UNIVERSITY OF CALIFORNIA,
IRVINE

Genomics and Epidemiological Analysis of Melanoma Laterality

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Biomedical Engineering

by

Winnie Zhu Fan

Thesis Committee:
Associate Professor James P. Brody , Chair
Assistant Professor Chang C. Liu
Assistant Professor Jered B. Haun

2018

DEDICATION

To

Grandma

AND

Friends, family, and my partner,

For putting up with and sometimes even supporting my nonsense.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract of the Thesis

Genomics and Epidemiological Analysis of Melanoma Laterality

By

Winnie Fan

Master of Science in Biomedical Engineering

University of California, Irvine, 2018

Professor James P. Brody, Chair

Skin cancer is the most commonly diagnosed cancer in the United States and melanoma is considered the deadliest form of skin cancer. Although the environmental causes of melanomas are known, the molecular mechanisms involved are still being researched. Melanomas present more often on the left side of the body, but explanations for this laterality are conflicting and largely focused on epidemiological factors. In this thesis, both epidemiological and genetic factors affecting melanoma laterality are analyzed to explore how tumor laterality and patterning may arise in general. The Surveillance, Epidemiology, and End Results (SEER) and The Cancer Genome Atlas (TCGA) databases were used to analyze clinical cases of melanoma. Data analysis was conducted by calculating a laterality ratio of asymmetric melanomas and comparing how these ratios differ in epidemiological and genetic variables. A machine learning algorithm was also applied to predict which variables or groups of variables may determine laterality. Results showed that, as established, melanomas tend to exhibit left-sided laterality, but epidemiological factors alone are not good indicators of where tumors present. Genomics analysis revealed several genes and targets of interest. Genes involved in cell adhesion were consistently significant, but there was no conclusive evidence that a specific

gene or set of genes causes left-sided patterning. Although results did not reveal specific genetic targets as determinants of melanoma laterality, they prove that the methods used can be tools for analyzing tumor laterality in general and can help in predicting what variables are important in molecular mechanisms indicating laterality.

Chapter 1: Introduction

Of the three main types of skin cancer, melanoma is considered the deadliest, with a death toll eight times higher than basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) [39]. Many studies have shown that sun exposure, and more specifically UV radiation, is highly correlated with the development of skin cancer [35], though the exact molecular mechanisms at play are still being researched [39].

More skin cancer tumors are diagnosed on the left side than the right [2]. This tends to hold true across gender, age group, and primary location of tumor [21]. The exact cause of this is unknown, but previous studies have postulated that UV exposure patterning is a major cause. In particular, it has been suggested that driving patterns are the main reason for this left-sided bias, especially as many studies are done in the US, where drivers are more exposed to UV light on their left side [20].

However, more recent studies done in the UK, where drivers are predominantly exposed to UV light on their right side, show the same asymmetrical bias for skin cancer tumors on the left side [2]. Other studies of skin cancer laterality done outside the US show this same pattern [21]. Although development of left-sided or right-sided skin tumors in specific locations on the body may have to do with driving patterns [20], there may be more at play than UV exposure patterning in skin cancer asymmetry [2].

Developmental patterning may play some part in this, especially as embryology is asymmetrical and there are similar patterns of tumor bias in paired organs [2]. Nodal, a morphogen involved in several developmental functions including left-right asymmetry during embryogenesis, could be a contributing factor and is known to be secreted in aggressive

melanomas of the skin [2], [21], [22]. However, conclusive evidence in current research has not been found for Nodal being a major cause of asymmetrical distribution of skin cancer tumors.

In this thesis, the genomic and epidemiological basis of asymmetry in melanomas is investigated using two publicly available databases. The Cancer Genome Atlas (TCGA) provides a Genomic Data Commons Data Portal (GDC) that contains harmonized cancer datasets which were used to analyze the genomics of melanoma asymmetry [41]. The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI) provides information on cancer statistics and epidemiology and was used to analyze any epidemiological bases of melanomas [30].

Chapter 2: Methods

Surveillance, Epidemiology, and End Results (SEER) data collected from 1973 to 2014 was used to analyze trends in epidemiological and population data related to melanomas. This data was accessed after signing a Research Data Agreement in May 2017. The raw database files were downloaded and loaded into a local database using the SQLite program and a custom Python script.

From this database, tumors arising from melanoma were identified using the 9th revision of the 6th edition (2014) International Classification of Diseases list [30]. These cases were assigned a laterality ratio based on the number of tumors diagnosed on the left side of the body divided by the number of tumors diagnosed on the right side of the body. Cases where laterality was not assigned or unknown were not used. A total of 275,000 cases were identified. For each case captured as described above, sex, age at diagnosis, patient location, tumor size, vital status, and survival months were recorded. A rate ratio test was performed as described above for these recorded categories and a p-value was generated to observe whether or not differences in left and right diagnoses for these cases were significant.

To examine whether groups within categories were significantly different from each other, a z-score was generated based on the rate ratio and total number of cases based on sex, patient location, and vital status. To evaluate whether rate ratios were predicted by age at diagnosis, tumor size, or survival months, the data were fit to a local regression using the loess method.

The Cancer Genome Atlas (TCGA) provided a Genomic Data Commons (GDC) Data Portal that contained cancer datasets organized by primary site of tumor. From this data portal, we

isolated cases where the skin was the primary tumor site. At the time this was done, the only project available was the “TCGA-SKCM” project which contained 470 cases. Using available, anonymized clinical data from GDC, we classified the specific sites on the body where the tumor originated as being either on the left side, right side, or unknown. Any cases where tumor laterality was unknown were not used.

Sequencing data as well as other next generation sequencing information from TCGA are available on Google BigQuery, a serverless cloud platform, and BigQuery was used to query these databases. The specific TCGA cloud used was “isb-cgc” and the datasets used were from the “TCGA_hg38_data_v0” dataset. The cases we pulled from GDC were uploaded to a private BigQuery project and used to filter the publicly available datasets.

The primary language used to query BigQuery is SQL and it was used to query and filter the TCGA data for the “TCGA-SKCM” project by left or right designated laterality for data pertaining to copy number segment variation, DNA methylation, miRNA expression, protein expression, RNA gene expression, and somatic mutation location. Data from the listed categories were imported into R for analysis after merging tables designated “left” or “right” for each category.

It was assumed that the rate ratio of observations for each category followed a Poisson distribution for tumors found on both the left and right side. Primary analysis for the copy number segment variation and somatic mutation tables was conducted using a rate ratio test to determine if there were differences in the rates of occurrence per chromosome segment by gene or mean segment length on the left vs right side and we used a 95% confidence interval to

determine statistically significant differences in these tumors. A p-value was generated to gauge significance on this confidence interval.

For DNA methylation, miRNA expression, protein expression, and mRNA expression a z-score was generated to measure whether left-right differences in expression per probe, miRNA ID, protein, or gene were statistically significant relative to the number of cases associated with tumors of the left or right side of the body. A Bonferroni adjustment was performed to adjust p-values so that each test for each gene, protein, miRNA, or probe was treated as an individual test. Thus, the p-value was adjusted such that $p_{adj} = \frac{p_{original}}{n}$, where n is the number of tests performed (eg. number of proteins compared).

Machine learning algorithms were applied to the protein expression dataset as well as a means of testing whether specific proteins or groups of proteins could predict laterality. H2O, an open source machine learning and predictive analytics platform, was used to run these algorithms. The algorithms are implemented on top of H2O's distributed Map/Reduce framework and utilize the Java Fork/Join framework for multi-threading. The data is read in parallel and is distributed across the cluster and stored in memory in a columnar format in a compressed way. H2O's data parser has built-in intelligence to guess the schema of the incoming dataset and supports data ingest from multiple sources in various formats [40].

Chapter 3: Results

3.1 Sex, Tumor Primary Site, and Geographical Location of Patient

We examined 275,000 cases of melanoma diagnosed in the United States from 1973 to 2014 as recorded by the SEER program. We first examined sex to see if there were significant differences in laterality ratios between male and female patients. Prior studies have reported mixed results about how melanoma asymmetry and sex are related, with some studies suggesting that men have a higher incidence of left-sided melanomas based on driving patterns [21], [24].

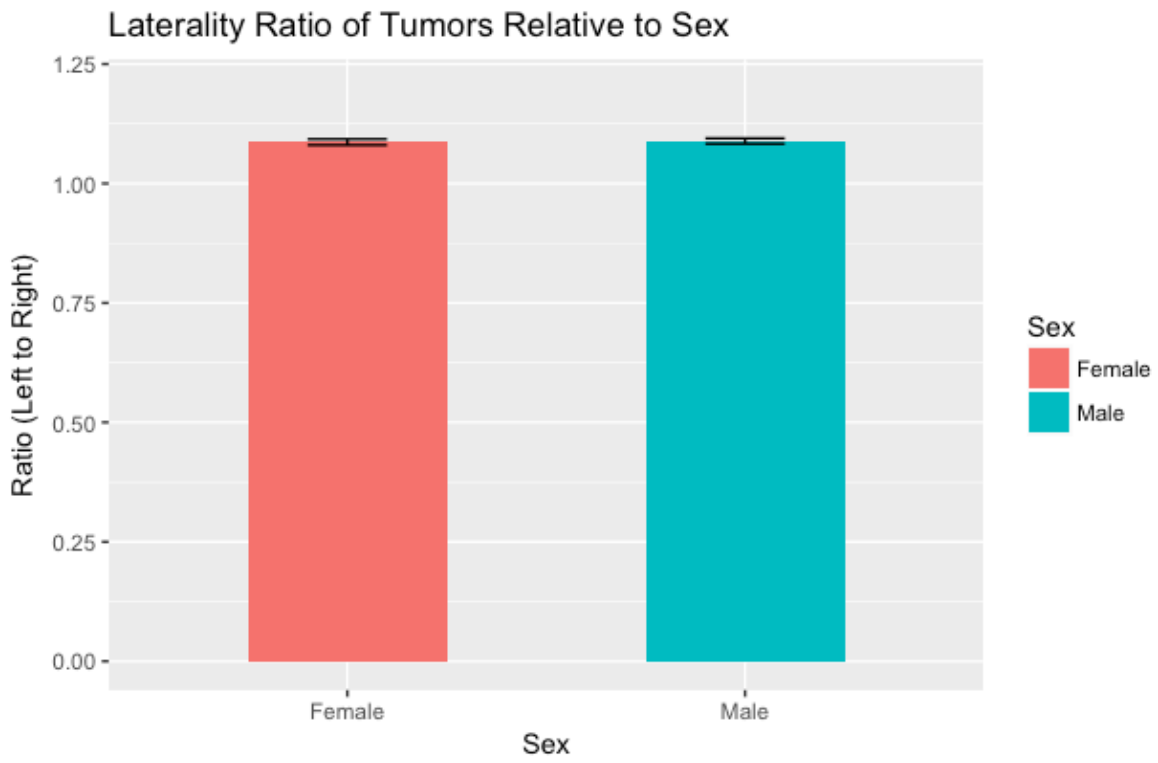

Figure 1: Left to right laterality ratio of melanomas found in female and male patients with error bars that represent ratio error.

