regulation is cancer. In cancer there are global changes to the epigenome, including alterations in gene expression, histone modifications, protein activation and DNA methylation patterns.[33] Common alterations in gene expression include activation of oncogenes and silencing of tumor suppressors. These conditions are advantageous for the uncontrolled growth of tumor cells.[33]

In cancer, hypomethylation is found genome wide, however certain CpG island promoters are hypermethylated.[34,35] DNA hypomethylation can cause oncogene activation and is integral in promoting genomic instability, encouraging both chromosomal rearrangements and retrotransposon translocation.[36–38] Conversely, hypermethylation has the opposite effect, acting to promote tumor suppressor silencing and tumor initiation, effectively causing the loss of DNA repair or apoptotic functions.[33,39]

Patients with cancer have genome wide changes in chromatin modifications. Common alterations found in cancer patients includes a loss of H4K16ac and H4K20me3, or an over expression of HDACs, causing a profound reduction of acetylation.[40,41] Chromosomal translocations can also occur in KAT proteins including *MOZ*, which is a prominently mutated protein in leukemia patients.[42] These are a fraction of potential modifications in epigenetic regulation that occur in cases of cancer.

1.1.2.2 NEURONAL DISORDERS

In addition to cancer, epigenetic dysregulation is implicated in other disorders. Hereafter alterations in the central nervous system (CNS) will be examined.

Neurodegenerative disorders have been linked to impairments in epigenetic mechanisms. One of these disorders is Alzheimer's disease, where there is a dysregulation of restrictive element 1-silencing transcription factor (*REST*).[43] REST proteins normally silence stress or apoptotic genes, however in Alzheimer's, REST is lost, leading to gene activation.[44] Alzheimer's features hypermethylation of CpG sites specific for genes, such as ankyrin 1, whose single nucleotide polymorphism (SNP) variants are risk factors for Alzheimer's.[43,45] Another neurodegenerative disorder is Huntington's disease, where patients have altered levels of 5-mc DNA methylation.[46] Furthermore patients with Huntington's have mutations in the huntingtin protein, which interacts

with the CBP acetyltransferase, causing a depletion in soluble or active CBP levels, therefore, leading to hypoacetylation.[47,48] Furthermore, in the brains of Huntington's disease patient there is an upregulation of HDAC's, particularly HDAC5.[49] A dysregulation in the equilibrium between KATs and HDACs is a common feature of Huntington's disease.[48]

Similar to neurodegenerative disorders, epigenetic dysregulation is found to cause developmental or intellectual disabilities. Rubinstein-Taybi syndrome is an intellectual disability and facial anomaly disorder that is caused by mutations in the *CBP* acetyltransferase.[50] Rett syndrome is another X-linked neurological disorder that impacts neuronal, motor and language development.[51] Most cases of Rett's are attributed to mutations in the methyl-CpG binding protein 2 (MECP2), a protein known to modulate transcription through binding to methylated regions of the DNA.[52,53] Mutations in the *MECP2* gene can demolish the proteins ability to interact with the corepressor complex NCOR-SMRT(nuclear receptor corepressor- silencing mediator or retinoic acid and thyroid hormone receptor), altering gene silencing.[54] Similarly, intellectual disability and global developmental delay syndromes caused by mutations in the MYST family of acetyltransferases have been well characterized. These disorders will be described in further detail in section 1.3.

1.2 Diamond Blackfan Anemia and the ATAC Complex

1.2.1 ATAC Complex

1.2.1.1 SAGA-Like Complexes

The Spt-Ada-GCN5-acetyltransferase (SAGA) complex was first discovered in yeast in 1997.[55] SAGA was initially characterized for its ability to act in transcriptional processes, its ability to effect histone modifications and its non-transcriptional roles in maintaining DNA integrity.[56] The SAGA complex is evolutionarily conserved; however, the complex diverges in metazoans, leading to two evolutionarily related complexes, SAGA and ATAC.[57] The metazoan SAGA complex more closely resembles its yeast ancestor compared to ATAC, however both SAGA and ATAC share a catalytic core. The core consists of four subunits, SAGA Associated Factor 29 (SGF29), Transcriptional Adaptor 3 (TADA3), KAT2A or KAT2B and TADA2A (ATAC) or TADA2B (SAGA).[58] ATAC and SAGA's remaining subunits are unique to their respective complexes.

While they may be evolutionarily connected, the roles SAGA and ATAC play in physiological and cellular processes are different. One difference between the two complexes is SAGAs ability to act as a deubiquitinase, specifically on histone H2B.[59] SAGAs deubiquitinase functions are mediated by the subunit Ubiquitin carboxy-terminal hydrolase 8 (UBP8).[59] Both ATAC and

SAGA are involved in stress induced signalling pathways, however the pathways and genes they interact and regulate differ. One stress pathway is the DNA damage response to UV damage, a key member of this pathway is tumour suppressor p53.[58] It was found in human U2OS cell lines treated with UV, that ADA2B, a SAGA specific subunit, was recruited to p53 dependent promoters (p21, Growth arrest and DNA-damage-inducible 45 alpha, p53 up-regulated modulator of apoptosis and BCL2 associated X apoptosis regulator), but not ATAC's ADA2A allowing SAGA, not ATAC, regulates p53.[60] Other stress pathways include MAPK (mitogen activated protein kinase) signalling cascades.[59] How these MAPK pathways are regulated depends upon which complex or protein they interact with. The SAGA complex is a part of the endoplasmic reticulum (ER) MAPK stress pathway.[59] In conditions of stress SAGA subunits and RNA Pol II localize to ER stress regulated genes *(glucose-regulated protein 78kDa, homocysteine inducible ER protein with ubiquitin like domain, C/EB homologous protein, endoplasmic reticulum resident protein 70)*.[61] Conversely, ATAC is able to regulate the osmotic stress pathway, specifically through regulating Jun-N-terminal kinase (JNK).[62] ATAC works to regulate the JNK signalling cascade in both conditions of normality and osmotic stress.[62] In normal conditions, c-Jun requires the ATAC complex to act as a transcriptional cofactor, where it occupies target genes *Jra* and *chickadee*, allowing for the acetylation of enhancer and promoter regions of these genes.[62] This indicates in homeostasis, the ATAC complex helps to activate transcription of JNK target genes.[62] However, in conditions of osmotic stress the ATAC complex plays an opposing role, acting to suppress *JNK* gene activation.[62] In order to suppress *JNK*, ATAC co-localizes to the *c-Jun* gene and recruits kinases like, mitogen-activated protein kinase kinase 4, JNK and Misshapen, which prevents the pathway from activating.[62]

Similar to their roles in stress responses, the acetyltransferase capabilities of ATAC and SAGA vary. Both SAGA and ATAC contain the same catalytic core, therefore some acetylation properties overlap, with KAT2A and KAT2B both functioning to acetylate histone H3K9.[58] However, the ATAC complex contains a second acetyltransferase subunit, KAT14. KAT14 works to acetylate histone H4 residues, particularly H4K16, and to a lesser extent H4K5/12.[58] While KAT14 in ATAC has histone H4 acetylation capabilities, these are generally weaker than those of KAT2A/B, indicating that strong histone H3 acetylation can be activated by both SAGA and ATAC, whereas weak histone H4 acetylation will only be found in the presence of ATAC.

1.2.1.2 ATAC Subunit Composition

The ATAC complex contains ten different subunits. Many of these subunits are considered to be epigenetic regulators and have unique roles in the context of the complex. ATAC can be thought of as two halves linked together through protein interactions. These two halves are the catalytic core, shared between SAGA and ATAC, and the ATAC specific half, likely linked through binding to WDR5.

In ATAC, the shared catalytic core features four subunits, TADA2A, TADA3, SGF29 and either KAT2A or KAT2B. TADA2A is ATAC specific, however a similar protein TADA2B is in the SAGA complex.[59] The role of TADA2A in ATAC is to stimulate KAT2A or KAT2Bs acetyltransferase capabilities, however this activation is weaker than that of TADA2B. [59] TADA2A has two common domains that affect chromatin function, SANT and SWIRM.[63] Furthermore, SANT domains are readers, both recognizing and binding to histones.[63] TADA2 proteins are mandatory for the acetyltransferase activity of KAT2A.[63] Like TADA2A, TADA3 is needed for efficient acetylation of histones by KAT2A and KAT2B.[61] TADA3 is able to interact with non-histone substrates such as p53. TADA3 is a co-factor of p53 and is required for p53s transcriptional and apoptotic functions.[64] SGF29 is another protein in the catalytic core. This protein features tandem Tudor domains, allowing SGF29 to specifically recognize methylated histones, particularly H3K4me2/3 and allows for recruitment of the ATAC or SAGA complexes to target sites.[65] The final members of the catalytic core are either KAT2A (GCN5) or KAT2B (PCAF, P300/CBP-associated factor). These KATs are highly related and interchangeable within the ATAC complex. Both KAT2A and KAT2B function as lysine acetyltransferases, working to acetylate histone H3 targets, with the highest affinity for H3K9. KAT2A and KAT2B contain bromodomains, thus they act as readers, and bind to acetylated histone lysine residues.[58]

The catalytic core is attached to the remaining ATAC subunits through interactions with WDR5. WDR5 is a histone reader and recognizes and binds to either modified or unmodified histone H3.[66] In addition to its roles as a histone reader, WDR5 also acts to mediate protein interactions. WDR5 has two different binding sites, the WDR5 binding motif (WBM) and the WDR5 interacting site (WIN).[66] Given similarities in binding motifs of proteins that interact with WDR5 it is likely that WDR5 will bind to TADA3 through the WIN site and Zinc finger ZZ-type containing 3 (ZZZ3) through the WBM motif, bridging the ATAC specific subunits to the catalytic core. This suggests WDR5 is integral for in structural formation of the complex.

The remaining five members of the complex are ATAC specific. ZZZ3 is one of these members, and is a histone H3 reader, whose binding properties are enhanced by the acetylation of H3K4.[67] The interactions between ZZZ3 and histone H3 are critical for ATACs acetylation and gene activation properties.[67] KAT14 is the remaining lysine acetyltransferase of the ATAC complex, and primarily acetylates histone H4K16.[68] Another notable role of KAT14 is its ability to act as a scaffold, helping to maintain complex stability and integrity.[58] When KAT14 is lost in drosophila, the ATAC complex will disassemble, highlighting its importance in complex integrity.[69] MAP3K12 binding inhibitory protein (MBIP) is a mammalian specific component of the ATAC complex, and was initially determined to be an inhibitor of the mitogen-activated protein kinase (MAPK) upstream kinase.[68] MBIP is believed to bind directly to KAT14 to promote the structural integrity of the ATAC complex.[69] Subunits YEATS domain containing 2 (YEATS2) and down-regulator of transcription 1 (DR1), form a heterodimer through the interaction of their histone fold domains. This heterodimer is a negative regulator of transcription, when recruited to a promoter alone they can interact with TBP (TATA-binding protein).[70] YEATS2 is a histone reader that selectively binds to both acetylated and crotonylated histone H3K27.[71] DR1 is the final component of the ATAC complex and it naturally forms a duplex with NC2α (Negative cofactor 2 α), however each subunit can function independently.[72] DR1 is found in high concentrations in growing cells when glucose is absent.[72] The formation of and the binding within the ATAC complex has yet to be determined.

1.2.1.3 ATAC Complex Functions

The ATAC complex serves numerous functions within cells. One notable role of the ATAC complex is acetylation of the chromatin allowing for transcriptional activity. ATAC is also involved in cell cycle progression.[73] When the complex is lost there is a dramatic decrease of cells in the S phase and an increase of cells in the G_2/M phase.[69] Certain members of the ATAC complex are critical in cell cycle transitions. There is a delay in M/ G_1 transitions when *ADA2A* or *ADA3* are lost, with the cells displaying defects in the midbody and centrosome, as well as, an increase in acetylation of both histone H4K16 and α-tubulin.[73] This confirms that ATAC is important in cell cycle phase transitions and when the complex is lost it leads to cell cycle arrest.[69] Throughout mitosis ATAC will localize to the mitotic spindle, thereby helping mediate progression through the cycle.[73] One important component of the cell cycle is Cyclin A/ cyclin dependent kinase 2 (CDK2). Cyclin A/CDK2 promotes the progression of cells through the S phase and stimulates the G_2 phase, where CDK2 acts to activate Cyclin A through phosphorylation.[69] Cyclin A/CDK2 are

important in early mitosis for the formation of the centrosome, but must be degraded through ATAC acetylation of cyclin A.[73] This degradation allows SIRT2 to become active which is required for proper progression of mitosis through the deacetylation of tubulin and H4K16.[73] This means SIRT2 activity is indirectly regulated by ATAC and ATAC is necessary for mitosis.[73]

Along with cell cycle regulation, the ATAC complex has a role in nucleosome sliding activity.[72] Nucleosome sliding refers to sliding or displacement of nucleosomes through modifications of histones and impacts the stability or the structure of nucleosomes.[72] Alterations in nucleosomes serve to adjust the chromatin structure causing transcriptional activation or repression.[72] While the ATAC complex helps facilitate nucleosome sliding activities of the SWI-SNF and ISWI remodeling complexes, ATAC itself cannot stimulate remodeling activity.[72]

The ATAC complex is a part of many cellular activities on its own, however it is also able to form additional complexes. ATAC can interact with the Mediator coactivator complex (MED). MED helps bridge proteins, such as transcription factors, to regulate the transcription of the preinition complex (PIC), through the loading RNA Pol II onto promoters.[74] When ATAC and MED interact they join together forming the meta-coactivator complex (MECO).[75] Assembly of the MECO complex occurs in certain types of cells, such as mouse embryonic stem cells or fibroblasts.[75] The interaction between the ATAC and MED complexes in MECO is mediated by the leucine zipper motif-containing protein 1 (LUZP1), however other factors likely contribute to the bridging of the complexes.[75] Within cells the MECO complex functions to regulate non-coding RNA genes that are transcribed by RNA Pol II.[75]

As well as being involved in the MECO complex, ATAC also interacts with a variety of other proteins including, Xeroderma Pigmentosum Complementation Group C (XPC), which is a DNA damage sensor, known for its roles as a nucleotide excision repair factor.[76] XPC also functions in the absence of stress, where it will co-localize with RNA Pol II at promoters in fibroblasts.[76] At these promoters there are high levels of E2F transcription factor 1 (E2F1), which interacts with XPC mediating their localization at promoters.[76] Through XPC's interaction with KAT2A, XPC is able to recruit the ATAC complex to XPC-dependent gene promoters.[76] This newly formed XPC-E2F1-KAT2A/ATAC complex at XPC-dependent gene promoters allows for the acetylation of histone H3K9.[77] This shows that XPC acts to remodel chromatin through the recruitment of ATAC, utilizing RNA Pol II to transcribe XPC-dependent genes.[76]

Another protein ATAC interacts with is actin. Actin is an key component of cells, helping to maintain the cytoskeleton and is involved in nuclear functions, such as gene expression.[78] In gene

expression, actins role is partially mediated through its ability to interact with RNA Pol II at transcriptional start sites and aids in the assembly of the PIC.[77,79] *In vitro,* actin is a negative regulator of the acetyltransferase properties of KAT14.[78] The ATAC complex interacts with actin, through direct binding of actin to the C-terminal end of hKAT14.[78] This interaction significantly reduces the catalytic capabilities of KAT14, resulting in a loss of H4 acetylation.[78]

1.2.1.4 ATAC Complex in Mammals

As previously described, the ATAC complex is involved in varying cellular processes, in addition to these functions ATAC is also critical in mammalian functioning. Certain subunits of the ATAC complex are implicated in different stages of development. *Kat2a* is integral for embryonic development in mice, and when lost there is embryonic lethality at E10.5, with the embryos displaying increased levels of apoptosis in dorsal mesodermal structures.[80] Knockout mouse models of *Kat14* have also been examined. In these models no mice survive to birth, with embryonic lethality occurring between day E8.5-11.[69] Knockout *Kat14* embryos displayed a profound loss of histone H3K9, H4K5/12/16 acetylation.[69] These embryos lack tissues of the developing heart, and presented with developmental delays, and increased levels of apoptosis.[69] *Kat14* null embryos also feature cell cycle abnormalities, with a twofold increase of G_2/M and G_1 phase cells and a decrease in S phase cells.[69] This indicates that the ATAC complex is important in cell cycle progression in mammals.

In mammalian cells, ATACs involvement in signaling cascades has been studied. When HeLa cells are treated with 12-O-tetradecanoylphorbol-13-acetate (TPA), an activator of the PKC signaling pathway, there is significant localization of ATAC subunits to three different immediate early genes, *c-Fos, Fos-related antigen 1* and *early growth response protein- 1*.[81] ATAC was found to be required for the regulation of these immediate early genes, in conditions of inactivation or stress activation.[81] However, there was no increase in localization of SAGA subunits, indicating only ATAC is brought to the promoter regions of TPA-induced genes. This represents one differential role in stress regulation between ATAC and SAGA in mammals.[81]

1.2.2 Diamond Blackfan Anemia

1.2.2.1 Clinical Presentation

Diamond Blackfan Anemia is a rare inherited blood disorder and is considered a congenital red cell aphasia.[82] In Diamond Blackfan anemia the bone marrow fails to produce mature erythrocytes, due to a defect in erythroid differentiation.[82] Impairments in the bone marrow likely stem from alterations in embryonic development, since patients are identified within the first year of life.[82]

Patients with Diamond Blackfan anemia present with anemia, which is characterized by a marked decrease in red blood cells, with near normal levels of neutrophils and platelets.[82] Other than anemia, patients clinical presentations may include: short stature, webbed necks, microcephaly, abnormal thumbs, and cleft lip or palate.[82] Diamond Blackfan anemia can cause other multi-systemic affects, leading to abnormalities in cardiac, genital and facial organization or functioning.[82] In addition to these phenotypes, Diamond Blackfan anemia is also a cancer predisposition syndrome and patients have an elevated risk of developing cancers like leukemia.[83]

Diamond Blackfan anemia is a rare disorder, with approximately seven cases identified per million births.[83] Seventy percent of Diamond Blackfan anemia cases are attributed to mutations in ribosomal protein genes.[83] The remaining thirty percent of cases of are idiopathic and have no known causes.[83] Common ribosomal protein mutations in Diamond Blackfan anemia include, *ribosomal protein SIP* (*RPS19*), *RPL5, RPL26* and *RPL11*.[83] Despite knowing these mutations occur, little is known about the mechanistic features of Diamond Blackfan anemia.

1.2.2.2 Mechanism of Pathogenesis

The exact mechanism of pathogenesis of Diamond Blackfan anemia is currently unknown, however there are hypotheses that speculate the mechanism of pathogenesis. First, Diamond Blackfan anemia is considered an intrinsic defect of the bone marrow, particularly in erythropoiesis. This explains why bone marrow transplants can be curative, since the microenvironment is not involved.[83–85] A major area of study in the pathogenic mechanism of Diamond Blackfan anemia is the tumor suppressor p53. p53 regulates expression of target genes, including *p21*.[86,87] Alterations in p53 target genes are found in erythroid cells of Diamond Blackfan anemia patients.[86,87] Patients with ribosomal protein gene mutations causing Diamond Blackfan anemia have an upregulation or increased activation of p53 in erythroid cells leading to decreases in erythroid cell proliferation.[86,87] Mutations in *RP* genes are thought to affect the RP's ability to sequester mouse double minute 2 homolog (MDM-2), an inhibitor of p53, thus p53 and MDM-2 will not bind and MDM-2 will not ubiquitinate p53.[86] Therefore proteasomal degradation of p53 will not occur, leading to higher levels and greater activation of p53.[86] The upregulated p53 accumulates in the erythroid lineage of hematopoietic cells, particularly within erythroid progenitor cells, ultimately causing apoptosis of the progenitor cells.[87] This mechanism explains why erythrocytes are selectively lost in Diamond Blackfan anemia.[87]

An additional complication to Diamond Blackfan anemia's pathophysiological mechanism is the imbalance between globin synthesis and free heme.[88,89] Diamond Blackfan anemia patients

have an abundance of free heme and reduced level of globin syntheses.[88,89] This imbalance leads to the production of reactive oxygen species and increased levels of apoptosis in erythroid progenitor cells.[88,89]

Depending on which RP is mutated in Diamond Blackfan anemia, there could be different mechanisms of pathogenesis. For example mutations in *RPS19* cause reductions in erythroid progenitor cell proliferation, whereas mutations in *RPL11* cause reductions of progenitor cell proliferation with increases in apoptosis and delayed differentiation.[86] However mutations in both proteins cause p53 pathway activation an activation and feature cell cycle arrest in G_0/G_1.[86]

The p53 pathway is not the only cascade to be affected in Diamond Blackfan anemia. Patients with Diamond Blackfan anemia have profound dysregulation of the TGF-β (transforming growth factor-beta) signaling pathway.[82,90] The TGF-β signalling pathway is known for its functions in the control of cellular proliferation and degradation.[91,92] Two specific functions of the TGF-β pathway are aiding in erythroid differentiation, as well as acting as a negative regulator of stem cell proliferation.[90] In Diamond Blackfan anemia patient induced pluripotent stem cells, there is a dramatic increases in JNK activation.[90] JNK is one member of the non-canonical TGF-β pathway.[90] This indicates that the TGF-β is either affected by Diamond Blackfan anemia, or is involved in pathogenesis of the Diamond Blackfan anemia phenotype.

1.2.3 Diamond Blackfan Anemia and KAT14 Rationale

Recently Dr. Pankaj B. Agrawal identified a patient with *de novo* heterozygous mutations in the *KAT14* gene, presenting with Diamond Blackfan anemia. We plan to assess if the mutations found in the patient are causing alterations to the KAT14 protein, namely KAT14's ability to act as an acetyltransferase, cellular localization and its binding capabilities and structural integrity within the ATAC complex. Furthermore, we wish to understand the pathogenic mechanisms of the mutated protein, and if they are related to any previously known theories of Diamond Blackfan anemia pathogenesis. In the future, we hope to bridge the understanding of how alterations in the ATAC complex affect TGF-β signaling and JNK, resulting in a novel form of Diamond Blackfan anemia. Future works aim to see if the mutations in KAT14 alter p53 signalling, which has previously been seen in Diamond Blackfan anemia patients. This discovery would link the ATAC complex to p53 regulation for the first time. Finally, in analyzing *KAT14* mutations and the ATAC complex as a whole, we hope to better understand the functions of KAT14 and the ATAC complex and their roles in development.

1.3 Global Developmental Delay Syndromes and WDR5

1.3.1 Global Developmental Delay and Intellectual Disability

1.3.1.1 Clinical Presentation and Diagnosis

Global development delay is characterized by a meaningful delay in at least two of the following areas: gross or fine motor development, cognition, daily activities, speech or language, social and personal.[93] Cases of global developmental delay are diagnosed before the age of five.[93] Although not part of the characterization of global developmental delay, many patients also have intellectual disabilities. Intellectual disability is not only a function of IQ.[94] Instead, intellectual disability can be diagnosed if three criteria are met: 1) there are impairments in intellectual functions noted by standardized intelligence tests and clinical assessments (problem-solving, learning, planning, etc.), 2) the impairments in adaptive functioning are great enough to prevent the fulfilment of civic duties or personal independence, and 3) these intellectual shortcomings present early on.[93]

Diagnosing both global developmental delay and intellectual disability early in life is beneficial to patients. This allows time to prevent complications associated with these disorders, since they often co-occur with multi-system affects, or larger problems in neuronal structure.[93] Understanding these problems can help improve the quality of life and overall life span of patients.[93] An example of a complication is seizures, which are common in global developmental delay disorders. If caregivers and physicians are aware of the possibility that negative side effects, they are better prepared to identify and manage them should they arise. Early diagnosis allows for treatments and support systems to be put in place ensuring for the safest and most normal life for the patients.[93] Such treatments could include specialized educational programs, speech therapy or physiotherapies to treat the language and motor delays, respectively.

1.3.1.2 Prevalence

Global developmental delay and intellectual disability are both found readily in patients, with one-to-three percent of the population are identified to have one or both of these disorders.[93] However, given a potential lack of testing, or the inability to obtain a diagnosis, it is likely these numbers are underreported. In areas where the population is of a lower socioeconomic status, levels of global developmental delay and intellectual disability will be higher than reported.

One family of protein complexes implicated in both intellectual disability and global developmental delay is the MYST family of acetyltransferases. The next section features an in-depth analysis of MYST complexes.

1.3.2 MYST Family Complexes

1.3.2.1 MYST Family of Acetyltransferases

The MYST (Moz, Ybf2/Sas3, Sas2, Tip60) family of KATs is named after the four founding members. In mammals there are five MYST proteins, KAT5, KAT6A, KAT6B, KAT7 and KAT8. Each member of the MYST family contains a hallmark MYST domain, is a conserved residue needed for both the catalytic and acetyl-CoA binding capabilities of the protein.[95] MYST proteins also feature a C2HC zinc finger.[96] The roles of each MYST protein and proteins within their respective multi-protein complexes will be discussed here after.

1.3.2.2 MYST Complex Composition and Functions

1.3.2.2.1 KAT5 Complexes

KAT5 (Tip60) is one member of the MYST family. Initially, KAT5 was noted for interacting with HIV-1 Tat.[97] KAT5, like the rest of the MYST proteins, is an acetyltransferase with the strongest affinity for histone H4K12. However, the protein can also acetylate histone H4K5/8/16, and H3K14.[97] KAT5 has the ability to act on non-histone targets, like p53, which requires acetylation for activation.[97] This enzymatic reaction is particularly important for activating p53s apoptotic and cell cycle arrest functions, helping to remove DNA damage.[97]

KAT5 is a member of the NuA4 (nucleosome acetyltransferase of H4) complex. In humans the NuA4 complex contains 16 different subunits: transformation/transcription domain associated protein (TRRAP), hDomino(p400), KAT5, enhancer of polycomb homolog 1 (EPC1/EPC-like), bromodomain containing 8 (BRD8), vacuolar protein sorting 72 homolog (YL-1), DNA methyltransferase 1-associated protein 1 (DMAP1), inhibitor of growth 3 (ING3), RuvB like AAA ATPase 1/2 (RUVBL1/2), YEATS domain containing 4 (GAS41), Actin, BRG1-associated factor 53A (BAF53a), MORF-related gene 15 protein (MRG15), Esa1 associated factor 6 (Eaf6) and MRG domain binding protein (MRGBP).[98] Many of these proteins act as epigenetic regulators. The proteins BRD8 (bromodomain), GAS41 (YEATS), KAT5 (Chromodomain) and ING3 (PHD Finger) all act as histone readers being able to recognize and bind to either modified or unmodified histone residues, allowing the complex to localize to the chromatin and exert transcriptional functions.[98] Subunits within the complex are known to contain domains important in chromatin remodelling including, hDomino (HAS/SWI2/SANT), and DMAP1 (SANT).[98] These domains allow the NuA4 complex to target specific promoter regions in order to activate transcription of select genes. [98]

The NuA4 complex has vast functions involving its many subunits. In humans NuA4 is an activator of transcription. This activation allows for cell cycle progression.[98] NuA4 plays a role in the DNA damage response, particularly in nonhomologous end joining.[98] Through the presence of the helicase subunits RUVBL1/2, the complex plays a role in homologous repair, through Holiday junction movement.[98] These are a few examples of NuA4 protein functions.

Along with the NuA4 complexes roles in cellular processes the complex also has importance in mammalian physiological functions. KAT5 is integral to mammalian development. With homozygous loss of *Kat5*, mice present with embryonic lethality near the blastocyst phase, with increases in cell death and an attenuation of cellular proliferation.[99] These results indicate that KAT5 is critical for cellular survival during embryogenesis. KAT5 is critical to CNS development, however, developmental models looking at the CNS have not been studied due to the embryonic lethal phenotype of *Kat5* knockouts.[100] KAT5 is found in the mammalian forebrain and hippocampus.[101] When adult mice undergo a conditional knockout of *Kat5* in the forebrain, the hippocampal CA1 region undergoes mass dysregulation of gene expression.[101] Additionally, hypoacetylation of histone H4K12 occurs at transcriptional start sites of many genes including those involved in neuroplasticity.[101] Ultimately, the loss of *Kat5* in mice forebrains lead to neuronal degradation in the hippocampus.[101]

While KAT5 is fundamentally critical in mice, other binding partners also have developmental importance. One NuA4 subunit, ING3, is traditionally a tumor suppressor protein. ING3 is critical for the formation of the prenatal brain and is required for mammalian embryonic development.[102] Like *Kat5*, when *Ing3* is lost in mice, embryonic lethality occurs, at E10.5.[102] These knockout mice display growth retardation, abnormal neural tube closure, developmental disorders and ectodermal differentiation is lost.[102] Both KAT5 and ING3 show the importance of the MYST family NuA4 complex in mammalian development and neurological functioning.

1.3.2.2.2 KAT6A AND KAT6B COMPLEXES

Another set of protein complexes in the MYST family of acetyltransferases are KAT6A and KAT6B. The MOZ/MORF or KA6A/KAT6B complexes are tetrameric, containing four subunits: Bromodomain and PHD finger-containing protein 1 (BRPF1), Inhibitor of Growth 5 (ING5), homolog of Esa1-associated factor 6 (hEAF6) and either KAT6A (MOZ) or KAT6B (MORF).[103] Both KAT6A and KAT6B acetylate histone H3, with the highest affinity for H3K23.[95,100,103] KAT6A and KAT6B serve as transcriptional coactivators of the MOZ/MORF complexes, since neither protein directly binds to the DNA.[104] The activator of the MOZ/MORF complex is the

multivalent chromatin reader BRPF1.[103,104] BRPF1 acts as a scaffold within the complex, bringing together either KAT6A or KAT6B with ING5 and EAF6.[103] Within the complex BRPF1 also acts as a reader containing two reader domains a Pro-Trp-Trp-Pro (PWWP) domain and a bromodomain, which are able to recognize and bind to unmodified or acetylated histone H3, respectively.[105–107] BRPF1 is vital in acetylation, acting through KAT6A and KAT6B to control acetylation of histone H3K23.[108] The subunit ING5 is an epigenetic regulator found to be highly expressed in stem cells and involved in DNA replication through direct interactions with MCM (mini-Chromosome-Maintenance) helicases.[109,110,98,111] Finally, the subunit hEAF6 is known to aid in the stabilization of the MOZ/MORFcomplexes.[98]

The MOZ/ MORF complexes are crucial in mammalian development and physiological functioning. KAT6A is important during embryonic development, acting to maintain fetal hematopoietic stem cells.[112,113] Loss of *Kat6a* in mice results in embryonic lethality at E14.5, with fetuses displaying altered hematopoiesis and a loss of blood.[112] The absence of *Kat6a* leads to senescence of neural stem cells, and causes altered brain development.[114] Similar to KAT6A, KAT6B is also imperative in development and in the regulation of neural stem cells.[115–117] *Kat6b* null mice, died prematurely, at approximately three weeks of age, with pups displaying abnormal brain development, craniofacial abnormalities and cerebral defects.[118] In addition to their roles in development, both *KAT6A* and *KAT6B* can be altered in cancer, both undergoing translocations in acute myeloid leukemia (AML), and are mutated in solid tumour malignancies.[25,119–121]

Like KAT6A/KAT6B, BRPF1 is also integral in mammalian development. BRPF1 is required for embryogenesis and when *Brpf1* is knocked out in mice embryonic lethality occurs at day E9.5.[122] BRPF1 is critical for neuronal development, including structures such as the cerebral cortex, hippocampus and the dentate gyrus.[123,124] Moreover, BRPF1 is vital neural stem cells regulation, and neocortex organization.[123] When *Brpf1* is knocked out in mice it leads to abnormal neogenesis and agenesis of the corpus callosum.[124] Knockout *Brpf1* mice have alterations in histone modifications, with a deficiency in both histone H3K23 acetylation and propionylation.[108,125,126] *Kat6a*, *Kat6b* and *Brpf1* mouse models highlight the importance of the MOZ/MORF complexes in mammalian development.

1.3.2.2.3 KAT7 COMPLEXES

KAT7, or HBO1 (HAT binding to ORC1) is a part of two different types of tetrameric complexes. This lysine acetyltransferase was initially identified through its interaction with the origin recognition complex protein 1 (ORC1). KAT7 can acetylate both histone and non-histone

targets, such as proteins in the ORC complex, including: ORC2, cell division cycle 6 and MCM2. The acetylation of ORC complex proteins initiates DNA replication.[127,128] KAT7 is found within two different HBO1 containing complexes, both featuring ING4/5, mEAF6, and either jade family of PHD finger (JADE) 1/2/3 or BRPF2/3. The main purpose of the JADE and BRPF subunits is to change the selectivity of KAT7.

The first KAT7 complex features differential JADE proteins. JADE is a histone reader containing two PHD domains which recognize and bind to tri-methylated H3K4.[129] JADE1 is the most common JADE protein interacting with KAT7 and binds directly to the chromatin to specifically encourage KAT7 acetylation of histone H4K5/8/12.[130,131]

The second KAT7 complex features BRPF homologs, with KAT7 having the highest affinity for BRPF2. KAT7 binds to BRPF2's N-terminal region, which is both necessary and sufficient for the interaction.[132,133] This interaction stabilizes BRPF2 protein levels and inhibits protein degradation.[132,133] Much like BRPF1, BRPF2 also contains bromodomains. These bromodomains are histone readers, binding to acetylated histone H3 and directing KAT7 to acetylate histone H3, particularly H3K14.[105,130] In some cases BRPF3 can form a tetrameric complex with KAT7, however BRPF2-KAT7 complexes are more abundant. BRPF3 is not well characterized due to its non-essential roles in development and survival, which is opposite to its paralogs BRPF1/2.[134]

Proteins ING3/4/5, and mEAF6, are found in all KAT7 complexes and their functions have been discussed in sections 1.3.2.2.1 and 1.3.2.2.2 . In both complexes, BRPF2 and JADE protein homologs act as scaffolds, bringing together KAT7, ING3/4/5 and mEAF6.[130]

KAT7 complexes have diverse cellular functions. Histone H3K14 acetylation by KAT7-BRPF2 complexes regulates immune system development.[135] When the *Kat7-Brpf2* complex in mice is lost, maturation of peripheral T cells halts, and T cell survival is diminished.[135] Conversely KAT7-JADE complexes are important in DNA replication and associate with the MCM complex, which assembles onto the replication origins on the chromatin.[136] The interaction with the MCM complex is mediated through the co-activating capabilities of KAT7 on chromatin licensing and DNA replication factor 1.[132]

Similar to other MYST family members, KAT7 complex subunits are integral for development. The KAT7-BRPF2 complexes acetyltransferase capabilities on histone H3K14 are especially critical for survival. When KAT7 is lost, there is a ninety percent reduction histone H3K14 acetylation causing a reduction in gene expression of genes regulating embryonic patterning.[137] The loss of *Kat7* in mice caused developmental arrest and lead to embryonic lethality between day

E10.5-11.5.[137] These mice did not feature alterations in DNA replication or cell proliferation in development.[137] Through this knockout mouse model, KAT7 was deemed critical in the regulation of embryonic patterning.

BRPF2 is also essential to development. The KAT7-BRPF2 complex helps regulate erythropoiesis in development.[133] When *Brpf2* is knocked out in mice embryonic lethality occurs at E15.5, and a global decrease in histone H3K14 acetylation is found.[133] Furthermore, these embryos displayed growth retardation, and in some cases either abnormal optic lenses, or a lack of fusion of the neural tube.[133] These knock out embryos had anemia, due to failed fetal hematopoiesis, and a reduction in of erythroblasts.[133] BRPF2 is essential for differentiation of erythroid cells, and this function is dependent on the interaction between BRPF2 and KAT7.[133]

As well as playing a role in hematopoiesis, BRPF2 is also critical in neuronal functioning. BRPF2 is abundantly expressed in the mammalian brain particularly in the proximal dendrites and perikaryal cytosol of CNS neurons.[138] *BRPF2* is highly expressed in early stages of embryonic development, with the highest concentration in early neuroblasts and neuroepithelial layers.[138] Utilizing a mono-allelic inactivation of *Brpf2* in mice brain development was studied.[139] Mice with the inactivation presented with brain abnormalities, including a reduction in volume of the amygdala and the striatum and a dysregulation of the neurotransmitters dopamine and serotonin.[139,140] Within the striatum, there was an overall loss of neurons, particularly in spiny medium sized neurons.[139] Similar to the embryonic knock out, *Brpf2* mice with monoallelic inactivation displayed a hypoacetylation of histone H3K14 within the cerebrum.[140] These mice feature behavioural changes, such as, increased aggression, less passive interactions, a lack of affinity for social interactions and deficits in both long term memory (lack freezing after fear conditioning), and working memory.[140] *Kat7* inactivated mice had imbalances with excitatory and inhibitory synaptic functions, with the mice featuring a cerebral dysfunction of GABAergic neurotransmission.[140] These examples show BRPF2 is critical for embryonic development, and neuronal functioning.

The function of JADE1 in development is well studied. JADE1 is expressed in mammalian embryogenesis, with high expression in mouse embryonic tissues such as placental components, extraembryonic ectoderms, trophoblasts, and neural progenitors.[141] JADE1 also regulates genes apart of developmental structures, including the tail bud and primitive streak.[141] Through studying the dynamics of a fusion product, mJade1-b-Gal, it appears that JADE1 is involved in the processes of elongation and determination in the anterior posterior axis.[141]

1.3.2.2.4 KAT8 COMPLEXES

Similar to KAT7, KAT8 is also a part of two complexes, the male-specific lethal (MSL) complex and the non-specific lethal (NSL) complex.

