236523 - Introduction to Bioinformatics - Spring 21 – HW4

Alon Hacohen | 311587653 & Roy Bernea | 316315829

01/07/2021

Question 1

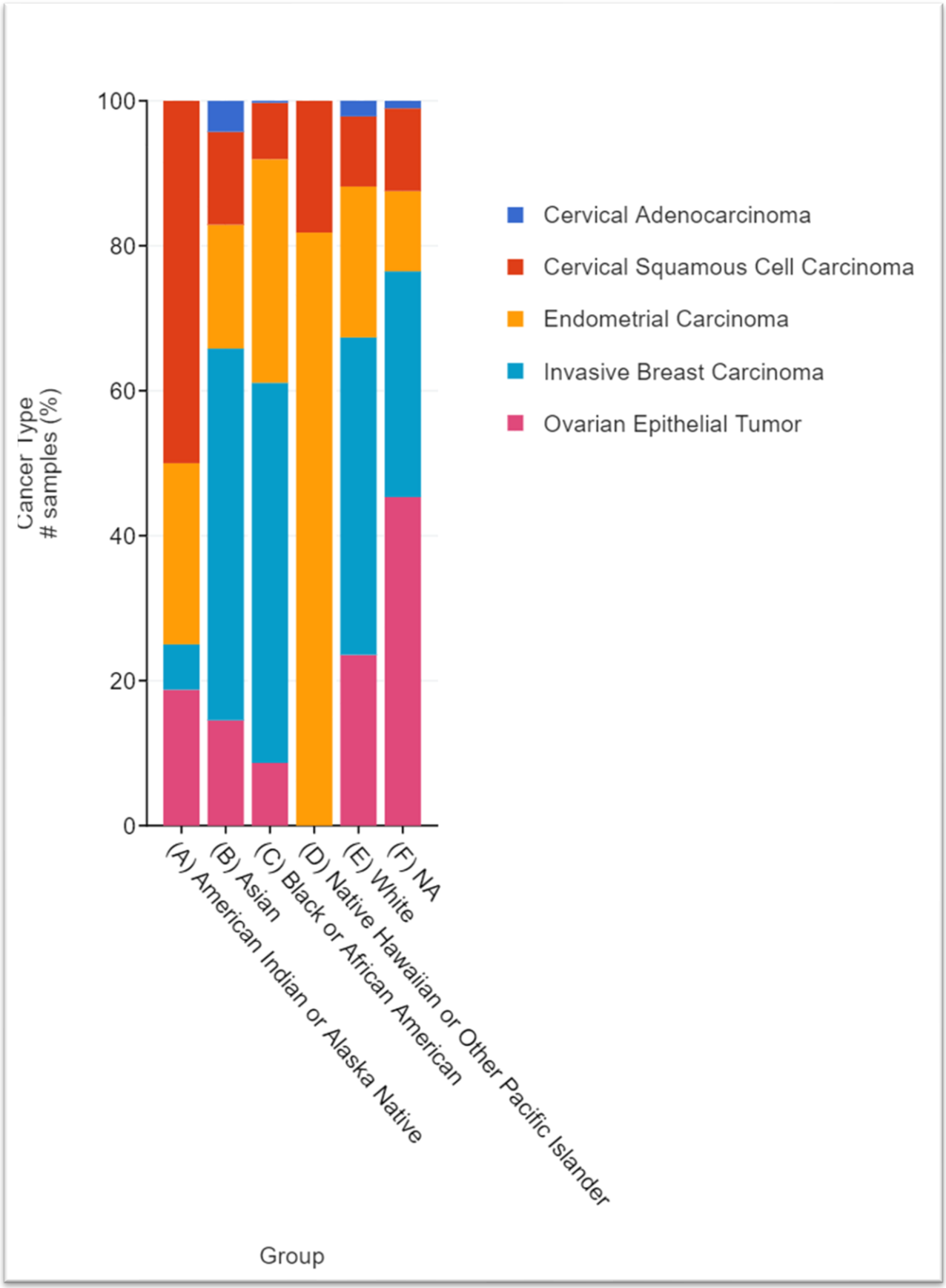
Q1.1

Comparing the four cancers by patients' race, and modifying the comparison accordingly to the question, we could have claimed that the Native Hawaiian or Other Pacific Islander race group is predisposed to a specific type of cancer – which is ***Endometrial cancer***.
But, and that is a major disclaimer – the sample size is very small. As we will show in a few rows, the dataset contains only 11 patients of such race, so we would not feel comfortable using the results for such claim.