The 125,272 female patients enrolled in the study have a left to right ratio of approximately 1.087 with a standard error of 0.00614 and a p-value of 8.68×10^{-49}. The 149,728

male patients enrolled have a left to right ratio of approximately 1.089 with a standard error of 0.00563 and a p-value of 6.31×10^{-61}.

We next examined whether the primary site of the melanoma affected laterality ratios. Sites were coded using the *International Classification of Diseases for Oncology, Third Edition*. Tumor sites that were not specifically designated skin cancer tumor sites were excluded.

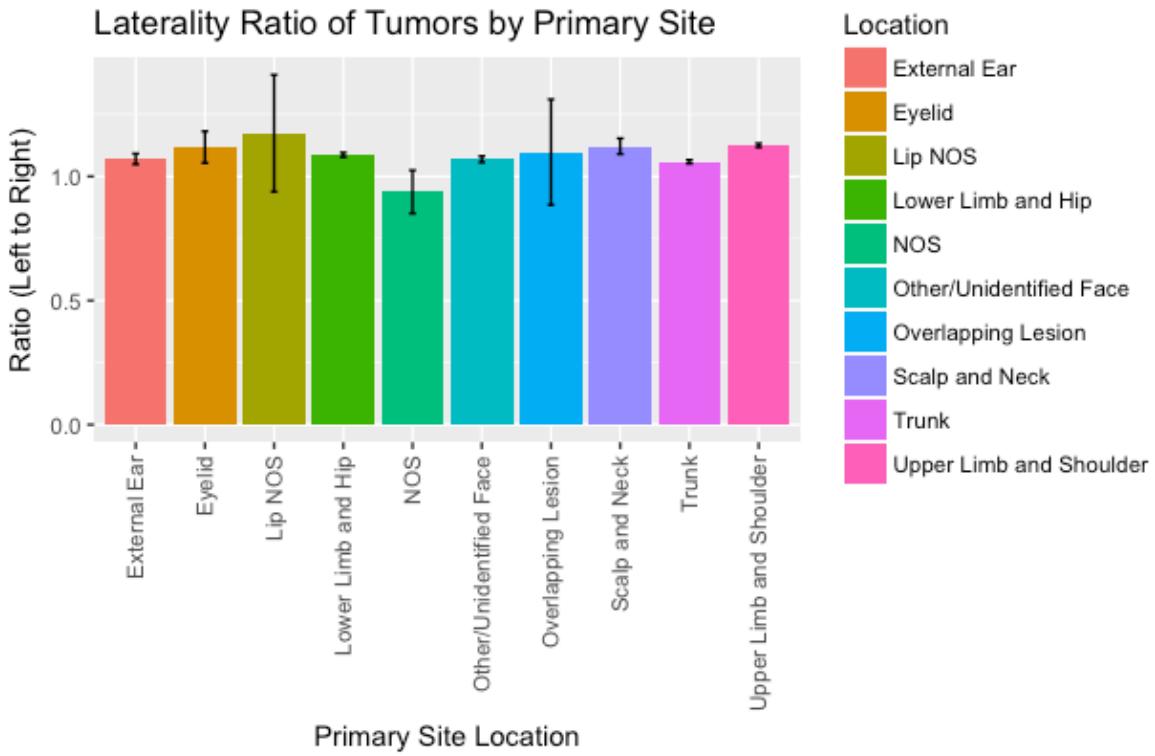

Figure 2: Left to right laterality ratios of melanomas by primary clinical site with error bars. Sites where there were fewer cases tended to have larger error bars.

A log transform of the laterality ratios was performed to more accurately differences in laterality ratio for data visualization.

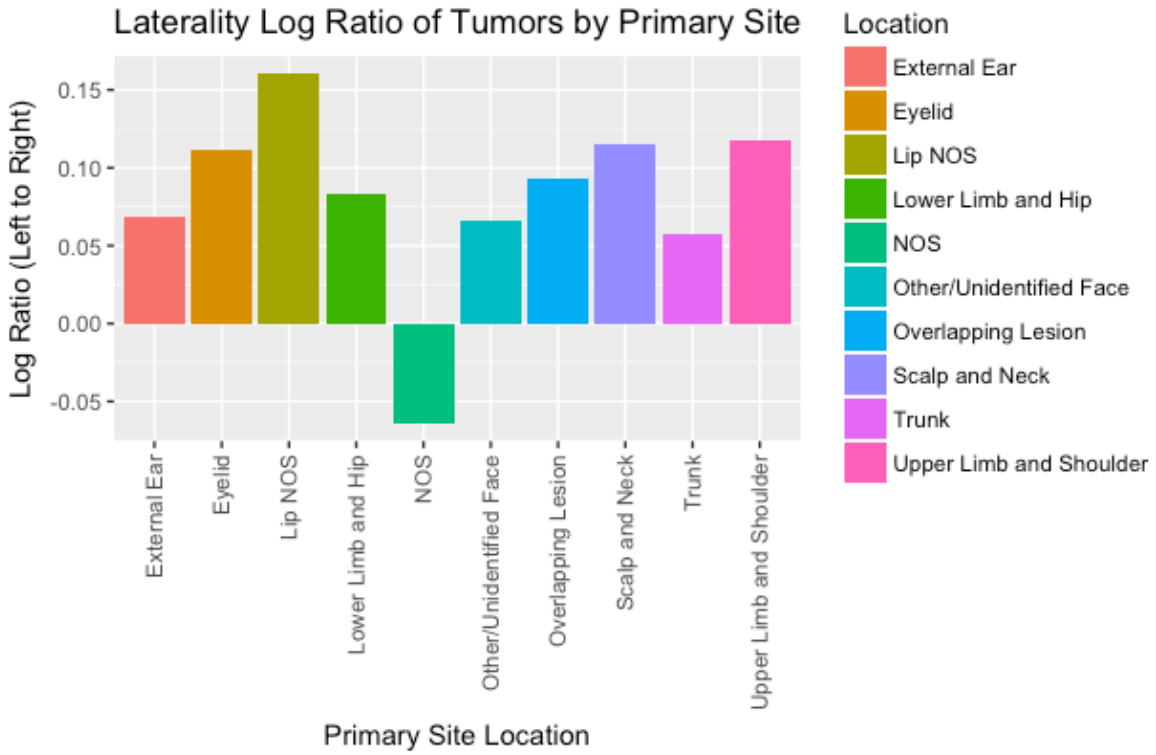

Figure 3: Log transform of laterality ratios by primary clinical site. The log transform more clearly shows the differences in laterality by site.

The average left to right laterality ratio was 1.086 with a standard deviation of 0.0623. When compared with the average, tumors classified as Not Otherwise Specified (NOS) had a significantly lower laterality ratio and tumors classified as Lip NOS had a significantly higher laterality ratio using a 95% confidence interval.

The 467 cases of NOS tumors had a laterality ratio of 0.9378 with a standard error of 0.08683 and the 100 cases of Lip NOS tumors had a laterality ratio of 1.174 with a standard error of 0.2355.

We also examined whether location of the patient affected laterality ratios. This was dependent of sites that are registered with SEER and thus, only participating states were included in this analysis.

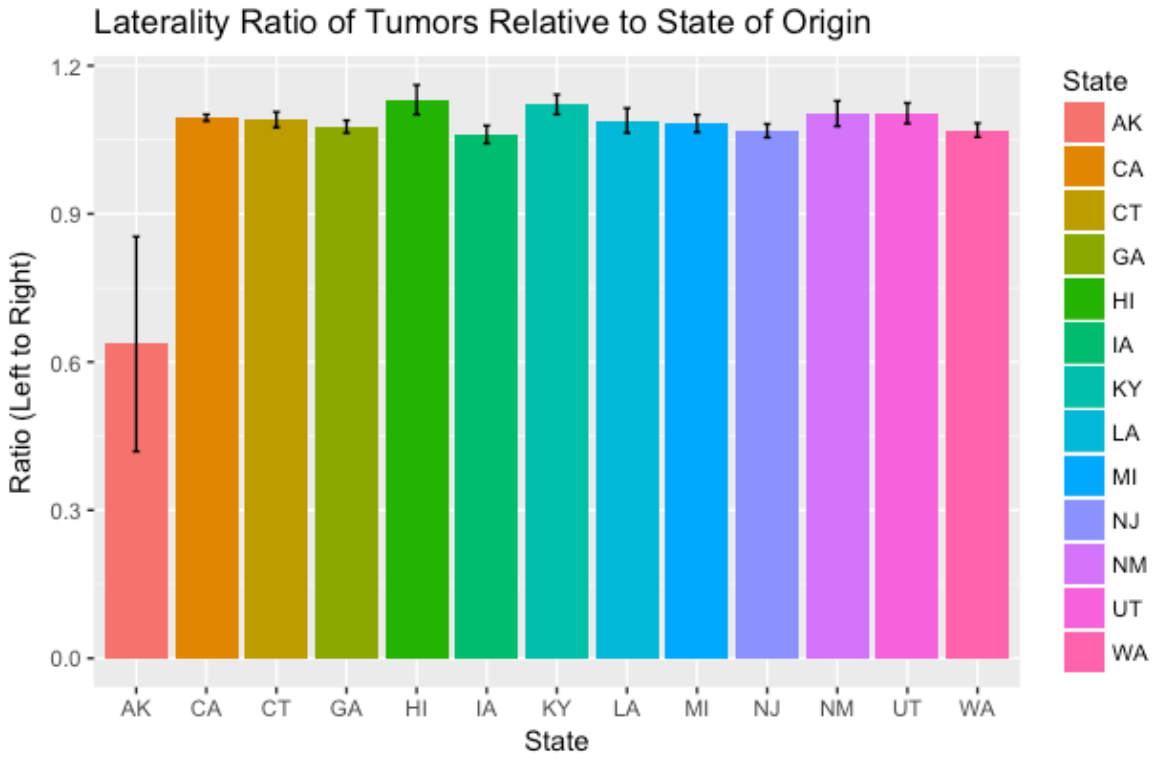

Figure 4: Left to right laterality ratios of melanomas found in patients by state. States where there were fewer cases tended to have larger error bars.

A log transform of the laterality ratios was performed to more accurately differences in laterality ratio for data visualization.