The MSL complex was initially identified in drosophila and is required for appropriate X chromosome compensation. In drosophila, males do not have a Y chromosome instead they have a single X chromosome, which is different to the females X chromosomes.[142] In order to rectify this issue the male X chromosome must be upregulated.[142] Within the MSL complex there are five subunits, KAT8, male-specific lethal 1/2/3 (MSL1/2/3) and maleless (MLE).[143] These five proteins are critical for survival in male drosophila, and play an important role in dosage compensation.[143] The MSL complex specifically binds to the male X chromosome and acetylates histone H4K16, helping mediate chromosome puffing and the up regulation of genes to compensate for the loss of a chromosome.[144] Two non-coding RNAs on the X chromosome called *roX1/2* attract the MSL complex, causing localized spreading of the MSL complex on male X chromosomes.[145] The MSL complex is thought to be targeted to active chromatin, and is generally co-localized with RNA Pol II proteins.[144]

While the MSL complexes most notable role is in drosophila dosage compensation, it has alternative functions in mammals. In mammals the MSL complex consists of the core KAT8 and MSL1/2/3 subunits.[146] The MSL complex in humans, is integral for acetylation of histone H4K16, similar to the complexes role in drosophila.[146] However, unlike in drosophila, hMSL is not involved in male characteristics, and does not specifically associate with either the X or the Y chromosome, instead hMSL is involved in global acetylation levels of H4K16.[146]

The NSL complex is another KAT8 containing complex that primarily acts to acetylate histone H4K5/K8/K16.[147–149] As a whole, the NSL complex helps regulate cellular processes. Subunits in the complex are found to be enriched at the promoters of constitutively active housekeeping genes.[147] This enrichment at promoters occurs to regulate the initiation of RNA Pol II transcription.[147] Promoters that NSL subunits localize around, include members of the PIC, indicating, the NSL complex is integral for the efficient recruitment of RNA Pol II to target genes.[147] In addition to initiating RNA Pol II, the NSL complex helps regulate transcription on a genome-wide level.[150] Due to the NSL complexes catalytic properties, it is believed the NSL complex acetylates histone H4K16 allowing for NURF to localize at target promoters.[151] Another attribute of the NSL complex is its ability to regulate signalling networks that are required for

proper cellular homeostasis, and thus localize at promoters and enhancers of microtubules and mitochondria, both of which are key in cellular functioning.[152]

Another facet of the NSL complex is its ability to interact with other protein complexes. The best example of this comes from the dynamics between NSL and the MLL/SET complexes. The MLL/SET family of complexes are methyltransferases, acting to methylate histone H3K4. Both complexes have two shared subunits, WDR5 and HCF1.[153] The NSL complex has be found to promote the methyltransferase capabilities of the MLL/SET complexes, in an histone H4K16 acetylation dependent manner.[153] In order to exert its affect, the NSL complex acts upstream of MLL/SET complex, meaning changes in MLL/SET complex do not affect the NSL complexes ability to act as an acetyltransferase.[153] Given the relationship between the NSL and MLL/SET complexes, it suggests that the interplay between these complexes may be important in coordinating gene transcription or regulation.[153]

The NSL complex is made up of nine subunits. Within the complex KAT8 functions as the acetyltransferase. KAT8 binds to a protein called KANSL1 (KAT8 Regulatory NSL Complex Subunit). KANSL1 acts as the scaffolding member, tethering the complex together, and helps potentiate the catalytic activities of KAT8. [148,152] In the NSL complex, KANSL1 binds to the histone reader WDR5.[66,154] The interaction between KANLS1 and WDR5 is integral for the complexes ability to interact with target gene promoters.[148] WDR5 is a part of many different epigenetic complexes, and is recognized for its value in facilitating protein-protein interactions. In the complex, WDR5 bridges the connection between KANSL1 and KANSL2. KANSL2 is not a well characterized protein. The drosophila homolog of KANSL2, dNSL2 is critical for the viability of drosophila.[155] KANSL2 has been shown to be upregulated in glioblastoma cells, and is responsible for helping to regulate the stemness of these cancer cells.[156] KANSL3 is the final KANSL protein in the complex, and is required for KAT8 to bind to mitochondrial DNA, for the regulation of respiratory genes involved in oxidative phosphorylation.[157] Through KANSL3, the NSL complex helps to control the regulation of both nuclear and mitochondrial DNA.[157]

KANSL3 binds to another subunit in the NSL complex, O-linked-N-acetylglucosamine transferase 1(OGT1).[158] OGT1 is a specific transferase that catalyzes the addition of *O*-Linked β-*N*-acetylglucosamine onto serine or threonine residues.[159] OGT1 *O*-GlcNAcylates KANSL3, which stabilizes KANSL3's interaction in the NSL complex.[158] By stabilizing the interaction between KANSL3 and the NSL complex, OGT1 helps regulate the NSL complexes global histone acetylation properties.[158]

OGT proteins interact with another NSL subunit, host cell factor 1 (HCF1). [160] HCF1 is a transcriptional co-regulator, and can be *O*-GlcNAcylated.[160] HCF1 acts as a histone reader and writer, although it cannot bind to the DNA directly, it can interact with other chromatin binding proteins.[161] One of HCF1's roles is acting as a cofactor of the E2F (E2 factor) transcription factor, which functions in the G1/S phase transition to help regulate the cell cycle.[161]

Another protein in the NSL complex is microspherule protein 1 (MCRS1). MCRS1 is a microspherule protein, and localizes in nucleoli, specifically in electron dense bodies.[162] The protein acts in cilium formation and is an important factor in cargo transport to the centrosome in a dynein-dependent manner.[162] MCRS1 is found in other complexes, including the chromatin remodelling complex, INO80.[163]

The final subunit of the NSL complex is PHD finger protein 20 (PHF20), a histone reader containing both tudor and PHD finger domains.[164] These domains can identify and bind to H3K4me2 allowing for the acetylation capabilities of the NSL complex to proceed.[164] PHF20 is also able to interact with methylated residues of p53, regulating the stabilization and activation of p53.[165] This interplay between methylation, the NSL complex and p53 is a potential mechanism for rapidly spreading open chromatin, through histone H4K16 acetylation.[164] Through the wide array of subunits, the NSL complex is involved in a variety of cellular processes, which are critical to cellular performance.

The NSL complex is integral in mammalian development. KAT8 is a critical component of embryogenesis and when *Kat8* is lost in mice embryogenesis fails, leading to death prior to implantation.[166] These knockout embryos display a profound loss of histone H4K16 acetylation, abnormal aggregation of the chromatin, DNA fragmentation and increased levels of apoptosis.[166] Reduced H4K16 acetylation has genome wide effects, such as defective double strand break repair or impairments to the DNA damage response.[167] Mice, featuring *Kat8* purkinjie cell knockouts were also generated and displayed cerebellar dysfunction, ataxia and defects in motor coordination indicating *Kat8* is critical for post mitotic survival and maintenance of purkinjie cells.[168] Cerebral knockouts of *Kat8* in mice leads to a profound loss of both histone H4K16 acetylation and propionylation in the cerebrum.[32] These mice all die before three weeks of age, however abnormalities can be found beginning on E16.5.[32] A cerebrum specific loss of *Kat8* also causes dysfunction in the development of neural stem and progenitor cells (NSPC).[32] Along with the alterations in NSPCs, the hippocampus and neocortex show cerebral hypoplasia, aberrant neurogenesis, a dysregulation of proliferation in cerebrocortical neuronal epithelia and a dramatic

increase in apoptosis.[32] These mouse models of *Kat8* indicate *Kat8* is integral in embryogenesis, purkinjie cell maintenance and neuronal development as a whole.

KANSL1 mouse models have also been studied. Mice with homozygous mutations in the *Kansl1* gene display impairments in memory, which may be attributed to alterations in hippocampal cells.[169] Hippocampal cells in the *Kansl1* mutant mice show an upregulation in genes that control neurogenesis, synaptic transmission and overall chromatin organization.[169] These mutant mice suffer from a dysregulation in promoters controlling CA1 neuronal cells.[169] Similar to KAT8, KANSL1 is imperative in neuronal development of mammals.

PHF20 is also developmentally important in mammals. A knockout *Phf20* mouse model led to lethality the first day after birth, with mice displaying abnormalities including: growth defects, skeletal composition, hematopoietic abnormalities and lymphocyte development problems.[170] Although PHF20 is not critical to the maintenance of global acetylation levels of histone H4K16 it does work downstream of acetylation to regulate the transcription of NSL gene targets.[170]

Other members of the NSL complex, MCRS1, OGT1 and HCF1 are also involved in developmental processes. When *Mcrs1* is mutated in mice, early embryonic lethality occurs to a failure in implantation, abnormal inner cell masses in the blastocysts and a reduction of cells in the epiblast lineage . In an *Ogt* deletion mouse model embryonic lethality occurs and in a model featuring cardiomyocyte-specific loss of *Ogt* severe symptoms of heart failure were found.[171,172] The self-renewing capacity of embryonic stem cells is decreased with the loss of *Ogt* in mice.[173] Mouse models of *Hcf1* have also been analyzed. In epiblast specific *Hcf1* knockout mice embryonic lethality occurs at E8.5.[174] These *Hcf1* null mice had unique features including a loss of the gastrulation phase developmental arrest at E6.5, alterations in the cell cycle and endoderm migration patterns.[174] As one can see, many of the members of the NSL complex are integral for development and neuronal processes.

1.3.2.3 MYST Complexes and Human Disorder

1.3.2.3.1 KAT5 Complex Mutations in Humans

Subunits of the NuA4 complex are important in humans, with abnormalities in these proteins resulting disease. One example of this comes from alterations in KAT5 levels. In cases of Alzheimer's disease there is an imbalance between levels of KAT5 and HDAC2.[175] To confirm this notion post-mortem human hippocampal tissue was studied from those with Alzheimer's, where levels of KAT5 were found to be decreased in both neuronal and glial cells.[175] In this study KAT5s cellular localization was altered, with KAT5 missing from the nuclei.[175] The lack of KAT5

in the Alzheimer patients hippocampal tissue corresponds to a reduction in KAT5 specific acetylation, including acetylation of histone H4K5/12/16.[175] While there is a loss of KAT5 an abundance of HDAC2 is found, indicating deacetylation is far more prevalent, and the balance between acetylation and deacetylation is lost.[175] It is thought that the increased levels of HDAC2 and loss of KAT5 mediate effects on repressing plasticity genes, through amyloid precursor proteins, before the formation Aβ plaques.[175] Although most of the models currently used to study the role of KAT5 in Alzheimer's disease are in drosophila, there is some evidence that increasing the levels of KAT5 may help to re-establish the balance between acetylation and deacetylation in the brain, helping to restore cognition and brain morphology.[175]

Recently, *KAT5* has been implicated in a different neuronal disorder. Three patients with *de novo* heterozygous missense mutations in *KAT5* have been identified.[176] These patients present with global developmental delay, intellectual disability, and features such as cerebral malformations and seizures.[176] A notable phenotype of these patients is the presence of facial dysmorphisms such as a flat facial profile, a round face and down slanting palpebral fissures.[176] Through genome editing, patient *KAT5* variants were introduced into the human cell line K562, where mutant KAT5 proteinz had impaired acetylation capabilities, leading to an education in histone H4 acetylation, similar to patients with Alzheimer's.[176] Both Alzheimer's and this novel global developmental delay phenotype arise when KAT5 is impaired or mutated showing the proteins importance in neuronal functioning.

Another member of the NuA4 complex that is critical in humans is TRRAP. Within the NuA4 complex, TRRAP acts as a scaffold to recruit the complex to the chromatin.[177] To date 24 patients with *de novo* missense variants in *TRRAP* have been identified.[177] In these patients two clinical presentations arose. The first is a multi-systemic syndrome, where patients had a range of malformations in the heart, brain, kidney and gastrointestinal systems.[177] The second clinical presentation includes patients with autism spectrum disorders and epilepsy.[177] A commonality between the two presentations was the presence of global developmental delay, intellectual disability and facial dysmorphisms, such as up slanted palpebral fissures, wide nasal bridge or thin upper lips.[177]

1.3.2.3.2 MOZ-MORF Complex Mutations in Humans

Similar to mutations in *TRRAP*, alterations in the MOZ-MORF complexes can also lead to neurodevelopmental disorders. Mutations in *KAT6A* causing global developmental delay were first reported in 2015.[178,179] Initially ten individuals with *de novo* heterozygous mutations in *KAT6A*

and one with a microdeletion including *KAT6A* were identified.[178,179] All the *de novo* mutations identified occurred in the C-terminal region of *KAT6A,* with certain locations in the gene being found to be hotspots for mutations.[178,179,180] A large portion of mutations, including those in hotspots, are located within region encoding the genes acidic domain. However, mutations can occur throughout the *KAT6A* gene.[180]

Following the discovery of the KAT6A Syndrome 86 additional cases were discovered.[178–190] Universally found features among patients include: global developmental delay, intellectual disabilities, speech and motor delays.[178–190] Other features found include: microcephaly, neonatal hypotonia, strabismus, and feeding defects.[178–190] Facial dysmorphisms such as broad nasal tip and thin upper lips were also present in many patients.[178–190] Varying multi-system defects have been discovered in patients with *KAT6A* mutations and are highlighted in table 2.

Another category of disorders similar to the KAT6A syndrome involves mutations in *KAT6B*. Depending on the location of the mutation in *KAT6B* one of four disorders can occur: Genitopatellar Syndrome (GPS), Say-Barber-Biesecker-Young-Simpson variant of Ohdo Syndrome (SBBYS), a mixed phenotype containing features of both disorders or a subtype of disorder that cannot be specified.

GPS was first defined in 2000.[191] Key characteristics of the disorder include, renal or genital abnormalities, absent patella, developmental delay, intellectual disability and dysmorphic facial features.[191] The GPS phenotype is associated with severe complications. When first defined, it was noted that three children with GPS died in their first few years of life, either spontaneously or due to respiratory distress.[191] Two pregnancies were also terminated due to renal anomalies and microencephaly.[191] Certain features patients present with are particularly damaging, including agenesis of the corpus callosum, and numerous skeletal abnormalities.[191]

In 2012 through exome sequencing, mutations in *KAT6B* were determined to cause GPS.[192,193] The identified mutations in *KAT6B* cluster in the C-terminus of the final exon, within the region encoding the acidic domain.[192,194] These *de novo* truncating mutants do not undergo nonsense-mediated decay (NMD), thus a truncated KAT6B protein is produced lacking the conserved methionine and serine rich domains, and the entire transactivation domain.[192,193,194] Moreover, mutations causing GPS lead to a significant reduction in global acetylation levels of histone H3 and H4.[193] In all, 42 cases of GPS have been identified.[191–193,195–208]

Simpson and Young first noted SBBYS in 1987, in a patient with congenital heart defects, mental handicaps, hypothyroidism and abnormal facial features, such as a bulbous nose.[209]

Following this case, Say, Barber and Biesecker identified patients presenting with developmental delay, blepharophimosis and abnormal facial and dental features.[210,211] Other common features found in SBBYS include a mask like face, bulbus nose, low set ears, patellae problems, and long thumbs and big toes.[212]

Mutations resulting in SBBYS are found throughout the *KAT6B* gene.[213] SBBYS can arise when a mutation occurs in the first 1200 codons.[213] In this case, no truncated KAT6B protein is produced, due to NMD.[213] SBBYS mutations can also occur in the most distal part of the gene, generating truncated protein variants that retain both the acidic and transactivation domain.[213] Since SBBYS's discovery, over of 90 cases have been identified. [202,206,209–212,214–229]

Both GPS and SBBYS occur in different regions of the *KAT6B* gene, thus are considered to be opposite ends of a spectrum of KAT6B disorders.[194,213] Researchers have noted a mixed phenotype arises when mutations occur in other regions of *KAT6B*.[213] Individuals who have these mixed phenotype mutations have hallmark features from both SBBYS and GPS, demonstrating these two disorders are more alike than they originally appeared.[230] Vlckova created an initial system to classify KAT6B disorders: Group One Disorders where mutations occur within the first 1205 codons, and exclusively lead to SBBYS, Group Two Disorders, with mutations occurring between codons 1205-1350 producing GPS, Group Three Disorders where mutations are from codons 1350-1520 leading to mixed phenotype, and finally Group Four Disorders, where mutations between 1520-2073 primarily cause SBBYS.[213,230] In 2016 Radvanskzy went on to generate a fifth group by dividing Vlckoa's original group one into two separate groups, where one would be for variants that lead to haploinsufficiency and have the potential to generate early truncating variants.[194] Whereas, the second group would be for other mutations occurring early in the coding sequence that lead to NMD.[194,230] In total 12 patients with *KAT6B* mutations have been described with mixed phenotypes.[194,202,213,215,230–233]

In 2017 mutations in the gene encoding *BRPF1* were identified in individuals with developmental delays.[108,234] Forty-one individuals have identified with *de novo*, mosaic or dominant inherited mutations in *BRPF1*.[108,125,234–238] Another five individuals were found to have large 3p25 deletions encompassing *BRPF1* and *SETD5* (SET Domain containing 5), a putative histone methyltransferase.[234] Key features found in patients with *BRPF1* mutations include global developmental delay, language and motor delays, intellectual disability, hypotonia, as well as skeletal and visual abnormalities.[235] Another important feature is facial dysmorphisms, including a flat facial profile, down slanting palpebral fissures, small or round ears, and a round face.[235]

Mutations resulting in BRPF1-associated intellectual disability are haploinsufficient, with only one DNA strand mutated.[236] Molecular analysis shows mutations in *BRPF1* lead to reductions in levels of histone H3K23 acetylation and propionylation, due to defects in mutant BRPF1 interacting with and activating KAT6A or KAT6B.[108,125] There is no consistent place where mutations in *BRPF1* occur, however common regions include in the histone and DNA binding domain, in or near the PHD domains, in the ING5/MEAF6 binding domains, and in the C-terminal region.[108,234–236]

There are similarities among all mutations that disrupt the MOZ/MORF complex. KAT6A mutations disrupt histone H3 acetylation, causing increases in H3K18 and decreases in H3K9 acetylation.[178] Additionally, global decreases in histone H3 and H4 acetylation are found in individuals with GPS.[193] Finally, mutations in BRPF1 alter acetylation patterns, changing levels of histone H3K23 acetylation.[108] A common theme among mutated regions in all three genes is they are all evolutionarily conserved. This suggests that the areas in which the mutations are occurring are important to gene function.

1.3.2.3.3 KAT7 Complex Mutations in Humans

Like other MYST complexes, KAT7 complexes have also been implicated in human disease. One of these disorders is AML. In AML, KAT7 acts as a regulator of leukemia stem cell maintenance.[239] Part of the regulation capabilities of KAT7 on leukemia stem cells is due to the KAT7-BRPF2 complexes ability to acetylate histone H3K14, allowing *Hox* genes to be highly expressed.[239] Another disorder that features changes in KAT7 is rheumatoid arthritis. Rheumatoid arthritis patient synovial fibroblasts feature an upregulation of KAT7, which contributes to the stimulation of proinflammatory cytokines.[240] When further studying this phenomenon in human tissue, overexpression of KAT7 leads to the induction of proinflammatory cytokines (TGF-β, interleukin 6), and promoted cellular differentiation.[240] This overexpression was found to promote T_H17 cell migration.[240] Given KAT7's roles in induction of proinflammatory cytokines, and both cellular differentiation and migration, it points to a role in KAT7 enhancing the progression and pathology of rheumatoid arthritis.[240]

Besides KAT7, BRPF2 is also involved in human disorder. A specific marker on *BRPF2* increases the susceptibility to two mental disorders, schizophrenia and bipolar affective disorder.[138] Schizophrenia is characterized by the affected individual's inability to normally interpret reality.[241] Symptoms include delusions, hallucinations or disorganized thinking.[241] Bipolar affective disorder features the oscillation between two extreme moods, depression and

mania.[242] In a Scottish cohort, approximately ten percent of people affected with schizophrenia or bipolar affective disorder have a specific haplotype of *BRPF2*.[138] Across control and affected individuals 11 SNPs were identified in *BRPF2*, six of these were associated with schizophrenia.[243] Further proving the notion that *BRPF2* is involved in schizophrenia, a large genome wide association study meta-analysis on schizophrenia found *BRPF2* to be the most significantly associated gene with the disorder.[244] The schizophrenia phenotype has also been recapitulated as a mouse model, utilizing monoallelic inactivation of the *BRPF2* gene.[140]

1.3.2.3.4 KAT8 COMPLEX ALTERATIONS IN HUMANS

Not only have mouse models proven the need for the NSL complex integrity in development, but human patients have been found with disorders similar to those seen in mice.

Patients have been identified carrying mutations in the *KAT8* gene.[32] In all, nine patients have been identified, each displaying a similar phenotype with a hallmark of global developmental delay and intellectual disability.[32] Eight of these patients featured heterozygous *de novo* mutations in *KAT8*, and one individual with a biallelic variant.[32] All of the individuals with *KAT8* mutations were unrelated, however three individuals share the same mutation (c.269A>G).[32] Along with the phenotype hallmarks many patients also suffered from epilepsy, neurological abnormalities, changes in cranial shape and facial dysmorphisms.[32] Patients may present with multisystem effects, including alterations in vision, ears, hands and feet, or cardiac abnormalities.[32] A full delineation of phenotypes can be found in table 2 . These patients with *KAT8* mutations exemplify KAT8s importance in development.

Another subunit of the NSL complex mutated in humans is *KANSL1*. Mutations in the *KANSL1* gene cause of Koolen-de Vries syndrome, a global developmental delay and intellectual disability disorder.[245] The *KANSL1* mutations are *de novo*, and affect a single allele, therefore this phenotype is caused by *KANSL1* haploinsufficiency.[245] Koolen-de Vries syndrome can also be a result of 17q21.31 microdeletions, where part of *KANSL1* is deleted.[245] Both the mutation and microdeletion of *KANSL1* share similar clinical features, including hypotonia, seizures, craniofacial, heart, skeletal and genito-urinary defects.[246] A further breakdown of the clinical presentation can be found in table 2. Patients with Koolen-de Vries syndrome may have severe CNS malformations, including thinning of the corpus callosum or hydrocephalus.[247] Currently 125 total cases of Koolen-de Vries syndrome have been identified, with over 120 of these being postnatal.[245,247–271] Postnatal and total cases need to be differentiated, since malformations in CNS development can be detected in utero, some parents may decide to terminate the pregnancy.[247] This

can cause a discrepancy between the prevalence of Koolen-de Vries syndrome in the population and the prevalence of *KANSL1* mutations in utero.

While Koolen-de Vries syndrome and KAT8 syndromes are well characterized, other NSL subunits are also linked to human mutations. Ten individuals have been identified with mutations in *OGT*, and an additional three had *OGT* and Mediator of RNA polymerase II transcription subunit 12 (*MED12*) mutations.[272–276] These patients present with X-linked intellectual disability features including, global developmental delay, intellectual disability, neurological abnormalities, and facial dimorphisms.[272,273,276] Patients may also suffer from other multi-system effects including brain cardiac, GI, or visual abnormalities.[276] In patients the mutations appear to cause a reduction in the viable OGT protein produced.[276] Three individuals presented with X-linked intellectual disability when missense mutations in *MED12* occur.[274] These *MED12* mutants co-segregated with mutant gene variants of *OGT*, and were found in three brothers.[274] Each of the brothers also presented with global developmental delay and intellectual disability, as well as other hallmark phenotypes seen in the other *OGT* patients.[274] Finally, two additional cases of *OGT* mutants are missense variants occurring in the region encoding *OGT*'s catalytic domain in two monozygotic female twins.[275] These twins have altered X gene inactivation, and the mutations cause a reduction in OGT's stability impacting OGT's substrate binding regions and feature a similar clinical presentation to the other *OGT* mutant patients.[275] These patients exemplify why OGT is critical for neuronal development in humans.

HCF1 is another subunit of the NSL complex implicated in neurodevelopmental disorders. 24 individuals have been identified with genetic mutations in *HCF1*, including one family with mental retardation, X-linked 3 (MRX3).[277–280] Another name for HCF1, is HCF1C, which stands for the C terminal end of HCF1. Patients with mutations in *HCF1* are found to have neurological impairments, including intellectual disability, global developmental delay, and dimorphic facial features.[277] These mutations can either be gain or loss of function and have opposite effects on neural progenitor cells (NPC), axonal growth, and nuclear localization.[278] For example loss of *HCFC1* causes an over proliferation of NPCs, with decreased differentiation, whereas over expression of HCFC1 causes a lack of proliferation and increases differentiation of NPC.[278] *HCFC1*overexpression mutations also cause increases in astrocyte production.[277]

Fifteen patients with *HCF1C* mutations present with X-Linked Cobalamin disorder.[279,280] Cobalamin disorder is a rare metabolic disorder, in which there is a lack of vitamin B12 derivative (cobalamin) metabolism.[280] Two major cofactors are needed for vitamin B12 metabolism, methyl-

cobalamin (MeCbl) and adenosylcobalamin (AdoCbl).[280] Mutations in these cofactors or other enzymes important for cobalamin metabolism can cause the disorder.[279] Patients with Cobalamin present with multi-systemic effects, including hematological, hepatic, ocular or cardiac defects.[280] The patients with Cobalamin disorder and *HCF1* mutations feature the typical cobalamin disorder phenotype with more severe neurological abnormalities, like developmental delay, epilepsy, facial dysmorphisms brain abnormalities and microcephaly.[279,280] In addition to these individual cases, a family with MRX3 has been identified with an over expression mutation in *HCFC1*.[277] Affected members of this family were male, indicating this is an X-linked intellectual disability.[277] The clinical presentation of the family includes intellectual disability, autistic traits, behavioural problems, and absence seizures.[277] *HCFC1* gene mutations all impact development. Furthermore, *HCF1* and NSL complex mutations all lead to global developmental delay and intellectual disability.

1.3.3 WDR5: A Multifaceted Protein Bridge

1.3.3.1 WDR5 Structure

One important member of the NSL complex is WDR5. WDR5 is a cellular multitasker, and a part of many complexes. *WDR5* is evolutionarily conserved, with all multicellular organisms having nearly identical *WDR5* paralogs.[66] WDR5's protein structure consists of a seven WD40 domains each serving as a β propeller that assembles along the outside of the protein, while two binding clefts are featured on the top and bottom.[66] WDR5's two binding regions are a shallow binding cleft called the WDR binding motif (WBM) and the WDR5 interacting site (WIN), a deep binding cavity that relies on arginine interactions.[66] These binding regions interact with proteins, allowing WDR5 to mediate protein-protein and DNA-protein interactions.[66] Through this binding WDR5 acts as a histone reader, recognizing and binding to histone H3 through the WIN site.[66] WDR5 does not bind to histone H3 carrying repressive modifications, but binds to unmodified histones or those carrying active modifications, such as methylation.[281] WDR5 also binds to transcription factors like MYC proteins who bind in the WBM cleft.[282] Together MYC and WDR5 co-localize to the chromatin, allowing MYC to interact with target genes, while preventing MYC from associating with genes that activate tumorigenesis.[282] WDR5 is also involved in mitosis, localizing to the midbody and mitotic spindle through stable WIN site.[283] Mitotic defects occur if the WDR5 WIN site is not present in mitosis.[284]

In addition to belonging to the NSL complex, WDR5 also serves as a member of the SET1/MLL family of complexes and nucleosome remodeling and deacetylase (NuRD), which are described in the following section.

1.3.3.2 WDR5 COMPLEXES

1.3.3.2.1 SET1/MLL COMPLEXES

The SET1/MLL family of complexes promote transcriptional activation through the di and tri-methylation of histone H3K4.[285] The term SET1/MLL represents two related methyltransferases, with SET1 found in yeast, and its homolog, the MLL family found in mammals.[286,287] These methyltransferases all contain the catalytic SET domain, which is responsible for enzymatic activity.[288] Despite there being six members of the MLL family in mammals, these proteins have their own specialized functions.[286] One major function attributed to MLL complexes is the maintenance of *HOX* genes expression patterns.[286,289]

SET1 or MLL methyltransferases are often in complexes with at least three other subunits, retinoblastoma-binding protein 5 (RbBP5), absent, small or homeotic 2 (ASH2), and WDR5.[290] Within the complex the association of ASH2, RbBP5 and WDR5 provides structural integrity, with MLL's controlling specificity.[290] In this scaffolding structure, WDR5 acts as a binding intermediate with RbBP5 binding to the WBM site and the SET1/MLL proteins binding to the WIN site.[154,291] Mammalian SET1/MLL complexes HMT activity is stimulated by the WRAD protein module consisting of the proteins WDR5, RbBP5, ASH2 and DumPY-30 homolog (DPY-30).[292] The WRAD module regulates complex interactions such as protein-protein, protein-DNA or protein-RNA interactions.[292] One protein-DNA interaction involves recruiting the SET1/MLL complex to the chromatin allowing for changes in gene expression.[292] The WRAD module helps target the complex to transcription factors, like Oct4, which causes the activation of ESC self-renewal pathways.[292,293]

Alterations in SET1/MLL complexes have negative consequences in humans. When one member of the SET1/MLL family, *MLL2*, is mutated a disorder called Kabuki syndrome arises.[294] The *MLL2* mutations are autosomal dominant and pathogenic, causing a clinical presentation of intellectual disability and abnormalities in skeletal, renal and cardiac systems.[295] Kabuki syndrome patients have distinctive facial features, including everted lower lids, long palpebral fissures, prominent ears, depressed nasal tip and arched eyebrows.[295] *MLL2* mutations are the prominent cause of Kabuki syndrome but mutations in *KDM6A* can also lead to this disorder.[296] All mutations causing Kabuki syndrome have a similar phenotype, however certain hallmarks are more readily

found when a certain gene is mutated. For example long palpebral fissures are found in all cases of Kabuki syndrome caused by *KDM6A* mutations, whereas this feature is less common in patients with *MLL2* mutations.[296] To date, there are 238 individuals identified with Kabuki syndrome caused by *MLL2* mutations.[296]

1.3.3.2.2 NuRD Complex

Another WDR5 containing complex is the co-repressor complex NuRD. NuRD acts to deacetylate histones, tightening the binding affinity between DNA and histone octamers making the DNA transcriptionally inactive. Through NuRD's nucleosome remodeling capabilities the complex prevents transcription factors from binding to gene enhancers or promoters, causing further repression.[297,298] The complex consists of six or more components, Chromodomain-helicase-DNA-binding protein 3/4 (CHD3/CHD4), methyl-CpG binding domain protein 2/3 (MBD2/3), metastasis-associated proteins 1/2/3 (MTA1/2/3), HDAC1 and HDAC2, Rbbp4/7, and GATA binding protein 2a/b (Gata2a/b).[299] In the complex two different enzymatic subunits exist, HDAC1/2 are involved in deacetylation and CHD3/4 perform nucleosome repositioning through the use of ATP.[300] Other subunit functions include; MTA isoforms which mediates the binding of HDAC1 to the DNA or transcriptions factors, Rbbp proteins which act as scaffolds and bind to histone H4, and Gata2a/b directly binds to MBD2/3.[301–303] MBD2 and MBD3 have different properties. Only MBD2 is able to efficiently bind to DNA with methylated cytosines.[304] Three different isoforms of MBD3 exist, MBD3A, MBD3B and MBD3C. When MBD3C is in the NuRD complex, it binds to WDR5's WIN site, adding a seventh subunit to the complex.[305]

The NuRD complex containing MBD3C and WDR5 has distinct functions. In embryonic stem cells the binding between WDR5 and MBD3C is necessary for MBD3C-NuRDs activity.[305] MBD3 is required for silencing pluripotency genes in embryonic stem cells, allowing for embryonic stem cell differentiation.[306] The addition of WDR5 into the MBD3C-NuRD complex changes the target genes the complex regulates by co-localizing at distal promoters and enhancers to modify gene expression.[305]

Through analyzing different complexes in which WDR5 is present, it exemplifies the versatile nature of WDR5.

1.3.4 WDR5 and Global Developmental Delay Rationale

Seven subjects with global developmental delay have been identified to have *de novo* mutations in the gene encoding *WDR5.* We speculate these patients belong to an emerging group of global developmental delay disorders, caused by alterations to the NSL complex. Patients with

WDR5 mutations have a similar phenotype to those with mutations in the NSL complex subunits *KAT8*, *KANSL1, HCF1* and *OGT1*. The accumulation of global developmental delay phenotypes in patients with NSL complex mutations indicates the mutations in *WDR5* may be responsible for the patient phenotype. To study the mutations in *WDR5*, we aim to see if mutated gene produces protein variants that differ from wild type WDR5 proteins in terms of cellular location, protein binding or structural integrity and acetylation properties within the NSL complex.

Global developmental delay syndromes caused by mutations within the NSL complex are a part of a larger family of MYST complex global developmental delay syndromes, which feature patients with mutations in the MOZ/MORF (KAT6A/B) and KAT5 complexes. With evidence found in both mice and human cohorts, it shows that the MYST family complexes are integral in developmental and neuronal processes, and through analyzing the NSL complex it is our intention to further prove this notion.

1.4 Rational For Project

Technology and science are constantly evolving, which significantly impacts our understanding of disorders and diseases. The dramatic increase in the availability of new information, coupled with advancements in genetic testing, provides new insights into a number of diseases and disorders that are attributed to changes at the genetic level.

One gene type that has damaging effects when altered are epigenetic regulators. Over 500 epigenetic regulators have been identified, with these genes playing fundamental roles in chromatin alteration leading to changes in gene expression.[108,307] *De novo* mutations in epigenetic regulators are frequently identified through genetic testing. However, often little is known about these new classes of mutations.

The aim of my master's project is to elucidate patient mutations occurring in epigenetic regulators that are part of lysine acetyltransferase complexes. In studying these mutations, my aim is to determine if the gene mutations cause variation in protein stability, enzymatic activities, subcellular localization or complex organization, when compared to the wild type protein. If there are differences between wild type and variant proteins produced, it is likely that the genetic mutations are responsible for the clinical presentation of the patients. If we can determine how the mutations are impacting the protein, we may be able to further define the pathogenic mechanisms leading to the disorder. By undertaking an in-depth analysis of lysine acetyltransferase mutations, we hope to further illustrate the importance of lysine and histone acetylation in human development.

1.5 Figures and Legend

Figure 1

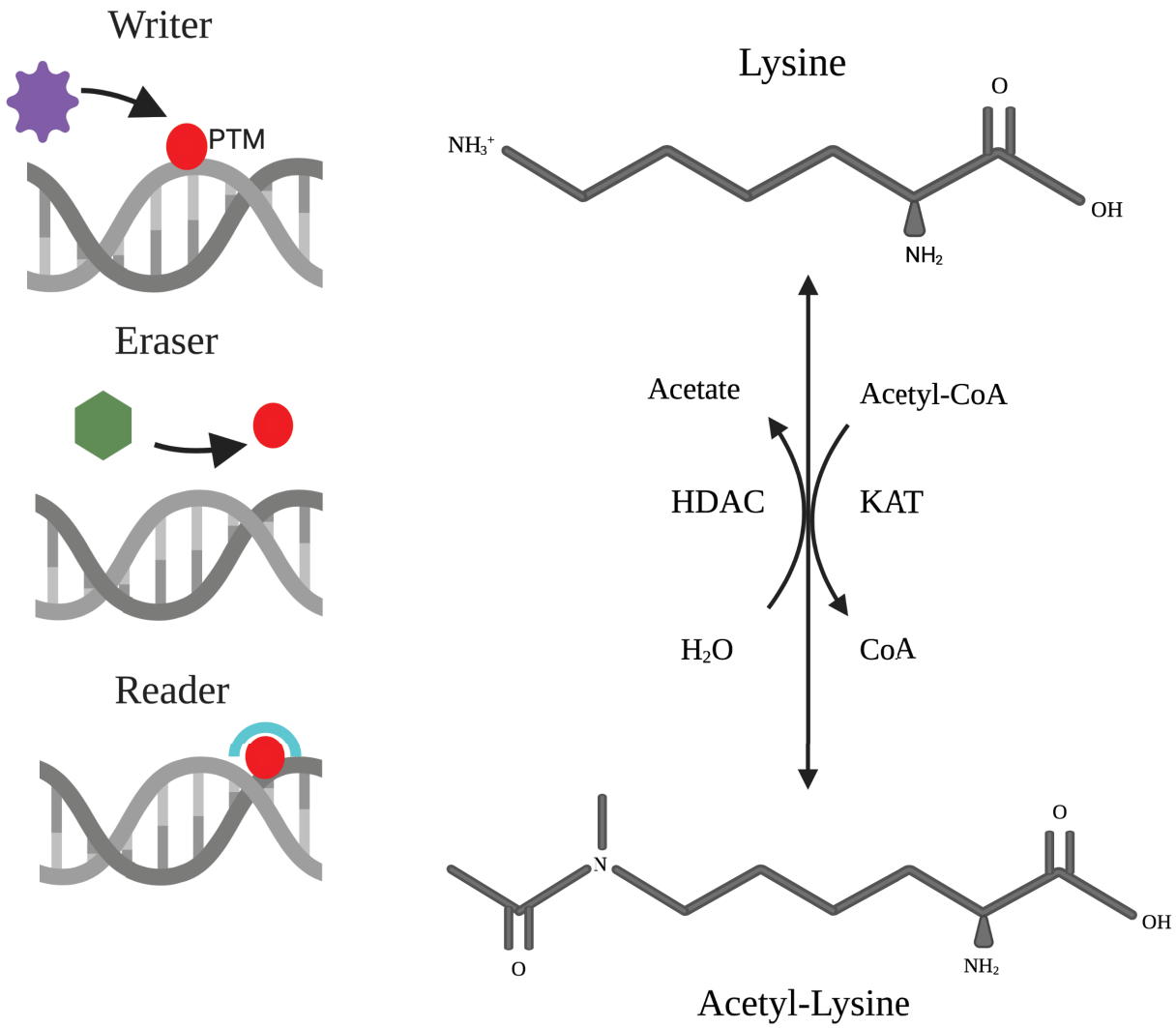


Figure 1. Acetylation catalyzed by KATs and deacetylation be HDACs

(A) Illustrations of histone reading writing and erasing. Writers add PTM's onto histones, while erasers remove them. Readers are able to recognize cellular targets, controlling the specificity of writers and erasers. Image generated with BioRender.com.

(B) Illustration of acetylation and deacetylation of lysine residues. Image generated with BioRender.com.

