The Effect Of Sudemycin Treatment On Alternative Splicing In Primary Murine Hematopoietic Cells Expressing Mutant *U2af1*

Edgar A. Campbell IV

INTRODUCTION

The myelodysplastic syndromes (MDS) are the most common cause of bone marrow failure in adults (Cogle et al. 2011). MDS are a heterogeneous group of hematopoietic stem cell disorders. Hematopoiesis refers to the process that generates all the cells that constitute the blood. Every blood cell originates as a hematopoietic stem cell. These cells reside in the bone marrow, the tissue found in the interior of bones. Hematopoietic stem cells proliferate to form progenitor cells, which differentiate into the different components of blood. The components of blood include erythrocytes (red blood cells), lymphocytes (B cells, T cells, NK cells), and granulocytes (neutrophils, eosinophils, basophils)(Cumano et al. 2001). MDS are characterized by a disruption in hematopoiesis. Cells fail to differentiate because of this disruption. As a result, patients with MDS usually present with anemia or other cytopenias (Tefferi et al. 2009). Up to 30% of individuals with MDS will progress to highly chemotherapy-resistant secondary acute myeloid leukemia (sAML) (Greenberg et al. 1997). Despite the prevalence and deadliness of MDS and sAML, treatments for these diseases have not changed in decades. One reason for his is that the genetic events that initiate the pathogenesis of these diseases were unknown (Ley et al. 2008).

DNA is a molecule that encodes the genetic information used in the development and function of every living organism known to man. The structure of DNA consists of nucleotides and two phosphate sugar strands that wind together to create a double helix. The genetic information is based on a code made by the order of the four nucleotides that comprise DNA. DNA sequencing is the process that determines the precise order of these nucleotides. Sanger et al. (1977) developed a method to determine the sequence of DNA. His method utilized labeled dideoxynucleotides during *in vitro* DNA replication. These dideoxynucleotides terminate DNA replication when they are incorporated into the newly formed DNA strand. In a DNA replication reaction, these dideoxynucleotides will incorporate randomly in the strand, creating many strands of different lengths. These DNA stands can then be separated based on size and the labels on the dideoxynucleotides can be read in order to give the sequence of DNA. In the ensuing decades, improvements were made in the types of labels on the dideoxynucleotides and in the machinery used to read the labels, but the basis of Sanger sequencing remained unchanged. Sanger sequencing was used to sequence the human genome, a project that lasted 13 years and cost nearly $2.7 billion (IHGSC 2004). Despite the benefits, the cost of sequencing using the Sanger method severely limited its use in biomedical research.

In 2005, the first major advancement in sequencing technology, since Sanger developed his method in the late 1970s, was made. This new approach in sequencing was dubbed “massively parallel sequencing” and differed significantly from Sanger’s method (Mardis 2011). Although different companies introduced a number of sequencing platforms, all massively parallel sequencers share a basic design. First the DNA being sequenced is fragmented and synthetic DNA adaptors are ligated to the ends. These fragments are then amplified on a solid surface. The solid

surface differs depending on the sequencing platform being used. After amplification, the sequencing machines perform and repeat a number of carefully orchestrated and automated steps that detect the sequence of the DNA. One commonly used massively parallel sequencing platform is the Illumina Hi-Seq 2000. The Illumina sequencers use a glass surface embedded with synthetic DNA primers. The synthetic primers ligated to the DNA fragment attach to the primers embedded on the glass. While attached to the glass plate, the DNA is amplified using PCR. This amplification creates clusters of DNA copies attached to the glass plate. These clusters are each composed of only one sequence. Fluorescently labeled nucleotides are then added, one at a time, on each DNA fragment. The nucleotides contain a chemical blocker that stops the DNA replication. This blocker blocks DNA replication in much the same way that dideoxynucleotides stopped the reaction in Sanger sequencing. Each nucleotide incorporates one at a time because of this blocking group. A single nucleotide on each strand is added, detected by laser, and recorded by camera after each incorporation event. Since DNA cluster contains the same sequence, the cluster gives a single, amplified signal. After the nucleotide is detected, the blocking group is chemically removed and the process is repeated (Mardis 2011). Because each nucleotide added requires many steps, it takes about eight days to read a 100-150 base pair DNA fragment. Despite the long run time and short reading length, the Illumina platform can detect enough DNA fragment island to read the entire human genome three times over in a single eight day run (Mardis 2011). Massively parallel sequencing has enabled genomic sequencing to be done in reasonable timeframe that is many orders of magnitude quicker than previous methods. Since this can all be done on one machine in a week's time, the cost to sequence a genome has also dropped significantly to about $30,000, as of early 2011 (Mardis 2011).

With sequencing costs drastically lowered and possible in a reasonable timeframe, whole genome sequencing approaches were used to identify genetic mutations in the cancer genomes of AML (Welch et al. 2011, Mardis et al. 2009, Ley et al. 2008, Link et al. 2011, Ley et al. 2010). Massively parallel whole genome sequencing offered an unbiased approach to discover genetic mutations in cancer genomes. This new approach is contrasted with the initial approach to identifying mutations in cancer that utilized polymerase chain reactions (PCR) to investigate genes that researchers thought might be mutated (Pao et al. 2004, Sjoblom et al. 2006, Wood et al. 2007). The new massively parallel approach has enabled researchers to begin searching for mutations within cancer genomes. The technique to finding cancer mutations begins with sequencing a cancer patient's normal genome. The normal genome is usually obtained from the skin or any other tissue from the patient that is not cancerous. In addition to this genome, the genome obtained from the patient's cancerous tissue is also sequenced. The two genomic sequences are then compared side by side. Differences in the sequence indicate mutations. This was the approach used by Graubert et al. (2012) to detect mutations within the MDS and MDS-derived sAML genomes of patients. A recurrent missense mutation was found within MDS patients. This mutation was located in the coding region of the gene *U2AF1* at amino acid position 34. Graubert et al. (2012) determined the frequency of this mutation within MDS patients by completely sequencing the entire coding region of *U2AF1* in bone marrow and skin samples from 150 MDS patients. Of these patients, 13 individuals (8.7%) were identified with a mutation in position 34 (Fig. 1). Of the thirteen patients identified with a mutation,

eleven harbored a mutation that resulted in an amino acid change from serine to phenylalanine (S34F). The remaining two had a mutation that resulted in an amino acid change of serine to tyrosine (S34Y).

U2AF1 is the highly conserved small (35kDa) subunit of the U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) (Web & Wise 2004). U2AF1 forms a heterodimer with the large subunit (65 kDa) U2AF2 to form U2AF (Wu et al. 1999). U2AF is involved in pre-mRNA processing, a process called splicing. Splicing is a complex process that is ubiquitous in eukaryotic cells. Splicing requires over 150 proteins, an array of small nuclear ribonucleoprotein particles (snRNPs), and a series of small nuclear RNAs (snRNAs) (Will et al. 2009, Wahl et al. 2011). These molecules, referred collectively as the spliceosome, act in concert to achieve crucial post-transcriptional processing of the pre-mRNA. In eukaryotes, DNA, the molecule that encodes all the genetic information needed by living organisms to develop and maintain life, resides in the nucleus of the cell. A molecule called RNA polymerase copies the information encoded in DNA to create a complementary RNA strand referred to as pre-messenger RNA (pre-mRNA). This pre-mRNA contains coding regions called exons and non-coding regions call introns. Introns must be excised out and the exons must be ligated together to form a mature mRNA strand. The major snRNAs that make up the spliceosome are the molecules U1, U2, U4/U6, and U5 (Marz et al. 2008). These snRNAs catalyze the major steps in the splicing process, but they do not act alone. There are a number of regulating and auxiliary proteins that facilitate the splicing reactions. U2AF1 is one of these auxiliary proteins. In order for an intron to be sliced out, the snRNA U2 must attach to the 3' end of the intron. To find the 3' end of the intron, U2 relies on U2AF to find and recognize it. The U2AF1 subunit of U2AF recognizes the AG splice acceptor dinucleotide present at the 3' end of every intron (Wu et al. 1999). This is a very important function for U2AF1 and partly explains why the sequence and structure has remained conserved throughout eukaryotic evolution (Web & Wise 2004).

