Research Article

An In Silico Analysis of CRISPR Therapy Targeting NPM1 Mutations in Acute Myeloid Leukemia

Alexandra T. Hedrick

Charlotte Country Day School, Charlotte, North Carolina, United States

ABSTRACT

Acute myeloid leukemia (AML) is a blood cancer typically affecting myeloid cells in the bone marrow. It is most common in adults and is characterized by its aggressive nature and relatively high likelihood of relapse. One of the most frequent mutations observed in patients with AML is a mutated NPM1 gene, which may lead to aberrant cytoplasmic localization of the NPM1 protein. FLT3 mutations are also commonly linked to NPM1 mutations and often signify poor prognosis in individuals with AML. Though research is being done into the prevention of these mutations, there are still very few treatments specifically targeting NPM1-mutated AML. In this study, a hypothetical CRISPR/Cas9 gen editing strategy was proposed as a possible path to study approaches for therapy to target mutant NPM1 genes and subsequent mutations in FLT3. Techniques such as transmembrane domain analysis and multiple protein alignment were used to explore protein characteristics such as localization and amino acid conservation patterns. A hypothetical CRISPR/Cas9 gene editing, polymerase chain reaction (PCR), and gel electrophoresis were simulated in exon 12 (a region commonly associated with NPM1 mutations) to propose a future research path for investigating the effects of mutations in the NPM1 gene. The goal was to generate a strategy to investigate whether a precise gRNA applied as a CRISPR/Cas9 gene editing could be a potential solution to limit the effect of NPM1 and FLT3 mutated proteins, therefore decreasing the development and severity of AML in the future.

Keywords: AML; CRISPR/Cas9; Nucleophosmin (NPM1); FLT3; PCR; Gel Electrophoresis; Cancer

INTRODUCTION

The human body contains billions of cells, which develop and differentiate to form the basis of tissues and organs. Various tissues and organs have different timelines for cell turnover. One of the reasons for cell division is to replace dead cells. Cells replicate their DNA to divide and replace dead cells when possible. The proliferation and apoptosis (programmed cell death) of cells usually occur in a controlled manner. Cancer is a disease characterized by abnormal cellular growth that spreads to other areas of the body. The carcinogenic process often begins when mutations occur in the genes that control the homeostatic mechanisms related to proliferation and apoptosis in cells. These

Corresponding author: Alexandra T. Hedrick, E-mail: alexhedrick@icloud.com.

Accepted September 15, 2025
https://doi.org/10.70251/HYJR2348.35249266

mutations can include point mutations, which involve single-nucleotide changes, nonsense mutations, which result in a premature stop codon, and frameshift mutations, which result from nucleotide insertion or deletion (1). External factors can also increase the likelihood of mutations, most notably risk factors such as smoking, exposure to radiation, and an unhealthy diet. A mutation is just the first step in the development of a cancer. Cells contain checkpoints throughout the cell cycle that are meant to identify DNA mutations to halt the proliferation of damaged cells and induce cell repair or apoptosis (2). Cells also have proto-oncogenes and tumor suppressor genes to control cell division. In the event of a mutation, proto-oncogenes can become carcinogenic oncogenes where the growth protein is overstimulated, resulting in rapid cell division. Conversely, mutations in tumor suppressor genes result in the gene becoming inactivated, which similarly leads to an overproduction of mutant cells. As these cells continue to divide, the mutant cells form colonies of carcinogenic cells that could give rise to tumors (3).

AML is a form of cancer affecting the bone marrow and hematopoietic stem cells. AML accounts for roughly 1% of all cancers and occurs much more frequently in adults over the age of 45 (4). In most cases, AML is caused by random genetic mutations in new blood stem cells made in the bone marrow, resulting in the multiplication of damaged myeloid cells, which would ordinarily develop into white or red blood cells (5). Though less frequent, AML can also result from hereditary mutations, previous chemotherapy treatments, and chemical exposure. AML is further classified into six different subtypes—AML with recurrent genetic abnormalities, AML with myelodysplasia-related changes, Therapy-related myeloid neoplasms, AML Not Otherwise specified, Myeloid Sarcoma, and Myeloid proliferations related to Down Syndrome—and three categories of severity—favorable, intermediate, and advanced. Given its varied specifications, there is a wide range of symptoms relating to different types of AML (6). Some patients may experience increased fatigue, a recurrence of infection, or pains in their bones and joints due to the disease's target of bone marrow and blood cells. (7).

One of the clearest ways to diagnose leukemia is with a blood sample. A blast count of over 20% in adults, particularly, is usually a strong indicator of AML, though the disease can occur even in people with far lower blast counts (8). Blast cells are immature cells that develop from stem cells before becoming mature cells. While they represent a normal step in the development of blood cells in the bone marrow, blasts exist in the early stages of cell development and thus should not be found in the bloodstream with mature blood cells (9). In the case of a stable blast count, karyotyping can be used to visualize cell chromosomes to look for cytogenetic abnormalities or chromosomal errors. This technique is also useful to subcategorize AML diagnoses into their level of severity based on a person's karyotype (8). The most common treatment option for most types of AML consists of multiple rounds of multi-agent cytotoxic chemotherapy; however, less intense hypo-methylating agents (HPMs) are also utilized, especially in more advanced cases where aggressive chemotherapy is too harsh. HPMs work to undo the process of DNA methylation by re-expressing mutated genes that have been silenced by promoter methylation, leading to the development of cancer. Unfortunately, AML has a high relapse rate, as any remaining leukemic stem cells (LSCs) can potentially repopulate due to their high mitotic activity (10).

The most common genetic lesion in AML patients is a mutated nucleophosmin (NPM1) gene, a protein centrally located in the nucleolus, which travels between the nucleus and the cytoplasm. In 2017, as the most common type of mutation linked to AML patients (seen in approximately 30% of adult cases), the World Health Organization (WHO) recognized NPM1-mutated AML as its own distinct category under the classification of hematopoietic neoplasms (11). NPM1 is a multifaceted protein involved in several important biological processes, some of which are still being discovered. As a nucleolar chaperone protein, it works to regulate both protein folding and chromatin regulation by utilizing its ability to shuttle between the cytoplasm and the nucleus. It also plays a key role in the processing of pre-ribosomal RNA (pre-rRNA) and the subsequent maturation to rRNA and ribosome biogenesis (the production of ribosomes) in the nucleolus by using its chaperone abilities to prevent protein accumulation during assembly (12). As a key part of controlling the cell cycle, NPM1 binds with the tumor suppressor gene, TP53, to help promote its stability so it can initiate cell cycle arrest or apoptosis when interacting with mutant cells. It also assists the protein repair gene APE1 to promote endonuclease activity that stops the proliferation of damaged DNA (13).

NPM1 is a gene comprised of 12 exons on chromosome 5q35. It codes for a protein comprised of 294 amino acids. NPM1 contains a nuclear export signal (NES) in its N-terminus domain and two nuclear

localization signals (NLS and NoLS) in its Central domain and C-terminus domain, respectively, all of which aid its functionality as a chaperone protein. Exon 12 is the most frequent location for NPM1 mutations. Almost 80% of NPM1 mutations are Type A, which consists of a tetranucleotide duplication of its previous TCTG sequence. While the 7 other common types of NPM1 mutations insert various tetranucleotide pairings in different locations on exon 12, all result in frameshift mutations that cause an extension of its normal 294 amino acid chain by 4 amino acids (total of 298). Most importantly, all mutations change the position of the tryptophan amino acid at position 288 or 290, which can cause aberrant cytoplasmic NPM1 localization through the loss of the NoLS and the exertion of a stronger NES, which causes accumulation in the cytoplasm. It is believed that the accumulation of mutant NPM1 genes initiate the leukemic state by upregulating the class I homeobox (HOX) genes, which are key for the production and renewal of blood stem cells (14).

NPM1 mutations in AML are always heterozygous, meaning they carry both the mutated NPM1 and wild-type NPM1. One limitation of therapies targeting NPM1c+ (the mutated form of NPM1 in AML, which causes the protein to become mislocalized in the cytoplasm) has to do with the difficulty of differentiating the mutant and wild-type form of NPM1, the latter of which is necessary for the survival of the cell (15). While it is better understood how NPM1c+ mutations can trigger leukemogenesis, scientists are still working to understand the mechanisms that cause NPM1c+ AML specifically. Given the heterozygous nature of the mutation, one avenue of exploration stems from the idea that the remaining wild type NPM1 allele gets weakened by its mutant counterpart and is then relocated to the nucleus and cytoplasm, where it forms heterodimers with the NPM1c+. The second option is that the NPM1c+ allele forms a new mutant NPM1 protein—localized in the cytoplasm—which dislocates other interacting proteins and molecules to the cytoplasm, therefore restricting their intended function as well (12). In the case of NPM1c+, one such affected molecule is the proteases caspase 6 and 8, which become impaired so that they cannot initiate apoptosis of the cell and mediate myeloid differentiation, a key process to differentiate hematopoietic stem cells in the bone marrow (16).

Commonly linked to NPM1c+ mutations are mutations in the FLT3, a cytokine receptor that belongs to the receptor tyrosine kinase class III. The two most common FLT3 mutations are internal tandem duplication (FLT3-ITD) in the area that codes for the regulatory juxta-membrane domain, and the deletion of codons 835 and 836 (FLT3-D835) in the tyrosine kinase domain of the gene (17). Several studies have explored the correlation between NPM1c+/FLT3-ITD mutations and prognosis in AML patients. Patients with NPM1c+ mutations without FLT3-ITD are generally seen to have far more favorable prognoses than those with both mutations. It is also recognized that generally NPM1c+ mutations occur before FLT3-ITD mutations, which have around a 40% likelihood of developing in patients with the NPM1c+ mutation (18). Sole NPM1c+ mutations have been found easier to treat because these mutations do not occur in normal tissues, thus making them easier to target, without the compounding variable of the FLT3-ITD mutation (19). In patients with the FLT3-ITD mutation, researchers are working to develop FLT3 inhibitors, though they are limited by genes such as XPO1 and BCL2, which can inhibit the efficacy of these drugs (20). However, in most AML cases, even patients with the more favorable NPM1c+ positive/FLT3-ITD negative combination, there is still an almost 50% relapse rate. In the case of relapse, there are fewer treatments proven to be effective at targeting reappeared NPM1c+ and other mutations (18).

