JOURNAL OF
EMERGING INVESTIGATORS

Article

Investigation of unknown causes of uveal melanoma uncovers seven recurrent genetic mutations

Hersh Nanda[1], Michael T. Barrett[2]
[1]BASIS Chandler, Chandler, Arizona
[2]Department of Hematology and Medical Oncology, Mayo Clinic, Phoenix, Arizona

SUMMARY

Uveal melanoma (UM) is a rare subtype of melanoma but the most frequent primary cancer of the eye in adults. The goal of this study was to research the genetic causes of UM through a comprehensive frequency analysis of base-pair mismatches in patient genomes. It was hypothesized that chromosomes 15 and 16, due to their role in the expression of eye color, would exhibit UM-associated mutations. The U.S. National Institute of Health (NIH) provided genomic data from surgically resected eye enucleations and liver metastases from 32 UM patients, which we individually analyzed to identify nucleotide base-base mismatches by comparing DNA sequences from cancerous cells to reference DNA sequences. After identifying mismatches we performed a Pareto analysis of cross-patient data to identify chromosomes with recurrent and non-recurrent genetic mutations across all samples. We then studied the mutated gene functions to evaluate causal links to cancer, such as for genes that encoded known tumor suppressor proteins. We discovered a total of 130 genetic mutations, including seven recurrent mutations, with most mutations occurring in chromosomes 3 and X. Recurrent mutations varied from 8.7% to 17.39% occurrence in the UM patient sample, with all mutations identified as missense. These findings suggest that UM is a recessive heterogeneous disease with selective homozygous mutations. Notably, this study has potential wider significance because the seven genes targeted by recurrent mutations are also involved in other cancers. As a result, immunotherapy is a highly promising treatment for uveal melanoma due to the disease's heterogeneous nature.

INTRODUCTION

Melanoma is a life-threatening malignancy that affects melanocytes (pigment-producing cells) found throughout the body (1). Melanomas are of two types: Cutaneous and non-cutaneous. Cutaneous melanomas, which account for approximately 95% of all melanomas, originate in the pigment-producing cells of the skin (2). Non-cutaneous (non-skin) melanomas affect other regions of the body including the eyes and mucous membranes, such as those present in nasal passages and the oral cavity (2, 3).

Although it is a rare disease, uveal melanoma (UM) is the most common form of non-cutaneous melanoma and is the most frequent primary cancer of the eye in adults (4, 5). UM is known to affect approximately 7000 individuals worldwide annually with incidence rates ranging from 0.2 to 0.3 per million individuals in African and Asian populations to up to 6 per million in white populations (5, 6). Diagnosis usually occurs at age 60, and this cancer is more prevalent among Caucasians (6). In most cases, UM forms in the choroid, the vascular layer of the eye lying between the sclera and the retina. Symptoms exhibited by UM patients include but are not limited to variable and painless visual disturbances, discoloration of the iris, change in the shape of the pupil, or loss of peripheral vision (7).

One aspect that contributes to the lethal nature of melanomas, including UM, is the risk of metastasis. Metastasis is the migration of cancerous cells through the bloodstream, which leads to the development of tumors elsewhere in the body (8). These tumors can cause tissue damage and other widespread effects which accelerate poor outcomes for cancer patients (8). In (UM), the liver is the organ most commonly affected by metastasis, which occurs in 80% of cases (9, 10). Although regular metastasis can be detected, undetectable micrometastases may occur, in which a small collection of cancerous cells spread to other parts of the body via the lymphovascular system (11). Micrometastasis poses a significant risk to all UM patients, which is why patients should receive immediate treatment after diagnosis.

Currently, the causes of UM are still unclear. Some studies have revealed that the DNA present within the cancerous cells showed alterations on chromosomes 1, 3, 6, and 8, but these conclusions can be researched more extensively (12, 13). Studies have also found that genes *BAP1*, *SF3B1*, *GNAQ*, and *GNA11* seem to play a role in the development of UM (12, 13, 14). Although these studies did uncover some genetic errors that may be responsible for UM, they analyzed chromosomal rearrangements rather than specific nucleotide alterations. Currently, there is a lack of effective therapies for the treatment of UM (10).

This study aimed to identify specific genetic mutations resulting from base pair discrepancies that contribute to the development of UM using data analysis, statistical

techniques, and gene ontology (GO) research with an NIH dataset of 32 samples including 23 primary and 9 metastatic samples (14). Following the analysis, we observed 130 unique genetic mutations in all the UM patients, with 7 recurrences in the primary samples and no recurrences in the metastasis samples. We identified seven genes – *ALG1L2*, *DMD*, *IL1RAPL2*, *KIA0825*, *LOC440040*, *NXF2*, and *PHYHD1* – to have the most recurrent mutations. The identification of these frequently mutated genes and the assessment of their possible causal link to UM is essential for further understanding the disease and developing effective treatment options.

RESULTS

Within this study, genetic alterations resulting from chromosomal rearrangements were disregarded except base pair mismatches causing missense, nonsense, or frameshift mutations. Some patients had experienced metastasis of cancerous cells through the bloodstream; we considered data from their sample genomes separately to identify genetic mutations that could promote and accelerate metastasis. The basic procedures for this study involved the comparison of a mutated DNA sequence (found in cancerous somatic cells) with the matched normal sequence (found in healthy, unaffected somatic cells).