The S34F/Y U2AF1 mutation found by Graubert et al. (2012) changes the serine normally found at position 34 into either a phenylalanine or a tyrosine. Both of these resulting amino acids are large and aromatic. Position 34 is located within the first zinc-finger domain of U2AF1. It is not known for sure what the function of the U2AF1's zinc-finger is, but it is known that it is structurally similar to both the mouse and human ZFP36 family of zinc-fingers. This family of zinc-fingers is known to bind RNA (Hudson et al. 2004, Lai et al. 2002, Liang et al. 2008). The position analogous to U2AF1's position 34 in ZFP36L2, a ZFP36 zinc-finger, binds to RNA through hydrogen bonding (Hudson et al. 2004). Serine is an amino acid is capable of forming hydrogen bonds with polar substances. This suggests that the zinc-finger in U2AF1 may bind with RNA and that the serine in position 34 is critical for that interaction. The S34F/Y mutation does not affect the amino acids that coordinate the structure of the zinc-finger. The mutation also does not result in a non-sense frame shift. The S34F/Y does not result in a loss of function for U2AF1, which is important to know, since the homozygous loss of U2AF1 function is lethal in many organisms (Golling et al. 2002, Rudner et al. 1996, Zorio & Blumenthal 1999).

Graubert et al. (2012) concluded that the S34F mutation may result in subtle increases in splicing efficiency or induce changes in isoform expression. The way the mutation will do this is by inducing alternative splicing. Alternative splicing is splicing that utilizes alternative splice sites. In normal splicing, introns are excised out and exon are retained and ligated together. Pre-mRNA that under

goes alternative splicing can result in exons that are excised out with introns. Introns can also fail to be excised. Splicing can also happen at cryptic splice sites, which are sites within the exon or intron that can undergo splicing. Each of these results is an altered transcriptome, which can contribute directly to cancer or may contribute indirectly by reengaging some other pathway (Grosso et al. 2008, David & Manley 2010). Alternative splicing has been previously related to levels of normal U2AF1 with cells (Fu et al. 2011, Kralovicova & Vorechovsky 2010, Pacheco et al. 2006). Graubert et al. (2012) report that the S34F mutation promotes exon skipping in *in vitro* assays. The S34F U2AF1 mutation may be responsible for the observed alteration of splicing in MDS.

In addition to the U2AF1 mutation, mutations have been found in other spliceosome genes in MDS patients (Visconte et al. 2011, Papaemmanuil et al. 2011, Yoshida et al. 2011) (Fig. 2). These mutations are believed to produce alterations in splicing that directly and/or indirectly contribute to the development of cancer. The majority of MDS patients who harbor a spliceosome mutation tend to only have one mutation (Graubert et al. 2012). It is believed that while one mutation can lead to the development of cancer, multiple mutations within the spliceosome will be deleterious for the cell. Based on this assumption, we have proposed a possible molecular targeted cancer therapy to treat MDS. A targeted therapy refers to the new generation of cancer drugs that are designed to interfere with a specific molecular target that has a critical role in tumor growth (Sawyers 2004). The U2AF1 mutation is believed to alter the spliceosome, which leads to tumor growth. It is also believed that the spliceosome can only handle one major alteration before become deleterious for the cell. If this is true, tumor cells with the U2AF1 mutation have a spliceosome that is more susceptible to interference than the spliceosomes of cells that do not express the mutation. We have chosen to target the spliceosome in our investigation of a novel, cancer therapy.

Two products derived from nature have been shown to interfere with the spliceosome. A molecule named FR901464, is a drug isolated from the fermentation broth of a strain of *Pseudomonas sp.* which was isolated from a sample of Japanese soil (Nakajima et al 1996). Pladienolide B is a drug isolated from a strain of *Streptomyces platensis* (Kotake et al 2007). Both of these compounds interact and disrupt the function of the spliceosome by targeting and modulating SF3B1 (Kaida et al. 2007, Kotake et al. 2007). SF3B1 is a subunit of the SF3 spliceosome complex. This complex is believed to prevent inappropriate nucleophilic attack on the pre-mRNA by other members of the spliceosome (Lardelli et al. 2010). Like *U2AF1*, mutations within *SF3B1* have also been discovered within MDS cancer genomes (Visconte et al. 2011, Papaemmanuil et al. 2011, Malcovati et al. 2010). Interference of SF3B1 by these two natural products causes cell growth to stall in the G1 and the G2/M cell cycles of growth. Despite this ability, both of these compounds make poor candidates for treatment options. One reason for this is that both compounds are chemically complex. Pladienolide B contains 9 stereocenters and FR901464 contains 10. The high number of stereocenters makes synthesis very complicated. For example, FR901464 requires 40 independent reaction steps to be synthesized, making it impractical for production (Albert et al. 2007). Another reason is that both of these compounds are also very unstable and readily degraded in biological fluids.

The natural products FR901464 and Pladienolide B have very similar mechanisms of actions, but very dissimilar structures. Despite this dissimilarity, Lagisetti et al. (2008) identified the key structural

components that these two unrelated compounds share. A new class of molecules was designed based on this consensus pharmacore called Sudemycins (Lagisetti et al. 2008, Lagisetti et al. 2009) (Fig. 3). Sudemycins are structurally less complex than their natural analogues, containing 6 fewer stereocenters. Sudemycins have better stability and are not degraded within biological fluids (Albert et al. 2009). Sudemycins also have the same effects as the natural products. Sudemycins inhibited splicing, arrested cell growth in the G1 and G2/M phases, and modulated the function of SF3B1 (Lagisetti et al. 2008, Lagisetti et al. 2009). Treatment with Sudemycin in cell lines caused an accumulation of aberrantly spliced gene products.

Sudemycin has been shown to effectively modulate splicing by targeting SF3B1. We believe that Sudemycin's effects on splicing make it a good candidate for a potential targeted cancer therapy. As stated before, cells expressing the S34F *U2AF1* mutation cause alterations in splice isoforms. It is believed that this environment provides a selective advantage for tumor cell to grow and proliferate. We also believe that perturbation of this splice environment could be deleterious to the cell and cause a selective disadvantage for cells expressing these mutations. It is unknown what effect Sudemycin has on alternative splicing in cells expressing the S34F *U2AF1* mutation. We hypothesize that Sudemycin will modulate the alternative splicing induced by the S34F U2AF1 mutation.

METHODS

S34F U2AF1 Expression in Murine Hematopoietic Cells

Femurs and tibias were collected from C57BL/6 mice. Bone marrow was collected from these bones and pooled together in a single solution. From the bone marrow, c-kit positive cells were selected. C-kit is a cell surface marker present on hematopoietic stem cells and progenitor cells. These c-kit positive cells were cultured with media enriched with stem cell factor (SCF), Flt-3, interleukin-3 (IL3), thromboprotein (TPO), and other cytokines necessary for growth.

