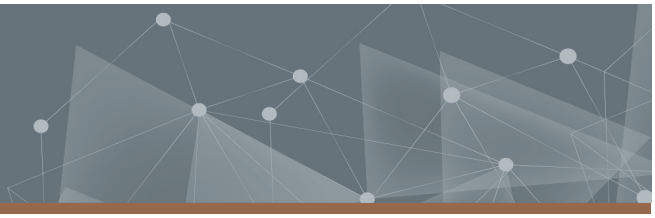
CHALMERS
UNIVERSITY OF TECHNOLOGY



Identifying genetic biomarkers predicting response to immunotherapy in non-small cell lung cancer

Master's thesis in Biotechnology

Louise Stauber Näslund

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING
DIVISION OF SYSTEMS AND SYNTHETIC BIOLOGY

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING
DIVISION OF SYSTEMS AND SYNTHETIC BIOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

Master's Thesis 2022

Identifying genetic biomarkers predicting response to immunotherapy in non-small cell lung cancer

Louise Stauber Näslund

CHALMERS
UNIVERSITY OF TECHNOLOGY

Identifying genetic biomarkers predicting response to immunotherapy in non-small cell lung cancer

Louise Stauber Näslund

Supervisor: Anna Rohlin, Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital
Examiner: Aleksej Zelezniak, Department of Biology and Biological Engineering, Division of Systems and Synthetic Biology
Master's thesis in Biotechnology 2022

Department of Clinical genetics and genomics
Sahlgrenska University Hospital
Gothenburg, Sweden 2022

Written in LaTeX
Gothenburg, Sweden 2022

Identifying genetic biomarkers predicting response to immunotherapy in non-small cell lung cancer

Louise Stauber Näslund
Department of Clinical Genetics and Genomics
Sahlgrenska University Hospital

Abstract

Background Immunotherapy has revolutionized the treatment of non-small cell lung cancer (NSCLC) in the last decade. However, not all patients respond to immunotherapy and current biomarkers used for patient selection are not optimal. Therefore, new and better biomarkers are urgently needed. This study aims to find genetic biomarkers predicting immunotherapy response in NSCLC patients.

Method Tumor DNA was sequenced for 44 NSCLC patients undergoing immunotherapy. Variants in 597 genes from GATC Biotech's OncoPanel All-in-One were assessed *in silico* and the genetic landscape was characterized. Kaplan-Meier analysis using log-rank test was used to assess the association of the top frequently mutated genes with survival and Cox regression was used to adjust for patients-related factors. Association with immunotherapy response was evaluated using Pearson's chi-squared test.

Results Patients with *KRAS* mutation and *KRAS/LRP1B* co-mutation were identified to be associated with prolonged survival (p=0.033 and p=0.022) and a trend for preferable immunotherapy response was observed. Patients with a low number of variants classified as pathogenic, likely pathogenic and "VUS+" was also found to be associated with survival (p=0.032) and were more likely to be responders of immunotherapy compared to patients with a high number of these variants (p=0.020).

Conclusion This project has further supported the role of *KRAS* as a potential predictive biomarker of immunotherapy response and has provided evidence for the *KRAS/LRP1B* co-mutation and the number of classified variants as potential biomarkers. Further studies including more patients may find additional results supporting the presented findings.

Keywords: NSCLC, genetic biomarkers, immunotherapy, immunotherapy response

Acknowledgements

I would like to thank my supervisor Anna Rohlin for begin supportive, helpful and engaged throughout this project. I would also like to thank Ella Äng Eklund and Maria Yhr for your help and support, as well as all other individuals involved in the BioLung project. Lastly, I would like to thank the patients who make this kind of research possible.

Contents

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Aim . . . 2
- 1.3 Delimitations . . . 2

2 Theory **3**
- 2.1 Immunotherapy . . . 3
  - 2.1.1 Immune checkpoint inhibitors . . . 3
  - 2.1.2 Established biomarkers of ICI response . . . 4
- 2.2 Survival analysis . . . 5
- 2.3 Variant types and cancer gene terminology . . . 7

3 Methodology **8**
- 3.1 Study cohort . . . 8
- 3.2 Ethics approval and confidentiality . . . 8
- 3.3 Samples and tumor sequencing . . . 8
- 3.4 Identification of genetic biomarkers . . . 9
  - 3.4.1 Classification of variants . . . 9
  - 3.4.2 Variants located in driver genes . . . 12
  - 3.4.3 Identification of frequently mutated genes . . . 12
  - 3.4.4 Statistical analysis . . . 13

4 Results **14**
- 4.1 Clinical characteristics of NSCLC patients . . . 14
- 4.2 Somatic mutation landscape in NSCLC . . . 15
- 4.3 Technical analysis and classification of tumor DNA variants . . . 16
- 4.4 Overall survival analysis . . . 17
  - 4.4.1 Individual genes . . . 17
  - 4.4.2 Co-mutations . . . 20
  - 4.4.3 Groups of variants . . . 22

5 Discussion **27**
- 5.1 Evaluation and classification of variants . . . 27
- 5.2 Characterization of genetic landscape of NSCLC patients . . . 27
- 5.3 Genes associated with survival and response . . . 28
- 5.4 Co-mutations associated with survival and response . . . 28
- 5.5 Groups of variants associated with survival and response . . . 29

5.6 Limitations 30
5.7 Future prospects and societal impact 30

6 Conclusions **31**

Appendix A Kaplan-Meier curves **38**

1 Introduction

This project is a part of a study conducted at the Department of Clinical Genetics and Genomics at Sahlgrenska University Hospital.

1.1 Background

Lung cancer is the second most diagnosed cancer worldwide, with approximately 2.2 million (11.4 %) new cases in 2020, and remains the leading cause of cancer death, with an estimated 1.8 million (18 %) deaths [1]. Despite modern treatment, including molecular targeted therapies against driver oncogenes such as *ALK* fusions and *EGFR* mutations, the five-year survival is still unsatisfactory as patients diagnosed with lung cancer during 2010-2014 had a five-year survival estimate of only $10-20$ % in most countries [2][3].

Non-small cell lung cancer (NSCLC) accounts for nearly 85 % of all lung cancer cases and encompasses the subtypes lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) [4]. The majority of patients with NSCLC have distant metastases by the time of diagnosis, and as a result the cure rates are low and the risk for progression and relapse are high [4][5].

In the last few years, treatment strategies for patients with NSCLC have changed considerably following the introduction of immunotherapy targeting immune checkpoint programmed cell death-1 (PD-1) and programmed cell death-1 ligand (PD-L1). Immune checkpoint inhibition (ICI) has achieved remarkable clinical results due to improved overall survival and durable responses for NSCLC patients without actionable driver mutations [6][7][8]. Unfortunately, only a minority of NSCLC patients experience durable clinical benefit of ICI treatment and many patients relapse in short time frame or experience life threatening immunotoxicity [9][10][11].

Due to the heterogeneity of ICI response, it is important to be able to select patients with a high likelihood of clinical response. However, predictive biomarkers of PD-1 and PD-L1 inhibition are few and often not optimal. To date, only PD-L1 expression and tumor mutation burden (TMB) are approved for clinical use as biomarkers for ICI treatment of NSCLC, although TMB has not yet been approved by the EMA [6]. Despite being promising, PD-L1 expression and TMB are far from optimal and new biomarkers are urgently needed.

In recent years, several new biomarkers have been under investigation. ranging from genetic and immunologic biomarkers to tumor-derived components in the blood. The focus in this thesis are genetic biomarkers, where broad genomic sequencing approaches, including whole-exome sequencing have been used to investigate the

genetic landscape of tumors to identify genetic patterns or specific mutations among patients and in individual tumors that can be used as biomarkers for ICI response.

1.2 Aim

The aim of this project is to characterize the genetic landscape of patients with NSCLC in order to identify genetic biomarkers predicting immunotherapy response. This will be done by first evaluating and interpreting DNA variants, followed by an analysis of the most frequently mutated genes. Finally, survival analysis will be used to associate mutations in individual genes, co-mutations, and different groups of variant data to survival and immunotherapy response.

1.3 Delimitations

This project will screen for variants in 597 genes found in GATC Biotech's Onco-Panel All-in-One. Only synonymous, non-synonymous, short frameshift indels, splice variants in the +1/+2 or -1/-2 position, inframe, and stop/start-loss variants will be evaluated. Only variants with an allele frequency $\geq$5 % will be included in the analysis due to the the poor quality of the Formalin-Fixed Paraffin-Embedded (FFPE) tumor samples used for sequencing.

2 Theory

This section describes important concepts and terminology related to the project.

2.1 Immunotherapy

Immunotherapy aims to use the patients own immune system to kill cancer cells. For patients with NSCLC, one type of immunotherapy is immune checkpoint inhibition (ICI), meaning antibody inhibition of the immune checkpoints PD-1 or PD-L1. Currently, the FDA has approved two PD-1 inhibitors (nivolumab and pembrolizumab) and two PD-L1 inhibitors (atezolizumab and durvalumab) for use as NSCLC immunotherapy treatment [12].

2.1.1 Immune checkpoint inhibitors

T cells are involved in the immune response against bacteria, fungi, viruses, parasites, and tumors. Although necessary, they can cause inflammation, which can lead to autoimmunity or immunopathology and must therefore be tightly regulated. The immune system is regulated by regulatory cells from the innate and adaptive immune system and immune checkpoints that control T-cell activation. The regulatory cells and immune checkpoints are often enhanced during cancer or inflammation to suppress and evade the immune system, which has made them important therapeutic targets [13].

Immune checkpoints have distinct receptors and ligands. Relevant for this project is the receptor PD-1 and its ligand PD-L1 [13]. PD-1 is mainly expressed on the surface of T cells and PD-L1 is mostly expressed on different types of cancer cells, such as lung cancer and melanoma [14]. PD-1/PD-L1 ligation suppresses T-cell function and is the key mechanism of cancer cells to evade the immune system [12]. By blocking PD-1 or PD-L1 using anti-PD-1/PD-L1 antibodies, the T-cell suppression is avoided, meaning the T-cell regain their ability to kill cancer cells (Figure 1) [13].

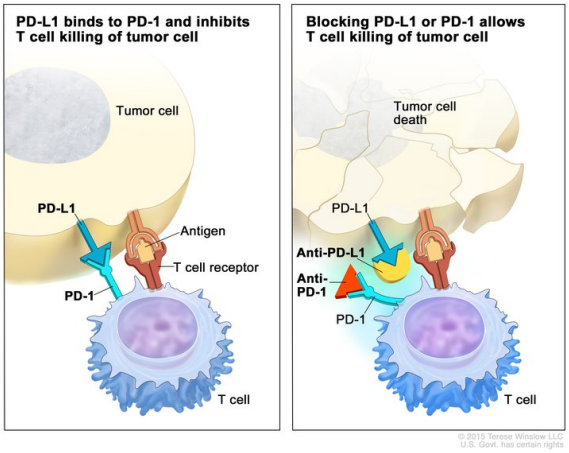


Figure 1: PD-1/PD-L1 ligation suppresses T cell function and inhibits T cells killing tumor cells. Introducing antibodies blocking either PD-1 or PD-L1 allows the T cells to kill the tumor cells. (From the National Cancer Institute © (2015) Terese Winslow LLC).

2.1.2 Established biomarkers of ICI response

Biomarkers are measurable biological characteristics that can be used to predict immunotherapy response and be used as a basis for treatment decisions by selecting patients with a high likelihood of clinical response to treatment [15]. To date, PD-L1 and tumor mutation burden (TMB) are the only approved predictive biomarkers for PD-1/PD-L1 inhibition [6].

PD-L1 expression on tumour tissue is determined using immunohistochemistry and is represented as a tumor proportion score (TPS). TPS is the percentage of viable tumor cells showing partial or complete membrane PD-L1 staining in relation to all viable tumor cells and is given as $<1\,\%$, $1\,\%$-$49\,\%$, and $\geq 50\,\%$ [16][17]. PD-L1 positivity is defined as a TPS of $\geq 1\,\%$ and has been associated with significantly longer progression-free survival (PFS) and overall survival (OS) [18][19]. Currently, a TPS of $\geq 50\,\%$ is used to select stage IV NSCLC patients for first-line monotherapy using pembrolizumab [20].

Tumor mutation burden (TMB) is in this project defined as the total number of non-synonymous mutations per coding area of a tumor genome in either tumor tissue or blood, and is reported as mutations per megabase (Mb) of DNA. Tumors with a higher number of somatic mutations have been associated with a greater generation of neoantigens and subsequent development of immunogenicity. It has therefore been suggested that patients with a high TMB respond better to ICI compared to patients with low TMB , which has been supported by reports of longer PFS and OS

in patients with a TMB higher than 10 mutations/Mb [21][22][23]. Currently, the FDA has approved TMB as an agnostic biomarker for selecting patients for treatment with pembrolizumab using a cut-off of $\geq$10 mutations/Mb [6].

Although promising, PD-L1 expression and TMB are far from optimal biomarkers. Studies have shown that patients negative for PD-L1 expression can present durable response to immunotherapy and improved OS [18]. In addition, there is no consistent standard for measuring PD-L1 expression in tumor cells. Different definitions of PD-L1 positivity, detection platforms, and evaluation systems makes it difficult to compare PD-L1 expressions between clinical trials [24]. Similar issues can be found for TMB, where there is no standard assessment across research and clinical studies. There are also many factors that can influence the TMB measurement, including the quality and quantity of the sample, sequencing platform, included mutations, genome coverage, bioinformatic pipeline, and the definition of what is considered high or low TMB [25].

