Regulation of Proviral Expression

and Post-Translational Modifications

in Embryonic Cells

Andreia Lee

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

Abstract

Regulation of Proviral Expression

and Post-Translational Modifications in Embryonic Cells

Andreia Lee

Moloney Murine Leukemia Virus (M-MLV) proviral DNA is transcriptionally

silenced in mouse embryonic cells by a repressor complex containing tripartite-motif-

containing 28 (Trim28). Trim28 depends on post-translational modifications, such as

sumoylation and phosphorylation, and its interactions with several co-repressor proteins

to regulate its repressive activity. YY1 is one such Trim28 co-repressor protein, recently

found to tether the Trim28 silencing complex to the M-MLV promoter. Here, we

investigated the biochemical interaction of Trim28 and YY1, and the role of sumoylation

and phosphorylation of Trim28 in mediating M-MLV silencing. Experiments probing the

binding of YY1 and Trim28 *in vitro* suggested that their interaction occurs indirectly.

Mutational studies demonstrated that the RBCC domain of Trim28 is sufficient for

interaction with YY1 while the acidic region 1 and zinc fingers of YY1 were necessary

and sufficient for its interaction with Trim28. Additionally, we found that the K779

residue was critical for Trim28-mediated silencing of M-MLV in embryonic cells.

The repressor complex that silences M-MLV is very large and likely consists of

many protein subunits. A few proteins contained in the repressor complex have been

identified, including Trim28, but the identity of most of the components forming the

repressor complex are unknown. We detected a new form of the complex that is of even

high molecular weight and likely contains additional associated cofactors. We reported an

approach for purifying this larger repressor complex and identified new candidates for cofactors that may potentially function in the silencing of M-MLV.

We also examined the regulation of sumoylation in embryonic cells. Sumoylation conjugation is a post-translational modification that affects a diverse range of processes and is important for embryo survival. Overall inhibition of the SUMO pathways results in embryonic lethality demonstrating the importance of the SUMO pathways for embryonic viability; however, our understanding of SUMO function in embryos at the cellular and molecular level is still greatly lacking. We demonstrated that SUMO1 cannot be overexpressed in embryonic carcinoma and embryonic stem cells and that SUMO1 overexpression is prevented at a post-transcriptional level. This occurred specifically for SUMO1 and not for SUMO2 overexpression. Furthermore, blocking conjugation or increasing the deconjugation of SUMO1 to substrates significantly improved SUMO1 overexpression. The results indicate that the overexpression of SUMO1 protein, in itself, is tolerated in embryonic cells, but the accumulation of substrate(s) modified by SUMO1 appears to be strongly prevented by an embryonic-specific post-transcriptional mechanism.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like to thank my mentor, Stephen Goff, for taking me on as a student and allowing me this opportunity for growth and exploration. Steve always made time to talk to me, allowing me to interrupt him at any moment to discuss my experiments, and I would always leave his office with newfound clarity and excitement for my project. He taught me how to trust my data and interpret it with accuracy, without attachment to a result. Steve is a brilliant scientist that leads from his passion and gives his lab members the space to be independent and creative thinkers. I could not have wished for a better mentor and work environment for my Ph.D.

I would like to thank my committee members, Carol Prives and Songtao Jia, for taking the time to provide valuable feedback on my projects over the years. They have continuously offered me new ideas, suggestions, and fresh perspectives on my experiments. Their discussions and guidance were invaluable, and I am grateful to have been able to learn from them. I would also like to thank Vincent Racaniello and Uttiya Basu for generously agreeing to be on my thesis committee and for seeing me through this last stage of my Ph.D.

I would like to thank my colleagues, whom I prefer to call my friends. My days with the Goff lab have been enjoyable because of their support and laughter. Conversations with my fellow Goff lab members have given me hope and joy in the times when I felt most defeated. I would like to first thank Oya for the constant personal and professional guidance, long chats about everything and anything, and being a person of impeccable integrity. Oya has had the pleasure and burden of sharing a lab bay with me and seeing me at my best and worst. Thank you for being there for both times and for showing me true friendship. I would

Chapter 1 : Introduction

1.1 Retroviruses

Impact of retroviruses

Retroviruses are enveloped RNA viruses, named for the distinctive aspect of their replication termed reverse transcription, in which their single-stranded RNA genome is converted into a double-stranded DNA intermediate. The DNA intermediate is then integrated into the host DNA and thereby becomes a stable component of the host genome. Integration of the viral DNA into the host genome results in a mutagenesis event, which can have severe consequences. Depending on the site of integration, this event can disrupt or upregulate the expression of neighboring genes. Retroviruses are also capable of acquiring and adapting cellular genes into their genome during their replicative cycle. This capability can have serious consequences, as certain transduced cellular genes can become oncogenic, leading to the transmission of these virally activated oncogenes to initiate tumor formation. Additionally, if viral genes are integrated into germ cells, the viral genome can get passed down to the offspring in a Mendelian fashion. These inherited retroviruses are called endogenous retroviruses, and vertebrates have accumulated a shocking number of endogenous retroviruses, comprising as much as 8% of the DNA in humans and 10% of the DNA in mice (Lander et al. 2001; Stocking & Kozak 2008). Some endogenous retroviruses that provide beneficial functions to the host have been co-opted by the host genome, demonstrating the important influence endogenous retroviruses have in the evolution of the human genome. Furthermore, retroviruses can cause serious diseases in humans such as the

progression to AIDS in HIV-1 infected patients. Our knowledge of retroviruses was critical for addressing the HIV-1 epidemic and for the development of treatments that are available today. According to the World Health Organization, 36.7 million people continue to live with HIV/AIDS, and 1.1 million people died from AIDS worldwide in just 2015 alone. As such, HIV continues to be a global health issue that further underscores the necessity of retroviral research. In summary, studies of retroviruses have been critical for progressing the fields of virology and cancer research, for understanding our own genomes, and for developing retroviral therapeutics.

A brief history of retroviruses

Retroviruses were first described by Vilhelm Ellermann and Oluf Bang in 1908 and by Peyton Rous in 1911 (Rous 1911). These early studies used cell-free filtrates to transmit avian cancer from diseased animals to healthy animals, demonstrating the presence of a virus. The viruses Ellermann and Bang and Rous described are now known as avian leukosis virus (ALV), and Rous sarcoma virus (RSV), respectively. These early studies were followed by the discovery of mammalian oncogenic retroviruses. The mouse mammary tumor virus (MMTV) was discovered by John Bittner in 1936, and the murine leukemia virus (MLV) was discovered by Ludvik Gross in 1951 (Bittner 1936; Gross 1951).

The next breakthrough in the field came in 1958 when Howard Temin and Harry Rubin established the focus assay, which allowed viral infection and transformation to be studied at the cellular level (Temin & Rubin 1958). Before then, tumor viruses were primarily studied at the organismal level, which was costly, time-consuming, and difficult to

control. The availability of a cell-based model opened the doors for a more quantitative and detailed study of the virus, facilitating important advances in the field of retrovirology.

The most critical and impactful turning point in the field arrived with the discovery of the reverse transcriptase enzyme. Up until this point, researchers remained puzzled as to how retroviruses replicate because viral infections were sensitive to inhibitors of DNA synthesis. Temin first hypothesized that retroviral replication was mediated by a DNA intermediary. This was called the proviral hypothesis and was initially met with a great deal of skepticism until Howard Temin and David Baltimore provided proof for this idea with the landmark discovery of the retroviral reverse transcriptase enzyme in 1970 (Baltimore 1970; Temin & Mizutani 1970). The reverse transcriptase enzyme explained how the DNA intermediate was made from the RNA genome, and importantly, it revised the central dogma of molecular biology, which stated that genetic information travels from DNA to RNA to protein.

The study of retrovirus-induced tumors also led to the discovery of the first oncogene, *src*. This oncogene was discovered in the genome of the RSV and was thought to be a gene originating from the retrovirus. In 1979, Michael Bishop and Harold Varmus made the important discovery of a cellular *src* gene (*c-src*) and demonstrated that viral *src* (*v-src*) is actually derived from *c-src* (H Oppermann 1979). These studies explained that retroviruses were capable of transducing oncogenic genes that originated from host cells. Beyond the study of virology, this study sparked great interest in the role of oncogenes in cancer.

The next major event in the history of retrovirology came with the discovery of the human immunodeficiency virus type 1 (HIV-1), the causative agent of AIDS (Barre-Sinoussi et al. 1983). The AIDS epidemic propelled HIV-1 to the forefront of the retrovirology field and HIV-1 has now become the most studied retrovirus to date. Major work in the study of

HIV-1 has led to important developments in our understanding of retroviruses and in the development of retroviral therapeutics.

Retroviral Classification

Retroviruses were initially classified by virion morphology or host range, but the current taxonomy of retroviruses organizes the *Retrovirdae* family into seven genera based on the sequence relatedness of the *pol* gene, the most well-conserved retroviral gene. The seven genera are alpharetroviruses, betaretroviruses, gammaretroviruses, deltaretroviruses, epsilonretroviruses, lentiviruses, and spumaviruses.

Another broader way of classifying retroviruses is by their genomic content. The common viral genes of all retroviruses are the *gag, pro, pol,* and *env* genes. Retroviruses that only contain this basic set of genes are called "simple" retroviruses. The alpha-, beta-, gamma-, and epsilon-retroviruses are considered simple retroviruses, and murine leukemia virus (MLV) is an example of a simple retrovirus from the gamma-retrovirus genus and a main topic of this dissertation. Retroviruses that contain additional accessory genes are called "complex" retroviruses. Accessory genes code for additional viral products that aid in the replication cycle. The lentiviruses, spumaviruses, and deltaretroviruses are considered complex retroviruses, and HIV-1 is an example of a complex retrovirus of the lentivirus genus.

Endogenous retroviruses

Endogenous retroviruses (ERV) are the result of ancient viral infections of germ cells that have been passed down over generations. While the majority of ERVs are defective and are slowly lost from the host genome, a significant number of ERVs in some species have

4

remained active and are capable of expression and replication (Stocking & Kozak 2008).

Endogenous retroviruses make up about 8% of the human genome and 10% of the mouse

genome (Chinwalla et al. 2002; Lander et al. 2001). As the focus of this dissertation is on

mouse retroviruses, we will briefly describe the mouse ERVs (mERV). mERVs are classified

by phylogenic analysis of the RT domain into three classes. Class I mERVs are one of the

most well-studied ERVs and are close relatives of the gammaretroviruses, class II mERVs

are most related to lentiviruses, alpharetroviruses, betaretroviruses, and deltaretroviruses, and

class III ERVs are most closely related to the spumaretroviruses (Stocking & Kozak 2008).

Retroviral genome

The genomes of retroviruses vary in length and complexity and are 7–12 kb in size.

The minimum and necessary genes found in all retroviruses are the *gag, pro, pol,* and *env*

genes, and these genes are always found in the order listed. The *gag* gene codes for the

internal structural proteins of the virus, the *pro* gene codes for a viral protease, the *pol* gene

codes for the reverse transcriptase and integrase enzymes, and the *env* gene codes for the

surface and transmembrane proteins of the virion.

Accessory genes of complex retroviruses are found between *pol* and *env*, downstream

of *env*, or overlapping these two genes. Some examples of well-studied accessory genes are

those found in HIV-1: *tat, rev, nef, vif, vpr,* and *vpu*. The *tat* gene codes for a transcription

activator, *rev* codes for a regulator of splicing and RNA transport, *nef* codes for a protein

involved in viral assembly, *vif* and *vpr* code for proteins that enhance virion infectivity, and

vpu codes for a membrane protein that facilitates virion release from host cells.

The retroviral genome also contains terminal noncoding regions, which are important for replication. The 5' end contains the R and U5 regions, and the 3' end contains the U3 and R regions. In the process of reverse transcription, these regions are reorganized such that two identical elements called the long terminal repeat (LTR) are formed at the terminal ends of the double-stranded viral DNA (Figure 1-1). The LTRs contain the noncoding regions in the following order: U3, R, and U5. The promoter and enhancer sequences are found in U3, and the site of transcription initiation occurs at the U3/R boundary. The poly(A) addition signal is found in U3 and R, and the site of pol(A) addition is at the R/U5 boundary. Two additional elements that are important for replication are an 18-base region called the primer binding site (PBS) and the polypurine tract (PPT). The 3' end of a tRNA containing a complementary sequence binds to the PBS and is used to prime reverse transcription of the RNA genome to form the first DNA strand, or "minus" strand. The amino acid linked to the complementary tRNA is used to designate the sequence of a PBS. For example, if a PBS sequence uses the proline tRNA to prime reverse transcription, then it is called a PBS_{pro}. The PPT sequence serves as the RNA fragment used to prime the second DNA strand, or "plus" strand. Additional details on the use of these sites for the formation of the dsDNA copy of the RNA genome will be further described in the reverse transcription section.
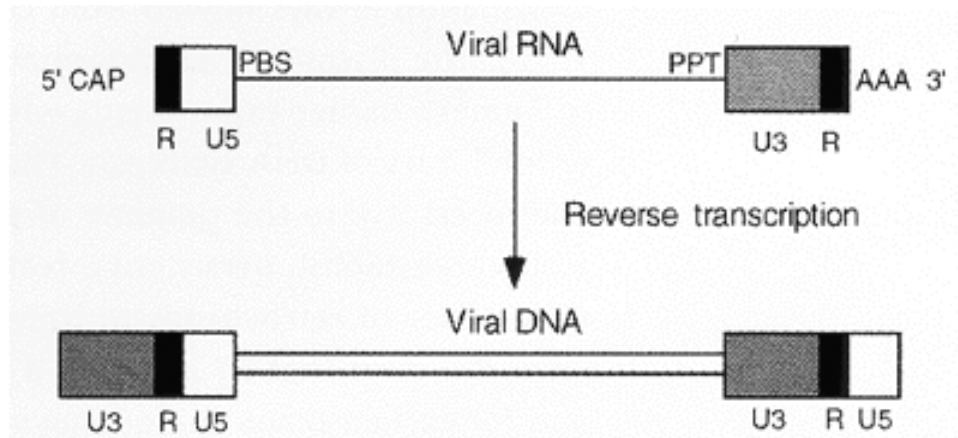
Figure 1-1 Retroviral genome
The viral RNA genome is shown in the first line. The terminal noncoding regions at the 5'
and 3' ends contain the R and U5 regions and the U3 and R regions, respectively. The PBS
and PPT sequences, necessary for reverse transcription of the viral RNA genome, are
marked. The dsDNA copy of the viral genome is shown in the second line. Two identical
LTR elements containing the noncoding regions, U3, R, and U5, are formed in the process of
reverse transcription. Sequences important for transcription, such as the enhancer, promoter,
and poly(A) addition signal, are located in the LTRs. Image taken from (Coffin, et al. 1997).

Retroviral Gene Expression

Retroviral particles are typically 80-120 nm in diameter. At their center, retroviruses harbor two copies of a single-stranded RNA genome and virally-encoded enzymes in a protein capsid shell. An outer plasma membrane, derived from the host cell, encapsulates the capsid core to form a viral particle (Figure 1-2). The plasma membrane is embedded with retroviral glycoprotein complexes, which are important for viral entry into host cells.

The Gag protein is the precursor protein derived from *gag*. Gag contains at least three domains, the matrix (MA), capsid (CA), and nucleocapsid (NC). Gag is proteolytically cleaved during the last stages of the retroviral replication cycle into mature MA, CA, and NC subunits. MA is found at the amino-terminal end of Gag and is crucial for targeting Gag to the plasma membrane. CA is the central domain of Gag, and it forms the capsid shell. NC is at the carboxy-terminal end of Gag and is found in the capsid core, associated with the RNA genome. Additional proteins may also be formed from the Gag precursor in some viruses, but the MA, CA, and NC are the minimal products formed from the Gag protein.

The *pro* and *pol* genes encode all the viral enzymes. The *pro* gene encodes the viral protease (PR), which is responsible for cleaving the viral polyproteins into their mature components, and the *pol* gene encodes the reverse transcriptase (RT) and integrase (IN) enzymes. RT has both DNA polymerase and RNase H activity and is responsible for synthesizing viral DNA, and IN catalyzes the integration of viral DNA into the host genome.

The primary product of *env* forms the Env polyprotein precursor, which is glycosylated and cleaved to form the mature surface protein (SU) and the transmembrane protein (TM). The SU and TM proteins are linked by disulfide bonds and form the glycoprotein complexes embedded in the plasma membrane of the virion. SU is a large
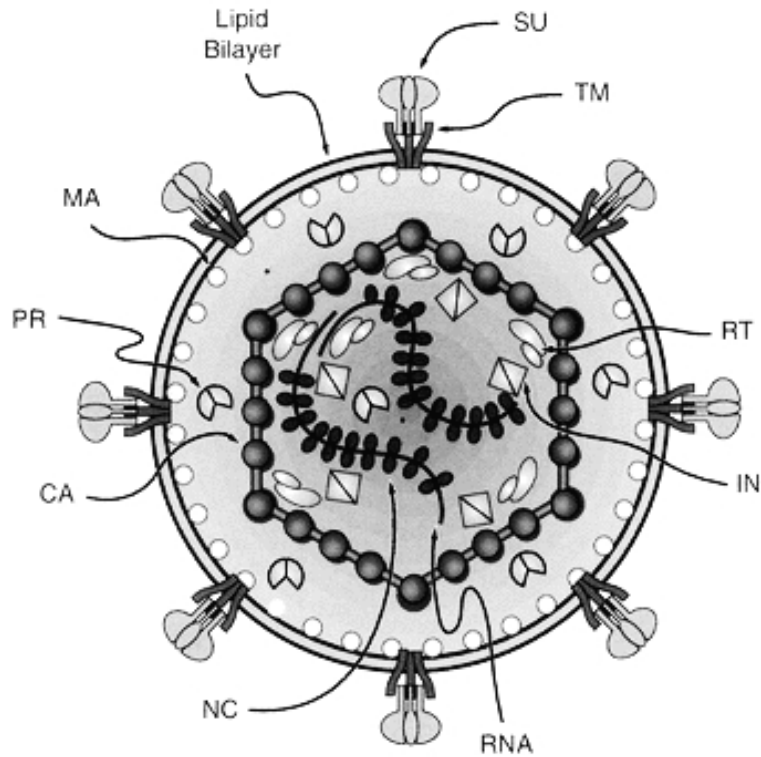
Figure 1-2: Structure and components of a retroviral particle.
The components that form a virion are indicated. MA, CA, and NC are the internal structural proteins formed from the Gag precursor. PR, RT, and IN are the viral enzymes formed from Pro and the Pol precursor. SU and TM are the viral plasma membrane proteins formed from the Env precursor. Image taken from (Coffin, et al. 1997).

glycosylated protein found on the surface of virions and is responsible for interacting with host receptors to gain entry into a cell. SU triggers TM to fuse the viral and host membranes, thereby mediating viral entry into a host cell.

Retroviral replication cycle

The retroviral replication cycle can be cleanly divided into the entry and exit phases. The entry phase consists of viral entry, reverse transcription, nuclear entry, and integration, and the exit phase consists of viral gene expression, assembly, and viral budding (Figure 1-3). The two phases are separated by the establishment of the integrated viral DNA, which is referred to as the provirus. Another distinction is that the entry phase mostly depends on virally-encoded proteins whereas the exit phase mostly depends on the cellular machinery.

Entry and Uncoating

Viral entry is initiated with the binding of SU to specific cell surface receptor molecules of a host cell. Therefore, the expression of the receptor defines the host range and cell tropism of a virus. For example, the Envelope of the ecotropic Moloney murine leukemia virus (M-MLV) binds to the mouse amino acid transporter, mCAT (Maddon et al. 1986; Kim et al. 1991). Binding of SU to a cell surface receptor activates TM, resulting in the fusion of the viral and cell membranes and entry of the capsid core into the cytoplasm. The capsid core now begins to disassemble in a process called uncoating.
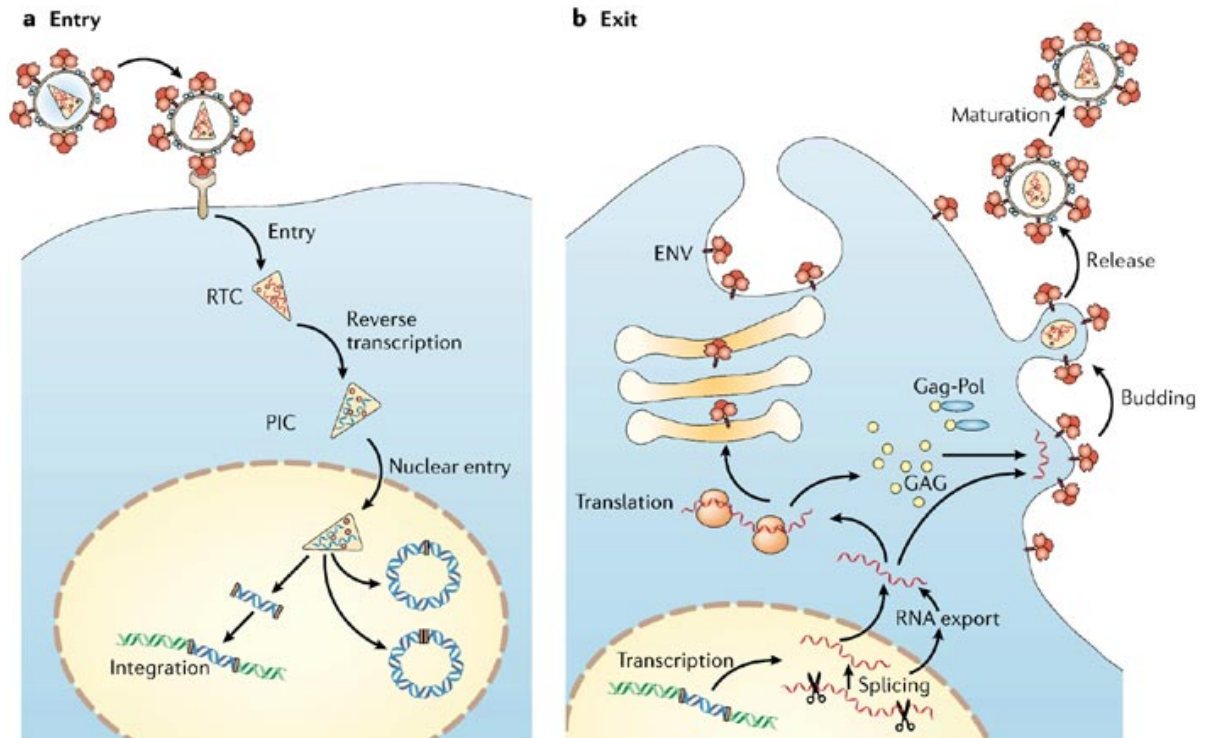
Figure 1-3 The entry and exit phases of the retroviral replication cycle
A schematic of the entry and exit phases of the retroviral replication cycle are shown. (A) The major sequence of events in the entry phase consists of viral entry, reverse transcription, nuclear entry, and integration. (B) The major sequence of events in the exit phase consists of transcription, assembly, and budding. Image taken from (Goff 2007).

Reverse Transcription

Following release of the viral core into the cytoplasm, RT is activated and initiates

reverse transcription of the RNA genome (Figure 1-4). It is not completely clear how RT is

activated, but it is thought that uncoating of the capsid core and exposure to

deoxyribonucleotides are involved (Luban et al. 1993; Goff 2001). The viral particle carries

two copies of the single-stranded RNA genome, and retroviral DNA synthesis can occur

from either strand. The 3' end of a partially unwound cellular tRNA, annealed to the RNA

genome during virion assembly, serves as primer to initiate DNA synthesis. RT synthesizes

the minus-strand DNA until it reaches the 5' end of the genomic RNA, at which point, it

digests the RNA strand of the RNA:DNA duplex. The minus-strand DNA "jumps" to the 3'

end of the RNA genome where it contains an R sequence that complements the 3' end of the

DNA strand, a process referred to as strand transfer. After strand transfer occurs, minus-

strand DNA synthesis resumes and proceeds through the PBS sequence near the 5' end of the

RNA. As elongation proceeds, RNase H digests the template RNA strand except for a short

polypurine tract, the PPT RNA segment. The PPT is resistant to RNase H degradation and

serves as the primer for plus-strand DNA synthesis. Plus-strand synthesis proceeds into a

portion of the tRNA still present at the 5' end of the minus strand. This step is followed by

RNase H degradation of the tRNA primer, which creates a PBS overhang on the plus-strand

DNA. A second strand transfer occurs in which the PBS overhang anneals to the PBS of the

minus strand. This allows the plus-strand to complete DNA synthesis, resulting in a dsDNA

copy of the RNA genome with identical LTR sequences at the 5' and 3' ends.
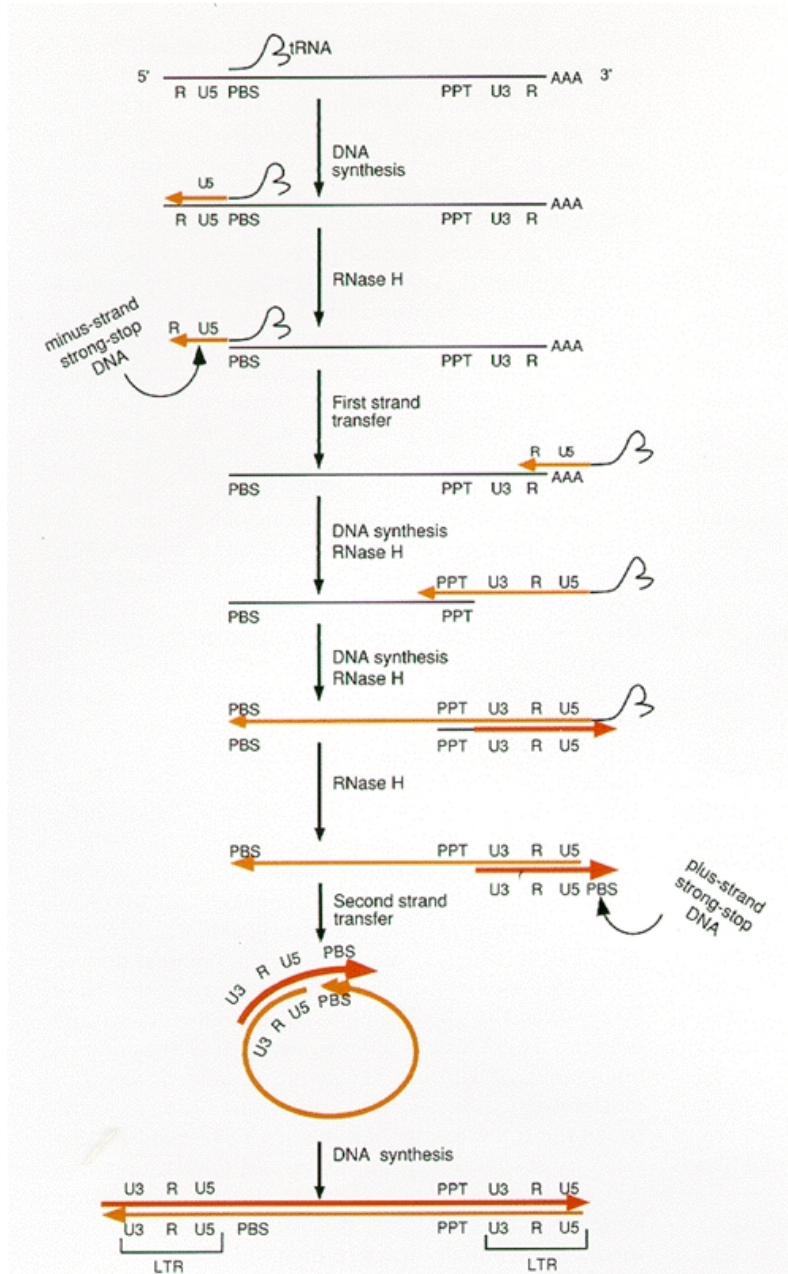
Figure 1-4 Reverse transcription of the retroviral genome.
Reverse transcription of the RNA genome involves a complicated sequence of steps outlined here. These steps include a series of DNA synthesis and RNA degradation events, catalyzed by RT, and two strand-transfer events. Image taken from (Coffin, et al. 1997).

Nuclear entry and Integration

The newly synthesized DNA product moves from the cytoplasm to the nucleus via association with IN and host cell factors that form a complex known as the pre-integration complex (PIC). Some retroviral PICs can only enter the nucleus after breakdown of the nuclear envelope during cell division, while others can enter through active transport. Upon migrating into the nucleus, IN cleaves two bases from the 3' ends of both viral DNA strands to create 5' overhangs. Subsequently, the PIC targets host DNA where IN catalyzes the breaking and joining of host DNA to viral DNA in a transesterification reaction. The last step of integration involves the repair of two short single-stranded gaps and base pair mismatches that result from insertion of the viral DNA. The integrated viral DNA is now called a provirus, and its integration into the host genome is irreversible.

Transcription and Splicing

Following integration, the provirus behaves like a cellular gene and mostly relies on cellular machinery for gene expression and assembly of progeny virions. The duplicate LTRs formed from reverse transcription contain *cis*-acting elements that direct viral gene expression, recognized by the host cell RNA polymerase II and transcriptional cofactors. Regulatory sequences for initiating transcription, such as the promoter and enhancer elements, are found in the U3 region of the 5' LTR. The control elements for posttranscriptional processing are found in the U3, R, and U5 of the 3' LTR. Transcription of the provirus, 5' end capping, and 3' end polyadenylation of the viral transcript proceeds as it would for any cellular gene. There is only one promoter used for the expression of *gag, pro, pol,* and *env,* and a single primary transcript is formed. This full-length transcript serves as

the mRNA for the Gag and Gag-Pro-Pol polyproteins, as well as the genomic RNA to be packaged into progeny viral particles. A fraction of the full-length transcript is spliced into subgenomic viral RNA, resulting in transcripts for generating the Env polyprotein and accessory proteins. Simple retroviruses undergo a single splicing event to generate the transcript encoding Env, while complex retroviruses can undergo multiple splicing events to generate transcripts for Env and the additional accessory products (Figure 1-5).

Synthesis, Assembly, and Budding

The Env polyprotein is formed from a spliced transcript, the Gag protein is formed from the unspliced transcript, and the viral enzymes are produced as fusions proteins to Gag in a Gag-Pro-Pol polyprotein. Translation of the unspliced transcript produces the Gag polyprotein in highest abundance, but regulated frameshift or readthrough events allow the ribosome to bypass the *gag* stop codon and form the Gag-Pro-Pol polyprotein. These polyproteins are later cleaved during viral maturation into their individual subunits.

The viral proteins are translated in different regions of the cell, but they assemble at the plasma membrane when forming new viral progeny. The Gag and Gag-Pro-Pol polyproteins are synthesized on free polyribosomes in the cytoplasm and are subsequently directed to the cytoplasmic side of the plasma membrane by Gag. Gag is also responsible for bringing two copies of the viral RNA genome to the site of assembly and for initiating budding of the plasma membrane. The Env polyprotein is synthesized on membrane-bound polyribosomes, is delivered through the golgi apparatus and endoplasmid reticulum to the plasma membrane, and finally migrates laterally to the sites of budding. Budding from the host cell results in the packaging of the viral proteins, RNA genome, and cellular tRNA to
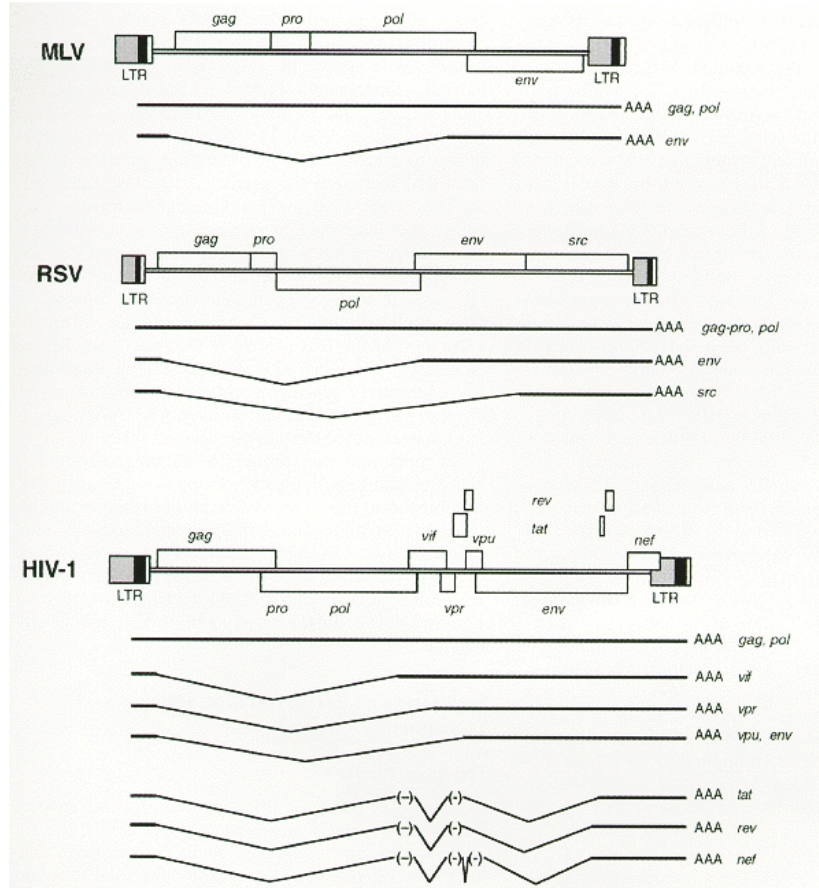
Figure 1-5 Alternative splicing patterns for MLV, RSV, and HIV-1 transcripts.
MLV is a simple retrovirus that only undergoes a single splicing event to generate the
transcript encoded by the *env* gene. RSV is also a simple retrovirus that undergoes two
splicing events, generating two subgenomic RNA products encoded by the *env* and *src* genes.
HIV-1 is a complex retrovirus that utilizes multiple splicing events to generate RNA
templates for the expression of Env and its accessory products. Image taken from (Coffin, et
al. 1997).

form "immature" virions. Following the release of the virion from the host cell, the polyproteins are cleaved into their respective subunits by viral protease to produce the mature virion (Figure 1-6). The mature virion can now infect another host cell and repeat the retroviral replication cycle.

Murine Leukemia Virus

Ludwik Gross first discovered MLV in 1951 and demonstrated that the virus transmitted leukemia to newborn mice (Gross 1951; Gross 1957). Gross's initial discovery of MLV and ensuing studies helped spark great interest in the field of tumor virology and established MLV as a model system for studying retroviruses. MLV is a prototype of the gammaretrovirus genus and is considered a simple retrovirus. MLV does not infect non-dividing cells and generally does not cause cell death of host cells upon infection. (T Roe 1993) There are various strains of MLV, including Friend, Rauscher, and Moloney. We used the Moloney strain in the studies described here.

Receptors

MLVs can be organized into four subgroups according to their receptor usage. Ecotropic MLV can only infect host cells of mouse or rat origin, xenotropic MLV can infect cells of non-murine origin, and polytropic and amphotropic MLV can infect cells of both murine and non-murine origin. Most ecotropic MLVs use the ubiquitously expressed murine cationic amino acid transporter (mCAT-1), amphotropic MLVs use a sodium-dependent inorganic phosphate transporter (PiT2), and xenotropic and polytropic MLVs use the Xpr1 molecule (Albritton et al. 1989; Battini et al. 1999; D. G. Miller & A. D. Miller 1994). Additional receptors can also be utilized by exchanging the natural envelope proteins of
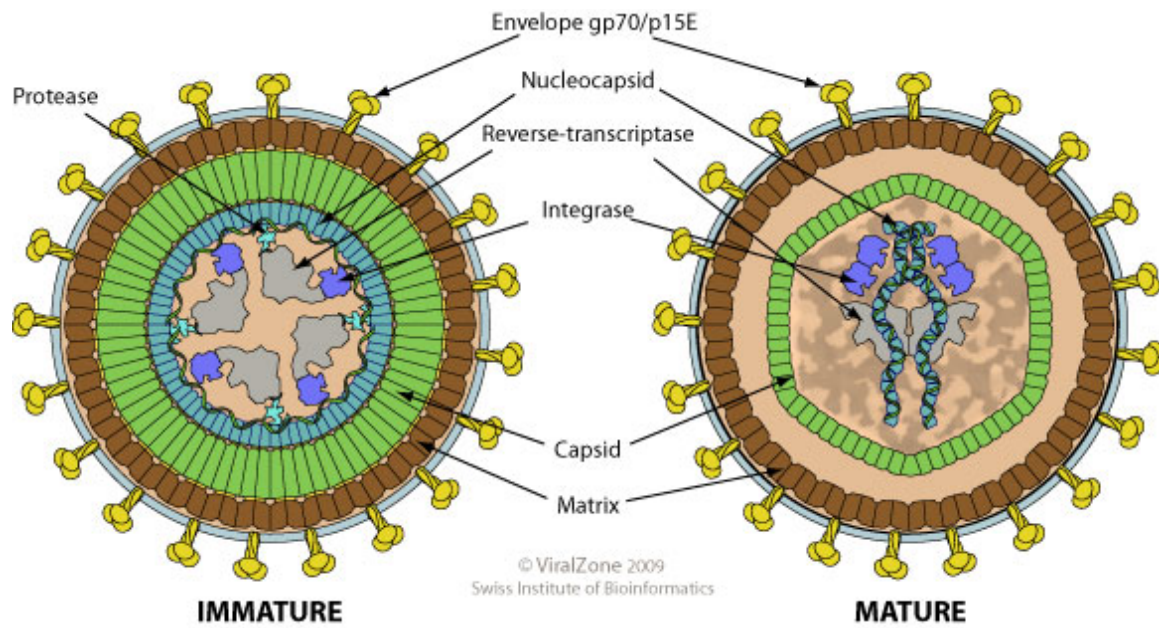
Figure 1-6 Structure of the immature and mature viral particle
The immature viral particles prior to viral polyprotein cleavage is shown on the left, and the mature viral particle is shown on the right. Image taken from ViralZone:www.expasy.org/viralzone, Swiss Institute of Bioinformatics.