A gene therapy could be the future of AML treatment. With the CRISPR/Cas 9, it would be possible to correct errors in the genome (such as those caused by genetic mutations) and to turn on/off genes of interest (21). The CRISPR/Cas9 gene editing technique has two main components: a guide RNA (sgRNA) and the Cas9 enzyme. This endonuclease gets attached to the sgRNA, which recognizes specific DNA sequences. With the sgRNA as a tool to identify the target gene, the Cas9 enzyme can then make a cut in the DNA as long as it recognizes a protospacer adjacent motif (PAM) sequence. After the DNA is cut, CRISPR can be used for nucleotide insertion/deletion with the goal of knocking out the gene (causing it to become nonfunctional). It may also utilize homology-directed repair mechanism to repair damaged DNA by inserting sequences. While CRISPR technology exhibits promising results for the future of gene editing, it is not without its limitations. There is a high reported frequency of off-target effects (OTEs) where healthy genes may be altered or mutated (22).

The aim of this study is to design a hypothetical CRISPR/Cas9 gene editing approach to target AML-related genetic mutations for future studies. In looking at the correlation between NPM1c+ mutations and

subsequent FLT3-ITD mutations, I will design a CRISPR approach to hypothetically examine the possibility of impairing the functionality of the NPM1c+ mutated protein with the hope that it may prevent the potential spread of NPM1c+ mutations. By targeting the NPM1 gene, I am also hypothesizing that this editing might alter the occurrence of a second FLT3-ITD mutation, given that an NPM1c+ mutation is thought to occur before FLT3-ITD.

METHODS AND MATERIALS

Protein sequence retrieval

The National Center for Biotechnology Information (NCBI) provides access to biomedical and genomic information, including amino acid sequences of proteins. Protein sequences taken from NCBI>Protein>FASTA.

The amino acid sequence of the human NPM1 protein, FLT3 protein, and their mutated forms were obtained in FASTA format from the NCBI's protein database. It was used to get the amino acid FASTA sequences of NPM1 and FLT3 for other animals, including mus musculus (mouse), pan troglodytes (chimpanzee), gallus gallus (chicken), danio rerio (zebrafish), and mixophyes fasciolatus (great barred frog). Additionally, NCBI's gene database provided both the view of NPM1 on Chromosome 5 and the gene FASTA sequence for NPM1.

Transmembrane prediction

The Transmembrane Helix Prediction Model (TMHMM) is a program under the Department of Health Technology to help predict transmembrane helices in proteins using an amino acid sequence. I submitted the protein sequences to run the analysis for the prediction.

The FASTA sequences were pasted into TMHMM to generate the prediction plots for the transmembrane domain analyses of NPM1, FLT3, and their two mutant forms.

Multiple Sequence alignment

Clustal Omega is a sequence alignment program that generates alignments for three or more protein, DNA, or RNA sequences. Using Clustal Omega with parameters ClustalW with character counts, the FASTA sequences for the animals were aligned, and a multiple protein alignment table was generated to determine the conservation of amino acids in each protein throughout the evolutionary process.

CRISPR/Cas9 gene editing strategy design and simulated PCR/DNA gel electrophoresis

Ensembl contains genome information to aid research in genomics, evolution, and protein transcription and sequencing. Its technology enables researchers to annotate genes, collect disease data, and predict transcriptional regulatory function. SnapGene is a design platform that enables scientists to design, visualize, and document Polymerase Chain Reactions and cloning procedures. CRISPR Finder (Sanger) is a program allowing researchers to identify interspaced short palindromic repeats (CRISPR) and Cas genes from submitted sequencing data.

The FASTA sequence of NPM1 was pasted into SnapGene's DNA sequence section to visualize the NPM1 gene, and the enzymes on the gene were hidden from view. Using Ensembl, the longest NPM1 gene (323 aa—protein coding) was selected, which contains a total of 12 intron and 12 exon regions. Exon regions 11 and 12 were copied into SnapGene to visually mark these segments in the gene. The gap in between was labeled intron 11. The NPM1 human gene was then searched in CRISPR Finder (Sanger) to find the best sgRNA to cut in the exon 12 region, based on the sgRNA with the least off-target sequences. The database only showed sgRNA up to exon 11 (length 611 in CRISPR Finder; however, the chosen sequence goes into the exon 12 region identified in Ensembl). After finding the best sgRNA, the corresponding sequence was also mapped in SnapGene. A PAM sequence was identified after the sgRNA (it must come before or after the guide), and the potential cut site for Cas9 protein was also marked between the guide sequence and the PAM site. The cut is expected to cause a mutation in the NPM1 gene. In a wet lab experiment, millions of cells would be given the Cas9 protein and sgRNA, and the results would be tested by isolating the cell DNA, amplifying it with PCR, and sending it for sequencing after gel electrophoresis.

PCR is an in vitro DNA synthesis technique used to amplify small DNA segments. This technique is a fast and inexpensive way to produce multiple copies of a specific DNA fragment for lab analysis, such as CRISPR (23). In the lab, DNA extracted from an organism is too long to be directly analyzed, especially in gel electrophoresis. Thus, PCR utilizes the polymerase enzyme and two primers to amplify the sequences of the gene of interest so that changes in a specific gene (such as nucleotide insertions and deletions of 200 bp or less) may be visible. In a PCR reaction, double-stranded DNA (dsDNA) is heated to roughly 90-96°C

to denature the DNA helix and separate the two strands into single-stranded DNA (ssDNA). The next step, known as annealing, requires two primers which bind to their complementary nucleotide bases on the single strands of DNA at the beginning and end of the target gene (one primer amplifies the strand complementary to the top in the forward direction, the other primer amplifies the strand complementary to the bottom in the reverse direction). The primers are chains of roughly 20 nucleotides that are specific to the target gene. Then, beginning at the primer, the DNA polymerase enzyme will read the template DNA and match it with the complementary nucleotides up to the second primer, in the process of elongation. Afterwards, the process can be repeated, doubling the target sequence each time to create multiple copies of the fragment (24).

Whereas PCR generates copies of the gene, Sanger sequencing is used to read these DNA sequences. This allows researchers to understand nucleotide variation in the sample, such as the presence of additions or deletions. The fundamental basis of Sanger sequencing is the following: first, the dsDNA is denatured into ssDNA. Instead of two primers binding to the strand, only one primer binds to one end of the corresponding sequence. Next, four polymerase solutions are added, which contain dNTPs (A, G, C, or T) that create a complementary strand to the template DNA. Traditional early Sanger sequencing uses one of four ddNTPs (either ddATP, ddGTP, ddCTP, or ddTTP), which terminate the chain reaction and complete synthesis (25). Later, the technology developed a way of labelling the dNTPs with fluorescent probes. After every dNTP joined to the reaction, the fluorescence was captured by a camera to be recorded as an image, followed by analysis of nucleotides corresponding to their fluorophore color. Currently, there are massively developed sequencing technologies where the basic method has been enhanced rapidly.

DNA gel electrophoresis is used to separate DNA fragments ranging from 100bp to 25 kb[1] by size. The DNA gets loaded into wells in the gel. Due to the DNA's negatively charged phosphate backbone, when an electric current is applied, the DNA will travel to the positively charged anode. The size of the DNA fragment determines how far it will travel in the gel, with the shortest fragments traveling the farthest (26).

Though not all these techniques were applied practically in this experiment, PCR results were simulated in a gel view. Primers were added in SnapGene to identify the target segment of DNA that would be amplified with PCR. Then, a sample gel was modeled to demonstrate possible outcomes of Cas9 cut on exon 12 in NPM1.

RESULTS

Transmembrane domain Prediction analysis of wild-type NPM1 and FLT3 versus mutants

Proteins can be structural, enzymatic, or both. These functions may be disrupted as an outcome of mutations in the gene. The structural protein requires the hydrophobic segment to be inserted into membrane structures. Here, I did TMHMM analysis for NPM1 and FLT3, both for wild-type and mutant (Figures 1-4), in order to see the TMHMM domain prediction of each protein and understand whether mutant versions may develop mislocalization.

Multiple protein alignment of wild-type NPM1 and FLT3

As we share DNA sequence with other species, this reflects on protein levels as well. In order to understand which amino acids are conserved and which are not across different species, one can run Clustal Omega multiple protein alignment analysis. I compared the protein sequence of NPM1 and FLT3 across zebrafish, chicken, mouse, chimpanzee, and human (Figures 4-5). Some amino acids have a more conserved score compared to others. If the mutation affects a conserved amino acid, this could lead to a severe outcome compared to non-conserved amino acids.

Design of CRSIPR/Cas9 strategy to target NPM1 gene

I aimed to hypothetically target the exon 12 of the NPM1 gene (Figure 7) in order to develop a potential gene editing approach for studying a potential prevention effect regarding the NPM1 mutation in AML. Gene editing can be considered as a therapy for AML, and approaching NPM1 gene may lead to the prevention of AML progression and relapse. I targeted exon 12 of NPM1 gene, and the sequence is visualized in SnapGene to represent the targeted region of the exon and the predicted gRNA (Figure 8) via Crispr Finder (Figure 9). Hypothetically, the gRNA and the Cas9 enzyme were considered to be given to the cells, and the potential mutation outcome is simulated on a DNA gel electrophoresis (Figure 10). This outcome is solely hypothetical and the mutations were decided based on expectation for training and presentation purpose.

