Annals of Case Reports
Bearden H, et al. Ann Case Rep: 11: 102628
www.doi.org/10.29011/2574-7754.102628
www.gavinpublishers.com

OPEN ACCESS

GAVIN PUBLISHERS

Research Article

The Effects of NUP98 Expression Levels Across Acute Myeloid Leukemia

Hailey Bearden*, Ruth Gelista, Vidhyaa Murugan

Bioinformatics at Baylor University, USA

***Corresponding author**: Hailey Bearden, Bioinformatics at Baylor University, USA

Citation: Bearden H, Gelista R, Murugan V. **(2026)**. The Effects of NUP98 Expression Levels Across Acute Myeloid Leukemia. Ann Case Report. 11: 2628. DOI: 10.29011/2574-7754.102628

Received: 20 May 2026; **Accepted:** 26 May 2026; **Published:** 28 May 2026

Abstract

NUP98 is a gene that normally regulates for part of the nuclear pore complex (NPC) and found on chromosome 11 in humans [1]. However, NUP98 is currently known for its potential role in Acute Myeloid Leukemia (AML). AML is cancer of the blood that is fast-growing and disproportionately affects pediatric patients. AML is a more aggressive cancer in general and its metastasis to other tissues and organ systems can be fatal in many cases due to its nature of being a blood-related cancer. Survival rates in AML patients also tend to be low in spite the age of medical advancement that is present today. Survival rates can be as low as thirty percent (Cleveland Clinic 24). The low survival rate, lack of understanding around the mechanisms of NUP98 that cause it to negatively impact AML, and the lack of effective treatments make the investigation between NUP98 and AML a necessity as AML rates increases.

The results of the different analysis showed a few different things. AML and NUP98 have some statistical significance when mortality rates are compared. However, demographic results showed no statistical significance in relation to NUP98 levels of expression. Insert aim 1 ideas. The evolutionary relationship between NUP98 in humans and mice showed that there is some conservation of the genes across species. This is helpful in terms of determining whether mice are useful in translational research from mice to humans.

The research performed showed the NUP98 levels do tend to be higher in patients with worse mortality outcomes. The statistical significance in this research was more difficult to interpret as many results were non-significant.

Research regarding NUP98 expression and its role in breast and pancreatic cancer is novel and should be further investigated as mortality was not present to compare against NUP98 expression. Finally, there does seem to be conservation of NUP98 between humans and the model species currently used to study AML. However, more work should be done on a larger scale to evaluate NUP98 conservation across several mammalian species.

Introduction

Nucleoporin 98 (NUP98) gene is a known "*structural component of the nuclear pore complex (NPC)*" that is found on chromosome eleven in humans [1]. NUP98 is part of a large multiprotein structure that runs along the nuclear membrane and is made up of around thirty proteins [1].

Currently, NUP98 is known for its association with Acute Myeloid Leukemia. It is known to be "*fused to at least 28 different partner genes*" in patients that have some sort of blood-related cancer [1]. The genome rearrangement believed to be related to this relationship are both translocations and inversions [1]. These translocations create fusion proteins [2]. Additionally, NUP98 is a "*dynamic nucleoporin*" that can move around and between different areas of the cytoplasm and nucleus [2]. Within AML specifically, this presents a series of problems. Its dynamic nature makes errors in transcription common that result in translocations and fusions. For example, HoxA9 (is a common homeodomain transcription factor that is involved in these fusions [2]. This transcription factor is usually responsible for binding DNA [2]. When studied in mouse models, it was seen that the NUP98/HoxA9 fusion was able to increase the malignancy of blood-related cancers that were

previously also induced by NUP98/HoxA9 [2].

This sets up another area of study. NUP98 in general is currently studied in mice as the model organism to compare to humans. Mice are commonly used in scientific and in clinical research as model organisms to compare to humans due to their relatively low cost, the fact that they do not take up majority of space, and several of mouse and human genetic areas overlap. There are ongoing studies suggesting that mice would also be good clinical representation of testing the mechanisms that could be behind NUP98's effect on diseases like acute myeloid leukemia. However, there is a gap in the literature in the sense that the cancer research community has not dug into the evolutionary relationship between mice and humans specifically regarding NUP98 and comparing its conservation across the two species.

Finally, NUP98's mechanisms and overall regulation affecting AML are still very unclear. However, a lot of research currently being done in regard to the gene points to indications that NUP98 with other fusions, such as NUP98-NSD1 could "*significantly impact the outcome of patients with AML*" [3]. The same study that made this observation also concluded that another fusion, NUP98-NSD1 and FLT3-ITD "could confer a poor prognostic effect" [3]. Overall, the survival rate with patients diagnosed with AML is around thirty percent (Cleveland Clinic 24). This is a lower survival rate considering that AML is a rarer cancer, and disproportionality affects pediatric patients. The gap in understanding if there is a measurable effect in expression levels impacting survival rates also needs to be further looked into.

Whether NUP98 mutations could be related to other malignant cancers is still unclear. Breast cancers are known to have the poorest outcomes and lack targeted therapies. NUP98's oncogenic role in breast cancer has potential for further investigation with it being known of potential dysfunction in the regulatory processes of the nuclear pore complex [4]. As a novel biomarker, NUP98 is thought to be involved in oncogenic translocations and could alter other aspects of transport. Similarly, the gene regulatory proteins in the nucleoporins may contribute to malignant phenotypes beyond then just as markers. Not only is this relevant in the recent cancer research in breast cancer patients but also sought to have potential for investigation in pancreatic cancer patients. Pancreatic cancer, for example, is seeing increased diagnosis in patients in recent years. The American Cancer Society states that the SEER stages combined five-year relative survival rate for Pancreatic Cancer is about thirteen percent. With a slim survival rate and high relation to this gene and more in the nuclear pore complex, the investigation of solely the expression of NUP98 is pivotal in understanding more of cancer development. If NUP98 expression levels are high in breast and pancreatic cancer, it can suggest that pathogenically such gene mutation might impact various cancer types similar to its impact on the mechanistic development of AML.

The goal of this research project was to research NUP98 and assess if there is a difference in the levels of NUP98 expression and whether patients with higher NUP98 expression experienced different mortality outcomes. Aim 1 is centered around two fast acting malignant cancers, pancreatic and breast cancer, compared to AML. Aim 2 is that the expression levels across AML will be compared between patient pools that had lower NUP98 expression and those that had higher NUP98 expression to measure if differences in expression rates resulted in different mortality outcomes. Aim 3 serves to compare and understand if NUP98 is genetically conserved human and mice species through sequence analysis and qualitative analysis.