Murine stem cell virus (MSCV), was used to transduce these cells with *U2AF1-IRES-GFP*, an expression vector designed to introduce the *U2AF1* gene into the cells. This expression vector is designed to express two genes: *U2AF1* and *GFP* through the use of the internal ribosomal entry site (*IRES*). *IRES* allows the ribosome to initiate protein translation in the middle of mRNA. *GFP* is a gene that encodes for green fluorescent protein (GFP). This is a protein that fluoresces green under blue light. When the successfully transduced cell expresses the *U2AF1* gene, the *IRES* allows GFP to be expressed as well. Under a blue light, cells that were successfully transduced can be easily distinguished by the green fluorescence. The c-kit positive bone marrow cell culture was split into three cultures and transduced with one of three expression vectors. One expression vector contained a wild-type U2AF1 gene along with IRES and GFP. This is the normal U2AF1 found within humans. The second expression vector contained U2AF1 with the S34F mutation that was discovered and identified in 8.7% of MDS patients. This vector also contained IRES and GFP. The third type of vector only contained IRES and GFP as a control. Cells were transduced twice with these vectors. The first transduction occurred initially after culture and the second transduction occurred 24 hours later.

After 24 hours of growth following the second transduction, the cells were sorted by flow cytometry. Flow cytometry is a technique that uses lasers to detect fluorescence of each cell. Cells are suspended in solution and ran through small tubes in the flow cytometer. The flow

cytometer then uses the laser to detect the fluorescence of each cell, one by one. The flow cytometer can then separate the cells based on a number of parameters, including fluorescence. Cells that were successfully transduced by MSCV were separated from the other cells to create cultures that expressed either wild-type U2AF1, mutant U2AF1, or an empty vector (Fig. 4).

Treatment with Sudemycin

After an additional 24 hours following GFP sort, the each culture was treated with either Sudemycin or a control. A concentration of 75nM Sudemycin in DMSO was used to treat the cells. This concentration was chosen based on previous unpublished experiments showing the most significant difference in cell death in cells expressing either mutant or wild-type U2AF1. A solution of DMSO was used a control. Cells were treated for 6 hours.

Generation of cDNA

After drug treatment, the RNA was isolated from each culture (Qiagen). The isolated RNA is primarily mRNA transcripts generated by the cell to be translated into protein. Unlike the DNA, the mRNA contains information as to how the gene was spliced. Since RNA is relatively unstable and subject to degradation, reverse transcriptase (Qiagen) was used to translate the mRNA back into DNA. This DNA is referred to as coding DNA or cDNA. cDNA was generated from both Sudemycin and DMSO treated cells from all three virus types.

RT-PCR

DNA primers were designed to amplify the FMR1 gene (Fig. 5). FMR1 is a gene endogenous to C57BL/6 mice and is biologically irrelevant gene to MDS. It was chosen for the PCR assay based on its ability to be alternatively spliced. Primers were designed using UC Santa Cruz's *in silico* PCR generator. The forward primer was complementary to a region in Exon 17 of *FMR1* while the reverse matched the sequence in Exon 18. Amplification of this region of *FMR1* produces three products of different size. The largest product of 160 base pairs (bps) results from normal splicing of the intron between Exon 17 and 18. The next largest product or 120 bps results from a splice event that occurs on the upstream cryptic splice site within Exon 18. The smallest product of 85 bps results from a splice event occurring on the downstream cryptic splice site. Each PCR reaction was repeated for a total of three technical repeats.

Densitometry

These PCR products can be easily discerned when run on a 10% polyacrylamide gel. Densitometry using Image J software was used to quantify the intensity of the bands within the gel. This quantification was used to calculate the percentage each splice event occurred [(Transcript Band)/(Total Sum of Three Bands) * 100%].

RESULTS

Untreated Cells

cDNA from cells transduced by MSCV to express mutant or wild-type U2AF1 was used in a *FMR1* PCR assay. These cells were not treated with either 75nM Sudemycin or the DMSO control. Three separate PCR reactions were run on 10% polyacrylamide gels (Fig. 6).

The amount of the 120 bp transcript was quantified and the percentage it contributed to the total amount of transcripts was calculated and summarized in graph form in Figure 7. The 120 bp transcript constituted less than 20% of the total transcripts measured in the PCR product of cells expressing wild-type U2AF1 and an empty vector. The 120 bp transcript constituted about 25% of the total transcripts measured in the PCR product of cells expressing S34F U2AF1 (Fig. 7).

The amount of the 85 bp transcript was quantified and the percentage of it contributed

to the total amount of transcripts was calculated and summarized in graph form in Figure 8. The 85 bp transcript constituted less than 15% of the total transcripts measured in the PCR product of cells expressing wild-type U2AF1 and an empty vector. The 120 bp transcript constituted less than 20% of the total transcripts measured in the PCR product of cells expressing S34F U2AF1 (Fig. 8).

6-Hour Drug-Treated Cells

cDNA from cells transduced by MSCV to express mutant or wild-type U2AF1 was used in a *FMR1* PCR assay. These cells were treated with either 75nM Sudemycin or the DMSO control for 6 hours. Three separate PCR reactions were run on 10% polyacrylamide gels (Fig. 9).

The amount of the 120 bp transcript was quantified and the percentage it contributed to the total amount of transcripts was calculated and summarized in graph form in Figure 10. Of the cells treated with DMSO for 6 hours, the 120 bp transcript constituted less than 20% of the total transcripts for cells expressing both wild-type U2AF1 and the empty vector. The 120 bp transcript constituted about 25% of the total transcripts in cells expressing S34F U2AF1 (Fig. 10). These results are similar to the results seen in the untreated cells summarized in Figure 7.

Of the DMSO-treated cells, the 85 bp transcript constituted about 15% of the total transcripts measured in both the cells expressing wild-type U2AF1 and the empty vector. The cells expressing S34F U2AF1 produced more of the 85 bp transcript than the other cells at over 20% (Fig 11). These results are similar to the results seen in the untreated cells summarized in Figure 8.

When treated with 75 nM Sudemycin for 6 hours, the 120 bp transcript's contribution decreased from about 20% to about 10% of the total transcripts in cells expressing empty vector (n=3; p=0.1056). The 120 bp transcript's contribution decreased significantly from about 20% to about 10% of the total transcripts in cells expressing wild-type U2AF1 (n=3; p=0.0089). In cells expressing S34F U2AF1, the 120 bp transcript's contribution also decreased significantly from about 25% to about 20% of the total transcripts measured in the PCR product (n=3; p=0.0022) (Fig. 10).

A similar result in the 75 nM Sudemycin-treated cells is seen in regards to the 85 bp transcript. When treated with 75 nM Sudemycin for 6 hours, the 85 bp transcript's contribution significantly decreased from about 15% to just over 5% of the total transcripts in cells expressing empty vector (n=3; p=0.0001). The 85 bp transcript's contribution also decreased significantly from about 15% to just over 5% of the total transcripts in cells expressing wild-type U2AF1 (n=3; p=0.0012). In cells expressing S34F U2AF1, the 120 bp transcript's contribution decreased significantly from about 20% to about 10% of the total transcripts measured in the PCR product (n=3; $p<0.0001$) (Fig. 11).

DISCUSSION

Untreated Cells

Cells expressing wild-type U2AF1 and the empty vector show a similar alternative splicing pattern when untreated with either 75 nM Sudemycin or DMSO. The 120 bp transcript constituted about 20% of the total transcripts for both virus types. The 85 bp transcript was also similar at about 15%. This suggests that overexpression of wild-type U2AF1 does not cause an alter alternative splicing patterns. When cells express S34F U2AF1, the percentages of both cryptically spliced transcripts increase significantly (120bp: n=3 p=0.008; 85bp: n=3 p=0.0068). Expression of S34F U2AF1 increases cryptic splice site use in FMR1 when compared to wild-type U2AF1 expression. This suggests that S34F U2AF1 effectively alters alternative splicing.