Figure 2

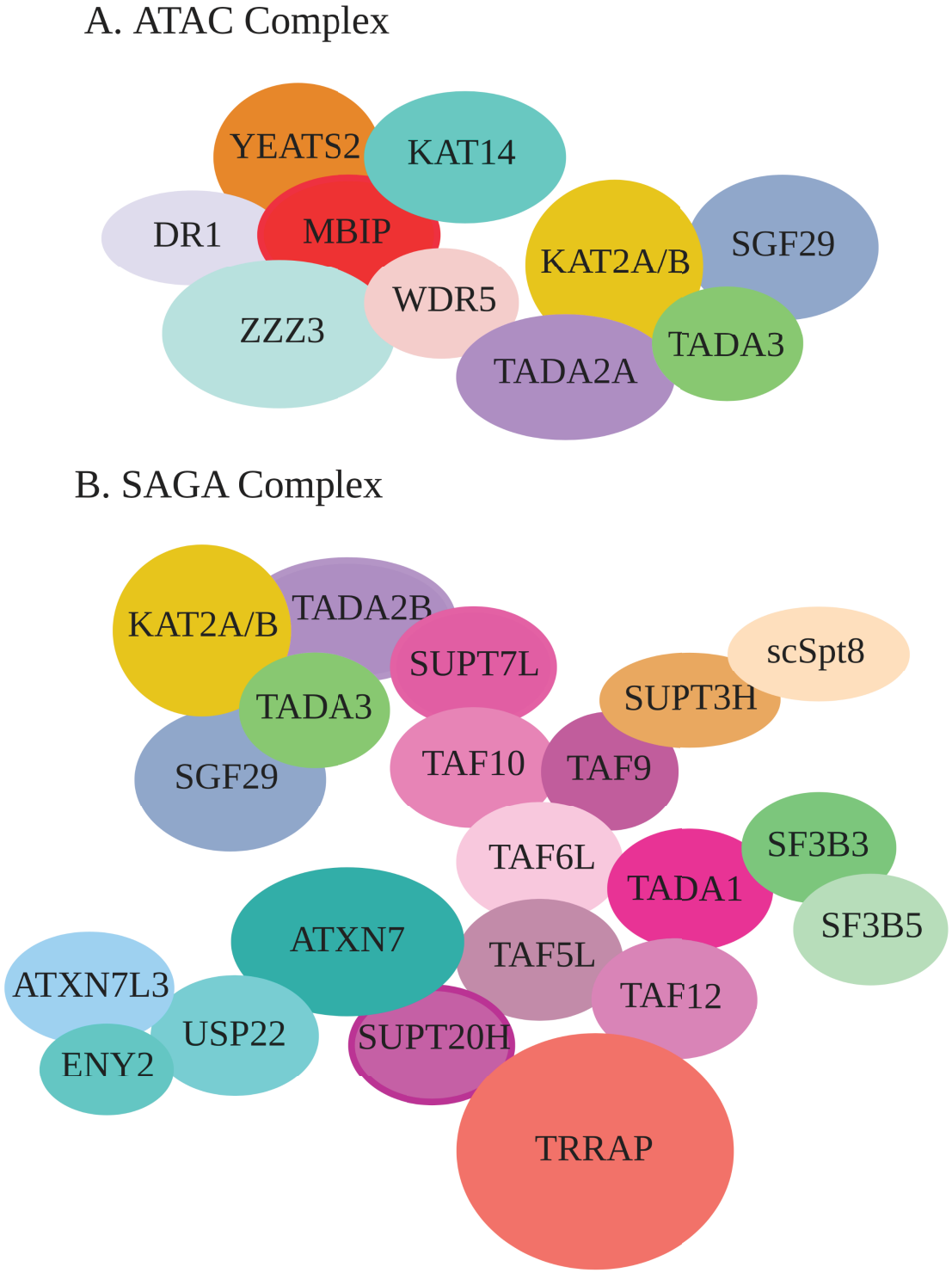


Figure 2. The ATAC and SAGA Complexes:

Illustration of the subunit composition of the (A) ATAC complex and (B) SAGA complex. Image generated with BioRender.com.

Figure 3

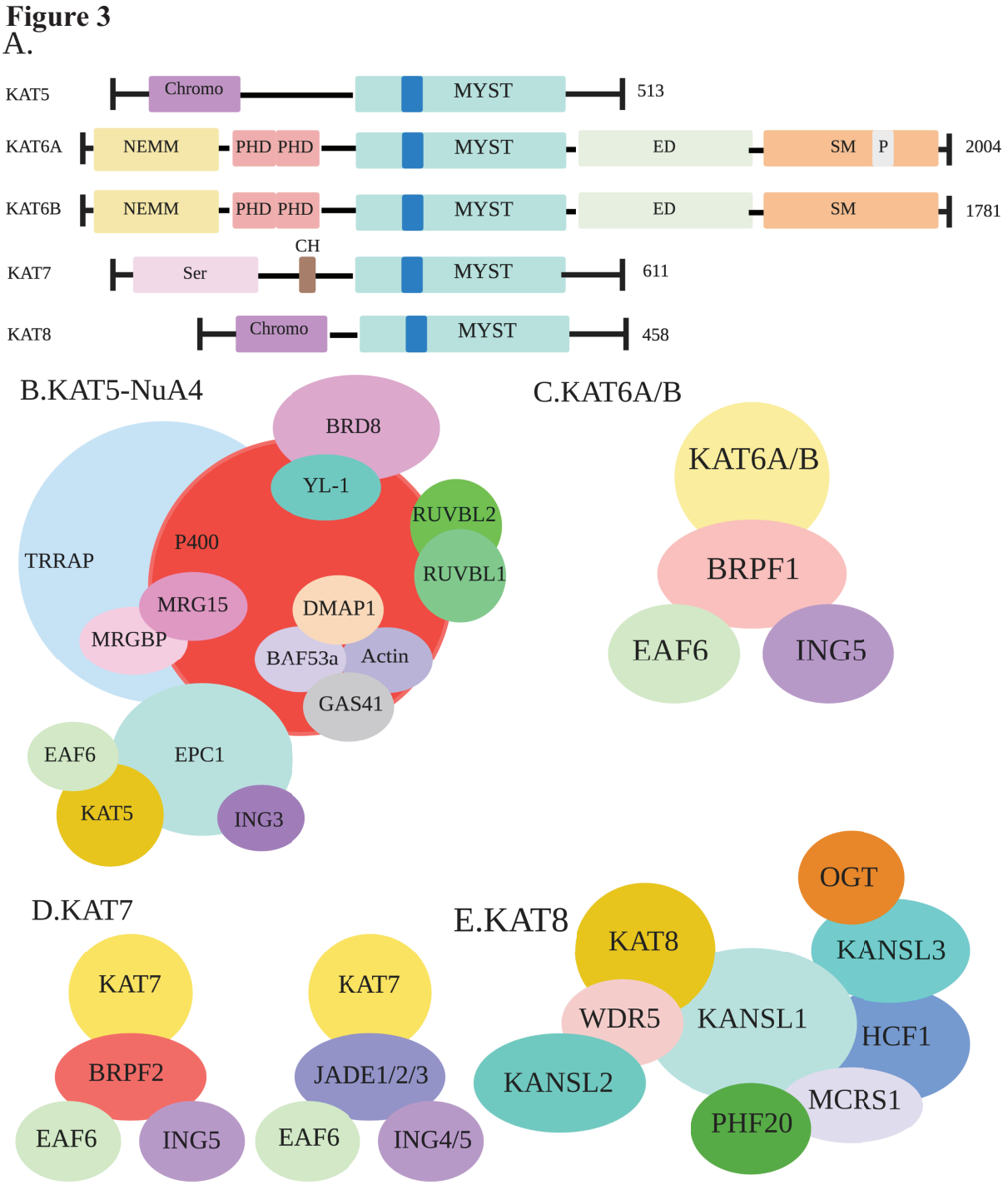


Figure 3. MYST KAT domain organization and MYST complex subunits

(A) Schematic representation of the MYST complexes: KAT5, KAT6A, KAT6B, KAT7, KAT8. Image generated with BioRender.com.

Illustrations of the subunit composition of (B) NuA4-KAT5 complex, (C) KAT6A/B complexes, (D) KAT7 complexes and (E) NSL-KAT8. Images generated with BioRender.com

Figure 4

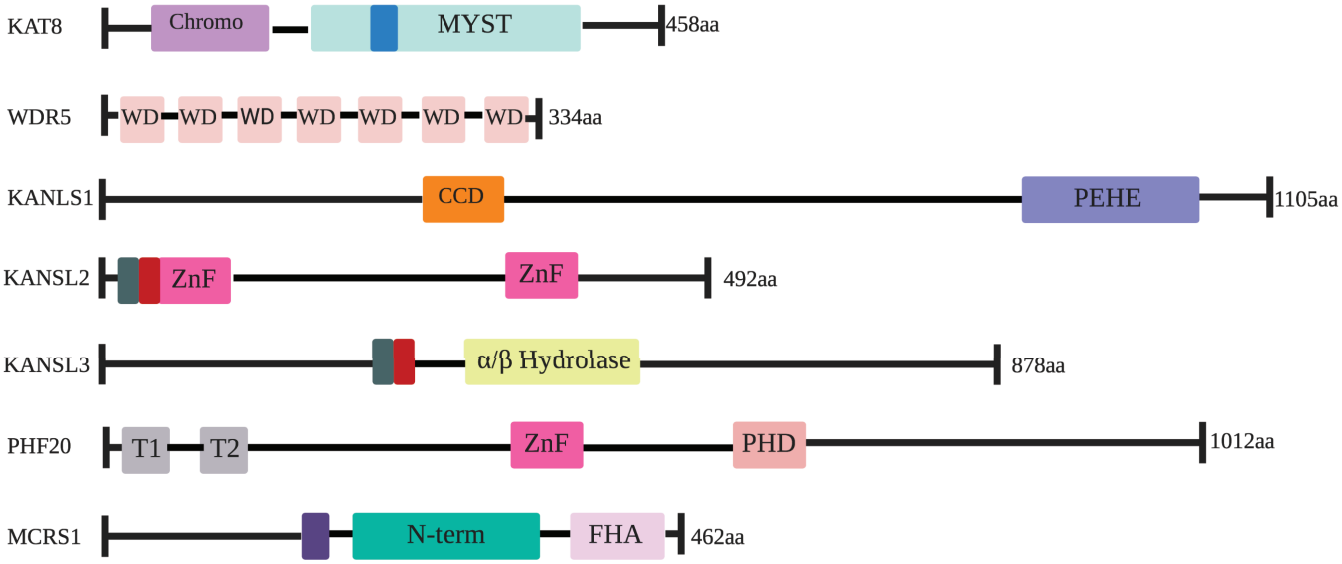


Table 1

DIAMOND BLACKFAN ANEMIA SYMPTOMS		
PHYSICAL ABNORMALITIES	**MULTI-SYSTEMIC EFFECTS**	**PREDISPOSITIONS**
• Microcephaly • Hypertelorism • Ptosis • Small or low set ears • Webbed neck • Broad and flat nose bridge • Micrognathia • Cleft palate • Small shoulder blades • Short stature • Finger abnormalities	• Anemia • Fatigue • Weakness • Horseshoe kidney • Structural heart defects • Hypospadias • Leukopenia • Pure red cell aphasia • Erythroid hypoplasia • Neurodevelopmental delay • Renal agenesis	• Acute Myeloid Leukemia • Myelodysplastic syndrome • Cataracts • glaucoma

Figure 4. NSL Domain Organization

Schematic representation of the NSL complex proteins: KAT8, KANSL1, KANSL2, KANSL3, WDR5, PHF20 and MCRS1. Image generated with BioRender.com

Table 1. Diamond Blackfan anemia symptoms and predispositions[82,308]

Table 2. Clinical Features of MYST Complex Mutations													
Complex	**NSL**	**NSL**	**NSL**	**NSL**	**SET1/MLL**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**NuA4**	**NuA4**
Common Mutations Genes	**KANSL1 (125)**	**KAT8 (9)**	**OGT1 (10)**	**HCF1 (24)**	**KMT2D (238)**	**KAT6A (86)**	**KAT6B SSBYS (90)**	**KAT6B GPS (42)**	**KAT6B Intermediate (15)**	**KAT6B NOS (3)**	**BRPF1 (41)**	**KAT5 (3)**	**TRRAP (24)**
Males	55/125	3/9	8/10	24/24		46/86	42/90	22/42	6/15	1/3	21/41	2/3	9/24
Females	64/125	6/9	2/10			40/86	47/90	20/42	9/15	2/3	19/41	1/3	15/24
Global Developmental Delay													
Developmental Delay	114/125	9/9	10/10	21/24	82/238	86/86	73/90	21/42	14/15	3/3	36/41	3/3	24/24
Intellectual Disability	83/125	6/9	10/10	6/24	238/238	65/86	75/90	21/42	9/15	3/3	37/41	3/3	17/20
Delayed Motor Development	90/125	9/9	5/10	2/24		28/86	23/90	14/42	7/15		29/41		6/24
Speech Delay or Disorder	93/125	9/9	4/10	2/24		81/86	33/90	9/42	11/15		30/41	2/3	14/24
Neurophysiological Disorders												3/3	
Behavioural Problems	32/125	1/9	2/10	1/24		23/86	6/90		3/15	1/3	8/41	3/3	5/24
Autism	6/125	2/9				12/86	1/90		1/15		3/41		4/24
ADHD	14/125	1/9				4/86	2/90			1/3	6/41	2/3	4/24
Neurological													
Hypotonia	98/125	1/9	2/10	6/24	154/238	63/86	63/90	14/42	9/15	1/3	20/41		8/24
Brain Abnormalities	46/125	5/9	3/10	3/24		20/86	17/90	17/42	4/15	1/3	10/41	3/3	10/24
Epilepsy/ Seizures	52/125	7/9	1/10	16/24		10/86	3/90	5/42	4/15	1/3	9/41	3/3	5/24
Microcephaly	8/125	1/9	3/10	8/24	96/238	31/86	5/90	30/42	5/15		10/41	2/3	7/24
Cerebral Hypoplasia/Atrophy	1/125		1/10			1/86		1/42			1/41	2/3	6/18
Corpus Callosum Agenesis/Hypoplasia	23/125		1/10	2/24		4/86	4/90	37/42	9/15		3/41	2/3	1/24
Movement Disorder/ Choreoathetosis			1/10	5/24		2/86							
Muscoskeletal/ Limb Defects													
Abnormal Cranial Shape	19/125	5/9		1/24		8/86	1/90						
Joint Hyperlaxity/Nontraumatic Dislocation	61/125	1/9			44/238	3/86	11/90	3/42	3/15		6/41		
Positional Deformity Feet	28/125												
Scoliosis/Kyphosis	35/125	1/9				3/86	2/90	3/42	2/15			1/3	3/24
Traceo/Laryngomalacia	15/125					6/86	3/90	7/42	4/15		1/41		3/24
Contractures/ Flexion Deformities	5/125	2/9				2/86	26/90	32/42	10/15		1/41		1/24
Limb Deformitites						4/86	1/90						
Lower Limb Hypperreflexia								1/42					5/24
Clinodactyly	2/125	1/9	8/10			8/86	5/90		4/15		11/41	2/3	5/24
Malposition of Toes	2/125	1/9		1/24		2/86	11/90		2/15		1/41		2/24
Spinal Deformities	2/125		1/10			2/86	1/90	1/42	1/15		5/41		1/24
Camptodactyly				1/24		2/86	1/90		4/15				
Talipes (Club Foot)	4/125					4/86	3/90	30/42	8/15		1/41		
Brachymetacarpia/ Brachydactyly	5/125				127/238	4/86					6/41	1/3	1/24
Craniofacial Dysmorphisms								2/42					
Patellar Agenesis or Hypoplasia								33/42	6/15				
Patellar Abnormalities							15/90		2/15				1/24
Thoracic Anomalies	13/125	2/9				5/86	1/90		1/15				
Osteopenia							1/90						
General Skeletal Abnormalities							2/90						1/24
Cardiac Defects	45/125	1/9	2/10	2/24	212/238	41/86	35/90	23/42	12/15	1/3	2/41	1/3	10/15
Genital/ Anal Defects			3/10				14/90	11/42	2/15	1/3			5/24
Cryptorchidism	29/55			1/24		4/86	19/90	16/42	5/15	1/3	2/41	2/3	2/9
Scrotal Hypoplasia	1/55			1/24			3/90	16/42	3/15				1/9
Bilateral Testicular Ectopia				1/24									
Hypospadias	3/55		2/10	2/24	2/238	2/86	5/90		2/15			1/3	
Micropenis	3/55		1/10	2/24		1/86		1/42	2/15	1/3			2/9
Undescended Testes	4/55		1/10	1/24		6/86		1/42					
Ambiguous Genitalia				1/24			1/90	1/42					
Clitormegaly								10/42	3/15				
Hypolasia of Labia Minora or Majora							1/90	13/42	3/15				
Anal Abnormalities	1/125					3/86	2/90	12/42	2/15	1/3			

Table 2. Clinical Features of MYST Complex Mutations													
Complex	**NSL**	**NSL**	**NSL**	**NSL**	**SET1/MLL**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**NuA4**	**NuA4**
Common Mutations Genes	**KANSL1 (125)**	**KAT8 (9)**	**OGT1 (10)**	**HCF1 (24)**	**KMT2D (238)**	**KAT6A (86)**	**KAT6B SSBYS (90)**	**KAT6B GPS (42)**	**KAT6B Intermediate (15)**	**KAT6B NOS (3)**	**BRPF1 (41)**	**KAT5 (3)**	**TRRAP (24)**
Renal Defects							1/90	11/42		2/3			5/17
Hydroephrosis	15/125					4/86		25/42	2/15		1/41	1/3	3/24
Ureterohydronephrosis								3/42					
Multiple Renal Cysts								5/42	2/15			1/3	
Recurrent Urinary Tract Infections	3/125											1/3	
Vesico-Uretral Reflex	11/125								1/15			1/3	2/24
Cystic Dysplasia of Kidney	1/125			1/24				2/42					2/24
Horseshoe Kideny						1/86						1/3	
Renal Agenesis										1/3			
Nephromegaly										1/3			
Facial Dimorphisms								8/42					19/24
Bulbus/ Broad Nasal Tip	79/125	2/9		2/24		58/86	12/90		4/15	1/3	4/41	1/3	3/24
Tubular/ Bulbous Nose	79/125		2/10			1/86	17/90	22/42	4/15			1/3	
Broad Nasal Root or Bridge	8/125	1/9	1/10			6/86	8/90	5/42	5/15	1/3	16/41		3/24
Flat Nasal Bridge/ Root	2/125					2/86	12/90		4/15				2/24
Long Nose				2/24									
Flat Nose	2/125			1/24									
Long Face	63/125	1/9	2/10			4/86		1/72	1/15				
Round Face	1/125						1/90				14/41	2/3	1/24
Coarse Face	7/125	1/9				3/86		14/72	1/15				1/24
Full Cheeks	2/125					1/86	8/90		1/15				
Flat Flacial Profile	1/125			2/24							14/41	2/3	
Everted Lower Lip	24/125												
Epicanthal Folds (Singular or Bilateral)	54/125	2/9				10/86	7/90		4/15	1/3	6/41	1/3	3/24
Inverted Epicanthal Folds							5/90						
Upslanting Palpebral Fissures	57/125	1/9		1/24		3/86	1/90	1/72	1/15		3/41		4/24
Downslanting Palpebral Fissures	8/125		2/10			1/86	6/90		3/15		10/41	2/3	1/24
Short Palpebral Fissures	24/125	1/9				3/86	5/90	1/42	5/15	2/3	2/41		4/24
Long Palpebral Fissures													
Long Philtrum	13/125	1/9	3/10	2/24	186/238		6/90	1/42	2/15		2/41		
Short Philtrum	11/125					2/86	1/90	1/42	1/15		4/41		2/24
Everted Eyelids													
Infraorbital Folds				1/24	149/238		1/90	2/42					
Prominent Occular Globes				1/24									
Hypertelorism	13/125	1/9	2/10	2/24		7/86	11/90	3/42	4/15		9/41		7/24
Cleft Palate/Lip	11/125			2/24		2/86	19/90	5/42	7/15	1/3		2/3	5/24
Arched Eyebrows		1/9			129/238	6/86	1/90		1/15				3/24
Lateral, Sparse or Notched Eyebrows	7/125	1/9			134/238	5/86	7/90	2/42			1/41	2/3	3/24
Short Columella	1/125				65/238		1/90	1/42	1/15		1/41		
Depressed Nasal Bridge	2/125	3/9			111/238		4/90	1/42	1/15		2/41	2/3	3/24
Prognathism	6/125				111/238	2/86			1/15			2/3	1/24
Almond Shaped Eyes	1/125	1/9				1/86						1/3	1/24
Macrostomia												1/3	1/24
Thick Lower Lip	2/125					3/86			1/15			1/3	
Full Lips	7/125	2/9	3/10	2/24		1/86					1/41		
Thin Upper Lip	8/125	1/9	2/10			43/86	4/90	1/42	5/15		1/41		6/24
Thin Lips		1/9		1/24					1/15		1/41		
Small Mouth	1/125	1/9	1/10			3/86	13/90		2/15		3/41		1/24
Prominent (Broad) Chin	16/125					1/86						1/3	
Small Chin	5/125						1/90	2/42			1/41		
Bitemporal Narrowing (Narrow Forehead)	1/125	1/9	1/10			11/86	2/90	2/42					2/24
Frontal Bossing/ Large Forehead	33/125	1/9	2/10			8/86							
Mask-Like Face							28/90	1/42	10/15				
SBBYSS Typical Face							53/90		1/15				
Dolichocephaly	5/125		2/10	1/24			1/90		3/15		1/41		2/24

Table 2. Clinical Features of MYST Complex Mutations													
Complex	**NSL**	**NSL**	**NSL**	**NSL**	**SET1/MLL**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**MOZ/MORF**	**NuA4**	**NuA4**
Common Mutations Genes	**KANSL1 (125)**	**KAT8 (9)**	**OGT1 (10)**	**HCF1 (24)**	**KMT2D (238)**	**KAT6A (86)**	**KAT6B SSBYS (90)**	**KAT6B GPS (42)**	**KAT6B Intermediate (15)**	**KAT6B NOS (3)**	**BRPF1 (41)**	**KAT5 (3)**	**TRRAP (24)**
Micrognathia	5/125	2/9				9/86	9/90	13/42	3/15	1/3	4/41		3/24
Microretognathia	2/125					2/86	1/90	1/42	5/15				
Retinogranathia	1/125		2/10			2/86	3/90	1/42	3/15		3/41		
Craniosynostasis						1/86							
Visual Impairments								4/42		1/3			4/24
Strabismus	38/125	2/9		1/24		41/86	4/90	3/42	3/15		4/41	1/3	2/24
Ptosis/ Blepharophimosis	35/125	1/9				13/86	48/90	2/42	13/15	2/3	30/41		
Hypermetropia	34/125	3/9	1/10	1/24		6/86	3/90		1/15		1/41	1/3	
Myopia	2/125		2/10			13/86	9/90		1/15	1/3		1/3	1/24
Epiblepharon												1/3	
Retinal Degradation				1/24									
Refraction Problems											4/41		
Nystagmus	3/125	1/9	2/10			2/86	8/90				2/41		
Amblyopia	0.25		1/10			7/86	6/90				4/41		3/24
Hearing Impairment	18/125		2/10	1/24		7/86	20/90	4/42	7/15	1/3			4/24
Miscellaneous					62/238								
Frequent Ear or Other Recurrent Infections	8/125		1/10			30/86	11/90	2/42	1/15	1/3			4/24
Large Dysplastic, Prominent or Cupped Ears	36/125	3/9				5/86			1/15		1/41	1/3	2/24
Low Set Ears	23/125	3/9	2/10	1/23	159/238	26/86	12/90	5/42	7/15	1/3	4/41	1/3	4/24
Simplified Ears	3/125			1/2			2/90				2/41		
Small/ Round Ears	7/125			1/2		2/86	6/90	2/42	2/15	1/3	3/41		
Edematous Hand				1/24									
Fetal Fingertip Pads						1/86							7/23
Short Stature	6/125		1/10	3/24		2/86	1/90				7/41		
Dental Abnormalities	19/125		1/10		224/238	22/26	39/90	2/42	4/15	1/3	4/41		8/24
Feeding Difficulties	62/125	3/9	2/10	1/24	108/238	64/86	62/90	16/42	13/14	1/3	15/41		8/24
Dysplastic Nails					94/238	1/86							3/24
Lacrimal-Duct Aplasia or Anomalies	1/125				86/238	1/86	9/90		5/15				4/24
Accessory Nipple			2/10			4/86							3/24
Hypothyroidism	2/125		1/10			3/86	10/90	5/42	7/15	1/3	2/41		
Thyroid Anomalies							22/90	3/42	1/15				
Long Thumbs or Big Toes							41/90	5/42	9/15		2/41		
Long Figers or Toes	27/125	1/9	2/10			1/86	8/90		2/15		1/41		1/24
Frontal Hair Upsweep			2/10						1/15		3/41		2/24
Premature Death	2/125			3/24		2/86	3/90	4/42	2/15		1/41		
Elevated Plasma Homocystine				9/24				5/42					
Urine Methymalonic Acid				12/24									
Failure to Thrive	5/125	1/9	1/10	6/24		8/86	4/90	3/42	1/15		3/41		1/24
Gastroesophageal Reflux	1/125		1/10			35/86	4/90	4/42	3/15		2/41		2/24
Constipation	3/125	1/9	1/10			27/86	5/90				3/41		2/24
Pulmonary Hypoplasia								3/42					
Respiratory Distress/ Disorder	11/125		1/10	1/24		13/86	4/90	4/42	4/15	1/3			1/24
Sleep Disturbances		2/9	1/10			20/86	1/90				4/41	3/3	3/24
Intrauterine Growth Resriction				2/24									
Defective MeCbl Synthesis													
Chronic Otitis Media Associated with Cholesteatoma	11/125					5/86	2/90						2/24
Neuroblastoma								1/42					
Neutropenia				1/24		2/86							
Food Allergies						5/86							

Table 2: Clinical presentation of MYST complex mutations: Subjects with each disorder were found in the following: KANSL1[245,247–263,270,271,309,310], KAT8[32], OGT1[272,273,275,276], HCF1[277–280], KMT2D[296], KAT6A[178–190,311], KAT6B SSBYS[202,206,209–212,214–229], KAT6B GPS[191–193,195–208], KAT6B Intermediate[194,206,213,215,230–233,312], KAT6B NOS[206], BRPF1[108,125,234–238], KAT5[176], TRRAP[177]

CHAPTER 2: MATERIALS AND METHODS

2.1 PARTICIPANTS

Subjects carrying *WDR5* mutations were found by our colleagues, Versput and colleagues at the Radbound University on the Netherlands, through either the Max Planck Institute for psycholinguistics, or through gene matching sites including GeneMatcher, Radboudumc in house exome sequencing database, the VKGL database, ClinVar denovo-db and Pubmed. In total seven subjects have been identified.[1] Participants or guardians gave consented to share or publish data through the Castor database.[1]

The subject carrying *KAT14* mutations was found by a clinician, Dr. Pankaj B. Agrawal in Boston, whose patient had the mutations. The patient or guardian has given consent to share their information.

2.2 WHOLE EXOME SEQUENCING (WES) AND MUTATION ANALYSIS

Sequencing of DNA coding regions was completed by external sources.

Subjects with *WDR5* mutations had their primary variant analysis and the protocol for analysis published: patient A[313], patient B[314], patient C/G[315], patient D[1], patient E[316], patient F[317]. Verseput and collogues performed further analysis, studying the location of the mutation and potential affects with AlamutVisual 2.10 and clustal alignment from Uniprot.[1] SIFT, PolyPhen-2 and CADD v1.4 used for prediction of pathogenicity.[18] GnomAD was used to ensure healthy controls lacked these variants.[1]

WES of subjects with *KAT14* mutations were provided by Dr. Pankaj B. Agrawal, whose patient has Diamond Blackfan anemia. Mutations were analyzed using the PolyPhen-2 software to understand if the effects had the potential to be damaging.[318] In addition to the PolyPhen-2 , I-Mutant was used to assess potential changes in free energy and stability of the mutants.[319]

2.3 PLASMIDS AND SITE DIRECTED MUTAGENESIS

Previously, WDR5, NSL1, NSL2 and YEATS2 were engineered onto pcDNA 3.1-Ha, and KAT8, histone H3 and KAT2A/B were engineered onto pcDNA3.1-FLAG. Both pcDNA3.1-Ha and FLAG are derived from expression vector pcDNA3.1 (Invitrogen). WDR5 is engineered on pcDNA3.1-FLAG and mCherry, KAT14 onto pcDNA3.1-FLAG, mCherry and HA, and DR1, ZZZ3, MBIP, SGF29, TADA2A and TADA3 are onto pcDNA3.1-Ha. Template DNA (Genetic Perturbation Services) underwent PCR amplification with Pfu high-fidelity DNA polymerase

(Roche) Restriction sites were cleaved in the DNA fragment, then ligated to expression vectors. (enzymes: NEB) Subsequent transformation into DH5α and growth in liquid culture occurred, followed by plasmid purification (Qiagen). Plasmids were sequenced for verification. Reagent amounts and reaction were determined through modification of NEB's sub cloning protocols.

Mutations in WDR5 and KAT14 were generated via PCR with DNA polymerase Pfu and mutant strands selected with Dpn1 (NEB). Mutations being sequenced for verification.

For both the generation of plasmids and mutant variants PCR amplification with gene or mutant specific primers were used. Primers were designed utilizing IDT Biotechnology and subsequently ordered from the company. Primers utilized to generate wild type plasmids were: for KAT14 CSRP2BP-1F (5'- GGG TTT GAA TTC GAT AGT AGC ATC CAC C- 3'), CSRP2BP-1R (5'-GGAG AAG GAG ATG GGC TTG T-3'), CSRP2BP-2F (5'-GCC GAG GCC GCA GGC CAG AT-3') and CSRP2BP-2R (5- GGG TTT GGA TCC GCG CCG GAG CCT CAG AAA G-3'); SGF29-F (5'-GGG AAA GAA TTC CCC CTG TAG ACA-3') and SGF29-R (5'- AAA TTT AAG CTT TGC CAG GCG GCA-3'); MBIP-F (5'- GGG AAA GGA TCC GCT GCT GCC ACG-3') and MBIP-R (5'-GGG AAA GGT ACC TCA TGG AAG GTG GTG-3'); DR1-F (5'- AAA AGA GGA TCC GCT TCC TCG TCT-3') and DR1-R (5'- GGG AGC AAG CTT TCA GAT ATC ATC ATC-3'); TADA3-F (5'- AAA GGG GGT ACC AGT GAG TTG AAA-3') and TADA3-R (5'-GGG AAA AAG CTT CTA CCC ATC CAG-3'); ZZZ3-1F (5'-GGA AAG AAT TCG CTG CTT CCC GA-3'), ZZZ3-1R (5'- TGG TGC CTG CCT AAA CTT TCA AGG-3'), ZZZ3-2F (5'-CTT TGA ATC AGA TCA TGT GGC AC-3') and ZZZ3-2R (5'-GGG TTT AAG CTT TCA TCT GTT TGC-3'); Kansl3-F (5'-TTT CCC GAA TTC GCC CAC CGG GGT GGG GAG AG-3') and KANSL3-R (5'- GCA CTC TCC TTG CAC ACT TG-3'); TADA2A-F (5'-AAA TAG GGA TCC GAC CGT TTG GGT-3') and TADA2A-R (5'- GGG CGC AAG CTT TTA GCC TTT AGT-3').

2.4 Cell Culture and Transfections

HEK293 cells were cultured and passaged in Dulbecco's modified Eagle's medium (DMEM, Gibco) with 10% heat-inactive fetal bovine serum (FBS, Sigma), 100μg/ml streptomycin, and 100 units/ml penicillin (P/S, Gibco).

For transfections, cells were seeded the day before at $2x10^6$ cells per 6cm dish. Between 1.5-2.5μl Lipofectamine 2000 was used per 1μg of plasmid DNA. Generally, 12-18μg of DNA and 15μl of Lipofectamine 2000 were used, in antibiotic free DMEM/FBS media. Mixture was

incubated at 37°C for 5 to 16 hours, media was changed to DMEM with FBS and P/S and further incubated until 48 hours post transfection.

2.5 Immunoprecipitation, Immunoblotting and Acetylation Assays

Tagged plasmids were transfected into HEK293 cells, the different conditions were: 1) FLAG-tagged H3, HA-tagged WDR5, 2) FLAG-tagged WDR5, 3) FLAG-tagged KAT8, HA-tagged WDR5, NSL1, NSL2, NSL3, 4) FLAG-tagged WDR5, HA-tagged NSL1, NSL2, 5) FLAG-tagged KAT14, 6) FLAG-tagged KAT14, HA-tagged KAT14, 7) FLAG-tagged GCN5, HA-tagged WDR5, KAT14, MBIP, DR1, YEATS2, ZZZ3, SGF29, TADA3, TADA2A. Wild type WDR5 and KAT14 were replaced with mutant variants for analysis. 48 hours post transfection cells were washed twice with PBS and soluble protein extracts were prepped for anti-FLAG M2-agarose (Sigma) affinity purification, as previously described by Yan.[108] FLAG peptide (Sigma F3290) was utilized to elute bound proteins from the M2 agarose beads (Millipore Sigma A2220). Both whole cell extracts and affinity purified proteins were prepared for immunoblotting and detection with anti-FLAG (Sigma F3165), anti-Ha (Biolegend) or specific histone antibodies to determine acetyltransferase abilities (anti-H3 Abcam ab1791, anti-H3K4ac EMD Millipore 07-539, anti-H3K9ac Abcam ab10812, anti-H3K14ac EMD Millipore 07-353, anti-H3K18ac EMD Millipore 07-354, anti-H3K23ac EMD Millipore 07-355, anti-H4 Abcam ab18253, anti-H4K5ac Millipore Sigma 07-327, anti-H4K8ac Millipore Sigma 07-328, anti-H4K12ac Millipore Sigma 07-595, anti-H4K16ac Millipore Sigma 07-329, anti-H4K16pr PTM Biolabs PTM210).

Histone acetylation assays were performed based on a modified protocol by Pelletier.[320] Affinity purified acetyltransferase complexes underwent acetylation reactions, containing 5x buffer A, purified water, histone or nucleosome substrates , acetyl-CoA and purified protein complex. Reactions were incubated at 37°C for 1 hour before they were prepared for immunoblotting. Detection was completed with anti-acetyl lysine antibodies specific for certain histone lysine residues, as well as anti-FLAG and Ha antibodies.

2.6 Red and Indirect Fluorescence Microscopy

Utilized for the analysis of subcellular locations. Plasmids for mCherry WDR5 or KAT14 with a control of GFP-histone H3 were transfected into HEK293 cells. Cells were taken for analysis after approximately 16 hours post transfection, using the florescence microscope (Axio Observer Z1, Zeiss).

A similar process was used to stain cells. With the transfection identical to the above. After 24 hours cells were washed twice with PBS and allowed to dry for 3 minutes. Following this cells

were covered in ice cold 50:50 acetone: methanol and placed in -20 ^{o}C for 20 minutes to allow for fixation of the cells. Cells were washed twice with PBS before a dilution of DAPI in PBS covered the cells for ten minutes. This dilution was washed off twice with PBS and analysis using the florescence microscope (Axio Observer Z1, Zeiss) occurred.

2.7 Image J and Statistical Analysis

The Image J software was used to analyze optical densities of immunoblot protein bands. Immunoblot scans were uploaded into Image J and optical densities for each band were measured using the gel analysis tool. Measurements were exported to excel.

Statistical analysis was performed in excel utilizing an unpaired 2-tailed Student's t test. Statistical significance is considered by p values, where $p < 0.05$. Graphs generate with excel.

Chapter 3: Analysis of KAT14 Mutations

3.1 Identification of KAT14 Mutation

Through a collaboration with Dr. Pankaj B. Agrawal, a physician in Boston, a patient diagnosed with Diamond Blackfan anemia underwent WES, finding the patient harbors two unique SNPs in the *KAT14* gene. The patient has two different SNPs, one on each allele. To determine if these variants have the potential to cause the Diamond Blackfan anemia phenotype they must be analyzed. The first variant occurs at p.Arg567Ter, an arginine to a termination codon, and the second variant is an arginine to a tryptophan at p.Arg596Trp. (Figure 1a)

Both mutations found are suspected to be pathogenic. The p.Arg567Ter *KAT14* variant is expected to cause an early truncating variant. This truncation causes loss of 200 amino acids in the C-terminus, where the histone acyltransferase domain lies. The truncated KAT14 protein produced will lack the HAT domain, thus the proteins enzymatic functions are expected to be altered. Since approximately a quarter of the protein will not be translated other structural or stabilizing properties may be disrupted. The second mutation, p.Arg596Trp, is a missense mutation. To predict the possible effects of the mutation the software PolyPhen-2 was used. The PolyPhen-2 program predicted the p.Arg596Trp variant is probably damaging, with a score of 1, on a scale of 0 to 1, with 1 being the most damaging. This arginine residue is evolutionarily conserved across the kingdom Animalia, indicating the residue is essential to protein function.

In understanding that both *KAT14* mutations likely impact function, structure or stability of KAT14, it is possible that they responsible for causing the Diamond Blackfan anemia phenotype found in the patient, thus we decided to proceed with further analysis of the mutations.

3.2 Generation of the ATAC Complex and KAT14 Variants

KAT14 FLAG-tagged plasmids were generated through PCR amplification of two KAT14 fragments with restriction sites, standard cloning procedures were then followed. The insertion of the full KAT14 gene into the FLAG vector was confirmed by colony PCR, restriction digestion and validated by sequencing. This process was repeated to generate HA-tagged plasmids of the ATAC complex, namely, DR1, MBIP, SGF29, ZZZ3, TADA2A and TADA3. FLAG-tagged KAT14 was then cleaved and ligated into both mCherry and HA tagged vectors.

Mutations in the KAT14 cDNA were generated via site directed mutagenesis, with primers designed to specifically insert the patient mutation. For KAT14 p.Arg567Ter, this correlates to a DNA location of 2136 with a change from a cytosine changed to a thymine (CGA to TGA), and for KAT14 p.Arg596Trp the mutation occurs at 2223bp with a change from a cytosine to a thymine (CGG to TGG). Mutation insertion was confirmed with sequencing.