From our comparison analysis, we got the following table:

Patient ID	Sample ID	Cancer Type	Mutation Count	Fraction Genome Altered	Race Category
TCGA-AJ-A3EL	TCGA-AJ-A3EL-01	Endometrial Carcinoma	7393	0.0008	Native Hawaiian or Other Pacific Islander
TCGA-AX-A3G4	TCGA-AX-A3G4-01	Endometrial Carcinoma	36	0.4534	Native Hawaiian or Other Pacific Islander
TCGA-BS-A0T9	TCGA-BS-A0T9-01	Endometrial Carcinoma	58	0.0504	Native Hawaiian or Other Pacific Islander
TCGA-BS-A0TA	TCGA-BS-A0TA-01	Endometrial Carcinoma	472	0.0005	Native Hawaiian or Other Pacific Islander
TCGA-BS-A0U9	TCGA-BS-A0U9-01	Endometrial Carcinoma	36	0.6511	Native Hawaiian or Other Pacific Islander
TCGA-BS-A0UL	TCGA-BS-A0UL-01	Endometrial Carcinoma	388	0.0001	Native Hawaiian or Other Pacific Islander
TCGA-BS-A0UV	TCGA-BS-A0UV-01	Endometrial Carcinoma	8975	0.1249	Native Hawaiian or Other Pacific Islander
TCGA-BS-A0V7	TCGA-BS-A0V7-01	Endometrial Carcinoma	59	0.0001	Native Hawaiian or Other Pacific Islander
TCGA-BS-A0V8	TCGA-BS-A0V8-01	Endometrial Carcinoma	76	0.5415	Native Hawaiian or Other Pacific Islander
TCGA-EK-A3GJ	TCGA-EK-A3GJ-01	Cervical Squamous Cell Carcinoma	201	0.2528	Native Hawaiian or Other Pacific Islander
TCGA-ZJ-A8QR	TCGA-ZJ-A8QR-01	Cervical Squamous Cell Carcinoma	123	0.2595	Native Hawaiian or Other Pacific Islander

In addition, we got the following 100% stacked barplot generated from this analysis:


Cervical Adenocarcinoma
Cervical Squamous Cell Carcinoma
Endometrial Carcinoma
Invasive Breast Carcinoma
Ovarian Epithelial Tumor
Cancer Type
samples (%)
100
80
60
40
20
0
(A) American Indian or Alaska Native
(B) Asian
(C) Black or African American
(D) Native Hawaiian or Other Pacific Islander
(E) White
(F) NA
Group

Looking at these results (the table and the plot), we may notice that the distribution for different cancer types among the *Native Hawaiian or Other Pacific Islander* race group differs greatly from its equivalent among other race group.
Even more than that – we may notice that over 80% of the patients in this race group suffer from Endometrial Carcinoma, which might lead us to determine that the race group seems to be predisposed to this specific type of cancer.
Once again, it is important to add that this is **not conclusive**. In order to claim such thing with higher confidence, we could have used a larger dataset or investigate other factors that could have been the cause for that correspondence.

The bottom line is that we **will not claim** that the Native Hawaiian or Other Pacific Islander race group is predisposed to a specific type of cancer, as we need more data to determine.

Q1.2

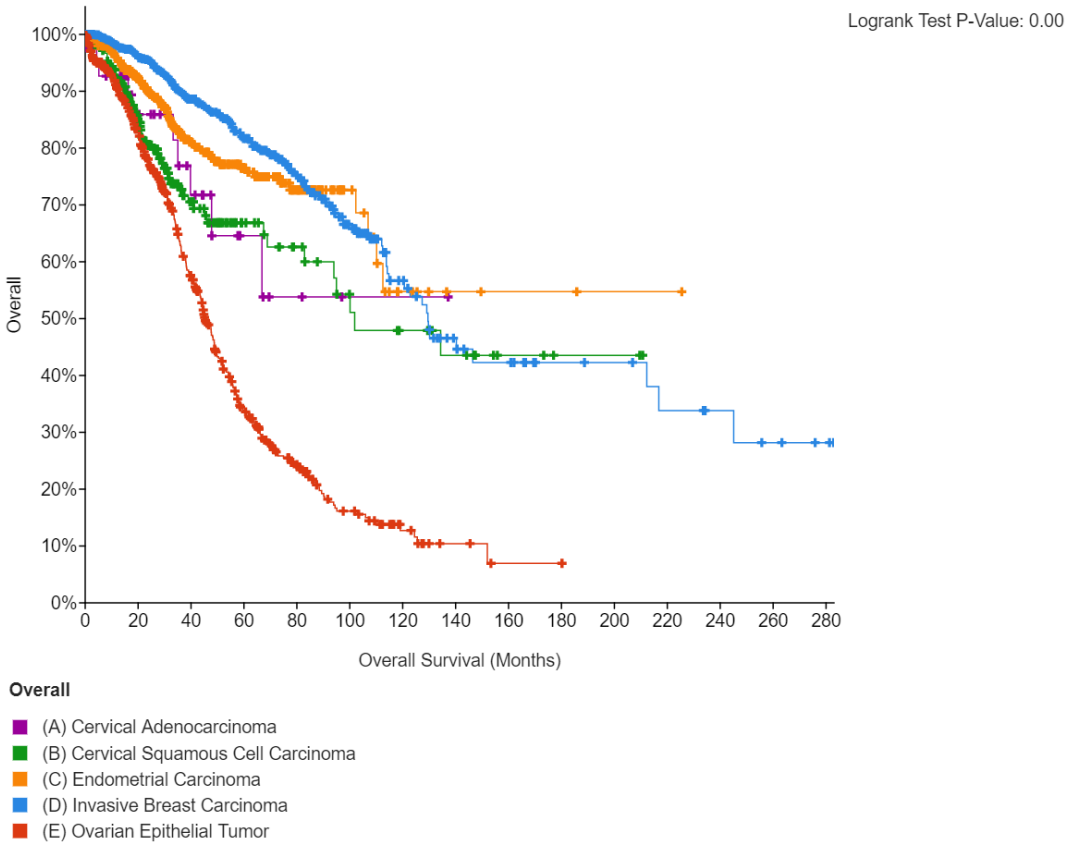
Comparing between overall survival statuses of the patients in the four cancers by cancer type, we see that the tumor type with the lowest median of overall survival time is **Ovarian Epithelial cancer**.