RecName: Full=Nucleophosmin; Short=NPM; AltName: Full=Nucleolar phosphoprotein B23; AltName: Full=Nucleolar protein NO38; AltName: Full=Numatrin

UniProtKB/Swiss-Prot: P06748.2

GenPept Identical Proteins Graphics

```
>sp|P06748.2|NPM_HUMAN RecName: Full=Nucleophosmin; Short=NPM; AltName: Full=Nucleolar
phosphoprotein B23; AltName: Full=Nucleolar protein NO38; AltName: Full=Numatrin
MEDSMDMDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELHIVEAEAMNYEGS
PIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVAVEEDAESEDEEEEDVKLLSISG
KRSAPGGGSKVPQKKVKLAADEDDDDDEEDDDEDDDDDFDDEEAEEKAPVKKSIRDTPAKNAQKSNQN
GKDSKPSSTPRSKGQESFKKQEKTPKTPKGPSSVEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTD
QEAIQDLWQWRKSL
```

TMHMM result

```
# sp|P06748.2|NPM_HUMAN Length: 294
# sp|P06748.2|NPM_HUMAN Number of predicted TMHs:  0
# sp|P06748.2|NPM_HUMAN Exp number of AAs in TMHs: 0.00136
# sp|P06748.2|NPM_HUMAN Exp number, first 60 AAs:  0
# sp|P06748.2|NPM_HUMAN Total prob of N-in:        0.03239
sp|P06748.2|NPM_HUMAN   TMHMM2.0        outside      1   294
```

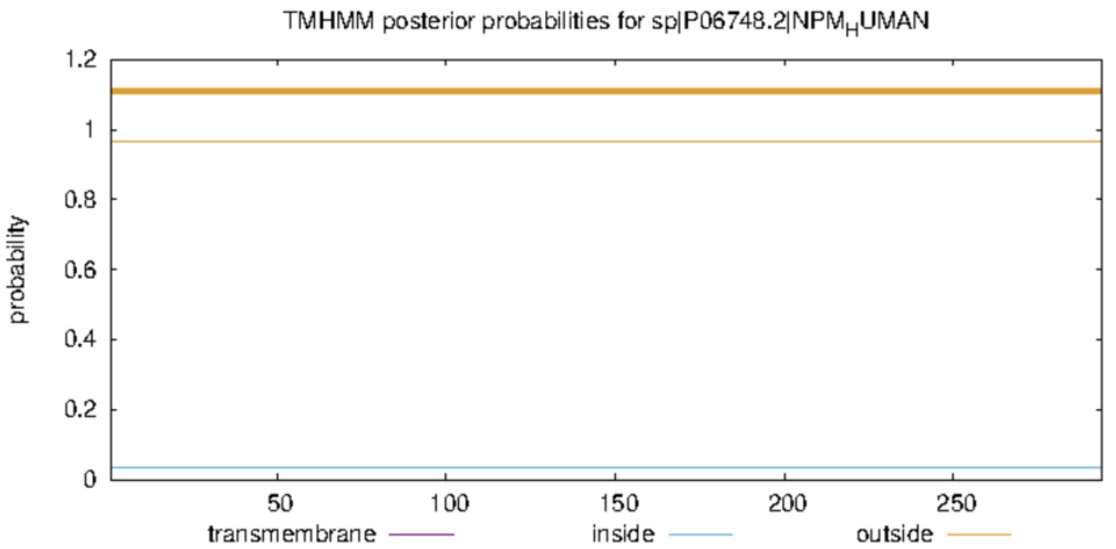


Figure 1. NPM1 doesn't contain a transmembrane domain. Upper image. NPM1 protein sequence in FASTA format taken from NCBI>Protein. **Lower image.** The prediction plot of NPM1 transmembrane domain analysis. The y-axis shows the certainty of each segment in the predicted NPM1 model. The probability between 1-1.2 represents the N-best prediction. A low probability displays possible weak transmembrane helices within the protein. The x-axis shows what parts of a protein sequence will be inside the membrane (blue), outside the membrane (yellow), or span across the membrane (purple). Given that all 294 amino acids are predicted to be outside, the prediction is that the protein contains no transmembrane domain (27).

mutant nucleolar phosphoprotein B23 [Homo sapiens]

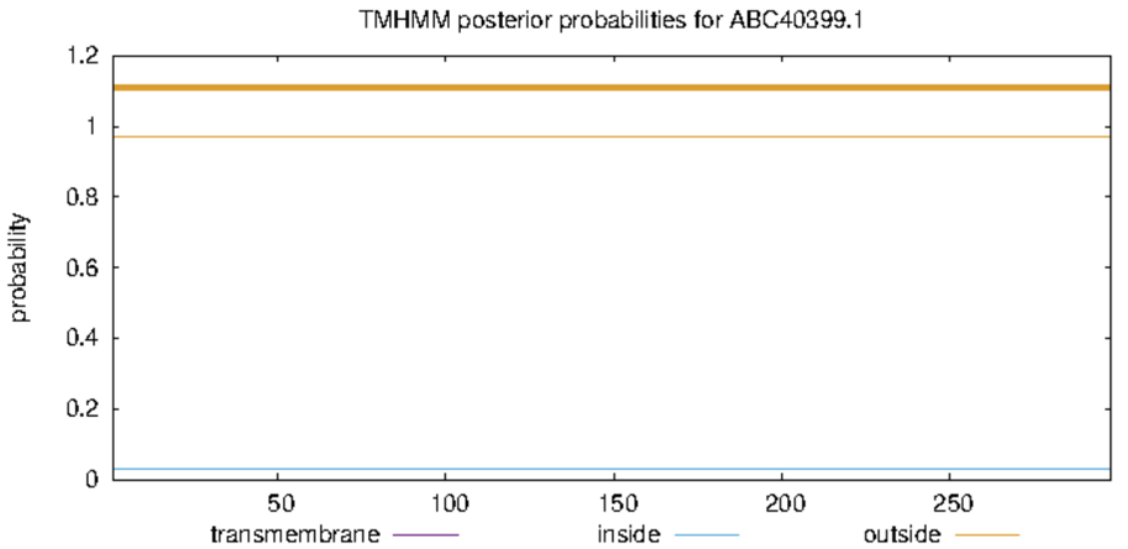
GenBank: ABC40399.1

GenPept Identical Proteins Graphics

```
>ABC40399.1 mutant nucleolar phosphoprotein B23 [Homo sapiens]
MEDSMDMDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELHIVEAEAMNYEGS
PIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVAVEEDAESEDEEEEDVKLLSISG
KRSAPGGGSKVPQKKVKLAADEDDDDDEEDDDEDDDDDFDDEEAEEKAPVKKSIRDTPAKNAQKSNQN
GKDSKPSSTPRSKGQESFKKQEKTPKTPKGPSSVEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTD
QEAIQDLCMAVEEVSLRK
```

TMHMM result

```
# ABC40399.1 Length: 298
# ABC40399.1 Number of predicted TMHs:  0
# ABC40399.1 Exp number of AAs in TMHs: 0.00133
# ABC40399.1 Exp number, first 60 AAs:  0
# ABC40399.1 Total prob of N-in:        0.03141
ABC40399.1      TMHMM2.0        outside      1   298
```



Note: B23 is another name for the NPM1 protein.

Figure 2. Mutated NPM1 (NPM1mut) doesn't contain a transmembrane domain. **Upper image**. Mutant NPM1 protein sequence in FASTA format taken from NCBI>Protein. **Lower image**. The prediction plot of mutant NPM1 transmembrane domain analysis. The y-axis shows the certainty of each segment in the predicted mutant NPM1 model. Given the N-best probability that all 298 amino acids are outside, the prediction is that the protein contains no transmembrane domain.

receptor-type tyrosine-protein kinase FLT3 precursor [Homo sapiens]

NCBI Reference Sequence: NP_004110.2

GenPept Identical Proteins Graphics

```
>NP_004110.2 receptor-type tyrosine-protein kinase FLT3 precursor [Homo sapiens]
MPALARDGGQLPLLVVFSAMIFGTITNQDLPVIKCVLINHKNNDSSVGKSSSYPMVSESPEDLGCALRPQ
SSGTVYEAAAVEVDVSASITLQVLVDAPGNISCLWVFKHSSLNCQPHFDLQNRGVVSMVILKMTETQAGE
YLLFIQSEATNYTILFTVSIRNTLLYTLRRPYFRKMENQDALVCISESVPEPIVEWVLCDSQGESCKEES
PAVVKKEEKVLHELFGTDIRCCARNELGRECTRLFTIDLNQTPQTTLPQLFLKVGEPLWIRCKAVHVNHG
FGLTWELENKALEEGNYFEMSTYSTNRTMIRILFAFVSSVARNDTGYYTCSSSKHPSQSALVTIVEKGFI
NATNSSEDYEIDQYEEFCFSVRFKAYPQIRCTWTFSRKSFPCEQKGLDNGYSISKFCNHKHQPGEYIFHA
ENDDAQFTKMFTLNIRRKPQVLAEASASQASCFSDGYPLPSWTWKKCSDKSPNCTEEITEGVWNRKANRK
VFGQWVSSSTLNMSEAIKGFLVKCCAYNSLGTSCETILLNSPGPFPFIQDNISFYATIGVCLLFIVVLTL
LICHKYKKQFRYESQLQMVQVTGSSDNEYFYVDFREYEYDLKWEFPRENLEFGKVLGSGAFGKVMNATAY
GISKTGVSIQVAVKMLKEKADSSEREALMSELKMMTQLGSHENIVNLLGACTLSGPIYLIFEYCCYGDLL
NYLRSKREKFHRTWTEIFKEHNFSFYPTFQSHPNSSMPGSREVQIHPDSDQISGLHGNSFHSEDEIEYEN
QKRLEEEEDLNVLTFEDLLCFAYQVAKGMEFLEFKSCVHRDLAARNVLVTHGKVVKICDFGLARDIMSDS
NYVVRGNARLPVKWMAPESLFEGIYTIKSDVWSYGILLWEIFSLGVNPYPGIPVDANFYKLIQNGFKMDQ
PFYATEEIYIIMQSCWAFDSRKRPSFPNLTSFLGCQLADAEEAMYQNVDGRVSECPHTYQNRRPFSREMD
LGLLSPQAQVEDS
```

TMHMM result

```
# NP_004110.2 Length: 993
# NP_004110.2 Number of predicted TMHs:  2
# NP_004110.2 Exp number of AAs in TMHs: 44.6266300000002
# NP_004110.2 Exp number, first 60 AAs:  2.38881
# NP_004110.2 Total prob of N-in:        0.11879
NP_004110.2     TMHMM2.0        outside      1   541
NP_004110.2     TMHMM2.0        TMhelix    542   564
NP_004110.2     TMHMM2.0        inside     565   673
NP_004110.2     TMHMM2.0        TMhelix    674   696
NP_004110.2     TMHMM2.0        outside    697   993
```