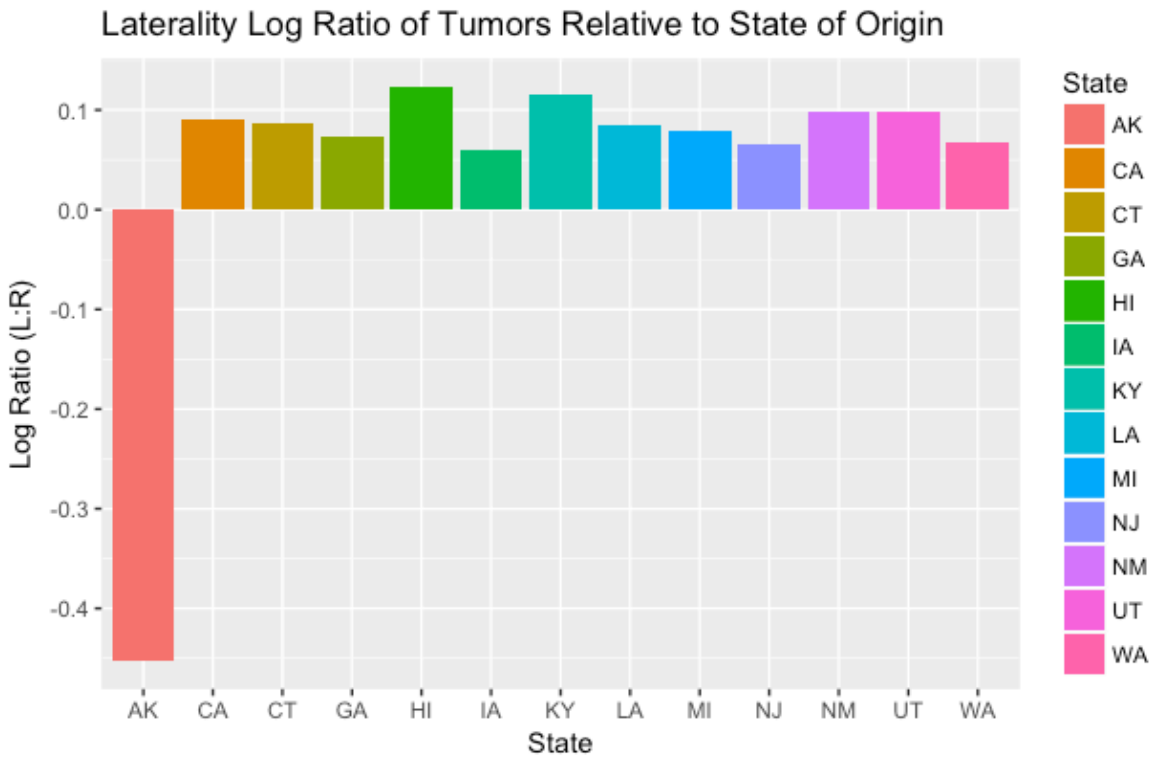

Figure 5: Log transform of left to right laterality ratios of melanomas found in patients by state. The log transform more clearly shows the differences in laterality by state.

The average laterality ratio across all states that have a SEER registry is 1.056 with a standard deviation of 0.1277. Compared to the average laterality ratio, patients from Alaska showed a significant difference in laterality ratio. Of the 36 enrolled patients in Alaska, the laterality ratio was 0.6364 with a standard error of 0.2175.

3.2 Vital Status and Survival Months

To examine whether there were differences in melanoma laterality in terms of survivability, we examined the vital status of patients enrolled as well as how long a patient survived in months based on study contact information [25]. The study data encompassed a total of 503 months and unknown survival times were coded 9999. For the purposes of our study, patients with unknown survival times were excluded.

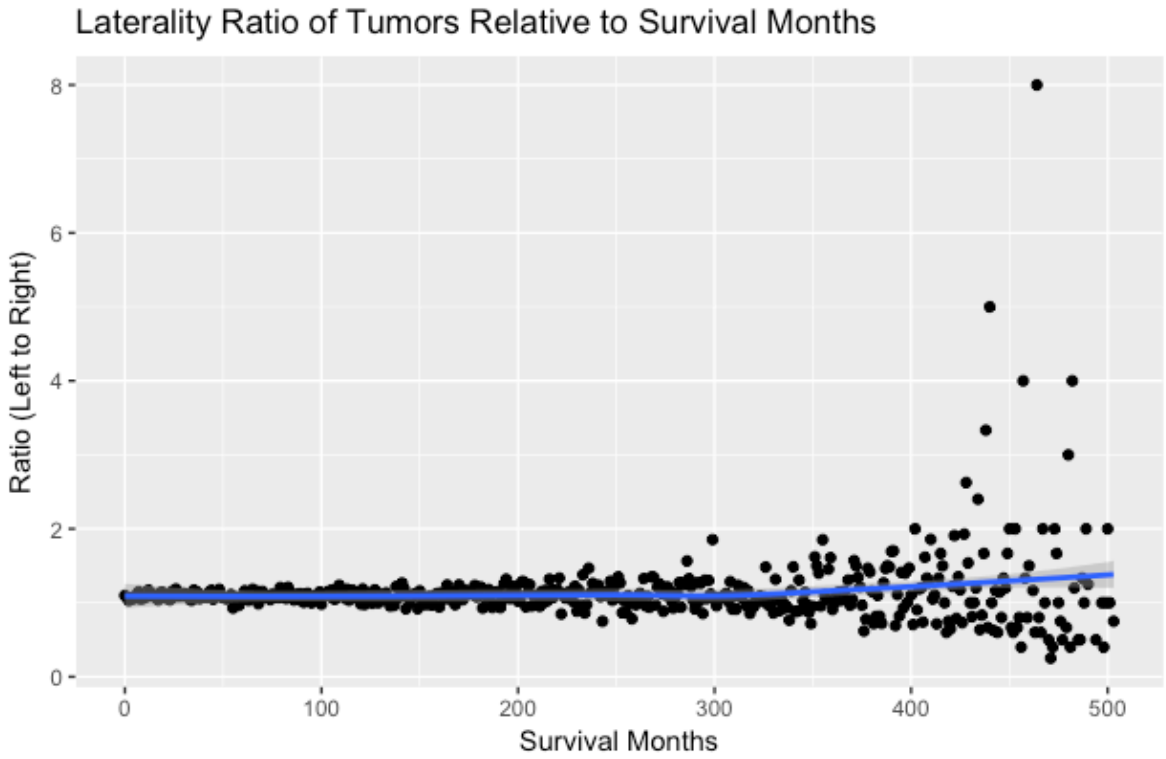

Figure 6: Left to right laterality ratio of tumors by patient survival months. The blue line within the shaded gray space represents the local polynomial regression (LOESS) fit of the data.

The average left to right laterality ratio for survival months is 1.145 and it appears that there is a higher incidence of left-sided melanomas regardless of survival time. A log transform was applied to better represent the variability of data at increased survival times.

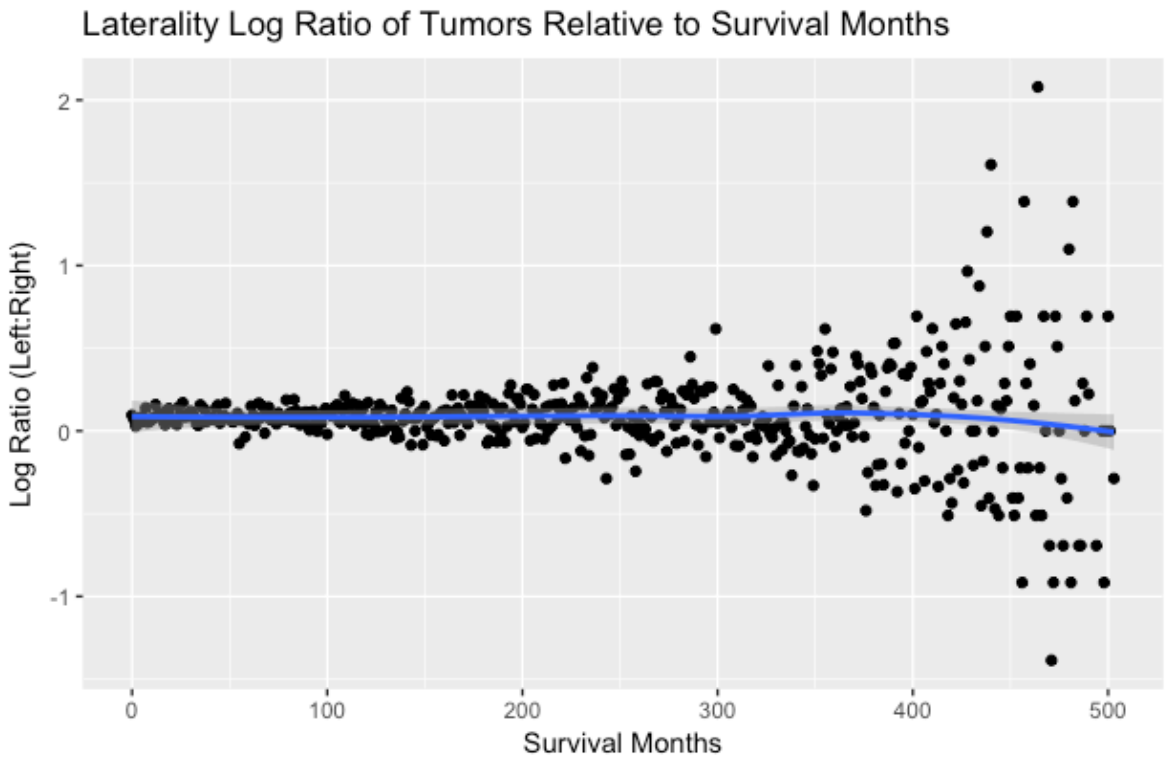

Figure 7: Log transform of left to right ratio of tumors by patient survival months.

The average left to right laterality ratio for survival months is 1.145 and it appears that there is a higher incidence of left-sided melanomas regardless of survival time. A log transform was applied to better represent the variability of data at increased survival times. Because the SEER program is ongoing, patients who have survived the longest represent a smaller fraction of the total patient population.

A linear regression was performed to gauge how well survival time predicted laterality differences. The overall slope of the best fit line was found to be 0.0004 with a p-value < 0.01. This implies that there is relatively little change as survival time increases or decreases.

To further examine if survivability affects melanoma laterality ratios, we looked at the vital status of patients enrolled in the study. Patients were classified as either “alive” or “dead” as of their study follow-up.

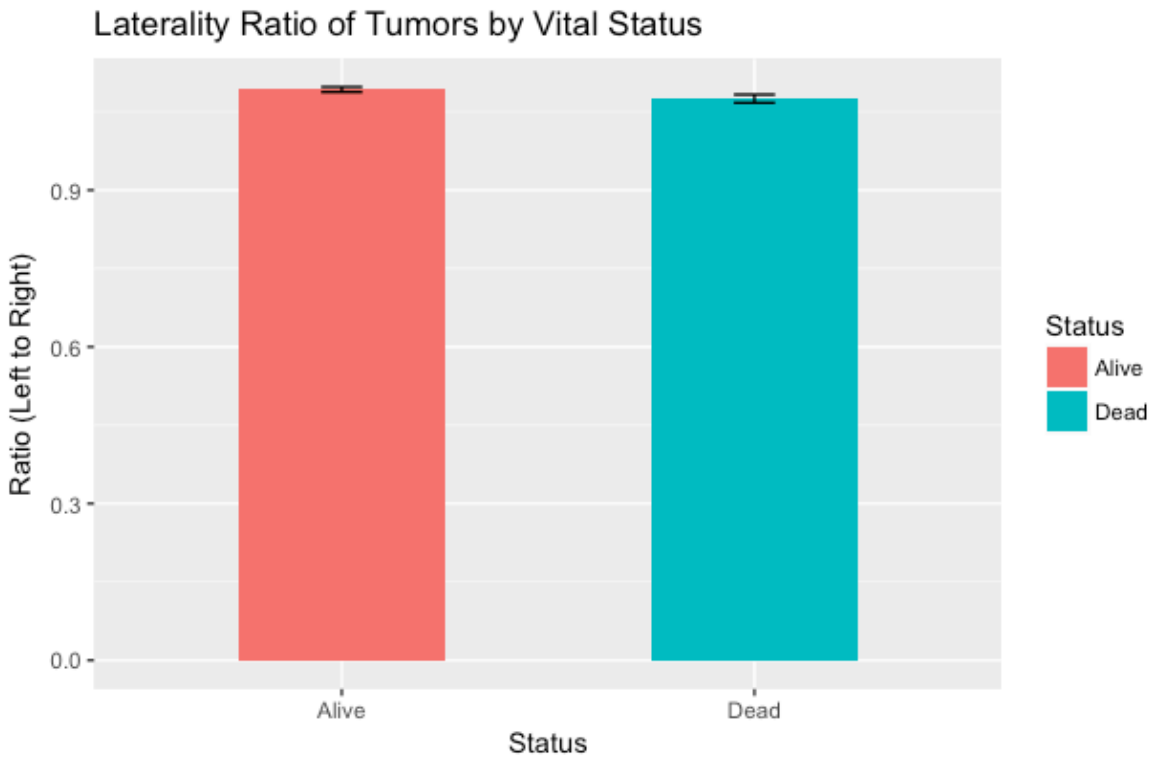

Figure 8: Left to right laterality ratio of melanomas in patients classified as either “alive” or “dead” as of their study follow up contact with error bars.

