Journal of
Translational Medicine

RESEARCH

Open Access

Prioritizing gut microbial SNPs linked to immunotherapy outcomes in NSCLC patients by integrative bioinformatics analysis

Muhammad Faheem Raziq[1,2], Nadeem Khan[2], Haseeb Manzoor[2], Hafiz Muhammad Adnan Tariq[2], Mehak Rafiq[3], Shahzad Rasool[3], Masood Ur Rehman Kayani[2*] and Lisu Huang[1,4*]

Abstract

Background The human gut microbiome has emerged as a potential modulator of treatment efficacy for different cancers, including non-small cell lung cancer (NSCLC) patients undergoing immune checkpoint inhibitor (ICI) therapy. In this study, we investigated the association of gut microbial variations with response against ICIs by analyzing the gut metagenomes of NSCLC patients.

Methods Strain identification from the publicly available metagenomes of 87 NSCLC patients, treated with nivolumab and collected at three different timepoints (T0, T1, and T2), was performed using StrainPhlAn3. Variant calling and annotations were performed using Snippy and associations between microbial genes and genomic variations with treatment responses were evaluated using MaAsLin2. Supervised machine learning models were developed to prioritize single nucleotide polymorphisms (SNPs) predictive of treatment response. Structural bioinformatics approaches were employed using MUpro, I-Mutant 2.0, CASTp and PyMOL to access the functional impact of prioritized SNPs on protein stability and active site interactions.

Results Our findings revealed the presence of strains for several microbial species (e.g., *Lachnospira eligens*) exclusively in Responders (R) or Non-responders (NR) (e.g., *Parabacteroides distasonis*). Variant calling and annotations for the identified strains from R and NR patients highlighted variations in genes (e.g., *ftsA*, *lpdA*, and *nadB*) that were significantly associated with the NR status of patients. Among the developed models, Logistic Regression performed best (accuracy > 90% and AUC ROC > 95%) in prioritizing SNPs in genes that could distinguish R and NR at T0. These SNPs included Ala168Val (*lpdA*) in *Phocaeicola dorei* and Tyr233His (*lpdA*), Leu330Ser (*lpdA*), and His233Arg (*obgE*) in *Parabacteroides distasonis*. Lastly, structural analyses of these prioritized variants in *objE* and *lpdA* revealed their involvement in the substrate binding site and an overall reduction in protein stability. This suggests that these variations might likely disrupt substrate interactions and compromise protein stability, thereby impairing normal protein functionality.

Conclusion The integration of metagenomics, machine learning, and structural bioinformatics provides a robust framework for understanding the association between gut microbial variations and treatment response, paving the way for personalized therapies for NSCLC in the future. These findings emphasize the potential clinical

*Correspondence:
Masood Ur Rehman Kayani
m.kayani@sines.nust.edu.pk
Lisu Huang
lisuhuang@zju.edu.cn
Full list of author information is available at the end of the article

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 2 of 14

implications of microbiome-based biomarkers in guiding patient-specific treatment strategies and improving immunotherapy outcomes.

Keywords NSCLC, Gut microbiome, Immunotherapy, Immune checkpoint inhibitors (ICIs), Anti-PD-1, Microbial strains, Genomic variants

Introduction

Lung cancer is the most prevalent type of cancer (11.6% of the total cancer cases) as well as the leading cause of mortalities globally, accounting for ~ 18.4% of total cancer-related fatalities [1]. This form of cancer can primarily be divided into two types based on the histology of the cancer cells, i.e., small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), which is the most common and accounts for 85–90% of all lung cancer cases [2]. The lack of a reliable platform for early-stage diagnosis, combined with the delayed onset of symptoms during disease progression, severely limits treatment options and significantly reduces the survival rates of patients [3]. Moreover, current treatment methods often yield poor outcomes in NSCLC patients [4]. This underscores the urgent need for novel strategies to facilitate early diagnosis and improve the efficacy of existing therapeutic approaches.

Immunotherapy has recently emerged as a therapeutic option for treating advanced-stage NSCLC patients. Immune checkpoint inhibitors (ICIs), such as anti-PD-1 antibodies like nivolumab, block inhibitory signals cancer cells use to evade immune recognition [5]. However, response rates to ICIs can vary greatly among patients, and a better understanding of the underlying factors that modulate the response is needed [6]. Human gut microbiome has emerged as a key player in influencing the efficacy of immunotherapy, including ICIs. Gut commensals have been shown to influence the maturation and activation of immune cells, including T-cells [7].

The PD-L1 ligand on tumor cells binds to the PD-1 receptor on T-cells, inhibiting their ability to attack tumors. ICI therapy disrupts this interaction by using anti-PD-1 antibodies to activate T-cells and restore their tumor-fighting function [5]. Specific bacterial species, such as *Akkermansia muciniphila* and *Faecalibacterium prausnitzii*, have been linked to improved immunotherapy responses [8]. Though the mechanisms are still under investigation, these bacteria enhance immune recognition of cancer cells by boosting dendritic and T-cell activity, with *Akkermansia muciniphila* promoting T-helper 17 cell differentiation [9–11]. Additionally, they stimulate anti-inflammatory cytokine production, downregulate PD-L1 expression on tumor cells to enhance T-cell activation, and produce short-chain fatty acids (SCFAs) like butyrate, which support gut barrier integrity and regulate immune-related gene expression [5]. By maintaining gut health and preventing systemic inflammation, *Akkermansia muciniphila* helps sustain immune function, underscoring the importance of microbiome-based strategies for optimizing immunotherapy [10].

In addition to *Akkermansia muciniphila* and *Faecalibacterium prausnitzii*, several other microbial species and their metabolic products may also significantly influence immunotherapy outcomes. For example, *Bacteroides uniformis* has been linked to the enhancement of immune surveillance and the destruction of cancer cells, potentially due to its role in regulating immune pathways [12]. Similarly, *Lachnospira eligens*, part of the *Lachnospiraceae* family, is known for its ability to promote the differentiation of *T-helper 17* cells, which are crucial for the body's anti-tumor immune response [10]. Microbial metabolites such as SCFAs, like butyrate and propionate, play a key role in modulating immune functions, including T-cell activation and anti-inflammatory effects, which could improve immunotherapy responses [13]. SCFAs produced by bacteria like *Phocaeicola vulgatus* and *Bacteroides uniformis* may help create a more favorable tumor microenvironment, enhancing treatment outcomes [14]. Furthermore, metabolites derived from the tryptophan metabolic pathway, which are influenced by gut bacteria, can modulate immune activity, with certain indole derivatives showing potential to enhance the efficacy of immune checkpoint inhibitors [15]. On the other hand, species like *Parabacteroides distasonis* and *Bacteroides fragilis*, found in higher abundance in non-responders (NR) than responders (R), may contribute to immune suppression via their lipopolysaccharide (LPS) production, which can induce chronic inflammation and undermine effective immune responses [16]) (Supplementary Table 1).