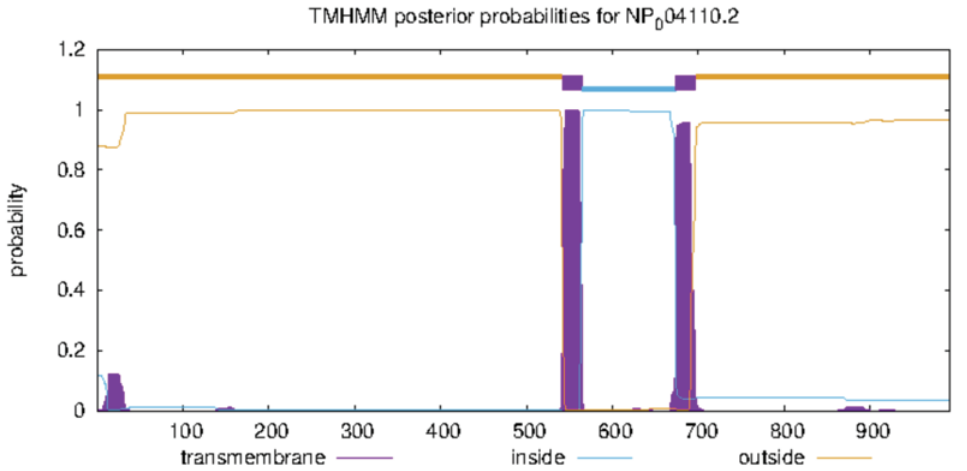


Figure 3. FLT3 contains 2 transmembrane domains and 1 possible weak TM domain, though relatively unlikely. **Upper image**. FLT3 protein sequence in FASTA format taken from NCBI>Protein. **Lower image**. The prediction plot of FLT3 transmembrane domain analysis. The y-axis shows the certainty of each segment in the predicted FLT3 model. Based on the model, the N-best prediction is that between amino acids (aa) 1-541, the loop region of FLT3 is outside the membrane. Between aa 542-564, there is a transmembrane domain. There is another loop region between aa 565-673 inside the membrane. There is another transmembrane helix between aa 674-696. The remaining aa from 697-993 in FLT3 are predicted to be outside the membrane. There is also an estimated 2 aa within the first 60 aa in FLT3, which is represented by the purple segment at the beginning of the prediction plot. But the prediction is quite weak.

FLT3-ITD, partial [Homo sapiens]

GenBank: WHF58414.1

GenPept Identical Proteins Graphics

```
>WHF58414.1 FLT3-ITD, partial [Homo sapiens]
QFRYESQLQMVQVTGSSDNEYFYVDFREYEYDLKWEFPTENLQFGKVLGSGAFGKVMNATAYGISKTGVS
IQVAVKMLK
```

TMHMM result

```
# WHF58414.1 Length: 79
# WHF58414.1 Number of predicted TMHs:  0
# WHF58414.1 Exp number of AAs in TMHs: 0.26114
# WHF58414.1 Exp number, first 60 AAs:  0.12049
# WHF58414.1 Total prob of N-in:        0.43119
WHF58414.1      TMHMM2.0        outside      1    79
```

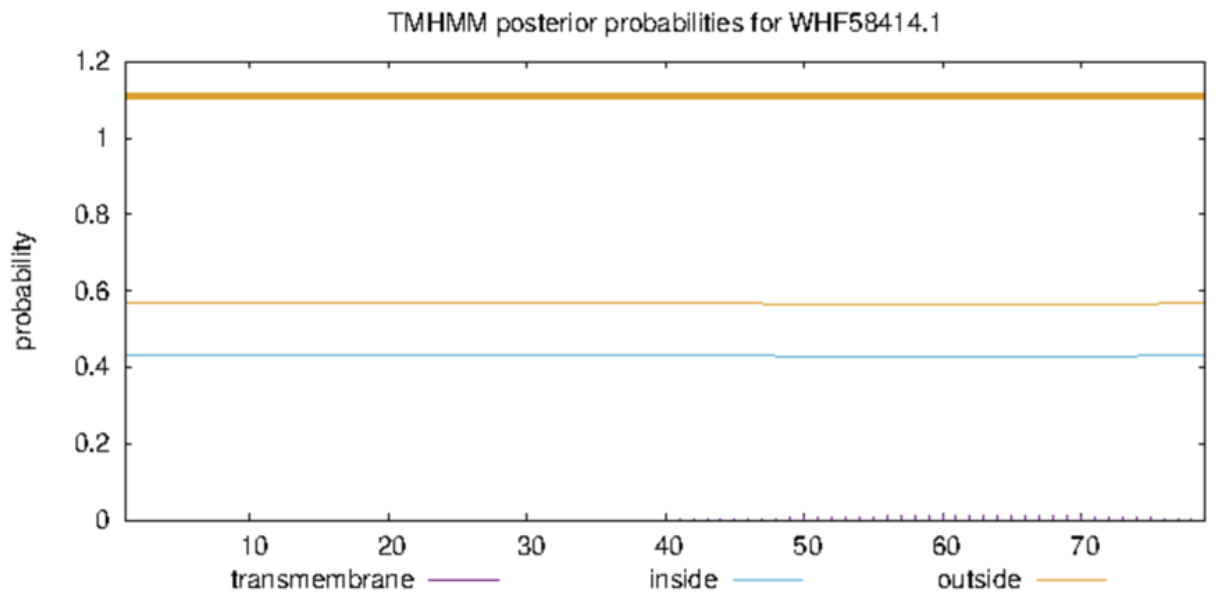
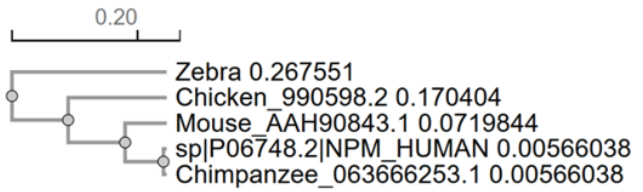


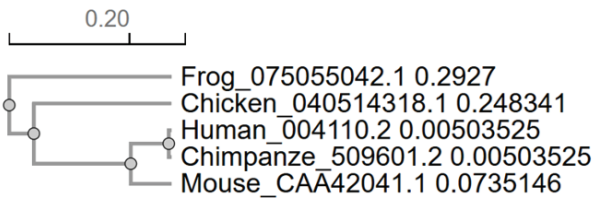
Figure 4. FLT3-ITD does not contain a transmembrane domain. Upper image. FLT3-ITD protein sequence in FASTA format taken from NCBI>Protein. **Lower image**. The prediction plot of FLT3-ITD transmembrane domain analysis. The y-axis shows the certainty of each segment in the predicted FLT3-ITD model. Given the N-best probability is that all 79 amino acids are outside in the cytosol, the prediction is that the protein contains no transmembrane domain. This conclusion is the same even for the less likely probability that FLT3 is inside the lumen region of the organelle.




```
CLUSTAL O(1.2.4) multiple sequence alignment


Zebra                        --------MDLEQMGPQTFLYGCELKAGKDITFNPEDDDYDHQLSVRMACVDPSTKDELN       52
Chicken_990598.2             MEDSAMDMESMGPLRPQTFLFGCELKAEKEYQFKVDDEENEHQLSLRTVTLGAGAKDELH       60
Mouse_AAH90843.1             MED-SMD-MNMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELH       58
sp|P06748.2|NPM_HUMAN        MED-SMD-MDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELH       58
Chimpanzee_063666253.1       MED-SMD-MDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELH       58
                                      .:  : **.:*:****** *:  *: :::: :****:* . :. .:****:

Zebra                        VVEIEGQDSEGQKVKAVLATLKPSTLPSVCLGGFEITPPVVFRLRTGSGPVHISGQHLVI       112
Chicken_990598.2             VVEAEALDYEGNPTKVVLASLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVYVSGQHLVA       120
Mouse_AAH90843.1             IVEAEAMNYEGSPIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVA       118
sp|P06748.2|NPM_HUMAN        IVEAEAMNYEGSPIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVA       118
Chimpanzee_063666253.1       IVEAEAMNYEGSPIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVA       118
                             :** *. : **.  *..**:** *. *:*.***********:**: *****::******

Zebra                        MGGDQSFDEEEEEEEEETVMTSKKRPA--LSTPKPSKKMKMDEEDDDDDD-EDDDDDDDD       169
Chicken_990598.2             LEEEPESEDEE-EDT-KIGNASTKRPASGGGAKTPQKKPKLSEDDEDDDEDED--DDEDD       176
Mouse_AAH90843.1             VEEDAESEDEDEEDV-KLLGMSGKRSAPGGGNKVPQKKVKLDEDDEDDDED-DEDDEDDD       176
sp|P06748.2|NPM_HUMAN        VEEDAESEDEEEEDV-KLLSISGKRSAPGGGSKVPQKKVKLAADEDDDDDDEEDDDEDDD       177
Chimpanzee_063666253.1       VEEDAESEDEEEEDV-KLLSISGKRSAPGGGSKVPQKKVKLAADEDDDDDDEEDDDEDDD       177
                             :  : . ::*: *:  :    * ** *   .   *.** *:  :::***:  :  *::**

Zebra                        EDDDEEDESEEESPVKEKKA---PAKPKTPTQNGKGPKPSTPAKQTP---QKGKKEQTPK       223
Chicken_990598.2             EDDLDDDEEEIKTPMKKPAREPAGKNMQKAKQNGKDSKPSTPASKTKTPDSKKDKSLTPK       236
Mouse_AAH90843.1             DDDFDEEETEEKVPVKKSVRDTPAKNAQKSNQNGKDLKPS-TPR-SKGQESFKKQEKTPK       234
sp|P06748.2|NPM_HUMAN        DDDFDDEEAEEKAPVKKSIRDTPAKNAQKSNQNGKDSKPSSTPR-SKGQESFKKQEKTPK       236
Chimpanzee_063666253.1       DDDFDDEEAEEKAPVK----------------------------KGQESFKKQEKTPK       207
                             :** :::* * : *:*                              .  .:. ***

Zebra                        TPQTPRTLADIKSKMMESVAKGVSLPKVQLKFENYVNNCFKGTDPKVVEELWKWRQTVK 282
Chicken_990598.2             TPKVPLSLEEIKAKMQASVDKGCSLPKLEPKFANYVKNCFRTEDQKVIQALWQWRQTL- 294
Mouse_AAH90843.1             TPKGPSSVEDIKAKMQASIEKAH---------------------------------- 257
sp|P06748.2|NPM_HUMAN        TPKGPSSVEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTDQEAIQDLWQWRKSL- 294
Chimpanzee_063666253.1       TPKGPSSVEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTDQEAIQDLWQWRKSL- 265
                             **: * :: :**:**  *: *.
```