The left to right laterality ratio of melanomas for patients classified as “alive” is 1.093 with a standard error of 0.00489 and 1.075 with a standard error of 0.00785 for patients classified as “dead”. No significant difference was found in laterality ratios between these two groups.

3.3 Age at Diagnosis

To examine whether age has an effect on melanoma asymmetry, we used data about the patients’ age at diagnosis for skin cancer. Patients whose ages were unknown were excluded.

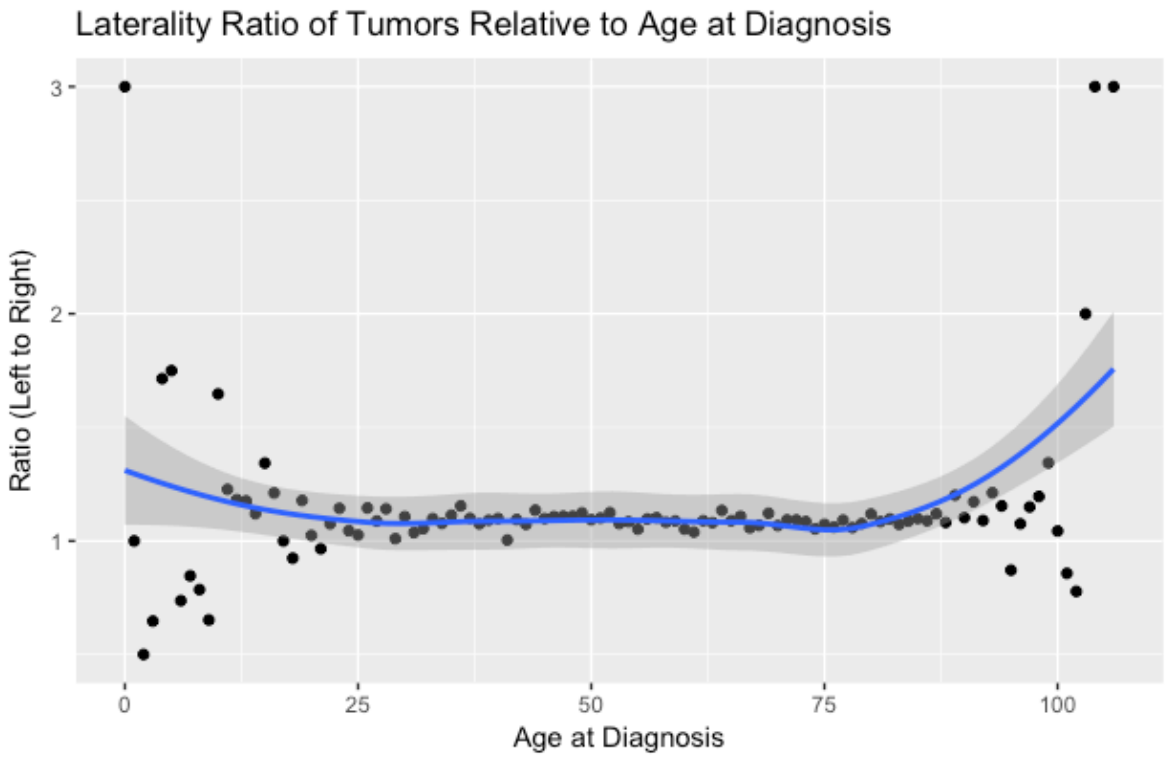

Figure 9: Left to right laterality ratio of tumors by patient age at diagnosis. The blue line within the shaded gray space represents the local polynomial regression (LOESS) fit of the data.

The mean left to right laterality ratio was 1.1485 with a standard deviation of 0.3667. Because the bulk of the patients enrolled in the SEER program were older patients (between approximately 50-80 years old), there was little data for very young patients and very old patients. This may explain the greater variance in data for the youngest and oldest patients in the study. A log transform was applied to better represent the variability of data.

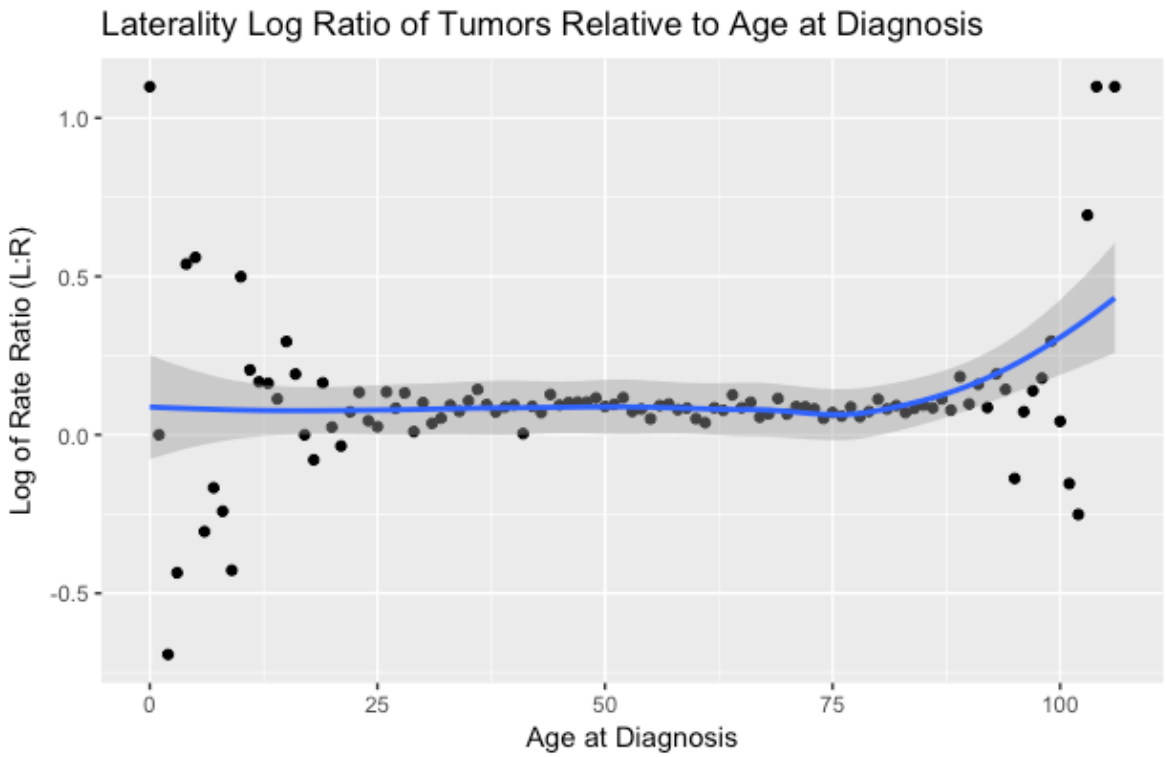

Figure 10: Log transform of left to right laterality ratio of melanomas by patient age at diagnosis.

A linear regression was run on the data to examine the relationship between age and laterality. No statistically significant relationship was found.

3.4 Tumor Size

To examine whether tumor growth and size might be affected by melanoma asymmetry, we looked at the relationship between tumor size and melanoma laterality.

The classifications for tumor size were based on a coding system applied to cases from 2004 onward and covered a total of 168,388 cases in our study. To see if there were any trends as tumor size increased or decreased, only tumors with a known size were included. This further narrowed the number of cases to 61,073. Size was given in millimeters for a range of sizes between 1 and 988 mm.

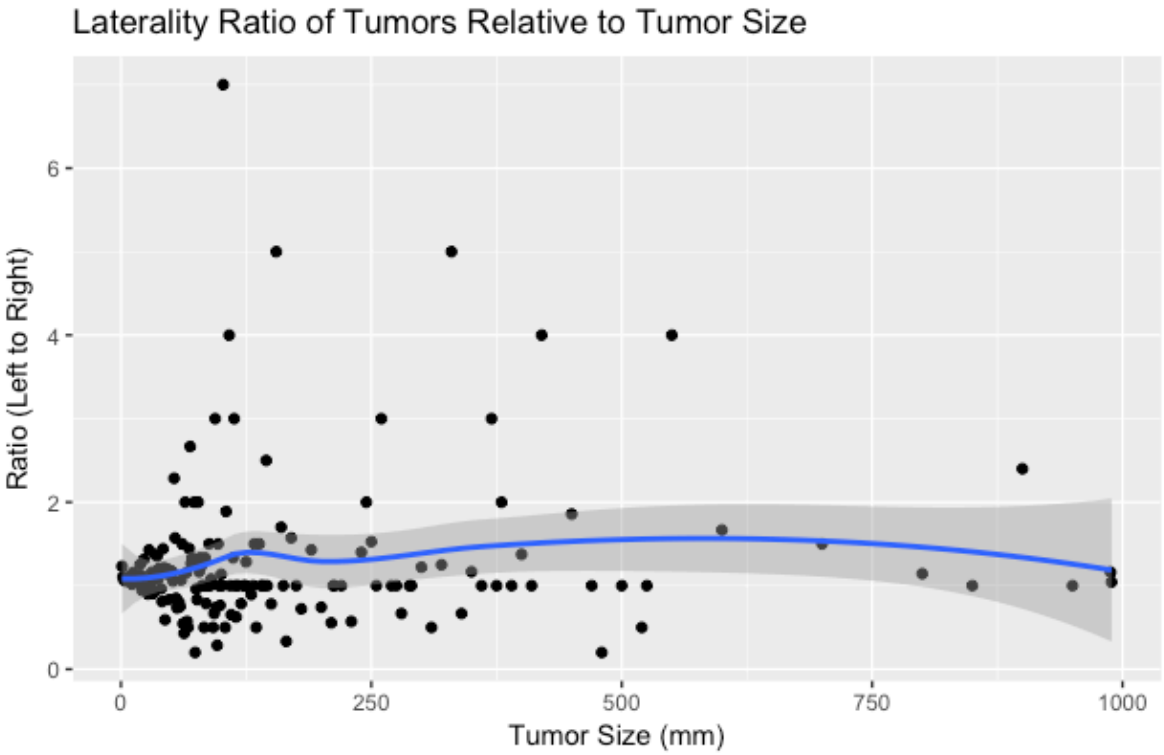

Figure 11: Left to right laterality ratio of melanomas relative to tumor size. All tumors that were coded with a known measurement for size are included with most of the data heavily skewed towards smaller measurements.

