WDR5 is a Conserved Regulator of Protein Synthesis Gene Expression

By

Audra Foshage Bryan

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cell and Developmental Biology

January 31, 2020

Nashville, Tennessee

Approved:

William P. Tansey, Ph.D.

Chin Chiang, M.D., Ph.D.

Ethan Lee, M.D., Ph.D.

Jin Chen, MD., Ph.D.

Scott W. Hiebert, Ph.D.

To my beloved husband, Miles R. Bryan III, Ph.D., who has been infinitely supportive, understanding, loving, kind, and caring throughout the ups and the downs of the graduate school process

and

To my amazing family, for their support, understanding, love, wisdom, and guidance, including my father Gerald Foshage, my sister Alexandra Foshage Williams, my mother Virginia Becker, and my step-father Richard Becker

and

Of course to God, without whom, none of this would have been possible.

ACKNOWLEDGEMENTS

This thesis work would not have been possible without the financial support of Vanderbilt University Graduate School and the Department of Cell and Developmental Biology. Training grants supporting Audra Foshage Bryan (A.F.B.) include CA210429 and GM008554. Additionally, this work has been supported by the National Institutes of Health [CA200709 to William P. Tansey (W.P.T.), Chemical Biology Consortium Contract No. HHSN261200800001E to Stephen W. Fesik (S.W.F.), CA119925 to A.F.B., Erin R. Aho (E.R.A.), and W.P.T., CA225065 to Alissa duPuy Guarnaccia (A.d.G.), CA009582 to A.d.G., CA095103 to W.P.T.]. This work was also supported by grants from the Robert J. Kleberg, Jr., and Helen C. Kleberg Foundation [to W.P.T. and S.W.F.], The TJ Martell Foundation [to W.P.T. and S.W.F.], the Edward P. Evans Foundation [to W.P.T.], the Rally Foundation for Childhood Cancer Research Fellowship [to April M. Weissmiller (A.M.W.)], the Open Hands Overflowing Hearts co-funded research fellowship [to A.M.W.], the American Association for Cancer Research Basic Cancer Research Fellowship [to A.M.W.]. The VANTAGE Shared Resource is supported by the National Institutes of Health [CA068485, EY08126, and RR030956].

I am grateful to my mentor, William P. Tansey, Ph.D., for all he has taught me, to the members of my laboratory, and all my collaborators at Vanderbilt University. Thank you to the members of my Dissertation Committee, who have provided advice and guidance throughout my graduate school career, especially my Committee Chair, Dr. Chin Chiang.

A huge thank you is also owed my family and friends. Thank you to my father Gerald Foshage, my mother Virginia Becker, my step-father Richard Becker, my sister Alexandra Foshage Williams (also for the edits), my brother-in-law Cole Williams, my mother-in-law Kyong Bryan, my father-in-lab Miles R. Bryan Jr., my brother-in-law Dean Bryan, my sister-in-law Melissa Bryan, and my nieces Aubrey and Ariana Bryan. Thank you all for supporting me, listening to me, helping me, and not asking me when I was going to *finally* finish my Ph.D. Thank you to my friends and workout partners, who have helped me stay positive, healthy, optimistic, and sane: Kelley Doss, Jordyn Wilcox, and Derrick Johnson. Thank you to my kittens, Samantha, Wadsworth, and Charles, who have provided endless cuddles and helped me get through the darkest of times. Thank you from the depths of my heart to my amazingly supportive husband, Miles R. Bryan III, Ph.D. You have provided me guidance, love, a listening ear and caring hugs, and you have been the ultimate role model throughout this process. You have gotten me water, food, computer chargers, and blankets when I needed them, and have given me strength (via cuddles) when I needed it the most. You are my rock, my best friend, my soul mate, my confidant, and my biggest supporter. Most importantly, I would like to thank God, who has used this experience to draw me closer to Him and to teach me countless priceless lessons, and through whom all things are possible.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter I

Introduction

Preface

WDR5 is a remarkably highly-conserved WD40-repeat protein that shares over 90% sequence identity among vertebrates - in fact, human and mouse WDR5 are identical [1]. WD repeats are minimally conserved regions of ~40 amino acids, and according to RefSeq, WD40-repeats are typically bracketed by glycine-histidine and tryptophan-aspartic acid (GH-WD) residues, which can facilitate the formation of heterotrimeric or multi-protein complexes. WDR5 contains seven of these repeats and has been identified as a critical scaffolding component of multiple multi-protein complexes [2]. WDR5 has two well-defined interfaces, to which more than two dozen proteins have been shown to bind: the WDR5-binding motif (WBM) - a shallow, hydrophobic cleft on one surface of WDR5 - and the WDR5-interacting (WIN) site - an arginine-binding cavity on the opposite surface of WDR5 (Fig 1.1). The utilization of these two pockets by various interaction partners (Fig 1.3) lends WDR5 the flexibility to be able to function differentially in discrete biological contexts. WDR5 is most notably known to act as a scaffold, allowing for the assembly of multi-protein complexes, such as the MLL/SET-type histone methyltransferases (HMTs) that catalyze histone H3 lysine 4 (H3K4) di- and tri-methylation, and the non-specific lethal (NSL) and Ada2-containing (ATAC) histone acetyltransferase (HAT) complexes [2, 3]. While the role of WDR5 as a core scaffolding component of histone methyltransferase complexes has been extensively studied, emerging evidence indicates that WDR5 has a much broader role in the cell. The results of my thesis research indicate one such role for WDR5, in which, as part of a chromatin-bound complex, WDR5 helps regulate the transcription of genes involved in ribosome biogenesis, RNA processing, and translation. I hypothesize that WDR5 uses the WIN site to bind to chromatin, while using the WBM site to bind an E-Box-binding transcription factor. Further, I posit that by recruiting transcription factors to chromatin, WDR5 is able to assist in the regulation of ribosome protein gene transcription. Chapters III-VII will detail the results of my thesis work, which explores and challenges these hypotheses, and will provide evidence for another conserved role for chromatin-bound WDR5.

The History of WDR5

WDR5 was first identified in 2001, during the study of osteoblastic differentiation [4]. Differential display analysis performed by Gori, et. al. revealed the induction of a ~3 kb mRNA during osteoblastic differentiation, which they

dubbed BIG-3 (BMP-2-induced gene 3kb) [4]. In this seminal work, this novel WD40-repeat protein, BIG-3, was cloned and characterized, found to be induced by BMP-2 treatment, and shown to dramatically accelerate osteoblastic differentiation in stably transfected MC3T3E1 cells [4]. Within a few years, as the prevalence and importance of WD40 proteins was further uncovered, BIG-3 was renamed WDR5, in order to more appropriately reflect the architecture of the protein [2]. Throughout the early 2000's, the role of WDR5 in development and cellular differentiation was further explored. Studies of skeletal development in mice showed that WDR5 is able to promote cellular differentiation and proper bone formation when over-expressed [4, 5]. WDR5 was found to be expressed in immortalized marrow stromal cells, osteoblasts, osteocytes and chondrocytes [4], to also accelerate chondrocyte differentiation in cell culture models, to developmentally be expressed in osteoblasts as well as in proliferative and hypertrophic chondrocytes during endochondral bone formation [4, 5]. WDR5 was shown to accelerate osteoblast differentiation and maturation by activating the canonical Wnt signaling pathway during skeletal development [6]. In contrast to the increased skeletal structure observed upon WDR5 over-expression [6, 7], silencing of *WDR5* in the limbs of a developing embryo proved to severely impair bone development [8].

Around the time that mammalian BIG-3/WDR5 was identified, its *Saccharomyces cerevisiae* homologue, Swd3, was identified as a member of a newly-characterized histone methyltransferase complex. Called COMPASS (Complex of Proteins Associated with SET1) in yeast, this complex is the homolog of the mammalian SET1 and MLL (mixed lineage leukemia) complexes which were known to catalyze H3K4 di- and tri-methylation [9-11]. Then, in 2005, the Allis laboratory further connected WDR5 to epigenetics, showing that WDR5 can directly associate with methylated histone H3, the mark that is catalyzed by the SET1/MLL protein complexes [10, 12]. They went on to deplete WDR5 in human cells, and determine that this causes the expression of developmentally-essential *HOX* genes to be reduced, and in *Xenopus* embryos, which causes a decrease in H3K4 methylation and developmental defects [12]. It was work in mouse embryonic stem cells (mESCs) that really cemented the connection between WDR5 and development - WDR5 was found to be highly expressed in pluripotent mESCs, but upon differentiation, WDR5 levels decreased [13]. Further, perturbing WDR5 expression in mESCs impaired differentiation and repressed the transcription of self-renewal genes [13].

These studies linked the observed WDR5 developmental phenotypes to epigenetics and set the stage for the further exploration of the role of WDR5 in epigenetic modifying complexes. In fact, we now know that WDR5 is a critical scaffolding component of multiple chromatin remodeling complexes, as well as histone reader, writer, and eraser complexes [2]. As early investigations into WDR5 revealed its ability to read Histone 3 Lysine 4 (H3K4) methylation marks, with a preference for H3K4 dimethylated sequences [12, 14], researchers turned to the structure of WDR5 to

determine whether there were certain features of WDR5 that could lend the protein the ability to bind proteins with sequence as the H3 tail. Structural studies initially revealed that the WDR5 protein consists of seven typical WD40 repeat domains, allowing it to form a seven-bladed propeller fold with each blade containing a four-stranded anti-parallel sheet, and suggesting that WDR5 has many exposed surfaces [15]. Since WDR5 is particularly amenable to structural interrogation, more than 60 unique structures of WDR5 have been able to be deposited into the Protein Data Bank (PDB) over the past decade [2]. This ability to solve the structure of WDR5 alone and in complex with co-factors and inhibitors reflects the importance of structural biology to understanding both the canonical and non-canonical functions of WDR5, and to developing novel inhibitors able to block interactions at the surface of the protein. Structural analyses showed that while arginine 2 (R2) of the histone H3 tail engages the WIN site of WDR5, the lysine 4 does not [1, 16-18]. Further, phenylalanine 133 and 263 residues, within what was termed the WIN site, of WDR5 were found to form a phenylalanine clamp around the R2 of H3 [16-18]. WIN site residues include: Ala65, Ser91, Asp107, Phe133, Tyr191, Tyr260, Phe263. In 2008, WDR5 was found to also use this clamp to bind the WIN motif (A-R-A/S/T; Fig 1.1 and 1.2) in SET1/MLL (mixed lineage leukemia) proteins, revealing that WDR5 can not bind H3 while scaffolding an intact SET1/MLL complex [19-21]. WDR5 was then determined to be essential for the SET1/MLL complexes to thoroughly write H3K4 di- and tri-methylation marks, working with RBBP5, ASH2L, and DPY30 to stimulate the histone methyltransferase activity of these complexes [22-25]. Further, WDR5 depletion led to a decrease in H3K4Me3, and inhibitors of the interaction between MLL and WDR5 have shown that WDR5 is required for cancer cell growth [26-29].

Structural analyses also revealed a second site on WDR5, available for proteins to bind. WBM site residues include: Asn225, Tyr228, Leu240, Phe266, Val268, Gln289. This shallow cleft, on the opposite side of WDR5 as the WIN site, came to be known as the WDR5-binding motif (WBM) site (Fig 1.1 and 1.2). More than two dozen proteins that directly interact with WDR5 have been mapped with precision, and bind to either the WIN site or the WBM site. This has led to the general belief that the repeating use of these two sites by various WDR5-interaction partners is what allows WDR5 to function discriminately in different biochemical contexts. In fact, more recent studies have identified WDR5 as an important member of the NSL (non-specific lethal) complex, in which the KANSL1-WDR5 interaction appears to be required for efficient recruitment of the NSL complex to chromatin, and therefore the writing of acetylation marks on histone H4 lysine 16 [30-33]. In the NSL complex, WDR5 associates with MOF, KANSL3, MCRS1, PHF20, HCF-1, OGT, and E2F6, binds KANSL1 via the WIN site, and binds KANSL2 via the WBM site, indicating that WDR5 is a mutually exclusive component of both histone methyltransferase complexes and a histone acetyltransferase complex [34-36]. WDR5 can also use the WIN site to bind MBDC3 in the nucleosome remodeling and deacetylase (NuRD)

complex [37-39]. In this chromatin-associated protein complex, WDR5 helps both erase histone acetylation marks and remodel chromatin. Less well-studied roles for WDR5 also include its interactions with other histone acetyltransferase complexes, such as the ATAC (Ada2a-containing) complex [40, 41], histone methyltransferase complexes, such as PRC1.6 [42-44], and chromatin remodeling complexes, via association with CHD8 [45, 46] and INO80 [32, 47]. WDR5 can bind to the Polycomb protein Cbx8 to maintain histone H3K4 trimethylation on Notch-network gene promoters [48], can form a complex with HDAC1, HDAC2, arginine-glutamic acid dipeptide repeats (RERE) protein and histone methyltransferase G9a to regulate retinoic acid signaling and embryonic symmetry [49], and can directly interact with the transcription factor (TF) MKL1 and the chromatin remodeling protein BRG1 to enhance MKL1-mediated promoter activities of pro-inflammatory genes [50]. WDR5 has also recently been shown to bind to long non-coding RNAs, such as HOTTIP, NeST, linc1405, and BLACAT2, to facilitate long non-coding RNA-mediated histone H3K4 trimethylation and gene transcription [51-54]. Since WDR5 is a critical component of so many varying multi-protein complexes, it may be surprising that WDR5 can also bind to sequence-specific transcription factors. In fact, WDR5 has been shown to directly bind and promote the activity of the transcription factors Oct4 [13], Pitx2 [55], Twist1 [56, 57], HSF2 [58], and MYC [59-61]. Utilizing the interaction with MYC as an example, the Tansey laboratory has shown yet another molecular function for WDR5 — WDR5 facilitates the recruitment of MYC to target gene chromatin, with MYC binding directly to the WBM in WDR5 [60]. Furthermore, the requirement for MYC to directly contact both the WBM of WDR5 and a multitude of sites in chromatin simultaneously [60], indicates that a significant portion of chromatin-bound WDR5 is in complex with MYC and therefore not available to bind within its other known methyltransferase, acetyltransferase, deacetylase, and chromatin remodeling complexes.

While WDR5 has been identified as a critical component of multiple chromatin-binding complexes, more recently it has also been found to moonlight off chromatin. In 2015, WDR5 was detected at the midbody and found to regulate abscission [62], and in 2017, WDR5 was determined to regulate KIF2A localization, via the WIN site, to the mitotic spindle in order to ensure chromosome congressional and proper spindle assembly during mitosis [63]. Localization of WDR5 to the midbody and mitotic spindle in dividing cells has been shown to be dependent on the integrity of the WIN site [62], and is thought to functionally relevant, as WDR5 WIN site mutants fail to rescue mitotic defects associated with the knock-down of WDR5 [64]. Further, two WDR5-containing complexes, the NSL and ATAC complexes, have been shown to associate with microtubules and regulate mitotic integrity [65, 66]. While a mechanistic understanding of how WDR5 functions in these roles is yet to be uncovered, it is clear that WDR5 has important functions both on and off chromatin.

WDR5 in Cancer

WDR5 has been determined to play a role in various types of cancer. For example, aberrant WDR5 expression has been implicated in leukemia [67], breast cancer [68], and bladder cancer [69, 70]. Further, WDR5 has been shown to promote cancer initiation and progression, cancer cell invasion, and metastasis [71]. Therefore, multiple laboratories have hypothesized that if highly potent drug-like inhibitors of WDR5 can be discovered, they could have a tremendous impact in the clinic.

WDR5 has been well-established as an oncogenic target for MLL-rearranged (MLLr) leukemias [72]. Via the WIN site, WDR5 has been shown to recruit MLL to regulatory enhancers, enriched in E-twenty-six (ETS) family transcription factor binding sites, leading to the activation of leukemogenic genes and leukemia [73]. As will be discussed throughout this document, I have determined that WDR5 binds to TSS-Proximal regions, regardless of cell type interrogated. Further, in certain cell types, there is a population of TSS-Distal WDR5. While my analysis has been directed towards the conserved sites of WDR5 binding to chromatin, all of which are TSS-Proximal, I hypothesize that WDR5 binds to enhancers when located TSS-Distally. Therefore, these reports by Krivtsov, et al. support this hypothesis by providing evidence that the WIN site of WDR5 is also utilized for the recruitment of WDR5 to regulatory enhancers. WDR5 has also been shown to interact with C/EBPα p30, the most common type of CEBPα mutation (CEBPα is the CCAAT/enhancer binding protein α, and is mutated in 9% of acute myeloid leukemia), at genomic regions enriched of histone H3K4 trimethylation marks [27]. In this complex, WDR5 was determined to be essential for C/EBPα p30-dependent leukemia cell self-renewal, myeloid differentiation block, and leukemogenesis [27]. Via the use of WIN site inhibitors in leukemia cells, including MLLr cancers, the WIN site of WDR5 has been determined to be critical for leukemia cells, however the reason these cells are sensitive to WIN site inhibition has yet to be determined [72, 74]. Initial experiments performed with early WIN site inhibitors showed a decrease in H3K4 methylation at the *HOX* genes in leukemia cells, which led to the hypothesis that WIN site inhibitors kill these cells by starving them of the vital *HOX* gene products necessary to maintain the malignant and stem-like state [26]. However, further experiments, performed by the Tansey lab, have determined that the action of more refined WIN site inhibitors occurs through the rapid and persistent decreases in protein synthesis genes. Therefore, I posit that leukemia cell sensitivity has little to do with MLL1, *HOX* genes, or even H3K4 methylation, but instead, is related to the ability of MLL-fusion oncoproteins to induce the transcription of genes connected to ribosomal and nucleolar processes, an important part of their transcriptional repertoire [74-76].

Recent studies have also revealed that WDR5 performs key roles in the tumorigenesis and progression of cancer. Via the WBM, WDR5 binds to MYC proteins, one of the most prevalent families of oncoprotein transcription factors in cancer, which induce tumorigenesis by regulating the transcription of genes involved in the cell cycle, apoptosis, metabolism, genome stability, and cellular growth, proliferation, and metabolism [77]. The MYC-WDR5 interaction has been shown to be essential for the recruitment of MYC to chromatin, MYC target gene transcription, and the initiation and maintenance of MYC-driven caners [59-61]. Thus far, this interaction between WDR5 and MYC has been determined to be crucial in lymphoma, neuroblastoma, and pancreatic ductal adenocarcinomas [59-61, 78]. For example, in two publications, the Tansey laboratory has shown that the MYC-WDR5 interaction was critical for tumor initiation and maintenance [60, 61]. In these papers, we utilized a mutant form of MYC that can not bind WDR5 (this mutation is in the region of MYC that binds the WBM of WDR5, and is thus called WBM-MYC) in distinct mouse experiments: 1) a mouse allograft model showed that while WT-MYC drives tumorigenesis, WBM-MYC does not [60], and 2) a Burkitt's lymphoma system in which WT-MYC is allowed to drive tumorigenesis, then WBM-MYC is induced, showed that while WT-MYC allows for tumor maintenance and progression, the induction of WBM-MYC causes tumor regression [61]. Further, via screening of patient-derived pancreatic ductal adenocarcinoma xenografts and genetically engineered mouse model-derived allografts, WDR5 was identified as a top tumor maintenance gene for, and found to be considerably over-expressed in, human pancreatic ductal adenocarcinomas [78]. Mechanistically, WDR5 was shown to interact with c-Myc in order to sustain proper DNA replication and drive pancreatic ductal adenocarcinoma cell proliferation, tumor initiation, and progression [78].

WDR5 has also been shown to play a role in 1) prostate cancer, 2) breast cancer, 3) bladder cancer, 4) gastric cancer, and 5) colon cancer. 1) The androgen receptor is a key factor in the progression of prostate cancer to castration-resistance, and upon androgen stimulation, PKN1 induces both histone H3 threonine 11 phosphorylation (H3T11P) and WDR5 chromatin association at androgen receptor target gene loci [79, 80]. As WDR5 is known to interact with H3T11P, it is thought that through this interaction, WDR5 enhances MLL recruitment, and consequently H3K4me3, at androgen receptor target gene loci, allowing for the transcriptional induction of androgen receptor target genes [79, 80]. WDR5 has also been found to be over-expressed in human prostate cancer tissues, and thus, has been implicated in both prostate cancer cell proliferation and castration resistance [79, 80]. 2) A loss-of-function screen of 60 epigenetic regulators identified the Polycomb protein Cbx8 as a key regulator of mammary carcinoma both *in vitro* and *in vivo* [81]. WDR5 has been shown to associate with Cbx8 in order to maintain H3K4 trimethylation at Notch-network gene promoters [81]. Further, this interaction between WDR5 and Cbx8 was deemed to be required for Notch signaling activation, the

maintenance of oncogenic NOTCH signaling, and breast cancer tumorigenesis [48, 81]. 3) WDR5 has been found to be upregulated in bladder cancer tissues, and this high level of WDR5 expression positively correlates with advanced disease stage and poor patient survival [70]. In bladder cancer, WDR5 has been found to up-regulate cyclin B1, cyclin E1, cyclin E2, UHMK1, MCL1, BIRC3, and Nanog gene transcription, via the induction of H3K4 trimethylation, resulting in bladder cancer cell proliferation, self-renewal, chemo-resistance to cisplatin *in vitro*, and tumor growth *in vivo* [70]. 4) WDR5 over-expression in gastric cancer cell lines and human tumor tissues has been associated with poor patient survival rates [82]. Mechanistically, WDR5 has been shown to induce H3K4 trimethylation at the Cyclin D1 gene promoter, therefore driving Cyclin D1 gene transcription and gastric cancer cell proliferation [82]. 5) WDR5 over-expression in colon cancer cell lines and human tumor tissues has also been associated with poor patient survival [29]. WDR5 has been shown to decrease the phosphorylation of the histone protein H2AX and induce H3K4 trimethylation, leading to colon cancer cell proliferation and survival [29]. Further, WDR5 depletion sensitizes colon cancer cells to radiation-induced DNA damage [29]. While WDR5 has clearly been determined to play specific roles in specific cancer cell lines, there are, as of yet, no studies that compare the common roles of WDR5 across multiple cancer cell lines. Therefore, two of the goals of this thesis research have been to suss out the conserved sites of WDR5 binding to chromatin and the genes WDR5 commonly regulates, regardless of cancer cell type interrogated.

WDR5 has also been shown to be able to induce cancer cell invasion and metastasis. One important step in cancer invasion and metastasis is the epithelial-mesenchymal transition of cancer cells. Under hypoxic conditions in non-small cell lung carcinoma cells, WDR5 has been found to interact with HDAC3 in order to activate mesenchymal gene transcription and initiate hypoxia-induced epithelial-mesenchymal transition [56]. In lung squamous cell carcinoma and breast carcinoma cells, WDR5 is recruited by protein arginine methyltransferase 5 (PRMT5) complexes with MEP50/ WDR77 to target gene promoters, allowing WDR5 to induce H3K4 trimethylation and resulting in target gene transcription and cancer cell invasion [83]. WDR5 has also been shown to complex with TWIST1 and induce HOXA9 gene transcription - TWIST1 and HOXA9 are enriched in primary human prostate cancer tissues and even further over-expressed in metastatic tissues [57]. By complexing with TWIST1, WDR5 is thought to drive prostate cancer cell migration, invasion, and metastasis [57]. Further, high levels of the lncRNA BLACAT2 in human bladder cancer tissues has been associated with lymphangiogenesis and lymphatic metastasis, and WDR5 has been found to directly interact with BLACAT2 [54]. This interaction between WDR5 and BLACAT2 was thought to allow WDR5 to upregulate the expression of the critical lymphangiogenesis factor VEGF-C, therefore driving lymphangiogenesis and lymphatic metastasis [54]. Finally, WDR5 has been shown to bind to the lncRNA GClnc1, which is up-regulated in human gastric

cancer tissues [84]. In complex with GClnc1, WDR5 was thought to regulate the transcription of oncogenes, such as SOD2, and consequently induce gastric cancer cell proliferation, invasion, and metastasis [84].

Due to these varied roles for WDR5 in cancer, WDR5 has begun to be assessed as a novel target for cancer therapy. Further, in the past decade, it has become apparent that epigenetic regulatory proteins, like WDR5 has been proposed to be, can be targeted by small molecule inhibitors for therapeutic benefit, and currently there are dozens of small molecule epigenetic inhibitors in various stages of clinical trial, targeting histone code writers, readers, and erasers [85]. In fact, WDR5 is currently thought to be a promising therapeutic target in a number of bloodborne and solid cancers [26-28]. While inhibitor discovery efforts for WDR5 are still in the early phases, both the WIN and WBM sites on WDR5 have been proposed to be potential sites in which small molecule inhibitors (SMIs) of WDR5 may prove efficacious. For example, the Tansey laboratory, in collaboration with the Fesik laboratory here at Vanderbilt University, has recently published our discovery and refinement of a WDR5 SMI that targets the WIN site of WDR5 [74]. This WIN site inhibitor, named C6, is able to displace WDR5 from chromatin and reduce the transcription of WDR5-bound genes [74]. Another example is the macrocyclic peptidomimetic, MM-401, which engages the WIN site with a relatively high affinity (K_d ~1 nM) by mimicking the arginine of the WIN motif [26]. MM-401 has been shown to inhibit the HMT function of MLL1 complexes in vitro, consistent with the requirement of the WDR5-MLL1 interaction for robust methyltransferase activity [86]. Dou and colleagues originally proposed that WIN-site inhibitors would be efficacious against tumors expressing *MLL1* gene rearrangements (MLLr; common in acute myelogenous leukemia (AML)), as MLLr cancers typically retain one wild-type copy of *MLL1*, and are uniquely dependent on the HMT activity of WT-MLL1-complexes for survival [87]. In fact, Cao, et. al. found that MM-401 is highly selective against MLLr cancer cells *in vitro*, in which it depletes H3K4me3 at *HOXA* genes, driving cellular differentiation and apoptosis [26]. However, in 2017 Chen, et. al. determined that MLLr cancers actually rely on MLL2, rather than MLL1 [88]. While WDR5 also complexes with MLL2, MLL2 does not require WDR5 binding to perform its methyltransferase activity, unlike MLL1 [86]. Therefore, it is unlikely that the sensitivity of MLLr cancers to MM-401 is due to the inhibition of MLL1-mediated H3K4 methyltransferase function - as of yet, we are still unsure of how MM-401 selectively inhibits MLLr cells. A third example is OICR-9429, a SMI which binds the WIN site with a K_d of ~100 nM, inhibits MLL1-HMT activity *in vitro*, and can inhibit cancer cell lines in culture [89, 90]. OICR-9429 has been shown to inhibit AML cells with the oncogenic C/EBPα p30 isoform, which, as was mentioned above, binds WDR5 [27]. Further, cells expressing gain-of-function p53 mutants (oncogenic mutations which comprise the largest group of *TP53* mutations in human cancer) have also been shown to be highly sensitive to

OICR-9429 [28]. While there is no direct contact between oncogenic p53 variants and WDR5, this sensitivity has been said to result from the selective induction of MLL1 expression by the oncogenic p53 variants, the impact of which is mitigated by WIN site blockade and inhibition of MLL1-HMT activity [2]. As p53 gain of function mutants are featured so often across cancer types, these findings indicate the potentially huge therapeutic impact WDR5 inhibitors could have. As such, the Tansey laboratory also believes that targeting the other interface of WDR5, the WBM site, is another important avenue for WDR5 inhibition that should be explored. Further, we have proposed that WBM site inhibitors may also have utility as anti-cancer agents by virtue of their ability to block the MYC–WDR5 interaction, allowing them to thwart MYC function in cancer cells as well [60]. While as of yet no small molecule WBM site inhibitors have been reported, it seems like it is just a matter of time before these become available.

Outstanding Questions About WDR5

At first glance, the WDR5 protein does not seem to be anything special. Its composition of seven WDR40 repeats is shared by multiple scaffolding proteins [91], and the WD40 repeat domain itself is one of the top ten most common interacting domains across eukaryotic proteomes [92]. With the high abundance of WD40 proteins in the cell, we might expect that each chromatin-regulatory complex would have its own, dedicated WD40 member. However, WDR5 has been found to be used in multiple chromatin-associating, chromatin-regulatory, and gene-regulatory complexes, which act in differing ways. WDR5 has also been shown to interact with RNA molecules within the nucleus, and the mitotic spindle, midbody, and microtubules in the cytoplasm [54, 65, 66, 84]. As a scaffolding protein in various complexes, both in the nucleus and cytoplasm, WDR5 helps control cell division, pluoripotency, and development, drive epigenetic regulation, and regulate transcription, tumorigenesis and the progression of cancer. While WDR5 plays an essential role in normal biology, recently the role of WDR5 in cancer has been of particular interest in the field. For example, within the past decade, WDR5 has been shown to play a key role in multiple cancer types, including the methylation and acetylation of histones, the recruitment of proteins to enhancers, the transcription of oncogenes, and the regulation of transcription via interaction with oncogene transcription factors [71]. The ability of WDR5 to regulate transcription via interaction with oncogene transcription factors has been of particular interest to the Tansey laboratory, as we study one such WDR5-bound oncogene, MYC. Through interactions with oncogene transcription factors like MYC, WDR5 assists in the transcriptional regulation of genes essential for cell cycle progression, DNA replication, cell proliferation, cell growth, cell survival, tumorigenesis, tumor progression, tumor invasion, and tumor metastasis, all in a variety of cells of differing origins. These

recent discoveries have also led to the identification of small molecule inhibitors of WDR5, which are currently being optimized and tested as potential novel anticancer agents.

While we know that all direct interactions with WDR5, which allow WDR5 to help perform its aforementioned roles in cells, occur via one of two conserved sites (the WIN site or the WBM site), it is unlikely that we have yet identified all WDR5 binding partners. Therefore, it is imperative that further research be performed to identify novel WDR5 binding proteins, including unbiased protein immunoprecipitation and mass spectrometry experiments. These experiments should be performed in a variety of normal and cancer cell types in order to determine whether the interactions are common to healthy and diseased cells, and whether the same pathways, oncogenic or otherwise, are regulated by WDR5 in multiple, or all, cells. This repeated use of the same two binding sites not only makes WDR5 an intriguing protein to study and provides WDR5 the flexibility to mutually exclusively premiere in numerous protein complexes, but also makes the study of WDR5 quite challenging. Thus, it is likely to take years to fully appreciate the roles of WDR5 in their entirety and determine which roles are the most reverent to normal and disease states. Additionally, while small molecule inhibitors of WDR5 are currently being refined, whether these will prove to be useful cancer therapies, and if so, which cancer types they will be efficacious in, requires further research. Continued synthesis and screening of more potent WDR5 inhibitors, with ideal pharmacokinetics and bioavailability *in vivo*, is necessary to determine whether WDR5 inhibitors will be suitable for clinical trials. Simultaneous basic and clinical research on WDR5 will most effectively move the field forward, as learning more about the basic functions of WDR5 will also help inform us of the potential for WDR5 inhibitors to be efficacious, identify which cancer cell types will be sensitive to WDR5 inhibition, and determine whether WDR5 inhibitors will be able to be used in the treatment of cancer without also deleteriously affecting healthy cells.

Further, despite the importance of WDR5, as indicated by the common repurposing of WDR5, its broad range of molecular functions, the regulation of chromatin and gene transcription by WDR5, and its role in cancer, a systematic study of conserved WDR5-binding sites has yet to be performed. Therefore, I decided to determine where WDR5 binds across multiple species, cell lines, and cancer types, and investigate how WDR5 affects gene transcription at conserved WDR5-bound genes. The results of my thesis research identify conserved WDR5 binding sites and indicate that WDR5 helps regulate the transcription of these conserved target genes, which are involved in ribosome biogenesis, RNA processing, and translation. I hypothesize that WDR5 uses the WIN site to bind to chromatin, while using the WBM site to bind E-Box-binding transcription factors, and it is via the recruitment of these transcription factors to chromatin that I

posit WDR5 is able to assist in the regulation of ribosome protein gene transcription. The following chapters will explore these hypotheses and provide further evidence for the importance of WDR5 for gene regulation, regardless of cell type.

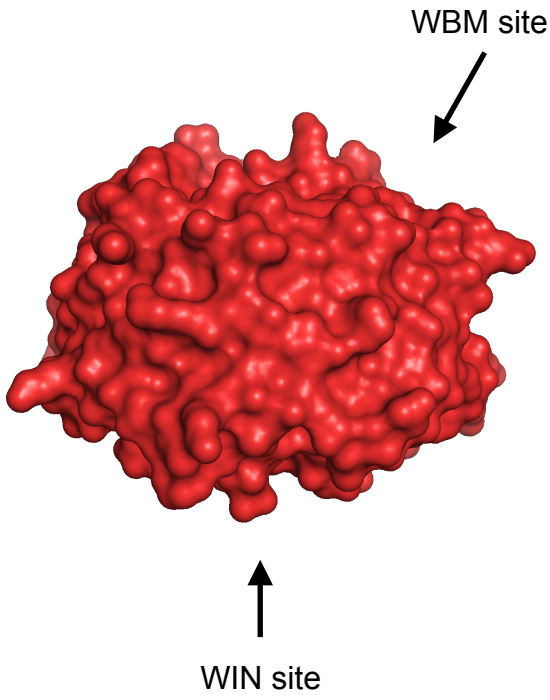


Figure 1.1. Surface structure of WDR5, in the same orientation as Figure 1.1 *Left*. In this orientation, the top face contains the WDR5-Binding Motif (WBM) site, and the bottom face contains the WDR5-Interacting (WIN) site.
Source: PDB ID 2H14.

Characterized WIN Motifs	
WDR5	EAARAQP
MLL1	GSARAEV
MLL2	GCARSEP
MLL3	GCARSEP
MLL4	GAARAEV
SET1A	GSARSEG
SET1B	GCARSEG
H3	--ARTKQ
KANSL1	VAARTRP
MBD3C	GAARCRV
KIF2A	GSARARP

Characterized WBM Motifs	
c-MYC	EEEIDVVSV
N-MYC	NEEIDVVTV
L-MYC	EEEIDVVTV
RBBP5	DEEVDVTSV
KANSL2	SDDLDVVGD

Figure 1.2. The proteins currently characterized, and determined to be direct interaction partners of WDR5, have similar motifs. The WIN motif and WBM motif sequences for these WDR5-interacting proteins are shown above. Notably, the WIN motifs are all centered on an arginine, which is flanked by an arginine upstream and either an arginine, serine, thymidine, or cytosine downstream. The WBM motif is comprised of a specific combination of acidic and hydrophobic residues, with "EEIDVV" being the most common and conserved WBM motif sequence. Residues highlighted in red are identical, while those in blue are homologous. Sequences provided by the following sources: Homo sapiens (NP_438172.1), Mus musculus (NP_543124.1), Xenopus tropicalis (NP_001011411.1), Drosophila melanogaster (NP_524984.1), Caenorhabditis elegans (Q17963.1), Trichoplax adhaerens (XP_002109498.1).

Binding Proteins	Functions	References [71]
SET1A/B	Inducing histone H3K4 mehtylation and target gene transcription	[11]
MLL1/2/3/4	Inducing histone H3K4 mehtylation and target gene transcription	[20, 22, 26]
KANSL1/2	Inducing histone H4 lysine 16 acetylation and recruiting the NSL complex to chromatin	[34, 35]
CUL4-DDB1	Inducing histone H3K4 mehtylation and acting as an adaptor for CUL4-DDB1 ligase-mediated substrate recognition and proteolysis	[3]
c-MYC & N-MYC	Inducing MYC target gene transcription and tumorigenesis	[59, 60, 61]
Oct4	Inducing transcriptional activation of pluripotency genes	[13]
CHD8	Inducing HOXA2 gene transcription and remodeling chromatin	[45, 46]
INO80	Chromatin remodeling	[32, 47]
GCN5	Modifying chromatin structure and regulating gene transcription	[41]
HDAC1/2 and G9a and RERE	Regulating retinoic acid signaling and embryonic symmetry	[49]
HDAC3	Increasing H3K4 trimethylation and mesenchymal gene expression	[56]
MKL1	Enhancing MKL1-mediated pro-inflammatory gene transcription	[50]
HSF2	Inducing HSF2 target genes	[58]
Pitx2	Inducing smooth muscle cell marker gene transcription and cell differentiation	[55]
Cbx8	Inducing Notch gene expression and tumorigenesis, and maintaining histone H3K4 trimethylation	[48]
Twist1	Inducing HOXA9 gene transcription and prostate cancer cell migration, invasion, and metastasis	[56, 57]
MBDC3	Nucleosome and chromatin remodeling and histone deacetylation	[37, 38, 39]

Binding RNAs	Functions	References [71]
HOTTIP	Inducing histone H3K4 methylation at the HOXA gene locus and HOXA gene over-expression	[51]
NeST	Inducing histone H3K4 methylation and IFN-y gene transcription, as well as susceptibility to viral and bacterial pathogens	[51]
Linc1405	Activating Mesp1 gene transcription	[53]
GClnc1	Inducing the transcription of oncogenes (ie: SOD2) and gastric cell proliferation, invasion, and metastasis	[84]
BLACAT2	Inducing VEGF-C expression, lymphangiogenesis, and lymphatic metastasis	[54]

Figure 1.3. Table1 - table of the proteins and long non-coding RNAs that bind WDR5. Protein/RNA functions are listed, according to the references cited.

CHAPTER II

Materials and Methods

Methods

Cell lines

MV4:11 (male), K562 (female), and Be2C (Be(2)C, male) cell lines were cultured in RPMI-1640 media with 10% FBS. HEK293 (female) cells were maintained in DMEM supplemented with 10% FBS. LoVo (male) cells were cultured in DMEM media supplemented with 10% FBS, 1% L-glutamine, and 1% Nonessential amino acids. MC38 (MC-38, female) and 3T3 (NIH3T3, male) cell lines were maintained in DMEM with 10% Bovine Serum. All media was supplemented with 100 IU/ml Penicillin and 100 µg/ml streptomycin. All cell lines were cultured at 37°C with 5% CO2 and split every 2-4 days; suspension cells were maintained at a cell density of between 1×10^5 and 1×10^6 cells/ml. RPMI-1640 media (Corning, 10-040-CV) contained L-glutamine. DMEM media (Corning, 10-013-CV) contained 4.5g/L glucose, L-glutamine, and sodium pyruvate.

Antibodies

Antibodies used in this study were: (i) IgG (2729, Cell Signaling), (ii) α-WDR5 (D9E1I, Cell Signaling), (iii) α-WDR5 (A302-429A, Bethyl), (iv) α-MYC (Y69, Abcam, ab32072), (v) α-MYCN (51705, Cell Signaling; western blot) (vi) α-N-MYC (Dr. Huck-Hui Ng, Serum; ChIP), (vii) α-GAPDH–HRP (MA5-15738-HRP, ThermoFisher), and (viii) goat α-rabbit IgG Fc secondary antibody (31463, ThermoFisher).

Endogenous WDR5 Expression Western Blots

HEK293, K562, MV4:11, LoVo, Be2C, 3T3, and MC38 cells were washed in PBS then lysed for 5 minutes on ice in 250 µL Kischkel buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, PMSF, Na_3VO_4, and Complete PIC). Whole cell extracts were sonicated at 25% power for 10s then clarified by centrifugation. Protein concentrations were quantified using the BioRad Bradford assay. Laemmli Sample Buffer was added and samples were boiled for 10 minutes before running on a 10% Poly-Acrylimide gel and transferring to PVDF membrane (PerkinElmer). Membranes were blocked in 5% milk in TBST for 20 minutes then probed with WDR5 primary antibody ([1:2,000], Cell Signaling, WDR5 (D9E1I) Rabbit mAb #13105) overnight. Anti-Rabbit HRP ([1:5,000], Goat anti-Rabbit IgG Secondary Antibody

HRP, Pierce #31463) was used with film and Supersignal West Pico PLUS (Pierce) to visualize signal. GAPDH-HRP ([1:50,000], Pierce #MA5-15738-HRP; RRID:AB_2537659) was used to detect GAPDH loading controls.

WDR5 Expression post-SMI Treatment Western Blots

K562/LoVo cells were treated for 4/16 hours with 2 μM/25 μM C6nc or C6, or a 0.1% DMSO. Cells were washed in PBS then lysed for 5 minutes on ice in 250 μL Kischkel buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, PMSF, Na_3VO_4, and Complete PIC). Whole cell extracts were sonicated at 25% power for 10s then clarified by centrifugation. Protein concentrations were quantified using the BioRad Bradford assay. Laemmli Sample Buffer (5 mL 1M Tris pH 6.8, 9 mL Glycerol, 2.3 g SDS, 1 mL BME, Bromophenol Blue to taste, water to 20 mL) was added and samples were boiled for 10 minutes before running on a 10% Poly-Acrylimide gel and transferring to PVDF membrane (PerkinElmer). Membranes were blocked in 5% milk in TBST for 20 minutes then probed with with WDR5 primary antibody ([1:2,000], Cell Signaling, WDR5 (D9E1I) Rabbit mAb #13105) overnight. Anti-Rabbit HRP ([1:5,000], Goat anti-Rabbit IgG Secondary Antibody HRP, Pierce #31463) was used with film and Supersignal West Pico PLUS (Pierce, 34580) to visualize signal. GAPDH-HRP ([1:50,000], Pierce #MA5-15738-HRP; RRID:AB_2537659) was used to detect GAPDH loading controls.

Chromatin Immunoprecipitation

LoVo cells were treated with 0.1% DMSO, 25 μM C6, or 25 μM C6nc for sixteen hours, then cross-linked with 0.75% formaldehyde (in PBS pH 7.4) at room temperature for 10 minutes. The reaction was quenched with 125 mM glycine for 10 minutes at room temperature, after which cells were washed twice with ice cold PBS. Cells were collected in ice cold PBS, pelleted, and lysed on ice in Nuclear Lysis Buffer A (1M HEPES pH 7.9, 1M KCl, 0.5M EDTA, 0.4% NP-40, PMSF, Na_3VO_4, and Complete Protease Inhibitor Cocktail (PIC) (Roche)) for 5 min. 750 μL of Lysis Buffer A was used per $1x10^7$ cells. Cells were once again pelleted and further lysed in Formaldehyde Lysis Buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100 (Fisher Bioreagents, BP151-500, Lot 180850), 1% SDS, PMSF, Na_3VO_4, and Complete PIC) using 100-250 μL of buffer per $1x10^7$ cells, on ice, for 15 minutes. Chromatin was sheared by 17-24 minute sonication (BioRuptor, Diagenode, UCD-200; on highest setting, alternating between 30 s on/30 s off) to yield a mean chromatin size of ~250 bp, and debris cleared by centrifugation. Sheared chromatin was diluted 10-fold in Formaldehyde Lysis Buffer without SDS (to achieve a final concentration of 0.1% SDS) before immunoprecipitation, rotating overnight at 4°C using 0.8 μL of IgG (Cell Signaling, Rabbit, Cat# 2729) or 4 μL of anti-WDR5 (Cell Signaling, (D9E1I) Rabbit mAb #13105) antibody. Chromatin from 10 million cells was used per immunoprecipitation reaction. The

next day, Protein-A agarose beads (Roche, 11 134 515 001) were washed three times in Formaldehyde Lysis Buffer without SDS, blocked for 30 min (10 mg BSA (Affymetrix, pH 7, 9048-46-8) and 100 µL Formaldehyde Lysis Buffer without SDS per 15 µL bed volume of Protein-A agarose beads) at room temperature, and added (100 µL per reaction) to the immunoprecipitation reactions for 4 hours at 4°C. Immune complexes were washed (by rotating beads for 5 minutes at 4°C with 1 mL of each buffer) sequentially with Low Salt Wash buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, PMSF, and Complete PIC), High Salt Wash Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, PMSF, and Complete PIC), LiCl Wash buffer (10 mM Tris pH 8.9, 250 mM LiCl, 1 mM EDTA, 1% Triton X-100, PMSF, and Complete PIC) and twice with TE (10 mM Tris pH 8.0, 1 mM EDTA, PMSF, and Complete PIC). Protein–DNA complexes were de-crosslinked overnight at 65°C in 50 µL Elution Buffer (10 mM Tris pH 8.0, 1 mM EDTA, 0.1% SDS, 20 µg Proteinase K (Clontech, 740506), 200 mM NaCl). Proteinase K was heat inactivated for 20 minutes 95°C then 150 µL of TE was added. 2 µL of DNA was used in a 15 µL PCR reaction using KAPA SYBR FAST qPCR Master Mix 2X Universal and Q-PCR using primers described in the **Oligonucleotide Table** below, then quantified on an Eppendorf Realplex2 Mastercycler in technical triplicate. ChIP signals were calculated as percent input. ChIP experiments were completed in biological quadruplicate with error bars representing the standard error of the mean.

Chromatin Immunoprecipitation–Sequencing

Cells were cross-linked with 0.75% formaldehyde (in PBS pH 7.4) at room temperature for 10 minutes. The reaction was quenched with 125 mM glycine for 10 minutes at room temperature, after which cells were washed twice with ice cold PBS. Cells were collected in ice cold PBS, pelleted, and lysed on ice in Nuclear Lysis Buffer A (1M HEPES pH 7.9, 1M KCl, 0.5M EDTA, 0.4% NP-40, PMSF, Na_3VO_4, and Complete Protease Inhibitor Cocktail (PIC) (Roche)) for 5 min. 750 µL of Buffer A was used per $1x10^7$ cells. Cells were once again pelleted and further lysed in Formaldehyde Lysis Buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100 (Fisher Bioreagents, BP151-500, Lot 180850), 1% SDS, PMSF, Na_3VO_4, and Complete PIC) using 100-250 µL of buffer per $1x10^7$ cells, on ice, for 15 minutes. Chromatin was sheared by 17-24 minute sonication (BioRuptor, Diagenode, UCD-200; on highest setting, alternating between 30 s on/30 s off) to yield a mean chromatin size of ~250 bp, and debris cleared by centrifugation. Sheared chromatin was diluted 10-fold in Formaldehyde Lysis Buffer without SDS (to achieve a final concentration of 0.1% SDS) before immunoprecipitation, rotating overnight at 4°C using 0.8 µL of IgG (Cell Signaling, Rabbit, Cat# 2729), 4 µL of anti-WDR5 (Cell Signaling, (D9E1I) Rabbit mAb #13105) antibody, 4 µL of anti-MYC (Y69, Abcam, ab32072) antibody, or 5 µL of anti-N-MYC (Dr. Huck-Hui Ng) antibody. Chromatin from 10 million cells was used per immunoprecipitation

reaction. The next day, Protein-A agarose beads (Roche, 11 134 515 001) were washed three times in Formaldehyde Lysis Buffer without SDS, blocked for 30 min (10 mg BSA (Affymetrix, pH 7, 9048-46-8) and 100 µL Formaldehyde Lysis Buffer without SDS per 15 µL bed volume of Protein-A agarose beads) at room temperature, and added (100 µL per reaction) to the immunoprecipitation reactions for 4 hours at 4°C. Immune complexes were washed (by rotating beads for 5 minutes at 4°C with 1 mL of each buffer) sequentially with Low Salt Wash buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, PMSF, and Complete PIC), High Salt Wash Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, PMSF, and Complete PIC), LiCl Wash buffer (10 mM Tris pH 8.9, 250 mM LiCl, 1 mM EDTA, 1% Triton X-100, PMSF, and Complete PIC) and twice with TE (10 mM Tris pH 8.0, 1 mM EDTA, PMSF, and Complete PIC). Protein–DNA complexes were de-crosslinked overnight at 65°C in 50 µL Elution Buffer (10 mM Tris pH 8.0, 1 mM EDTA, 0.1% SDS, 20 µg Proteinase K (Clontech, 740506), 200 mM NaCl). The following day, three immunoprecipitation reactions per condition were pooled (a total of 3 x 10^7 cellular equivalents per sample) and protein was removed by two rounds of phenol chloroform (Sigma) extraction. DNA was precipitated by adding 45 µL of 3M NaOAc pH 5.5, 10 mg glycogen (Roche, 10901393001), and 3 volumes of 100% ethanol. After 30 minute incubation at -80C, DNA was pelleted via centrifugation for 10 minutes at 16 x g at 4°C. DNA pellets were air-dried then resuspended in 35 µL TE, 30 µL of which was used for next generation sequencing (NGS) library preparation. Indexed libraries were made using the DNA Ultra II Library Prep Kit for Illumina (New England BioLabs, Inc., E7645). Library quality was assessed using the 2100 Bioanalyzer (Agilent) and libraries were quantitated using KAPA Library Quantification Kits (KAPA Biosystems). Pooled libraries were subject to 75 bp single-end sequencing according to the manufacturer's protocol (Illumina NextSeq500). Sequencing was performed by the Vanderbilt Technologies for Advanced Genomics (VANTAGE) Shared Resource. Used Bcl2fastq2 Conversion Software (Illumina) to generate de-multiplexed Fastq files.