We observed most of the genetic mutations on chromosomes 1, 3, 5, 9, 11, and X, which confirms previous findings of chromosomes 1 and 3 being involved in UM (12, 13), and identifies chromosomes 5, 9, 11, and X as potential new factors (**Figure 1**). To further analyze chromosome involvement, we performed an in-depth Pareto analysis of all identified genetic mutations. A Pareto chart is a bar chart in which the bars are ordered from highest frequency of occurrence to lowest frequency of occurrence (15). This cause analysis tool can be used to measure the frequency of problems or causes in a process (16). Although this technique is typically used in statistical decision-making, it helped visualize quantitative data on the frequency of genetic mutations across all samples in the study. The Pareto analysis showed that chromosomes 3 and X exhibited the most genetic mutations with a combined total of 75 across all primary samples (**Figure 2**).

Somatic single-nucleotide variant (SNV) files were examined from the data package. Each data file contained information on the patient's whole genome, including the locus, zygosity, and variation type of each DNA base pair, as well as the genes identified during sequencing (**Table 1**). The two columns outlined in bold include the reference (normal) DNA sequence and the allele 2 (mutated) DNA sequence (**Table 1**). To identify base-base mismatches, we integrated an automated IF function to compare each cell from the reference DNA sequence to the corresponding cell in the allele 2 sequence to flag mutations as Xs in the inserted 'Match' column. Subsequently, we utilized this analysis method for all remaining primary (PUM) and metastases (MUM) patient files.

All mutations that occurred were noted with their respective chromosome number, zygosity, and gene name. We also considered the gender of each patient to determine if UM is an autosomal or sex-linked disease. This assessment did not involve the variation type of each mutation including base substitutions, deletions, and insertions, as that had already been investigated by other studies. However, the application of a data filter on the variation type (labeled "varType" in **Table 1**) to sort for chromosomal rearrangements coupled with the IF function from this procedure allowed for the reaffirmation of *GNAQ*, *GNA11*, and *BAP1* as plausible genetic drivers for

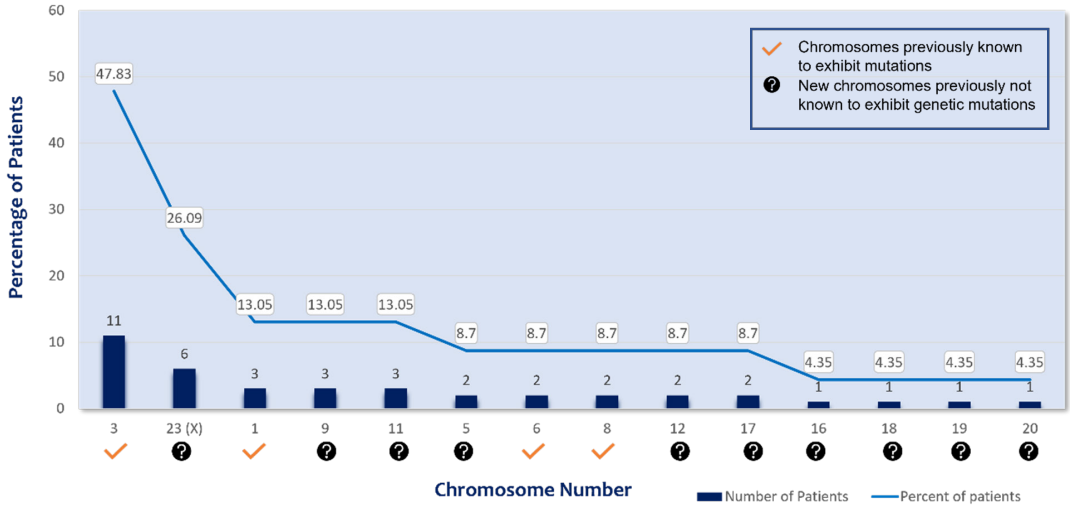

Figure 1. Bar chart of frequency of chromosomes with mutated genes in primary UM samples. This figure displays the number and percentage of primary UM patients in which each chromosome exhibited one or more genetic mutations. As shown, chromosomes 1, 3, 6, and 8, which were confirmed by previous studies to be associated with UM, were found to have genetic mutations in 8.7% - 47.83% of patients. Chromosomes 5, 9, 11, 12, 16, 18, 19, 20, and X, however, which exhibited mutations in 4.35% - 26.09% of patients, are newly identified and have not had any confirmed association with UM.