Methods

Aim 1:

Transcript expression data and case information for the gene NUP98 were obtained from publicly available RNA-seq case files on NCI GDC Data Portal. Patient samples were selected based on tissue origin, disease type, experimental strategy, primary site, and open access. Breast and pancreatic samples were selected at random with 20-30 samples per group to be compared to AML samples. TPM (transcripts per million) were quantified to explore expression levels across cancer types for 20-30 cases. A nonparametric two-tailed Mann-Whitney U test was conducted to compare NUP98 TPM expression between two independent patient pools. The null hypothesis that could be examined is that NUP98 expression distribution does not differ between cancer groups. This test was ideal because it does not require normally distributed data and is appropriate for comparing expression values to discern differences between two independent groups [5-10]. NUP98 values quantified as TPM were analyzed without log2 transformation because the test analysis did not require normality. TPM values were compared in a box plot of all the different cancers and their distribution. Microsoft Excel and a Mann-Whitney U Test Calculator from the Social Sciences Statistics website were used for the analysis and collection of data.

Aim 2:

Data for much of this project was obtained from the National Institutes of Health's (NIH) National Cancer Institute (NCI). Both clinical and transcriptomic information were obtained for Acute Myeloid Leukemia (AML) and Pancreatic Cancer. The vital status, demographic variables such as race, gender, ethnicity, and age were collected along with the transcripts per million (TPM) for the NUP98 gene. Each patient was selected randomly from the entire database through a random number generator. Differences in gene expression between deceased and living patients were assessed

by a Mann-Whitney U test within JASP's statistical software and an unpaired t-test using BioRender. Then, associations between the demographic categories and mortality status were evaluated with chi square tests in JASP. The significance level was defined at a =9.95 and each test was two tailed (checked for a significant difference from the baseline).

All data used for the Acute Myeloid Leukemia patients was recorded in an Excel file and extracted progressively to test various conditions, NUP98, and demographic outcomes.

Bar chart visualizations were created in Microsoft Excel, including bar charts for the demographic distributions, and a boxplot was made for NUP98 expression in comparison to mortality status. The figures aided the support of statistical findings and used to illustrate these two findings.

BioRender and an unpaired t-test was the last form of statistical analysis run. Thirty patients were sorted based on their vital status for the NUP98 TPM to be compared. A one variable test was run using an unpaired t-test to test for significance between the expression levels and vital status.

Aim 3:

To measure the potential evolutionary relationship between humans and mice for NUP98 conservation, a few databases were visited to look for ortholgical relationships. The European Bioinformatics Institute's (EMBL-EBI) Ensembl was first used to determine if humans (Homo sapiens) and mice (Mus musculus) are orthologues for NUP98, and if so, to what extent. After those results were determined, then the cDNA alignments were downloaded in a FASTA file. Then, EMBL-EBI was used to run a multiple sequence alignment (MSA) using Clustal Omega. The FASTA file from Ensembl was used and uploaded to receive an alignment output onto EMBL-EBI. The Clustal Omega allowed for the alignments to be viewed, as well as a small phylogenetic tree. Afterwards, BLAST was run using the same FASTA file from Ensembl. Finally, the National Institute of Health (NIH) portal for both mice and humans was accessed to compare the positions of NUP98.

From here, the analysis of how conserved NUP98 was done qualitatively. Figures from each collected from each to further analyze where NUP98 is expressed across both species.

Results

Aim 1:

To compare the gene NUP98 expression across cancer groups with TPM values, a nonparametric Mann Whitney U test sought to address whether expression distribution differs. The box plot (Figure 1c) displays overall TPM expression values: AML samples have higher expression than Breast and Pancreatic cancer. AML showed variable expression compared to the other groups' range. Breast cancer displayed the lowest TPM expression values, IQR, and mean (Figure 1a). In a comparison between AML vs Pancreatic, the value of U is 50, z-score is -4.04399, provided p-value is < .00001, and sample size (N) is 20. In a comparison between AML vs Breast, the value of U is 32, z-score is -4.53089, provided p-value is < .00001, and sample size (N) is 20. In a comparison between Pancreatic vs Breast, the value of U is 188, z-score is 3.8661, provided p-value is 0.0001, and sample size (N) is 30. Between AML vs. Pancreatic samples, a difference in expression was observed, evident in the small p-value ($U = 50$, $Z = -4.04$, $p < 0.00001$; $n = 20$ per group). The U value of 50 and a z-score of -4.04 indicates a strong separation in expression ranks between groups. In a comparison between AML vs. Breast samples, a difference was observed, evident in the sample p-value ($U = 32$, $Z = -4.53$, $p < 0.00001$; $n = 20$ per group). A low U value and large z-score indicate strong separation in expression distributions. With a $p < 0.05$ for both comparisons, the null hypothesis could be rejected based on the provided results of Mann-Whitney U Test Calculator from the Social Sciences Statistics website.

Aim 2:

Vital status and NUP98 transcripts per million (TPM) were analyzed on a box plot (Figure 1b). Among the deceased, the minimum was found to be 31.32. The first quartile range was 31.32, and the third quartile was 197.5. The minimum was 31.32; the maximum was 271.80. The median was 111.99 and the mean was 137.19. The distribution was wide with a large interquartile range (Q3-Q1=102.16) which indicates the high variability in the NUP98 expression among the deceased patients. Also, the mean is higher than the median, which could suggest a right skewed distribution due to some patients with very high TPM values.

In the "*alive*" group, the minimum was 8.51, the Q1 was 73.91, the median was 103.19, the max was 147.29 and the mean was 93.28. In the "*alive*" group, there was a narrower IQR of 49.44 which indicates more consistency in expression levels. The mean is lower than the median, therefore, there is a small, left skew.

Overall, deceased patients had a higher mean and median TPM value for NUP98 expression than that of living patients. The deceased group has a greater variability and higher max values. These visual differences and the Mann-Whitney U test ($U=78.0$, $p=0.161$) confirm that the difference is not statistically significant.

A chi-squared test comparing the mortality status across individuals and their racial categories yielded a chi-squared value of 3400, df=5, p=0.639 (Figures 3, 4). There was no significant association found between race and status.

The mortality differences among gender were assessed with a chi-squared test resulting in a value of 2.143, df=1, and p=0.143 (Figures 7, 8). There were overall more deceased female patients. However, the results of the test were not statistically significant.

The ethnicity categories of Hispanic/Latino, not Hispanic/Latino, and unknown were analyzed and tested against mortality status. The chi square test produced a value of 2.360, df-2, p=0.307 (Figures 5, 6). This indicated no significant relationship.