The mean left to right laterality ratio was 1.262 with a standard deviation of 0.8470. A linear regression of the data showed no statistically significant relationship existed between tumor size and laterality ratio.

The bulk of the cases captured were for patients with tumors smaller than 100 mm. To account for this skew, we looked at the 59,785 cases where tumors were 100 mm or less. This represented approximately 98% of the filtered cases.

The average left to right ratio of this filtered dataset was 1.112 with a standard deviation of 0.4152. A log transform of the data was performed to account for discrepancies in laterality ratios. A linear regression was performed and showed no statistically significant relationship between tumor size and laterality ratio.

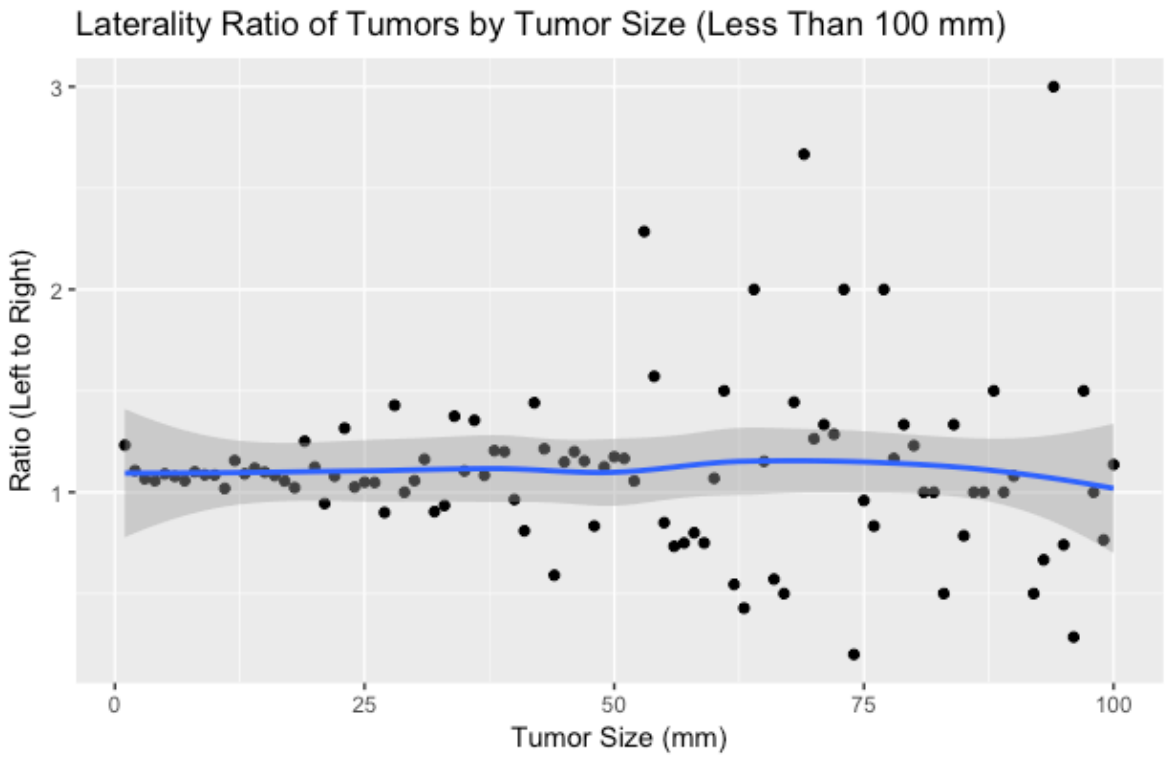

Figure 12: Left to right laterality ratio of melanoma cases relative to tumor size for all tumors 100 mm or less. This data captures 98% of all tumors measured in the SEER database and allows for a representation of data variation that excludes especially large patient tumors.

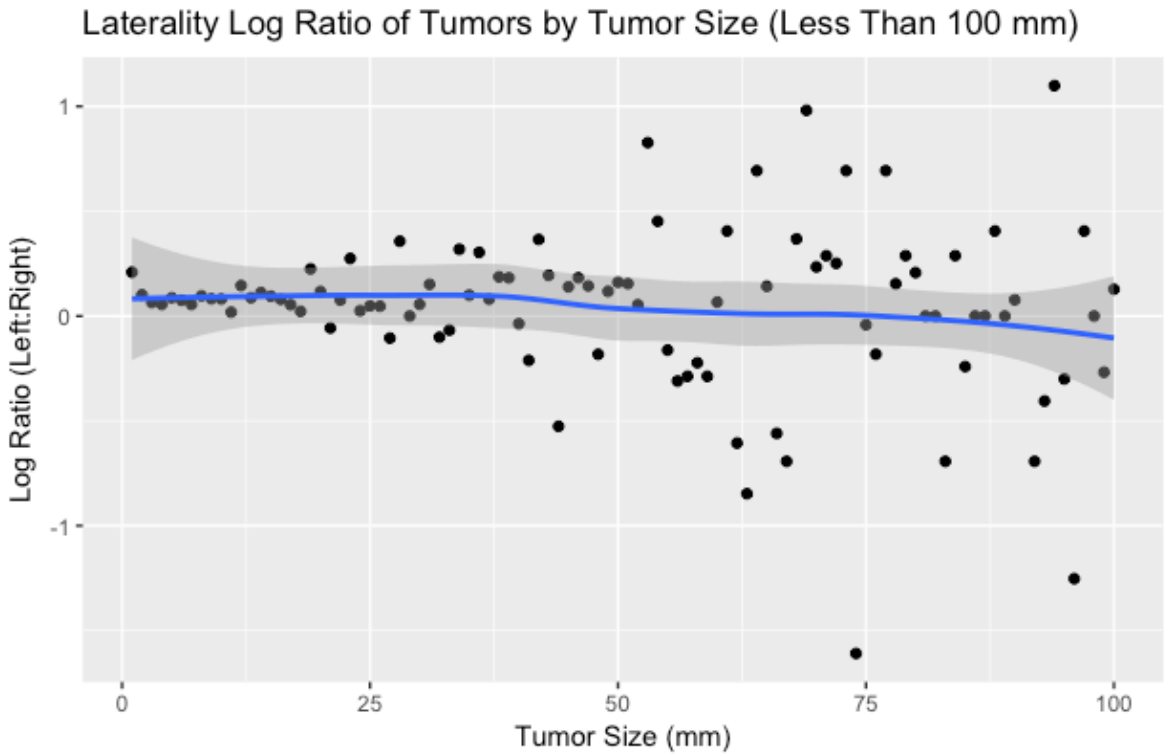

Figure 13: Log transform of the left to right laterality ratio of melanoma cases relative to tumor size for all tumors 100 mm or less

3.5 mRNA Gene Expression

To examine the genetic basis of asymmetric melanoma formation we first looked at the 10,237 cases available in the TCGA mRNA expression data table and used 470 skin cancer tumor cases to assess asymmetry in melanomas by gene expression. From these cases, 58,837 genes were identified as being expressed in melanomas. Measures of gene expression were taken from an HT-Seq raw read count and quantified as Fragments per Kilobase of transcript per Million mapped reads (FPKM) and FPKM-UQ (upper quartile normalization).

We used FPKM-UQ, which is the number of reads of mRNA divided by length of gene and total number of reads mapped to protein coding genes in the 75^{th} percentile, as a measure of gene expression for our analysis.

From the 58,837 genes identified, we obtained z-scores to determine if there were statistically significant differences in gene expression in melanomas on the right side of the body versus the left side of the body. Given the total number of genes, we determined that genes with a z-score of higher than 3 or lower than -3 could have statistically significant differences in gene expression based on laterality.

After filtering for $-3 > z\text{-scores} > 3$, we identified 51 genes of interest. These genes were characterized using the UnitProtKB search engine and from the original 51, only 22 were found in the database. The rest were assumed to be pseudogenes or genes that had not yet been characterized.

From this list of 22, 15 genes are implicated in post-translational modification and 18 genes have some form of coding sequence diversity. There were also 3 genes from the list that are

implicated in differentiation (LRRC34, PDLIM7, and DIAPH2). Ribosomal protein genes were conspicuous on the unfiltered list, but these genes were not well characterized.

When a Bonferroni adjustment was applied to the p-values, there was no statistically significant difference in laterality for any particular gene.

3.6 Protein Expression

To further examine the genetics of melanoma, we used the cases available in the TCGA Protein Expression data table, 470 skin cancer tumor cases were used to assess asymmetry in melanoma protein expression. After filtering out non-melanoma related cases, 187 proteins were found to be expressed in melanomas.

Z-scores for each protein were calculated to compare differences in protein expression on the right versus the left side of the body. Six proteins had a z-score greater than 2 or less than -2. Differences in protein expression were compared to differences in mRNA expression and three genes were found to have -1 > z-scores > 1 (PTEN, ATM, and ABL1).

Table 1: Comparison of gene expression laterality in protein and mRNA

Gene name	Preferentially expressed laterality		Chromosome	Function
	In protein	In mRNA		
BAD	Right	Left	11	Cell death
PTEN	Right	Right	10	Tumor suppression
ATM	Right	Right	11	DNA damage signaling
CLDN7	Right	Left	17	Cell adhesion
ABL1	Left	Left	9	Multiple cell processes
YAP1	Left	Left	11	Organ size control, tumor suppression, apoptosis

Note: The genes listed have the largest discrepancy in protein expression depending on laterality and all have a z-score of greater than 2 or less than -2.

When a Bonferroni adjustment was applied to the p-values, there was no statistically significant difference in laterality for any particular protein.

To further examine how proteins impact melanoma laterality, we used the predictive analytics and machine learning platform h2o to explore whether a protein or group of proteins might predict for melanoma asymmetry.

The protein dataset used for predictive modeling encompassed all protein expression cases categorized as skin cancer cases where tumor laterality was known. A training dataset was generated by randomly selecting 75% of this dataset and a validation dataset was created from the remaining 25% of this dataset.

Several machine learning algorithms were tested and the gradient boosting machine (GBM) algorithm produced the highest AUC of 0.59902 when using cross validation metrics that randomly cross-sampled the training and validation sets.

