Analysis Report

Overall Question:

Can mutations be discovered that are associated with treatment response?

1. Data Loading and Preparation:

The dataset containing cancer genomic data from 50 patients was loaded.

Each patient's tumor mutations were stored in Mutation Annotation Format (MAF) files.

Clinical information, including response to treatment, was loaded from the sample information file.

2. Mutation Filtering:

Mutations classified as "Silent" were filtered out, focusing on nonsynonymous mutations that result in changes to the produced protein.

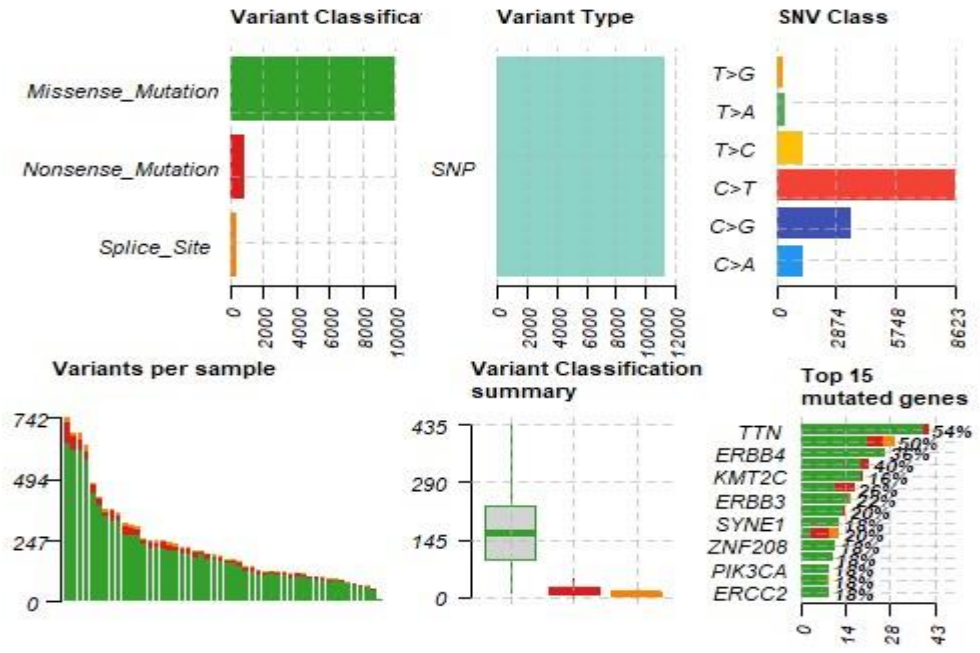

Fig 1: PlotSummary

3. Common Mutations:

The 15 most common mutations were identified based on the combination of gene name (Hugo_Symbol) and protein change. To infer more about the type of mutations and most mutated gene, oncoplot was used.