6-Hour 75 nM Sudemycin-Treated Cells

DMSO-treated cells showed a similar alternative splicing pattern as cells that were treated with nothing. DMSO-treated cells that expressed either wild-type U2AF1 or the empty vector showed very similar alternatively spliced transcript percentages to each other. This suggests that, similar to untreated cells, wild-type U2AF1 does not alter alternative splicing when treated for 6 hours with DMSO. The similarity of the DMSO result to the untreated results suggests that DMSO does not elicit an effect on alternative splicing.

Cells treated with 75 nM Sudemycin showed a significant decrease in alternative splicing in FMR1 for all virus types when compared to DMSO-treated cells. Cells expressing wild-type U2AF1 showed a similar decrease in cells expressing the empty vector when compared to DMSO-treated cells. Both virus types saw a decrease in 120 bp percentage from about 20% to about 10% (n=3; V p=0.1056; WT p=0.0089) and a significant decrease in the percentage of the 85 bp transcript from about 15% to about 5% (n=3; V: p=0.0001; W: p=0.0012). This suggests that Sudemycin modulates splicing.

Cells expressing S34F U2AF1 show a significant decrease in alternative splicing in *FMR1* after a 6-hour treatment with 75 nM Sudemycin when compared to cells treated with DMSO for the same time period (120bp: p=0.0022; 85bp: p<0.0001). Expression of S34F U2AF1 increased alternative splice in *FMR1* in both untreated cells and DMSO treated cells. The decrease in transcript levels did not reach levels seen by both wild-type U2AF1 and empty vector expressing cells. These results suggest that Sudemycin treatment can modulate alternative splicing induced by S34F U2AF1.

CONCLUSION AND FUTURE DIRECTIONS

The S34F *U2AF1* mutation alters alternative splicing in *FMR1*. Treatment with 75 nM of Sudemycin modulates this alteration in alternative splicing in *FMR1*.

Splice modulators, such as Sudemycin, warrant further investigation as a novel therapy for people harboring *U2AF1* mutations.

In the future, RNA sequencing will be used to identify any biologically relevant genes affected by both S34F U2AF1 mutation and Sudemycin modulation.

References

Albert, B. J., McPherson, P. A., O'Brien, K., Czaicki, N. L., Destefino, V., Osman, S., Li, M., Day, B. W., Grabowski, P. J., Moore, M. J., Vogt, A., Koide, K. 2009. Meayamycin inhibits pre-messenger RNA splicing and exhibits picomolar activity against multidrug-resistant cells. Molecular Cancer Therapeutics 8: 2308–2318.

Albert, B. J., Sivaramakrishnan, A., Naka, T., Czaicki, N. L., Koide, K. 2007. Total syntheses, fragmentation studies, and antitumor/antiproliferative activities of FR901464 and its low picomolar analogue. Journal of the American Chemical Society 129: 2648–2659.

Cogle, C. R., Craig, B. M., Rollison, D. E., List, A. F. 2011. Incidence of the myelodysplastic syndromes using a novel claims-based algorithm: high number of uncaptured cases by cancer registries. Blood 117: 7121-7125.

Cumano, A. & Godin, I. 2001. Pluripotent hematopoietic stem cell development during embryogenesis. Current Opinion in Immunology 13(2): 166-171.

Fan, L., Lagisetti, C., Edwards, C. C., Webb, T. R., & Potter, P. M. 2011. Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. ACS Chemical Biology, 6(6), 582-589.

Golling, G. et al. 2002. Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. Nature Genetics 31: 135–140.

Graubert, T. A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C. L., Shao, J., Krysiak, K., Harris, C. C., Koboldt, D.C., Larson, D. E., McLellan, M. D., Dooling, D. J., Abbott, R. M., Fulton, R. S., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., Grillot, M., Baty, J., Heath, S., Frater, J. L., Nasim, T., Link, D. C., Tomasson, M. H., Westervelt, P., DiPersio, J. F., Mardis, E. R., Ley, T. J., Wilson, R. K., Walter, M. J. 2011. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nature Genetics 44: 53–57.

Greenberg, P., Cox, C., LeBeau, M. M. , et al. 1997. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 89: 2079-2088.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. Nature 431(7011): 931–945.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., Tani, T., Horinouchi, S., Yoshida, M. 2007. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nature Chemical Biology 3: 576–583.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., Mizui, Y. 2007. Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nature Chemical Biology 3: 570–575.

Lagisetti, C., Pourpak, A., Goronga, T., Jiang, Q., Cui, X., Hyle, J., Lahti, J. M., Morris, S. W., Webb, T. R. 2009. Synthetic mRNA splicing modulator compounds with in vivo antitumor activity. Journal of Medicinal Chemistry 52: 6979–6990.

Lagisetti, C., Pourpak, A., Jiang, Q., Cui, X., Goronga, T., Morris, S. W., Webb, T. R. 2008. Antitumor compounds based on a natural product consensus pharmacophore. Journal of Medicinal Chemistry 51: 6220–6224.

Lardelli, R. M., Thompson, J. X., Yates, J. R., Stevens, S. W. 2010. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA 16(3): 516-528.

Ley, T.J. et al. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456, 66–72.

Ley, T.J. et al. 2010. DNMT3A mutations in acute myeloid leukemia. The New England Journal of Medicine 363: 2424–2433.

Link, D. C. et al. 2011. Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML. The Journal of the American Medical Association 305: 1568–1576.

Malcovati, L., Papaemmanuil, E., Bowen, D. T., Boultwood, J., Della Porta, M. G., Pascutto, C., Travaglino, E., Groves, M. J., Godfrey, A. L., Ambaglio, I., Gallì, A., Da Vià, M. C., Conte, S., Tauro, S., Keenan, N., Hyslop, A., Hinton, J., Mudie, L. J., Wainscoat, J. S., Futreal, P. A., Stratton, M. R., Campbell P. J., Hellström-Lindberg, E., Cazzola, M. 2011. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood 118(24): 6239–6246.

Mardis, E. R. 2011. A decade's perspective on DNA sequencing technology. Nature 470(7333): 198-203.

Mardis, E. R. et al. 2009. Recurring mutations found by sequencing an acute myeloid leukemia genome. The New England Journal of Medicine 361: 1058–1066.

Marz, M., Kirsten, T., Stadler, P. F. 2008. Evolution of spliceosomal snRNA genes in metazoan animals. Journal of Molecular Evolution 67: 594–607.

Nakajima, H. et al. 1996. New antitumor substances, FR901463, FR901464 and FR901465. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. The Journal of Antibiotics (Tokyo) 49: 1196–1203.

Pao, W. et al. 2004. EGF receptor gene mutations are common in lung cancers from ‘‘never smokers’’ and are associated with sensitivity of tumors to gefitinib and erlotinib. Proceedings of the National Academy of Sciences of the United States of America 101, 13306–13311.

Papaemmanuil, E. et al. 2011. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. New England Journal of Medicine 365: 1384–1395.

Rudner, D. Z., Kanaar, R., Breger, K. S., Rio, D. C. 1996. Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects. Proceedings of the National Academy of Sciences of the United States of America 93, 10333–10337.

Sanger, F., Nicklen, S., Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74 (12): 5463–5467.

Sawyers, C. 2004. Targeted cancer therapy. Nature 432(7015): 294-297.

Sjoblom, T. et al. 2006. The consensus coding sequences of human breast and colorectal cancers. Science 314: 268–274.

Tefferi, A. & Vardiman, J. W. 2009. Myelodysplastic syndromes. New England Journal of Medicine 361(19): 1872-1885.