The ATAC complex plasmids, along with previously generated plasmids (KAT2A-FLAG, KAT2B-FLAG, WDR5-HA, YEATS2-HA) and the KAT14 mutants were used in the following experiments to test the differences in KAT14.

3.3 KAT14 and Mutated Variants Protein Stability and Size

To begin assessing individual differences between the wild type and mutant KAT14 proteins a transfection of *KAT14* was performed. Variant p.Arg567Ter is expected to affect protein size, generating a truncated protein since ~200 base pairs are lost from the coding sequence. This alteration may affect structural integrity or stability. A loss of stability would be depicted as a reduction protein concentration or weaker band intensity in western blots. p.Arg567Ter may also express reduced protein concentrations due to NMD. In cases of nonsense mutations, like p.Arg567Ter, the NMD pathway can degrade truncated proteins protecting cells against gain-of-function expression.

The program I-Mutant was used to study the p.Arg596Trp variant. This software predicts p.Arg596Trp has decreased stability, with a DDG value predication of -0.63, representing a large decrease in protein stability. Therefore a reduction in protein production of KAT14 p.Arg596Trp is expected.

For this analysis KAT14 plasmids were transfected into HEK293 cells and purified through CO-IP. (Figure 1b)

Interesting differences in the variants were found. As expected, the p.Arg567Ter variant caused truncated protein production. This protein is ~ 40 KDa smaller than that of the wild type KAT14 protein. The truncation corresponds to the loss of C-terminal translation, however, the protein was

smaller than anticipated. Despite the small size of the truncated p.Arg567Ter protein, it appears that the variant is unaffected by NMD. In looking at the truncated protein a higher concentration was present than wild type KAT14 indicating NMD is not degrading the truncating variant, and the protein maintains its stability.

As predicted, the mutation p.Arg596Trp produces a less stable protein, as shown by the reduction in protein produced.

To assess the statistical significance of our results a t-test was performed. Immunoblot images were inputted into Image J where the optical densities of each protein band were determined. (Figure 1c) In all 11 observations were used, a large number was chosen due to a lack of a loading control, such as tubulin or actin. Both antibodies were used, however did not reliably leave bands, therefore by increasing the number of observations it should eliminate slight loading errors. The optical densities of the CO-IP purified KAT14 variants were uploaded into excel where a two-sided t-test was performed. Both KAT14 mutants p.Arg567Ter and p.Arg596Trp, optical densities were compared to that of wild type KAT14, and they were found to have significantly different optical densities. While wild type KAT14's mean density was 9505, p.Arg567Ter had a significantly higher density, of 27185, and p.Arg596Trp had significantly lower density, of 5619, having p-values of 0.0004303 and 0.0440616, respectively. Since both p-values are under the threshold of 0.05, we determine that there are statistically significant changes in the levels of KAT14 protein produced by the mutant variants.

Both the decreased stability of the p.Arg596Trp protein and decreased size of p.Arg567Ter are significant differences compared to the wild type KAT14 protein, indicating these variants may cause significant changes *in vivo*. To further explore the variants, both enzymatic and structural properties need to be assessed.

3.4 Self-Binding Capabilities of KAT14

Many properties of the KAT14 protein have yet to be delineated. One such subject area is the binding properties of KAT14. Within the ATAC complex it is thought that KAT14 directly binds to MBIP, however interactions between KAT14 and the remaining subunits are unknown. Many proteins have the ability to bind to themselves. Through a literature review we found that KAT14 has never been assessed for this self-binding capability, therefore we decided to analyze this ourselves. For this a transfection of KAT14 FLAG-tagged and KAT14 HA-tagged plasmids occurred with subsequent purification through Co-IP. If KAT14 is able to bind to itself, during the

Co-IP FLAG-tagged KAT14 will pull down the HA-tagged KAT14, thus on a western blot you will see KAT14 on both membranes stained with HA or FLAG tagged antibodies. (Figure 2)

In figure two, we see a band representing KAT14 on both HA and FLAG membranes, for both the whole cell extract and the IP, indicating KAT14 does bind to itself. In addition to testing wild type KAT14 proteins, KAT14 mutants self-binding properties were also assessed, with FLAG-tagged wild type KAT14 and HA-tagged mutant KAT14 plasmids. In the results both the whole cell extract and the CO-IP contained the HA-tagged KAT14 variants. This shows that the mutant p.Arg567Ter and p.Arg596Trp proteins are able to bind to themselves. One difference between the variants is the protein levels of the mutants. Despite having equal amounts of FLAG-tagged wild type KAT14, there was a reduction in the KAT14 p.Arg567Ter protein compared to its counterparts. This differs from what was found when KAT14 variants were purified alone, where p.Arg567Ter had a significantly higher protein concertation and p.Arg596Trp had a reduction in protein concentration. The reduction in truncated protein concentration indicates the termination mutation affects KAT14 self-binding and by altering KAT14's protein structure it causes the interaction to be less stable. The exact protein residues required for KAT14 self-binding are unknown, therefore it is possible residues within the final 200 amino acids of KAT14 are required for self-binding efficiency.

Given KAT14 self-binding is a newly uncovered protein property the biological relevance is unknown. Therefore we cannot speculate what negative consequences could arise due to a reduction in self-binding of the KAT14 p.Arg567Ter variant.

3.5 KAT14 Acetylation Properties

One of KAT14's most notable roles is as a histone acetyltransferase. KAT14 is known to acetylate histone H4, with affinity for H4K16. The ATAC complex as a whole, has been shown to acetylate both histone H3 and H4, with KAT2A or KAT2B responsible for histone H3 acetylation. To date there has been limited research done in terms of assessing KAT14's acetylation properties in isolation, without the presence of the ATAC complex.

Given the location of both mutations in KAT14, one right before the acetylation domain and the other creating a truncated variant lacking the acetylation domain, we anticipate alterations in the acetylation capabilities of both KAT14 protein variants.

To delineate KAT14 acetylation *in vitro* histone acetylation assays were performed with purified KAT14 proteins, acetyl CoA, and HeLa histones. Subsequently immunoblotting was performed to visualize the results. (Figure 3a) Results are depicted, with representations of control levels, KAT14-Flag and H4 and with graphical representations of levels (optical densities) of

specific acetylated lysine residues (Figure 3b-i) Additionally, the optical densities were utilized to create a ratio of acetyl-CoA over non-acetyl-CoA containing conditions, showing the differences in acetylation capabilities of the KAT14 variant proteins. (Figure 3j-q) These ratios have a value of one if there was no difference between the acetylation and control, indicating acetylation was not occurring. Values greater than one indicate histone acetylation was catalyzed by KAT14. Information for both graphs was taken from an n of three, with outliers (an optical density with a threshold difference of 10000, to a maximum of one outlier) removed. Statistical significance was assessed through a student's t-test describes in the methodology.

In the image we see that all three KAT14 variants maintained their ability to acetylate histone H4K5, with significant levels of acetylation compared to the non-acetylated control for each variant found, all having p-values below 0.001. In the graph one sees a much higher level of histone H4K5 being acetylated with wild type KAT14. Using the ratios, the biggest difference is seen the termination variant, with the acetylation capacity being reduce by 20 percent, compared to wild type.

Both histone H4K8/K16 acetylation showcased dramatic differences. The acetylation of histone H4K8 by variant p.Arg567Ter was reduced to approximately half of KAT14 wild type levels, whereas p.Arg596Ter was not able to acetylate this residue. Similarly, acetylation of histone H4K16 by the termination variant saw minor reductions, whereas acetylation by the p.Arg596Trp variant was lost.

Acetylation of histone H3 was also assessed, at lysine residues K4, K14, K18 and K23. Similar to the trends seen in histone H4 acetylation, there were marked decreases in histone H3 acetylation. The p.Arg567Ter variant caused decreases in H3K4/K14/K23 acetylation, with levels of both histone H3K14 and H3K23 being the same their non-acetyl-CoA control. Variant p.Arg596Trp caused reductions in acetylation of H3K4/K14/K18/K23, with levels of acetylated histone H3K23 completely lost.

Compared to wild type KAT14 both p.Arg567Ter and p.Arg596Trp protein variants had a generalized reduction or loss of acetylation, indicating these mutations negatively impact the acetylation capabilities of KAT14. These losses in acetylation could contribute to the pathogenesis of the patient.

In the literature, little was known about mammalian KAT14's acetylation properties, particularly which lysine residues the protein acts upon. In assessing the wild type KAT14 proteins acetylation properties we see that KAT14 was able to increase acetylation of each residue studied,

H3K4/K14/K18/K23 and H4K5/K8/K16. The greatest increases in acetylation were found on histone H4K5 and H4K8, with ratios of 5.3 and 5. Statistically, the differences between the wild type acetyl-CoA and the non-acetyl-CoA control were significant in both histone HK5 and H3K23 acetylation with p-values less than 0.001. This information helps delineate KAT14 acetylation properties, and contradicts the notion, that within the ATAC complex only KAT2A or KAT2B are able to acetylate histone H3 targets.

3.6 KAT14 Subcellular Localization

Another property that could be impacted by KAT14 mutations is subcellular localization. Subcellular localization changes can impact a proteins ability to complete their cellular functions or could indicate changes in binding. To assess potential changes wild type, p.Arg567Ter and p.Arg596Trp KAT14 cDNAwas ligated into mCherry-tagged plasmids and transfected into HEK293 cells with a control plasmid of GFP-tagged histoneH3. Approximately 24 hours post transfection the cells were analyzed with a fluorescence microscope (Axio Observer Z1, Zeiss). The control of histone H3 was chosen due its consistent location, within the nucleus, therefore we can compare where the KAT14 protein localizes in relation to histone H3. If the mCherry and GFP signals overlap KAT14 proteins are nuclear, if the signals do not overlap than KAT14 proteins are cytosolic and if there is a mixture of overlapping and non-overlapping signals it indicates that KAT14 protein is pancellular.

In assessing the images, one sees a clear overlap between histone H3 and wild type KAT14 signals, indicating KAT14 is a nuclear protein. (Figure 4) This nuclear localization is supported by the ATAC complex's roles in gene transcription and nucleosome sliding.[62,72] Both KAT14 variants p.Arg567Ter and p.Arg596Trp have a similar pattern of localization, with overlapping green and red fluorescent signals, indicating they are also nuclear. Since wild type and variant KAT14 proteins localize to the nucleus it appears that the subcellular localization is not affected.

Although localization was not affected, transfection efficiency was. Mimicking results from section 3.3 there was a large reduction in the concentration of p.Arg596Trp protein compared to both wild type and p.Arg567Ter variant proteins. This reduction was seen in each experimental repeat, despite consistent plasmid concentrations and experimental procedures. This reduction in the p.Arg596Trp protein is highlighted when assessing protein levels against the histone H3 control. These results further prove the p.Arg596Trp mutant protein has reduced stability, affecting protein quantities or that there is a reduction in protein translation by the variant.

3.7 ATAC Complex and the Complexes Acetylation Properties

To further understand potential differences between the wild type and mutant KAT14 proteins, their interaction within the ATAC complex was assessed. Given KAT14 p.Arg567Ter causes a truncated protein and KAT14 p.Arg596Trp protein levels are dramatically reduced, there could be alterations in both the stability and acetylation capacity of the ATAC complex.

To generate the ATAC complex, DNA plasmids of HA-tagged wild type or mutant KAT14 were transfected into HEK293 cells with the remaining ATAC subunit plasmids: FLAG- tagged GCN5 and HA-tagged SGF29, TADA2A, TADA3, WDR5, MBIP, DR1, YEATS2, ZZZ3. After 48 hours cells were coimmunoprecipitated and visualized through immunoblotting.

Unfortunately, given the large size of the complex, the full ATAC complex was never generated through CO-IP, despite trying numerous plasmid concentrations, buffers, and CO-IP methods. However partial complexes were generated on numerous occasions. Immunoblots of these purified partial complexes consistently showed protein bands corresponding to GCN5, KAT14, WDR5, SGF29, and generally showed levels of DR1 and ZZZ3. Figure 5b and figure 5c show to depictions of the results. In both these depictions it appears that in the CO-IP of ATAC with wild type KAT14 there are higher protein concentrations of subunits, such as WDR5, SGF29 and ZZZ3. Additionally, there was a dramatic reduction in the protein levels of KAT14 p.Arg567Ter compared to the other KAT14 variants. Given the higher concentration of these subunits with consistent levels of GCN5 it may indicate that the stability of the ATAC complex is affected by the KAT14 mutations. In addition to the CO-IP, figure 5b depicts whole cells extract levels from the ATAC transfection. In this image higher protein concentration is noted, with an increased number of subunits, including DR1, YEATS2, TADA2A. In addition, TADA2A, TADA3 and MBIP all have approximately the same protein size, and therefore may actually be represented in the image but be hidden by the bands for WDR5. In the future reconstituting the whole ATAC complex is necessary to determine specific complex changes in the presence of variant KAT14 proteins, however this data suggests there are changes in the ATAC complex changes when KAT14 is mutated.

In addition to the CO-IP, whole cell lysates, or whole cell extracts were collected pre-CO-IP and visualized through immunoblotting. With the whole cell extracts, levels of histone acetylation were assessed. This was done by probing the whole cell extract immunoblots with specific histone antibodies to assess levels of acetylation within the cells. (Figure 5a) To analyze acetylation levels ratios were taken for the amount of an acetylated histone, such as histone H4K5ac over the non-

acetylated histone control, histone H4. In doing this, with an n of three, differences among the ATAC complexes with different KAT14 variants were noted. In the histone H3 lysine's tested, both mutant KAT14 containing complexes had reduced histone H3K4 and H3K14 acetylation levels compared to their wild type counterpart, while an increase in acetylation was seen on histone H3K23 by the variants. More variation was noted in histone H4, in general the mutant KAT14 containing complexes have higher acetylation than the wild type KAT14 containing complex.

Since the whole cell extract was used, these results are not conclusive. By using the whole cell extract, acetylation levels cannot be solely contributed to the transfected ATAC complex, other cellular complexes could be contributing to the levels of acetylation seen. In addition, the enzymatic capabilities of the ATAC complex were not assessed with the method. To properly assess acetylation a HAT assay, similar to what was done in section 3.5 needs to be performed. Although these results are not complete, the differences in acetylation between KAT14 and KAT14 mutant containing ATAC complexes should be explored further in the future.

CHAPTER 4: ANALYSIS OF WDR5 MUTATIONS

We believe patients with de novo mutations in *WDR5* belong to a new group of global developmental delay disorders, caused by alterations to the NSL complex. Patients with mutation in the *WDR5* gene have a similar phenotype to those with mutations in other subunits in the NSL complex, namely *KAT8* and *KANSL1*. Global developmental delay syndromes caused by these mutations are a part of a larger family of MYST complex global developmental delay syndromes.

4.1 IDENTIFICATION OF WDR5 MUTATIONS

Recently, seven different patients with *de novo* mutations in *WDR5* have been identified by our collaborators in the Netherlands.[1] Four different missense variants were identified. Each of the following variants were found in one individual, p.Ala169Pro, p.Arg196Cys and p.Asn213Asp, and an additional three unrelated individuals have the variant p.Thr208Met. The seventh individual was found to have the possible splicing variant, c.445-3C>T. *WDR5* itself is extremely intolerant to loss of function mutations, indicating mutations that completely disrupts WDR5's ability to function are not widely found in society. Therefore during embryonic development a fetus with *WDR5* loss of function mutations is unlikely to make it to term. Therefore, we suspect the patients in our cohort will not have loss of function *WDR5* variants, however given the clinical presentation we suspect the *WDR5* variants may be pathogenic.

In order to predict the potential pathogenicity of the missense mutations PolyPhen-2 was utilized. Three of the WDR5 missense variants, p.Ala169Pro, p.Arg196Cys and p.Thr208Met were predicted to be probably damaging, with scores of 1.000, 1.000, and 0.997 out of 1.000. All three of these amino acids are evolutionarily conserved in eukaryotic organisms. The final missense variant p.Asn213Asp was predicted to be benign with a score of 0.013 out of 1. This aspartic acid at 213, is frequently found in the same position through evolution, however can be replaced with other amino acids, with the most frequent substitution being glutamic acid. One reason this variant may be benign is due to the similarity of aspartic acid and asparagine. Both amino acids are a similar in size but have different charges in their side chains.

These variants were also analyzed through two other means, CADD-scores and SIFT by our collogues.[1] SIFT scores are used to predict if protein functions could be affected by substitutions of amino acids, in the case our missense variants, p.Ala169Pro, p.Arg196Cys and p.Thr208Met had a score of 0 and predicted to be deleterious, whereas the p.Asn213Asp variant had a score of 0.12 and predicted to be tolerated by the WDR5 protein. Finally, CADD scores were used to predict deleterious effects of SNPs, each of the missense variants gave a score over 22, indicating they are likely to have negative effects. The splice variant had a predicted score of 13 indicating it may have deleterious effects. CADD scores above 20 are predicted to be damaging, whereas scores under 10 are unlikely to have effects.

Changes in stability can also be utilized to predict potential consequences of protein mutations. When proteins become less stable it can impact protein levels and functions. To predict variant protein stability a program called I-Mutant was used, giving a predicted DDG value that measures the free energy change and a prediction of stability. DDG values above 0.5 and below -0.5 indicate a large increase or decrease in stability, and values between -0.5 and 0.5 indicate a neutral stability. For our variants both p.Arg196Cys and p.Asp213Asn were predicted to have large decreases in stability, with DDG values being -0.95 and -1.08. The remaining two variants were predicted to have neutral stability with DDG values for p.Ala169Pro being 0.14, and -0.22 for p.Thr208Met. The predicted large decreases in stability of the p.Arg196Cys and p.Asn213Asp variants indicates they may be damaging, and could result in decreased protein concentration, increased WDR5 degradation or affect WDR5-complex integrity .

Since each of the *WDR5* missense mutations are likely damaging, we went on further to assess the location of the mutations, and what functions of WDR5 they may be impacting. We noticed that each of the mutations occur on or near the fourth WD40 domain of WDR5, as seen in Figure

6a. Particular functions of the fourth WD40 domain are unknown, thus impacts of the mutations in this region cannot be predicted. The mutations are also located near an evolutionarily conserved threonine (191), that is important in WDR5 WIN site binding. Therefore the mutations may alter WDR5's ability to bind to histone H3 or KANSL1.

Through prediction that *WDR5* patient mutations occur near an important binding site, in a similar region and are probably damaging it leads us to believe that these WDR5 variants are responsible for the clinical phenotype found in patients.

4.2 WDR5 MUTANTS PATIENT PHENOTYPE AND NSL COMPLEX

Seven individuals with global developmental delay were reported by our collogues in the Netherlands. Each of these individuals presented with global developmental and speech delays, and six of the seven individuals have intellectual disabilities. These patients also displayed other multi-system phenotypes, including musculoskeletal, cardiac defects, autism or ADHD. Nurological problems were also found, such as hypotonia, epilepsy and brain MRI abnormalities.

Given the similar phenotypes and the predications that the variant WDR5 proteins are likely damaging we speculate that the phenotype is a result of these mutations.

Intriguingly, two complexes known to contain WDR5 have previously been implicated in global developmental delay disorders, namely the NSL and the MLL1/SET complexes. Mutations in the NSL complex subunits, *KANSL1*, *KAT8, OGT1* and *HCF1*, result in a phenotype that closely resembles the phenotype of patients with *WDR5* mutations. Patients with mutations in *KANSL1, KAT8* and *WDR5* all have the hallmarks of global developmental delay, speech and motor delays, and intellectual disability. In addition, many patients suffer from behavioural problems, autism or neurological problems such as hypotonia, seizures or brain abnormalities. Due to the similarities between our cohort of individuals with *WDR5* mutations and those affected with NSL complex mutations we chose to focus on how the *WDR5* mutations impact the NSL complex.

It is worth mentioning that members of the, WDR5 containing, MLL1/SET complex have been linked to intellectual disability syndromes. Namely mutations in *KMTD2* are the cause of Kabuki syndrome, which presents with a distinctive facial profile, and intellectual disability. However, the disorder lacks the presence of developmental delay and motor or speech delays, which are hallmark features of the *WDR5* mutant patients. Furthermore no specific abnormal facial profile was found in patients with *WDR5* mutations. This information lead to us studying the WDR5 mutations in relation to the NSL complex instead of the SET1/MLL complex.

4.3 Generation of NSL Complex and WDR5 Mutant Variants

The KANSL3 HA-tagged plasmid was generated in the same fashion as the ATAC complex plasmids, with an initial PCR amplification of KANSL3, followed by standard cloning procedures, with colony PCR, restriction digestions, and sequencing confirming the successful integration of KANSL3 into the HA vector.

The four missense mutations in *WDR5* were generated through site directed mutagenesis, following the same protocol as KAT14. The first mutation, p.Ala169Pro occurs at DNA base pair 679 and switches a guanine to a cytosine (GCT to CCT); the second mutation p.Arg196Cys occurs at DNA base pair 760 and is a switch from a cytosine to a thymine (CGC to TGC); the third mutation p.Thr208Met changes a cytosine to a thymine at DNA location 796 (ACG to ATG); finally p.Asn213Asp changes an adenine to a guanine at base pair 811 (AAC to GAC). After the generation of the plasmids results were confirmed via sequencing.

WDR5 HA-tagged wild type and mutant plasmids were cleaved, and subsequently ligated into both FLAG and mCherry vectors. After plasmid preparation results for WDR5's insertion into the FLAG vector was confirmed through immunoprecipitation and immunoblotting. WDR5's insertion into mCherry was confirmed though restriction digestions.

With the successful generation of KANSL3 and the WDR5 mutant and wild type plasmids, along with the previously generated NSL complex plasmids, we began to use these plasmids to further study the differences in the WDR5 variants.

4.4 WDR5 and Mutant Variants Protein Stability and Size

We began by exploring the individual properties of each WDR5 protein variant in relation to wild type WDR5. Since no patient presented with nonsense mutation large differences in protein size were not expected. However, software analysis predicted the WDR5 missense proteins would impact stability, thus changes in protein concentration may occur. To assess this FLAG-tagged WDR5 variants were transfected into HEK293 cells, purified through CO-IP, and imaged through immunoblotting. (Figure 6b)

To assess differences between the WDR5 variant proteins, western blot images were uploaded into Image J, where optical densities were determined for each protein. Optical densities were inputted into excel, where statistical analysis occurred and graphs were generated. (Figure 6c) Eight observations occurred, and like section 3.3, loading control antibodies did not reliably work, therefore a larger n was chosen to ensure that the phenomenon seen was accurate. In looking at the mean optical densities, one sees a dramatic reduction in protein levels of three WDR5 variants,

p.Arg196Cys, p.Thr208Met and p.Asn213Asp, with an approximate 50 percent reduction in protein concentration compared to wild type WDR5. Whereas, p.Ala169Pro showed an approximately 30 percent increase in protein concentration compared to wild type WDR5. From the stability production program, I-Mutant, mutant p.Ala169Pro was predicted to have little to no effect on WDR5's stability, which is exemplified in our results, with higher protein levels of this variant. The switch from arginine to proline may positively impact stability due to the nature of proline. Proline contains a unique pyrrolidine ring adding structural rigidity to proteins, which helps to strengthen structural integrity.[321] While p.Ala169Pro may have increased structural stability compared to wild type WDR5, the rigidity of the proline may impede the WDR5 protein folding confirmation or the ability of WDR5 to facilitate protein-protein interactions. The remaining variants, p.Arg196Cys, p.Thr208Met and p.Asn213Asp dramatically reduce protein concentration, which indicates a loss of protein stability or a reduction in production. WDR5 is extremely intolerant to loss of function mutations, therefore this significant reduction in WDR5 protein levels caused by p.Arg196Cys, p.Thr208Met and p.Asn213Asp mutants may be responsible for the changes seen *in vivo*.

4.5 WDR5 AND MUTANT VARIANTS BINDING TO HISTONE H3

One of WDR5's binding partners is histone H3, which binds WDR5's WIN site. Due to the location of the WDR5 mutations binding to histone H3 may be impacted. To test this a transfection was completed with FLAG-tagged histone H3 and HA-tagged WDR5 variants, followed by Co-IP. Each of the variants were able to bind to histone H3, indicating no mutant completely lost histone H3 binding and therefore maintained some reader functions. (Figure 6d) To confirm protein concentration, Image J was used to quantify the optical densities of each protein band. Following this, the data was inputted into excel where a protein ratio of WDR5 to histone H3 was calculated, with histone H3 representing the loading control since the concentration of histone H3 directly impact the amount of WDR5 purified during the CO-IP. This experiment was repeated eight times, and for analysis two outliers were removed from each set of ratios (highest and lowest values), leaving six observations. (Figure 6e) After a two-sided t-test was performed to assess the statistical significance, showing that only one of the mutants, WDR5 p.Asn213Asp was statistically different to the wild type WDR5, with a p-value of 0.03. In looking at the trends, it shows that there are reduced levels of variant WDR5 protein compared to wild type WDR5, with the levels of WDR5 p.Asn213Asp being the smallest. As shown in Figure 6d-e.

In addition to performing CO-IP of histone H3 and the WDR5 variants, subcellular localization of WDR5 was also assessed. Given the ability of WDR5 to act within different complexes, we wanted to assess levels of localization within the nucleus, where histone H3 is located, to see if there are changes in histone H3-WDR5 binding. To do this both GFP tagged histone H3 and mCherry tagged WDR5 were transfected into HEK293 cells and visualized with a fluorescence microscope (Axio Observer Z1, Zeiss) after 24 hours. (Figure 7) In looking at the merged images one sees an overlap of green and red fluorescence signals, indicating that both histone H3 and the WDR5 variants are found within the nucleus. One difference between the green and red fluorescence signals is the presence of extremely bright green dots within the green nuclear signals. These bright signals represent nucleolar localization. The nucleolus is a separate part of the nucleus that functions in ribosomal synthesis. Neither wild type or variant WDR5 have particular elevations within the nucleolus and do not feature the cytoplasmic localization that occurs in the absence of histone H3. This confirms the immunoblotting results that WDR5 and histone H3 readily bind, and binding is not abolished in mutant WDR5 proteins.

4.6 Effects of WDR5 Mutant Variants on the NSL Complex

Along with the individual protein properties of WDR5, impacts on the NSL complex were also assessed. We began by generating a minimal protein complex containing a subset of the NSL complex, including KANSL1, KANSL2, KANSL3, KAT8 and WDR5. These proteins are responsible for the majority of the stabilizing aspects of the complex, with both KANSL1, WDR5 and KANSL2 binding, while KAT8 is required for stability and acetylation capabilities. The missing components, OGT1, HCF1 and PHF20, are not integral for stability or acetylation, therefore to simplify experiments they were left out. To generate the complex, HEK293 cells were transfected with plasmids of KANSL1, KANSL2, KANSL3, KAT8 and either wild type or mutant WDR5. The protein complexes were subsequently purified, and a western blot completed to assess protein concentrations. In looking at results we notice a few things.

Firstly, the band for KANSL2 was often absent. Despite adding varying levels of KANSL2 the protein did not appear in many of the western blot images. Initially, the plasmid being used was sent for sequencing, determining the sequence was correct. The KANSL2 protein appeared in two separate western blots, twice in the wild type NSL complex and once in both p.Ala169Pro and p.Arg196Cys, respectively. In these instances, the KANSL2 bands were faint, suggesting KANSL2 is likely present in the complex, but weakly expressed, therefore not reliably showing in the western blot. This could indicate the protein is not being transfected at the same rate as its

counterparts. Given the limitation of unreliable KANSL2 bands, there is a lack of imaging data and no conclusion can be drawn about the binding between wild type and mutant WDR5 to KANSL2. Since the complex does not seem to be affected when KANSL2 was absent, with the other proteins stably forming a complex, it leads us believe weak expression is likely causing the KANSL2 loss and therefore continue on with the analysis.

The experiment was designed with FLAG-tagged KAT8 , therefore during the CO-IP proteins that directly or indirectly bind to KAT8 within the complex were pulled down and found within the IP. The four remaining proteins, KAT8, WDR5, KANSL1 and KANSL3 were found in the IP for each WDR5 condition. (Figure 8a) In the image we see relatively similar amounts of each of the subunit proteins. KANSL1 was found in each complex showing mutations in WDR5 did not completely abolish WDR5s ability to bind to KANSL1. Similarly, the NSL complex was able to form despite the mutations in WDR5, meaning these mutations did not sabotage complex stability, which would be seen if KAT8 had been found alone within the IP. However, the stability of the complex and WDR5's ability to bind to KANSL1 could still be affected.

To test if differences between the wild type and mutant WDR5's were significant t-tests were completed. A two-sided t-test was used, with nine observations (the two highest and lowest outliers were removed). The t-test was performed utilizing lane ratios, of WDR5 levels over KAT8, since KAT8 was flag tagged and used to purify the complex, levels of WDR5 are dependent on how much KAT8 was purified. (Figure 8b) In performing the t-tests comparing levels of wild type WDR5 to each of the mutants, we saw three of the mutants, p.Ala169Pro, p.Arg196Cys, and p.Thr208Met have higher levels of WDR5 than the wild type, with p.Arg196Cys having significantly more with a p value of 0.024. Mutant p.Asn213Asp had approximately equivalent levels of WDR5 to the wild type. This indicates that the variant p.Arg196Cys has a greater protein concentration than the wild type WDR5 and may highlight an increased stability of these mutants within the NSL complex.

In addition to looking at levels of WDR5, levels of its binding partner KANSL1 were also assessed. (Figure 8c) Similar to the analysis above, utilizing eight observations, (some smudges into the KANSL1 bands during imaging, producing less usable images), and a protein ratio of the optical densities of KANSL1 over WDR5. This ratio was chosen because levels of KANSL1 are directly dependent on levels of WDR5, these proteins are binding partners, therefore automatically if there are higher levels of WDR5 it should correlate to an increase in levels of KANSL1. Therefore, to see if there are changes in KANSL1 stability we need to use WDR5 as a control. We

find there are no significant changes between the protein concentration of KANSL1 when binding to the WDR5 variants in the NSL complex. There are slightly elevated levels of KANSL1 when both p.Ala169Pro and p.Arg196Cys are present, but both have p- values above 0.05, indicating insignificance. Therefore, it seems that the variations in WDR5 are not impacting the stability of the NSL and not affecting KANSL1's binding within the complex.

To further test if WDR5's binding is impacted by the mutations we decided to test the WDR5 variants ability to bind to their binding partners, KANSL1 and KANSL2 alone. To do this FLAG-tagged WDR5 variants were transfected with HA-tagged KANSL1 and KANSL2 followed by CO-IP and immunoblotting. (Figure 9) We noticed that the expression of the WDR5 variants was approximately the same as within the NSL complex, however changes were noted in their binding partners. When WDR5 is not present there is a dramatic reduction of both KANSL1 and KANSL2 in the whole cell lysate. This loss of KANSL1 and KANSL2 counters the high expression of KANSL1 and weak expression of KANSL2 when either wild type or mutant WDR5 protein is present. The alterations in KANSL1 and KANSL2 expression in the absence of WDR5 indicates that WDR5 necessary for the stability or the production of KANSL1/KANSL2. Compared to studying the minimal NSL complex, by studying these binding partners we are able to tell that KANSL2 is present, however exhibits exceedingly weak expression levels. Given there are minimal differences between the WDR5 variants binding abilities to KANSL1 and KANSL2 it could indicate that neither of WDR5's binding sites, WIN or WBM are affected by the *de novo* mutations.

4.7 NSL Complex Acylation Properties

One important property of the NSL complex is its ability to acetylate histones. The lysine acetyltransferase in the complex, KAT8, has a particular affinity for acetylating histone H4. Patient mutations in histone acetyltransferase complexes are known to a decrease acetylation, as seen in BRPF1, KAT6B and KAT8.[32,108,193] In order to test if the mutations in WDR5 affect NSL's acetylation capabilities a histone acetylation assay was performed, with the purified NSL complex, acetyl-CoA and HeLa histones. For each mutant and wild type NSL complex, two conditions were assessed, one with the presence of acetyl-CoA, and one without, therefore we expect an increase in acetylation to only occur in the presence of acetyl-CoA, resulting in increased optical densities. Results for the experiment are found in figure 10a. Figure 10b-d, show the corresponding graphs of the optical density from the immunoblot bands, with data accumulated from up to five repeated experiments, a maximum of one outlier was removed from the experiment (threshold for removal

was 10000 from either the second highest or lowest value). Statistical significance in acetylation of a particular variant was given if a student's t-test for a variant, between acetyl-CoA containing and non-containing conditions, if the p-value were <0.05 Mean values of optical density were used to generate a ration of variant with acetyl-CoA present over variant without acetyl-CoA, with a value in one representing no differences in acetylation and values greater than one indicative of acetylation. (Figure 10e-g)

We studied three different residues on histone H4K5/8/16. (Figure 10) For histone H4K5, each wild type and variant WDR5 containing NSL complex was able to sufficiently acetylate this residue compared to their non-acetyl-CoA containing counterpart. This acetylation was considered to be statistically significant, with p values less than 0.05. In comparing the ratio of mean optical densities to wild type levels of acetylation both p.Ala169Pro and p.Asn213Asp have comparable or greater levels of acetylation than wild type, whereas NSL complexes containing WDR5 p.Arg196Cys and p.Thr208Met have a dramatic reduction in acetylation levels, approximately 80% less the wild type NSL complex. This indicates that both mutations WDR5 p.Arg196Cys and p.Thr208Met decrease NSL's capability to acetylate histone H4K5.

Large changes in acetylation were also found on histone H4K8. There was two and a half times more acetylated histone H4K8 in the wild type WDR5 acetyl-CoA containing condition compared to the control. A large change in acetylation was also seen in mutant p.Ala169Pro, where levels of acetylation were near that of wild type, with only a ten percent reduction. However, reductions were noticed in the remaining mutations with the loss of acetylation being over 50, 50 and 56 percent for p.Arg196Cys, p.Thr208Met and p.Asn213Asp, respectively.

The largest levels of acetylation of histone H4K16 are in the presence of wild type WDR5 NSL complex. While all four variants were able to acetylate histone H4K16, there was a reduction in acetylation capacity of both mutant p.Ala169Pro and p.Thr208Met equaling approximate a 15 and 30 percent decrease compared to the wild type NSL complex.

When compared to wild type WDR5, each WDR5 variant as part of the NSL complex negatively affected acetylation of at least one histone H4 lysine residues. This shows the WDR5 mutant protein are in fact different than wild type WDR5 and impact the NSL complex functioning and dysregulate acetylation. Similar to mutations in BRPF1, or KAT8, these mutations in WDR5 causing alterations in acetylation could lead to pathogenicity predicted by the software analysis.[32,108]

In addition to acetylation KAT8 has been shown to catalyze other PTMs. KAT8 when in the MSL complex catalyzes propionylation of histone H4K16.[32] Patients with *de novo* mutations in *KAT8* causing global developmental delay, have reduced levels of histone H4K16 acetylation and propionylation.[32] These reductions in acetylation are similar to our patients with *WDR5* muations.[32] To assess if the NSL complex, is implicated in propionylation, and if mutations in WDR5 cause a deregulation in histone H4K16pr, a propionylation assay was performed. This assay utilized the same protocol as the acetylation assay, bar the substitution of acetyl-CoA for propionyl-CoA. The immunoblot for histone H4K16pr did not contain any bands, where the control blots of H4 and FLAG-KAT8 had bands present. (Figure 11) Even with longer exposures no bands were seen. This shows that KAT8 within the NSL complex is either incapable of propionylation of histone H4K16, or it occurs at such trace amounts that it was not detected through imaging. To confirm that the NSL complex is not involved in propionylation further tests should be undergone, however these results illustrate that KAT8's propionylation capabilities are unique to the MSL complex.

4.8 WDR5 Subcellular Localization

Another protein property that may be affected by mutations in WDR5 is subcellular localization. To assess subcellular localization mCherry WDR5 variants or a control of GFP histone H3 were transfected into HEK293 cells and allowed to grow for 24 hours. Post transfection the cells were fixed with an acetone: methonal mixture and stained with DAPI before imaging with a fluorescence microscope (Axio Observer Z1, Zeiss). DAPI is a known nuclear stain, which is reflected by the complete overlap of blue fluorescence with the green histone H3 signal. (Figure 12) In looking at wild type WDR5 the red fluoresce of WDR5 surrounds the nuclear DAPI signal in the merged image. In the individual red fluorescence signal there are circular red rings with holes in the center, showing WDR5 is not in the nucleus. Since there is no overlap between the nuclear signal and the WDR5 signal it indicates that WDR5 is cytoplasmic.

In assessing the variants both p.Ala169Pro and p.Arg196Cys have the same cytosolic pattern of localization as the wild type WDR5. However, both WDR5 p.Thr208Met and p.Asn213Asp are slightly different. While the majority of these variant proteins are in the cytosol, there are multiple examples within each photo where the variants are in both the nucleus and cytosol. This indicates these variants are pancellular, located throughout the cell, instead of solely cytosolic. These changes in subcellular localization may cause alterations in WDR5s cellular functions, although

since both WDR5 p.Thr208Met and p.Asn213Asp are still found in the cytosol there should be no inhibition in performance of the WDR5 variants on the basis of cellular location.

DISCUSSION

The maintenance of global patterns of epigenetic modifications are integral to life. Modification patterns and the genes sustaining them are evolutionarily conserved. Alterations in these genes or the epigenome at large can have detrimental impacts, ranging from cancer to neurological disorders. Within this thesis two different genes in histone acetylation complexes, *KAT14* and *WDR5*, were analyzed. Our aim is to understand the effects *de novo* mutations have in protein and epigenomic properties and their impacts on patients.

In the gene encoding *KAT14* two different *de novo* mutations, one on each genetic locus, were identified in a patient with Diamond Blackfan anemia. Diamond Blackfan anemia is a rare genetic blood disorder, leading to congenital red cell aphasia.[82] Whole exome sequencing determined the patient did not have alternations in ribosomal protein genes commonly associated with Diamond Blackfan anemia, leading to the hypothesis that one or both of the mutations in *KAT14* could be responsible for the clinical phenotype. Both *KAT14* mutations, p.Arg567Ter and p.Arg596Trp, are predicted to be pathogenic. p.Arg567Ter leads to the production of a truncated protein missing the HAT domain, and p.Arg596Trp alters an evolutionarily conserved residue, indicating this residue is likely essential to protein function.