2.2 Survival analysis

The main assessment in many cancer studies is survival time, which is the time from a starting point to an event of interest, often death. The difficulty of this type of data is that censoring often occurs, making the survival data incomplete. Censoring can happen when: (a) a patient has not experienced the event before the last follow-up date; (b) a patient is lost to follow-up; or (c) a patient drops out or experiences another event that makes further follow-up impossible. Another problem is that survival data is rarely normally distributed, rendering many methods of analysis non-applicable. Therefore, special methods called survival analysis are necessary, which typically include univariate Kaplan-Meier survival analysis, log-rank tests, and multivariate Cox (proportional hazards) regression [26].

Kaplan-Meier survival analysis nonparametrically estimates the survival function (also called cumulative survival or survival probability) from observed survival times. The survival function describes the probability that an individual survives from the time of origin to a specified future time. Mathematically, the survival function is described as

$$S(t_i) = S(t_{i-1})(1 - \frac{d_i}{n_i})$$

where the probability of being live at time t_i, called $S(t_i)$, is determined by the probability of being alive at time t_{i-1}, called $S(t_{i-1})$, the number of patients alive just before time t_i, called n_i, and the number of events at time t_i, called d_i. At time $t_0 = 0$, the probability is $S(0) = 1$. By dividing the time period into smaller

intervals, the survival function can be calculated by multiplying the probability of surviving from one interval to the next, with the assumption that events happen independently from each other. The survival function is often plotted against time in Kaplan-Meier curves, which allows for estimations of median and mean survival times. Since survival data is often skewed, the estimated median is usually reported rather than the mean. Survival between groups can then be compared using the log-rank test, which is one of the most widely used nonparametric tests to compare survival between groups [26].

Cox (proportional hazards) regression is the most widely used multivariate method for analyzing survival data in the medical field. Statistical models, like Cox regression, is a way to assess survival with respect to multiple patient-related factors (covariates), such as age or smoking history, that could potentially affect the survival time. Additionally, they can provide an estimated effect of each factor [27]. The Cox regression model is a multiple linear regression of the logarithm of the hazard model and describes the relation between the hazard model and a set of covariates. In turn, the hazard model describes the probability that a patient under observation experiences the event around a certain time point. Cox regression is mathematically written as

$$h(t) = h_0(t) * exp(b_1x_1 + b_2x_2 + ... + b_px_p)$$

where the hazard function $h(t)$ is dependent on a set of covariates (x_1,x_2,...,x_p) whose impact is described by the respective regression coefficients (b_1,b_2,...,b_p). $h_0(t)$ is the baseline hazard, and $exp(b_i)$ is the hazard ratio (HR), which describes the relation between the probability of events in a treatment group compared to the probability of evens in a reference group. Covariates with a hazard ratio >1 are negatively associated with survival and covariates with a hazard ratio <1 are positively associated with survival [27].

Related to survival analysis is the association between a factor, such as a mutation in a gene, and immunotherapy response. In this project, this is evaluated using Pearson's chi-squared test, which provides a p-value and an odds ratio (OR). The OR represents the odds of an outcome based on a particular exposure and is given as a number between 0 and infinity. In this case, the outcome is immunotherapy response, and the exposure is a mutation in a gene of interest. In the scope of this project, an odds ratio (OR) <1 means that the mutated gene is associated with lower odds of immunotherapy response and an odds ratio above 1 that the mutated gene is associated with higher odds of immunotherapy response [28].

2.3 Variant types and cancer gene terminology

Throughout this thesis, variant, alteration and mutation are used interchangeably to describe nucleotides that differ in the sequenced tumor samples compared to a reference genome.

There are different types of variants evaluated in this project, including non-synonymous, synonymous, frameshift, inframe, stop-and startloss, and splice site mutations. Non-synonymous mutations alters the protein sequences and include missense and nonsense mutations. Missense mutations replace an amino acid with another and nonsense mutations introduce a stop codon. Synonymous mutations are mutations that do not alter the protein sequence and frameshift mutations can either be insertions or deletions and shifts the way the DNA is read. Inframe mutations are deletions where the reading frame is preserved. Stop- and startloss mutations are mutations that affect the termination and initiation codon, respectively. Splice site mutations occur at the boundary of exons and introns, and can disrupt how the gene is read by disrupting existing ones or creating new splice sites. Splice site mutations in a position marked "+" occur in the nucleotides just after the exon, and those marked "-" occur just before the exon [29][30]. In addition, mutations can be divided into driver and passenger mutations, where driver mutations cause the initiation and proliferation of cancer and passenger mutations do not [31].

Cancer genes can also be divided into oncogenes and tumor suppressor genes, where oncogenes are genes that are related to the formation and growth of cancer and tumor suppressor genes help regulate the growth cells [31].

3 Methodology

This project encompasses data analysis of tumor DNA NGS data. Sample collection and sequencing was done previous to the project.

3.1 Study cohort

Patients with NSCLC (stage III or IV) suitable for immunotherapy, were recruited from Gothenburg and Skövde, Sweden, during 2019-2021. Previous treatments, age, sex, or smoking status were not exclusion criteria. Patients were followed from the start of the immunotherapy treatment through, at most, 5 cycles of immunotherapy. At 10 months after the first treatment, the patients were divided into responders and non-responders. Responders were defined as patients with partial response, stable disease or complete response, and non-responders was defined as patients with progressive disease. Patients that had passed away before the 10 month cut-off were denoted as non-responders.

3.2 Ethics approval and confidentiality

The study was approved in 2018 by the Central Ethical Review Board in Gothenburg (diary number 953-18). Written informed consent was obtained from each participant and all personal data was gathered previous to this project and was regulated according to the General Data Protection Regulation (EU 2016/670). The obtained samples were registered in a biobank according to the Law of biobanks in healthcare (SFS 2002:297). In total, 54 patients were included in the study this project is apart of.

3.3 Samples and tumor sequencing

Formalin-fixed paraffin-embedded (FFPE) tumor samples were taken from archival tumor tissue obtained before the start of immunotherapy treatment for each patient. The FFPE tumor material was sequenced for 597 cancer genes in GATC Biotech's OncoPanel All-in-One (v2) by Eurofins Genomics (Europe Sequencing Gmb, Germany) using Genome Sequencer Illumina HiSeq (San Diego, CA, USA). Quality assessment of the raw data, mapping and variant calling was also performed by Eurofins Genomics. Genomic DNA from blood samples from each patient was also sequenced in order to filter out germline mutations in the tumor DNA, as to only retain somatic mutations. The data was then filtered further to only include variants with an allele frequency of $\geq$5 % and a population frequency $<$1 % (to filter out many

of the (likely) benign variants). In total, 44 out of 54 patients had tumor material sequenced and were included in this project.

3.4 Identification of genetic biomarkers

In this project, variants in individual genes, co-mutations, and groups of variants were evaluated as biomarkers. Manual evaluation and classification of the variants was the first step of the analysis.

3.4.1 Classification of variants

The classification of the tumor DNA variants was done according to the ComPerMed workflow created by Froyen et al. [32], illustrated in Figure 2. ComPerMed uses the five biological classes of the ACMG & AMP Standards and Guidelines published by Richards et al. [33] and includes the classes "Pathogenic", "Likely Pathogenic", "Variant of Unknown Significance (VUS)", "Likely Benign", and "Benign".

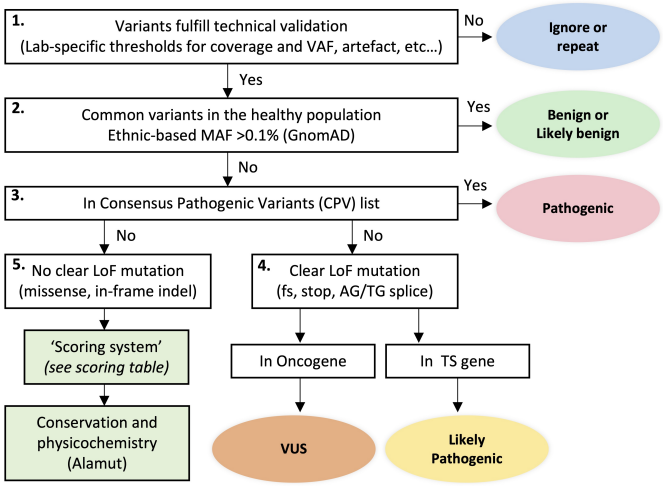


Figure 2: ComPerMEd workflow for biological classification of somatic variants, generated by Froyen et al. [32]. **(1)** Technical filtering and manual evaluation in IGV. **(2)** Check if variants are present in healthy population databases. **(3)** Check if variants are previously known to be pathogenic using the Consensus Pathogenic Variant (CPV) list. **(4)** Check if variants have a clear loss of function mutation, in which case it is checked if the variant is located in a tumor suppressor gene or in an oncogene. **(5)** Variants with no clear loss of function mutation is further classified using a scoring system and if scoring 1.5, evaluated in Alamut for nucleotide conservation and physicochemical effect. Adapted from Froyen et al. [32]

Initially, the variants were manually evaluated in Integrative Genomics Viewer (IGV) (v2.11.9) [34] to remove sequencing artefacts missed during previous filtering. Hg38 was used as the reference genome. The evaluation was done based on: distribution in forward and reverse reads, where variants were evaluated as artefacts if they only appeared in one direction or if the distribution was unbalanced; sequencing depth; and mapping, where variants were evaluated as artefacts if the mapping contained many surrounding errors or looked like structural errors. All artefacts were discarded from further analysis.

The remaining variants were then checked for their presence in the healthy population (Figure 2-Box 2). Using the gnomAD database (v2.1.1 and v3.1.1) (`https://gnomad.broadinstitute.org/`), the ethnic-based minor allele frequency (MAF) was determined for each variant, where variants with an ethnic-based MAF $\geq$**0.1 %** and $<$1 % were directly classified as "Likely Benign" and those $\leq$1 % as "Benign". Variants with an MAF $<$**0.1 %** or not reported in the gnomAD database required further classification and were checked to see if they were previously known to be pathogenic using the Consensus Pathogenic Variant (CPV) list (Figure 2-Box 3) (full list is found in Froyen et al. [32]).

Variants included in the CPV list were directly classified as "Pathogenic", while variants not included in the CPV list were evaluated for having a clear loss of function (LoF) mutation (Figure 2- Box 4). LoF mutations included frame shift, introduction of stop/start codon, loss of stop/start codon, and $\pm$ 1,$\pm$ 2 splice sites. Variants with a LoF mutation were then checked for being located in an oncogene, in which case the variant was classified as a "VUS", or in a tumor suppressor (Ts) gene, in which case the variant was classified as "Likely Pathogenic". To find out if the variant was located in an oncogene or a Ts gene, the Ts & Oncogene list by Froyen et al. [32] was used, as well as the Cancer Gene Census (CGC) catalogue in COSMIC (`https://cancer.sanger.ac.uk/census`).

Variants without a clear LoF mutation (Figure 2- Box 5), for example missense or in-frame indel mutations, were further classified using a scoring system (Table 1). The scoring table was modified compared to Froyen et al. [32]. The first parameter scored the variants based on the total number of entries for that specific amino acid change in that specific position in the Catalogue of Somatic Mutations in Cancer (COSMIC, `https://cancer.sanger.ac.uk/cosmic`). For solid tumors, a score of "+2" was given to variants with a total number of entries $\geq$ **50** and a score of "0" was given for variants with entries below 10. Intermediate numbers were given a score of "+1".

The second parameter described by Froyen et al., relates to the theoretical prediction tools SIFT and MutationTaster. In this project, however, the number of

Table 1: Scoring Table for the biological variant classification of non-loss-of-function variants.

Parameter	Score +2	Score +1	Score +0.5	Score 0	Score -1
Total # of entries of that particular AA change at that position in COSMIC	≥ 50	$50 > x > 10$	-	≤ 10	-
Prediction scores in VarSome	-	$\geq 75\%$ Damaging	$75\% > x \geq 50\%$ Damaging	$< 50\%$ Damaging	-
Harmful in functional studies (VarSome)	-	-	Yes	Not reported	No
Described in Varsome (ACMG classification) and ClinVar	-	-	As (Likely) Pathogenic	Not described/ unknown	As (Likely) Benign

Variants with a final score of ≥ 2 were classified as "Likely Pathogenic". Variants with a final score < 2 were classified as "VUS". Adapted from Froyen et al. [32].

theoretical prediction tools examined were expanded to encompass those included in the database VarSome (`https://varsome.com/`). A score of "+1" was given to variants where more than $75\,\%$ of the prediction tools predicted the variant to be damaging. Variants predicted to be damaging by between $50-75\,\%$ were given a score of "0.5", and variants predicted to be damaging by $<50\,\%$ were given a score of "0". The number of prediction tools available ranged from 3 to 19 between variants and was not taken into consideration when scoring the variants.

The third parameter was used to determine if the variant had been previously mentioned as harmful or not in functional studies. For this, VarSome was used to find relevant publications for each variant. If the variant had been reported as harmful in functional studies a score of "0.5" was given, if not reported a score of "0" was given and if a variant was found to not be harmful a score of "-1" was given.

The fourth parameter related to if the variant had been previously classified or described in a genomic database. In contrast to what is described by Froyen et al [32], both VarSome (ACMG classification) and ClinVar were taken into consideration in this project and a score was given for each database depending on the stated classification. Variants classified as "(Likely) Pathogenic" in VarSome or ClinVar were given a score of "0.5" and variants of "Unknown significance" or not described were given a score of "0". Variants described as "(Likely) Benign" were given a score of "-1".