MLV for those of another virus. This process is called pseudotyping and is advantageous for altering host tropism and increasing gene delivery by retroviral vectors. A commonly used pseudotyping envelope protein is the glycoprotein of vesicular stomatitis virus (VSVg) because it can transduce a variety of cells from different species and is relatively stable to physical handling.

Fv1 gene

MLVs can also be categorized by their susceptibility to restriction by the *Fv1* gene. Frank Lilly's lab was the first to describe the *Fv1* locus and its ability to control MLV infectivity (Theodore Pincus 1971). Two major *Fv1* alleles, $Fv1^n$ or $Fv1^b$, determine resistance to MLV strains. MLVs blocked by the protein coded by the $Fv1^n$ allele are classified as B-tropic, and MLVs blocked by the protein coded by the $Fv1^b$ gene are classified as N-tropic. Expression of both alleles ($Fv1^{n/b}$) blocks infection by both N-tropic and B-tropic MLV. Some MLV retroviruses, such as M-MLV, are not blocked by the proteins coded by either *Fv1* alleles and are classified as NB-tropic MLV. Ironically, the *Fv1* gene encodes an ancient and degenerate retroviral Gag protein. The Gag produced by the *Fv1* gene binds CA, restricting the replication cycle at a point between reverse transcription and nuclear entry of the PIC. Following the discovery of *Fv1*, many more inhibitory host factors have been identified and these will be further described in the following section.

1.2 M-MLV Silencing in Embryonic Cells

Host Factors with Inhibitory Activity to Retroviruses

Host factors that interact with retroviruses can function to either support or inhibit the retroviral replication cycle. Some supporting host factors were addressed in the discussion of the retroviral replication cycle above, such as the cellular machinery used in proviral transcription and translation. However, host cells have also evolved mechanisms to protect themselves from retroviral infections. Inhibitory host factors target various stages of the retroviral replication cycle. Some such inhibitory factors include the *Fv1* gene product, APOBEC3G, TRIM5α, Trim28, ZAP, and tetherin (Figure 1-7). The *Fv1* gene product was the first host restriction factor identified, found to block nuclear import of the pre-integration complex for certain strains of MLV (Theodore Pincus 1971). APOBEC3G restricts HIV-1 replication by creating deoxycytidine to deoxyuridine mutations on the minus-strand DNA during reverse transcription (Sheehy et al. 2002; Goff 2004). Some species-specific variants of TRIM5α interfere with viral uncoating of HIV-1 and certain strains of MLV (Stremlau et al. 2004). Trim28 silences transcription of exogenous and endogenous MLV in mouse embryonic cells (Wolf & Goff 2007; Rowe et al. 2010). ZAP targets viral RNA for degradation, thereby restricting replication of MLV as well as the replication of some alphaviruses such Sindbis and Ebola virus (Gao 2002). Tetherin prevents the release of progeny HIV-1 particles from the surface of infected cells (Neil et al. 2008). Collectively, these inhibitory host factors demonstrate various strategies host cells developed to restrict retroviral replication. In our studies, we focused on the mechanism responsible for transcriptional silencing of M-MLV in mouse embryonic cells.
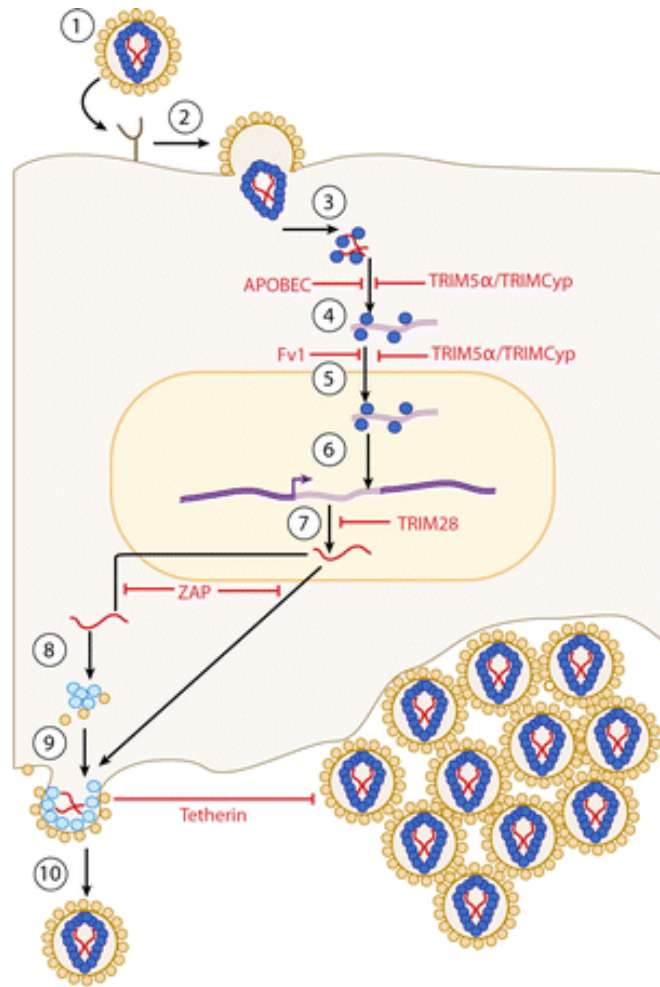
Figure 1-7 Host factors that inhibit retroviral replication
Major stages of the retroviral replication cycle are numbered from 1-10. Some inhibitory host factors and the stages at which they block the replication cycle are indicated in red. APOBEC3G inhibits reverse transcription, the *Fv1* product restricts nuclear entry, Trim5α can interfere with viral uncoating, Trim28 inhibits transcription, ZAP degrades viral mRNA, and tetherin prevents the release of viral particles. Image taken from (Wolf & Goff 2008).

M-MLV Silencing in Embryonic Cells

The Moloney murine leukemia virus (M-MLV) is a prototypical retrovirus that depends heavily on the host cell machinery for its replication. While M-MLV can carry out a productive infection in most cell types, virus replication is potently restricted in embryonic stem (ES) cells and embryonic carcinoma (EC) cells (Teich et al. 1977). M-MLV entry, reverse transcription, and viral DNA integration into the host genome occurs normally (Stewart et al. 1982), but the replication cycle is blocked at the transcriptional level (Niwa et al. 1983; Gautsch & Wilson 1983). Transcriptional silencing of M-MLV in embryonic cells is thought to be attributable to the absence of enhancer protein function (Linney et al. 1984; Hilberg et al. 1987) and the repressive activity of a *trans*-acting silencing complex (Loh et al. 1990; Petersen et al. 1991). The repressor complex mediates immediate proviral silencing in addition to establishing permanent gene silencing by *de novo* methylation of the provirus (Stewart et al. 1982; Leung et al. 2011; Niwa et al. 1983).

Cis-acting Elements

Several *cis*-acting elements contribute to the silencing of M-MLV in embryonic cells. First, enhancer elements are inactive in embryonic cells because of the absence of compatible transcription factors that bind to these sites (Linney et al. 1984; Hilberg et al. 1987). Thus, insertion of enhancers derived from viruses that express well in embryonic cells, improves M-MLV expression. Secondly, repressive elements play a significant role in limiting proviral expression. The PBS_{pro} element is a major site of repression, and a single base-pair mutation in the PBS called the B2 mutation (PBS_{B2}) is sufficient to relieve proviral repression

(Barklis et al. 1986; Loh et al. 1987). Swapping the PBS$_{pro}$ for the PBS$_{gln}$ sequence also de-represses M-MLV (Grez et al. 1990). Moreover, the PBS$_{pro}$ element is capable of silencing the provirus independent of the position and orientation of the site, indicating that the repressive function occurs at the DNA level (Loh et al. 1990). Interestingly, the primary function of the PBS is to bind a complementary 3' tRNA sequence onto the viral RNA genome to prime viral DNA synthesis, but the repression and DNA synthesis priming activity at this site appear to function independently of each other. Another known site for repression occurs at the negative coding region (NCR) located in the 5' LTR of M-MLV (Flanagan et al. 1989). Deletion of the NCR relieves methylation of the provirus and silencing of M-MLV in embryonic cells (Schlesinger et al. 2013). Thus, silencing activity at both the PBS and NCR elements are needed to achieve maximal M-MLV silencing in embryonic cells.

Trans-acting Proteins

The presence of _trans_-acting factors was first shown from the ability to saturate M-MLV repression with increasing amounts of competitor DNA containing the PBS region (Loh et al. 1988). The _trans_-acting factors that bind the PBS sequence are referred to as the repressor binding site (RBS) complex, and this complex specifically binds to the PBS$_{Pro}$ sequence and not the PBS$_{gln}$ or PBS$_{B2}$ sequences that relieve repression (Loh et al. 1990; Wolf & Goff 2007). Recent studies have identified some key components of the _trans_-acting silencing complex. Tripartite motif-containing 28 (Trim28) is a major repressor protein that mediates repressive activity on both the PBS and NCR elements (Wolf & Goff 2007; Schlesinger et al. 2013), and zinc finger protein 809 (ZFP809) and yin yang 1 (YY1) are

DNA-binding proteins that recruit Trim28 to these elements (Flanagan et al. 1992; Wolf & Goff 2009; Schlesinger et al. 2013).

Trim28

Trim28 is also known as KRAB-associated protein-1 (KAP1) or transcriptional intermediary factor 1β (TIF1β) and is a ubiquitously expressed transcriptional repressor protein that serves as a scaffold for additional co-repressor proteins. Protein-DNA binding assays using embryonic cell lysates show that Trim28 binds to the restrictive PBS_{pro} site, but not the PBS_{B2} or PBS_{gln} sites, and knock down of Trim28 relieves embryonic repression of M-MLV (Wolf & Goff 2007). Trim28 is also recruited to the NCR element, which leads to the enrichment of repressive markers at this site (Schlesinger et al. 2013). Additionally, Trim28 is responsible for silencing endogenous retroviruses in embryonic cells (Rowe et al. 2010; Matsui et al. 2010). Trim28 is enriched on certain endogenous retroviral genes and knockdown of Trim28 results in the upregulation of these ERVs. Trim28 also restricts HIV-1, but it inhibits HIV-1 integration rather than its transcription by mediating the deacetylation of IN (Allouch et al. 2011). In addition to retroviral silencing, Trim28 has important functions in embryonic development, the regulation of T-cell activation, and mediating the DNA damage response, amongst others (Messerschmidt et al. 2012; Ziv et al. 2006; Chikuma et al. 2012). Thus, Trim28 regulates a variety of cellular pathways, including retroviral expression.

Trim28 repressor complex

Trim28 serves as a scaffold for the nucleosome remodeling deacetylase (NuRD) complex, heterochromatin protein 1 (HP1), and the histone-lysine N-methyltransferase

ESET, which collectively contribute to heterochromatin formation and histone modifications to maintain long-term gene silencing (Matsui et al. 2010; Wolf, Cammas, et al. 2008; Schultz 2002; Schultz et al. 2001; B Le Douarin 1996). The NuRD complex contains several subunits, including chromodomain-helicase-DNA-binding protein 3 (CHD3) and the histone deacetylases 1 and 2 (HDAC1/2). CHD3 remodels nucleosomes by altering histone-DNA contacts, and HDAC removes acetylation modifications from histones. This coupled activity facilitates heterochromatin formation and leads to gene silencing (Denslow & Wade 2007). ESET trimethylates lysine 9 of histone H3 (H3K9me3), which is a mark of heterochromatin and a binding site for HP1. HP1 is a key structural protein in the assembly of the repressor complex and stabilizes the heterochromatin domain (Maison & Almouzni 2004). Trim28 interacts with a large family of repressor transcription factors called Krüppel-associated box zinc fingers (KRAB-ZNFs), which recruit Trim28 and its co-repressors to specific DNA sequences. Trim28 interacts with these co-repressor proteins using several domains, including the ring B1-B2 coiled-coil (RBCC) domain, chromoshadow domain, plant homeodomain (PHD), and bromodomain. The KRAB-ZFPs interact with the RBCC domain, HP1 interacts with the chromoshadow domain, and CHD3 and ESET interact with the PHD and bromodomain (Friedman et al. 1996; Lechner et al. 2000; Schultz et al. 2001; Schultz 2002). Thus, Trim28 is the central repressor factor in the M-MLV silencing complex, linking repressor function to specific gene targets.

Trim28 modifications

Trim28 activity is regulated by multiple post-translational modifications. In particular, sumoylation and phosphorylation modifications on Trim28 have both been shown

to regulate Trim28 repressive activity in the context of the DNA damage response pathway, cell cycle progression, and viral latency (Ziv et al. 2006; Goodarzi et al. 2011; Ivanov et al. 2007; Y.-K. Lee et al. 2007; Benjamin Rauwel 2015; Li et al. 2007). Trim28 contains six lysine residues that can be conjugated by the SUMOs: K554, K575, K676, K750, K779, and K804. Of these six, K554, K779, and K804 are previously reported to be the most important sites for Trim28 repressive activity (Ivanov et al. 2007; Y.-K. Lee et al. 2007; Mascle et al. 2007; Goodarzi et al. 2011). In a yeast two-hybrid assay, the interaction of Trim28 with ESET was disrupted by a mutation at K676, and Trim28 interaction with CHD3 and ESET was disrupted by a double mutation at K676 and K779 (Ivanov et al. 2007). In mammalian cells, mutating K554, K779, and K804 to arginine results in chromatin relaxation and in transcriptional derepression of the p21 promoter and of a reporter gene in a Gal4-based system (Y.-K. Lee et al. 2007; Goodarzi et al. 2011; Mascle et al. 2007). Interestingly, the PHD of Trim28 functions as an E3 ligase and contributes to its auto-sumoylation as well as to the sumoylation of other proteins (Ivanov et al. 2007; Liang et al. 2011).

The SUMO family consists of four members – SUMO1, SUMO2, SUMO3, and SUMO4. While most of the work on Trim28 sumoylation has been done with SUMO1, SUMO2 has also been specifically implicated in Trim28-mediated viral silencing. A recent genome-wide siRNA screen for proviral silencing factors in embryonic cells identified SUMO2 and the SUMO2 conjugating enzymes, but not SUMO1 or SUMO3 (B. X. Yang et al. 2015). This study also shows that SUMO2 knock down (KD) abolishes Trim28 binding to the LTRs of endogenous retroviruses, and Trim28 KD reduces SUMO2 enrichment on proviral elements, suggesting that SUMO2 modification of Trim28 can regulate its proviral silencing activity.

Phosphorylation of Trim28 abrogates its repressive activity and reduces SUMO conjugation of Trim28 (Li et al. 2007; Benjamin Rauwel 2015; Goodarzi et al. 2011). Specifically, phosphorylation of two residues, S824 and S473, reduces Trim28 repressive activity by perturbing the interaction of Trim28 with CHD3 and HP1, respectively (Goodarzi et al. 2011; Chang et al. 2008). Trim28 is also involved in establishing human cytomegalovirus (HCMV) latency in hematopoietic stem cells, and phosphorylation of Trim28 at the S824 residue relieves latency (Benjamin Rauwel 2015). While these studies demonstrate that post-translational modifications on Trim28 regulate its repressive activity, it is unknown if the reported modifications are also critical for Trim28-mediated repression of M-MLV in embryonic cells.

KRAB-ZNF family

As Trim28 cannot directly bind to DNA, it depends on DNA-binding proteins to bridge its interaction to specific DNA sequences. The KRAB-ZNF family constitutes the largest family of transcription regulators in human and mice and is known to recruit Trim28 to sites of gene silencing (Emerson & J. H. Thomas 2009). KRAB-ZFPs bind to specific DNA sequences by their zinc fingers at their C-termini, and recruit Trim28 by their KRAB domain at their N-termini (Wolfe et al. 2000). Interestingly, Trim28 is enriched on the genes of KRAB-ZNFs, suggesting a possible auto-regulatory role for the KRAB-ZNFs that involves Trim28 (O'Geen et al. 2007). In addition to embryonic-specific silencing of retroviruses, KRAB-ZNFs are important for the establishment of global embryonic DNA methylation (Quenneville et al. 2012; Wolf & Goff 2009).

ZFP809

ZFP809 is a member of the KRAB-ZNFs, containing a KRAB domain at its N-terminus and seven zinc fingers at its C-terminus. ZFP808 interacts with Trim28 from its KRAB domain and recognizes the PBS_{pro} sequence of M-MLV by its zinc fingers (Wolf & Goff 2009). Mutation of the PBS region or knockdown of ZFP809 significantly reduces M-MLV silencing in embryonic cells (Wolf & Goff 2009; Barklis et al. 1986). ZFP809 protein is specifically expressed in embryonic cells and does not normally express well in differentiated cells, but expression of a truncated ZFP809 protein was capable of blocking M-MLV expression in differentiated cells. Our lab recently demonstrates that ZFP809 is ubiquitinated and rapidly degraded in differentiated cells, but not in embryonic cells (Cheng Wang, unpublished results). This finding provides one explanation for the embryonic-specific silencing activity occurring on M-MLV.

ZFP809 is not the only DNA-binding protein responsible for recruiting Trim28 to M-MLV. Trim28 is enriched on other regions distant from the PBS domain, and introducing the PBS_{B2} mutation does not relieve all repressive activity (Schlesinger et al. 2013). Moreover, M-MLV containing a PBS_{lys1-2} sequence or endogenous retroviruses containing a PBS_{phe} sequence were also silenced by the Trim28 silencing complex, although ZFP809 cannot bind these PBS sequences (Wolf, Hug, et al. 2008; Rowe et al. 2010). These observations suggest the presence of a PBS-independent silencing mechanism that utilizes other ZFPs to recruit Trim28 to M-MLV. Recently, we demonstrated that YY1 is another such zinc-finger protein that recruits Trim28 to the 5'LTR of M-MLV (Schlesinger et al. 2013).

YY1

YY1, also known as the upstream conserved region binding protein (UCRBP), is a well-conserved and ubiquitously expressed protein that derives its name from its unique ability to function as both a transcriptional activator and repressor (Shi et al. 1997). YY1's activator function is roughly mapped to the N-terminal end, and its repressor function is mapped to the zinc fingers at its C-terminal end (Shi et al. 1997). YY1 contains four zinc fingers that recognize a consensus binding motif found in the promoters of a wide variety of cellular and viral genes, suggesting its involvement in a vast range of cellular functions (Hyde-DeRuyscher et al. 1995). Some notable cellular genes regulated by YY1 include c-myc, c-fos, and IFN-γ (Shi et al. 1997). YY1 also modulates apoptosis, cellular proliferation, and cancer progression from its interaction with several significant cell cycle proteins such as p53, Mdm2, Sp1 and retinoblastoma (Rb) (Gordon et al. 2005; J. S. Lee et al. 1993). Interestingly, YY1 also interacts with RNA. In the context of X-chromosome inactivation in females, YY1 tethers the noncoding RNA, Xist, to the X- inactivation center to mediate silencing of one X-chromosome (Jeon & J. T. Lee 2011).

In addition to its cellular functions, YY1 is also implicated in regulating retroviral expression. We became interested in investigating YY1 in the context of M-MLV silencing in embryonic cells because of an early report by Flanagan and colleagues demonstrating YY1 binding to the NCR element (Flanagan et al. 1989; Flanagan et al. 1992). In this study, YY1 was shown to down-regulate M-MLV transcription in a dose-dependent manner. YY1 was also found to associate with LSF (late simian virus 40 transcription factor) and HDAC1 to mediate HIV-1 latency in CD4+ cells (Romerio et al. 1997; Coull et al. 2000). Moreover, unpublished observations from our lab suggest HTLV-1 is positively regulated by YY1

interaction with the HTLV-1 RNA (Gary Wang, unpublished results). Thus, YY1 is a multifaceted protein that interacts with DNA, RNA, and proteins to regulate cellular and viral functions.

These studies led us to investigate the involvement of YY1 in the silencing of M-MLV in embryonic cells. In a study led by Sharon Schlesinger, we demonstrated that endogenous YY1 protein specifically binds to the NCR element and that YY1 KD or deletion of the NCR element results in the de-repression of M-MLV in embryonic cells (Schlesinger et al. 2013). This is further supported by ChIP experiments and bisulfite analysis demonstrating that YY1 binding is associated with increased H3K9me3 and DNA methylation marks on M-MLV. We also found that intracisternal A particles (IAPs), encoded by a family of class II mERVs, are upregulated in the absence of YY1, demonstrating that YY1 is involved in the silencing of both exogenous and endogenous retroviruses. Unlike ZFP809, YY1 does not contain a KRAB domain, but co-immunoprecipitation experiments show that YY1 also binds and recruits Trim28 to the provirus. Taken together, these studies demonstrate that the embryonic-specific silencing of M-MLV depends on Trim28, ZFP809, and YY1.

One unexplored aspect of the YY1-Trim28 silencing complex is the nature of the biochemical interaction of the two proteins. As YY1 does not contain a domain predicted to interact with Trim28, it is unknown how this interaction occurs. An intriguing finding demonstrated that the YY1-Trim28 interaction occurs in embryonic cells but not in differentiated cells (Figure 1-8) (Schlesinger et al. 2013), even though YY1 and Trim28 are expressed in both cell types. This suggests that the YY1-Trim28 interaction could be differentially regulated at the post-translational level, which could explain a portion of the
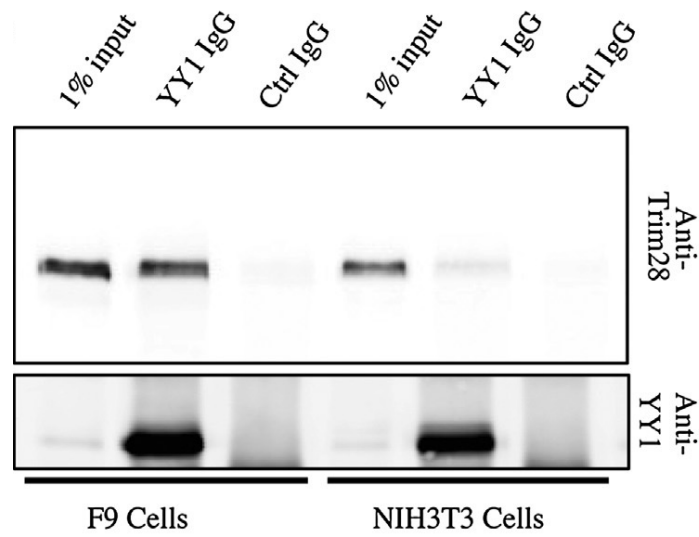
Figure 1-8 Detection of Trim28 with the immunoprecipitation of YY1
Trim28 is shown to interact with YY1 in F9 cells, but not in NIH3T3 cells. Anti-YY1 antibody was used to immunoprecipitate YY1 from nuclear extracts of the F9 embryonic carcinoma cell line and of the NIH3T3 differentiated cells line. Anti-YY1 immunoprecipitates were analyzed by western blot with anti-Trim28 and anti-YY1 antibodies. 1% of total lysates used for the immunoprecipitation reactions are shown in the input lane as a comparison. Image taken from (Schlesinger et al. 2013).

embryonic cell-specificity of M-MLV silencing. A possible mechanism for regulating this interaction includes the presence of supporting cofactors or absence of inhibitory cofactors. Another possibility is that post-translational modifications regulate the YY1-Trim28 interaction. There are no strong candidates for possible cofactors involved in the interaction, but there are a number of studies indicating the importance of sumoylation and phosphorylation modifications for regulating Trim28 interactions and repressive activity, as previously described (Ziv et al. 2006; Goodarzi et al. 2011; Ivanov et al. 2007; Y.-K. Lee et al. 2007; Benjamin Rauwel 2015; Li et al. 2007). In Chapter 3, we investigated the biochemical interaction of YY1 and Trim28 and the role of Trim28 modifications in the silencing of M-MLV. We mapped the domains involved in the YY1-Trim28 interaction and demonstrated that YY1 and Trim28 do no interact *in vitro*. Sumoylation and phosphorylation modifications were not found to be necessary for the YY1-Trim28 interaction, but the K779 residue in Trim28 was necessary for the silencing of M-MLV. Our results are consistent with what has been reported for the importance of K779 in other settings, but we did not find the K554, K804, S473, or S824 residues to be important for Trim28 repression of MLV, suggesting that there are significant differences between the Trim28-dependent repressive mechanisms occurring on M-MLV in embryonic cells and those involved in the DNA repair response, cell cycle progression, and HCMV latency (Y.-K. Lee et al. 2007; Ivanov et al. 2007; Goodarzi et al. 2011; Benjamin Rauwel 2015).

1.3 Small Ubiquitin-Like Modifiers (SUMO)

Our interest in the regulation of Trim28 silencing activity by SUMO modifications led us to an interesting finding in which we observed a strong block to ectopic SUMO1 expression in embryonic cells, but not in differentiated cells. We have investigated this finding further and our studies are reported in Chapter 5. Here, we provide a more in-depth background on SUMO and what is currently known about SUMO function in embryonic cells.

Introduction to SUMO

The conjugation of SUMO to protein substrates is a post-translational modification that impacts a diverse range of cellular processes such as transcriptional regulation, RNA processing, viral repression, the DNA damage response, and protein localization (Geiss-Friedlander & Melchior 2007; Verger et al. 2003). The downstream consequences of SUMO conjugation are quite varied and substrate specific. Sumoylation or desumoylation can result in altered protein-protein interactions, protein-DNA interactions, protein stability, protein trafficking, or protein activity (Willson 2009). Several important mammalian proteins are modified and regulated by sumoylation including RanGAP1, p53, c-Jun, Trim28, and HDAC1, and the list of proteins modified by SUMO continues to expand (Matunis 1996; Muller 2000; Mascle et al. 2007). Recent proteomic studies estimate that up to 15% of human proteins may be modified by SUMO, revealing a large body of additional proteins and pathways that have yet to be investigated for SUMO involvement (Hendriks & Vertegaal 2016). Moreover, aberrations in the SUMO pathway can result in tumor development and progression, heart defects, and the Alzheimer's disease (Flotho & Melchior 2013; Kessler et

al. 2012; L. Lee et al. 2013; Wang et al. 2011). Thus, our growing understanding of SUMO

and its vast roles in the cell continues to highlight the importance of this modification for

normal cellular function.

SUMO Family Members

The SUMO family consists of four members: SUMO1, SUMO2, SUMO3, and

SUMO4. SUMO4 is only expressed in the kidney, dendritic cells, and macrophages, so we

do not address SUMO4 in our studies (Geiss-Friedlander & Melchior 2007; Bohren et al.

2004). SUMO2 and SUMO3 are often referred to as SUMO2/3, because they are 97%

identical and are thought to modify the same substrates and affect the same cellular

functions. On the other hand, SUMO1 only shares 50% sequence identity with SUMO2/3

and demonstrates distinct cellular functions from SUMO2/3 (Geiss-Friedlander & Melchior

2007). SUMO2/3 also differs in its ability to form polymeric chains on substrates whereas

SUMO1 does not (Tatham et al. 2001). Another difference is that SUMO2/3 predominantly

exists as free or non-conjugated proteins, whereas SUMO1 is predominantly found

conjugated to protein targets (Saitoh 2000). Moreover, some substrates demonstrate

preferences for one isoform while others are conjugated by either isoforms (Zhang et al.

2008). However, it is unclear how relevant these differences are as SUMO2/3 was found to

compensate for SUMO1 in SUMO1 knockout (KO) mice (Zhang et al. 2008).

SUMO Conjugation

The SUMOs resemble ubiquitin in structure and mechanics, as their name implies.

Similar to ubiquitination, sumoylation involves an enzymatic cascade of three reactions,

resulting in SUMO conjugation to a substrate via an isopeptide bond. Targets of sumoylation

commonly contain the tetrapeptide consensus motif, Ψ-K-x-D/E, where Ψ is a hydrophobic

residue, K is the lysine that directly conjugates to SUMO, x is any amino acid, and D/E is an

acidic residue. However, 50% of SUMO conjugation occurs on lysine residues that do not

adhere to this consensus sequence, demonstrating that this motif is not a requirement for

sumoylation (Chung et al. 2004). Prior to SUMO conjugation, the SUMOs are cleaved by

SUMO specific proteases (SENPs) to expose diglycine residues at its C-terminus, producing

the mature SUMOs. Subsequently, the SUMOs can be covalently conjugated to lysine

residues of substrates from a cascade of reactions directed by three enzymes – the activating

enzyme, SAE1/2 (E1), the conjugating enzyme Ubc9 (E2), and one of many ligases (E3).

SENPs also function in the deconjugation of the SUMOs and are largely responsible for the

rapid cycles of conjugating and deconjugating SUMOs from its substrates. Six SENP

proteases are found in mice and human, numbered SENP1, 2, 3, 5, 6, and 7. SENP1 is

predominantly responsible for deconjugating SUMO1, and SENPs 2, 3, and 5 are

predominantly responsible for deconjugating SUMO2/3 (Sharma et al. 2013; Gong & Yeh

2006). Deconjugation by the SENP family is a key process for regulating the steady-state

levels of a SUMO-modified substrate, which generally makes up less than 5% of the total

substrate (Willson 2009). Even so, the small amount of SUMO-modified substrates is enough

to exert large effects on cellular processes.

Role of SUMO in Embryogenesis

SUMO KO studies demonstrate that SUMO conjugation is essential for embryonic

development. Ubc9 KO results in early embryo death at the postimplantation stage,

indicating that the SUMO pathways are collectively important for embryonic development

(Nacerddine et al. 2005). SUMO1 KO mice develop normally, but mice deficient in SUMO2/3 are not viable (Zhang et al. 2008; Liangli Wang 2014). Moreover, SENP1 KO is embryonic lethal but can be rescued by genetically reducing SUMO1 levels, suggesting that maintaining proper levels of SUMO1-modified substrates is critical for embryo survival (Sharma et al. 2013). Collectively, the KO studies suggest that SUMO2/3 protein levels are important for embryos, while SUMO1 is dispensable and may even be harmful if SUMO1-modified substrate levels are upregulated.

While these knockout studies demonstrate the importance of global sumoylation for embryonic viability and development, the role of SUMO in embryos at the molecular and cellular level has been relatively unstudied. Some studies have explored the effects of SUMO conjugation on the pluripotency-associated transcription factors, Oct4, Sox2, and Nanog. Sumoylation of Oct4 at its two conserved SUMO sites leads to an increase in Oct4 stability, DNA binding, and transactivation (Wei et al. 2007), and elevated levels of SUMO-modified Oct4 and Sox2 decreases Nanog expression (Wu et al. 2012). These studies suggest that sumoylation of such embryonic proteins could have significant downstream effects in embryonic function, but this has yet to be shown.

Despite mounting evidence indicating the importance of sumoylation for fundamental cellular processes, very little is known about this modification in the context of embryonic cell function. In Chapter 5, we investigated the effects of SUMO overexpression in mammalian cell lines and showed that embryonic cells, but not differentiated cells, cannot overexpress SUMO1 at the protein level. We reported that elevated SUMO1 conjugation activity was highly unfavorable in embryonic cells and provided evidence for a post-transcriptional mechanism for tightly regulating SUMO1-modified substrate levels in

embryonic cells. These studies underscored the importance of regulating SUMO expression and conjugation activity for embryonic cell function.

Chapter 2 : Materials and Methods

Cell culture and RNA interference

F9, PCC4, E14, Rat2, 293T, HeLa, and NIH3T3 cells were cultured in DMEM media with 10% FBS, 100 U/mL penicillin, 0.05 mM streptomycin, and 2 mM L-glutamine. E14 cells were cultured with in DMEM media with 15% FBS, 20 mM HEPES, 0.1 mM nonessential amino acids, 0.1 mM 2-mercaptoethanol, 100 U/mL penicillin, 0.05 mM streptomycin, 2 mM L-glutamine, and leukemia inhibitory factor was added fresh at the time of culture. Plates were coated with 0.1% gelatin prior to plating embryonic cells. Cells treated with pharmacological agents in Chapter 3 were treated at the following concentrations: 20 μM chloroquine, 10 μM KU55933, 20 μM chloroquine and 10 μM KU55933, 0.1 μM rapamycin, or 200ng/ml Neocarzinostatin (NCS). Cells treated with pharmacological agents in Chapter 5 were treated at the following concentrations: 10 μM MG132 (EMP Millipore), 100 μM chloroquine (Sigma), 1 μM retinoic acid (Sigma), 10 μM - 100 mM hydrogen peroxide (H_2O_2) (Sigma), and 10 – 100 μM ginkolic acid (Abcam).

RNA interference

RNAi KD was performed with pLKO.1 shRNA (Sigma or GE Dharmacon). The target sequence used for mTrim28 KD, mYY1 KD, mUbc9 KD, and mSENP1 KD are listed below:

       mYY1 KD: CGACGGTTGTAATAAGAAGTT

       mTRIM28 KD: CCGCATGTTCAAACAGTTCAA

       Ubc9 KD 1: AGGCCAGCTATCACCATCAAA

       Ubc9 KD 2: TGTTCAAGCTACGGATGCTTT

       Ubc9 KD 3: AGGCCAGCTATCACCATCAAA

Ubc9 KD 4: TGGCACAATGAACCTGATGAA

Ubc9 KD 5: GCCTACACAATTTACTGCCAA

SENP1 KD 1: CGGGCCTTTGTAGATTTCCTA

SENP1 KD 2: CGCAAAGACATTCAGACTCTA

SENP1 KD 3: GCCATATTTCCGAAAGCGAAT

SENP1 KD 4: CCAGCCTATCGTCCAGATTAT

SENP1 KD 5: CGTAACGGTAACCAGGATGAA

Plasmid construction

Trim28, YY1, SUMO1, SUMO2, RanGAP1, and SENP1 cDNA were cloned into the pLVX-EF1a-IRES vector with the drug resistance markers, neomycin, hygromycin, or puromycin, or the ZsGreen1 reporter gene (Clontech). ZsGreen1 was cloned from the pLVX-CMV-ZsGreen1 vector (Clontech). Silent mutations at the YY1 and Trim28 shRNA target were created to prevent the knockdown of the exogenous proteins. SUMO and phosphorylation mutant Trim28 constructs were introduced using primers containing the desired mutations in overlapping PCR reactions or the GENEART Site-Directed Mutagenesis System (Thermo Fischer Scientific). Trim28, YY1, and HP1 were cloned into the pQE80L vector (Qiagen) for the generation of recombinant proteins.

Virus production

Viruses were produced in 293T cells. 3.5×10^6 293T cells were plated in a 10 cm plate. The following day, 293T cells were transfected with 8 µg of a retroviral transfer plasmid, 4 µg of a vector contain the *VSV-G* envelope gene, and 4 µg of a plasmid containing the *gag* and *pol* genes. MLV-GFP viruses were produced by co-transfection of pNCA-GFP, pCMV-intron,

and pMD.G in 293T cells, as previously described (Lim et al. 2002; Wolf & Goff 2007). Lentiviruses were produced by co-transfection of pLVX-EF1a-IRES or pLKO.1 vectors with pCMVΔr8.2 and pMD.G in 293T cells, as previously described (Schlesinger et al. 2013). Transfections were carried out using Polyethylenimine (PEI). Virions were harvested 48 hours and 72 hours after transfection, and cells were transduced with virus for 3 hours with 8 μg/ml of polybrene.

Cell colony formation assay and measuring viral titer

10^5 F9 cells were plated on 6-well plates. The following day, cells were infected with virus for three hours, and replaced with fresh media. 48 hours later, antibiotic was added to the cells, and surviving cells were allowed to form colonies over 2 weeks. When colonies were visible, cells were washed with D-PBS, and fixed with 100% chilled methanol for 10 minutes. Cells were died using Giemsa staining and colonies were counted. This method was also used to determine viral titers, but 293T or HeLa cells were used instead.

GFP expression from MLV vector

F9 cells were plated at $2x10^5$ cells per well of a 6-well dish and transduced with virus preparations containing an MLV vector expressing GFP (pNCA-GFP) at an MOI of 1. 48 hours after infection, cells were trypsinized, washed with D-PBS, and re-suspended in D-PBS supplemented with 2% FBS. Cells were examined for GFP expression by flow cytometry on the Guava flow cytometer (EMD Millipore) and analyzed with FlowJo software (TreeStar).

Recombinant protein expression and purification

BL21 cells (NEB) were transformed with pQE80L vectors expressing recombinant proteins.

Bacterial cultures were grown with ampicillin and 1 mM IPTG for 4 hours. Cells were

harvested and lysed with Buffer A (6 M GuHCl, 100 mM NaH_2PO_4, 10 mM Tris-Cl, 5 mM

B-mercaptoethanol, 10 mM imidazole, pH 8.0) for 15 minutes. Lysates were clarified by

centrifugation at 16.1 RCF for 1 h. The lysates were mixed with pre-washed Ni-NTA

Agarose beads (Qiagen) for 30 minutes and then loaded onto a centrifuge column (Pierce).

Beads were washed 4 times with buffer A and recombinant protein was eluted with elution

buffer (Buffer A with 100 mM imidazole). Recombinant protein was dialyzed into folding

buffer (Golebiowski et al 2011) or PBS using the Slide-a-Lyzer Dialysis Cassette (Peirce)

(Golebiowski et al. 2011).

F9 total cell lysate preparation

Cells were grown to confluent in 10-cm dishes and collected and washed in ice cold D-PBS

per immunoprecipitation reaction. Ice-cold 0.1% NP40 lysis buffer (0.1% NP40, 250 mM

NaCl, 20 mM Na_3PO_4, pH 7.0, 30 mM $Na_4P_2O_7$, 5 mM EDTA, 10 mM NaF) with 1x

complete protease inhibitor (Roche) was added to the pellet at 2x the cell pellet volume. Cells

were lysed on ice for 30 minutes and the lysates were clarified by centrifugation for 15

minutes at 14,000 RPM at 4°C.