K562 + SMI Treatment Chromatin Immunoprecipitation–Sequencing

K562 cells were grown to 10^6 cells/ml and treated for 4 hours with 0.1% DMSO, 2 µM C6, or 2 µM C6nc. Cells were concentrated to 4x10^6 cells/ml in PBS and cross-linked with 1% formaldehyde for 10 minutes at room temperature followed by quenching with 125 mM glycine for 10 minutes. Cells were washed with ice-cold PBS and lysed in 250 µL Formaldehyde Lysis Buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1% SDS, and Roche Complete PIC) per 10^7 cells and incubated on ice for 10 minutes. Chromatin was sheared (BioRuptor, Diagenode, UCD-200; on highest setting, alternating between 30 s on/30 s off) to achieve an average fragment size of ~250 nucleotides; cellular debris was then cleared through centrifugation for 10 minutes at 16 x g at 4°C. Sheared chromatin

was diluted 10-fold in Formaldehyde Lysis Buffer without SDS (to achieve a final concentration of 0.1% SDS) before immunoprecipitation overnight at 4°C using 6 µg of IgG (Cell Signaling, Rabbit, Cat# 2729) or 8 µL of anti-WDR5 (Cell Signaling, (D9E1I) Rabbit mAb #13105) antibody. Samples were rotated overnight at 4°C. The next day, Protein-A agarose beads (Roche, 11 134 515 001) were washed three times in Formaldehyde Lysis Buffer without SDS, blocked for 30 min (10 mg BSA and 100 µL Formaldehyde Lysis Buffer without SDS per 15 µL bed volume of Protein-A agarose beads) at room temperature. 100 µL blocked bead slurry was added to each reaction and rotated for 4 hours at 4°C. Immune complexes were washed (by rotating beads for 5 minutes at 4°C with 1 mL of each buffer) sequentially: once with Low Salt Wash buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton), High Salt Wash Buffer (20 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton), LiCl Wash buffer (25 mM LiCl, 1 mM EDTA, 1% Triton) and twice with TE (pH 8.0). Washed chromatin-bound beads were resuspended in 50 µL TE, 5 µL 1% SDS, and 20 µg Proteinase K and incubated overnight at 65°C. The following day, 300 µL TE was added and protein removed by phenol chloroform extraction. DNA was precipitated by adding 36 µL 3M NaOAc pH 5.2, 10 mg glycogen, and 1 mL 100% ethanol and centrifugation for 10 minutes at 16k x g at 4°C. DNA pellets were washed once with 70% ethanol and air-dried. DNA pellets were suspended in 100 µL TE and used for next generation sequencing (NGS) library preparation. Indexed libraries were made using the DNA Ultra II Library Prep Kit for Illumina (New England BioLabs, Inc., E7645). Library quality was assessed using the 2100 Bioanalyzer (Agilent) and libraries were quantitated using KAPA Library Quantification Kits (KAPA Biosystems). Pooled libraries were subject to 50 bp single-end sequencing according to the manufacturer's protocol (Illumina HiSeq 3000). Sequencing was performed by the Vanderbilt Technologies for Advanced Genomics (VANTAGE) Shared Resource. Bcl2fastq2 Conversion Software (Illumina) was used to generate de-multiplexed Fastq files.

ChIP-Sequencing analysis

ChIP-Seq reads were aligned to the genome hg19 or mm10 using Bowtie2 [93] after adaptor trimming by Cutadapt [DOI: 10.14806/ej.17.1.200]. Peaks were called using MACS2 with a q value of 0.01 [94]. Peak sets overlapping and quantification were determined by Diffbind [Stark R, Brown G (2011)]. DiffBind: differential binding analysis of ChIP-Seq peak data. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf]. Peaks were annotated using Homer [95] command annotatePeaks, and enriched motifs were identified by Homer command findMotifsGenome with the default region size and the motif length (http://homer.ucsd.edu/homer/). DAVID [96] was

used for all reported Functional GO analyses. Gene Set Enrichment Analysis (GSEA) [97] was performed to evaluate the enrichment of WDR5 binding genes in the repressed genes in response to 2 μM C6 treatment (RNA-Seq) in K562.

K562 + SMI Treatment RNA-Sequencing

K562 cells were treated for 72 hours with 2 μM C6, 2 μM C6nc or 0.1% DMSO. Cells were washed in PBS then lysed in 500 ml Trizol (Ambion by Life Technologies, 15596018). RNA was isolated using the Zymo Research Direct-zol RNA MiniPrep kit with on-column DNase digestion following the manufacturer's instructions. Library preparation with rRNA depletion and paired-end 150 base pair sequencing on an Illumina HiSeq was performed by GENEWIZ. Prior to sequencing, RNA integrity was assessed by 2100 Bioanalyser (Agilent) and concentration was assayed by Qubit. RNA-Seq for K562 cells treated with C6nc, C6 and DMSO was completed with 5 biological replicates.

RNA-Sequencing analysis

After adapter trimming by Cutadapt [98], RNA-Seq reads were aligned to the human reference genome using STAR [99], and quantified by featureCounts [100]. Read counts were normalized by the Relative Log Expression (RLE) method. Differential analysis were performed by DESeq2 [101], which determined the log2 fold changes, Wald test p-values, and adjusted p-value (FDR) by the Benjamini-Hochberg procedure. Significantly changed genes were assessed with an FDR<0.05 and a |log2FC|>1.

LoVo SLAM-Sequencing

Nascent RNA was labeled using the SLAMseq Kinetics Kit–Anabolic Kinetics Module (061, Lexogen). LoVo cells were pretreated for one hour with 25 μM C6, 25 μM C6nc, or 0.1% DMSO prior to a three hour labeling with 1 mM 4-Thiouridine (S4U) (with continued treatment with 25 μM C6, 25 μM C6nc, or 0.1% DMSO). Following the manufacturer's instructions, total RNA was isolated, alkylated with iodoacetamide, flash frozen, and shipped to Lexogen for analysis. After quality control analyses, libraries were prepared (250 ng RNA per sample) using Lexogen's QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina, following the User Guide (015UG009V0251) recommendations. Sequencing was performed by Lexogen on an Illumina NextSeq 500 system, using the SR75 High Output Kit. The SLAMdunk analysis pipeline [102] was used to analyze SLAMseq sequencing data. SLAM-Seq completed with three biological replicates.

Proliferation assay

White, opaque, flat-bottomed 96-well plates were used, seeded with 5,000 cells per well. LoVo cells were treated with 0.1% DMSO vehicle only and five two-fold dilutions of WDR5 inhibitor. Final DMSO concentration was 0.1% in all compound treatment experiments. Each concentration of inhibitor was tested in triplicate wells and three biological replicates were performed. Plates were incubated at 37^{o}C for three days and cells quantified using the CellTiter-Glo Luminescent Kit (Promega) according to the manufacturer's instructions. Raw luminescence values were normalized to the DMSO vehicle only wells and PRISM software was used to generate GI_{50} values. Error bars on proliferation curves represent standard error of the mean.

Table 2. Oligonucleotide Table

Table of all oligonucleotides used in this study.

Oligonucleotides	Assay	Sequence
hLAMP1 FWD	ChIP	TCACTTTCTCCCGCCACTAC
hLAMP1 REV	ChIP	ACCCTGGACCCACGTGAC
hSNHG15 FWD	ChIP	CGCCACTGAACCCAATCC
hSNHG15 REV	ChIP	TCTAGTCATCCACCGCCATC
hSNHG15 Gene Body (-ve) FWD	ChIP	AATTATGTGTCCAGGGTTGC
hSNHG15 Gene Body (-ve) REV	ChIP	CACCGGCTTCTATATTCCAC
hPUM1 FWD	ChIP	TATGAAGGGACAATCTGCTC
hPUM1 REV	ChIP	AATCCATCTTCATCCTACCG
hRPS6 FWD	ChIP	GAGACCCTTCTCCACCTAAA
hRPS6 REV	ChIP	CGAGTGTTAGACTGGGTTTG
hEIF4G1 FWD	ChIP	GGCAAATCCCATGTGCTC
hEIF4G1 REV	ChIP	CGTAGTCCACAACCATTTCC
hRNPS1 FWD	ChIP	GATGTAAGTTGGGGCGGAAT
hRNPS1 REV	ChIP	GAGGAGTGGACCGGCTTC
hRPL23 FWD	ChIP	AGATGTCGAAGCGAGGTGAG
hRPL23 REV	ChIP	GGCCTGAAGGAGAGCAAAG
hRPL35 FWD	ChIP	CTTGTGCAGCAATGGTGAGA
hRPL35 REV	ChIP	GCCTAGGTGGCAGATAGAATC
hRPL37 FWD	ChIP	GCAGAAGCGAGATGGTGAGT
hRPL37 REV	ChIP	CCCCAAGCACAGCAAACAG
hRPL5 FWD	ChIP	CCTGCAGGTCTCTGTCGAG
hRPL5 REV	ChIP	GGCATACGGGCAAGAAAAG
hRPS24 FWD	ChIP	TTGGCTGTCTGAAGATAGATCG
hRPS24 REV	ChIP	CGCGTGCCTATAGCTCAAGT
hCCT7 FWD	ChIP	TTCCAAAATGATGGTGAGTG
hCCT7 REV	ChIP	AGAGGGTCCTACAGAGCAAG
hRPS12 FWD	ChIP	TCTGAAGACTGCCCTCATCC
hRPS12 REV	ChIP	CTTGGGTGGCAGTTTTGTTC
hRPS5 FWD	ChIP	GGGCGAGATTACTGCATAGC
hRPS5 REV	ChIP	CTGTTTCCCTGCTCGACCT
hRPL10 FWD	ChIP	GCAAGAGTTCTACGCCCAAG
hRPL10 REV	ChIP	CACATGCGCAGATCAGAGAG
hMETTL1 FWD	ChIP	GCATGGCTGCGTCATTAACT
hMETTL1 REV	ChIP	GAGTCTCGGCTGCCATGAT
hSURF6 FWD	ChIP	GGGTGATAGAGGCACTGAGG
hSURF6 REV	ChIP	GATTAGCCAAGCCTGACTCC

Table 3. Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Normal Rabbit IgG antibody (ChIP)	Cell Signaling Technology	Cat# 2729; lot# 7; RRID:AB_1031062
Rabbit anti-WDR5 DE91I antibody (ChIP and western blot)	Cell Signaling Technology	Cat# 13105; lot# 1; RRID:AB_2620133
Anti-WDR5 antibody (ChIP)	Bethyl	A302-429A
Anti-MYC Y69 antibody (ChIP and western blot)	Abcam	ab32072
Anti-N-MYC antibody (ChIP)	Laboratory of Dr. Huck-Hui Ng	Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008.
Anti-MYCN antibody (western blot)	Cell Signaling Technology	51705
Anti-GAPDH-HRP (western blot)	Pierce	Cat# MA5-15738-HRP; RRID:AB_2537659
Goat anti-Rabbit IgG Secondary Antibody HRP (western blot)	Pierce	Cat# 31463
Protein-A Agarose Beads (ChIP)	Roche	Cat# 11134515001
Chemicals, Peptides, and Reagents		
Complete Protease Inhibitor Cocktail (PIC), EDTA-free	Roche	Cat# 11873580001
Formaldehyde	Sigma-Aldrich	Cat# 47608
C6	Aho, Wang et al. 2019	VU0808641
C6nc	Aho, Wang et al. 2019	VU0817566
DMSO	Sigma	Cat# D8418
Tween-20	Sigma	Cat# P2287
NP-40 Alternative	Calbiochem	Cat# 492016
Triton X-100	Fisher BioReagents	BP151-500
KAPA SYBR FAST qPCR Master Mix 2X Universal	Kapa Biosystems	Cat# KK4602
SuperSignal West Pico PLUS Chemiluminescent Substrate	ThermoFisher	Cat# 34580
Trizol	Ambion by Life Technologies	Ref# 15596018

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Reagents Continued		
Direct-zol RNA MiniPrep kit	Zymo Research	Cat# R2050
DNA Ultra II Library Prep Kit for Illumina	New England BioLabs, Inc.	Cat# E7645
Phenol:Chloroform:Isoamyl Alcohol	Sigma	Cat# P2069
DMEM, 1X (with 4.5g/L glucose, L-glutamine & sodium pyruvate)	Corning	Ref# 10-013-CV
RPMI 1640, 1X (with L-glutamine)	Corning	Ref# 10-040-CV
Fetal Bovine Serum (FBS)	Denville Scientific, Inc.	Cat# FB5001-H
L-glutamine	Corning	Cat# MT25005CI
Non-essential Amino Acids	Gibco	Cat# 11140-050
Bovine Serum	Gibco	Ref# 16170-078
100 IU/ml Penicillin and 100 µg/ml streptomycin	Gibco	Cat# 15140122
Glycine	RPI	Cat# G36050-5000
BSA	Affymetrix	Cas# 9048-46-8
Proteinase K	Clontech	Cat# 740506
Glycogen	Roche	10901393001
DMSO	Sigma	Cat# D8418
PVDF	PerkinElmer	Cat# nef1002
RNAse A	Clontech (EMD)	Cat# 740505
Deposited Data		
ChIP-seq: LoVo, Be2C, MC38, and 3T3 SLAM-seq: LoVo	GEO	GSE136451
ChIP-seq: K562 and MV4:11 RNA-seq: K562 and MV4:11 PRO-seq: MV4:11	GEO	GSE115377

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Commercial Assays		
CellTiter-Glo Proliferation assay kit	Promega	Cat# G7572
SLAMseq Kinetics Kit – Anabolic Kinetics Module	Lexogen	Cat# 061
KAPA Library Quantification Kits	KAPA Biosystems	KK4824; 07960140001
Bio-Rad Protein Assay Dye Reagent Concentrate (Bradford assay)	Bio Rad	Cat# 5000006
Experimental Models: Cell Lines		
K562	ATCC	Cat# CCL-243; RRID:CVCL_0004
LoVo	ATCC	Cat# CCL-229; RRID:CVCL_0399
Be(2)C (Be2C)	ATCC	Cat# CRL-2268; RRID:CVCL_0529
NIH3T3 (3T3)	ATCC	Cat# CRL-1658; RRID:CVCL_0594
MC-38 (MC38)	Lonza	Lonza, 893; CVCL_B288
MV4:11	ATCC	Cat# CRL-9591; RRID:CVCL_0064
HEK293	ATCC	Cat# CRL-1573; RRID:CVCL_0045
Software and Algorithms		
Prism 7	GraphPad	https://www.graphpad.com/scientific-software/prism/
SLAMdunk (SLAMseq Data Analysis Pipeline)	Lexogen	https://www.lexogen.com/wp-content/uploads\2017/11/063UG147V0100-SLAMdunk-User-Guide.pdf
Bowtie2	[94] Langmead, B., et al., Genome Biol, 2009.	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Cutadapt	[99] Martin, M., EMBnet. journal, 2011.	DOI:10.14806/ej.17.1.200
MACS2	[95] Feng, J., et al., Nat Protoc, 2012.	
Diffbind	Stark R, Brown G, 2011. *DiffBind: differential binding analysis of ChIP-Seq peak data*	http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf]
DAVID	[97] Huang da, W., B.T. Sherman, and R.A. Lempicki. Nature protocols, 2009.	https://david.ncifcrf.gov/home.jsp
HOMER	[96] Heinz, S., et al., Mol Cell, 2010.	http://homer.ucsd.edu/homer/
STAR	[100] Dobin, A., et al., Bioinformatics, 2013.	https://academic.oup.com/bioinformatics/article/29/1/15/272537
featureCounts	[101] Liao, Y., G.K. Smyth, and W. Shi, Bioinformatics, 2014.	https://academic.oup.com/bioinformatics/article/30/7/923/232889
DESeq2	[102] Love, M.I., W. Huber, and S. Anders, Genome Biol, 2014.	

CHAPTER III

Comparison of the WDR5–Chromatin Association in Disparate Cell Types

Abstract

WDR5, a highly-conserved nuclear protein, is known to perform scaffolding functions for multiple chromatin-associating complexes. Efforts to fully understand WDR5 have previously been complicated by its many functions in the nucleus and thwarted by a lack of knowledge of the conserved gene networks that are under its control. Therefore, to better understand the roles WDR5 plays in the cell, we first must thoroughly investigate where WDR5 binds to chromatin. This chapter will detail the binding of WDR5 across the genome in a variety of cell lines. Chromatin immunoprecipitation, followed by sequencing (ChIP-Seq), was performed on two mouse cell lines and four human cell lines. Comprehensive investigation into the location, intensity, and distribution of WDR5 binding across these multiple species, cell lines, and cancer types indicated common WDR5-peak characteristics, binding sequences, and gene localization. Together, these data reveal the conservation of WDR5 binding to TSS-proximal regions (Transcriptional Start Site), replete with sequence-specific transcription factor motifs, and at genes connected to protein synthesis across cell types, species, and cancers.

Introduction

WDR5 functions in a variety of processes within the nucleus and on chromatin. WDR5 has been established to be a critical scaffolding component of multiple chromatin remodeling complexes, as well as epigenetic regulating complexes [2]. Early investigations into WDR5 revealed its ability to bind sites of Histone 3 Lysine 4 (H3K4) methylation [12, 14, 18], and structural analyses showed that residues within the WIN site of WDR5 form a phenylalanine clamp around Arginine 2 of H3 [16-18]. WDR5 also uses this clamp to bind the WIN motif in SET1/MLL proteins, revealing that WDR5 can not bind H3 while scaffolding an intact SET1/MLL complex [19-21]. WDR5 is essential for the SET1/MLL complexes to thoroughly di- and tri-methylate H3K4, working with RBBP5 (which binds the WBM), ASH2L, and DPY30 to stimulate the histone methyltransferase activity of these complexes [22-25]. More recent studies show WDR5 is an important member of the NSL (non-specific lethal) complex, in which the KANSL1-WDR5 interaction is required for efficient recruitment of the NSL complex to chromatin, and therefore the acetylation of H4K16 [30-33]. In the NSL complex, WDR5 associates with MOF and E2F6, binds KANSL1 via the WIN site, and binds KANSL2 via the WBM

site, indicating that WDR5 is a mutually exclusive component of both histone methyltransferase complexes and histone acetyltransferase complexes [34, 35]. Less well-studied roles for WDR5 also include its interactions with other histone acetyltransferase complexes, such as the ATAC (Ada2a-containing) complex [40, 41], histone methyltransferase complexes, such as PRC1.6 [42-44], and chromatin remodeling complexes, via association with CHD8 [45, 46]. WDR5 can bind to the Polycomb protein Cbx8 to maintain histone H3K4 trimethylation on Notch-network gene promoters [48], and can form a complex with HDAC1, HDAC2, arginine-glutamic acid dipeptide repeats (RERE) protein, and histone methyltransferase G9a to regulate retinoic acid signaling and embryonic symmetry [49]. WDR5 can also bind to and promote the activity of sequence-specific transcription factors such as Oct4 [13], HSF2 [58], and MYC [59, 60]. Clearly, the multi-functional, mutually-exclusive roles of WDR5 as part of various types of chromatin-associating complexes make it difficult to fundamentally understand the role of WDR5, to tease apart the significance of WDR5 stably bound to chromatin, and to establish the predominant biological setting in which WDR5 operates. In order to determine the consequences of WDR5 stably binding chromatin and deduce the conserved, primary transcriptional responses to this binding, a thorough investigation of WDR5 localization across multiple cell types and species is necessary.

Over the past fifteen years, the location of WDR5 on chromatin has been evaluated, as is outlined below. However, as there has been no systemic investigation of WDR5 across contexts, there is little consensus on where WDR5 commonly binds to chromatin regardless of cell type. In Cell (2005), Wysocka, et. al. revealed that WDR5 directly associates with H3K4me2 nucleosomes in HeLa cells [12]. Ang, et. al., in Cell, 2011, used mouse ES cells to show that WDR5 co-localizes with Oct4, H3K4me3, and Rbbp5 [13]. WDR5, Rbbp5, and H3K4me3 binding regions were largely located within Refseq promoters over-represented in gene-rich chromosomal regions [13]. In fact, 9,303 WDR5, Rbbp5 and H3K4me3 co-associated target genes were identified, many with bivalent domains, and fell into developmental processes, neurogenesis, embryogenesis, and mesoderm and ectoderm development GO (Gene Ontology) enrichment categories [13]. This paper also showed that Oct4 is required for WDR5 recruitment to self-renewal-associated gene promoters and maintenance of robust H3K4 trimethylation - and via this direct binding to Oct4 and subsequent H3K4 trimethylation at Oct4 target gene promoters, WDR5 activates the transcription of Oct4 target genes [13]. Wang, et. al., in Nature, 2011, determined that HOTTIP directly binds WDR5 and targets WDR5/MLL complexes across the *HOXA* locus, driving histone H3K4 trimethylation and gene transcription in distally derived cells [51]. WDR5 and MLL1 densely occupied an extended region of the 5′ *HOXA* cluster, coincident with the H3K4me3 domain, and exhibited defined peaks near the TSSs of multiple 5′ *HOXA* genes [51]. WDR5 was also required for 5′ *HOXA* gene expression, including HOTTIP [51].

In Cell Stem Cell (2012), Kloc and Ivanova show that one-third of all WDR5/H3K4me3 peaks overlap with MOF peaks, in ES cells, and that the joint MOF/WDR5/H3K4me3 peaks are located around the TSS [103]. In PLoS Genetics (2012), Hopkin, et. al. determined that GRHL3 directly interacts with WDR5 in differentiated keratinocytes and that 88% of genes that contain a GRHL3 peak also have an overlapping WDR5 peak, supporting their hypothesis that GRHL3 recruits WDR5 to regulatory regions of epidermal differentiation genes in differentiated NHEK cells [104]. For both WDR5 and GRHL3 there was statistically significant enrichment in occupancy at promoters, compared to the average genomic distribution. GO analysis of these co-occupied targets revealed enrichment for processes like cell differentiation, positive regulation of gene expression, regulation of programmed cell death, cell-cell adhesion, and regulation of lipid biosynthetic processes, all important components of epidermal keratinocyte differentiation [104]. Zhao, et. al., in Cell, 2013, used ChIP-Seq in ESCs to determine that roughly 30% of WDR5 bound genes were bound by PHF20, and that both WDR5 and PHF20 were bound to the *Oct4* promoter and several key epigenetic factor genes, including *Baf155*, *Brg-1,* and *Sall4,* [105]. The majority of WDR5 and PHF20 binding sites mapped to the gene body (coding and intron regions) and the 5′ proximal region, and both WDR5 and PHF20 binding peaks centered on the TSS within a 7 kb region (from -2 to +5 kb). GO analysis showed that WDR5 and PHF20 binding targets were enriched for genes involved in cell and organ developmental process, embryonic development, and cell differentiation in ESCs [105].

Cheng, et. al., in Molecular Cell, 2014, provided evidence that WDR5 co-localizes with H3K4me3, Sin3A, ING1, LSD1, MLL1, and Menin at the majority of active genes in C2C12 myoblasts [106]. WDR5 bound to muscle genes, *Myog*, *Acta1*, and *Mybph* irrespective of H3K4 methylation status, and to a portion of the *Hoxa* genes [106]. Kim, et. al., in Molecular Cell, 2014, found that WDR5 interacts with H3T11P, facilitates the recruitment of the MLL1 complex, and subsequently aids in the trimethylation of H3K4 in LNCaP cells [107]. Upon androgen stimulation, H3T11P-marked regions increased six-fold and H3T11P recruited WDR5 to one third of H3T11P-enriched promoters [107]. The authors went on to claim that WDR5 is a critical epigenomic integrator of histone phosphorylation and methylation in LNCaP cells with androgen stimulation - WDR5 and H3T11P were enriched at the *IGF2R* and *RPL13A* androgen-responsive gene promoters, sites also occupied by Androgen Receptor [107]. Riggi, et. al., in Cancer Cell, 2014, showed that WDR5 localizes to 88% of activated EWS-FLI1 sites in SKNMC and A673 Ewing sarcoma cells [108]. EWS-FLI1 expression in SKNMC and mesenchymal stem cells leads to nucleosomal rearrangement and WDR5 recruitment - following this, H3K4me1 and H3K27ac enhancer marks are deposited and chromatin remodeling complexes are recruited in order to induce de novo active enhancers at DNA repeats lacking regulatory functions in other contexts [108].

In *Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC* (Molecular Cell, 2015), the Tansey laboratory mapped WDR5 binding in HEK293 cells and determined that WDR5 displayed a marked preference for promoters, but showed significantly higher enrichment in intergenic regions than c-MYC [60]. We concluded that there is a widespread and intimate overlap of MYC and WDR5 binding sites on chromatin, as ~80% of c-MYC binding sites overlapped with WDR5 [60]. Additionally, we determined that ~80% of the c-MYC/WDR5 co-localized genes contained the E-Box motif and, using high-resolution ChIP mapping, we showed that MYC and WDR5 distribute identically across *SNHG15*, with the signals for both proteins peaking at an E-Box within exon 1, just upstream of the H3K4me3 peak [60]. Sun, et. al., in Cancer Research, 2015, indicated that WDR5 binds, with N-MYC, to N-MYC target gene promoters rich in E-Boxes, and preferentially modulates histone H3K4 trimethylation and transcriptional activation of N-MYC target genes, including *MDM2* and *cyclin E1* in Be2C cells [59].

Xu, et. al., in Cell Discovery, 2016, determined that 50% of WDR5 binding occurs at intergenic and intronic regions in primary murine MLL-AF9 cells [73]. 40% of WDR5 targets were also bound by MLL1, and significant enrichment of WDR5, as well as H3K4 mono-, di-, and tri-methylation, was found at or near MLL1 peak centers [73]. Furthermore, WDR5, MLL1 and H3K4me1/2 were enriched at promoter/TSS-proximal and enhancer regions, and MLL1-WDR5 direct target genes include those involved in cell signaling, transcription, hypoxia, hematopoiesis and myeloid differentiation, as determined by GO analysis [73]. Ge, et. al., in Oncotarget, 2016, showed that more than half of WDR5 peaks are located in intergene regions and ~6% of peaks are in promoter and enhancer regions in both human RS4;11 ALL and THP-1 AML leukemia cells [67]. Genes to which WDR5 is bound in these leukemia cells fall into the following GO categories: multi-oncogenic signaling, apoptosis, transcriptional regulation, histone modification, cell proliferation and apoptosis, cell adhesion and metabolism [67]. In both RS4;11 and THP-1 cell lines, WDR5 peaks significantly overlapped with H3K4me3 peaks, and in Nalm6 B-ALL, U937 AML, and primary B-ALL and AML cells, WDR5 significantly bound to - and H3K4me3 was enriched in - the promoter region of Tyrosine-protein kinase Lyn (*Lyn*), B-cell CLL/lymphoma 9 (*BCL9*), Ras-related protein Rab-28 (*RAb28*), Cluster of Differentiation 93 (*CD93*), Mediater complex subunit 24 (*MED24*), and *RBM22* [67]. Sun, et al., in Cancer Discovery, 2016, determined that WDR5 binds to 966 gene promoters in BGC823 cells [84]. 147 of these genes were occupied by both WDR5 and KAT2A and the promoter regions of these genes were rich in H3K4me3 and H3K9Ac, the respective histone methylation and acetylation products of the WDR5 and KAT2A complex [84].

In Molecular Cell, 2019, Scelfo, et. al., showed that WDR5 localizes to Polycomb group ring finger protein promoter regions in wild-type mouse ESCs [109]. In *Displacement of WDR5 from Chromatin by a WIN Site Inhibitor with*

Picomolar Affinity (Cell Reports 2019), the Tansey laboratory detailed WDR5 binding in MV4:11 cells, indicating that WDR5-binding sites are predominantly promoter-proximal and often occur within 500 bp downstream of the TSS [74]. Among WDR5-bound genes, we observed a strong biological clustering under GO terms connected to protein synthesis and the ribosome, and determined that WDR5 binds a specific subset of translation initiation factors and ribosome protein genes, corresponding to genes encoding ~40% of the small and ~70% of the large ribosome subunit proteins [74]. Clearly the localization of WDR5 on chromatin has been of interest in the field for the past fifteen years, however there is yet to be a comprehensive analysis of common WDR5 bound genes across cell types. WDR5 has been reported to bind to different areas on chromatin and a variety of GO categories, but whether there is a common set of genes to which WDR5 binds and regulates, and therefore common enriched GO categories, is not yet known. This lack of knowledge about conserved binding and regulation via WDR5 is due to the fact that there is no systematic analysis of WDR5 across contexts, the problem that previous studies have mixed up WDR5 binding with other components of methyltransferase/ acetyltransferase/chromatin remodeling complexes, and the use of different reagents and antibodies across these publications.

While the total binding of WDR5 genome-wide has previously been assessed in a variety of cell lines, there are currently no studies that survey the overlap of WDR5 binding across cell lines, using the same antibody and reagents. Previous reports for total WDR5 binding sites vary greatly, between cell types and antibodies used, and include: 10,915 in mouse embryonic stem cells (ESCs) [13], 833 in LNCaP cells [107], 7,774 genes in ESCs [105], and 17,171 in HEK293 cells [60] using the Bethyl A302-429A antibody, 1,490 in RS4;11 cells, 515 in THP-1 cells [67]**,** and 48,269 in NHEK cells [104] using the Abcam 2C2 antibody (ab56919), 5,549 in BGC823 cells [84] and 158 in MV4:11 cells [74] using the Cell Signaling Technologies (CST) D9E1I antibody, and 13,075 peaks in MLL-AF9 (transduced mouse bone marrow cells) cells [73] using the anti-rabbit antibody made by the Dou laboratory. The lack of published data overlapping WDR5 across cell lines, using the same reagents, has made it impossible to determine the conserved sites of WDR5 binding, and therefore to identify potential roles for chromatin-bound WDR5.

Therefore, in an effort to better understand the importance and conserved functions of this cellular multitasker, I first set out to determine where WDR5 binds in the genome, which genes WDR5 binds to and regulates, and the extent to which this is conserved. To identify the conserved sites in chromatin to which WDR5 is bound, I performed ChIP-Seq in five cell lines: two mouse lines - MC38, a colon adenocarcinoma cell line, and 3T3, an immortalized fibroblast cell line - and three human lines - LoVo, a colorectal adenocarcinoma cell line, K562, a chronic myelogenous leukemia cell line, and Be2C, a neuroblastoma cell line. I also incorporated data from a recent paper, in which Aho, et. al. published ChIP/PRO/

RNA-Seq in MV4:11 cells, a biphenotypic B myelomonocytic leukemia cell line [74]. Importantly, all ChIP-Seq experiments were performed in the same laboratory, with the same reagents, and the same antibody.

Results

While total binding of WDR5 genome-wide has previously been assessed in a variety of cell lines, there are currently no published studies that survey the overlap of WDR5 binding across cell lines, using the same antibody and reagents. A thorough, well-controlled experiment like this is important, as available, published, ChIP-seq data for WDR5 has been collected using multiple WDR5 antibodies and have been performed by multiple laboratories. Through the use of the same antibody and reagents across experiments, we can more appropriately control for experimental error and antibody variation, and therefore produce more accurate, reliable data. As mentioned previously in this chapter, the reported total number of WDR5 binding sites varies greatly between cell types and antibodies used. In order to determine the ideal WDR5 antibody to perform chromatin immunoprecipitation (ChIP) experiments with, the Tansey laboratory tested two previously used, commercially available antibodies: the the Bethyl A302-429A antibody and CST D9E1I antibody. Using ChIP-Seq (ChIP followed by sequencing) Dr. Gregory Caleb Howard identified 158 WDR5 binding sites using the Bethyl antibody and 149 using the CST antibody in MV4:11 cells, 99 of which were bound by WDR5 using both antibodies, in side-by-side experiments [74]. Further, while our paper in 2015 reported 17,171 WDR5 binding sites in HEK293 cells with the Bethyl antibody [60], our repetition of these experiments with the CST antibody identified far fewer WDR5 binding sites in HEK293 cells. I also completed ChIP-Seq and identified a greater number of WDR5 binding sites using the Bethyl antibody, versus the CST antibody, in Be2C cells. Additionally, results of preliminary ChIP experiments performed by Dr. Gregory Caleb Howard indicate that Flag-WDR5 (detected via use of a Flag antibody with Flag-epitope tagged WDR5) and Millipore-detected WDR5 (MilliPore 1075 WDR5 antibody used) overlap with CST-detected WDR5 at all loci probed, but do not overlap with Bethyl-detected WDR5 at every locus. I postulate that the increased number of WDR5 binding sites identified using the Bethyl antibody is due to off-target antibody binding, as there is a significant number of sites to which the Bethyl antibody indicates WDR5 binding, but there is no Flag-WDR5 peak. Therefore, I have used the CST WDR5 (D9E1I) antibody for these studies.

To begin to understand the general biological context in which WDR5 operates, I compared the genomic location of WDR5 in six cell lines of varying cell type, species, and cancer background, with different levels of WDR5 expression. Mouse and human WDR5 are identical, which allowed me to study the conservation of WDR5 binding across the genome of two species. This further justified the use of the CST WDR5 antibody (D9E1I - it binds to WDR5 well in both mouse

and human cells) for this analysis, and permitted me to directly compare my results. By incorporating binding across two species, I can more fully investigate the evolutionary conservation of WDR5 binding to target genes in chromatin. As the Tansey laboratory has assessed WDR5 binding in a non-cancer-derived human cell line, HEK293 [60], I chose to use 3T3 (NIH3T3) cells, a transformed (spontaneously immortalized) fibroblast cell line, as a non-cancer-derived mouse cell line. I also chose to incorporate various cancer-derived cell lines for three reasons: 1) In addition to WDR5, the Tansey laboratory studies the well-known oncogene MYC. 2) MYC dysregulation in cancer cells drives a gene expression program that, mechanistically, has been shown to require H3K4 trimethylation at target gene promoters [110]. WDR5 has been shown to have an essential role in the trimethylation of H3K4 [12, 37, 51]. Therefore, the relationship between WDR5-mediated K3K4me3 and MYC transcriptional programs will be important for a better understanding of the roles both proteins play in cancer, as well as their use as therapeutic targets. 3) Recent studies have revealed that WDR5 plays key roles in tumorigenesis and the progression of a variety of cancers [29, 48, 59, 60, 70, 78, 82, 107, 111].

MC38, a mouse colon adenocarcinoma cell line, was paired with 3T3 cells, as the mouse cancer-cell of choice. Furthermore, LoVo cells, human colorectal adenocarcinoma cells, were chosen as the human cell line to pair with the MC38 mouse cell line. Colon/colorectal adenocarcinoma cell lines were chosen due to the well-established role c-MYC plays in these cancers [112-117]. In BMC Cancer (2018), Neilsen, et. al. state that their data “demonstrate a clear role for WDR5 in colon cancer and future studies should examine its potential to serve as a therapeutic target in cancer” [29]. Be2C, a human neuroblastoma cell line, was chosen because in Cancer Research (2015), Sun, et. al. identified WDR5 as a key cofactor for N-MYC-regulated transcriptional activation [59]. In this paper, they set a precedent that the interaction between WDR5 and N-MYC is important for tumorigenesis, show that WDR5 binds to N-MYC, but they do not assess genome-wide binding of WDR5 and N-MYC. Use of a neuroblastoma cell line also allows me to incorporate both c-MYC and N-MYC driven cancer cell lines into my studies. K562, a chronic myelogenous leukemia cell line was paired with our recently published data in MV4:11 cells [74] because most experiments testing the efficacy and consequences of using WDR5 inhibitors have focused on heme malignancies, specifically those driven by MLL-fusion oncoproteins (ie: MV4:11 cells have an MLL-AF4 gene fusion) [76]. K562 cells, on the other hand, are derived from a heme malignancy with a BCR-ABL1 gene fusion. Using two cells lines from heme malignancies with different well-known oncogene fusions allows for greater coverage of this cancer type.

ChIP-Seq performed across these six cell lines, two species, and five cancer types revealed varying amounts of WDR5 binding across the genome. ChIP-Seq reads were aligned to the genome hg19 or mm10 and peaks were called using MACS2, with a q value of 0.01 [94]. Figure 2.1 indicates the total number of WDR5 peaks per cell line, as well as

the total number of genes to which the WDR5 peaks are assigned. To determine which genes were associated with called peaks, these rules were followed: If a peak fell within 2kb upstream or anywhere in the open reading frame of a gene, that gene was assigned to that peak. If one peak fell within this range for more than one gene, multiple genes were assigned to that peak. Total WDR5 peaks are as follows, listed from highest to lowest: MC38: 2,870, 3T3: 2,428, LoVo: 2,162, K562: 525, Be2C: 253, and MV4:11: 158 (Fig 2.1). The number of genes associated with these peaks, as defined by nearest genes assignment, reflects a similar distribution: MC38: 2,078, 3T3: 2,229, LoVo: 1,383, K562: 648, Be2C: 330, and MV4:11: 231 (Fig 2.1). Figure 2.2 shows the heat maps of WDR5 ChIP-Seq peak intensity in each line. Yellow pixels indicate the presence of called WDR5 ChIP-Seq peak signal. For MC38, 3T3, K562, Be2C, and MV4:11 cells, signals that were present in two out of three replicates are graphed (and used for further analysis) - for LoVo, signals present in two out of two replicates are graphed. The combined average of normalized peak intensity is indicated, split into 100-bp bins, and ± 2 kb around the center of peaks is shown. Peaks are ranked based on MV4:11 cells.

Figure 2.3 shows the endogenous, steady-state levels of the WDR5 protein in all six cell lines. Importantly, the total number of WDR5 peaks called in each cell line does not correlate with the amount of WDR5 protein expressed in each. In fact, there appears to be a trend toward anti-correlation, with MV4:11 cells having the highest, and MC38 the lowest, levels of WDR5 protein. Currently, we do not have a clear understanding of how WDR5 protein expression is regulated in the cell. While WDR5 expression has been shown to be upregulated under hypoxic conditions [71], and WDR5 is upregulated by N-MYC in neuroblastoma cells [59], there is yet to be a defined, conserved mechanism for WDR5 regulation. It may prove that MYC is responsible for controlling the expression of WDR5 in all cell types, or perhaps WDR5 levels are regulated in response to signaling pathways, such as mTOR, Akt, IGF, Wnt, TGF-β, etc. Although peak number and expression levels do not correlate, total peak number does, however, correlate with the distribution of WDR5 relative to TSSs. The majority of WDR5 is TSS-proximal (within +/-500bp of the TSS; also defined as Promoter-TSS, Intron, and 5'UTR) in K562, Be2C, and MV4:11 cells, the lines with fewer WDR5 binding events, and is distributed fairly equally between 500 bp upstream and 500 bp downstream of the TSS (Fig 2.4). In MC38, 3T3, and LoVo cells, the lines with more WDR5 binding events, the majority of WDR5 is split equally between TSS-proximal and TSS-distal sites (more than +/-5,000bp away from the TSS; also defined as Intergenic). Additionally, the distribution of WDR5 binding events in MC38, 3T3, and LoVo cells between 500 bp upstream, 500 bp downstream, >5000 bp upstream, and >5000 bp downstream of the TSS is fairly equivalent. These results are graphed in Figure 2.4 - the top graph quantifies the percent of total peaks per cell line that fall within ± 5 kb of the TSS. ChIP-Seq peaks for all six cell lines are plotted according to distance from nearest TSS, binned by region: +/- 0-500, 500-1000, 1000-5000, and >5000 bp away

from TSS. On the bottom, Figure 2.4 quantifies the percent of total peaks per cell line that are assigned to the following gene elements, according to HOMER [95](command annotatePeaks): 3' UTR, 5' UTR, Exon, Intergenic, Intron, Non-coding, Promoter-TSS, and TTS (Transcription Termination Site). The majority of peaks for all six cell lines fall within three categories, Intergenic, Intron, and Promoter-TSS. Most MC38 peaks are located in Intergenic and Intronic regions, with a smaller but significant amount in the Promoter-TSS region. Most 3T3 peaks are split roughly equally between Intergenic, Intronic, ant Promoter-TSS regions. Most LoVo peaks are located in Intergenic and Promoter-TSS regions, with a smaller but significant amount in the Intronic region. Most K562 peaks are located in the Promoter-TSS region, with the remaining peaks fairly equally split between Intronic, Intergenic, 5' UTR, and Exonic regions. In both Be2C and MV4:11 cells, most peaks are located in the Promoter-TSS and Intronic regions, with the remaining peaks roughly equally split between Intergenic, 5' UTR, and Exonic regions.

Despite the differing number of WDR5 binding sites in these different cellular contexts, four recurring themes are apparent. 1) The average TSS-Proximal peak zenith across all six cell lines is just downstream of the TSS (Fig 2.5). In mouse cells, the zenith is slightly closer to the TSS center, while in human cells, the zenith is ~50 bp farther downstream (Fig 2.5). 2) Upon further inspection, it became apparent that all six cell lines exhibit a bimodal distribution, with two separate WDR5 populations, one on either side of the TSS (Fig 2.5). This distribution results from genes having either one upstream or one downstream WDR5 site - few genes show a WDR5 peak at both locations. 3) Known Motif analysis, completed on the WDR5 ChIP-Seq peak sequences from all six cell lines, shows that WDR5 most conspicuously localizes to the canonical E-Box motif (CACGTG), as well as E-Box variants (CANNTG; where 'N' can be A, T, C, or G) (Fig 2.6, 2.7). WDR5 also localizes to sequences of multiple transcription factors. 4) Functional GO analysis of WDR5-bound genes identified categories involved in translation, ribosome biogenesis, and protein synthesis to be enriched in all six cell lines (Fig 2.8).

In Figure 2.5, TSS-proximal (as defined by HOMER [95]) ChIP-Seq peaks per cell line were averaged and graphed. Averaged signal, from 2500 bp upstream to 2500 bp downstream of the TSS, is plotted, allowing for visualization of the average TSS-proximal ChIP-Seq peak shape and distribution in all six cell lines. Fragment depth was normalized to the maximum peak read for each cell line. On average, all six peaks span from ~800bp upstream of the TSS to ~700bp downstream of the TSS. The overall WDR5 peak zenith however, and therefore the location of the majority of the TSS-proximal WDR5, is downstream of the TSS. In mouse cells (MC38 and 3T3) the peak zenith is ~0-50 bp downstream of the TSS, while in human cells (LoVo, Be2C, K562, MV4:11) it is ~50-100bp downstream. The downstream TSS-proximal WDR5 peak boundary is quite sharp and is consistent across all six cell lines. The upstream

TSS-proximal WDR5 peak boundary, however, is much more variable. This is, in part, due to the two different WDR5 populations that are visible: the smaller one ~300–500 bp upstream of the TSS and the larger one one ~50–100 bp downstream of the TSS. Importantly, the majority of genes have either a TSS-proximal peak upstream of the TSS or one downstream of the TSS - very few genes display two sites of WDR5 binding.

Figures 2.6 and 2.7 display the results of known motif analysis of WDR5 ChIP-Seq peaks. HOMER [95] was used (command findMotifsGenome), with the default region size and the motif length, to determine the known-motifs present within WDR5 ChIP-Seq peak sequences. E-Boxes (CANNTG) were the top motif, the most often and consistently represented across all six cell lines. Figure 2.6 shows the top-ranked E-Box motif for each cell line. The motif sequence is shown above, with the cell line, p-value, motif name, and % of target sequences containing the motif listed beneath. The top E-Box motif identified in each cell line is as follows — MC38: 1e-53, E-Box, 4.2% - 3T3: 1e-151, E-Box, 9.7% - LoVo: 1e-264, E-Box, 14.8% - K562: 1e-75, E-Box, 20.6% - Be2C: 1e-36, bHLHE40, 29.3% - MV4:11: 1e-2, USF1, 10.8%. While the '% of target sequences contains the motif ' may seem small, there were multiple E-Box motifs called per cell line. Therefore, the total percent of target sequences containing an E-Box motif is much greater per cell line. In MC38 cells, 21 (out of 94) different identified motifs included an E-Box, each ranging from 1.57% to 52.2% of target sequences containing the motif. In 3T3 cells, 16 (out of 85) different identified motifs included an E-Box, each ranging from 7.13% to 46.99% of target sequences containing the motif. In LoVo cells, 17 (out of 67) different identified motifs included an E-Box, each ranging from 10.55% to 28.31% of target sequences containing the motif. In K562 cells, 17 (out of 33) different identified motifs included an E-Box, each ranging from 7.43% to 37.33% of target sequences containing the motif. In Be2C cells, 16 (out of 26) different identified motifs included an E-Box, each ranging from 15.42% to 35.57% of target sequences containing the motif. In MV4:11 cells, 3 (out of 6) different identified motifs included an E-Box, each ranging from 8.23% to 10.76% of target sequences containing the motif. In similar fashion, Figure 2.7 shows the top four ranked motifs for each cell line. The motif names for the top four motifs per cell line are as follows — 3T3: Bach1, NF-E2, MafF, and E-Box - LoVo: E-Box, STAT5, BMYB, and ZFX - MC38: Bach1, MafF, E-Box, and Foxo1 - K562: E-Box, CRE, E2F7, and E2F4 - Be2C: bHLHE40, CRE, E2F6, and E2F7 - MV4:11: Foxh1, USF1, E2F1, and Nanog. WDR5 binds to motifs to which known sequence-specific transcription factors bind (in data, but not all shown in Fig 2.7), including MafF, YY1, E2F1/3/4/6/7, GABPA/B, NFY, Nanog, STAT5, BMYB. WDR5 also localizes to sequences of multiple transcription factors that bind E-Boxes, including MYC, USF1/2, MITF, bHLHE40, TFE3, NPAS, and HIF-1b.

Figure 2.8 shows the Functional GO enrichment analysis of WDR5-bound genes in each cell line. DAVID was used to determine which pathways are enriched [118]. DAVID provides typical batch annotation and gene-GO term enrichment analysis to highlight the most relevant GO terms associated with a given gene list [118]. GO results are ranked based on the Modified Fisher Exact P-Value, EASE Score — the smaller the P-Value, the more enriched the gene set is in the corresponding annotation term [118]. In Figure 2.8, GO results are reported as $-\log_{10}$(P-Value) values. In the top graph, the top five conserved GO categories for each line are shown: 'Peptide Biosynthetic Process,' 'Translation,' 'Ribosome,' 'Structural Constituent of Ribosome,' and 'Cytosolic Large Ribosomal Subunit.' The numbers to the right of each bar indicate the number of genes in each category. For example, the highest ranked common GO category in all six cell types is 'Peptide Biosynthetic Process.' 82 WDR5-bound genes in MC38 cells fall into this category, 60 in 3T3 cells, 62 in LoVo cells, 106 in K562 cells, 112 in Be2C cells, and 138 in MV4:11 cells. In the bottom graph, the top eight GO categories for each line are shown, with the numbers in italics indicating the number of genes in each category.

Discussion

Thorough identification of WDR5 binding sites across the genome is critical to understanding the ways in which WDR5 functions as a cellular multitasker. While certain sites of WDR5 binding are cell-type specific, allowing for cell-type specific functions, the identification of conserved WDR5 binding sites will help us understand the broad, over-arching roles of WDR5. Acting as a scaffold, WDR5 is crucial for the function of multiple histone methyltransferase, acetyltransferase, and chromatin remodeling complexes, binds transcription factors and long non-coding RNAs, and plays critical roles in gene transcription and tumorigenesis [71]. As such, the deconvolution of the role for WDR5 on chromatin has been difficult. Recently, WDR5 has also been proposed as a novel target for cancer therapy. The motivation for targeting WDR5 is based on its over-expression in multiple [67-70, 82, 107, 119-121] cancers, its involvement in malignant processes such as the epithelial to mesenchymal transition [56] and cell motility [122], and its ties to oncogenic drivers such as MLL-fusion oncoproteins [26] and MYC [59, 60, 78]. In addition, multiple small molecule inhibitors (SMIs) have been discovered that bind to the WIN site in WDR5, one of which has been shown to remove WDR5 from chromatin [27, 74, 89, 90]. The function and effects of these inhibitors have mainly been studied in heme malignancies [26, 74, 86, 87, 89, 90], specifically those driven by MLL-fusion oncoproteins, however they have been proposed to have efficacy in a range of cancers. In order to fully comprehend the consequences of removing WDR5 from chromatin via these SMIs, we must first have a comprehensive list of sites to which WDR5 binds across cell lines. As these SMIs are refined and make their way towards clinical trials, thoroughly understanding the role of chromatin-bound WDR5 becomes

more and more critical. Again, determining the sites and genes to which WDR5 is consistently bound is of increasing importance - this knowledge can help assess what types of cancer may be susceptible to WDR5 inhibition, identify effects we can expect from WDR5 inhibition (ie: cell death, senescence, necrosis, or nucleolar stress), and establish a profile for the on-target action of WIN site inhibitors.