Through experimental analysis KAT14 variant proteins were shown to be significantly different to the wild type KAT14. Firstly, the termination mutation produces a protein that is 40 KDa smaller than the wild type protein, however there is an up regulation of protein production compared to wild type. The second mutation, p.Arg596Trp, produces substantially less protein than its wild type KAT14 counterpart, meaning either protein stability or production is affected. The mutations also impacted the proteins self-binding capabilities. Through this thesis it was discovered that KAT14 was able to bind to itself. While each variant was also able to bind to wild type KAT14, the termination variant, p.Arg567Ter, was less abundant than both wild-type and p.Arg596Trp counterparts. Therefore, KAT14 p.Arg567Ter causes a reduction in self-binding, which may mean that within the 200 amino acid residues lost, there are specific regions critical for self-binding.

Acetylation properties of KAT14 are also altered by the mutations. Both variants cause a reduction in acetylation of histone H3 and H4, with the biggest reductions in histone H4K8/16 and H3K23 acetylation. A generalized reduction in acetylation was seen throughout histone marks.

The alterations in KAT14 stability and protein size combined with the reduction of histone acetylation mean it is likely that both p.Arg567Ter and p.Arg596Trp KAT14 mutations cause a deregulation of epigenetic programs and impact development. Since the patient has one mutation, p.Arg567Ter or p.Arg596Trp, on each genetic locus, the consequences we have seen *in vitro* will have a much greater impact, given no compensation or haploinsufficiency mechanisms can occur. For these reasons it seems likely that both p.Arg567Ter and p.Arg596Trp KAT14 are pathogenetic and involved in the causing the disease progression of the patient.

Acetylation modifications are incredibly versatile and involved in a width breadth of functions. Most notably histone acetylation is seen as a transcriptional activation mark, allowing the decompaction of the chromatin, or the change from heterochromatin to euchromatin. Acetylated lysine residues are known to have unique roles including; histone H4K12ac in DNA repair and mitotic or meiotic progression, histone H3K9ac and H3K14ac's involvement in nuclear receptor co-activation and histone H4K5ac is linked to cell cycle progression.[322] Histone H4K16ac has been found to prevent the condensation of 11nm nucleosomes into the 30nm chromatin fibers, which allows the chromatin to maintain its open formation, and therefore enhances transcriptional activation.[323]

Changes in histone acetylation have been shown to have wide effects in mammals, ranging from neurological disorders to alterations in hematopoietic stem cell (HSC) development. Hematopoiesis refers to the development of blood cells, a process which occurs throughout the life span. HSCs are the cell type from which all blood cell lineages are derived. BRPF1, a histone reader found in the MOZ/MORF complexes is integral to the maintenance of HSC's. When *Brpf1* is deleted in mice HSC it leads to lethality pre-weaning.[126] These pups displayed aplastic anemia and bone marrow failure, indicating BRPF1is critical to HSC functioning.[126] *Brpf1* null pups featured reduced proliferation of erythroid progenitors, thus there were fewer progenitors and less red blood cells.[126] Acetylation levels were also impacted within the bone marrow, with *Brpf1* loss leading to reductions of histone H3K9/K14 /K23 acetylation, with minor effects on histone H4K16ac.[126] Thee substantial changes in histone H3 acetylation in the bone marrow of *Brpf1* knockout mice lead to the bone marrow failure seen.

Like BRPF1, the patient we are studying with two *de novo* mutations in the gene encoding *KAT14* displays the importance of acetylation in the bone marrow and proper HSC functioning. The patient studied with *KAT14* mutations has a rare bone marrow disorder, Diamond Blackfan anemia, which is characterized by red blood cell aphasia. Diamond Blackfan anemia does not

affect all blood cell lineages, it only affects the erythroid lineage, indicating HSC are likely not affected. In cases of Diamond Blackfan anemia caused by ribosomal protein mutations, erythroid progenitor cell levels are reduced through an increase in apoptosis.[87] This is similar to the reduction in erythroid progenitors seen in the *Brpf1* mutant pups.[87,126] In studying the patient KAT14 mutant proteins we see a similar reduction in acetylation, with histone H4K16, H3K14 and H3K23 acetylation being affected, as well as additional losses of acetylation on H4K5, H4K8, H3K4 and H3K18. Akin to the HSC of *Brpf1* knockout mice, the patient being studied featured KAT14 mutations on both gene loci, with each locus having a separate mutation. Since both mutations alter structure, stability and the acetylation properties of KAT14, a compensation model will not occur, instead there is a greater loss of function leading to a negative impacts on the individual. Due to the similarities in the reduction of histone acetylation seen in the KAT14 mutants and *Brpf1* knockout pups, it seems reasonable to investigate if this loss of acetylation in the KAT14 variants is occurring in the bone marrow.[126]

In future studies an animal model similar to the one generated by You et al. should be utilized, expressing the KAT14 mutations p.Arg567Ter and p.Arg596Trp, ideally creating a double mutant mouse to better mimic the patients genotype.[126] Two *Kat14* mutant mouse models should be generated, one specifically selecting for HSC, and another selecting for erythroid progenitor cells. These models will help to further elucidate the roles of KAT14 within the bone marrow and help determine more succinctly if the KAT14 mutations are leading to the pathogenesis of Diamond Blackfan anemia seen within our patient. In this hypothetical mouse model, one would expect the pups to suffer anemia, with a reduction of red blood cells and erythroid progenitor cells. These mice would also be expected to have altered levels of histone acetylation, likely mimicking that seen in our *in vitro* experiments. If the criteria are met in mice, and bone marrow failure occurs, one could conclude that KAT14 is critical in bone marrow functioning and the process of erythropoiesis.

In Diamond Blackfan anemia caused by ribosomal protein mutations, patients display gene dysregulation. Two genes, *p53* and *JNK* are implicated in potential mechanisms of pathogenesis for Diamond Blackfan anemia. One theory of pathogenesis is that Diamond Blackfan anemia causes defects within the bone marrow, and ultimately prevents the p53 inhibitor MDM-2 from interacting with p53.[86] This inability to inhibit p53 leads to an accumulation of p53 within erythroid progenitor cells, causing apoptosis.[86] Therefore, there is a reduction in erythroid progenitor cells and thus a decrease in erythroid proliferation, causing the red blood cell shortage

seen in patients. While KAT14, and the ATAC complex have yet to be implicated in p53 regulation, it's evolutionarily related complex, SAGA has. ADA2B, a subunit of SAGA, is recruited to p53 dependent promoters, and controls p53 regulation, however, its homolog, ADA2A in the ATAC complex has not been found at these promoters.[58] Despite ATAC not currently being implicated in p53 regulation, in a *Kat14* mutant mouse model, potential alterations in p53 should be assessed. Changes in p53 with this model could lead to a better understanding of the ATAC complex and help validate the theory that dysregulation of p53 is causing the mechanism of pathogenesis of Diamond Blackfan anemia.

The second protein that should be examined is the TGF-β signaling pathway protein JNK. In the past the TGF-β cascade has been shown to have a role in erythroid differentiation and in the negative regulation of stem cells proliferation.[90] Pluripotent stem cells of patients with Diamond Blackfan anemia have dramatically increased levels of JNK, which indicates that these stem cells may not be able to undergo proliferation, since JNK negatively regulates proliferation.[90] Pluripotent stem cells are unique due to their ability to become any cell type, or to remain and propagate as stem cells. This means the patient stem cells can develop into erythroid progenitors, or hematopoietic stem cells. An accumulation of JNK in erythroid progenitors would prevent proliferation and the generation of mature erythrocytes and therefore cause a reduction in red blood cells, similar to what was seen in the *Brpf1* HSC knockout mice.[90,126] KAT14 and the ATAC complex have been shown to regulate JNK.[62] ATAC works to activate JNK target gene transcription in housekeeping functions, however once osmotic stress occurs, ATAC acts to suppress JNK, preventing pathway activation.[62] Given ATAC's role in JNK regulation, mutations in KAT14 could prevent the suppression of JNK activation, leading to increased levels of JNK. This dysregulation of JNK seen in Diamond Blackfan anemia patient pluripotent stem cells leads to an interesting link between Diamond Blackfan anemia and KAT14, which should be further assessed.

This dysregulation of the TGF-β may be in part responsible for the mechanism of pathogenesis of Diamond Blackfan anemia, particularly in our patient with KAT14 mutations. These two potential mechanisms of pathogenesis for Diamond Blackfan anemia having links to KAT14 should be further studied. In generating a *Kat14* double mutant mouse model, one could assess the levels of gene expression for both p53 and JNK, as well as related proteins, such as downstream members of the p53 and TGF-β pathways. This could be done through RNA-seq of the mutant model's bone marrow and verified using RT-qPCR. Additionally, apoptosis of erythroid

progenitors, HSC and pluripotent stem cells could be assessed to determine if there is a reduction in cells and exactly which stage of erythropoiesis is ultimately affected leading to the clinical phenotype seen in our patient.

In vivo confirmation of deregulation of histone acetylation by mutant KAT14, and assessment of KAT14's ability to regulate JNK and p53 will provide a greater understanding behind the mechanistic progression of Diamond Blackfan anemia. By further understanding the pathogenic mechanisms behind Diamond Blackfan anemia, more appropriate therapies to manage the clinical phenotypes associated with Diamond Blackfan anemia could be developed.

KAT14 is often studied in relation to the ATAC complex, therefore many properties of the protein itself are unknown. While studying the mutations in KAT14, discoveries about the protein were made. First, one of KAT14's binding partners is itself. This self-binding capacity has yet to be mentioned in the literature. Since the binding efficiency of the p.Arg567Ter truncated protein was reduced, it is likely that either stabilizing interactions supporting self-binding or part of the KAT14 binding region occurs within this lost 200 codons.

In addition to self-binding, a further understanding of KAT14's role as a histone acetyltransferase was developed. From the literature, it is understood that KAT14 acts as a histone H4 acetyltransferase, with affinity for H4K16 and to a lesser extent H4K5. Any histone H3 acetyltransferase capabilities of the ATAC complex were attributed to either KAT2A or KAT2B. However, through acetylation assays of KAT14 alone it was determined that KAT14 itself is able to stimulate histone H3 acetylation. KAT14's acetylation of histone H3K23 was significant, with acetylation of H3K4, H3K14 and H3K18 also being stimulated. Both KAT2A/B are known to predominantly exert their effects on H3K9, therefore, the acetylation of other histone H3 residues by the ATAC complex may fully or partially be due to KAT14. To further understand the differences between KAT14 and KAT2A/KAT2B's acetylation properties additional histone acetylation assays should be performed. By performing these assays with KAT14, KAT2A and KAT2B alone, and then both KAT14 and KAT2A/B as a part of the full ATAC complex one would have a better understanding of each KAT's unique acetylation properties and whether binding in the complex enhances acetylation. Though needing further study, KAT14 appears to have a more global role in acetylation than it was initially credited for.

In addition to understanding the individual roles of KAT2A/B and KAT14 in acetylation, looking at acetylation by the ATAC complex when the KAT14 mutations are present is necessary. Through our preliminary work on the complex, it appears that KAT14 mutations result in

decreased expression of ATAC complex subunits, such as SGF29 and YEATS2. Furthermore, there is a dramatic reduction of KAT14 p.Arg567Ter within the ATAC complex. Acetylation of ATAC whole cell lysates also appeared to be dysregulated by KAT14 mutations. In the future purification of the entire complex and subsequent acetylation assays are needed to understand the full scope of the effects caused by the KAT14 variants. These initial results indicate the KAT14 mutations impact the composition, structure and acetylation properties of the ATAC complex.

The second gene studied in this thesis was *WDR5*. WDR5 is a histone reader and functions to facilitate protein-protein interactions in many complexes. Seven unrelated patients with *de novo* mutations in the *WDR5* gene were identified through WES. All patients with *WDR5* mutations present with a global developmental delay syndrome with hallmarks of intellectual disability and motor and speech delay. Patients are also feature a wide range of multi-system effects including skeletal or limb deformaties, cardiac defects, and neurological problems, such as hypotonia, brain abnormalities. Patients with similar clinical presentations have been found with mutations in genes apart of WDR5's NSL complex, leading to the hypothesis that the patient mutations in *WDR5* are responsible for a dysregulation of the NSL complex, ultimately leading to the global developmental delay phenotype. Of the seven patients, six presented with four different missense mutations, p.Ala169Pro, p.Arg196Cys, p.Thr208Met and p.Asn213Asp which were used for further study. Utilizing three predictive software's to determine potential pathogenicity, each of the four variants were predicted to be pathogenic or destabilizing in one or more of these programs.

All four of the *WDR5* mutations occur in or near the region encoding the fourth WD40 domain, which makes up the fourth β-propeller of the protein structure. The role of the fourth β-propeller is unknown, therefore we cannot determine if the mutations affect the propellers functions, however overall confirmation of the protein structure may be altered. These mutations may also affect WDR5's ability to bind to proteins through the WIN site. One genomic location important to WIN site binding is an evolutionarily conserved threonine at 191. Given the missense mutations proximity to the WIN binding site and the mutations clustering in or near the fourth WD40 domain it is likely there is a either a structural or functional change in the WDR5 mutant proteins.

To further assess how the protein variants are different to the wild type WDR5, mutant and wild type plasmids were generated and underwent transfection to see if stability was affected. Three WDR5 variants, p.Arg196Cys, p.Thr208Met and p.Asn213Asp were found to negatively impact the protein concentration of WDR5. Each of these variants produced approximately 50 percent less WDR5 protein than wild type WDR5. This indicates that either less protein is being

produced by these variants or that the protein produced is not stable, and therefore being degraded upon production. In either case WDR5 protein loss could cause major effects, with less WDR5 available to function in one of its many roles throughout the body.

WDR5 acts as a histone reader, binding to histone H3. Each WDR5 variant impeded WDR5's ability to bind to histone H3. There was a seven percent reduction in binding of p.Ala169Pro, a 20 percent reduction in p.Arg196Cys, a 25 percent reduction for p.Thr208Met, and p.Asn213Asp had a 50 percent reduction in histone H3 binding. This mirrors the results seen with the WDR5 protein alone, with minimal differences between p.Arg196Cys and the wild type WDR5 protein and dramatic differences in p.Arg196Cys, p.Thr208Met and p.Asn213Asp. The role of a histone reader is to locate residues on histones which need to be modified, determine which modification is needed and can also interact with the surrounding sequence to help facilitate the interaction or modification.[324] The reduction in WDR5 binding to histone H3 caused by variants p.Arg196Cys, p.Thr208Met and p.Asn213Asp shows that these mutations affect WDR5's ability to act as a histone reader. This could lead to changes in modifications, alterations in WDR5's ability to facilitate histone-protein interactions or cause decreases in transcription since modifications signaling transcriptional activation may be missing. Alterations in transcription, and chromosome condensation may greatly impact protein composition within the cells and throughout an organism.

One WDR5 containing complex is the NSL complex, which is a part of the larger family of enzymatic MYST. The NSL complex has largely been studied in flies and mice, where it is best known for being enriched at constitutively active housekeeping genes and aiding in the regulation of RNA Polymerase II transcription.[147] Within the NSL complex WDR5 plays a role in structural integrity, binding to both KANSL1 and KANSL2. Since WDR5 is important in NSL's structural integrity we wondered if the mutations in WDR5 cause alterations to the NSL complex. In performing a transfection and a CO-IP it was found that levels of proteins within the complex remained stable, with minimal differences between levels of WDR5 and its binding partner KANSL1 when the mutants were present. This indicates the WDR5 mutations have minimal impacts on the NSL complex's stability and the mutations are likely not affecting KANSL1 binding. The other binding partner of WDR5 in the NSL complex is KANSL2. KANSL2 was not stably pulled down in any condition therefore no conclusions can be made about the impact of WDR5 mutations on KANSL2 binding. In a subsequent experiment, WDR5 and purified with its two binding partners, KANSL1 and KANSL2. Through this experiment it was shown that when WDR5 was absent, KANSL2 was not produced, and there was a dramatic reduction in the protein

levels of KANSL1, within the whole cell extract. This shows WDR5 is required for the expression of both KANSL1 and KANSL2.

The most notable role of the NSL complex is its ability to act as a lysine acetyltransferase The NSL complex is a KAT complex, and functions as a major regulator of H4K16 acetylation, but has also been known to acetylate both histone H4K5 and H4K8.[148] It has previously been shown that mutations within histone reader proteins, such as BRPF1, a protein a part of the MOZ/MORF complex, can impact the histone acetylation properties of the complexes they are a part of.[108] Similar to the mutations in BRPF1 proteins, we found WDR5 mutations also lead to histone hypoacetylation. Mutations in WDR5 caused dysregulation of histone H4 acetylation, with mutant p.Ala169Pro decreasing levels of H4K5ac, mutants p.Ala169Pro, p.Arg196Cys and p.Thr208Met decreasing levels of H4K16ac and mutants p.Ala169Pro, p.Arg196Cys, p.Thr208Met and p.Asn213Asp decreasing levels of H4K8ac. Additionally mutants p.Arg196Cys and p.Thr208Met abolished H4K5ac.

Recently, insights were gained on the NSL complex's histone acetylation capabilities in mammals.[149] KAT8, the acetyltransferase of the NSL complex is also a part of the MSL complex. One difference in KAT8s interactions with these complexes is that KAT8 associates more stably in the NSL complex compared to the MSL complex.[149] These complexes vary on their affinity to acetylate different lysine residues. The MSL complex predominately catalyze acetylation of histone H4K16ac, whereas the NSL complex preferentially acetylates H4K5 and H4K8, with a much weaker affinity for H4K16.[149] Similar results were noted in the KAT assays performed in this thesis, with wild type NSL complexes having the greatest levels of acetylation on histone H4K5 and H4K8, with lesser levels of H4K16 acetylation seen.

A similar experiment was performed to assess the propionylation capabilities of the NSL complex, remarkably the NSL complex was not able to propionylate H4K16. KAT8 as a part of the MSL complex, serves to both acetylate and propionylate histone H4K16.[32] This lack of propionylation activity of the NSL complex helps distinguish the two separate KAT8 complexes, MSL and NSL. Given KAT8s importance within the NSL complex, further studies to see how patient mutations in KAT8 affect the NSL complexes acetylation properties levels may help to further elucidated these complex patterns of acetylation.

Global histone hypoacetylation is commonly found when MYST complex proteins are mutated. *De novo* mutations in *KAT5* causing global developmental delay impact acetylation, and alterations in KAT5 leading to Alzheimer's disease cause a reduction in histone H4K5, H4K12

and H4K16 acetylation.[175,176] Within the MOZ/MORF complexes *de novo* mutation in *KAT6A, KAT6B* and *BRPF1* cause global developmental delay and a dysregulation in acetylation of histone H3 and H4. *KAT6A* mutations increase H3K18ac, and decrease H3K9ac, while *BRPF1* mutations decrease levels of H3K23ac.[108,178,193]

Mouse models mimicking the loss of MYST complex members also experienced changes in acetylation. When *Kat5* is lost in mice, there's a decrease in histone H4K12ac at transcription start sites, *Kat7* loss leads to a 90 percent reduction in H3K14ac and *Kat8* mutant mice show reductions in H4K16ac.[32,101,137] This dysregulation in histone acetylation of patients with global developmental delay and mouse models of global developmental delay show the importance of maintaining acetylation levels in neuronal development and functioning. These changes in acetylation found when MYST complex mutations occur are similar to the changes seen in our four missense WDR5 variants.

Our *WDR5* patients had a remarkably similar clinical presentations to those affected with MYST complex mutations, specifically in genes *KAT5*, *TRRAP*, *KAT6A/B*, *BRPF1*, *KAT8*, *KANSL1*, *OGT1* and *HCF1*. All mutant phenotypes shared a hallmark of global development delay, with intellectual disability, and speech and motor delay. These patients also shared many similar facial deformities, neurological and brain abnormalities and a wide range of multi-systemic effects. With the similarities in both phenotype and the presence of hypoacetylation it indicates that the *WDR5* mutations are likely part of the pathogenesis causing global developmental delay. In the future mouse models could be used to confirm WDR5 protein mutations cause changes in acetylation *in* vivo. Since *WDR5* is intolerant to loss of function, the generation of a complete knockout of *Wdr5* in mice is not an option, however a neuronal or cerebrum specific *Wdr5* knockout mouse or mice with the specific missense mutations seen in patients could be generated. These mice should be assessed for changes in neuronal development and the viability of neuronal stem cells, since *Kat8* cerebrum knockout mice have cerebral hypoplasia and alterations in neural stem and progenitor cells.[32]

The NSL complex plays an important role in regulating methylation through communication with the MLL/SET complexes. Specifically, NSL helps to coordinate and regulate transcription of the MLL/SET complexes. Proper functioning of the NSL complex is required for MLL/SET methylation.[153] Loss of the NSL complex causes a reduction of histone H4K16ac and H3K4me2.[153] One similarity among both complexes is the shared subunit, WDR5. Through the study of *WDR5* patient mutations in the NSL complex, we have seen alterations in histone H4 acetylation levels,

however methylation levels in the presence of the mutations were not assessed. The WDR5 variants could impact the MLL/SET complexes due to the dysregulation of acetylation caused by the WDR5 variants in NSL. Dysregulation of acetylation could impede NSL's ability to promote MLL/SET's methylation, and therefore cause a reduction in H3K4me2. Another possibility is that mutations in *WDR5* have a greater negative impact on MLL/SET, since the complex loses the upstream regulation by the NSL complex combined with potential structural or functional losses in the MLL/SET complex that are caused by the mutations within WDR5. In the future, assessing how and if the mutations in WDR5 impact H3K4me2 and the MLL/SET complex would help understand the mechanism of pathogenies of the WDR5 mutants.

Despite losses in WDR5 potentially affecting histone methylation patterns, the effects of acetylation were focused on. This selective focus was due to the clinical presentation of the patients found with genomic mutations in *WDR5*. The phenotype of the *WDR5* patients was exceptionally similar to those with Koolen-de Vries Syndrome, featuring mutations in *KANSL1* and patients with *KAT8* mutations, in addition to patients suffering from mutations in other MYST complexes.[32,246] Patients with mutations in SET1/MLL, particularly within the *KMT2D* gene have been identified with Kabuki Syndrome, however a large majority of these patients do not present with global developmental delay, or speech and motor delay which are hallmarks features of the patients with *WDR5* mutations.[296] Kabuki syndrome patients have distinct facial dimorphisms, including highly arched eyebrows, everted lower eyelids, depressed nasal tip and prominent ears.[296] These unique facial dimorphisms found in Kabuki syndrome patients have not been identified in *WDR5* patients. In looking at table 2, one sees that there is a greater overlap between features found in MYST mutation patients, these phenotypic features resemble those found in the *WDR5* patients, however there is similarities in the clinical presentations between WDR5, MYST and SET1/MLL patients. Therefore, it is more likely that alterations in the NSL complex occur from the *WDR5* mutations and thus impact clinical presentation, therefore studied in depth.

Although a large role of the NSL complex is to control histone modifications, it also has roles in housekeeping functions. Constitutive activation is found in most genes that NSL is bound to.[147] The promoter regions of these bound genes exhibit increases in histone activating modifications and the presence of the NSL complex is necessary for the recruitment of RNA Pol II particularly at DNA replication-related element.[147] Alterations in NSL complex proteins downregulate 27 different essential genes including genes involved in RNA splicing, vacuole homeostasis,

translation, telomere elongation and mitochondrial translation.[149] Due to NSL's role in housekeeping functions the complex is integral for the survival of cells.[149]

Many subunits within the NSL complex, including WDR5 are required for cells survival.[149] With the importance of the NSL complex in cell survival, it is likely that the mutations studied impacting WDR5, and the acetylation properties of the NSL complex will have downstream effects on RNA Pol II transcription and on the regulation of varying housekeeping genes. While cell death may not occur due to the WDR5 protein variants, unlike *KANSL2, KANSL3* and *KAT8*, these mutations did impact H4K5ac and H4K8ac.[149] Both H4K5ac and H4K8ac marks occur at transcriptional start site of human genes allowing for the binding of RNA Pol II.[149] With the mutations causing a significant loss of acetylation at both of these marks it indicates that the ability of RNA Pol II to bind at target promoters may be lost or reduced when WDR5 mutations occur. In the future to determine if the WDR5 mutations are impacting RNA Pol II and the transcription of housekeeping genes, WDR5 mutations could be stably inserted into human cells through CRISPR, followed by RNA sequencing to understand the global transcription profile. Genes previously determined to be down-regulated by *KANSL2* and *KANSL3* should be studied to see if expression is altered.[149] ChIP-seq should be performed to determine if alterations in *WDR5* lead to a reduction in H4K5ac and H4K8ac at the transcriptional start sites of NSL target genes, indicating a reduction in expression.

Through this additional research on *WDR5* mutations and the NSL complex a better picture of the NSL complex's roles in mammalian development will be understood. Most research on the NSL complex has been done in drosophila, thus a deeper understanding would fill the void in mammalian research. While both acetylation properties of *de novo* mutations in *KAT8, WDR5, OGT1* have been assessed, it would be intriguing to further assess these properties in *de novo KANSL1* and *HCF1* gene variants.[1,32,158] Neither *HCF1* nor *KANS1* mutations have had their acetylation properties fully delineated, however these mutations cause a similar global developmental delay. By fully understanding how these patient mutations alter the NSL complex greater insights on the functioning and importance of the NSL complex will be gained.

Frequent identification of patients harboring mutations within MYST complex is occurring, however these disorders are still rare on a population level. The lack of patients presenting with mutations in each gene makes it hard to draw definitive conclusions on key phenotypic features of patients and on the mechanistic pathways of dysregulation. Despite differences seen among individuals with these mutations, as a collective there are many similarities including the hallmarks

of global developmental delay, distinctive facial dimorphisms, other multi-system complications, as well as a dysregulation of histone acetylation and changes to neuronal development. By looking at MYST mutations as a family of related disorders it provides a greater understanding of the similarities and differences among the individual MYST complexes. This could lead to better insights about the roles of each complex within development and the potential to understand if similar pathways or acetylation markers are being deregulated. An in depth understanding of the MYST family as a whole could be used to increase the population size of interest allowing for more definitive and meaningful information to be generated from the patients. By treating these individual disorders as a greater family of disorders it is likely to have larger impacts on the individuals affected. One benefit of understanding this complex of related disorders, is the potential development of a framework for doctors to identify these disorders. Utilizing a larger population base would make doctors more aware of these disorders and more likely to recognize the disorders thus ensuring better care and quicker sequencing requests. In addition to better patient identification, a larger group of disorders attracts more attention and thus increases the odds of research being done on the complexes and incentivizes the development of therapeutics to industry. In understanding the minute differences and similarities among dysregulation in these MYST mutants' therapeutic treatments could be optimized to be the most beneficial to patients. In all there are many pros to studying MYST mutations, including those in WDR5, as a collective whole.

The identification of rare genetic disorders, such as those described in our patients with *KAT14* and *WDR5 de novo* variants, is happening more readily. This is due to advancements made in sequencing technology, allowing for faster and more cost-effective methods of genomic sequencing, and the ability to process and analyze the sequencing data. Whole genome or exome sequencing is an important tool for many reasons and is a driving force in the advancement of precision or personalized medicine. Precision medicine aim's to use one's genetic profile to prevent or treat diseases. Facets of precision medicine include genetic testing, diagnostic measures including screening or the development of therapeutics designed to optimize treatment efficacy.

Screening for genetic predispositions is a common preventative measure. This occurs to see if certain genetic features, such as CNVs or SNPs, that predict an increased risk for a disease are present. One well known example is genetic testing for cancer predisposition, with markers such as mutations in BRCA1 and BRCA2 known to increase the risk of cancer, particularly breast cancer.[325] Another common use of genetic testing is to determine one's risk for being a carry of a

genetic disorder. This often coincides with the use of a genetic counsellor to help determine if potential children will inherit a disorder and can be done with both partners to predict if any pre-existing risks that may arise in future children.

Despite the usefulness of genetic screening for prevention or preconception screening to identify predispositions to disorders, the identification of *de novo* mutations, such as the individuals studied with *KAT14* and *WDR5* mutations would not benefit from this. Since *de novo* mutations arise spontaneously neither of these screening methods would account for these mutations. The most useful method of genetic testing is to assess if *de novo* mutations occur prenatal screening in utero or assessing after birth when the onset of symptoms or distinctive phenotypes are recognized. In the case of prenatal screening, it is generally used to assess chromosomal changes, such as trisomy or neural tube defects. Although this screening is useful, it neglects the assessment of more specific genetic alterations that cause disorders. Understanding possible genetic causes for disease as early as possible is important in mitigating impacts. For example, phenylketonuria, is caused by a mutation within chromosomal 12 resulting in abnormal phenylalanine metabolism, with affected infants presenting with growth failure or global developmental delay.[326] However due to screening in newborns, children with phenylketonuria undergo dietary modifications to prevent symptoms from arising.[326] Early screening could have massive impacts for individuals affected by genetic developmental problems, such as Diamond Blackfan anemia or global developmental delay, which both arise during embryonic development. The multisystemic nature of both the disorders studied may greatly benefit from early detection to help administer treatments quickly to minimize impacts on growth or neurological and cardiac problems.

With the increasing reliance on screening and the expanding range of genetic disorders identified, in the future it may be possible to assess for a larger panel of genetic alterations in prenatal screening. An advantage of this would be the development of therapies to counteract or alleviate the severity of the developmental disorders found within the developing fetus. For example, a mutation in a KAT gene may result in dysregulation of acetylation, and thus a deficit in transcription of important developmental genes leading to disorder. Therefore the development of therapeutics to counteract the loss of acetylation or additions of the downregulated proteins may help compensate for the mutations. A similar concept for individuals recognized with a genetic disorder to develop therapies to counteract epigenetic dysregulation would also be useful.

This tailoring of therapeutics to an individual based on their genetic profile is a notable goal of precision medicine. It involves generating a drug regimen personalized to the patient, taking into account a person's genetic and proteomic background, what cell types are affected, the microbiome, among lifestyle and family factors to develop an optimized treatment plan. For individuals suffering from both global developmental delay syndromes and from Diamond Blackfan anemia this may be an ideal form of treatment, given both disorders can be caused due to multiple different mutations, which may alter different downstream pathways. Additionally, both feature have wide range of multi-system effects that are unlikely to be aided by only one drug. By understanding the mechanisms of pathogenesis of both disorders, and why body systems are affected appropriate treatments could be designed to address individual needs.

With the ever-evolving landscape of technology it is optimistic to think precision medicine will become the standard of care, especially in the treatment of genetic developmental disorders. For this to become common place and offer the greatest benefit to affected individual's further characterization of genetic mutations is crucial. By understanding the impacts mutant variants, such as *KAT14* and *WDR5,* and their mechanisms of pathogenesis we will be better able to treat affected individuals. However, with an abundance of new genes being implicated in diseases this pursuit of delineating the genetic underpinnings of disease is far from complete and offers countless possibilities for further research.

CONCLUSION

The goal of this thesis was to delineate the epigenetic mechanisms underlying genetic disorders. Through the course of the project both *KAT14* and *WDR5* gene varients were studied as subunits a part of ATAC and NSL complexes.

Our results showed that *KAT14* mutants had detrimental impacts on the size and stability of the KAT14 protein, leading to changes in self binding and substantial changes in acetylation capabilities. Through preliminary research it appears that KAT14 mutations impact the formation of the ATAC complex, particularly the nonsense variant, and affect acetylation. Given the diverse and likely damaging nature of these KAT14 mutations it seems they are responsible for the clinical presentation of Diamond Blackfan anemia found in the patient. These mutations highlight a new genetic mutation responsible for Diamond Blackfan anemia and may be useful in better understanding the pathogenic mechanisms involved with the disorder.

The second gene studied in the project was *WDR5*. Our results showed the four *de novo* mutations impacted protein concentration of WDR5 and reduced WDR5's ability to bind to histone H3. This

suggests that WDR5's ability to act as a histone reader is diminished by these mutations. Although the stability of the NSL complex was not impacted, the acetylation capabilities of the complex were greatly reduced by the mutations. Additionally, subcellular localization of certain WDR5 mutants shifted from being cytosolic to pancellular, which may alter the function of WDR5. *WDR5* patient global developmental delay phenotypes closely resemble other identified MYST complex disorders and further show the importance of MYST complexes in development.

ILLUSTRATIONS

Figure 1

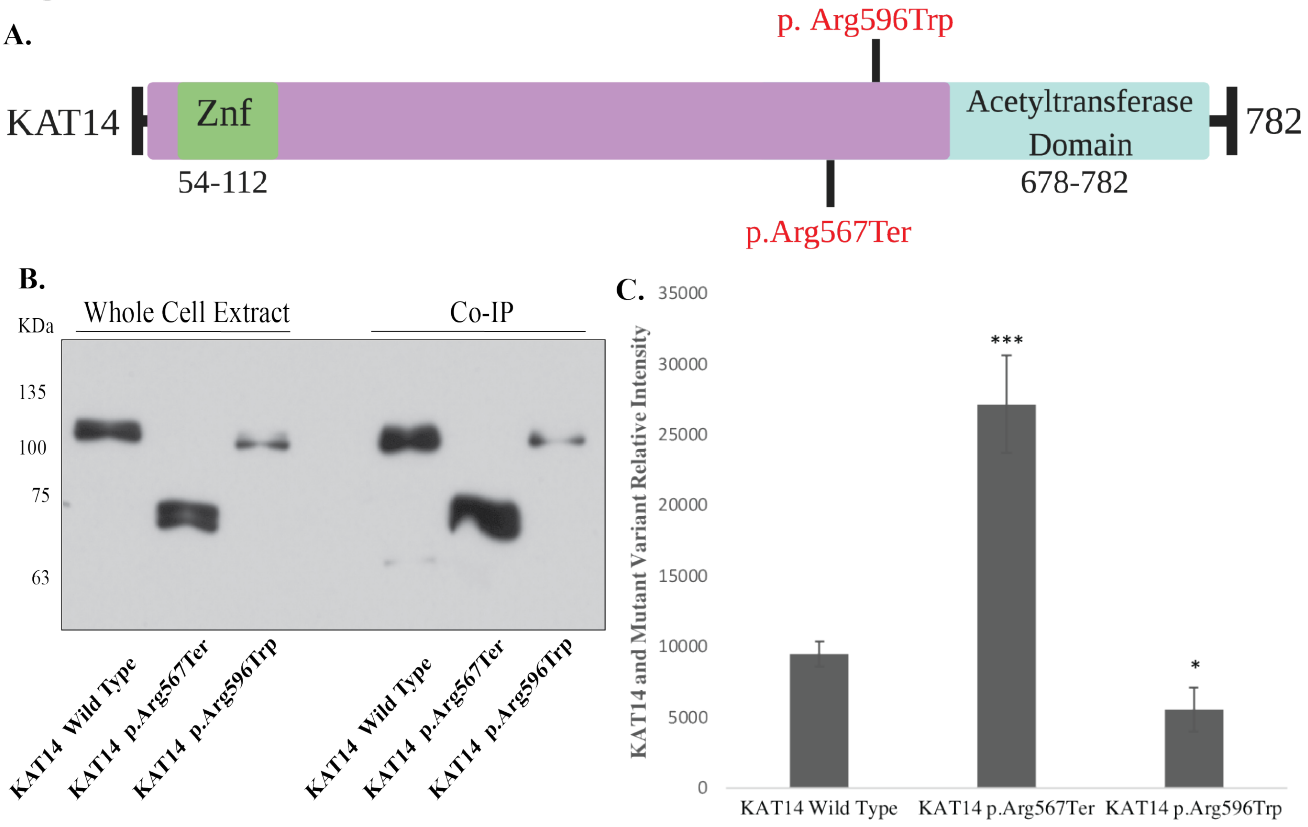


Figure 1. Domain organization of KAT14 and variants from individual with Diamond Blackfan Anemia, CO-IP confirmation of protein abnormalities in KAT14 variants.

(A) Schematic representation of the KAT14 protein, with both *de novo* variants. Both mutations occur near the KAT domain of KAT14 near the C-terminus of the protein. The catalytic properties of KAT14 are found in the C-terminus, from residues 678-782. Image generated in BioRender.com.

(B) FLAG-tagged KAT14 wild type and mutant variants underwent transfection in HEK293 cells, followed by IP. As mutational analysis predicted, KAT14 p.Arg567Ter produced a truncated variant, and KAT14 p.Arg596Trp produced decreased levels of protein.

(C) Analysis of KAT14 immunoblots were repeated 11 times. Mean values of optical densities are presented with standard deviations. Variant protein levels were compared to those of wild type levels. $p<0.05$ is considered statistically significant.*** equals $p \leq 0.001$.