We chose to add the following table:

Survival Type	Median months survival in (B) Cervical Squamous Cell Carcinoma (95% CI) ▲	Median months survival in (C) Endometrial Carcinoma (95% CI)	Median months survival in (D) Invasive Breast Carcinoma (95% CI)	Median months survival in (E) Ovarian Epithelial Tumor (95% CI)
Overall	101.82 (93.99 - NA)	NA	129.57 (114.80 - 245.09)	45.14 (43.36 - 48.79)

From the table, we understand that the lowest median months survival for any type of cancer (among the four cancers chosen for our analysis) is ***45.14 months***, and is received for patients with the **Ovarian Epithelial Tumor**.

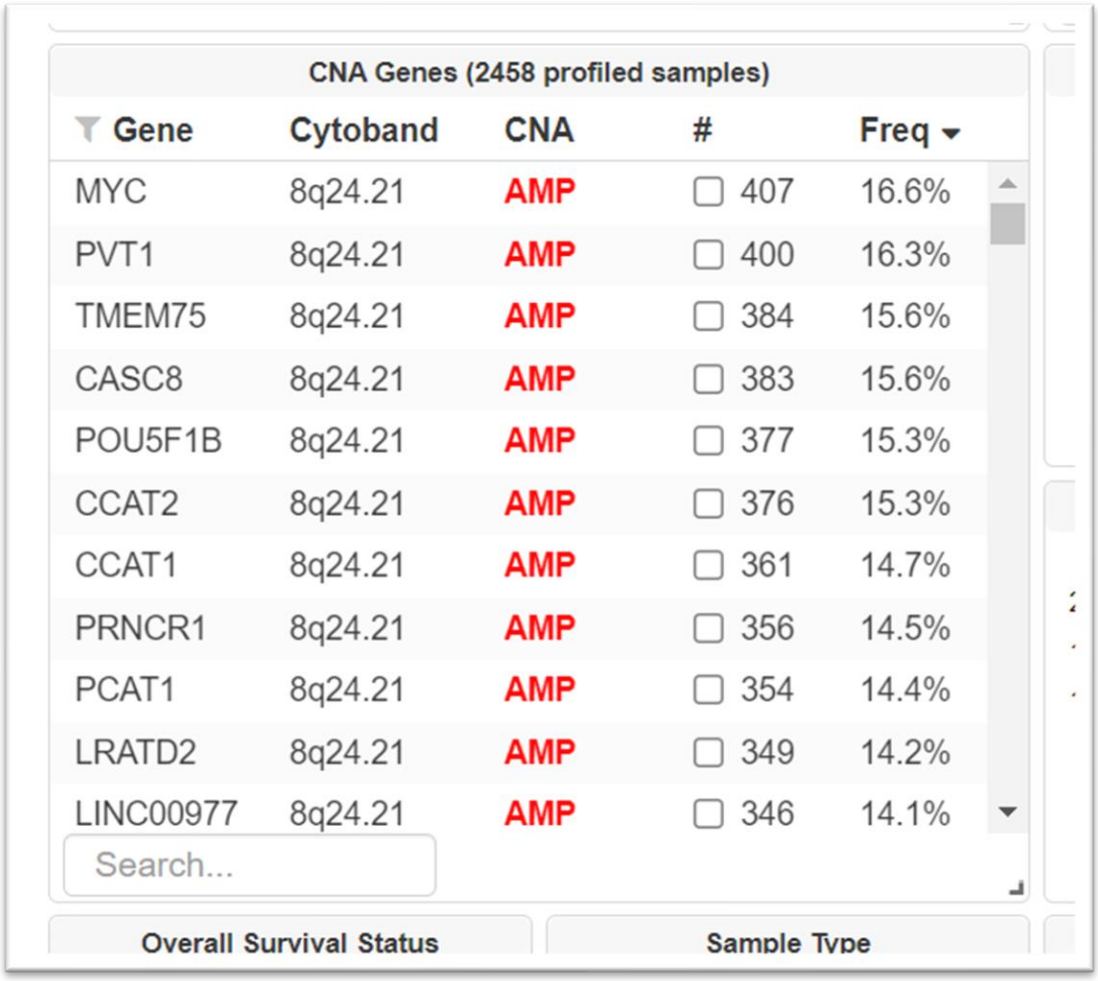
Here is another graph that helps us come to this conclusion:



Q1.3

The gene with the highest frequency of copy number alterations (i.e. CNA) in this combined study is **MYC** (with a frequency of ***16.6%***).

This can be easily found with looking at the following table:



MYC gene's functional role in the cell is encoding a multifunctional, nuclear phosphoprotein that controls a variety of cellular functions, including cell cycle, cell growth, apoptosis, cellular metabolism and biosynthesis, adhesion, and mitochondrial biogenesis – as MYC (MYC Proto-Oncogene, BHLH Transcription Factor) is a protein coding gene.

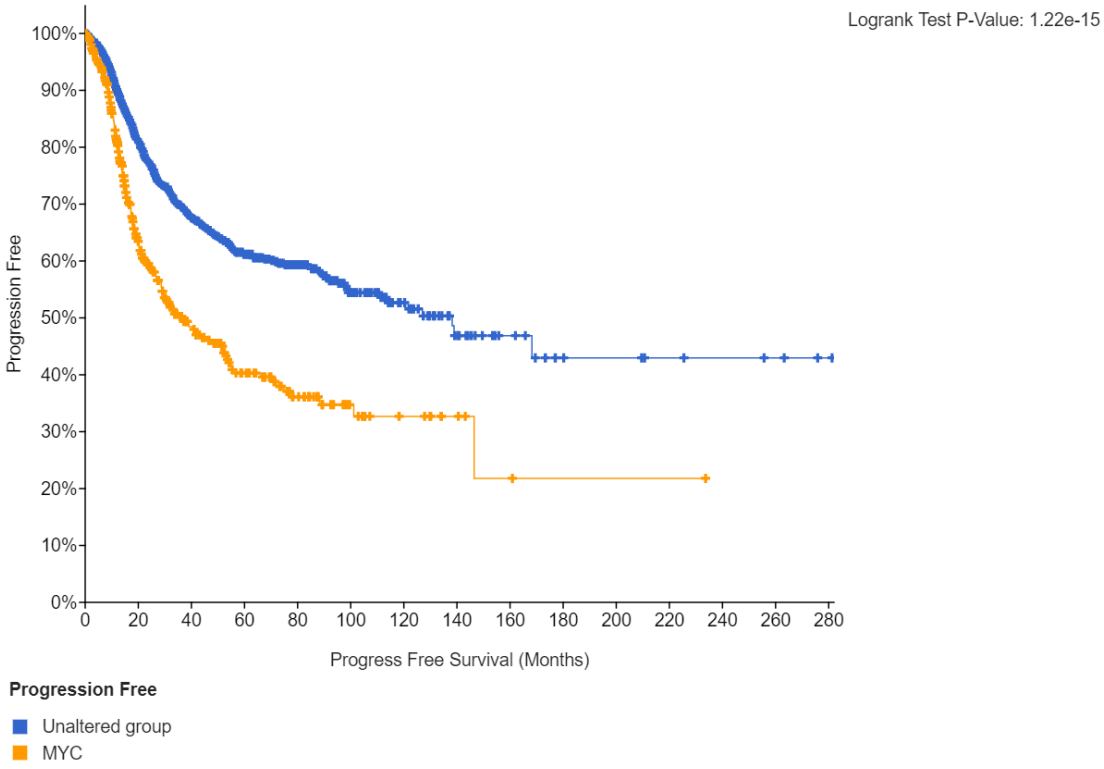
As we read a little online about the gene, we discovered that MYC has several diseases associated with it.