Co-immunoprecipitation

Cell lysates were diluted to 400 µl total volume in lysis buffer, per IP reaction. 2% of total

lysate volume was saved for input lanes, and remaining lysate was incubated with 4 µg anti-

YY1 antibody (sc281, Santa Cruz Biotechnology) or rabbit control antibody (sc-2027, Santa

Cruz Biotechnology) for 16 hr in 4°C. Prewashed protein A/G dynabeads (Thermo Fischer Scientific) were added to the lysates and incubated for 1 hour in 4°C. For anti-myc IP experiments, lysates were incubated with 20 µl of anti-myc beads (Pierce) for 1 hour in 4°C. For *in vitro* co-IP experiments, recombinant proteins were incubated with antibody or beads in 400 µl of lysis buffer. 10 µl of prewashed anti-Flag beads (M8823, Sigma) were used for the IP of Flag-rHP1 protein. Beads were washed 3x with lysis buffer and bound proteins were eluted and analyzed by Western blot.

Lysate preparation and immunoprecipitation of sumoylated substrates

Flag-tagged SUMO proteins were co-expressed with HA-tagged Trim28 in 293T cells. Cell lysates were prepared in SDS lysis buffer (5% SDS, 30% glycerol, 0.15M Tris-HCl, pH 6.8), diluted 1:4 with 0.5% NP40/PBS, added to pre-washed anti-HA magnetic beads (Pierce), and incubated overnight at 4°C in rotation. Beads were washed with 0.5% NP40/PBS three times and bound proteins were removed with 1x SDS sample buffer and boiling for 5 minutes at 95°C. Co-IP of proteins was analyzed by Western blot.

DNA precipitation of binding proteins

50 ul of pre-washed streptavidin coated dynabeads (ThermoFisher) were incubated with 50 pmols of annealed biotinylated oligonucleotide in 200 µl binding buffer (1 M NaCl, 5 mM Tris, 1 mM EDTA) for 10 minutes. The same biotinylated oligonucleotides sequences used for EMSA reactions were used here. Beads were washed twice and resuspended with 200 µl EMSA buffer (10 mM Tris, 50 mM KCl, 2.5% glycerol, 5mM $MgCl_2$, and 0.05% NP40). 2 µg poly(dI:dC) (Sigma) and 2 µg rYY1 protein or F9 cell lysate was added to the beads and incubated for 30 minutes. Beads were washed 3 times with EMSA buffer (100mM KCl, 5%

glycerol, 10mM MgCl$_2$, 0.1% NP40, and 10mM Tris, pH 7.5) and the bound proteins were eluted and analyzed by SDS-Page.

F9 nuclear extract preparation

All procedures were performed on ice. Cells collected, washed with D-PBS, resususpended in Dignam buffer A (10 mM HEPES pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT Protease inhibitor cocktail (Roche)), and lysed with 15-20 strokes of a dounce homogenizer. Nuclei were pelleted by centrifugation at 4300 x g for 5 minutes, and resuspended in Buffer C (20 mM HEPES pH 7.9, 1.5 mM MgCl2, 25% v/v glycerol, 0.2 mM EDTA, 0.5 mM DTT, Protease inhibitor cocktail (Roche)). Nuclei were lysed with 15-20 strokes of a dounce homogenizer, and nuclear extracts were clarified by centrifugation.

EMSA

Nuclear extracts were prepared as previously described. Double-stranded DNA probes were end labeled with biotin and used for EMSA reactions with the LightShift Chemiluminescent EMSA Kit (cat. #20148; Thermo Scientific) according to the manufacturer's instructions. For the detection of supershifts, antibody was added 30 min prior to adding the probe. Protein complexes were analyzed by a 5% native polyacrylamide gels and transferred to 0.45 µM pore nitrocellulose membrane for 40 minutes at 380 mAmps in 4°C. Membranes were immediately crosslinked using the UV Crosslinker (Fischer Scientific) for 1 minute prior to blocking. The biotinylated oligonucleotides were purchased from IDT and their sequences are listed below:

> NCR: AGCTTAAGTAACGCCATTTTGCAAGGCA
>
> NCRm: AGCTTAAGTAATACGGCTATGCAAGGCA

PBS: GGGGGCTCGTCCGGGATCGGGAGACCCC

B2: GGGGGCTCGTCCGAGATCGGGAGACCCC

BS2 + PBS: GTCTTTCATTTGGGGGCTCGTCCGGGAT

BS2 + B2: GTCTTTCATTTGGGGGCTCGTCCGAGAT

BS2: CTACCCGTCAGCGGGGGTCTTTCATTTG

BS2 scram: CTACCCGTCAGCGGGGGTCTTTCGGCTA

ΔBS2 + PBS: GCGGGGGTCTTGGGGGCTCGTCCGGGAT

ΔBS2 + B2: GCGGGGGTCTTGGGGGCTCGTCCGAGAT

Precipitation of protein complexes by ammonium sulfate

By gradual addition of powdered $(NH_4)_2SO_4$, F9 nuclear extract was raised to 25%
saturation, incubated for 30 min on ice, and precipitated proteins were removed by
centrifugation at 13,000 RPM for 10 minutes and resuspended in Dignam D. The remaining
extract sample was then raised to 40% $(NH_4)_2SO_4$ saturation, incubated for 30 min on ice,
and the precipitated proteins was centrifuged and dissolved in Dignam D. This was repeated
for the collection of precipitated proteins in nuclear extract samples brought to 80%
$(NH_4)_2SO_4$ saturation. $(NH_4)_2SO_4$ was removed from the protein samples by desalting in an
Amicon 100 kD column. Fractions were monitored for binding activity to the BS2 + PBS
probe by EMSA.

Sucrose gradient and immunoprecipitation

A 10-40% sucrose gradient was prepared using BioComp's Gradient Master in EMSA buffer
(see EMSA protocol). F9 nuclear extract was slowly loaded onto the sucrose gradient and
centrifuged in a Beckman SW55 rotor for 10 hours at 40,000 RPM and 4°C. 12 fractions

were collected, and the proteins in the fractions were examined for binding to the BS2 + PBS probe by EMSA. Fractions found to contain the RBS or large RBS complex were pooled. Pooled samples were divided into two and anti-Trim28 or control antibody was added. Samples and antibody were incubated overnight in rotation at 4°C. Prewashed protein A beads were added and incubated for 1 hour in rotation at 4°C. Bound proteins were washed 3 times with EMSA buffer (see EMSA protocol), removed from beads, and resolved on a 4-20% gradient gel.

Protein identification by mass spectrometry

Gel slices containing the proteins samples were excised, proteins were subjected to tryptic digest followed by peptide identification by lc-ms/ms using a hybrid high-resolution quadrupole time-of-flight electrospray mass spectrometer. Results were analyzed using the MASCOT database search tool (Matrix Science).

PCR

PCR reactions were conducted using the KOD Hot Start DNA Polymerase (EMD Millipore) kit according to manufacturer's instructions. For qRT-PCR reactions, RNA was extracted from cells using the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. cDNA was synthesized using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufacturer's instructions. For qPCR reactions, total DNA was isolated using DNeasy kit (Qiagen) according to the manufacturer's instructions. DNA isolates or cDNA was combined with FastStart universal SYBR Green Master (Roche) containing 300 nM of indicated primers. qPCR was performed in 96-well plates using LightCycler96 (Roche) with the following reaction conditions: 10 min at 95°C, followed by

45 cycles of 30 s at 95°C, 30 s at 60°C and 30 s at 72°C. Primer sequences used for PCR,

qPCR, and qRT-PCR reactions are listed below:

PCR primers for amplifying transgene inserts in the pLVX-EF1 vector

       pLVX-F: TCAAGCCTCAGACAGTGGTTC

       pLVX-R: ACCCCTAGGAATGCTCGTCAAGAA

q-PCR primers

       SUMO1-F: ACAAGGTTACTAGTTCTAGAATGTCTGA

       SUMO1-R: GTATCTCACTGCTATCCTGTCCAATAA

       hGAPDH F: ACATCATCCCTGCCTCTAC

       hGAPDH R: TCAAAGGTGGAGGAGTGG

       mCyclophilin A-F: GCAGGTCCATCTACGGAGAGAAA

       mCyclophilin A-R: GTCAACAGATCCCATTCACTGTTTCTTA

       Puro-F: GCCGCGCAGCAACAGAT

       Puro-R: CGCTCGTAGAAGGGGAGGTT

qRT-PCR primers:

       mSUMO1-F: ATTGGACAGGATAGCAGTGAGA

       mSUMO1-R: TCCCAGTTCTTTCGGAGTATGA

       mSUMO2-F: TGGAGTAAAGTAGCAGGCTCCCTTT

       mSUMO2-R: ACTAATGAAAGCCTATTGTGAAC

       mGAPDH-F: AACGACCCCTTCATTGAC

       mGAPDH-R: TCCACGACATACTCAGCAC

       mSENP1-F: ATTATCACTCAGATAACCCTTCCTCAGA

       mSENP1-R: ACTTTGACCAAAGGTTCTTACGTCA

mUbc9-F: AAGGGGACTCCATGGGAAGG

mUbc9-R:CTCCAGGATGGACAGGCACA

Detection of SSEA-1 expression by flow cytometry

$2x10^4$ cells were collected per sample for SSEA-1 staining. Cells were spun down and

resuspended in 200µl of 10% FBS/D-MEM medium and 5 ul of alexa fluor 488 anti-SSEA-1

antibody (125609, Biolegend). Cells were incubated on ice and protected from light for 30

minutes. After incubation, cells were washed once with cold D-PBS and resuspended in 200

µl of D-PBS. Cells were analyzed by flow cytometry on the Guava flow cytometer (EMD

Millipore) and analysis of data was done using the FlowJo software (TreeStar). The

percentage of cells expressing GFP is reported relative to the percent of NIH3T3 cells

expressing GFP infected in parallel.

Fluorescence-activated cell sorting (FACS)

F9 cells were transduced with the SUMO1-ZsGreen or SUMO1GG-ZsGreen vectors. Cells

demonstrating greater ZsGreen expression intensity relative to untreated F9 cells were

selected by FACS (FACSAria Cell Sorter, BD Biosciences). Data were acquired on an

automated cell analyser (LSR II; BD Biosciences) and analysed with FlowJo software

(Treestar). Collected cells were allowed to recover and multiply over 8-10 days before

repeating the sorting procedure. Cells were selected over three rounds of FACS.

Antibodies

Antibodies used for Western blots were as follows: anti-Trim28 20C1 (ab22553, Abcam),

anti-YY1 C-20 (sc281, Santa Cruz Biotechnology), anti-HA.11 (901515, BioLegend), anti-

myc 71D10 (2278, Cell Signaling Technology), anti-myc 9E10 (sc-40, Santa Cruz

Biotechnology), anti-Flag M2 (F3165, Sigma-Aldrich), pS824-Trim28 (ab70369), anti-SUMO antibody (sc-9060, Santa Cruz Biotechnology), anti-Oct3/4 (H-134, Santa Cruz Biotechnology), and anti-β-actin (A1978, Sigma). Antibodies used for co-IP experiments are as follows: anti-YY1 C-20 (sc281, Santa Cruz Biotechnology), anti-Trim28 antibody (ab10484, Abcam), and rabbit control antibody (sc-2027, Santa Cruz Biotechnology). Antibodies used for EMSA shifts were as follows: anti-YY1 C-20 (sc281, Santa Cruz Biotechnology), anti-Trim28 20C1 (ab22553, Abcam), and rabbit control antibody (sc-2027, Santa Cruz Biotechnology). The antibody used for flow cytometry was AF488 anti-SSEA-1 antibody (125609, Biolegend).

Chapter 3 : Characterization of the Interaction Between Trim28 and YY1 in Transcriptional Silencing of Moloney Murine Leukemia Provirus

In a previous study, we demonstrate that YY1 mediates the silencing of M-MLV by recruiting Trim28 to the negative coding region (NCR) in the 5'LTR of the provirus (Schlesinger et al. 2013). Co-IP experiments revealed that YY1 interacts with Trim28, although YY1 does not contain a domain predicted to interact with Trim28. This interaction was detected in the F9 embryonic cell line but not in the NIH3T3 differentiated cell line. The results from this study provided the foundation for the work we report here, in which we further examined the nature of the YY1-Trim28 biochemical interaction and investigated possible mechanisms for regulating the YY1-Trim28 interaction and silencing activity.

Recombinant YY1 and TRIM28 do not interact in vitro

To test for a direct interaction between YY1 and Trim28, we conducted *in vitro* co-immunoprecipitation (co-IP) experiments using recombinant proteins expressed in *E. coli*. Trim28 has been demonstrated to interact directly with HP1, so this interaction was used as a positive control (B Le Douarin 1996). We expressed recombinant His$_6$-tagged YY1, His$_6$-tagged Trim28, or His$_6$-Flag-tagged HP1 in bacteria and purified the recombinant proteins by nickel affinity chromatography. rYY1 was incubated with anti-YY1 antibody in 0.1% NP40 lysis buffer, followed by the addition of protein A/G beads and rTrim28. In parallel, rHP1 was incubated with anti-Flag beads, followed by the addition of rTrim28. Bound proteins were removed from the beads and analyzed by Western blot (Figure 3-1). We did not detect
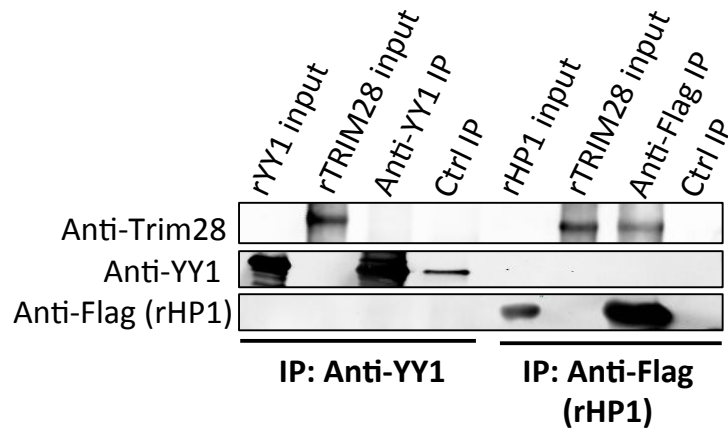
Figure 3-1 *In vitro* **co-IP monitoring interaction between rYY1 and rTrim28.**
rYY1 was immunoprecipitated using anti-YY1 antibody, and rabbit control antibody was
used in the control IP. Protein and antibody complexes were incubated with protein A/G
beads in 0.1% NP40 cell lysis buffer (see methods) and bound proteins were removed from
beads and analyzed by Western blot. An *in vitro* co-IP with rHP1 and rTrim28 was
conducted alongside as a positive control. Flag-rHP1 was immunoprecipitated using anti-flag
beads, and the corresponding control IP mixed rTrim28 with anti-flag beads in the absence of
rHP1. 10% of total recombinant protein used for each IP was loaded in the input lanes.

rTrim28 with the IP of rYY1 using this approach. Analysis of the control antibody IP showed some traces of rYY1, probably due to non-specific binding, but significantly less than the levels of rYY1 detected in the YY1 IP lane. In contrast, we detected rTrim28 with the IP of rHP1. We performed similar experiments with an alternative binding buffer (EMSA buffer, see methods) or longer incubation periods, but the various conditions we attempted did not result in a detectable interaction between rYY1 and rTrim28. We also generated GST-tagged rTrim28 protein, immunoprecipitated GST-Trim28 with glutathione beads, and incubated bound rTrim28 protein with rYY1. Proteins were removed from the beads and analyzed by Western blot probed with anti-YY1 antibody. We did not detect rYY1 with the IP of rTrim28 with this approach either (data not shown). The inability for rYY1 and rTrim28 to co-immunoprecipitate *in vitro* suggests that this interaction may occur indirectly.

One possible explanation for the failure to detect an interaction is that the YY1 antibody could have blocked rTrim28 interaction with rYY1 because of the order of addition of the components in which we conducted our co-IP experiments. To test for this possibility, we conducted the *in vitro* co-IP with a different order of protein addition. rYY1 and rTrim28 were first incubated together at high concentrations, and rYY1 was subsequently immunoprecipitated using anti-YY1 antibody and protein A/G beads. Bound proteins were removed from the beads and analyzed by Western blot probing for Trim28. We did not detect rTrim28 in the rYY1 IP using this approach either (data not shown), suggesting that rYY1 and rTrim28 most likely interact indirectly in the cell.

Trim28 RBCC domain is necessary for its interaction with YY1

As we could not detect an interaction between recombinant YY1 and Trim28 *in vitro,* we continued further examination of the YY1-Trim28 interaction detected by immunoprecipitation from mammalian cell lysates. To determine the domains on Trim28 responsible for the YY1-Trim28 interaction, we conducted co-IP experiments with HA-tagged Trim28 mutants in the embryonic carcinoma F9 cell line, in which M-MLV silencing occurs. As embryonic cells are difficult to transfect and can silence a variety of promoters, all knockdown and overexpression studies were conducted by lentiviral transduction mediating expression of Trim28 and a drug resistance gene, driven by the EF1α promoter. Cells expressing exogenous Trim28 were selected for stable expression of the drug resistance gene, followed by knock down (KD) of endogenous Trim28 by shRNA. We found that endogenous Trim28 expression returned after prolonged culture, likely because Trim28 is critical for long-term embryonic cell function and survival (Cammas et al. 2000). Thus, Trim28 was knocked down immediately prior to using the cells in our various assays, and KD levels were confirmed by Western blot.

To determine which Trim28 domain(s) were necessary for its interaction with YY1, we expressed mutant forms of Trim28 with deletions of its major domains, generating Trim28$^{1\text{-}628}$ (with deletion of the PHD/Bromodomain), Trim28$^{\Delta441\text{-}522}$ (with deletion of the chromoshadow domain, denoted "PxVxL" for its pentapeptide sequence), and Trim28$^{388\text{-}834}$ (with deletion of the RBCC domain) (Figure 3-2A). An HA-Trim28^{WT} cell line was also produced as a positive control. Cell lysates were prepared, and a fraction of the total lysate was saved for the input lane of SDS PAGE gels, while the remaining lysate was incubated with anti-YY1 antibody or control antibody. Antibodies and their interacting proteins were
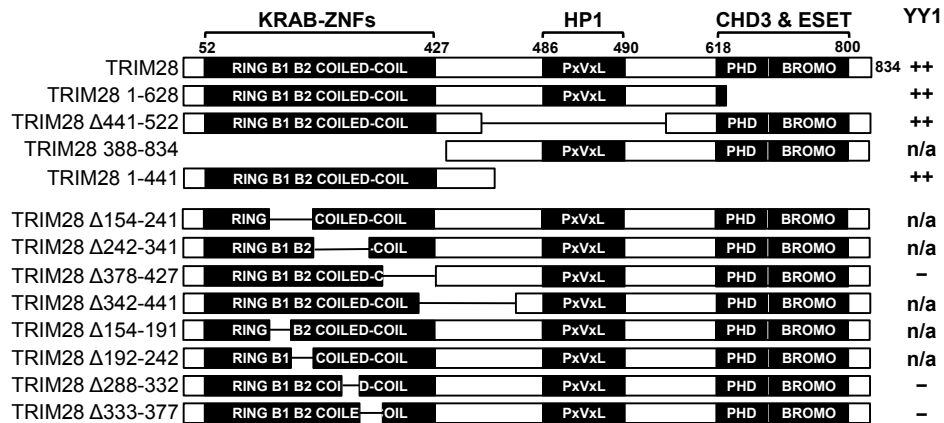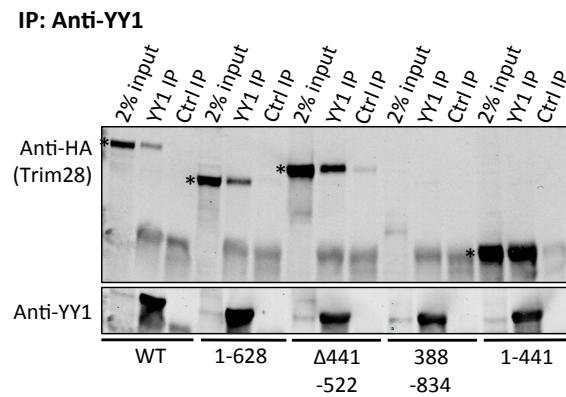
Figure 3-2 Co-IP of YY1 with Trim28 mutants containing deletions of major domains
(A) Schematic of the Trim28 wild-type and mutant proteins. All mutants were HA-tagged at the N-terminus. Domains known to interact with other repressor partners are noted above the domain. The Trim28 mutants that interact with endogenous YY1 are indicated in the right column. A strong interaction is noted by "++", no interaction is indicated by "-", and mutants that displayed poor expression are noted by n/a. (B) HA-Trim28 constructs were overexpressed in F9 cells, YY1 was immunoprecipitated with an anti-YY1 antibody, and bound proteins were examined by Western blot probed with anti-HA antibody. The bands corresponding to the Trim28 mutants are noted in the input lanes by an asterisk.

immunoprecipitated by protein A/G beads, displayed on SDS PAGE and transferred to Western blots, and probed for HA-Trim28 (Figure 3-2B). All mutant constructs expressed well except for Trim28$^{388-835}$, the construct missing the RBCC domain. Trim28^{WT}, Trim28$^{1-628}$, and Trim28$^{\Delta441-522}$ were found to interact with endogenous YY1, indicating that the PHD/Bromodomain and chromoshadow domain are not necessary for YY1 interaction. We did not detect HA-Trim28 or YY1 in any of the IPs with the control antibody, demonstrating the specificity of the anti-YY1 antibody. To confirm the importance of the RBCC domain, we expressed a Trim28 mutant containing only the RBCC domain (Trim28$^{1-441}$) (Figure 3-2A). Cell lysates were prepared, YY1 was immunoprecipitated, and bound proteins were examined by Western blot probing for HA-Trim28. The Trim28$^{1-441}$ mutant was detected with the IP of YY1 (Figure 3-2B), demonstrating that the RBCC domain is sufficient for Trim28 interaction with YY1.

We attempted to narrow down the Trim28 domains necessary for YY1 interaction further by making smaller deletions in the RBCC domain. We expressed Trim28 constructs with 100 amino acid region deletions in the RBCC (Trim28$^{\Delta154-241}$, Trim28$^{\Delta242-341}$, Trim28$^{\Delta378-427}$, Trim28$^{\Delta342-441}$) (Figure 3-2A). Lysates were prepared, YY1 was immunoprecipitated, and bound proteins were examined for mutant Trim28 interaction with YY1 by Western blot. Of these mutants, only Trim28$^{\Delta378-427}$ expressed well, as indicated in the 2% input lanes for each mutant cell line. We did not detect Trim28$^{\Delta378-427}$ interaction with YY1, in contrast to Trim28^{WT}, which expressed well and interacted with YY1 (Figure 3-3A). To determine if smaller deletions in the RBCC could improve Trim28 expression, we expressed Trim28 constructs with 50 amino acids region deletions (Trim28$^{\Delta154-191}$, Trim28$^{\Delta192-242}$, Trim28$^{\Delta288-332}$, Trim28$^{\Delta333-377}$) (Figure 3-2A). YY1 was immunoprecipitated

A
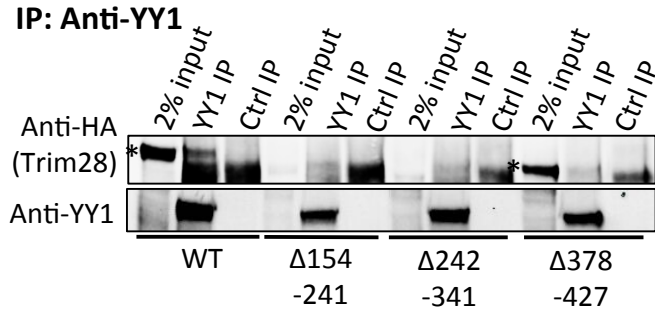
B
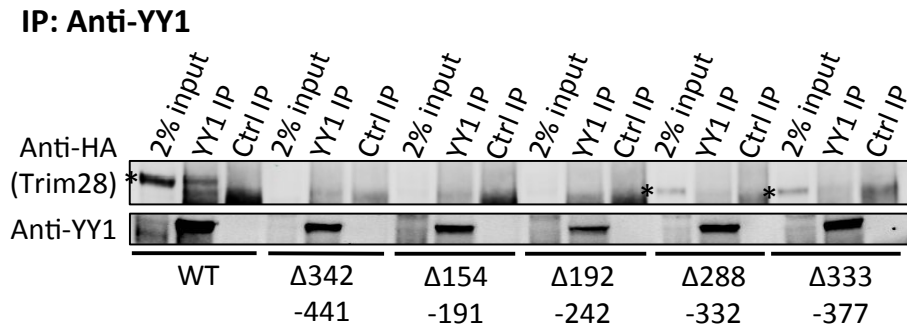
Figure 3-3 Co-IP of YY1 and Trim28 mutants containing deletions in the RBCC domain
(A) Trim28 mutants containing 100 amino acid deletions in the RBCC domains were expressed in F9 cells. Lysates were prepared, YY1 was immunoprecipitated with an anti-YY1 antibody, and bound proteins were examined by Western blot probed with anti-HA antibody. The bands corresponding to the Trim28 mutants are noted to the left of the input lanes by an asterisk. (B) Trim28 mutants containing 50 amino acid deletions in the RBCC domain (except Δ342-441, which has a 100 amino acid deleted region) were expressed in F9 cells and examined for YY1 interaction with the Trim28 in the same fashion as conducted for the previous Trim28 mutants. The bands corresponding to the Trim28 mutants are noted in to the left of the input lanes by an asterisk.

from total cell lysates, and bound proteins were examined for mutant Trim28 interaction with

YY1 by Western blot. These mutants also showed low or no expression, as indicated in the

input lanes for each mutant Trim28 cell line (Figure 3-3B). These experiments suggest that

an intact RBCC domain is important for Trim28 expression and that deletions in this region

may be disruptive to its structure. Our results demonstrate that the smallest region of Trim28

found to interact with YY1 was the RBCC domain.

YY1 acidic region 1 and zinc fingers are necessary for Trim28 interaction

To determine the domains on YY1 necessary for interaction with Trim28, a panel of

myc-tagged YY1 mutants were expressed in F9 cells and tested for their interaction with

endogenous Trim28. The major domains of YY1, starting from the N-terminus, are the acidic

region 1 (AR1), acidic region 2 (AR2), GA-rich region (GA), and the four zinc fingers

(ZNFs). We were unable to produce a construct with a deleted AR1 region, but constructs

with deletions of each of the other regions were made: $YY1^{\Delta54-144}$ (with deletion of the AR2),

$YY1^{\Delta151-209}$ (with deletion of the GA), $YY1^{\Delta177-273}$ (with deletion of the spacer), and $YY1^{1-295}$

(with deletion of the ZNFs) (Figure 3-4A). The empty vector was used as a negative control,

and a YY1 wild-type construct ($YY1^{WT}$) was produced and used as a positive control. F9

cells were transduced with the empty vector, or the $YY1^{WT}$ or mutant YY1 vectors and were

selected for stable expression of the drug resistant gene. Lysates were prepared and

exogenous YY1 was immunoprecipitated using anti-myc beads. Bound proteins were

removed from the beads and analyzed by Western blot probed with an anti-Trim28 antibody

(Figure 3-4B). IP of myc-YY1, in the $YY1^{WT}$ cell lysate, demonstrated Trim28 interacted

with exogenous YY1. Trim28 was undetectable with the IP of myc in the empty vector cell

56

A

B

IP: Anti-Myc

57
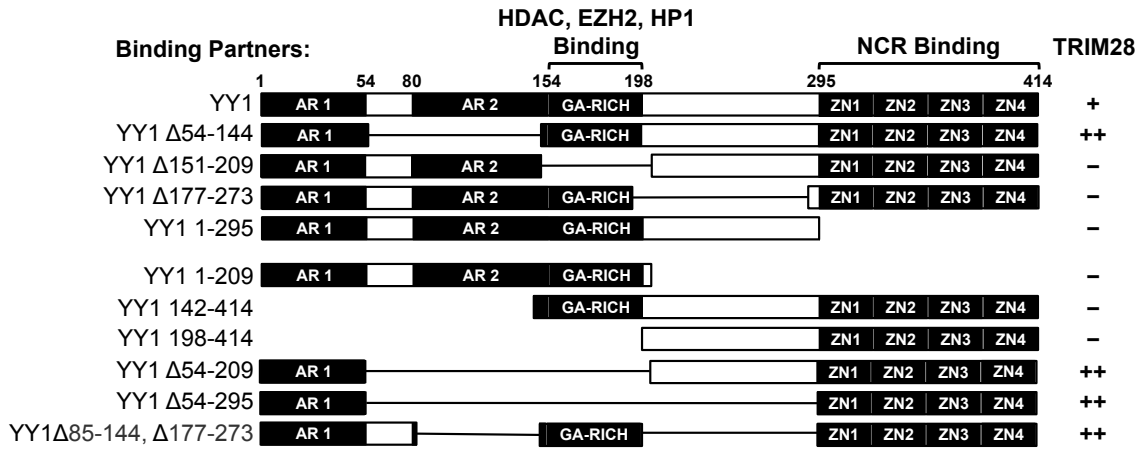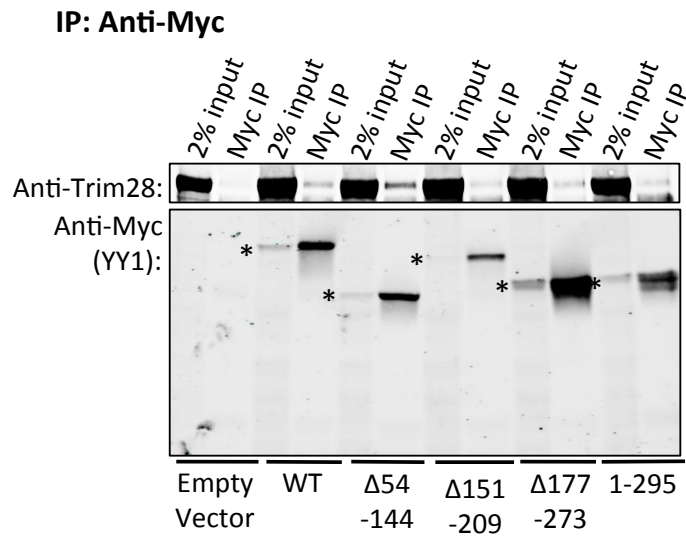
C
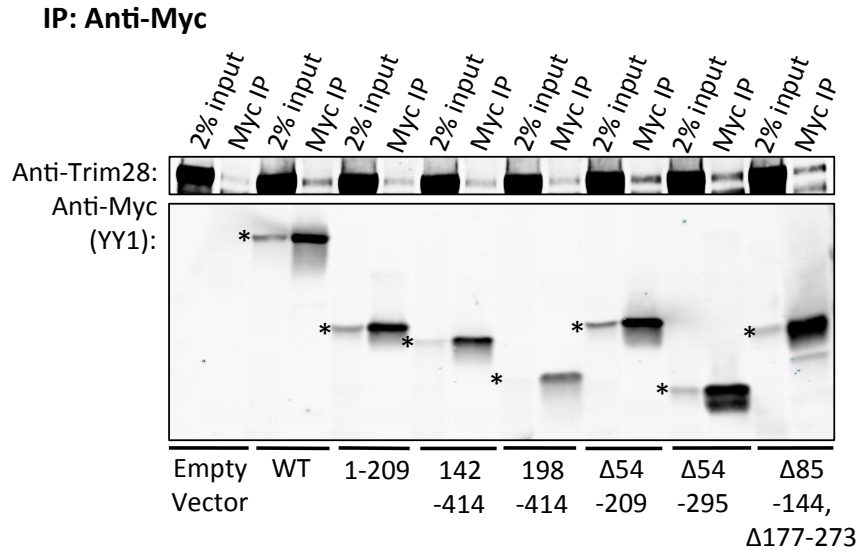
Figure 3-4 Co-IP of Trim28 and YY1 mutants
(A) Schematic of the YY1 wild-type and mutant proteins. All mutants were myc-tagged at the N-terminus. Domains known to interact with other repressor partners are noted above the domain. The YY1 mutants found to interact with endogenous Trim28 are indicated in the right column. A strong interaction is noted by "++", a weak interaction is noted by "+", and no interaction is indicated by "-". (B) Myc-YY1 constructs were overexpressed in F9 cells, and cell lysates were prepared for co-IP experiments. Myc-YY1 proteins were immunoprecipitated with anti-myc beads, and bound proteins were examined by Western blot probed with an anti-Trim28 antibody. The bands corresponding to the YY1 mutants are noted to the left of the input lanes by an asterisk. Mutants analyzed on this blot include the four mutants in which the AR2, GA, spacer, and ZNFs regions were independently deleted. (C) Mutants analyzed on this blot include the mutants in which two or more regions were deleted.

lysates. IPs of the myc-YY1 mutants showed that Trim28 interacted with YY1 in the absence of the AR2 domain (YY1$^{\Delta 54\text{-}144}$), indicating that this domain is not necessary for the interaction. However, Trim28 proteins were undetectable in the IPs of myc-YY1$^{\Delta 151\text{-}209}$, myc-YY1$^{\Delta 177\text{-}273}$, and myc-YY1$^{1\text{-}295}$, demonstrating that these mutants do not interact with Trim28. We constructed additional mutants containing only the N-terminal domains (YY1$^{1\text{-}209}$), only the C-terminal domains (YY1$^{142\text{-}414}$ and YY1$^{198\text{-}414}$), or both terminal domains (YY1$^{\Delta 54\text{-}209}$, YY1$^{\Delta 54\text{-}295}$, and YY1$^{\Delta 85\text{-}144,\ \Delta 177\text{-}273}$) of YY1 (Figure 3-4A). F9 cells were transduced with these constructs and selected for stable expression of the drug resistance gene. Cell lysates were prepared and myc-tagged YY1 mutants were immunoprecipitated and examined for their interaction with endogenous Trim28 by Western blot (Figure 3-4C). As shown in the previous Western blot, Trim28 was undetectable following myc immunoprecipitation in the empty vector cell lysate but was detected in the YY1^{WT} cell lysate. We did not detect Trim28 in the IPs of mutants containing only the N-terminal domains (YY1$^{1\text{-}209}$) or only the C-terminal domains (YY1$^{142\text{-}414}$, YY1$^{198\text{-}414}$), indicating that these mutants did not interact with endogenous Trim28. On the other hand, Trim28 interacted with the YY1 mutants containing both terminal domains (YY1$^{\Delta 54\text{-}209}$, YY1$^{\Delta 54\text{-}295}$, and YY1$^{\Delta 85\text{-}144,\ \Delta 177\text{-}273}$), revealing that both terminal domains of YY1 are necessary and sufficient for interaction with Trim28. YY1$^{\Delta 54\text{-}295}$ was the mutant with the fewest domains that still successfully interacted with Trim28. This demonstrated that the AR2, GA, and spacer are dispensable, but the combination of the AR1 and ZNFs are necessary for YY1 interaction with Trim28.

An additional observation to note is that the interaction between ectopic YY1^{WT} and Trim28 was weaker relative to the interaction we detected between endogenous YY1 and

Trim28. One possibility is that the myc-tag at the N-terminus interfered with $YY1^{WT}$ interaction with Trim28. Moving the myc-tag to the C-terminus of $YY1^{WT}$ did not improve its interaction with Trim28, making this possibility less likely (data not shown). Another possibility is that there is a limiting level of endogenous Trim28 available to interact with exogenous YY1, because all available Trim28 molecules are already interacting with endogenous binding partners. To test this possibility, we expressed exogenous $YY1^{WT}$ with the KD of endogenous YY1 or with the overexpression of Trim28, but these conditions also did not improve the interaction. It is also possible that expression levels of exogenous $YY1^{WT}$ are lower than that of endogenous YY1 and that the levels of Trim28 detected with the IPs of the respective YY1 proteins are actually proportional to its expression levels. We are currently testing for this possibility by quantifying and comparing the ratios of Trim28 to YY1 protein detected in the IPs of YY1 and by increasing the expression of ectopic $YY1^{WT}$.