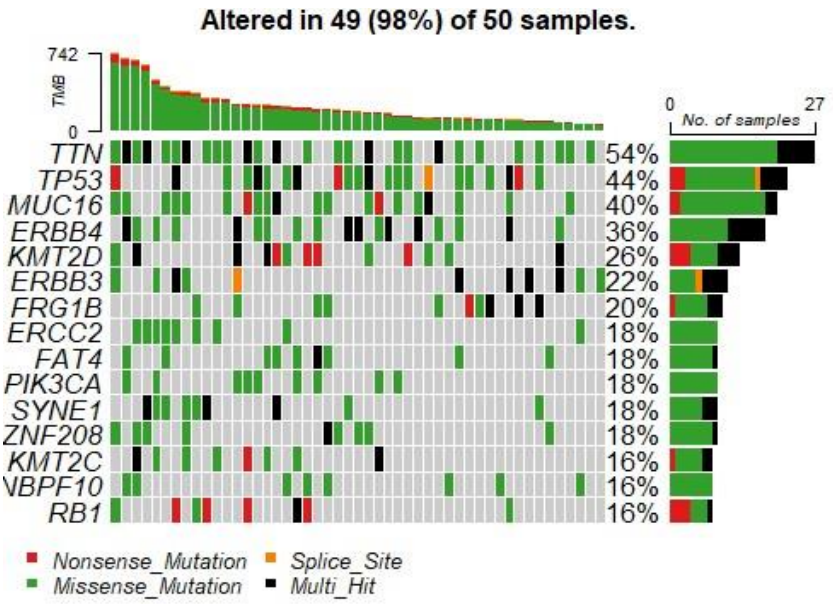

Fig 2: Oncoplot

4. Statistical Analysis:

A statistical test (Fisher's exact test) was performed to explore if any mutated genes are enriched in patients who responded or not.

Clinical Enrichment Analysis:

Responders and non-responders were 25 for each group (split into 2 patient count)

Out of the 134 genes analyzed, 7 genes were found to be enriched.

Enrichment was determined based on a significance level of $p < 0.05$ and odds ratio > 1.

This finding suggests a potential association between these genes and the clinical feature of interest (e.g., treatment response).

Interpretation:

The enrichment of these genes indicates that they may play a role in the biological processes underlying the clinical feature being studied.

The odds ratio greater than 1 suggests that the likelihood of observing mutations in these genes is higher in the group of interest compared to the rest of the sample.

5. Scatter Plot:

A scatter plot was created to visualize the number of mutated patients (x-axis) and the results from the statistical test (y-axis).

The plot did not show a clear association between the number of mutations and treatment response. And hence we used plotEnrichmentResults function from maftools to get a better visualization and identify the enriched genes.