Finally, an unpaired t-test was run using BioRender measuring NUP98 expression levels in Transcripts per Million (TPM) against patient vital status (alive and dead) (Figures 9, 10-16). The results were statistically significant with a two-tailed p-value of 0.0394 when compared to an alpha value of 0.05. The mean was 93.28 in alive patients, 137.32 in dead patients. The standard deviation in living patients was 73.91 and, in the deceased, it was 95.3.

Aim 3:

The Ensembl query showed that humans and mice have the same peptide length for NUP98: 1800 aa. The percentage of cDNA in humans was 89% and in mice it is 88%. The percent coverage between both species was 99%. The query clarified that NUP98 is located on chromosome 11 in humans, but it is located on chromosome 7 in mice. (Figure 1d)

Ensembl also generated a larger phylogenetic tree visualizing the broad evolutionary relationship between several species and the presence of NUP98 across them. The main relationships were between several species of primates as well as different rodents that were mostly mice, but not under the species '*musculus*'. (Figure 2).

EMBL-EBI resulted in a FASTA file of nucleotides bases that resulted in a phylogenetic analysis across several species comparing coparallel alignments. Another figure mapped the protein coding regions across mice and humans that included the location of NUP98.

The clustal2 alignment from the nightingale showed several conserved nucleotide across both the mice and the humans. The colors that match up to each other indicate conservation across the two sequences. There were more alignments that matched than otherwise.

	TPM vs Cancer Types (Box Plot)		
	Pancreatic	Breast	AML
Mean	67.17	49.86	115.49
Median	69.31	51.73	104.94
IQR	24.81	22.46	39.61

Figure 1a: TPM vs. Pancreatic, Breast, and AML Cancers in relation to TPM expression. The data shows the results of the box plot for each cancer type.

Mann Whitney U test			
	AML vs. Pancreatic	AML vs Breast	Pancreatic vs Breast
U value	50	32	188
Z-score	-4.043	-4.531	3.866
N	20	20	30

Figure 1b: Relevant results of the Mann Whitney U test between AML vs. Pancreatic, AML vs. Breast, and Pancreatic vs Breast. The calculated U value, z score, and sample size (N) were recorded.

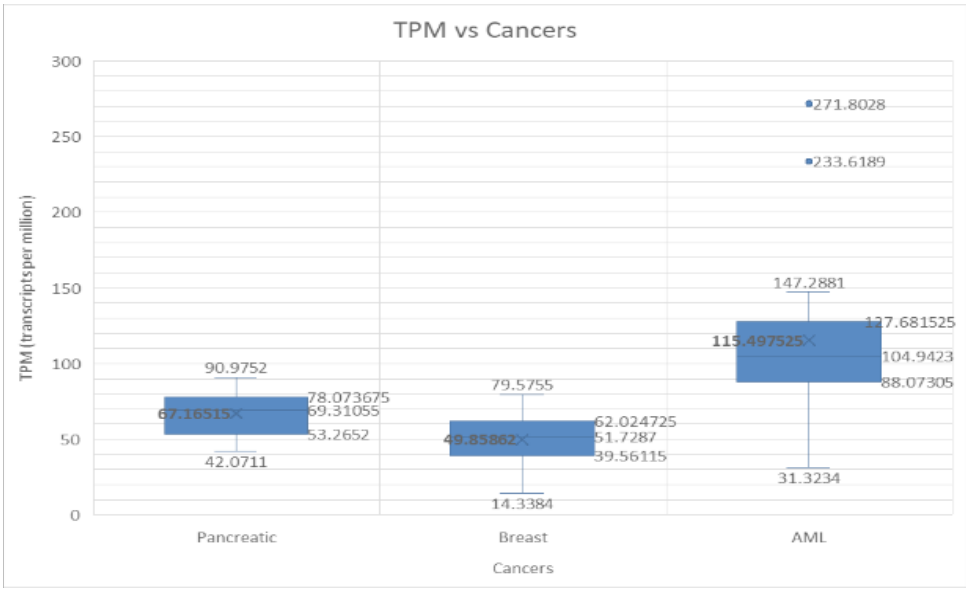


Figure 1c: TPM vs. Pancreatic, Breast, and AML Cancers in relation to TPM expression using a box plot. The different cancer types are in the x-axis and the TPM are recorded in the y-axis.

Cancer Type	TPM
Pancreatic	80.2403
Pancreatic	68.6462
Pancreatic	42.0711
Pancreatic	50.3712
Pancreatic	66.335
Pancreatic	70.3689
Pancreatic	78.6352
Pancreatic	55.1771
Pancreatic	75.9176
Pancreatic	63.7774
Pancreatic	52.6279
Pancreatic	59.8289
Pancreatic	76.3891
Pancreatic	71.9183
Pancreatic	80.7667
Pancreatic	49.9959
Pancreatic	90.9752
Pancreatic	87.0339
Pancreatic	52.2522
Pancreatic	69.9749
Pancreatic	71.3048
Pancreatic	82.6838

Pancreatic	56.79
Pancreatic	82.7332
Pancreatic	77.1069
Pancreatic	99.4474
Pancreatic	75.9367
Pancreatic	61.1756
Pancreatic	46.9935
Pancreatic	60.7837
Breast	60.2644
Breast	62.6935
Breast	50.4111
Breast	74.2016
Breast	44.3851
Breast	50.5331
Breast	39.4677
Breast	57.7258
Breast	39.8415
Breast	79.5755
Breast	52.9243
Breast	38.7405
Breast	64.6402
Breast	31.415
Breast	41.3741

Breast	54.0753
Breast	53.1388
Breast	62.6115
Breast	14.3384
Breast	24.815
Breast	55.0757
Breast	29.5488
Breast	43.5215
Breast	49.6429
Breast	65.0503
Breast	45.184
Breast	54.9833
Breast	33.4998
Breast	92.3796
Breast	66.1774
AML	111.992
AML	116.785
AML	97.1467
AML	92.6368
AML	31.3234
AML	233.619
AML	142.395
AML	271.803
AML	100.221
AML	96.7045
AML	84.5798
AML	91.6492
AML	49.9883
AML	125.433
AML	109.663
AML	117.501
AML	73.9079
AML	147.288
AML	86.881
AML	128.431

Figure 1d: Transcripts per million (TPM) for Pancreatic, Breast, and AML Cancers. Collected as raw counts from NCI GDC Data Portal.