Residue K779 of Trim28 is necessary for M-MLV silencing

SUMO modification of Trim28 was reported to be necessary for several Trim28 interactions (Goodarzi et al. 2011; Ivanov et al. 2007), and it is possible that sumoylation of Trim28 is important for the YY1-Trim28 interaction as well. To test this, we designed Trim28 constructs with mutations at the six lysine residues reported to be SUMO-modified (Figure 3-5A). We mutated the lysine residues to arginine to eliminate the SUMO conjugation site while retaining the same positive charge. Each site was mutated independently ($Trim28^{K554R}$, $Trim28^{K575R}$, $Trim28^{K676R}$, $Trim28^{K750R}$, $Trim28^{K779R}$, and $Trim28^{K804R}$), or all six collectively ($Trim28^{6KR}$). In addition to the seven Trim28 mutants, $Trim28^{WT}$ and an empty vector were included as controls. As previously done, F9 cells were

A

B

TRIM28 KD

C

IP: Anti-YY1

F9   EV   WT   K554R K676R K705R K804R

TRIM28 KD

IP: Anti-YY1

WT   EV   K575R K779R 6KR
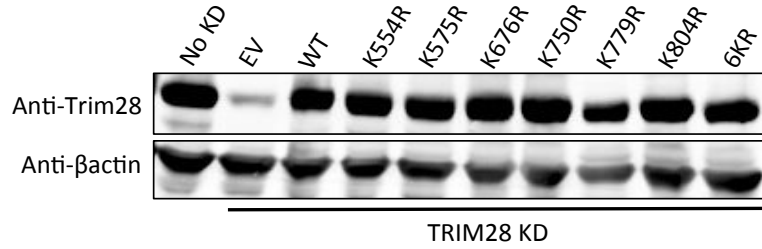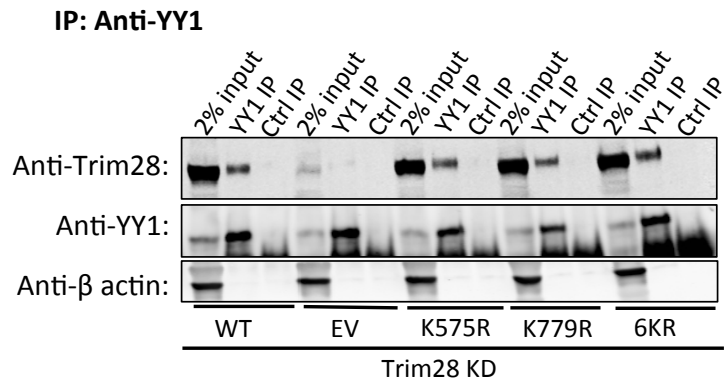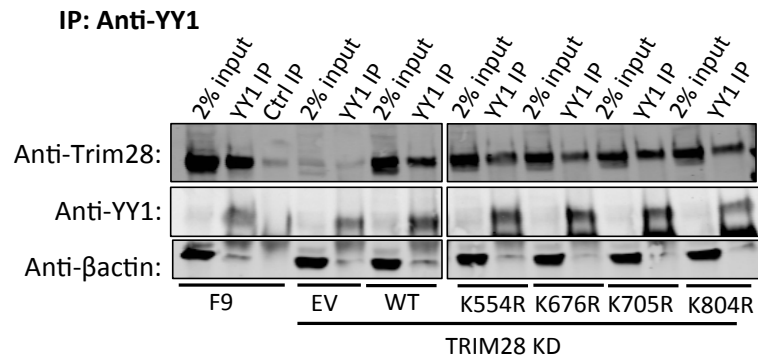
Trim28 KD

D
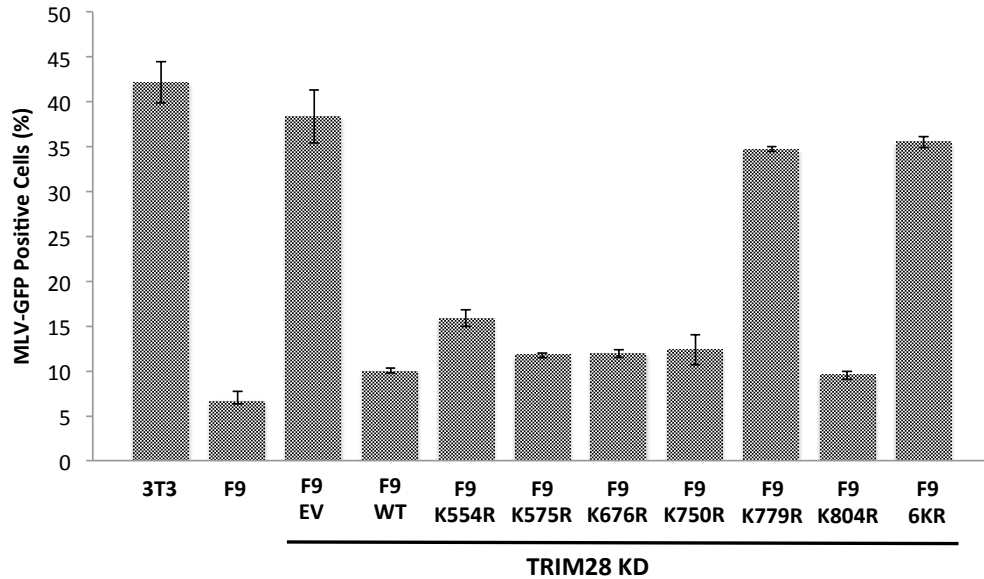
Figure 3-5 Mutations of major sumoylation sites on Trim28
(A) Schematic of Trim28 protein with the lysine residues that are sumoylated indicated. (B) Mutant Trim28 was expressed and endogenous Trim28 was knocked down in F9 cells, and cell lysates were prepared and analyzed by Western blot. The "No KD" lane shows basal Trim28 levels in F9 cells. The empty vector (EV) lane shows the level of endogenous Trim28 expression in all KD lines, and the remaining lanes show expression levels of exogenous Trim28. (C) Western blots monitoring interaction of mutant Trim28 with YY1 by co-IP. YY1 was pulled-down with anti-YY1 antibody and the western blot was probed for Trim28. Co-IP results for four mutants are show in the top panel and for the 3 remaining mutants on the bottom panel. (D) NIH3T3, untreated F9 cells, and Trim28 mutant lines were infected with the M-MLV-GFP virus, and GFP expression was measured by flow cytometry. Uninfected cells were used to set our gates for counting "GFP-positive" expressing cells. Cells infected with the M-MLV-GFP vector were counted as GFP-positive if GFP expression was greater than that detected in the uninfected cells. NIH3T3 represent maximal M-MLV expression levels, and F9 cells represent maximal repression of M-MLV expression. We report the percentage of GFP-positive cells detected. Values represent the mean percentage of three biological replicates, and error bars represent the standard error mean.

transduced with the empty vector, or a Trim28^{WT} or Trim28 mutant vector and selected for

stable expression of the drug resistance gene. Subsequently, endogenous Trim28 was

knocked down by shRNA. Lysates were prepared and analyzed by Western blot for Trim28

expression (Figure 3-5B). We detected Trim28 in the untreated F9 cells (No KD), reflecting

the basal levels of endogenous Trim28, and we detected only very low levels of Trim28 in

the KD cells expressing the Trim28 shRNA and lacking the Trim28 cDNA (the empty vector

line). This indicated that endogenous Trim28 was knocked down well in our Trim28 KD

lines. We also detected high levels of Trim28 in the Trim28 mutant-expressing cell lines,

demonstrating that all our mutants expressed well. YY1 was immunoprecipitated with an

anti-YY1 antibody from cell lysates, and bound proteins were examined for the presence of

exogenous Trim28 by Western blot (Figure 3-5C). Control co-IP reactions demonstrate

background levels of Trim28 detected when lysates were incubated with non-specific

antibodies. Trim28 was detected with the IP of YY1 in untreated F9 cell lysate (F9) but not

with the IP of YY1 in the Trim28 KD cell lysate (EV), demonstrating that endogenous

Trim28 had been knocked down to levels such that any interaction with YY1 was not

detectable. All mutant Trim28 proteins, including Trim28^{6KR}, were expressed at comparable

levels and interacted with YY1 to similar degrees as Trim28^{WT}, indicating that the six lysine

residues examined on Trim28 are not necessary for the YY1-Trim28 interaction.

It has been reported that sumoylation modifications on Trim28 are required for its

transcriptional repressive activity (Y.-K. Lee et al. 2007; Goodarzi et al. 2011; Mascle et al.

2007). We asked if the sites of sumoylation were also required for silencing of M-MLV

transcription in embryonic cell lines. To investigate this, we infected the Trim28 mutant lines

with an M-MLV-GFP virus, in which a GFP reporter gene driven by the M-MLV LTR is

packaged into a virus. Cells were infected at a constant multiplicity of infection, and M-MLV expression was measured two days after infection by flow cytometry. Cells that demonstrated greater GFP expression levels relative to uninfected cells were counted as GFP-positive cells (Figure 3-5D). Knockdown of endogenous Trim28 was confirmed at the time of M-MLV-GFP infection by Western blot, such that any M-MLV silencing detected could be attributed to the activity of the exogenous Trim28. M-MLV-GFP expression level measured in the differentiated line, NIH3T3, represented the maximal M-MLV expression level and that in the F9 cells represented maximal M-MLV repression level. Cells with Trim28 knocked down and carrying the empty vector construct (EV) contained roughly the same percentage of GFP-positive cells as that measured in NIH3T3 cells, underscoring the importance of Trim28 for M-MLV silencing. The Trim28^{WT} expressing cells showed very strong repression of M-MLV-GFP, demonstrating that exogenous Trim28 rescues the repressive phenotype very effectively. Cells expressing five of the Trim28 mutants –Trim28^{K55R}, Trim28^{K575R}, Trim28^{K676R}, Trim28^{K750R}, Trim28^{K804R} – displayed similar percentages of GFP-positive cells to that measured in Trim28^{WT} cells, indicating that these sites were not critical for M-MLV silencing. However, the repressive activity of the Trim28^{K779R} and Trim28^{6KR} mutants were significantly compromised, as indicated by the high percentage of GFP-positive cells detected in cells expressing these two mutants. The fact that the Trim28^{K779R} mutant showed the same loss of function as the Trim28^{6KR} mutant indicates that the K779R mutation makes up for most, if not all, of the derepression detected in the 6KR mutant. Also, the similar percentages of GFP-positive cells detected in the Trim28^{K779R} and empty vector expressing cells indicates that mutation of this residue results in close to no Trim28 repressive function on M-MLV. The importance of K779 for Trim28 repressive activity in this setting is

consistent with what has been reported in previous studies of Trim28 silencing in other settings. In contrast, our results indicating that K554 and K804 are not critical for Trim28 silencing activity of M-MLV is distinct from what has been reported for Trim28-mediated transcriptional silencing of other gene targets (Ivanov et al. 2007; Y.-K. Lee et al. 2007; Mascle et al. 2007).

Trim28 is modified by SUMO2 at the same sites as those used by SUMO1

The vast majority of studies on the effects of sumoylation modifications on Trim28 have involved the SUMO1 protein, but it is unknown if SUMO1, SUMO2/3, or all isoforms can mediate Trim28 repressive activity. A recent study implicating SUMO2 involvement in Trim28-mediated viral silencing prompted us to investigate whether SUMO2 could be conjugated to Trim28 and if it conjugates the same lysine residues as those conjugated by SUMO1 (B. X. Yang et al. 2015). To confirm that the six reported lysine residues on Trim28 were the only available sites for SUMO1 conjugation, we co-expressed HA-tagged Trim28^{WT} or Trim28^{6KR} with Flag-tagged SUMO1 or SUMO1^{GG}, a mutant form of SUMO1 that cannot be conjugated to any substrate, in 293T cells. Cell lysates were prepared using a lysis buffer that prevents SUMO deconjugation, HA-Trim28 was immunoprecipitated with anti-HA beads, and precipitated proteins were examined for SUMO1 modifications by Western blot probing for Flag-SUMO1. We detected a ladder of SUMO1-modified Trim28 products in the Trim28^{WT} expressing cells and not in the Trim28^{6KR} expressing cells, demonstrating that Trim28 is modified by SUMO1 at one or more of the 6 lysine residues that were mutated in the 6KR mutant (Figure 3-6A). As expected, SUMO1^{GG} was not detected with the IP of

A
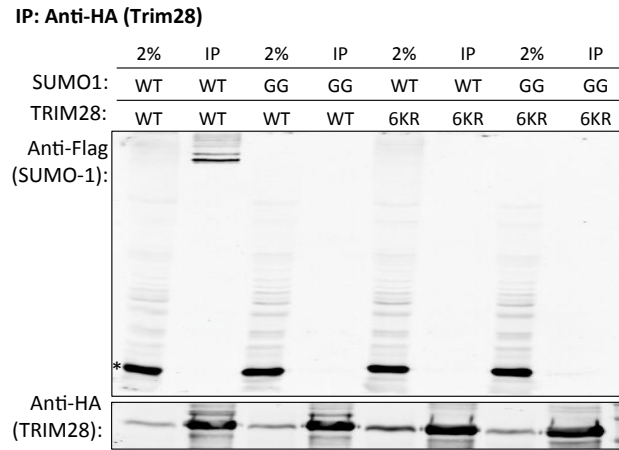
B
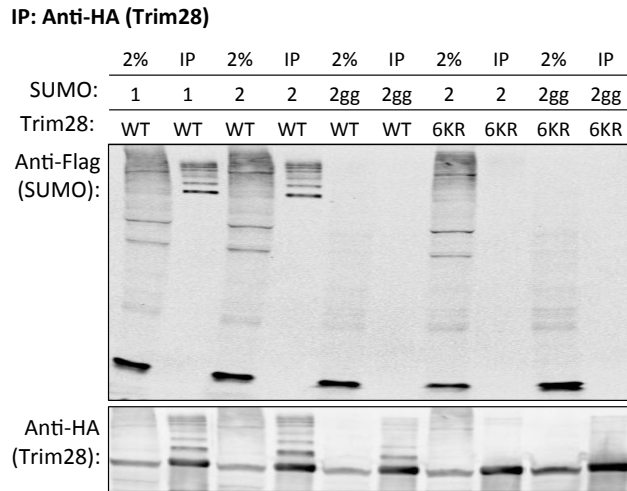
Figure 3-6 Detecting for SUMO conjugation of Trim28
(A) Immunoprecipitation of HA-tagged Trim28 and probing for Flag-tagged SUMO1. Wild-type Flag-SUMO1 (WT) or Flag-SUMO1GG (GG) was stably co-expressed with wild-type HA-Trim28 (WT) or HA-Trim28^{6KR} (6KR) in 293T cells. The input lanes demonstrate the expression levels of Flag-SUMO1, Flag-SUMO1GG, HA-Trim28, or HA-Trim28^{6KR}. Trim28 was immunoprecipitated using anti-HA magnetic beads and bound proteins were analyzed by Western blot probed with anti-Flag antibodies to detect sumoylation levels on Trim28. (B) Immunoprecipitation of HA-tagged Trim28 and probing for Flag-tagged SUMO2. SUMO2 is indicated by "2" and SUMO2^{GG} is indicated by "2gg," and immunopreciptiation of HA-Trim28 with SUMO1 was included as a control in in the first two lanes.

Trim28^{WT} or Trim28^{6KR}. These results confirmed that the six reported lysine residues on Trim28 encompass all available sites that can be conjugated by SUMO1.

To determine if SUMO2 can be conjugated to Trim28 and whether it conjugates to the same lysine residues as those conjugated by SUMO1, we co-expressed HA-Trim28^{WT} and HA-Trim28^{6KR} with either Flag-SUMO2 or Flag-SUMO2^{GG} in 293T cells. As previously done, cell lysates were prepared, HA-Trim28 was immunoprecipitated with anti-HA beads, and precipitated proteins were analyzed by Western blot probing for the Flag-tag. We detected a ladder of SUMO2-modified Trim28^{WT} products, similar to the ladder of SUMO1-modified Trim28^{WT} products, but SUMO2 conjugation to Trim28^{6KR} was not detected (Figure 3-6B). As expected, SUMO2^{GG} was not detected with the IP of either Trim28 proteins. These results indicate that one or more of the six reported lysine residues on Trim28 are targets for both SUMO1 and SUMO2 conjugation, opening the possibility for M-MLV silencing to be mediated by SUMO1, SUMO2, and possibly, other modifications that can potentially occur at the K779 residue.

Major phosphorylation sites of Trim28 are not involved in the YY1-Trim28 interaction or derepression of M-MLV

The interaction of Trim28 with co-repressor proteins such as HP1 and CHD3, is disrupted by phosphorylation modifications on S473 or S824 (Ivanov et al. 2007; Goodarzi et al. 2011; Bolderson et al. 2012; Chang et al. 2008). To test whether phosphorylation of these sites could also regulate the YY1-Trim28 interaction, we generated a panel of Trim28 phosphorylation mutants and tested them for binding to YY1. We constructed four Trim28 phosphorylation mutants – an alanine substitution mutation at S473 (S473A) or S824

(S824A), or a phosphomimetic mutation at S473 (S473E) or S824 (S824D) (Figure 3-7A).

The alanine mutations rendered those sites unavailable for phosphorylation while the

phosphomimetic mutations mimicked constitutively phosphorylated sites. The empty vector,

Trim28^{WT}, or a Trim28 mutant were expressed in F9 cells, followed by endogenous Trim28

KD and preparation of lysates. To measure expression levels of mutant Trim28 constructs,

lysates were analyzed by Western blot probing for Trim28. We detected Trim28 in the

untreated F9 cells (No KD), reflecting the basal levels of endogenous Trim28, but only very

low levels of Trim28 in the cells expressing the empty vector and Trim28 shRNA (EV). This

indicated that endogenous Trim28 was knocked down well in our Trim28 KD lines. We also

detected high levels of exogenous Trim28 expression for all mutants except for Trim28^{S473A}

(Figure 3-7B). YY1 was immunoprecipitated with an anti-YY1 antibody from cell lysates,

and bound proteins were examined for the presence of exogenous Trim28 by Western blot.

Trim28^{WT} and all mutant Trim28 proteins interacted with YY1 to a similar degree (Figure

3-7C), indicating that the S473E, S824A, and S824D mutations did not measurably impact

the YY1-Trim28 interaction.

As phosphorylation modifications were also reported to interfere with Trim28

transcriptional repressive activity (Li et al. 2007; Benjamin Rauwel 2015), we investigated

the importance of the phosphorylation sites for the silencing of M-MLV using the MLV-GFP

reporter assay. As previously described, the Trim28 phosphorylation mutant cell lines were

infected with an M-MLV-GFP vector at a constant multiplicity of infection and the

percentage of GFP-positive cells was measured by flow cytometry (Figure 3-7D). The

percentage of GFP-positive cells was 79% for NIH3T3 cells, 8% in untreated F9 cells, and

69% in the F9 cells expressing the empty vector and Trim28 shRNA (EV). The Trim28

A

B
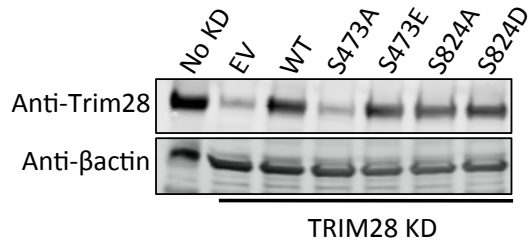
C
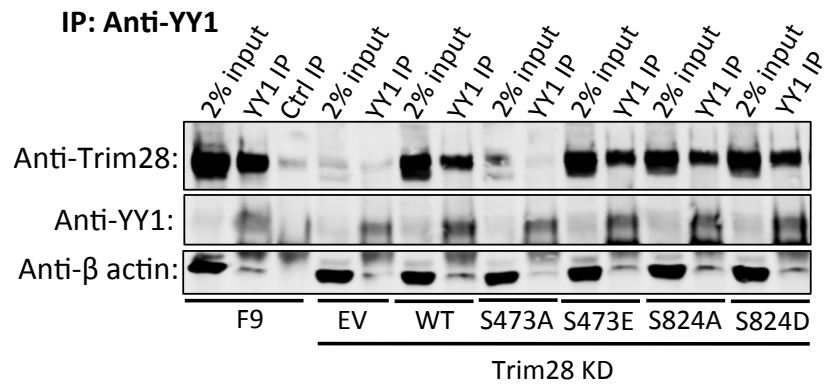
D
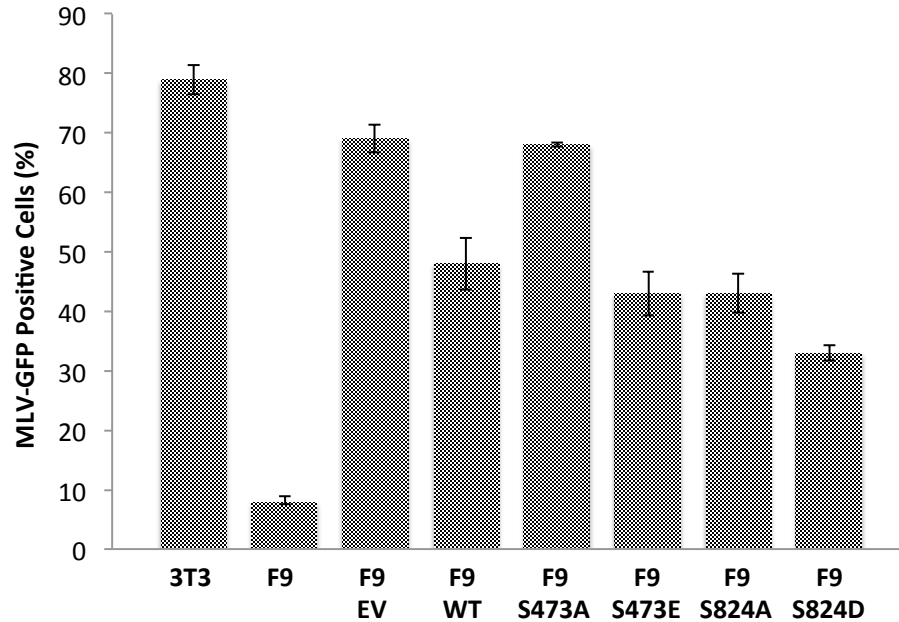
Figure 3-7 Mutation of two major phosphorylation sites on Trim28
(A) Schematic of Trim28 protein with the major phosphorylation sites marked. Serine sites were mutated to alanine or to an acidic residue that mimics a constitutively phosphorylated residue. (B) Mutant Trim28 was expressed and endogenous Trim28 was knocked down in F9 cells, and cell lysates were prepared and analyzed by Western blot. The "No KD" lane shows basal Trim28 levels in F9 cells. The empty vector (EV) lane shows the level of endogenous Trim28 expression in all KD lines, and the remaining lanes show expression levels of exogenous Trim28. (C) Western blot monitoring interaction of mutant Trim28 with YY1 by co-IP. YY1 was pulled-down with anti-YY1 antibody and the western blot was probed for Trim28. (D) NIH3T3, untreated F9 cells, and Trim28 mutant lines were infected with the M-MLV-GFP virus, and GFP expression was measured by flow cytometry. Uninfected cells were used to set our gates for counting "GFP-positive" expressing cells. Cells infected with the M-MLV-GFP vector were counted as GFP-positive if GFP expression was greater than that detected in the uninfected cells. NIH3T3 represent maximal M-MLV expression levels, and F9 cells represent maximal repression of M-MLV expression. We report the percentage of GFP-positive cells detected. Values represent the mean percentage of three biological replicates, and error bars represent the standard error mean.

mutant expressing cell lines expressed M-MLV-GFP to the same levels as Trim28^{WT}, indicating that these sites were not critical for regulating Trim28-mediated repression of M-MLV in embryonic cells. These results are distinct from those reported in studies of Trim28 regulation of HCMV latency and cell cycle progression (Li et al. 2007; Benjamin Rauwel 2015).

To test the importance of phosphorylation for the YY1-Trim28 interaction and M-MLV silencing using a different approach, we utilized pharmacological agents reported to alter phosphorylated Trim28 (pTrim28) levels. Chloroquine is a protease inhibitor, KU55933 is an ATM kinase inhibitor, and rapamycin is an mTOR inhibitor. Chloroquine was reported to increase pTrim28 levels, whereas KU55933 and rapamycin act to decrease pTrim28 levels (Benjamin Rauwel 2015). Neocarzinostatin (NCS), a potent DNA damaging agent shown to increase p^{S824}Trim28 levels (Ziv et al. 2006), was used as a positive control. As NCS treatment results in a significant amount of DNA damage, we did not use this as one of our experimental conditions. Embryonic cells can also be induced to differentiate by retinoic acid (RA) treatment (Jiao et al. 2012), and differentiation has been reported to decrease the phosphorylation of Trim28 (Seki et al. 2010). Therefore we also measured p^{S824}Trim28 levels in F9 cells treated with RA.

We attempted to check the basal and induced p^{S824}Trim28 levels in NIH3T3 and F9 cells by Western blot. p^{S473}Trim28 levels could not be checked due to the unavailability of a good antibody to this modified site. Lysates were prepared from NIH3T3s, F9s, and F9 cells differentiated by RA treatment and analyzed for p^{S824}Trim28 levels by Western blot; however, we did not detect p^{S824}Trim28 in any cell lysates (Figure 3-8A). Next, we tried altering phosphorylation levels by treating F9 cells with chloroquine, KU55933, chloroquine

A
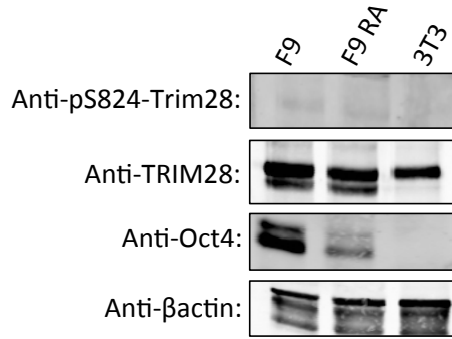
B
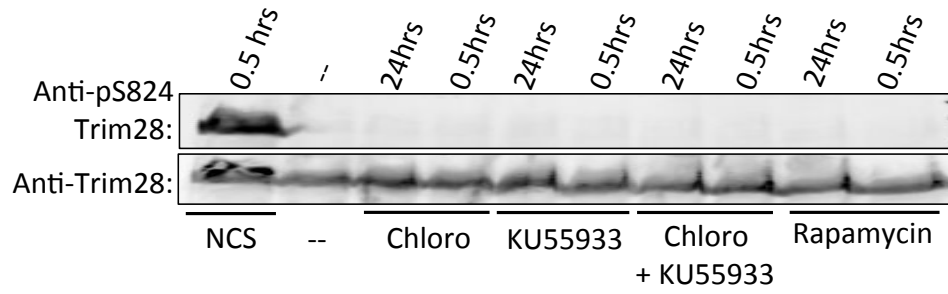
C
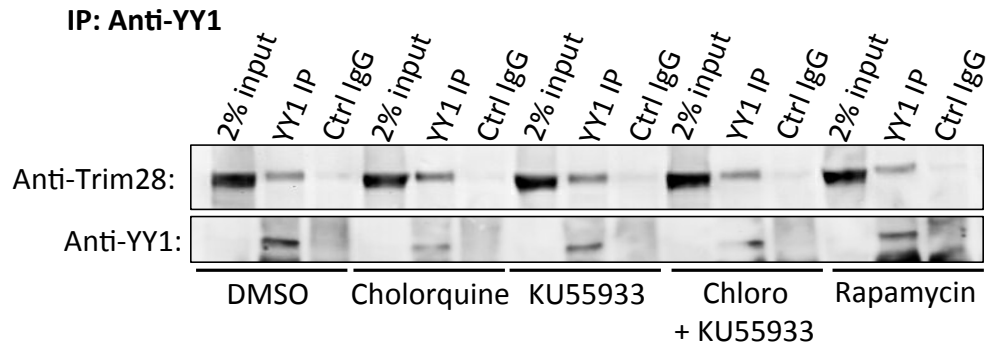
D
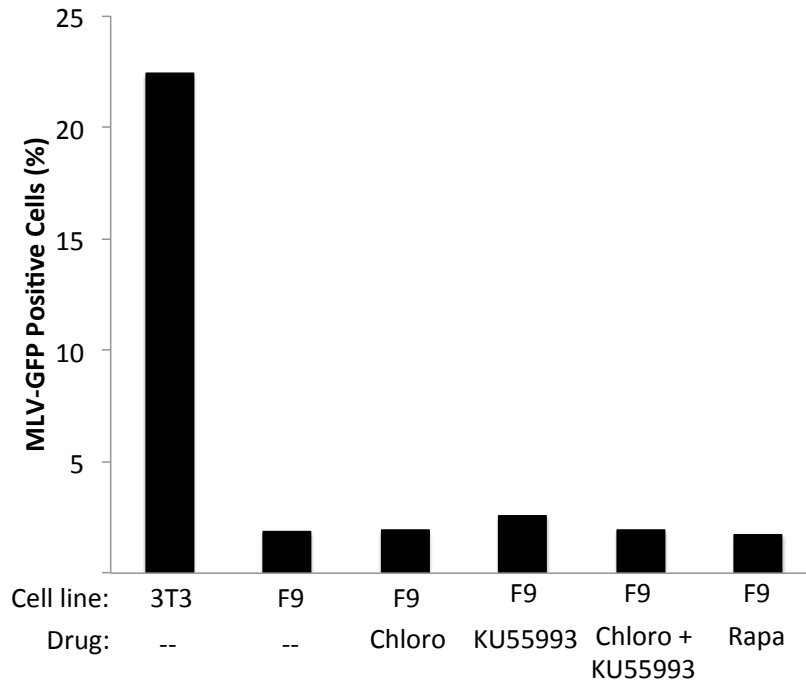
Figure 3-8 Treatment with pharmacological agents that alter phosphorylation levels
(A) F9, NIH3T3, and F9 RA treated cells were analyzed by Western blot probed with anti-pS824-Trim28 (ab70369). (B) Western blot analysis of cells treated with 20 µM chloroquine, 10 µM KU55933, 20 µM chloroquine and 10 µM KU55933, or 0.1 µM rapamycin for 30 minutes or 24 hours. Cells treated with 200 ng/ml of Neocarzinostatin (NCS) for 30 minutes was used as a positive control. (C) YY1 was immunoprecipitated with anti-YY1 antibody and detected for interaction with Trim28 by Western blot in F9 cells treated with drugs for 30 minutes. (D) F9 cells treated with drugs were infected with the M-MLV-GFP vector and the percentages of GFP-positive cells were measured by flow cytometry.

and KU55933, or rapamycin for 30 minutes or 24 hours. F9 cells treated with NCS for 30 minutes were used as a positive control. Lysates were prepared from the drug-treated cells and p^{S824}Trim28 levels were measured by Western blot. p^{S824}Trim28 levels remained undetectable after 30 minutes and 24 hours of drug treatments and was only detected in the NCS-treated cells (Figure 3-8B).

Even though we could not detect changes in phosphorylation in most settings, we nevertheless examined the effects of drug treatment on the YY1-Trim28 interaction and M-MLV silencing activity. Lysates were prepared from F9 cells treated with DMSO, chloroquine, KU55933, chloroquine and KU55933, or rapamycin for 24 hours, YY1 was immunoprecipitated, and bound proteins were analyzed by Western blot. The Trim28 interacted with YY1 to similar levels across all drug-treated and untreated cells (Figure 3-8C), indicating that treatment of F9 cells with the various drugs did not affect the YY1-Trim28 interaction. To determine if drug treatment affected M-MLV silencing activity, we infected cells with M-MLV-GFP virus after 30 minutes of drug treatment, as conducted in previous studies (Benjamin Rauwel 2015). Cells expressing higher levels of GFP relative to uninfected cells were counted as GFP-positive. Expression levels in NIH3T3-infected cells represented maximal expression and in untreated F9-infected cells represented maximal repression. The percentage of GFP-positive cells measured for cells treated with the various drug conditions were similar to that measured for untreated cells (Figure 3-8D), indicating that Trim28-mediated silencing activity on M-MLV was not affected by drug treatment. However, we could not confirm that drug treatment altered p^{S824}Trim28 levels as reported, so it is also possible that the drugs failed to generate an effect in our hands.

YY1-Trim28 complex is fragile and cannot be easily isolated

The results of the *in vitro* co-IP experiments raised the possibility that the YY1-Trim28 interaction depended on additional cofactors. So far, several cofactors have been studied for their involvement in the YY1-Trim28 interaction. Our previous work shows that the YY1-Trim28 interaction does not depend on DNA, RNA, or the presence of ZFP809 (Schlesinger et al. 2013). In this work, we have also eliminated SUMO and phosphorylation modifications on Trim28 as possible cofactors regulating the interaction (Figure 3-5C and Figure 3-7C), but other Trim28 modifications or YY1 modifications still remain viable options. Another possibility is that additional proteins bridge the YY1-Trim28 interaction.

Isolating the YY1-Trim28 complex could potentially identify additional bridging proteins, and several approaches can be utilized to isolate protein complexes. One method is to fractionate F9 nuclear extracts and monitor fractions for binding activity to M-MLV probes by EMSA, as previously done by Daniel Wolf in the purification of the silencing complex that bound the PBS element (Wolf & Goff 2007). Fractions containing binding activity to the M-MLV probes can be assayed by mass spectrometry to determine the components of the purified complex. We attempted to identify bridging proteins using this procedure. The first step was to detect the YY1-Trim28 complex by EMSA using an NCR probe. We incubated F9 nuclear extracts with a biotinylated NCR or NCR mutant probe, in which the YY1 binding site was scrambled. The lysates and probe mixture were displayed on a native gel, transferred to Western blot, and examined for the migration of the biotinylated probe. The NCR probe, and not the NCR mutant probe, was shifted when incubated with F9 lysates, demonstrating the presence of proteins binding specifically to the NCR sequence. (Figure 3-9A). When F9 nuclear extract was incubated with the NCR probe and anti-YY1

A
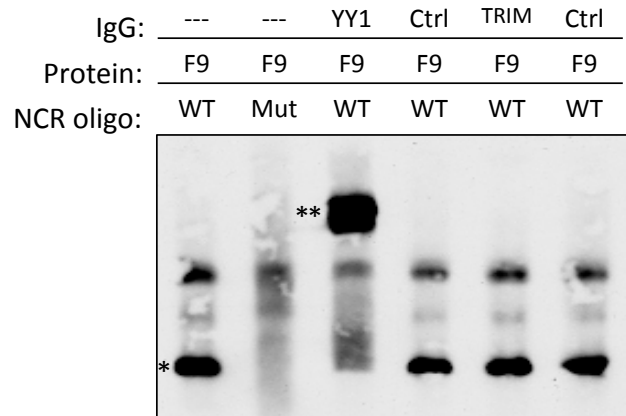
B
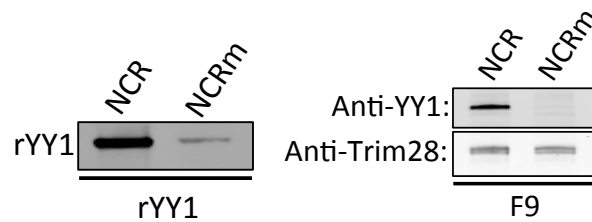
Figure 3-9 Isolating the YY1-Trim28 complex
(A) F9 nuclear lysates were incubated with biotinylated NCR (WT) or NCR mutant (Mut) oligonucleotides. Lysates and oligonucleotides were displayed on a non-denaturing gel, transferred to a membrane, and examined for the migration of the biotinylated olionucleotides. A shift of the NCR probe is noted with an asterisk and a supershift of the probe is noted with two asterisks. (B) DNA precipitation assay using biotinylated NCR or NCR mutant (NCRm) oligonucleotides bound to streptavidin beads. Beads were mixed with rYY1 protein as a control (left blot) and with F9 lysate (right blot). Proteins were removed from beads and displayed on Western blot to detect for the presence of YY1 and Trim28.

antibody, we detected a supershift of the NCR probe, indicating the presence of YY1 in this binding complex. However, when F9 lysates was incubated with the NCR probe and anti-Trim28 antibody, the NCR probe was not supershifted, indicating that we were detecting YY1 but not the YY1-Trim28 complex. We were discouraged from proceeding with further purification without indication of the retention of Trim28 in the shift complex.

The second approach we attempted was to isolate the YY1-Trim28 complex by binding to NCR oligonucleotides. Biotinylated NCR or NCR mutant oligonucleotides were immobilized to streptavidin-coated beads and incubated with F9 nuclear extract or with rYY1 as a positive control. Beads were washed with cell lysis buffer and bound proteins were removed and analyzed for the presence of the YY1-Trim28 complex by Western blot. We show that rYY1 was precipitated with the NCR, but not with the NCR mutant oligonucleotides, indicating the specificity of YY1 for the NCR sequence. We also detected endogenous YY1 in the nuclear extract proteins precipitated with the NCR, but not the NCR mutant oligonucleotide. Low levels of Trim28 were also detected, but similar levels of Trim28 were detected in the proteins precipitated with both the NCR and NCR mutant oligonucleotides, indicating that these are only background levels of Trim28 nonspecifically binding to the beads (Figure 3-9B). This result indicates that the YY1-Trim28 complex is unlikely to be isolated using this method either. The inability to isolate the YY1-Trim28 complex through the various methods attempted, suggests that the YY1-Trim28 complex is fragile and may be difficult to purify biochemically.