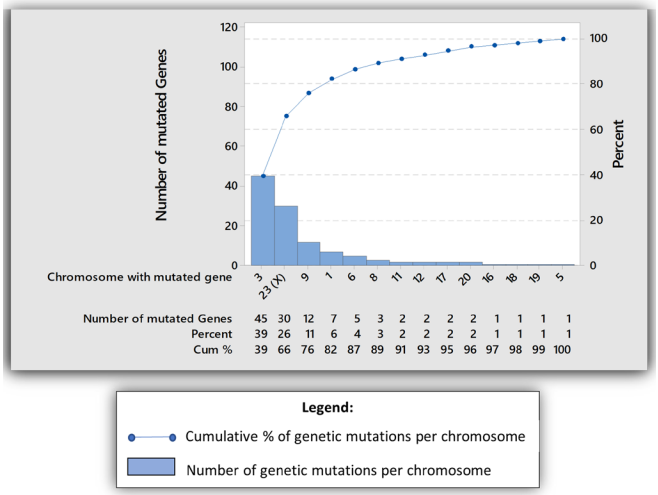

Figure 2. Pareto chart of frequency of genetic mutations per chromosome in primary UM samples. This figure displays the extent of chromosomal involvement across the primary UM patient sample in descending order of mutation count. The analysis revealed that chromosomes 3 and X exhibited the most genetic mutations with a combined total of 75 recurrent and non-recurrent, accounting for approximately 66% of all identified mutations across all primary samples.

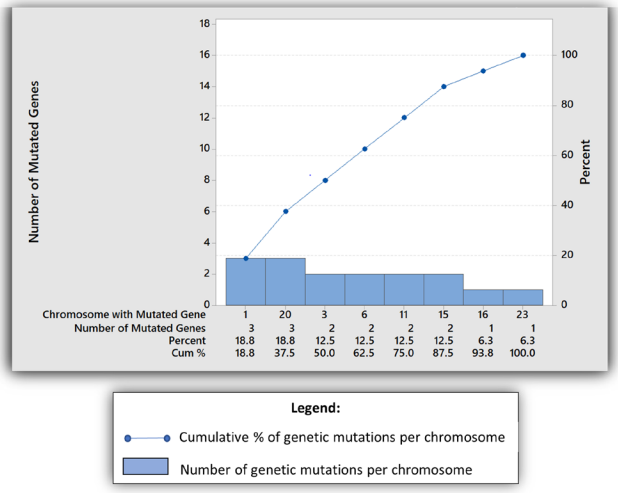

Figure 3. Pareto chart of frequency of genetic mutations per chromosome in metastasis UM samples. This figure displays the extent of chromosomal involvement across the metastasis UM patient sample in descending order of mutation count. It was found that chromosomes 1 and 20 exhibited the most genetic mutations with a combined total of 6 non-recurrent, accounting for approximately 37.5% of all identified mutations across all metastatic samples.

Table 1. Snapshot of uveal melanoma primary sample data.

Locus	Ploidy	Chromosome	Begin	End	Zygosity	VarType	Match	Reference	Allele2Seq	Gene Name
656203	2	chr1	1E+08	1E+08	hom	snp	X	C	T	NA
5356442	2	chr6	7E+07	7E+07	half	snp	X	G	?	NA
5561068	2	chr6	1E+08	1E+08	hom	snp	X	G	T	FOXO3;
9276693	2	chr11	1E+08	1E+08	hom	snp	X	C	T	NA
9363079	2	chr11	1E+08	1E+08	hom	snp	X	G	T	PCSK7;
9379683	2	chr11	1E+08	1E+08	hom	snp	X	A	G	NA
9400430	2	chr11	1E+08	1E+08	hom	snp	X	A	G	DCPS;
12932227	2	chr20	2E+06	2E+06	hom	snp	X	C	T	SIRPB1;
12932231	2	chr20	2E+06	2E+06	hom	snp	X	C	A	SIRPB1;
12932233	2	chr20	2E+06	2E+06	hom	snp	X	C	A	SIRPB1;
12932279	2	chr20	2E+06	2E+06	half	snp	X	T	?	SIRPB1;
12932666	2	chr20	2E+06	2E+06	half	snp	X	T	?	SIRPB1;
12932778	2	chr20	2E+06	2E+06	hom	snp	X	A	G	SIRPB1;
12935032	2	chr20	2E+06	2E+06	hom	snp	X	A	G	NA
12970932	2	chr20	1E+07	1E+07	hom	snp	X	A	C	PAK7;
12984366	2	chr20	1E+07	1E+07	half	snp	X	T	?	NA
12996659	2	chr20	2E+07	2E+07	hom	snp	X	A	G	MACROD2;
13011746	2	chr20	2E+07	2E+07	hom	snp	X	C	T	NA
13641551	2	chrX	336920	3E+05	half	snp	X	C	?	PPP2R3B;
13653396	2	chrX	1E+06	1E+06	half	snp	X	C	?	NA
13653496	2	chrX	1E+06	1E+06	half	snp	X	ACACGCGTGA	?	NA
13653842	2	chrX	1E+06	1E+06	half	snp	X	C	?	NA
13663423	2	chrX	2E+06	2E+06	half	snp	X	C	?	NA
13663425	2	chrX	2E+06	2E+06	half	snp	X	GA	?	NA
13663427	2	chrX	2E+06	2E+06	half	snp	X	T	?	NA
13663433	2	chrX	2E+06	2E+06	half	snp	X	C	?	NA

NOTE: *This table snapshot displays raw data from a sample primary UM Excel file. Red box: "Reference" and "Allele2Seq" columns, which are the matched normal and mutated DNA sequences, respectively. Green box: "Match" column with an imported IF function to analyze the reference and allele 2 sequences, showing Xs whenever there is a mismatch between the reference and allele2seq columns, thus signifying a mutation.*

UM.