Figure 5. Multiple protein alignment for NPM1 protein. Upper image. Phylogenetic tree of NPM1 gene in zebrafish, chicken, mouse, human, and chimpanzee. The tree represents the evolutionary relationship relating to these five animals with respect to NPM1 amino acid conservation. The "length" of the branches (horizontal distance) represents the evolutionary distance between species; a longer distance shows more change. In order of conservation, humans are most similar to the chimpanzee, then the mouse, then the chicken, then the zebrafish. **Lower image.** The protein sequence alignment of NPM1 across the 5 species. The numbers on the right represent the number of amino acids from the beginning of the alignment. An asterisk (*) below the row indicates complete amino acid conservation among species. A colon (:) indicates conservation between groups of strongly similar properties as below - roughly equivalent to scoring > 0.5 in the Gonnet PAM 250 matrix. A period (.) indicates conservation between groups of weakly similar properties as below - roughly equivalent to scoring =< 0.5 and > 0 in the Gonnet PAM 250 matrix (28, 29). If there is nothing underneath, this represents a non-conserved mutation. The Clustal X color scheme is used to indicate certain amino acid properties. Blue for hydrophobic amino acids, red is for positively charged amino acids, magenta is for negatively charged amino acids, and green is for polar amino acids (30). Results indicate various segments of conservation (marked with an asterisk) and segments of non-conservation (colon and period), mostly linked with chicken and zebrafish. Conserved NPM1 segments are the same color. Semi-conserved segments vary by one color (property).

A



B

```
Frog_075055042.1        GEPFVIRCWAMYNTYKFNIHLNIESKSK--MEYNEESCYQLDGSMVRVVYASSPAMESGI   347
Chicken_040514318.1     GEPLLIRCRAVYRNYKFRINVSFENKEVQQNRRFEGLEYQTDHSAIRTQYMFTPAARKGD   324
Human_004110.2          GEPLWIRCKAVHVNHGFGLTWELENKALEEGNYFEMSTYSTNRTMIRILFAFVSSVARND   324
Chimpanze_509601.2       GEPLWIRCKAVHVNHGFGLTWELENKALEEGNYFEMSTYSTNRTMIRILFAFVSSVARND   324
Mouse_CAA42041.1        GEPLWIRCKAIHVNHGFGLTWELEDKALEEGSYFEMSTYSTNRTMIRILLAFVSSVGRND   325
                        ***: *** *:: .: * :  .:*.*        *   *. : : :*       :   .

Frog_075055042.1        SGSVTCLSSIHPNKSVKVTILDKGFIEMNTSSEQYDIDRKEDLCFVVQFKAYPAGRCAWI   407
Chicken_040514318.1     SGRYTCSSTAHPNQTALVTVLEKGFINITDSREDFEIGD-EEFCFEVNFTAYPPVKCMWL   383
Human_004110.2          TGYYTCSSSKHPSQSALVTIVEKGFINATNSSEDYEIDQYEEFCFSVRFKAYPQIRCTWT   384
Chimpanze_509601.2       TGYYTCSSSKHPSQSALVTIVEKGFINATNSSEDYEIDPYEEFCFSVRFKAYPQIRCTWT   384
Mouse_CAA42041.1        TGYYTCSSSKHPSQSALVTILEKGFINATSSQEEYEIDPYEKFCFSVRFKAYPRIRCTWI   385
                        :*  ** *: **.::. **:::****: . * *:::*.  *.:** *.*.***  :* *

Frog_075055042.1        HQSNTFPCAFLMDENGEST-SRYCNHNHVPGEYLFYLENDDVQLNKTFTLNILQTPQVQC   466
Chicken_040514318.1     FSQKTFPCKQSYSVDGHSISSKFCNHQHRSGIYMFYAENGDTAMTKKFTLYVRRKPEVTM   443
Human_004110.2          FSRKSFPCEQKGLDNGYSI-SKFCNHKHQPGEYIFHAENDDAQFTKMFTLNIRRKPQVLA   443
Chimpanze_509601.2       FSRKSFPCEQKGLDNGYSI-SKFCNHKHQPGEYIFHAENDDAQFTKMFTLNIRRKPQVLA   443
Mouse_CAA42041.1        FSQASFPCEQRGLEDGYSI-SKFCDHKNKPGEYIFYAENDDAQFTKMFTLNIRKKPQVLA   444
                        ..  :***      :* *  *::*:*::  * *:*: **.*. :.* *** : :.*:*

Frog_075055042.1        KMLSSKRVSCNSQGYPSSTWLWLRSAGE-SLNCTEKITEGITYGSL-ERTTDVWIATASL   524
Chicken_040514318.1     QL-LFKQISCIADSYPASSWVWRHCPDLNSTNCTEEIKEGIQNFLPERRSLGSWISSSIL   502
Human_004110.2          EA-SASQASCFSDGYPLPSWTWKKCSDK-SPNCTEEITEGVWNRKANRKVFGQWVSSSTL   501
Chimpanze_509601.2       EA-SASQASCFSDGYPLPSWTWKKCSDK-SPNCTEEITEGVWNRKANRKVFGQWVSSSTL   501
Mouse_CAA42041.1        NA-SASQASCSSDGYPLPSWTWKKCSDK-SPNCTEEIPEGVWNKKANRKVFGQWVSSSTL   502
                        :    .: ** ::.**  :* * :. .  * ****:* **:      .:  . *:::: *

Frog_075055042.1        DINTTAETYAICGCAINSAGMACAETLLHA-------YAQNNFIVLIAIGVCVLSIALLF   577
Chicken_040514318.1     DVKETATTFSVECCASNLVGSNCRKSFINMSVAGTVSSLPDNAALCVFAGFIFLLIVFLC   562
Human_004110.2          NMSEAIKGFLVKCCAYNSLGTSCETILLNS--PGPFPFIQDNISFYATIGVCLLFIVVLT   559
Chimpanze_509601.2       NMSEAIKGFLVKCCAYNSLGTSCETMLLNS--PGPFPFIQDNISFYATIGVCLLFIVILT   559
Mouse_CAA42041.1        NMSEAGKGLLVKCCAYNSMGTSCETIFLNS--PGPFPFIQDNISFYATIGLCLPFIVVLI   560
                        ::. :     :  ** *  *  *   :::           :*  .    *. .  *..*

Frog_075055042.1        ALVLSKYKKKYKYESQLRMIQYMESSDNEYIYIDFSVLEYDLKKWEFPRKNLEFGDTIGS   637
Chicken_040514318.1     ALICHKYKKQFRYESQLQMMQVIGPLDNEYIYVDFREYEYDL-KWEFPRENLEFGQVLGS   621
Human_004110.2          LLICHKYKKQFRYESQLQMVQVTGSSDNEYFYVDFREYEYDL-KWEFPRENLEFGKVLGS   618
Chimpanze_509601.2       LLICHKYKKQFRYESQLQMVQVTGSSDNEYFYVDFREYEYDL-KWEFPRENLEFGKVLGS   618
Mouse_CAA42041.1        VLICHKYKKQFRYESQLQMIQVTGPLDNEYFYVDFRDYEYDL-KWEFPRENLEFGKVLGS   619
                         *:  ****:::*****:*:*     ****:*:**   **** ******:*****..:**
```