Chapter 4 : Detection, Isolation, and Identification of Larger RBS Complex

Detection of a larger RBS complex with an extended PBS probe

Early studies on the silencing of M-MLV in embryonic cells utilized exonuclease III protection assays and electrophoretic mobility shift assays (EMSA) to demonstrate the presence of a *trans*-acting silencing complex on the PBS element (Loh et al. 1990; Petersen et al. 1991). This complex is also referred to the repressor binding site (RBS) complex and contains Trim28 and ZFP809 (Wolf & Goff 2007; Wolf & Goff 2009). The RBS complex specifically binds to probes containing the PBS_{Pro} sequence (Loh et al. 1990; Petersen et al. 1991), but not to probes containing sequences permissive for M-MLV expression (PBS_{B2} and $PBS_{Gln,}$) (Grez et al. 1990; Barklis et al. 1986; Wolf & Goff 2007). Interestingly, we found a site containing the YY1 binding motif just upstream of the PBS site (Shi et al. 1997). We called this site "binding site 2" (BS2) and investigated whether YY1 and the RBS complex could bind to it. To assess this, we designed various biotinylated oligonucleotide probes spanning the region containing the BS2 and/or the PBS sequence and detected for their interaction with protein complexes in F9 nuclear extracts by EMSA. Biotinylated probes containing the following sequences were made: the PBS sequence alone (PBS), the PBS_{B2} mutant sequence (BS), the BS2 and PBS sequences (BS2 + PBS), the BS2 and PBS_{B2} sequences (BS2 + B2), the BS2 sequence alone (BS2), the BS2 sequence scrambled (BS2 scram), the sequence upstream of the BS2 and the PBS sequence (ΔBS2 + PBS), and the sequence upstream of the BS2 and the PBS_{B2} sequence (ΔBS2 + PBS_{B2}) (Figure 4-1A). The B2 probe served as the negative control for the PBS probe, the BS2 + B2 probe served as the

A

```
M-MLV:       CTACCCGTCAGCGGGGGTCTTTCATTTGGGGGGCTCGTCCGGGATCGGGAGACCCC
PBS:                                      GGGGGCTCGTCCGGGATCGGGAGACCCC
B2:                                       GGGGGCTCGTCCGAGATCGGGAGACCCC
BS2 + PBS:                   GTCTTTCATTTGGGGGGCTCGTCCGGGAT
BS2 + B2:                    GTCTTTCATTTGGGGGGCTCGTCCGAGAT
BS2:         CTACCCGTCAGCGGGGGTCTTTCATTTG
BS2 scram:   CTACCCGTCAGCGGGGGTCTTTCGGCTA
ΔBS2 + PBS:                  GCGGGGGTCT-----------TGGGGGCTCGTCCGGGAT
ΔBS2 + B2:                   GCGGGGGTCT-----------TGGGGGCTCGTCCGAGAT
```

B
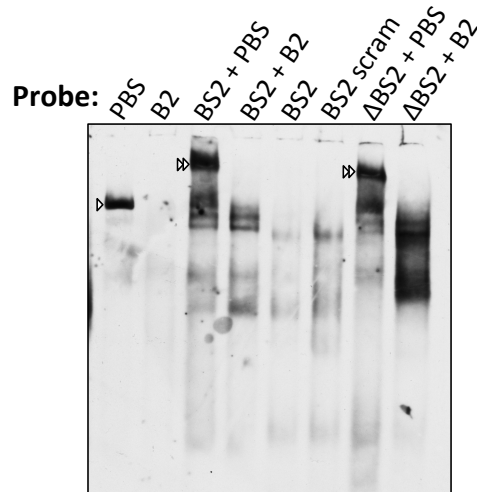
Figure 4-1 Detection of the large RBS complex
(A) A fragment of the M-MLV sequence containing the BS2 and PBS sequences is shown. The sequence of the BS2 site is underlined with a solid line and the sequence of the repressor-binding site contained within the PBS sequence is underlined with a dotted line. Sequences for the various biotinylated probes created are listed below the M-MLV sequence. Probes containing the PBS_{B2} (B2) sequences served as the negative control for the probes containing the PBS sequence. Probes containing the BS2 scrambled sequence (BS2 scram) or a sequence upstream of the BS2 (ΔBS2 + B2) were used as negative controls for probes that contained the BS2 sequence. Mutations are indicated with red font and deletions are indicated with a hash mark line in place of the deleted sequence. (B) F9 cell nuclear extracts were incubated with the indicated biotinylated probes for 30 minutes, resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, UV crosslinked, probed with a streptavidin-HRP conjugate, and incubated with a substrate to detect for HRP. Probes that shifted with the RBS and larger RBS complexes are marked with a single and double arrowhead, respectively.

negative control for the BS2 + PBS probe, the BS2 scram probe served as the negative control for the BS2 probe, and the ΔBS2 + PBS$_{B2}$ probe served as the negative control for the ΔBS2 + PBS probe. Nuclear extracts were prepared from F9 cells and incubated with the probes in a standard binding buffer. The nuclear extract and bound probe were resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, UV crosslinked, probed with a streptavidin-HRP conjugate, and incubated with a substrate to detect HRP. As seen in previous studies, the PBS probe shifted with the RBS complex while the B2 probe did not (Figure 4-1B). We also found that the BS2 + PBS probe shifted with a complex that ran at a higher molecular weight than the RBS complex, whereas the BS2 + B2 probe did not. The necessity for the wild-type PBS sequence suggests that the higher molecular weight complex contained the RBS complex; hence, we referred to this complex as the "larger RBS complex." However, we found that the BS2 and BS2 scram probes did not shift when incubated with F9 nuclear extracts, indicating the necessity of the PBS site for detection of the larger RBS. Surprisingly, we found that the ΔBS2 + PBS probe, which contained the PBS site and a sequence upstream of the BS2, was also capable of shifting with the larger RBS complex, whereas the ΔBS2 + B2 did not. We created additional probes with various non-specific sequences upstream of the PBS, and found that these probes also shifted with the larger RBS complex (data not shown). These results demonstrate that the larger RBS complex is not specific for the BS2 + PBS sequence, but rather, depends on the presence of the PBS and the addition of extra nucleotides upstream of the PBS, apparently of any sequence.

Larger RBS complex contains Trim28 but not YY1

To determine if YY1 and Trim28 were contained in the larger RBS complex, we conducted binding reactions in which the BS2 + PBS probes were incubated with F9 nuclear extracts and anti-YY1 antibody or anti-Trim28 antibody. Binding reactions containing nuclear extracts and the PBS and BS2 + PBS probes, without antibody addition, were conducted as positive controls. Binding reactions containing nuclear extract and the B2, BS2 + B2, and ΔBS2 + B2 probes were conducted as negative controls. Nuclear extracts were resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, and examined for migration of the DNA probes (Figure 4-2A). As we had shown before, the PBS and BS2 + PBS probes shifted with the RBS and larger RBS complexes, respectively, whereas the negative control probes (B2 and BS2 + B2) did not. The ΔBS2 + PBS probe also shifted to the molecular weight of the larger RBS complex, as we had previously seen. We found that the addition of anti-Trim28 antibody to a binding reaction containing the BS2 + PBS probe and nuclear extract supershifted the probe, indicating the presence of Trim28 in the RBS and larger RBS complex. However, the addition of anti-YY1 antibody to a binding reaction containing the BS2 + PBS probe and nuclear extract did not supershift either the RBS or larger RBS complexes. These results may indicate that YY1 is absent from these complexes or that the YY1 antibody failed to bind YY1 protein within the context of the complex. Thus, we were unable to demonstrate the presence of YY1 in these complexes. A technical difficulty of this experiment was that the larger RBS complex runs at a very high molecular weight, making it difficult to resolve from a shifted larger RBS complex. To improve separation of the large RBS complex and the shifted larger RBS complex, we increased the amount of time we ran the gel from four hours to ten hours. This provided a better separation
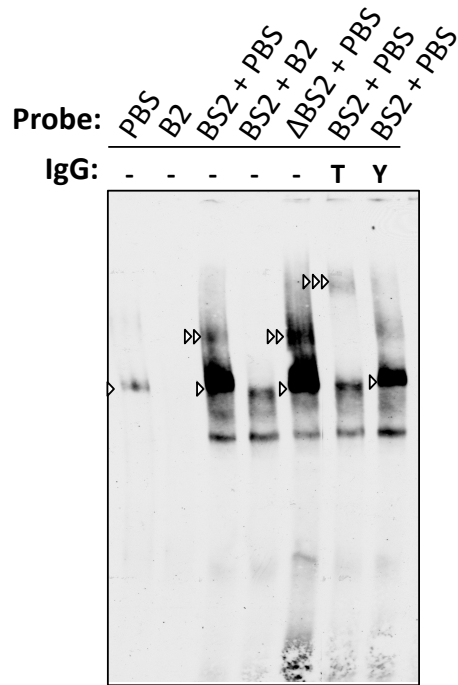
Figure 4-2 Large RBS Complex contains Trim28
F9 cell nuclear extract was incubated with the indicated biotinylated probe for 30 minutes, resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, UV crosslinked, probed with a streptavidin-HRP conjugate, and incubated with a substrate to detect for HRP. Binding reactions that included anti-Trim28 antibody (T), anti-YY1 antibody (Y) or no antibody ("-") are indicated. Probes that shifted with the RBS and larger RBS complexes are marked with a single and double arrowhead, respectively. Probes that shifted with a supershifted complex are marked with a triple arrowhead.

of the higher molecular weight complexes but also decreased the amount of complex detected. Thus, we were unable to improve the detection of the supershifted probes under the conditions necessary to separate the large molecular weight complexes.

Factors identified in the larger RBS complex

Although we did not identify YY1 in the larger RBS complex, we were interested to identify any additional factors contained in the larger RBS complex. To identify these factors, we tried two methods for purifying the larger RBS complex. The first was to separate out the complex by ammonium sulfate $((NH_4)_2SO_4)$ precipitation, the method used to originally precipitate the RBS complex by Daniel Wolf and colleagues (Wolf & Goff 2007). F9 nuclear extract was prepared and treated with increasing concentrations of ammonium sulfate. Precipitated proteins were collected by centrifugation after 25%, 40%, and 80% saturation by ammonium sulfate. The precipitated proteins were re-dissolved, ammonium sulfate was removed by a centrifugal filter unit, and proteins were examined for the presence of the larger RBS complex by EMSA (Figure 4-3A). F9 nuclear extracts that had not been treated with ammonium sulfate were used as a positive control (untreated F9 lysate). F9 nuclear extract or protein complexes fractionated by ammonium sulfate precipitation were incubated with the PBS, B2, BS2 + PBS, or BS2 + B2 probes. Protein and probe binding reactions were resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, and detected for the DNA probes. Binding reactions with the F9 nuclear extract demonstrated a shift of the PBS and BS2 + PBS probes, to the molecular weights of the RBS and large RBS complex, respectively. From the precipitated proteins, only proteins complexes

A

| Untreated F9 lysate | 25% (NH$_4$)SO$_4$ | 40% (NH$_4$)SO$_4$ | 80% (NH$_4$)SO$_4$ |

B
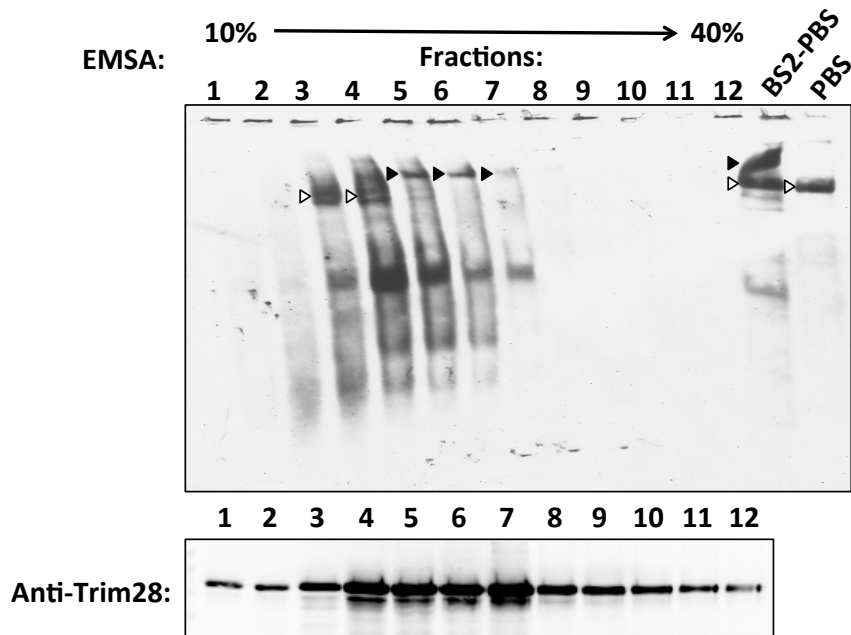
Figure 4-3 Isolation of the large RBS complex
(A) Protein complexes were precipitated from F9 nuclear extract by saturating nuclear extract samples with 25%, 40%, and 80% ammonium sulfate. At each saturation step, precipitated proteins were collected by centrifugation, resuspended, and passed through a buffer exchange column. F9 cell nuclear extracts or proteins fractionated by ammonium sulfate precipitation w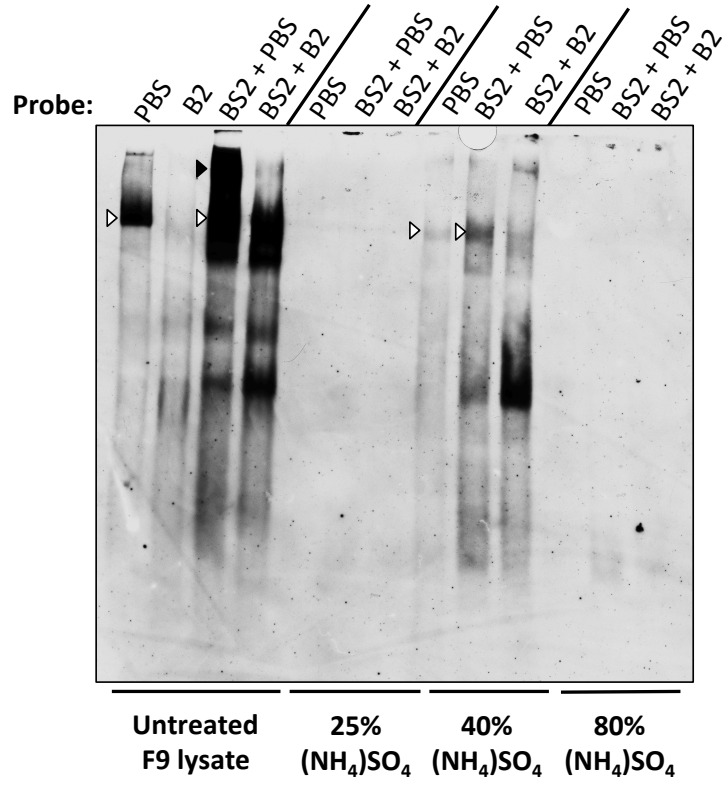ere incubated with the indicated biotinylated probe for 30 minutes, resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, UV crosslinked, probed with a streptavidin-HRP conjugate, and incubated with a substrate to detect for HRP. Probes that shifted with the RBS and larger RBS complexes are marked with white and black arrowhead, respectively. (B) F9 nuclear extract protein complexes were separated by velocity sedimentation using a 10-40% sucrose gradient. 12 fractions were collected and examined for the presence of the larger RBS complex by EMSA. 8% of each fraction was incubated with the BS2 + PBS probe. F9 nuclear extracts were incubated with the PBS and BS2 + PBS probes as a positive control (last two lanes). Nuclear extracts and probes were resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, and examined for the migration of the probes.

precipitated at 40% ammonium sulfate saturation demonstrated binding to the PBS and BS2 + PBS probes. Both probes shifted to the molecular weight of the RBS complex but neither shifted to the molecular weight of the large RBS complex. These results show that the ammonium sulfate precipitation method does not precipitate the larger RBS complex as effectively as it does the RBS complex. Perhaps this suggests that the larger RBS complex is more susceptible to disintegration in the high salt concentrations and requires gentler methods for isolation.

To try a gentler method for isolating the larger RBS complex, we separated F9 nuclear extract protein complexes by velocity sedimentation on a 10-40% sucrose gradient. Fractions were collected, and samples of each fraction were incubated with the BS2 + PBS probe and examined for the presence of the larger RBS complex by EMSA. Total F9 nuclear extracts were incubated with the PBS or BS2 + PBS probes as a positive control. Proteins and bound probes were resolved on a non-denaturing gel, transferred to a nitrocellulose membrane, and examined for the migration of the probes. The results suggested that it was possible to separate distinct complexes with distinct components. In particular, we detected the smaller RBS complex in fractions with lower sucrose densities and the larger RBS complex in fractions with higher sucrose densities (Figure 4-3B, above). Fractions were also examined by Western blot for the presence of Trim28. Trim28 was most abundant in the same fractions in which we found to contain both the RBS and larger RBS complexes (Figure 4-3B, lower blot). We separately pooled fractions containing either the RBS complex or fractions containing the large RBS complex and immunoprecipitated Trim28 using a polyclonal Trim28 antibody. Immunoprecipitation of proteins using a control antibody was also conducted for both pooled fractions as a negative control. Protein A/G beads were added

to the nuclear extract and antibody complex, and bound proteins were removed from the beads and analyzed by mass spectrometry. A list of proteins detected with the IP of Trim28 was compiled, and proteins also found in the control antibody IPs were eliminated as non-specific binding proteins. The remaining list demonstrates 27 proteins in the larger RBS complex and 22 proteins in the RBS complex as candidates that may interact with Trim28 (Table 1). Several proteins such as KHDR1, SK2L2, LC7L3, and ARGL1 were notable for their connections to viral or transcriptional functions and warrant further investigation for their involvement in the silencing of M-MLV. These proteins are further described in the discussion section.

In summary, we have identified a large complex that contains Trim28 and the RBS complex, and we demonstrate an effective approach for purifying the large complex. We have provided a list of proteins identified in the purified complexes by mass spectrometry, which can serve to identify new repressor cofactors involved in the transcriptional silencing of M-MLV in embryonic cells.

Chapter 5 : The Regulation of SUMO1-modified Substrates in Mouse Embryonic Cells

SUMO1 cannot be overexpressed in mouse embryonic cells

SUMO1 can be successfully overexpressed in differentiated cells (Fukuda et al. 2009; Bossis & Melchior 2006), and there are no studies that suggest that SUMO1 would not be overexpressed just as effectively in embryonic cells. Our interest in SUMO1 expression in embryonic cells developed from an observation that attempts to induce SUMO1 overexpression in the F9 embryonic carcinoma cell line were not successful – SUMO1 expression was highly inefficient. The basis for this problem was not clear.

As embryonic cells are difficult to transfect and can silence a variety of promoters, we examined a number of gene constructs introduced into these cells. In one series of experiments, we used the pLVX lentiviral vector to overexpress Flag-tagged SUMO constructs by transduction. In this system, the expression of the transgene is driven by the EF1α promoter, followed by an IRES element and a drug resistance gene, which allows the SUMO and the drug resistance genes to be expressed as a single bicistronic transcript and translated separately. Lentiviral particles were produced by transfection in 293T cells, supernatants containing virions were collected, passed through a 0.45 µ filter, and were used to infect F9 cells. We designed constructs containing either the wild-type SUMO1 gene or a truncated form of SUMO1 that cannot be conjugated to any substrate (termed the SUMO1GG mutant). F9 cells transduced with the Flag-SUMO1 or Flag-SUMO1GG vectors were selected for stable expression of the drug resistance gene, and exogenous SUMO1 expression levels were analyzed by Western blot probing for the Flag-tag. We observed that

SUMO1 expression was strikingly less efficient than SUMO1GG expression in F9 cells (Figure 5-1A, left panel). This unexpected result led us to try overexpressing SUMO1 in other embryonic cell lines. PCC4 is a mouse embryonic carcinoma cell line, and E14 is a mouse embryonic stem cell line. We transduced both cell lines with the SUMO1 and SUMO1GG vectors, selected for stable expression of the drug resistance gene, and assessed SUMO1 protein expression by Western blot. Again, we found SUMO1 was poorly expressed in both PCC4 and E14 cells, while SUMO1GG expressed well (Figure 5-1A, center and right panels). To test if SUMO1 could be overexpressed in differentiated cells, we transduced the mouse NIH3T3 and human 293T cell lines with the SUMO1 and SUMO1GG vectors, selected for stable expression of the drug resistance marker, and measured Flag-SUMO1 expression in total cell lysates by Western blot. In contrast to the embryonic cells, the differentiated cells expressed similar levels of SUMO1 and SUMO1GG (Figure 5-1B). These results indicate that embryonic cells, but not differentiated cells, have a block against overexpression of SUMO1.

Retinoic acid (RA) treatment of embryonic cells activates a cascade of changes in the chromatin structure and transcription that result in the differentiation of embryonic cells (Jiao et al. 2012). To determine if differentiation of embryonic cells by RA treatment could improve SUMO1 overexpression, F9 cells transduced with the SUMO1 or SUMO1GG vectors were selected for the stable expression of the SUMO vectors and were treated with RA for five or seven days. Cell lysates were prepared and analyzed by Western blot probing for the expression of exogenous SUMO1. RA treatment for five or seven days did not improve SUMO1 overexpression in F9 cells (Figure 5-2A, left). These results suggested that SUMO1 overexpression could not be improved by RA treatment, or that cells needed to be
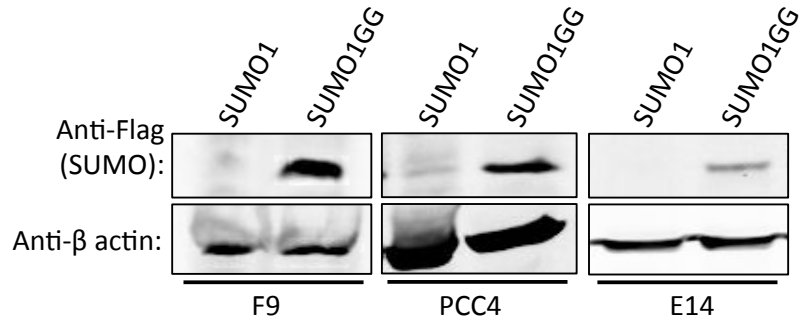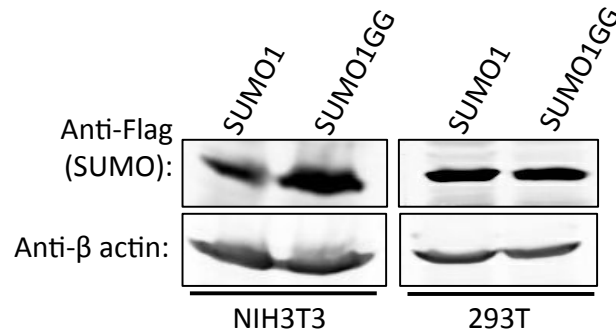
Figure 5-1 SUMO1 cannot be overexpressed in embryonic cells
(A) Western blot displaying expression levels of Flag-SUMO1 or Flag-SUMO1GG in F9, PCC4, and E14 cells transduced with and selected for stable expression of the SUMO vectors by drug selection. (B) Western blot displaying expression levels of Flag-SUMO1 or Flag-SUMO1GG in NIH3T3 and 293T cells transduced by and selected for stable expression of the SUMO vectors by drug selection.

A
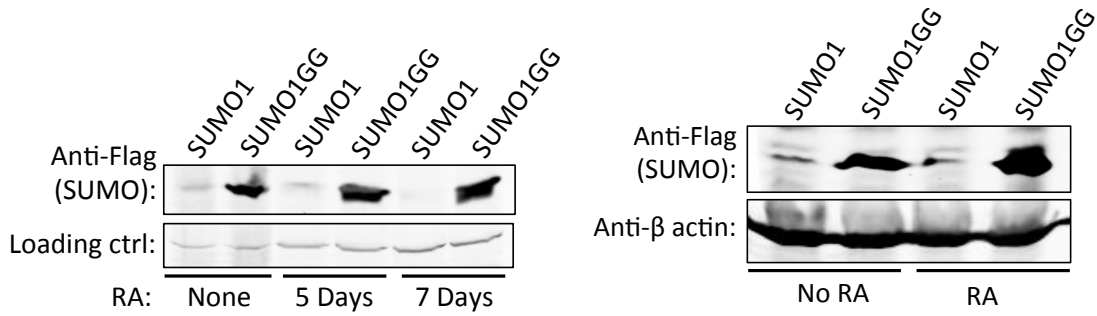
B
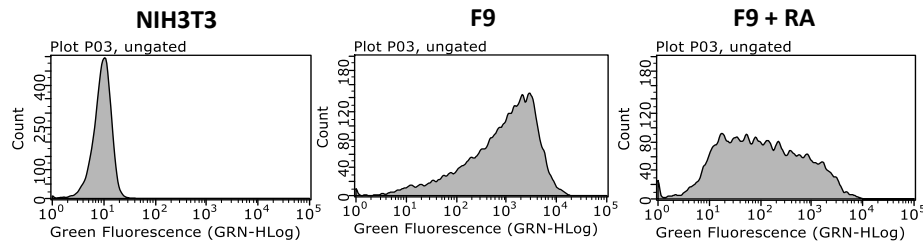
Figure 5-2 SUMO1 overexpression levels in embryonic cells differentiation by retinoic acid

(A) Western blot displaying expression levels of SUMO1 or SUMO1GG in F9 cells transduced with and selected for the expression of the SUMO vectors, followed by retinoic acid (RA) treatment for 5 or 7 days (left). Western blot displaying expression levels of SUMO1 or SUMO1GG in F9 cells treated, or untreated, with RA for 3 days prior to transduction with and selection for the expression of the SUMO vectors (right). (B) NIH3T3 cells, F9 cells, and F9 cells treated with RA for 3 days were stained with the AF488 SSEA-1 antibody, and expression levels of the SSEA-1 cell surface marker were measured by flow cytometry. Each flow cytometry histogram represent 5,000 collected events, and the expression levels are shown on a logarithmic scale.

differentiated prior to SUMO1 overexpression. To test the latter possibility, we treated F9 cells with RA for three days prior to transduction with the SUMO1 or SUMO1GG vectors. Lysates were prepared and analyzed for SUMO1 overexpression by Western blot. We found that exogenous SUMO1 was not expressed well regardless of prior RA treatment (Figure 5-2A, right), indicating that differentiation by RA treatment was insufficient to alter SUMO1 expression levels in F9 cells. To confirm that cells were indeed differentiated as result of the RA treatment, F9 cells treated with RA for three days were examined for expression of the cell-surface embryonic stem cell marker, SSEA-1, by flow cytometry. We also measured SSEA-1 expression levels on NIH3T3 cells as a negative control and on untreated F9 cells as a positive control. The flow cytometry histograms demonstrate that NIH3T3 cells express very low levels of SSEA-1, untreated F9 cells express high levels of SSEA-1 and F9 cells treated with RA express moderate levels of SSEA-1 (Figure 5-2B). These results demonstrate a clear shift of the F9 cell population towards lower SSEA-1 expression levels with RA treatment, though not to the extent that is detected in fully differentiated NIH3T3 cells. Prolonging RA treatment to five and eight days also did not lower the SSEA-1 expression levels on F9 cells any further (data not shown). These results indicate that RA treatment may only partially differentiate F9 cells, and that the level of differentiation achieved by RA treatment is insufficient for achieving the ability to the overexpress SUMO1 in these cells.

One possible explanation for the apparently low levels of exogenous SUMO1 expression in embryonic cells was that SUMO1 overexpression was causing cell death and only cells that maintained low SUMO1 expression were surviving. To test this, we transduced F9 cells with equal titers of the SUMO1 or SUMO1GG containing virions and selected cells for expression of the drug resistance marker over two weeks. Surviving cells

formed colonies and were counted as a readout for survival with the expression of the SUMO vectors. We found that cells transduced with the SUMO1 and SUMO1GG vectors formed roughly equivalent numbers of colonies (Figure 5-3), demonstrating that the expression of the SUMO1 vector was not causing more cell death than expression of the control vector.

To check if SUMO1 could be expressed in the first few days after transduction, we measured SUMO expression after transient transduction with no drug selection. F9 cells were transduced with the SUMO constructs, and lysates were prepared 72 hours after transduction and were examined for SUMO1 overexpression by Western blot. We detected some expression of SUMO1 after transient transduction, but SUMO1 levels were still lower relative to SUMO1GG (Figure 5-4). The transient expression experiments suggest that a low level of SUMO1 overexpression may be attained, but even at this point we see lower levels of SUMO1 expression relative to SUMO1GG expression, indicating that an activity is limiting expression of the wild-type SUMO but not a non-conjugatable SUMO. We also probed for SUMO1 and SUMO1GG levels by Western blot 24 and 48 hours after transduction, but SUMO1GG expression levels were still too low to detect at these time points (data not shown).

To further investigate if embryonic cells can tolerate low to moderate overexpression of SUMO1, we selected for SUMO1 overexpressing cells by a different selection marker. We cloned the ZsGreen reporter gene into the SUMO1 and SUMO1GG vectors, in place of the drug resistance genes. F9 cells were transduced with these constructs and selected for expression of the ZsGreen reporter by fluorescence-activated cell sorting (FACS). Cells expressing ZsGreen above the background of uninfected cells were counted as ZsGreen-positive and were collected. Cells were sorted three times in succession at 8, 16, and 26 days
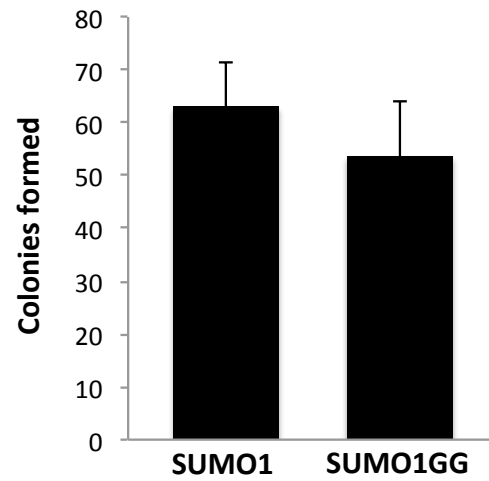
Figure 5-3 Tittering SUMO1 and SUMO1GG viruses on F9 cells
F9 cells were transduced with equal quantities of the Flag-SUMO1 or Flag-SUMO1GG vectors and selected for stable expression of the vectors by drug selection over 2 weeks. Surviving cells formed colonies and were quantified. Values represent the mean of three biological replicates, and error bars represent the standard error mean.
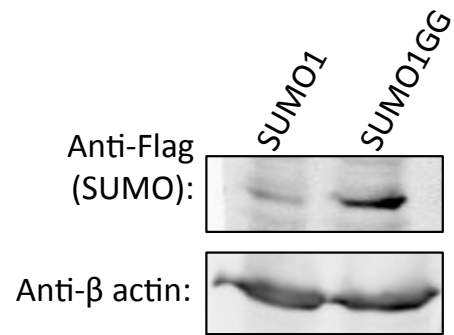
Figure 5-4 Transient transduction of the SUMO vectors in embryonic cells
F9 cells were transduced with the Flag-SUMO1 or Flag-SUMO1GG vectors with no drug selection. Lysates were prepared three days after transduction and analyzed for exogenous SUMO1 or SUMO1GG expression by Western blot.

after transduction to continue increasing the purity of ZsGreen-expressing cells over the three sorts. After each sorting event, cells were allowed to recover and expand over several days, and ZsGreen expression was confirmed by flow cytometry. The proportion of cells expressing the ZsGreen reporter was less than 1% 6 days after transduction for both cell lines. By 29 days after transduction, we were able to recover populations in which 43% of the cells were scored as ZsGreen-positive for the SUMO1-ZsGreen expressing cell population and 39.5% for the SUMO1GG-ZsGreen expressing cell population (Figure 5-5A and B). Cell lysates were prepared from the sorted populations and examined for SUMO1 and SUMO1GG expression by Western blot. SUMO1 and SUMO1GG expression were undetectable in the cells selected by the ZsGreen reporter gene (Figure 5-5C). The lack of SUMO1GG expression was unexpected, but can possibly be explained by our sorting strategy. Cells selected for expression of the ZsGreen reporter were not under a sufficient amount of pressure to select for SUMO expression levels high enough to be detected by Western blot, whereas we can detect SUMO1GG in the cells selected for expression by the drug selection marker.

Expression levels of endogenous SUMO are too low to detect in murine cell lines

These observations led us to ask whether endogenous SUMO1 is also expressed at higher levels in differentiated cells than in embryonic cells. To assess endogenous SUMO1 protein expression levels, we prepared cell lysates from the differentiated murine cell lines, Rat2 and NIH3T3, and embryonic cell lines F9, PCC4, and E14. We also prepared lysates from F9 cells expressing the SUMO1GG vector as a positive control. Lysates were examined for SUMO1 expression by Western blot probed with an anti-SUMO1 antibody

A
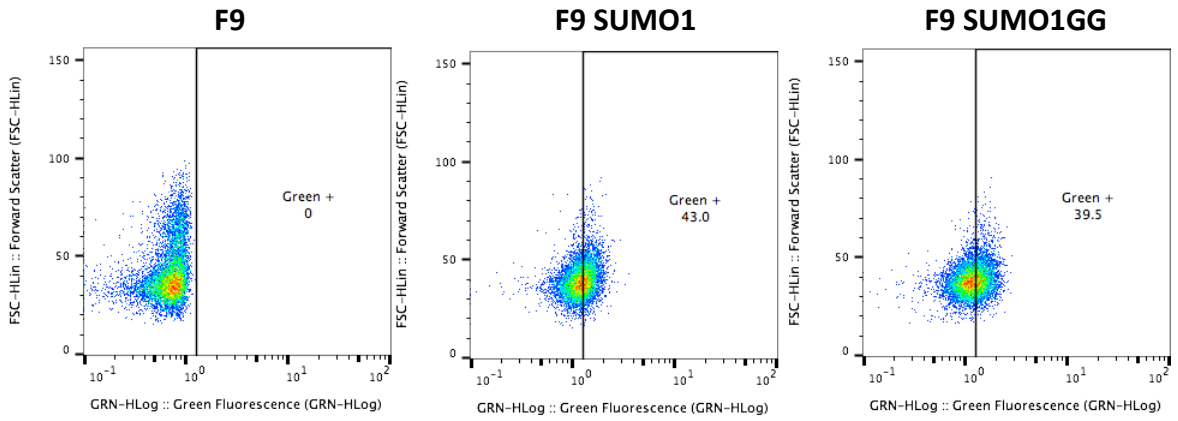
B
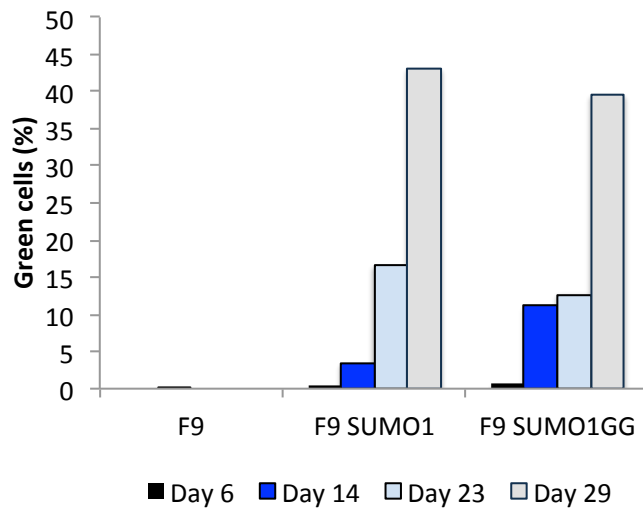
C
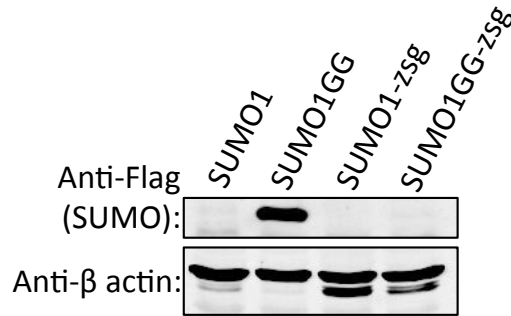
Figure 5-5 Selection for expression of the SUMO1 vector by a ZsGreen reporter
(A) The levels of ZsGreen expression were measured by flow cytometry. 5,000 events were collected for each sample and were analyzed by the FlowJo software. Events are shown on a dot plot measuring forward scatter (FSC) and ZsGreen expression plot. Cells expressing higher levels of ZsGreen relative to the uninfected cells were counted as ZsGreen-positive. An example of flow cytometry dot plots measuring ZsGreen expression 29 days after transduction is shown. (B) The progression in the proportion of ZsGreen-expressing cells for each cell line was monitored over the course of the 29 days. The percentage of ZsGreen-expressing cells quantified by flow cytometry at 6, 14, 23, and 29 days after transduction are shown in the bar graph. (C) Western blot probing for the expression of SUMO1 or SUMO1GG in cells selected for SUMO1-ZsGreen and SUMO1GG-ZsGreen expression by FACS. Lysates from F9 cells transduced with and selected for the expression of the SUMO vectors containing the drug resistance marker gene were used as controls.

(Figure 5-6A). Endogenous SUMO1 protein levels were too low to detect in all cell lines tested, except for the positive control, demonstrating that basal SUMO1 protein expression levels in these cell lines fall below the detection limit by Western blot. To assess endogenous SUMO1 mRNA levels, we isolated RNA from the aforementioned cell lines, synthesized cDNA and measured endogenous SUMO1 expression by quantitative reverse transcriptase PCR (qRT-PCR) using primers that span an exon-exon junction. SUMO2 expression was also measured as a comparison. Endogenous SUMO1 and SUMO2 transcript levels were low in all cell lines and could not be reliably measured, except for in the positive control (Figure 5-6B). Thus, we were unable to compare endogenous SUMO1 protein and transcript levels between differentiated and undifferentiated cells due to our inability to measure basal SUMO1 expression levels either by Western blot or qRT-PCR. These results demonstrate that basal SUMO1 levels are found in very low levels across a variety of murine cell lines.