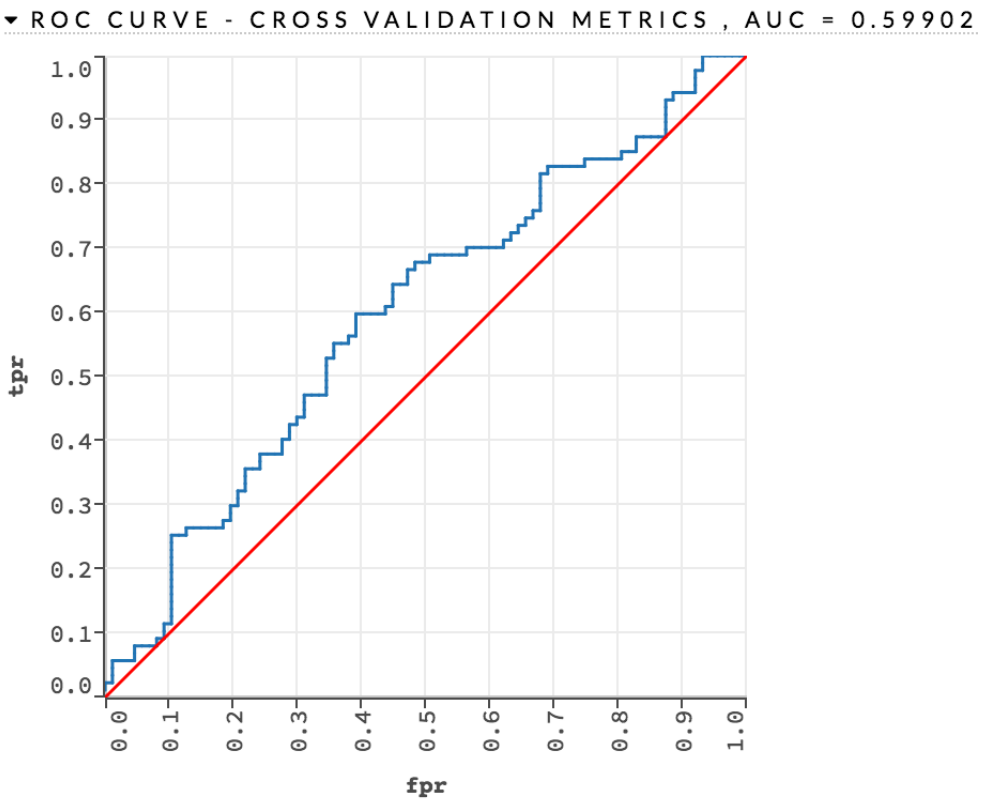

Figure 14: ROC curve produced from cross validation of protein expression training and test sets using a gradient boosting machine (GBM) algorithm. The blue line represents the false positive rate vs true positive rate of the GBM algorithm prediction for tumor laterality based on protein expression.

The GBM algorithm was also able to rank proteins in order of importance of predictive ability.

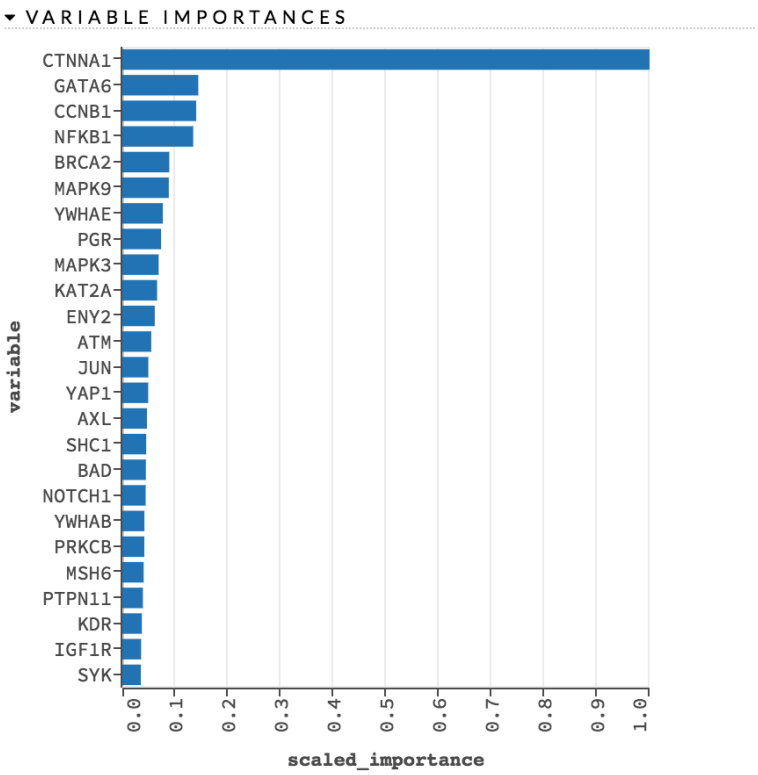

Figure 15: Scaled importance of genes by how well they predict for laterality of skin cancer tumors by protein expression. CTNNA1 is shown to be of critical importance when using the GBM algorithm.

We also tested other potentially important machine learning algorithms, i.e. deep learning and random distributed forest algorithms. Although those algorithms resulted in models with lower AUCs, they tended to rank CTNNA1 as the most important variable in predicting for tumor laterality.

3.7 miRNA Expression

We next looked a miRNA expression as a means of further investigating the effects of protein expression on melanoma asymmetry as miRNAs are known to inhibit protein expression [19]. From the cases available in the TCGA Protein Expression data table, 470 skin cancer tumor cases were used to assess asymmetry in melanoma miRNA expression. After filtering out non-melanoma related cases, 1,881 miRNAs were found to be expressed in melanomas.

Z-scores for each miRNA identified were calculated to compare differences in miRNA expression on the right versus the left side of the body. Three miRNAs had a z-score greater than 3 or less than -3.

The IDs for these miRNAs were searched on mirdb.org and gene targets for each miRNA were identified. We found three genes that were targets for more than one miRNA. The only gene target that was specifically targeted asymmetrically by miRNA was SIGLEC1, which codes for a cell adhesion protein, sialoadhesin. SIGLEC1 is targeted by both hsa-mir-151b and hsa-mir-6868 and both miRNAs had higher expression in right-sided melanomas.

Table 2: miRNA targets with significant differences in lateral expression

miRNA IDs	Gene target	miRNA laterality	Chromosome	Function
hsa-mir-151b hsa-mir-6838	SIGLEC1	Right	20	Cell adhesion
hsa-mir-4506 hsa-mir-6838	TXNRD1	Both	12	Redox homeostasis, cytoskeletal processes
hsa-mir-4506 hsa-mir-6838	C2orf43	Both	2	Lipid metabolism

Note: No genes were targeted by more than two miRNAs.

When a Bonferroni adjustment was applied to the p-values, there was no statistically significant difference in laterality for any particular miRNA.

3.8 Copy Number Variation

Next, we looked at the effects of genomic variation and mutations on melanoma asymmetry, starting with genomic repeats. From the cases available in the TCGA Copy Number Segment data table, 470 skin cancer tumor cases were used to assess asymmetry in copy number variation (CNV). Microarray data obtained using Affymetrix SNP 6.0 was used to identify repeated genomic regions and infer how many copies (copy number) of these repeats.

The Genomic Data Commons (GDC) further transformed copy number values into mean segment values so that diploid repeats have a mean value of 0, amplified regions have a positive value, and deletions will have negative values [5].

From the obtained data, we were able to determine the mean segment values for specific regions of the 22 somatic chromosomes and the X chromosome for tumors found on the right and left side of the body.

A z-score was calculated for the 85,529 regions where there were incidences of CNVs. From this initial dataset, we filtered out regions where the z-score was infinity or not able to be calculated. The majority of these filtered regions had only one case on one side of the body. After filtering the data, there were 278 regions with CNVs on both sides of the body.

A Bonferroni adjustment was applied to the p-value for these 278 regions and 60 regions had statistically significant differences in their mean segment values. However, most of these regions had relatively few cases with the region specific CNV. From these regions, chr7:53,250,694-53,251,092 had the largest amount of cases with four cases on the left side of the body and three on the right side of the body. When this region was visualized using the UCSC genome browser, no known genes were identified in this region.

The region with the highest amount of cases in general was chr22: 16934932-48940621. Because this region encompasses 32,005,690 basepairs, many genes, including genes implicated in cancer, were identified using the genome browser. A comparison of the mean segment value on the left and right side of the body for this region was performed and showed no significant changes. This was confirmed by the adjusted p-value and a KS test.

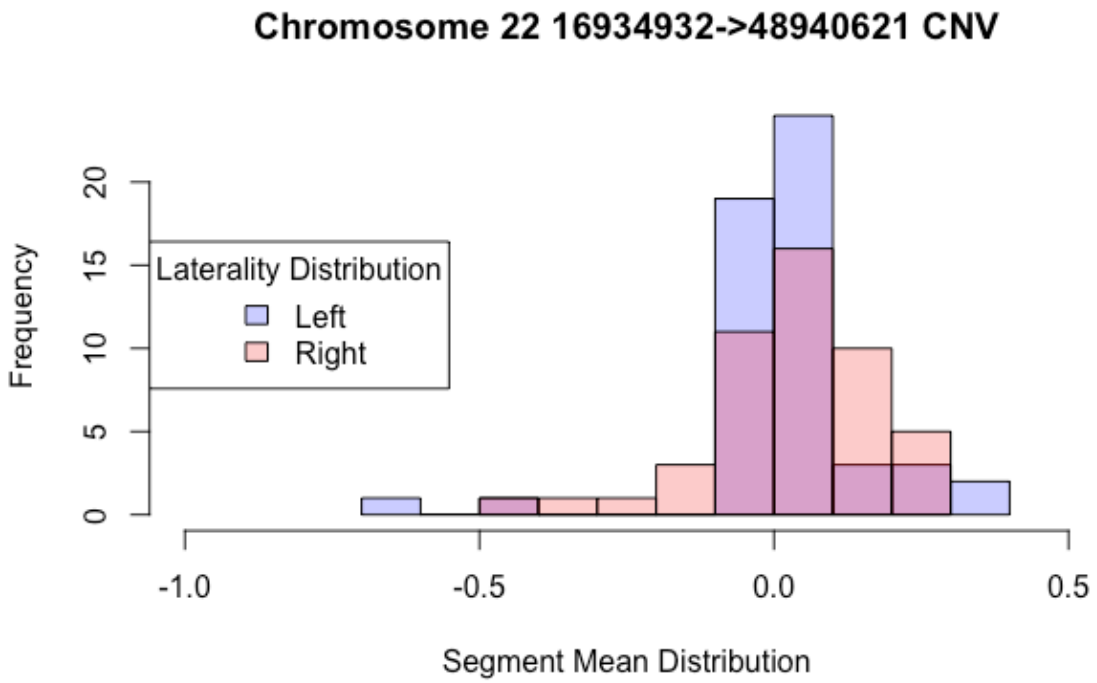

Figure 16: Comparison of the distribution of the mean segment value of copy number variation (CNV) on the left vs right side of the body for the most frequently affected region, Chromosome 22 position 16944932 – 48940621. Cases with positive segment mean values are amplified and cases with negative segment mean values are deleted. Genomic regions with the largest difference in mean segment value are often seen in only a few cases.

This comparison did show that this region tended to have a slightly positive segment mean, implying that this region is repeated in melanomas in general.

3.9 Somatic Mutations

To further explore genomic changes, we looked at the somatic mutations dataset. From the 470 skin cancer cases we used to assess asymmetry in melanomas, we found 9,805 recorded incidences of somatic mutations within these cases.

After running a rate ratio test and obtaining p-values to assess significant differences in expression of these mutations on the right versus left side of the body. We found one mutation with a statistically significant different number of cases between the left and right side of the body. This mutation was found on chromosome 9 at position 12793787 and involved a single

nucleotide polymorphism (SNP) where a cytosine was replaced with a thymine. This affects the gene DPM2, which is involved in protein modification. This SNP occurred 12 times on the left side of the body and 2 times on the right side of the body.