Once all samples were surveyed for mutations, we conducted a Pareto analysis of chromosome involvement on the data which was consolidated and sequentially ordered as shown (**Table 2**). The analysis revealed that the highest frequency of patients exhibited genetic alterations to chromosome 3, accounting for nearly 50 percent of all genomes analyzed. Chromosome X was also found to have a high frequency of mutated genes and exhibited alterations in more than a quarter of the patients analyzed. Chromosomes 1, 9, and 11 follow with the third-highest frequency of ~13 percent, and all other chromosomes were affected in 2 or fewer patients.

Another Pareto analysis was conducted to determine the frequency of mutated genes in UM patients. As shown, the analysis uncovered seven genes that recurred in two or more UM patients (**Table 3**). ALG1L2 showed to have the highest recurrence, appearing to be mutated in 4 out of 23 (17.39%) patients. We only performed the Pareto analysis of mutated genes for the primary samples because the results revealed no recurring genetic mutations among the patients whose cells had undergone metastasis (**Figure 4**). Since this study identified hundreds of non-recurrent missense mutations, this high rate of recurrence suggests a link for these genes to cancer. To confirm this inference, we researched the functions of all the mutated genes and studied the recurrent ones extensively (**Table 4**). A gene ontology (GO) analysis of recurrent genetic mutations was conducted to explore the biochemical pathways, associated genes, and dysfunctional proteins that may contribute to the progression of UM.

From the gene ontology analysis, we found that *ALG1L2* encodes a putative glycosyltransferase protein which is responsible for transferring glycosyl groups (18, 19). *DMD* codes for dystrophin (26, 27). *IL1RAPL2* encodes an interleukin receptor accessory protein which is integral to immune response (26, 27). *KIAA0825* shares interactions with several genes including *PAX4*, *E2F1*, and *E2F4*, which

Table 2. Frequency of mutated genes per chromosome in primary UM samples (listed in descending order of frequency per patient).

Chromosomes Exhibiting Genetic Mutations	Number of Mutated Genes	Number of Patients	Percentage of Patients
3	45	11	47.83
X	30	6	26.09
1	7	3	13.05
9	12	3	13.05
11	2	3	13.05
5	1	2	8.70
6	5	2	8.70
8	3	2	8.70
12	2	2	8.70
17	2	2	8.70
16	1	1	4.35
18	1	1	4.35
19	1	1	4.35
20	2	1	4.35

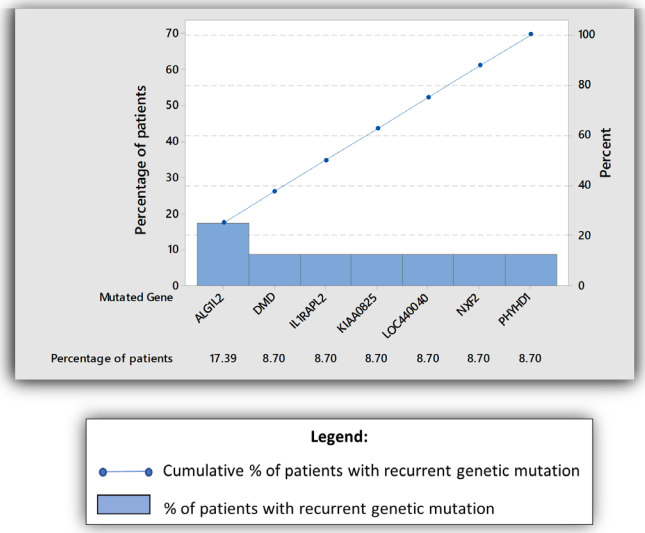

Figure 4. Pareto analysis of the frequency of recurrent genetic mutations in primary UM samples. This figure displays the most recurrent genetic mutations in decreasing order of frequency across all primary UM samples (we did not identify any recurrent mutations in the metastatic samples). There was a total of seven recurrently mutated genes: *ALG1L2, DMD, IL1RAPL2, KIAA0825, LOC440040, NXF2*, and *PHYHD1*. Gene *ALG1L2* was mutated in 17.39% of patients, making it the most frequently mutated gene across all UM samples.

code for transcription factors and control the behavior of other tumor suppressor proteins (26, 27). Long noncoding RNA *LOC440040* is a pseudogene of Glutamate Metabotropic Receptor 5 (GRM5) on chromosome 11. GRM5 is responsible for the regulation of neural network activity and synaptic plasticity (26, 27). *NXF2* encodes a nuclear RNA export protein that is responsible for RNA and nucleotide binding (26, 27). Nuclear RNA export factor 2, specifically, is involved in the export of mRNA from the nucleus to the cytoplasm (26, 27). *PHYHD1* encodes phytanoyl-CoA dioxygenase domain-containing protein 1, which has a role in the epigenetic regulation of gene transcription (29). *PHYHD1* also interacts with the FH (fumarate hydratase) and CAT (catalase) proteins, which act as a tumor suppressor and promote cell growth respectively (26, 27).