In this study, the total number of WDR5 binding sites varied roughly 18x from the lowest to the highest number in the six cell lines characterized. This range is in agreement with previous reports of total WDR5 peaks across multiple cell types, which include 158 WDR5 binding sites in MV4:11 cells [74], 515 in THP-1 cells [67], 833 in LNCaP cells [107], 1,490 in RS4;11 cells [67], 5,549 in BGC823 cells [84], 7,774 genes in ESCs [105], 10,915 in mouse ESCs [13], 13,075 in MLL-AF9 cells [73], 17,171 in HEK293 cells [60], and 48,269 in NHEK cells [104]. Despite this variation, total WDR5 binding sites did not correlate with WDR5 protein levels across the six cell types. Therefore, in cells where WDR5 is more abundant there are not necessarily more sites to which WDR5 binds (ie: MV4:11 cells have the highest level of endogenous WDR5 but the smallest amount of WDR5 peaks). This could indicate that in cells with higher steady-state levels of WDR5 there is more free WDR5 or a greater number of extra-chromosomal WDR5 complexes. While I can not rule out the possibility that WDR5 could serve different functions in its free form or in altered complexes when in high abundance, I can at least say that extra steady-state WDR5 does not absolutely correlate with additional WDR5 binding sites on chromatin. I conclude that while there are substantial differences in the number of WDR5 binding sites on chromatin, these differences are not driven by variations in WDR5 expression across cell lines.

Classification of WDR5 binding sites has revealed traits common to WDR5 binding sites across cell lines. In agreement with published data in certain cell lines [67, 73, 108], there is a population of WDR5 that binds TSS-distally in a subset of cell lines I investigated: MC38, 3T3, and LoVo. These three cell lines were also the three with the highest numbers of WDR5 binding sites - while the absolute number of promoter-proximal sites was generally preserved in these cells, additional sites in gene-distal (intergenic) regions appeared. This allowed me to identify a trend across all six cell lines: as the number of WDR5 binding sites increases so does the chance that these will occur in non-promoter (TSS-distal/intergenic) regions. So while there is not necessarily more steady-state WDR5 in cells with more WDR5 binding sites, there is some mechanism that causes WDR5 to increasingly bind TSS-distally. This same trend is obvious in RS4;11 leukemia cells, which over-express the WDR5 protein, but have a relatively small (~1,500) number of WDR5 binding sites [67]. While further investigation into how TSS-distal WDR5 is bound to chromatin is required to determine this mechanism, I posit that WDR5 is at enhancers when found TSS-distally, as this trend is reminiscent of enhancer invasion by MYC - a process in which MYC can be thought of as essentially saturating promoter proximal sites at normal levels of

expression, and then invading enhancers to control new gene expression patterns as its expression passes an oncogenic threshold [123]. In the case of WDR5, although its transcriptional influence could be magnified by increased enhancer binding, the number of WDR5 binding sites is disconnected from the average amount of WDR5 protein. This observation indicates that there must be some other determinant, extrinsic from WDR5, that dictates its gene-distal chromatin binding patterns. It is also provocative in light of studies showing that WDR5 is overexpressed in a variety of cancers [67-70, 82, 107, 119-121], where its over-expression often correlates with poor clinical outcomes. If levels of WDR5 do not predict either the number of sites bound by WDR5, nor the relative level of WDR5 at those sites (compare peak-read graphs in Figure 2.5), the impact of enhanced WDR5 expression in cancer cells is unlikely to manifest itself at WDR5-bound target genes. Perhaps WDR5 over-expression in cancer is irrelevant to the malignant state, or perhaps its impact is on events that do not to require stable association of WDR5 with chromatin, such as H3K4 methylation [74] or non-transcriptional "moonlighting" processes [2].

Across all six cell lines interrogated, however, the majority of WDR5 was TSS-proximal. These results are in accordance with previous characterization of WDR5 as a TSS-proximal binding protein [50, 67, 73, 74, 84, 105, 109]. This study allowed me to paint a portrait of a prototypical "universal" WDR5-bound gene. Key characteristics of this prototype are a high WDR5 signal by ChIP, localization of WDR5 to sequences immediately downstream of the TSS, and of course a connection to protein synthesis, most commonly by encoding a subunit of the ribosome. It is impossible to know if the high ChIP signal of WDR5 at RPGs reflects increased occupancy, increased stoichiometry, or increased epitope accessibility. When bound TSS-proximally, WDR5 bound to regions of open chromatin (‘open’ as defined by [124]) enriched in transcription factor motifs, including those for MYC, E2F1/4/6/7, Bach1, STAT5, BMYB, YY1, MafF, and Nanog (Fig 2.7). As WDR5 is known to facilitate the formation of multi-protein complexes, I posit that WDR5 acts as a scaffold for various TF complexes at TSS-proximal sites upstream of the TSS. When localized downstream of the TSS, I propose that WDR5 is scaffolding transcriptional elongation complexes or intronic enhancer complexes. The multi-pocketed nature of WDR5 [16, 17, 60] may allow WDR5 to act as a scaffold for these differing complexes. For example, a sequence-specific TF could bring WDR5 to chromatin by binding to the WIN site - then, using the WBM site, WDR5 could recruit other proteins to chromatin. More specifically, we have shown that WDR5 recruits MYC to chromatin via the WBM site [60]. We do not yet know if or how WDR5 is getting to chromatin while in complex with MYC, but we do know that the WIN site is available and is structurally unaltered by the binding of MYC to WDR5 [60]. Thus, another sequence-specific TF would have access to the WIN site, could recruit WDR5 to chromatin, then subsequently, WDR5 could use the WBM to recruit MYC.

Newly identified proteins that bind to WDR5 are constantly being published, and thus, it is not unlikely that there are various WDR5-containing complexes yet to be discovered. Below, I outline nine potential TF complexes that, based on motif analysis, may be present in the cell, bound to chromatin at WDR5-binding sides, and include WDR5.

1) WDR5-MOF-Nanog: I determined that WDR5 binds to the Nanog motif in K562, MV4:11, and MC38 cells. WDR5 has been shown to associate TSS-proximally with MOF [34, 35, 103], while MOF has been shown to associate with Nanog at ~80% of Nanog-target gene promoters [103]. Further, it has been suggested that MOF may serve as a coactivator of Nanog-mediated transcription [103]. Therefore, a complex could exist in which WDR5, MOF, and Nanog associate with each other and WDR5 helps regulate the transcription of Nanog-target genes.
2) Altered DREAM: The DREAM complex includes both BMYB and E2F4 and coordinates the expression of cell cycle-dependent G2/M genes [125]. I determined that WDR5 peaks include motifs for both BMYB and E2F4 - LoVo and MC38 cells had the BYMB motif and LoVo, Be2C, K562, 3T3, and MC38 cells had the E2F4 motif. While WDR5 is not identified as a component of the canonical DREAM complex, an altered DREAM complex may exist, in which WDR5 helps coordinate the expression of specific subset of cell cycle-dependent genes.
3) p300-BMYB-MLL-WDR5: the KIX domain in p300 has been shown to interact with both MYB proteins and MLL simultaneously, in a complex that is important for oncogenesis [126]. Each MYB protein regulates a distinct set of genes, and notably, BMYB regulates the CCND1 gene, which encodes the G1/S regulator Cyclin D1 [126]. As I identified the BMYB motif in WDR5 peak sequences, BMYB could be the link through which p300 could bind chromatin, recruit MLL, and therefore WDR5, to form a novel complex in which WDR5 contributes to the control of the G1/S transition in LoVo and MC38 cells.
4) YY1-CTCF-WDR5: I identified the YY1 binding motif within the WDR5 peak sequences in 3T3, LoVo, and K562 cells. YY1 binds to the N-terminus of CTCF and together they impact chromosomal looping and epigenetic regulation, including H3K27Me3 [127]. WDR5 has also been shown to associates with CTCF [127]. As WDR5 has already been identified to interact with one polycomb complex [43, 44], and polycomb complexes are known to aid in the trimethylation of H3K27, it would not be surprising to find WDR5 in another complex that also trimethylates H3K27. As CTCF and YY1 are known to be involved in chromosomal looping, it is also possible that some of the TSS-distal WDR5 population could bind to chromatin through CTCF then YY1, at distal YY1 motifs, and help regulate looping and the resulting transcription. Furthermore, the YY1 motif has been reported to be present within intronic enhancers [128, 129] - perhaps WDR5 is in complex with YY1 when found downstream of the TSS, as part of an intronic enhancer complex.

5) YY1-CTCF-CHD8-WDR5: Another potential complex that may be tethered to chromatin via YY1 could include: CTCF, as YY1 binds the N-terminus of CTCF [127], CHD8, as CHD8 binds the DNA-binding domain of CTCF [127], and WDR5, as WDR5 can bind CHD8 [45, 46]. In Molecular and Cellular Biology (2008), Thompson, et. al. show that WDR5 and CHD8 directly interact, without the presence of MLL peptides [45]. They therefore suggest that WDR5-CHD8 may be present outside of the MLL complex [45]. Perhaps the complex WDR5 and CHD8 are in includes CTCF and YY1.
6) E2F1-MLL-Set1-WDR5-HCF-1: The E2F1 motif was identified in WDR5 binding sites in K562, Be2C, MV4:11, and MC38 cells. E2F1 is already known to recruit a complex that includes MLL, Set1, HCF-1, and WDR5, to activate the expression of its target genes in S phase [130]. Perhaps this complex is present in K562, Be2C, MV4:11, and MC38 cells at sites of WDR5 binding that include the E2F1 motif.
7) As part of PRC1.6: I identified the E2F6 motif inWDR5 binding sites in 3T3, MC38, LoVo, Be2C, and K562 cells. WDR5 has been reported to interact with E2F6 in the PRC1.6 methyltransferase complex [34, 42-44]. Perhaps this complex is present in 3T3, MC38, LoVo, Be2C, and K562 cells at sites of WDR5 binding that include the E2F6 motif.
8) As part of the NSL complex: WDR5 has previously been identified as an important member of the NSL complex, in which the KANSL1-WDR5 interaction appears to be required for efficient recruitment of the NSL complex to chromatin, and therefore the acetylation of H4K16 [30-33]. In the NSL complex, WDR5 associates with MOF and E2F6, binds KANSL1 via the WIN site, and binds KANSL2 via the WBM site. As I identified the E2F6 motif in WDR5 binding sites in 3T3, MC38, LoVo, Be2C, and K562 cells, it is possible that the NSL complex is present in these cells at sites of WDR5 binding that include the E2F6 motif.
9) WDR5-KLF3: The KLF3 motif was identified in WDR5 binding sites in LoVo, 3T3, and MC38 cells. WDR5 reportedly directly binds to KLF3, and is proposed to allow the functional domain of KLF3 to specify target gene selection [131]. KLF3 is known to bind to promoters, enhancers, and other control regions of target genes, function as a transcriptional repressor of genes involved in adipogenesis, erythropoiesis, and B cell development [132]. WDR5 may directly bind KLF3 in LoVo, 3T3, and MC38 cells and help regulate KLF3 target genes.

The genes to which WDR5 binds significantly cluster into common GO categories, including different steps in translation and RNA processing, localization of proteins to membranes, gene expression, and ribosome biogenesis. These categories are not only consistently found across all six cell lines, but are also the most enriched in every cell line. This conservation, across cell types and evolution, indicates the importance of WDR5 binding to genes involved in these processes. In light of these results, I posit that the overarching role for WDR5, across evolution, is to regulate the

expression of genes involved in translation, RNA processing, and ribosome biosynthesis. This hypothesis will be further addressed in the following chapters - chapter IV will discuss the specific genes that are consistently bound by WDR5, and Chapters V and VI will identify the genes that are regulated by WDR5.

Overall, this chapter identifies and characterizes WDR5 binding sites across MC38, 3T3, LoVo, Be2C, K562, and MV4:11 cells. This analysis indicates that as the total number of WDR5 binding sites increase, WDR5 is progressively found at more TSS-distal sites. WDR5 is consistently bound TSS-proximally, to the region of genes that is involved in transcriptional regulation, across all six cell types. The genes to which WDR5 is bound in all six cells play a role in translation, RNA processing, and ribosome biosynthesis.

	MC38	3T3	LoVo	K562	Be2C	MV4:11
Peaks	2,870	2,428	2,162	525	253	158
Genes	2,078	2,229	1,383	648	330	231

Figure 2.1. Results of ChIP-Seq analyses of WDR5 in six cell lines, showing (*top ro*w) the number of WDR5 peaks, and (*bottom row*) the number of genes assigned to those peaks in each line.

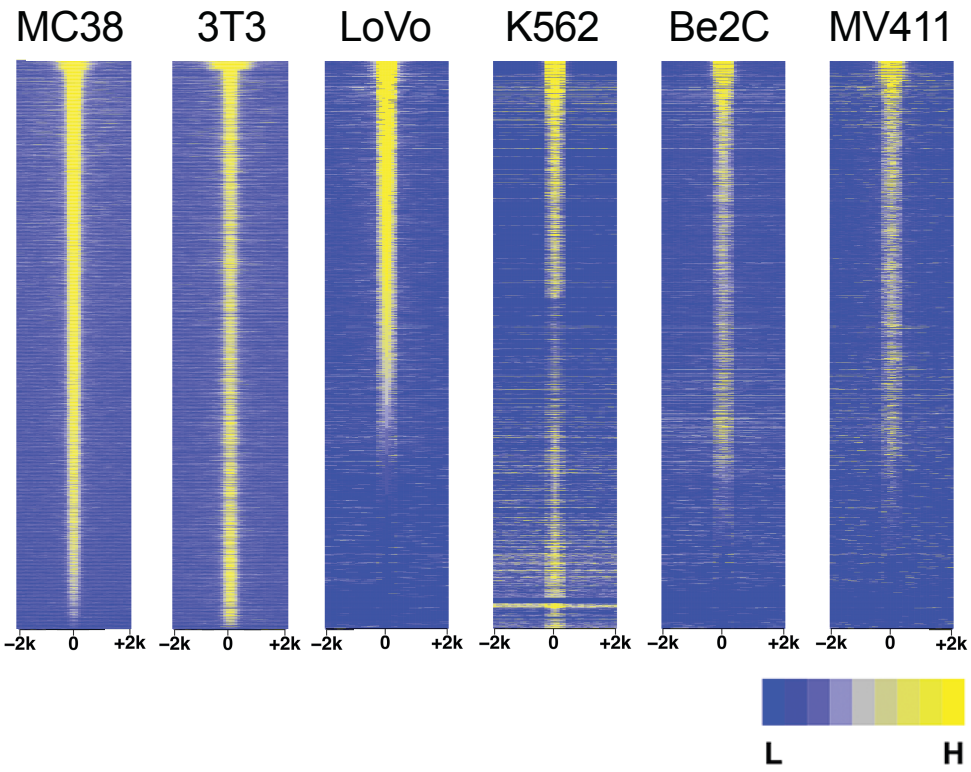


Figure 2.2. Heat maps of WDR5 ChIP-Seq peak intensity in each line. The figure represents the combined average of normalized peak intensity in 100-bp bins ± 2 kb around the center of peaks. Peaks are ranked based on MV4:11 cells.

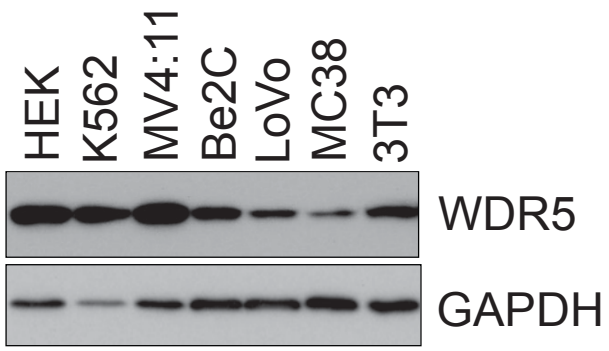


Figure 2.3. Immunoblotting of steady-state WDR5 levels in the indicated cell lines. GAPDH is a loading control.

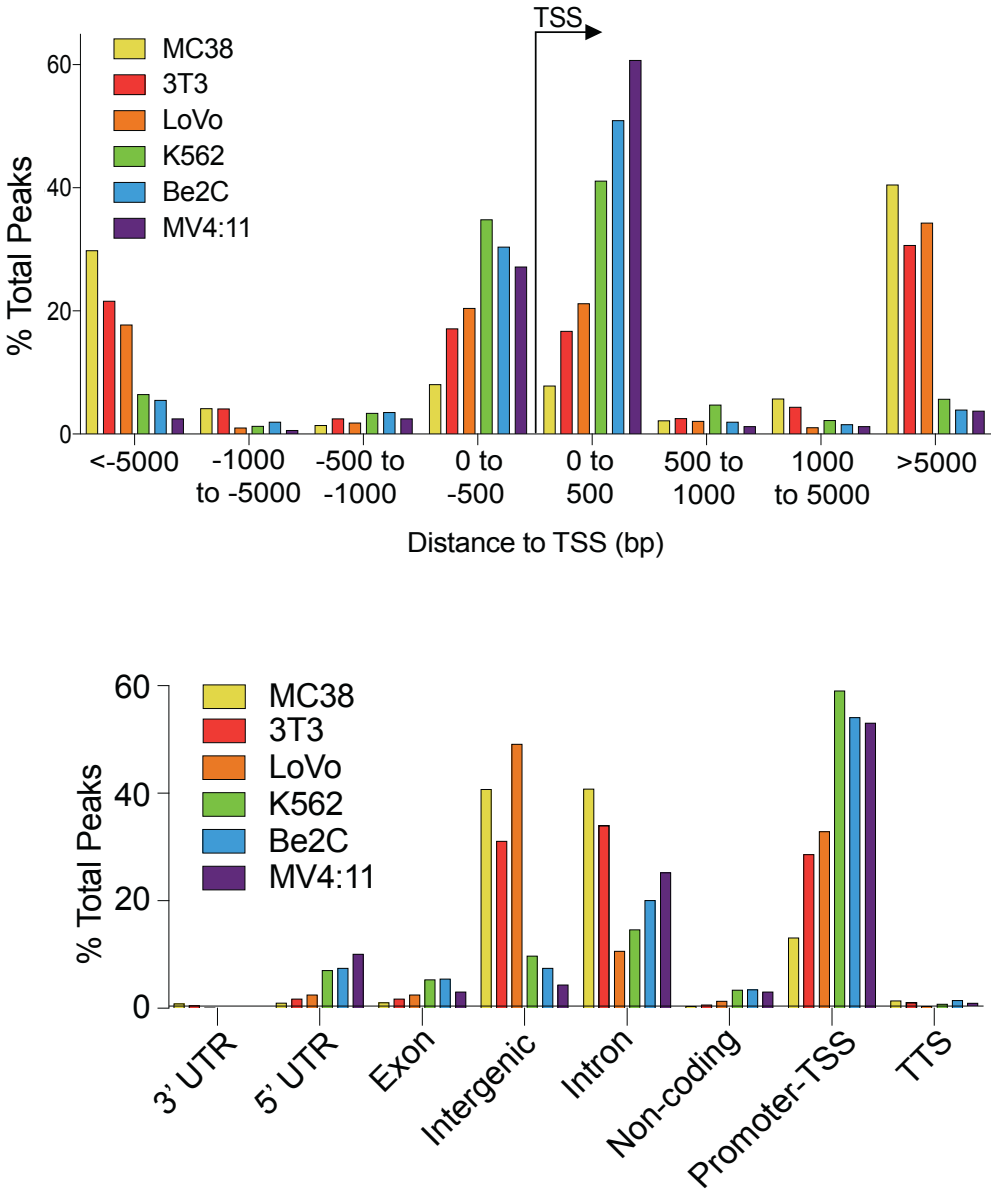


Figure 2.4. (*Top*) Distribution of WDR5 binding sites, relative to annotated TSS. ChIP-Seq peaks for all six cell lines are plotted according to distance from nearest TSS, binned by region: +/- 0-500, 500-1000, 1000-5000, and >5000 bp away from TSS. (*Bottom*) Assignment of WDR5 binding sites to functional gene elements, according to HOMER. Graph shows, for each cell line, the percentage of peaks that fall into each of the designated categories. TSS: Transcription Start Site; TTS: Transcription Termination Site.

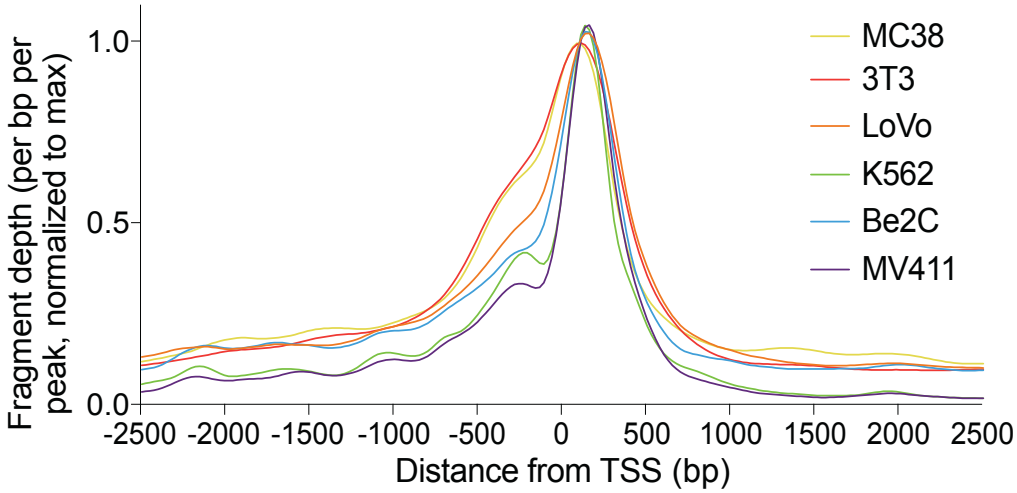


Figure 2.5. Averaged TSS-proximal ChIP-Seq peak shape and distribution in all six cell lines. Fragment depth is normalized to the maximum peak read for each cell line.

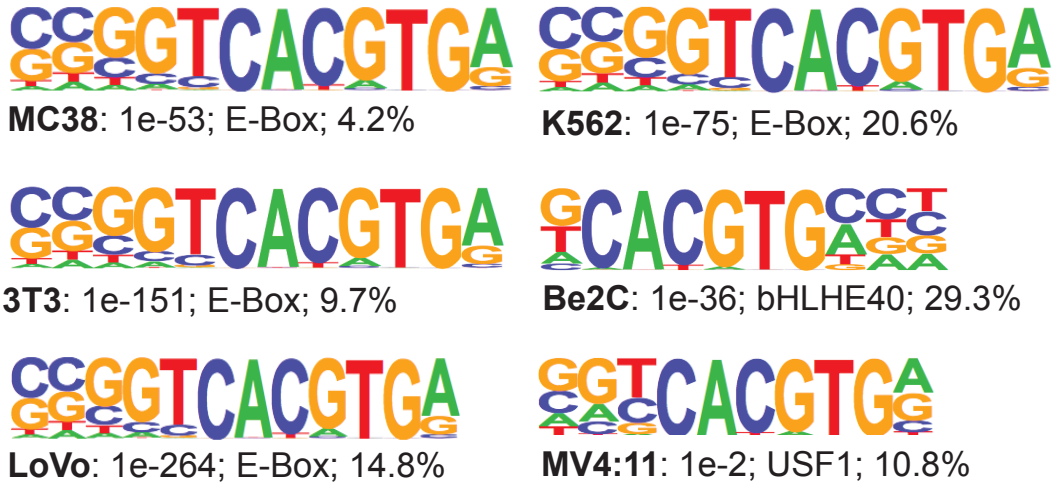


Figure 2.6. Results of known-motif analysis of WDR5 ChIP-Seq peaks, showing the top-ranked E-box motif for each line. p-value, motif name, and % of target sequences containing the motif are listed beneath each motif sequence.

Figure 2.7. Results of known-motif analysis of WDR5 ChIP-Seq peaks, showing the top four ranked motifs for each line. p-value, motif name, and % of target sequences containing the motif are listed beneath each motif sequence.

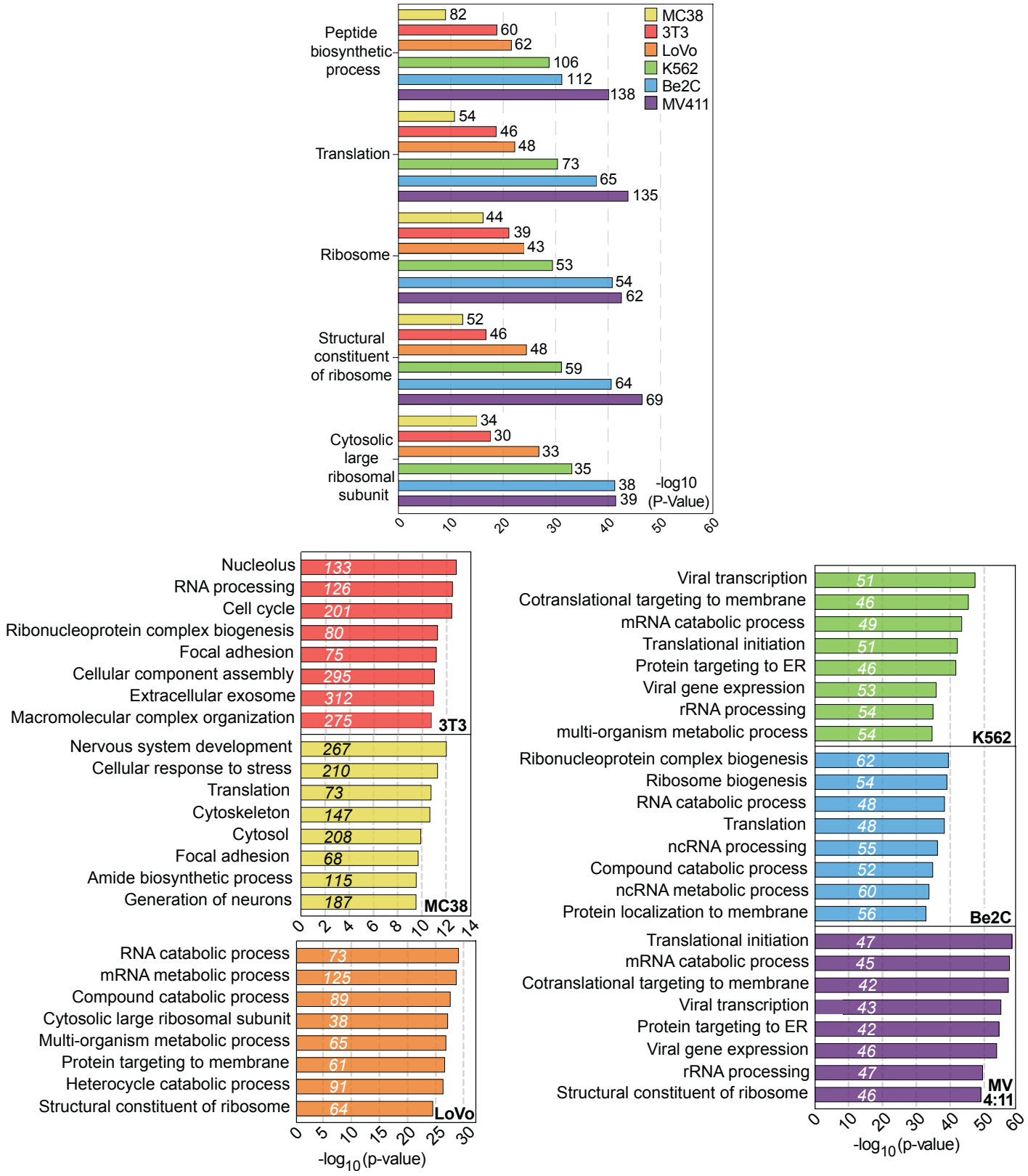


Figure 2.8. (*Top*) GO enrichment analysis of WDR5-bound genes in each cell line. The top five conserved GO categories for each line are shown; numbers to the right of each bar indicate the number of genes in each category.
(*Bottom*) GO enrichment analysis of WDR5-bound genes in each cell line. The top eight GO categories for each line are shown; numbers in italics indicate the number of genes in each category.

CHAPTER IV

Conservation of WDR5 Binding to Chromatin in Disparate Cell Types

Abstract

Deciphering and de-convoluting the role of chromatin-bound WDR5 has been challenging, due to the various nuclear complexes in which WDR5 is a core component, including, but not limited to, histone writer, reader, eraser, and remodeling complexes, such as the NSL, NuRD, ATAC, and MLL complexes. Further, many of these chromatin-modifying complexes, while known to associate with chromatin, may do so in a fleeting, unstable manner. In order to determine the overarching, evolutionarily conserved role for stable (ChIP-able), chromatin-bound WDR5, we must compare WDR5-bound loci across multiple cell lines and species. In this chapter, I will identify the conserved sites of WDR5 binding, and the conserved genes to which WDR5 binds. The ChIP-Seq experiments, performed on two mouse cell lines and four human cell lines and introduced in Chapter III, were further analyzed and used to determine where WDR5 binds regardless of cell line, type, or species interrogated. Together these data reveal a common set of WDR5-bound genes in human cells and mouse cells. 94 genes were identified as evolutionarily conserved sites of WDR5-binding - these genes are enriched in translation, ribosome biogenesis, and RNA processing, and over half are ribosome protein genes.

Introduction

In Chapter III, the general characteristics of WDR5 binding to chromatin across all six cell types were established. A prototypical WDR5-target gene was defined and a common set of GO categories was identified. These results raised the question, does WDR5 bind to different genes that all fall within the same GO category, or is WDR5 binding the same genes across cell types? In this chapter, I address the question of whether these common characteristics derive from a conserved set of WDR5-bound loci.

In order to determine the evolutionarily conserved role for chromatin-bound WDR5, we must first identify the genes to which WDR5 stably binds across multiple cell lines and species. While the genes to which WDR5 binds in certain cellular contexts have been evaluated, as of yet there has been no systemic identification of conserved WDR5-bound genes, regardless of cell type. The following includes the results from publications in which specific WDR5-bound genes have been identified - I propose that many of those focused on in these studies are cell-type specific. In *Drosophila*,

WDR5 is proposed to be important for the efficient targeting of the NSL complex to target genes [133]. Using KANSL1 mutant flies, the authors show that the KANSL1–WDR5 interaction is required for proper NSL complex assembly, fly viability, and efficient recruitment of the NSL complex to target promoters, including Sec5, CG15011, Patj, tho2, CG6506, Act57B, CG5992, and OdsH [133]. In ES cells, WDR5 is proposed to interact with HOTTIP and NeST to target the MLL complex to MLL-target genes [51, 52, 134], and has ben shown to localize to Polycomb group ring finger protein promoter regions [109] and to overlap with MOF at MOF-target genes [103]. WDR5 has also been shown to associate with Oct4 at Oct4-target genes, localize to several key epigenetic factor genes such as *Baf155*, *Brg-1*, and *Sall4*, and bind to genes involved in cell and organ developmental process, embryogenesis, neurogenesis, and cell differentiation in ES cells [13, 105]. Questionably, in distally derived cells, WDR5 was found to bind with HOTTIP across the *HOXA* locus [51]. In differentiated keratinocytes, WDR5 was shown to directly interact with GRHL3 at GRHL3-target genes involved in cell differentiation, gene expression, programmed cell death, cell-cell adhesion, and lipid biosynthetic processes, and GRHL3 was proposed to recruit WDR5 to regulatory regions of epidermal differentiation genes in differentiated NHEK cells [104]. In C2C12 myoblasts, WDR5 was shown to bind to muscle genes, *Myog*, *Acta1*, and *Mybph*, co-localize with H3K4me3, Sin3A, ING1, LSD1, MLL1, and Menin at the majority of active genes, and bind to a portion of the *Hoxa* genes [106]. In LNCaP cells, WDR5 has been proposed to interact with H3T11P to facilitate the recruitment of the MLL1 complex to MLL-target genes, including *IGF2R* and *RPL13A* [107]. WDR5 was found to localize to activated EWS-FLI1 sites in SKNMC and A673 Ewing sarcoma cells [108]. WDR5 binds to c-MYC target genes in HEK293 cells [60] and to N-MYC target genes, including *MDM2* and *cyclin E1*, in Be2C cells [59]. In primary murine MLL-AF9 cells MLL1-WDR5 direct target genes were proposed to include those involved in cell signaling, transcription, hypoxia, hematopoiesis and myeloid differentiation, as determined by GO analysis [73]. In both human RS4;11 ALL and THP-1 AML leukemia cells WDR5 has been shown to bind to genes involved in multi-oncogenic signaling, apoptosis, transcriptional regulation, histone modification, cell proliferation and apoptosis, cell adhesion and metabolism [67]. In Nalm6 B-ALL, U937 AML, and primary B-ALL and AML cells, WDR5 was found at *Lyn*, *BCL9, RAb28, CD93, MED24*, and *RBM22* [67]. WDR5 has been shown to bind a portion of KAT2A target genes in BGC823 cells [84]. In MV4:11 cells, WDR5 binds a specific subset of translation initiation factors and ribosome protein genes, corresponding to genes encoding ~40% of the small and ~70% of the large ribosome subunit proteins, and [74]. While the genes to which WDR5 binds in disparate cell types has clearly been of interest in the field, there has been no systematic identification of WDR5-bound genes, regardless of cell type, with the same reagents and antibody. Therefore, I have completed such analysis, and will detail in this chapter, the genes to which WDR5 is bound across two species, six cell

lines, and five cancer types. Due to the results discussed below, I believe the previously identified WDR5-bound genes and gene-sets in ES, distally derived, differentiated keratinocytes, NHEK, C2C12, LNCaP, SKNMC, A673, HEK293, Be2C, MLL-AF9, RS4;11, THP-1, Nalm6, U937, and BGC823 cells that have been focused on (outlined above) are largely cell-type specific, while the conserved WDR5-bound genes are those involved in translation, RNA processing, and ribosome biogenesis.

Results

In order to determine the conserved sites of WDR5 binding across the genome, I utilized the ChIP-Seq experiments described in Chapter III to identify sites to which WDR5 was bound in six cell lines. Due to genome sequence differences, I first compared human and mouse data individually, and then overlaid shared gene assignments. In Figure 3.1, two Venn diagrams indicate the shared WDR5 binding sites across the four human and the two mouse cell lines. On the left, Be2C, LoVo, MV4:11, and K562 peaks/genes are overlapped, identifying 103 common genes to which WDR5 binds across all four human cell lines (Fig 3.1). On the right, 3T3 and MC38 peaks/genes are overlapped, identifying 1,161 common genes to which WDR5 binds across both mouse cell lines (Fig 3.1). To make these graphs, WDR5 peaks were first assigned to genes. To determine which genes were associated with called peaks, these rules were followed: If a peak fell within 2kb upstream or anywhere in the open reading frame of a gene, that gene was assigned to that peak. If one peak fell within this range for more than one gene, multiple genes were assigned to that peak. The assigned gene names were used to create the Venn diagram.

Figure 3.2 (*top*) includes four graphs. The peaks corresponding to the 103 common human genes are plotted in each of the human cell lines. Below, one graph shows the peaks corresponding to the 1,161 common mouse cell lines, with both cell lines plotted on the same graph. As the common peaks were TSS-proximal, all graphs indicate WDR5 peak signal from 2,500 bp upstream to 2,500 bp downstream of the TSS. Across cell lines and species, the general profiles of these common peaks are highly similar, with particularly sharp definition at sites immediately downstream of the TSS, indicating an almost invariant positioning of WDR5 at 3' TSS-proximal locations. More variation is apparent at 5' TSS-proximal sites. Figure 3.3 (*top*) includes four graphs - each contains all the called WDR5 peaks per human cell line indicated, plotted based on peak read intensity. Below these curves, in navy, the peaks that are assigned to the 103 common human genes are indicated. Figure 3.3 (*bottom*) includes two graphs - each contains the called WDR5 peaks per mouse cell line indicated, plotted based on peak read intensity. Below these curves, in navy, the peaks that are assigned to the 1,161 common mouse genes are indicated. Below these curves, in red, the peaks that are assigned to the 94 conserved

human and mouse genes (see Fig 3.4) are indicated. Comparing the intensity of these common peaks among all sites bound by WDR5 in any particular line, I observed that, in human cells, common WDR5 peaks generally have the highest amount of WDR5 signal (Fig 3.3, *top,* navy), whereas in the two mouse lines there is no correlation between common gene identity and intensity (Fig 3.3, *bottom,* navy), as to be expected from the similar number of WDR5 binding events in these lines (Fig 2.1). However, when the peaks corresponding to the 94 conserved genes were identified in the mouse cell lines, I observed that conserved WDR5 peaks generally have the highest amount of WDR5 signal (Fig 3.3, *bottom,* red).

Overlaying all six sets of gene assignments, I identified 94 conserved WDR5-bound genes (Fig 3.4 and 3.5), which, due to multiple gene assignment, correspond to 74 unique human and 76 unique mouse WDR5-binding sites. Figure 3.4 includes two Venn diagrams. The top shows the overlap between the 103 common human genes and the 1,161 common mouse genes to which WDR5 binds. The bottom shows the overlap between all human and mouse genes to which WDR5 binds, indicates how many WDR5-bound genes are unique to each cell line, and shows all other combinations. While there are 94 genes to which WDR5 binds regardless of cell line interrogated, there are 144 genes in K562 cells, 38 in Be2C, 674 in LoVo, 17 in MV4:11, 892 in 3T3, and 845 genes in MC38 cells to which WDR5 binds only in that cell type. These numbers correlate with the total number of WDR5 binding events in these cell lines (Fig 2.1). As is further discussed in Figure 3.8, the 94 conserved genes significantly cluster into common GO categories, while the cell-type-specific genes do not significantly cluster into any meaningful GO categories. Figure 3.5 lists the 94 conserved genes, indicates their location in the human and mouse genome, and includes a description of the function of each gene. The listed genes are grouped by common function.

In Figure 3.6, the peaks assigned to the 94 conserved genes are averaged and graphed, per cell line. As all 94 sites are TSS-proximal, peak reads from 1,500 bp upstream to 1,500 bp downstream of the TSS are plotted. The fragment depth is plotted per base pair per peak, normalized to the max peak read signal per cell line. There is little variation in the upstream and downstream peak boundaries, and at these conserved genes, the WDR5 peak zenith is just downstream of the TSS (Fig 3.6). There is no common DNA motif shared amongst all these conserved binding sites, and enrichments detected by HOMER [95] using known motif analysis (with the default region size and the motif length) were modest, likely due to the small number of sequences involved. However, since the E-Box motif was so highly represented in WDR5 binding sites across all six cell lines (Figures 2.7 and 2.8), I investigated the presence of E-Boxes within the sites assigned to the 94 conserved genes. Figure 3.7 shows how many canonical (CACGTG) and non-canonical (CANNTG) E-Boxes were present in the 74 human or 76 mouse WDR5 binding sites, corresponding to the 94 conserved genes. Almost all conserved genes contained at least one E-Box, and roughly one-third of the conserved genes contained a canonical E-

Box. In fact, 68 of the 74 human sites contained an E-Box (any CANNTG), 22 of which were perfect (CACGTG; Fig 3.7, *left*); 72 of the 76 mouse sites contained an E-Box (any CANNTG), 25 of which were perfect (CACGTG; Fig 3.7, *right*). Interestingly, these 94 conserved genes are sites of high WDR5 intensity in both human and mouse cells (Fig 3.3, red lines). Figure 3.10 also shows a representative subset of WDR5-bound RPGs, with all E-Boxes within the WDR5 peaks indicated by type (color coded) and location.

Functional GO analysis, performed using DAVID [118], revealed that these 94 conserved genes are highly enriched in those connected to protein synthesis, RNA processing, protein targeting, and translation as is indicated in Figure 3.8. In fact, the top 12 GO categories include 'structural constituent of ribosome,' 'nuclear transcribed mRNA catabolic process,' 'rRNA processing,''cotranslational protein targeting to membrane,' 'protein targeting to ER,' 'translation,' and 'translational initiation.' Further analysis of the specific genes that made up these top-ranked categories revealed that most include ribosome protein genes (RPGs). In fact, roughly half of the common WDR5 binding sites occur within the same subset of small and large subunit RPGs, as is indicated in Figure 3.9. These RPGs (Fig 3.9) are also bound by WDR5 in published ChIP-Seq data from prostate [107] and gastric [84] cancer cell lines. Notably these RPGs, WDR5 peaks span a significant portion, if not all, of the first exon and intron. Figure 3.10 shows a representative subset of WDR5-bound RPGs - the gene structure of *RPL*s bound by WDR5 are shown, with the detected WDR5 peaks noted above in green bars. WDR5 peaks are shown for two human cell lines, Be2C and LoVo, and the two mouse cell lines, MC38 and 3T3.

Taken together, these data identify a set of 94 broadly-conserved WDR5-bound genes, reveal that the conserved WDR5-binding loci are typically high intensity sites located immediately downstream of the TSS, and determine that the conserved WDR5-bound genes are overtly linked to protein synthesis.

Discussion

In order to begin to determine the overarching, evolutionarily conserved role for chromatin-bound WDR5, I have compared stable WDR5 bound-loci across multiple cell lines and species. Due to genome sequence differences, I began by identifying the common WDR5-bound genes in human cell lines (103 genes) and mouse cell lines (1,161 genes) separately (Fig 3.1). While I identified significantly more WDR5 common peaks in mouse cells, I contributed this to two main factors: 1) There were more total WDR5 peaks in MC38 and 3T3 cell lines than in each of the human cell lines, and 2) I only compared two mouse cell lines, while I compared four human cell lines, therefore increasing the chance that any mouse gene would also be bound by WDR5 in only one other cell line. Importantly, when the peaks assigned to the 103

common human genes and the 1,161 common mouse genes were averaged and plotted (Fig 3.2), I was able to determine that all six common WDR5 peak distribution curves looked the same. This indicates that WDR5 is binding to the same location within the common genes, in all cell types examined. Additionally, as the shape of the common WDR5 peak distribution curves were highly similar, and differences in peak distribution curves can be indicative of differences in bound complexes [135, 136], is it likely WDR5 is in the same complex in all cell lines examined. This further provides evidence for the conserved biological role and mechanistic function of WDR5 across cell/cancer types and species.

Across all six cell lines, five cancer cell types, and two species, I identified 94 genes to which WDR5 stably binds (Fig 3.4). Many of these genes are RPGs, snoRN(A/D)s, and translation initiation factors. When plotted according to peak intensity, the majority of common and conserved WDR5 binding sites were determined to be those of high intensity (Fig 3.3). While I can not fully exclude the possibility that I am preferentially picking up the ChIP signal from the high-intensity WDR5 binding sites, I do show that some of the conserved binding sites are of low-intensity, which argues against this possibility. Also, while they are not all conserved, there are a significant number of cell-line specific low-intensity WDR5 peaks in each cell line. Furthermore, in independent ChIP-qPCR replicates, I determined that the pattern of WDR5 binding - the presence of particular WDR5 peaks in certain cell lines, and the lack of the same peak in other cell lines - remains true. This indicates that the variations in WDR5 peak totals are not simply caused by different ChIP-efficiencies in the individual experiments, and the cell line-dependent differences in total WDR5 binding is real.

I also determined that the E-Box motif is highly represented in conserved WDR5 binding sites (Fig 3.7). WDR5 has no known DNA binding domain and we currently do now know how WDR5 is getting to its target genes in chromatin. Therefore, one potential mechanism for WDR5 localization on chromatin is WDR5 in complex with other E-Box binding transcription factors, which are allowing WDR5 to localize to the majority of these 94 conserved genes. A slew of transcription factors are known to bind to the E-Box motif, while in complex with other proteins. For example, CLOCK, BMAL, and NPAS2, circadian rhythm proteins [137, 138], bHLHE40, a circadian rhythm, myogenesis, and hypoxia responsive protein [139], MyoD and MyoG, myogenesis proteins [140, 141], TCF3/E47, a tissue-specific and differentiation protein [142], USF1, USF2, and TFE-3 [143-145], MITF, a cellular differentiation, proliferation, and survival protein [145], MAX, a transcriptional activator and repressor, [145], and HIF1A and ARNT, hypoxia responsive proteins [146, 147], are all known to bind the E-Box motif in chromatin. As of yet, WDR5 has not been shown to bind any of these transcription factors, so while I can not rule out this possible mechanism, I believe there is better evidence for WDR5 being bound to chromatin through interactions between its WIN site and the other proteins, as I discussed in Chapter III.

Alternatively, I propose that the E-Box motif is present in these sites because chromatin-bound WDR5 recruits and stabilizes E-Box-specific transcription factors through its WBM. For example, WDR5 is known to use the WBM to directly bind and recruit MYC to chromatin, the oncogene transcription factor that preferentially localizes to E-Box motifs in chromatin [59, 60]. It is possible that WDR5 is recruiting MYC to these 94 specific sites in chromatin, and helping MYC carry out its role, regulating the transcription of target genes. This will be further explored in Chapters VII and VIII. WDR5 could also recruit the other aforementioned E-Box specific transcription factors to the 94 conserved genes under specific conditions: 1) CLOCK, BMAL, and/or NPAS2 during the different phases of the circadian rhythm cycle, 2) bHLHE40 during the circadian rhythm cycle, muscle differentiation, and/or in response to hypoxia, 3) MyoD or MyoG during muscle differentiation, 4) TCF3/E47 in a tissue-specific manner and during certain stages of differentiation, 5) MITF during different stages of differentiation, the cell cycle, or mitosis, 6) MAX differentially in both a repressive and activating manner, or 7) HIF1A or ARNT in response to hypoxia. As the levels of WDR5 are known to be highest during initial stages of development [2, 10, 12, 13], it would not be surprising to find that WDR5 recruits all these proteins to its conserved targets (or potentially different subsets of conserved targets), each during a different stage of development.

GO analysis on the 94 conserved genes confirmed the findings from Chapter III, that WDR5 consistently binds to genes involved in ribosome biogenesis, RNA processing, and translation (Fig 3.8). While there are genes that do not fall within these categories (see Fig 3.5), there was an obvious, significant enrichment in conserved WDR5 binding sites occurring at RPGs (Figures 3.5 and 3.9). The links between WDR5 and protein synthesis genes, particularly RPGs, are profound and intriguing. One particularly notable feature is the location of WDR5 binding at these sites (Fig 3.10): The region bound by WDR5 is downstream of the TSS, situated within the crucial +1 nucleosome [148] - the dynamics and composition of which can profoundly impact transcriptional processes [149] - and is contained entirely within transcribed intronic sequences. The location of WDR5 at these sites is thus akin to enhancer location, leading us to speculate that these elements may be intronic enhancers [150]. Indeed, several intronic enhancers have been described within mammalian RPGs [128, 129, 151, 152], the boundaries of which encompass the conserved WDR5 binding sites defined here. One possibility is that WDR5 acts as part of a rheostat at these intronic enhancers, fine-tuning the transcription of select RPGs according to cellular needs.