SUMO1 overexpression is prevented at the post-transcriptional level

The mechanism responsible for restricting SUMO1 overexpression could be regulating SUMO1 expression on any of several levels, including at the level of transcription, post-transcriptional mRNA processing and stability, translation, or post-translational protein stability. Although unlikely, in some rare events retroviral vectors are unstable and can delete genes from the viral genome, making it possible that some of our cells selected for the SUMO1 vector contained the drug resistance gene and not the SUMO1 transgene. To rule out this possibility, we first examined the SUMO1 transgene. We isolated DNA from F9 cells infected with the empty vector, or the SUMO1 or SUMO1GG vectors that had been selected for stable expression of the drug selection marker. SUMO1 plasmid DNA was used as a

A
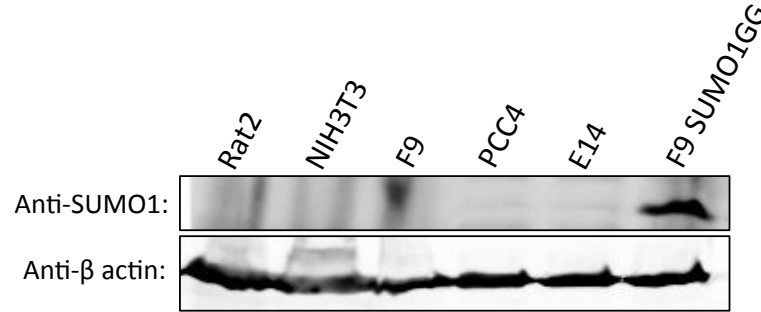
B
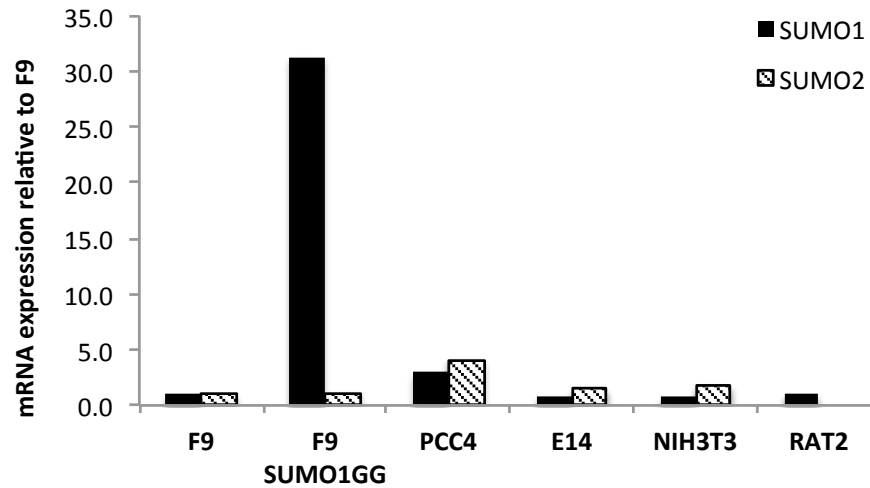
Figure 5-6 Endogenous SUMO1 expression levels
(A) Western blot probing for endogenous SUMO1 protein in lysates prepared from differentiated (Rat2 and NIH3T3) and undifferentiated (F9, PCC4, and E14) cell lines. Lysate prepared from F9 cells expressing the SUMO1GG vector was used as a positive control. (B) Measurement of endogenous SUMO1 and SUMO2 mRNA levels in undifferentiated (F9, PCC4, E14) and differentiated (NIH3T3, Rat2) cell lines by qRT-PCR. RNA isolated from F9 cells transduced with the SUMO1GG vector was used as a positive control. Expression levels were normalized to the housekeeping gene, GAPDH, and were reported relative to F9 cells.

positive control. SUMO1 and SUMO1GG transgene levels were measured by qPCR using primers spanning an exon-exon junction such that endogenous SUMO1 DNA would not be amplified, and only the SUMO1 DNA introduced by the viral vector would be detected. We calculated similar SUMO1 and SUMO1GG transgene levels relative to the puromycin gene in F9 cells (Figure 5-7A). SUMO1 DNA was not detected in F9 cells expressing the empty vector (F9 EV), and SUMO1 amplified well in the plasmid control (SUMO1 DNA). This experiment demonstrates that the difference between the SUMO1 and SUMO1GG expression levels were not due to rare events in which the SUMO vectors did not successfully transduce the SUMO transgene.

Next, we checked if the integrated transgene contained any mutations that could prevent SUMO1 expression. Genomic DNA was isolated from F9 cells transduced with the SUMO1 or SUMO1GG vectors, and the SUMO1 insert was amplified by PCR using primers that spanned the EF1α promoter and IRES region (Figure 5-7B, top). PCR amplification of the SUMO1 insert was also conducted with plasmid DNA as a positive control and with DNA isolated from uninfected F9 cells as a negative control. PCR of DNA extracted from F9 cells transduced with the SUMO1, SUMO1GG, or empty vector demonstrated PCR products that were the same size as the PCR products amplified from the corresponding plasmid controls (Figure 5-7B, bottom). The PCR products amplified from F9 cells were purified and sequenced. The sequences did not contain any detectable mutations in the SUMO1 transgene, indicating that the SUMO1 transgene remained intact in F9 cells.

To examine exogenous SUMO transcript levels, we isolated RNA from F9 and NIH3T3 cells transduced with and selected for the stable expression of the SUMO1 and SUMO1GG vectors. RNA was also isolated from untreated F9 and NIH3T3 cells as negative

A
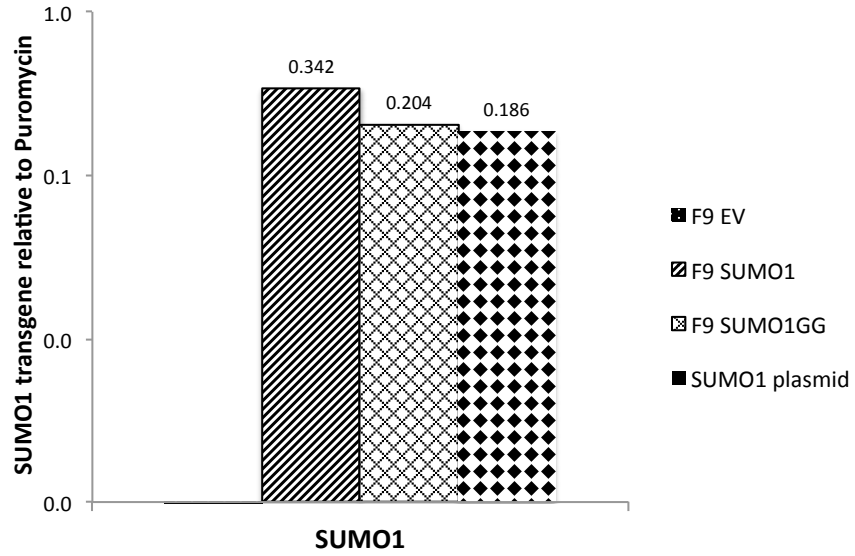
B
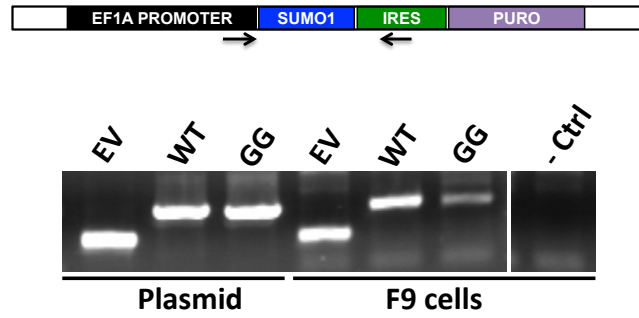
C
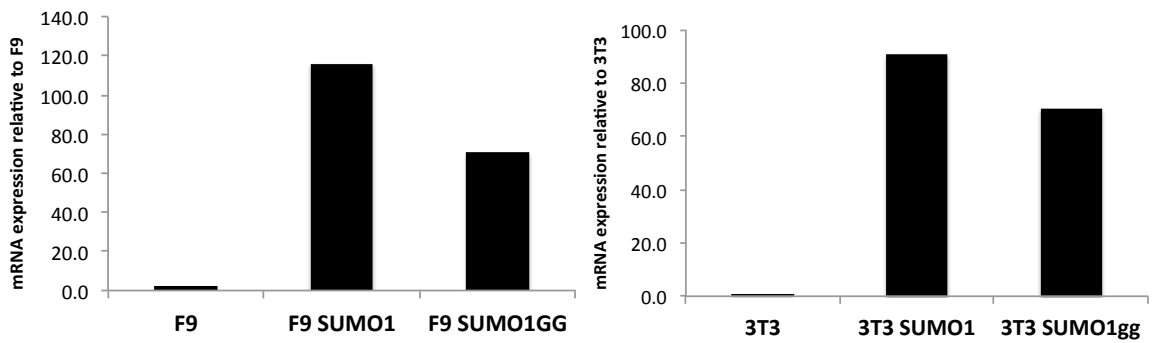
D
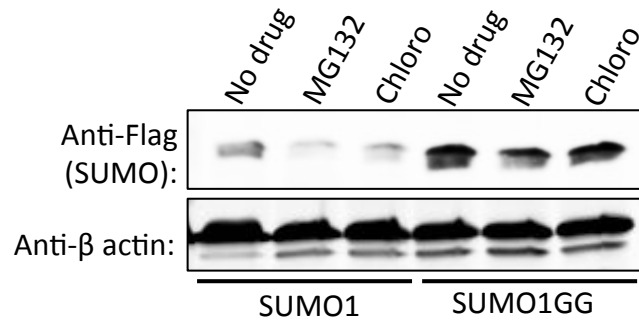
Figure 5-7 Post-transcriptional mechanism for preventing SUMO1 expression in embryonic cells

(A) Genomic DNA was isolated from F9 cells transduced with the SUMO1 or SUMO1GG vectors or the empty vector. SUMO1 plasmid DNA was used as a positive control. SUMO1 and SUMO1GG transgene levels were measured by qPCR and were quantified relative to the puromycin resistance gene levels. (B) Primer locations used for PCR are indicated on a schematic of the SUMO1 vector (above). An agarose gel showing PCR products amplified from the empty vector, SUMO1, or SUMO1GG plasmid DNAs and DNAs isolated from F9 cells transduced with the empty vector, SUMO1, or SUMO1GG vectors (below). PCR using DNA isolated from untreated F9 cells was included as a negative control. (C) Quantification of exogenous SUMO1 and SUMO1GG mRNA levels in F9 and NIH3T3 cells by qRT-PCR. Expression levels were normalized to GAPDH, and were reported relative to untransduced cells. (D) F9 cells were transduced with the SUMO1 or SUMO1GG vectors and treated with 10 µM MG132 or 100 µM chloroquine for 4 hours, 3 days after transduction. Lysates were prepared and were examined for exogenous SUMO1 and SUMO1GG expression by Western blot.

controls. SUMO1 and SUMO1GG mRNA levels were measured by qRT-PCR using primers overlapping a SUMO1 exon-exon junction. We detected roughly similar levels of exogenous SUMO1 and SUMO1GG transcripts in both F9 cells and in NIH3T3 cells (Figure 5-7C), suggesting that differences in the expression of these two constructs in embryonic cells must be occurring at the post-transcriptional level.

One possible mechanism for preventing expression at the post-transcriptional stage is by protein degradation. Protein degradation can occur through the proteasomal or lysosomal pathway, which can be blocked by MG132 and chloroquine, respectively (Shintani 2004; Rock et al. 1994). To determine if either pathway degrades SUMO1, we transduced F9 cells with the SUMO1 or SUMO1GG vectors, and 72 hours later treated cells with MG132 or chloroquine for 4 hours. Cell lysates were prepared and analyzed for SUMO1 expression by Western blot. We found that SUMO1 expression levels did not improve after 4 hours of treatment with either drug (Figure 5-7D), and prolonging drug treatment to 8, 12, or 24 hours did not make a difference in the SUMO1 expression levels (data not shown). By 24 hours of drug treatment, we observed some cell death in the chloroquine and MG132 treated cells, indicating that longer time points would not be feasible or result in reliable measurements. These results suggest that the overexpression of SUMO1 is not likely regulated by the proteasomal and lysosomal pathways.

SUMO2 can be overexpressed in embryonic and differentiated cells

To determine if SUMO2 is also poorly expressed in embryonic cells, we cloned Flag-tagged SUMO2 and SUMO2GG into the same pLVX lentiviral vector used for producing the SUMO1 constructs. F9, E14, and 293T cells were transduced with the SUMO2 or

SUMO2GG vectors, selected for stable expression of the drug resistance gene, and cell

lysates were prepared and examined for SUMO2 and SUMO2GG expression by Western

blot. F9 cells expressed SUMO2 and SUMO2GG proteins at roughly similar levels, while

SUMO1 protein expression levels were dramatically lower than that for SUMO1GG, as seen

before (Figure 5-8A, left blot). E14 cells also expressed SUMO2 and SUMO2GG at similar

levels, while SUMO1 was poorly expressed relative to SUMO1GG (Figure 5-8A, right blot).

293T cells transduced with the SUMO2 and SUMO2GG vectors expressed both constructs

well and at similar levels to SUMO1GG, as expected (Figure 5-8B). These results indicate

that poor SUMO expression in embryonic cells occurs specifically for SUMO1, and not for

all the SUMO family members. While we did not test for the overexpression profile of

SUMO3, we suspect that it would reflect the same trends we demonstrated for SUMO2 given

their high similarity.

Reducing SUMO1 conjugation activity improves SUMO1 expression

SUMO1GG is missing the last six amino acids of SUMO1, which includes the di-

glycine residues necessary for conjugation to substrates. Although unlikely, it was possible

that the last six amino acids are toxic or are a target for post-transcriptional regulation. To

rule out this possibility, we created a construct with the full-length Flag-tagged SUMO1

containing alanine substitutions at the di-glycine residues (SUMO1AA). F9 cells were

transduced with the SUMO1 or SUMO1AA vector, and lysates were prepared and examined

for SUMO expression by Western blot. As we had seen with SUMO1GG, SUMO1AA also

expressed well relative to the poor overexpression of SUMO1 in embryonic cells (Figure

5-9A). This result demonstrates that the last few amino acids do not negatively impact

A
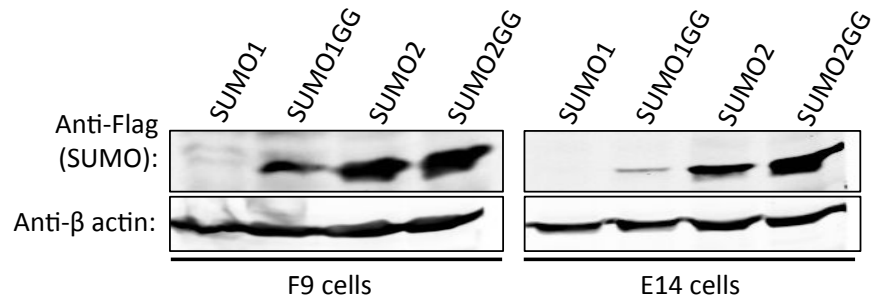
B
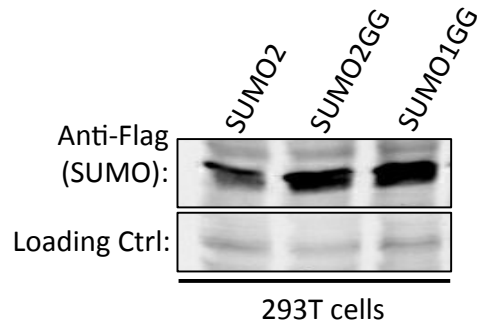
Figure 5-8 SUMO2 overexpression in embryonic and differentiated cells
(A) Western blot displaying expression levels of SUMO1, SUMO1GG, SUMO2, and
SUMO2GG in F9 and E14 cells transduced with and selected for stable expression of the
SUMO vectors. (B) Western blot displaying expression levels of SUMO2, SUMO2GG, and
SUMO1GG in 293T cells transduced with and selected for stable expression of the SUMO
vectors.

106

A
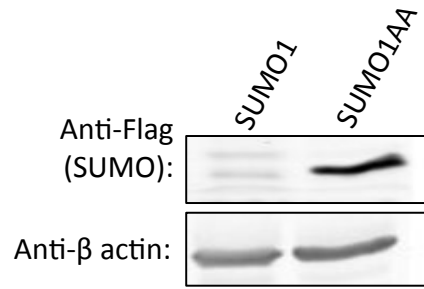
B
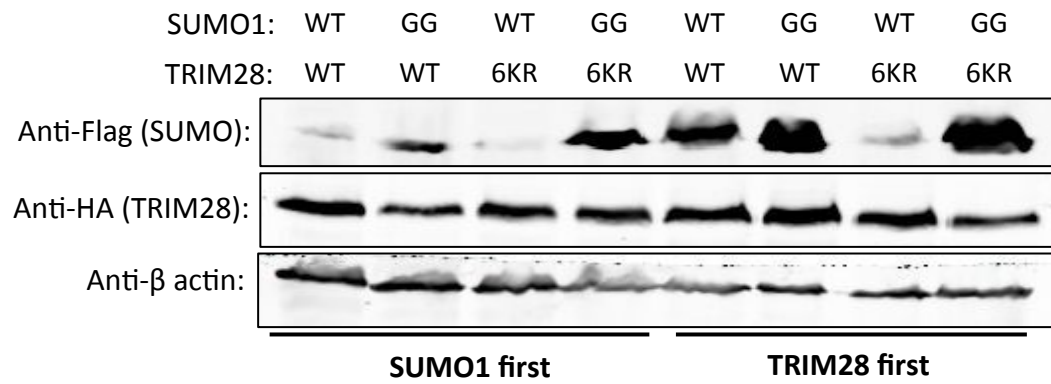
C
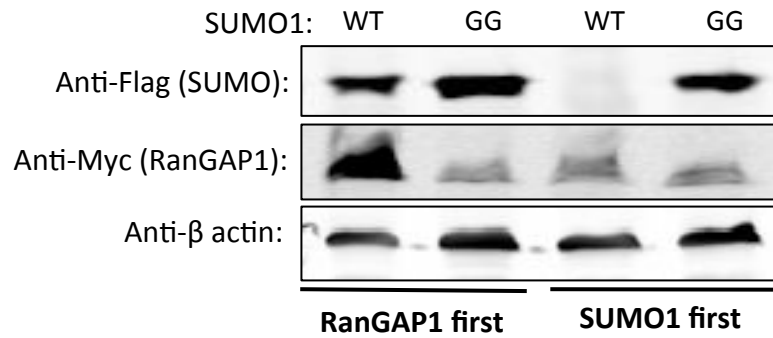
D
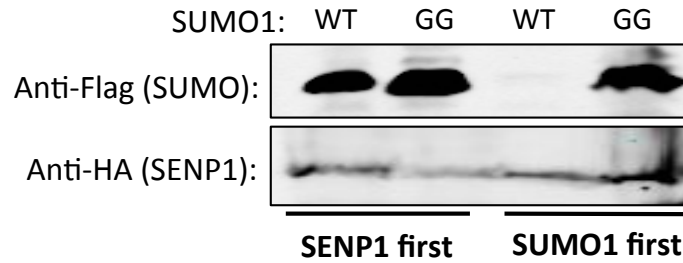
Figure 5-9 Blocking or reducing sumoylation improves SUMO1 overexpression in embryonic cells

(A) Western blot displaying expression levels of SUMO constructs in F9 cells transduced with and selected for stable expression of the Flag-SUMO1 or Flag-SUMO1AA vectors by drug selection. (B) Western blot displaying expression levels SUMO constructs in F9 cells co-expressing Flag-SUMO1 (WT) or Flag-SUMO1GG (GG) and HA-Trim28 (WT) or HA-Trim28^{6KR} (6KR). The first four lanes show expression levels detected in cells transduced with the SUMO vectors first, and the last four lanes show expression levels detected in cells transduced with the Trim28 vectors first. (C) Western blot analysis of F9 cells co-expressing Flag-SUMO1 or Flag-SUMO1GG and HA-SENP1. The first two lanes show expression levels detected in cells transduced with the HA-SENP1 vector first, and the last two lanes show expression levels detected in cells transduced with the SUMO1 vectors first.

SUMO1 overexpression, and suggests it is the conjugation of SUMO1 to substrates in embryonic cells that causes SUMO1 overexpression to be restricted.

If elevated SUMO1-modified substrates are preventing the overexpression of SUMO1, then reducing levels of some SUMO1-modified substrates should allow SUMO1 to be overexpressed. To achieve the reduction of SUMO1 conjugation of endogenous substrates, one approach we tested was to co-express SUMO1 with a substrate that could act as a scaffold or "sponge" for exogenously expressed SUMO1. Tripartite motif-containing 28 (Trim28) is a ubiquitously expressed transcriptional repressor protein that is abundantly expressed in embryonic cells and displays six lysine residues that can be conjugated by SUMO (Ivanov et al. 2007). HA-tagged Trim28 was cloned into a pLVX lentiviral vector containing the neomycin resistance gene and was co-expressed with the Flag-tagged SUMO vectors containing the puromycin resistance gene. We also created a Trim28 construct in which the six lysine residues were mutated to arginine, rendering the mutant unavailable for SUMO conjugation (Trim28^{6KR}). F9 cell lines were prepared in which the SUMO construct was introduced first, followed by the Trim28 construct. Other cell lines were prepared in the opposite order in which the Trim28 construct was introduced first, followed by the SUMO construct. Cell lysates were prepared from these cells and SUMO1 expression levels were assessed by Western blot. Cells transduced with the SUMO vectors first, before transduction with the Trim28 vectors, continued to show weak expression of SUMO1 in the presence of exogenous Trim28^{WT} or Trim28^{6KR} proteins (Figure 5-9B). As expected, SUMO1GG was well expressed in the presence of either Trim28 protein. However, cells transduced with the Trim28 vectors before transduction with the SUMO vectors yielded different results. Here, the SUMO1 overexpression was greatly improved in the background of Trim28^{WT}

overexpression, whereas SUMO1 overexpression did not improve in the background of Trim28^{6KR} expression (Figure 5-9B). Thus, the availability of an overexpressed substrate for sumoylation relieved the lack of expression and allowed for much higher SUMO1 expression, and the identical substrate lacking the lysines for SUMO addition did not do so. This result indicates that SUMO1 conjugation to endogenous substrates is a likely cause for the lack of SUMO1 overexpression in embryonic cells and that embryonic cells can accumulate free SUMO1 protein with no obvious consequences, so long as it is not conjugated to endogenous substrates. An important point here was that SUMO1 overexpression was only improved in embryonic cells when the SUMO sponge was present at the time of SUMO1 transduction, but not after cells had already been selected for SUMO in the absence of the SUMO sponge. This suggests that the mechanism for maintaining low SUMO1-conjugated substrate levels in embryonic cells may not be easily reversed.

We wanted to confirm that the rescue of SUMO1 expression in the presence of high levels of Trim28 was not due to the impact of Trim28 itself on embryonic cells, but instead resulted from its ability to sequester SUMO1 away from endogenous substrates. To evaluate this possibility, we repeated the previous experiments using a different SUMO sponge. Ran GTPase-activating protein 1 (RanGAP1) is a trafficking protein that is commonly used in SUMO1 studies, as one of the earliest proteins found to be a target of SUMO1 conjugation (Matunis 1996). Myc-tagged RanGAP1 was cloned in the pLVX lentiviral vector containing the neomycin resistance gene and was co-expressed with the Flag-SUMO vectors containing the puromycin resistance genes. As previously described for the Trim28 constructs, RanGAP1 was overexpressed in F9 cells before or after transduction with the SUMO constructs. Lysates were prepared from these cells, and Flag-SUMO1 expression levels were

examined by Western blot. We found that SUMO1 overexpression markedly improved with the co-expression of RanGAP1, but only when RanGAP1 was overexpressed prior to SUMO1 overexpression (Figure 5-9C). These results provide further evidence to support the notion that SUMO1 conjugation of endogenous substrates, and not the accumulation of free SUMO1 protein, causes the down-regulation of SUMO1 in embryonic cells.

In another approach to reduce the levels of SUMO1 conjugation to substrates, we tested the overexpression of SENP1. SENP1 is a protease that deconjugates SUMO1 from its substrates. The overexpression of SENP1 could potentially improve SUMO1 overexpression as the SUMO sponges had done. A cDNA encoding an HA-tagged SENP1 was cloned into the pLVX vector containing the neomycin resistance gene and was co-expressed with the Flag-SUMO vectors containing the puromycin resistance genes. As previously described for Trim28 and RanGAP1, SENP1 was overexpressed in F9 cells before or after transduction with the SUMO constructs. Lysates were prepared and were analyzed for SUMO1 expression by Western blot. We found that SENP1 overexpression prior to transduction with the SUMO vectors dramatically improved SUMO1 overexpression levels, almost reaching SUMO1GG expression levels (Figure 5-9D). However, SUMO1 overexpression levels did not improve when this order was reversed. Consistent with our previous results, we demonstrate that increasing the deconjugation activity of SUMO1 from substrates allows for SUMO1 to be overexpressed and that these effects are not seen when we increased the deconjugation activity in cells that have already been selected for SUMO.

Collectively, we found that reducing SUMO1 conjugation by mutation of the di-glycine residues, the co-expression of a SUMO sponge, or the overexpression of SENP1 significantly improved SUMO1 overexpression. These results indicate that the post-

111

transcriptional mechanism(s) for reducing SUMO1 overexpression is only elicited upon the accumulation of SUMO1-modified proteins and not from the accumulation of free SUMO1 protein itself. Moreover, this mechanism may be irreversible or have some long-term effects as reducing the levels of SUMO1-modified proteins after this mechanism has been set in place did not improve SUMO1 overexpression.

Supplemental data

Additional attempts to reduce or block SUMO1 conjugation activity

Ubc9 is responsible for conjugating all SUMO proteins to their substrates, so knockdown of Ubc9 would affect all SUMO pathways. To investigate if we could reduce the levels of SUMO1 conjugation to endogenous substrates from this approach, we knocked down Ubc9 by transducing F9 cells with pLKO-1 lentiviral vectors containing a drug resistance gene and shRNAs targeting Ubc9. Cells were subjected to drug selection, but only two out of five Ubc9 KD lines survived the drug treatment. RNA was isolated from surviving cells, Ubc9 mRNA expression levels were measured by qRT-PCR using primers that spanned a Ubc9 exon-exon junction, and expression levels were normalized to GAPDH. Ubc9 expression levels were decreased by 54% and 47% for the two Ubc9 KD lines, relative to expression levels in cells expressing the non-targeting shRNA vector (Figure 5-10). This demonstrates that obtaining strong Ubc9 KD levels in F9 cells may not be feasible. Given that Ubc9 knockout mice are embryonic lethal, it is also likely that F9 cells with greater Ubc9 KD levels did not survive.

Treatment of cells by hydrogen peroxide (H_2O_2) is reported to result in the crosslinking and inactivation of the SAE2 and Ubc9 enzymes (Bossis & Melchior 2006). To determine if H_2O_2 could decrease levels of sumoylated substrates, we treated 293T cells expressing the SUMO1 vector with various concentrations of H_2O_2 (0.01 - 100 mM). Only cells treated with H_2O_2 at the 1 mM concentration demonstrated a detectable decrease in sumoylated protein levels (Figure 5-11A); however, H_2O_2 treatment at this concentration was too toxic to the cells to conduct any further experiments.
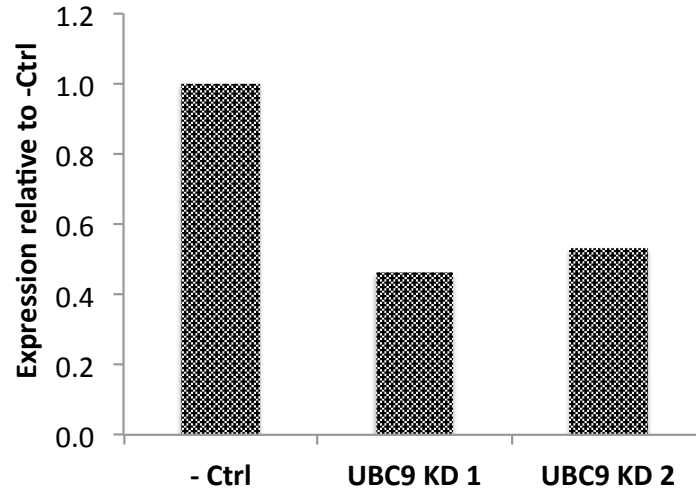
Figure 5-10 Ubc9 mRNA expression levels in Ubc9 KD cells
mRNA expression levels of Ubc9 were measured in F9 cells transduced with the pLKO-1 vectors containing Ubc9 shRNAs, by qRT-PCR. Ubc9 expression was normalized to the housekeeping GAPDH gene and reported relative to the negative control (-Ctrl) cells, which expressed a non-targeting shRNA.

A
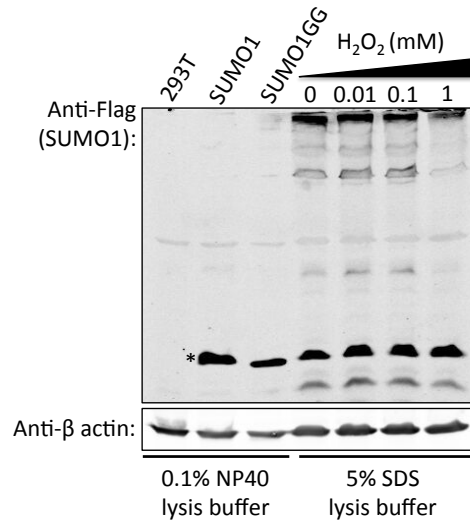
B
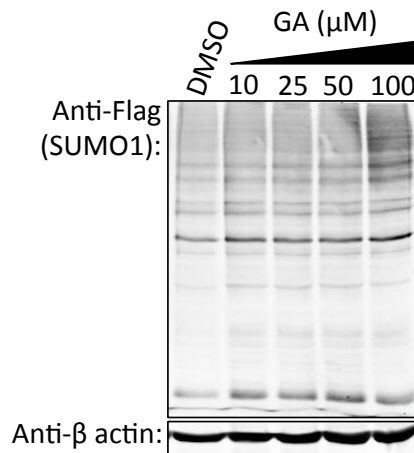
Figure 5-11 Pharmacologic agents for inhibiting SUMO conjugation
(A) 293T cells expressing the SUMO1 vector were treated with 10 μM – 100 mM H_2O_2 for one hour. Cells treated at 10 and 100 mM H_2O_2 died after treatment and were discarded. Lysates were prepared for the remaining cells using the SDS lysis buffer, which prevents the deconjugation of SUMO from substrates. As a control, lysates were also prepared from untreated 293T cells and 293T cells expressing the SUMO vectors with 0.1% NP40 lysis buffer, which does not prevent the deconjugation of SUMO from substrates. Proteins in lysates were analyzed by Western blot probing for the Flag-tag. (B) 293T cells expressing the SUMO1 vector were treated with 10 μM – 100 μM ginkgolic acid for 4 hours. Lysates were prepared using the SDS lysis buffer and analyzed by Western blot probing for the Flag-tag.

115

Ginkgolic acid treatment is shown to directly bind the E1 enzyme (SAE1/SAE2), thereby preventing SUMO conjugation to substrates (Fukuda et al. 2009). We assessed if preventing the SUMO pathways by ginkolic acid treatment could improve SUMO1 overexpression. To determine the appropriate dosage of GA to use in our experiments, 293T cells expressing the Flag-SUMO1 vector were treated with DMSO or GA at various concentrations (10 – 100 µM) for 4 hours. Lysates were prepared using an SDS lysis buffer that inhibits SUMO deconjugation and the levels of SUMO1-modified endogenous substrates were examined by Western blot probing for the Flag-tag. Fukuda and colleagues demonstrate that endogenous protein sumoylation levels are decreased after 4 hours of 50 µM GA treatment, but we did not detect this in cells treated with 50 or 100 µM GA treatment relative to the DMSO treated cells (Figure 5-11B). Prolonged GA treatment before or after F9 cells were transduced with the SUMO1 or SUMO1GG vectors did not improve SUMO1 overexpression levels relative to the untreated cells (data not shown).

SENP1 cannot be effectively knocked down in embryonic cells

To determine if SENP1 KD could have the opposite effect of SENP1 overexpression, we transduced F9 cells with pLKO-1 lentiviral vectors containing shRNAs targeting SENP1. Cells transduced with a pLKO.1 vector containing a non-target shRNA control were used as a negative control. To measure SENP1 expression levels in the various SENP1 KD lines, we measured SENP1 mRNA levels by qRT-PCR using primers that span an SENP1 exon-exon junction. SENP1 levels were normalized to the housekeeping gene GAPDH and are reported relative to cells expressing a non-targeting shRNA vector (Figure 5-12). We were not able to get a significant reduction in SENP1 mRNA levels, with any of the five shRNA constructs

116

used. Given that SENP1 knockout mice are embryonic lethal, this result suggests that SENP1 may be essential for the survival of embryonic cells and possibly the reason for why we were unsuccessful in knocking down SENP1 mRNA levels in F9 cells.
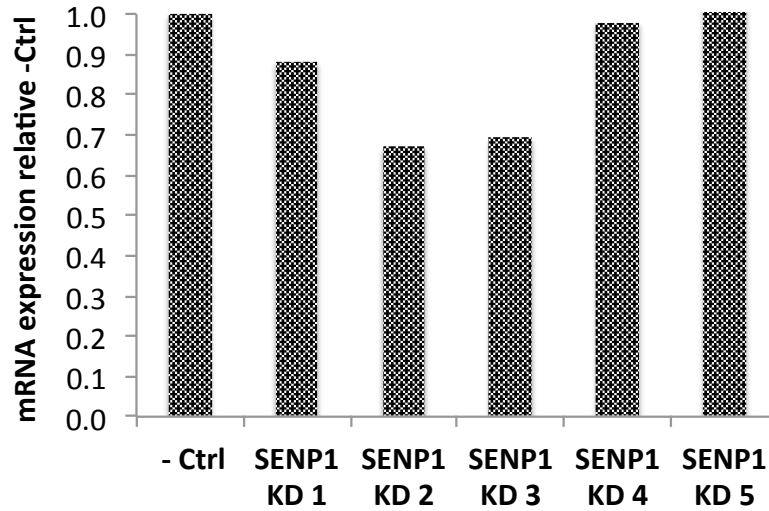
Figure 5-12 SENP1 knockdown is highly inefficient in embryonic cells
mRNA levels of SENP1 were measured in F9 cells transduced with the pLKO-1 vectors
containing SENP1 shRNAs, by qRT-PCR. SENP1 expression was normalized to the
housekeeping GAPDH gene and reported relative to the negative control (-Ctrl) cells, which
expressed a non-targeting shRNA.

Chapter 6 : Discussion

Trim28 and YY1 in Silencing Proviral DNA of M-MLV

Trim28 is a key repressor protein in the complex that mediates transcriptional silencing of the M-MLV provirus (Wolf & Goff 2007; Rowe et al. 2010). M-MLV silencing specifically occurs in mouse embryonic cells although Trim28 is ubiquitously expressed across all developmental stages, raising the question of how Trim28-mediated silencing of M-MLV is regulated. The identification of the zinc-finger proteins, ZFP809 and YY1, responsible for recruiting Trim28 to the provirus, provided some clues. The full length ZFP809 protein can only be expressed in embryonic cells, such that the PBS-dependent silencing mechanism is absent in differentiated cells by limitation of ZFP809 expression (Wolf & Goff 2009). Our lab has recently elucidated the mechanism for this, showing that ZFP809 protein is stable in embryonic cells but is ubiquitinated and rapidly degraded in differentiated cells (Cheng Wang, unpublished data). The selectivity of YY1 activity seems to result from a different mechanism. YY1 is expressed well in both embryonic and differentiated cells, but the YY1-Trim28 interaction is detected preferentially in embryonic cells (Schlesinger et al. 2013), suggesting that this interaction is regulated. Sumoylation and phosphorylation modifications on Trim28 were found to regulate Trim28 interactions and repressive activity in the context of the DNA damage response, cell cycle progression, and HCMV latency (Ivanov et al. 2007; Goodarzi et al. 2011; Bolderson et al. 2012; Chang et al. 2008; Li et al. 2007; Benjamin Rauwel 2015). In Chapter 3, we have shed light on the YY1-Trim28 interaction by mapping out the interacting domains of both proteins and investigating possible explanations for the regulation of this interaction. Furthermore, we determined that

the K779 residue on Trim28 is necessary for mediating M-MLV silencing. Based on other studies of Trim28 functions, this residue is likely the site of modification by SUMO addition.

Several functional domains of Trim28 are known to interact with critical co-repressor proteins. Trim28 contains an RBCC domain, chromoshadow domain, PHD domain, and bromodomain. The chromoshadow domain, also called the HP1 box, interacts with HP1 (B Le Douarin 1996), the PHD-Bromodomain interacts with CHD3 and ESET (Schultz et al. 2001; Schultz 2002), and the RBCC domain interacts with the repressive KRAB domain of KRAB-ZNFs, such as ZFP809 (Friedman et al. 1996; Wolf & Goff 2009). Our findings demonstrated that the RBCC domain of Trim28 is sufficient for interaction with YY1, despite YY1 not containing a KRAB domain (Figure 3-2B). The YY1 interaction with the RBCC domain of Trim28 may involve other KRAB domain-containing proteins that serve as bridging proteins. This notion is consistent with our inability to detect a direct interaction between YY1 and Trim28 (Figure 3-1).

A particular feature of YY1 is that it can behave as a transcriptional activator or repressor, depending on the context and the partners available for interaction (Shi et al. 1997). The repressor domains of YY1 have been mapped to the GA-rich region and the ZNFs, while the activating domains have been mapped to the N-terminal domains, AR1 and AR2 (M. J. Thomas & Seto 1999). Thus far, it is known that the GA-rich region of YY1 is required for its interaction with several co-repressor proteins, including the histone deacetylase 1 (HDAC1), enhancer of zeste homolog 2 (EZH2), and the HP1 proteins (W. M. Yang et al. 1996; Morey et al. 2012). Our experiments showed that Trim28 interaction with YY1 depended on the AR1 and ZNFs (Figure 3-4B and C). The involvement of the ZNFs is consistent with what is thought to be the repressive domains of YY1, but the involvement of

AR1 was surprising. As the AR domains of YY1 are responsible for its transcriptional activator activity, the necessity for this region within a repressor complex was unexpected (Bushmeyer et al. 1995). Furthermore, we consistently detected a weaker interaction between exogenous YY1^{WT} and Trim28 relative to the interaction detected between endogenous YY1 and Trim28. Moving the myc-tag, knocking down endogenous YY1, or overexpressing Trim28 did not improve this interaction. We are currently attempting to improve the YY1-Trim28 interaction by increasing ectopic YY1^{WT} expression levels. However, a continued inability to improve this interaction may suggest that exogenous YY1 is not as suitable as endogenous YY1 for forming the silencing complex needed for the YY1-Trim28 interaction to occur.