Table 3. Frequency of recurrent genetic mutations in primary UM samples (listed in descending order of frequency per patient)

Mutated Gene	Number of Patients	% of Patients
ALG1L2	4	17.39
KIAA0825	2	8.70
PHYHD1	2	8.70
LOC440040	2	8.70
NXF2	2	8.70
IL1RAPL2	2	8.70
DMD	2	8.70

Table 4. Information on location, function, and frequency of genes targeted by recurrent mutations in primary UM samples.

Mutated Gene	Cytogenic band	Gene Function	Freq.
ALG1L2 (ALG1, Chitobiosyldiphospho-dolichol Beta-Mannosyltransferase Like 2)	3q22.1	Protein Coding gene. Gene Ontology (GO) annotations related to this gene include transferase activity, transferring glycosyl groups.	17.39%
DMD Dystrophin	Xp21.2-p21.1	Protein Coding gene. Diseases associated with DMD include Muscular Dystrophy, Duchenne Type and Muscular Dystrophy, Becker Type. Gene Ontology (GO) annotations related to this gene include calcium ion binding and structural constituent of cytoskeleton.	8.7%
IL1RAPL2 Interleukin 1 Receptor Accessory Protein Like 2	Xq22.3	Protein Coding gene. Diseases associated include Cold Urticaria and Anterior Scleritis. Gene Ontology (GO) annotations related to this gene include interleukin-1 receptor activity and interleukin-1, Type II, blocking receptor activity.	8.7%
KIAA0825	5q15	Protein coding gene. Possible risk factor in Type II Diabetes and associated with high levels of glucose in the blood.	8.7%
LOC440040 Glutamate Metabotropic Receptor 5 Pseudogene	11p11.12	Pseudogenes play essential roles in gene regulation of their parent genes, and many pseudogenes are transcribed into RNA. Pseudogenes regulate tumor suppressors and oncogenes [13].	8.7%
NXF2 Nuclear RNA Export Factor 2	Xq22.1	Protein Coding gene. Among its related pathways are RNA transport and Transport of Mature Transcript to Cytoplasm. Gene Ontology (GO) annotations related to this gene include RNA binding and nucleotide binding.	8.7%
PHYHD1 Phytanoyl-CoA Dioxygenase Domain Containing 1	9q34.11	Protein Coding gene. Gene Ontology (GO) annotations related to this gene include dioxygenase activity.	8.7%

DISCUSSION

From our study, it can be concluded that UM is a heterogeneous disease characterized by non-recurrent mutations. This heterogeneous nature is a common characteristic of most types of aggressive melanomas and serves as essential information for predicting the appropriate form of treatment for UM patients. The Pareto analysis uncovered that the frequency of the recurrent genetic mutations in our UM cohort ranged from 2 to 4 out of 23 primary samples (8.7% - 17.39%), which is significant considering the size of the human genome.

ALG1L2, the most frequently mutated gene, was altered in 4 patients out of the 23 primary samples. This frequency is quite high considering that there are an estimated ~25,000 genes in the human genome, and thus the recurrences were likely not due to random chance (17). Additionally, *ALG1L2*'s encoded glycosyltransferase protein plays a critical role in determining the structure, stability, and function of a protein (18). If its expression were inhibited, this could potentially contribute to the onset of UM. For example, if tumor suppressor proteins were synthesized incorrectly due to the absence or deformation of glycosyltransferase, then this could consequently affect tumor suppression pathways.

DMD exhibited recurrent mutations in samples 19 and 21. *Although DMD* deletions are commonly associated with muscular dystrophy, mutations in *DMD* could alter the structure of dystrophin, which has been found to suppress myogenic tumors and prevent metastasis of cancerous cells (20, 21). *DMD* mutations could therefore contribute to myogenic tumor growth in the ciliary body, a ring-shaped muscle located behind the iris, thus facilitating the development of ciliary body melanoma (22). Additionally, research has shown that abnormal dystrophin levels are indicative of *DMD* involvement in the pathogenesis of several other cancers and melanomas (23).

Mutations in the protein-coding gene *IL1RAPL2* were also found in two patients. *IL1RAPL2*'s encoded interleukin receptor accessory protein may facilitate tumor development when mutated, as the primary function of interleukins is to modulate cell growth, differentiation, and activation during inflammatory responses (24). Additionally, chronic inflammation can often damage DNA and promote carcinogenesis, and interleukin signaling in cancerous cells has been researched as a critical factor in cancer development, progression, and control (25).

KIAA0825 was also identified as a recurrently mutated gene, but little is known about its gene product. However, it is known that *KIAA0825*'s shared genetic interactions and indirect influence on transcription factors and tumor suppressor proteins may serve as a plausible cause for the development of UM or other cancers.

LOC440040, which was mutated in samples 1 and 20, as previously mentioned, is a pseudogene for GRM5. According to the Gene Expression Database, GRM5 is expressed in the visual system and helps modulate synapse activity through glutamate signaling. Dysregulation of transporters and dysfunctional glutamate receptors can adversely affect glutamate signaling beyond the central nervous system, which could promote cancer development in visual pathways (28).