Despite growing evidence linking the gut microbiome to immunotherapy outcomes in NSCLC, several limitations remain. Most studies focus on taxonomic composition without prioritizing specific microbial strains or patient subgroups that may exhibit distinct therapy responses. Additionally, strain-level profiling is often lacking, limiting the identification of functionally relevant microbial variations, such as single nucleotide polymorphisms (SNPs), that may influence immunotherapy efficacy. Addressing these gaps is essential for refining patient stratification and

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 3 of 14

optimizing microbiome-based therapeutic strategies to improve immunotherapy success rates in NSCLC.

While specific gut microbes, such as *Akkermansia muciniphila* and *Faecalibacterium prausnitzii*, have been linked to improved responses, microbiome-targeted interventions remain largely unexplored. Future research should focus on evaluating clinical efficacy, optimal formulations, and mechanisms of action of microbiome-based strategies, including probiotics, prebiotics, and microbiome transplantation, to enhance immunotherapy outcomes.

In this study, we sought to identify gut microbial variations present in NSCLC patients to establish their connection with the variable treatment responses. We integrated metagenomic analysis, statistical methods, and machine learning (ML) models to prioritize variants that classify R and NR in ICI therapy. We also performed structural bioinformatics on reference and variant proteins to analyze the potential effects of the prioritized SNPs on the structure as well as their function. The overall workflow of our study is shown in Fig. 1. We hope our findings will contribute to expanding the current knowledge of the link between gut microbiome, underlying genomic variations, and their impact on the efficacy of immunotherapy in NSCLC patients.

Materials and methods

Metagenomic data acquisition and preprocessing

The metagenomic shotgun sequencing data used for this study was obtained from the European Nucleotide Archive (ENA) using accession number PRJEB22863 [17] via the SRA Toolkit (v3.0.0) (https://github.com/ncbi/sra-tools). This dataset includes 87 NSCLC patients who were treated with nivolumab. Metagenomic samples were collected at three timepoints: T0 (baseline, before treatment initiation), T1 (after one month of treatment), and T2 (after two months of treatment). The dataset consisted of 65, 38, and 15 samples corresponding to T0, T1, and T2, respectively. Importantly, each sample corresponds to a single patient at a given time point. The male to female patient ratio is 2.11:1 (59:28) which concludes that male patients are more than twice as prevalent as female patients in this dataset.

Based on treatment response, a total of 118 samples were classified into R (59) and NR (59). Few patients had samples at multiple time points (T0, T1, and/or T2), but their response category (R or NR) was determined based on their final treatment outcome. Some patients who initially responded at T0 exhibited progression or resistance at later time points (T1 and/or T2), while others maintained their initial response throughout treatment. In these metagenomic samples, the average number of raw reads per sample was around 20 ± 3.46 Mbp (Supplementary Table 2).
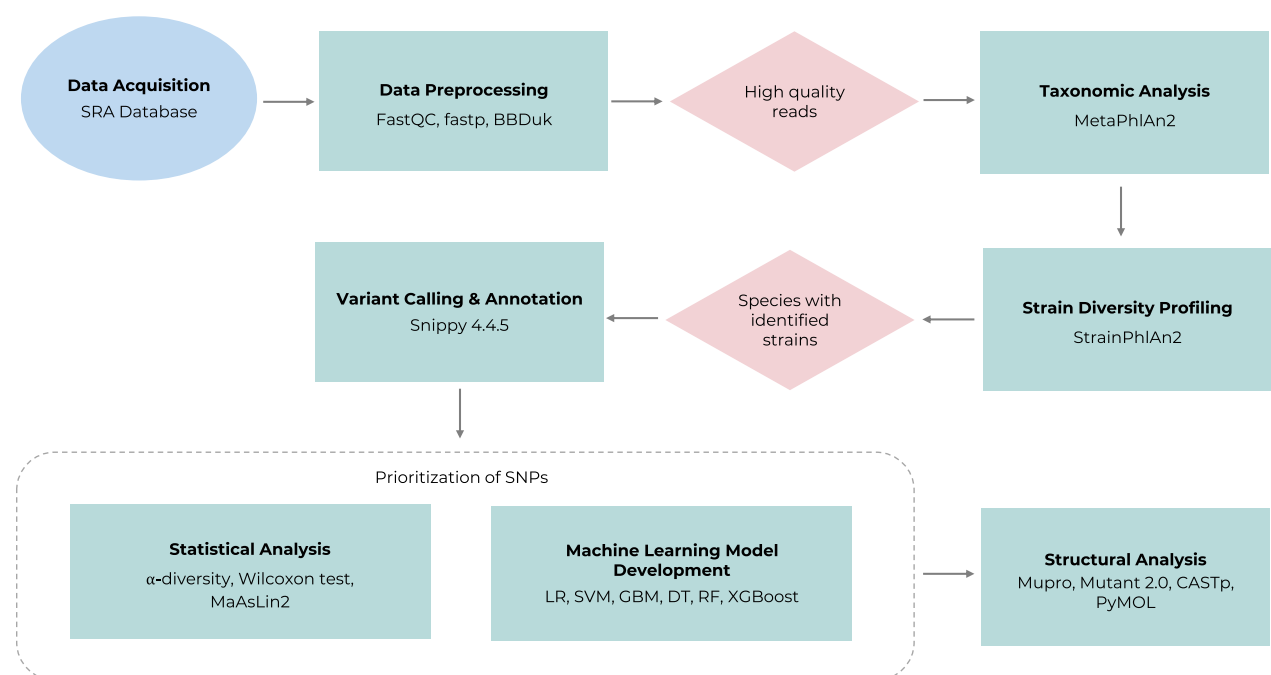

Fig. 1 Overview of the study workflow for prioritizing gut microbial SNPs linked to immunotherapy outcomes in NSCLC patients by integrative bioinformatics analysis

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 4 of 14

FastQC (v0.12.1) (https://www.bioinformatics.babra ham.ac.uk/projects/fastqc/) was used for checking the quality of the raw reads. After which, adapter trimming and removal of low-quality reads were conducted using the fastp (v0.23.4) [18]) with the following parameters: -trim_front1 5 (for trimming first 5 bases of each read), -length_limit 265 (for setting the maximum allowed read length to 265), and -length_required 100 (to discard reads shorter than 100), respectively. Bases above position 265 had extremely poor-quality scores. Therefore, we kept the maximum read length to 265 in this study. Finally, BBduk (BBMap—Bushnell B.—sourceforge.net/projects/ bbmap/) was used to remove host-derived sequences by mapping metagenomic reads to the human reference genome (GRch38) and discarding mapped reads. The average number of reads per sample in the post-processed data was $\sim 18.5 \pm 2.69$ Mbp.