Cancer being one of those, we learned that MYC might induce tumorigenesis by evading multiple tumor-suppressing checkpoint mechanisms, including proliferative arrest, apoptosis, and/or senescence. On MYC suppression these barriers are restored, enabling sustained tumor regression.

Q1.4

Relying on this comparison/survival analysis, while querying the combined studies (for the MYC gene found in 1.3), we can say that in the four cancers, the progression free survival data differs significantly with the altered MYC gene versus the unaltered genes group.

We decided to have a look at the following graph:



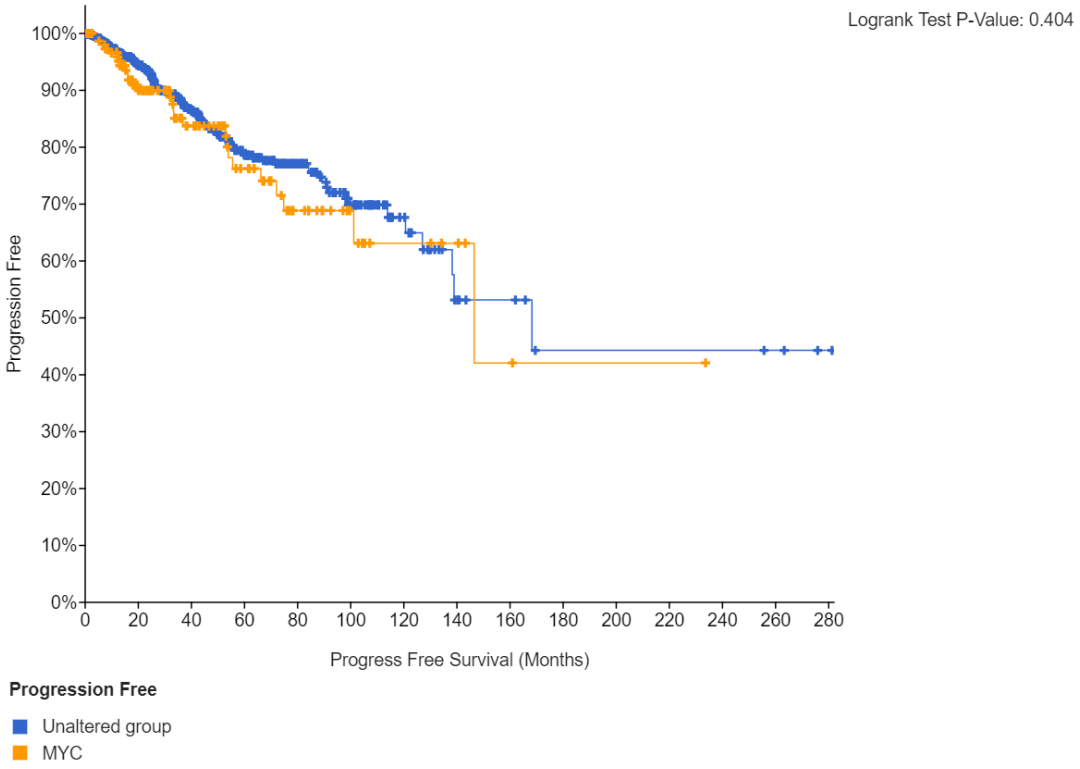
This graph tells us that the percentage of progression free survival among patient with the altered MYC gene is much lower than the equivalent percentage among the unaltered group, over time. *For example, we see that after 200 months the difference is between around 20% to over 40%.*