NXF2, which was mutated in samples 3 and 17, plays a significant role in transporting mRNA for transcription. Therefore, issues with mRNA transport can have a cascading effect on transcription and protein synthesis, which could potentially correlate with the development of UM. Diseases associated with *NXF2* include Progesterone-Receptor Negative Breast Cancer and Spermatogenic Failure (26, 27).

Finally, *PHYHD1*, mutated in samples 19 and 22, regulates gene transcription and is associated with proteins which, if altered, may accelerate unregulated cell division or tumor growth.

It was also found that most of the non-recurrent genes that were mutated coded for transcription factors and tumor suppressors – proteins that, when defective, are known to cause other cancers.

The genetic mutations were homozygous, meaning that UM is a recessive disorder, which explains the rarity of this cancer. Although two recurrent mutations target genes on the X chromosome, no pattern was observed between patient gender and the identified mutations, thus suggesting that UM is an autosomal disease. All mutations were missense mutations, resulting from single base mismatches within the protein and non-protein-coding regions of the DNA. Across all samples, we observed 15 genes (8 pseudogenes and 7 RNA genes) that were mutated in non-protein-coding regions. Only one of these genetic mutations, present on LOC44004, was recurrent. The other two types, frameshift and nonsense mutations, were not found within any of the samples that were analyzed. All recurrent genetic mutations that were found in this study (**Figure 5**) have not been listed in any source of literature, suggesting that these are potentially new mutations responsible for UM, and likely responsible for other cancers (the same genetic mutations are often responsible for multiple types of cancers, such as genetic mutations in *BRAF*, *V600E*, *TP53*, and *CDKN2A*, which have a high recurrence in cutaneous melanomas).

Our work narrows down the best treatment approach for people suffering from UM, as our results have shown that UM is caused by several unique genetic mutations. With this heterogeneity, targeted therapy may not be the best approach. Targeted therapies are developed to target and inhibit the function of a specific gene or defective protein that results from genetic mutations. For example, Larotrectinib is a medication for solid tumors that inhibits TRK (tyrosine kinase), a protein that promotes cancer (30). TRK is produced because of the fusion of two genes due to an underlying genetic mutation in neurotrophic receptor tyrosine kinase (NTRK). The best treatment option predicted for UM could include immunotherapy, which prevents disease by stimulating and enhancing the immune system to fight against dysfunctional proteins caused by mutations. Immunotherapy would be effective in reducing the cumulative effects of all these harmful genes and proteins. In contrast, targeted therapy could be used to inactivate specific genes and pathways which are implicated in UM.

Additionally, the recurrent genetic mutations identified in this study can be activated in mice models to assess the phenotypic consequences and potentially causational relationship with UM (31).

MATERIALS AND METHODS

All data for this experiment was collected from Complete Genomics Inc., which hosts publicly available data published by the National Center for Biotechnology Information (NCBI). Sample genome data was publicly available and provided in the "Supplemental Data" section of a 2016 whole-genome sequencing study which is accessible through the National Institute of Health (NIH) National Library of Medicine (14). This data was obtained from 32 UM patients, and the data package included 32 data files (23 Primary Uveal Melanoma (PUM) files, and 9 Metastases Uveal Melanoma (MUM) files). The UM patient data was in Excel format, and each file included a sample reference DNA sequence (matched normal) and a mutated allele 2 sequence from each patient. Each file also contained specific data on each subject, including gender, date of data collection, and sample source. Microsoft Excel was used in the analysis of patient genome data to identify and locate genetic mutations. Minitab 2018 software was then used to perform a Pareto analysis to analyze the frequency of all genetic mutations and chromosomes involved.

Analysis methodology

We used the following procedure to examine each SNV file using Microsoft Excel: Initially, we applied data filters to all column headings to efficiently sort and organize genome data. Next, we inserted a new column labeled "Match" between the "reference" (matched normal sequence) column and "allele2Sequence" (mutated sequence) column (**Table 1**). It is important to note that the "allele1Sequence" column could be ignored, as it only contained nitrogenous bases that were complementary to the reference sequence. The 'Match' column was not present within the data and was created with an automated IF function to compare each cell from the reference sequence with the corresponding cell from the allele 2 sequence. Whenever there was a mismatch or instance where the logical test within the IF function failed, the function automatically flagged each mutation with an 'X' symbol in the corresponding cell within the 'Match' column. This IF function in the Match column was applied to the entire patient genome to search for mutations.

After the comparison was complete, we filtered the Match column to show all detected base pair mismatches, which were indicated by an 'X' in the 'Match' column (**Table 1**). All '?' in the allele 2 sequence indicated that there was an unknown base present, therefore we disregarded those mismatches as mutations due to ambiguity. Additionally, we filtered out all empty cells or "blanks" in the allele 2 sequence, as those cells were haploid and only had one allele. For each observed mutation, recurrent or non-recurrent, we noted the corresponding zygosity, variation type, and the name of the gene that was affected (this information was provided in the "Zygosity," "varType," and "Gene Name" columns, respectively). We followed this analysis procedure for both the primary (PUM) and metastases (MUM) files.