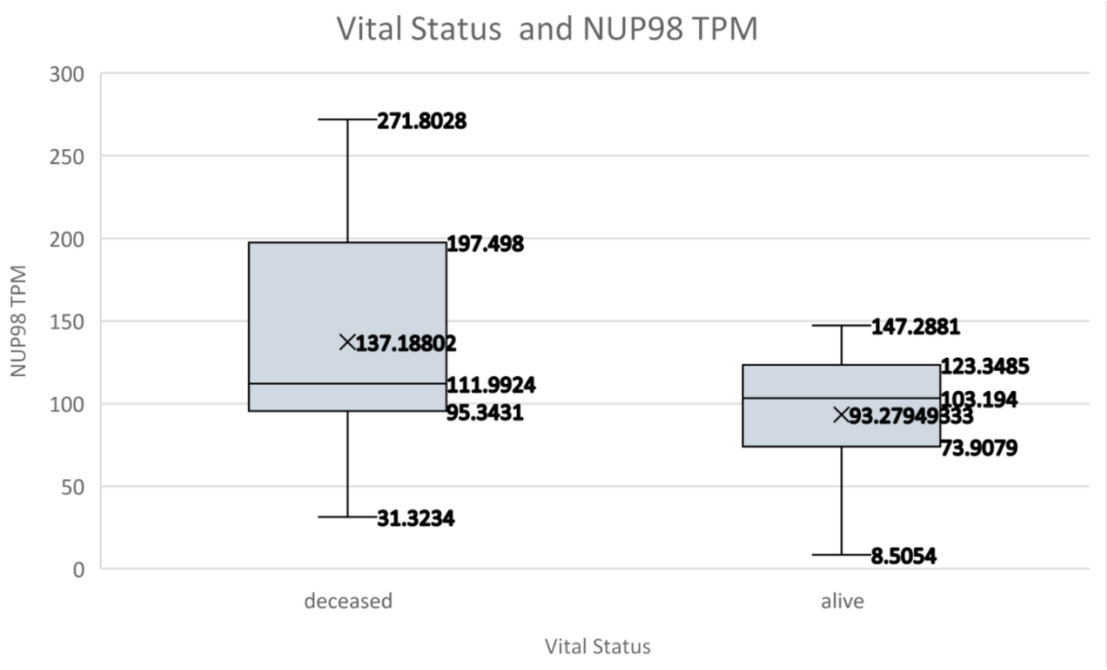
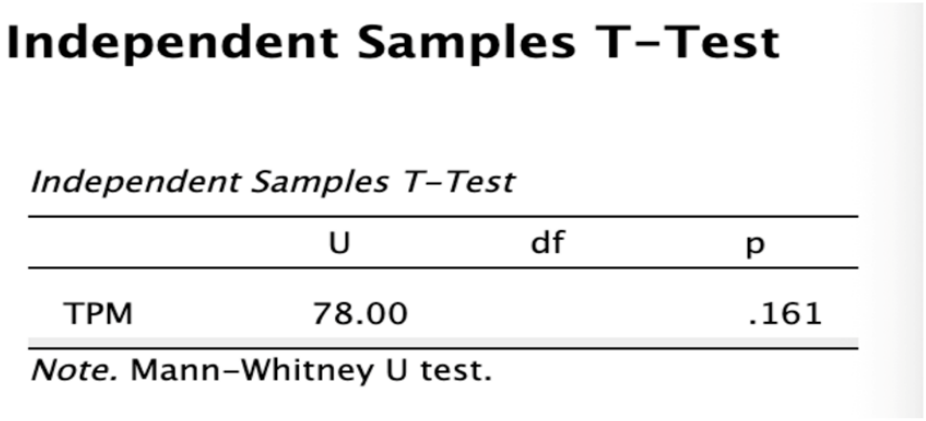


Figure 2: Box plot comparing vital status vs. NUP98 TPM Expression Levels. The plot on the left measures the deceased patients, and the plot on the right measures patients who are alive. The x-axis shows assigned vital status. The y-axis shows NUP98 expression in Transcripts per Million (TPM).



Independent Samples T-Test

Independent Samples T-Test

	U	df	p
TPM	78.00		.161

Note. Mann-Whitney U test.

Figure 3: The table shows an independent samples T-test that was run as a Mann-Whitney U test. The p-value was 0.161 and the alpha value was 0.05.

Contingency Tables

Contingency Tables

	Column 5						
Column 3	asain	black/african american	native american/alaska native	native hawsiian/pacific islander	unknown	white	Total
alive	0	2	1	1	0	11	15
deceased	1	3	1	0	1	9	15
Total	1	5	2	1	1	20	30

Note. Each cell displays the observed counts

Chi-Squared Tests

	Value	df	p
X^2	3.400	5	.639
N	30		

Note. Continuity correction is available only for 2x2 tables.

Figure 4: The table shows an independent sample Chi-Squared Tests and contingency table that was run comparing status to racial background of the patient. The p-value was 0.619 and the Chi-Squared Value was 3.400.

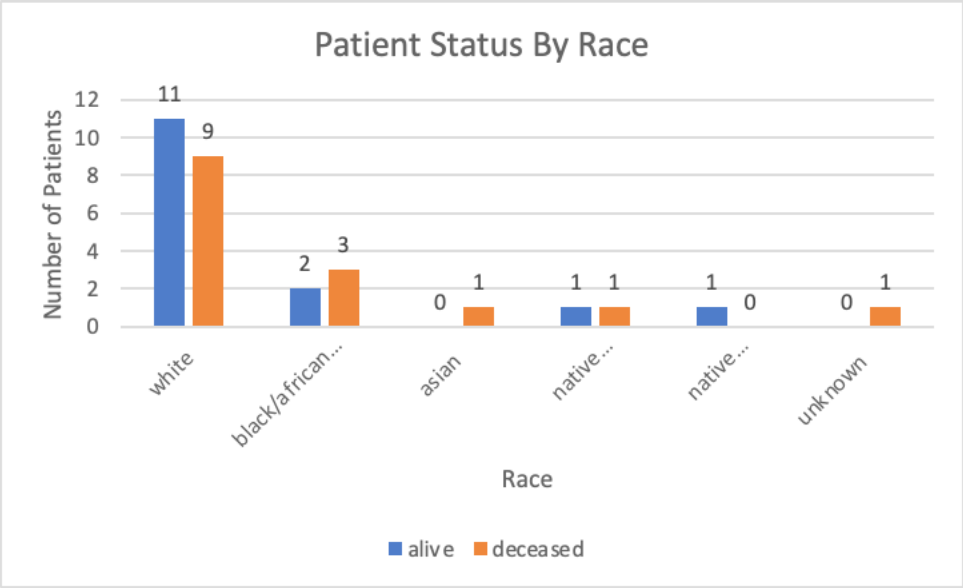


Figure 5: The graph compares the patient status based on racial backgrounds. See Figure 3.