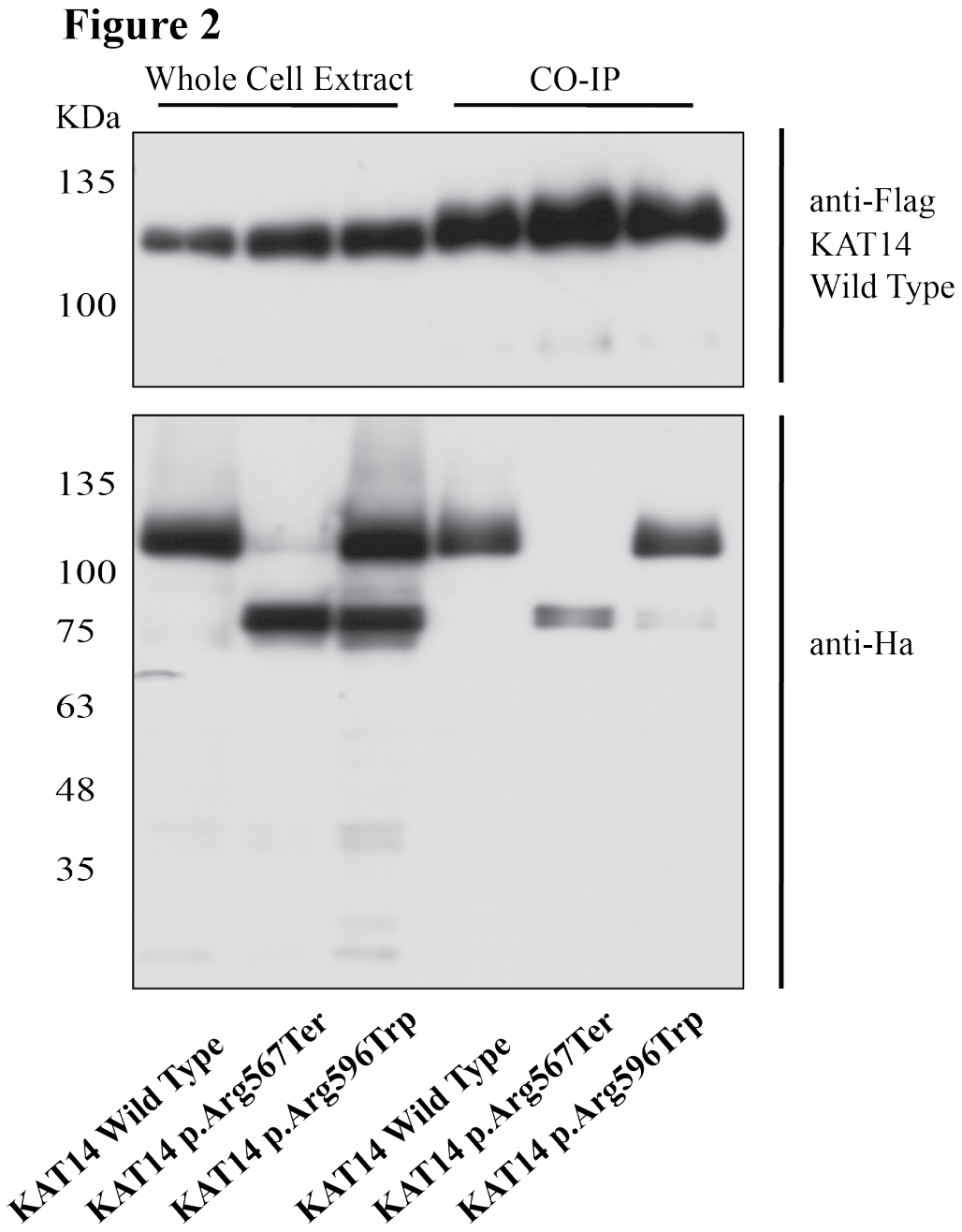


Figure 2. KAT14 and variant self-binding capabilities

(A) FLAG-tagged wild type KAT14 and HA-tagged KAT14 variants were transfected into HEK293 and underwent CO-IP. Both variants and wild type KAT14 had the ability to bind to FLAG-tagged wild type KAT14.

Figure 3

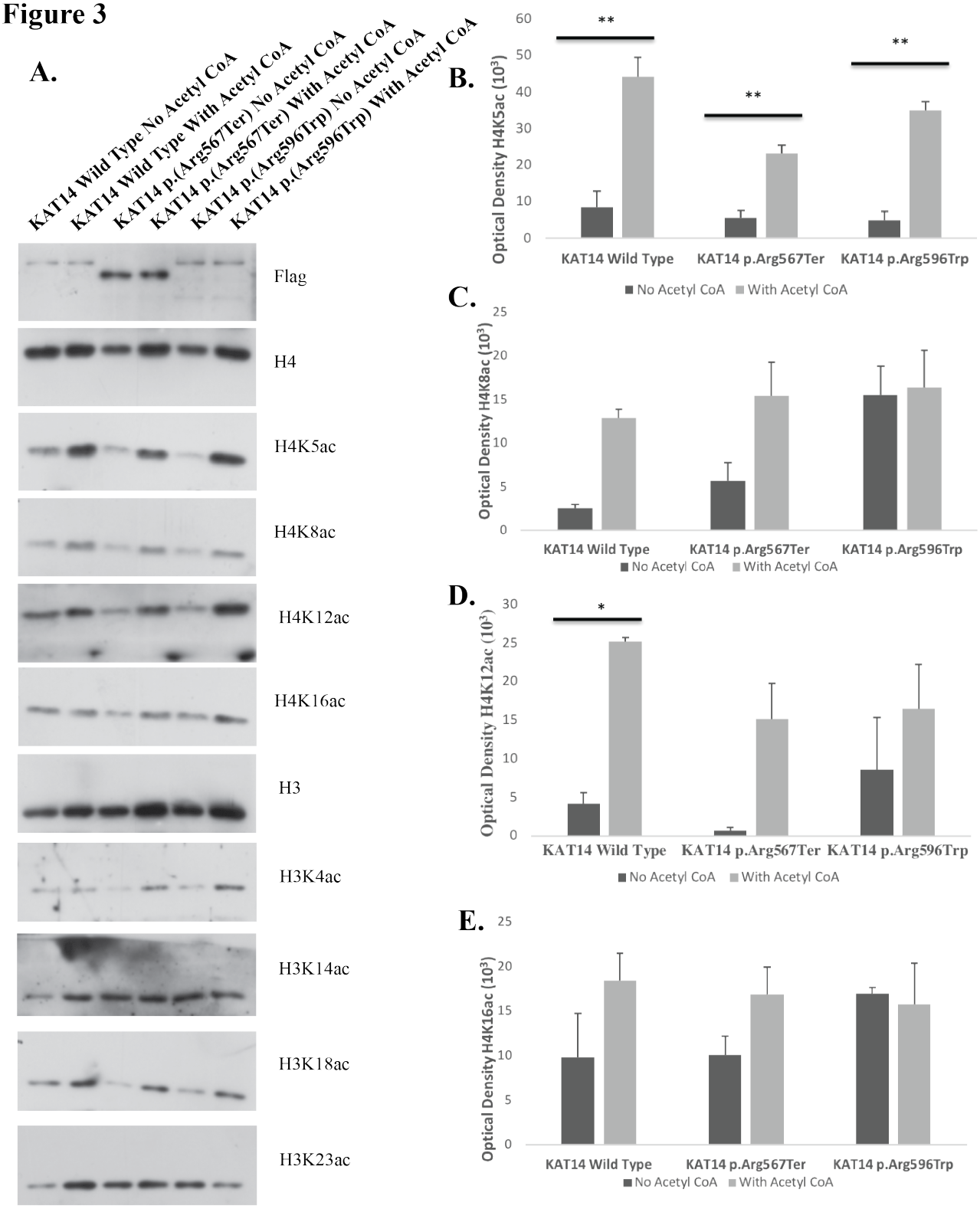

A.
KAT14 Wild Type No Acetyl CoA
KAT14 Wild Type With Acetyl CoA
KAT14 p.(Arg567Ter) No Acetyl CoA
KAT14 p.(Arg567Ter) With Acetyl CoA
KAT14 p.(Arg596Trp) No Acetyl CoA
KAT14 p.(Arg596Trp) With Acetyl CoA
Flag
H4
H4K5ac
H4K8ac
H4K12ac
H4K16ac
H3
H3K4ac
H3K14ac
H3K18ac
H3K23ac
B.
Optical Density H4K5ac (10³)
KAT14 Wild Type
KAT14 p.Arg567Ter
KAT14 p.Arg596Trp
No Acetyl CoA
With Acetyl CoA
C.
Optical Density H4K8ac (10³)
D.
Optical Density H4K12ac (10³)
E.
Optical Density H4K16ac (10³)

Figure 3

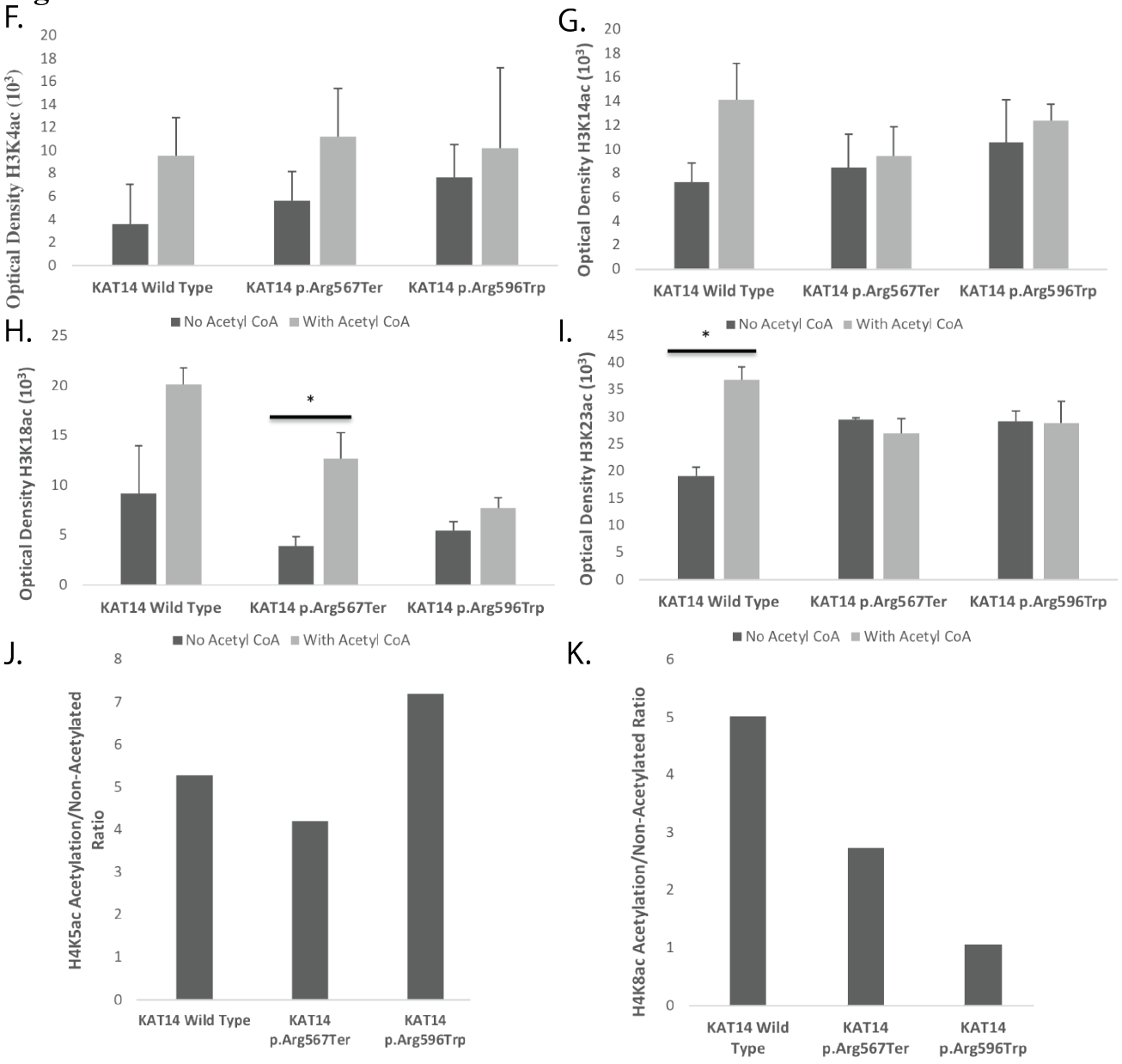
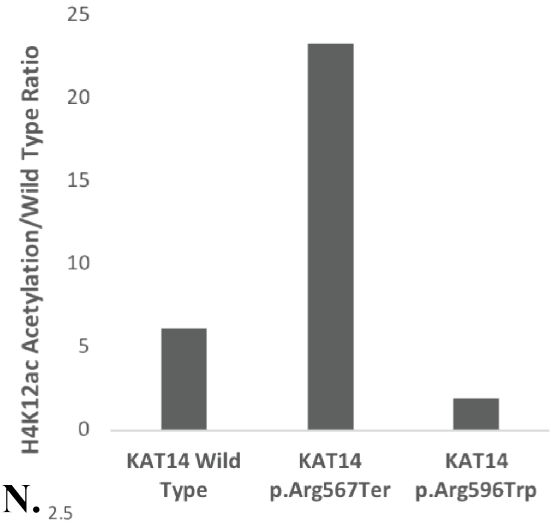
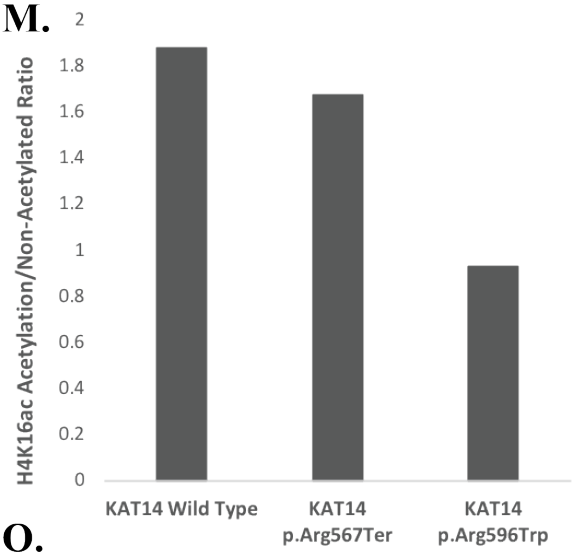

F.
Optical Density H3K4ac (10^3)
KAT14 Wild Type
KAT14 p.Arg567Ter
KAT14 p.Arg596Trp
No Acetyl CoA
With Acetyl CoA
G.
Optical Density H3K14ac (10^3)
H.
Optical Density H3K18ac (10^3)
*
I.
Optical Density H3K23ac (10^3)
J.
H4K5ac Acetylation/Non-Acetylated Ratio
K.
H4K8ac Acetylation/Non-Acetylated Ratio

Figure 3

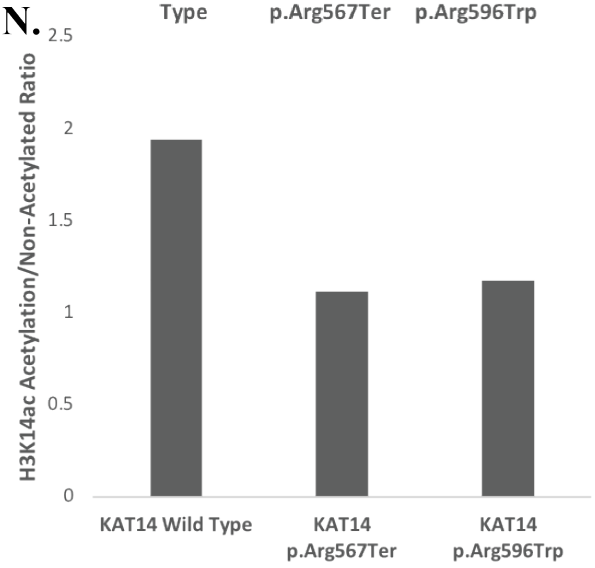
L.



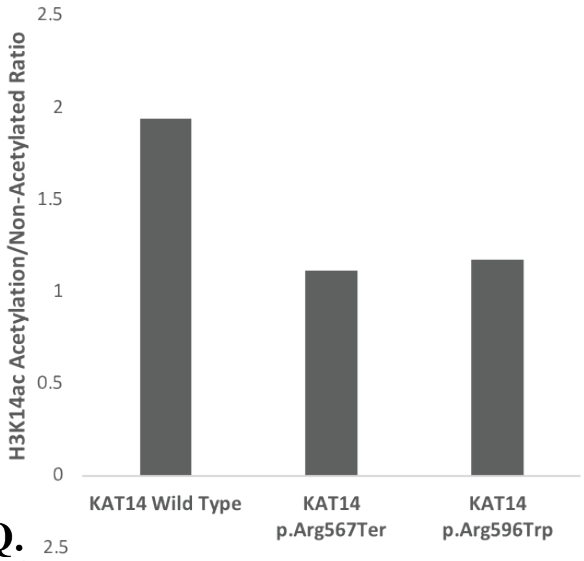
M.



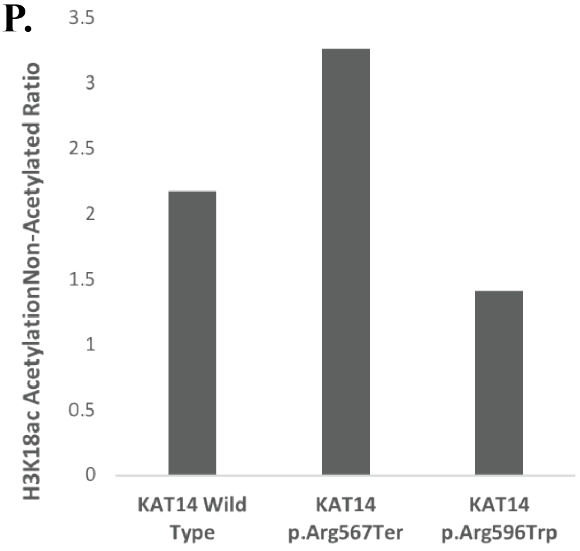
N.



O.



P.



Q.

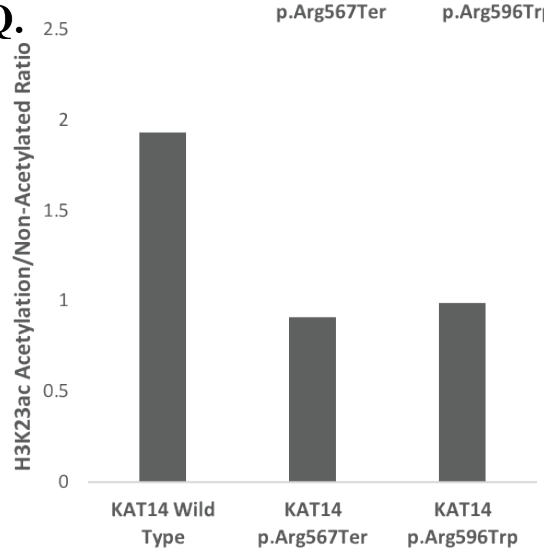


Figure 3. KAT14 variants histone acetylation

(A) Histones acetylation assays were performed with the HeLa histones utilized as substrates, for affinity purified proteins (from Figure 1 B.). Histone acetylation was assessed through the use of specific histone H3 and H4 antibodies.

(B) (C) (D) (E) (F) (G) (H) (I) Optical densities of each wild type and variant KAT14 acetylation levels were compared to non-acetylated controls. Figures represent the changes in acetylation over three experiments represented in the immunoblots.

(J) (K) (L) (M) (N) (O) (P) (Q) Ratios of acetylated over non-acetylated variants were calculated. The ratios show vast changes in acetylation between the wild type and KAT14 variants.

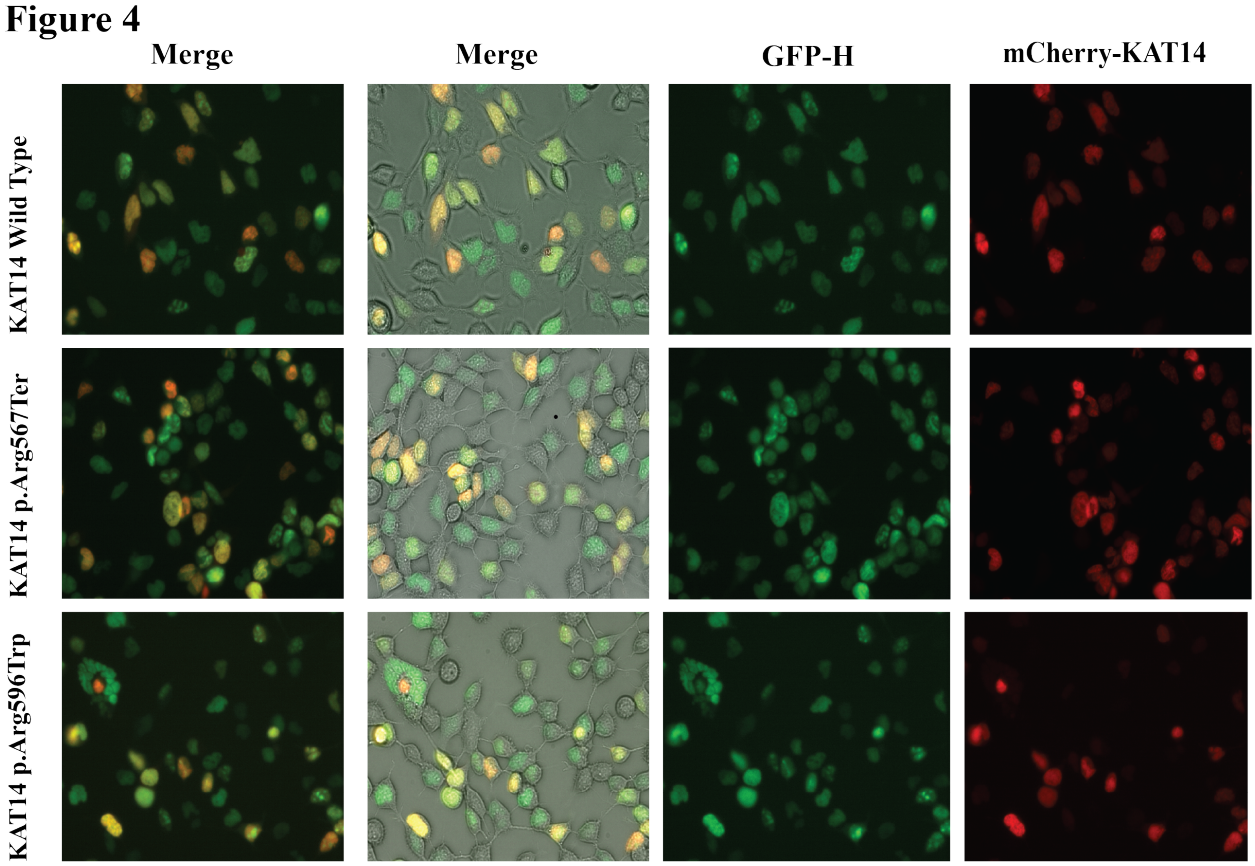


Figure 4. KAT14 and variants subcellular localization

mCherry-tagged KAT14 variants and GFP-tagged histone H3 were transfected into HEK293 cells and visualized through fluorescence microscopy. The merged images between the GFP and mCherry signals show the overlap between histone H3 and the KAT14 variants. KAT14 p.Arg596Trp levels were reduced by approximately half compared to its wild type counterpart.

Figure 5.

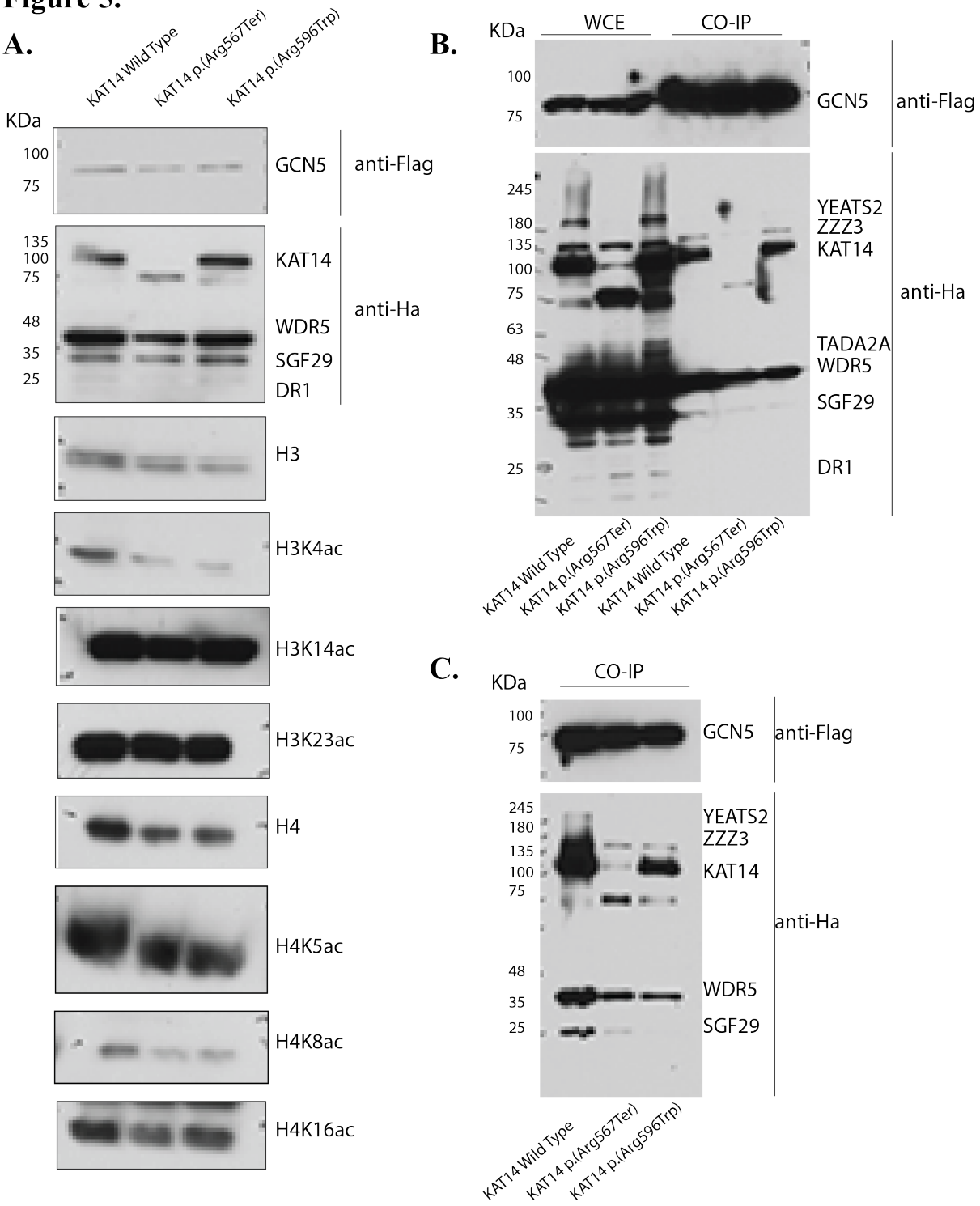


Figure 5: ATAC Complex and whole cell extract acetylation

(A) FLAG-tagged GCN5 and HA-tagged ATAC subunits underwent transfection in HEK293. Whole cell extracts were assessed for protein levels, and histone acetylation levels.

(B) (C) Two separate trials of ATAC complex transfection, showing ATAC subunits.

Figure 6

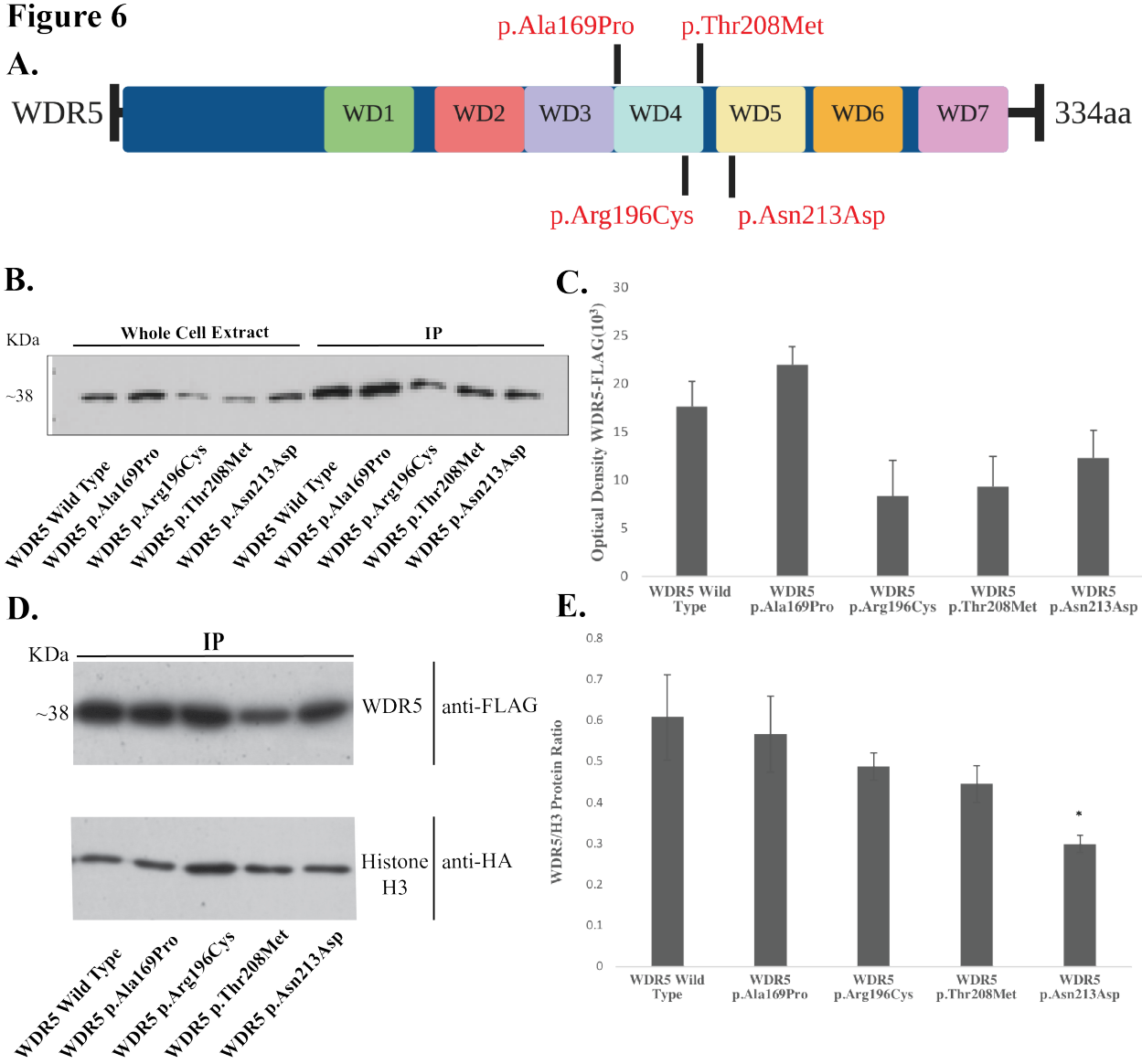


Figure 6. WDR5 domain organization with *de novo* variants causing global developmental delay syndromes, IP representation of WDR5 mutants and WDR5 histone H3 binding

(A) Schematic representation of the WDR5 protein, with the four *de novo* variants. Image generated using BioRender.com.

(B) FLAG-tagged WDR5 variants underwent transfection in HEK293 cells, followed by IP. Differences were seen in protein levels between the variants. (C) Analysis of immunoblots (n=8) Mean values of optical densities are presented with standard deviations as error bars.

(D) FLAG-tagged histone H3 and HA-tagged WDR5 wild type and mutant variants underwent transfection in HEK293 cells, followed by IP. (E) Analysis of immunoblots (n=6). Mean WDR5-histone H3 ratios, presented with standard deviation as error bars. * equals $p \leq 0.05$

Figure 7.

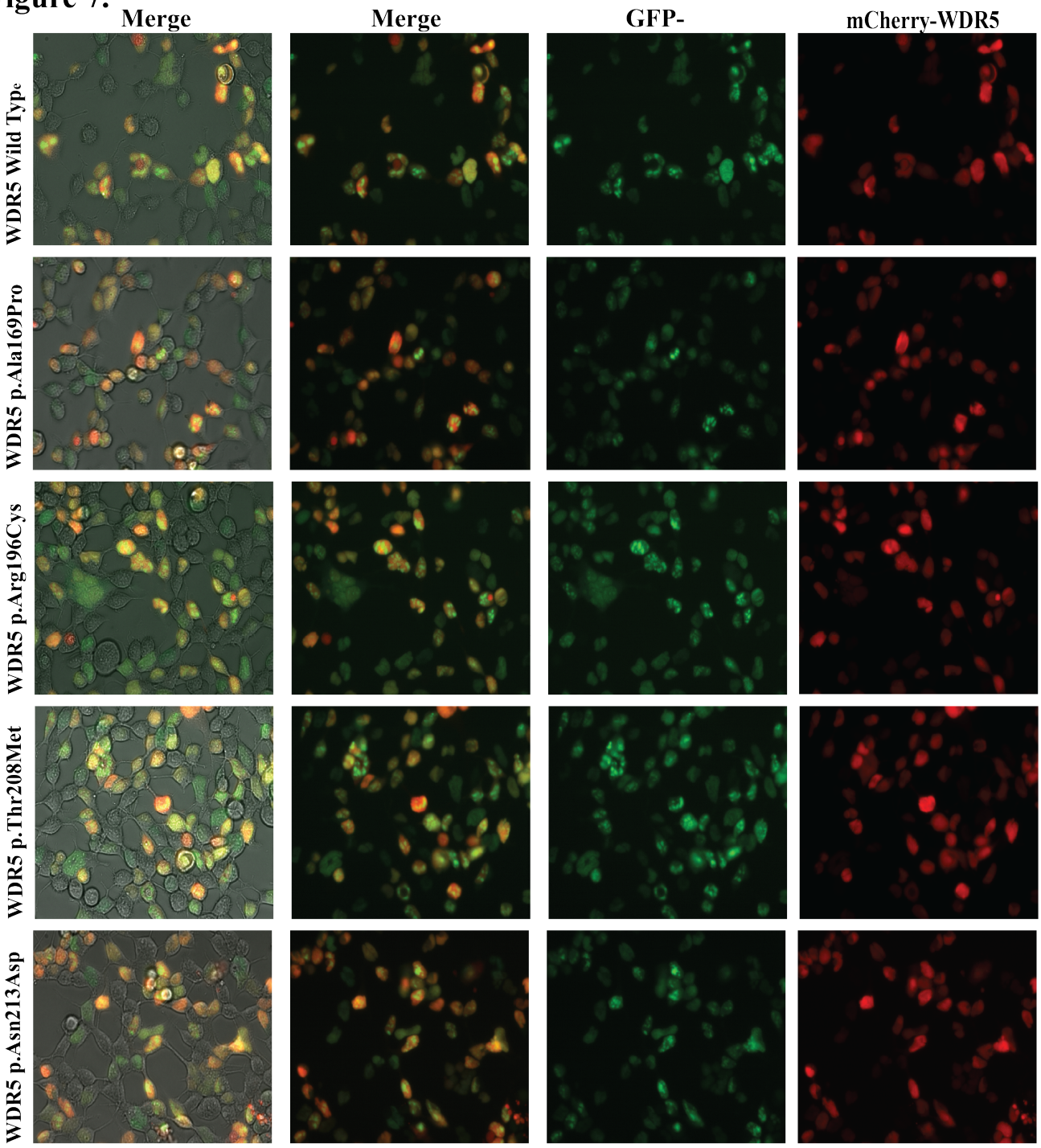


Figure 7. WDR5- histone H3 sub cellular localization

(A) mCherry tagged WDR5 variants and GFP tagged H3 were transfected into HEK293 cells and visualized through fluorescence microscopy. The merged images between the GFP and mCherry signals show the overlap between histone H3 and the WDR5 variants.

Figure 8

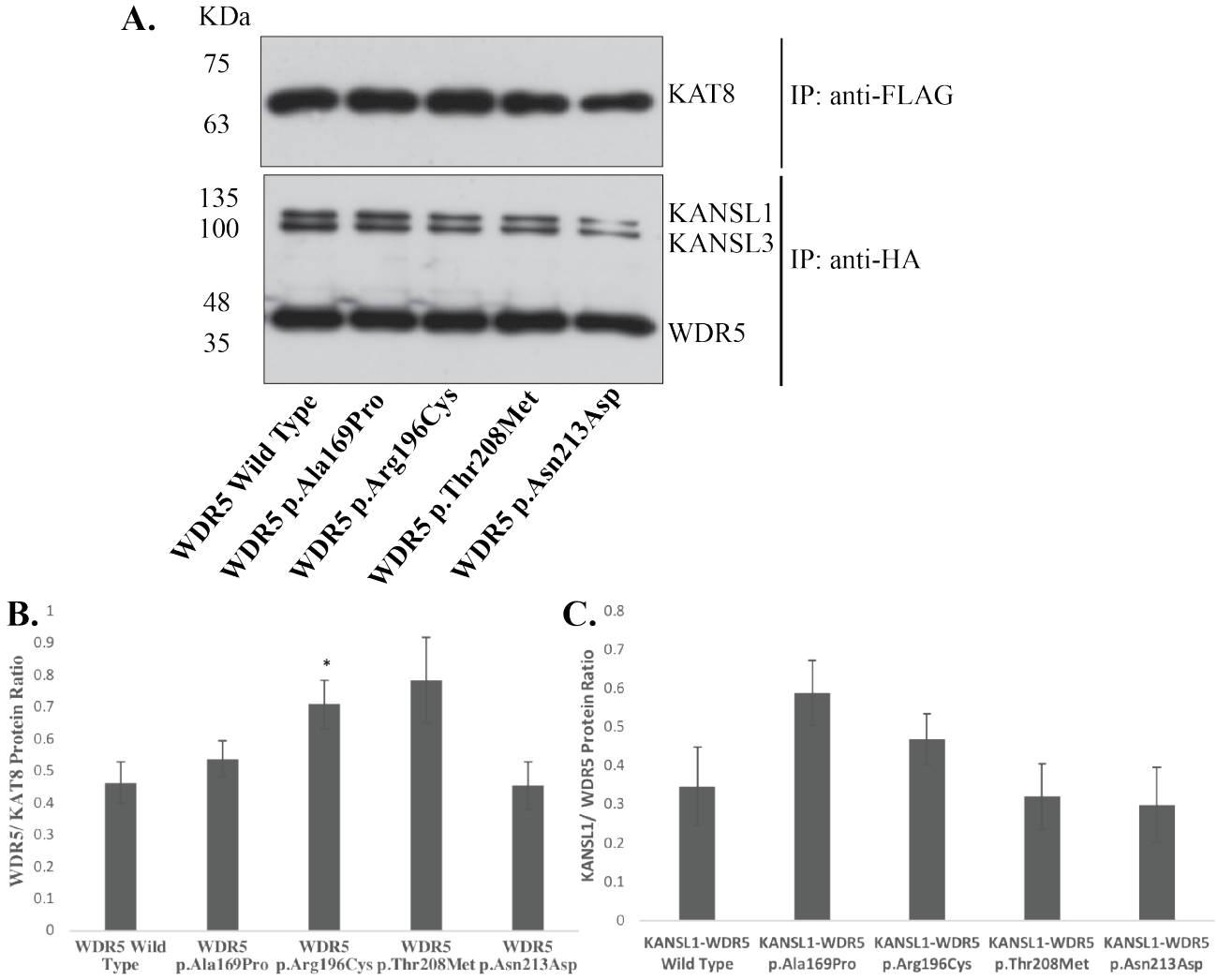


Figure 8. NSL Complex CO-IP

(A) FLAG-tagged KAT8 and HA-tagged KANSL1, KANSL2 and wild type & mutant WDR5 variants underwent transfection in HEK293 cells, followed by IP.