Pareto analysis

After identifying mutations across all patient samples, we

performed an in-depth Pareto analysis to assess the frequency of chromosomes with observed genetic mutations, as well as the frequency of individual recurrent genetic mutations. We started by listing all chromosomes with mutated genes sequentially from chromosome 1 to 23 (X), along with the count of mutated genes on each chromosome. We counted the number of patients with one or more observed mutations (**Table 2**), and then using the Minitab software, performed a Pareto analysis of chromosomes with mutated genes by listing the chromosomes in descending order by a frequency distribution of mutation count in all patients from the primary sample (non-metastatic). For example, chromosome 3 exhibited the highest number of mutations – a total of 45 – and therefore it was the first bar in the Pareto chart (**Figure 2**). This frequency analysis method was then repeated for patients with metastasis (**Figure 3**).

We subsequently performed a Pareto analysis of mutated genes in all patients from the primary sample by listing mutated genes in descending order by a frequency distribution of mutation count (**Figure 4**). For example, a mutation on ALG1L2 was observed in four patients, thus equating to a 17.39% recurrence. Following Pareto analysis, we researched the gene functions of all mutated genes (**Figure 5**) to investigate links between gene function and anomalies that may result in cancer, and more specifically, UM (**Table 4**).

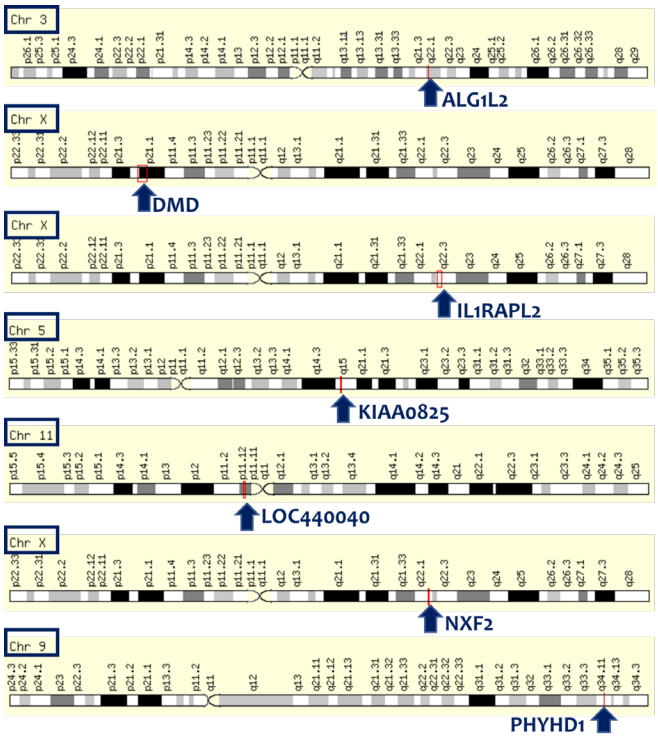

Figure 5. Genes targeted by recurrent mutations and respective locations. All recurrently mutated genes - *ALG1L2, DMD, IL1RAPL2, KIAA0825, LOC440040, NXF2*, and *PHYHD1* - have their chromosomal locations displayed above.

ACKNOWLEDGEMENTS

We would like to thank the anonymous judges at Arizona Science and Engineering Fair (AZSEF 2019) and Intel International Science and Engineering Fair (ISEF 2019) whose questions and feedback on our research helped improve this paper.

Received: September 10, 2021
Accepted: August 7, 2022
Published: August 25, 2022

REFERENCES

1. Mayo Foundation for Medical Education and Research. (2022, June 18). "*Melanoma.*" Mayo Clinic. Retrieved July 2, 2022
2. Kim, H. S., Kim, E. K., Jun, H. J., *et al.* (2010). "*Noncutaneous malignant melanoma: A prognostic model from a retrospective Multicenter Study.*" *BMC Cancer, 10*(1). doi.org/10.1186/1471-2407-10-167
3. Chattopadhyay, Chandrani, *et al.* "*Uveal Melanoma: From Diagnosis to Treatment and the Science in Between.*" Cancer, vol. 122, no. 15, 2016, pp. 2299–2312., doi.org/10.1002/cncr.29727.
4. Doherty RE, Alfawaz M, Francis J, *et al.* Genetics of Uveal Melanoma. In: Scott JF, Gerstenblith MR, editors. "*Noncutaneous Melanoma.*" Brisbane (AU): Codon Publications; 2018 Mar. Chapter 2.
5. Amaro, A., Gangemi, R., Piaggio, F. *et al.* (2017). *"The biology of uveal melanoma. Cancer metastasis reviews,"* 36(1), 109–140. doi.org/10.1007/s10555-017-9663-3
6. Helgadottir H, Höiom V. "*The genetics of uveal melanoma: current insights.*" Appl Clin Genet. 2016; 9:147-55. Published 2016 Sep 6. doi:10.2147/TACG.S69210
7. Kivela T. "*The epidemiological challenge of the most frequent eye cancer: retinoblastoma, an issue of birth and death.*" Br J Ophthalmol. 2009;93(9):1129–1131. doi: 10.1136/bjo.2008.150292
8. Kaliki, S, and C L Shields. "*Uveal Melanoma: Relatively Rare but Deadly Cancer.*" Eye, vol. 31, no. 2, 2016, pp. 241–257., doi.org/10.1038/eye.2016.275.
9. "*Metastatic Melanoma: Melanoma Types.*" Cancer Treatment Centers of America, 21 Sept. 2021, www.cancercenter.com/cancer-types/melanoma/types/metastatic-melanoma.
10. Carvajal, Richard D, *et al.* "*Metastatic Disease from Uveal Melanoma: Treatment Options and Future Prospects.*" British Journal of Ophthalmology, vol. 101, no. 1, 2016, pp. 38–44., doi.org/10.1136/bjophthalmol-2016-309034.
11. Hoorens, A. (2022). "*Detection and significance of micrometastases and isolated tumor cells in lymph nodes of colorectal cancer resections.*" The Lymphatic System in Colorectal Cancer, 115–130. doi.org/10.1016/b978-0-12-