Contingency Tables

Contingency Tables

Column 3	Column 6			
	hispanic / latino	not hispanic/latino	unknown	Total
alive	1	14	0	15
deceased	3	11	1	15
Total	4	25	1	30

Note. Each cell displays the observed counts

Chi-Squared Tests

	Value	df	p
X^2	2.360	2	.307
N	30		

Note. Continuity correction is available only for 2x2 tables.

Figure 6: The table shows an independent sample Chi-Squared Tests and contingency table that was run comparing status to ethnic background of the patient. The p-value was 0.307 and the Chi-Squared Value was 2.360.

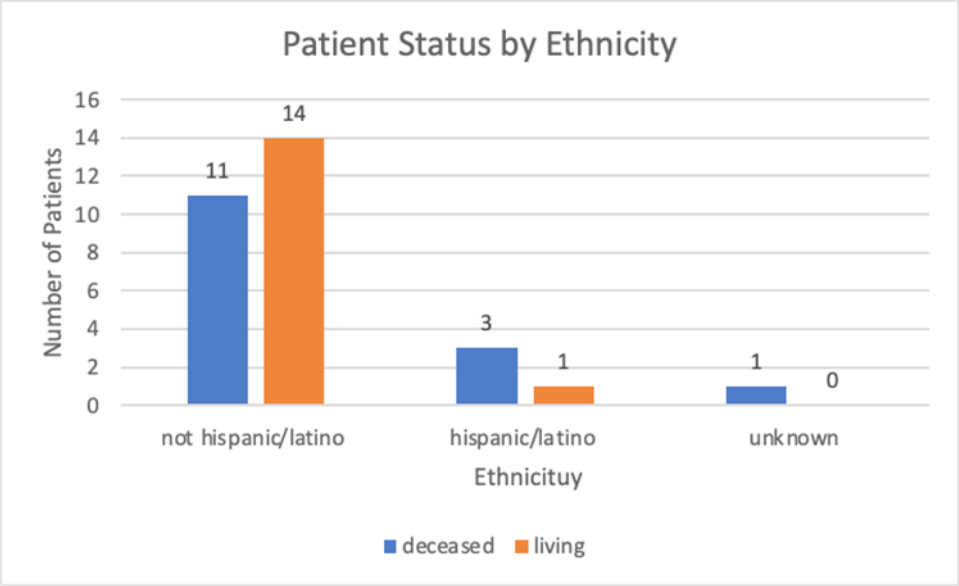


Figure 7: The graph compares patient status to their ethnicity. See Figure 5.

Contingency Tables

Contingency Tables

	Column 7		
Column 3	female	male	Total
alive	6	9	15
deceased	10	5	15
Total	16	14	30

Note. Each cell displays the observed counts

Chi-Squared Tests

	Value	df	p
X^2	2.143	1	.143
N	30		

Figure 8: The table shows an independent sample Chi-Squared Tests and contingency table that was run comparing status to gender background of the patient. The p-value was 0.143 and the Chi-Squared Value was 2.143.

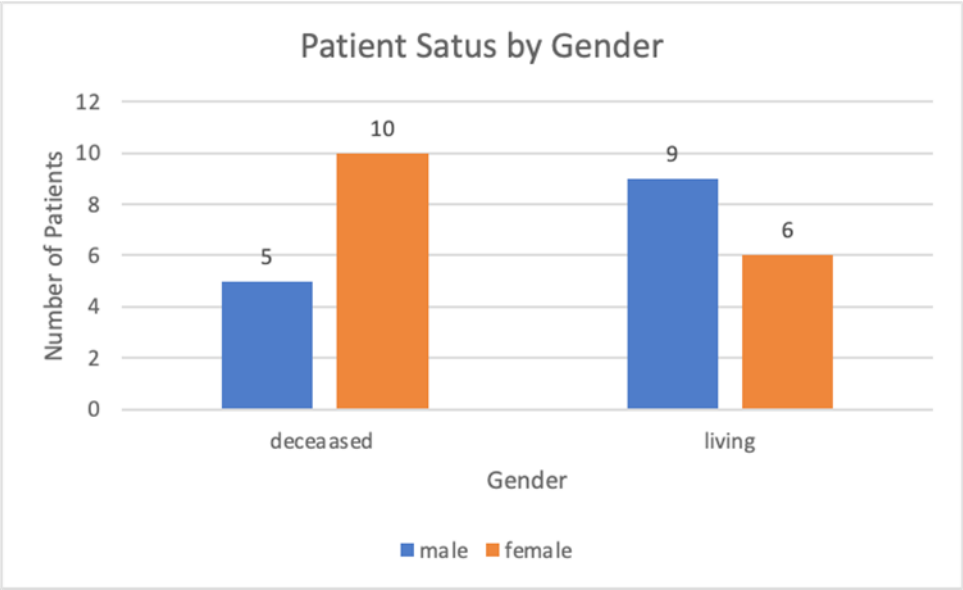


Figure 9: The graph compares the patient status to their gender. See figure 7.

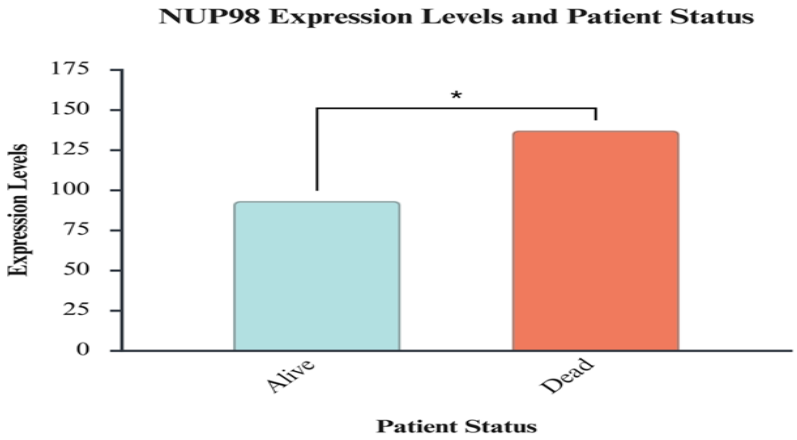


Figure 10: T-test results from BioRender, analyzing expression levels of NUP98 and patient status.

Unpaired t-test

Unpaired t-test					
Unpaired t-test					
t	DF	p-value (two-tailed)	Difference Between Means (A - B) ± SEM	95% Confidence Interval	Cohen's d
-2.1609	28.0000	0.0394*	43.9085 ± 20.3194	-85.5309 to -2.2861	0.7891
Descriptive statistics					
Category	Mean	SD	SEM		
Alive	93.2795	38.1013	9.8377		
Dead	137.1880	68.8582	17.7791		
Category	Minimum	25th percentile	Median	75th percentile	Maximum
Alive	8.5054	73.9079	103.1940	123.3485	147.2881
Dead	31.3234	95.3431	111.9924	197.4980	271.8028
Shapiro-Wilk normality test					
Category	W	p-value	Passed normality test (alpha=0.05)?	n	
Alive	0.9347	0.3199ns	Yes	15	
Dead	0.8915	0.07066ns	Yes	15	
Levene test of equal variances					
F	DFn	DFd	p-value	Are the SDs significantly different?	
2.0615	1.0000	28.0000	0.1621ns	No	

Figure 11: Unpaired T test examining status and various descriptive categories.