A final score was then calculated from the four parameters and variants with a score of ≥ 2 were classified as "Likely Pathogenic" and variants with a score < 2 were classified as "VUS". In order to differentiate variants classified as VUS, variants with a score "1.5" were evaluated in Alamut for nucleotide conservation among species and physicochemical effect of the amino acid change. This was also a modification

compared to the workflow described by Froyen et al. [32]. Variants with a high conservation and moderate to high physicochemical effect were scored an additional 0.5 points and were classified as "VUS+", which in this project means variants classified as VUS that are more towards a likely pathogenic classification.

Exceptions to the ComPerMed workflow were done for *TP53*, *BRCA1* and *BRCA2* according to the procedure by Froyen et al [32]. Variants in *TP53* that were found in OncoKB (`https://www.oncokb.org/`) were classified as "(Likely) Pathogenic". Truncating and splice site *TP53* variants not found in OncoKB were classified as "Likely Pathogenic", unlike what is described by Froyen et al [32]. Missense variants not found in OncoKB were assessed using the International Agency for Research in Cancer (IARC) TP53 database (`http://p53.iarc.fr/`)), which compiles information on human TP53 variations in relation to cancer [35], and Seshat (`http://vps338341.ovh.net/`). Variants not found in any of the listed resources were classified as "VUS".

Variants in BRCA1 and BRCA2 were evaluated as described by Froyen et al [32], meaning that clear LoF mutation were classified as "Pathogenic" and other variants were evaluated using the following databases: ARUP (`http://www.arup.utah.edu/database/BRCA/`), InterVar (`http://wintervar.wglab.org/`), ClinVar (`https://www.ncbi.nlm.nih.gov/clinvar/`), Enigma (`https://brcaexchange.org/`), and LOVD (`https://databases.lovd.nl/shared/genes`).

3.4.2 Variants located in driver genes

After classification, the variants were evaluated for being located in driver genes or not, using information about known driver genes from the Cancer Gene Census (`https://cancer.sanger.ac.uk/census`). Genes that were not previously known to be drivers, were then searched in a database developed by Dietlein et al [36], that predicts if the gene is a driver or not based on unusual nucleotide context. If the gene was found in either database as a driver for NSCLC (lung adenocarcinoma or lung squamous cell carcinoma) or lung cancer, it was denoted as a driver. Genes found in either database as a driver for other types of cancer excluding NSCLC or lung cancer and genes not found in either database were not denoted as drivers NSCLC.

3.4.3 Identification of frequently mutated genes

Based on the variant data for the cohort, a TXT file annotated by the hg38 reference genome was generated, which only included three columns: sample ID, mutated gene name and the variant types included in this project. The TXT file was then visualized in a waterfall plot using the R package GenVisR, which ranks mutated genes in

descending order of mutation frequency. Only genes with a mutation frequency over 20 % were displayed.

3.4.4 Statistical analysis

Survival was estimated using the Kaplan-Meier method and differences in overall survival between groups was assessed using log-rank tests (Mantel-Cox). Overall survival (OS) was defined as the time from the start of immunotherapy to death or to the last follow-up date. Multivariate Cox (proportional hazards) regression was conducted to adjust for patient-related factors such as age, gender, smoking history, PD-L1 expression, and TMB. Pearson's chi-squared test was used to assess the association to immunotherapy response. The data analysis was done using IBM SPSS (v.28.0.1.1).

4 Results

4.1 Clinical characteristics of NSCLC patients

From 2019 to 2021, 44 patients with NSCLC undergoing immunotherapy had tumor DNA sequenced and were included in this project. Of these 44 patients, 19 (43.2 %) were male and 25 (56.8 %) female. 11 patients (25.0 %) were current smokers, 29 (65.9 %) previous smokers, and 4 were non-smokers (9.1 %). Among the patients, 20 were aged under 70 (45.4 %) and 24 over 70 (54.5 %). The most common subtype pf NSCLC in the cohort was LUAD, accounting for 75.0 % (33 of 44) of all the cases and LUSC for 25.0 % (11 of 44). There were 8 patients (18.2 %) with stage III NSCLC and 36 (81.8 %) with stage IV. Table 2 summarizes the clinical characteristics of the patients.

Table 2: Clinical characterisation of patient cohort

	n (%)
All patients	44 (100 %)
Gender	
Male	19 (43 %)
Female	25 (57 %)
Age	
≤70	20 (45 %)
>70	24 (55 %)
Stage	
III	8 (18 %)
IV	36 (82 %)
Histology	
Adenocarcinoma	33 (75 %)
Squamous cell carcinoma	11 (25 %)
Smoking status	
Current	11 (25 %)
Previous	29 (66 %)
Never	4 (9 %)
ICI response	
Responder	25 (57 %)
Non-responder	19 (43 %)

4.2 Somatic mutation landscape in NSCLC

Based on the waterfall plot of the 44 patients in the cohort, 10 genes were found to have a gene mutation frequency above 20 % (*TP53*, *LRP1B*, *CSMD3*, *KRAS*, *KMT2C*, *FAT3*, *ADGRB3*, *TRRAP*, *SPTA1*, and *FLT1*), where *TP53* had the highest mutation frequency (29 of 44, 65.9 %) (Figure 3). The majority of patients (37 of 44, 84.0 %) had at least one mutation among of the most frequently mutated genes and the most common mutation type was missense mutations.

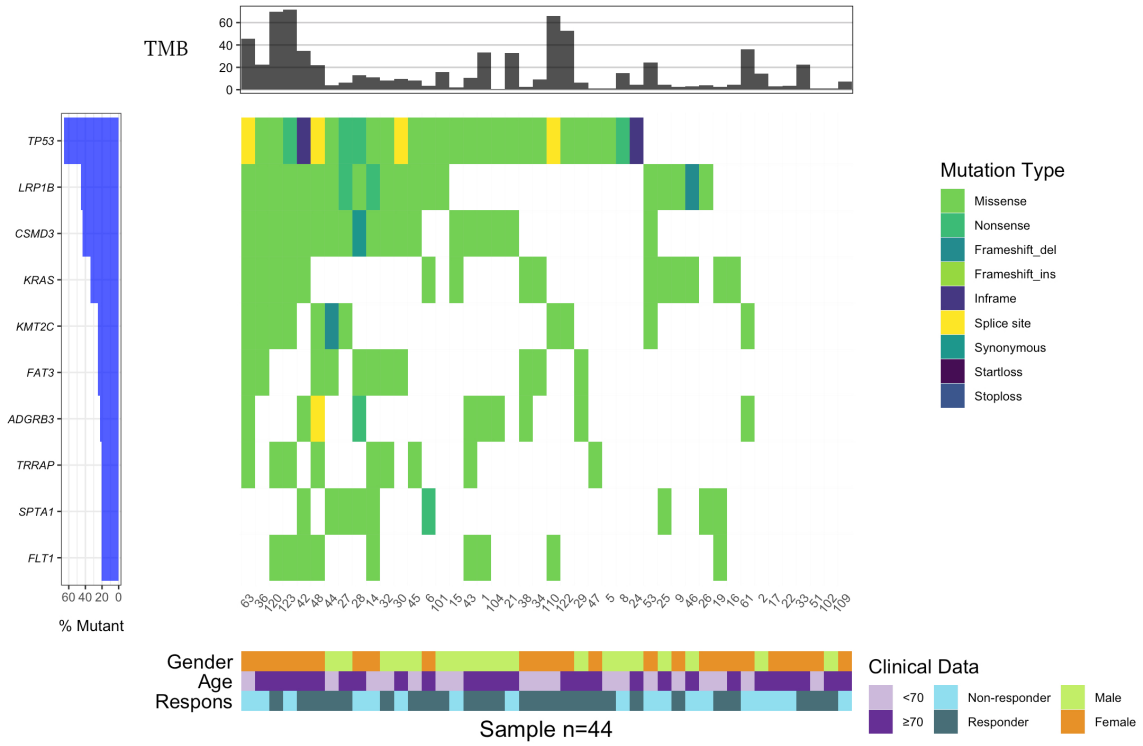


Figure 3: Waterfall plot displaying the most frequently mutated genes in 44 NSCLC patients. Genes with a mutation frequency greater than 20 % were arranged in descending order of mutation frequency (left panel) and different mutation types were represented by different colors (right panel). TMB for each patient is included in the top panel and clinical data in the lower panel.

The cohort consists of patients diagnosed with LUAD (33 of 44, 75 %) and LUSC (11 of 44, 25 %) and the most frequently mutated genes were found to differ between the diagnoses (Figure 4). Among patients with LUAD, the most mutated genes were *TP53* (18 of 33, 56 %) followed by *LRP1B* (16 of 33, 48 %), and included the gene *ATM* (8 of 33, 24 %), which was not frequently mutated among LUSC

patients (Figure 4A). For patients with LUSC, the most frequently mutated genes were, *TP53* (10 of 11, 90.9 %) followed by *CSMD3* (7 of 11, 64 %), and included genes not frequently mutated in LUAD patients such as *CUBN* (5 of 11, 45 %), *ERBB4* (4 of 11, 36 %), *SETBP1* (3 of 11, 27 %), and *PREX2* (3 of 11, 27 %) (Figure 4B).

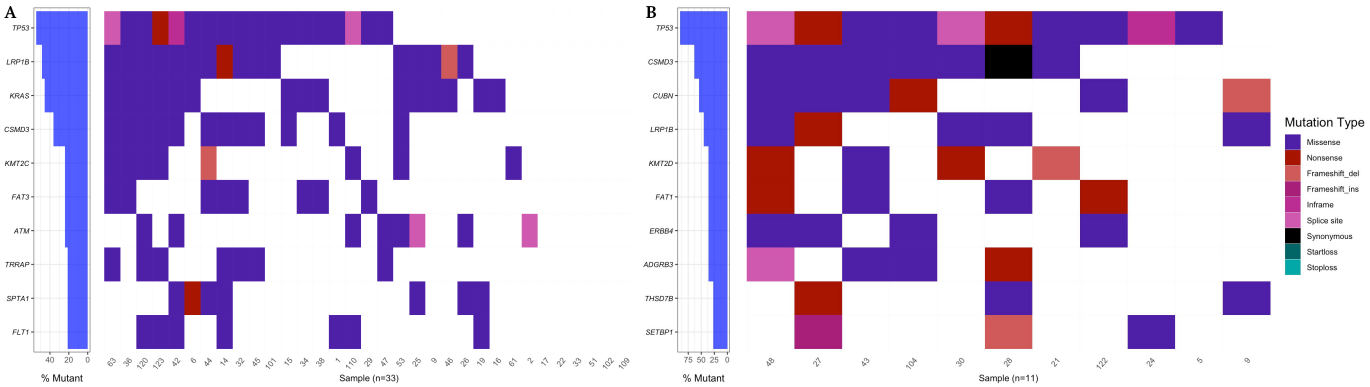


Figure 4: Waterfall plot of the genetic landscape of patients diagnosed with **(A)** LUAD and **(B)** LUSC. Mutation frequency for genes with frequency higher than 20 % are is displayed in descending order in the left panels. Mutation type is represented by different colors.

4.3 Technical analysis and classification of tumor DNA variants

In total, 2572 variants from 44 patients were evaluated in IGV after filtration using a 5 % allele frequency cut-off. Of these, 1633 variants (63.5 %) were evaluated as artefacts and discarded from further analysis. The remaining 939 variants (36.5 %) were evaluated as true and included in the final analysis, where 18 variants (1.91 %) were classified as "Pathogenic", 134 (14.3 %) as "Likely pathogenic", 755 (80.4 %) as "VUS" and 6 variants (0.64 %) were classified as "Benign" or "Likely Benign". From the scoring table containing 737 variants, 59 variants classified as "VUS" were evaluated for nucleotide conservation and physicochemical effect in Alamut, out of which 26 variants (2.77 %) were re-classified as "VUS+".

The majority of the pathogenic variants were located in *KRAS* (15 of 18, 83.3 %), and many of the likely pathogenic variants were found in *TP53* (28 of 134, 20.9 %). Variants classified as "VUS" were mainly found in *CSMD3* (30 of 755, 3.97 %), *LRP1B* (23 of 755, 3.04 %), *KMT2C* (18 of 755, 2.38 %) and *FAT3* (14 of 755, 1.85 %). Half of the variants classified as "Benign" and "Likely benign" were located in *BRCA1* and *BRCA2*. Among the genes classified as "VUS+", no gene was found in majority.

4.4 Overall survival analysis

Kaplan-Meier survival analysis was conducted for the most frequently mutated genes, which included *TP53*, *LRP1B*, *CSMD3*, *KRAS*, *FAT3*, *KMT2C ADGRB3*, *TRRAP*, *SPTA1*, and *FLT1*. Cox regression analysis was preformed for genes with significant association to survival to adjust for patient-related factors. A p-value < 0.05 was deemed significant. Pearson's chi-squared test was used to find associations between genes and immunotherapy response.

4.4.1 Individual genes

Wildtype *TP53* (15 of 44, 34.1 %) was found to be significantly associated with improved overall survival (log-rank test p=0.015, Figure 5A). The estimated mean survival was longer for wildtype *TP53* compared to mutated *TP53* (29 vs 19 months), and the estimated median survival was 24 months for mutated *TP53* and was not found for wildtype *TP53*. The association between *TP53* and survival did not remain significant when taking patient related factors into account (Cox proportional hazards regression, HR 3.23, [95 % CI, 0.802-12.924], p=0.098, Figure 5B). There was a trend towards response in patients with wildtype *TP53*, however it was not significant (Pearson's chi-squared test, p=0.450, OR=0.618, Figure 5C).