Figure 6 (to be continued)

```
Frog_075055042.1        GAFGNVVTATAYGISKPGVFLQVAVKMLKENPGASEKEALMSELKMMTQIGHHENIVNLL   697
Chicken_040514318.1     GAFGKVVNATAYGISKAGDSVQVAVKMLKEKSDTSEKDALMSELKMMTHIGSHENIVNLL   681
Human_004110.2          GAFGKVMNATAYGISKTGVSIQVAVKMLKEKADSSEREALMSELKMMTQLGSHENIVNLL   678
Chimpanze_509601.2      GAFGKVMNATAYGISKTGVSIQVAVKMLKEKADSSEREALMSELKMMTQLGSHENIVNLL   678
Mouse_CAA42041.1        GAFGRVMNATAYGISKTGVSIQVAVKMLKEKADSCEKEALMSELKMMTHLGHHDNIVNLL   679
                        ****.*:.******** *  :*********: .:.*::**********::* *:******

Frog_075055042.1        GACTTSGPIYLIFEYCRHGDLLNYLKTNREMFHKTWSDVIEDNNFIFYHNSNQDEKFRDR   757
Chicken_040514318.1     GACTVSGPIYLIFEYCCYGDLLNYLRSKREKFHWTLTDIFKQHNFSFYHNICSEQNSRKG   741
Human_004110.2          GACTLSGPIYLIFEYCCYGDLLNYLRSKREKFHRTWTEIFKEHNFSFYPTFQSHPNSSMP   738
Chimpanze_509601.2      GACTLSGPIYLIFEYCCYGDLLNYLRSKREKFHRTWTEIFKEHNFSFYPTFQSHPNSSMP   738
Mouse_CAA42041.1        GACTLSGPVYLIFEYCCYGDLLNYLRSKREKFHRTWTEIFKEHNFSFYPTFQAHSNSSMP   739
                        **** ***:******* :*******:::** ** * :::::::** ** .   . :

Frog_075055042.1        AVTTVSNCSYTWIN----TT---Q--EHDKNPHHWVLSPEILKTKEMKYQ---NIRKYDE   805
Chicken_040514318.1     AHLEYGVNTDLYGEDELKATQTSQNINMTSGPNGIALCS---EEDEIKCA---SRQMDEE   795
Human_004110.2          GSREVQIH---------------PDSDQISGLHGNSFHS---ED-EIEYENQKRLEE--E   777
Chimpanze_509601.2      GSREVQIH---------------PDSDQISGLHGNSFHS---ED-EIEYENQKRLEE--E   777
Mouse_CAA42041.1        GSREVQLH---------------PPLDQLSGFNGNSIHS---ED-EIEYENQKRLAEEEE   780
                        .                        :  .. :   :     :  *::          *

Frog_075055042.1        EDLNILTFEDLLYFSYQVAKGMEFLESKMCIHRDLAARNILVTDEKVAKICDFGLARDIV   865
Chicken_040514318.1     EDFCVLTFEDLLCFAYQVAKGMEFLESKLCIHRDLAARNILVTHGKVMKICDFGLARDVV   855
Human_004110.2          EDLNVLTFEDLLCFAYQVAKGMEFLEFKSCVHRDLAARNVLVTHGKVVKICDFGLARDIM   837
Chimpanze_509601.2      EDLNVLTFEDLLCFAYQVAKGMEFLEFKSCVHRDLAARNVLVTHGKVVKICDFGLARDIM   837
Mouse_CAA42041.1        EDLNVLTFEDLLCFAYQVAKGMEFLEFKSCVHRDLAARNVLVTHGKVVKICDFGLARDIL   840
                        **: :******* *:*********** * *:********:***. ** **********::

Frog_075055042.1        NDSNYVVKGNARLPVKWMSPESIFEGIYTTRSDVWSYGVLLWEIFSLGVNPYPGMLVNAK   925
Chicken_040514318.1     NDSNYIVRGNVRLPVKWMAPESLFEGMYTMKSDVWSYGILLWEIFSLGVNPYPGIQVDAN   915
Human_004110.2          SDSNYVVRGNARLPVKWMAPESLFEGIYTIKSDVWSYGILLWEIFSLGVNPYPGIPVDAN   897
Chimpanze_509601.2      SDSNYVVRGNARLPVKWMAPESLFEGIYTIKSDVWSYGILLWEIFSLGVNPYPGIPVDAN   897
Mouse_CAA42041.1        SDSSYVVRGNARLPVKWMAPESLFEGIYTIKSDVWSYGILLWEIFSLGVNPYPGIPVDAN   900
                        .**.*:*:**.*******:***:***:** :*******:***************: *:*:

Frog_075055042.1        FYKLLQSGYKMDQPFYATDEIYFLMQSCWAFDCKKRPSFSQLVSFLGYQISNTEALVKAS   985
Chicken_040514318.1     FYKLIKSGFKMDRPYYATKHVYYVMQSCWALDSRRRPSFSHLVSSLACQLVEAEGAVYQN   975
Human_004110.2          FYKLIQNGFKMDQPFYATEEIYIIMQSCWAFDSRKRPSFPNLTSFLGCQLADAEEAMYQN   957
Chimpanze_509601.2      FYKLIQNGFKMDQPFYATEEIYIIMQSCWAFDSRKRPSFPHLTSFLGCQLADAEEAMYQN   957
Mouse_CAA42041.1        FYKLIQSGFKMEQPFYATEGIYFVMQSCWAFDSRKRPSFPNLTSFLGCQLAEAEEAMYQN   960
                        ****::.*:**::*:***. :* :******:*.::**** :*.* *. *: ::*  :  .

Frog_075055042.1        DAV-----------------------------------       988
Chicken_040514318.1     MKKNASPPKSNNKT-NPRVSKAEESLLSPPLL--------      1006
Human_004110.2          VDGRVSECPHTYQNRRPFSREMDLGLLSPQAQVEDS----      993
Chimpanze_509601.2      VDGCVSECPHTYQNRRPFSREMDSGPLSPQAQVEDS----      993
Mouse_CAA42041.1        MGGNVPEHPSIYQNRRPLSREAGSEPPSPQAQVKIHRERS      1000
```

Figure 6. Multiple protein alignment for FLT3 protein. (A) Phylogenetic tree of FLT3 gene in the great barred frog, chicken, mouse, human, and chimpanzee. In order of conservation, humans are most similar to the chimpanzee, then the mouse, then the chicken, then the great barred frog. **(B)** The sequence alignment of FLT3 across five species. Results indicate various segments of conservation (marked with asterisk) and segments of non-conservation (colon and period), mostly linked with chicken and frog. Conserved segments all have the same amino acid properties. Semi-conserved FLT3 sequences typically only vary by one color (property).

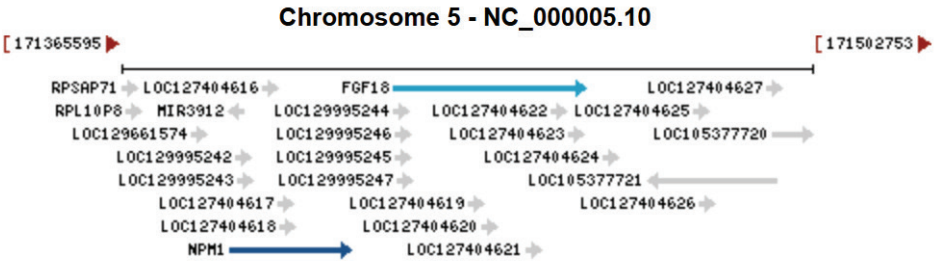


Figure 7. View of NPM1 gene location on human Chromosome 5 via NCBI. The gray arrows indicate potential other genes or pseudo-genes that are not found in humans but may be found as genes in other organisms. The colored arrows indicate human genes. NPM1 is represented by the dark blue arrow and should be read left to right based on the arrow direction. This view also indicates there is another gene, FGF18, which is represented by the light blue arrow and is also found on Chromosome 5; however, there is no significant overlap with NPM1.

Exon ID	ID	Location	Seq	Off targets
ENSE00001907503	1001394851	5:171410505-171410527	ATTTCTTTTTTTTTTTTTCCAGG	0: 5 1: 39 2: 666 3: 6011 4: 43590
	1001394852	5:171410522-171410544	TCCAGGCTATTCAAGATCTCTGG	0: 1 1: 0 2: 0 3: 32 4: 117
	1001394853	5:171410523-171410545	CCAGGCTATTCAAGATCTCTGGC	0: 1 1: 0 2: 0 3: 8 4: 72
	1001394854	5:171410528-171410550	CTATTCAAGATCTCTGGCAGTGG	0: 20 1: 5 2: 6 3: 12 4: 101
	1001394855	5:171410531-171410553	TTCAAGATCTCTGGCAGTGGAGG	0: 19 1: 9 2: 4 3: 22 4: 192

Figure 8. CRISPR Finder sgRNA sequences (some of them) for NPM1 gene. View of different possible sgRNA sequences starting from the end of intron 11. CRISPR Finder algorithm identifies guides that can be designed for a specific exon region in a gene, as well as potential OTEs. The ideal sgRNA would target only one specific region of NPM1 and not any other place in the gene. This would be represented in the Off-targets section by 0:1—indicating only one region with the exact nucleotide sequence of the sgRNA—and zeros next to the remaining off-target possibilities: 1, 2, 3, 4. A non-zero number next to the 1, for example, means that there is another segment on the gene that only differs from the sgRNA sequence by one nucleotide. The sequence boxed in red represents the best sgRNA sequence with the least off-target possibilities. There are no other places on the gene with the exact sequence of the guide, or a sequence that varies by 1 or 2 nucleotides. 8 other areas differ from the guide by 3 nucleotides, and 72 differ from the guide by 4 nucleotides.

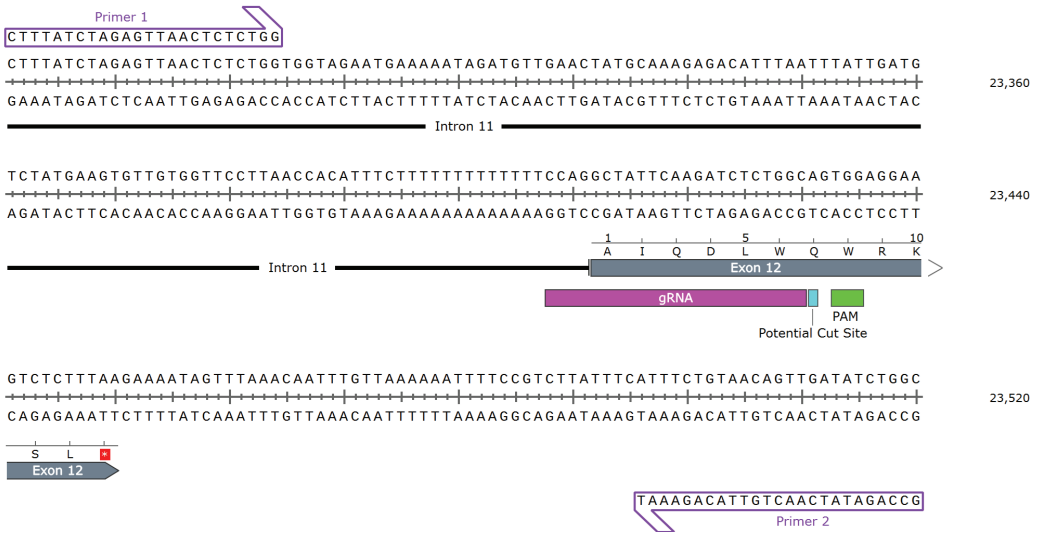


Figure 9. CRISPR/Cas9 Strategy presented via SnapGene. The target area on NPM1 is exon 12 due to the high occurrence of mutations in this area for NPM1c+. The most optimal sgRNA from CRISPR Finder (ID: 1001394853) was placed under its corresponding sequence on NPM1. The PAM sequence was identified a few nucleotides from the end of the sgRNA sequence as TGG. The blue box identifies a potential cut site for the Cas9 protein after recognizing sgRNA sequence. Primer 1 will be read left to right and aligns with the bottom strand of DNA. Primer 2 will be read from right to left and aligns with the top strand of DNA. Primers 1 and 2 will amplify the gene region between the two primers (including exon 12, where sgRNA and PAM sequence are located) to allow the polymerase enzyme to read the DNA template.