Strain diversity profiling, variant calling, and annotation

MetaPhlAn3 ([19], p. 3) was used to profile the microbial species present in gut metagenome samples. Next, StrainPhlAn3 [20] was used for identifying strains for the highly abundant species with the following parameters: –phylophlan_mode fast (fast phylogenetic analysis), -mutation_rates (mutation rates table for each of the aligned markers), -marker_in_n_samples 20 (threshold defining the minimum absolute number of samples for a marker to be primary), and -sample_with_n_markers 20 (threshold defining the minimum absolute number of markers for a sample to be primary). In StrainPhlAn, identified marker sequences for each of the 25 species underwent alignment using MUSCLE [21].

For each of the species for which StrainPhlAn3 reported the presence of strains, reference genomes were retrieved from NCBI RefSeq. Genomic variants, including single nucleotide polymorphisms (SNPs), multiple nucleotide polymorphisms (MNPs), complex mutations, deletions, and insertions, were profiled using Snippy (https://github.com/tseemann/snippy) by aligning high-quality reads against the respective reference genomes.

Statistical analysis and machine learning model development

A-diversity was estimated using indices including the Shannon index, richness, and evenness, with the R library "vegan". The Wilcoxon test was used to compare the differences in α-diversity between the two treatment response groups. Additionally, MaAsLin2 [22] was used for the identification of associations between genes harboring variations and treatment response.

Next, we developed supervised machine learning models for prioritizing genomic variations at T0 which are predictive of NR status in patients treated with ICIs. The classifiers used included logistic regression (LR), support vector machine (SVM), decision tree (DT), extreme gradient boosting (XGBoost), gradient boosting machine (GBM), and random forest (RF). Data was split into 80/20 train and test datasets, followed by the encoding of categorical data. To enhance model performance, hyperparameter tuning was conducted using GridSearchCV with three-fold cross-validation. For the SVM classifier, regularization strength (C) was tested at multiple values (0.1, 1, 10, 100), with a linear kernel selected for simplicity and efficiency. For XGBoost, the number of boosting rounds (n_estimators) was optimized at 50, 100, and 200, while tree depth (max_depth) was tested at values of 3, 5, and 7. Additionally, the learning rate was fine-tuned at 0.01, 0.1, and 0.2, along with subsample ratios of 0.7, 0.8, and 0.9 to control for overfitting. Similarly, for the decision tree model, tuning focused on maximum tree depth (None, 10, 20, 30), the minimum number of samples required for a split (2, 5, 10), and the minimum samples per leaf node (1, 2, 4). The best-performing model was selected based on the highest accuracy, and its feature importance scores were analyzed. The SNPs were filtered from the genomic variations that occurred in the NR-associated microbial genes belonging to the T0. The SNPs data was given as input to the various above-mentioned ML classifiers to prioritize them based on their importance values.

Protein sequence retrieval and comparative structural analysis

For the genes predictive of NR status at T0, identified by machine learning models, reference protein sequences were retrieved from UniProt using accession numbers A0A3L7ZQF3 and A0A412ZJ10. The reference structures of the retrieved proteins were manually mutated with the SNPs prioritized by ML and evaluated for changes in residue interactions using the PyMOL [23]. Additionally, the active sites of the proteins were predicted using CASTp to assess the involvement of reference and mutated residues in active site formation [24]. These analyses were conducted to compare reference and mutated structures and to examine any changes in residue interactions, particularly those involved in active site formation.

Next, we analyzed and compared the stability of the reference and mutant forms using the MUpro (https:// mupro.proteomics.ics.uci.edu) and I-Mutant 2.0 (https:// folding.biofold.org/i-mutant), respectively [25, 26]. MUpro uses an SVM model to classify and predict the stability of query proteins. The tool was trained on 615 single-site variations obtained from 42 different proteins. Similarly, I-Mutant 2.0 is trained on 14,468 protein structures and employs an SVM-based classification to predict protein stability when variations are introduced.

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 5 of 14

Results

Dynamics of strain-level diversity in NSCLC patients

MetaPhlAn3 was used for profiling the taxonomic composition in NSCLC patients, which identified 1,655 microbial species in NSCLC patient samples (Supplementary Table 3). Among these, 410 species were identified in T0, 374 in T1, and 252 in T2, with *Phocaeicola vulgatus*, *Bacteroides uniformis*, *Faecalibacterium prausnitzii*, *Phocaeicola dorei*, and *Parabacteroides distasonis* among the most abundant species. Stratifying the data by treatment response, 1377 species were identified in R whereas 1371 species were identified in NR. In R, the most abundant species were *Phocaeicola vulgatus*, *Bacteroides uniformis*, *Faecalibacterium prausnitzii*, *Phocaeicola dorei*, and *Akkermansia muciniphila*, suggesting their potential in positively influencing treatment response. Conversely, in NR, *Bacteroides uniformis*, *Phocaeicola vulgatus*, *Faecalibacterium prausnitzii*, *Parabacteroides distasonis*, and *Phocaeicola dorei* were most abundant. Next, we sought to determine the associations between species with treatment response using MaAsLin2. The results revealed significant positive associations of *Phocaeicola vulgatus* and *Bacteroides caccae* with the R group of patients, with coefficients of 3.49 and 2.91, respectively, and an FDR of 0.14 for both (Supplementary Fig. 1).

We selected the 25 most abundant microbial species, based on their relative abundance, to perform strain diversity profiling using StrainPhlAn3. The results indicated the presence of strains for 21 species out of 25 (top 10 abundant species with strains shown in Fig. 2a). Strains for *Akkermansia muciniphila*, *Phocaeicola copri*, *Bacteroides stercoris,,* *Phocaeicola dorei*, *Faecalibacterium prausnitzii*, *Bacteroides uniformis*, and *Phocaeicola vulgatus* were present across most of the timepoints,

while strains for *Roseburia faecis* and *Eubacterium rectale* were found only at T0, whereas *Bacteroides fragilis* and *Alistipes finegoldii* were specific to T2 in NR. Stratifying by treatment response, strains in R were identified at all three timepoints for *Phocaeicola vulgatus*, *Bacteroides uniformis*, while NR showed the strains for *Bacteroides stercoris*, *Phocaeicola dorei*, *Faecalibacterium prausnitzii* and *Bacteroides uniformis*. Strains unique in R belonged to *Parabacteroides merdae*, *Lachnospira eligens*, and *Escherichia coli* whereas for NR the unique strains belonged to *Bacteroides fragilis*, *Alistipes finegoldii* and *Parabacteroides distasonis* (Fig. 2b) (Supplementary Table 4).