(B) Analysis of NSL immunoblots (n=5). Mean WDR5-KAT8 ratios are presented with standard deviation used as error bars. $p \leq 0.05$ is considered statistically significant. * equals $p \leq 0.05$

(C) Analysis of NSL immunoblots (n=5). Mean KANSL1-WDR5 ratios are presented with standard deviation used as error bars. $p \leq 0.05$ is considered statistically significant.

Figure 9

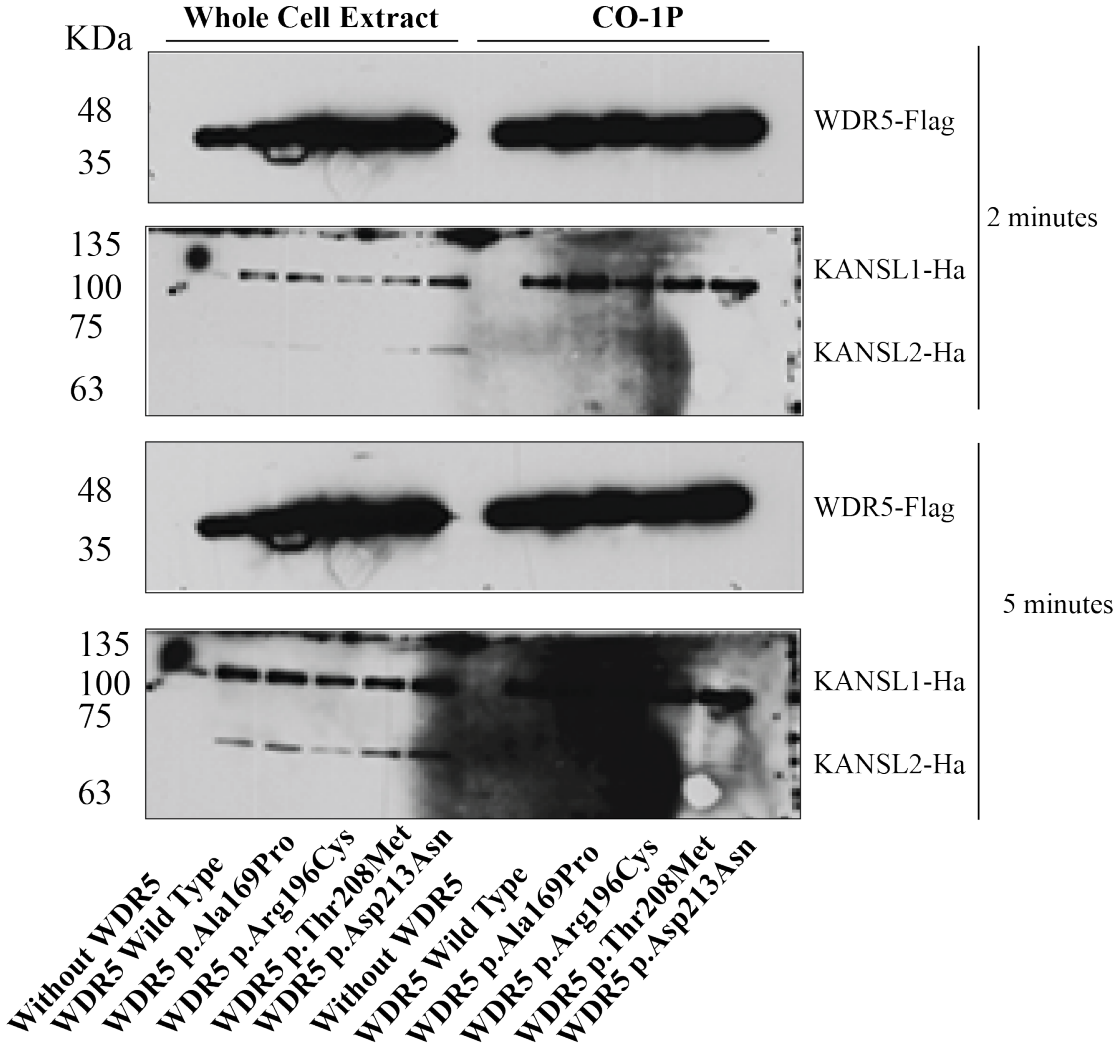


Figure 9. WDR5 binding to KANSL1 and KANSL2

HA-tagged KANSL1 & KANSL2 and FLAG-tagged wild type & mutant WDR5 variants underwent transfection in HEK293 cells, followed by IP.

Figure 10

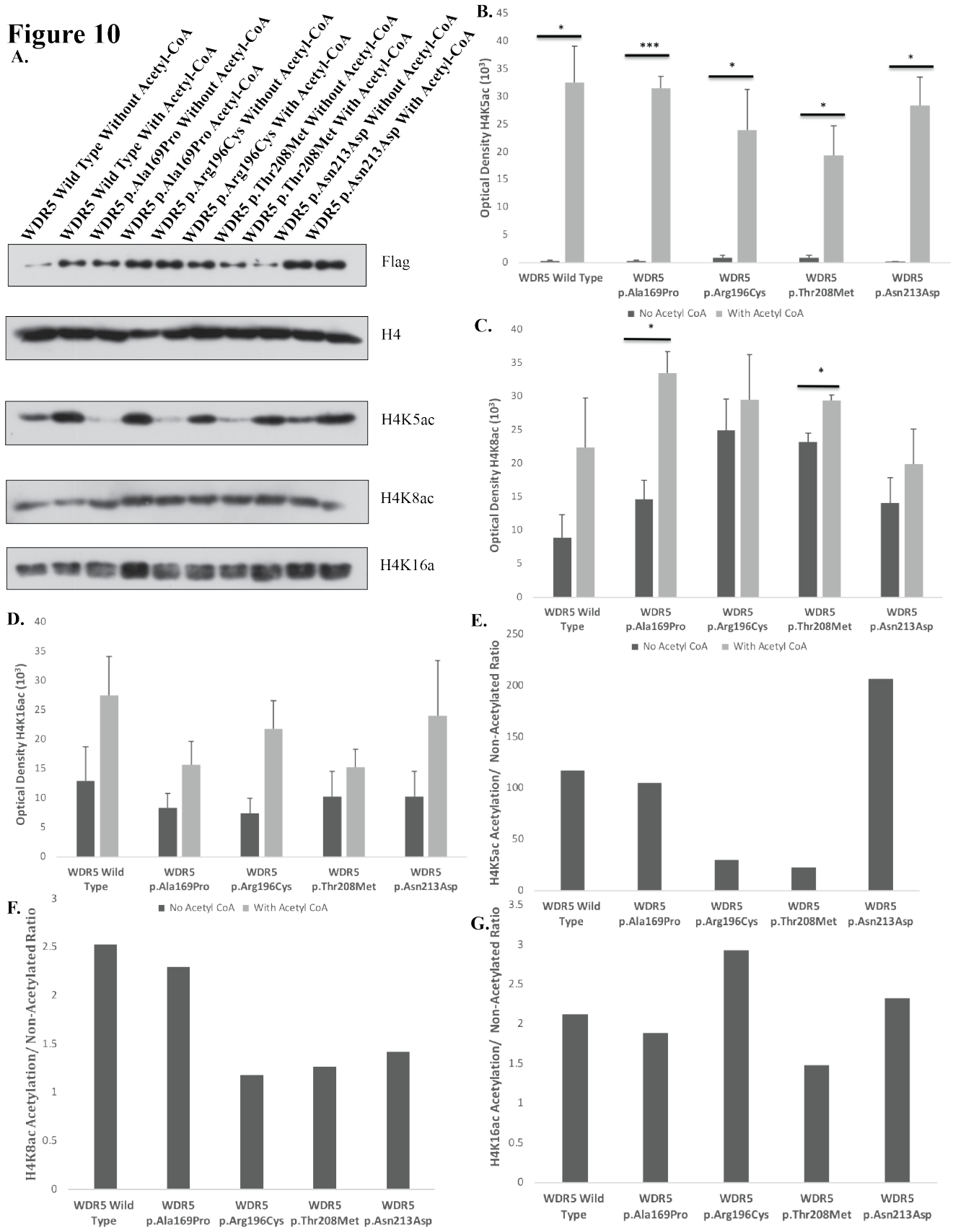

A.
WDR5 Wild Type Without Acetyl-CoA
WDR5 Wild Type With Acetyl-CoA
WDR5 p.Ala169Pro Without Acetyl-CoA
WDR5 p.Ala169Pro Acetyl-CoA
WDR5 p.Arg196Cys Without Acetyl-CoA
WDR5 p.Arg196Cys With Acetyl-CoA
WDR5 p.Thr208Met Without Acetyl-CoA
WDR5 p.Thr208Met With Acetyl-CoA
WDR5 p.Asn213Asp Without Acetyl-CoA
WDR5 p.Asn213Asp With Acetyl-CoA
Flag
H4
H4K5ac
H4K8ac
H4K16a
B.
Optical Density H4K5ac (10³)
WDR5 Wild Type
WDR5 p.Ala169Pro
WDR5 p.Arg196Cys
WDR5 p.Thr208Met
WDR5 p.Asn213Asp
No Acetyl CoA
With Acetyl CoA
C.
Optical Density H4K8ac (10³)
D.
Optical Density H4K16ac (10³)
E.
H4K5ac Acetylation/ Non-Acetylated Ratio
F.
H4K8ac Acetylation/ Non-Acetylated Ratio
G.
H4K16ac Acetylation/ Non-Acetylated Ratio

Figure 10. NSL acetylation assays

(A) Histones acetylation assays were performed with the HeLa histones utilized as substrates, for affinity purified proteins (from Figure 7a.). Histone acetylation was assessed through the use of specific histone H4 antibodies.

(B) (C) (D) Optical densities of each wild type and variant WDR5 acetylation levels were compared to non-acetylated controls. Figures represent the changes in acetylation in five experiments, represented in the immunoblots. $p \leq 0.05$ is considered statistically significant. * equals $p \leq 0.05$, *** equals ≤ 0.001.

(E) (F) (G) Ratios of acetylated over non-acetylated variants were calculated. The ratios show vast changes in acetylation between the wild type and WDR5 variants.

Figure 11.

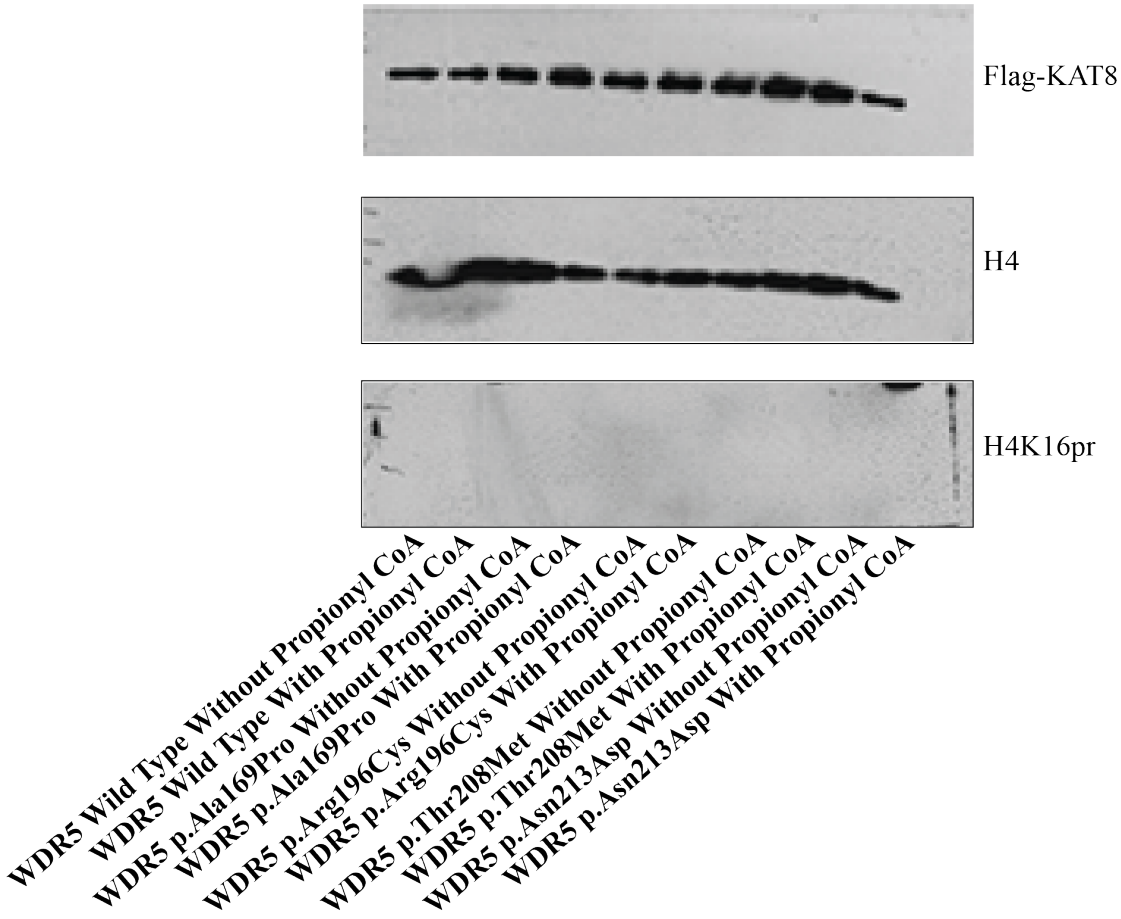


Figure 11. NSL propionylation assays

Histones propionylation assays were performed with the HeLa histones utilized as substrates, for affinity purified proteins (from Figure 7). Histone propionylation was assessed through the use of specific histone H4 antibodies.

Figure 12

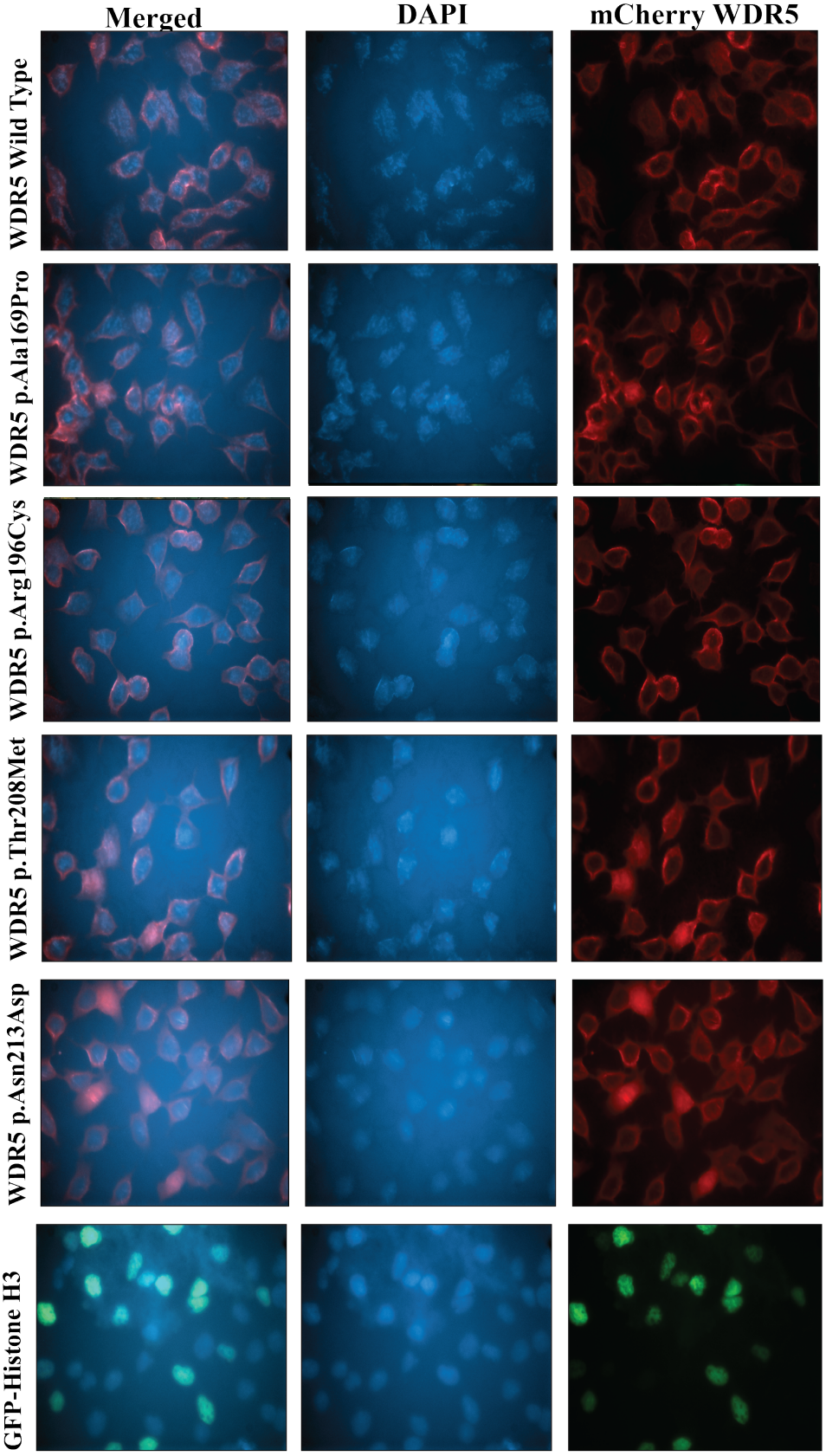


Figure 12. WDR5 and variants subcellular localization: mCherry tagged WDR5 variants were transfected into HEK293 cells prior to fixation and DAPI staining. Cells were visualized through fluorescence microscopy. The merged images between the DAPI and mCherry signals show the differences in signals between the nuclear DAPI staining and the mCherry WDR5 variants.

REFERENCES

1. Verseput, J. *et al.* A clustering of missense mutations in the crucial chromatin modifier WDR5 defines a new neurodevelopmental disorder. (2018).
2. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589**, 2914–2922 (2015).
3. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *GENES Dev.* **23**, 781–783 (2009).
4. Treviño, L. S., Wang, Q. & Walker, C. L. Phosphorylation of epigenetic "readers, writers and erasers": Implications for developmental reprogramming and the epigenetic basis for health and disease. *Prog. Biophys. Mol. Biol.* **118**, 8–13 (2015).
5. Lothrop, A. P., Torres, M. P. & Fuchs, S. M. Deciphering post-translational modification codes. *FEBS Lett.* **587**, 1247–1257 (2013).
6. Biswas, S. & Rao, C. M. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur. J. Pharmacol.* **837**, 8–24 (2018).
7. Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1**, 90 (2011).
8. Fitzpatrick, D. R. & Wilson, C. B. Methylation and demethylation in the regulation of genes, cells, and responses in the immune system. *Clin. Immunol.* **109**, 37–45 (2003).
9. Ferreira, H. J. & Esteller, M. Non-coding RNAs, epigenetics, and cancer: tying it all together. *Cancer Metastasis Rev.* **37**, 55–73 (2018).
10. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* **29**, 288–299 (2007).
11. Llave, C. Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science* **297**, 2053–2056 (2002).

12. Valencia-Sanchez, M. A. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**, 515–524 (2006).

13. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).

14. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).

15. Clapier, C. R. & Cairns, B. R. The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.* **78**, 273–304 (2009).

16. Becker, P. B. & Hörz, W. ATP-Dependent Nucleosome Remodeling. *Annu. Rev. Biochem.* **71**, 247–273 (2002).

17. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **6**, 838–849 (2005).

18. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **48**, 491–507 (2012).

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

20. Yang, X.-J. & Seto, E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* **26**, 5310–5318 (2007).

21. MacDonald, V. E. & Howe, L. J. Histone acetylation: Where to go and how to get there. *Epigenetics* **4**, 139–143 (2009).

22. Allis, C. D. *et al.* New Nomenclature for Chromatin-Modifying Enzymes. *Cell* **131**, 633–636 (2007).

23. Glordano, A. & Avantaggiati, M. p300 and CBP: partners for life and death. *J. Cell. Physiol.* **181**, 218–230 (1999).

24. Kimura, A., Matsubara, K. & Horikoshi, M. A Decade of Histone Acetylation: Marking Eukaryotic Chromosomes with Specific Codes. *J. Biochem. (Tokyo)* **138**, 647–662 (2005).

25. Borrow, J. *et al.* The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat. Genet.* **14**, 33–41 (1996).

26. Kamine, J., Elangovan, B., Subramanian, T., Coleman, D. & Chinnadurai, G. Identification of a Cellular Protein That Specifically Interacts with the Essential Cysteine Region of the HIV-1 Tat Transactivator. *Virology* **216**, 357–366 (1996).

27. Reifsnyder, C., Lowell, J., Clarke, A. & Pillus, L. Yeast SAS silencing genes and human genes associated with AML and HIV–1 Tat interactions are homologous with acetyltransferases. *Nat. Genet.* **14**, 42–49 (1996).

28. Seto, E. & Yoshida, M. Erasers of Histone Acetylation: The Histone Deacetylase Enzymes. *Cold Spring Harb. Perspect. Biol.* **6**, a018713–a018713 (2014).

29. Bird, A. W. *et al.* Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature* **419**, 411–415 (2002).

30. Hasan, S., Hassa, P. O., Imhof, R. & Hottiger, M. O. Transcription coactivator p300 binds PCNA and may have a role in DNA repair synthesis. *Nature* **410**, 387–391 (2001).

31. Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nat. Rev. Mol. Cell Biol.* **18**, 90–101 (2017).

32. Li, L. *et al.* Lysine acetyltransferase 8 is involved in cerebral development and syndromic intellectual disability. *J. Clin. Invest.* **130**, 1431–1445 (2020).

33. Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31**, 27–36 (2010).

34. Riggs, A. D. & Jones, P. A. 5-Methylcytosine, Gene Regulation, and Cancer. in *Advances in Cancer Research* vol. 40 1–30 (Elsevier, 1983).

35. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).

36. Eden, A. Chromosomal Instability and Tumors Promoted by DNA Hypomethylation. *Science* **300**, 455–455 (2003).

37. Howard, G., Eiges, R., Gaudet, F., Jaenisch, R. & Eden, A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* **27**, 404–408 (2008).

38. Wilson, A. S., Power, B. E. & Molloy, P. L. DNA hypomethylation and human diseases. *Biochim. Biophys. Acta BBA - Rev. Cancer* **1775**, 138–162 (2007).

39. Jones, P. A. & Laird, P. W. Cancer-epigenetics comes of age. *Nat. Genet.* **21**, 163–167 (1999).

40. Halkidou, K. *et al.* Upregulation and Nuclear Recruitment of HDAC1 in Hormone Refractory Prostate Cancer. *The Prostate* **59**, 177–189 (2004).

41. Fraga, M. F. *et al.* Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* **37**, 391–400 (2005).

42. Yang, X.-J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* **32**, 959–976 (2004).

43. De Jager, P. L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).

44. Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer's disease. *Nature* **507**, 448–454 (2014).

45. Chi, S. *et al.* Association of Single-Nucleotide Polymorphism in ANK1 with Late-Onset Alzheimer's Disease in Han Chinese. *Mol. Neurobiol.* **53**, 6476–6481 (2016).

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

47. Glajch, K. E. & Sadri-Vakili, G. Epigenetic Mechanisms Involved in Huntington's Disease Pathogenesis. *J. Huntingt. Dis.* **4**, 1–15 (2015).

48. *Neuroepigenomics in Aging and Disease*. vol. 978 (Springer International Publishing, 2017).

49. Hoshino, M. *et al.* Histone deacetylase activity is retained in primary neurons expressing mutant huntingtin protein: HDAC in neurons of Huntington's disease. *J. Neurochem.* **87**, 257–267 (2003).

50. Petrif, F. *et al.* Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376**, 348–351 (1995).

51. Chahrour, M. & Zoghbi, H. Y. The Story of Rett Syndrome: From Clinic to Neurobiology. *Neuron* **56**, 422–437 (2007).

52. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).

53. Lyst, M. J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* **16**, 261–275 (2015).

54. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013).

55. Grant, P. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *GENES&DEVELOPMEN* **11**, 1640–1650 (1997).

56. Smith, S. The SAGA continues...to the end. *Mol. Cell* **35**, 256–258 (2009).

57. Guelman, S. *et al.* Host Cell Factor and an Uncharacterized SANT Domain Protein Are Stable Components of ATAC, a Novel dAda2A/dGcn5-Containing Histone Acetyltransferase Complex in Drosophila. *Mol. Cell. Biol.* **26**, 871–882 (2006).

58. Spedale, G., Timmers, H. T. M. & Pijnappel, W. W. M. P. ATAC-king the complexity of SAGA during evolution. *Genes Dev.* **26**, 527–541 (2012).

59. Lee, K. K., Florens, L., Swanson, S. K., Washburn, M. P. & Workman, J. L. The Deubiquitylation Activity of Ubp8 Is Dependent upon Sgf11 and Its Association with the SAGA Complex. *Mol. Cell. Biol.* **25**, 1173–1182 (2005).

60. Gamper, A. M., Kim, J. & Roeder, R. G. The STAGA Subunit ADA2b Is an Important Regulator of Human GCN5 Catalysis. *Mol. Cell. Biol.* **29**, 266–280 (2009).

61. Nagy, Z. *et al.* The Human SPT20-Containing SAGA Complex Plays a Direct Role in the Regulation of Endoplasmic Reticulum Stress-Induced Genes. *Mol. Cell. Biol.* **29**, 1649–1660 (2009).

62. Suganuma, T. *et al.* The ATAC Acetyltransferase Complex Coordinates MAP Kinases to Regulate JNK Target Genes. *Cell* **142**, 726–736 (2010).

63. Riss, A. *et al.* Subunits of ADA-two-A-containing (ATAC) or Spt-Ada-Gcn5-acetyltrasferase (SAGA) Coactivator Complexes Enhance the Acetyltransferase Activity of GCN5. *J. Biol. Chem.* **290**, 28997–29009 (2015).

64. Wang, T. *et al.* hADA3 is required for p53 activity. *EMBO J.* **20**, 6404–6413 (2001).

65. Bian, C. *et al.* Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation: Sgf29 functions as an H3K4me2/3 binder in SAGA. *EMBO J.* **30**, 2829–2842 (2011).

66. Guarnaccia, A. & Tansey, W. Moonlighting with WDR5: A Cellular Multitasker. *J. Clin. Med.* **7**, 21 (2018).

67. Mi, W. *et al.* The ZZ-type zinc finger of ZZZ3 modulates the ATAC complex-mediated histone acetylation and gene activation. *Nat. Commun.* **9**, 3759 (2018).

68. Fukuyama, K. *et al.* MAPK Upstream Kinase (MUK)-binding Inhibitory Protein, a Negative Regulator of MUK/Dual Leucine Zipper-bearing Kinase/Leucine Zipper Protein Kinase. *J. Biol. Chem.* **275**, 21247–21254 (2000).

69. Guelman, S. *et al.* The Double-Histone-Acetyltransferase Complex ATAC Is Essential for Mammalian Development. *Mol. Cell. Biol.* **29**, 1176–1188 (2009).

70. Wang, Y.-L., Faiola, F., Xu, M., Pan, S. & Martinez, E. Human ATAC Is a GCN5/PCAF-containing Acetylase Complex with a Novel NC2-like Histone Fold Module That Interacts with the TATA-binding Protein. *J. Biol. Chem.* **283**, 33808–33815 (2008).

71. Zhao, D. *et al.* YEATS2 is a selective histone crotonylation reader. *Cell Res.* **26**, 629–632 (2016).

72. Suganuma, T. *et al.* ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat. Struct. Mol. Biol.* **15**, 364–372 (2008).

73. Orpinell, M. *et al.* The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. *EMBO J.* **29**, 2381–2394 (2010).

74. Thomas, M. C. & Chiang, C.-M. The General Transcription Machinery and General Cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–178 (2006).

75. Krebs, A. R. *et al.* ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes. *EMBO Rep.* **11**, 541–547 (2010).

76. Bidon, B. *et al.* XPC is an RNA polymerase II cofactor recruiting ATAC to promoters by interacting with E2F1. *Nat. Commun.* **9**, 2610 (2018).

77. Hofmann, W. A. *et al.* Actin is part of pre-initiation complexes and is necessary for transcription by RNA polymerase II. *Nat. Cell Biol.* **6**, 1094–1101 (2004).

78. Viita, T. *et al.* Nuclear actin interactome analysis links actin to KAT14 histone acetyl transferase and mRNA splicing. *J. Cell Sci.* **132**, jcs226852 (2019).

79. Sokolova, M. *et al.* Nuclear Actin Is Required for Transcription during Drosophila Oogenesis. *iScience* **9**, 63–70 (2018).

80. Xu, W. *et al.* Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. *Nat. Genet.* **26**, 229–232 (2000).

81. Nagy, Z. *et al.* The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. *Cell. Mol. Life Sci.* **67**, 611–628 (2010).

82. Diamond-Blackfan Anemia. (2020).

83. Da Costa, L., Narla, A. & Mohandas, N. An update on the pathogenesis and diagnosis of Diamond–Blackfan anemia. *F1000Research* **7**, 1350 (2018).

84. Bonsi, L. *et al.* M-07e cell bioassay detects stromal cell production of granulocyte-macrophage colony stimulating factor and stem cell factor in normal and in Diamond-Blackfan anemia bone marrow. *STEM CELLS* **11**, 131–134 (1993).

85. Santucci, M. *et al.* Long-term bone marrow cultures in Diamond-Blackfan anemia reveal a defect of both granulomacrophage and erythroid progenitors. *Exp. Hematol.* **27**, 9–18 (1999).

86. on behalf of the DBA group of Société d'Hématologie et d'Immunologie pédiatrique (SHIP) *et al.* Primary hematopoietic cells from DBA patients with mutations in RPL11 and RPS19 genes exhibit distinct erythroid phenotype in vitro. *Cell Death Dis.* **3**, e356–e356 (2012).

87. Dutt, S. *et al.* Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells. *Blood* **117**, 2567–2576 (2011).

88. Doulatov, S. *et al.* Drug discovery for Diamond-Blackfan anemia using reprogrammed hematopoietic progenitors. *Sci. Transl. Med.* **9**, eaah5645 (2017).

89. Heijnen, H. F. *et al.* Ribosomal Protein Mutations Induce Autophagy through S6 Kinase Inhibition of the Insulin Pathway. *PLoS Genet.* **10**, e1004371 (2014).

90. Ge, J. *et al.* Dysregulation of the Transforming Growth Factor β Pathway in Induced Pluripotent Stem Cells Generated from Patients with Diamond Blackfan Anemia. *PLOS ONE* **10**, e0134878 (2015).

91. Shi, Y. & Massague, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685–700 (2003).

92. Ikushima, H. & Miyazono, K. TGF-β signal transduction spreading to a wider field: a broad variety of mechanisms for context-dependent effects of TGF-β. *Cell Tissue Res.* **347**, 37–49 (2012).

93. Bélanger, S. A. & Caron, J. Evaluation of the child with global developmental delay and intellectual disability. *Paediatr. Child Health* **23**, 403–410 (2018).

94. Moeschler, J. B., Shevell, M., & COMMITTEE ON GENETICS. Comprehensive Evaluation of the Child With Intellectual Disability or Global Developmental Delays. *PEDIATRICS* **134**, e903–e918 (2014).

95. Avvakumov, N. & Côté, J. The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene* **26**, 5395–5407 (2007).

96. Thomas, T. & Voss, A. K. The Diverse Biological Roles of MYST Histone Acetyltransferase Family Proteins. *Cell Cycle* **6**, 696–704 (2007).

97. Zhang, X., Wu, J. & Luan, Y. Tip60: Main Functions and Its Inhibitors. *Mini-Rev. Med. Chem.* **17**, 675–682 (2017).

98. Doyon, Y. & Côté, J. The highly conserved and multifunctional NuA4 HAT complex. *Curr. Opin. Genet. Dev.* **14**, 147–154 (2004).

99. Hu, Y. *et al.* Homozygous disruption of the *tip60* gene causes early embryonic lethality. *Dev. Dyn.* **238**, 2912–2921 (2009).

100. Voss, A. K., Collin, C., Dixon, M. P. & Thomas, T. Moz and Retinoic Acid Coordinately Regulate H3K9 Acetylation, Hox Gene Expression, and Segment Identity. *Dev. Cell* **17**, 674–686 (2009).

101. Urban, I. *et al.* TIP60/KAT5 is required for neuronal viability in hippocampal CA1. *Sci. Rep.* **9**, 16173 (2019).

102. Fink, D. *et al.* Loss of Ing3 Expression Results in Growth Retardation and Embryonic Death. *Cancers* **12**, 80 (2019).

103. Ullah, M. *et al.* Molecular Architecture of Quartet MOZ/MORF Histone Acetyltransferase Complexes. *Mol. Cell. Biol.* **28**, 6828–6843 (2008).

104. Yang, X.-J. & Ullah, M. MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene* **26**, 5408–5419 (2007).

105. Lalonde, M.-E. *et al.* Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes Dev.* **27**, 2009–2024 (2013).

106. Qin, S. *et al.* Recognition of Unmodified Histone H3 by the First PHD Finger of Bromodomain-PHD Finger Protein 2 Provides Insights into the Regulation of Histone

Acetyltransferases Monocytic Leukemic Zinc-finger Protein (MOZ) and MOZ-related factor (MORF). *J. Biol. Chem.* **286**, 36944–36955 (2011).

107. Wu, H. *et al.* Structural and Histone Binding Ability Characterizations of Human PWWP Domains. *PLoS ONE* **6**, e18919 (2011).

108. Yan, K. *et al.* Mutations in the Chromatin Regulator Gene BRPF1 Cause Syndromic Intellectual Disability and Deficient Histone Acetylation. *Am. J. Hum. Genet.* **100**, 91–104 (2017).

109. Campos, E. I., Chin, M. Y., Kuo, W. H. & Li, G. Biological functions of the ING family tumor suppressors. *Cell. Mol. Life Sci.* **61**, 2597–2613 (2004).

110. Wang, F. *et al.* ING5 activity in self-renewal of glioblastoma stem cells via calcium and follicle stimulating hormone pathways. *Oncogene* **37**, 286–301 (2018).

111. Mulder, K. W. *et al.* Diverse epigenetic strategies interact to control epidermal differentiation. *Nat. Cell Biol.* **14**, 753–763 (2012).

112. Katsumoto, T. MOZ is essential for maintenance of hematopoietic stem cells. *Genes Dev.* **20**, 1321–1330 (2006).

113. Thomas, T. Monocytic leukemia zinc finger protein is essential for the development of long-term reconstituting hematopoietic stem cells. *Genes Dev.* **20**, 1175–1186 (2006).

114. Perez-Campo, F. M. *et al.* MOZ-Mediated Repression of $p16^{INK4a}$ Is Critical for the Self-Renewal of Neural and Hematopoietic Stem Cells: MOZ: Key Regulator of Proliferation versus Senescence. *STEM CELLS* **32**, 1591–1601 (2014).

115. Merson, T. D. *et al.* The Transcriptional Coactivator Querkopf Controls Adult Neurogenesis. *J. Neurosci.* **26**, 11359–11370 (2006).

116. Sheikh, B. N., Dixon, M. P., Thomas, T. & Voss, A. K. . Querkopf Is a Key Marker of Self-Renewal and Multipotency of Adult Neural Stem Cells. *Development* **139**, (2012).

117. Thomas, T. & Voss, A. K. QUERKOPF, A HISTONE ACETYLTRANSFERASE, IS ESSENTIAL FOR EMBRYONIC NEUROGENESIS. *Front. Biosci.* **9**, 24–31 (2004).

118. Thomas, T., Voss, A. K., Chowdhury, K. & Gruss, P. Querkopf, a MYST family histone acetyltransferase, is required for normal cerebral cortex development. *Development* **127**, 2537–2548 (2000).

119. Panagopoulos, I. Fusion of the MORF and CBP genes in acute myeloid leukemia with the t(10;16)(q22;p13). *Hum. Mol. Genet.* **10**, 395–404 (2001).

120. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45**, 478–486 (2013).

121. Moore, S. D. P. *et al.* Uterine Leiomyomata with t(10;17) Disrupt the Histone Acetyltransferase ***MORF*** . *Cancer Res.* **64**, 5570–5577 (2004).

122. You, L., Chen, L., Penney, J., Miao, D. & Yang, X.-J. Expression atlas of the multivalent epigenetic regulator Brpf1 and its requirement for survival of mouse embryos. *Epigenetics* **9**, 860–872 (2014).

123. You, L. *et al.* The Lysine Acetyltransferase Activator Brpf1 Governs Dentate Gyrus Development through Neural Stem Cells and Progenitors. *PLOS Genet.* **11**, e1005034 (2015).

124. You, L. *et al.* Deficiency of the Chromatin Regulator Brpf1 Causes Abnormal Brain Development. *J. Biol. Chem.* **290**, 7114–7129 (2015).

125. Yan, K. *et al.* Deficient histone H3 propionylation by BRPF1-KAT6 complexes in neurodevelopmental disorders and cancer. *Sci. Adv.* **6**, eaax0021 (2020).

126. You, L. *et al.* BRPF1 is essential for development of fetal hematopoietic stem cells. *J. Clin. Invest.* **126**, 3247–3262 (2016).

127. Chen, X., Liu, G. & Leffak, M. Activation of a human chromosomal replication origin by protein tethering. *Nucleic Acids Res.* **41**, 6460–6474 (2013).

128. Chitalia, V. C. *et al.* Jade-1 inhibits Wnt signalling by ubiquitylating β-catenin and mediates Wnt pathway inhibition by pVHL. *Nat. Cell Biol.* **10**, 1208–1216 (2008).