Wahl, M. C., Will, C. L., Luhrmann, R. 2009. The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701–718.

Webb, C. J. & Wise, J. A. 2004. The splicing factor U2AF small subunit is functionally conserved between fission yeast and humans. Molecular and Cellular Biology 24: 4229–4240.

Welch, J. S. et al. 2011. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. The Journal of the American Medical Association 305: 1577–1584.

Will, C. L., Schneider, C., MacMillan, A. M., Katopodis, N. F., Neubauer, G., Wilm, M. et al. 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. The EMBO Journal 20: 4536–4546.

Wood, L. D. et al. 2007. The genomic landscapes of human breast and colorectal cancers. Science 318: 1108–1113.

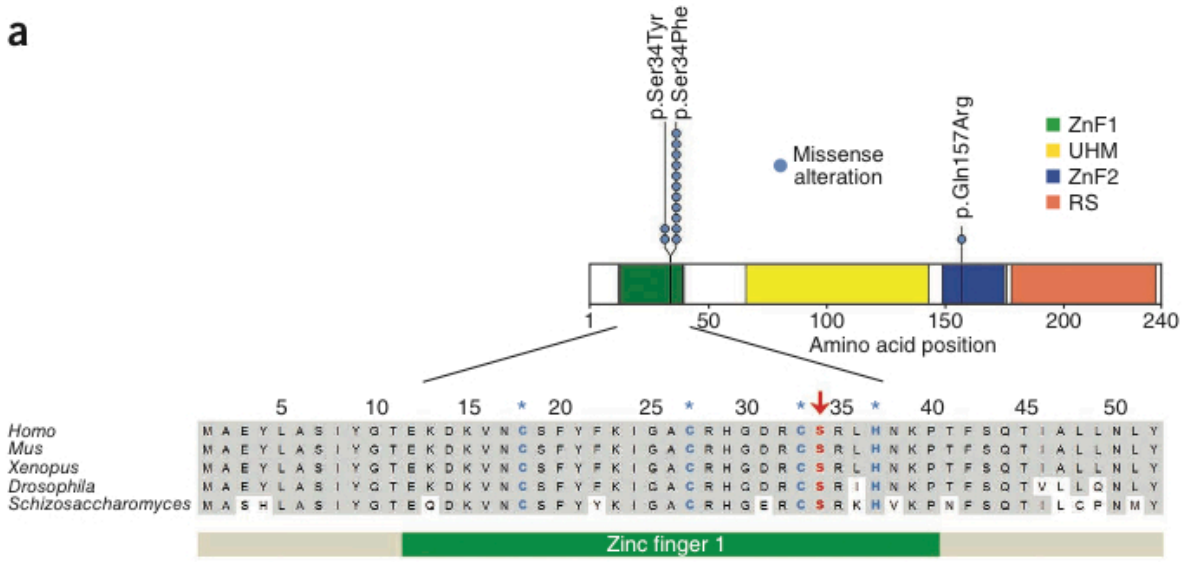
Wu, S., Romfo, C. M., Nilsen, T. W., Green, M. R. 1999. Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature 402: 832–835.

Yoshida, K. et al. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478: 64–69.

Zorio, D. A. & Blumenthal, T. 1999. U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in Caenorhabditis elegans. RNA 5: 487–494.

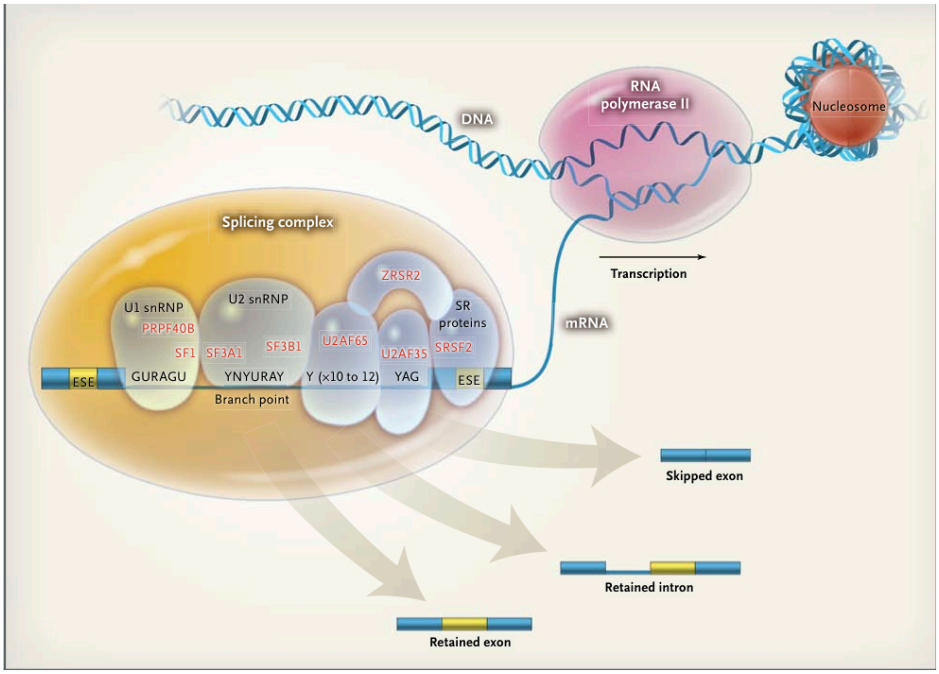
Acknowledgements

My involvement with this project would not have been possible without the opportunity provided to me by Dr. Cherilynn Shadding and the *Opportunities in Genomics Research (OGR) Program*. I give special thanks to Dr. Shadding, Mrs. Wallace, and The Genome Institute for two amazing summers at Washington University in St. Louis School of Medicine. I would also like to thank Dr. Cara Lunn for mentoring my first summer and the rest of the Walter Lab for the initiation into the U2AF1 project. I give special thanks to Dr. Manorama Tripathi for her mentorship of this project and to Dr. Tim Graubert for his guidance.



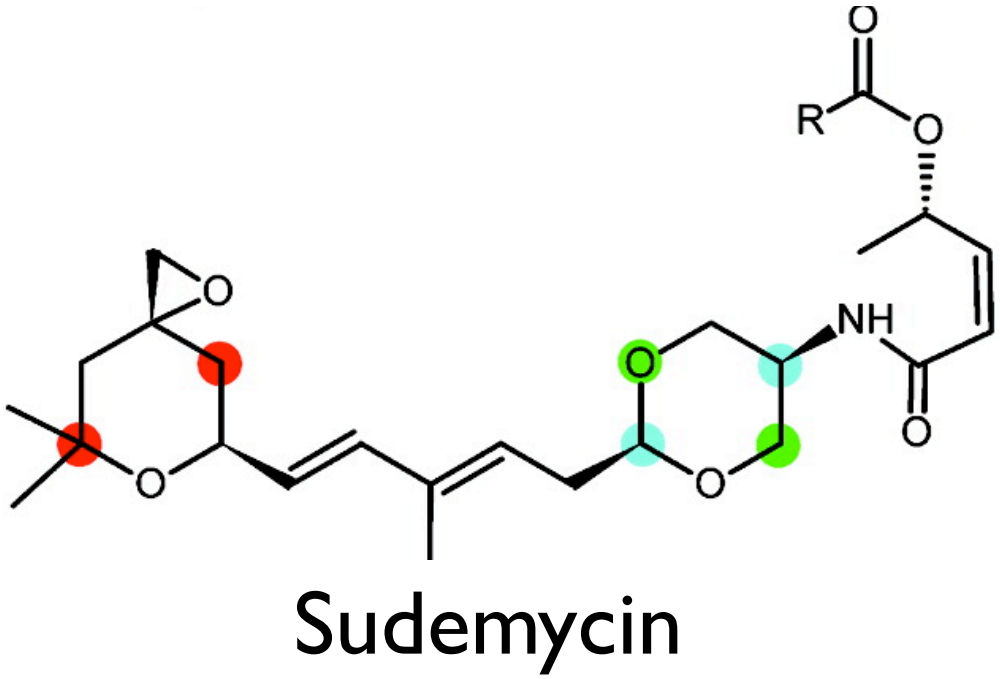
Graubert et al. *Nat. Gen.* 2011

Figure 1. U2AF1 mutations found in individuals with MDS. Missense mutations were detected in codons 34 of U2AF1. The ZnF1 (zinc finger 1), UHM (U2AF homology motif), ZnF2 (zinc finger 2) and RS (arginine-serine rich) domains are shown. The amino acid sequence of the ZnF1 domain is highly conserved (gray shaded area). The zinc coordinating and mutated residues are shown in blue (asterisks) and red (arrow), respectively.