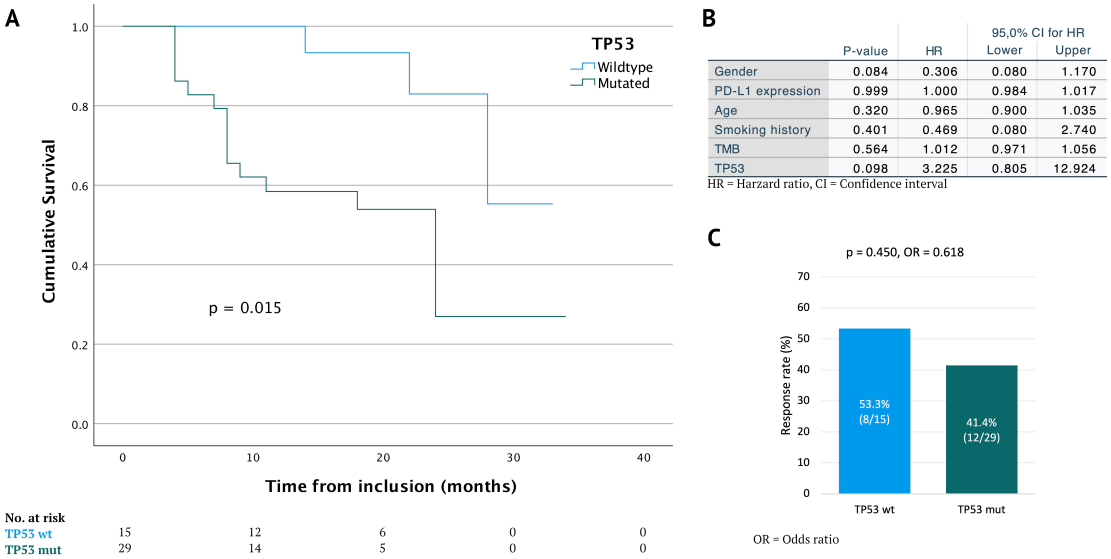


Figure 5: Association of *TP53* mutation with survival and immunotherapy response in NSCLC cohort.**(A)** Kaplan-Meier curve showing overall survival of patients with wildtype (n=15) and mutated (n=29) *TP53*. The p-value shown was determined from the log-rank test. **(B)** Cox regression analysis adjusting for gender, PD-L1 expression, age, smoking history and TMB. *TP53* wildtype used as reference. **(C)** Association between *TP53* (wildtype and mutated) and immunotherapy response was determined from a Pearson's chi-squared test.

KRAS mutation (15 of 44, 34.1 %) showed a significant association with improved overall survival (log-rank test p=0.033, Figure 6A). The estimated mean survival was shorter for wildtype *KRAS* compared to mutated (18 vs 29 months), and the estimated median survival was 24 months for wildtype *KRAS* and was not found for mutated *KRAS*. In multivariate Cox regression, mutated *KRAS* did not remain significantly associated with survival (HR 0.435, [95 % CI, 0.083-2.288], p=0.326, Figure 6B). There was a trend for immunotherapy response among patients with mutated *KRAS*, however, it was not significant (Pearson's chi-squared test, p=0.450, OR=1.629, Figure 6C).

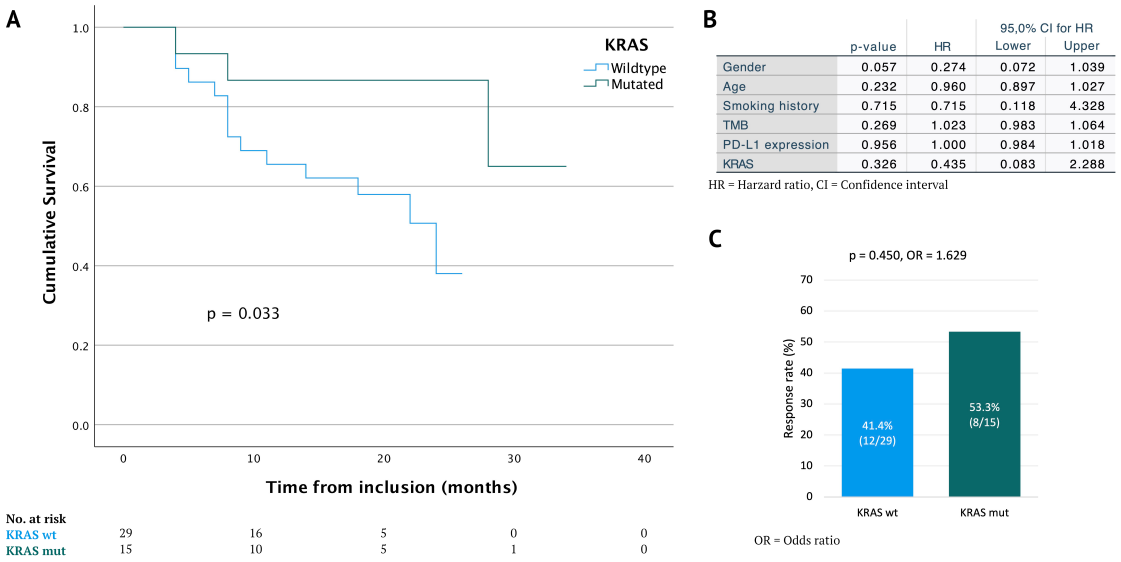


Figure 6: Association of *KRAS* mutation with survival and immunotherapy response in NSCLC cohort.**(A)** Kaplan-Meier curve showing overall survival of patients with wildtype (n=29) and mutated (n=15) *KRAS*. The p-value shown was determined from the log-rank test. **(B)** Cox regression analysis adjusting for gender, PD-L1 expression, age, smoking history and TMB. *KRAS* wildtype used as reference. **(C)** Association between *KRAS* (wildtype and mutated) and immunotherapy response was determined from Pearson's chi-squared test.

In Kaplan-Meier survival analysis, mutation in *LRP1B* (20 of 44, **45.4 %**) was significantly associated with improved overall survival (log-rank test p=0.033, Figure 7A). The estimated mean survival for wildtype *LRP1B* was 19 months compared to 29 months for mutated *LRP1B* and the estimated median survival was 24 months for wildtype *LRP1B* but was not found for mutated *LRP1B*. In multivariate Cox regression, mutated *LRP1B* did not remain significantly associated with survival (HR 0.072, [**95 %** CI, 0.114-1.098], p=0.072, Figure 7B), and there was no significant association between immunotherapy response and wildtype or mutated *LRP1B* (Pearson's chi-squared test, p=0.956, OR=0.967, Figure 7C).

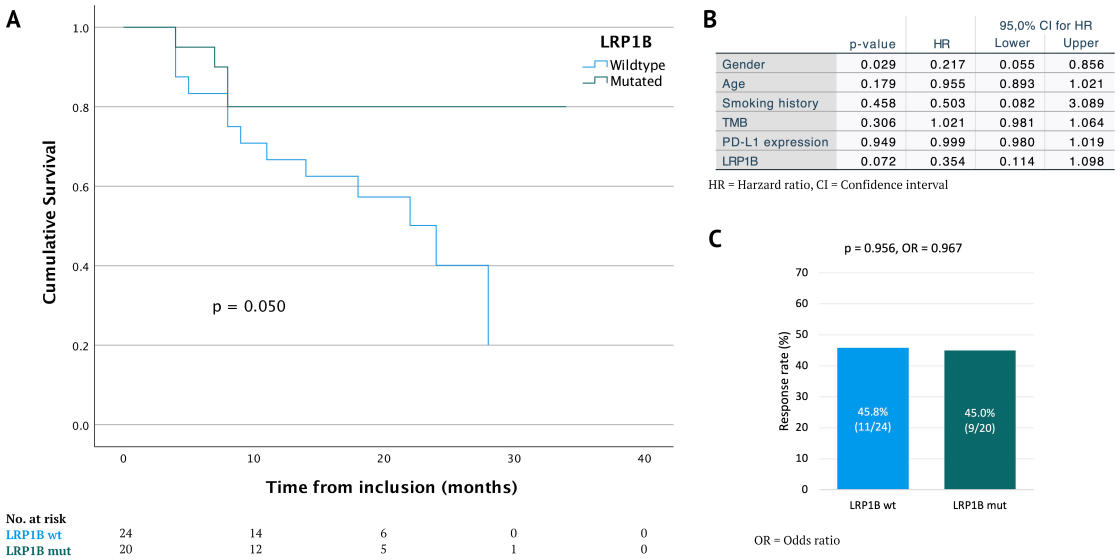


Figure 7: Association of *LRP1B* mutation with survival and immunotherapy response in NSCLC cohort.**(A)** Kaplan-Meier curve showing overall survival of patients with wildtype (n=24) and mutated (n=20) *LRP1B*. The p-value shown was determined from the log-rank test. **(B)** Cox regression analysis adjusting for gender, PD-L1 expression, age, smoking history and TMB. *LRP1B* wildtype used as reference. **(C)** Association between *LRP1B* (wildtype and mutated) and immunotherapy response was determined from Pearson's chi-squared test.

Other genes included in the Kaplan-Meier analysis were not significantly associated with survival (*CSMD3* p=0.989, *ADGRB3* p=0.730, *FAT3* p=0.828, *KMT2C* p=0.544, *FLT1* p=0.851, *SPTA1* p=0.180, *TRRAP* p=0.931, *KEAP1* p=0.647) (Appendix A, Figure A1 and A2).

4.4.2 Co-mutations

Different co-mutations of the top mutated genes in this cohort were analyzed. In Kaplan-Meier survival analysis, co-mutations between *LRP1B* and *KRAS*, and *LRP1B* and *TP53* were found to be significantly associated with survival (p=0.022, Figure 8A; p=0.036 Figure 8B). Significant association with survival was also found for patients with *TP53*, *KRAS*, and *LRP1B* co-mutation (p=0.027, Figure 8D). *TP53* and *KRAS* co-mutation was not found to be significantly associated with survival (p=0.099, Figure 8C). Other co-mutations tested were not significantly associated with survival (*TP53* and *KEAP1* p=0.833; *TP53* and *KMT2C* p=0.356; *KRAS* and *KEAP1* p=0.700; *KRAS* and *KMT2C* p=0.304) (Appendix A, Figure A3).

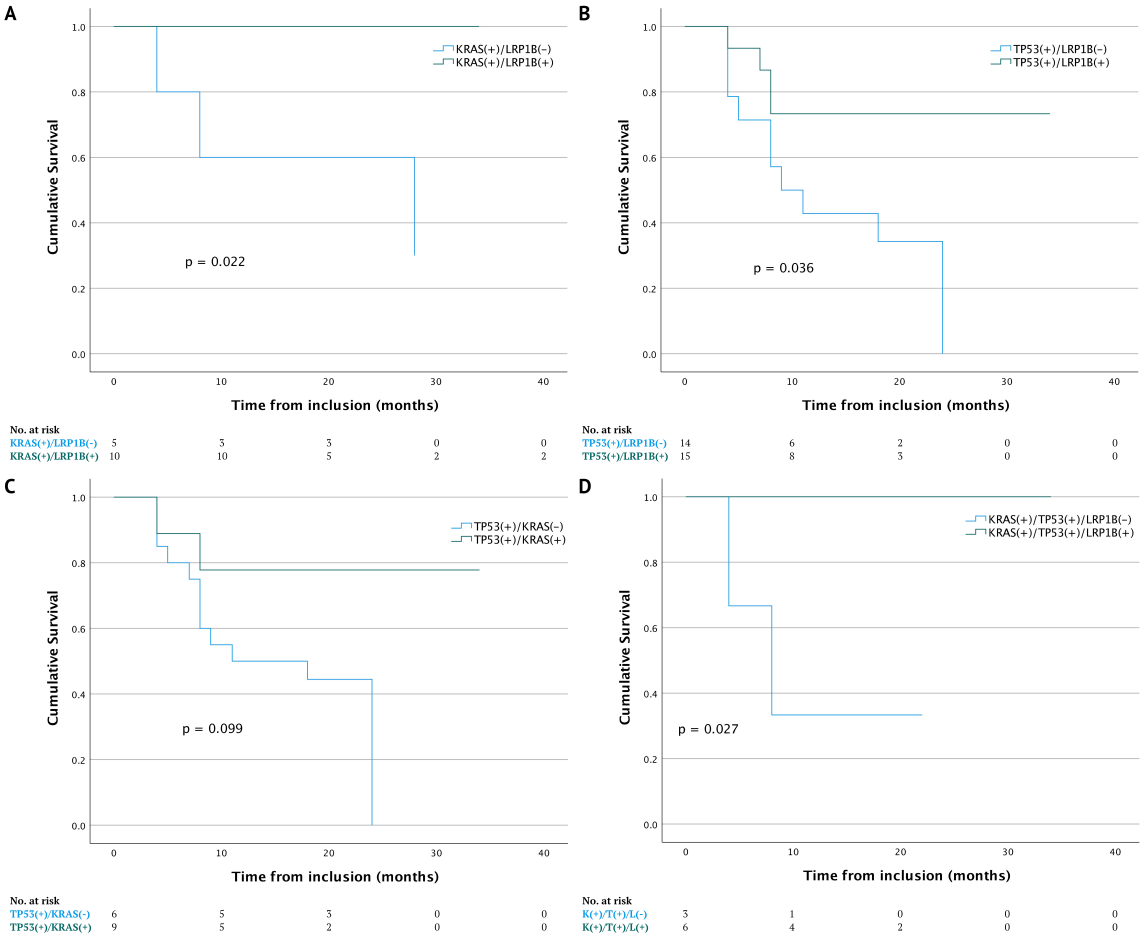


Figure 8: Kaplan-Meier curve showing overall survival of patients with **(A)** KRAS mutation (KRAS(+)) and wildtype (n=5) or mutated (n=10) LRP1B (p=0.022); **(B)** TP53 mutation (TP53(+)) and wildtype (n=14) or mutated (n=15) LRP1B (p=0.036) ,**(C)** TP53 mutation (TP53(+)) and wildtype (n=6) or mutated (n=9) KRAS (p=0.099); **(D)** TP53 mutation (TP53(+)), KRAS mutation (KRAS(+)), and wildtype (n=3) or mutated (n=6) LRP1B (p=0.027).