Why some RPGs would use WDR5 for this purpose, and others not, is unclear, but it could reflect specific aspects of the transcriptional and post-transcriptional control of individual subunits. For example, the transcription of certain RPGs may require an extra boost of transcription factor stability (which could be offered by WDR5) to allow for appropriate levels of RPG mRNA to be synthesized. It is also possible that certain RPG mRNA is less stable, and

therefore needs to be synthesized at different rates under certain cellular conditions to maintain the correct level of RPG mRNA - WDR5 may respond to some stimulus and bind more/less at RPGs to help regulate the concentration of RPG mRNA in a cell. In addition, ribosome construction, from many small subunits, provides a challenge to cells, which must keep many independent RPs in a particular stoichiometry. The specific set of WDR5-bound RPGs could reflect this cellular necessity to balance overall ribosome subunit protein levels. The need for well-balanced ribosome subunits has been well established, as ribosome deficiency results in strong cellular effects - in humans, ribosome deficiencies or mutations can lead to ribosomopathies [153], tissue-specific defects [153], altered gene expression signatures [154], accumulation of ribosome-free ribosomal proteins [155], and nucleolar stress [156]. In fact, I have shown that WDR5 binds to the loci of multiple RPGs known to play a role in the cellular response to nucleolar stress, including RPL11, RPL5, RPS7, RPL23, RPS3, and RPL26 (Fig 3.9). These specific RPGs have been shown to play a role in both the p53-dependent and the p53-independent response pathways to nucleolar stress. RPL11 and RPL5 have been found to be essential for p53 up-regulation in response to nucleolar stress [157], while RPS7 and RPL23 have been shown to bind and inhibit the E3 ligase activity of Hdm2 toward p53, but are not required for the p53 response [158]. RPS3 can bind to p53 or MDM2 and RPL26 binds to p53 mRNA to enhance p53 translation [157]. Alternatively, upon nucleolar stress in cells with inactive p53, free RPL11 can bind to MDM2, causing the release of E2F-1 and its subsequent degradation, thereby inhibiting cell proliferation [157]. RPS14 and RPL11 also suppress cell proliferation in a p53-independent pathway by negatively regulating c-MYC transcriptional activity [157].

It has also been shown that small and large ribosomal subunits depend on each other for stability and accumulation [154]. Small versus large ribosomal subunit deficiencies cause distinct phenotypes and gene expression signatures — for example, studies using RPG mutants have shown that mutations in RPLs lead to an up-regulation of protein-mediated protein catabolism, while mutations in RPSs lead to an increase in the translation of genes involved in ribosome biogenesis [154]. As I have determined that WDR5 regulates certain small *and* large ribosomal subunits across multiple cell types, this may indicate that WDR5 helps regulate not only the overall balance of RPGs as a whole, but also the RPL to RPS ratio. WDR5 may help control RPL and RPS stoichiometry, allowing cells to appropriately degrade excess RP subunits in a proteasome-dependent process and drive appropriate cellular growth via ribosome biogenesis.

Why some RPGs are bound by WDR5 could also be related to non-canonical ribosome protein functions, such as extra-ribosomal activities [155] and specialized ribosome formation [159], which impose unique demands for transcriptional control of specific ribosome protein subunits. WDR5 may regulate the production of certain RPGs that have extra-ribosomal activities. For example, RPL3, whose locus I have shown is bound by WDR5 (Fig 3.9), not only

plays a role in within the ribosome, but also has extra-ribosomal functions that affect the cellular response to nucleolar stress. The key extra-ribosomal role of RPL3 is to arrest cell cycle progression and to induce apoptosis [157]. RPL3 can induce G1 arrest through the activation of p21 gene transcription and apoptosis by molecular pathways involving p21 and regulate of DNA repair in p21-dependent and p21-independent manners [157]. Further, it has been proposed that the composition of ribosomes may be different in different cell types [153]. Perhaps WDR5 provides different cells the flexibility to enhance the production of certain RPGs in a cell-type specific manner, to allow for cell-type specific ribosome complexes. Alternatively, WDR5 could be necessary for the enhanced transcription of essential RPGs, while the transcription of cell-type specific ribosome's RPGs could be controlled in another manner.

The next step to determine the overarching, evolutionarily conserved role for chromatin-bound WDR5, is to identify which of the 94 conserved genes are controlled by WDR5, and respond to WDR5 binding. Chapters V and VI will delve into the experiments performed to answer this question - Chapter V will introduce the SMI used to reduce WDR5 binding on chromatin and Chapter VI will explore the consequences of this reduction.

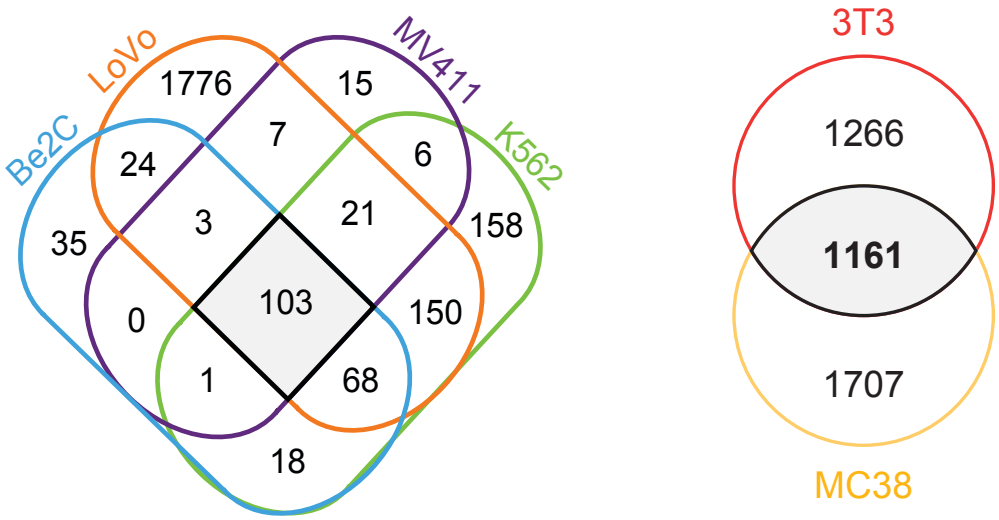


Figure 3.1. (*Left*) Venn diagram, showing the overlap of WDR5 peaks in ChIP-Seq data from all four human cell lines: Be2C, LoVo, MV4:11, and K562. (*Right*) Venn diagram, showing the overlap of WDR5 peaks in ChIP-Seq data from the two mouse cell lines: 3T3 (NIH3T3) and MC38.

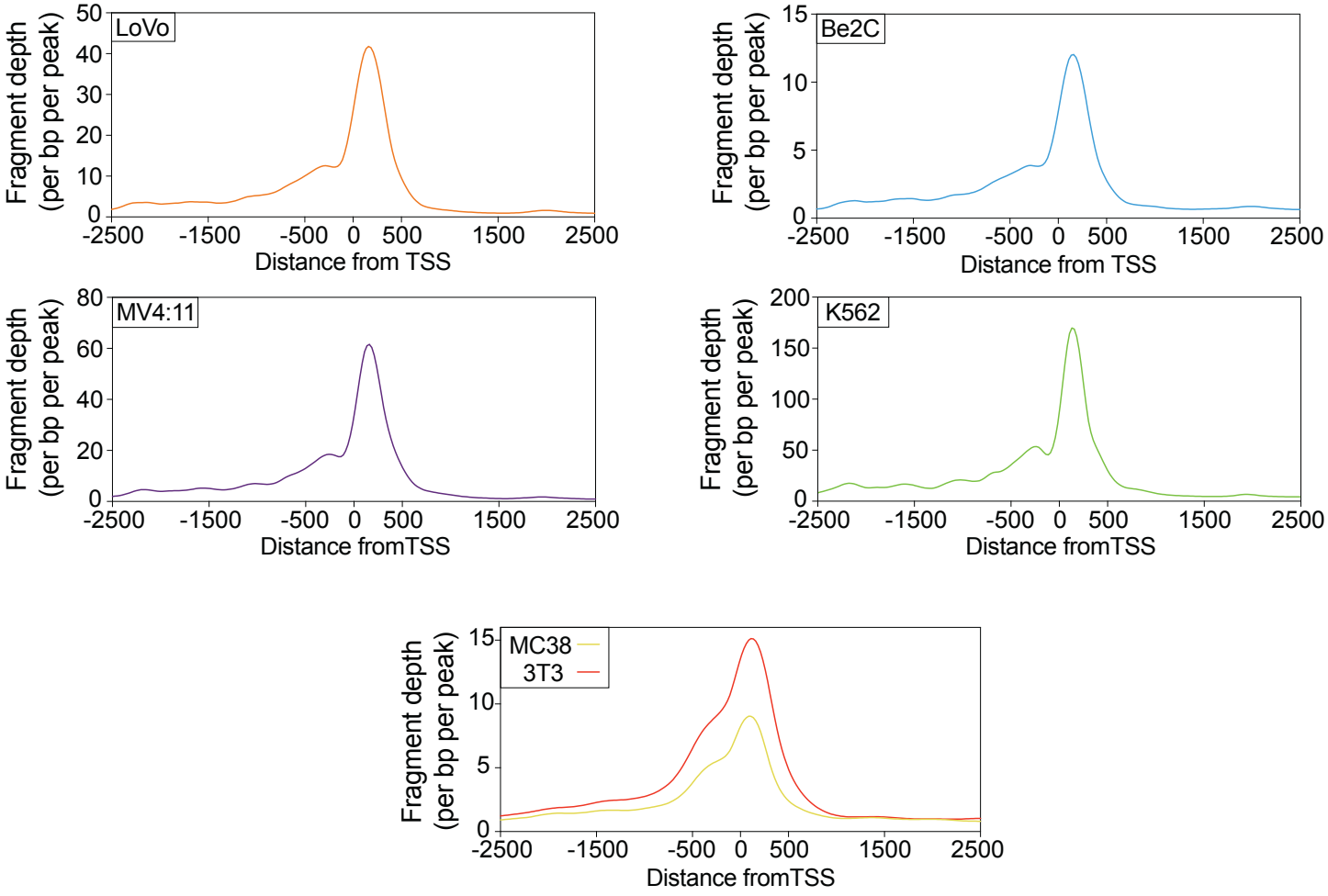


Figure 3.2. (*Top*) Averaged ChIP-Seq peak distribution for WDR5 at the 103 conserved sites/genes in human cells (K652, MV4:11, LoVo, and Be2C). (*Bottom*) Averaged ChIP-Seq peak distribution for WDR5 at the 1161 conserved sites/genes in mouse cells (3T3 and MC38).

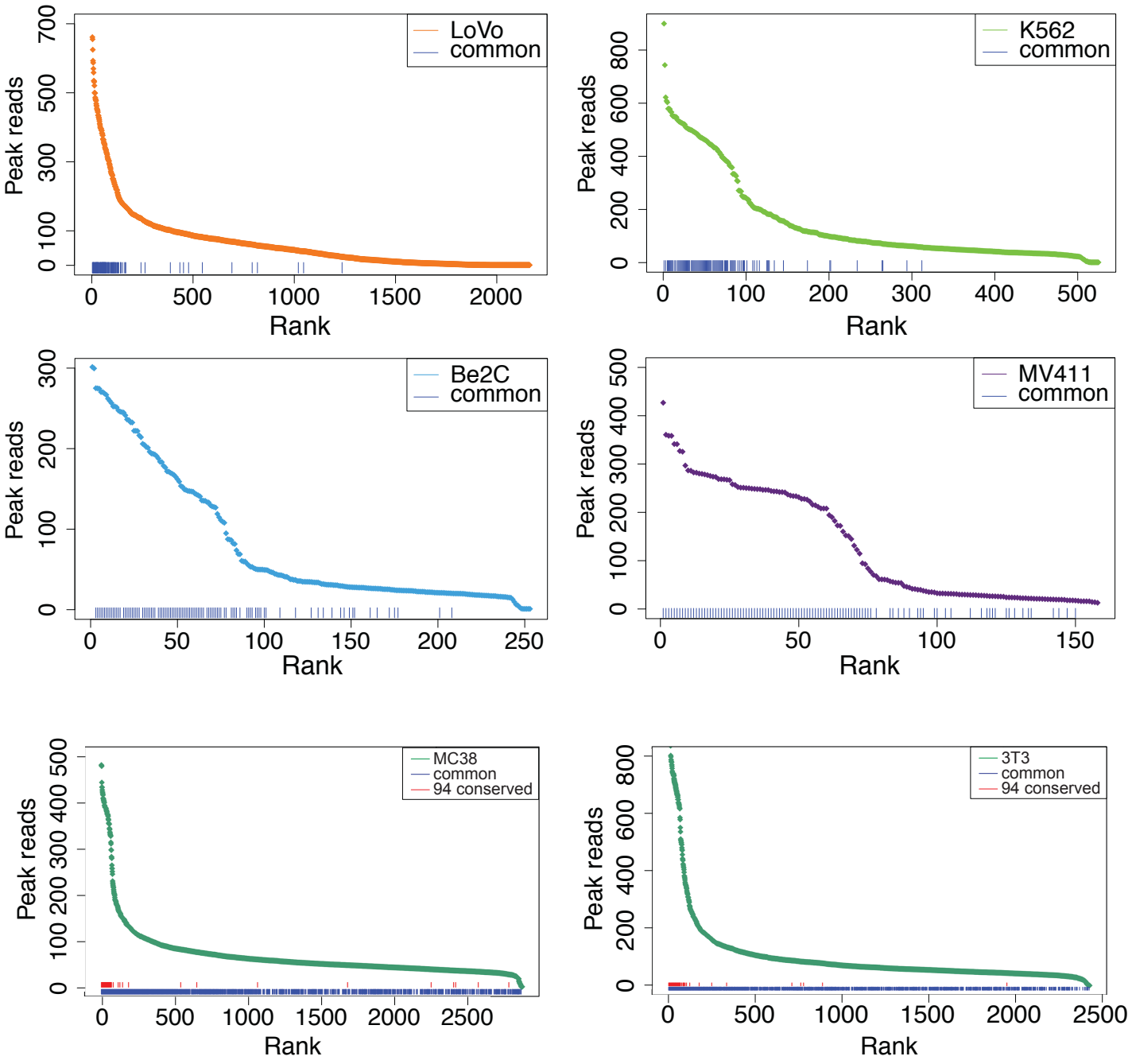


Figure 3.3. (*Top*) WDR5 peaks in all four human cell lines were ranked according to peak intensity; the blue lines at the bottom of each graph denote one of the 103 conserved human WDR5 peaks. (*Bottom*) WDR5 peaks in two mouse cell lines were ranked according to peak intensity; the blue lines at the bottom of each graph denote one of the 1161 conserved mouse WDR5 peaks. The red lines above these indicate one of the 94 conserved WDR5-bound genes.

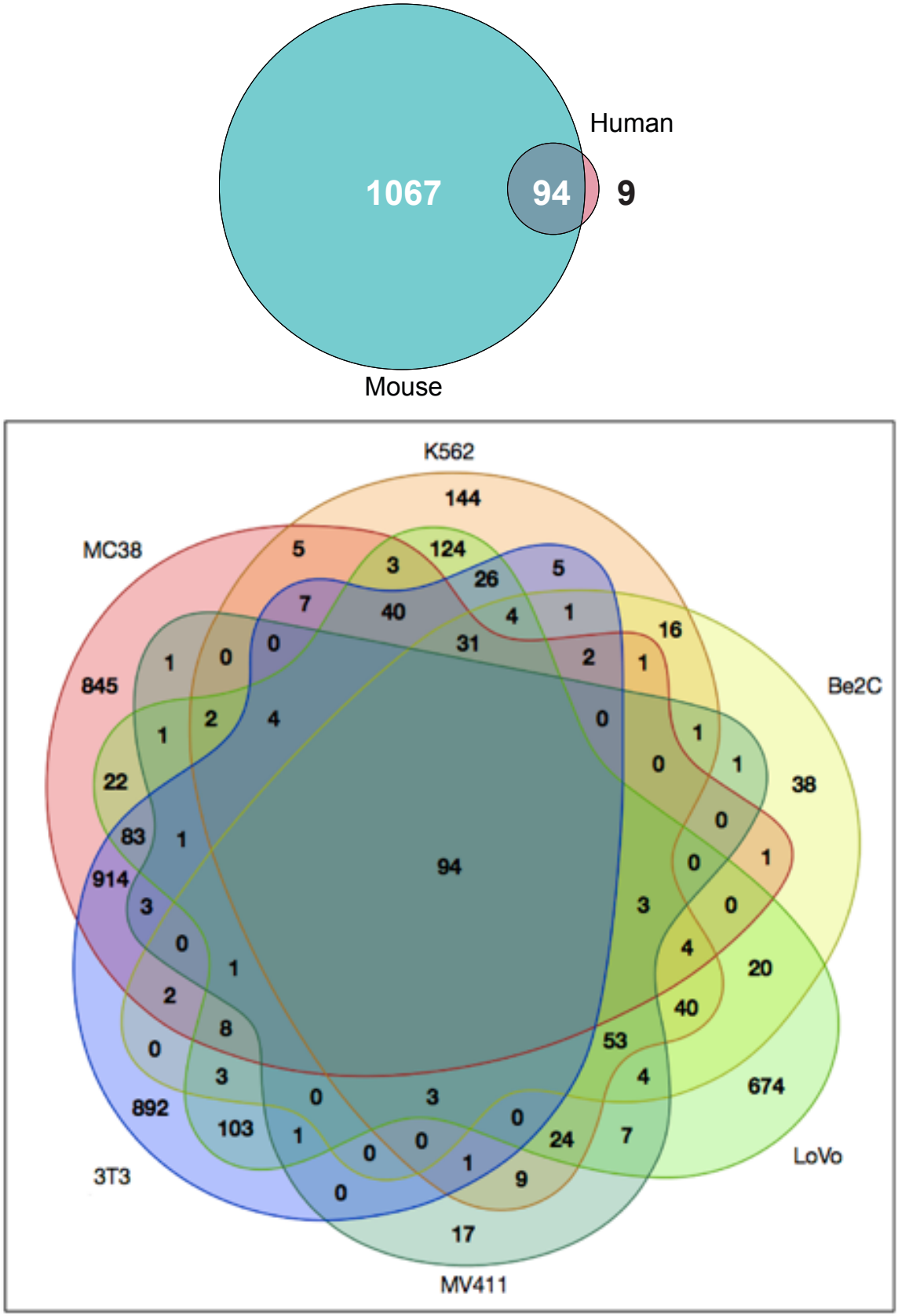


Figure 3.4. Venn diagrams, showing the overlap of genes bound by WDR5 in human and mouse cells. (*Top*) The 103 conserved human peaks/genes overlapped with the 1161 conserved mouse sites/genes. (*Bottom*) All sites/genes in all six cell lines represented and all overlaps are shown.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
AAGAB	Chromosome 15: 67,201,033-67,255,195 reverse strand	Chromosome 9: 63,602,660-63,644,588 forward strand	UniProt Function Summary	May be involved in endocytic recycling of growth factor receptors such as EGFR
CSNK1D	Chromosome 17: 82,239,023-82,273,731 reverse strand	Chromosome 11: 120,961,749-120,991,330 reverse strand	UniProt Function Summary	Essential serine/threonine-protein kinase that regulates diverse cellular growth and survival processes including Wnt signaling, DNA repair and circadian rhythms. It can phosphorylate a large number of proteins. Casein kinases are operationally defined by their preferential utilization of acidic proteins such as caseins as substrates. Phosphorylates connexin-43/GJA1, MAP1A, SNAPIN, MAPT/TAU, TOP2A, DCK, HIF1A, EIF6, p53/TP53, DVL2, DVL3, ESR1, AIB1/NCOA3, DNMT1, PKD2, YAP1, PER1 and PER2.
CSNK1E	Chromosome 22: 38,290,691-38,318,084 reverse strand	Chromosome 15: 79,417,856-79,455,566 reverse strand	UniProt Function Summary	Casein kinases are operationally defined by their preferential utilization of acidic proteins such as caseins as substrates. Can phosphorylate a large number of proteins. Participates in Wnt signaling. Phosphorylates DVL1 and DVL2. Central component of the circadian clock. In balance with PP1, determines the circadian period length, through the regulation of the speed and rhythmicity of PER1 and PER2 phosphorylation. Controls PER1 and PER2 nuclear transport and degradation. Inhibits cytokine-induced granuloytic differentiation.
LIAS	Chromosome 4: 39,458,587-39,485,109 forward strand	Chromosome 5: 65,391,497-65,410,693 forward strand	UniProt Function Summary	Catalyzes the radical-mediated insertion of two sulfur atoms into the C-6 and C-8 positions of the octanoyl moiety bound to the lipoyl domains of lipoate-dependent enzymes, thereby converting the octanoylated domains into lipoylated derivatives.
NACA	Chromosome 12: 56,712,428-56,731,628 reverse strand	Chromosome 10: 128,035,575-128,048,637 forward strand	UniProt Function Summary	Prevents inappropriate targeting of non-secretory polypeptides to the endoplasmic reticulum (ER). Binds to nascent polypeptide chains as they emerge from the ribosome and blocks their interaction with the signal recognition particle (SRP), which normally targets nascent secretory peptides to the ER. Also reduces the inherent affinity of ribosomes for protein translocation sites in the ER membrane (M sites). May act as a specific coactivator for JUN, binding to DNA and stabilizing the interaction of JUN homodimers with target gene promoters.
NKIRAS1	Chromosome 3: 23,891,660-23,946,591 reverse strand	Chromosome 14: 18,271,136-18,284,003 forward strand	UniProt Function Summary	Atypical Ras-like protein that acts as a potent regulator of NF-kappa-B activity by preventing the degradation of NF-kappa-B inhibitor beta (NFKBIB) by most signals, explaining why NFKBIB is more resistant to degradation. May act by blocking phosphorylation of NFKBIB and mediating cytoplasmic retention of p65/RELA NF-kappa-B subunit. It is unclear whether it acts as a GTPase. Both GTP- and GDP-bound forms block phosphorylation of NFKBIB.
PPP2R2A	Chromosome 8: 26,291,508-26,372,680 forward strand	Chromosome 14: 67,014,056-67,072,444 reverse strand	UniProt Function Summary	The B regulatory subunit might modulate substrate selectivity and catalytic activity, and also might direct the localization of the catalytic enzyme to a particular subcellular compartment.
YWHAE	Chromosome 17: 1,344,275-1,400,222 reverse strand	Chromosome 11: 75,732,869-75,765,845 forward strand	UniProt Function Summary	Adapter protein implicated in the regulation of a large spectrum of both general and specialized signaling pathways. Binds to a large number of partners, usually by recognition of a phosphoserine or phosphothreonine motif. Binding generally results in the modulation of the activity of the binding partner (By similarity). Positively regulates phosphorylated protein HSF1 nuclear export to the cytoplasm.
ACAP2	Chromosome 3: 195,274,745-195,443,044 reverse strand	Chromosome 16: 31,092,412-31,201,245 reverse strand	UniProt Function Summary	GTPase-activating protein (GAP) for ADP ribosylation factor 6 (ARF6)
CCT4	Chromosome 2: 61,868,085-61,888,671 reverse strand	Chromosome 11: 22,990,519-23,003,780 forward strand	UniProt Function Summary	Component of the chaperonin-containing T-complex (TRiC), a molecular chaperone complex that assists the folding of proteins upon ATP hydrolysis. The TRiC complex mediates the folding of WRAP53/TCAB1, thereby regulating telomere maintenance. As part of the TRiC complex may play a role in the assembly of BBSome, a complex involved in ciliogenesis regulating transports vesicles to the cilia. The TRiC complex plays a role in the folding of actin and tubulin (Probable).
CCT7	Chromosome 2: 73,233,420-73,253,021 forward strand	Chromosome 6: 85,451,514-85,468,475 forward strand	UniProt Function Summary	Component of the chaperonin-containing T-complex (TRiC), a molecular chaperone complex that assists the folding of proteins upon ATP hydrolysis. The TRiC complex mediates the folding of WRAP53/TCAB1, thereby regulating telomere maintenance. The TRiC complex plays a role in the folding of actin and tubulin (Probable).
RDH10	Chromosome 8: 73,294,602-73,325,281 forward strand	Chromosome 1: 16,105,774-16,133,734 forward strand	UniProt Function Summary	Retinol dehydrogenase with a clear preference for NADP. Converts all-trans-retinol to all-trans-retinal. Has no detectable activity towards 11-cis-retinol, 9-cis-retinol and 13-cis-retinol.
ARPC5L	Chromosome 9: 124,862,130-124,877,733 forward strand	Chromosome 2: 39,005,348-39,015,877 forward strand	UniProt Function Summary	May function as component of the Arp2/3 complex which is involved in regulation of actin polymerization and together with an activating nucleation-promoting factor (NPF) mediates the formation of branched actin networks.
IQCG	Chromosome 3: 197,889,077-197,960,142 reverse strand	Chromosome 16: 33,012,683-33,056,218 reverse strand	UniProt Function Summary	Component of the nexin-dynein regulatory complex (N-DRC), a key regulator of ciliary/flagellar motility which maintains the alignment and integrity of the distal axoneme and regulates microtubule sliding in motile axonemes. Binds calmodulin when cellular Ca2+ levels are low and thereby contributes to the regulation of calcium and calmodulin-dependent protein kinase IV (CAMK4) activity; contributes to the regulation of CAMK4 signaling cascades. Required for normal axoneme assembly in sperm flagella, normal sperm tail formation and for male fertility.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
IQCH	Chromosome 15: 67,254,786-67,502,260 forward strand	Chromosome 9: 63,421,455-63,602,493 reverse strand	UniProt Function Summary	May play a regulatory role in spermatogenesis.
RNF151	Chromosome 16: 1,966,823-1,968,975 forward strand	Chromosome 17: 24,715,839-24,718,063 reverse strand	UniProt Function Summary	May be involved in acrosome formation of spermatids.
BCL2L1	Chromosome 20: 31,664,452-31,723,989 reverse strand	Chromosome 2: 152,780,668-152,831,728 reverse strand	UniProt Function Summary	Potent inhibitor of cell death. Inhibits activation of caspases. Appears to regulate cell death by blocking the voltage-dependent anion channel (VDAC) by binding to it and preventing the release of the caspase activator, CYC1, from the mitochondrial membrane. Also acts as a regulator of G2 checkpoint and progression to cytokinesis during mitosis.
COG4	Chromosome 16: 70,480,568-70,523,558 reverse strand	Chromosome 8: 110,846,600-110,882,227 forward strand	UniProt Function Summary	Required for normal Golgi function. Plays a role in SNARE-pin assembly and Golgi-to-ER retrograde transport via its interaction with SCFD1
PRADC1	Chromosome 2: 73,228,010-73,233,239 reverse strand	Chromosome 6: 85,446,810-85,451,970 reverse strand	UniProt Function Summary	A secreted glycoprotein with unknown function.
SPHK2	Chromosome 19: 48,619,291-48,630,717 forward strand	Chromosome 7: 45,709,467-45,718,002 reverse strand	UniProt Function Summary	Catalyzes the phosphorylation of sphingosine to form sphingosine-1-phosphate (SPP), a lipid mediator with both intra- and extracellular functions. Also acts on D-erythro-dihydrosphingosine, D-erythro-sphingosine and L-threo-dihydrosphingosine. Binds phosphoinositides. In contrast to prosurvival SPHK1, has a positive effect on intracellular ceramide levels, inhibits cells growth and enhances apoptosis. In mitochondria, is important for cytochrome-c oxidase assembly and mitochondrial respiration. The SPP produced in mitochondria binds PHB2 and modulates the regulation via PHB2 of complex IV assembly and respiration. In nucleus, plays a role in epigenetic regulation of gene expression. Interacts with HDAC1 and HDAC2 and, through SPP production, inhibits their enzymatic activity, preventing the removal of acetyl groups from lysine residues with histones. Upregulates acetylation of histone H3-K9, histone H4-K5 and histone H2B-K12. In nucleus, may have an inhibitory effect on DNA synthesis and cell cycle. In mast cells, is the main regulator of SPP production which mediates calcium influx, NF-kappa-B activation, cytokine production, such as TNF and IL6, and degranulation of mast cells (By similarity). In dopaminergic neurons, is involved in promoting mitochondrial functions regulating ATP and ROS levels (By similarity). Also involved in the regulation of glucose and lipid metabolism (By similarity).
FAF1	Chromosome 1: 50,437,028-50,960,267 reverse strand	Chromosome 4: 109,676,588-109,963,960 forward strand	UniProt Function Summary	Ubiquitin-binding protein. Required for the progression of DNA replication forks by targeting DNA replication licensing factor CDT1 for degradation. Potentiates but cannot initiate FAS-induced apoptosis (By similarity).
LRSAM1	Chromosome 9: 127,451,486-127,503,501 forward strand	Chromosome 2: 32,925,216-32,961,614 reverse strand	UniProt Function Summary	E3 ubiquitin-protein ligase that mediates monoubiquitination of TSG101 at multiple sites, leading to inactivate the ability of TSG101 to sort endocytic (EGF receptors) and exocytic (HIV-1 viral proteins) cargos. Bacterial recognition protein that defends the cytoplasm from invasive pathogens. Localizes to several intracellular bacterial pathogens and generates the bacteria-associated ubiquitin signal leading to autophagy-mediated intracellular bacteria degradation (xenophagy).
RNF220	Chromosome 1: 44,405,194-44,651,724 forward strand	Chromosome 4: 117,271,463-117,497,052 reverse strand	UniProt Function Summary	E3 ubiquitin-protein ligase that promotes the ubiquitination and proteasomal degradation of SIN3B (By similarity). Independently of its E3 ligase activity, acts as a CTNNB1 stabilizer through USP7-mediated deubiquitination of CTNNB1 promoting Wnt signaling.
UBE2I	Chromosome 16: 1,308,880-1,327,018 forward strand	Chromosome 17: 25,261,916-25,274,622 reverse strand	UniProt Function Summary	Accepts the ubiquitin-like proteins SUMO1, SUMO2, SUMO3, SUMO4 and SUMO1P1/SUMO5 from the UBLE1A-UBLE1B E1 complex and catalyzes their covalent attachment to other proteins with the help of an E3 ligase such as RANBP2, CBX4 and ZNF451. Can catalyze the formation of poly-SUMO chains. Necessary for sumoylation of FOXL2 and KAT5. Essential for nuclear architecture and chromosome segregation. Sumoylates p53/TP53 at 'Lys-386'.
FAM208A	Chromosome 15: 28,422,172-28,422,373 forward strand	Chromosome 14: 27,428,834-27,483,555 forward strand	UniProt Function Summary	Component of the HUSH complex, a multiprotein complex that mediates epigenetic repression. The HUSH complex is recruited to genomic loci rich in H3K9me3 and is required to maintain transcriptional silencing by promoting recruitment of SETDB1, a histone methyltransferase that mediates further deposition of H3K9me3, as well as MORC2. Also represses L1 retrotransposons in collaboration with MORC2 and, probably, SETDB1, the silencing is dependent of repressive epigenetic modifications, such as H3K9me3 mark. Silencing events often occur within introns of transcriptionally active genes, and lead to the down-regulation of host gene expression. The HUSH complex is also involved in the silencing of unintegrated retroviral DNA by being recruited by ZNF638: some part of the retroviral DNA formed immediately after infection remains unintegrated in the host genome and is transcriptionally repressed.
HMGN1	Chromosome 21: 39,342,315-39,349,647 reverse strand	Chromosome 16: 96,120,618-96,127,729 reverse strand	UniProt Function Summary	Binds to the inner side of the nucleosomal DNA thus altering the interaction between the DNA and the histone octamer. May be involved in the process which maintains transcribable genes in a unique chromatin conformation. Inhibits the phosphorylation of nucleosomal histones H3 and H2A by RPS6KA5/MSK1 and RPS6KA3/RSK2.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
EEF2	Chromosome 19: 3,976,056-3,985,463 reverse strand	Chromosome 10: 81,176,631-81,182,498 forward strand	UniProt Function Summary	Catalyzes the GTP-dependent ribosomal translocation step during translation elongation. During this step, the ribosome changes from the pre-translocational (PRE) to the post-translocational (POST) state as the newly formed A-site-bound peptidyl-tRNA and P-site-bound deacylated tRNA move to the P and E sites, respectively. Catalyzes the coordinated movement of the two tRNA molecules, the mRNA and conformational changes in the ribosome.
EEF1G	Chromosome 11: 62,559,596-62,574,086 reverse strand	Chromosome 19: 8,967,041-8,978,479 forward strand	UniProt Function Summary	Probably plays a role in anchoring the complex to other cellular components.
EIF3D	Chromosome 22: 36,510,855-36,529,184 reverse strand	Chromosome 15: 77,958,998-77,970,813 reverse strand	UniProt Function Summary	mRNA cap-binding component of the eukaryotic translation initiation factor 3 (eIF-3) complex, a complex required for several steps in the initiation of protein synthesis of a specialized repertoire of mRNAs. The eIF-3 complex associates with the 40S ribosome and facilitates the recruitment of eIF-1, eIF-1A, eIF-2:GTP:methionyl-tRNAi and eIF-5 to form the 43S pre-initiation complex (43S PIC). The eIF-3 complex stimulates mRNA recruitment to the 43S PIC and scanning of the mRNA for AUG recognition. The eIF-3 complex is also required for disassembly and recycling of post-termination ribosomal complexes and subsequently prevents premature joining of the 40S and 60S ribosomal subunits prior to initiation. The eIF-3 complex specifically targets and initiates translation of a subset of mRNAs involved in cell proliferation, including cell cycling, differentiation and apoptosis, and uses different modes of RNA stem-loop binding to exert either translational activation or repression. In the eIF-3 complex, EIF3D specifically recognizes and binds the 7-methylguanosine cap of a subset of mRNAs.
EIF4G1	Chromosome 3: 184,314,495-184,335,358 forward strand	Chromosome 16: 20,668,313-20,692,884 forward strand	UniProt Function Summary	Component of the protein complex eIF4F, which is involved in the recognition of the mRNA cap, ATP-dependent unwinding of 5'-terminal secondary structure and recruitment of mRNA to the ribosome.
EIF4G3	Chromosome 1: 20,806,292-21,176,888 reverse strand	Chromosome 4: 137,993,022-138,208,508 forward strand	UniProt Function Summary	Probable component of the protein complex eIF4F, which is involved in the recognition of the mRNA cap, ATP-dependent unwinding of 5'-terminal secondary structure and recruitment of mRNA to the ribosome. Thought to be a functional homolog of EIF4G1.
LUC7L2	Chromosome 7: 139,340,359-139,423,457 forward strand	Chromosome 6: 38,551,334-38,609,470 forward strand	UniProt Function Summary	May bind to RNA via its Arg/Ser-rich domain.
PUM1	Chromosome 1: 30,931,506-31,065,991 reverse strand	Chromosome 4: 130,663,321-130,781,564 forward strand	UniProt Function Summary	Sequence-specific RNA-binding protein that acts as a post-transcriptional repressor by binding the 3'-UTR of mRNA targets. Binds to an RNA consensus sequence, the Pumilio Response Element (PRE), 5'-UGUANAUA-3'. Mediates post-transcriptional repression of transcripts via different mechanisms: acts via direct recruitment of the CCR4-POP2-NOT deadenylase leading to translational inhibition and mRNA degradation. Also mediates deadenylation-independent repression by promoting accessibility of miRNAs. Following growth factor stimulation, phosphorylated and binds to the 3'-UTR of CDKN1B/p27 mRNA, inducing a local conformational change that exposes miRNA-binding sites, promoting association of miR-221 and miR-222, efficient suppression of CDKN1B/p27 expression, and rapid entry to the cell cycle. Acts as a post-transcriptional repressor of E2F3 mRNAs by binding to its 3'-UTR and facilitating miRNA regulation. Represses a program of genes necessary to maintain genomic stability such as key mitotic, DNA repair and DNA replication factors. Its ability to repress those target mRNAs is regulated by the lncRNA NORAD (non-coding RNA activated by DNA damage) which, due to its high abundance and multitude of PUMILIO binding sites, is able to sequester a significant fraction of PUM1 and PUM2 in the cytoplasm. Involved in neuronal functions by regulating ATXN1 mRNA levels: acts by binding to the 3'-UTR of ATXN1 transcripts, leading to their down-regulation independently of the miRNA machinery. Plays a role in cytoplasmic sensing of viral infection. In testis, acts as a post-transcriptional regulator of spermatogenesis by binding to the 3'-UTR of mRNAs coding for regulators of p53/TP53. Involved in embryonic stem cell renewal by facilitating the exit from the ground state: acts by targeting mRNAs coding for naive pluripotency transcription factors and accelerates their down-regulation at the onset of differentiation (By similarity). Binds specifically to miRNA MIR199A precursor, with PUM2, regulates miRNA MIR199A expression at a postranscriptional level.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
RACK1	Chromosome 5: 181,236,909-181,248,096 reverse strand	Chromosome 11: 48,800,332-48,806,434 forward strand	UniProt Function Summary	Scaffolding protein involved in the recruitment, assembly and/or regulation of a variety of signaling molecules. Interacts with a wide variety of proteins and plays a role in many cellular processes. Component of the 40S ribosomal subunit involved in translational repression. Involved in the initiation of the ribosome quality control (RQC), a pathway that takes place when a ribosome has stalled during translation, by promoting ubiquitination of a subset of 40S ribosomal subunits. Binds to and stabilizes activated protein kinase C (PKC), increasing PKC-mediated phosphorylation. May recruit activated PKC to the ribosome, leading to phosphorylation of EIF6. Inhibits the activity of SRC kinases including SRC, LCK and YES1. Inhibits cell growth by prolonging the G0/G1 phase of the cell cycle. Enhances phosphorylation of BMAL1 by PRKCA and inhibits transcriptional activity of the BMAL1-CLOCK heterodimer. Facilitates ligand-independent nuclear translocation of AR following PKC activation, represses AR transactivation activity and is required for phosphorylation of AR by SRC. Modulates IGF1R-dependent integrin signaling and promotes cell spreading and contact with the extracellular matrix. Involved in PKC-dependent translocation of ADAM12 to the cell membrane. Promotes the ubiquitination and proteasome-mediated degradation of proteins such as CLEC1B and HIF1A. Required for VANGL2 membrane localization, inhibits Wnt signaling, and regulates cellular polarization and oriented cell division during gastrulation. Required for PTK2/FAK1 phosphorylation and dephosphorylation. Regulates internalization of the muscarinic receptor CHRM2. Promotes apoptosis by increasing oligomerization of BAX and disrupting the interaction of BAX with the anti-apoptotic factor BCL2L. Inhibits TRPM6 channel activity. Regulates cell surface expression of some GPCRs such as TBXA2R. Plays a role in regulation of FLT1-mediated cell migration. Involved in the transport of ABCB4 from the Golgi to the apical bile canalicular membrane. Promotes migration of breast carcinoma cells by binding to and activating RHOA
RNPS1	Chromosome 16: 2,253,116-2,268,397 reverse strand	Chromosome 17: 24,414,565-24,425,901 forward strand	UniProt Function Summary	Part of pre- and post-splicing multiprotein mRNP complexes. Auxiliary component of the splicing-dependent multiprotein exon junction complex (EJC) deposited at splice junction on mRNAs. The EJC is a dynamic structure consisting of core proteins and several peripheral nuclear and cytoplasmic associated factors that join the complex only transiently either during EJC assembly or during subsequent mRNA metabolism. Component of the ASAP and PSAP complexes which bind RNA in a sequence-independent manner and are proposed to be recruited to the EJC prior to or during the splicing process and to regulate specific excision of introns in specific transcription subsets. The ASAP complex can inhibit RNA processing during in vitro splicing reactions. The ASAP complex promotes apoptosis and is disassembled after induction of apoptosis. Enhances the formation of the ATP-dependent A complex of the spliceosome. Involved in both constitutive splicing and, in association with SRP54 and TRA2B/SFRS10, in distinctive modulation of alternative splicing in a substrate-dependent manner. Involved in the splicing modulation of BCL2L1/Bcl-X (and probably other apoptotic genes); specifically inhibits formation of proapoptotic isoforms such as Bcl-X(S); the activity is different from the established EJC assembly and function. Participates in mRNA 3'-end cleavage. Involved in UPF2-dependent nonsense-mediated decay (NMD) of mRNAs containing premature stop codons. Also mediates increase of mRNA abundance and translational efficiency. Binds spliced mRNA 20-25 nt upstream of exon-exon junctions.
SERBP1	Chromosome 1: 67,407,810-67,430,415 reverse strand	Chromosome 6: 67,238,176-67,297,736 forward strand	UniProt Function Summary	May play a role in the regulation of mRNA stability. Binds to the 3'-most 134 nt of the SERPINE1/PAI1 mRNA, a region which confers cyclic nucleotide regulation of message decay. Seems to play a role in PML-nuclear bodies formation.
SF3B3	Chromosome 16: 70,523,791-70,577,670 forward strand	Chromosome 8: 110,810,239-110,846,7 reverse strand	UniProt Function Summary	Involved in pre-mRNA splicing as a component of the splicing factor SF3B complex, a constituent of the spliceosome. SF3B complex is required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branchpoint sequence (BPS) in pre-mRNA. Sequence independent binding of SF3A/SF3B complex upstream of the branch site is essential, it may anchor U2 snRNP to the pre-mRNA. May also be involved in the assembly of the 'E' complex. Belongs also to the minor U12-dependent spliceosome, which is involved in the splicing of rare class of nuclear pre-mRNA intron
SNRPB	Chromosome 20: 2,461,634-2,470,853 reverse strand	Chromosome 2: 130,171,414-130,179,403 reverse strand	UniProt Function Summary	Plays role in pre-mRNA splicing as core component of the SMN-Sm complex that mediates spliceosomal snRNP assembly and as component of the spliceosomal U1, U2, U4 and U5 small nuclear ribonucleoproteins (snRNPs), the building blocks of the spliceosome. Component of both the pre-catalytic spliceosome B complex and activated spliceosome C complexes. Is also a component of the minor U12 spliceosome. As part of the U7 snRNP it is involved in histone pre-mRNA 3'-end processing
ZC3H10	Chromosome 12: 56,118,260-56,127,514 forward strand	Chromosome 10: 128,541,963-128,547,774 reverse strand	UniProt Function Summary	Specific regulator of miRNA biogenesis. Binds, via the C3H1-type zinc finger domains, to the binding motif 5'-GCAGCGC-3' on microRNA pri-MIR143 and negatively regulates the processing to mature microRNA.
RPL10A	Chromosome 6: 35,468,401-35,470,785 forward strand	Chromosome 17: 28,328,471-28,331,033 forward strand	UniProt Function Summary	Component of the large ribosomal subunit.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
RPL11	Chromosome 1: 23,691,742-23,696,835 forward strand	Chromosome 4: 136,028,265-136,053,428 reverse strand	UniProt Function Summary	Component of the ribosome, a large ribonucleoprotein complex responsible for the synthesis of proteins in the cell. The small ribosomal subunit (SSU) binds messenger RNAs (mRNAs) and translates the encoded message by selecting cognate aminoacyl-transfer RNA (tRNA) molecules. The large subunit (LSU) contains the ribosomal catalytic site termed the peptidyl transferase center, which catalyzes the formation of peptide bonds, thereby polymerizing the amino acids delivered by tRNAs into a polypeptide chain. The nascent polypeptides leave the ribosome through a tunnel in the LSU and interact with protein factors that function in enzymatic processing, targeting, and the membrane insertion of nascent chains at the exit of the ribosomal tunnel. As part of the 5S RNP/5S ribonucleoprotein particle it is an essential component of the LSU, required for its formation and the maturation of rRNAs. It also couples ribosome biogenesis to p53/TP53 activation. As part of the 5S RNP it accumulates in the nucleoplasm and inhibits MDM2, when ribosome biogenesis is perturbed, mediating the stabilization and the activation of TP53. Promotes nucleolar location of PML.
RPL12	Chromosome 9: 127,447,674-127,451,406 reverse strand	Chromosome 2: 32,961,559-32,965,345 forward strand	UniProt Function Summary	Binds directly to 26S ribosomal RNA.
RPL13	Chromosome 16: 89,560,657-89,566,828 forward strand	Chromosome 8: 123,102,350-123,105,244 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L13E family of ribosomal proteins. It is located in the cytoplasm.
RPL15	Chromosome 3: 23,916,545-23,924,374 forward strand	Chromosome 14: 18,267,823-18,271,391 reverse strand	Entrez Gene Summary	This gene encodes a member of the L15E family of ribosomal proteins and a component of the 60S subunit. This gene shares sequence similarity with the yeast ribosomal protein YL10 gene.
RPL17	Chromosome 18: 49,488,453-49,492,523 reverse strand	Chromosome 18: 74,998,558-75,003,381 fo rward strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL18	Chromosome 19: 48,615,328-48,619,184 reverse strand	Chromosome 7: 45,715,457-45,720,836 forward strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL23	Chromosome 17: 38,847,860-38,853,764 reverse strand	Chromosome 11: 97,777,527-97,782,437 reverse strand.	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L14P family of ribosomal proteins. It is located in the cytoplasm.
RPL24	Chromosome 3: 101,681,091-101,686,71 reverse strand	Chromosome 16: 55,966,275-55,971,435 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L24E family of ribosomal proteins. It is located in the cytoplasm.
RPL26	Chromosome 17: 8,377,516-8,383,250 reverse strand	Chromosome 11: 68,901,583-68,906,989 forward strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL27	Chromosome 17: 42,998,273-43,002,959 forward strand	Chromosome 11: 101,442,298-101,445,529 forward strand	UniProt Function Summary	Component of the large ribosomal subunit. Required for proper rRNA processing and maturation of 28S and 5.8S rRNAs.
RPL29	Chromosome 3: 51,993,522-51,995,942 reverse strand	Chromosome 9: 106,429,454-106,431,568 forward strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL3	Chromosome 22: 39,312,882-39,320,389 reverse strand	Chromosome 15: 80,077,791-80,091,868 reverse strand	UniProt Function Summary	The L3 protein is a component of the large subunit of cytoplasmic ribosomes.
RPL31	Chromosome 2: 101,002,229-101,024,032 forward strand	Chromosome 1: 39,367,842-39,371,911 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L31E family of ribosomal proteins. It is located in the cytoplasm.
RPL32	Chromosome 3: 12,834,485-12,841,588 reverse strand	Chromosome 6: 115,805,505-115,808,747 reverse strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L32E family of ribosomal proteins. It is located in the cytoplasm.
RPL34	Chromosome 4: 108,620,566-108,630,412 forward strand	Chromosome 3: 130,726,831-130,730,398 reverse strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL35	Chromosome 9: 124,857,880-124,861,981 reverse strand	Chromosome 2: 39,001,580-39,005,624 reverse strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL35A	Chromosome 3: 197,949,980-197,956,610 forward strand	Chromosome 16: 33,056,453-33,060,189 forward strand	UniProt Function Summary	Required for the proliferation and viability of hematopoietic cells. Plays a role in 60S ribosomal subunit formation. The protein was found to bind to both initiator and elongator tRNAs and consequently was assigned to the P site or P and A site.
RPL36	Chromosome 19: 5,674,947-5,691,875 forward strand	Chromosome 17: 56,613,416-56,614,243 forward strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL37	Chromosome 5: 40,825,262-40,835,222 reverse strand	Chromosome 15: 5,116,645-5,119,140 forward strand	UniProt Function Summary	Binds to the 23S rRNA.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
RPL37A	Chromosome 2: 216,498,189-216,579,180 forward strand	Chromosome 1: 72,711,290-72,713,813 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L37AE family of ribosomal proteins. It is located in the cytoplasm. The protein contains a C4-type zinc finger-like domain.
RPL38	Chromosome 17: 74,203,582-74,210,655 forward strand	Chromosome 11: 114,668,524-114,672,331 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L38E family of ribosomal proteins. It is located in the cytoplasm.
RPL39	Chromosome X: 119,786,504-119,791,630 reverse strand	Chromosome X: 37,082,520-37,085,402 reverse strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L38E family of ribosomal proteins. It is located in the cytoplasm.
RPL41	Chromosome 12: 56,116,590-56,117,967 forward strand	Chromosome 10: 128,548,114-128,549,305 reverse strand	UniProt Function Summary	This protein is the smallest and one of the most basic of the proteins in liver ribosomes.
RPL5	Chromosome 1: 92,831,983-92,841,924 forward strand	Chromosome 5: 107,900,502-107,909,005 forward strand	UniProt Function Summary	Component of the ribosome, a large ribonucleoprotein complex responsible for the synthesis of proteins in the cell. The small ribosomal subunit (SSU) binds messenger RNAs (mRNAs) and translates the encoded message by selecting cognate aminoacyl-transfer RNA (tRNA) molecules. The large subunit (LSU) contains the ribosomal catalytic site termed the peptidyl transferase center (PTC), which catalyzes the formation of peptide bonds, thereby polymerizing the amino acids delivered by tRNAs into a polypeptide chain. The nascent polypeptides leave the ribosome through a tunnel in the LSU and interact with protein factors that function in enzymatic processing, targeting, and the membrane insertion of nascent chains at the exit of the ribosomal tunnel. As part of the 5S RNP/5S ribonucleoprotein particle it is an essential component of the LSU, required for its formation and the maturation of rRNAs. It also couples ribosome biogenesis to p53/TP53 activation. As part of the 5S RNP it accumulates in the nucleoplasm and inhibits MDM2, when ribosome biogenesis is perturbed, mediating the stabilization and the activation of TP53.
RPL6	Chromosome 12: 112,405,190-112,418,838 reverse strand	Chromosome 5: 121,204,481-121,209,241 forward strand	UniProt Function Summary	Component of the large ribosomal subunit. (Microbial infection) Specifically binds to domain C of the Tax-responsive enhancer element in the long terminal repeat of HTLV-I.
RPL7	Chromosome 8: 73,290,242-73,295,789 reverse strand	Chromosome 1: 16,101,295-16,104,662 reverse strand	UniProt Function Summary	Component of the large ribosomal subunit. Binds to G-rich structures in 28S rRNA and in mRNAs. Plays a regulatory role in the translation apparatus; inhibits cell-free translation of mRNAs.
RPL7L1	Chromosome 6: 42,879,618-42,889,925 forward strand	Chromosome 17: 46,773,907-46,782,672 reverse strand	Gene Cards Summary	RPL7L1 (Ribosomal Protein L7 Like 1) is a Protein Coding gene. An important paralog of this gene is RPL7.
RPL8	Chromosome 8: 144,789,765-144,792,587 reverse strand	Chromosome 15: 76,904,078-76,906,314 forward strand	UniProt Function Summary	Component of the large ribosomal subunit.
RPL9	Chromosome 4: 39,452,521-39,458,949 reverse strand	Chromosome 5: 65,388,364-65,391,444 reverse strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L6P family of ribosomal proteins. It is located in the cytoplasm.
RPS10	Chromosome 6: 34,417,454-34,426,125 reverse strand	Chromosome 17: 27,630,418-27,636,669 reverse strand	UniProt Function Summary	Component of the 40S ribosomal subunit.
RPS14	Chromosome 5: 150,442,635-150,449,739 reverse strand	Chromosome 18: 60,747,098-60,778,546 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 40S subunit. The protein belongs to the S11P family of ribosomal proteins. It is located in the cytoplasm.
RPS15	Chromosome 19: 1,438,358-1,440,494 forward strand	Chromosome 10: 80,292,453-80,294,114 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 40S subunit. The protein belongs to the S19P family of ribosomal proteins. It is located in the cytoplasm. This gene has been found to be activated in various tumors, such as insulinomas, esophageal cancers, and colon cancers.
RPS15A	Chromosome 16: 18,781,295-18,790,383 reverse strand	Chromosome 7: 118,104,372-118,116,188 reverse strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 40S subunit. The protein belongs to the S8P family of ribosomal proteins. It is located in the cytoplasm.
RPS2	Chromosome 16: 1,962,058-1,964,841 reverse strand	Chromosome 17: 24,718,116-24,721,929 forward strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 40S subunit. The protein belongs to the S5P family of ribosomal proteins. It is located in the cytoplasm. This gene shares sequence similarity with mouse LLRep3. It is co-transcribed with the small nucleolar RNA gene U64, which is located in its third intron.
RPS24	Chromosome 10: 78,033,732-78,056,813 forward strand	Chromosome 14: 24,487,125-24,496,959 forward strand	UniProt Function Summary	Required for processing of pre-rRNA and maturation of 40S ribosomal subunits.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
RPS3	Chromosome 11: 75,399,515-75,422,280 forward strand	Chromosome 7: 99,477,896-99,483,738 reverse strand	UniProt Function Summary	Involved in translation as a component of the 40S small ribosomal subunit. Has endonuclease activity and plays a role in repair of damaged DNA. Cleaves phosphodiester bonds of DNAs containing altered bases with broad specificity and cleaves supercoiled DNA more efficiently than relaxed DNA. Displays high binding affinity for 7,8-dihydro-8-oxoguanine (8-oxoG), a common DNA lesion caused by reactive oxygen species (ROS). Has also been shown to bind with similar affinity to intact and damaged DNA. Stimulates the N-glycosylase activity of the base excision protein OGG1. Enhances the uracil excision activity of UNG1. Also stimulates the cleavage of the phosphodiester backbone by APEX1. When located in the mitochondrion, reduces cellular ROS levels and mitochondrial DNA damage. Has also been shown to negatively regulate DNA repair in cells exposed to hydrogen peroxide. Plays a role in regulating transcription as part of the NF-kappa-B p65-p50 complex where it binds to the RELA/p65 subunit, enhances binding of the complex to DNA and promotes transcription of target genes. Represses its own translation by binding to its cognate mRNA. Binds to and protects TP53/p53 from MDM2-mediated ubiquitination. Involved in spindle formation and chromosome movement during mitosis by regulating microtubule polymerization. Involved in induction of apoptosis through its role in activation of CASP8. Induces neuronal apoptosis by interacting with the E2F1 transcription factor and acting synergistically with it to up-regulate pro-apoptotic proteins BCL2L11/BIM and HRK/Dp5. Interacts with TRADD following exposure to UV radiation and induces apoptosis by caspase-dependent JNK activation.
RPS4X	Chromosome X: 72,255,679-72,277,248 reverse strand	Chromosome X: 102,184,941-102,189,394 reverse strand	Entrez Gene Summary	This gene encodes ribosomal protein S4, a component of the 40S subunit. RPS4 is the only ribosomal protein known to be encoded by more than one gene, this gene and ribosomal protein S4, Y-linked (RPS4Y). The 2 isoforms encoded by these genes are not identical, but are functionally equivalent. RPS4 belongs to the S4E family of ribosomal proteins. This gene is not subject to X-inactivation.
RPS6	Chromosome 9: 19,375,715-19,380,254 reverse strand	Chromosome 4: 86,854,660-86,857,412 reverse strand	UniProt Function Summary	May play an important role in controlling cell growth and proliferation through the selective translation of particular classes of mRNA.
RPS7	Chromosome 2: 3,575,205-3,580,920 forward strand	Chromosome 12: 28,630,854-28,635,953 reverse strand	UniProt Function Summary	Required for rRNA maturation.
RPS8	Chromosome 1: 44,775,251-44,778,779 forward strand	Chromosome 4: 117,153,827-117,156,243 reverse strand	Entrez Gene Summary	This gene encodes a ribosomal protein that is a component of the 40S subunit. The protein belongs to the S8E family of ribosomal proteins. It is located in the cytoplasm. This gene is co-transcribed with the small nucleolar RNA genes U38A, U38B, U39, and U40, which are located in its fourth, fifth, first, and second introns, respectively.
SNHG15	Chromosome 7: 44,983,023-44,986,961 reverse strand	Chromosome 11: 6,525,591-6,528,760 reverse strand	Entrez Gene Summary	This gene represents a snoRNA host gene that produces a short-lived long non-coding RNA. This non-coding RNA is upregulated in tumor cells and may contribute to cell proliferation by acting as a sponge for microRNAs.
SNHG17	Chromosome 20: 38,420,587-38,435,361 reverse strand	Chromosome 2: 158,353,700-158,361,580 reverse strand	Gene Cards Summary	SNHG17 (Small Nucleolar RNA Host Gene 17) is an RNA Gene, and is affiliated with the non-coding RNA class.
SNHG9	Chromosome 16: 1,964,959-1,965,509 forward strand	Chromosome 17: 24,719,531-24,719,965 reverse strand	Gene Cards Summary	SNHG9 (Small Nucleolar RNA Host Gene 9) is an RNA Gene, and is affiliated with the non-coding RNA class.
SNORA21	Chromosome 17: 38,852,863-38,852,994 reverse strand	Chromosome 11: 97,781,639-97,781,775 reverse strand	Gene Cards Summary	SNORA21 (Small Nucleolar RNA, H/ACA Box 21) is an RNA Gene, and is affiliated with the snoRNA class.
SNORA64	Chromosome 16: 1,962,973-1,963,106 reverse strand	Chromosome 17: 24,720,772-24,720,905 forward strand	Gene Cards Summary	SNORA64 (Small Nucleolar RNA, H/ACA Box 64) is an RNA Gene, and is affiliated with the snoRNA class.
SNORA78	Chromosome 16: 1,965,184-1,965,310 forward strand	Chromosome 17: 24,719,676-24,719,803 reverse strand	Gene Cards Summary	SNORA78 (Small Nucleolar RNA, H/ACA Box 78) is an RNA Gene, and is affiliated with the snoRNA class.
SNORA7A	Chromosome 3: 12,840,312-12,840,450 reverse strand	Chromosome 6: 115,807,975-115,808,114 reverse strand	Gene Cards Summary	SNORA7A (Small Nucleolar RNA, H/ACA Box 7A) is an RNA Gene, and is affiliated with the snoRNA class.
SNORA9	Chromosome 7: 44,985,378-44,985,510 reverse strand	Chromosome 11: 6,528,033-6,528,165 reverse strand	Gene Cards Summary	SNORA9 (Small Nucleolar RNA, H/ACA Box 9) is an RNA Gene, and is affiliated with the snoRNA class.
SNORD15A	Chromosome 11: 75,400,391-75,400,538 forward strand	Chromosome 7: 99,482,785-99,482,932 reverse strand	Gene Cards Summary	SNORD15A (Small Nucleolar RNA, C/D Box 15A) is an RNA Gene, and is affiliated with the snoRNA class.
SNORD43	Chromosome 22: 39,319,050-39,319,113 reverse strand	Chromosome 15: 80,082,847-80,082,954 reverse strand	Entrez Gene Summary	Intronic regions of ribosomal protein genes can harbor noncoding small nucleolar RNAs (snoRNAs), like SNORD43, which are generated during pre-mRNA processing. snoRNAs form part of the small nucleolar ribonucleoprotein particles (snoRNPs) involved in pre-rRNA processing and modification. snoRNAs of the box C/D class, like SNORD43, function in 2-prime-O-ribose methylation of rRNAs.