Ebert & Bernard; *NEJM* 2011

Figure 2. The S34F mutation is one of many mutation associated with MDS that are found in the spliceosome shown here. The S34F U2AF1 mutation results in an increase in aberrantly spliced gene products. Retention of introns, skipping of exons, and utilization of cryptic splice sites are all ways genes can be aberrantly spliced. These aberrantly spliced products create an environment that is believed to cause tumor growth.



Sudemycin

Fan et al. *ACS Chem Bio* 2011

Figure 3. A summary structure for the class of synthetic splice modulators known as Sudemycins. Structure is color-coded highlighting of the atoms that have been modified in order to remove chirality and to give a less complex, and more chemically stable structure when compared to FR901464. Red – gain of symmetry and loss of a chemically destabilizing OH group; green – loss of a methyl group and/or atom type change to enhance solubility; cyan – gain of symmetry.

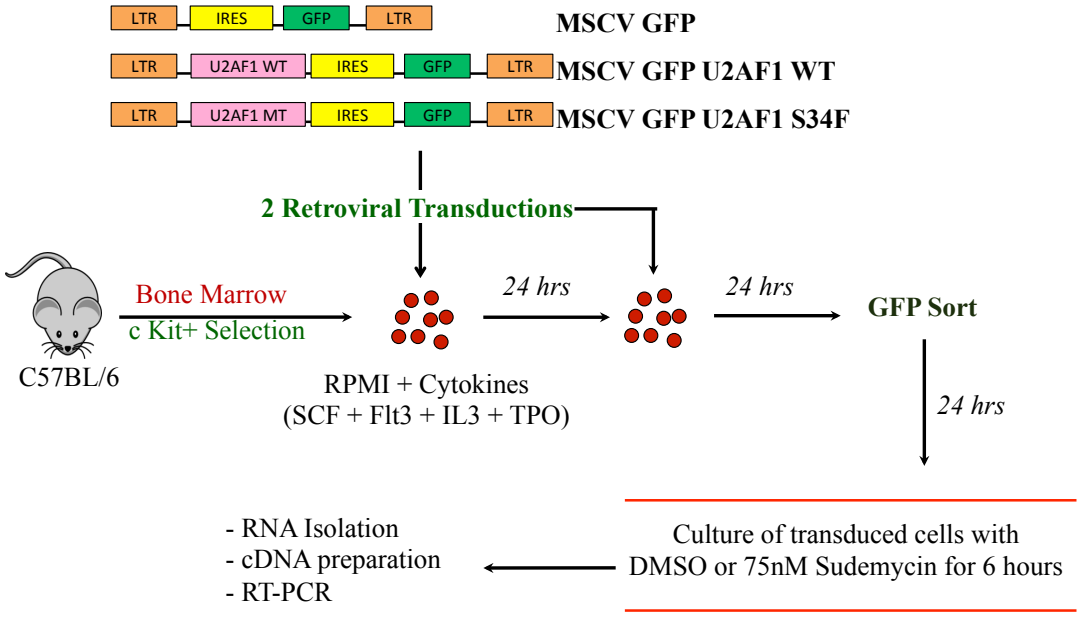


Figure 4. Overview of method used to generate murine hematopoietic cells that express either mutant or wild-type U2AF1.