Profiling gut microbial genomic variants in NSCLC patients and their association with treatment response

Next, we selected the seven microbial species that showed the presence of strains across most of the timepoints in R and NR, for the identification of genomic variants that could potentially be linked with alternate treatment responses. These included *Akkermansia muciniphila*, *Phocaeicola dorei*, *Bacteroides stercoris*, *Bacteroides uniformis*, *Facalibacterium prausnitzii*, and *Phocaeicola vulgatus*. The only exception was *Parabacteroides distasonis*, which was chosen due to its presence across all timepoints in NR (Fig. 2b).

A total of 83,583 variations were identified in R and NR samples collectively as shown in Supplementary Fig. 2, including 63,737 SNPs, 17,553 complex mutations, 1803 MNPs, 278 deletions, and 216 insertions. Among these, 35,615 and 47,969 variations were identified from R and NR, respectively (Fig. 3a). Among these, *Akkermansia muciniphila* showed 4197 vs. 5939 variations (2903 vs. 4245 SNPs), *Phocaeicola dorei* had 1971 vs. 2,719 variations (1633 vs. 2357 SNPs), *Bacteroides stercoris* showed
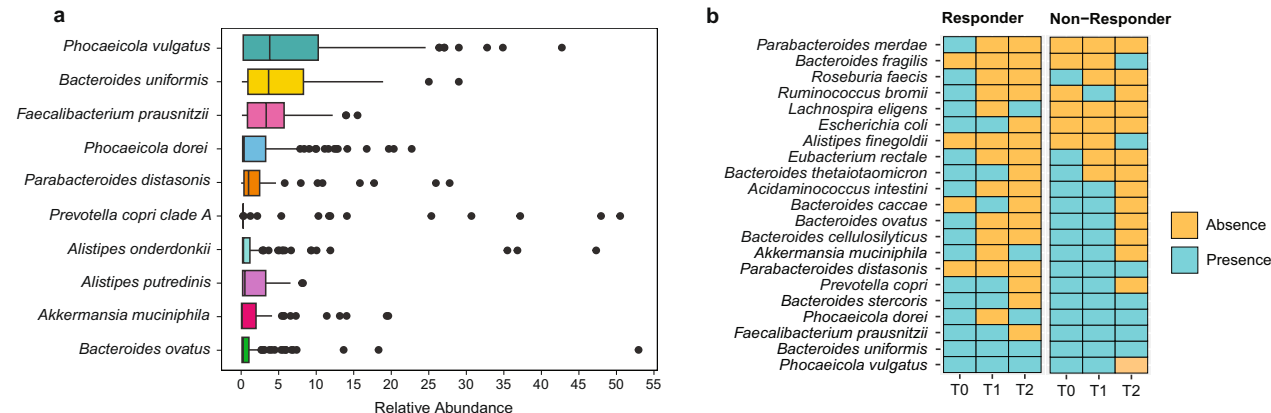

Fig. 2 Distribution of species with identified strains in R and NR. **a** The box plot shows the top 10 most abundant species based on their relative abundances. **b** The presence/absence plot illustrating the presence and absence of species with identified strains across different timepoints in R and NR
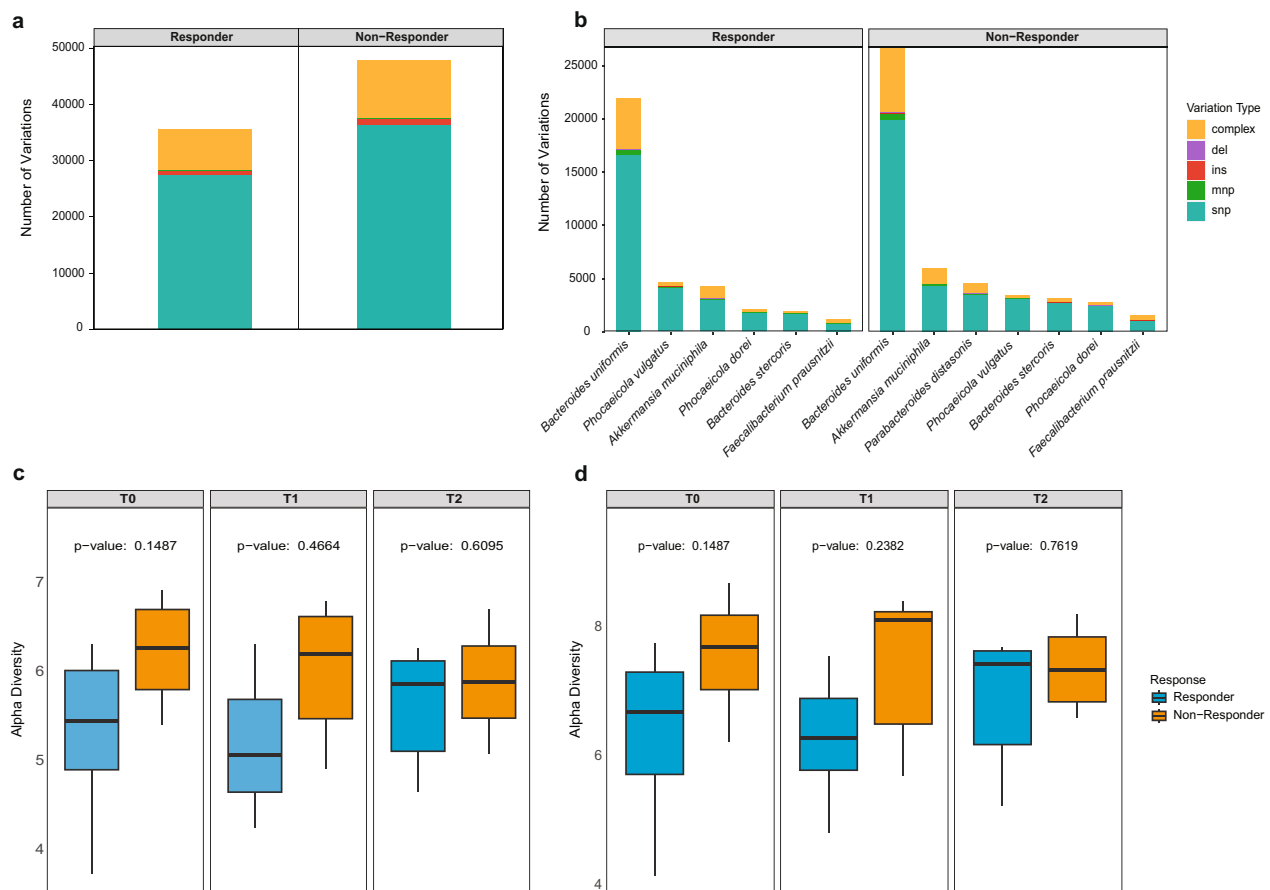
Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 6 of 14

Fig. 3 Number of genomic variations profiled in R vs NR along with α-diversity of genes and genomic variants. **a** The stacked bar plot demonstrating the total number of variations identified in R vs NR constituted with type of variations. **b** Total number of variations profiled in common species with identified strains comparing them in R and NR stacked with type of variations. **c, d** The trend of α-diversity of genes and variants counts across different timepoints (T0, T1 and T2) in R and NR to ICIs along with their p-values

1801 vs. 3177 variations (1531 vs. 2653 SNPs), *Bacteroides uniformis* exhibited 21,964 vs. 26,810 variations (16,636 vs. 19,876 SNPs), *Faecalibacterium prausnitzii* contributed 1113 vs. 1532 variations (679 vs. 907 SNPs), and *Phocaeicola vulgatus* displayed 4574 vs. 3291 variations (4027 vs. 2,938 SNPs), in R and NR, respectively but *Parabacteroides distasonis* had all the 4507 variations in NR (3351 SNPs) (Fig. 3b).