Gene	Human Locus (from Ensembl)	Mouse Locus (from Ensembl)	Description Source	Description
SNORD55	Chromosome 1: 44,775,864-44,775,943 forward strand	Chromosome 4: 117,155,770-117,155,848 reverse strand	Gene Cards Summary	SNORD55 (Small Nucleolar RNA, C/D Box 55) is an RNA Gene, and is affiliated with the snoRNA class.
SNORD58B	Chromosome 18: 49,491,664-49,491,729 reverse strand	Chromosome 18: 75,001,058-75,001,123 forward strand	Gene Cards Summary	SNORD58B (Small Nucleolar RNA, C/D Box 58B) is an RNA Gene, and is affiliated with the snoRNA class.
SNORD68	Chromosome 16: 89,561,434-89,561,517 forward strand	Chromosome 8: 123,103,038-123,103,123 forward strand	Gene Cards Summary	SNORD68 (Small Nucleolar RNA, C/D Box 68) is an RNA Gene, and is affiliated with the snoRNA class.
SNORD96A	Chromosome 5: 181,241,814-181,241,892 reverse strand	Chromosome 11: 48,802,032-48,802,112 forward strand	Gene Cards Summary	SNORD96A (Small Nucleolar RNA, C/D Box 96A) is an RNA Gene, and is affiliated with the snoRNA class.

Figure 3.5. Table 4 - table of the 94 conserved genes, across all human and mouse cell lines. The locus for each gene, in each species is listed, as is a description of the gene. List is clustered by gene function.

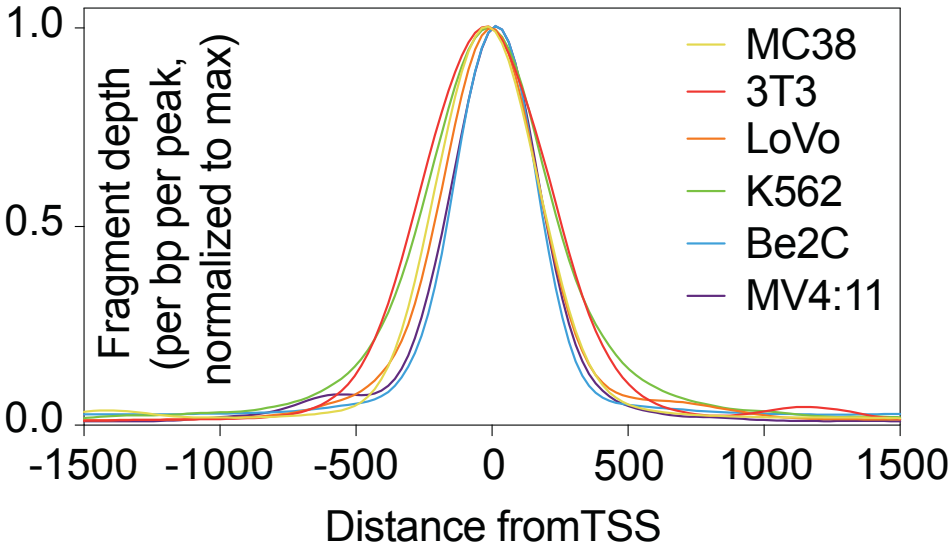


Figure 3.6. Averaged ChIP-Seq peak distribution for WDR5 at the 94 genes bound by WDR5 in all six cell lines, relative to the TSS. Peak signal was normalized to the maximum peak signal per cell line.

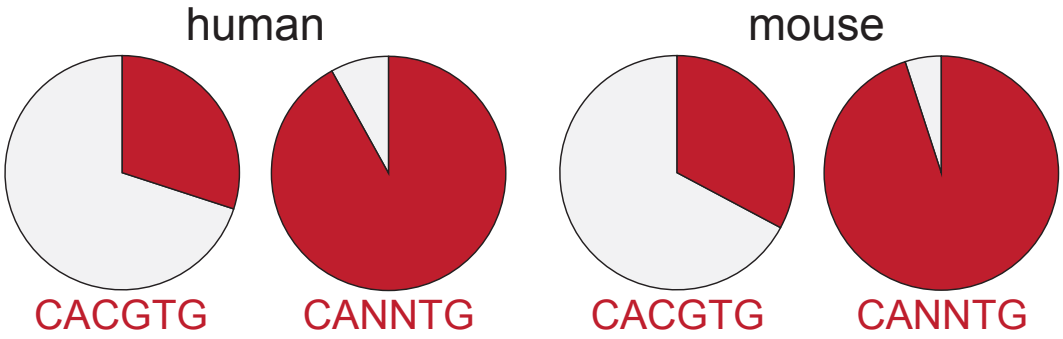


Figure 3.7. Representation of canonical (CACGTG) or non-canonical (CANNTG) E-box motifs in the 74 human (*left*), or 76 mouse (*right*) WDR5 binding sites corresponding to the 94 common genes. Charts only indicate the presence of at least one canonical or non-canonical E-Box per site. In some sites there are more than one E-Box motif, but this is not indicated here. Further, a non-canonical motif is defined here as any motif with CANNTG, including CACGTG, and where the 'N' can be either A, T, C, or G.

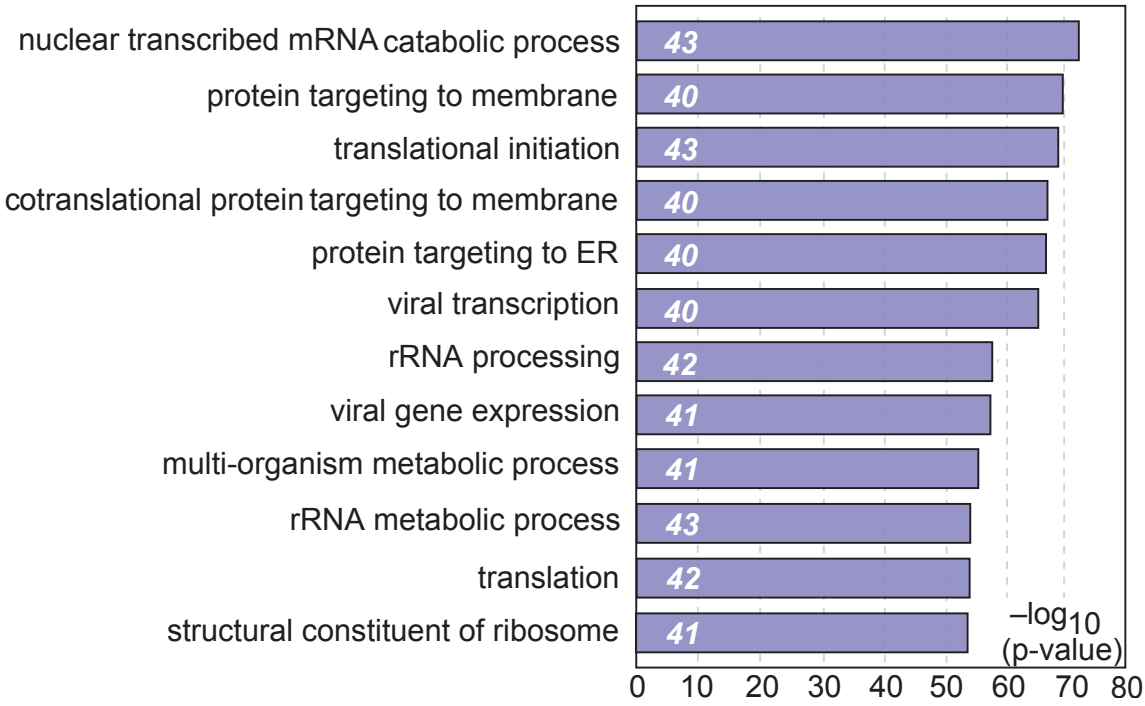


Figure 3.8. Functional GO enrichment analysis of the 94 conserved WDR5-bound genes. The top 12 GO categories are shown; numbers in italics show the number of genes in each category. GO analysis was performed using DAVID (Huang da, Sherman et al. 2009).

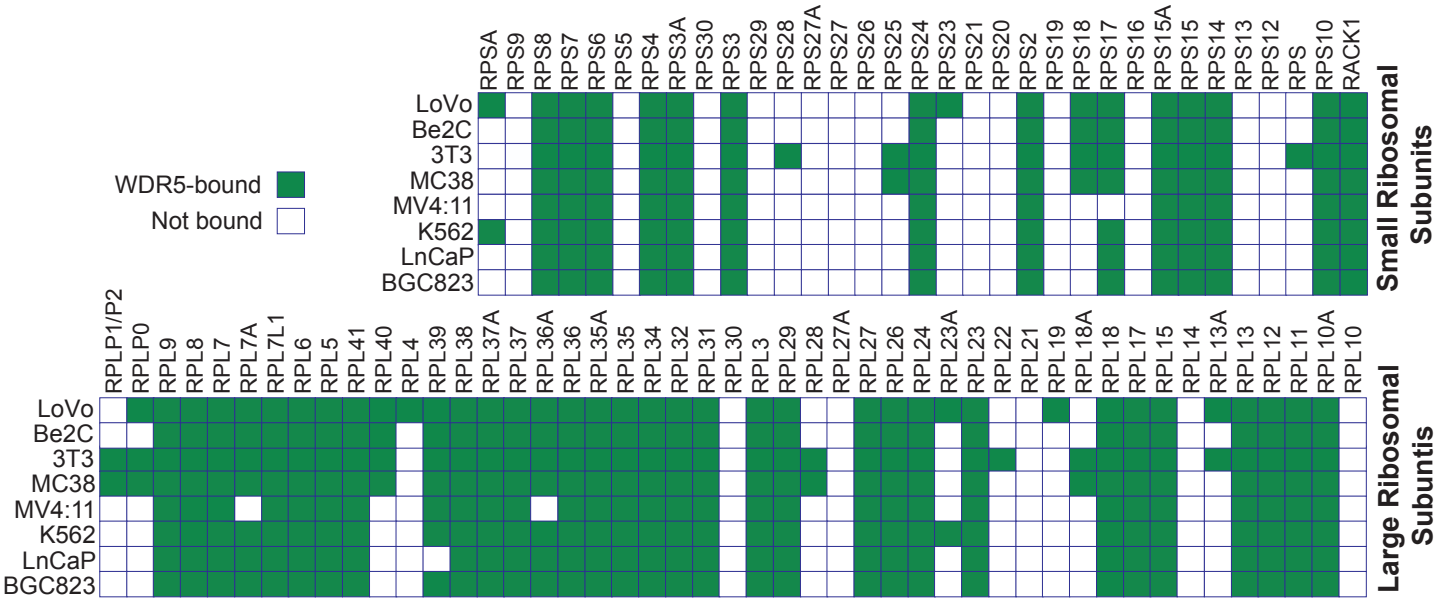
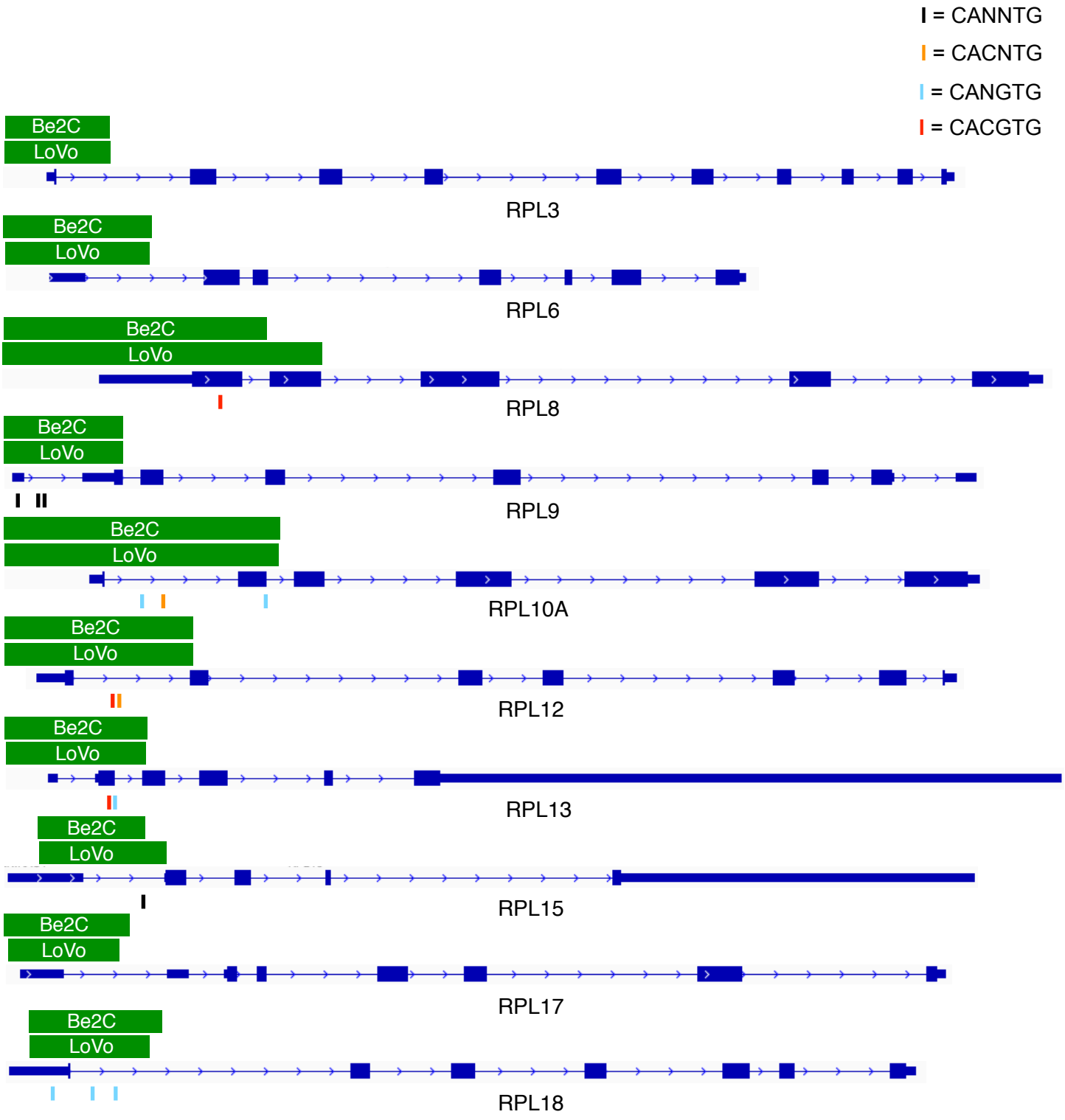
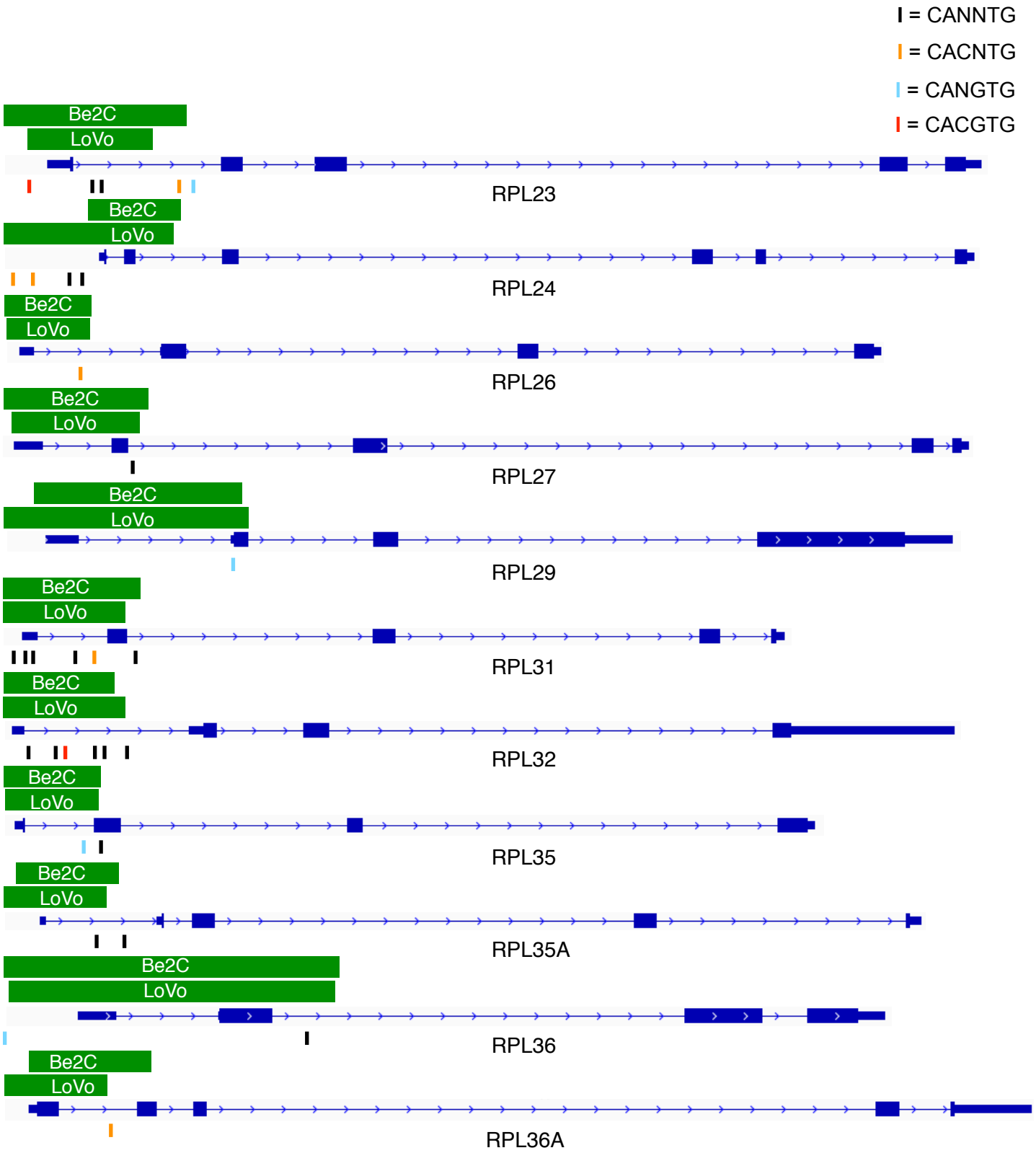
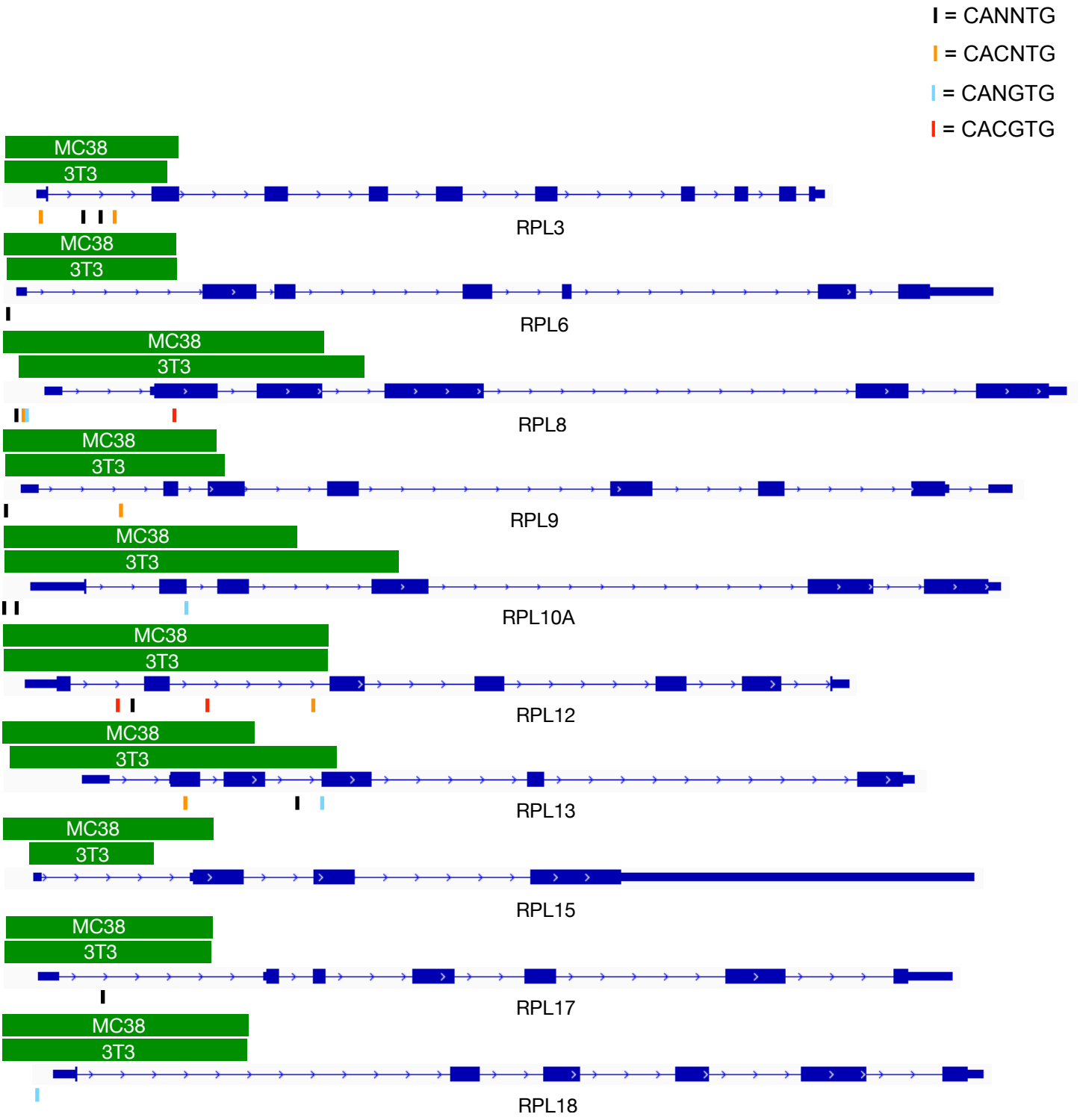


Figure 3.9. Ribosomogram, showing small (*top*) and large (*bottom*) ribosome subunit RPGs; a green box indicates whether WDR5 is bound top each RPG in the indicated cell type. LoVo, Be2C, MC38, 3T3, and K562 ChIP-seq data are from this study. MV4:11 ChIP-seq data are taken from Aho, Wang, et.al. 2019. ChIP-seq data in LnCaP and BGC823 are from GSE55279 (WDR5 ChIP-Seq, EtOH treatment) and GSE63763 (WDR5 NC_ChIPSeq), respectively.


I = CANNTG
I = CACNTG
I = CANGTG
I = CACGTG
Be2C
LoVo
RPL3
Be2C
LoVo
RPL6
Be2C
LoVo
RPL8
Be2C
LoVo
RPL9
Be2C
LoVo
RPL10A
Be2C
LoVo
RPL12
Be2C
LoVo
RPL13
Be2C
LoVo
RPL15
Be2C
LoVo
RPL17
Be2C
LoVo
RPL18


I = CANNTG
I = CACNTG
I = CANGTG
I = CACGTG
Be2C
LoVo
RPL23
RPL24
RPL26
RPL27
RPL29
RPL31
RPL32
RPL35
RPL35A
RPL36
RPL36A


I = CANNTG
I = CACNTG
I = CANGTG
I = CACGTG
MC38
3T3
RPL3
MC38
3T3
RPL6
MC38
3T3
RPL8
MC38
3T3
RPL9
MC38
3T3
RPL10A
MC38
3T3
RPL12
MC38
3T3
RPL13
MC38
3T3
RPL15
MC38
3T3
RPL17
MC38
3T3
RPL18

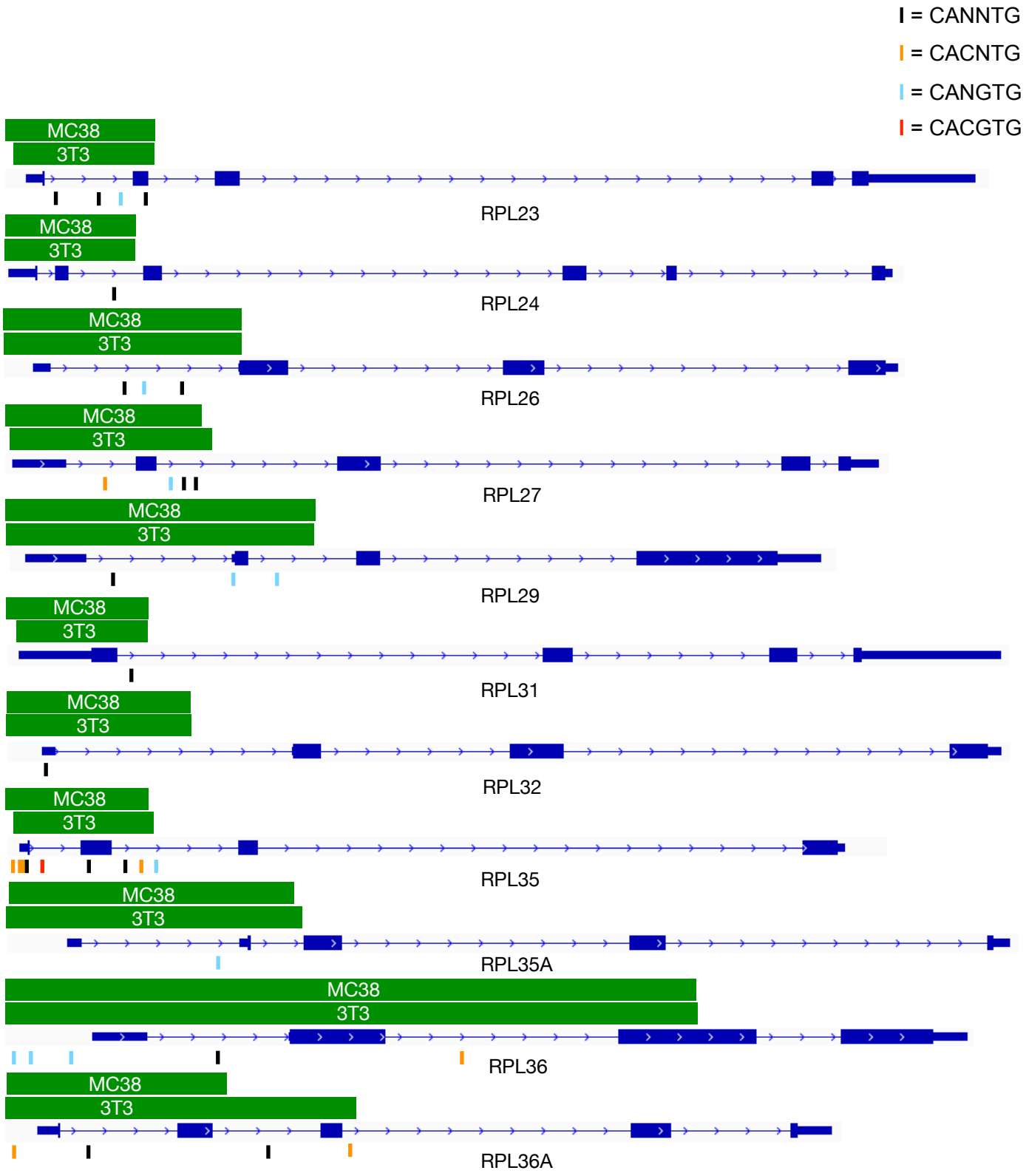


Figure 3.10. WDR5 peaks (green bars) are plotted above IGV screen shots of RP genes in the human cell lines Be2C and LoVo, and the mouse cell lines MC38 and 3T3. Below, the location of E-Boxes is indicated by colored lines - black: CANNTG, orange: CACNTG, blue: CANGTG, and red: CACGTG (where N can be A, T, C, or G).

CHAPTER V

Binding of WDR5 to Conserved Genes is WIN Site-Dependent

Abstract

Having identified a set of conserved WDR5-bound loci, I next asked whether the interaction of WDR5 with these conserved genes is WIN site-dependent. Utilizing novel small molecule inhibitors (SMIs) of WDR5, I attempted to disrupt the interaction between WDR5 and chromatin. As the WIN site of WDR5 has previously been shown, by the Tansey laboratory, to be necessary for WDR5 to bind chromatin, I used our recently published WIN-site WDR5 inhibitor in two ChIP experiments. First, I used ChIP-qPCR to test WDR5 binding in LoVo cells treated with this SMI. Finding a reduction of WDR5 at all sites interrogated, ChIP-Seq was used to test WDR5 binding in K562 cells treated with this SMI. Significant global reduction of WDR5 was apparent, and WDR5 binding at all 94 conserved genes was reduced. From this analysis, I conclude that the WIN site plays a general role in linking WDR5 to chromatin in various cancer cell lines, especially at the conserved sites of WDR5 binding, and by extension that WIN site inhibitors function broadly to displace WDR5 from conserved target genes.

Introduction

Chapters III and IV have outlined the foundation of my thesis work. The discovery of chromatin-bound WDR5 across cell types and identification of conserved WDR5 binding sites has indicated the importance of WDR5 localization to loci of genes involved in ribosome biogenesis, RNA processing, and translation. Therefore, determining which WDR5-bound genes are responsive to WDR5 localization became paramount to understanding the overarching, evolutionarily conserved role for chromatin-bound WDR5. While the ideal experiment would be to completely remove WDR5 from the cell and determine the changes in transcript levels, such an experiment was not feasible. WDR5 is an essential gene, and therefore, creating a stable cell line without WDR5 expression is not possible. Additionally, at the beginning of my thesis work in 2014, the now infamous dTAG system (the inducible, tag-based degradation system) was not yet invented. Therefore, I decided to take advantage of the fact that the WIN site of WDR5 is utilized for its known interactions with chromatin. WDR5 has been shown to use the WIN site to bind methylated histone H3 [12, 14, 18], to bind with MLL(1-4) as part of the MLL complexes [19-21], to bind with SET1A or SET1B as part of the SET (also known as COMPASS) complexes [19-21], to bind with KANSL1 as part of the NSL complex [133], and to bind with MBD3C as part of the

NuRD complex [39]. There is evidence that these interactions, via the WIN site, are critical to recruitment of WDR5 to chromatin. For example, the WIN site has been shown to be critical for robust HMT activity — through its WIN site interaction with SET1/MLL proteins that interact with chromatin, WDR5 is thought to be able to bind to chromatin and induce HMT activity [86]. WDR5-WIN site interaction with KANSL1 also appears to be required for efficient recruitment of the NSL complex with chromatin, as mutations that disrupt the KANSL1-WDR5 interact disrupt NSL binding to chromatin [86]. In addition, multiple WIN site inhibitors have been shown to alter the effects of chromatin-bound WDR5. MM-401, a macrocyclic peptidomimetic, inhibits the HMT function of MLL1 complexes in MLL-AF9 cells and in vitro, consistent with the requirement of the WDR5-WIN site interaction with MLL1 for robust binding to chromatin and methyltransferase activity [2, 86]. Using C3 and C6, Aho, et. al. showed that WIN site SMIs rapidly displace WDR5 from chromatin and act as potent inhibitors of MLL1-driven HMT activity in MV4:11 cells [74]. Grebien, et al. also showed that a chemically-distinct WIN site SMI, OICR-9429, inhibits MLL1-HMT activity in vitro and shows inhibition of cancer cell lines in culture (including human AML cells with N-terminal C/EBPα mutations) [160].

As an ongoing collaboration between the Tansey and Fesik laboratories at Vanderbilt University has resulted in the discovery of the aforementioned small molecule inhibitors that displace WDR5 from chromatin, C3 and C6, I was able to use these SMIs to probe whether WDR5-bound genes are responsive to WDR5-WIN site binding to chromatin. Earlier this year, in *Displacement of WDR5 from Chromatin by a WIN Site Inhibitor with Picomolar Affinity*, the Tansey laboratory introduced C3 and C6, and their respective negative control compounds C3nc and C6nc, to explore the cellular consequences of WIN site blockade [74]. C3 was their first-generation chemical probe, with a K_d of 1.3 nM, and C3nc has the same molecular weight as C3 but binds WDR5 with reduced affinity due to the regioisomeric fluorine atom that clashes with the protein [74]. A further improvement in affinity was achieved by occupying the S_4 pocket with compound C6 — C6 has a K_d of ~100 pM and served as their second-generation chemical probe [74]. The negative control, C6nc, was developed by adjusting the attachment point of the S_2 imidazole-imine "warhead" to yield a compound with a >1,000-fold reduction in binding affinity [74]. Using C3 and C6, they showed that by blocking the WIN site on WDR5, WDR5 is rapidly displaced from chromatin [74]. However, these inhibitors were only characterized in one cell line, MV4:11, and therefore, the authors can only hypothesize these SMIs will work similarly in other cell lines. Thus, I set out to determine whether these SMIs have the same effect in multiple cancer cell lines. Use of these SMIs as chemical probes allowed me to further interrogate the mechanism of WDR5 binding to chromatin and determine whether the WIN site of WDR5 is necessary for this binding across cell types. The results discussed in this chapter provide evidence for the use of WIN site SMIs in determining which WDR5-bound genes are WDR5-WIN-site-dependent, which will be discussed in chapter VI.

Results

To determine whether the WIN site of WDR5 mediates the interaction between WDR5 and chromatin in additional cell lines, I used C6 and C6nc in LoVo and K562 cells. I first utilized gene-specific ChIP in LoVo cells to interrogate WDR5 binding to a select set of the conserved genes. ChIP conditions were tested and optimized using the WDR5 antibody or the control antibody, IgG, in LoVo cells treated overnight with DMSO, 25 μM C6nc, or 25 μM C6 (data not shown). Co-precipitating DNA was detected by Q-PCR and expressed as a percentage of signal in input chromatin. As IgG signal was still consistently low, regardless of treatment with DMSO, C6nc, or C6, DMSO was chosen to be used with IgG as the control treatment in the following quadruplicate experiments. ChIP for WDR5 was completed in biological quadruplicate in LoVo cells treated overnight with DMSO, 25 μM C6nc, or 25 μM C6. As C6 was used in these experiments as a tool compound, and LoVo cells do not exhibit a growth deficit or apoptotic response to high levels of C6 (Fig 4.2), I wanted to use a higher concentration of C6 than was used by Aho, et. al. [74]. I chose 25 μM because I wanted to determine whether WDR5 can be completely displaced from chromatin using a higher dose of C6, as a small population of WDR5 remained on chromatin in MV4:11 cells treated with 2 μM C6 [74]. To address potential off-target effects, I also treated with the control compound, C6nc, at 25 μM. Genes to determine ChIP signal at were chosen based on ChIP-Seq results in LoVo cells. Since WDR5 bound SNHG15, RPS6, PUM1, LAMP1, EIF4G1, RNPS1, RPL23, RPL35, RPL37, RPL5, RPS24, and CCT7, some of which are shown in Figure 4.1, these genes were used to determine the effects of C6 on chromatin-bound WDR5. As WDR5 did not bind RPS12, RPS5, RPL10, METTL1, or SURF6, these were used negative control genes. A site within the gene body of SNHG15 was also used as a negative control. Figure 4.1 shows representative screen shots of normalized WDR5 ChIP-Seq data in LoVo cells, in the Integrated Genomics Viewer (IGV) program. 6 of the 94 conserved genes are shown. The data range was set as indicated on the left in each screen shot, and each image encompasses the entire gene of interest, with the genomic region, in bp, noted at the top of each screen shot. For each gene, the direction of transcription is indicated with blue arrows and the exons are indicated by blue boxes.

While C6 had an obvious effect on WDR5 at positive control loci, as is shown in Figure 4.2, no signifiant signal was detected for WDR5 at all negative control loci in LoVo cells. One-Way ANOVA, followed by Dunnett's Post-Hoc Test, was performed on signals from each positive control gene to determine the statistical significance of WDR5 displacement upon C6/C6nc vs. DMSO treatment. C6nc only significantly reduced the signal of WDR5 at one gene, RNPS1, however C6nc did so with a p-value of $P<0.05$, while C6 reduced signal much more significantly, with a p-value of $P<0.0001$. In contrast to C6nc, C6 significantly reduced the interaction of WDR5 with chromatin at all twelve genes interrogated, including SNHG15, RPS6, PUM1, LAMP1, EIF4G1, RNPS1, RPL23, RPL35, RPL37, RPL5, RPS241, and

CCT7. These changes in chromatin association are not accompanied by changes in WDR5 protein levels. Figure 4.3 shows the steady-state levels of WDR5 in LoVo cells after overnight treatment with DMSO, 25 μM C6nc, or 25 μM C6.

To extend these findings globally, and test the effects of WIN site inhibition in a third cell line, ChIP-Seq was performed by Dr. Gregory Caleb Howard in biological triplicate in K562 cells treated with DMSO, C6nc, or C6 for four hours — the results were comparable to MV4:11 [74] and LoVo cell findings. Figure 4.4 (*top*) shows a scatterplot of normalized average read counts for WDR5 binding peaks in K562 cells treated for four hours with DMSO, 2 μM C6nc, or 2 μM C6, as determined by ChIP-Seq. These experiments were performed using 2 μM C6 in order to more accurately compare this insensitive leukemia cell line (K562, shown to be insensitive to C6 in [74]) to a sensitive leukemia cell line (MV4:11, shown to be sensitive to C6 in [74]; 2 μM C6 was used to perform ChIP-seq experiments in [74]). Peaks are ranked based on read counts in DMSO-treated cells. As is obvious from this graph, every WDR5-bound peak in K562 cells is affected by C6 treatment - I observed a widespread reduction in the interaction of WDR5 with chromatin, the average magnitude of which was ~2.5-fold - again confirming that the binding of WDR5 to chromatin is WIN-site-dependent. High-intensity WDR5 peaks, which include the majority of the peaks assigned to the 94 conserved genes, are the most significantly reduced upon C6 treatment. A subset of peaks are reduced with C6nc, but not significantly (see Fig 4.6) and not nearly to the extent that they are reduced with C6. To determine whether the interaction of WDR5 with the 94 conserved genes is WIN-site dependent, I looked specifically at the K562 ChIP-Seq peaks that were assigned to the 94 conserved genes. Figure 4.4 (*bottom*) shows the heat maps for these peaks, with the WDR5 ChIP-Seq peak intensity per treatment in K562 cells indicated. Yellow pixels indicate the presence of called WDR5 ChIP-Seq peak signal. Signals that were present in two out of three replicates are graphed (and used for further analysis). The combined average of normalized peak intensity is indicated, split into 100-bp bins, and ± 2 kb around the center of peaks is shown. Genes are ranked based on DMSO-treated cells. Clearly, the amount of WDR5 bound to all 94 conserved genes is reduced. Additionally, as is indicated in Figure 4.5, there was no change in the steady state levels of WDR5 in K562 cells upon inhibitor treatment for four hours with DMSO, 2 μM C6nc, or 2 μM C6.