Column1	Case ID	Status	TPM of NUP98	demographic race	demographic ethnicity	demographic gender	age
1	00b9bfd0-95c0-59c6-afa4-44d13563a7e6	deceased	111.9924	white	hispanic / latino	male	7
2	00d795da-4d30-54dd-948d-3d2b3aba7beb	deceased	116.7853	black/african american	not hispanic/latino	female	11
3	0113b21f-c0d9-5e77-89b1-57a78f0f9a1e	deceased	97.1467	white	hispanic / latino	male	14
4	092b4b9f-af9b-52ce-9b30-df6291079960	deceased	92.6368	unknown	not hispanic/latino	female	23
5	f6f2dbad-3a44-46ef-9734-3f97da791d42	deceased	31.3234	white	not hispanic/latino	female	1
6	0bb8d24b-8544-4707-a116-2e3ca0e5dc10	deceased	233.6189	white	not hispanic/latino	male	2
7	12022d3a-6891-4216-a40d-e0f5155a0691	deceased	142.3953	white	not hispanic/latino	female	0
8	16ba295f-59c3-4d55-a6a1-ba2e85cd2f76	deceased	271.8028	white	not hispanic/latino	male	7
9	17872859-4139-57d0-99ea-574709cfd8cf	deceased	100.2208	black/african american	unknown	female	11
10	1db797e3-1101-4917-ad3e-db75b2cc9bc1	deceased	96.7045	native american/alaska native	hispanic / latino	female	17
11	2a813b7c-64b2-509f-ac64-ad9db5a56a7e	deceased	81.4784	black/african american	not hispanic/latino	female	11
12	2d2ac607-c4c9-4e9a-802e-ce9c28268f8b	deceased	244.609	white	not hispanic/latino	female	8
13	308a27f3-45f3-48fd-8b81-ac5fc4a00643	deceased	144.2649	white	not hispanic/latino	female	4
14	34e9517a-2b8f-5b11-ac7f-cc18cdb6bab3	deceased	197.498	asain	not hispanic/latino	female	16
15	41ded9f7-bbf4-4671-99d4-c55b50b53dbb	deceased	95.3431	white	not hispanic/latino	male	16
16	00e49a81-96ab-5f9b-b28c-4bc9e9e65544	alive	84.5798	white	not hispanic/latino	male	0
17	0116cc2c-50b9-4284-a257-386262715003	alive	91.6492	white	not hispanic/latino	male	27
18	018eaf84-b4e2-5946-85b2-085a4739155a	alive	49.9883	black/african american	not hispanic/latino	female	17
19	02fe6803-6894-4835-a86f-ddb3729b10c3	alive	125.4334	white	not hispanic/latino	female	7
20	12ba6092-9a34-530f-a193-42c10c11382d	alive	109.6638	white	not hispanic/latino	male	10
21	144ba1e6-9c5b-4b2a-bb62-284c7e8169ee	alive	117.5012	white	not hispanic/latino	female	15
22	1c631572-dde0-5d71-8c74-6155059d9f20	alive	73.9079	white	hispanic / latino	male	17
23	20822766-ed02-51ff-9d8f-3328dfbefbf2	alive	147.2881	white	not hispanic/latino	male	14
24	2d88a221-4e9f-49f4-b1f6-f607d1788171	alive	86.881	white	not hispanic/latino	female	6
25	43917f75-488d-52ca-9eb3-48ad6e161a1d	alive	128.4309	white	not hispanic/latino	male	4
26	1e5c8323-383d-51a0-9199-1b9504b29c7e	alive	8.5054	white	not hispanic/latino	female	6
27	26f32ebb-797a-44fb-89c0-00d0b5c11130	alive	35.5996	black/african american	not hispanic/latino	male	0
28	368350b3-8a0b-4738-8c65-719dedbe70f9	alive	103.194	native american/alaska native	not hispanic/latino	male	17
29	396161ed-8983-5d0d-ab8d-8dc24fe2b415	alive	113.2213	native hawsiian/pacific islander	not hispanic/latino	male	3
30	5f47a02f-6500-4525-85ce-fcd580457156	alive	123.3485	white	not hispanic/latino	female	19

Figure 12: AML patient data chart for conditions analyzed.

Type: 1-to-1 orthologues

Species	Gene ID	Peptide ID	Peptide length	% identity (cDNA)	% coverage	Genomic location
Human (*Homo sapiens*)	ENSG00000110713	ENSP00000316032	1800 aa	89 %	99 %	11:3671083-3797792
Mouse (*Mus musculus*)	ENSMUSG00000069550	ENSMUSP00000148115	1816 aa	88 %	99 %	7:101766605-101859383

Figure 13: From Ensembl comparing Homo sapiens and Mus Musculus for NUP98 orthologue compatibility as well as genomic location.