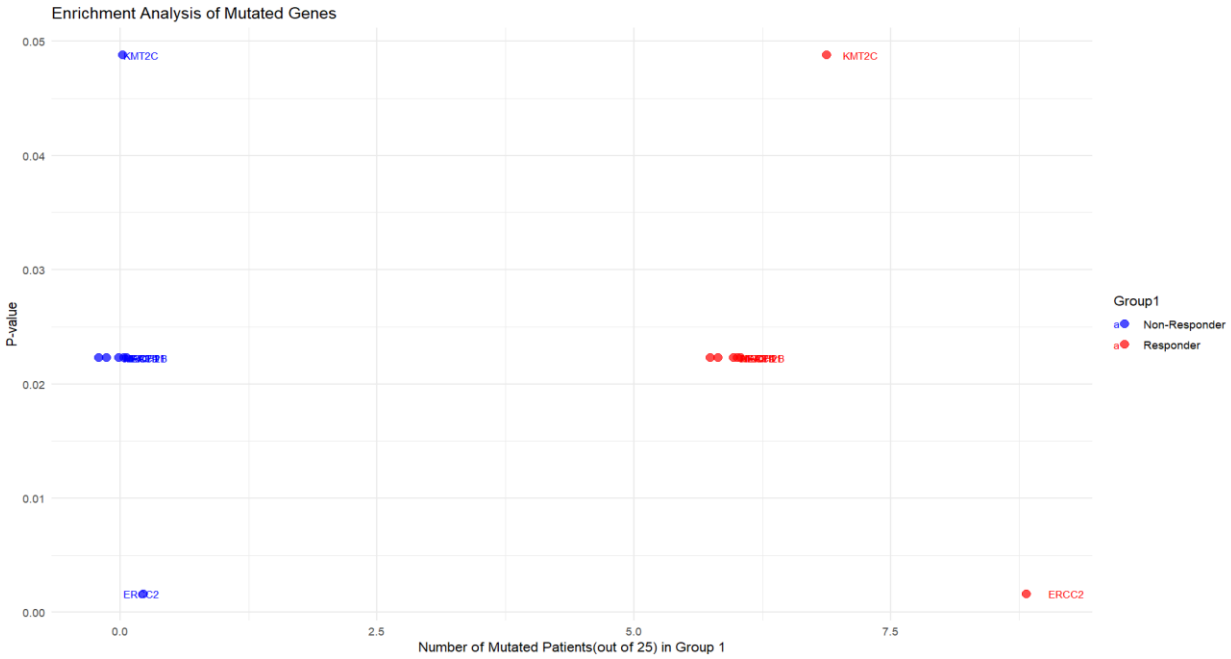

Fig 3 Scatterplot

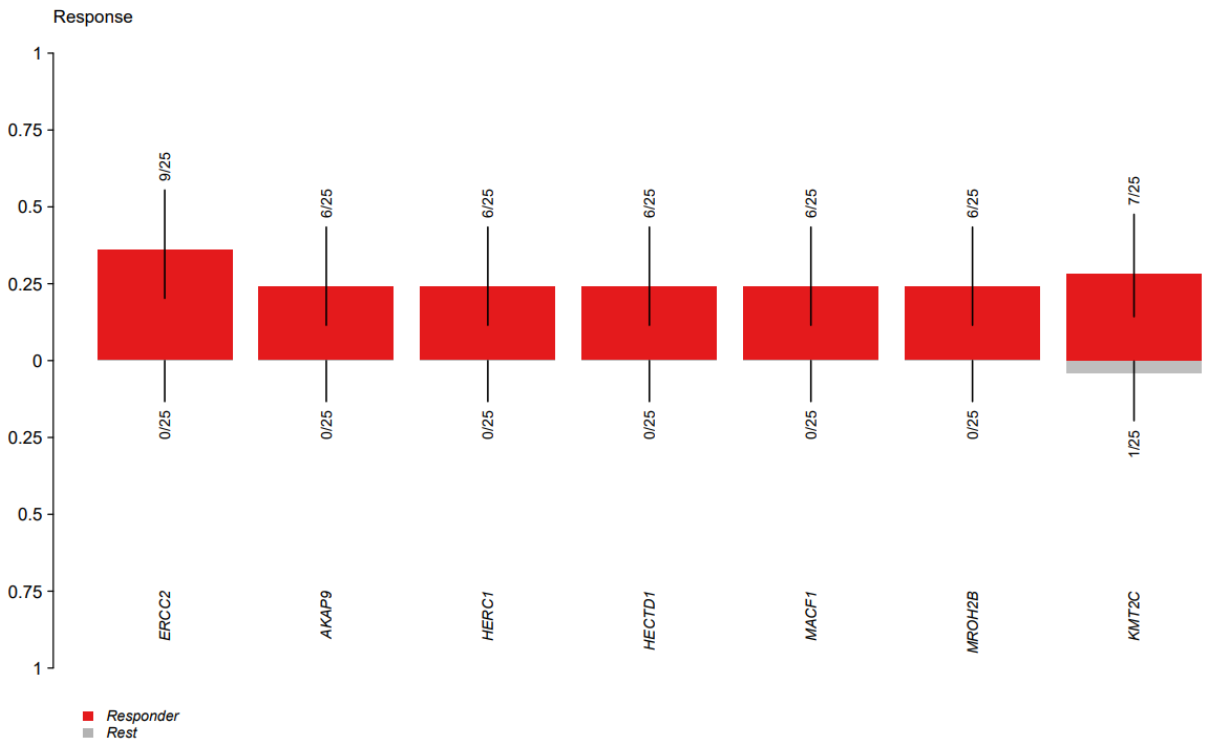

Fig 4 plotEnrichmentResults.

6. Mutation Burden Analysis:

The mutation burden was calculated for the nonsynonymous mutations per megabase. It suggests that, on average, each megabase of genomic sequence in the studied samples contains 0.03 mutations. This value provides a baseline for understanding the overall mutation rate in thegenomic region analyzed.