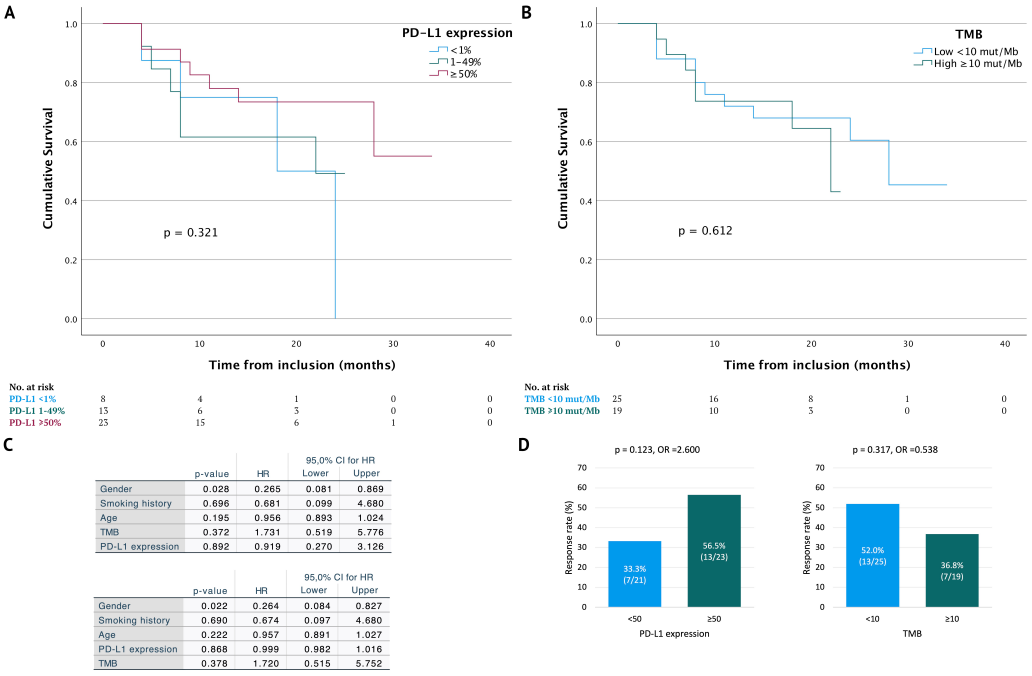
In multivariate cox regression analysis, patients with *TP53* and *LRP1B* co-mutation, and *TP53* and *KRAS* co-mutation were positively associated with survival, however, it was not significant (HR 0.322, [95 % CI, 0.084-1.226], p=0.097; HR 0.559, [95 % CI, 0.079-3.983], p=0.559). Cox regression could not be preformed on patients with *KRAS* and *LRP1B* co-mutation or *KRAS*, *LRP1B* and *TP53* co-mutation due to too few events (data not shown).

Associations between the co-mutations and immunotherapy response were made

for co-mutations with significant association with survival as well as the *TP53/KRAS* co-mutation. For the *KRAS/LRP1B* co-mutation, there was a trend for immunotherapy response among patients with the co-mutation, however it was not significant (Pearson's chi-squared test p=0.464, OR=2.250). In patients with the *TP53/LRP1B* co-mutation there was a trend for being non-responders (Pearson's chi-squared test p=0.362, OR=0.500) and among patients with the *TP53/KRAS* co-mutation there was no trend between the groups (Pearson's chi-squared test p=0.822, OR=1.200). No association with response could be made for the *TP53/KRAS/LRP1B* co-mutation due to too few cases.

4.4.3 Groups of variants

Currently, PD-L1 expression and TMB are the only predictive biomarkers used for patient selection. Kaplan-Meier analysis showed no association between PD-L1 or TMB and survival (log-rank test p=0.321, Figure 9A; log-rank test p=0.612, Figure 9B). Neither PD-L1 expression or TMB were found significantly associated with survival in multivariate cox regression analysis (HR 0.919, [95 % CI, 0.270-3.126], p=0.892, Figure 9C (top panel); HR 1.720, [95 % CI, 0.515-5.752], p=0.378, Figure 9C (lower panel)). There was a trend for immunotherapy response among patients with a PD-L1 $\geq$50 % and among patients with a TMB $<$ 10 mut/Mb, however, it was not significant (Pearson's chi-squared test p =0.123, OR=2.600; p=0.317, OR=0.538,9D)



	p-value	HR	95,0% CI for HR Lower	Upper
Gender	0.028	0.265	0.081	0.869
Smoking history	0.696	0.681	0.099	4.680
Age	0.195	0.956	0.893	1.024
TMB	0.372	1.731	0.519	5.776
PD-L1 expression	0.892	0.919	0.270	3.126

	p-value	HR	95,0% CI for HR Lower	Upper
Gender	0.022	0.264	0.084	0.827
Smoking history	0.690	0.674	0.097	4.680
Age	0.222	0.957	0.891	1.027
PD-L1 expression	0.868	0.999	0.982	1.016
TMB	0.378	1.720	0.515	5.752

Figure 9: Association of PD-L1 expression and TMB with survival and immunotherapy response. Kaplan-Meier curve showing overall survival of patients with **(A)** a PD-L1 expression of: $<1\,\%$ (n=8); $1-49\,\%$ (n=13); and $\geq 50\,\%$ (n=23), and **(B)** patients with low (<10 mut/Mb, n=25) or high ($\geq$10 mut/Mb, n=19) TMB. The p-values shown were determined from the log-rank test. **(C)** Cox regression analysis of PD-L1 expression (top panel) and TMB (lower panel). **D** Association between PD-L1 expression (left panel) and TMB (right panel) and immunotherapy response was determined from Pearson's chi-squared test.

Different groupings of the variant data was made in an effort to investigate alternative biomarkers in regards to TMB. In this project, TMB included all non-synonymous mutations in the tumor genome, regardless of their classification. Initially, the survival based on the number of variants in each classification was analyzed. Kaplan-Meier analysis showed that the number of VUS variants were less associated with survival compared to the other classifications (number of pathogenic variants, log-rank test p=0.110, Figure 10A; number of likely pathogenic variants, log-rank test p=0.004, Figure 10B; number of "VUS+", log-rank test p=0.176, Figure 10C; number of VUS, log-rank test p=0.481, Figure 10D).

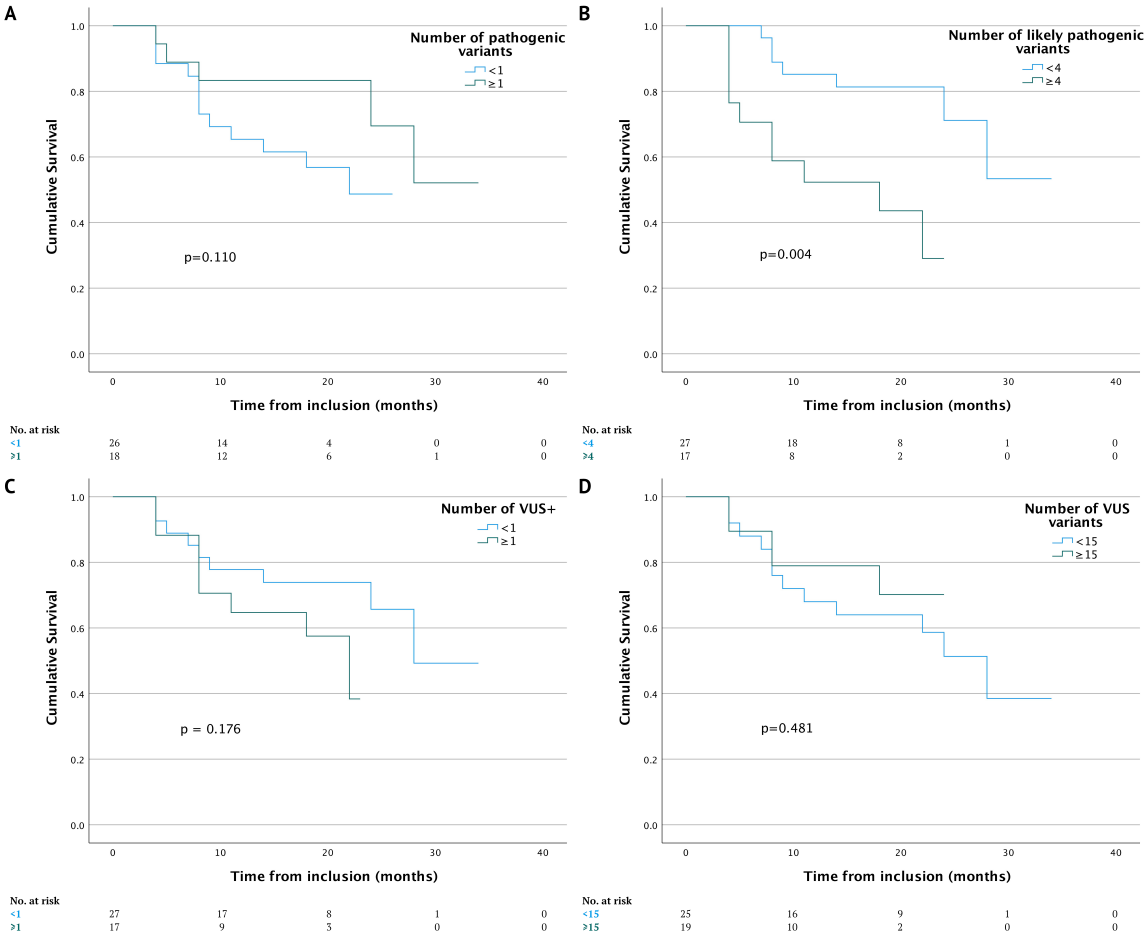


Figure 10: Kaplan-Meier analysis for patients with **(A)** a high (≥1, n=18) or low (<1, n=26) number of pathogenic variants; **(B)** a high (≥4, n=17) or low (<4, n=27) number of likely pathogenic variants; **(C)** a high (≥1, n=17) or low (<1, n=27) number of "VUS+"; **(D)** a high (≥15, n=19) or low (<15, n=25) number of VUS.

The classified variants were then grouped together in different ways, where the combination of all classified variants, in a similar way as TMB, showed no association with survival or immunotherapy response (log-rank test p=0.684; Pearson's chi-squared test p=0.956, OR=0.976, Appendix A, Figure A4A and B). A "refined" version of TMB was then investigated, only containing the total number of pathogenic and likely pathogenic variants in each patient. Kaplan-Meier showed no significant association with survival (log-rank test p=0.079, Figure 11A), although there was a trend for improved survival among patients with less than five of these variants. Cox regression analysis showed no significant association with survival (p=0.080, Fig-

ure 11B). Having less than five variants classified as likely pathogenic and pathogenic showed a significant association with immunotherapy response (Pearson's chi-squared test p=0.009, OD=0.131, Figure 11C).

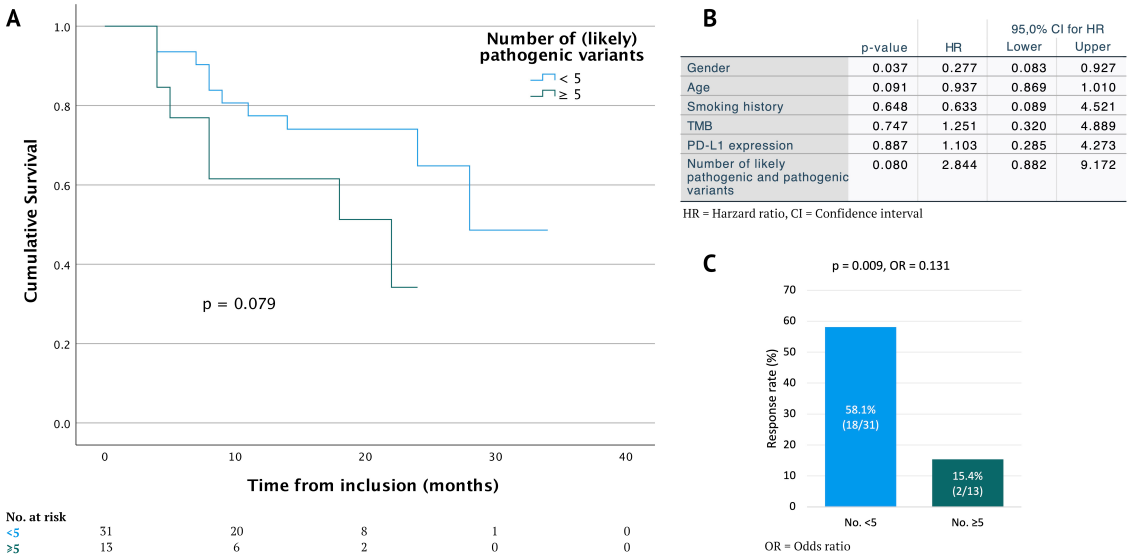


Figure 11: Kaplan-Meier curve showing overall survival for patients with a high (≥5, n=13) or low (<5, n=31) number of variants classified as likely pathogenic and pathogenic.**(B)** Cox regression analysis adjusting gender, PD-L1 expression, age, smoking history and TMB. A number of variants less than five was used as reference. **(C)** Association between the number of variants and immunotherapy response was determined from Pearson's chi-squared test.