We compared the α-diversity among different time points for R and NR groups based on genes and genomic variations. For genes, no statistically significant differences in α-diversity were observed between the R and NR groups at T0, T1 and T2 with their respective p-values as 0.1487, 0.2382, and 0.7619. However, within the NR group, α-diversity showed a continuous decrease from T0 to T2 (Fig. 3c). For genomic variants, a similar trend was observed, with no significant differences in α-diversity between R and NR groups with the p-values of 0.1487, 0.4664, and 0.6095 at T0, T1 and T2, respectively. In the NR group, α-diversity continuously decreased from T0

to T2, whereas in the R group, α-diversity dropped at T1 and showed a slight increase at T2 (Fig. 3d).

Next, we performed association testing using MaAsLin2 to identify the genes associated with R and NR statuses (Supplementary Table 5). The *lpdA* (dihydrolipoyl dehydrogenase) gene was negatively associated with R, showing a coefficient of − 0.967 and an FDR of 0.2335. Similarly, *nadB* (ʟ-aspartate oxidase) exhibited a negative association with R, having a coefficient of − 0.591 and an FDR of 0.2335. The genes *sufD* (Fe-S cluster assembly protein SufD) and *uxaC* (glucuronate isomerase) also showed a negative association with R, resulting in coefficients of − 1.14 and − 0.952, respectively, both having an FDR of 0.2335. Additionally, *ftsA* (cell division protein FtsA), *obgE* (GTPase ObgE), *rhaT* (ʟ-rhamnose-proton symporter), and *xylE* (ᴅ-xylose transporter XylE) were found to be significantly positively associated with NR, showing coefficients of − 0.586, − 0.363, − 0.521, and − 0.654, respectively, and FDR values around ∼ 0.2354. Hence, R exhibited lower association levels of these

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 7 of 14

genes compared to NR (Fig. 4). The variant calling identified higher genomic variations in these associated genes linked with NR, such as *ftsA* showing 35 variations in NR compared to 7 in R, and *sufD* displaying 119 variations in NR compared to 41 in R.

Machine learning models differentiate R and NR based on prioritized SNPs

The machine learning models developed in this study showed high performance in distinguishing R and NR based on gut microbial SNPs at baseline. We hypothesized that the genes found to be strongly associated with NR, might harbor variations that may play a role in patients being NR to ICI therapy. The performances of ML models were evaluated in terms of accuracy, precision, recall, F1-score, and ROC AUC (Table 1). The LR and XGB models performed similarly with an accuracy of 91%, precision of 94%, recall of 89%, and an F1-score of 91%, but LR showed the highest ROC AUC of 0.96 followed by XGB with 0.95 (Fig. 5 and Supplementary Figure 3). Using the LR model, we investigated the SNPs that had the most significant effect on determining the response to immunotherapy. All prioritized SNPs are listed in Table 2 belonging to *Phocaeicola dorei* and *Parabacteriodes distasonis* with the highest importance values.

The other models like RF, GBM, and DT also showed good results with ROC AUC of 0.96, 0.94, and 0.92, respectively (Supplementary Fig. 3). The models highlighted the microbial genes, including *lpdA*, *sufD*,
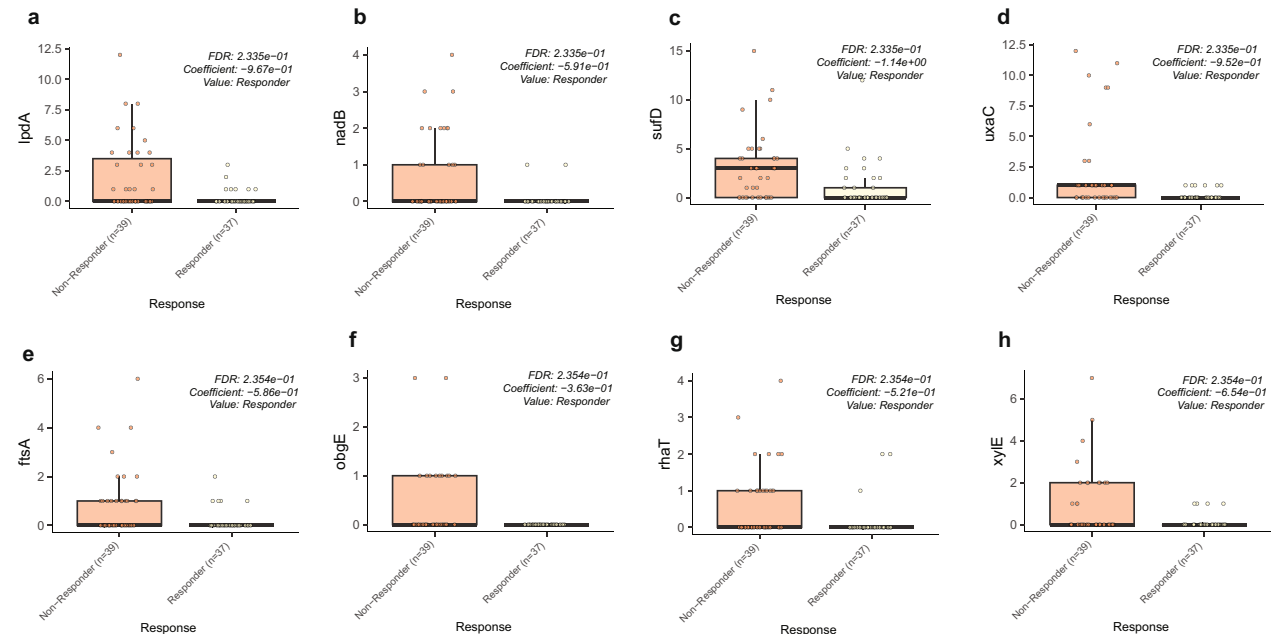
Fig. 4 Genes identified by MaAsLin2 associated with treatment response. **a–c** The genes *xylE*, *rhaT*, and *obgE* show a negative association with R and are found in NR. **d–f** The genes ftsA, *sufD*, and *uxaC* are inversely associated with R and are significantly present in NR. **g, h** The genes *lpdA* and *nadB* are found to be negatively associated with R and are observed in NR