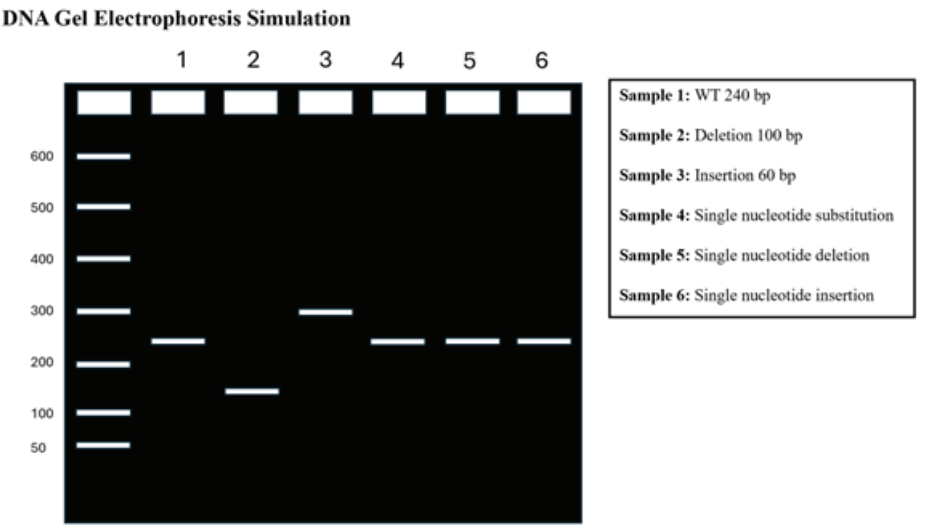


Figure 10. Hypothetical expected outcome simulated on DNA gel electrophoresis. A simulated image demonstrates the possible results of Cas9 cut on the exon 12 segment of NPM1. Shorter DNA fragments will run farther to the positive anode of the gel. The left-most column is a size marker to scale the bands of the samples. Column 1 corresponds with Sample 1, detailed in the legend to the right. Sample 1 is the wild-type (non-mutant version) of the NPM1 gene segment from primer 1 to primer 2 (on Figure 9). The deletion of 100 bp (Sample 2) results in a shorter DNA fragment of 140 bp, which is closer to the bottom of the gel. The addition of 40 bp (Sample 3) results in a longer DNA fragment of 300 bp, which is closer to the top of the gel. The single nucleotide substitution (Sample 4) results in a fragment of 240 bp, the single nucleotide deletion (Sample 5) results in a fragment of 239 bp, and the single nucleotide insertion (Sample 6) results in a fragment of 239 bp, all of which are not possible to differentiate on a gel.

DISCUSSION

AML accounts for about 1% of all cancers and about 1 out of 3 leukemias in adults in the USA. In 2025, it is predicted that roughly 22,010 American people will receive an AML diagnosis and about 11,090 American people will die from the disease (31). Currently, AML has a survival rate well under 50%, and this percentage gets significantly lower for individuals older than 70 (32). AML is also characterized by high rates of relapse. While treatments such as chemotherapy, immunotherapy, hypomethylating agents, and other targeted drugs are used to prevent the recurrence of AML, these methods can cause debilitating side effects like nausea, bleeding, and increased infection risk (33, 34). Research into AML is necessary for the continued development of treatment that can lead to complete remission with a lower risk of relapse and/or need for maintenance therapy.

NPM1 was chosen as the gene of interest for this study due to the fact that mutations involving NPM1 are frequently found in patients with AML. It is known that NPM1 mutations alone do not cause leukemia. Instead, they help initiate leukemia by disrupting other genes such as HOXA and MEIS1, which are transcription factors with functions necessary to maintain healthy cell cycles (35). HOXA cluster genes and MEIS1 play a significant role in hematopoiesis (the development of blood cells and other blood components) and are often linked to AML because their mutations promote leukemia cell proliferation and survival (36). A high expression level of these genes is found in almost half of patients with AML, usually where NPM1c+ mutations are also found. This suggests that NPM1c+ could drive the oncogenic gene expression of HOXA cluster genes and MEIS1 (37). Another mutation that occurs in over half the cases of NPM1-mutated AML is FLT3-ITD, the other gene of interest chosen for this study. FLT3-ITD is often associated with poor prognosis and is believed to occur after NPM1c+ mutations. Given the role of NPM1c+ as a facilitator for other gene mutations directly linked to AML with poor diagnosis, I wanted to hypothesize whether CRISPR gene editing could be used studying a potential therapy approach that could target NPM1c+ mutations and help prevent the development of other mutations directly related to AML, like FLT3-ITD in the future. I am aware that my strategy was not tested at bench for this study; however, I wanted to propose that this CRISPR/Cas9 editing design will be applicable for future opportunities. I aimed to generate a strategy and simulated the expected outcomes as a proposed experimental plan for the future. I also analyzed several other aspects, such as transmembrane domain prediction and multiple protein alignment of these genes, to present the features of the protein products.

Analyzing transmembrane domain prediction allowed me to gain an insight for the potential localization of wild-type and mutant NPM1 and FLT3. Figures 1-4 indicate if and where there are transmembrane domains for NPM1, FLT3, and their mutant forms, NPM1c+ and FLT3-ITD. This analysis was done because if the wild-type forms of the proteins contain transmembrane domains that are subsequently erased during a mutation, that protein may likely no longer be localized to the membrane. This may account for accumulation in the cytosol or some other area of the cell. The plots on Figures 1 and 2 indicate there is no change in the prediction of transmembrane domain for NPM1 or its mutant form, and that they do not have transmembrane domains. This is another way to show that NPM1 is not a structural transmembrane protein. Additionally, given this data, NPM1c+ does not accumulate in the cytoplasm due to a loss of a transmembrane domain. In Figure 3 plot, FLT3 is shown to contain two hydrophobic regions from amino acids 542-564 and 674-696, which are the areas where the protein is may likely to form transmembrane/intramembrane domains. These regions are most likely to be inserted in the lipid membrane because they cannot face the cytosol or lumen regions of organelles, which contain water. Given FLT3's identity as a receptor tyrosine kinase class III protein, this data can be understood as FLT3's transmembrane/intramembrane domain(s) help anchor the receptor to the cell membrane, which is essential for its intracellular signaling functions (38). Figure 4 shows that FLT3-ITD is 79 amino acids long. Since its wild-type protein consists of well over 900 amino acids, this data suggests that part of what happens when FLT3 mutates to FLT3-ITD could be eliciting the coding of an artificial stop codon or early termination. Because this stop codon occurs so early in the protein, many of the codons that make up the protein will not be read. This is likely to result in a big change to FLT3 functionality, if not making the protein completely non-functional. This drastic effect could also contribute to one reason why FLT3-ITD is linked to poor prognoses in AML patients.

The multiple alignment data for FLT3 and NPM1 are important to display the evolutionary conservation

of amino acids in the protein. Conserved amino acids (marked with *) and other conserved amino acids are especially important when looking at this data because they are likely more essential to the protein function and may correlate to more aggressive behavior in the disease when mutated. On the other hand, non-conserved and even semi-conserved regions may have little to no effect if the mutation targets these amino acids. Extending to heterogeneity among patients, the conservation data may be used to indicate prognosis or give scientists data on the severity of a particular mutation. A patient responding well to a particular drug, meaning the aggressiveness of the cancer is lower, might be due to a mutation in a non-conserved region of the protein instead of a conserved region. Further analysis of NPM1 and FLT3-ITD mutations in cases of AML is necessary to determine the relative importance of a mutation in a conserved amino acid within the protein. This analysis would involve looking at the number of patients with NPM1 and/or FLT3-ITD mutations in conserved versus non-conserved regions of the protein, to then determine the correlation between poor prognosis and mutations in conserved regions.

I designed a potential gene editing strategy for NPM1. There are no other clashing human genes with NPM1 in the same region on Chromosome 5 (Figure 7). In the context of this hypothetical design using CRISPR/Cas9 gene editing, not having another colliding gene is favorable because if genes collide due to localization in the same chromosome region, it will be more likely that sgRNA could edit genes outside the target gene. Given there is little overlap between NPM1 and FGF18—the other gene displayed on Chromosome 5—hypothetically targeted sgRNA editing for NPM1 mutations is unlikely to accidentally alter another gene. The sgRNA selected in NPM1 (Figure 8) is still not entirely optimal because the off-target sequences only differ from the sgRNA sequence by 3 or 4 nucleotides. Off-target effects are one of the biggest challenges preventing more widespread use of CRISPR-based gene editing in treating patients with different illnesses and conditions. In an off-target event, Cas9 cuts at an untargeted area, which can have adverse effects on the cells. While *in silico* tools such as CRISPR Finder are helpful for quantifying the risk for off-target cuts, there is still the possibility of these databases missing a matching or similar sequence (39).

There are several potential solutions to decrease the likelihood of off-target events. Double-stranded breaks (DSBs) utilize Cas9 nickases (mutated Cas9 nucleases) and two sgRNA to cut two opposite strands of DNA from a relatively close distance. Though this requires having two sgRNAs at the right area of the gene with PAM sequences, it would result in fewer off-target effects (39). Another strategy would be to use more unique PAM sites to decrease the number of sequences that match the sgRNA. StCas9 recognizes the site 5'-NGGNG-3". This site does exist in the middle of exon 12 in NPM1, so it could potentially be a future target (40). The area between primer 1 and primer 2 (Figure 9) is what would be amplified with PCR in a wet lab scenario. It is important to note that in the simulated gel containing potential results of a Cas9 cut (Figure 10), there are other outcomes and mutations that may occur related to the number of nucleotide bases inserted, deleted, or changed. While the gel provides valuable insight into whether mutations occur, the next step would be to use techniques like Sanger sequencing to understand what bases change and how these changes impact the protein in its entirety in an actual wet lab trial. Some single nucleotide changes may not edit the protein at all. In these cases, gene editing could be considered less successful because if the purpose is to mimic NPM1c+ and generate other NPM1 mutations to investigate the effects of different mutations and do comparisons, one ideally may need to have changes that impact the protein.