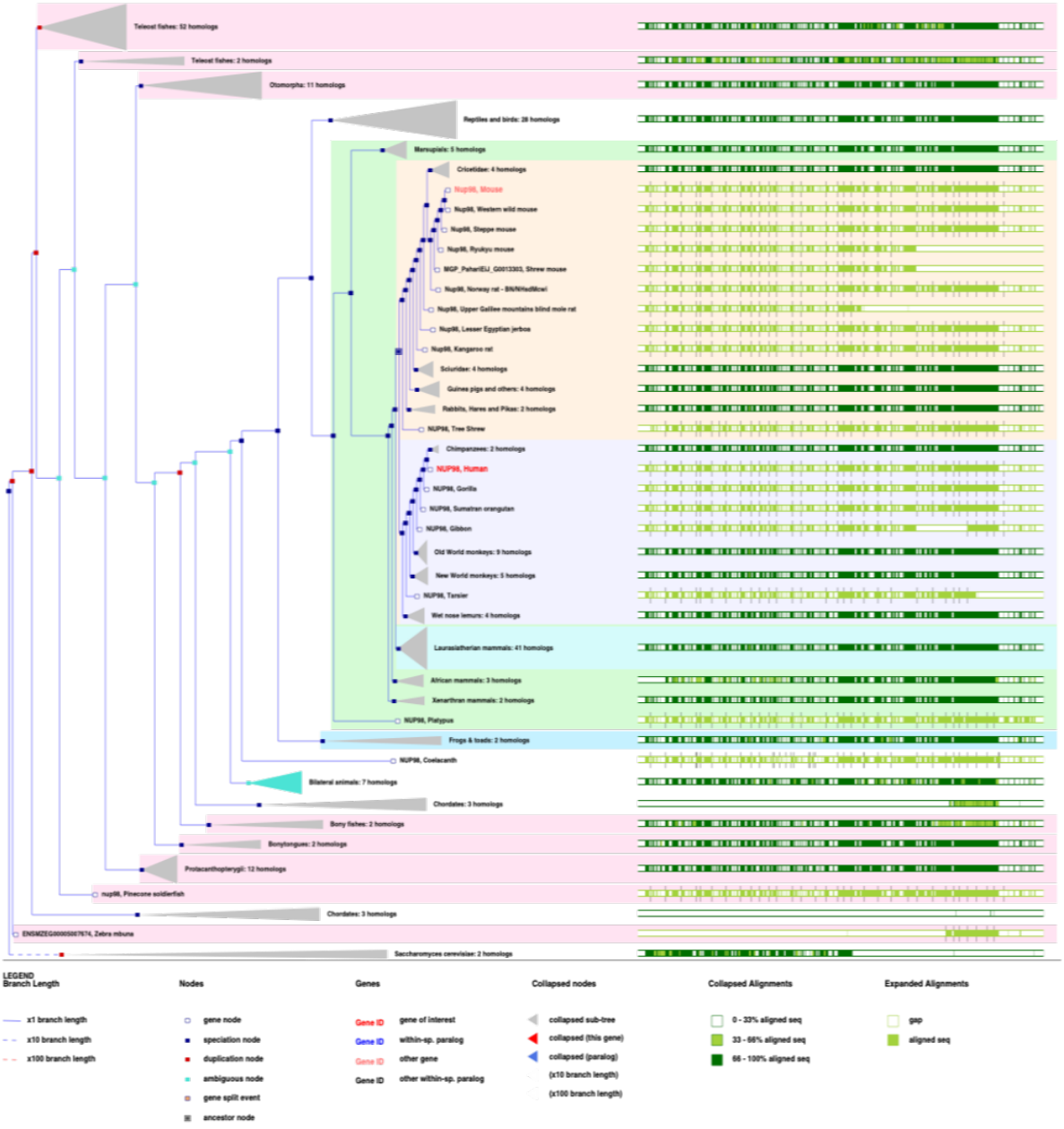


Figure 14: Phylogenetic analysis between Homo sapiens and Mus musculus along with expression of NUP98.

Figure 15: Genome coding regions being compared at the location of NUP98 in humans, specifically the protein that is highlighted in green towards the bottom of the figure.

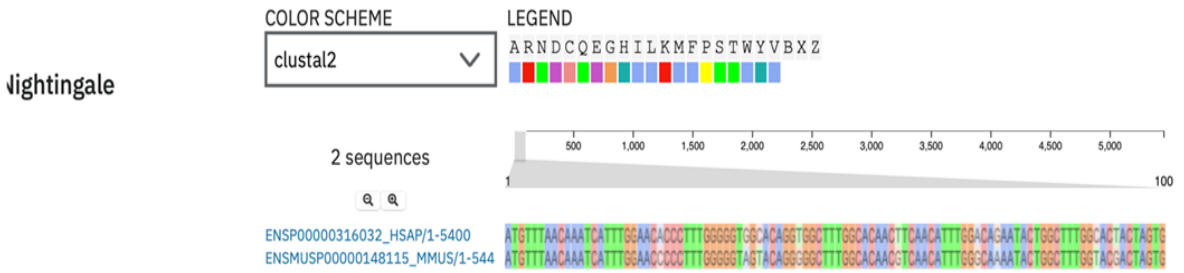


Figure 16: The sequential analysis performed between Homo sapiens and Mus musculus using Clustal.

Discussion

Aim 1:

AML and two different cancers were compared to assess whether NUP98 expression differs between these groups based on TPM counts. AML showed higher NUP98 TPM expression compared to pancreatic and breast cancer groups. This can led us to potentially reject the null hypothesis that there is no difference comparing the outcomes of cancer groups with the gene expressed in high and lower levels. However, definitive findings on gene expression between cancers are not stated due to data limitations. These findings simply suggest that exploring further how NUP98 expression patterns will differ across malignancies is relevant in cancer research. A higher NUP98 expression in AML, compared to the other cancer groups, is consistent with our understanding of the involvement of NUP98 in leukemogenesis. Higher transcript levels of AML, and similarly elevated levels in breast cancer reflect a prominent role NUP98 plays in malignancy and disease progression. In a comparison of pancreatic and breast cancer samples, variable NUP98 expression is evident in the box plot figure. This suggests that NUP98 may play a different mechanistic role in the body, compared to AML. This analysis provides a direct look into TPM-based comparison of NUP98 expression across cancer types, an approach that is absent in literature apart from AML and how solid tumor studies are not diversified.

However, this approach lacks a deep understanding of differential gene expression due to being limited to TPM raw counts and lack of the use of linear models. The tests conducted explore the NUP98 distribution across groups and do not represent significant differential expression and correlation of the NUP98 gene to pancreatic and breast cancer [5-10]. A small sample size also reflects the lack of suggesting any clinical outcomes or significant contributions to the effect of mutation in the NUP98 gene in other cancers.

Future research should compare differential gene expression analysis and statistical testing across various cancer types with NUP98. The nuclear pore complex requires various pathways of regulation and significant biomarkers, that if mutated, could affect the malignant phenotype of cancer. NUP96 is another gene, similar to NUP98, that could be analyzed to understand the pathogenic effects of mutations in AML, breast, and pancreatic cancer in regulatory pathways. Comparison between the differential gene expression of hematologic malignancies and solid tumors could also be explored to see how tissue source, organ-specific signaling pathways, and NUP gene transcription could be impacted.

Aim 2:

This study investigated whether NUP98 expression levels are correlated with mortality outcomes in patients with Acute Myeloid Leukemia (AML). In the study, the deceased group showed higher median and mean TPM values for the expression of the NUP98 gene expression compared to the living group. The difference between the two groups was not statistically significant suggested by a Mann-Whitney test (results being $U = 78.0$, $p = 0.161$). The boxplot analysis revealed a wide distribution and variability in NUP98 expression among the deceased patients. Among the deceased patient group, there was a right-skewed pattern which suggests that some of the individuals may have elevated expression of the gene of interest. There was overlap in the interquartile ranges between groups, so this indicates that the NUP98 gene expression alone does not reliably distinguish mortality status.