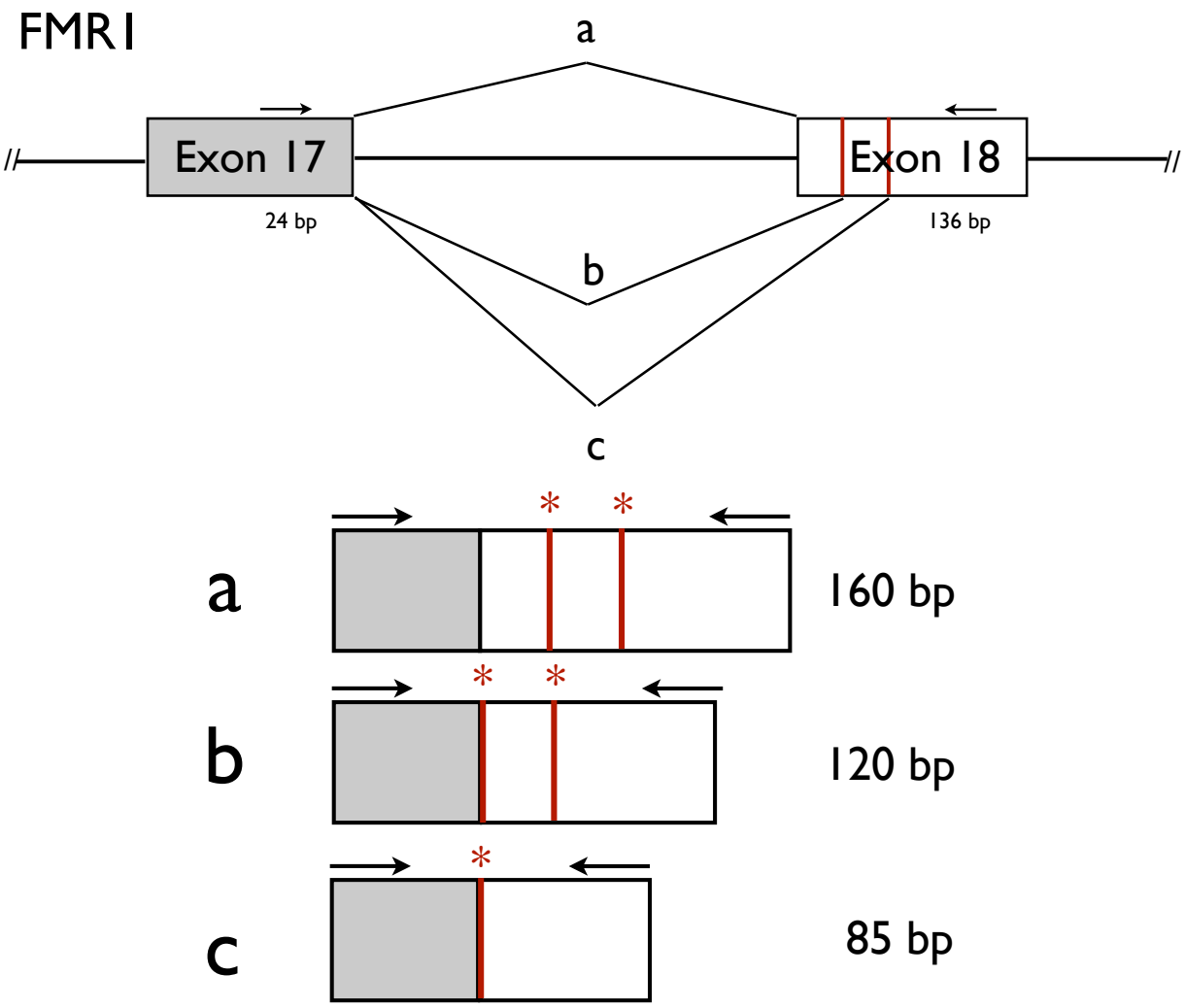


Figure 5. *FMR1*, endogenous to C57BL/6 mice, is shown. The forward primer sits on Exon 17, 24 bps upstream of the 3' end on the exon. The reverse primer sits 136bp downstream the 5' beginning of the exon. Three PCR products are seen using these primers to amplify cDNA obtained from retrovirally transduced murine hematopoietic cells that were treated with either 75nM of Sudemycin or DMSO control. Transcript (a) is a 160 bp product and is he result of normal splice events occurring at the intron-exon junction. Transcripts (b) and (c) are products that resulted from splice events that occurred at cryptic splice sites within Exon 18 highlighted in red. Transcript (b) is a 120 bp product that results from a splice event occurring at the upstream cryptic splice site. Transcript (c) is a 85 bp product that results from a splice event occurring at the downstream cryptic splice site.

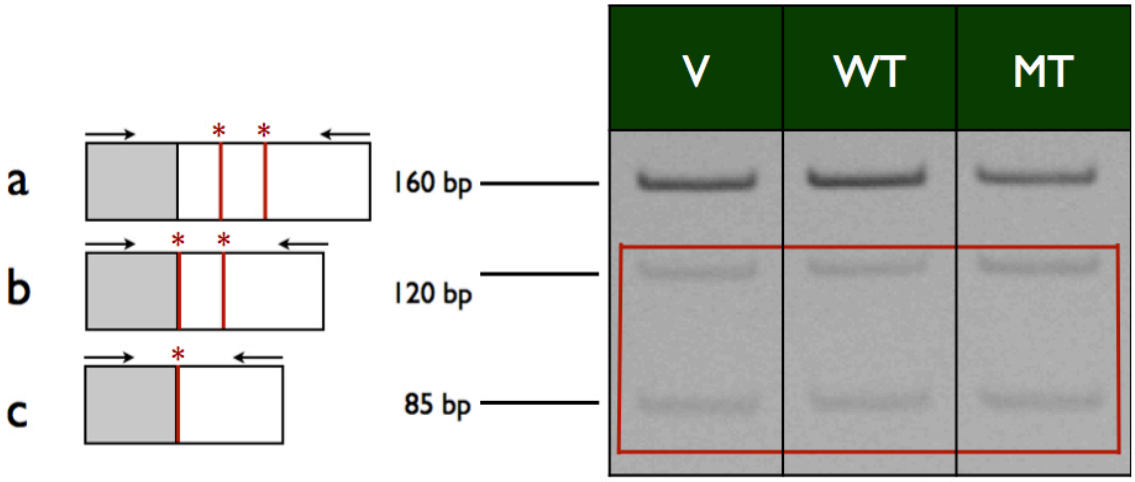


Figure 6. cDNA was obtained from c-kit+ murine bone marrow cells that either expressed empty vector+GFP, wild-type U2AF1+GFP, or S34F mutant U2AF1+GFP. These cells were not treated with either Sudemycin or DMSO. The region spanning Exons 17 and 18 in the *FMR1* gene was amplified using PCR and ran on a 10% polyacrylamide gel. Three transcripts were seen. Transcript (a) utilized normal splice sites, while transripts (b) and (c), highlighted in the red box, utilized cryptic splice sites within exon 18.

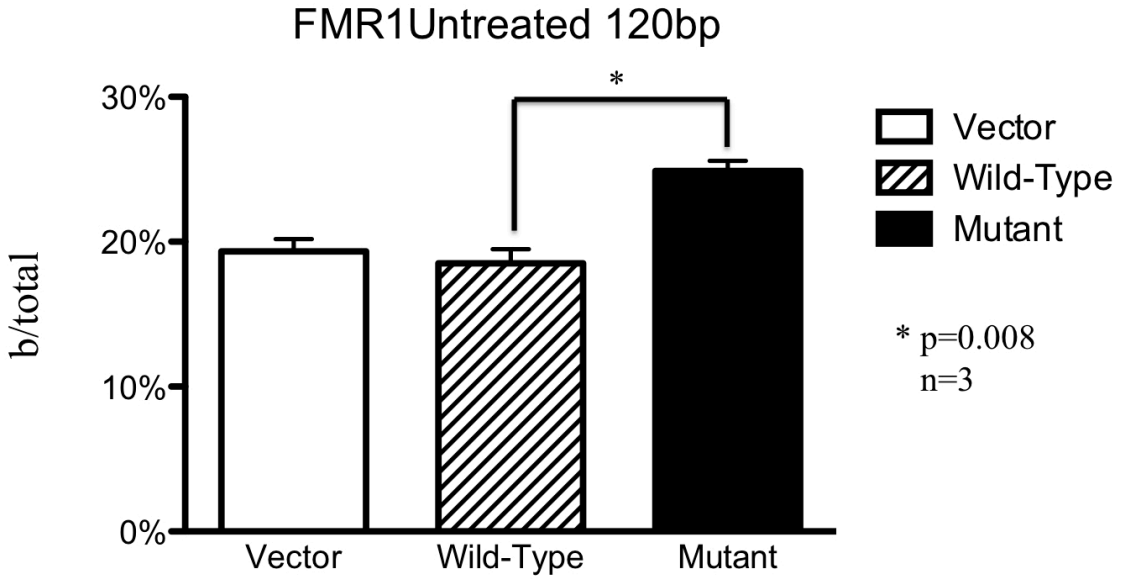


Figure 7. Image J was used to quantify the percentage of transcript abundance after in the polyacrylamide gels seen in Figure 6. Here, the percentage of the 120 bp cryptic splice site transcript (b) was calculated for all 3 virus types. A significant increase in utilization of the upstream cryptic splice site was observed in cells expressing S34F mutant U2AF1 when compared to cells expressing wild-type U2AF1. n=3, p=0.0008.

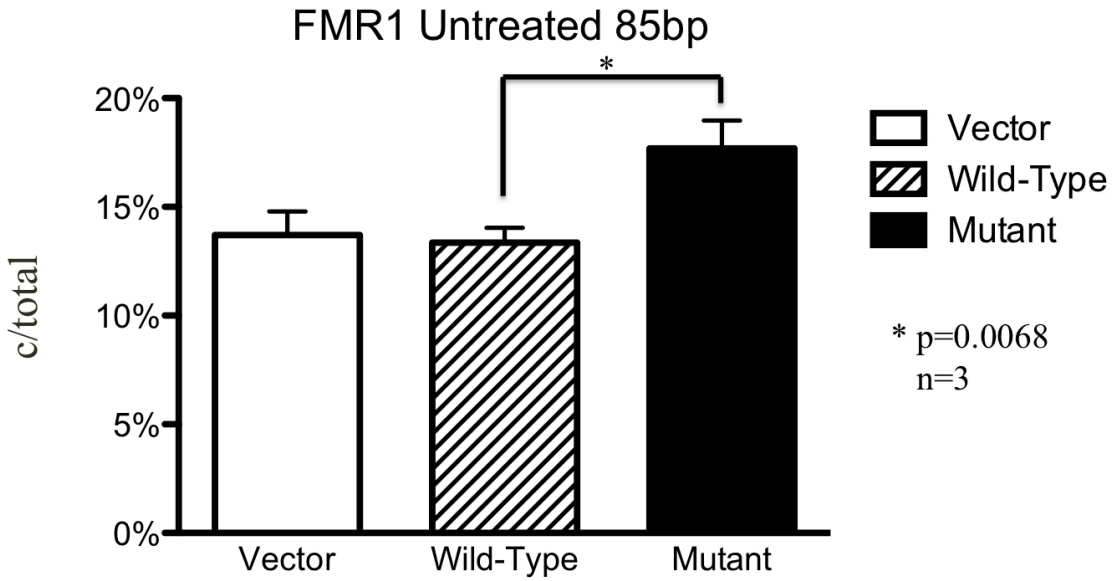


Figure 8. Image J was used to quantify the percentage of transcript abundance in the polyacrylamide gels seen in Table 6. Here, the percentage of the 85bp cryptic splice site transcript (c) was calculated for all 3 virus types. A significant increase in utilization of the downstream cryptic splice site was observed in cells expressing S34F mutant U2AF1 when compared to cells expressing wild-type U2AF1. n=3, p=0.0068.