Next, we looked at the mutations that occurred the most frequently. The most frequent mutation was a SNP where adenine is replaced by a thymine at position 140753336 of chromosome 7. There are 73 cases on the left side of the body with this mutation and 71 cases on the right side of the body with this mutation. This affects the BRAF gene, which is involved in the MAPK signaling pathway.

The next most frequent occurrence of a mutation was a SNP where thymine was replaced with cytosine on chromosome 1, position 114713908. This occurred in 24 cases on the left side of the body and 18 cases on the right side of the body. This mutation affects the NRAS gene, which is implicated in tumor formation.

3.10 DNA Methylation

The last means of genomic variation we looked at was DNA methylation. DNA methylation is known to modify gene expression, typically repressing or silencing gene transcription [10]. DNA methylation has also been of interest in cancer immunotherapy [11].

DNA methylation data was obtained by TCGA using the Illumina Human Methylation 27k and 450k platforms and beta values were calculated as the ratio between the methylated intensity and total array intensity. For the purposes of this study, this beta value was used as an approximation of the percent methylation of the site corresponding to a given probe.

From the 470 skin cancer cases we used to assess asymmetry in melanomas, 396,065 probed sites were found to show some percentage of methylation on at least one side of the

body. A z-score was calculated for these probes. Five probed sites were found to have a z-score higher than 4 or less than -4. However, after performing a Bonferroni adjustment, no probed sites were shown to have statistically significant differences in methylation on the left vs right side of the body.

The five sites with the highest and lowest z-scores were searched in the isb-cgc platform reference dataset, specifically in the GDC_hg38_methylation_annotation dataset, to obtain information about the genes and chromosomes associated with their probe sites. This data was searched in UniProtKB to obtain more information about these genes.

Table 3: Gene targets with significant differences in methylation laterality

Probe ID	Gene	Methylation Laterality	Chromosome	Function
cg01103812	ITPR2	Left	12	Calcium mediated signaling
cg02204630	NMUR1	Left	2	GPCR and calcium mediated signaling
cg13978325	RXRA	Right	9	Transcription regulator
cg22812178	ZFYVE26	Right	14	Cytokinesis, DNA damage repair
cg25828445	NA	Left	12	NA

Note: Probe cg25828445 did not target an area where any specific gene could be identified. Probe IDs were matched to gene names using the TCGA platform reference dataset and protein names and functions were found using UniProtKB.

Chapter 4: Discussion

4.1 Epidemiological Factors Affecting Melanoma Asymmetry

Our study found that overall, across all epidemiological and clinical variables observed, there is an increased incidence of melanomas on the left side of the body. This is consistent with past studies that have reported the same [2], [21], [22].

When looking at sex differences, there was no statistically significant difference in the left to right laterality ratio of melanomas in men and women (Figure 1). Prior studies have reported mixed results about how melanoma asymmetry and sex are related, with some studies suggesting that men have a higher incidence of left-sided melanomas based on driving patterns [20], [24]. Studies that examined the effects of sex as well as other factors support evidence that there is no significant difference in melanoma laterality based on sex [4].

Although the data we used was derived from patients in the United States, the observation that there are generally more left-sided melanomas holds true in other countries as well [21].

One of the long-held explanations for this left-sided bias has been that there is more UV exposure on the left side of the body, particularly when driving. However, studies from countries where drivers would typically drive on the right side of the road, also show a left-sided bias [21].

To further investigate the potential environmental effects of UV exposure, we looked at tumor primary sites. If UV exposure while driving was a significant contributor to the left-sided asymmetry of melanoma, we'd expect to see a significant increase in the left-right laterality ratio in areas of the body exposed to UV light when driving i.e. upper limbs, arms, or face. Although there is some evidence for higher laterality ratios in this area [20], [31] we only found

statistically significant differences in laterality ratios in sites classified as “Not Otherwise Specified” (Figures 2 and 3).

This suggests that UV exposure while driving plays a relatively small part in the left-sided asymmetry of melanoma. To further examine how UV exposure might affect melanoma laterality, we looked at the locations in the US where patients in the SEER study were admitted. Although the 13 states where patients were enrolled did cover environmental differences in UV exposure, we only found that the 36 cases in Alaska had a statistically significant difference in laterality ratio (Figures 4 and 5). Patients enrolled in Alaska tended to have more melanomas on the right side of the body, but the standard error of comparison was high and the sample size was small.

Other studies of laterality by patient location have suggested that there is a consistent bias towards left-sided melanomas across multiple countries in different hemispheres [21]. It appears that although UV exposure is a major contributing factor to melanoma development in general, patterns of exposure in different regions do not necessarily explain the predominance of left-sided melanomas.

We also looked at survivability (Figures 6, 7, and 8), age at diagnosis (Figures 9 and 10), and tumor size (Figures 11, 12, and 13) as potential indicators of melanoma laterality, but found no statistically significant results. Data on age and tumor size are mixed depending on where the data was obtained [20], [21] and it should be noted that ranges approaching significance were typically for smaller populations. This lack of consistency indicates that although these factors may play some role in melanoma laterality, there is no conclusive evidence that they are predictors or indicators of melanoma asymmetry.

4.2 Genetic Determinants of Melanoma Asymmetry

Knowing that epidemiological and environmental factors are not conclusive indicators of melanoma laterality, we next examined potential genetic factors that might affect asymmetric melanoma development. With the advent of next generation sequencing (NGS), cancer genomics has become a viable means of discovering the underlying mechanisms of tumor formation, including the potential to find the underlying cause of tumor asymmetry, and a potential means of discovering new treatments [6], [18].

We first examined mRNA and protein expression as a means of gauging what genes and proteins might be indicators of melanoma asymmetry. After conducting a Bonferroni adjustment, there were no genes and proteins that played a statistically significant role in melanoma asymmetry. However, we were able to compare mRNA expression to protein expression to gauge differences in how mRNA was translated in proteins with the largest differences in expression in left versus right-sided melanomas.

This comparison showed that BAD and CLDN7 did not have the same expression pattern in melanoma laterality (Table 1) and might be potential translational targets. However, research has shown that BAD is unlikely to be a key factor in promoting apoptosis of cancerous cells [3] and has similar levels of expression in cancerous melanocytes and normal melanocytes [15]. CLDN7 does play some role in the survival of melanoma cells [16]. Furthermore, CLDN7 is part of the Claudin family, which has been speculated to have some effect on left-right patterning in xenopus models [8].

When we used the protein dataset to model which proteins might predict for tumor laterality, CTNNA1 consistently appeared as the most important protein in predicting laterality.

CTNNA1 codes for catenin alpha-1, a cadherin-associate protein, which plays a role in cell adhesion, cell-cell interaction, and morphogenesis [33]. Cadherin expression plays a role in malignant tumors [33] and CTNNA1 expression tends to be lower in leukemia-initiation stem cells [7] and is mutated in colon cancer cells [38].

Cadherins and catenins are also involved in melanocyte development and transformation and their interactions may play some role in melanoma formation [27]. Furthermore, CTNNB1, which is in the same family of catenins as CTNNA1, has been identified as a protein that is frequently mutated in metastatic melanoma [26]. These studies in combination with our observation that CTNNA1 may be involved in melanoma laterality may prove to be good groundwork for further research into the role of cadherin-catenin interactions on tumor laterality.

To further investigate how protein expression might affect melanoma laterality, we examined miRNA expression as a means of gauging what genes might be target for miRNA silencing and thus inhibit protein expression. Although only three miRNAs were identified as having significant differences in expression on the left vs right side of the body, each miRNA targeted many genes. Of the targeted genes, three were found to be targeted by more than one miRNA and only one gene, SIGLEC1, was targeted specifically on the right side of the body (Table 2).

SIGLEC1 codes for sialoadhesin, also known as CD169, which is part of a major class of mammalian glycan-binding proteins and is known to be expressed by macrophages [9]. Typically, SIGLEC1 functions in the processes of cell-cell adhesion and cell-pathogen interactions. However, SIGLEC1 is also known to be expressed in tumor-associated

macrophages and has been implicated in the initiation, development, and metastasis of many tumors [32]. In melanoma, however, although it is clear that SIGLEC1 contributes to the immune response of cancer, there is no clear connection between SIGLEC1 expression and melanoma progression [23]. Information about the role SIGLEC1 plays in right-left patterning or its effects on melanoma laterality is largely unknown, but may present an opportunity for further research.

We next looked at how mutations might affect, melanoma asymmetry, starting with genomic repeats. After a Bonferroni adjustment, we identified 60 regions that had statistically significant differences in segment copy numbers. Most of these 60 regions had a relatively small amount of total cases where the segments were identified as having varying copy numbers. The segment with the largest total amount of cases was on chromosome 7, position 53,250,694-53,251,092. This region tended to be deleted in cases on both the right and left side of the body, but was shown to have significantly more deletions on the right side of the body. The region is relatively small and no known genes were identified in this region. Changes in chromosome 7 are known to play role in cancer [34], but more research is needed to determine if there is a link between any cancer-causing factors and where tumors present in the body.

The genomic region that had the most cases in which there was some variation in the region was on chromosome 22, position 16934932-48940621. Cases on both the right and left side of the body showed amplification of this region (Figure 15). The region is relatively large, so many genes were identified within it. Although chromosome 22 has previously been implicated in cancer, how mutations in this gene melanoma is not well characterized [1].

We further examined the role of mutations in melanoma asymmetry by looking at specific somatic mutations and were able to identify a SNP that had a statistically significant difference in incidence of left versus right-sided melanomas. The SNP affects the DPM2 gene and was found to be more prevalent in left-sided melanomas.

DPM2 is a subunit of DPM, dolichol-phosphate mannose synthase, and has been implicated in disorders of glycosylation such as muscular dystrophy [13], glaucoma [17], and prostate cancer [12]. However, not much is known about its role in melanoma or how it might affect tumor laterality. This presents a potential avenue for further research based on our finding that it may play some role in melanoma laterality.

The most common somatic mutations were found to be in the BRAF and NRAS gene sequences, which are already known to play a role in melanoma formation [14]. We found no statistically significant differences in the occurrence of these mutations on the left versus right side, but there is potential to investigate how they might interact with each other and PTEN to affect melanoma laterality based on prior research [14] and our finding that PTEN protein expression is much higher in right-sided melanomas.

The last potential avenue of genomic modification we looked at was DNA methylation, which tends to downregulate or silence transcription. Although there was no statistical significance in difference of number of probes found on the left and right-sided tumors, two genes of interest were identified.

A probe that was preferentially found in left-sided tumors was found to affect ITPR2, which plays a role in a variety of functions including tumor-suppression [37]. We also found a probe preferentially found in right-sided tumors that affects the RXRA gene, which codes for a retinoic

acid receptor (Table 3). Retinoic acid has been found to play some role in left-right patterning [29] and may affect melanoma differentiation [28]. The effect of RXRA on melanoma laterality is not well researched, however, and is a potential avenue of continued research.

4.3 Conclusions

Potential epidemiological causes of melanoma asymmetry have been discussed in the past with mixed results. Our study found that when comparing one variable at a time, there is no conclusive evidence that any single factor is a major determinant of melanoma laterality. All that can be said is that across these multiple clinical and epidemiological variables, there is a significant preference for melanoma development on the left side of the body. Furthermore, patterns of UV exposure associated with driving are not good predictors of melanoma development. It is likely that there are genetic factors at play—which may or may not interact with epidemiological variables—that are better indicators of what might cause melanomas to develop on one side of the body versus the other.