824297-1.00011-7
12. "*Rare Disease Database: Ocular Melanoma*" (2018), (2022, June 18). National Organization for Rare Disorders (NORD), rarediseases.org/rare-diseases/ocular-melanoma/
13. Decatur CL, Ong E, Garg N, *et al*. "*Driver Mutations in Uveal Melanoma: Associations with Gene Expression Profile and Patient Outcomes*." JAMA Ophthalmol. 2016 Jul 1;134(7):728-33. doi: 10.1001/jamaophthalmol.2016.090
14. Royer-Bertrand B, Torsello M, Rimoldi D, *et al*. "*Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing*." Am J Hum Genet. 2016 Nov 3;99(5):1190-1198. doi: 10.1016/j.ajhg.2016.09.008. Epub 2016 Oct 13.
15. "*Interpret the Key Results for Pareto Chart*." (2022, June 18). Minitab 18 Support, support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/quality-tools/how-to/pareto-chart/interpret-the-results/key-results/
16. "*What Is a Pareto Chart?*" (2022, June 18). ASQ, asq.org/quality-resources/pareto.
17. International Human Genome Sequencing Consortium. "*Finishing the Euchromatic Sequence of the Human Genome*." Nature, vol. 431, no. 7011, 2004, pp. 931–945., doi.org/10.1038/nature03001.
18. Ashburner *et al*. "*Gene ontology: tool for the unification of biology*." Nat Genet. May 2000;25(1):25-9
19. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. Jan 2021;49(D1):D325-D334.
20. Tayeb, Mohammed T. "*Deletion Mutations in Duchenne Muscular Dystrophy (DMD) in Western Saudi Children*." Saudi Journal of Biological Sciences, vol. 17, no. 3, 2010, pp. 237–240., doi.org/10.1016/j.sjbs.2010.04.008.
21. Wang, Yuexiang, *et al*. "*Dystrophin Is a Tumor Suppressor in Human Cancers with Myogenic Programs*." Nature Genetics, vol. 46, no. 6, 2014, pp. 601–606., doi.org/10.1038/ng.2974.
22. "*NCI Dictionary of Cancer Terms*." National Cancer Institute, www.cancer.gov/publications/dictionaries/cancer-terms/def/ciliary-body.
23. Jones, Leanne, *et al*. "*The Duchenne Muscular Dystrophy Gene and Cancer*." Cellular Oncology, vol. 44, no. 1, 2020, pp. 19–32., doi.org/10.1007/s13402-020-00572-y.
24. Justiz Vaillant, Angel A. and Ahmad Qurie. "*Interleukin*." StatPearls, StatPearls Publishing, 30 August 2021.
25. Briukhovetska, Daria, *et al*. "*Interleukins in Cancer: From Biology to Therapy*." Nature Reviews Cancer, vol. 21, no. 8, 2021, pp. 481–499.
26. Ashburner, M *et al*. "*Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*." Nature genetics vol. 25,1 (2000): 25-9. doi:10.1038/75556
27. Gene Ontology Consortium. "*The Gene Ontology resource: enriching a GOld mine*." Nucleic acids research vol. 49, D1 (2021): D325-D334. doi:10.1093/nar/gkaa1113
28. Yi, Haowei, *et al*. "*Glutamate in Cancers: From Metabolism to Signaling*." The Journal of Biomedical Research, 2020, p. 1., doi.org/10.7555/jbr.34.20190037.
29. Losman, Julie-Aurore, *et al*. "*2-Oxoglutarate-Dependent Dioxygenases in Cancer*." Nature Reviews Cancer, vol. 20, no. 12, 2020, pp. 710–726., doi.org/10.1038/s41568-020-00303-3.
30. Riedel, Richard. "*Faculty Opinions Recommendation of Efficacy of Larotrectinib in Trk Fusion-Positive Cancers in Adults and Children*." Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature, 2018, doi.org/10.3410/f.732725726.793552159.
31. Gurumurthy, Channabasavaiah B, and Kevin C Kent Lloyd. "*Generating mouse models for biomedical research: technological advances*." Disease models & mechanisms vol. 12,1 dmm029462. 8 Jan. 2019, doi:10.1242/dmm.029462