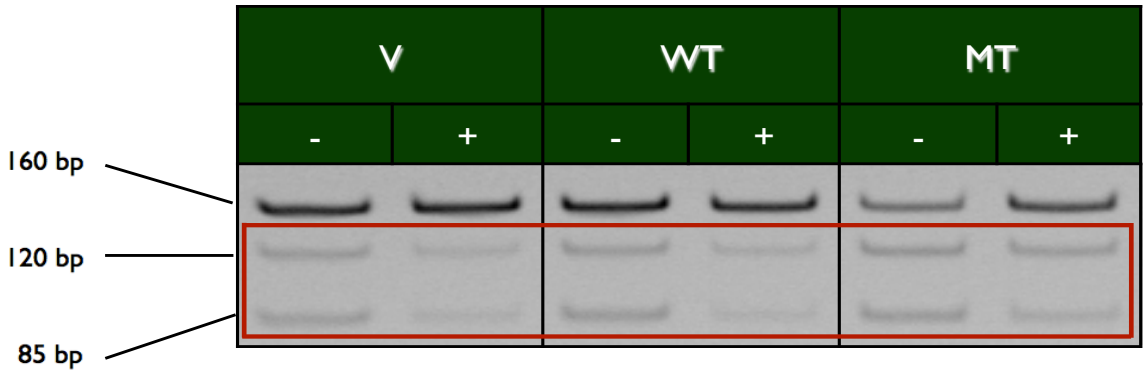


Figure 9. cDNA was obtained from c-kit+ murine bone marrow cells that either expressed empty vector+GFP, wild-type *U2AF1*+GFP, or S34F mutant *U2AF1*+GFP. These cells were treated with either DMSO (-) or 75 nM Sudemycin (+) for 6 hours. The region spanning Exons 17 and 18 in the *FMR1* gene was amplified using PCR and ran on a 10% polyacrylamide gel. Three transcripts were seen. Transcript (a) utilized normal splice sites, while transcripts (b) and (c), highlighted in the red box, utilized cryptic splice sites within Exon 18.

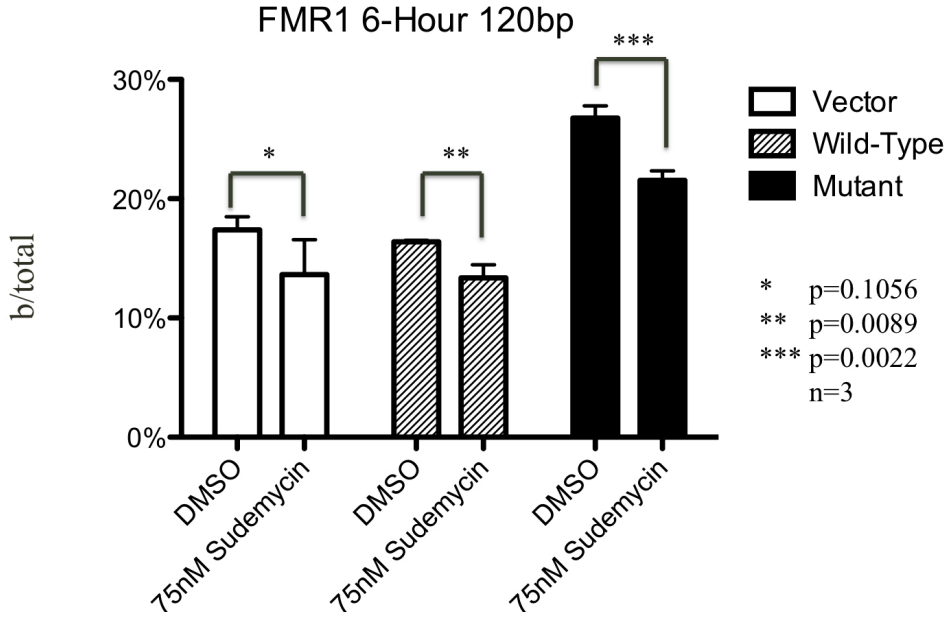


Figure 10. Image J was used to quantify the percentage of transcript abundance in the 10% polyacrylamide gels seen in Figure 9. Here, the percentage of the 120 bp cryptic splice site transcript (b) was calculated for all three virus types and both treatments. A significant decrease in utilization of the upstream cryptic splice site was observed in cells expressing both S34F mutant *U2AF1* and wild-type *U2AF1* after treatment with 75nM Sudemycin when compared to cells treated with DMSO. n=3, WT p=0.0089, MT p=0.0022.

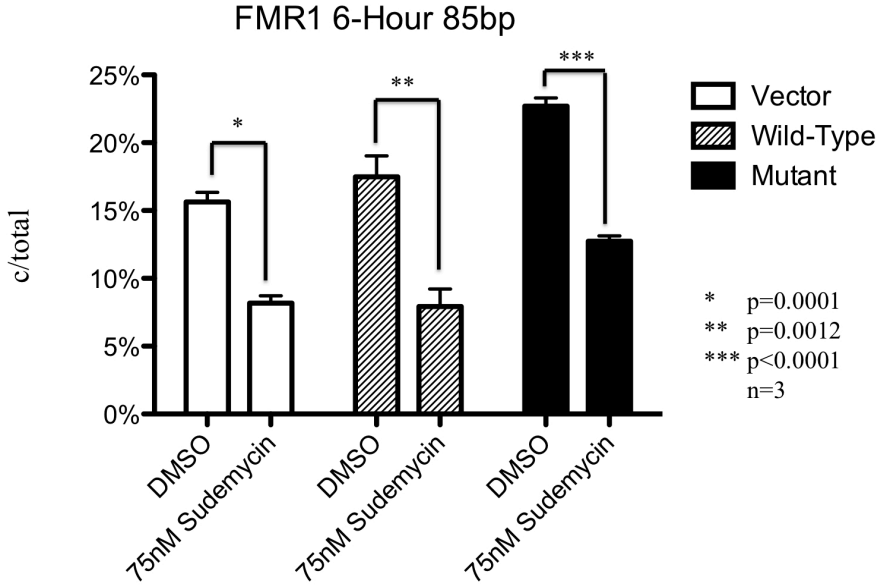


Figure 11. Image J was used to quantify the percentage of transcript abundance in the polyacrylamide gels seen in Figure 9. Here, the percentage of the 85 bp cryptic splice site transcript (c) was calculated for all 3 virus types and both treatments. A significant decrease in utilization of the downstream cryptic splice site was observed in cells expressing both S34F mutant *U2AF1* and wild-type *U2AF1* after treatment with 75nM Sudemycin when compared to cells treated with DMSO. n=3, WT p=0.0012, MT p<0.0001.