One possible complication in our studies is the presence of YY2, an isotype of YY1. YY2 cross-reacts with YY1 antibodies, binds to some of the same sequences that YY1 occupies, and is 86.4% identical to YY1 in its zinc fingers (Klar & Bode 2005; Klar 2010; Chen et al. 2010; Nguyen et al. 2004). Thus, what was previously believed to be the silencing activities of YY1, alone, may actually be representative of the combined repressive activities of YY1 and YY2. It is also possible that YY2 interacts with Trim28 and is immunoprecipitated with the anti-YY1 antibody. However, YY2 does not contain the AR1 domain demonstrated to be necessary for Trim28 interaction, making this less likely.

Several studies have demonstrated the importance of sumoylation and phosphorylation modifications for Trim28 interaction and repressive activity (Ivanov et al. 2007; Goodarzi et al. 2011; Bolderson et al. 2012; Chang et al. 2008; Li et al. 2007; Benjamin Rauwel 2015). We investigated the need for the sites of these modifications in the context of silencing M-MLV in embryonic cells. None of the sites tested were found to

be responsible for regulating the YY1-Trim28 interaction. Thus, the basis for the regulated

interaction between these proteins remains unknown. Functional tests, however, show that

the K779 residue was necessary for Trim28-mediated silencing of M-MLV (Figure 3-5D). A

study using a yeast-two hybrid assay previously demonstrated that the K779R mutation does

not perturb binding of the Trim28 PHD/Bromo domain to CHD3 or ESET, but mutations at

both K779 and K676 eliminates CHD3 and ESET binding (Ivanov et al. 2007). We suspect

that the single K779R mutation may be capable of disrupting its interaction with ESET,

CHD3, or both, in our settings, as we found this mutation to greatly reduced Trim28

silencing of M-MLV. In addition to the K779 residue, previous studies have highlighted the

importance of the K554 and K804 residues for the repressive activity of Trim28 in the DNA

damage response pathway and in the regulation of cell cycle genes (Y.-K. Lee et al. 2007;

Goodarzi et al. 2011; Mascle et al. 2007). In contrast to what has been reported in these

settings, the K554 and K804 residues were not found to be critical for the silencing of M-

MLV when mutated independently. It is still possible that these sites have redundant

repressive functions, thereby requiring the combined mutation of multiple residues to disrupt

silencing activity. Another difference between previous studies and our study is that Trim28

phosphomimetic mutations at S473 or S824 had no impact on Trim28 silencing of M-MLV

(Figure 3-7D), whereas phosphomimetic mutations at these sites abrogates Trim28 repressive

activity in the context of HCMV latency, the DNA damage response, and the expression of

cell cycle genes (Goodarzi et al. 2011; Chang et al. 2008; Benjamin Rauwel 2015). Notably,

Trim28^{S473A} did not express well, indicating the importance of this single residue for proper

expression. A caveat of these results is that our functional assay only tested for short-term M-

MLV silencing, as expression of the M-MLV-GFP reporter is measured 48 hours after

infection. Thus, it is possible that mutations at the K554, K804, S473, or S824 residues impact long-term silencing, but short-term silencing of M-MLV is only dependent on the K779 residue. The findings highlight differences in the requirement for sumoylation modification on Trim28 in the context of differentiated cells, as opposed to the context of embryonic cells.

We have also investigated which of the SUMO proteins can conjugate to Trim28 at the six lysine residues examined. The lysine residues on Trim28 can be modified by any of several modifications that target lysine residues (Gill 2004; Enchev et al. 2014), but evidence from previous studies pointed to SUMO1 and SUMO2 as the most probable options for mediating Trim28 repressive activity (Ivanov et al. 2007; Goodarzi et al. 2008; Goodarzi et al. 2010; Goodarzi et al. 2011; Liangli Wang 2014; B. X. Yang et al. 2015). Most studies addressed the role of SUMO1 in Trim28-mediated repression, and specifically tested for activity in differentiated cells and in the context of the DNA damage or cell cycle pathways. Very few studies have explored the role of Trim28 modification by SUMO2. A recent genomic-wide siRNA screen showed that SUMO2 and its associated sumoylation factors are necessary for silencing of endogenous retroviruses in embryonic cells (B. X. Yang et al. 2015). Moreover, this study demonstrates that the sets of endogenous retroviral genes upregulated from Trim28 KD and SUMO2 KD greatly overlap, implicating the two proteins in a co-dependent repressive function. We showed that both SUMO1 and SUMO2 can be conjugated to Trim28 and that mutation of the six lysine residues prevented conjugation by both SUMO1 and SUMO2 (Figure 3-6A and B). It is currently unknown if all SUMOs, one of the SUMOs, or possibly another modification mediates Trim28 silencing activity at the

K779 residue. It will be important to distinguish between these possibilities in future studies of the Trim28 silencing activity.

In conclusion, we have mapped the interacting domains on Trim28 and YY1 and demonstrated the necessity of the K779 residue on Trim28 for mediating M-MLV silencing. We demonstrated that the RBCC domain of Trim28 is sufficient for its interaction with YY1 and that the AR1 and the zinc fingers of YY1 were necessary and sufficient for its interaction with Trim28. Moreover, our *in vitro* experiments suggest that this interaction occurs indirectly. The K779 residue on Trim28 was shown to be necessary for its repression of M-MLV, but the other sumoylation and phosphorylation sites were dispensable. Thus, our results suggest that silencing of M-MLV in embryonic cells utilizes a Trim28-mediated silencing mechanism different from what has been described in studies of DNA damage, cell cycle progression, and HCMV latency. Further studies aimed at elucidating the consequences of Trim28 modifications at the various residues in different pathways and cell types will be important for developing a more comprehensive understanding of how Trim28 silencing activity is regulated.

Large RBS Complex

A major target of M-MLV repression is a 17-nucleotide sequence, contained in the 18-nucleotide primer binding site (PBS) element (Barklis et al. 1986; Loh et al. 1987). Protein-DNA binding assays with embryonic nuclear extracts demonstrate the presence of a complex binding to DNA probes containing the PBS element (Loh et al. 1990; Petersen et al. 1991; Wolf & Goff 2007). This complex is also referred to as the repressor binding site (RBS) complex and was isolated by Daniel Wolf and colleagues. The PBS probe designed by

Wolf to detect the RBS complex was 28 nucleotides long, including the 17-nucleotide

repressive sequence contained in the PBS and 11 nucleotides found downstream of the PBS

(Wolf & Goff 2007). However, we found a putative YY1 binding site in the 5 nucleotides

upstream of the PBS sequence (5'-CATTT-3'), and we refer this site as the "BS2." If YY1

binds to this site, it would potentially not have been detected in past protein-DNA binding

studies using the original PBS probe. To investigate if YY1 protein and the RBS complex

could bind together to a probe containing the BS2 and PBS sequences, we designed probes

which contained the PBS sequence and the 11 nucleotides found upstream of it. We called

this probe the BS2 + PBS (Figure 4-1A). EMSA experiments in which we tested for protein

binding to the BS2 + PBS probe demonstrated a large molecular weight complex interacting

with the BS2 + PBS sequence that migrated slower than the conventional RBS complex by

native gel electrophoresis (Figure 4-1B). A probe containing the BS2 and the mutant PBS_{B2}

sequences (BS2 + B2) or using a probe containing only the BS2 site did not bind this large

molecular weight complex, demonstrating its specificity for the repressive wild-type PBS

sequence. We refer to this slower-migrating complex as the "larger RBS complex."

Surprisingly, a BS2 + PBS probe with a scrambled BS2 sequence was still capable of shifting

with the larger RBS complex, demonstrating that its binding to the BS2 + PBS probe was not

dependent on the identity of the BS2 sequence – merely on the presence of extra base pairs.

We also found that the larger complex contained Trim28, but we could not detect YY1 in our

assays (Figure 4-2). These experiments suggest that there are likely to be additional factors

that associate with the RBS complex and that these factors require extra nucleotides upstream

of the PBS to support its binding.

Although we could not determine if the larger RBS complex contained YY1, the detection of the larger RBS complex presented an opportunity to search for additional co-repressor proteins that may also participate in the silencing of M-MLV. To identify these proteins, we used two methods for purifying the larger RBS complex. The first method was to purify protein complexes in F9 nuclear extracts by ammonium sulfate precipitation, the method used by Daniel Wolf and colleagues to purify the RBS complex (Wolf & Goff 2007). Applying this method, we were able to detect the RBS complex in the 40% ammonium sulfate fraction but were unable to detect the larger RBS complex in any fraction (Figure 4-3A). This indicated that precipitation by ammonium sulfate was not suitable for the isolation of the larger RBS complex, possibly because it was more fragile than the RBS complex or disintegrated in the high salt concentrations. The second method we tried was to fractionate protein complexes in F9 nuclear extract by velocity sedimentation through a sucrose gradient. Fractions from these gradients were examined for the presence of the larger RBS complex by EMSA using the BS2 + PBS probe. We detected the RBS and larger RBS complexes in these fractions, with some partial separation achieved by differential sedimentation (Figure 4-3B), demonstrating that separation of protein complexes by a sucrose gradient was an effective method for recovering the larger RBS complex. The fractions that contained the RBS and larger RBS complexes were also found to contain the highest amounts of Trim28, as expected. These results indicate that previous methods used for purifying the RBS complex may have resulted in loss of certain binding factors that are more sensitive to the procedures involved in protein precipitation by ammonium sulfate. The fractions containing the RBS or larger RBS complexes were pooled and used to immunoprecipitate Trim28. Bound proteins were analyzed by mass spectrometry, resulting in

126

a list of 22 and 27 proteins detected in the samples containing the RBS or larger RBS complexes, respectively (Table 1). KHDR1, SK2L2, LC7L3, and ARGL1 were some of the proteins detected that are worth highlighting. KHDR1 was detected in the samples containing the RBS and larger RBS complexes; SK2L2 was detected only in the sample containing the larger RBS complex; and LC7L3 and ARGL1 were only detected in the sample containing the RBS complex.

Studies on KHDR1 and SK2L2 indicate a link between these proteins and retroviruses. KHDR1 (KH domain-containing, RNA-binding, signal transduction-associated protein 1), also known as SAM68, is reported to be involved in the export of HIV-1 RNA from the nucleus (Modem 2005). However, this study describes KHDR1 as a supportive retroviral factor rather than an inhibitory factor. SK2L2 (superkiller viralicidic activity 2-like) is reported to interact with the Rev accessory protein of HIV-1 in a proteomics screen, but its effects on HIV-1 replication have not been examined (Naji et al. 2012). LC7L3 and ARGL1 may also be interesting to investigate further because they are associated with transcriptional activities. LC7L3 (Luc7-like protein 3) is a zinc finger protein found to bind DNA regulatory elements (Shipman et al. 2006), and GO annotation suggests the ARGL1 (Arginine and glutamate-rich protein 1) protein is involved in cellular transcription (Baltz et al. 2012). Further studies investigating these genes and others in our list can serve as a good starting point for identifying additional co-repressor proteins functioning in retroviral repression.

As expected, Trim28 was detected in both samples containing the RBS and large RBS complex, but we did not find other proteins known to interact with Trim28 and silence M-MLV, such as ZFP809. The absence of ZFP809 from our mass spectrometry list may be

explained by the fact that ZFP809 is not very abundant in F9 cells, and a large amount of starting nuclear extract was required to identify ZFP809 (Wolf & Goff 2009). Our fractionation methods could be repeated with a larger quantity of nuclear extract to detect these less abundant proteins.

Thus, we have demonstrated the detection of a larger RBS complex that likely contained the RBS complex and additional binding factors. The larger RBS complex contained Trim28 and specifically interacted with a PBS probe containing extra nucleotides upstream of the PBS sequence. We also demonstrated a method for partially enriching for the larger RBS complex and presented a list of proteins that were detected in association with the complex.

Regulation of SUMO1-modified Substrates in Embryonic Cells

SUMO conjugation regulates a variety of basic cellular functions including chromosome segregation, transcription, and the cell cycle (Hay 2005). Not surprisingly, the SUMO pathways are necessary for cell viability in yeast, nematodes, and high eukaryotic cells (Hay 2005) and for embryonic viability of mammals as demonstrated by knockout studies of the SUMO components in mice (Sharma et al. 2013; Liangli Wang 2014; Nacerddine et al. 2005). In Chapter 5, we provide evidence for a post-transcriptional mechanism that prevents the accumulation of SUMO1-modified substrates in embryonic cells. This mechanism acts specifically against the overexpression of SUMO1 and not SUMO2/3, and some effects of this mechanism appear to be irreversible by the methods we explored.

SUMO1 can be effectively overexpressed in differentiated cells (Fukuda et al. 2009; Bossis & Melchior 2006), and we suspected that SUMO1 could be overexpressed just as well in embryonic cells. Unexpectedly, we found that SUMO1 overexpression was highly inefficient in both embryonic carcinoma and embryonic stem cells (Figure 5-1A and B), whereas SUMO1 was overexpressed well in differentiated cells, as other studies had previously shown. One possible explanation for the lack of detectable ectopic SUMO1 expression in embryonic cells was that the overexpression of SUMO1 was toxic or resulted in greater cell death in embryonic cells, and cells that survived transduction and stable selection of the SUMO1 vector had a unique ability to maintain low levels of SUMO1. In an assay for survival of embryonic cells carrying the SUMO vectors, we counted roughly similar number of cells surviving transduction with and selection for the SUMO1 vector as we did for cells transduced with and selected for an equivalent quantity of the SUMO1GG vector (Figure 5-3). This suggested that the lack of SUMO1 overexpression in embryonic cells was not due to greater cell death of SUMO1 overexpressing cells, but rather, the result of an embryonic-specific response to elevated levels of SUMO1 protein or SUMO1 conjugation.

Endogenous SUMO1 is predominantly found conjugated to substrates and not as "free" SUMO1 (Saitoh 2000), and it was possible that either elevated levels of free SUMO1 or of SUMO1-modified proteins was highly unfavorable or selected against in embryonic cells. We found that SUMO1 mutants that are unable to conjugate to substrates (SUMO1GG and SUMO1AA) were expressed in embryonic cells just as well as they were expressed in differentiated cells (Figure 5-1A and B, and Figure 5-9A), indicating that the non-conjugatable SUMO1 protein is permitted to accumulate in embryonic cells and that the

functional activity of SUMO1 is what prevents it from being overexpressed in embryonic cells. To assess this further, we blocked or reduced SUMO1 conjugation to endogenous substrates using various approaches. One approach was to co-express SUMO1 with substrates that could act as a "SUMO sponge" for excess SUMO1 protein expression. We chose Trim28 and RanGAP1 as SUMO sponges because Trim28 displays multiple sites that are targets of sumoylation (Ivanov et al. 2007), and RanGAP1 was one of the first substrates found to be conjugated by SUMO1 (Hay 2005). In the background of Trim28 and RanGAP1 overexpression, SUMO1 overexpression greatly improved in embryonic cells (Figure 5-9B and C). This suggested that SUMO1 overexpression in embryonic cells is prevented as a response to or result of the accumulation of unknown endogenous substrates conjugated by SUMO1, and not the overexpression of SUMO1 protein *per se*. In another approach, we overexpressed SUMO1 with SENP1, the enzyme responsible for deconjugating SUMO1 from its substrates (Sharma et al. 2013). Consistent with our previous results, we found that in the background of SENP1 overexpression, SUMO1 overexpression was greatly improved (Figure 5-9D). These results indicate that the accumulation of SUMO1-modified substrates could be prevented by the upregulation of SENP1 deconjugating activity, which allows SUMO1 to be overexpressed in embryonic cells.

An important aspect of these experiments was that SUMO1 overexpression only improved when the SUMO sponge or SENP1 were overexpressed prior to transduction with the SUMO1 vector. Reversing this order did not improve SUMO1 overexpression (Figure 5-9B, C, and D), SUMO1 modification of substrates are reported to have long-term effects even when SUMO1 is no longer conjugated to the substrate (Hay 2005), and perhaps

SUMO1 modification of certain embryonic substrates creates long-term effects that are not immediately reversible by the approaches we utilized.

SUMO1 overexpression can be potentially regulated at any of several stages including at the level of transcription, mRNA processing and stability, translation, or post-translational processing and degradation. Rare events can also occur in which the retroviral vector deletes genes from the viral genome, making it possible that our cells contained the drug resistance marker and not the SUMO1 transgene. Embryonic cells transduced with the SUMO1 or SUMO1GG vectors showed equivalent levels of the SUMO transgenes (Figure 5-7A), suggesting that the inability to overexpress SUMO1 was not due to of rare events in which the SUMO1 gene is deleted from the retroviral vectors. We also detected similar levels of the SUMO transcripts, indicating that the mechanism for reducing SUMO1-modified proteins was acting at some post-transcriptional level (Figure 5-7C). Next, we checked if protein degradation by the proteosomal and lysosomal pathways could be involved by treating embryonic cells transduced with SUMO1 or SUMO1GG vectors with inhibitors of both pathways (Figure 5-7D). However, we did not find evidence to support the involvement of these pathways. Other possibilities include a mechanism that regulates SUMO1 mRNA processing and stability or SUMO1 translation, and additional studies investigating these possibilities will be required for understanding how embryonic cells regulate the steady-state levels of SUMO1-modified proteins.

Another important aspect of this mechanism is that it was specific to SUMO1 and not to the other SUMO family members. SUMO2 was overexpressed in both embryonic and differentiated cells as efficiently as SUMO2GG and SUMO1GG (Figure 5-8). This result highlights a notable difference between SUMO1 and SUMO2/3 that specifically occurs in

embryonic cells. SUMO2 may not conjugate to the same substrates that lead to SUMO1 down-regulation, or SUMO2 conjugation of these substrates leads to different consequences. One SUMO1 knockout study raises the possibility that the reason SUMO1 is dispensable for normal mouse development is because its loss is compensated by SUMO2/3 (Zhang et al. 2008). While this is possible for certain substrate targets, our results indicate important differences for these modifications in embryonic cells; we are seeing distinct differences in the response to SUMO1 versus SUMO2 conjugation. There may be differences in the bulk extent of conjugation of SUMO1 versus SUMO2. Because endogenous SUMO1 is normally found conjugated to substrates, it is possible that the overexpression of SUMO1 is prevented because it is more actively conjugated to substrates whereas the overexpression of SUMO2 might accumulate more readily as free SUMO2 proteins.

In summary, we provide evidence for a post-transcriptional mechanism that acts to maintain low steady-state levels of SUMO1-modified proteins. Our results are consistent with a previous studies in which increases of SUMO1-modified substrate levels by SENP1 knockout was embryonic lethal, but can be rescued from genetically lowering SUMO1 levels (Zhang et al. 2008; Sharma et al. 2013). In Chapter 5, we demonstrate the importance of regulating steady-state levels of SUMO1-modified proteins at the cellular level. We provide evidence for an embryonic-specific post-transcriptional mechanism that is elicited from the increase of SUMO1-modified proteins and results in the downregulation of SUMO1 overexpression. Identifying the SUMO1-modified substrate(s) that lead(s) to SUMO1 down-regulation and determining the stage at which SUMO1 overexpression is prevented will be critical for further understanding this mechanism and the function of SUMO1 in embryonic cells.

Concluding remarks

We have investigated two embryonic-specific mechanisms in these studies. In Chapter 3 and 4, we examined transcriptional silencing factors involved in the repression of M-MLV expression. We mapped the domains necessary for interaction between YY1 and Trim28 and we demonstrate the necessity of the K779 residue on Trim28 to mediate repression of M-MLV. Furthermore, we demonstrate the detection and enrichment of a larger RBS complex and provide a list of proteins identified within this complex. Elucidating the impact of Trim28 modification at the K779 residue and identification of additional repressor cofactors will be important for developing a more detailed understanding of the silencing mechanism acting on M-MLV.

In Chapter 5, we investigated a previously unreported mechanism for regulating the levels of SUMO1-modified substrates in embryonic cells. We demonstrated that the accumulation of SUMO1-modified substrates down-regulated the expression of SUMO1 in a post-transcriptional manner, and this occurred specifically for SUMO1 but not SUMO2. This mechanism has been largely unexplored, and there are several gaps that remain to be filled. Determining the post-transcriptional stage in which SUMO1 is downregulated and identifying potential embryonic substrates that may be involved in the response leading to the downregulation of SUMO1 expression will be fundamental to understanding this mechanism.

References

Albritton, L.M. et al., 1989. A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. *Cell*, 57(4), pp.659–666.

Allouch, A. et al., 2011. The TRIM Family Protein KAP1 Inhibits HIV-1 Integration. *Cell Host & Microbe*, 9(6), pp.484–495.

B Le Douarin, A.L.N.J.M.G.H.I.F.J.R.L.P.C., 1996. A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *The EMBO journal*, 15(23), p.6701.

Baltimore, D., 1970. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature*, 226(5252), pp.1209–1211.

Baltz, A.G. et al., 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular Cell*, 46(5), pp.674–690.

Barklis, E., Mulligan, R.C. & Jaenisch, R., 1986. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell*, 47(3), pp.391–399.

Barre-Sinoussi, F. et al., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220(4599), pp.868–871.

Battini, J.L., Rasko, J.E. & Miller, A.D., 1999. A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: possible role in G protein-coupled signal transduction. *Proceedings of the National Academy of Sciences of the United States of America*, 96(4), pp.1385–1390.

Benjamin Rauwel, S.M.J.M.C.A.K.I.B.D.T., 2015. Release of human cytomegalovirus from latency by a KAP1/TRIM28 phosphorylation switch. *eLife*, 4.

Bittner, J.J., 1936. Some Possible Effects Of Nursing On The Mammary Gland Tumor Incidence In Mice. *Science*, 84(2172), P.162.

Bohren, K.M. et al., 2004. A M55V Polymorphism in a Novel SUMO Gene (SUMO-4) Differentially Activates Heat Shock Transcription Factors and Is Associated with Susceptibility to Type I Diabetes Mellitus. *Journal of Biological Chemistry*, 279(26), pp.27233–27238.

Bolderson, E. et al., 2012. Kruppel-associated Box (KRAB)-associated Co-repressor (KAP-1) Ser-473 Phosphorylation Regulates Heterochromatin Protein 1 (HP1-) Mobilization and DNA Repair in Heterochromatin. *Journal of Biological Chemistry*, 287(33), pp.28122–28131.

Bossis, G. & Melchior, F., 2006. Regulation of SUMOylation by Reversible Oxidation of SUMO Conjugating Enzymes. *Molecular Cell*, 21(3), pp.349–357.

Bushmeyer, S., Park, K. & Atchison, M.L., 1995. Characterization of functional domains within the multifunctional transcription factor, YY1. *The Journal of biological chemistry*, 270(50), pp.30213–30220.

Cammas, F. et al., 2000. Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. *Development*, 127(13), pp.2955–2963.

Chang, C.-W. et al., 2008. Phosphorylation at Ser473 regulates heterochromatin protein 1 binding and corepressor function of TIF1beta/KAP1. *BMC Molecular Biology*, 9(1), p.61.

Chen, L. et al., 2010. Genome-wide analysis of YY2 versus YY1 target genes. *Nucleic Acids Research*, 38(12), pp.4011–4026.

Chikuma, S. et al., 2012. TRIM28 prevents autoinflammatory T cell development in vivo. *Nature Immunology*, 13(6), pp.596–603.

Chinwalla, A.T. et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*.

Chung, T.L. et al., 2004. In Vitro Modification of Human Centromere Protein CENP-C Fragments by Small Ubiquitin-like Modifier (SUMO) Protein: Definitive Identification Of The Modification Sites By Tandem Mass Spectrometry Analysis Of The Isopeptides. *Journal of Biological Chemistry*, 279(38), pp.39653–39662.

Coffin, J.M. et al, 1997. Retroviruses. Cold Spring Harbor Laboratory Press.

Coull, J.J. et al., 2000. The Human Factors YY1 and LSF Repress the Human Immunodeficiency Virus Type 1 Long Terminal Repeat via Recruitment of Histone Deacetylase 1. *Journal of Virology*, 74(15), pp.6790–6799.

Denslow, S.A. & Wade, P.A., 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene*, 26(37), pp.5433–5438.

Emerson, R.O. & Thomas, J.H., 2009. Adaptive Evolution in Zinc Finger Transcription Factors S. Myers, ed. *PLoS Genetics*, 5(1), p.e1000325.

Enchev, R.I., Schulman, B.A. & Peter, M., 2014. Protein neddylation: beyond cullin–RING ligases. *Nature Reviews Molecular Cell Biology*, 16(1), pp.30–44.

Flanagan, J.R. et al., 1992. Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. *Molecular and Cellular Biology*, 12(1), pp.38–44.

Flanagan, J.R. et al., 1989. Negative control region at the 5' end of murine leukemia virus long terminal repeats. *Molecular and Cellular Biology*, 9(2), pp.739–746.

Flotho, A. & Melchior, F., 2013. Sumoylation: A Regulatory Protein Modification in Health and Disease. *Annual Review of Biochemistry*, 82(1), pp.357–385.

Friedman, J.R. et al., 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & Development*, 10(16), pp.2067–2078.

Fukuda, I. et al., 2009. Ginkgolic acid inhibits protein SUMOylation by blocking formation of the E1-SUMO intermediate. *Chemistry & biology*, 16(2), pp.133–140.

Gao, G., 2002. Inhibition of Retroviral RNA Production by ZAP, a CCCH-Type Zinc Finger Protein. *Science*, 297(5587), pp.1703–1706.

Gautsch, J.W. & Wilson, M.C., 1983. Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. *Nature*, 301(5895), pp.32–37.

Geiss-Friedlander, R. & Melchior, F., 2007. Concepts in sumoylation: a decade on. *Nature Reviews Molecular Cell Biology*, 8(12), pp.947–956.

Gill, G., 2004. SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes & Development*, 18(17), pp.2046–2059.

Goff, S.P., 2007. Host factors exploited by retroviruses. *Nature Reviews Microbiology*, 5(4), pp.253–263.

Goff, S.P., 2001. Intracellular trafficking of retroviral genomes during the early phase of infection: viral exploitation of cellular pathways. *The journal of gene medicine*, 3(6), pp.517–528.

Goff, S.P., 2004. Retrovirus restriction factors. *Molecular Cell*, 16(6), pp.849–859.

Golebiowski, F.M. et al., 2011. Efficient overexpression and purification of active full-length human transcription factor Yin Yang 1 in Escherichia coli. *Protein Expression and Purification*, 77(2), pp.198–206.

Gong, L. & Yeh, E.T.H., 2006. Characterization of a Family of Nucleolar SUMO-specific Proteases with Preference for SUMO-2 or SUMO-3. *Journal of Biological Chemistry*, 281(23), pp.15869–15877.

Goodarzi, A.A. et al., 2008. ATM Signaling Facilitates Repair of DNA Double-Strand Breaks Associated with Heterochromatin. *Molecular Cell*, 31(2), pp.167–177.

Goodarzi, A.A., Jeggo, P. & Lobrich, M., 2010. The influence of heterochromatin on DNA double strand break repair: Getting the strong, silent type to relax. *DNA Repair*, 9(12),

pp.1273–1282.

Goodarzi, A.A., Kurka, T. & Jeggo, P.A., 2011. KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. *Nature Publishing Group*, 18(7), pp.831–839. Available at: http://www.nature.com/doifinder/10.1038/nsmb.2077.

Gordon, S. et al., 2005. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*, 25(8), pp.1125–1142.

Grez, M. et al., 1990. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. *Proceedings of the National Academy of Sciences*, 87(23), pp.9202–9206.

Gross, L., 1957. Development and serial cellfree passage of a highly potent strain of mouse leukemia virus. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, 94(4), pp.767–771.

Gross, L., 1951. "Spontaneous" Leukemia Developing in G3H Mice Following Inoculation, In Infancy, with AK-Emkemic. *Experimental Biology and Medicine*, 76(1), pp.27–32.

H Oppermann, A.D.L.H.E.V.L.L.J.M.B., 1979. Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (src). *Proceedings of the National Academy of Sciences of the United States of America*, 76(4), p.1804.

Hay, R.T., 2005. SUMO. *Molecular Cell*, 18(1), pp.1–12.

Hendriks, I.A. & Vertegaal, A.C.O., 2016. A comprehensive compilation of SUMO proteomics. *Nature Reviews Molecular Cell Biology*, 17(9), pp.581–595.

Hilberg, F. et al., 1987. Functional analysis of a retroviral host-range mutant: altered long terminal repeat sequences allow expression in embryonal carcinoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, 84(15), pp.5232–5236.

Hyde-DeRuyscher, R.P., Jennings, E. & Shenk, T., 1995. DNA binding sites for the transcriptional activator/repressor YY1. *Nucleic Acids Research*, 23(21), pp.4457–4465.

Ivanov, A.V. et al., 2007. PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain Required for Gene Silencing. *Molecular Cell*, 28(5), pp.823–837.

Jeon, Y. & Lee, J.T., 2011. YY1 Tethers Xist RNAto the Inactive X Nucleation Center. *Cell*, 146(1), pp.119–133.

Jiao, R.-Q., Li, G. & Chiu, J.-F., 2012. Comparative proteomic analysis of differentiation of

mouse F9 embryonic carcinoma cells induced by retinoic acid. *Journal of cellular biochemistry*, 113(6), pp.1811–1819.

Kessler, J.D. et al., 2012. A SUMOylation-Dependent Transcriptional Subprogram Is Required for Myc-Driven Tumorigenesis. *Science*, 335(6066), pp.348–353.

Kim, J.W. et al., 1991. Transport of cationic amino acids by the mouse ecotropic retrovirus receptor. *Nature*, 352(6337), pp.725–728.

Klar, M., 2010. It is not necessarily YY1--the frequently forgotten Yin-Yang-2 transcription factor. *Proceedings of the National Academy of Sciences*, 107(52), pp.E190–E190.

Klar, M. & Bode, J., 2005. Enhanceosome Formation over the Beta Interferon Promoter Underlies a Remote-Control Mechanism Mediated by YY1 and YY2. *Molecular and Cellular Biology*, 25(22), pp.10159–10170.

Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860–921.

Lechner, M.S. et al., 2000. Molecular Determinants for Targeting Heterochromatin Protein 1-Mediated Gene Silencing: Direct Chromoshadow Domain-KAP-1 Corepressor Interaction Is Essential. *Molecular and Cellular Biology*, 20(17), pp.6449–6465.

Lee, J.S., Galvin, K.M. & Shi, Y., 1993. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proceedings of the National Academy of Sciences of the United States of America*, 90(13), pp.6145–6149.

Lee, L. et al., 2013. SUMO and Alzheimer's Disease. *NeuroMolecular Medicine*, 15(4), pp.720–736.

Lee, Y.-K. et al., 2007. Doxorubicin down-regulates Kruppel-associated box domain-associated protein 1 sumoylation that relieves its transcription repression on p21WAF1/CIP1 in breast cancer MCF-7 cells. *The Journal of biological chemistry*, 282(3), pp.1595–1606.

Leung, D.C. et al., 2011. Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proceedings of the National Academy of Sciences*, 108(14), pp.5718–5723.

Li, X. et al., 2007. Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in regulating KAP1-mediated transcriptional repression. *The Journal of biological chemistry*, 282(50), pp.36177–36189.

Liang, Q. et al., 2011. Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *Journal of immunology (Baltimore, Md. : 1950)*, 187(9), pp.4754–4763.

Liangli Wang, C.W.S.Z.P.M.W.P.W.Y., 2014. SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. *EMBO reports*, 15(8), pp.878–885.

Lim, D., Orlova, M. & Goff, S.P., 2002. Mutations of the RNase H C Helix of the Moloney Murine Leukemia Virus Reverse Transcriptase Reveal Defects in Polypurine Tract Recognition. *Journal of Virology*, 76(16), pp.8360–8373.

Linney, E. et al., 1984. Non-function of a Moloney murine leukaemia virus regulatory sequence in F9 embryonal carcinoma cells. *Nature*, 308(5958), pp.470–472.

Loh, T.P., Sievert, L.L. & Scott, R.W., 1990. Evidence for a stem cell-specific repressor of Moloney murine leukemia virus expression in embryonal carcinoma cells. *Molecular and Cellular Biology*, 10(8), pp.4045–4057.

Loh, T.P., Sievert, L.L. & Scott, R.W., 1988. Negative regulation of retrovirus expression in embryonal carcinoma cells mediated by an intragenic domain. *Journal of Virology*, 62(11), pp.4086–4095.

Loh, T.P., Sievert, L.L. & Scott, R.W., 1987. Proviral sequences that restrict retroviral expression in mouse embryonal carcinoma cells. *Molecular and Cellular Biology*, 7(10), pp.3775–3784.

Luban, J. et al., 1993. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell*, 73(6), pp.1067–1078.

Maddon, P.J. et al., 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*, 47(3), pp.333–348.

Maison, C. & Almouzni, G., 2004. HP1 and the dynamics of heterochromatin maintenance. *Nature Reviews Molecular Cell Biology*, 5(4), pp.296–305.

Mascle, X.H. et al., 2007. Sumoylation of the Transcriptional Intermediary Factor 1beta (TIF1beta), the Co-repressor of the KRAB Multifinger Proteins, Is Required for Its Transcriptional Activity and Is Modulated by the KRAB Domain. *Journal of Biological Chemistry*, 282(14), pp.10190–10202.

Matsui, T. et al., 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, 464(7290), pp.927–931.

Matunis, M.J., 1996. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *The Journal of cell biology*, 135(6), pp.1457–1470.

Messerschmidt, D.M. et al., 2012. Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. *Science*, 335(6075), pp.1499–1502.

Miller, D.G. & Miller, A.D., 1994. A family of retroviruses that utilize related phosphate

transporters for cell entry. *Journal of Virology*, 68(12), pp.8270–8276.

Modem, S., 2005. Sam68 is absolutely required for Rev function and HIV-1 production. *Nucleic Acids Research*, 33(3), pp.873–879.

Morey, L. et al., 2012. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell stem cell*, 10(1), pp.47–62.

Muller, S., 2000. c-Jun and p53 Activity Is Modulated by SUMO-1 Modification. *Journal of Biological Chemistry*, 275(18), pp.13321–13329.

Nacerddine, K. et al., 2005. The SUMO Pathway Is Essential for Nuclear Integrity and Chromosome Segregation in Mice. *Developmental Cell*, 9(6), pp.769–779.

Naji, S. et al., 2012. Host Cell Interactome of HIV-1 Rev Includes RNA Helicases Involved in Multiple Facets of Virus Production. *Molecular & Cellular Proteomics*, 11(4), pp.M111.015313–M111.015313.

Neil, S.J.D., Zang, T. & Bieniasz, P.D., 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*, 451(7177), pp.425–430.

Nguyen, N. et al., 2004. Molecular Cloning and Functional Characterization of the Transcription Factor YY2. *Journal of Biological Chemistry*, 279(24), pp.25927–25934.

Niwa, O. et al., 1983. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. *Cell*, 32(4), pp.1105–1113.

O'Geen, H. et al., 2007. Genome-Wide Analysis of KAP1 Binding Suggests Autoregulation of KRAB-ZNFs. *PLoS Genetics*, 3(6), p.e89.

Petersen, R., Kempler, G. & Barklis, E., 1991. A stem cell-specific silencer in the primer-binding site of a retrovirus. *Molecular and Cellular Biology*, 11(3), pp.1214–1221.

Quenneville, S. et al., 2012. The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development. *Cell reports*, pp.1–8.

Rock, K.L. et al., 1994. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell*, 78(5), pp.761–771.

Romerio, F., Gabriel, M.N. & Margolis, D.M., 1997. Repression of human immunodeficiency virus type 1 through the novel cooperation of human factors YY1 and LSF. *Journal of Virology*, 71(12), pp.9375–9382.

Rous, P., 1911. A Sarcoma Of The Fowl Transmissible By An Agent Separable From The

Tumor Cells. *The Journal of experimental medicine*, 13(4), pp.397–411.

Rowe, H.M. et al., 2010. KAP1 controls endogenous retroviruses in embryonicstem cells. *Nature*, 463(7278), pp.237–240. Available at: http://www.nature.com/doifinder/10.1038/nature08674.

Saitoh, H., 2000. Functional Heterogeneity of Small Ubiquitin-related Protein Modifiers SUMO-1 versus SUMO-2/3. *Journal of Biological Chemistry*, 275(9), pp.6252–6258.

Schlesinger, S. et al., 2013. Proviral silencing in embryonic cells is regulated by Yin Yang 1. *Cell reports*, 4(1), pp.50–58.

Schultz, D.C., 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development*, 16(8), pp.919–932.

Schultz, D.C., Friedman, J.R. & Rauscher, F.J.3., 2001. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & Development*, 15(4), pp.428–443.

Seki, Y. et al., 2010. TIF1 regulates the pluripotency of embryonic stem cells in a phosphorylation-dependent manner. *Proceedings of the National Academy of Sciences*, 107(24), pp.10926–10931.

Sharma, P. et al., 2013. Senp1 Is Essential for Desumoylating Sumo1-Modified Proteins but Dispensable for Sumo2 and Sumo3 Deconjugation in the Mouse Embryo. *Cell reports*, 3(5), pp.1640–1650.

Sheehy, A.M. et al., 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, 418(6898), pp.646–650.

Shi, Y., Lee, J.S. & Galvin, K.M., 1997. Everything you have ever wanted to know about Yin Yang 1...... *Biochimica et biophysica acta*, 1332(2), pp.F49–66.

Shintani, T., 2004. Autophagy in Health and Disease: A Double-Edged Sword. *Science*, 306(5698), pp.990–995.

Shipman, K.L. et al., 2006. Identification of a family of DNA-binding proteins with homology to RNA splicing factors. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 84(1), pp.9–19.

Stewart, C.L. et al., 1982. De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, 79(13), pp.4098–4102.