However, although many genes and chromosomal regions of interest were identified, their impact on melanoma laterality is still largely unknown. Many genes of interest are already known to play some role in tumor suppression or formation or are known to regulate transcription of such proteins. One of the more surprising observations was that cell-adhesion and structural proteins may also influence laterality and may have some relation to left-right patterning at development. The question of why these genes seem to influence or be influenced by the location of melanoma is one that remains unanswered and it is more likely than not that multiple genes are at play in the determination of where a melanoma presents. As genomics research progresses and analytical methods become more powerful, we will have

more clinical information and more access to tools for analysis. It is possible that in the future we will be able to do more with more information and use that to start making inferences about how human genetics affect melanoma and other tumor tissue laterality.

References

[1] P. G. Buckley *et al.*, “A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications,” *Hum Mol Genet*, vol. 11, no. 25, pp. 3221–3229, Dec. 2002.

[2] M. Gorman, A. Hart, and B. Mathew, “A Left-Sided Prevalence of Lentigo Maligna: A UK Based Observational Study and Review of the Evidence,” *Dermatology Research and Practice*, 2015. [Online]. Available: https://www.hindawi.com/journals/drp/2015/310270/. [Accessed: 15-Jan-2018].

[3] Y. F. Wang, C. C. Jiang, K. A. Kiejda, S. Gillespie, X. D. Zhang, and P. Hersey, “Apoptosis Induction in Human Melanoma Cells by Inhibition of MEK Is Caspase-Independent and Mediated by the Bcl-2 Family Members PUMA, Bim, and Mcl-1,” *Clin Cancer Res*, vol. 13, no. 16, pp. 4934–4942, Aug. 2007.

[4] K. G. Paulson, J. G. Iyer, and P. Nghiem, “Asymmetric lateral distribution of melanoma and Merkel cell carcinoma in the United States,” *Journal of the American Academy of Dermatology*, vol. 65, no. 1, pp. 35–39, Jul. 2011.

[5] “Bioinformatics Pipeline: Copy Number Variation Analysis - GDC Docs.” [Online]. Available: https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/. [Accessed: 01-Mar-2018].

[6] L. Chin, J. N. Andersen, and P. A. Futreal, “Cancer genomics: from discovery science to personalized medicine,” *Nature Medicine*, vol. 17, no. 3, pp. 297–303, Mar. 2011.

[7] T. X. Liu *et al.*, “Chromosome 5q deletion and epigenetic suppression of the gene encoding α-catenin (*CTNNA1*) in myeloid cell transformation,” *Nature Medicine*, vol. 13, no. 1, pp. 78–83, Jan. 2007.

[8] I. Gupta and A. Ryan, “Claudins: unlocking the code to tight junction function during embryogenesis and in disease,” *Clinical Genetics*, vol. 77, no. 4, pp. 314–325, Apr. 2010.

[9] Y. Saito *et al.*, “Clinicopathological and prognostic significance of CD169+ lymph node sinus macrophages in patients with malignant melanoma,” *Cancer Immunol Res*, p. canimm.0180.2014, Aug. 2015.

[10] B. Jin, Y. Li, and K. D. Robertson, “DNA Methylation,” *Genes Cancer*, vol. 2, no. 6, pp. 607–617, Jun. 2011.

[11] “DNA Methylation Inhibitor Triggers Immune Response in Cancer,” *The Cancer Genome Atlas - National Cancer Institute*. [Online]. Available: https://cancergenome.nih.gov/researchhighlights/researchbriefs/antivirus. [Accessed: 01-Mar-2018].

[12] S. Kim *et al.*, “Dolichol phosphate mannose synthase (*DPM1*) mutations define congenital disorder of glycosylation Ie (CDG-Ie),” *J Clin Invest*, vol. 105, no. 2, pp. 191–198, Jan. 2000.

[13] R. Barone *et al.*, “DPM2-CDG: a muscular dystrophy-dystroglycanopathy syndrome with severe epilepsy,” *Ann. Neurol.*, vol. 72, no. 4, pp. 550–558, Oct. 2012.

[14] V. K. Goel, A. J. F. Lazar, C. L. Warneke, M. S. Redston, and F. G. Haluska, “Examination of Mutations in BRAF, NRAS, and PTEN in Primary Cutaneous Melanoma,” *Journal of Investigative Dermatology*, vol. 126, no. 1, pp. 154–160, Jan. 2006.

[15] E. Selzer, H. Schlagbauer-Wadl, I. Okamoto, H. Pehamberger, R. Pötter, and B. Jansen, "Expression of Bcl-2 family members in human melanocytes, in melanoma metastases and in melanoma cell lines.," *Melanoma Res*, vol. 8, no. 3, pp. 197–203, Jun. 1998.

[16] A. Kaushik, Y. Bhatia, S. Ali, and D. Gupta, "Gene Network Rewiring to Study Melanoma Stage Progression and Elements Essential for Driving Melanoma," *PLOS ONE*, vol. 10, no. 11, p. e0142443, Nov. 2015.

[17] J. L. Wiggs and L. R. Pasquale, "Genetics of glaucoma," *Hum. Mol. Genet.*, vol. 26, no. R1, pp. R21–R27, 01 2017.

[18] K. S. M. Smalley, K. L. Nathanson, and K. T. Flaherty, "Genetic Subgrouping of Melanoma Reveals New Opportunities for Targeted Therapy," *Cancer Res*, vol. 69, no. 8, pp. 3241–3244, Apr. 2009.

[19] J. Seo, D. Jin, C.-H. Choi, and H. Lee, "Integration of MicroRNA, mRNA, and Protein Expression Data for the Identification of Cancer-Related MicroRNAs," *PLOS ONE*, vol. 12, no. 1, p. e0168412, Jan. 2017.

[20] J.-L. Bulliard, S. Ess, A. Bordoni, I. Konzelmann, and F. Levi, "Left-Sided Excess in the Laterality of Cutaneous Melanoma," *Arch Dermatol*, vol. 144, no. 4, pp. 556–558, Apr. 2008.

[21] D. H. Brewster, M.-J. D. Horner, S. Rowan, P. Jelfs, E. de Vries, and E. Pukkala, "Left-sided excess of invasive cutaneous melanoma in six countries," *European Journal of Cancer*, vol. 43, no. 18, pp. 2634–2637, Dec. 2007.

[22] T. Gambichler *et al.*, “Left-sided laterality of Merkel cell carcinoma in a German population: more than just sun exposure,” *J Cancer Res Clin Oncol*, vol. 143, no. 2, pp. 347–350, Feb. 2017.

[23] M.-R. Muhsin-Sharafaldine, S. C. Saunderson, A. C. Dunn, and A. D. McLellan, “Melanoma growth and lymph node metastasis is independent of host CD169 expression,” *Biochemical and Biophysical Research Communications*, vol. 486, no. 4, pp. 965–970, May 2017.

[24] G. M. Dores, M. M. Huycke, and S. S. Devesa, “Melanoma of the skin and laterality,” *Journal of the American Academy of Dermatology*, vol. 64, no. 1, pp. 193–195, Jan. 2011.

[25] “Months Survived Based on Complete Dates - SEER.” [Online]. Available: https://seer.cancer.gov/survivaltime/. [Accessed: 01-Mar-2018].

[26] B.-H. Neves de Oliveira, C. Dalmaz, and F. Zeidán-Chuliá, “Network-Based Identification of Altered Stem Cell Pluripotency and Calcium Signaling Pathways in Metastatic Melanoma,” *Med Sci (Basel)*, vol. 6, no. 1, Mar. 2018.

[27] A. Jouneau, Y.-Q. Yu, M. Pasdar, and L. Larue, “Plasticity of Cadherin–Catenin Expression in the Melanocyte Lineage,” *Pigment Cell Research*, vol. 13, no. 4, pp. 260–272, Aug. 2000.

[28] K. Demary, L. Wong, J. S. Liou, D. V. Faller, and R. A. Spanjaard, “Redox Control of Retinoic Acid Receptor Activity: A Novel Mechanism for Retinoic Acid Resistance in Melanoma Cells,” *Endocrinology*, vol. 142, no. 6, pp. 2600–2605, Jun. 2001.

[29] S. Wasiak and D. Lohnes, “Retinoic Acid Affects Left–Right Patterning,” *Developmental Biology*, vol. 215, no. 2, pp. 332–342, Nov. 1999.

[30] “SEER Data, 1973-2014.” [Online]. Available: https://seer.cancer.gov/data/. [Accessed: 15-Jan-2018].

[31] C. Lesage, C. Barbe, A. Le Clainche, F.-X. Lesage, P. Bernard, and F. Grange, “Sex-Related Location of Head and Neck Melanoma Strongly Argues for a Major Role of Sun Exposure in Cars and Photoprotection by Hair,” *Journal of Investigative Dermatology*, vol. 133, no. 5, pp. 1205–1211, May 2013.

[32] B. Mohan and P. R. Crocker, “Sialylation and Immune Surveillance of Cancer by Siglecs,” in *Glycosignals in Cancer: Mechanisms of Malignant Phenotypes*, K. Furukawa and M. Fukuda, Eds. Tokyo: Springer Japan, 2016, pp. 125–138.

[33] Y. Furukawa *et al.*, “Structure, expression and chromosome assignment of the human catenin (cadherin-associated protein) alpha 1 gene (CTNNA1),” *CGR*, vol. 65, no. 1–2, pp. 74–78, 1994.

[34] L. W. Hillier *et al.*, “The DNA sequence of human chromosome 7,” *Nature*, vol. 424, no. 6945, pp. 157–164, Jul. 2003.

[35] B. K. Armstrong and A. Kricker, “The epidemiology of UV induced skin cancer,” *Journal of Photochemistry and Photobiology B: Biology*, vol. 63, no. 1, pp. 8–18, Oct. 2001.

[36] “The type 2 inositol 1,4,5-trisphosphate receptor, emerging functions for an intriguing Ca2+-release channel.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4465056/. [Accessed: 27-Mar-2018].

[37] Vervloessem, “The type 2 inositol 1,4,5-trisphosphate receptor, emerging functions for an intriguing Ca2 +-release channel - ScienceDirect.” [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S0167488914004340?via%3Dihub. [Accessed: 27-Mar-2018].

[38] S. J. Vermeulen *et al.*, "The αE-catenin gene (***CTNNA1***) acts as an invasion-suppressor gene in human colon cancer cells," *Oncogene*, vol. 18, no. 4, pp. 905–915, Jan. 1999.

[39] F. Liu-Smith, J. Jia, and Y. Zheng, "UV-Induced Molecular Signaling Differences in Melanoma and Non-melanoma Skin Cancer," in *Ultraviolet Light in Human Health, Diseases and Environment*, Springer, Cham, 2017, pp. 27–40.

[40] "Welcome to H2O 3 — H2O 3.18.0.5 documentation." [Online]. Available: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/welcome.html. [Accessed: 10-Apr-2018].

[41] "What is Cancer Genomics?," *The Cancer Genome Atlas - National Cancer Institute*. [Online]. Available: https://cancergenome.nih.gov/cancergenomics/whatisgenomics/whatis. [Accessed: 15-Jan-2018].