Q1.5

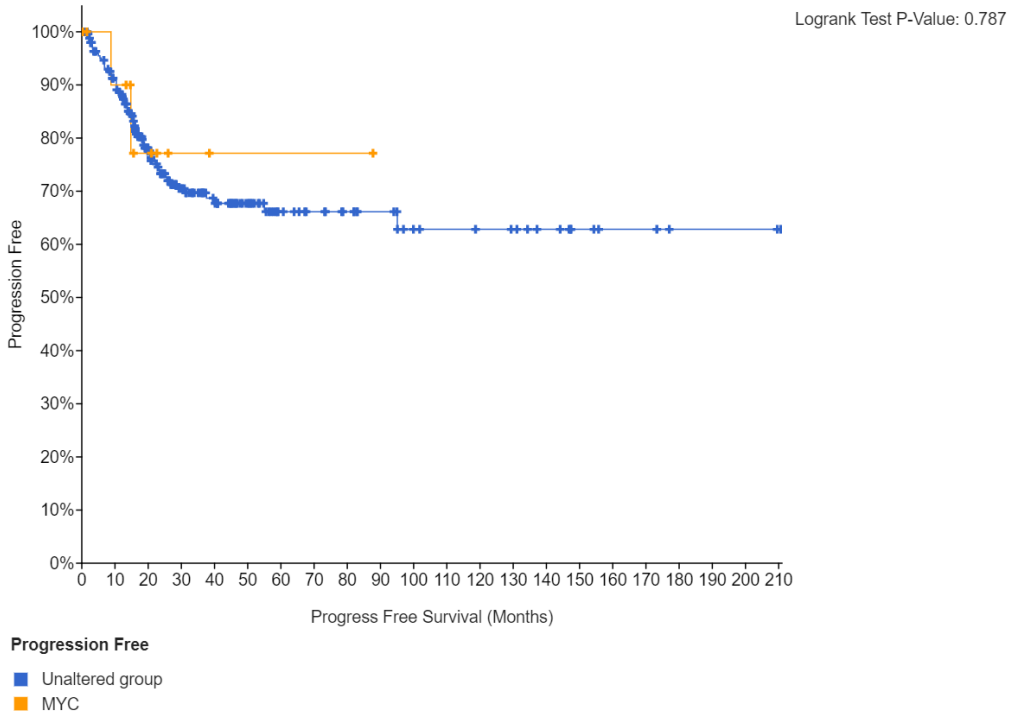
In this section, we have browsed each cancer type separately for its progression free plot of the MYC gene versus the unaltered genes group.

You may enjoy them as well:

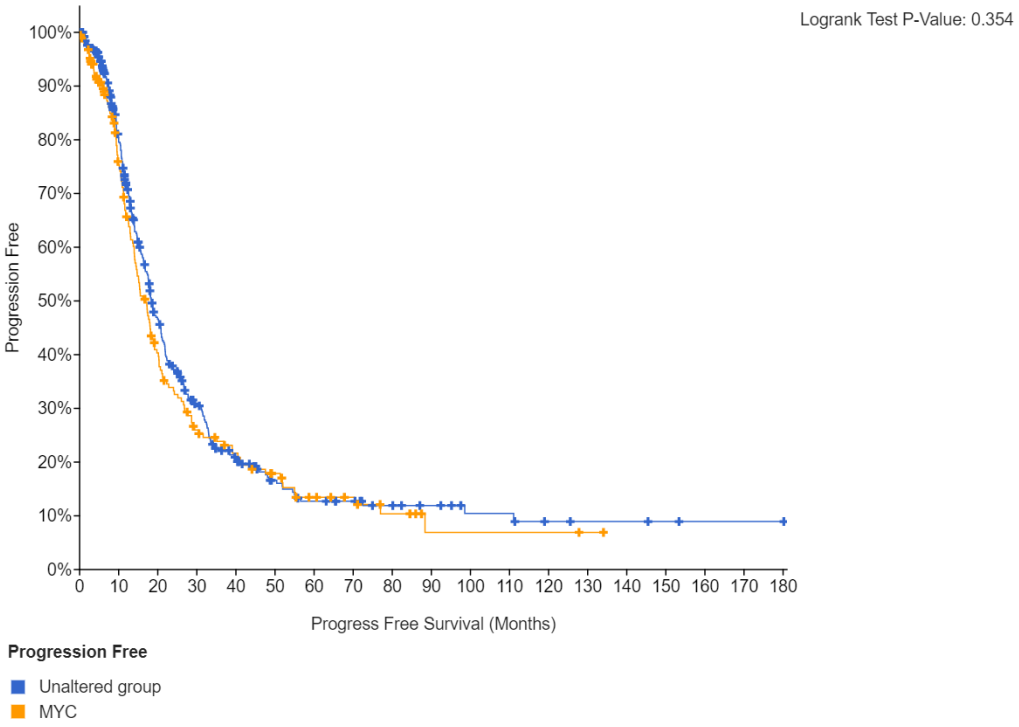
Breast Invasive Carcinoma-



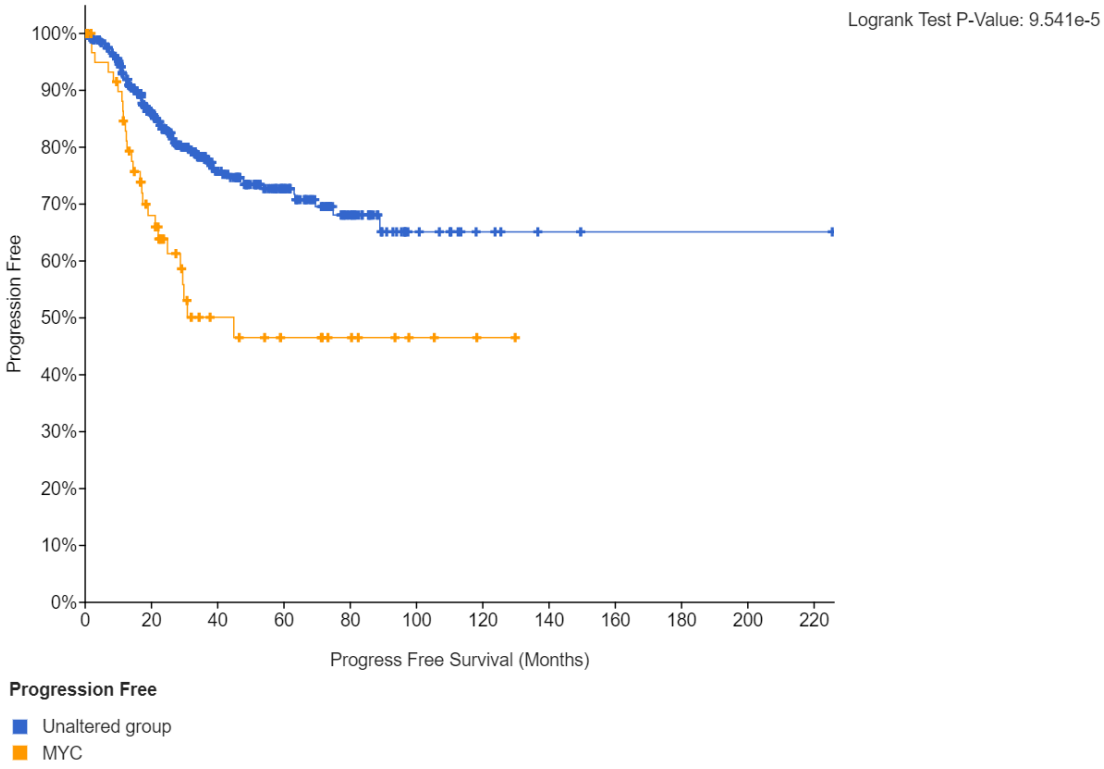
Cervical Squamous Cell Carcinoma-



Ovarian Serous Cystadenocarcinoma-



Uterine Corpus Endometrial Carcinoma-



In this section, we have browsed each cancer type separately for its progression free plot of the MYC gene versus the unaltered genes group.

Relying on the plots in front of us, it seems that the MYC gene may be connected to the progression free survival of one type of cancer significantly more than the other three – and that is the ***Uterine Corpus Endometrial Carcinoma***.

In 1.4, we saw that the progression free survival of patients with the MYC gene altered is lower than the equivalent of patients in the unaltered group. Browsing for each the connection between the gene and each cancer individually helps us come to an important conclusion-

It seems that the difference in the progression free survival percentage it caused by the Uterine Corpus Endometrial Carcinoma specifically, as the other three types of cancer do not demonstrate the same reduction in percentage.