Figure 4.6 (*top*) delves into the statistical significance of C6 and C6nc treatment of K562 cells. A box and whisker plot, showing the $\log_2$-fold change in WDR5 ChIP-Seq peak intensity in K562 cells, compares C6nc and C6 treatments. The difference in signal for each peak is represented as a dot in the scatter plot. The box extends from the 25th to the 75th percentile, with the median marked by the middle line; whiskers extend from minimum to maximum points. Using the Wilcoxon test, I was able to determine: 1) there is no significant difference in ChIP-Seq peak signal for cells treated with C6nc versus DMSO, and 2) there is a significant difference in ChIP-Seq peak signal for cells treated with C6 versus

DMSO, with a p-value of P<0.0001. Figure 4.6 (*bottom*) further shows the impact of C6 on WDR5 binding to chromatin, with IGV screenshots of normalized ChIP-Seq data for WDR5 in K562 cells treated for four hours with DMSO, 2 μM C6nc, or 2 μM C6. Each image encompasses 7,308 bp from the indicated genomic regions, therefore, only part of the *RPL5*, *RPS24*, and *PUM1* genes are shown. The same data range was set for all samples and for each gene, direction of transcription is indicated with red arrows. Included is a representative selection of WDR5-bound genes, including a subset of the 94 conserved genes.

Discussion

In this chapter, I have provided further evidence that WIN site of WDR5 mediates the interaction between WDR5 and chromatin. A novel SMI, C6, displaces WDR5 from two cell lines, not previously tested: LoVo and K562. Utilizing C6, I have also determined where WDR5 is displaced from chromatin when targeted with a WIN site inhibitor.

In LoVo cells, I determined that by inhibiting the WIN site, WDR5 binding is reduced at genes I previously identified as bound by WDR5 via ChIP-Seq (Figures 4.1, 2.2, and 2.3). Included in these genes are some of conserved RPGs. In K562 cells, inhibition of the WIN site leads to global displacement of WDR5 from chromatin. Additionally, inhibition of WDR5 leads to a reduction in all peaks assigned to the 94 conserved genes, and therefore, interaction of WDR5 with the conserved genes is dependent on on the WIN site. These data are consistent with the previously reported reduction in global WDR5 binding in MV4:11 cells treated with C6 [74]. In all three cell lines, LoVo, K562, and MV4:11, treatment with C6 leads to displacement of WDR5 from chromatin without a change in steady-state protein levels (Figures 4.3-4.6 and [74]). Therefore, it is unlikely that C6 is working by effecting the stability, transcription, or translation, of the WDR5 protein. Rather, it is most likely that C6 is sterically inhibiting WDR5 from binding to other proteins by taking up residence in the WIN site. By blocking the WIN site pocket, we believe that C6 is preventing WDR5 from binding to the, as of yet, unknown proteins which are responsible for mediating the interaction between WDR5 and chromatin. It is interesting to note that in LoVo and K562 cells (Figures 4.2, 4.4, and 4.6), and in the Tansey laboratory's previous work in MV4:11 [74] cells, C6 is unable to completely displace WDR5 from chromatin at the conditions used, even at high concentrations (*e.g.,* 25 μM). While I can not confirm that there is full WDR5 binding by inhibitors, I do see a similar amount of residual WDR5 bound to chromatin across cell types, whether treated with as low as 2 μM or as high of 25 μM C6. Thus, for the prototypical WDR5-bound gene there are likely two modes of association of WDR5 with chromatin; one that is sensitive to C6, and another that is not. Two potential explanations are outlined below:

1) One possibility is that C6 is able to compete with some proteins that link WDR5 to chromatin through the WIN site, but that higher affinity WIN binding partners are refractory to competition. For example, it is clear that WDR5-WIN site affinity for C6 is greater than its affinity for MLL1, as indicated by C6 acting as a potent inhibitor of MLL1-driven HMT activity in MV4:11 cells [74]. However, I can not rule out the possibility that there are other proteins that bind WDR5 with a greater affinity than MLL1 and C6. Based on the enrichment of DNA-motifs, many of which are bound by transcription factors, in my WDR5 ChIP-seq data (Fig 2.8), it is possible that there DNA-binding proteins and/or transcription factors that bind to and recruit WDR5 to chromatin via the WIN site. These yet to be determined proteins could have a greater affinity for the WIN site than C6, and therefore remain bound to WDR5 upon C6 treatment, keeping a certain population of WDR5 bound to chromatin.

2) A second possibility is that there may be two distinct anchoring mechanisms for WDR5 on chromatin, only one of which is WIN-site mediated. We do not know the molecular mechanism of WDR5 recruitment to its target genes, but the precise and conserved registration of WDR5 at prototype genes (Figures 2.6 and 3.6) and the enrichment of DNA-motifs at these sites (Fig 2.8) suggests that there may be a DNA-binding protein that recruits WDR5 to chromatin. It is possible that WDR5 is initially recruited to chromatin through interactions between the WBM and a DNA-binding protein, and then subsequently additional WDR5 molecules load on via the WIN site. Alternatively, the binding pocket to which other identified WDR5-binding proteins bind and the mechanism for WDR5 recruitment to chromatin in other complexes — such as with CHD8 [45, 46], CBX8 [48], or INO80 [32, 47], with the long non-coding RNAs NeST [52], GClnc1 [84], or HOXD-AS1 [161], or as part of the ATAC [40, 41], WHHERE [49], or PRC1.6 [42-44] complexes — has yet to be determined, and the WBM may be the site these proteins bind and use to recruit WDR5 to chromatin. In addition, it has been previously established that WDR5 can selectively interact with certain proteins and long non-coding RNAs (such as HOTTIP) using its WBM site [51, 60, 133, 134, 162]. It is possible that these other proteins and/or long non-coding RNAs are able to directly or indirectly recruit WDR5 to chromatin without the use of the WIN site. Therefore, C6 treatment would not effect instances of WBM-dependent WDR5 binding. Clearly, a deeper understanding of the biochemical context in which WDR5 associates with chromatin is needed to address this issue.

Through these experiments, I have begun to answer my overarching questions: what genes are controlled by WDR5 and what are the primary transcriptional consequences of WDR5 binding to chromatin? I have determined that I have the right tools to ask these questions, as I have identified conserved genes bound by WDR5 and have confirmed that C6 does indeed work in these cell lines to reduce the levels of chromatin-bound WDR5. While WDR5 is bound to a conserved set of 94 genes, binding only suggests regulation. In order to determine the primary cellular effects of WDR5

binding to chromatin, changes in translation must be measured. Are all 94 genes that are bound by WDR5 actually regulated by WDR5? In Chapter VI, I will now use the tools and data I have established in Chapters III, IV, and V to determine which of the 94 conserved WDR5-bound genes are true WDR5 target genes, and what the effects of stable, chromatin-bound WDR5 are, through identifying the primary transcriptional consequences of displacing WDR5 from chromatin.

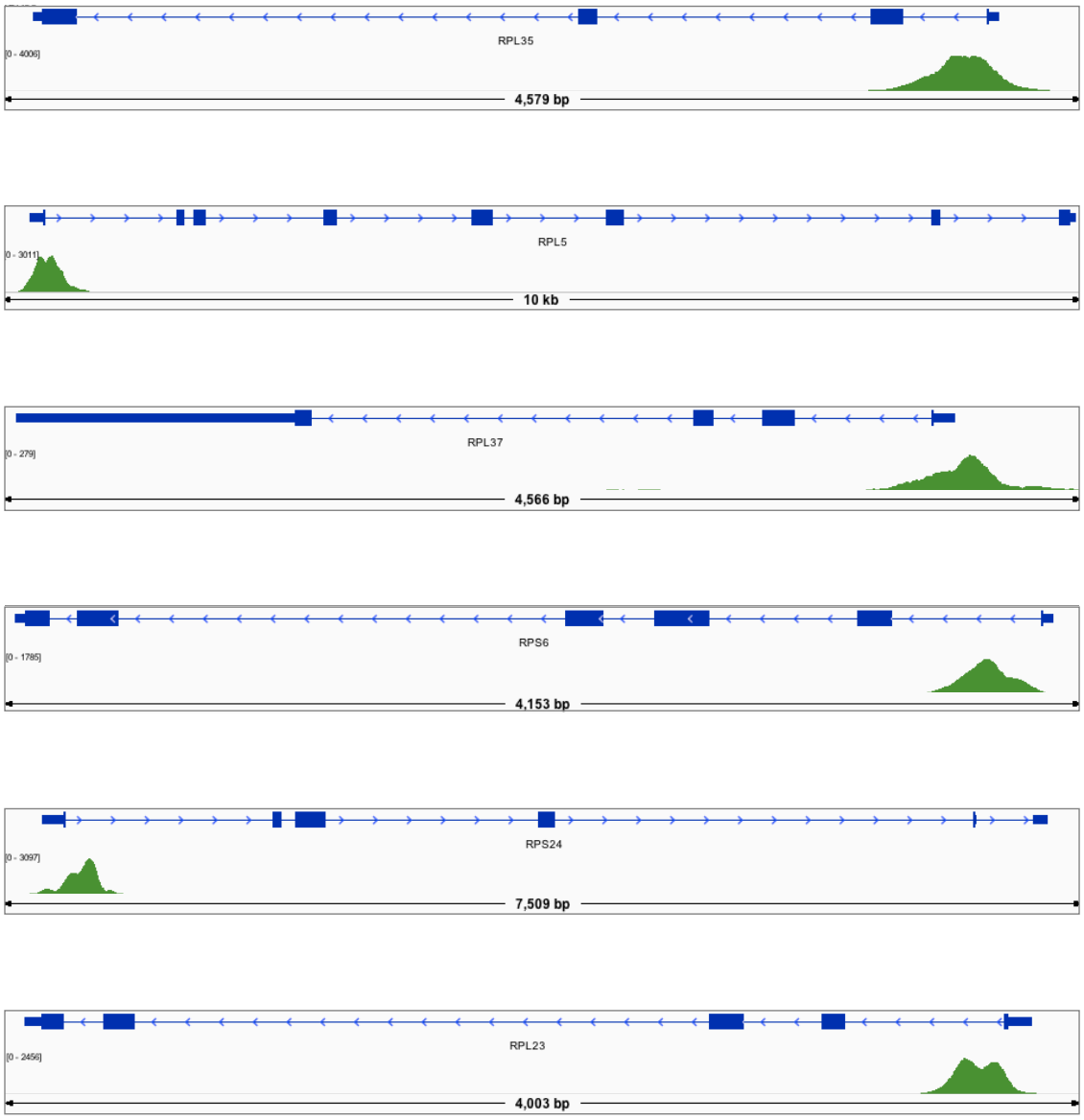
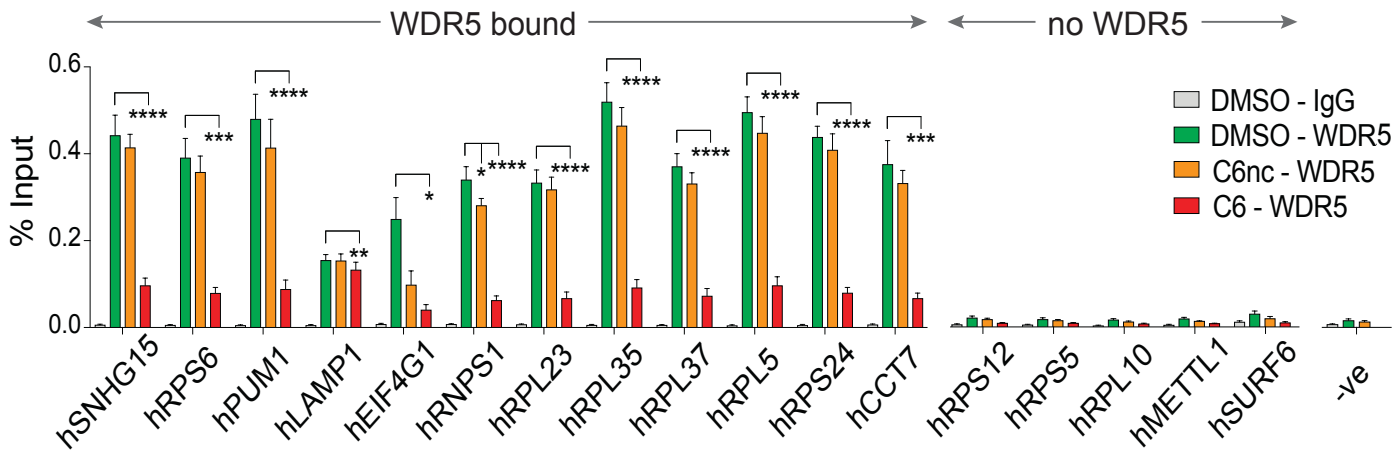


Figure 4.1. Representative screen shots of normalized WDR5 ChIP-seq peaks in LoVo cells. The figure shows six Integrated Genomics Viewer (IGV) screenshots of normalized ChIP-Seq data for WDR5. The data range was set as indicated on the left, [0-X], in each screen shot. Each image encompasses the entire gene of interest, with the genomic region, in bp, noted at the top of each screen shot. For each gene, the direction of transcription is indicated with blue arrows and the exons are indicated by blue boxes. The 5' and 3' UTR are indicated by slightly thinner blue boxes at either end of the gene.



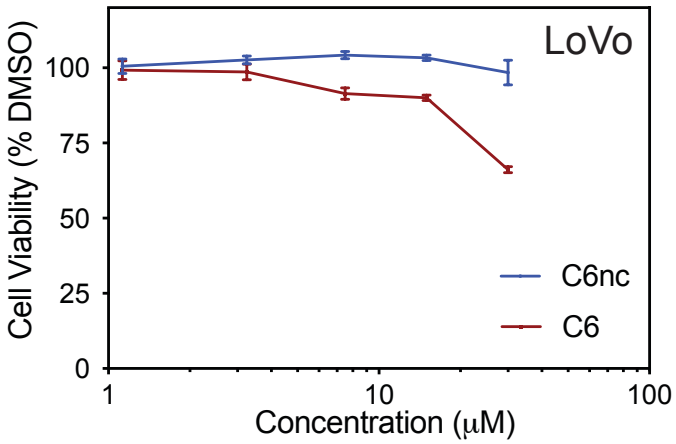


Figure 4.2. (*Top*) ChIP for WDR5 was performed in LoVo cells treated with DMSO, or 25 μM C6nc or C6, for 16 hours. Co-precipitating DNA was detected by Q-PCR and expressed as a percentage of signal in input chromatin. Genes are grouped according to whether or not they are bound by WDR5 in our LoVo ChIP-Seq data. "-ve" corresponds to a primer set within the *SNHG15* gene body that does not bind WDR5. Data are presented as mean ± SEM, $n = 4$ independent ChIP experiments.

(*Middle*) Dose response of LoVo cells to C6. Cells were treated with increasing concentrations of C6 or C6nc, up to 30 μM, for three days, before Cell Titer Glo was performed to assess cell viability. GI_{50} for C6 in LoVo cells was determined to be > 30 μM. Error bars represent SEM. $n = 3$.

(*Bottom*) One-Way ANOVA followed by Dunnett's Post-Hoc Test was performed on WDR5 ChIP data from each gene to determine the statistical significance of WDR5 displacement upon C6/C6nc vs DMSO treatment.
* = p<0.05, ** = p<0.01, *** = p<0.001, **** = p<0.0001.
The 'h' before the gene name indicates the primers used were designed to target human loci.

	RM One-Way ANOVA WDR5 ChIP in LoVo cells			**Dunnett's Post-Hoc Test**	
Gene of Interest	**F (DFn, DFd)**	**P value**	**P value summary**	**DMSO vs. C6nc**	**DMSO vs. C6**
hSNHG15	F (2, 6) = 111.6	P<0.0001	****	ns	****
hRPS6	F (2, 6) = 54.66	P=0.0001	***	ns	***
hPUM1	F (2, 6) = 66.55	P<0.0001	****	ns	****
hLAMP1	F (2, 6) = 17.24	P=0.0033	**	ns	**
hEIF4G1	F (2, 6) = 6.49	P=0.0316	*	ns	*
hRNPS1	F (2, 6) = 120.6	P<0.0001	****	*	****
hRPL23	F (2, 6) = 151.8	P<0.0001	****	ns	****
hRPL35	F (2, 6) = 130.6	P<0.0001	****	ns	****
hRPL37	F (2, 6) = 166.8	P<0.0001	****	ns	****
hRPL5	F (2, 6) = 169.5	P<0.0001	****	ns	****
hRPS24	F (2, 6) = 107.3	P<0.0001	****	ns	****
hCCT7	F (2, 6) = 47.73	P=0.0002	***	ns	***

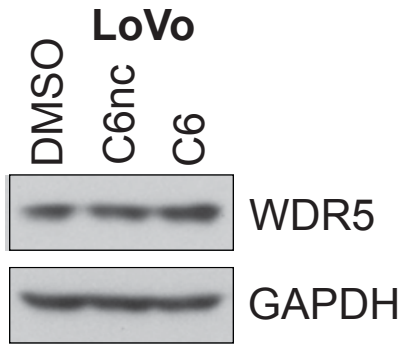


Figure 4.3. Immunoblotting of steady-state WDR5 levels in LoVo cells treated with DMSO, or 25 μM C6nc or C6, for 16 hours. GAPDH is a loading control.

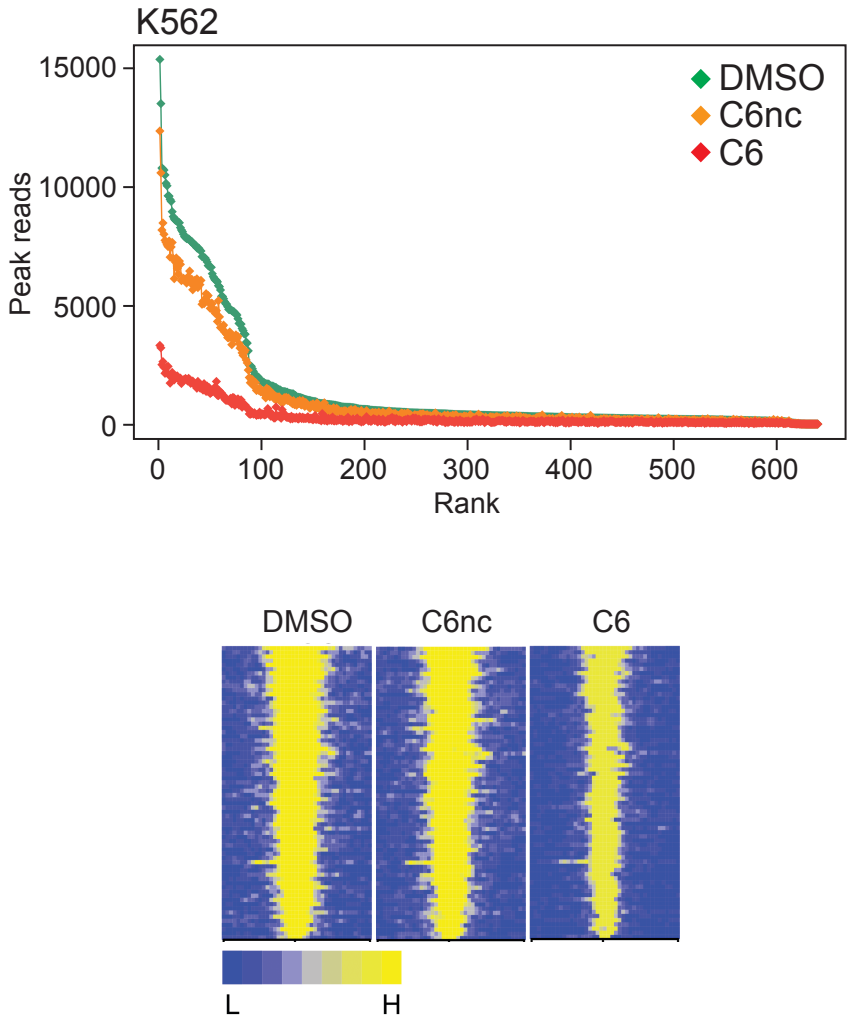


Figure 4.4. (*Top*) Scatterplot of normalized average read counts for WDR5 binding peaks in K562 cells treated for four hours with DMSO, 2 µM C6nc, or 2 µM C6, as determined by ChIP-Seq. Peaks are ranked based on read counts in DMSO-treated cells. $n = 3$ independent ChIP experiments. (*Bottom*) Heatmap of WDR5 ChIP-Seq peak intensity in K562 cells treated for four hours with DMSO, 2 µM C6nc, or 2 µM C6, representing the combined average of normalized peak intensity in 100-bp bins ± 2 kb around the center of peaks, for the 94 conserved WDR5-bound genes. Genes ranked based on DMSO-treated cells.

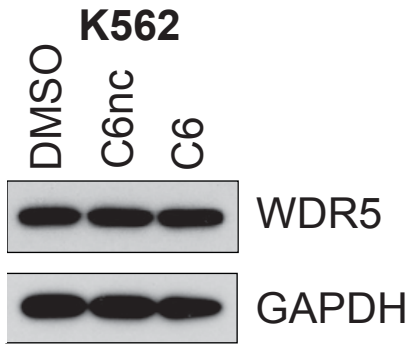


Figure 4.5. Immunoblotting of steady-state WDR5 levels in K562 cells treated with DMSO, or 2 μM C6nc or C6, for 4 hours. GAPDH is a loading control.

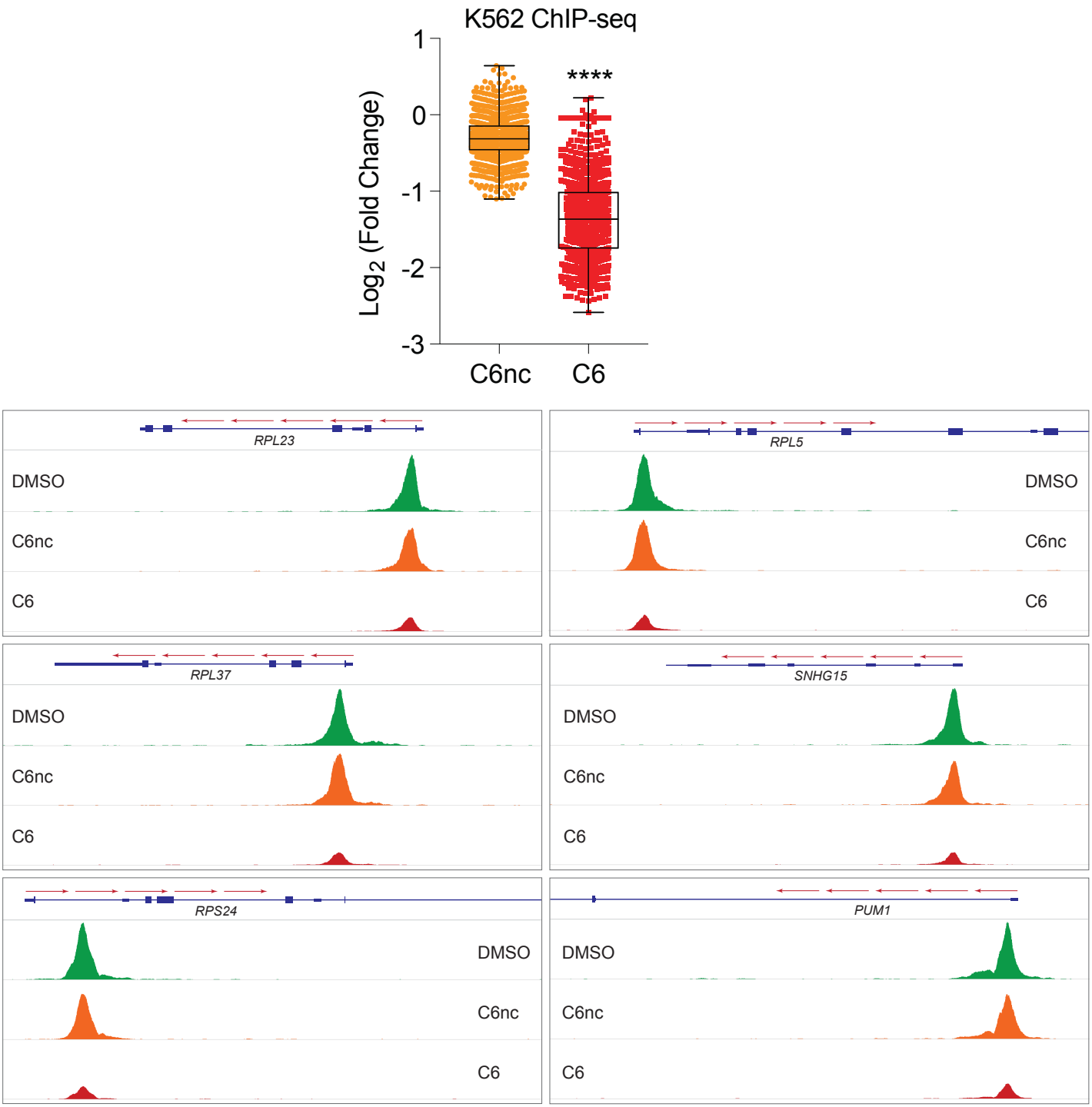


Figure 4.6. (*Top*) Box and whisker plot, showing the $\log_2$-fold change in WDR5 ChIP-Seq peak intensity in K562 cells, comparing C6nc and C6 treatments. The difference in signal for each peak is represented as a dot in the scatter plot. The box extends from the 25th to the 75th percentile, with the median marked by the middle line; whiskers extend from minimum to maximum points. Wilcoxon test shows a significant difference in the fold change of C6nc/DMSO versus C6/DMSO, **** = $p<0.0001$. (*Bottom*) Impact of C6 on WDR5 binding to chromatin in K562 cells. The figure shows six Integrated Genomics Viewer (IGV) screenshots of normalized ChIP-Seq data for WDR5 in K562 cells treated for four hours with DMSO, 2 µM C6nc, or 2 µM C6. The same data range was set for all samples. Each image encompasses 7,308 bp from the indicated genomic regions. For each gene, direction of transcription is indicated with red arrows. Only part of the *RPL5*, *RPS24*, and *PUM1* genes are shown. indicated by blue boxes. Exons are indicated by blue boxes, while the 5' and 3' UTR are indicated by slightly thinner blue boxes at either end of the gene.

CHAPTER VI

Select Protein Synthesis Genes are Bound by WDR5 in Multiple Cells Types and Inhibited by WIN Site Blockade

Abstract

As WIN site inhibition of WDR5 proved to displace a large portion of WDR5 from chromatin, I utilized a small molecule WIN site inhibitor in order to identify WDR5 target genes. Specifically, C6 was used to determine whether any of the 94 conserved WDR5-bound genes are transcriptionally regulated by WDR5 binding. In LoVo cells, ChIP-Seq coupled with SLAM-Seq allowed me to determine which genes are early, likely primary, WDR5 targets. In K562 cells, ChIP-Seq coupled with RNA-Seq allowed for the identification of genes that are long-term WDR5 targets. These experiments were combined with MV4:11 ChIP-Seq, PRO-Seq, and RNA-Seq results, published by the Tansey laboratory this year. Taken together, the resulting data support the notion that there is a conserved cohort of protein synthesis genes, mostly encoding large and small subunit ribosomal proteins, that are directly and positively regulated by WDR5 binding to chromatin.

Introduction

This chapter will detail which genes are commonly regulated by WDR5, via the WIN site, across multiple cancer cell lines, allowing for a more thorough understanding of the overarching functions of WDR5. In Chapters III, IV, and V, I outlined which genes are bound by WDR5 across cell types, and demonstrated the successful use of the WIN site SMI, C6, to displace WDR5 from chromatin. The critical question addressed here is whether WDR5 binding to a gene is associated with its transcriptional output.

C3 and C6 were recently used to show that WIN site inhibitors rapidly displace WDR5 from chromatin, independent of changes in H3K4 methylation, in MV4:11 cells [74]. Aho, et. al. also determined that C6 is a potent inhibitor of MLL1-driven HMT activity in MV4:11 cells and forecast that WIN site blockade could have utility against multiple cancer types [74]. However, in this publication, C6 was only characterized in this one cell line, and therefore, the authors can only hypothesize these SMIs will similarly affect transcription in other cell lines - they were also unable to identify a set of conserved genes controlled by WDR5, still leaving us with the inability to predict the primary transcriptional consequences of WIN site blockade. Utilizing C6 in my studies has not only helped me further interrogate

the mechanism of WDR5 binding to chromatin, ascertain whether the WIN site of WDR5 is necessary for this binding, and identify the conserved sites of WDR5 binding that are dependent on the WIN site of WDR5, but has helped determine which genes respond to WDR5 binding via the WIN site, and has added to the validity of this novel mode of WDR5 inhibition in cancer cells.

As WDR5 has been found to be over-expressed in cancer [67-70, 82, 107, 119-121], involved in malignant processes [26, 56, 59, 60, 78, 122], and is currently being assessed as a novel target for cancer therapies [27, 74, 89, 90], is it important to not only determine its evolutionarily conserved role when bound to chromatin, but also to examine the effects of displacing WDR5 from chromatin, if a WDR5 inhibitor is ever to be used in the clinic. While various cancer contexts have been proposed to be targets for WDR5 WIN site inhibitors, a lack of understanding of the conserved WDR5 target genes and of the primary effects of WIN site inhibitors continues to hamper their utility. In this study, a 'target gene' is considered to be any gene that is both bound by WDR5 and exhibits a rapid reduction/induction in transcription in response to WDR5 inhibition. The identification of genes regulated by WDR5 have been of interest in previous studies, but such studies are limited. In the human bladder cancer cell line, UM-UC-3, 136 genes were upregulated and 42 genes were downregulated in response to siRNA knock-down of WDR5 [70]. In these cells, the WDR5 regulated genes were mainly associated with transcription, the cell cycle, cell adhesion, and were anti-apoptotic [70]. In the human lymphoma cell line, MV4:11, 72 genes were upregulated and 462 genes were downregulated [74]. Using WIN site inhibitors, Aho, et al. displaced WDR5 from chromatin - this led to a decrease the expression of associated genes, causing translational inhibition, nucleolar stress, and p53 induction [74]. Only 1 WDR5-bound gene was induced by the removal of WDR5 from chromatin, while 59 genes were repressed - of the 59 repressed genes, 39 encoded subunits of the ribosome [74]. The additional analysis, outlined in this chapter, adds to the list of known WDR5-WIN-site regulated genes, and identifies which genes bound by WDR5 across multiple cell lines are early and long-term WDR5 targets.

Results

In order to answer the overarching question, which genes, bound by WDR5 across multiple cell lines, are targets of WDR5, I decided to determine which genes are regulated by WDR5 bound to chromatin via the WIN site. Since WDR5 can not be fully depleted in a cell without cell death (and the dTAG system of rapidly degrading a protein wasn't not invented in 2014 when I started my thesis research), I decided to use a transient, fast-acting method of displacing WDR5 from chromatin, which importantly did not displace chromatin-bound WDR5 completely - since some WDR5 remains bound to chromatin, we can say the effects on translation are not simply due to cell death in response to total loss of an

essential gene. As outlined in Chapter V, the SMI, C6, is able to do just that. Also, to avoid transcriptional complications from the induction of cell death, I decided to use cells that are not phenotypically sensitive to C6 treatment. Based on sensitivity (as defined in Aho, et al., 2019 [74]) I chose to further interrogate the effects of WDR5 inhibition, in order to determine which genes WDR5 regulates, in two human cell lines: LoVo (GI_{50}>25 μM; Fig 4.2) and K562 (GI_{50} ~25 μM). I chose K562 cells, as they offered an insensitive companion cell line to the published MV4:11 data [74]. I also chose LoVo cells, to be a companion cell line to both leukemia cell lines, in order to layer in another cancer cell type which also has dysregulated c-MYC (as was previously mentioned, the Tansey laboratory is also interested in the WDR5-MYC connection).

To examine primary transcriptional responses to chromatin-bound WDR5, I coupled C6 treatment with Thiol (SH)-Linked Alkylation for the Metabolic Sequencing of RNA (SLAM-Seq [163]), an approach that allows for the quantification of changes in newly-synthesized mRNAs after exposure to inhibitors. LoVo cells were pre-treated with DMSO or C6 for one hour, after which new mRNAs were labeled with 4-thiouridine (S4U) for three hours before harvest and completion of the SLAM-Seq pipeline. Cells were continuously exposed to DMSO and C6 for all four hours, before cells were collected and RNA was extracted. I chose this time point because within this timeframe, I expect to see both primary as well as early secondary effects of WIN site blockade, and I am most interested in the immediate effects of WDR5 WIN site inhibition. Using SLAM-Seq, I identified 48 genes that are induced, and 261 genes that are repressed by C6 treatment (Fig 5.1). Figure 5.1 shows a heatmap, displaying z-transformed gene expression for significantly changed genes in C6 versus DMSO (FDR < 0.05), for all three biological replicates (rep1–rep3) of SLAM-Seq, and examines the impact of four hours of DMSO or 25 μM C6 treatment of LoVo cells. As all three replicates were highly similar, all three were used to determine which transcripts responded to WDR5 inhibition. As fewer significant changes in nascent transcripts were detected at this time point than I expected — I previously determined there to be 2,162 sites of WDR5 binding across the LoVo genome (assigned to 1,383 genes) — I became interested in which categories of genes were being so rapidly regulated by chromatin-bound WDR5. As is displayed in Figure 5.2, the genes induced by C6 in LoVo cells showed weak clustering in fairly uninformative GO categories, such as ‘response to transition metal nanoparticle,’ ‘regulation of growth,’ ‘response to inorganic substance,’ ‘cytoplasm,’ ‘regulation of cell morphogenesis,’ and ‘nucleus,’ all of which had $-\log_{10}$(P-value)s of less than 6. Conversely, C6-repressed genes, however, clustered strongly in GO terms, including ‘nuclear transcribed mRNA catabolic process,’ ‘cotranslational protein targeting to membrane,’ ‘protein targeting to ER,’ ‘translational initiation,’ ‘rRNA processing,’ ‘ribosome,’ ‘translation,’ and ‘structural constituent of ribosome,’ all of which had $-\log_{10}$(P-value)s greater than 35 (Fig 5.2). Figure 5.2 (*top*) shows these top eight GO

enrichment clusters for gene transcripts significantly repressed by C6 treatment of LoVo cells, as determined by SLAM-Seq. Figure 5.2 (*bottom*) also shows the top six GO enrichment clusters for gene transcripts significantly induced by C6 treatment of LoVo cells, as determined by SLAM-Seq. Results are ranked by $-\log_{10}$(P-value) and the numbers in italics represent the number of repressed genes in each category.

As expected, based on the GO analysis of the 94 conserved genes, the genes positively regulated by the presence of WDR5 on chromatin included those connected to protein synthesis. Figure 5.3 indicates how many WDR5-bound genes in LoVo cells (determined by ChIP-Seq) are primary targets of WDR5, via its WIN site (as identified via SLAM-Seq experiments). 70 of the 1,383 genes bound by WDR5 in LoVo cells exhibit a significant repression in nascent transcript levels upon the displacement of WDR5 from chromatin (Fig 5.3, *top, left*). Only 7 of the 1,383 genes bound by WDR5 in LoVo cells exhibit a significant induction of nascent transcript levels upon the displacement of WDR5 from chromatin (Fig 5.3, *bottom, left*). Figure 5.3 also indicates how many of the 94 conserved WDR5-bound genes are primary targets of WDR5 in LoVo cells. ~50% (44 of the 94) of the WDR5-bound genes, conserved across cell lines (as identified in Chapter IV), are repressed in LoVo cells when the levels of chromatin-bound WDR5 decrease (Fig 5.3, *top, right*). None of the WDR5-bound genes, conserved across cell lines, are induced in LoVo cells when the levels of chromatin-bound WDR5 decrease (Fig 5.3, *bottom, right*).

Of the 70 primary WDR5 target genes in LoVo cells, 56% (39) were ribosome protein genes, and of the 44 conserved, primary target genes in LoVo cells, 80% (35) were ribosome protein genes. Therefore, unsurprisingly, GO analysis of the 70 WDR5-bound, C6-repressed gene transcripts in LoVo cells, indicated that primary WDR5 target genes are those involved in protein synthesis (Fig 5.4, *top*). Additionally, GO analysis of the 191 non-WDR5-bound, C6-repressed gene transcripts in LoVo cells indicated that indirect WDR5 target genes are also involved in translation and RNA processing, however they are involved in autophagy and nucleus organization as well (Fig 5.4, *bottom, left*). GO analysis of the 261 C6-repressed gene transcripts in LoVo cells without the 41 RPGs, indicated that the non-RPG WDR5 target genes are those involved in autophagy, RNA processing, and translation (Fig 5.4, *bottom, right*). Figure 5.4 (*top*) shows the top eight GO enrichment categories for WDR5-bound, C6-repressed gene transcripts in LoVo cells, which include 'cotranslational protein targeting to membrane,' 'protein targeting to ER,' 'nuclear transcribed mRNA catabolic process,' 'translational initiation,' 'viral transcription,' 'multi-organism metabolic process,' 'translation,' and 'ribosome.' Figure 5.4 (b*ottom, left*) shows the top eight GO enrichment categories for non-WDR5-bound, C6-repressed gene transcripts in LoVo cells, including 'translation,' 'autophagy,' 'macromolecular complex subunit organization,' 'nucleus organization,' 'ncRNA metabolic process,' 'RNA processing,' 'cellular component biogenesis,' and 'poly(A) RNA

binding. Figure 5.4 (b*ottom, right*) shows the top eight GO enrichment clusters for C6-repressed gene transcripts, minus RPG gene transcripts, in LoVo cells, which include 'autophagy,' 'poly(A) RNA binding,' 'nucleoplasm,' 'RNA processing,' 'cytoplasm,' 'membrane,' 'translation,' and 'protein binding.' In all three graphs, results are ranked by $-\log_{10}$(P-value) and numbers in italics represent the number of repressed genes in each category. In Figure 5.5, a Ribosomogram shows which of the small (*top*) and large (*bottom*) ribosome subunit RPGs are bound by WDR5 in LoVo and MV4:11 [74] cells, as determined by ChIP-Seq. It also shows which RPGs are directly repressed upon small molecule WIN-site inhibition, as determined by C6 treatment followed by SLAM-Seq in LoVo cells (Fig 5.1), and C3 treatment followed by PRO-Seq in MV4:11 cells [74].

To examine persistent transcriptional changes induced by chromatin-bound WDR5, in an additional cell line and one that can accompany the MV4:11 sensitive cell line, I utilized ChIP-Seq and RNA-Seq data obtained from experiments a post-doc in the Tansey laboratory, Dr. Caleb Howard, performed. K562 cells were treated with 2 μM C6, 2 μM C6nc, or DMSO for three days and RNA-Seq analysis performed (Fig 5.6). Cells were treated at this concentration for three days in order to more appropriately compare RNA-Seq results in K562, insensitive cells, to the published RNA-Seq results in MV4:11, sensitive cells, which were also treated for 3 days with 2 μM C6 before RNA-Seq was performed [74]. The negative control compound, C6nc, altered the expression of just one gene. The active compound, in contrast, had a much broader transcriptional impact - 65 genes were induced and 187 were repressed by C6 treatment (Fig 5.6). The long-term transcriptional changes observed in K562 cells mirrored the shorter term study in LoVo cells. In K562 cells, genes induced by C6 are not strongly clustered (Fig 5.7, *bottom*), but repressed genes are strongly enriched in GO terms connected to protein synthesis (Fig 5.7, *top*). Figure 5.7 (*top*) shows the top eight GO enrichment categories for repressed gene transcripts in K562 cells, which include 'SRP-dependent cotranslational protein targeting to membrane,' 'protein targeting to ER,' 'translational initiation,' 'viral transcription,' 'nuclear-transcribed mRNA catabolic process,' 'ribosome,' 'cytosolic large ribosome subunit,' and 'multi-organism metabolic process.' Figure 5.7 (*bottom*) shows the top five GO enrichment categories for induced gene transcripts in K562 cells, including 'oxygen transport,' 'heme biosynthetic process,' 'protoporphyrinogen IX biosynthetic processes,' 'porphyrin-containing compound biosynthetic processes,' and 'hemaglobin metabolic process.' Results are ranked by $-\log_{10}$(P-value) and the numbers in italics represent the number of repressed genes in each category.

Figure 5.8 indicates how many WDR5-bound genes in K562 cells (determined by ChIP-Seq) are long-term targets (long-term: persistently decreased by WIN site inhibition) of WDR5 (as identified via RNA-Seq experiments). In agreement with LoVo cell results, less than 5% of K562 WDR5-bound gene transcripts (3 out of 65) were induced in

response to WIN-site inhibition, while 28% of K562 WDR5-bound gene transcripts (52 out of 187) were repressed in response to WIN-site inhibition (Fig 5.8, *left*). 52 of the 648 total genes bound by WDR5 in K562 cells exhibit a significant repression in transcript levels upon the displacement of WDR5 from chromatin (Fig 5.8, *top, left*). Only 3 of the 648 total genes bound by WDR5 in K562 cells exhibit a significant induction in transcript levels upon the displacement of WDR5 from chromatin (Fig 5.8, *bottom, left*). Figure 5.3 also indicates how many of the 94 conserved WDR5-bound genes are targets of WDR5 in K562 cells. ~50% (46 of the 94) of the WDR5-bound genes, conserved across cell lines (as identified in Chapter IV), are repressed in K562 cells when the levels of chromatin-bound WDR5 decrease (Fig 5.8, *top, right*). None of the WDR5-bound genes, conserved across cell lines, are induced in K562 cells when the levels of chromatin-bound WDR5 decrease (Fig 5.8, *bottom, right*). In addition, gene set enrichment analysis of K562 genes modulated in response to C6 treatment (RNA-Seq), against all WDR5-bound genes in K562 cells (ChIP-Seq), indicates that there is a highly significant tendency for WDR5-bound genes to be transcriptionally repressed by C6 in K562 cells (Fig 5.9). In Figure 5.9, genes were ranked by $\log_2$-fold change, an FDR cutoff of q = 0.0 was applied, and an NES score of -2.92 was achieved. In Figure 5.10, a Ribosomogram shows which of the small (*top*) and large (*bottom*) ribosome subunit RPGs are bound by WDR5 in LoVo, K562, and MV4:11 [74] cells, as determined by ChIP-Seq. Further, it shows which RPGs are repressed upon small molecule WIN-site inhibition, as determined by C6 treatment followed by SLAM-Seq in LoVo cells (Fig 5.1), C6 treatment followed by RNA-Seq in K562 cells (Fig 5.6), and C3 treatment followed by PRO-Seq in MV4:11 cells [74]. These data indicate that there is a conserved cohort of protein synthesis genes, mostly RPGs, that are directly and positively regulated by WDR5 binding to chromatin.

Although WDR5 binds to, and is displaced from 94 conserved genes upon WIN inhibition, I determined that WDR5 regulates 33 conserved genes, across LoVo (SLAM-Seq), K562 (RNA-Seq), and MV4:11 (RNA-Seq) cells. These genes include: *RPL38, RPL35, CCT4, RPL8, RPL27, RPL5, RPL18, RPL15, RPL36, RPS14, SNRPB, RPL35A, RPL37A, RPS4X, RPL13, RPL31, RPL9, RPL41, EEF2, RPS24, CSNK1E, RPL11, RPL34, RPS3, RPL32, RPL3, RPS15, RPL29, RPL7, RPL24, RPL26, RPS10,* and *RPL6* (Fig 5.11). Notably, only four of these genes are not constituents of the ribosome. Further analysis of the regulation of genes was also performed in all three three cell types:

1) LoVo: Out of the 44 conserved genes that were repressed by C6 in LoVo cells (Fig 5.3, *top, right*), 35 were RPGs. The nine non-RPG genes are *SERBP1*, *EEF1G*, *EEF2*, *NACA*, *SNRPB*, *EIF4G1*, *EIF3D*, *CCT4*, and *CSNK1E*. This means that 50 conserved genes are not directly regulated by the association of WDR5 with chromatin. Interestingly, 16 of these 50 genes are located just upstream of or within an RPG. While I assigned WDR5 ChIP-Seq peaks to all nearby genes (the peak was assigned to a gene if it fell within 2 kb upstream of the TSS or within the open reading frame of the

gene), as I felt it inappropriately biased the data to only assign peaks to the first nearby gene, the overlap of ChIP-Seq and SLAM-Seq data has helped determine which genes WDR5 truly effects, when assigned to more than one gene. The following conserved, C6-repressed RP genes are also assigned to a conserved non-C6-repressed gene: *RPL18* and *SPHK2*, *RPL13* and *SNORD68*, *RPL9* and *LIAS*, *RPL23* and *SNORA21*, *RPS3* and *SNORD15A*, *RPL3* and *SNORD43*, *RPL32* and *SNORA7A*, *RPL35A* and *IQCG*, *RPL15* and *NKIRAS1*, *RPL35* and *ARPC5L*, *RPL41* and *ZC3H10*, *RACK1* and *SNORD96A*, *RPL7* and *RDH10*, as well as *RPS2* and *SNHG9*, *SNORA78*, and *SNORA64*.

2) K562: Out of the 46 conserved genes that were repressed by C6 (Fig 5.8, *top, right*), 34 were RPGs. The twelve non-RP genes are *EEF2*, *SNRPB*, *EIF4G3*, *EIF3D*, *CCT4*, *CSNK1E*, *SF3B3*, *RNF220*, *SNHG15*, *CCT7*, *SNORA21*, and *SNHG17*. This means that 48 conserved genes are not directly regulated by the association of WDR5 with chromatin. However, 15 of these 48 genes are located just upstream of or within an RPG. Again, the overlap of ChIP-Seq and RNA-Seq data has helped determine which genes WDR5 truly effects, when assigned to more than one gene. The following conserved, C6-repressed RP genes are also assigned to a conserved non-C6-repressed gene: *RPL18* and *SPHK2*, *RPL13* and *SNORD68*, *RPL9* and *LIAS*, *RPL23* and *SNORA21*, *RPS3* and *SNORD15A*, *RPL3* and *SNORD43*, *RPL32* and *SNORA7A*, *RPL35A* and *IQCG*, *RPL15* and *NKIRAS1*, *RPL35* and *ARPC5L*, *RPL7* and *RDH10*, *SNHG15* and *SNORA9*, *SF3B3* and *COG4*, *CCT7* and *PRADC1*, as well as *RPL12* and *LRSAM1*.

3) MV4:11: Almost all of the genes repressed by compound treatment in MV4:11 cells are bound by WDR5, and a majority encode ribosome subunits [74]. Moreover, within the RPGs, a consistent pattern of genes occupied by WDR5 scored as repressed in both PRO-Seq and RNA-Seq experiments, indicating that these genes are early and persistent transcriptional targets of WIN site blockade (Fig 5.10, [74]). As determined by RNA-Seq, 52 of the 94 conserved genes were repressed in MV4:11 cells. Out of the 52 conserved, C6-repressed, 38 were RPGs. The fourteen non-RPG genes are *EEF2*, *SNRPB*, *EIF4G3*, *EIF4G1*, *CCT4*, *CSNK1E*, *SF3B3*, *RNF220*, *CCT7*, *PUM1*, *SERBP1*, *YWHAE*, *RNPS1*, and *SNHG17*. This means that 42 conserved genes are not directly regulated by the association of WDR5 with chromatin. However, 18 of these 42 genes are located just upstream of or within an RPG. The overlap of ChIP-Seq and RNA-Seq data helped determine which genes WDR5 truly effects, when assigned to more than one gene. The following conserved, C6-repressed RPG peaks are also assigned to a conserved non-C6-repressed gene: *RPL18* and *SPHK2*, *RPL13* and *SNORD68*, *RPL9* and *LIAS*, *RPL23* and *SNORA21*, *RPS3* and *SNORD15A*, *RPL3* and *SNORD43*, *RPL32* and *SNORA7A*, *RPL35A* and *IQCG*, *RPL15* and *NKIRAS1*, *RPL35* and *ARPC5L*, *RPL7* and *RDH10*, *SF3B3* and *COG4*, *CCT7* and *PRADC1*, *RPL17* and *SNORD58B*, *RPS8* and *SNORD55*, as well as *RPS2* and *SHNG9*, *SNORA78*, and *SNORA64*. As

none of the snoR(A/D)s were repressed, chromatin-bound WDR5 does not specifically induce the expression of snoRNAs in MV4:11 cells.