Table 1 Performance matrix of various ML models

Model	Responders			Non-responders				
	Precision	Recall	F1-score	Precision	Recall	F1-score	Accuracy	AUC
Logistic regression	0.94	0.89	0.91	0.88	0.93	0.90	0.91	0.96
Random forest	0.89	0.87	0.88	0.89	0.91	0.90	0.89	0.96
Extreme gradient boosting	0.94	0.89	0.91	0.88	0.93	0.90	0.91	0.95
Gradient boosting machine	0.91	0.91	0.91	0.89	0.89	0.89	0.90	0.94
Decision tree	0.92	0.85	0.88	0.84	0.91	0.87	0.88	0.92

The performance matrix of different machine learning classifiers was integrated into the study to prioritize SNPs identified in T0 samples associated with treatment efficacy in NSCLC patients
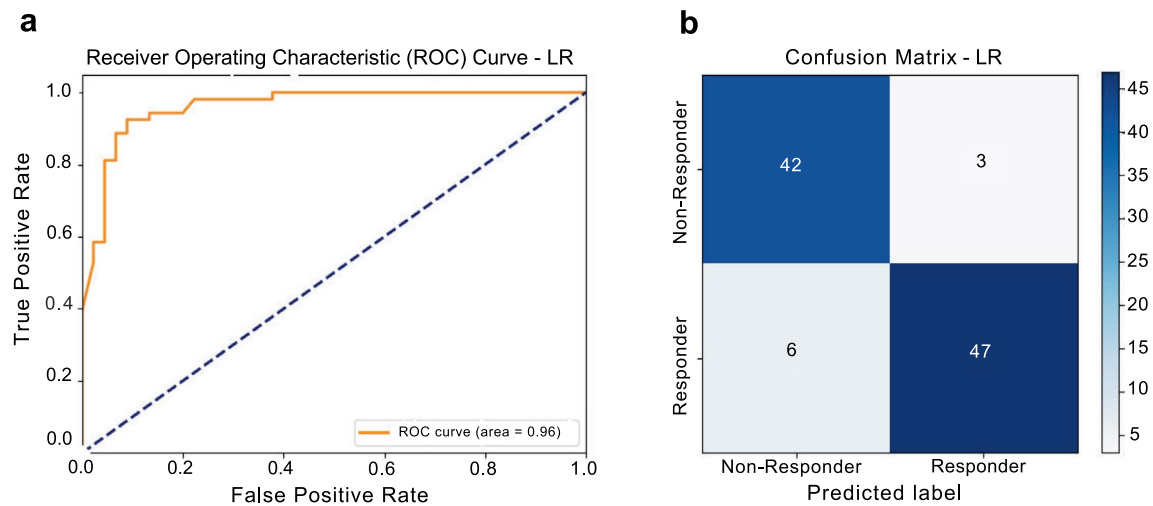
Fig. 5 The evaluation of best performing ML models. **a** The LR model's ROC curve shows optimal performance in prioritizing baseline SNPs linked to treatment response. **b** The confusion matrix of LR model records 42 true negatives, 06 false negatives, 03 false positives, and 47 true positives in prioritizing baseline SNPs

Table 2 Prioritized SNPs via ML in NR to ICIs at baseline (T0)

Species with identified strains	Missense variation	Genes	Category
Bacteroides dorei	c.503C > T p.Ala168Val	*lpdA*	Non-responder
Parabacteroides distasonis	c.697T > C p.Tyr233His		
	c.989T > C p.Leu330Ser		
	c.8G > A p.Ser3Asn	*uxaC*	
	c.698A > G p.His233Arg	*obgE*	

The genomic missense variants in two species with identified strains at baseline in the genes associated with treatment response were predicted by different machine learning models in R and NR to ICIs

Table 3 Microbial genes with overlapping variations prioritized by different machine learning models

Model	Genes
LR	*sufD, lpdA, uxaC,* and *obgE*
RF	*lpdA, sufD, uxaC,* and *nadB*
XGBoost	*sufD, lpdA, obgE,* and *rhaT*
GBM	*sufD, obgE, rhaT,* and *xylE*
DT	*sufD, lpdA, rhaT,* and *xylE*

The overlapping genes listed in which SNPs were prioritized by the respective machine learning models

uxaC, nadB, obgE, xylE, and *rhaT,* in which SNPs have been observed to be important (Table 3). It can be seen from the table that there is significant overlap in the genes containing prioritized SNPs to determine the model's prediction with mostly *sufD* being the most common one followed by *lpdA, obgE,* and *rhaT.*

Structural comparison of prioritized SNPs linked with NR

The impact of SNPs in the genes, prioritized through ML models at T0, was further investigated using structural bioinformatics approaches. For this, the reference structures for the respective proteins were retrieved and mutated with the prioritized SNPs and evaluated for changes in residue interactions, as described in Materials and Methods. Histidine-233 was identified as a critical residue contributing to the active site formation of the reference *obgE* protein. To assess the impact of an SNP, Histidine-233 was substituted with Arginine-233, and subsequent analysis revealed that the replaced residue retained its role in active site formation. Similarly, in the *lpdA* protein, Tyrosine-233 when mutated with Histidine-233, also contributed significantly to the formation of the active site. The involvement of mutated residues in active site formation indicates that the SNPs in these genes may have profound effect on the function by rendering these genes respond differently to their target substrates.

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 9 of 14

Furthermore, the effect of SNPs on protein stability revealed that the stability of variant proteins got compromised (Table 4), which can plausibly negatively impact their normal function. The thorough examination of reference and mutant proteins further provides evidence of the compromised stability in the SNP *lpdA* and *obgE*. In the normal *obgE*, the Histidine-233 residue forms three polar contacts with Glycine-231, Arginine-236, and Histidine-237. In contrast, the mutated Arginine-233 forms only two polar contacts with Arginine-236 and Histidine-237, which may account for its decreased stability (Fig. 6). Similarly, observing the reference and mutated structures of *lpdA*, three amino acids Alanine-168, Tyrosine-233, and Leucine-330 were mutated to Valine-168, Histidine-233, and Serine-330, respectively (Fig. 7). Comparative analysis of the reference and mutant structures of *lpdA* revealed that in the reference structure, Tyrosine-233 forms a polar contact with Glutamate-233 at a

bond length of 3.0 Å. However, in the mutant structure, where Tyrosine-233 is replaced by Histidine-233, the bond length is slightly reduced to 2.9 Å. This reduction increases the likelihood of molecular clashes between the interacting amino acid residues, potentially compromising the overall stability of the mutant structure. Therefore, the occurrence of SNPs in *obgE* and *lpdA*, may have a great effect in its overall stability and functioning respectively.