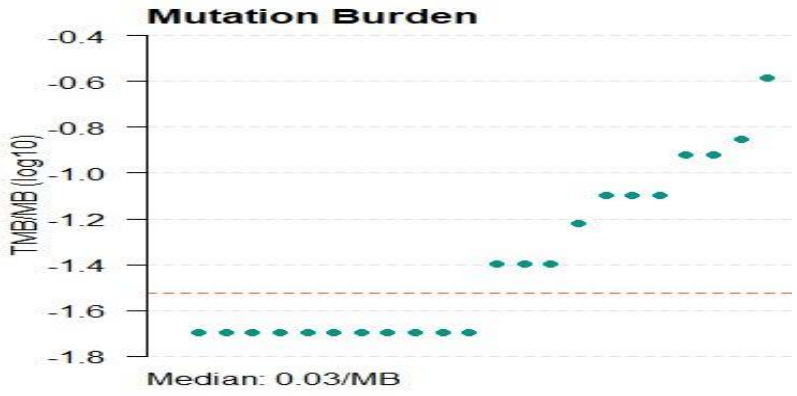

Fig 5 Nonsyn_MB

Number of mutants and wildtypes and significant difference between mutant and wildtype for the most enriched gene:

Total mutants: 175 , wildtype = 129 , mutant type = 46 (for 7 enriched genes)

Total mutants : 25, wildtype = 16 , mutant type = 9 (for most enriched genes)

Enriched Proteins using pfamDomains were CO504 and 7tm_.

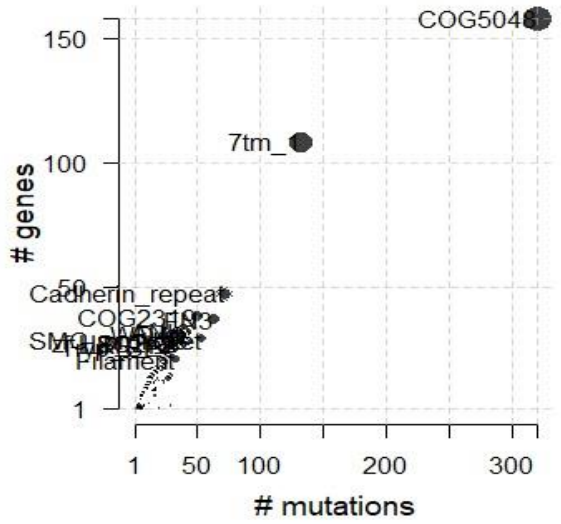

Fig 6 pfamDomain.

Fisher's exact test was performed to compare the number of mutant and wild-type samples for ERCC2.

No statistically significant difference was found in the number of mutations between the mutant and wild-type groups ($p > 0.05$).

Conclusion:

The analysis identified mutations that were enriched in certain genes, indicating a potential association with the studied clinical feature (treatment response).

However, when comparing these mutations between responders and non-responders, the enriched gene(s) showed a higher p-value, suggesting that these mutations may not be directly associated with treatment response in this dataset.

This highlights the complexity of the genetic factors influencing treatment response in cancer patients.

Further research and collaboration are essential to validate these findings and explore additional genetic and environmental factors that may play a role in treatment response.

Suggestions for Improvement:

Explore additional statistical tests or machine learning approaches to uncover potential associations.

Validate the findings using independent datasets to ensure reproducibility.

Final Thoughts:

The lack of statistically significant differences between wild type and mutant type for response suggests that, in this dataset and analysis, the treatment response may be independent of the mutations observed in the patients. This could mean that the treatment effectiveness is not influenced by these specific mutations, or that other factors not considered in the analysis (such as additional genetic variations, environmental factors, or treatment protocols) play a more significant role in determining treatment response.

It's important to interpret these results cautiously and consider them within the context of the specific dataset and analysis performed. Additional studies and analyses may be needed to further explore the relationship between mutations and treatment response in this cancer type.