Discussion

In order to determine the overarching, evolutionarily conserved role for chromatin-bound WDR5, I have compared stable (ChIP-able) WDR5 bound-loci, across multiple cell lines and species, with the changes in transcription that result from displacing WDR5 from chromatin. Overall, I have determined that WDR5 acts mainly to induce transcription at target genes, as indicated by most transcripts being repressed upon the displacement of WDR5 from chromatin. These experiments with C6 show that displacement of WDR5 from chromatin results, on average, in less than a two fold decrease in transcription at its target genes, which are mainly RPGs, implying that the function of WDR5 is not to turn these genes on or off, but to modulate their expression within a fairly narrow window. A chemically-distinct WIN site inhibitor, OICR-9429 [160], has also been shown to have similar effects on RPG mRNA levels [74]. Therefore, I confidently propose that the conserved role for chromatin-bound WDR5 is to precisely modulate the expression of protein synthesis and ribosome protein target genes.

To arrive at this conclusion, I first performed SLAM-Seq in C6 treated LoVo cells, an insensitive CRC cell line (Fig 4.2 and 5.1). SLAM-Seq proved successful in detecting changes in nascent mRNAs and allowing for the identification of two populations of transcripts: those induced by the removal of WDR5 from chromatin, and those repressed. There were far fewer induced genes, and while some GO clustering was possible, the results were vague and not highly significant (Fig 5.2, *bottom*). There were roughly 5.5x more repressed (than induced) genes identified - unsurprisingly, GO analysis showed that the genes that are positively regulated by the presence of WDR5 on chromatin in LoVo cells include those involved in protein synthesis (Fig 5.2, *top*).

By overlaying WDR5 ChIP-Seq data onto these mRNA changes, to separate primary from secondary effects, six main themes emerged. 1) WDR5-bound genes that respond to WIN site inhibition in a short timeframe are generally repressed. 70 of the genes repressed by C6 treatment in SLAM-Seq are bound by WDR5, compared to just 7 of the genes that are induced by C6 exposure (Fig 5.3, *left*). Focusing on the 94 conserved WDR5-bound genes (Fig 5.3, *right*), this relationship becomes clearer. Almost half (44) of the conserved WDR5-bound genes are repressed by C6 in LoVo cells, while none are induced (Fig 5.3, *right*). 2) WDR5-bound and C6-repressed genes are strongly enriched in those connected to protein synthesis (Fig 5.4, *top*). 3) Early secondary effects of WIN site inhibition (*i.e.,* at non-WDR5-bound genes) are generally not as tightly biologically clustered as the primary effects, although among the repressed genes there is modest enrichment in GO categories relating to autophagy and protein synthesis (Fig 5.4, *bottom, left*). 4) Primary WDR5 target

genes include a specific subset of small and large subunit RPGs (Fig 5.5). Indeed, 39 of the C6-repressed genes in LoVo cells are WDR5-bound RPGs, and 30 of the C6-repressed genes in MV4:11 cells are WDR5-bound RPGs (Fig 5.5, [74]) — 25 of these RPGs are primary WDR5 targets in both LoVo and MV4:11 cells (Fig 5.5). 5) While a significant subset of LoVo, C6-repressed genes are RPGs, there is a separate subset of primary and early-secondary genes that are repressed upon the removal of WDR5 from chromatin (Fig 5.4, *bottom, right*). As the GO clusters (Fig 5.4, *bottom, right*) indicate that these genes are highly similar in category, significance, and number of genes per category to the early secondary effects of WIN site inhibition (Fig 5.4, *bottom, left*), it is highly likely that the majority of the reduced genes that cluster are either A) RPGs, or B) late (or secondary) targets of WDR5, rather than primary, non-RPG targets of WDR5. 6) Roughly half (44) of the 94 conserved genes were repressed by C6 (Fig 5.3, *top, right*). This means that the other half (50) of the conserved genes are not directly regulated by the association of WDR5 with chromatin. Importantly, 16 of these 50 genes (12/50 are snoRNAs) are assigned to a WDR5 peak that is also assigned to an RPG. Therefore, although WDR5 binds 94 conserved genes, at least 16 of these were assigned to a peak because of their presence within/near an RPG, rather than being a true primary WDR5 target in a LoVo cell. Further, none of the snoR(A/D)s were repressed - so while the expression of specific snoRNAs has been linked to certain cancers, chromatin-bound WDR5 does not specifically induce the expression of any snoRNAs in LoVo cells.

Second, RNA-Seq on K562 cells, an insensitive leukemia cell line, identified changes in steady-state transcription ([74], Fig 5.6). Again, two populations were discovered: those induced by the removal of WDR5 from chromatin, and those repressed. There were far fewer induced genes, and while some GO clustering was possible, the results, though heme-centric, were not highly significant (Fig 5.7, *bottom*). There were roughly 3x more repressed (than induced) genes identified - and again, unsurprisingly, GO analysis showed that the genes that are positively regulated by the presence of WDR5 on chromatin in K562 cells include those involved in protein synthesis (Fig 5.7, *top*).

By overlaying WDR5 ChIP-Seq data onto these mRNA changes, to separate WDR5-bound from non-WDR5-bound genes with changes in transcription, three main themes emerged. 1) There is a highly significant tendency for WDR5-bound genes to be transcriptionally modulated by C6 in K562 cells (Fig 5.8). The predominant effect of C6 on WDR5-bound genes is repression - 52 of the repressed genes are bound by WDR5 in K562 cells, compared to just 3 of the induced genes (Fig 5.8, *left*). Additionally, half of the 94 conserved WDR5-bound genes are repressed by C6 in K562 cells (Fig 5.8, *right*), while none are induced. 2) WDR5 target genes include a specific subset of small and large subunit RPGs (Fig 5.10). Again, I observed a very specific pattern of WDR5-bound RPGs that are repressed by WIN site blockade in K562 cells, which overlaps strongly with those found to be repressed by C6 in LoVo and MV4:11 cells (Fig 5.10).

Indeed, 37 of the C6-repressed genes in K562 cells are WDR5-bound RPGs, 39 of the C6-repressed genes in LoVo cells are WDR5-bound RPGs, and 41 (as determined by either PRO-Seq, RNA-Seq, or both) of the C6-repressed genes in MV4:11 cells are WDR5-bound RPGs (Fig 5.10, [74]) — 32 of these RPGs are WDR5 targets in K562, LoVo, and MV4:11 cells (Fig 5.10). 3) Roughly half (46) of the 94 conserved genes were repressed by C6 (Fig 5.8, *top, right*). This means that the other half (48) of the conserved genes are not directly regulated by the association of WDR5 with chromatin. Importantly, 15 of these 48 genes (11/48 are snoRNAs) are assigned to a WDR5 peak that is also assigned to an RPG. Therefore, although WDR5 binds 94 conserved genes, at least 15 of these were assigned to a peak because of their presence within/near an RPG, rather than being a true primary WDR5 target in a K562 cell. As only one of the snoR(A/D)s was repressed, chromatin-bound WDR5 does not specifically induce the expression of multiple snoRNAs in K562 cells.

Interestingly, although WDR5 is displaced from chromatin by C6 at all peaks assigned to the 94 conserved genes. only about half of the prototypical WDR5-bound genes are transcriptionally-repressed by WIN site inhibitor. This is true in all three cell lines interrogated. In LoVo, K562, and MV4:11 cells, the transcriptional effects of WIN site inhibition at WDR5-bound loci are confined mostly to RPGs. This finding implies that the function of WDR5 is either not the same at all genes to which it binds (A), or there is some aspect of finer transcriptional regulation by WDR5 that is not apparent under the conditions used in our experiments (B). A) WDR5 has already been identified as part of multiple chromatin-associating complexes - it is possible that a certain population of stable, chromatin-bound WDR5 is in a SET1/MLL or NSL complex, for example, and therefore is helping affect chromatin methylation or acetylation, rather than translation. Alternatively, WDR5 could bind within the ORF of one gene, but act as an enhancer for a different gene. This scenario could account for both non-responsive WDR5-bound genes, and responsive non-WDR5 bound genes. Furthermore, there are a significant number of WDR5 binding sites that fall within the ORF of more than one gene. While this causes the WDR5 ChIP-seq peak to be assigned to both genes, WDR5 may only bind to the correct location, or be in the right confirmation, to affect the transcription of one of these genes. Also, just because a protein is bound to a gene does not mean that it will affect the transcription of the gene. B) WDR5 could be poised on chromatin, waiting to effect transcription in response to a stimulus. Such stimulus could come from a certain step in the cell cycle, a growth signaling pathway, or stress (ie: nutrient deprivation, toxicity, induction of apoptosis). WDR5 could be bound to chromatin to help regulate normal fluctuations in the cell, such as cell cycle regulation of protein synthesis capacity [164], to help a cell respond quickly to abnormal cellular events, such as during periods of altered mTOR signaling, or both. In fact, the genes to which WDR5 binds but that do not respond to C6 treatment are enriched in the following GO functional annotation

categories: cell cycle process, chromosome organization, DNA replication, mitotic cell cycle, cellular response to stress, and DNA repair. In order to determine whether WDR5 plays a role in the transcription of these genes under different conditions further experiments should be performed with C6, such as post cellular-stress induction, cell cycle synchronization, or DNA damage. Regardless, for the WDR5-bound genes that do transcriptionally respond to the WIN inhibitor in the experiments detailed above, the pattern of response is highly consistent, allowing me to forecast with high confidence precisely which RPGs will be repressed by a WIN inhibitor in any cellular setting.

Notably, only 4 of the 33 conserved WDR5 targets genes (Fig 5.11) are not constituents of the ribosome: *CCT4*, *CSNK1E*, *EEF2*, and *SNRPB*. According to UniProt, CCT4 is a component of the chaperonin-containing T-complex (TRiC), a molecular chaperone complex that assists the folding of proteins upon ATP hydrolysis, and mediates the folding of WRAP53/TCAB1, thereby regulating telomere maintenance [165]. CSNK1E is a casein kinase that utilizes acidic proteins, such as caseins, as substrates [165]. It can phosphorylate a large number of proteins, such as DVL1, DVL2, PER1, and PER2, participate in Wnt signaling, inhibit cytokine-induced granuloytic differentiation, and is central component of the circadian clock through the regulation of the speed and rhythmicity of PER1 and PER2 phosphorylation, nuclear transport, and degradation [165]. EEF2 catalyzes both the GTP-dependent ribosomal translocation step during translation elongation and the coordinated movement of the two tRNA molecules, the mRNA, and conformational changes in the ribosome [165]. SNRPB plays role in pre-mRNA splicing - it is a core component of the SMN-Sm complex that mediates spliceosomal snRNP assembly, a component of the spliceosomal U1, U2, U4 and U5 small nuclear ribonucleoproteins (snRNPs), a component of both the pre-catalytic spliceosome B complex and activated spliceosome C complexes, a component of the minor U12 spliceosome, and as a part of the U7 snRNP it is involved in histone pre-mRNA 3'-end processing [165]. Therefore, while the overarching, conserved role of chromatin-bound WDR5 in the cell lines interrogated is to modulate ribogenesis and translation (via select RPGs and EEF2), WDR5 also helps control the expression of genes involved in protein folding, telomere maintenance, Wnt signaling, the circadian clock, and pre-mRNA splicing.

Since WDR5 has been found to be over-expressed in cancer [67-70, 82, 107, 119-121], involved in malignant processes [26, 56, 59, 60, 78, 122], and is currently being assessed as a novel target for cancer therapies [27, 74, 89, 90], it is increasingly important to identify which genes are bonafide WDR5 target genes for multiple reasons. As scientists, we need to know which genes are commonly regulated by WDR5 in cancer cells to be able to more thoroughly understand the overarching functions of WDR5, to understand the effects of inhibiting WDR5, and to establish a set of “on-target” genes that are impacted by WDR5 inhibition. As WDR5 SMIs are improved and moved toward the clinic, it is important to

know which specific genes are bound and regulated by WDR5, so we can test the effects of the developing SMIs and differentiate between good, on target responses versus off target responses. This will assist in the refinement of WDR5 SMIs and may prove helpful in establishing the correct dosing with these SMIs. Binding alone is not a sufficient metric — just because gene is bound by WDR5 does not necessarily mean said gene is regulated by WDR5. In fact, there are such instances in the data discussed above, where a WDR5 peak is assigned to two genes, however changes in transcription in response to C6 treatment only occur with one of those genes. While the second gene could still be regulated via a non-WIN-site-dependent mechanism, it is still important to know which genes respond to a specific, targeted treatment — for example, it is important to know which genes respond to C6 treatment, so as better WIN site inhibitors are developed, on-target and off-target responses can continue to be assessed, and so we can better anticipate what effects occur if taken to clinical trials. Knowing which genes WDR5 regulates can also help determine which types of cancer may be susceptible to WDR5 inhibition - will a WDR5 inhibitor be broadly effective against all cancers, or will there be certain cancers which are particularly susceptible to WDR5 inhibition?

The ability to predict which genes will be bound by WDR5 in any cell type, and which will respond to WIN site blockade, has important ramifications for targeting WDR5 and/or utilizing WIN site inhibitors as anti-cancer therapies. Overall, WIN site inhibitors, in a similar fashion as the RNA polymerase I inhibitors that are currently being tested as potential chemotherapeutic agents [166], disrupt protein synthesis homeostasis. By thinking about WIN site inhibitors in this context, apart from their historical ties to MLL-fusion oncoproteins [26], it becomes clear that these SMIs may have much broader utility, particularly because aberrant protein synthesis is a common and characteristic theme in cancer [167]. Therefore, the development of more potent small molecule WDR5, WIN site inhibitors with true drug-like properties is of paramount importance, to both fully understand the therapeutic potential of these novel compounds and hopefully, one day treat cancer.

	LoVo	
	Up	Down
C6 v. DMSO	48	261

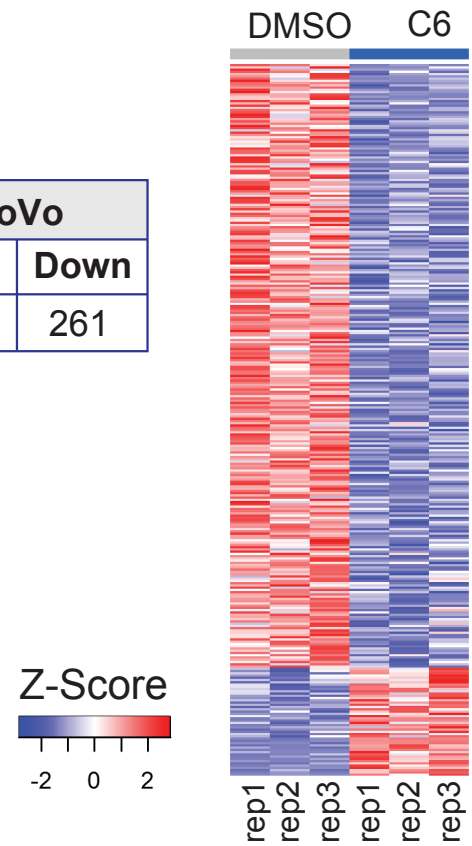


Figure 5.1. (*Left*) Results of SLAM-Seq analysis. Table shows the number of transcripts significantly (FDR < 0.05) altered by four hours of treatment of LoVo cells with 25 μM C6, compared to DMSO control. (*Right*) Heatmap, displaying z-transformed gene expression for significantly changed genes in C6 versus DMSO (FDR < 0.05) for all three replicates (rep1–rep3) of SLAM-Seq, examining the impact of four hours of DMSO or 25 μM C6 treatment of LoVo cells.

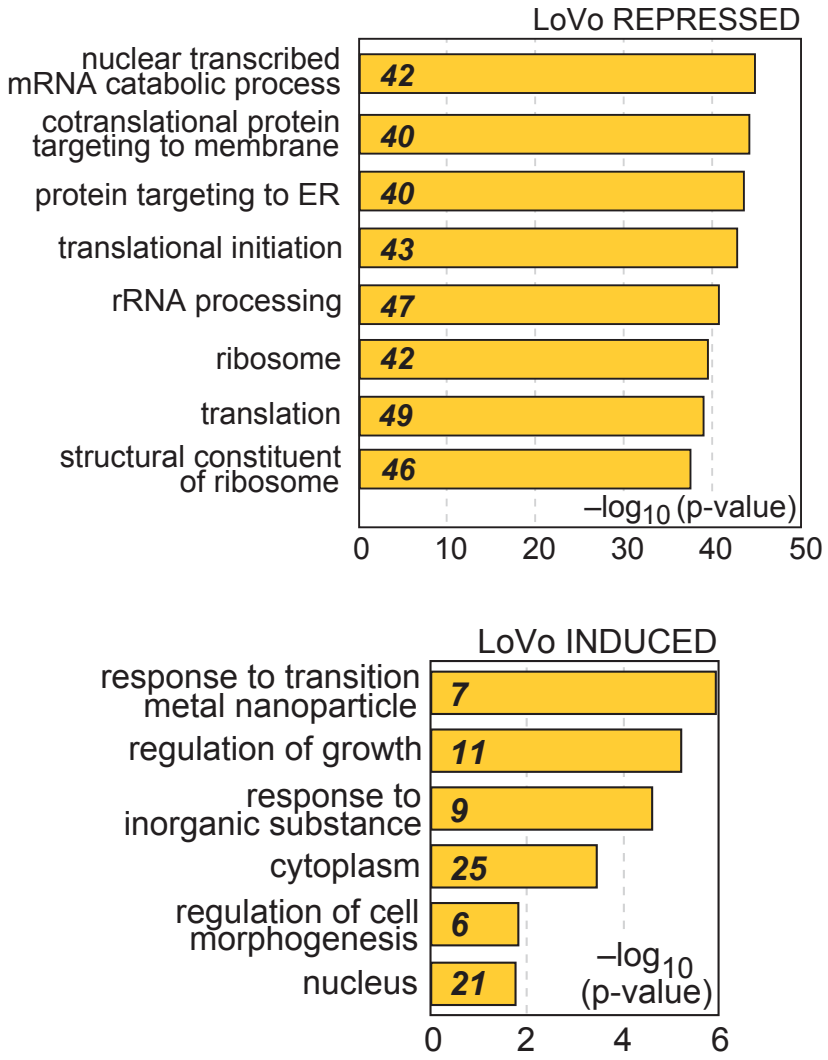


Figure 5.2. (*Top*) GO enrichment clusters for gene transcripts significantly repressed by C6 treatment of LoVo cells, as determined by SLAM-Seq. Numbers in italics represent the number of repressed genes in each category. (*Bottom*) GO enrichment clusters for gene transcripts significantly induced by C6 treatment of LoVo cells, as determined by SLAM-Seq. Numbers in italics represent the number of induced genes in each category.

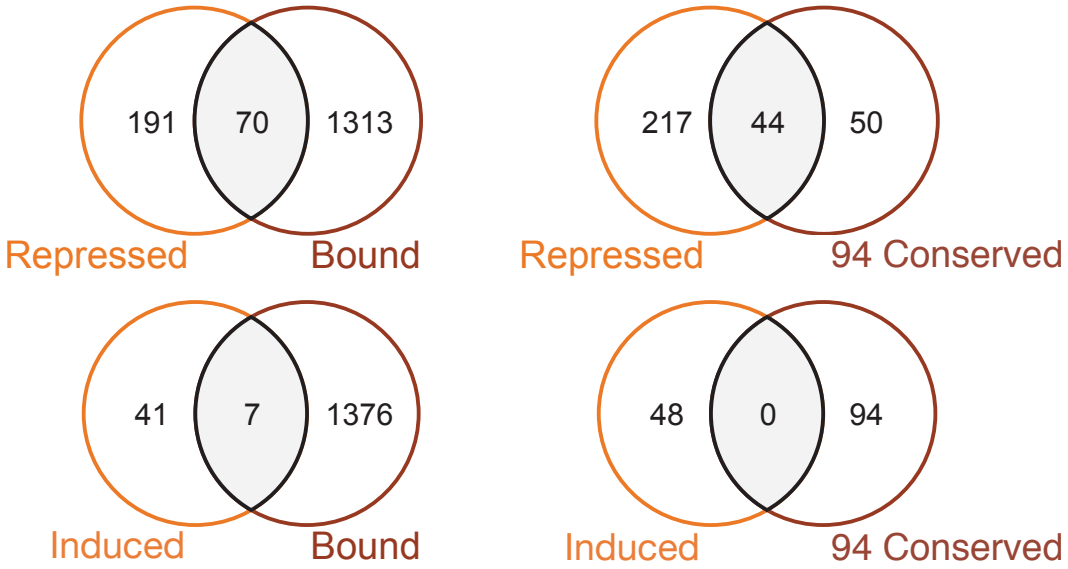


Figure 5.3. (*Left*) Venn diagrams, showing overlap of genes repressed (*top, left*) or induced (*bottom, left*) by C6 treatment of LoVo cells with genes bound by WDR5 in LoVo cells (as determined by ChIP-Seq). (*Right*) Venn diagrams, showing overlap of genes repressed (*top, right*) or induced (*bottom, right*) by C6 treatment of LoVo cells with the 94 conserved genes bound by WDR5 across all six cell types (as determined by ChIP-seq).

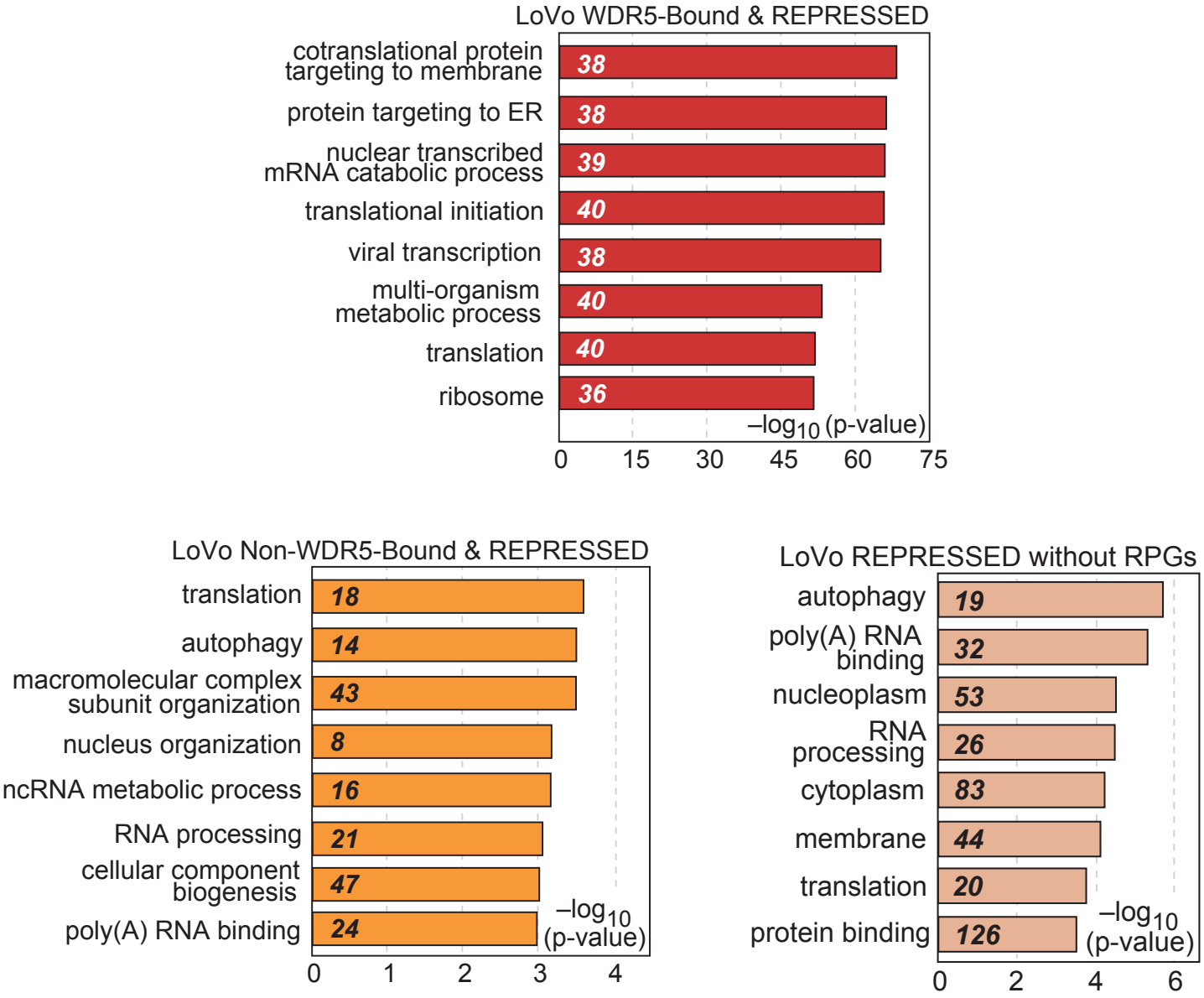


Figure 5.4. (*Top*) GO enrichment clusters for WDR5-bound (determined by ChIP-seq), C6-repressed (determined by SLAM-seq), gene transcripts in LoVo cells. Numbers in italics represent the number of repressed genes in each category. (*Bottom, left*) GO enrichment clusters for non-WDR5-bound (determined by ChIP-seq), C6-repressed (determined by SLAM-seq), gene transcripts in LoVo cells. Numbers in italics represent the number of repressed genes in each category. (*Bottom, right*) GO enrichment clusters for C6-repressed gene transcripts, minus RPG gene transcripts, in LoVo cells, as determined by SLAM-seq. Numbers in italics represent the number of repressed genes in each category.

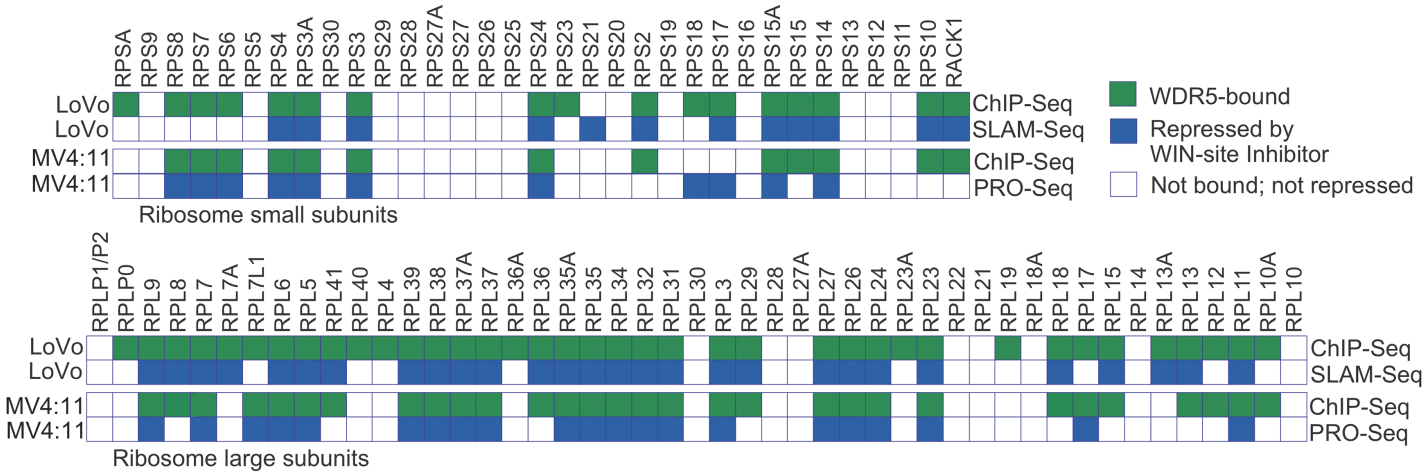


Figure 5.5. Ribosomogram, showing small (*top*) and large (*bottom*) ribosome subunit RPGs; a green box indicates whether WDR5 is bound to each RPG in the indicated cell type, a blue box indicates whether the gene is repressed by WIN site inhibition. LoVo cell data are from this study. MV4:11 data are taken from Aho, Wang, et.al. 2019.

	K562	
	Up	**Down**
C6nc v. DMSO	1	0
C6 v. DMSO	65	187

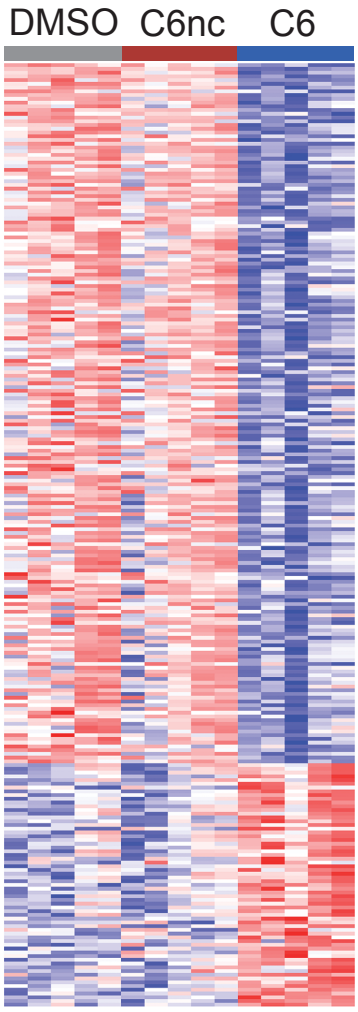


Figure 5.6. (*Left*) Table shows the number of transcripts significantly (FDR < 0.05) altered (in RNA-Seq analysis) by three days of treatment of K562 cells with 2 μM C6 or C6nc, compared to DMSO control. (*Right*) Heatmap, displaying z-transformed gene expression for significantly changed genes in C6 or C6nc versus DMSO (FDR < 0.05) for five biological replicates (rep1–rep5) of RNA-Seq, examining the impact of three days of DMSO, 2 μM C6nc, or 2 μM C6 treatment of K562 cells.

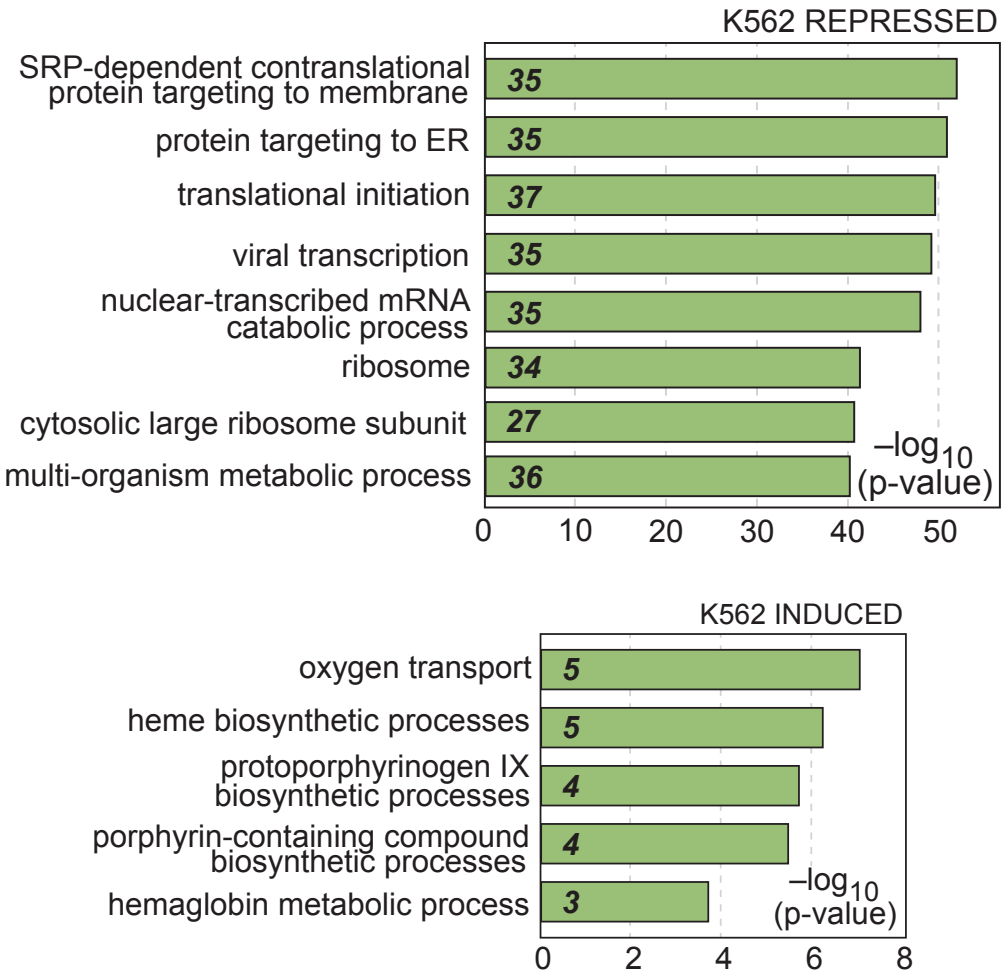


Figure 5.7. (*Top*) GO enrichment clusters for gene transcripts significantly repressed by C6 treatment of K562 cells, as determined by RNA-Seq. Numbers in italics represent the number of repressed genes in each category. (*Bottom*) GO enrichment clusters for gene transcripts significantly induced by C6 treatment of K562 cells, as determined by RNA-Seq. Numbers in italics represent the number of induced genes in each category.

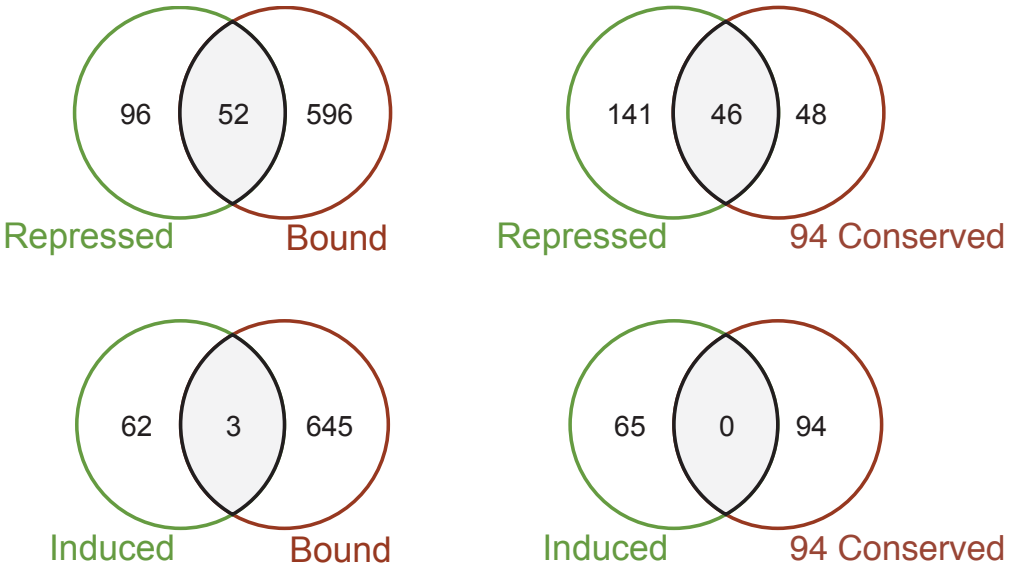


Figure 5.8. (*Left*) Venn diagrams, showing overlap of genes repressed (*top, left*) or induced (*bottom, left*) by C6 treatment of K562 cells with genes bound by WDR5 in K562 cells (as determined by ChIP-Seq). (*Right*) Venn diagrams, showing overlap of genes repressed (*top, right*) or induced (*bottom, right*) by C6 treatment of K562 cells with the 94 genes bound by WDR5 across all six cell types (as determined by ChIP-seq).

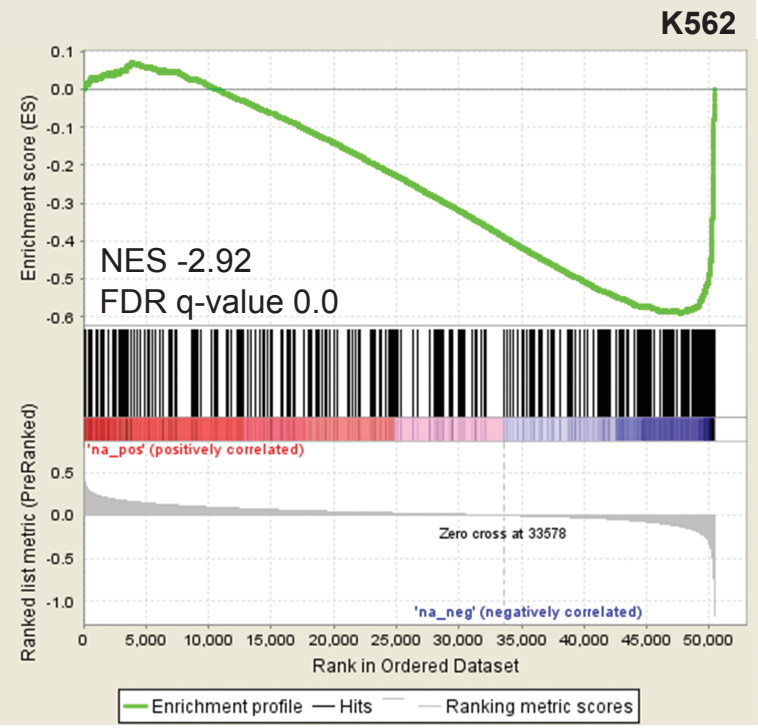


Figure 5.9. Gene set enrichment analysis showing the distribution of K562 genes repressed in response to 2 μM C6 treatment (RNA-Seq) against the list of all WDR5-bound genes in K562 cells (ChIP-Seq). FDR q = 0.0, genes ranked by $\log_2$-fold change.

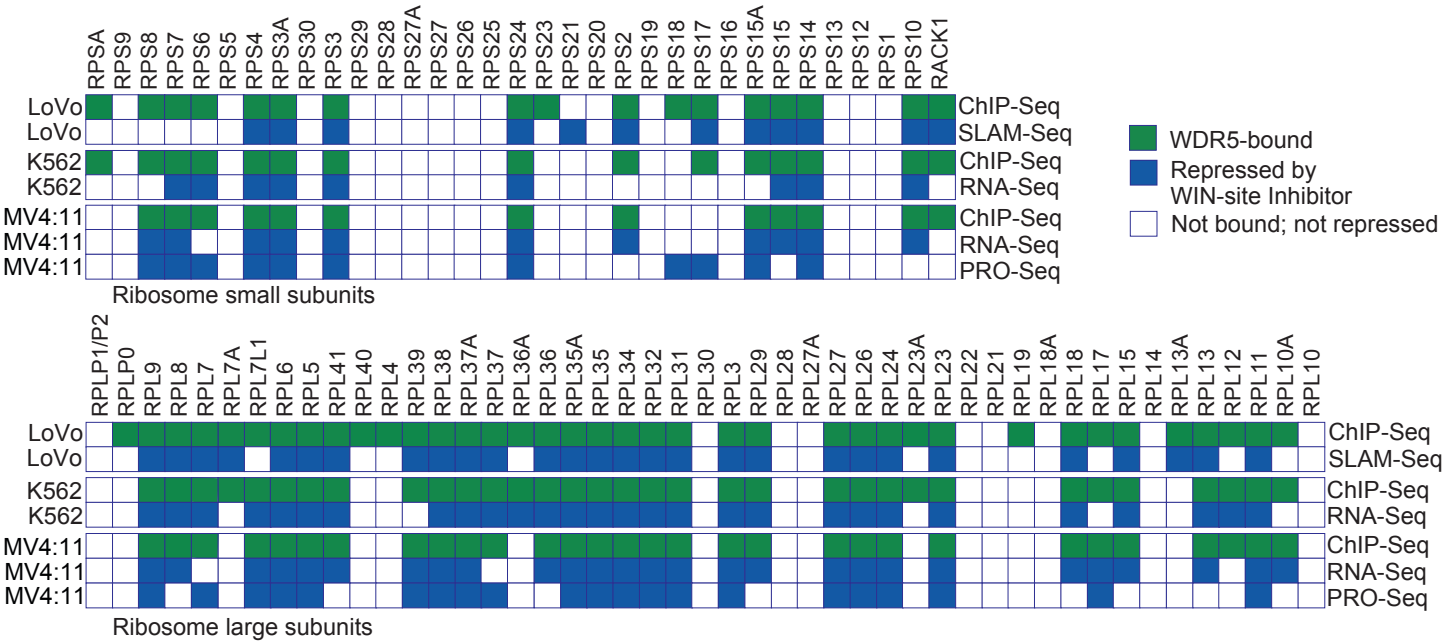


Figure 5.10. Ribosomogram, showing small (*top*) and large (*bottom*) ribosome subunit RPGs; a green box indicates whether WDR5 is bound to each RPG in the indicated cell type, a blue box indicates whether the gene is repressed by WIN site inhibition. The method of analysis is shown at the right of each row. K562 and LoVo cell data are from this study. MV4:11 data are taken from Aho, Wang, et.al. 2019.

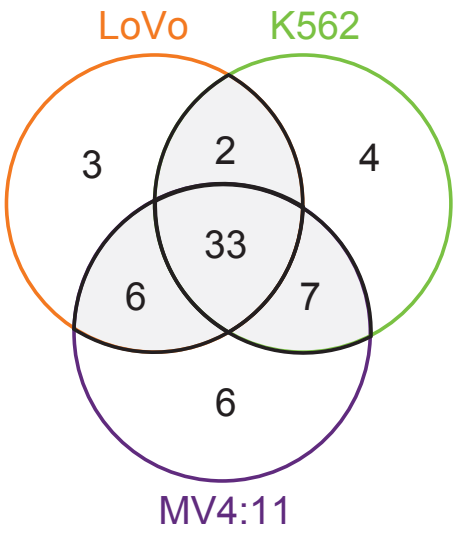


Figure 5.11. Overlap of the 44 early WDR5 target genes in LoVo cells, the 46 long-term WDR5 target genes in K562 cells, and the 52 long-term direct WDR5 target genes in MV4:11 cells. The 33 conserved genes are listed in the text. All other genes are listed in the table below. LoVo SLAM-Seq and K562 RNA-Seq data are from this study. MV4:11 RNA-Seq data are taken from Aho, Wang, et.al. 2019.

K562 + LoVo	LoVo + MV4:11	K562 + MV4:11	LoVo	K562	MV4:11
RPL37	RPL39	SNHG17	NACA	SNORA21	RNPS1
EIF3D	RPS15A	RNF220	RACK1	RPS6	PUM1
	SERBP1	CCT7	EEF1G	SNHG15	RPL17
	RPL23	EIF4G3		RPL12	RPL10A
	EIF4G1	RPS7			RPS8
	RPS2	RPL7L1			YWHAE
		SF3B3			

CHAPTER VII

Discussion

Preface

The conserved, highly structured nature of WDR5, and its recurrence in multiple chromatin-bound complexes, makes WDR5 an interesting, yet challenging protein to study. Further, the lack of rigorous study and identification of common WDR5 binding sites in chromatin has made it impossible to determine which conserved gene networks are under its control. Therefore, to begin to understand the general biological context in which WDR5 operates, I started my thesis work by completing ChIP-seq experiments for WDR5, using the same antibody and reagents across cell lines. I compared the genomic location of WDR5 in six cell lines of varying cell type, species, and cancer background, with different levels of WDR5 expression.While varying amounts of WDR5 binding across the genome per cell line were detected, common themes emerged across all six cell lines. 1) The majority of chromatin-bound is located WDR5 TSS-Proximally (Fig 2.4). 2) All six cell lines exhibit two separate TSS-Proximal WDR5 populations, one on either side of the TSS (Fig 2.5). 3) WDR5 most conspicuously localizes to the canonical E-Box motif (CACGTG), as well as E-Box variants, and also localizes to other binding motifs of multiple transcription factors. 4) WDR5-bound genes are involved in translation, RNA binding, ribosome biogenesis, and protein synthesis (Fig 2.8). These themes drove my further analyses and led me to hypothesize that WDR5 is recruited to chromatin via its WIN site, by a transcription factor (TF), to TSS-Proximal regions, in order to recruit other transcription factors via its WBM site, and therefore regulate the transcription of protein synthesis genes. Throughout this chapter I will discuss the data I have collected that support this hypothesis, as well as the additional experiments that should be performed to further bolster this hypothesis. In order to more clearly present my arguments, I will break down the hypothesis into the following sections: 1) WDR5 is recruited to chromatin via its WIN site, 2) by a transcription factor, 3) to TSS-Proximal regions, 4) in order to recruit other transcription factors via its WBM site, 5) and therefore regulate the transcription of protein synthesis genes.

WDR5 is Recruited to Chromatin Via its WIN Site

Chapter V outlines the evidence I have collected that further validates the utilization of the WIN site to tether WDR5 to chromatin. Using a novel SMI that tightly binds to the WIN site of WDR5 C6 (K_d ~100 pM), I show that WDR5 is displaced from chromatin upon treatment with C6 (Fig. 4.2 and 4.4). This is true in both cell lines interrogated, LoVo

(Fig 4.2), a human colorectal adenocarcinoma line, and K562 (Fig 4.4 and 4.6), a human chronic myelogenous leukemia line. These results are in agreement with the recently published results from the Tansey laboratory by Aho, et. al. [74], which show that C6 also displaces WDR5 from chromatin in MV:411 cells, a human biphenotypic B myelomonocytic leukemia cell line. Importantly, all C6-treated (as well as non-SMI-treated) ChIP-Seq experiments were performed in the same laboratory, with the same reagents, and the same antibody. This recurring ability of WIN site blockade to displace WDR5 from chromatin provides evidence for the hypothesis that WDR5 is linked to chromatin via the WIN site, regardless of cell type, and that this method of WDR5 recruitment to chromatin is conserved. This interaction between the WIN site of WDR5 and chromatin is likely to be indirect, or through the engagement of a WIN motif in a protein already bound to chromatin. While we can not yet say whether WDR5 binds other proteins, and then those proteins bind chromatin, or whether WDR5 binds to a chromatin-resident protein, we can say that the availability of the WIN site is necessary for this interaction. The reason I posit WDR5 is *recruited* to chromatin via the WIN site is two-fold: 1) because WDR5 has no known ability to directly bind DNA, nor does it have a DNA binding domain, and 2) because other TF motifs are present in WDR5 peak sequences. This second reason provides a nice segue into the next section.

WDR5 is Recruited to Chromatin by a Transcription Factor

Known motif analysis of WDR5 peak sequences across all six cell lines reveled two main, enriched categories. While one category is the E-Box motif specifically, the other is a slew of additional known transcription factor motifs. Therefore, I posit that WDR5 is recruited to chromatin, via its WIN site, by a transcription factor due to: 1) the presence of known transcription factor motifs across all six cell lines, 2) WDR5 already being known to use the WIN site to bind to proteins which bind to chromatin (ie: SET/MLL proteins), 3) the location of binding being TSS- Proximal, where TFs are known to bind. While the presence of these TF motifs within WDR5 peak sequences could just be coincidence, as TSS-Proximally it is likely one will find multiple TF motifs, further experiments are needed to determine whether a TF is recruiting WDR5 to chromatin. To start, one could synthesize WDR5 proteins and the transcription factors identified within WDR5 peak sequences (ie: Nanog, BMYB, E2F4, YY1, E2F1, E2F6, and KLF3). These could be used to perform *in vitro* experiments, to determine whether WDR5 can directly interact with any of these proteins. Simultaneously, co-immunoprecipitation (CoIP) experiments on cell lysates with over-expressed WDR5 could be performed — by pulling down WDR5 and probing for the TFOI (TF of Interest), it could be determined whether WDR5 and the TFOI interact indirectly. Further, this would determine whether WDR5 and the TFOI are capable of binding within a cellular context. If it is determined that, when over-expressed, WDR5 can interact with one of these TFOI, either directly or indirectly, one

could perform a CoIP on the endogenous proteins, in order to make sure the WDR5-TFOI interaction was not a side-effect of the large amount of WDR5 in cells with over-expressed WDR5. Alternatively, proteomics could be performed in an attempt to identify novel WDR5-binding partners. If proteomic experiments identified a novel WDR5-TFOI interaction, one should still complete the aforementioned *in vitro* and CoIP experiments in multiple cell lines in order to validate the proteomics and determine whether the WDR5-TFOI interaction is cell type specific or common across cell types. Next, one could treat cells with C6, as well as utilize WIN site mutants, to determine whether the WDR5-TFOI interaction is WIN-site dependent. Additionally, by mutating the DNA binding segment of the TFOI, whether WDR5 is capable of binding to chromatin (or at least to the same extent) without the TFOI being able to bind chromatin, could be tested. This should be paired with WDR5 WIN site mutants and WBM mutants. If the TFOI is capable of binding to chromatin without WDR5, but WDR5 can not bind to chromatin if the TFOI is mutated or its WIN site is mutated, then this will show that the TFOI is recruiting WDR5 to chromatin, specifically through the WIN site. While the protein that links WDR5 to chromatin could be one of the already characterized WIN site binders, there are thousands of proteins which contain the WIN motif, which makes it likely that the protein responsible for recruiting WDR5 to chromatin has yet to be identified. Further, this yet unknown protein is not necessarily a transcription factor. Therefore, it may be more prudent to start with a proteomic screen to identify novel WDR5 binding partners, in an attempt to determine the full complex in which WDR5 is bound when it localizes to chromatin.