Discussion

NSCLC is an aggressive type of lung cancer with a high mortality rate and the ability to evade the immune system leading to immune suppression. While immunotherapy has shown promising results in treating advanced stages of NSCLC, its effectiveness is limited to certain tumor types, and the high incidence of immune-related side effects varies between individuals [27]. The challenges

Table 4 Genomic variants' structure stability predicted by SVM-based classification models

Webserver	*obgE* (H233R)	*lpdA* (A168V)	*lpdA* (Y233H)	*lpdA* (L330S)
MUpro (Stability)	Decrease	Increase	Decrease	Decrease
I-Mutant (Stability)	Decrease	Increase	Decrease	Decrease

The predicted stability of *obgE* significantly decreases when Histidine is mutated by Arginine. Similarly, the overall stability of *lpdA* decreases when the wild type residues are mutated respectively
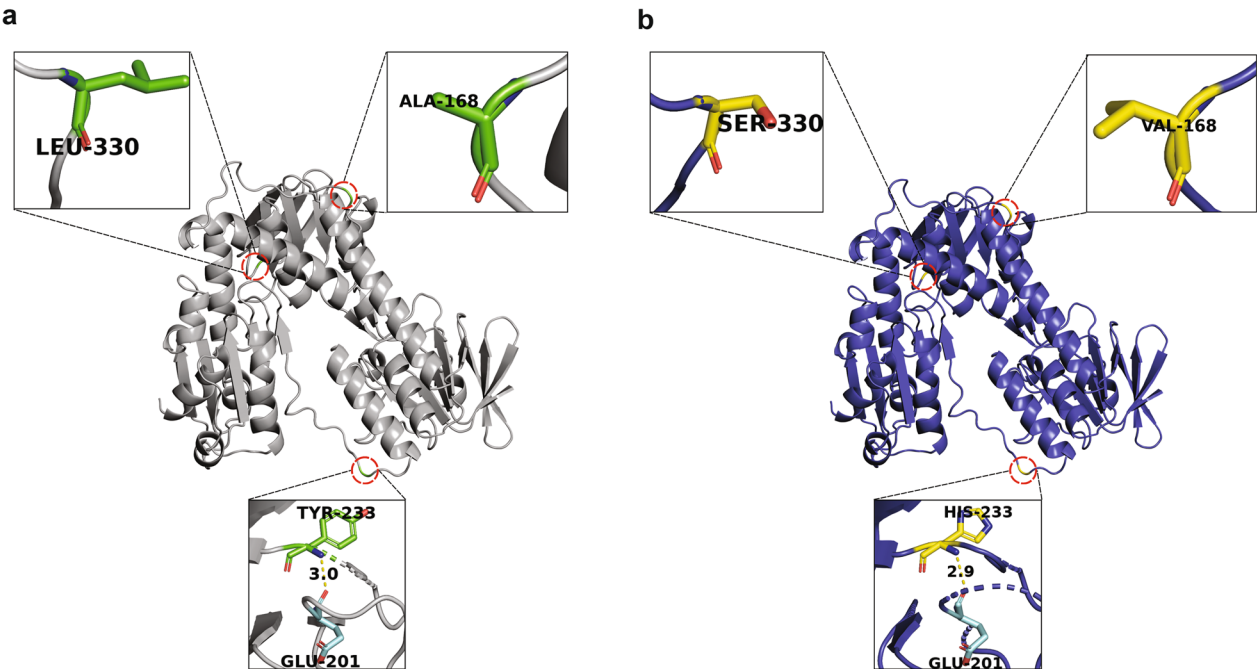

Fig. 6 *obgE* reference and mutated structures. **a** It showed the position and overall interactions of Histidine residue with its neighbouring residues. **b** It showed the position and interaction of the mutant protein. Three polar contacts are formed by Histidine-233 with Histidine-237 and Glysine-231 where nitrogen (purple) and oxygen (red) are actively involved. Similarly, two polar contacts are formed by the mutant residue respectively
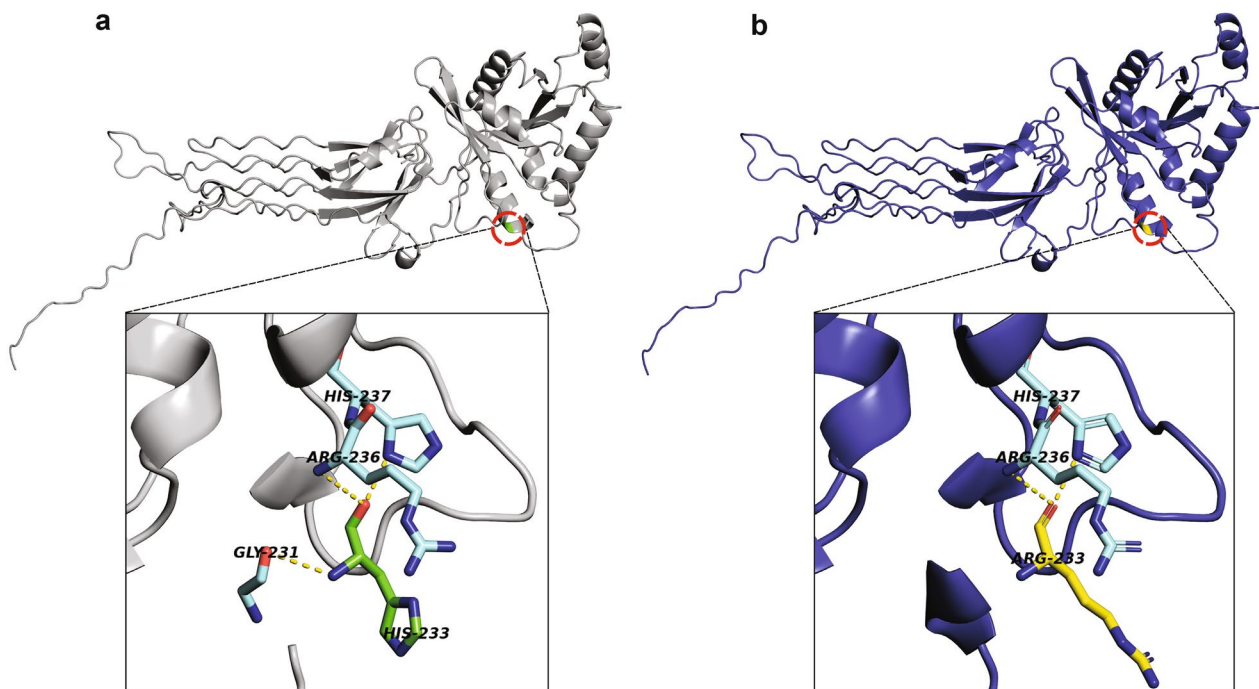
Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 10 of 14