The number of variants classified as "VUS+" were then added to the likely pathogenic and pathogenic variants. In Kaplan-Meier analysis, having less than 5 variants classified as likely pathogenic, pathogenic or "VUS+" was significantly associated with improved survival (log-rank test p=0.032, Figure 12A). Having less than five of these variants remained significantly associated with survival in cox regression analysis (HR 3.562, [95 % CI, 1.088-11.663], p=0.036, Figure 12B) and showed a significant association with immunotherapy response (Pearson's chi-squared test, p=0.020, OR=0.212, Figure 12C).

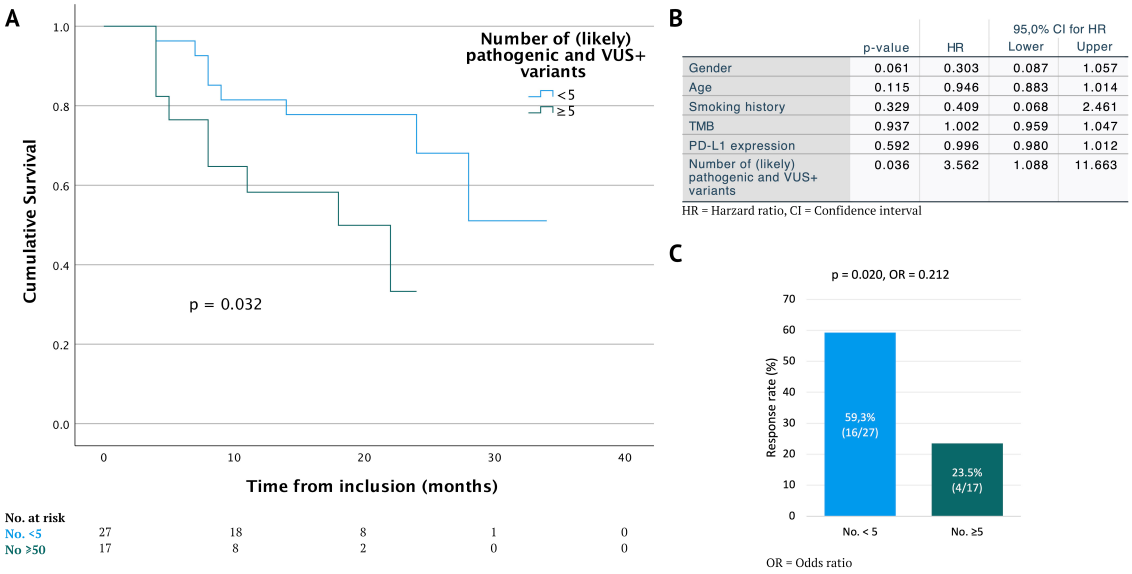


Figure 12: Association of the number of variants classified as likely pathogenic, pathogenic and "VUS+" with survival and immunotherapy response in NSCLC cohort.**(A)** Kaplan-Meier curve showing overall survival of patients with a low (<5, n=27) and high (≥5, n=17) number of these variants. The p-value shown was determined from the log-rank test. **(B)** Cox regression analysis adjusting for gender, PD-L1 expression, age, smoking history and TMB. Number of variants <5 was used as reference. **(C)** Association between the number of likely pathogenic, pathogenic and "VUS+" variants and immunotherapy response was determined from Pearson's chi-squared test.

Overall survival was also compared based on the number of variants in driver genes but no significant association to survival was found (Appendix A, Figure A5).

5 Discussion

The aim of this project was to characterize the genetic landscape of NSCLC patients in order to find genetic biomarkers associated with immunotherapy response. This was done by starting with a comprehensive evaluation and interpretation of DNA variants from sequencing data, followed by an analysis of the most frequently mutated genes. Survival analysis was then used to associate mutations in individual genes, co-mutations, and different groups of variant data to survival and immunotherapy response.

5.1 Evaluation and classification of variants

Initially, the DNA variants were manually evaluated in IGV to remove artefacts. This is considered good practice when working with FFPE-material since these samples are often of poor quality and the variant data often contain many sequencing errors due to the formalin-fixation [37]. The variants were then classified using a workflow developed by Froyen et al. [32], which is based on the ACMG AMP guidelines that details standards and guidelines for classification of germline variants [33], but contains modifications to better suit classification of somatic variants. In this project, the workflow from Froyen et al. was further modified to include more prediction tools, information about predicted classification from both VarSome (`https://varsome.com/`) and ClinVar (`https://www.ncbi.nlm.nih.gov/clinvar/`), and information from Alamut to differentiate variants classified as "VUS". The purpose of these modifications was to develop the classification of somatic variants further.

5.2 Characterization of genetic landscape of NSCLC patients

The genetic landscape of the patients was analyzed using a waterfall plot displaying the most frequently mutated genes in the cohort. Exploring the top frequently mutated genes is important to find the genes that are the most relevant to investigate as biomarkers, and in this case, it was important to find genes that had a high enough mutation frequency to get decent results since the cohort is relatively small. We found that the top frequently mutated genes for the cohort were *TP53*, *LRP1B*, *CSMD3*, and *KRAS*, and when the cohort was divided based on histologic subtype, we observed differences in the top mutated genes between patients diagnosed with LUAD and LUSC. For patients with LUAD, the top mutated genes were similar to what we found for the whole cohort and included *TP53*, *LRP1B*, *KRAS*, and *FAT3*. For patients with LUSC, *TP53* and *LRP1B* were among the top frequently mutated genes,

but we also observed genes not found in patients with LUAD or in the whole cohort, including *KMT2D*, *ERBB4* and *SETBP1*. The top mutated genes for the whole cohort and for the histologic subtypes are consistent with previous research, where *TP53* has been consistently reported as the highest mutated gene for many types of cancers, including both LUAD and LUSC [38][39][40][41]. Additionally, *KRAS*, *FAT3*, and *LRP1B* have all been found to be frequently mutated in LUAD patients, which is consistent with the findings in this project [41][38][42]. For patients with LUSC, some of the most commonly reported mutations are *TP53*, *ERBB4*, *KEAP1*, and *KRAS* [43], out of which only *KEAP1* was not found frequently mutated.

5.3 Genes associated with survival and response

To find genetic biomarkers predicting immunotherapy response in individual genes, we analyzed the association of mutated genes with survival and response. Survival analysis of the most frequently mutated genes showed that patients with mutated *KRAS* where more associated with immunotherapy response than patients with wild-type *KRAS* and was significantly associated with survival. This is in agreement with previous studies, suggesting that *KRAS* is a potential biomarker for immunotherapy response [44][45]. Additionally, mutated *LRP1B* was found significantly associated with survival, which is consistent with other studies [42][46], however, no association with response was found, indicating that mutated *LRP1B* alone is not a potential biomarker. Wildtype *TP53* was also significantly associated with survival and showed a trend to associate with immunotherapy response. In the literature, there are contradicting reports on *TP53*, where some studies indicate that wildtype *TP53* is associated with survival and better immunotherapy outcome [47], while others report that patients with mutated *TP53* displays improved survival and preferable immunotherapy response [45]. The role of *TP53* as a predictive biomarker therefore remains unclear.

5.4 Co-mutations associated with survival and response

In addition to looking at mutations in individual genes, co-mutations of the top frequently mutated genes were analyzed. A few of the these co-mutations showed significant association with survival. The co-mutations *TP53/LRP1B* and *KRAS/LRP1B* have not been mentioned much in the literature, however, both co-mutations were associated with survival, where patients with the *KRAS/LRP1B* were more likely to be responders and patients with the *TP53/LRP1B* more likely to be non-responders. This suggests that the *KRAS/LRP1B* could be a potential biomarker, however, the

sample size in this case is small and the association is not significant, meaning the results might not be representative of a larger population. The *TP53/LRP1B* co-mutation is likely not a predictive biomarker.

The *TP53/KRAS* co-mutation showed a trend towards improved survival. This is in agreement with the literature, where many studies have reported improved survival in patients harbouring the *TP53/KRAS* co-mutation compared to patients with no or single mutation in either of the genes [47][48]. However, in contradiction to previous studies, there was no association with response among patients with the *TP53/KRAS* co-mutation. One possible explanation for the discrepancy is the small study cohort and the small number of patients with this co-mutation. Furthermore, the *TP53/KRAS/LRP1B* co-mutation was also significantly associated with improved survival, however, due to the small sample size, no real conclusions can be drawn from the results.

5.5 Groups of variants associated with survival and response

Currently, only PD-L1 expression and TMB are used to guide treatment decisions among patients with NSCLC. It was therefore of interest to see the level of association of these biomarkers with survival and immunotherapy response in this cohort. We found that neither high or low TMB was associated with survival, nor was PD-L1 expression. There was an association, though not significant, with immunotherapy response among patients with a high PD-L1 expression, which is consistent with previous research [18] [19]. There was also an association for immunotherapy response among patients with a low TMB, which contradicts some reports [21][22][23], but there are also reports that TMB is not connected to response [49], suggesting that TMB in not a sufficient predictor of benefit from immunotherapy.

Different groups of the variant data were analyzed in an attempt to find an alternative biomarker to TMB. Unlike TMB, where a higher number of mutations have been associated with improved overall survival, having a low number of variants classified as likely pathogenic, pathogenic and "VUS+" was found to be significantly associated with survival, even after taking patients-related factors into account. There was also a significant association with immunotherapy response among patients with a low number of these variants. Similar results were found when looking only at the number of likely pathogenic and pathogenic variants, although no significant association to survival was found. Additionally, we found that looking at all variants, regardless of classification, and only variants classified as "VUS" showed no significant association with survival or immunotherapy response. This suggests, in contrary to TMB, that the classification of the variants is an important factor for predicting re-

sponse, more so than the total number of unclassified variants, and that the number of likely pathogenic, pathogenic and "VUS+" variants is a potential biomarker for response. However, it should also be considered that the underlying co-mutations among the pathogenic, likely pathogenic and "VUS+" variants may play a role in these results and not just the classification itself. We also found that generally, not all driver genes seem to be associated with survival or immunotherapy response. This conclusion is in agreement with previous results, since some of the pathogenic variants are also drivers.

Altogether, the results suggest that some individual genes and co-mutations can be more linked to survival and immunotherapy response than others, and could therefore potential biomarkers of ICI response. It should be noted, however, that these genes and co-mutations were less associated with response than PD-L1 expression, suggesting that they are not sufficient enough to replace already established biomarkers. The results also showed that patients with variants classified as likely pathogenic, pathogenic and "VUS+" in certain genes or in combination of genes seemed to respond to immunotherapy. This suggests that a "refined" TMB that considers the classification of variants may provide a more effective biomarker than the traditional TMB.

5.6 Limitations

One of the main limitations in this study was the relatively small cohort, which makes it difficult to obtain significant results that are representative for a larger population. Another limitation is that some of the included patients have had immunotherapy in combination with chemotherapy. This opens the possibility of observing the effects of chemotherapy instead of only immunotherapy, however, the cohort would be substantially smaller if only patients with monotherapy were included. An additional limitation is that we only looked at SNP/indels and not, for example, copy number variants, structural variants or fusion genes, which can also be important.

5.7 Future prospects and societal impact

Providing precision medicine tailored to individuals is a big the goal within healthcare, and the identification of genetic biomarkers predicting immunotherapy response is a major step towards this aim. In the future, combinations of many different types of biomarkers, including genetic, immunologic and tumor-derived biomarkers in the blood, may provide a clearer picture of what makes a patient respond or not respond to immunotherapy.

6 Conclusions

In conclusion, this project provides insight into the genetic landscape of NSCLC patients and the immunotherapeutic implications of mutations and co-mutations of frequently mutated genes. This project provides evidence that mutated *KRAS*, co-mutation of *KRAS*/*LRP1B*, and most importantly, counting the number of variants based on classification are potential biomarkers for predicting immunotherapy response among NSCLC patients.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, *et al.*, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021. DOI: 10.3322/caac.21660.

[2] A. Addeo, A. Passaro, U. Malapelle, G. L. Banna, V. Subbiah, and A. Friedlaender, "Immunotherapy in non-small cell lung cancer harbouring driver mutations," *Cancer treatment reviews*, vol. 96, 2021. DOI: 10.1016/j.ctrv.2021.102179.

[3] C. Allemani, T. Matsuda, V. Di Carlo, *et al.*, "Global surveillance of trends in cancer survival 2000–14 (concord-3): Analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries," *The Lancet*, vol. 391, no. 10125, pp. 1023–1075, 2018. DOI: 10.1016/S0140-6736(17)33326-3.

[4] A. Hallqvist, A. Rohlin, and S. Raghavan, "Immune checkpoint blockade and biomarkers of clinical response in non–small cell lung cancer," *Scandinavian Journal of Immunology*, vol. 92, no. 6, e12980, 2020. DOI: 10.1111/sji.12980.

[5] Swedish lung cancer registry. (2020), [Online]. Available: https://cancercentrum.se/vast/cancerdiagnoser/lunga-och-lungsack/kvalitetsregister (visited on 02/02/2022).

[6] A. Marabelle, M. Fakih, J. Lopez, *et al.*, "Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: Prospective biomarker analysis of the multicohort, open-label, phase 2 keynote-158 study," *The Lancet Oncology*, vol. 21, no. 10, pp. 1353–1365, 2020. DOI: 10.1016/S1470-2045(20)30445-9.

[7] D. Spigel, F. De Marinis, G. Giaccone, *et al.*, "Impower110: Interim overall survival (os) analysis of a phase iii study of atezolizumab (atezo) vs platinum-based chemotherapy (chemo) as first-line (1l) treatment (tx) in pd-l1–selected nsclc," *Annals of Oncology*, vol. 30, p. v915, 2019. DOI: 10.1093/annonc/mdz293.

[8] M. D. Hellmann, T.-E. Ciuleanu, A. Pluzanski, *et al.*, "Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden," *New England Journal of Medicine*, vol. 378, no. 22, pp. 2093–2104, 2018. DOI: 10.1056/NEJMoa1801946.