Demographic variables were examined to evaluate cofounding effects that they might have regarding AML patient mortality. A chi-squared test was run for race (χ^2=3.400, p=0.639), ethnicity (χ^2=2.360, $p = 0.307$) and gender ($\chi^2 = 2.143$, $p = 0.143$), and all failed to reach statistical significance and disprove the null hypothesis. In the data that was obtained randomly, more female patients were deceased than alive and there were some racial

patterns observed such as more white individuals living. There were also some ethnic trends observed within the curated graphs; however, these patterns were not distinguished enough and did not meet the threshold for significance, and the data could reflect the sampling variability rather than true associations.

The findings of the study suggest that gene NUP98 plays a significant role in leukemogenesis and the progression of acute myeloid leukemia. Its expression level along is not a significant biomarker to predict and project the outcome for patients with acute myeloid leukemia, but it could be a significant aspect of AML's progressive abilities when working in tandem with other genes or mutations.

The limited sample size of 30 individuals with acute myeloid leukemia may have reduced the statistical power of this study. It limited the results of TPM expression levels of NUP98 to be distributed but not enough samples to establish a correlation between NUP98 expression levels and mortality results. Further, it limited the analysis of demographic impacts on patients with acute myeloid leukemia. The heterogeneity of AML impacts significantly to the complexity of outcome prediction.

Future research should use larger and more diverse cohorts and consider the gene mutation status, disease severity, treatment history, and mutations involved with NUP98. Additional comparative studies across species, like comparative study between humans and mice, that could further clarify its functional role and inform transitional models. Ultimately, NUP98 will remain a gene of interest in AML research and the prognostic ability of the gene will require further investigation within a broader clinical and molecular context to aid in the treatment of those with AML, especially with a NUP98 mutation.

Aim 3:

The Ensembl results show that humans (Homo sapiens) and mice that are typically used to conduct clinical research (Mus musculus) are orthologues. The Type 1:1 analysis conducted showed an 88% match for NUP98 alignment for humans and 89% alignment (cDNA) for mice.

The human NUP98 and mouse copy of NUP98 are distantly related on figure 2c's phylogenetic tree. The distance between the two were unexpected though the overall relation was not. There was a 33-60% alignment between the two, showing there was a partial alignment between the two genes across these species. The exact alignment percentage was not given by this figure.

Figure 3c shows the different genes that exist on chromosome 11 in humans. This did not show NUP98 for mice as the copy of the gene on their chromosome 7.

Figure 4c showed that the Clustal alignment between the sequences coding for NUP98 across mice and humans.

Overall, the analyses show that there is a high match between the cDNA of both species. This is beneficial because the mice that were analyzed are the most commonly used species of mice for clinical-based research. Being able to ascertain that these two species are conserved is beneficial because it confirms that the research being performed on these mice is more likely to translate over into clinical trials for patients who have been diagnosed with AML specifically. The conservation of the gene within both species suggests that mice would likely be a good model going forward in clinical research.

Some difficulties that arose during this portion of the analysis is that while both humans and mice share conservation of NUP98, they are located on different chromosomes. Additionally, the BLAST that was run was initially difficult to narrow down; in the end, it focused on proteins rather than nucleotide analysis due to the program's algorithm.

In the future, the evolutionary conservation of NUP98 could be further investigated across other species in the future as well. The reason why this might be an area of interest is that if NUP98 is widely conserved across several species, then it might prove that it serves a vital regulatory role important to proteins and nuclear pore related regulation. If it could be proved that NUP98 does serve this role, then more research might be done towards uncovering the mechanisms that result from dysregulation when NUP98 experiences fusion with other genes. Additionally, if NUP98 and its sequence are not widely preserved across species, then NUP98 might be able to be proved as an easily mutated gene that is susceptible to disregulation. If that is the case, the AML link that is associated with mutated NUP98 might be explained by its susceptibility to being mutated, but further research should be performed before being certain.

Individual Contributions

Hailey: Contributions include the development and investigation of Aim 2, generating figures and graphs used in this study via software tools (JASP and Excel), analysis and interpretation of graphical and numerical results after Chi-squared and Mann-Whitney U tests were performed. The labeling of figures connected to aim 2, discussion, and a section of methods and results.

Vidhyaa: Contributions include the development and investigation of Aim 1, generating figures and graphs used in this study, analysis and interpretation of graphical and numerical results after conducting Mann Whitney-U Test and Microsoft Excel analysis. The small section of the introduction, writing for the methods, results, and discussion of aim 1.

Ruth: Contributions include the development and investigation of aim 3, generating figures used for Aim 3 and its analysis, analysis of the TPM values in aim 2 when measured against vital status of patients, generation of graphs for some of aim 2, analysis and interpretation of an unpaired t-test for Aim 2, the discussion, results, and a section of methods for aim 3. Additionally, some of the background literature review as well as construction of the introduction and abstract.

Acknowledgements

We would like to thank Dr. Mary Lauren Benton for her guidance and feedback during this project that she generously provided.

References

1. Gough SM, Slape CI, Aplan PD. (2011). NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. Blood. 118: 6247-6257.
2. Xu S, Powers MA. (2010). Nup98-homeodomain fusions interact with endogenous Nup98 during interphase and localize to kinetochores and chromosome arms during mitosis. Mol Biol Cell. 21: 1585-1596.
3. Sheikhi M, Siyadat P, Rostami M, Sadeghian MH, Zahiri E, et al. (2024). Prognostic importance of NUP98-rearrangements in acute myeloid leukemia: A systematic review and meta-analysis. Caspian J Intern Med. 15: 579-588.
4. Mullan PB, Bingham V, Haddock P, Irwin GW, Kay E. (2019). NUP98 - a novel predictor of response to anthracycline-based chemotherapy in triple negative breast cancer. BMC Cancer. 19: 236.
5. Zhao S, Ye Z, Stanton R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. RNA (New York, N.Y.), 26: 903-909.
6. Biorender. BioRender App. (n.d.). Accessed in Online.
7. Preston CC, Faustino RS. (2018). Nuclear Envelope Regulation of Oncogenic Processes: Roles in Pancreatic Cancer. Epigenomes, 2: 15.
8. Roy A, Narayan G. (2019). Oncogenic potential of nucleoporins in non-hematological cancers: recent update beyond chromosome translocation and gene fusion. Journal of cancer research and clinical oncology, 145: 2901-2910.
9. Social Science Statistics. (n.d.). Mann–Whitney U test. Accessed in Online.
10. Survival rates for pancreatic cancer. American Cancer Society. (n.d.). Accessed in Online.