Stocking, C. & Kozak, C.A., 2008. Endogenous retroviruses. *Cellular and Molecular Life*

Sciences (CMLS), 65(21), pp.3383–3398.

Stremlau, M. et al., 2004. The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys. *Nature*, 427(6977), pp.848–853.

T Roe, T.C.R.G.Y.P.O.B., 1993. Integration of murine leukemia virus DNA depends on mitosis. *The EMBO journal*, 12(5), p.2099.

Tatham, M.H. et al., 2001. Polymeric Chains of SUMO-2 and SUMO-3 Are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9. *Journal of Biological Chemistry*, 276(38), pp.35368–35374.

Teich, N.M. et al., 1977. Virus infection of murine teratocarcinoma stem cell lines. *Cell*, 12(4), pp.973–982.

Temin, H.M. & Mizutani, S., 1970. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature*, 226(5252), pp.1211–1213.

Temin, H.M. & Rubin, H., 1958. Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture. *Virology*, 6(3), pp.669–688.

Theodore Pincus, W.P.R.F.L., 1971. A Major Genetic Locus Affecting Resistance To Infection With Murine Leukemia Viruses : Ii. Apparent Identity To A Major Locus Described For Resistance To Friend Murine Leukemia Virus. *The Journal of experimental medicine*, 133(6), pp.1234–476.

Thomas, M.J. & Seto, E., 1999. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene*, 236(2), pp.197–208.

Verger, A., Perdomo, J. & Crossley, M., 2003. Modification with SUMO. *EMBO reports*, 4(2), pp.137–142.

Wang, J. et al., 2011. Defective sumoylation pathway directs congenital heart disease. *Birth defects research. Part A, Clinical and molecular teratology*, 91(6), pp.468–476.

Wei, F., Scholer, H.R. & Atchison, M.L., 2007. Sumoylation of Oct4 Enhances Its Stability, DNA Binding, and Transactivation. *Journal of Biological Chemistry*, 282(29), pp.21551–21560.

Willson, V., 2009. *SUMO Regulation of Cellular Processes*, Springer Science & Business Media.

Wolf, D. & Goff, S.P., 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, 458(7242), pp.1201–1204.

Wolf, D. & Goff, S.P., 2008. Host restriction factors blocking retroviral replication. *Annual*

review of genetics, 42, pp.143–163.

Wolf, D. & Goff, S.P., 2007. TRIM28 Mediates Primer Binding Site-Targeted Silencing of Murine Leukemia Virus in Embryonic Cells. *Cell*, 131(1), pp.46–57.

Wolf, D., Cammas, F., et al., 2008. Primer Binding Site-Dependent Restriction of Murine Leukemia Virus Requires HP1 Binding by TRIM28. *Journal of Virology*, 82(9), pp.4675–4679.

Wolf, D., Hug, K. & Goff, S.P., 2008. TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. *Proceedings of the National Academy of Sciences*, 105(34), pp.12521–12526.

Wolfe, S.A., Nekludova, L. & Pabo, C.O., 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics and biomolecular structure*, 29, pp.183–212.

Wu, Y. et al., 2012. SUMOylation Represses Nanog Expression via Modulating Transcription Factors Oct4 and Sox2. *PLoS ONE*, 7(6), p.e39606.

Yang, B.X. et al., 2015. Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell*, 163(1), pp.230–245.

Yang, W.M. et al., 1996. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23), pp.12845–12850.

Zhang, F.P. et al., 2008. Sumo-1 Function Is Dispensable in Normal Mouse Development. *Molecular and Cellular Biology*, 28(17), pp.5381–5390.

Ziv, Y. et al., 2006. Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nature Cell Biology*, 8(8), pp.870–876.

Appendix

TRIM28 Immunoprecipitation in pooled fractions 3-4 (RBS complex)

TOP2A_MOUSE	DNA topoisomerase 2-alpha OS=Mus musculus GN=Top2a PE=1 SV=2
CPSF6_MOUSE	Cleavage and polyadenylation specificity factor subunit 6 OS=Mus musculus GN=Cpsf6 PE=1 SV=1
NONO_MOUSE	Non-POU domain-containing octamer-binding protein OS=Mus musculus GN=Nono PE=1 SV=3
TIF1B_MOUSE	Transcription intermediary factor 1-beta OS=Mus musculus GN=Trim28 PE=1 SV=3
DDX3X_MOUSE	ATP-dependent RNA helicase DDX3X OS=Mus musculus GN=Ddx3x PE=1 SV=3
DDX5_MOUSE	Probable ATP-dependent RNA helicase DDX5 OS=Mus musculus GN=Ddx5 PE=1 SV=2
SFPQ_MOUSE	Splicing factor, proline- and glutamine-rich OS=Mus musculus GN=Sfpq PE=1 SV=1
PARP1_MOUSE	Poly [ADP-ribose] polymerase 1 OS=Mus musculus GN=Parp1 PE=1 SV=3
TOP2B_MOUSE	DNA topoisomerase 2-beta OS=Mus musculus GN=Top2b PE=1 SV=2
CPSF5_MOUSE	Cleavage and polyadenylation specificity factor subunit 5 OS=Mus musculus GN=Nudt21 PE=2 SV=1
TCOF_MOUSE	Treacle protein OS=Mus musculus GN=Tcof1 PE=1 SV=1
DDX17_MOUSE	Probable ATP-dependent RNA helicase DDX17 OS=Mus musculus GN=Ddx17 PE=2 SV=1
LAP2A_MOUSE	Lamina-associated polypeptide 2, isoforms alpha/zeta OS=Mus musculus GN=Tmpo PE=1 SV=4
G3P_MOUSE	Glyceraldehyde-3-phosphate dehydrogenase OS=Mus musculus GN=Gapdh PE=1 SV=2
CPSF7_MOUSE	Cleavage and polyadenylation specificity factor subunit 7 OS=Mus musculus GN=Cpsf7 PE=1 SV=2
PR40A_MOUSE	Pre-mRNA-processing factor 40 homolog A OS=Mus musculus GN=Prpf40a PE=1 SV=1
MYH9_MOUSE	Myosin-9 OS=Mus musculus GN=Myh9 PE=1 SV=4
RBM25_MOUSE	RNA-binding protein 25 OS=Mus musculus GN=Rbm25 PE=1 SV=2
DHX15_MOUSE	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Mus musculus GN=Dhx15 PE=2 SV=2
SP16H_MOUSE	FACT complex subunit SPT16 OS=Mus musculus GN=Supt16h PE=1 SV=2
SRRM1_MOUSE	Serine/arginine repetitive matrix protein 1 OS=Mus musculus GN=Srrm1 PE=1 SV=2
H2B1B_MOUSE	Histone H2B type 1-B OS=Mus musculus GN=Hist1h2bb PE=1 SV=3
ROA3_MOUSE	Heterogeneous nuclear ribonucleoprotein A3 OS=Mus musculus GN=Hnrnpa3 PE=1 SV=1
ITCH_MOUSE	E3 ubiquitin-protein ligase Itchy OS=Mus musculus GN=Itch PE=1 SV=2
SRSF3_MOUSE	Serine/arginine-rich splicing factor 3 OS=Mus musculus GN=Srsf3 PE=1 SV=1
SRSF1_MOUSE	Serine/arginine-rich splicing factor 1 OS=Mus musculus GN=Srsf1 PE=1 SV=3
NPM_MOUSE	Nucleophosmin OS=Mus musculus GN=Npm1 PE=1 SV=1
RPA1_MOUSE	DNA-directed RNA polymerase I subunit RPA1 OS=Mus musculus GN=Polr1a PE=1 SV=2
TF2H1_MOUSE	General transcription factor IIH subunit 1 OS=Mus musculus GN=Gtf2h1 PE=2 SV=2
SRRM2_MOUSE	Serine/arginine repetitive matrix protein 2 OS=Mus musculus GN=Srrm2 PE=1 SV=3
IF2P_MOUSE	Eukaryotic translation initiation factor 5B OS=Mus musculus GN=Eif5b PE=1 SV=2
FUS_MOUSE	RNA-binding protein FUS OS=Mus musculus GN=Fus PE=2 SV=1
TRAP1_MOUSE	Heat shock protein 75 kDa, mitochondrial OS=Mus musculus GN=Trap1 PE=1 SV=1
CDC5L_MOUSE	Cell division cycle 5-like protein OS=Mus musculus GN=Cdc5l PE=1 SV=2
SRSF5_MOUSE	Serine/arginine-rich splicing factor 5 OS=Mus musculus GN=Srsf5 PE=1 SV=2
SRSF2_MOUSE	Serine/arginine-rich splicing factor 2 OS=Mus musculus GN=Srsf2 PE=1 SV=4
COPB2_MOUSE	Coatomer subunit beta' OS=Mus musculus GN=Copb2 PE=2 SV=2
IF4A3_MOUSE	Eukaryotic initiation factor 4A-III OS=Mus musculus GN=Eif4a3 PE=2 SV=3
UBF1_MOUSE	Nucleolar transcription factor 1 OS=Mus musculus GN=Ubtf PE=1 SV=1
H4_MOUSE	Histone H4 OS=Mus musculus GN=Hist1h4a PE=1 SV=2
SMU1_MOUSE	WD40 repeat-containing protein SMU1 OS=Mus musculus GN=Smu1 PE=2 SV=2
SMD1_MOUSE	Small nuclear ribonucleoprotein Sm D1 OS=Mus musculus GN=Snrpd1 PE=1 SV=1
PRDX1_MOUSE	Peroxiredoxin-1 OS=Mus musculus GN=Prdx1 PE=1 SV=1
RU17_MOUSE	U1 small nuclear ribonucleoprotein 70 kDa OS=Mus musculus GN=Snrnp70 PE=1 SV=2
NOP56_MOUSE	Nucleolar protein 56 OS=Mus musculus GN=Nop56 PE=1 SV=2
KHDR1_MOUSE	KH domain-containing, RNA-binding, signal transduction-associated protein 1 OS=Mus musculus GN=Khdrbs1
K22O_MOUSE	Keratin, type II cytoskeletal 2 oral OS=Mus musculus GN=Krt76 PE=2 SV=1
SRP68_MOUSE	Signal recognition particle subunit SRP68 OS=Mus musculus GN=Srp68 PE=2 SV=2
CWC22_MOUSE	Pre-mRNA-splicing factor CWC22 homolog OS=Mus musculus GN=Cwc22 PE=1 SV=1
H2AX_MOUSE	Histone H2A.x OS=Mus musculus GN=H2afx PE=1 SV=2
SMD2_MOUSE	Small nuclear ribonucleoprotein Sm D2 OS=Mus musculus GN=Snrpd2 PE=1 SV=1
RBM14_MOUSE	RNA-binding protein 14 OS=Mus musculus GN=Rbm14 PE=1 SV=1
NUCL_MOUSE	Nucleolin OS=Mus musculus GN=Ncl PE=1 SV=2
H2A1F_MOUSE	Histone H2A type 1-F OS=Mus musculus GN=Hist1h2af PE=1 SV=3
SSRP1_MOUSE	FACT complex subunit SSRP1 OS=Mus musculus GN=Ssrp1 PE=1 SV=2
KIF4_MOUSE	Chromosome-associated kinesin KIF4 OS=Mus musculus GN=Kif4 PE=1 SV=3
SF3B3_MOUSE	Splicing factor 3B subunit 3 OS=Mus musculus GN=Sf3b3 PE=1 SV=1
LEO1_MOUSE	RNA polymerase-associated protein LEO1 OS=Mus musculus GN=Leo1 PE=1 SV=2
SK2L2_MOUSE	Superkiller viralicidic activity 2-like 2 OS=Mus musculus GN=Skiv2l2 PE=2 SV=1
SMD3_MOUSE	Small nuclear ribonucleoprotein Sm D3 OS=Mus musculus GN=Snrpd3 PE=1 SV=1
RED_MOUSE	Protein Red OS=Mus musculus GN=Ik PE=2 SV=2
PTK6_MOUSE	Protein-tyrosine kinase 6 OS=Mus musculus GN=Ptk6 PE=1 SV=1
FZD2_MOUSE	Frizzled-2 OS=Mus musculus GN=Fzd2 PE=2 SV=1
COPG1_MOUSE	Coatomer subunit gamma-1 OS=Mus musculus GN=Copg1 PE=2 SV=1
CD11B_MOUSE	Cyclin-dependent kinase 11B OS=Mus musculus GN=Cdk11b PE=1 SV=2
HSP7C_MOUSE	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
S30BP_MOUSE	SAP30-binding protein OS=Mus musculus GN=Sap30bp PE=2 SV=2
FYN_MOUSE	Tyrosine-protein kinase Fyn OS=Mus musculus GN=Fyn PE=1 SV=4

CDK13_MOUSE   Cyclin-dependent kinase 13 OS=Mus musculus GN=Cdk13 PE=1 SV=3
H32_MOUSE   Histone H3.2 OS=Mus musculus GN=Hist1h3b PE=1 SV=2
COPA_MOUSE   Coatomer subunit alpha OS=Mus musculus GN=Copa PE=1 SV=2
SRS10_MOUSE   Serine/arginine-rich splicing factor 10 OS=Mus musculus GN=Srsf10 PE=1 SV=2
CDC73_MOUSE   Parafibromin OS=Mus musculus GN=Cdc73 PE=2 SV=1
AGTRA_MOUSE   Type-1A angiotensin II receptor OS=Mus musculus GN=Agtr1a PE=1 SV=1
HNRH1_MOUSE   Heterogeneous nuclear ribonucleoprotein H OS=Mus musculus GN=Hnrnph1 PE=1 SV=3
CTR9_MOUSE   RNA polymerase-associated protein CTR9 homolog OS=Mus musculus GN=Ctr9 PE=1 SV=2
SRSF7_MOUSE   Serine/arginine-rich splicing factor 7 OS=Mus musculus GN=Srsf7 PE=1 SV=1
ACTG_MOUSE   Actin, cytoplasmic 2 OS=Mus musculus GN=Actg1 PE=1 SV=1
SR140_MOUSE   U2 snRNP-associated SURP motif-containing protein OS=Mus musculus GN=U2surp PE=1 SV=3
U2AF1_MOUSE   Splicing factor U2AF 35 kDa subunit OS=Mus musculus GN=U2af1 PE=1 SV=4

Control Antibody Immunoprecipitation in pooled fractions 3-4 (RBS complex)
TOP2A_MOUSE   DNA topoisomerase 2-alpha OS=Mus musculus GN=Top2a PE=1 SV=2
NONO_MOUSE   Non-POU domain-containing octamer-binding protein OS=Mus musculus GN=Nono PE=1 SV=3
SFPQ_MOUSE   Splicing factor, proline- and glutamine-rich OS=Mus musculus GN=Sfpq PE=1 SV=1
TOP2B_MOUSE   DNA topoisomerase 2-beta OS=Mus musculus GN=Top2b PE=1 SV=2
TCOF_MOUSE   Treacle protein OS=Mus musculus GN=Tcof1 PE=1 SV=1
DDX3X_MOUSE   ATP-dependent RNA helicase DDX3X OS=Mus musculus GN=Ddx3x PE=1 SV=3
DDX5_MOUSE   Probable ATP-dependent RNA helicase DDX5 OS=Mus musculus GN=Ddx5 PE=1 SV=2
PARP1_MOUSE   Poly [ADP-ribose] polymerase 1 OS=Mus musculus GN=Parp1 PE=1 SV=3
NPM_MOUSE   Nucleophosmin OS=Mus musculus GN=Npm1 PE=1 SV=1
CPSF6_MOUSE   Cleavage and polyadenylation specificity factor subunit 6 OS=Mus musculus GN=Cpsf6 PE=1 SV=1
LAP2A_MOUSE   Lamina-associated polypeptide 2, isoforms alpha/zeta OS=Mus musculus GN=Tmpo PE=1 SV=4
UBF1_MOUSE   Nucleolar transcription factor 1 OS=Mus musculus GN=Ubtf PE=1 SV=1
MYH9_MOUSE   Myosin-9 OS=Mus musculus GN=Myh9 PE=1 SV=4
PR40A_MOUSE   Pre-mRNA-processing factor 40 homolog A OS=Mus musculus GN=Prpf40a PE=1 SV=1
RBM25_MOUSE   RNA-binding protein 25 OS=Mus musculus GN=Rbm25 PE=1 SV=2
SP16H_MOUSE   FACT complex subunit SPT16 OS=Mus musculus GN=Supt16h PE=1 SV=2
DDX17_MOUSE   Probable ATP-dependent RNA helicase DDX17 OS=Mus musculus GN=Ddx17 PE=2 SV=1
SRRM1_MOUSE   Serine/arginine repetitive matrix protein 1 OS=Mus musculus GN=Srrm1 PE=1 SV=2
ROA3_MOUSE   Heterogeneous nuclear ribonucleoprotein A3 OS=Mus musculus GN=Hnrnpa3 PE=1 SV=1
CPSF5_MOUSE   Cleavage and polyadenylation specificity factor subunit 5 OS=Mus musculus GN=Nudt21 PE=2 SV=1
CD11B_MOUSE   Cyclin-dependent kinase 11B OS=Mus musculus GN=Cdk11b PE=1 SV=2
K1C10_MOUSE   Keratin, type I cytoskeletal 10 OS=Mus musculus GN=Krt10 PE=1 SV=3
RBM14_MOUSE   RNA-binding protein 14 OS=Mus musculus GN=Rbm14 PE=1 SV=1
PML_MOUSE   Protein PML OS=Mus musculus GN=Pml PE=1 SV=3
DDX46_MOUSE   Probable ATP-dependent RNA helicase DDX46 OS=Mus musculus GN=Ddx46 PE=1 SV=2
COPA_MOUSE   Coatomer subunit alpha OS=Mus musculus GN=Copa PE=1 SV=2
SRSF5_MOUSE   Serine/arginine-rich splicing factor 5 OS=Mus musculus GN=Srsf5 PE=1 SV=2
SRSF2_MOUSE   Serine/arginine-rich splicing factor 2 OS=Mus musculus GN=Srsf2 PE=1 SV=4
IF2P_MOUSE   Eukaryotic translation initiation factor 5B OS=Mus musculus GN=Eif5b PE=1 SV=2
CWC22_MOUSE   Pre-mRNA-splicing factor CWC22 homolog OS=Mus musculus GN=Cwc22 PE=1 SV=1
K2C71_MOUSE   Keratin, type II cytoskeletal 71 OS=Mus musculus GN=Krt71 PE=1 SV=1
SRSF3_MOUSE   Serine/arginine-rich splicing factor 3 OS=Mus musculus GN=Srsf3 PE=1 SV=1
EF1A1_MOUSE   Elongation factor 1-alpha 1 OS=Mus musculus GN=Eef1a1 PE=1 SV=3
RBM27_MOUSE   RNA-binding protein 27 OS=Mus musculus GN=Rbm27 PE=2 SV=3
COPB2_MOUSE   Coatomer subunit beta' OS=Mus musculus GN=Copb2 PE=2 SV=2
RU17_MOUSE   U1 small nuclear ribonucleoprotein 70 kDa OS=Mus musculus GN=Snrnp70 PE=1 SV=2
SMC2_MOUSE   Structural maintenance of chromosomes protein 2 OS=Mus musculus GN=Smc2 PE=1 SV=2
K22O_MOUSE   Keratin, type II cytoskeletal 2 oral OS=Mus musculus GN=Krt76 PE=2 SV=1
SRSF7_MOUSE   Serine/arginine-rich splicing factor 7 OS=Mus musculus GN=Srsf7 PE=1 SV=1
HSP7C_MOUSE   Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
SMD2_MOUSE   Small nuclear ribonucleoprotein Sm D2 OS=Mus musculus GN=Snrpd2 PE=1 SV=1
H2B1B_MOUSE   Histone H2B type 1-B OS=Mus musculus GN=Hist1h2bb PE=1 SV=3
U2AF1_MOUSE   Splicing factor U2AF 35 kDa subunit OS=Mus musculus GN=U2af1 PE=1 SV=4
DHX15_MOUSE   Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Mus musculus GN=Dhx15 PE=2 SV=2
H4_MOUSE   Histone H4 OS=Mus musculus GN=Hist1h4a PE=1 SV=2
RCC2_MOUSE   Protein RCC2 OS=Mus musculus GN=Rcc2 PE=2 SV=1
ACTG_MOUSE   Actin, cytoplasmic 2 OS=Mus musculus GN=Actg1 PE=1 SV=1
FUS_MOUSE   RNA-binding protein FUS OS=Mus musculus GN=Fus PE=2 SV=1
PRDX1_MOUSE   Peroxiredoxin-1 OS=Mus musculus GN=Prdx1 PE=1 SV=1
TF2H1_MOUSE   General transcription factor IIH subunit 1 OS=Mus musculus GN=Gtf2h1 PE=2 SV=2
FYN_MOUSE   Tyrosine-protein kinase Fyn OS=Mus musculus GN=Fyn PE=1 SV=4
SRRM2_MOUSE   Serine/arginine repetitive matrix protein 2 OS=Mus musculus GN=Srrm2 PE=1 SV=3
SNRPA_MOUSE   U1 small nuclear ribonucleoprotein A OS=Mus musculus GN=Snrpa PE=2 SV=3

```
  PAF1_MOUSE        RNA polymerase II-associated factor 1 homolog OS=Mus musculus GN=Paf1 PE=2 SV=1
  SRSF1_MOUSE       Serine/arginine-rich splicing factor 1 OS=Mus musculus GN=Srsf1 PE=1 SV=3
  SRP68_MOUSE       Signal recognition particle subunit SRP68 OS=Mus musculus GN=Srp68 PE=2 SV=2
  AGTRA_MOUSE       Type-1A angiotensin II receptor OS=Mus musculus GN=Agtr1a PE=1 SV=1
  S30BP_MOUSE       SAP30-binding protein OS=Mus musculus GN=Sap30bp PE=2 SV=2
  PR38A_MOUSE       Pre-mRNA-splicing factor 38A OS=Mus musculus GN=Prpf38a PE=1 SV=1
  RBM39_MOUSE       RNA-binding protein 39 OS=Mus musculus GN=Rbm39 PE=1 SV=2
  SSRP1_MOUSE       FACT complex subunit SSRP1 OS=Mus musculus GN=Ssrp1 PE=1 SV=2
  G3P_MOUSE         Glyceraldehyde-3-phosphate dehydrogenase OS=Mus musculus GN=Gapdh PE=1 SV=2
  RFC1_MOUSE        Replication factor C subunit 1 OS=Mus musculus GN=Rfc1 PE=1 SV=2
  COPG1_MOUSE       Coatomer subunit gamma-1 OS=Mus musculus GN=Copg1 PE=2 SV=1
  SMD1_MOUSE        Small nuclear ribonucleoprotein Sm D1 OS=Mus musculus GN=Snrpd1 PE=1 SV=1
  TF2H4_MOUSE       General transcription factor IIH subunit 4 OS=Mus musculus GN=Gtf2h4 PE=2 SV=1
```

TRIM28 Immunoprecipitation in pooled fractions 5-7 (large RBS complex)

```
  NONO_MOUSE        Non-POU domain-containing octamer-binding protein OS=Mus musculus GN=Nono PE=1 SV=3
  SFPQ_MOUSE        Splicing factor, proline- and glutamine-rich OS=Mus musculus GN=Sfpq PE=1 SV=1
  FUS_MOUSE         RNA-binding protein FUS OS=Mus musculus GN=Fus PE=2 SV=1
  DDX3X_MOUSE       ATP-dependent RNA helicase DDX3X OS=Mus musculus GN=Ddx3x PE=1 SV=3
  DDX5_MOUSE        Probable ATP-dependent RNA helicase DDX5 OS=Mus musculus GN=Ddx5 PE=1 SV=2
  CPSF5_MOUSE       Cleavage and polyadenylation specificity factor subunit 5 OS=Mus musculus GN=Nudt21 PE=2 SV=1
  CPSF6_MOUSE       Cleavage and polyadenylation specificity factor subunit 6 OS=Mus musculus GN=Cpsf6 PE=1 SV=1
  DDX17_MOUSE       Probable ATP-dependent RNA helicase DDX17 OS=Mus musculus GN=Ddx17 PE=2 SV=1
  LAP2A_MOUSE       Lamina-associated polypeptide 2, isoforms alpha/zeta OS=Mus musculus GN=Tmpo PE=1 SV=4
  TIF1B_MOUSE       Transcription intermediary factor 1-beta OS=Mus musculus GN=Trim28 PE=1 SV=3
  SRSF2_MOUSE       Serine/arginine-rich splicing factor 2 OS=Mus musculus GN=Srsf2 PE=1 SV=4
  CPSF7_MOUSE       Cleavage and polyadenylation specificity factor subunit 7 OS=Mus musculus GN=Cpsf7 PE=1 SV=2
  ROA3_MOUSE        Heterogeneous nuclear ribonucleoprotein A3 OS=Mus musculus GN=Hnrnpa3 PE=1 SV=1
  PARP1_MOUSE       Poly [ADP-ribose] polymerase 1 OS=Mus musculus GN=Parp1 PE=1 SV=3
  TOP2A_MOUSE       DNA topoisomerase 2-alpha OS=Mus musculus GN=Top2a PE=1 SV=2
  RBM14_MOUSE       RNA-binding protein 14 OS=Mus musculus GN=Rbm14 PE=1 SV=1
  SRSF3_MOUSE       Serine/arginine-rich splicing factor 3 OS=Mus musculus GN=Srsf3 PE=1 SV=1
  ROA1_MOUSE        Heterogeneous nuclear ribonucleoprotein A1 OS=Mus musculus GN=Hnrnpa1 PE=1 SV=2
  ITCH_MOUSE        E3 ubiquitin-protein ligase Itchy OS=Mus musculus GN=Itch PE=1 SV=1
  SRSF5_MOUSE       Serine/arginine-rich splicing factor 5 OS=Mus musculus GN=Srsf5 PE=1 SV=2
  HNRPF_MOUSE       Heterogeneous nuclear ribonucleoprotein F OS=Mus musculus GN=Hnrnpf PE=1 SV=3
  SRRM1_MOUSE       Serine/arginine repetitive matrix protein 1 OS=Mus musculus GN=Srrm1 PE=1 SV=2
  SRRM2_MOUSE       Serine/arginine repetitive matrix protein 2 OS=Mus musculus GN=Srrm2 PE=1 SV=3
  PR38A_MOUSE       Pre-mRNA-splicing factor 38A OS=Mus musculus GN=Prpf38a PE=1 SV=1
  ROA2_MOUSE        Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Mus musculus GN=Hnrnpa2b1 PE=1 SV=2
  THOC4_MOUSE       THO complex subunit 4 OS=Mus musculus GN=Alyref PE=1 SV=3
  RBM26_MOUSE       RNA-binding protein 26 OS=Mus musculus GN=Rbm26 PE=1 SV=2
  SRSF1_MOUSE       Serine/arginine-rich splicing factor 1 OS=Mus musculus GN=Srsf1 PE=1 SV=3
  HNRPM_MOUSE       Heterogeneous nuclear ribonucleoprotein M OS=Mus musculus GN=Hnrnpm PE=1 SV=3
  KHDR1_MOUSE       KH domain-containing, RNA-binding, signal transduction-associated protein 1 OS=Mus musculus GN=Khdrbs
  K2C74_MOUSE       Keratin, type II cytoskeletal 74 OS=Mus musculus GN=Krt74 PE=2 SV=1
  DDX46_MOUSE       Probable ATP-dependent RNA helicase DDX46 OS=Mus musculus GN=Ddx46 PE=1 SV=2
  PRDX1_MOUSE       Peroxiredoxin-1 OS=Mus musculus GN=Prdx1 PE=1 SV=1
  RSRC2_MOUSE       Arginine/serine-rich coiled-coil protein 2 OS=Mus musculus GN=Rsrc2 PE=2 SV=1
  RBM27_MOUSE       RNA-binding protein 27 OS=Mus musculus GN=Rbm27 PE=2 SV=3
  SRS10_MOUSE       Serine/arginine-rich splicing factor 10 OS=Mus musculus GN=Srsf10 PE=1 SV=2
  ACTG_MOUSE        Actin, cytoplasmic 2 OS=Mus musculus GN=Actg1 PE=1 SV=1
  IF4A3_MOUSE       Eukaryotic initiation factor 4A-III OS=Mus musculus GN=Eif4a3 PE=2 SV=3
  LMNA_MOUSE        Prelamin-A/C OS=Mus musculus GN=Lmna PE=1 SV=2
  PR40A_MOUSE       Pre-mRNA-processing factor 40 homolog A OS=Mus musculus GN=Prpf40a PE=1 SV=1
  G3P_MOUSE         Glyceraldehyde-3-phosphate dehydrogenase OS=Mus musculus GN=Gapdh PE=1 SV=2
  DHX15_MOUSE       Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Mus musculus GN=Dhx15 PE=2 SV=2
  CWC22_MOUSE       Pre-mRNA-splicing factor CWC22 homolog OS=Mus musculus GN=Cwc22 PE=1 SV=1
  TCOF_MOUSE        Treacle protein OS=Mus musculus GN=Tcof1 PE=1 SV=1
  HNRPK_MOUSE       Heterogeneous nuclear ribonucleoprotein K OS=Mus musculus GN=Hnrnpk PE=1 SV=1
  SRSF7_MOUSE       Serine/arginine-rich splicing factor 7 OS=Mus musculus GN=Srsf7 PE=1 SV=1
  H2A1F_MOUSE       Histone H2A type 1-F OS=Mus musculus GN=Hist1h2af PE=1 SV=3
  LC7L3_MOUSE       Luc7-like protein 3 OS=Mus musculus GN=Luc7l3 PE=1 SV=1
  ARGL1_MOUSE       Arginine and glutamate-rich protein 1 OS=Mus musculus GN=Arglu1 PE=1 SV=2
  AL1A3_MOUSE       Aldehyde dehydrogenase family 1 member A3 OS=Mus musculus GN=Aldh1a3 PE=2 SV=1
  HNRL1_MOUSE       Heterogeneous nuclear ribonucleoprotein U-like protein 1 OS=Mus musculus GN=Hnrnpul1 PE=1 SV=1
  SAFB1_MOUSE       Scaffold attachment factor B1 OS=Mus musculus GN=Safb PE=1 SV=2
  HNRPL_MOUSE       Heterogeneous nuclear ribonucleoprotein L OS=Mus musculus GN=Hnrnpl PE=1 SV=2
```

146

MFAP1_MOUSE	Microfibrillar-associated protein 1 OS=Mus musculus GN=Mfap1 PE=1 SV=1
ITPR3_MOUSE	Inositol 1,4,5-trisphosphate receptor type 3 OS=Mus musculus GN=Itpr3 PE=1 SV=3
HNRH2_MOUSE	Heterogeneous nuclear ribonucleoprotein H2 OS=Mus musculus GN=Hnrnph2 PE=1 SV=1
SNRPA_MOUSE	U1 small nuclear ribonucleoprotein A OS=Mus musculus GN=Snrpa PE=2 SV=3

Control Antibody Immunoprecipitation in pooled fractions 5-7 (large RBS complex

NONO_MOUSE	Non-POU domain-containing octamer-binding protein OS=Mus musculus GN=Nono PE=1 SV=3
DDX5_MOUSE	Probable ATP-dependent RNA helicase DDX5 OS=Mus musculus GN=Ddx5 PE=1 SV=2
SFPQ_MOUSE	Splicing factor, proline- and glutamine-rich OS=Mus musculus GN=Sfpq PE=1 SV=1
DDX3X_MOUSE	ATP-dependent RNA helicase DDX3X OS=Mus musculus GN=Ddx3x PE=1 SV=3
DDX17_MOUSE	Probable ATP-dependent RNA helicase DDX17 OS=Mus musculus GN=Ddx17 PE=2 SV=1
CPSF6_MOUSE	Cleavage and polyadenylation specificity factor subunit 6 OS=Mus musculus GN=Cpsf6 PE=1 SV=1
RBM14_MOUSE	RNA-binding protein 14 OS=Mus musculus GN=Rbm14 PE=1 SV=1
SRSF2_MOUSE	Serine/arginine-rich splicing factor 2 OS=Mus musculus GN=Srsf2 PE=1 SV=4
FUS_MOUSE	RNA-binding protein FUS OS=Mus musculus GN=Fus PE=2 SV=1
LAP2A_MOUSE	Lamina-associated polypeptide 2, isoforms alpha/zeta OS=Mus musculus GN=Tmpo PE=1 SV=4
TOP2A_MOUSE	DNA topoisomerase 2-alpha OS=Mus musculus GN=Top2a PE=1 SV=2
K2C1_MOUSE	Keratin, type II cytoskeletal 1 OS=Mus musculus GN=Krt1 PE=1 SV=4
K1C10_MOUSE	Keratin, type I cytoskeletal 10 OS=Mus musculus GN=Krt10 PE=1 SV=3
K2C75_MOUSE	Keratin, type II cytoskeletal 75 OS=Mus musculus GN=Krt75 PE=1 SV=1
HNRPF_MOUSE	Heterogeneous nuclear ribonucleoprotein F OS=Mus musculus GN=Hnrnpf PE=1 SV=3
RBM27_MOUSE	RNA-binding protein 27 OS=Mus musculus GN=Rbm27 PE=2 SV=3
CPSF5_MOUSE	Cleavage and polyadenylation specificity factor subunit 5 OS=Mus musculus GN=Nudt21 PE=2 SV=1
SRSF1_MOUSE	Serine/arginine-rich splicing factor 1 OS=Mus musculus GN=Srsf1 PE=1 SV=3
CPSF7_MOUSE	Cleavage and polyadenylation specificity factor subunit 7 OS=Mus musculus GN=Cpsf7 PE=1 SV=2
ROA3_MOUSE	Heterogeneous nuclear ribonucleoprotein A3 OS=Mus musculus GN=Hnrnpa3 PE=1 SV=1
SRRM1_MOUSE	Serine/arginine repetitive matrix protein 1 OS=Mus musculus GN=Srrm1 PE=1 SV=2
G3P_MOUSE	Glyceraldehyde-3-phosphate dehydrogenase OS=Mus musculus GN=Gapdh PE=1 SV=2
SRSF5_MOUSE	Serine/arginine-rich splicing factor 5 OS=Mus musculus GN=Srsf5 PE=1 SV=2
PRDX1_MOUSE	Peroxiredoxin-1 OS=Mus musculus GN=Prdx1 PE=1 SV=1
PARP1_MOUSE	Poly [ADP-ribose] polymerase 1 OS=Mus musculus GN=Parp1 PE=1 SV=3
THOC4_MOUSE	THO complex subunit 4 OS=Mus musculus GN=Alyref PE=1 SV=3
SRSF3_MOUSE	Serine/arginine-rich splicing factor 3 OS=Mus musculus GN=Srsf3 PE=1 SV=1
RBM26_MOUSE	RNA-binding protein 26 OS=Mus musculus GN=Rbm26 PE=1 SV=2
HNRPM_MOUSE	Heterogeneous nuclear ribonucleoprotein M OS=Mus musculus GN=Hnrnpm PE=1 SV=3
EF1A1_MOUSE	Elongation factor 1-alpha 1 OS=Mus musculus GN=Eef1a1 PE=1 SV=3
ROA2_MOUSE	Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Mus musculus GN=Hnrnpa2b1 PE=1 SV=2
DDX46_MOUSE	Probable ATP-dependent RNA helicase DDX46 OS=Mus musculus GN=Ddx46 PE=1 SV=2
U2AF1_MOUSE	Splicing factor U2AF 35 kDa subunit OS=Mus musculus GN=U2af1 PE=1 SV=4
HNRPK_MOUSE	Heterogeneous nuclear ribonucleoprotein K OS=Mus musculus GN=Hnrnpk PE=1 SV=1
SSRP1_MOUSE	FACT complex subunit SSRP1 OS=Mus musculus GN=Ssrp1 PE=1 SV=2
SRRM2_MOUSE	Serine/arginine repetitive matrix protein 2 OS=Mus musculus GN=Srrm2 PE=1 SV=3
PR40A_MOUSE	Pre-mRNA-processing factor 40 homolog A OS=Mus musculus GN=Prpf40a PE=1 SV=1
K22O_MOUSE	Keratin, type II cytoskeletal 2 oral OS=Mus musculus GN=Krt76 PE=2 SV=1
PR38A_MOUSE	Pre-mRNA-splicing factor 38A OS=Mus musculus GN=Prpf38a PE=1 SV=1
M21D2_MOUSE	Protein MB21D2 OS=Mus musculus GN=Mb21d2 PE=1 SV=1
SAFB1_MOUSE	Scaffold attachment factor B1 OS=Mus musculus GN=Safb PE=1 SV=2
ITCH_MOUSE	E3 ubiquitin-protein ligase Itchy OS=Mus musculus GN=Itch PE=1 SV=2
SNRPA_MOUSE	U1 small nuclear ribonucleoprotein A OS=Mus musculus GN=Snrpa PE=2 SV=3
IF4A1_MOUSE	Eukaryotic initiation factor 4A-I OS=Mus musculus GN=Eif4a1 PE=2 SV=1

Table 1 Proteins detected in RBS and large RBS complexes by Mass spectrometry
F9 cell lysates were separated by an velocity sedimentation on a 10-40% sucrose gradient, and fractions were collected and analyzed by EMSA using the BS2 + PBS probe to detect for the RBS and large RBS complexes. Fractions containing the RBS (fractions 3-4) or large RBS (fractions 5-7) complexes were pooled, separately, and protein complexes were immunoprecipitated with an anti-Trim28 antibody or control antibody. Precipitated proteins were resolved by SDS-Page, gel slices containing the proteins samples were excised, and proteins were subjected to tryptic digest followed by peptide identification by lc-ms/ms using

a hybrid high-resolution quadrupole time-of-flight electrospray mass spectrometer. Results were analyzed using the MASCOT database search tool (Matrix Science).