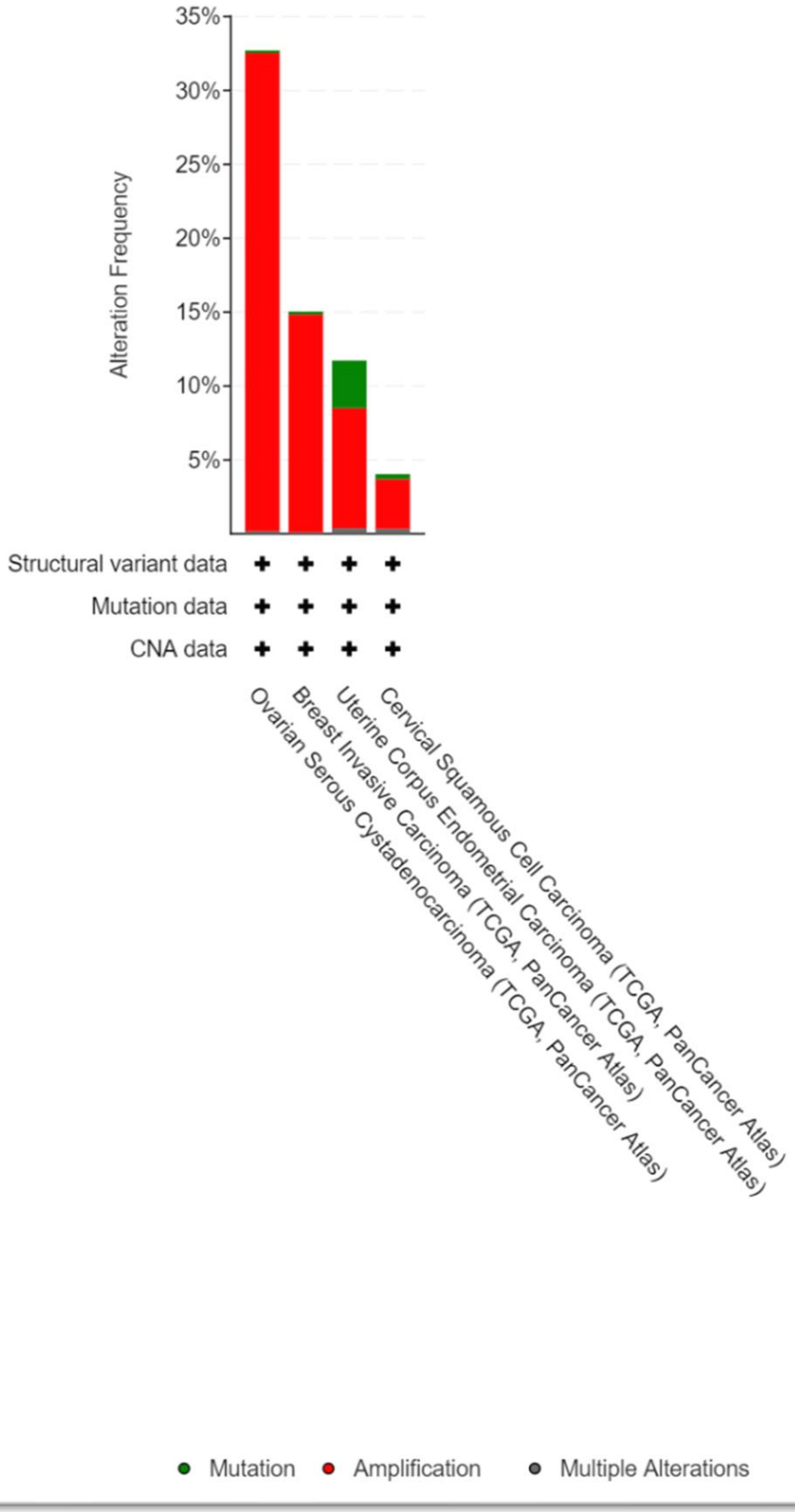
Q1.6

In total, the gene is altered in ~17% of queried patients/samples (as calculated in *cBioPortal*).

The alteration frequency in the MYC gene for all four cancers analyzed is seen in the **Ovarian Serous Cystadenocarcinoma**.
Observing the plot attached at the end of this section, we notice that the gene is altered over 30% of the patients with this type of cancer, while it is altered in less than 15% of patients with all other three types of cancer.

The **most abundant genetic alteration**, in all four cancers, is clearly **Amplification**.
As can be seen in the same plot attached, the main alteration in all four types of cancer is amplification (we may state that according to the color of the columns in the graph, which is red – representing amplification).

You may see the plot below:


35%
30%
25%
20%
15%
10%
5%
Alteration Frequency
Structural variant data
Mutation data
CNA data
Ovarian Serous Cystadenocarcinoma (TCGA, PanCancer Atlas)
Breast Invasive Carcinoma (TCGA, PanCancer Atlas)
Uterine Corpus Endometrial Carcinoma (TCGA, PanCancer Atlas)
Cervical Squamous Cell Carcinoma (TCGA, PanCancer Atlas)
Mutation
Amplification
Multiple Alterations

Question 2

Q2.1

let-7g is family of miRNAs.

miRNAs are non-coding RNAs that are involved in post-transcriptional regulation of gene expression in organisms. They affect both the stability and translation of mRNAs.

dysregulation of let-7g is known to be a factor of several types of cancer, such as lung and breast cancer.

Q2.2

Q2.2.1

The gene WDR82 overlaps let-7g.

Q2.2.3

It is clearly visible that the let-7g section is conserved very accurately through most species we checked, but the areas surrounding the let-7g are only conserved in apes.

Q2.2

Q2.3.1

TP53 & AGO2 tend to co-occur.

AGO2 & MIR-34A tend to be mutually exclusive.

TP53 & MIR-34A tend to co-occur.

Q2.3.2

The study we found is "The Role of TP53 in miRNA Loading Onto AGO2 and in Remodelling the miRNA–mRNA Interaction Network".

a cite from the study concerning TP53 & AGO2 tendency to co-occur:

"TP53 directly associated with AGO2, and induced and reduced loading of a subset of miRNAs, including the lethal 7 (let-7) miRNA family members, onto AGO2 in response to DNA damage."

Q2.3.3

In the study "TP53 regulates miRNA association with AGO2 to remodel the miRNA–mRNA interaction network" the writers discuss the connection between the pair and the let-7 family:

"This revealed a significant TP53-dependent regulatory effect on the association between AGO2 and the entire let-7 family apart from let-7d (Fig. 2A), and we validated these findings by RT-qPCR (Fig. 2B,C)."