[9] K. Suresh, J. Naidoo, C. T. Lin, and S. Danoff, "Immune checkpoint immunotherapy for non-small cell lung cancer: Benefits and pulmonary toxicities," *Chest*, vol. 154, no. 6, pp. 1416–1423, 2018. DOI: 10.1016/j.chest.2018.08.1048.

[10] M. Mencoboni, M. Ceppi, M. Bruzzone, *et al.*, "Effectiveness and safety of immune checkpoint inhibitors for patients with advanced non small-cell lung cancer in real-world: Review and meta-analysis," *Cancers*, vol. 13, no. 6, p. 1388, 2021. DOI: 10.3390/cancers13061388.

[11] W. Mu, L. Jiang, Y. Shi, *et al.*, "Non-invasive measurement of pd-l1 status and prediction of immunotherapy response using deep learning of pet/ct images," *Journal for Immunotherapy of Cancer*, vol. 9, no. 6, 2021. DOI: 10.1136/jitc-2020-002118.

[12] L. Wang, Y. Hu, S. Wang, J. Shen, and X. Wang, "Biomarkers of immunotherapy in non-small cell lung cancer," *Oncology letters*, vol. 20, no. 5, pp. 1–1, 2020. DOI: 10.3892/ol.2020.11999.

[13] L. Dyck and K. H. Mills, "Immune checkpoints and their inhibition in cancer and infectious diseases," *European journal of immunology*, vol. 47, no. 5, pp. 765–779, 2017. DOI: 10.1002/eji.201646875.

[14] X. Wang, F. Teng, L. Kong, and J. Yu, "Pd-l1 expression in human cancers and its association with clinical outcomes," *OncoTargets and therapy*, vol. 9, p. 5023, 2016. DOI: 10.2147/OTT.S105862.

[15] B. D. W. Group, A. J. Atkinson Jr, W. A. Colburn, *et al.*, "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework," *Clinical pharmacology & therapeutics*, vol. 69, no. 3, pp. 89–95, 2001. DOI: 10.1067/mcp.2001.113989.

[16] K. Takada, T. Okamoto, G. Toyokawa, *et al.*, "The expression of pd-l1 protein as a prognostic factor in lung squamous cell carcinoma," *Lung cancer*, vol. 104, pp. 7–15, 2017. DOI: 10.1016/j.lungcan.2016.12.006.

[17] H. Rizvi, F. Sanchez-Vega, K. La, *et al.*, "Molecular determinants of response to anti–programmed cell death (pd)-1 and anti–programmed death-ligand 1 (pd-l1) blockade in patients with non–small-cell lung cancer profiled with targeted next-generation sequencing," *Journal of Clinical Oncology*, vol. 36, no. 7, p. 633, 2018. DOI: 10.1200/JCO.2017.75.3384.

[18] A. Rittmeyer, F. Barlesi, D. Waterkamp, *et al.*, "Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (oak): A phase 3, open-label, multicentre randomised controlled trial," *The Lancet*, vol. 389, no. 10066, pp. 255–265, 2017. DOI: 10.1016/S0140-6736(16)32517-X.

[19] S. L. Topalian, F. S. Hodi, J. R. Brahmer, *et al.*, "Safety, activity, and immune correlates of anti–pd-1 antibody in cancer," *New England Journal of Medicine*, vol. 366, no. 26, pp. 2443–2454, 2012. DOI: 10.1056/NEJMoa1200690.

[20] M. Reck, D. Rodrıguez-Abreu, A. G. Robinson, *et al.*, "Pembrolizumab versus chemotherapy for pd-l1–positive non–small-cell lung cancer," *N engl J med*, vol. 375, pp. 1823–1833, 2016. DOI: 10.1056/NEJMoa1606774.

[21] G. Rossi, A. Russo, M. Tagliamento, *et al.*, "Precision medicine for nsclc in the era of immunotherapy: New biomarkers to select the most suitable treatment or the most suitable patient," *Cancers*, vol. 12, no. 5, p. 1125, 2020. DOI: 10.3390/cancers12051125.

[22] M. D. Hellmann, M. K. Callahan, M. M. Awad, *et al.*, "Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer," *Cancer cell*, vol. 33, no. 5, pp. 853–861, 2018. DOI: 10.1016/j.ccell.2018.04.001.

[23] M. D. Hellmann, T. Nathanson, H. Rizvi, *et al.*, "Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer," *Cancer cell*, vol. 33, no. 5, pp. 843–852, 2018. DOI: 10.1016/j.ccell.2018.04.005.

[24] G. T. Gibney, L. M. Weiner, and M. B. Atkins, "Predictive biomarkers for checkpoint inhibitor-based immunotherapy," *The Lancet Oncology*, vol. 17, no. 12, e542–e551, 2016. DOI: 10.1016/S1470-2045(16)30406-5.

[25] R. Büttner, J. W. Longshore, F. López-Rıos, *et al.*, "Implementing tmb measurement in clinical practice: Considerations on assay requirements," *ESMO open*, vol. 4, no. 1, e000442, 2019. DOI: 10.1136/esmoopen-2018-000442.

[26] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, "Survival analysis part i: Basic concepts and first analyses," *British journal of cancer*, vol. 89, no. 2, pp. 232–238, 2003. DOI: 10.1038/sj.bjc.6601118.

[27] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival analysis part ii: Multivariate data analysis–an introduction to concepts and methods," *British journal of cancer*, vol. 89, no. 3, pp. 431–436, 2003. DOI: 10.1038/sj.bjc.6601119.

[28] M. Szumilas, "Explaining odds ratios," *Journal of the Canadian academy of child and adolescent psychiatry*, vol. 19, no. 3, p. 227, 2010.

[29] National Human Genome Research Institute. "Talking glossary of genetic terms: Nih." (2022), [Online]. Available: https://www.genome.gov/genetics-glossary (visited on 05/23/2022).

[30] National Cancer Institute. "Nci dictionary of genetics terms." (2022), [Online]. Available: https://www.cancer.gov/publications/dictionaries/genetics-dictionary (visited on 05/23/2022).

[31] National Cancer Institute, "Nci dictionary of cancer terms." (2022), [Online]. Available: https://www.cancer.gov/publications/dictionaries/cancer-terms (visited on 05/23/2022).

[32] G. Froyen, M. Le Mercier, E. Lierman, *et al.*, "Standardization of somatic variant classifications in solid and haematological tumours by a two-level approach of biological and clinical classes: An initiative of the belgian compermed expert panel," *Cancers*, vol. 11, no. 12, p. 2030, 2019. DOI: 10.3390/cancers11122030.

[33] S. Richards, N. Aziz, S. Bale, *et al.*, "Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology," *Genetics in medicine*, vol. 17, no. 5, pp. 405–423, 2015. DOI: 10.1038/gim.2015.30.

[34] J. T. Robinson, H. Thorvaldsdóttir, A. M. Wenger, A. Zehir, and J. P. Mesirov, "Variant review with the integrative genomics viewer," *Cancer research*, vol. 77, no. 21, e31–e34, 2017. DOI: 10.1158/0008-5472.CAN-17-0337.

[35] L. Bouaoun, D. Sonkin, M. Ardin, *et al.*, "Tp53 variations in human cancers: New lessons from the iarc tp53 database and genomics data," *Human mutation*, vol. 37, no. 9, pp. 865–876, 2016. DOI: 10.1002/humu.23035.

[36] F. Dietlein, D. Weghorn, A. Taylor-Weiner, *et al.*, "Identification of cancer driver genes based on nucleotide context," *Nature genetics*, vol. 52, no. 2, pp. 208–218, 2020. DOI: 10.1038/s41588-019-0572-y.

[37] S. Q. Wong, J. Li, A. Y. Tan, *et al.*, "Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing," *BMC medical genomics*, vol. 7, no. 1, pp. 1–10, 2014. DOI: 10.1186/1755-8794-7-23..

[38] M. Zhu, L. Zhang, H. Cui, *et al.*, "Co-mutation of fat3 and lrp1b in lung adenocarcinoma defines a unique subset correlated with the efficacy of immunotherapy," *Frontiers in immunology*, vol. 12, 2021. DOI: 10.3389/fimmu.2021.800951.

[39] J. Gao, B. A. Aksoy, U. Dogrusoz, *et al.*, "Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal," *Science signaling*, vol. 6, no. 269, pl1–pl1, 2013. DOI: 10.1126/scisignal.2004088.

[40] C. Gao, H. Li, C. Liu, *et al.*, "Tumor mutation burden and immune invasion characteristics in triple negative breast cancer: Genome high-throughput data analysis," *Frontiers in Immunology*, vol. 12, p. 1328, 2021. DOI: 10.3389/fimmu.2021.650491.

[41] J. D. Campbell, A. Alexandrov, J. Kim, *et al.*, "Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas," *Nature genetics*, vol. 48, no. 6, pp. 607–616, 2016. DOI: 10.1038/ng.3564.

[42] H. Chen, W. Chong, Q. Wu, Y. Yao, M. Mao, and X. Wang, "Association of lrp1b mutation with tumor mutation burden and outcomes in melanoma and non-small cell lung cancer patients treated with immune check-point blockades," *Frontiers in immunology*, vol. 10, p. 1113, 2019. DOI: 10.3389/fimmu.2019.01113.

[43] R. S. Heist, L. V. Sequist, and J. A. Engelman, "Genetic changes in squamous cell lung cancer: A review," *Journal of Thoracic Oncology*, vol. 7, no. 5, pp. 924–933, 2012. DOI: 10.1097/JTO.0b013e31824cc334.

[44] E. A. Eklund, C. Wiel, H. Fagman, *et al.*, "Kras mutations impact clinical outcome in metastatic non-small cell lung cancer," *Cancers*, vol. 14, no. 9, p. 2063, 2022. DOI: 10.3390/cancers14092063.

[45] Z.-Y. Dong, W.-Z. Zhong, X.-C. Zhang, *et al.*, "Potential predictive value of tp53 and kras mutation status for response to pd-1 blockade immunotherapy in lung adenocarcinoma," *Clinical cancer research*, vol. 23, no. 12, pp. 3012–3024, 2017. DOI: 10.1158/1078-0432.CCR-16-2554.

[46] B. El Osta, M. Behera, S. Kim, *et al.*, "Characteristics and outcomes of patients with metastatic kras-mutant lung adenocarcinomas: The lung cancer mutation consortium experience," *Journal of thoracic oncology*, vol. 14, no. 5, pp. 876–889, 2019. DOI: 10.1016/j.jtho.2019.01.020.

[47] S. Assoun, N. Theou-Anton, M. Nguenang, *et al.*, "Association of tp53 mutations with response and longer survival under immune checkpoint inhibitors in advanced non-small-cell lung cancer," *Lung Cancer*, vol. 132, pp. 65–71, 2019. DOI: 10.1016/j.lungcan.2019.04.005.

[48] F. Zhang, J. Wang, Y. Xu, *et al.*, "Co-occurring genomic alterations and immunotherapy efficacy in nsclc," *NPJ Precision Oncology*, vol. 6, no. 1, pp. 1–12, 2022. DOI: 10.1038/s41698-021-00243-7.

[49] L. Paz-Ares, C. Langer, S. Novello, *et al.*, "Pembrolizumab (pembro) plus platinum-based chemotherapy (chemo) for metastatic nsclc: Tissue tmb (ttmb) and outcomes in keynote-021, 189, and 407," *Annals of Oncology*, vol. 30, pp. v917–v918, 2019. DOI: 10.1093/annonc/mdz394.

Appendix A Kaplan-Meier curves

Kaplan-Meier analysis for patients with wildtype or mutated; *CSMD3* (p=0.989), *ADGRB3* (p=0.730), *FAT3* (p=0.828) and *KMT2C* (p=0.544), showed no significant association with overall survival (Figure A1).