129. Hung, T. *et al.* ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation. *Mol. Cell* **33**, 248–256 (2009).

130. Han, J. *et al.* The scaffolding protein JADE1 physically links the acetyltransferase subunit HBO1 with its histone H3–H4 substrate. *J. Biol. Chem.* **293**, 4498–4509 (2018).

131. Calvi, B. R. HBO1: JADE1 at the cell cycle chromatin crossroads. *Cell Cycle* **13**, 2322–2322 (2014).

132. Lan, R. & Wang, Q. Deciphering structure, function and mechanism of lysine acetyltransferase HBO1 in protein acetylation, transcription regulation, DNA replication and its oncogenic properties in cancer. *Cell Mol. Life Sci.* **77**, 637–649 (2020).

133. Mishima, Y. *et al.* The Hbo1-Brd1/Brpf2 complex is responsible for global acetylation of H3K14 and required for fetal liver erythropoiesis. *Blood* **118**, 2443–2453 (2011).

134. Yan, K. *et al.* The Chromatin Regulator BRPF3 Preferentially Activates the HBO1 Acetyltransferase but Is Dispensable for Mouse Development and Survival. *J. Biol. Chem.* **291**, 2647–2663 (2016).

135. Newman, D. M., Voss, A. K., Thomas, T. & Allen, R. S. Essential role for the histone acetyltransferase KAT7 in T cell development, fitness, and survival. *J. Leukoc. Biol.* **101**, 887–892 (2017).

136. Miotto, B. & Struhl, K. HBO1 Histone Acetylase Activity Is Essential for DNA Replication Licensing and Inhibited by Geminin. *Mol. Cell* **37**, 57–66 (2010).

137. Kueh, A. J., Dixon, M. P., Voss, A. K. & Thomas, T. HBO1 Is Required for H3K14 Acetylation and Normal Transcriptional Activity during Embryonic Development. *Mol. Cell. Biol.* **31**, 845–860 (2011).

138. Severinsen, J. E. *et al.* Evidence implicating BRD1 with brain development and susceptibility to both schizophrenia and bipolar affective disorder. *Mol. Psychiatry* **11**, 1126–1138 (2006).

139. Qvist, P. *et al.* Brain volumetric alterations accompanied with loss of striatal medium-sized spiny neurons and cortical parvalbumin expressing interneurons in Brd1+/− mice. *Sci. Rep.* **8**, 16486 (2018).

140. Qvist, P. *et al.* The Schizophrenia-Associated BRD1 Gene Regulates Behavior, Neurotransmission, and Expression of Schizophrenia Risk Enriched Gene Sets in Mice. *Biol. Psychiatry* **82**, 62–76 (2017).

141. Tzouanacou, E., Tweedie, S. & Wilson, V. Identification of Jade1, a Gene Encoding a PHD Zinc Finger Protein, in a Gene Trap Mutagenesis Screen for Genes Involved in Anteroposterior Axis Development. *Mol. Cell. Biol.* **23**, 8553–8552 (2003).

142. Gelbart, M. E. & Kuroda, M. I. Drosophila dosage compensation: a complex voyage to the X chromosome. *Development* **136**, 1399–1410 (2009).

143. Straub, T. & Becker, P. B. Dosage compensation: the beginning and end of generalization. *Nat. Rev. Genet.* **8**, 47–57 (2007).

144. Laverty, C., Lucci, J. & Akhtar, A. The MSL complex: X chromosome and beyond. *Curr. Opin. Genet. Dev.* **20**, 171–178 (2010).

145. Kelley, R. L., Lee, O.-K. & Shim, Y.-K. Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex. *Mech. Dev.* **125**, 1009–1019 (2008).

146. Smith, E. R. *et al.* A Human Protein Complex Homologous to the Drosophila MSL Complex Is Responsible for the Majority of Histone H4 Acetylation at Lysine 16. *MOL CELL BIOL* **25**, 15 (2005).

147. Lam, K. C. *et al.* The NSL Complex Regulates Housekeeping Genes in Drosophila. *PLoS Genet.* **8**, e1002736 (2012).

148. Dias, J. *et al.* Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. *Genes Dev.* **28**, 929–942 (2014).

149. Radzisheuskaya, A. *et al.* Complex-dependent histone acetyltransferase activity of KAT8 determines its role in transcription and cellular homeostasis. *Mol. Cell* **81**, 1749-1765.e8 (2021).

150. Raja, S. J. *et al.* The Nonspecific Lethal Complex Is a Transcriptional Regulator in Drosophila. *Mol. Cell* **38**, 827–841 (2010).

151. Feller, C. *et al.* The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. *Nucleic Acids Res.* **40**, 1509–1522 (2012).

152. Sheikh, B. N., Guhathakurta, S. & Akhtar, A. The non-specific lethal (NSL) complex at the crossroads of transcriptional control and cellular homeostasis. *EMBO Rep.* **20**, (2019).

153. Zhao, X. *et al.* Crosstalk between NSL Histone Acetyltransferase and MLL/SET Complexes: NSL Complex Functions in Promoting Histone H3K4 Di-Methylation Activity by MLL/SET Complexes. *PLoS Genet.* **9**, e1003940 (2013).

154. Odho, Z., Southall, S. M. & Wilson, J. R. Characterization of a Novel WDR5-binding Site That Recruits RbBP5 through a Conserved Motif to Enhance Methylation of Histone H3 Lysine 4 by Mixed Lineage Leukemia Protein-1. *J. Biol. Chem.* **285**, 32967–32976 (2010).

155. Mendjan, S. *et al.* Nuclear Pore Components Are Involved in the Transcriptional Regulation of Dosage Compensation in Drosophila. *Mol. Cell* **21**, 811–823 (2006).

156. Ferreyra Solari, N. E. *et al.* The NSL Chromatin-Modifying Complex Subunit KANSL2 Regulates Cancer Stem–like Properties in Glioblastoma That Contribute to Tumorigenesis. *Cancer Res.* **76**, 5383–5394 (2016).

157. Chatterjee, A. *et al.* MOF Acetyl Transferase Regulates Transcription and Respiration in Mitochondria. *Cell* **167**, 722-738.e23 (2016).

158. Wu, D. *et al. O* -Linked *N* -acetylglucosamine transferase 1 regulates global histone H4 acetylation via stabilization of the nonspecific lethal protein NSL3. *J. Biol. Chem.* **292**, 10014–10025 (2017).

159. Wells, L. & Hart, G. W. O-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar. *FEBS Lett.* **546**, 154–158 (2003).

160. Gambetta, M. C. & Müller, J. A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. *Chromosoma* **124**, 429–442 (2015).

161. Zargar, Z. U. & Tyagi, S. Role of Host Cell Factor-1 in cell cycle regulation. *Transcription* **3**, 187–192 (2012).

162. Lee, S.-H. *et al.* MCRS1 associates with cytoplasmic dynein and mediates pericentrosomal material recruitment. *Sci. Rep.* **6**, 27284 (2016).

163. Su, J., Wang, F., Cai, Y. & Jin, J. The Functional Analysis of Histone Acetyltransferase MOF in Tumorigenesis. *Int. J. Mol. Sci.* **17**, 99 (2016).

164. Klein, B. J. *et al.* PHF20 Readers Link Methylation of Histone H3K4 and p53 with H4K16 Acetylation. *Cell Rep.* **17**, 1158–1170 (2016).

165. Cui, G. *et al.* PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nat. Struct. Mol. Biol.* **19**, 916–924 (2012).

166. Thomas, T., Dixon, M. P., Kueh, A. J. & Voss, A. K. Mof (MYST1 or KAT8) Is Essential for Progression of Embryonic Development Past the Blastocyst Stage and Required for Normal Chromatin Architecture. *Mol. Cell. Biol.* **28**, 5093–5105 (2008).

167. Sharma, G. G. *et al.* MOF and Histone H4 Acetylation at Lysine 16 Are Critical for DNA Damage Response and Double-Strand Break Repair. *Mol. Cell. Biol.* **30**, 3582–3595 (2010).

168. Kumar, R. *et al.* Purkinje cell-specific males absent on the first (*mMof*) gene deletion results in an ataxia-telangiectasia-like neurological phenotype and backward walking in mice. *Proc. Natl. Acad. Sci.* **108**, 3636–3641 (2011).

169. Arbogast, T. *et al.* Mouse models of 17q21.31 microdeletion and microduplication syndromes highlight the importance of Kansl1 for cognition. *PLOS Genet.* **13**, e1006886 (2017).

170. Badeaux, A. I. *et al.* Loss of the Methyl Lysine Effector Protein PHF20 Impacts the Expression of Genes Regulated by the Lysine Acetyltransferase MOF. *J. Biol. Chem.* **287**, 429–437 (2012).

171. O'Donnell, N., Zachara, N. E., Hart, G. W. & Marth, J. D. Ogt-Dependent X-Chromosome-Linked Protein Glycosylation Is a Requisite Modification in Somatic Cell Function and Embryo Viability. *Mol. Cell. Biol.* **24**, 1680–1690 (2004).

172. Medford, H. M. & Marsh, S. A. The role of O-GlcNAc transferase in regulating the gene transcription of developing and failing hearts. *Future Cardiol.* **10**, 801–812 (2014).

173. Jang, H. *et al.* O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network. *Cell Stem Cell* **11**, 62–74 (2012).

174. Minocha, S. *et al.* Epiblast-specific loss of HCF-1 leads to failure in anterior-posterior axis specification. *Dev. Biol.* **418**, 75–88 (2016).

175. Panikker, P. *et al.* Restoring Tip60 HAT/HDAC2 Balance in the Neurodegenerative Brain Relieves Epigenetic Transcriptional Repression and Reinstates Cognition. *J. Neurosci.* **38**, 4569–4583 (2018).

176. Humbert, J. *et al.* De Novo KAT5 Variants Cause a Syndrome with Recognizable Facial Dysmorphisms, Cerebellar Atrophy, Sleep Disturbance, and Epilepsy. *Am. J. Hum. Genet.* **107**, 564–574 (2020).

177. Cogné, B. *et al.* Missense Variants in the Histone Acetyltransferase Complex Component Gene TRRAP Cause Autism and Syndromic Intellectual Disability. *Am. J. Hum. Genet.* **104**, 530–541 (2019).

178. Arboleda, V. A. *et al.* De Novo Nonsense Mutations in KAT6A, a Lysine Acetyl-Transferase Gene, Cause a Syndrome Including Microcephaly and Global Developmental Delay. *Am. J. Hum. Genet.* **96**, 498–506 (2015).

179. Tham, E. *et al.* Dominant Mutations in KAT6A Cause Intellectual Disability with Recognizable Syndromic Features. *Am. J. Hum. Genet.* **96**, 507–513 (2015).

180. The DDD Study, *et al.* KAT6A Syndrome: genotype–phenotype correlation in 76 patients with pathogenic KAT6A variants. *Genet. Med.* **21**, 850–860 (2019).

181. Millan, F. *et al.* Whole exome sequencing reveals de novo pathogenic variants in KAT6A as a cause of a neurodevelopmental disorder. *Am. J. Med. Genet. A.* **170**, 1791–1798 (2016).

182. Gauthier-Vasserot, A. *et al.* Application of whole-exome sequencing to unravel the molecular basis of undiagnosed syndromic congenital neutropenia with intellectual disability. *Am. J. Med. Genet. A.* **173**, 62–71 (2017).

183. Satoh, C. *et al.* Three brothers with a nonsense mutation in KAT6A caused by parental germline mosaicism. *Hum. Genome Var.* **4**, 17045 (2017).

184. Elenius, V., Lähdesmäki, T., Hietala, M. & Jartti, T. Food allergy in a child with de novo KAT6A mutation. *Clin. Transl. Allergy* **7**, 19 (2017).

185. Murray, C. *et al.* Novel Causative Variants in DYRK1A, KARS, and KAT6A Associated with Intellectual Disability and Additional Phenotypic Features. *J. Pediatr. Genet.* **06**, 077–083 (2017).

186. Zwaveling-Soonawala, N. *et al.* Variants in *KAT6A* and pituitary anomalies. *Am. J. Med. Genet. A.* **173**, 2562–2565 (2017).

187. Trinh, J. *et al.* A KAT6A variant in a family with autosomal dominantly inherited microcephaly and developmental delay. *J. Hum. Genet.* **63**, 997–1001 (2018).

188. Efthymiou, S., Salpietro, V., Bettencourt, C. & Houlden, H. Paroxysmal Movement Disorder and Epilepsy Caused by a De Novo Truncating Mutation in KAT6A. *J. Pediatr. Genet.* **07**, 114–116 (2018).

189. Alkhateeb, A. & Alazaizeh, W. A Novel De Novo Frameshift Mutation in KAT6A Identified by Whole Exome Sequencing. *J. Pediatr. Genet.* **08**, 010–014 (2019).

190. Urreizti, R. *et al.* Five new cases of syndromic intellectual disability due to KAT6A mutations: widening the molecular and clinical spectrum. *Orphanet J. Rare Dis.* **15**, 44 (2020).

191. Cormier-Daire, V. *et al.* Genitopatellar syndrome: a new condition comprising absent patellae, scrotal hypoplasia, renal anomalies, facial dysmorphism, and mental retardation. 5.

192. Campeau, P. M. *et al.* Mutations in KAT6B, Encoding a Histone Acetyltransferase, Cause Genitopatellar Syndrome. *Am. J. Hum. Genet.* **90**, 282–289 (2012).

193. Simpson, M. A. *et al.* De Novo Mutations of the Gene Encoding the Histone Acetyltransferase KAT6B Cause Genitopatellar Syndrome. *Am. J. Hum. Genet.* **90**, 290–294 (2012).

194. Radvanszky, J. *et al.* Complex phenotypes blur conventional borders between Say-Barber-Biesecker-Young-Simpson syndrome and genitopatellar syndrome: Complex phenotypes blur conventional borders. *Clin. Genet.* **91**, 339–343 (2017).

195. Armstrong, L. Report of a new case of ‘genitopatellar’ syndrome which challenges the importance of absent patellae as a defining feature. *J. Med. Genet.* **39**, 933–934 (2002).

196. Bergmann, C., Spranger, S., Javaher, P. & Ptok, M. Genitopatellar syndrome, sensorineural hearing loss, and cleft palate. *Oral Maxillofac. Surg.* **15**, 103–106 (2011).

197. Brugha, R., Kinali, M., Aminu, K., Bridges, N. & Holder, S. Genitopatellar syndrome: a further case. *Clin. Dysmorphol.* **20**, 163–165 (2011).

198. Goldblatt, J., Wallis, C. & Zieff, S. A Syndrome of Hypoplastic Patellae, Mental Retardation, Skeletal and Genitourinary Anomalies with Normal Chromosomes. *Clin. Dysmorphol.* **2**, 91–93 (1988).

199. Penttinen, M., Koillinen, H., Niinikoski, H., Mäkitie, O. & Hietala, M. Genitopatellar syndrome in an adolescent female with severe osteoporosis and endocrine abnormalities. *Am. J. Med. Genet. A.* **149A**, 451–455 (2009).

200. Reardon, W. Genitopatellar Syndrome: A Recognizable Phenotype. *Am. J. Med. Genet.* **111**, 313–315 (2002).

201. Kim, B. R. *et al.* Genitopatellar Syndrome Secondary to *De Novo KAT6B* Mutation: The First Genetically Confirmed Case in South Korea. *Yonsei Med. J.* **60**, 395 (2019).

202. DDD study *et al.* Further delineation of the KAT6B molecular and phenotypic spectrum. *Eur. J. Hum. Genet.* **23**, 1165–1170 (2015).

203. Okano, S. *et al.* Genitopatellar syndrome: the first reported case in Japan. *Hum. Genome Var.* **5**, 8 (2018).

204. Hamaguchi, Y. *et al.* KAT6B-related disorder in a patient with a novel frameshift variant (c.3925dup). *Hum. Genome Var.* **6**, 54 (2019).

205. Takahashi, K. *et al.* A Japanese Patient with Genitopatellar Syndrome Transiently Presenting with Cardiac Intramural Cavity during the Neonatal Period. *Case Rep. Genet.* **2020**, 1–5 (2020).

206. Zhang, L. X. *et al.* Further delineation of the clinical spectrum of KAT6B disorders and allelic series of pathogenic variants. *Genet. Med.* **22**, 1338–1347 (2020).

207. Ferrando Meseguer, E. *et al.* A further patient with genitopatellar syndrome requiring multidisciplinary management. *Clin. Dysmorphol.* **29**, 193–196 (2020).

208. Knight, S. *et al.* Genitopatellar syndrome and neuroblastoma: The multidisciplinary management of a previously unreported association. *Pediatr. Blood Cancer* **65**, e27373 (2018).

209. Young, I. D. & Simpson, K. Unknown syndrome: abnormal facies, congenital heart defects, hypothyroidism, and severe retardation. *J. Med. Genet.* **24**, 715–716 (1987).

210. Say, B. & Barber, N. Mental retardation with blepharophimosis. *J. Med. Genet.* **24**, 511–511 (1987).

211. Biesecker, L. G. The Ohdo blepharophimosis syndrome: a third case. *J. Med. Genet.* **28**, 131–134 (1991).

212. Day, R. *et al.* A clinical and genetic study of the Say/Barber/Biesecker/Young-Simpson type of Ohdo syndrome. *Clin. Genet.* **74**, 434–444 (2008).

213. Vlckova, M. *et al.* A patient showing features of both SBBYSS and GPS supports the concept of a KAT6B-related disease spectrum, with mutations in mid-exon 18 possibly leading to combined phenotypes. *Eur. J. Med. Genet.* **58**, 550–555 (2015).

214. Fernández-Mayoralas, D. M., Calleja-Pérez, B., Álvarez, S. & Fernández-Jaén, A. Mutación de novo en KAT6B, síndrome Say-Barber-Biesecker-Young-Simpson y trastorno específico del lenguaje. *Neurología* S0213485319300854 (2019) doi:10.1016/j.nrl.2019.05.006.

215. Brea-Fernández, A., Dacruz, D., Eirís, J., Barros, F. & Carracedo, Á. Novel truncating variants expand the phenotypic spectrum of *KAT6B* -related disorders. *Am. J. Med. Genet. A.* **179**, 290–294 (2019).

216. Merritt, J., Hart, J. C. & LeGrow, T. L. Autism spectrum disorder in Say-Barber-Biesecker-Young-Simpson syndrome. *BMJ Case Rep.* bcr-2017-219930 (2017) doi:10.1136/bcr-2017-219930.

217. Yilmaz, R. *et al.* A recurrent synonymous *KAT6B* mutation causes Say-Barber-Biesecker/Young-Simpson syndrome by inducing aberrant splicing. *Am. J. Med. Genet. A.* **167**, 3006–3010 (2015).

218. Galletti, B. *et al.* Chronic Otitis Media Associated with Cholesteatoma in a Case of the Say-Barber-Biesecker-Young-Simpson Variant of Ohdo Syndrome. *Am. J. Case Rep.* **20**, 175–178 (2019).

219. Lyu, N., Shang, Q., Li, J., Ma, C. & Li, D. [A Case of SBBYSS Syndrome Caused by KAT6B Gene Variant]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **36**, 727–730 (2019).

220. Kim, Y. R. *et al.* Identifying the *KAT6B* Mutation via Diagnostic Exome Sequencing to Diagnose Say-Barber-Biesecker-Young-Simpson Syndrome in Three Generations of a Family. *Ann. Rehabil. Med.* **41**, 505 (2017).

221. Lundsgaard, M. *et al.* De novo ***KAT6B*** Mutation Identified with Whole-Exome Sequencing in a Girl with Say-Barber/Biesecker/Young-Simpson Syndrome. *Mol. Syndromol.* **8**, 24–29 (2017).

222. Yu, H.-C., Geiger, E. A., Medne, L., Zackai, E. H. & Shaikh, T. H. An individual with blepharophimosis-ptosis-epicanthus inversus syndrome (BPES) and additional features expands the phenotype associated with mutations in *KAT6B*. *Am. J. Med. Genet. A.* **164**, 950–957 (2014).

223. Szakszon, K. *et al.* De novo mutations of the gene encoding the histone acetyltransferase KAT6B in two patients with Say-Barber/Biesecker/Young-Simpson syndrome. *Am. J. Med. Genet. A.* **161**, 884–888 (2013).

224. Mendez, R., Delea, M., Dain, L. & Rittler, M. A novel pathogenic frameshift variant of KAT6B identified by clinical exome sequencing in a newborn with the Say–Barber–Biesecker–Young–Simpson syndrome. *Clin. Dysmorphol.* **29**, 42–45 (2020).

225. Yates, T. M. *et al.* Novel *KAT6B* proximal familial variant expands genotypic and phenotypic spectrum. *Clin. Genet.* **95**, 334–335 (2019).

226. Li, G. *et al. De Novo* Mutation of *KAT6B* Gene Causing Atypical Say–Barber–Biesecker–Young–Simpson Syndrome or Genitopatellar Syndrome. *Fetal Pediatr. Pathol.* **36**, 130–138 (2017).

227. Zhu, L., Lv, L., Wu, D. & Shao, J. KAT6B Genetic Variant Identified in a Short Stature Chinese Infant: A Report of Physical Growth in Clinical Spectrum of KAT6B-Related Disorders. *Front. Pediatr.* **8**, 124 (2020).

228. Clayton-Smith, J. *et al.* Whole-Exome-Sequencing Identifies Mutations in Histone Acetyltransferase Gene KAT6B in Individuals with the Say-Barber-Biesecker Variant of Ohdo Syndrome. *Am. J. Hum. Genet.* **89**, 675–681 (2011).

229. Preiksaitiene, E. *et al.* Features of *KAT6B* -related disorders in a patient with 10q22.1q22.3 deletion. *Ophthalmic Genet.* **38**, 383–386 (2017).

230. Lonardo, F. *et al.* Say-Barber-Biesecker-Young-Simpson syndrome and Genitopatellar syndrome: Lumping or splitting?: LONARDO ET AL. *Clin. Genet.* **95**, 253–261 (2019).

231. Marangi, G. *et al.* A novel truncating variant within exon 7 of *KAT6B* associated with features of both Say–Barber–Bieseker–Young–Simpson syndrome and genitopatellar

syndrome: Further evidence of a continuum in the clinical spectrum of *KAT6B* -related disorders. *Am. J. Med. Genet. A.* **176**, 455–459 (2018).

232. Niida, Y. *et al.* A Say-Barber-Biesecker-Young-Simpson variant of Ohdo syndrome with a *KAT6B* 10-base pair palindromic duplication: A recurrent mutation causing a severe phenotype mixed with genitopatellar syndrome: SBBYSS and GPS caused by *KAT6B* mutation. *Congenit. Anom.* **57**, 86–88 (2017).

233. Bashir, R. A. *et al.* Lin-Gettig syndrome: Craniosynostosis expands the spectrum of the *KAT6B* related disorders. *Am. J. Med. Genet. A.* **173**, 2596–2604 (2017).

234. Mattioli, F. *et al.* Mutations in Histone Acetylase Modifier BRPF1 Cause an Autosomal-Dominant Form of Intellectual Disability with Associated Ptosis. *Am. J. Hum. Genet.* **100**, 105–116 (2017).

235. Pode-Shakked, N. *et al. BRPF1* -associated intellectual disability, ptosis, and facial dysmorphism in a multiplex family. *Mol. Genet. Genomic Med.* e665 (2019) doi:10.1002/mgg3.665.

236. Demeulenaere, S. *et al.* Novel BRPF1 mutation in a boy with intellectual disability, coloboma, facial nerve palsy and hypoplasia of the corpus callosum. *Eur. J. Med. Genet.* **62**, 103691 (2019).

237. Naseer, M. I. *et al.* Novel Missense Variant in Heterozygous State in the BRPF1 Gene Leading to Intellectual Developmental Disorder With Dysmorphic Facies and Ptosis. *Front. Genet.* **11**, 368 (2020).

238. Keywan, C. *et al.* A de novo BRPF1 variant in a case of Sudden Unexplained Death in Childhood. *Eur. J. Med. Genet.* **63**, 104002 (2020).

239. MacPherson, L. *et al.* HBO1 is required for the maintenance of leukaemia stem cells. *Nature* **577**, 266–270 (2020).

240. Gao, S., Qi, X., Li, J. & Sang, L. Upregulated KAT7 in synovial fibroblasts promotes Th17 cell differentiation and infiltration in rheumatoid arthritis. *Biochem. Biophys. Res. Commun.* **489**, 235–241 (2017).

241. Mayo Clinic Staff. Schizophrenia. (2020).

242. Mayo Clinic Staff. Bipolar disorder. (2018).

243. Nyegaard, M. *et al.* Support of association between *BRD1* and both schizophrenia and bipolar affective disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **153B**, 582–591 (2010).

244. Aberg, K. A. *et al.* A Comprehensive Family-Based Replication Study of Schizophrenia Genes. *JAMA Psychiatry* **70**, 573 (2013).

245. Koolen, D. A. *et al.* Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat. Genet.* **44**, 639–641 (2012).

246. Koolen, D. A., Morgan, A. & de Vries, B. Koolen-de Vries Syndrome. 22.

247. Sauvestre, F. *et al.* Early fetal presentation of Koolen-de Vries: Case report with literature review. *Eur. J. Med. Genet.* **60**, 605–609 (2017).

248. Koolen, D. A. *et al.* Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J. Med. Genet.* **45**, 710–720 (2008).

249. Terrone, G. *et al.* A further contribution to the delineation of the 17q21.31 microdeletion syndrome: Central nervous involvement in two Italian patients. *Eur. J. Med. Genet.* **55**, 466–471 (2012).

250. Moreno-Igoa, M. *et al.* KANSL1 gene disruption associated with the full clinical spectrum of 17q21.31 microdeletion syndrome. *BMC Med. Genet.* **16**, 68 (2015).

251. DDD Study *et al.* The Koolen-de Vries syndrome: a phenotypic comparison of patients with a 17q21.31 microdeletion versus a KANSL1 sequence variant. *Eur. J. Hum. Genet.* **24**, 652–659 (2016).

252. Barone, C. *et al.* An additional clinical sign of 17q21.31 microdeletion syndrome: Preaxial polydactyly of hands with broad thumbs. *Am. J. Med. Genet. A.* **167**, 1671–1673 (2015).

253. Kitsiou-Tzeli, S. *et al.* Microdeletion and microduplication 17q21.31 plus an additional CNV, in patients with intellectual disability, identified by array-CGH. *Gene* **492**, 319–324 (2012).

254. Jaillard, S. *et al.* Identification of gene copy number variations in patients with mental retardation using array-CGH: Novel syndromes in a large French series. *Eur. J. Med. Genet.* **53**, 66–75 (2010).

255. Varela, M. C. *et al.* A 17q21.31 microdeletion encompassing the *MAPT* gene in a mentally impaired patient. *Cytogenet. Genome Res.* **114**, 89–92 (2006).

256. Wright, E. B., Donnai, D., Johnson, D. & Clayton-Smith, J. Cutaneous features in 17q21.31 deletion syndrome: a differential diagnosis for cardio–facio–cutaneous syndrome. *Clin. Dysmorphol.* **20**, 15–20 (2011).

257. Wray, C. D. 17q21.31 Microdeletion associated with infantile spasms. *Eur. J. Med. Genet.* **56**, 59–61 (2013).

258. Vlckova, M. *et al.* Monozygotic Twins with 17q21.31 Microdeletion Syndrome. *Twin Res. Hum. Genet.* **17**, 405–410 (2014).

259. Tan, T. Y. *et al.* Phenotypic expansion and further characterisation of the 17q21.31 microdeletion syndrome. *J. Med. Genet.* **46**, 480–489 (2009).

260. Sharkey, F. H. *et al.* 17q21.31 Microdeletion Syndrome: Further Expanding the Clinical Phenotype. *Cytogenet. Genome Res.* **127**, 61–6 (2008).

261. El Chehadeh-Djebbar, S. *et al.* 17q21.31 microdeletion in a patient with pituitary stalk interruption syndrome. *Eur. J. Med. Genet.* **54**, 369–373 (2011).

262. Dubourg, C. *et al.* Clinical and molecular characterization of 17q21.31 microdeletion syndrome in 14 French patients with mental retardation. *Eur. J. Med. Genet.* **54**, 144–151 (2011).

263. Digilio, M. C. *et al.* Hypopigmented skin patches in 17q21.31 microdeletion syndrome: expanding the spectrum of cutaneous findings. *Clin. Dysmorphol.* **23**, 32–34 (2014).

264. Shaw-Smith, C. *et al.* Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat. Genet.* **38**, 1032–1037 (2006).

265. Koolen, D. A. *et al.* A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat. Genet.* **38**, 999–1001 (2006).

266. Wagenstaller, J. *et al.* Copy-Number Variations Measured by Single-Nucleotide–Polymorphism Oligonucleotide Arrays in Patients with Mental Retardation. *Am. J. Hum. Genet.* **81**, 768–779 (2007).

267. Sharp, A. J. *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat. Genet.* **38**, 1038–1042 (2006).

268. Hoyer, J. *et al.* Molecular karyotyping in patients with mental retardation using 100K single-nucleotide polymorphism arrays. *J. Med. Genet.* **44**, 629–636 (2007).

269. Aradhya, S., Manning, M. A., Splendore, A. & Cherry, A. M. Whole-genome array-CGH identifies novel contiguous gene deletions and duplications associated with developmental delay, mental retardation, and dysmorphic features. *Am. J. Med. Genet. A.* **143A**, 1431–1441 (2007).

270. Zollino, M. *et al.* Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat. Genet.* **44**, 636–638 (2012).

271. Egloff, M. *et al.* 17q21.31 Microdeletion: Brain Anomalies Leading to Prenatal Diagnosis. *Cytogenet. Genome Res.* **144**, 178–182 (2014).

272. Vaidyanathan, K. *et al.* Identification and characterization of a missense mutation in the *O* -linked β- *N* -acetylglucosamine (*O* -GlcNAc) transferase gene that segregates with X-linked intellectual disability. *J. Biol. Chem.* **292**, 8948–8963 (2017).

273. Selvan, N. *et al. O* -GlcNAc transferase missense mutations linked to X-linked intellectual disability deregulate genes involved in cell fate determination and signaling. *J. Biol. Chem.* **293**, 10810–10824 (2018).

274. Bouazzi, H., Lesca, G., Trujillo, C., Alwasiyah, M. K. & Munnich, A. Nonsyndromic X-linked intellectual deficiency in three brothers with a novel *MED12* missense mutation [c.5922G>T (p.Glu1974His)]. *Clin. Case Rep.* **3**, 604–609 (2015).

275. Pravata, V. M. *et al.* Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability. *Proc. Natl. Acad. Sci.* **116**, 14961–14970 (2019).

276. Willems, A. P. *et al.* Mutations in *N* -acetylglucosamine (*O* -GlcNAc) transferase in patients with X-linked intellectual disability. *J. Biol. Chem.* **292**, 12621–12631 (2017).

277. Huang, L. *et al.* A Noncoding, Regulatory Mutation Implicates HCFC1 in Nonsyndromic Intellectual Disability. *Am. J. Hum. Genet.* **91**, 694–702 (2012).

278. Jolly, L. A. *et al.* HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain. *Hum. Mol. Genet.* **24**, 3335–3347 (2015).

279. Yu, H.-C. *et al.* An X-Linked Cobalamin Disorder Caused by Mutations in Transcriptional Coregulator HCFC1. *Am. J. Hum. Genet.* **93**, 506–514 (2013).

280. Gérard, M. *et al.* Multiple congenital anomalies in two boys with mutation in HCFC1 and cobalamin disorder. *Eur. J. Med. Genet.* **58**, 148–153 (2015).

281. Couture, J.-F., Collazo, E. & Trievel, R. C. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat. Struct. Mol. Biol.* **13**, 698–703 (2006).

282. Thomas, L. R. *et al.* Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. *Mol. Cell* **58**, 440–452 (2015).

283. Bailey, J. K. *et al.* WD Repeat-containing Protein 5 (WDR5) Localizes to the Midbody and Regulates Abscission. *J. Biol. Chem.* **290**, 8987–9001 (2015).

284. Ali, A., Veeranki, S. N. & Tyagi, S. A SET-domain-independent role of WRAD complex in cell-cycle regulatory function of mixed lineage leukemia. *Nucleic Acids Res.* **42**, 7611–7624 (2014).

285. Miller, T. *et al.* COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci.* **98**, 12902–12907 (2001).

286. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Mol. Cell* **25**, 15–30 (2007).

287. Briggs, S. D. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev.* **15**, 3286–3295 (2001).

288. Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* **6**, 227 (2005).

289. Glaser, S. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**, 1423–1432 (2006).

290. Dou, Y. *et al.* Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.* **13**, 713–719 (2006).

291. Dharmarajan, V., Lee, J.-H., Patel, A., Skalnik, D. G. & Cosgrove, M. S. Structural Basis for WDR5 Interaction (Win) Motif Recognition in Human SET1 Family Histone Methyltransferases. *J. Biol. Chem.* **287**, 27275–27289 (2012).

292. Ernst, P. & Vakoc, C. R. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief. Funct. Genomics* **11**, 217–226 (2012).

293. Ang, Y.-S. *et al.* Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. *Cell* **145**, 183–197 (2011).

294. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (2010).

295. Makrythanasis, P. *et al.* MLL2 mutation detection in 86 patients with Kabuki syndrome: a genotype-phenotype study: MLL2 mutation detection in 86 patients with Kabuki syndrome. *Clin. Genet.* **84**, 539–545 (2013).

296. Adam, M. P. *et al.* Kabuki syndrome: international consensus diagnostic criteria. *J. Med. Genet.* **56**, 89–95 (2019).

297. Denslow, S. A. & Wade, P. A. The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26**, 5433–5438 (2007).

298. Yildirim, O. *et al.* Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell* **147**, 1498–1510 (2011).

299. Basta, J. & Rauchman, M. The nucleosome remodeling and deacetylase complex in development and disease. *Transl. Res.* **165**, 36–47 (2015).

300. Bouazoune, K. The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J.* **21**, 2430–2440 (2002).

301. Roche, A. E. *et al.* The zinc finger and C-terminal domains of MTA proteins are required for FOG-2-mediated transcriptional repression via the NuRD complex. *J. Mol. Cell. Cardiol.* **44**, 352–360 (2008).

302. Murzina, N. V. *et al.* Structural Basis for the Recognition of Histone H4 by the Histone-Chaperone RbAp46. *Structure* **16**, 1077–1085 (2008).

303. Gnanapragasam, M. N. *et al.* p66 -MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc. Natl. Acad. Sci.* **108**, 7487–7492 (2011).

304. Hendrich, B. & Bird, A. Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *MOL CELL BIOL* **18**, 10 (1998).

305. Ee, L.-S. *et al.* An Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5. *Stem Cell Rep.* **8**, 1488–1496 (2017).

306. Reynolds, N. *et al.* NuRD Suppresses Pluripotency Gene Expression to Promote Transcriptional Heterogeneity and Lineage Commitment. *Cell Stem Cell* **10**, 583–594 (2012).

307. Lois, S., Blanco, N., Martínez-Balbás, M. & de la Cruz, X. The functional modulation of epigenetic regulators by alternative splicing. *BMC Genomics* **8**, 252 (2007).

308. *Diamond-Blackfan anemia*. https://rarediseases.info.nih.gov/diseases/6274/diamond-blackfan-anemia.

309. Dornelles-Wawruk, H. *et al.* Complex Phenotype Associated with 17q21.31 Microdeletion. *Mol. Syndromol.* **4**, 297–301 (2013).

310. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43**, 838–846 (2011).

311. Lin, Y.-F. *et al.* Diagnosis of Arboleda-Tham syndrome by whole genome sequencing in an Asian boy with severe developmental delay. *Mol. Genet. Metab. Rep.* **25**, 100686 (2020).

312. Herriges, J. C. Clinical and molecular cytogenetic characterization of a novel 10q interstitial deletion: a case report and review of the literature. 8 (2019).

313. Neveling, K. *et al.* A Post-Hoc Comparison of the Utility of Sanger Sequencing and Exome Sequencing for the Diagnosis of Heterogeneous Diseases. *Hum. Mutat.* **34**, 1721–1726 (2013).

314. Pajusalu, S., Reimand, T. & Õunap, K. Novel homozygous mutation in *KPTN* gene causing a familial intellectual disability-macrocephaly syndrome: *KPTN* -related syndrome. *Am. J. Med. Genet. A.* **167**, 1913–1915 (2015).

315. Tanaka, A. J. *et al.* Mutations in SPATA5 Are Associated with Microcephaly, Intellectual Disability, Seizures, and Hearing Loss. *Am. J. Hum. Genet.* **97**, 457–464 (2015).

316. Bouman, A. *et al.* Homozygous *DMRT2* variant associates with severe rib malformations in a newborn. *Am. J. Med. Genet. A.* **176**, 1216–1221 (2018).

317. Stellacci, E. *et al.* Clinical and functional characterization of two novel *ZBTB20* mutations causing Primrose syndrome. *Hum. Mutat.* **39**, 959–964 (2018).

318. Adzhubei, I. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

319. Capriotti, E. *i-Mutant Suite*. (ELIXIR-ITA-BOLOGNA).

320. Pelletier, N., Grégoire, S. & Yang, X. Analysis of Protein Lysine Acetylation In Vitro and In Vivo. *Curr. Protoc. Protein Sci.* **54**, (2008).

321. Choi, E. J. & Mayo, S. L. Generation and analysis of proline mutants in protein G. *Protein Eng. Des. Sel.* **19**, 285–289 (2006).

322. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* **7**, a025064 (2015).

323. Suganuma, T. & Workman, J. L. Signals and Combinatorial Functions of Histone Modifications. *Annu. Rev. Biochem.* **80**, 473–499 (2011).

324. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res.* **21**, 564–578 (2011).

325. Olopade, O. I., Grushko, T. A., Nanda, R. & Huo, D. Advances in Breast Cancer: Pathways to Personalized Medicine. *Clin. Cancer Res.* **14**, 7988–7999 (2008).

326. Williams, R. A., Mamotte, C. D. & Burnett, J. R. Phenylketonuria: An Inborn Error of Phenylalanine Metabolism. 11.