WDR5 is Recruited to TSS-Proximal Regions in Chromatin

In Chapter III, WDR5 peaks across cell types are characterized. When peaks were plotted, based on distance from TSS, it became clear that a large population of WDR5 is bound within 500 bp of the TSS, across all six cell lines (Fig 2.4 and 2.5). Additionally, upon classification of peak location by HOMER, a large majority of WDR5 was shown to bind within the Promoter-TSS region (Fig 2.4). While MC38, 3T3, and LoVo cells also exhibited a TSS-Distal population of WDR5, I propose that the WDR5 at these locations does not perform the same role as TSS-Proximal WDR5. In fact, the amount of TSS-Distal WDR5 correlates with the WDR5 protein level in the cell. This is reminiscent of enhancer invasion, as WDR5 essentially saturates promoter proximal sites and then invades distal sites upon higher levels of protein expression the binding of additional sites to chromatin, and therefore I propose that WDR5 is acting at distal enhancers when bound TSS-Distally. While the identification of enhancers has, historically, been difficult, new methods are constantly being developed to more accurately identify genomic enhancers. Therefore, in order to determine whether TSS-Distal WDR5 is binding at enhancers, there are several experiments that could be performed and combined. If so inclined,

one can perform ChIP-Seq for histone modifications and predict the location of active enhancers - areas with high levels of H3K4me1 and H3K27ac and a lack of H3K4me3 - however, these modifications can only be considered correlative and can only describe the chromatin state [168]. On the other hand, the use of a more quantitative experiment, similar to the SLAM-Seq experiment I used in LoVo cells, can allow for the mapping of RNA polymerase II (RNAP), and therefore can not only identify nascent RNA but also can identify sites of polymerase pausing. Additionally, this assay termed PRO-Seq (precision nuclear run-on sequencing), directly maps RNAP active sites (via mapping elongation-competent RNAP) with base pair resolution, provides transcriptional direction [169, 170], and allows for the accurate identification and quantification of enhancer RNAs in a single experiment [168]. Further, by performing Nascent RNA sequencing analysis (NRSA) on the PRO-Seq data, enhancers can be identified, quantified, annotated, and assigned to their potential targets [168]. Therefore, by overlapping PRO-Seq analyses and the ChIP-seq results from this study, WDR5-bound enhancers can be identified.

As there is a significant population of WDR5 at TSS-Proximal sites across all six cell lines, I focused on these sites, as they are more likely to be sites of conserved WDR5 binding and functioning. After assigning WDR5 peaks to genes (as peak location can not be compared across two species) and overlapping all six gene lists, I determined that 94 genes are commonly bound by WDR5, regardless of cell type (Fig 3.4). In fact, after cross-referencing the peaks to which these genes were assigned, all peaks were found to be TSS-Proximal (Fig 3.6). Therefore, this data suggests that the location for conserved chromatin-bound WDR5 is TSS-Proximal.

WDR5 is Recruited to Chromatin in Order to Recruit Other Transcription Factors Via its WBM Site

Currently, the only known function for WDR5 is to act as a scaffold for multi-protein complexes. However, the fact that WDR5 consistently localizes to TSS-Proximal regions indicates that WDR5 may be necessary for the transcriptional regulation of genes to which it binds. Therefore, it is conceptually most likely that, rather than actively regulating transcription itself, WDR5 is instead recruiting a transcription factor to regulate its target genes. One such example can be found in the studies that have been performed with MYC. WDR5 directly interacts with MYC via the WBM site [60], and MYC has been shown to co-bind with WDR5 on chromatin in HEK293 cells [60], Ramos cells [61], neuroblastoma cells [59]. The Tansey laboratory has also shown that the MYC-WDR5 interaction (via the WBM site) is crucial for the recruitment of MYC to genes in chromatin [60, 61]. Preliminary analyses of ChIP-Seq experiments I have performed for MYC in LoVo, Be2C, MC38, and 3T3 cells has revealed a significant overlap in MYC and WDR5 binding across all four cell lines. In addition, these pilot studies have shown that by displacing WDR5 with a WIN site inhibitor,

MYC is also displaced from chromatin. These data, along with the published data, further support the notion that WDR5 is first recruited to chromatin via its WIN site, then is able to use its WBM site to recruit a transcription factor, such as MYC.

As MYC has been shown to bind to WDR5 via the WBM site, it is plausible that other transcription factors also are capable of binding to WDR5 via the WBM site. This would also further explain why WDR5 consistently binds to so many transcription factor motifs, including E-Boxes. While MYC is a well-known E-Box-binding transcription factor, others E-Box-associating TFs include CLOCK, BMAL1, MYOG, MYOD, TCF3, USF1/2, NPAS(2), TFE3, MITF, and HIF-1a/b. Therefore, it is possible that WDR5 could also recruit any of these factors to chromatin via the WBM site. Further, as WDR5 was also shown to bind other transcription factor motifs, it is also possible that WDR5 is recruiting any of these other TFs to chromatin via the WBM, including Nanog, BYMB, E2F1/3/4/6/7, KLF3/4/5/6/9/14, Bach1/2, NF-E2, MafK, MafB, Nrf2, Jun-AP1, Fosl2, Fra1/2, JunB, AP-1, Atf3, BATF, Sp1/2/5, Fli1, Elk1/4, ELF1, Foxa2, Pit1, YY1, STAT1/3/5, Oct4, p73, p53, p63, ETS(1), ETV1, Ronin, GABPA, and/or HIF2a. In order to test this hypothesis, proteomic analysis of the WDR5 complex is likely to provide the most straightforward results. Proteomic screening of WT-WDR5, versus WBM-mutant-WDR5 and WIN-mutant-WDR5, would help identify not only the other proteins in complex with chromatin-bound WDR5, but also which proteins bind via the WBM site versus the WIN site. Follow up CoIP experiments (as outlined above) would help determine which proteins are necessary for the recruitment of WDR5 to chromatin (most likely via the WIN site), and which proteins are then recruited to chromatin by WDR5 (most likely via the WBM site).

WDR5 is Recruited to Chromatin to Regulate the Transcription of Protein Synthesis Genes

As the cell lines interrogated spanned two species, six cell lines, and five cancer types, I was able to identify a conserved set of WDR5-bound genes (Fig 3.1 and 3.4). The 94 genes to which WDR5 binds across all six mouse and human cell lines are involved in protein synthesis, translation, RNA binding, and ribosome biogenesis (Fig 3.5 and 3.8). Notably, a significant portion of these 94 genes are RPGs (Fig 3.9). WDR5 also localizes to RPGs in the prostate adenocarcinoma cell line, LnCaP [107], and the gastric carcinoma cell line, BGC823 [84]. However, localization does not necessarily indicate regulation. Therefore, I paired C6 treatment, to displace WDR5 from chromatin, with RNA-Seq and SLAM-Seq. Chapter VI outlines these experiments, which reveal that half of the 94 conserved genes exhibit a reduction in either nascent RNA levels, in LoVo cells, or steady-state RNA levels, in K562 cells, upon C6 treatment (Fig. 5.1, 5.3, 5.6,

and 5.8). These results are consistent with the published results in MV4:11 cells [74]. In fact, across LoVo, K562, and MV4:11 cells, 33 RPGs are consistently regulated by chromatin-bound WDR5 (Fig 5.10).

Further evidence that a prominent role for chromatin-bound WDR5 is to assist in the regulation of protein synthesis genes comes from the experiments performed with WDR5 and MYC. In fact, WDR5 has been shown to recruit MYC to chromatin [60, 61], allowing MYC to perform its role as a transcription factor. Further, multiple publications have shown that the MYC-WDR5 interaction is necessary for MYC to drive both tumor initiation and maintenance [59-61]. Therefore, if other TFOI are identified as WDR5-interactors, the necessity of the WDR5-TFOI interaction for transcriptional regulation should be tested.

WDR5 Inhibitors

I have utilized C6 as a tool compound throughout my studies — this SMI was discovered in collaboration with Dr. Fesik's laboratory, here at Vanderbilt University. This collaboration was established in order to discover, test, and refine small molecule inhibitors of WDR5. While the Fesik laboratory has worked on discovering SMIs that can bind either the WIN or WBM site, they have had much better luck developing strong WIN site binders. This is in part due to the differences in the structures of the two sites - the WIN site is a deep pocket, while the WBM site is more of a shallow cleft. The hope is that the potent WIN site inhibitors can be further refined and eventually moved to clinical trials. Experiments with the new classes of WIN site SMIs, performed in cells, will continue to help refine the molecules, determine expected on-target effects, and identify potential negative off-target effects. While it may seem counterintuitive to target an essential gene like WDR5 since, upon treatment, all cells in the body will be subject to small molecule inhibition, the finding that C6 is not toxic to all cancer cells suggests that the loss of WDR5 is distinctly different than WIN site inhibition of WDR5. This further predicts a potential therapeutic window may be available, in which WDR5 WIN site inhibitors can be used to treat cancer cells, addicted to high levels of RPGs and protein synthesis genes, but remain relatively ineffective in non-cancer cells.

The WIN site inhibitors discovered through this collaboration were tested, as other WDR5 inhibitors were tested, for their ability to inhibit MLL-1 mediate HMT activity. While C6 indeed proved to be a potent and selective inhibitor of MLL-1 mediate HMT activity, the mechanism of action for C6 is distinct. For example, there is no direct point of intersection between MLL-fusion oncoproteins and transcriptional processes controlled by WDR5 [76]. Therefore, although MLLr leukemia cells lines were shown to be sensitive to C6 [74], the molecular mechanism for this sensitivity remains unclear. The study by Aho, et. al. was thus able to disconnect the mechanism of action of C6 from the historical

connection of WIN site inhibitors to MLL-fusion oncoproteins, revealing a new avenue for the clinical use of these novel WIN site inhibitors [74]. Rather than being dependent on MLL-fusion oncoproteins for effectiveness against cancer cells, it seems that WIN site inhibitors will be able to affect cancer cells through their inhibition on protein synthesis gene and RPG expression. Protein synthesis has been established as a recurring feature of cancer [76], cancer cells have been proposed to be addicted to the increases in protein and ribosome levels [171], and ribosome biogenesis has even been suggested to be the 'Achilles heel' of cancer [172]. Oncoproteins, such as MYC and RAS, are known to drive cellular growth and proliferation, creating what has been called an "enhanced demand for protein synthesis in malignant cells that likely pushes proteostatic checkpoints to the brink of collapse" [76]. While targeting oncoproteins themselves would be ideal, unfortunately they have proven difficult to develop SMI against. For example, the MYC protein is highly unstructured, and therefore typical structure-activity relationship assays used to develop SMIs are impossible to perform with MYC. However, by targeting a protein, such as WDR5, that complexes with MYC, recruits MYC to chromatin, and has been shown to be necessary for MYC to carry out its role as a transcription factor and drive tumor formation and maintenance, we may now be able to affect MYC functions without requiring that an SMI binds directly to it. Therefore, I posit that WDR5 WIN site inhibitors will be effective against MYC-driven tumors, and potentially other oncogene-driven tumors. Further, as WIN site inhibitors restrict the supply and alter the balance of RP subunits, they could also be regarded a novel type of nucleolar-targeted therapy, if able to make it to the clinic. WIN site inhibitors may also inhibit other enzymatic complexes WDR5 is involved in, such as the NSL complex. Certain cancers may prove to uniquely rely on the histone methylation or acetylation marks deposited by MLL-1 or NSL complexes, both of which rely on WDR5 for full function. Therefore, WIN site inhibitors may effectively target transcription, translation, protein synthesis, the balance of ribosome protein subunits, and epigenetics.

As WDR5 has been shown to utilize the WIN site to scaffold multiple protein complexes, there may also be other side effects of WIN site inhibitors in cells. For example, optimized WIN site inhibitors will likely disrupt the SET/MLL complexes, and could therefore affect promoter and enhancer histone methylation, and thus, gene expression. This could occur globally, on a gene-specific/enhancer-specific basis, or even a cell-type specific basis. The function of the NSL, ATAC, MOF, and/or NuRD complexes could also be disrupted, which would affect histone acetylation. Specific chromatin remodeling complexes are also likely targets of WIN site inhibitors. Extra-chromosomal complexes may be affected by WIN site inhibitors as well. For example, WDR5 has been shown to rely on the WIN site to localize to both the mitotic spindle and midbody during human cell division, and use the WIN site to bind KIF2A in the cytoplasm [62, 63]. Therefore, WIN site inhibitors may also affect WDR5 outside the nucleus, causing the improper regulation of cell

division. Further, the NSL and ATAC complexes have been shown to associate with microtubules and affect mitotic integrity [65, 66]. As such, WDR5 may even have multiple effects in a cell just by disrupting the same complex, both in the nucleus and cytoplasm. Again, such side effects could occur in all cells or in a cell-type-specific manner. Therefore, as WDR5 WIN site SMIs are continued to be developed, side effects such as these should be monitored in various cell lines, and the effects of these SMIs on both cancerous and healthy cells should be determined. We hope a therapeutic window will be able to be identified, in which cancerous cells are sensitive to WIN site inhibition, but healthy cells are recalcitrant.

Although there are other potential cancer treatments on the horizon, such as epigenetic drugs [173], nucleolar targeted drugs [174], and rRNA poisons [175], because WIN site inhibitors act through a fundamentally different mechanism, they will likely have a different therapeutic window, different on-target toxicities, and a different spectrum of drug synergies. There also may be some cancers which prove sensitive to WIN inhibition, epigenetic drugs, and/or rRNA poisons, while others may be resistant. Further, as cancer has proven to be a resilient and quickly adapting beast, the more therapeutics available, the better. Therefore, the development of more potent small molecule WDR5-WIN site inhibitors with true drug-like properties is needed to move this brainchild to the clinic and fully understand the therapeutic potential of this new class of compounds. The further study of these enhanced WIN site inhibitors in cellular contexts may also prove to help us better understand WDR5 biology. Used as tool compounds, they could help us determine which cancer cells are sensitive to the translational inhibition and the protein synthesis and ribosomal protein imbalance caused by WIN site inhibition. They may also help uncover the reasoning behind the localization of WDR5 to a select subset of RPGs.

The Future of WDR5

While it may take years to fully understand and appreciate all the roles WDR5 plays in cells and which are most relevant to normal versus diseased cellular states, the results of these studies have set us on a path to find these answers more quickly. By identifying a conserved set of genes bound by WDR5 regardless of cell type, we can now confidently predict where WDR5 will be displaced from in any cell type, upon treatment with an inhibitor of WDR5. Further, we now know that WDR5 is linked to chromatin at these genes via its WIN site, and that the transcription of certain genes is dependent on WDR5. Therefore, we can predict that, regardless of cell type, WIN site inhibitors will be the most effective at displacing WDR5 from chromatin, and that this displacement will lead to a reduction in the transcription of protein synthesis genes, namely RPGs.

However, this study also raises many questions that should be the focus of further efforts to better understand the biology of WDR5, particularly if it is going to be continued to be considered as a novel target for cancer therapy. Some of the overarching questions I have at the end of these studies, which I hope the Tansey laboratory will investigate, include:

-What other proteins complex with WDR5 when WDR5 is bound to RPGs in chromatin?

-What protein(s) recruit WDR5 to chromatin? Is it indeed a transcription factor that recruits WDR5 to RPGs?

-Does WDR5 recruit other transcription factors, as it does MYC, to chromatin?

-What other proteins are in the MYC-WDR5 complex?

-Does WDR5 recruit other oncogenes to chromatin?

-Why does WDR5 only bind, and control the transcription of, a subset of RPGs?

-What is the biological significance of this biased control of a subset of RPGs by WDR5?

-Will WDR5, WIN site SMIs prove to be clinically relevant?

Therefore, were I to continue to work on WDR5, I would first perform the proteomic experiments, suggested above. I would follow these up with the CoIP experiments also outlined above. I would also pair these experiments with ChIP experiments for the identified proteins of interest (POI), to show that WDR5 does indeed co-localize with the POI. If possible, I would also complete ChIP-re-ChIP experiments to show that WDR5 does indeed bind chromatin with the POI in the same location, in the same cell. Hopefully these experiments would lead to the identification of the other proteins that complex with WDR5, when WDR5 is bound to chromatin. Additionally, these experiments coupled with WIN-mutants and WBM-mutants, could identify which protein(s) is responsible for recruiting WDR5 to chromatin via the WIN site, and which proteins WDR5 is able to recruit to chromatin via the WBM site. As the Tansey laboratory is also interested in the MYC-WDR5 interaction, I would be very interested to figure out what proteins also complex with MYC and WDR5. Further, if it is determined that WDR5 also recruits other oncogenes to chromatin, this would suggest that WDR5 inhibitors could prove efficacious in other oncogene-driven cancers.

In order to determine the biological significance of the biased control of a subset of RPGs by WDR5, I would first perform ChIP-Exo in order to 1) more precisely determine the sequence to which WDR5 binds, and 2) in hopes that with increased resolution, we would find that while WDR5 binds in the region of many TF motifs (as determined by ChIP-Seq peak sequence analysis discussed in the previous chapters), it binds to a unique set of nucleic acids that are not involved in TF motifs. However, if we determine that WDR5 binds to a specific TF motif, then ChIP-Exo results could be paired with the proposed proteomics experiments, in order to help identify the protein that recruits WDR5 to chromatin. After identifying the precise nucleic acids to which WDR5 binds at each RPG, I would then utilize CRISPR to edit the sequence

to which WDR5 binds. I would next ensure that the transcription of the RPG of interest is decreased, in response to the loss of the WDR5 binding site. I would perform these experiments in cancer cells, so I could then perform classical cancer biology assays, such as tracking cellular doubling time and the ability to proliferate, testing for apoptosis and senescence, and surveying for changes in the cell cycle, cellular metabolism, and cell growth. Based on the results of these assays, colony formation assays, cellular migration assays, and mouse allograft assays could also be completed. These experiments would help identify which RPGs require WDR5 binding for a functional, cellular phenotype, and determine whether WDR5-dependent regulation of any of these RPGs is critical for cancer cell function and/or survival. I would also pair these experiments with RPG knock-out experiments. Using CRISPR paired with the novel method of dTAG-ing proteins, I would dTAG an RPG, then rapidly degrade the RPG and test for the same functional readouts. This would help determine whether the loss of WDR5-binding at these genes leads to the same phenotype as loss of the RPG. I would test all RPGs to which WDR5 consistently binds, regardless of cell type, with these experiments. By completing the proposed experiments with all WDR5-bound RPGs, we can determine whether there are certain RPGs that drive a particular cellular phenotype (ie: cell death) or whether the reduction of even just one RPG (by loss of WDR5 binding) is able to affect cancer cells. If the reduction of just one RPG is not enough to cause a phenotype, the WDR5 binding sites for multiple RPGs could be perturbed in one cell line - if this is needed to observe a phenotype, then it is more likely that the imbalance of RP subunits is driving the phenotype, rather than the reduction of one single RPG. Therefore, these experiments may also further help elucidate the biological reasoning for WDR5 regulation of these specific RPGs. If the WDR5-bound RPGs, which are found to play a role in the viability of one cancer cell line, are also tested (using the above experiments) in other cancer cell lines, this could help determine whether these RPGs depend on WDR5 regardless of cancer type, or whether certain cancer cells are more susceptible to the perturbation of WDR5-binding at specific RPGs. These follow-up experiments could teach us about the conserved biology of WDR5, as well as help identify which cancers may be more sensitive to WDR5 inhibitors.

Finally, I think it is important to continue to test improved WIN site inhibitors across various cancer cell lines, to determine which cancer types WIN site inhibitors may be efficacious in, to identify the on-target and off-target effects of each new class of compounds, and to continue to inform drug refinement efforts. Side-by-side experiments with WIN site inhibitors and dTAG-ed WDR5 could help identify the differences between the removal of WDR5 from a cell versus the inhibition of WDR5. Only time and clinical trials will determine whether one of our WIN site inhibitors will prove to be an effective, novel cancer therapy.

REFERENCES

1. Schuetz, A., et al., Structural basis for molecular recognition and presentation of histone H3 by WDR5. EMBO J, 2006. **25**(18): p. 4245-52.
2. Guarnaccia, A.D. and W.P. Tansey, Moonlighting with WDR5: A Cellular Multitasker. J Clin Med, 2018. **7**(2).
3. Higa, L.A., et al., CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. Nat Cell Biol, 2006. **8**(11): p. 1277-83.
4. Gori, F., P. Divieti, and M.B. Demay, Cloning and characterization of a novel WD-40 repeat protein that dramatically accelerates osteoblastic differentiation. J Biol Chem, 2001. **276**(49): p. 46515-22.
5. Gori, F. and M.B. Demay, BIG-3, a novel WD-40 repeat protein, is expressed in the developing growth plate and accelerates chondrocyte differentiation in vitro. Endocrinology, 2004. **145**(3): p. 1050-4.
6. Gori, F., L.G. Friedman, and M.B. Demay, Wdr5, a WD-40 protein, regulates osteoblast differentiation during embryonic bone development. Dev Biol, 2006. **295**(2): p. 498-506.
7. Gori, F., E.D. Zhu, and M.B. Demay, Perichondrial expression of Wdr5 regulates chondrocyte proliferation and differentiation. Dev Biol, 2009. **329**(1): p. 36-43.
8. Zhu, S., et al., Wdr5 is required for chick skeletal development. J Bone Miner Res, 2010. **25**(11): p. 2504-14.
9. Miller, T., et al., COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A, 2001. **98**(23): p. 12902-7.
10. Roguev, A., et al., The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J, 2001. **20**(24): p. 7137-48.
11. Zhang, P., et al., The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. Nucleic Acids Res, 2012. **40**(9): p. 4237-46.
12. Wysocka, J., et al., WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell, 2005. **121**(6): p. 859-72.
13. Ang, Y.S., et al., Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell, 2011. **145**(2): p. 183-97.
14. Ruthenburg, A.J., C.D. Allis, and J. Wysocka, Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell, 2007. **25**(1): p. 15-30.
15. Trievel, R.C. and A. Shilatifard, *WDR5, a complexed protein.* Nat Struct Mol Biol, 2009. **16**(7): p. 678-80.
16. Couture, J.F., E. Collazo, and R.C. Trievel, Molecular recognition of histone H3 by the WD40 protein WDR5. Nat Struct Mol Biol, 2006. **13**(8): p. 698-703.
17. Han, Z., et al., Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. Mol Cell, 2006. **22**(1): p. 137-44.
18. Ruthenburg, A.J., et al., Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat Struct Mol Biol, 2006. **13**(8): p. 704-12.
19. Patel, A., V. Dharmarajan, and M.S. Cosgrove, Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. J Biol Chem, 2008. **283**(47): p. 32158-61.
20. Patel, A., et al., A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. J Biol Chem, 2008. **283**(47): p. 32162-75.
21. Dharmarajan, V., et al., Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. J Biol Chem, 2012. **287**(33): p. 27275-89.
22. Dou, Y., et al., Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.
23. Schuettengruber, B., et al., Trithorax group proteins: switching genes on and keeping them active. Nat Rev Mol Cell Biol, 2011. **12**(12): p. 799-814.
24. Shinsky, S.A., et al., Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation. J Biol Chem, 2015. **290**(10): p. 6361-75.
25. Li, Y., et al., Structural basis for activity regulation of MLL family methyltransferases. Nature, 2016. **530**(7591): p. 447-52.
26. Cao, F., et al., Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. Mol Cell, 2014. **53**(2): p. 247-61.
27. Grebien, F., et al., Pharmacological targeting of the Wdr5-MLL interaction in C/EBPalpha N-terminal leukemia. Nat Chem Biol, 2015. **11**(8): p. 571-578.
28. Zhu, J., et al., Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. Nature, 2015. **525**(7568): p. 206-11.
29. Neilsen, B.K., et al., WDR5 supports colon cancer cells by promoting methylation of H3K4 and suppressing DNA damage. BMC Cancer, 2018. **18**(1): p. 673.

30. 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, 2005. **25**(21): p. 9175-88.
31. Mendjan, S., et al., Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Mol Cell, 2006. **21**(6): p. 811-23.
32. Cai, Y., et al., Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J Biol Chem, 2010. **285**(7): p. 4268-72.
33. Raja, S.J., et al., The nonspecific lethal complex is a transcriptional regulator in Drosophila. Mol Cell, 2010. **38**(6): p. 827-41.
34. Dou, Y., et al., Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell, 2005. **121**(6): p. 873-85.
35. 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, 2013. **9**(11): p. e1003940.
36. Sheikh, B.N., S. Guhathakurta, and A. Akhtar, The non-specific lethal (NSL) complex at the crossroads of transcriptional control and cellular homeostasis. EMBO Rep, 2019. **20**(7): p. e47630.
37. Wysocka, J., et al., A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature, 2006. **442**(7098): p. 86-90.
38. Bode, D., et al., Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells. Mol Cell Proteomics, 2016. **15**(3): p. 878-91.
39. Ee, L.S., et al., An Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5. Stem Cell Reports, 2017. **8**(6): p. 1488-1496.
40. Suganuma, T., et al., ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. Nat Struct Mol Biol, 2008. **15**(4): p. 364-72.
41. Wang, Y.L., et al., 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, 2008. **283**(49): p. 33808-15.
42. Gao, Z., et al., PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell, 2012. **45**(3): p. 344-56.
43. Qin, J., et al., The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. Cell Stem Cell, 2012. **11**(3): p. 319-32.
44. Hauri, S., et al., A High-Density Map for Navigating the Human Polycomb Complexome. Cell Rep, 2016. **17**(2): p. 583-595.
45. Thompson, B.A., et al., CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. Mol Cell Biol, 2008. **28**(12): p. 3894-904.
46. Yates, J.A., et al., Regulation of HOXA2 gene expression by the ATP-dependent chromatin remodeling enzyme CHD8. FEBS Lett, 2010. **584**(4): p. 689-93.
47. Wang, L., et al., INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell Stem Cell, 2014. **14**(5): p. 575-91.
48. Chung, C.Y., et al., Cbx8 Acts Non-canonically with Wdr5 to Promote Mammary Tumorigenesis. Cell Rep, 2016. **16**(2): p. 472-486.
49. Vilhais-Neto, G.C., et al., The WHHERE coactivator complex is required for retinoic acid-dependent regulation of embryonic symmetry. Nat Commun, 2017. **8**(1): p. 728.
50. Xu, W., et al., MKL1 mediates TNF-alpha induced pro-inflammatory transcription by bridging the crosstalk between BRG1 and WDR5. J Biomed Res, 2019. **33**(3): p. 164-172.
51. Wang, K.C., et al., A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature, 2011. **472**(7341): p. 120-4.
52. Gomez, J.A., et al., The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. Cell, 2013. **152**(4): p. 743-54.
53. Guo, X., et al., A Linc1405/Eomes Complex Promotes Cardiac Mesoderm Specification and Cardiogenesis. Cell Stem Cell, 2018. **22**(6): p. 893-908 e6.
54. He, W., et al., Long noncoding RNA BLACAT2 promotes bladder cancer-associated lymphangiogenesis and lymphatic metastasis. J Clin Invest, 2018. **128**(2): p. 861-875.
55. Gan, Q., et al., WD repeat-containing protein 5, a ubiquitously expressed histone methyltransferase adaptor protein, regulates smooth muscle cell-selective gene activation through interaction with pituitary homeobox 2. J Biol Chem, 2011. **286**(24): p. 21853-64.
56. Wu, M.Z., et al., Interplay between HDAC3 and WDR5 is essential for hypoxia-induced epithelial-mesenchymal transition. Mol Cell, 2011. **43**(5): p. 811-22.

57. Malek, R., et al., TWIST1-WDR5-Hottip Regulates Hoxa9 Chromatin to Facilitate Prostate Cancer Metastasis. Cancer Res, 2017. **77**(12): p. 3181-3193.
58. Hayashida, N., Set1/MLL complex is indispensable for the transcriptional ability of heat shock transcription factor 2. Biochem Biophys Res Commun, 2015. **467**(4): p. 805-12.
59. Sun, Y., et al., WDR5 Supports an N-Myc Transcriptional Complex That Drives a Protumorigenic Gene Expression Signature in Neuroblastoma. Cancer Res, 2015. **75**(23): p. 5143-54.
60. Thomas, L.R., et al., Interaction with WDR5 promotes target gene recognition and tumorigenesis by MYC. Mol Cell, 2015. **58**(3): p. 440-52.
61. Thomas, L.R., et al., Interaction of the oncoprotein transcription factor MYC with its chromatin cofactor WDR5 is essential for tumor maintenance. Proc Natl Acad Sci U S A, 2019.
62. Bailey, J.K., et al., WD repeat-containing protein 5 (WDR5) localizes to the midbody and regulates abscission. J Biol Chem, 2015. **290**(37): p. 22447.
63. Ali, A., et al., MLL/WDR5 Complex Regulates Kif2A Localization to Ensure Chromosome Congression and Proper Spindle Assembly during Mitosis. Dev Cell, 2017. **41**(6): p. 605-622 e7.
64. Ali, A., S.N. Veeranki, and S. Tyagi, A SET-domain-independent role of WRAD complex in cell-cycle regulatory function of mixed lineage leukemia. Nucleic Acids Res, 2014. **42**(12): p. 7611-24.
65. Orpinell, M., et al., The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. EMBO J, 2010. **29**(14): p. 2381-94.
66. Meunier, S., et al., An epigenetic regulator emerges as microtubule minus-end binding and stabilizing factor in mitosis. Nat Commun, 2015. **6**: p. 7889.
67. Ge, Z., et al., WDR5 high expression and its effect on tumorigenesis in leukemia. Oncotarget, 2016. **7**(25): p. 37740-37754.
68. Dai, X., et al., WDR5 Expression Is Prognostic of Breast Cancer Outcome. PLoS One, 2015. **10**(9): p. e0124964.
69. Chen, X., et al., Gene expression profiling of WDR5 regulated genes in bladder cancer. Genom Data, 2015. **5**: p. 27-9.
70. Chen, X., et al., Upregulated WDR5 promotes proliferation, self-renewal and chemoresistance in bladder cancer via mediating H3K4 trimethylation. Sci Rep, 2015. **5**: p. 8293.
71. Lu, K., et al., The Histone H3 Lysine 4 Presenter WDR5 as an Oncogenic Protein and Novel Epigenetic Target in Cancer. Front Oncol, 2018. **8**: p. 502.
72. Krivtsov, A.V. and S.A. Armstrong, MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer, 2007. **7**(11): p. 823-33.
73. Xu, J., et al., MLL1 and MLL1 fusion proteins have distinct functions in regulating leukemic transcription program. Cell Discov, 2016. **2**: p. 16008.
74. Aho, E.R., et al., Displacement of WDR5 from Chromatin by a WIN Site Inhibitor with Picomolar Affinity. Cell Rep, 2019. **26**(11): p. 2916-2928 e13.
75. Garcia-Cuellar, M.P., et al., Leukemogenic MLL-ENL Fusions Induce Alternative Chromatin States to Drive a Functionally Dichotomous Group of Target Genes. Cell Rep, 2016. **15**(2): p. 310-22.
76. Aho, E.R., et al., Targeting WDR5: A WINning Anti-Cancer Strategy? Epigenet Insights, 2019. **12**: p. 2516865719865282.
77. Tansey, W.P., Mammalian MYC Proteins and Cancer. New Journal of Science, 2014. **2014**: p. 27.
78. Carugo, A., et al., In Vivo Functional Platform Targeting Patient-Derived Xenografts Identifies WDR5-Myc Association as a Critical Determinant of Pancreatic Cancer. Cell Rep, 2016. **16**(1): p. 133-147.
79. Prendergast, G.C. and E.B. Ziff, Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science, 1991. **251**(4990): p. 186-9.
80. Robinson, K.A. and J.M. Lopes, SURVEY AND SUMMARY: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res, 2000. **28**(7): p. 1499-505.
81. Conacci-Sorrell, M., C. Ngouenet, and R.N. Eisenman, Myc-nick: a cytoplasmic cleavage product of Myc that promotes alpha-tubulin acetylation and cell differentiation. Cell, 2010. **142**(3): p. 480-93.
82. Sun, W., F. Guo, and M. Liu, Up-regulated WDR5 promotes gastric cancer formation by induced cyclin D1 expression. J Cell Biochem, 2018. **119**(4): p. 3304-3316.
83. Chen, H., et al., A TGFbeta-PRMT5-MEP50 axis regulates cancer cell invasion through histone H3 and H4 arginine methylation coupled transcriptional activation and repression. Oncogene, 2017. **36**(3): p. 373-386.
84. Sun, T.T., et al., LncRNA GClnc1 Promotes Gastric Carcinogenesis and May Act as a Modular Scaffold of WDR5 and KAT2A Complexes to Specify the Histone Modification Pattern. Cancer Discov, 2016. **6**(7): p. 784-801.
85. Bennett, R.L. and J.D. Licht, *Targeting Epigenetics in Cancer.* Annu Rev Pharmacol Toxicol, 2018. **58**: p. 187-207.

86. Alicea-Velazquez, N.L., et al., Targeted Disruption of the Interaction between WD-40 Repeat Protein 5 (WDR5) and Mixed Lineage Leukemia (MLL)/SET1 Family Proteins Specifically Inhibits MLL1 and SETd1A Methyltransferase Complexes. J Biol Chem, 2016. **291**(43): p. 22357-22372.
87. Thiel, A.T., et al., MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. Cancer Cell, 2010. **17**(2): p. 148-59.
88. Chen, Y., et al., MLL2, Not MLL1, Plays a Major Role in Sustaining MLL-Rearranged Acute Myeloid Leukemia. Cancer Cell, 2017. **31**(6): p. 755-770 e6.
89. Bolshan, Y., et al., Synthesis, Optimization, and Evaluation of Novel Small Molecules as Antagonists of WDR5-MLL Interaction. ACS Med Chem Lett, 2013. **4**(3): p. 353-7.
90. Senisterra, G., et al., Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. Biochem J, 2013. **449**(1): p. 151-9.
91. Migliori, V., M. Mapelli, and E. Guccione, On WD40 proteins: propelling our knowledge of transcriptional control? Epigenetics, 2012. **7**(8): p. 815-22.
92. Stirnimann, C.U., et al., WD40 proteins propel cellular networks. Trends Biochem Sci, 2010. **35**(10): p. 565-74.
93. Langmead, B., et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 2009. **10**(3): p. R25.
94. Feng, J., et al., Identifying ChIP-seq enrichment using MACS. Nat Protoc, 2012. **7**(9): p. 1728-40.
95. Heinz, S., et al., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell, 2010. **38**(4): p. 576-89.
96. Huang da, W., B.T. Sherman, and R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols, 2009. **4**(1): p. 44-57.
97. Subramanian, A., et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
98. Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal, 2011. **17**: p. 10-12.
99. Dobin, A., et al., STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 2013. **29**(1): p. 15-21.
100. Liao, Y., G.K. Smyth, and W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, 2014. **30**(7): p. 923-30.
101. Love, M.I., W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol, 2014. **15**(12): p. 550.
102. Neumann, T., et al., Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. BMC Bioinformatics, 2019. **20**(1): p. 258.
103. Kloc, A. and N. Ivanova, Chromatin and pluripotency: the MYSTerious connection. Cell Stem Cell, 2012. **11**(2): p. 139-40.
104. Hopkin, A.S., et al., GRHL3/GET1 and trithorax group members collaborate to activate the epidermal progenitor differentiation program. PLoS Genet, 2012. **8**(7): p. e1002829.
105. Zhao, W., et al., Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination. Cell, 2013. **152**(5): p. 1037-50.
106. Cheng, J., et al., A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Mol Cell, 2014. **53**(6): p. 979-92.
107. Kim, J.Y., et al., A role for WDR5 in integrating threonine 11 phosphorylation to lysine 4 methylation on histone H3 during androgen signaling and in prostate cancer. Mol Cell, 2014. **54**(4): p. 613-25.
108. Riggi, N., et al., EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. Cancer Cell, 2014. **26**(5): p. 668-681.
109. Scelfo, A., et al., Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities. Mol Cell, 2019. **74**(5): p. 1037-1052 e7.
110. Guccione, E., et al., Myc-binding-site recognition in the human genome is determined by chromatin context. Nat Cell Biol, 2006. **8**(7): p. 764-70.
111. Mahajan, K., et al., ACK1/TNK2 Regulates Histone H4 Tyr88-phosphorylation and AR Gene Expression in Castration-Resistant Prostate Cancer. Cancer Cell, 2017. **31**(6): p. 790-803 e8.
112. Sikora, K., et al., c-myc oncogene expression in colorectal cancer. Cancer, 1987. **59**(7): p. 1289-95.
113. Smith, D.R., T. Myint, and H.S. Goh, Over-expression of the c-myc proto-oncogene in colorectal carcinoma. Br J Cancer, 1993. **68**(2): p. 407-13.
114. Castell, A. and L.G. Larsson, Targeting MYC Translation in Colorectal Cancer. Cancer Discov, 2015. **5**(7): p. 701-3.
115. Wiegering, A., et al., Targeting Translation Initiation Bypasses Signaling Crosstalk Mechanisms That Maintain High MYC Levels in Colorectal Cancer. Cancer Discov, 2015. **5**(7): p. 768-781.

116. Wu, T., et al., Co-inhibition of BET proteins and NF-kappaB as a potential therapy for colorectal cancer through synergistic inhibiting MYC and FOXM1 expressions. Cell Death Dis, 2018. **9**(3): p. 315.
117. Elbadawy, M., et al., Emerging Roles of C-Myc in Cancer Stem Cell-Related Signaling and Resistance to Cancer Chemotherapy: A Potential Therapeutic Target Against Colorectal Cancer. Int J Mol Sci, 2019. **20**(9).
118. Huang da, W., B.T. Sherman, and R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 2009. **4**(1): p. 44-57.
119. Tan, X., et al., PI3K/AKT-mediated upregulation of WDR5 promotes colorectal cancer metastasis by directly targeting ZNF407. Cell Death Dis, 2017. **8**(3): p. e2686.
120. Cui, Z., et al., Effect of high WDR5 expression on the hepatocellular carcinoma prognosis. Oncol Lett, 2018. **15**(5): p. 7864-7870.
121. Wu, Y., et al., Overexpression of WD repeat domain 5 associates with aggressive clinicopathological features and unfavorable prognosis in head neck squamous cell carcinoma. J Oral Pathol Med, 2018. **47**(5): p. 502-510.
122. Wang, P., et al., WDR5 modulates cell motility and morphology and controls nuclear changes induced by a 3D environment. Proc Natl Acad Sci U S A, 2018. **115**(34): p. 8581-8586.
123. Zeid, R., et al., Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. Nat Genet, 2018. **50**(4): p. 515-523.
124. Li, X., et al., Integrative analyses shed new light on human ribosomal protein gene regulation. Sci Rep, 2016. **6**: p. 28619.
125. Sadasivam, S. and J.A. DeCaprio, The DREAM complex: master coordinator of cell cycle-dependent gene expression. Nat Rev Cancer, 2013. **13**(8): p. 585-95.
126. George, O.L. and S.A. Ness, Situational awareness: regulation of the myb transcription factor in differentiation, the cell cycle and oncogenesis. Cancers (Basel), 2014. **6**(4): p. 2049-71.
127. Arzate-Mejia, R.G., F. Recillas-Targa, and V.G. Corces, *Developing in 3D: the role of CTCF in cell differentiation.* Development, 2018. **145**(6).
128. Chung, S. and R.P. Perry, Importance of introns for expression of mouse ribosomal protein gene rpL32. Mol Cell Biol, 1989. **9**(5): p. 2075-82.
129. Antoine, M. and P. Kiefer, Functional characterization of transcriptional regulatory elements in the upstream region and intron 1 of the human S6 ribosomal protein gene. Biochem J, 1998. **336 (Pt 2)**: p. 327-35.
130. Blais, A. and B.D. Dynlacht, E2F-associated chromatin modifiers and cell cycle control. Curr Opin Cell Biol, 2007. **19**(6): p. 658-62.
131. Chavalit, T., WDR5 is a novel partner of KLF3 and this interaction is important for KLF3 genomic localisation, in Biotechnology & Biomolecular Sciences. 2018, UNSW Sidney.
132. Pearson, R.C., A.P. Funnell, and M. Crossley, The mammalian zinc finger transcription factor Kruppel-like factor 3 (KLF3/BKLF). IUBMB Life, 2011. **63**(2): p. 86-93.
133. 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, 2014. **28**(9): p. 929-42.
134. Yang, Y.W., et al., Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. Elife, 2014. **3**: p. e02046.
135. Cremona, M.A., et al., Peak shape clustering reveals biological insights. BMC Bioinformatics, 2015. **16**: p. 349.
136. Starick, S.R., et al., ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors. Genome Res, 2015. **25**(6): p. 825-35.
137. Gekakis, N., et al., Role of the CLOCK protein in the mammalian circadian mechanism. Science, 1998. **280**(5369): p. 1564-9.
138. Haque, R., et al., CLOCK and NPAS2 have overlapping roles in the circadian oscillation of arylalkylamine N-acetyltransferase mRNA in chicken cone photoreceptors. J Neurochem, 2010. **113**(5): p. 1296-306.
139. Chang, J.H., et al., Myogenic basic helix-loop-helix proteins regulate the expression of peroxisomal proliferator activated receptor-gamma coactivator-1alpha. Endocrinology, 2006. **147**(6): p. 3093-106.
140. Shklover, J., et al., MyoD uses overlapping but distinct elements to bind E-box and tetraplex structures of regulatory sequences of muscle-specific genes. Nucleic Acids Res, 2007. **35**(21): p. 7087-95.
141. Ganassi, M., et al., Myogenin promotes myocyte fusion to balance fibre number and size. Nat Commun, 2018. **9**(1): p. 4232.
142. Ellenberger, T., et al., Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev, 1994. **8**(8): p. 970-80.
143. Giangrande, P.H., et al., Identification of E-box factor TFE3 as a functional partner for the E2F3 transcription factor. Mol Cell Biol, 2003. **23**(11): p. 3707-20.
144. Jiang, S., M.R. Galindo, and H.W. Jarrett, Purification and identification of a transcription factor, USF-2, binding to E-box element in the promoter of human telomerase reverse transcriptase (hTERT). Proteomics, 2010. **10**(2): p. 203-11.

145. Terragni, J., et al., The E-box binding factors Max/Mnt, MITF, and USF1 act coordinately with FoxO to regulate expression of proapoptotic and cell cycle control genes by phosphatidylinositol 3-kinase/Akt/glycogen synthase kinase 3 signaling. J Biol Chem, 2011. **286**(42): p. 36215-27.
146. Huffman, J.L., et al., The basic helix-loop-helix domain of the aryl hydrocarbon receptor nuclear transporter (ARNT) can oligomerize and bind E-box DNA specifically. J Biol Chem, 2001. **276**(44): p. 40537-44.
147. Wu, Y., et al., Reciprocal Regulation between the Circadian Clock and Hypoxia Signaling at the Genome Level in Mammals. Cell Metab, 2017. **25**(1): p. 73-85.
148. Schones, D.E., et al., Dynamic regulation of nucleosome positioning in the human genome. Cell, 2008. **132**(5): p. 887-98.
149. Weber, C.M., S. Ramachandran, and S. Henikoff, Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol Cell, 2014. **53**(5): p. 819-30.
150. Chorev, M. and L. Carmel, *The function of introns.* Front Genet, 2012. **3**: p. 55.
151. Hariharan, N., D.E. Kelley, and R.P. Perry, Equipotent mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements. Genes Dev, 1989. **3**(11): p. 1789-800.
152. Chung, S. and R.P. Perry, The importance of downstream delta-factor binding elements for the activity of the rpL32 promoter. Nucleic Acids Res, 1993. **21**(14): p. 3301-8.
153. McCann, K.L. and S.J. Baserga, Genetics. Mysterious ribosomopathies. Science, 2013. **341**(6148): p. 849-50.
154. Cheng, Z., et al., Small and Large Ribosomal Subunit Deficiencies Lead to Distinct Gene Expression Signatures that Reflect Cellular Growth Rate. Mol Cell, 2019. **73**(1): p. 36-47 e10.
155. Zhou, X., et al., Ribosomal proteins: functions beyond the ribosome. J Mol Cell Biol, 2015. **7**(2): p. 92-104.
156. Yang, K., J. Yang, and J. Yi, Nucleolar Stress: hallmarks, sensing mechanism and diseases. Cell Stress, 2018. **2**(6): p. 125-140.
157. Russo, A. and G. Russo, Ribosomal Proteins Control or Bypass p53 during Nucleolar Stress. Int J Mol Sci, 2017. **18**(1).
158. Fumagalli, S., et al., Suprainduction of p53 by disruption of 40S and 60S ribosome biogenesis leads to the activation of a novel G2/M checkpoint. Genes Dev, 2012. **26**(10): p. 1028-40.
159. Xue, S. and M. Barna, Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat Rev Mol Cell Biol, 2012. **13**(6): p. 355-69.
160. Grebien, F., et al., Pharmacological targeting of the Wdr5-MLL interaction in C/EBPalpha N-terminal leukemia. Nat Chem Biol, 2015. **11**(8): p. 571-8.
161. Gu, P., et al., lncRNA HOXD-AS1 Regulates Proliferation and Chemo-Resistance of Castration-Resistant Prostate Cancer via Recruiting WDR5. Mol Ther, 2017. **25**(8): p. 1959-1973.
162. Odho, Z., S.M. Southall, and J.R. Wilson, 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, 2010. **285**(43): p. 32967-76.
163. Muhar, M., et al., SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science, 2018. **360**(6390): p. 800-805.
164. Stumpf, C.R., et al., The translational landscape of the mammalian cell cycle. Mol Cell, 2013. **52**(4): p. 574-82.
165. UniProt, C., UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.
166. Penzo, M., et al., The Ribosome Biogenesis-Cancer Connection. Cells, 2019. **8**(1).
167. Vaklavas, C., S.W. Blume, and W.E. Grizzle, Translational Dysregulation in Cancer: Molecular Insights and Potential Clinical Applications in Biomarker Development. Front Oncol, 2017. **7**: p. 158.
168. Wang, J., et al., Nascent RNA sequencing analysis provides insights into enhancer-mediated gene regulation. BMC Genomics, 2018. **19**(1): p. 633.
169. Core, L.J., et al., Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet, 2014. **46**(12): p. 1311-20.
170. Hah, N., et al., Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. Proc Natl Acad Sci U S A, 2015. **112**(3): p. E297-302.
171. Ruggero, D., Revisiting the nucleolus: from marker to dynamic integrator of cancer signaling. Sci Signal, 2012. **5**(241): p. pe38.
172. van Sluis, M. and B. McStay, Ribosome biogenesis: Achilles heel of cancer? Genes Cancer, 2014. **5**(5-6): p. 152-3.
173. Roberti, A., et al., Epigenetics in cancer therapy and nanomedicine. Clin Epigenetics, 2019. **11**(1): p. 81.
174. Stepinski, D., The nucleolus, an ally, and an enemy of cancer cells. Histochem Cell Biol, 2018. **150**(6): p. 607-629.
175. Negi, S.S. and P. Brown, Transient rRNA synthesis inhibition with CX-5461 is sufficient to elicit growth arrest and cell death in acute lymphoblastic leukemia cells. Oncotarget, 2015. **6**(33): p. 34846-58.