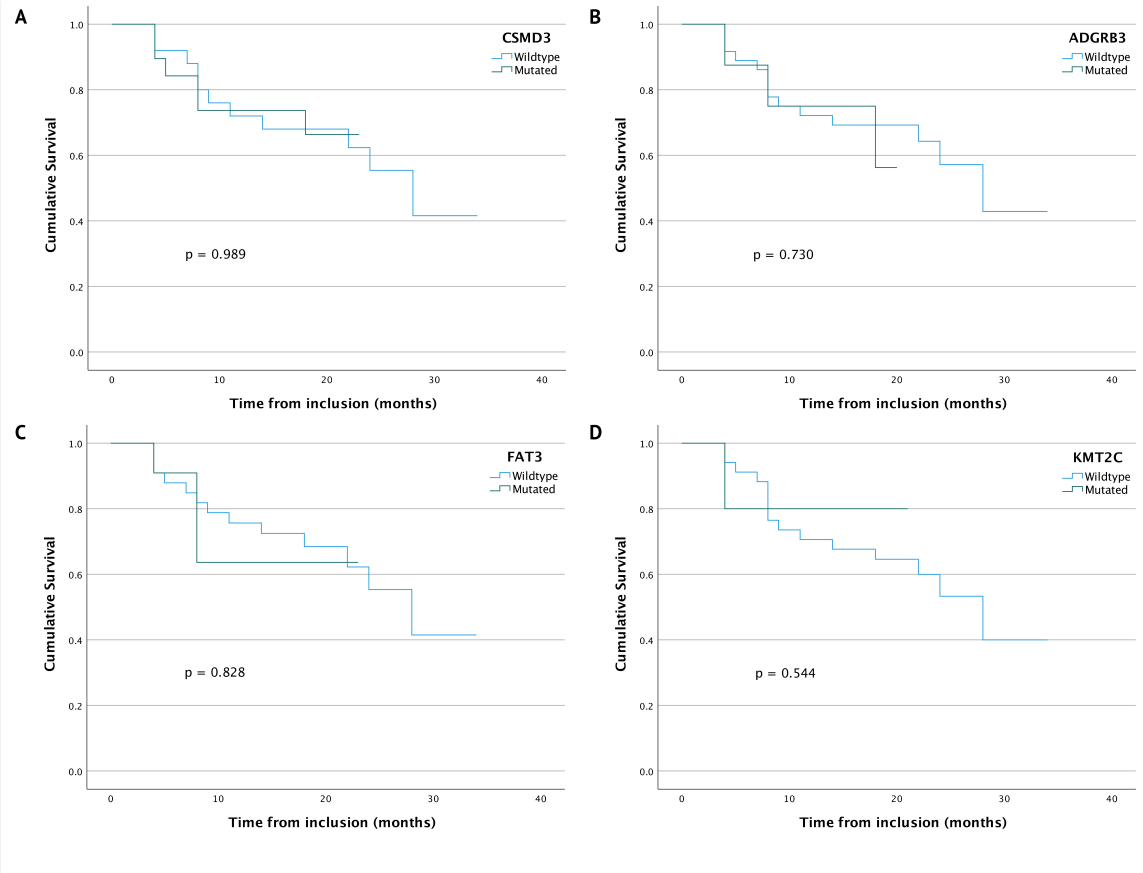


Figure A1: Kaplan-Meier curve showing overall survival for patients with **(A)** wildtype (n=25) and mutated (n=19) *CSMD3* **(B)** wildtype (n=36) and mutated (n=8) *ADGRB3* **(C)** wildtype (n=33) and mutated (n=11) *FAT3* **(D)** wildtype (n=34) and mutated (n=10) *KMT2C*

Kaplan-Meier analysis for patients with wildtype or mutated; *FLT1* (p=0.851), *SPTA1* (p=0.180), *TRRAP* (p=0.931) and *KEAP1* (p=0.647), showed no significant association with overall survival (Figure A2).

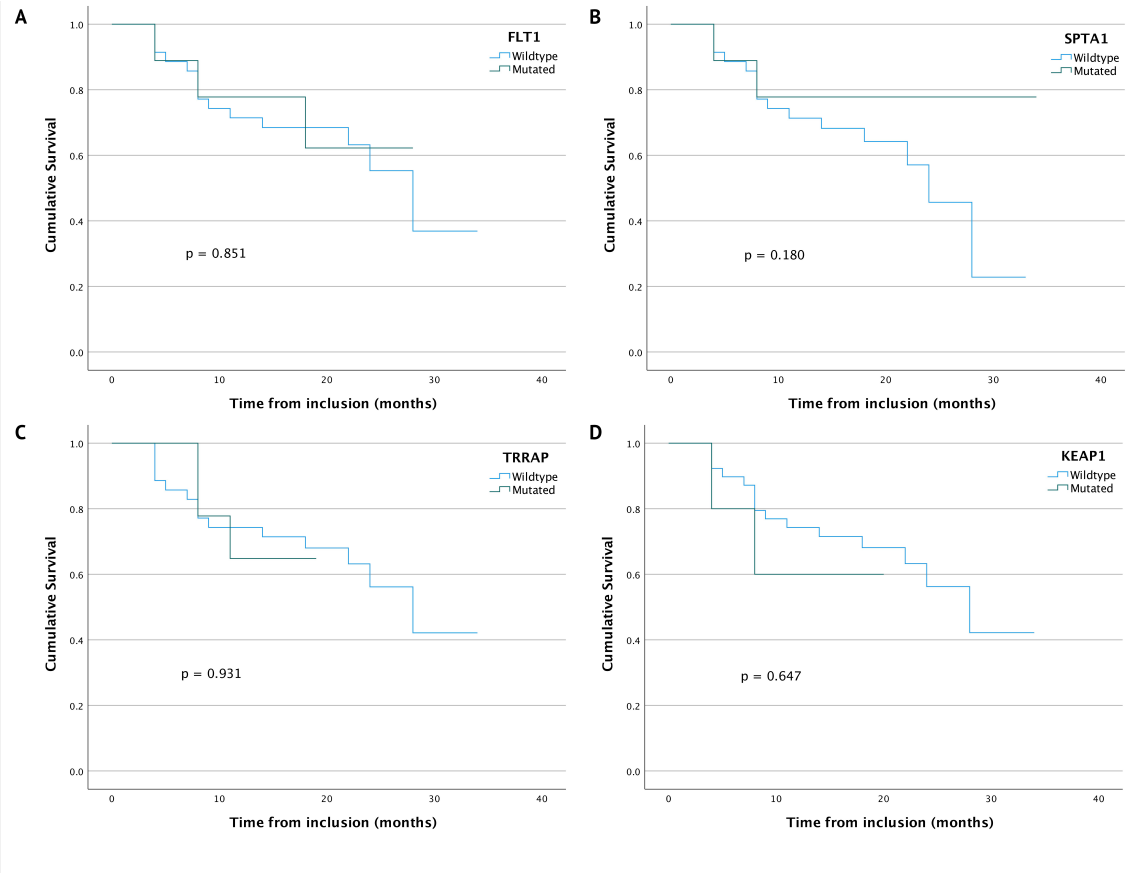


Figure A2: Kaplan-Meier curve showing overall survival for patients with **(A)** wildtype (n=35) and mutated (n=9) *FLT1* **(B)** wildtype (n=35) and mutated (n=9) *SPTA1* **(C)** wildtype (n=35) and mutated (n=9) *TRRAP* **(D)** wildtype (n=39) and mutated (n=5) *KEAP1*

Kaplan-Meier analysis was also conducted on different co-mutations, which were not significantly associated with survival (*TP53* and *KEAP1* p=0.833 (Figure A3A), *TP53* and *KMT2C* p=0.356 (Figure A3B), *KRAS* and *KEAP1* p=0.700 (Figure A3C), and *KRAS* and *KMT2C* p=0.304) (Figure A3D).

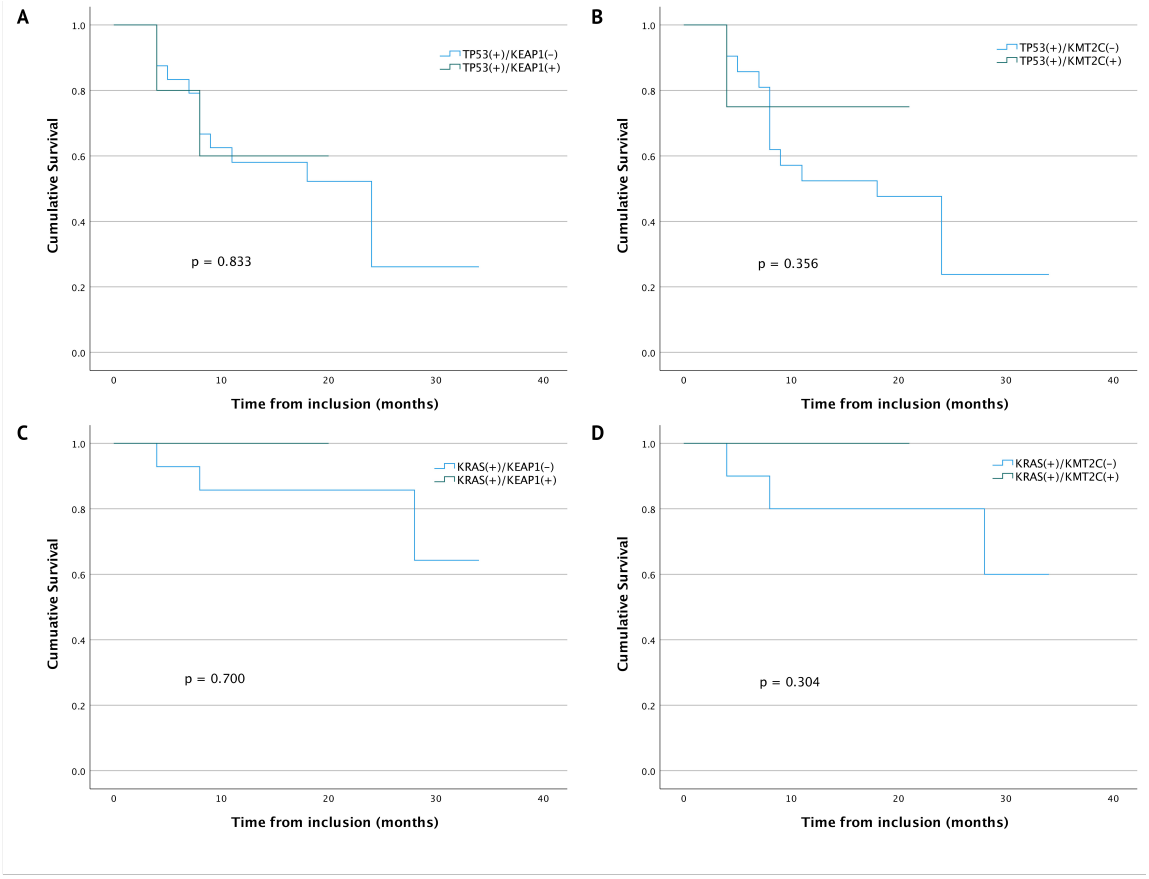


Figure A3: Kaplan-Meier curve showing overall survival for patients with **(A)** mutated *TP53* and wildtype (n=X) and mutated (n=X) *KEAP1* .**(B)** mutated *TP53* and wildtype (n=X) and mutated (n=X) *KMT2C*. **(C)** mutated *KRAS* wildtype (n=X) and mutated (n=X) *KEAP1* **(D)** mutated *KRAS* and wildtype (n=X) and mutated (n=X) *KMT2C*

Kaplan-Meier analysis of patients with a high (≥20) or low (<20) total number variants showed no significant association to survival (long-rank test p=0.684, Figure A4A). No significant association with immunotherapy response was found (p=0.824, OR=0.873, Figure A4B).

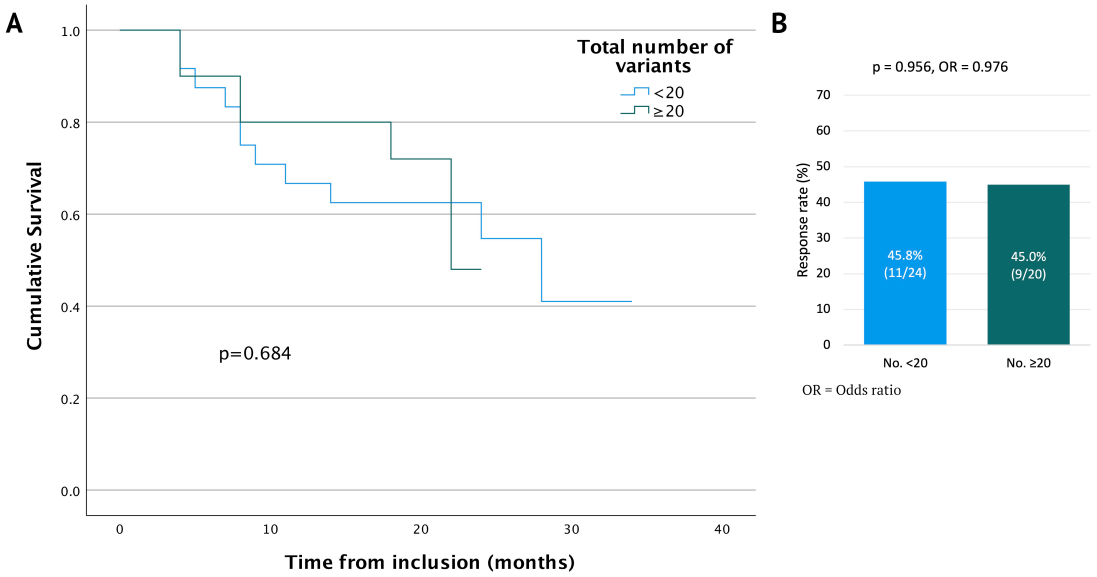


Figure A4: (A) Kaplan-Meier curve showing overall survival for patients with a high (≥20, n=20) or low (<20, n=24) total number of variants. **B** No association between the total number of variants and immunotherapy response was found.

Kaplan-Meier analysis of patients with a high (≥5) or low (<5) number of variants in driver genes showed no significant association to survival (p=0.860) (Figure A5).

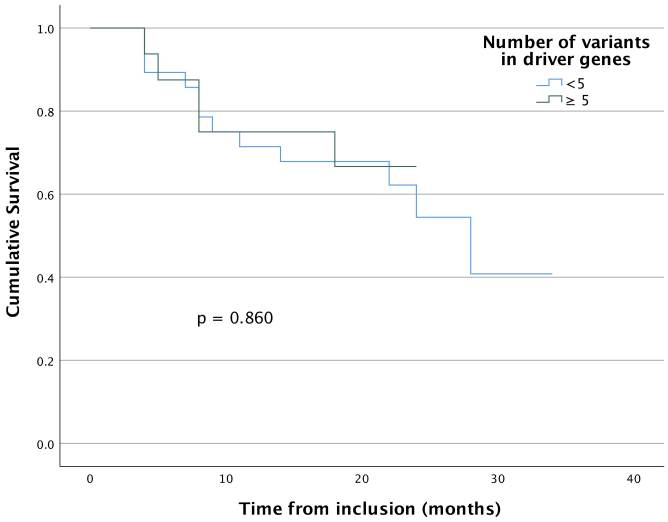


Figure A5: Kaplan-Meier curve showing overall survival for patients with a high ($\geq$5, n=16) or low ($<$5, n=28) number of variants in driver genes.