One limitation of targeting NPM1c+ mutations when considering treatment options for AML is the heterozygous nature of the gene. NPM1c+ is not viable as a homozygous mutation; thus, NPM1 retains one wild-type allele with the mutation. The most frequent expression of the mutation is a tetranucleotide insertion, but overall, the mutant form is very similar to its wild type (15). Given NPM1's role in cell survival, one worry is that targeting its mutant form with Cas9 could also damage the healthy allele in cancerous and non-cancerous cells.

This study proposes that NPM1 may be a candidate for investigating a preventative strategy for more severe mutations, such as FLT3-ITD, since this mutation is commonly linked to NPM1-mutated AML. However, this hypothesis relies on NPM1-mutated AML being caught before the FLT3-ITD mutation occurs, which could be challenging if the FLT3-ITD mutation develops shortly after NPM1. Additionally, studies have demonstrated diverse forms of FLT3-ITD mutants from the NPM1c+ mutation (41). Thus, it must be explored if Cas9 technology targeting NPM1c+ mutations at exon 12 could prevent all further FLT3-ITD mutants or just some versions in the future.

CONCLUSION

As over 20,000 American people will receive a new AML diagnosis within this year, ideally, treatment must be designed to prevent AML development and the likelihood of relapse that so often accompanies the disease. NPM1 may be useful for investigating a potential angle for a treatment as the mutated gene (NPM1c+) is linked to several other mutations that increase the severity and spread of AML, including FLT3-ITD. This study proposed a simulation of gene editing technology through an exploration in silico design. The next steps would be to explore potential gRNAs with lower chances for off-target effects and to model the CRISPR method on other organisms to examine whether it's truly effective in knocking out NPM1c+ in wet lab setting.

ACKNOWLEDGEMENTS

I thank Dr. Zeynep Öztürk (Ludwig Institute for Cancer Research, University of Oxford) for mentoring me during writing this paper.

CONFLICT OF INTERESTS

The author declares that there are no conflicts of interest regarding the publication of this article.

REFERENCES

1. John S. Bertram. The molecular biology of cancer. *Molecular Aspects of Medicine.* 2000; 21 (6): 167-223. doi: https://doi.org/10.1016/S0098-2997(00)00007-8.
2. Bruce N. Ames, *et al*. The causes and prevention of cancer. *Proceedings of the National Academy of Science of the United States of America*. 1995; 92 (12): 5258-5265. doi: 10.1073/pnas.92.12.5258.
3. Robert A. Weinberg. How Cancer Arises. Scientific American; 1996; 275 (3): 62-70. Accessed: https://www.jstor.org/stable/10.2307/24993349.
4. Key Statistics for Acute Myeloid Leukemia (AML). Available from: https://my.clevelandclinic.org/health/body/blast-cells (accessed on 2025-7-7).
5. What Is Acute Myeloid Leukemia (AML). Available from: https://www.cancer.org/cancer/types/acute-myeloid-leukemia/about/what-is-aml.html (accessed on 2025-7-7).
6. Ari Pelcovitis, *et al*. Acute Myeloid Leukemia: A Review. *Rhode Island Medical Journal*. 2020; 103: 38-40. PMID: 32236160.
7. Bob Lowenberg, *et al*. Acute Myeloid Leukemia. *New England Journal of Medicine*. 1999; 341 (14): 1051-1062. doi: 10.1056/NEJM199909303411407.
8. Bryan L. Betz, *et al*. Acute Myeloid Leukemia Diagnosis in the 21st Century. *Archives of Pathology & Laboratory Medicine*. 2010; 134 (10): 1427-1433. doi: https://doi.org/10.5858/2010-0245-RA.1.
9. Blast Cells: Purpose, Tests & Related Conditions. Available from: https://my.clevelandclinic.org/health/body/blast-cells (accessed on 2025-7-10).
10. Natalie Uy, *et al*. Hypomethylation Agents (HMA) Treatment for Myelodysplastic Syndromes: Alternatives in the Frontline and Relapse Settings. *Expert opinion on pharmacotherapy.* 2017; 18 (12): 1213. doi: 10.1080/14656566.2017.1349100.
11. Brunangelo Falini. NPM1-mutated acute myeloid leukemia: New pathogenetic and therapeutic insights and open questions. *American Journal of Hematology*. 2023; 98 (9): 1452-1464. doi: https://doi.org/10.1002/ajh.26989.
12. Shea Ping Yip, *et al*. The Multifunctional Nucleolar Protein Nucleophosmin/NPM/B23 and the Nucleoplasmin Family of Proteins. *The Nucleolus*. 2011; 15: 213-52. doi: 10.1007/978-1-4614-0514-6_10.
13. Brunangelo Falini, *et al*. NPM1-mutated acute myeloid leukemia: from bench to bedside. *Blood*. 2020; 136 (15): 1707-1721. doi: https://doi.org/10.1182/blood.2019004226.
14. Andrew Hindley, *et al*. Significance of NPM1 Gene Mutations in AML. *International Journal of Molecular Sciences.* 2021; 22 (18): 10040. doi: https://doi.org/10.3390/ijms221810040.
15. Lorenzo Brunetti, *et al*. Mutant NPM1 Maintains the Leukemic State through HOX Expression. *Cancer Cell*. 2018; 34 (3): 499-512. doi: 10.1016/j.ccell.2018.08.005.
16. Paul J. Coffer, The role of STATs in myeloid differentiation and leukemia. *Oncogene*. 2000; 19 (21): 2511-2522. doi: 10.1038/sj.onc.1203479.
17. Pradeep Singh Chauhan, *et al*. Mutation of NPM1 and FLT3 Genes in Acute Myeloid Leukemia and Their Association with Clinical and Immunophenotypic Features. *Disease Markers*. 2013; 35 (5): 581-588. doi: https://doi.org/10.1155/2013/582569.
18. Christian Thiede, *et al*. Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). *Blood*. 2006; 107 (10): 4011-4020. doi: 10.1182/blood-2005-08-3167.
19. Rong Wang, *et al*. Targeted therapy in NPM1-mutated AML: Knowns and unknowns. *Frontiers in Oncology*. 2022; 12: 972606. doi: https://doi.org/10.3389/fonc.2022.972606.
20. Tian Lin, *et al*. CRISPR screens in mechanism and target discovery for AML. *Heliyon*. 2024; 10 (8): e29382. doi: 10.1016/J.HELIYON.2024.E29382.
21. Melody Redman, *et al*. What is CRISPR/Cas9?. *Archives of Disease in Childhood - Education*

and Practice. 2016; 101 (4): 213-215. doi: 10.1136/ARCHDISCHILD-2016-310459.
22. Fathema Uddin, *et al*. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Frontiers in Oncology*. 2020; 10 (7): 1387. doi: 10.3389/FONC.2020.01387.
23. Polymerase Chain Reaction (PCR) Fact Sheet. Available from: https://www.genome.gov/about-genomics/fact-sheets/Polymerase-Chain-Reaction-Fact-Sheet (accessed 2025-8-2).
24. Tabitha M. Powledge. The polymerase chain reaction. *American Journal of Physiology - Advances in Physiology Education*. 2004; 28 (2): 44-50. doi: 10.1152/advan.00002.2004.
25. Sanger Sequencing: Introduction, Principle, and Protocol. The Genomics Service Company. 21 February, 2020. Available from: Sanger Sequencing: Introduction, Principle, and Protocol | CD Genomics Blog (accessed 2025-8-2).
26. Pei Yun Lee, *et al*. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments*. 2012; (62): 3923. doi: 10.3791/3923.
27. TMHMMGuide. Available from: TMHMMGuide - Research Computing Center Wiki (accessed on 2025-7-19).
28. Fabian Sievers, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. 2011; 7 (539). doi: 10.1038/msb.2011.75.
29. Clustal Omega FAQs. Available from: Clustal Omega FAQs - Job Dispatcher Documentation (accessed on 2025-7-19).
30. Clustal X Colour Scheme. Available from: Clustal Colour Scheme (accessed on 2025-7-21).
31. Key Statistics for Acute Myeloid Leukemia. Available from: Key Statistics for Acute Myeloid Leukemia (AML) | American Cancer Society (accessed 2025-7-19).
32. Cancer Stat Facts: Leukemia – Acute Myeloid Leukemia (AML). Available from: Acute Myeloid Leukemia — Cancer Stat Facts (accessed 2025-7-19).
33. Chemotherapy for Acute Myeloid Leukemia. Available from: Chemotherapy for Acute Myeloid Leukemia (AML) | American Cancer Society (accessed 2025-7-23).
34. Matteo Molica, *et al*. Maintenance therapy in AML: The past, the present and the future. *American Journal of Hematology*. 2019; 94 (11): 1254-1265. doi: https://doi.org/10.1002/ajh.25620.
35. Unnur Thorsteinsdottir, *et al*. Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. *Molecular and Cell Biology*. 2001; 21 (1): 224-234. doi: 10.1128/MCB.21.1.224-234.2001.
36. Jabra Zarka, *et al*. Nucleophosmin 1 Mutations in Acute Myeloid Leukemia. *Genes*. 2020; 11 (6): 649. doi: 10.3390/GENES11060649.
37. Hannah J. Uckelmann, *et al*. Mutant NPM1 Directly Regulates Oncogenic Transcription in Acute Myeloid Leukemia. *Cancer discovery*. 2023; 13 (3): 744-765. doi: 10.1158/2159-8290.CD-22-0366.
38. Soheil Meshinchi, *et al*. Structural and functional Alterations of FLT3 in Acute Myeloid Leukemia. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2009; 15 (13): 4263. doi: 10.1158/1078-0432.CCR-08-1123.
39. Congting Guo, *et al*. Off-target effects in CRISPR/Cas9 gene editing. *Frontiers in Bioengineering and Biotechnology*. 2023; 11: 1143157. doi: https://doi.org/10.3389/fbioe.2023.1143157.
40. Muhammad Naeem, *et al*. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells*. 2020; 9 (7): 1608. doi: 10.3390/cells9071608.
41. Namam Sharma, *et al*. NPM1 Mutations in AML—The Landscape in 2023. *Cancers*. 2023; 15 (4): 1177. doi: 10.3390/cancers15041177.