Fig. 7 *lpdA* reference structure and mutant structure. **a** The figure demonstrates residues of wild type and its interactions with neighbouring residues (if present). Tyrosine-233 makes a polar contact with Glutamate-201 making a bond length of 3.0 Å. Amino group of Tyrosine-233 and oxygen of Glutamate-201 is involved in polar contact. **b** The figure demonstrates residues of mutant type and its interactions with neighbouring residues (if present). Histadine-233 makes a polar contact with Glutamate-201 making a bond length of 2.9 Å. Amino group of Histadine-233 and oxygen of Glutamate-201 is involved in polar contact

of immunotherapy for NSCLC are numerous and complex. One of the primary challenges is the heterogeneity of NSCLC, which means that different patients may respond differently to immunotherapy. This heterogeneity can be attributed to several factors, such as tumor genotype, immune status, and tumor microenvironment [28]. Recent advancements in sequencing technologies have highlighted the influence of gut microbiome, including specific strains and genomic variations on immunotherapy outcomes [29]. However, further studies are needed to improve our understanding.

Specific microbial species identified in this study can potentially be associated with variable treatment efficacy between R and NR. In R, predominant microbial species included *Phocaeicola vulgatus*, *Bateroides uniformis*, *Faecalibacterium prausnitzii*, *Phocaeicola dorei*, and *Akkermansia muciniphila*, known for their positive effects on gut health and immune function, thus supporting favorable treatment outcomes [13]. Conversely, NR exhibited a higher abundance of *Parabacteroides distasonis* and *Phocaeicola dorei*, which are associated with less effective treatment responses and are more prevalent in this group [16]. Advanced strain-resolved metagenomic technique was employed to identify specific microbial strains within these top abundant species [15]. Strain

diversity analysis showed that strains for certain species, including *Phocaeicola dorei*, *Bacteroides uniformis*, and *Faecalibacterium prausnitzii*, were consistently present at all sampled timepoints in both R and NR, highlighting their stable influence within the microbiome. Notably, the abundance of *Phocaeicola dorei* has been linked to longer progression-free survival (PFS) in patients [14]. *Bacteroides uniformis* is considered to enhance immune recognition and destruction of tumor cells; its presence is associated with higher microbial diversity and may enhance the effectiveness of ICIs. *Faecalibacterium prausnitzii*, known for its anti-inflammatory properties, is associated with higher production of SCFAs which can modulate immune responses and enhance treatment efficacy [12]. Unique strains found exclusively in R included strains of *Lachnospira eligens*, *Escherichia coli*, and *Parabacteroides merdae*. The Lachnospiraceae family, to which *Lachnospira eligens* belongs, is known for boosting immune responses and enhancing immunotherapy effectiveness by promoting the differentiation of T-helper 17 cells, hence, essential for anti-tumor immunity [10]. Although typically considered pathogenic, certain *Escherichia coli* strains can beneficially affect the immune system, enhancing immunotherapy responses by influencing immune cell activity and promoting a conducive

tumor microenvironment [30]. *Parabacteroides merdae* has been explored for its role in impacting the gut microbiome's effect on immune responses [31]. On the other hand, *Parabacteroides distasonis* and *Bacteroides fragilis* were found specifically in NR. Studies suggested that NSCLC patients with lower levels of these bacteria had better responses to PD-1 immunotherapy, indicating that these bacteria might be linked to worse immunotherapy outcomes [32].

NSCLC is often influenced by specific genomic variations that affect treatment effectiveness. The genomic landscape of NSCLC includes various mutations that enable the development of targeted treatments. However, the response to immunotherapy, particularly with ICIs, can vary significantly among patients due to tumor heterogeneity and distinct genomic variations [33]. In this study, NR showed a higher number of genomic variations compared to R, suggesting a relationship between genomic diversity and treatment response. Several genes, including *lpdA*, *nadB*, *sufD*, *uxaC*, *ftsA*, *obgE*, *rhaT*, and *xylE* have been found to be significantly associated with treatment outcomes in NSCLC. These genes are involved in crucial biological processes and metabolic pathways that could help explain differences in treatment results. For instance, the gene *lpdA*, which is vital for the oxidative decarboxylation of pyruvate and other α-keto acids, supports tumor growth and survival in NSCLC by affecting cellular redox balance and energy metabolism [34]. The *nadB* gene, essential for NAD (nicotinamide adenine dinucleotide) biosynthesis, plays a key role in redox reactions necessary for cellular metabolism and energy generation. NAD metabolism disruptions, often seen in NSCLC, can alter cellular energy balance and encourage cancer cell proliferation [35]. The *sufD* gene, involved in forming iron-sulfur clusters which are key components for many enzymes, can influence mitochondrial function and genomic stability, potentially promoting cancer development [36]. The *uxaC* gene, which helps break down uronic acids, reflects the complex carbohydrate metabolism within tumor environments [37].

The *ftsA* gene is crucial for bacterial cell division, playing a role in cytokinesis [38]. The *obgE* gene, which codes for a GTP-binding protein, is linked to ribosome assembly, stress response, and cell cycle regulation. These functions are often modified in cancer to support tumor growth and resistance to therapy [39]. The genes *rhaT* and *xylE*, involved in sugar metabolism and transport, such as rhamnose and xylose, indicate shifts in sugar metabolism pathways that may affect the tumor environment and cellular metabolism [40]. The genomic variations occurring in the above-mentioned genes may affect their normal functions and hence may disrupt their associated pathways impacting the tumor microenvironment,

metabolic adaptability, and immune evasion mechanisms, ultimately influencing the efficacy of ICIs in NSCLC patients. These metabolic pathway disruptions associated with these genes could be key in determining NSCLC treatment effectiveness.

Regarding the α-diversity of microbial genes harboring variations, a contrasting pattern was observed. The fluctuations in R could be indicative of an initial perturbation followed by adaptation phase. Conversely, NR exhibited a steady decline in α-diversity across timepoints, suggesting a continuous reduction in gene diversity which might be linked to a lack of response to immunotherapy. However, the observed differences between the two groups at the three timepoints were statistically not significant. The decrease in α-diversity observed in NR may be partly due to the limited number of samples available at the T2 timepoint, which could have influenced the results. In R, the α-diversity of genomic variants exhibited an initial decline from T0 to T1, followed by a recovery and increase at T2, suggesting an adaptive genomic response to the immunotherapy over time. This increase could reflect the tumor's or microbiome's ability to adjust to treatment pressures. Conversely, NR demonstrated a steady decline in α-diversity from T0 to T2, indicating a lack of such adaptive changes. Despite this trend, NR maintained a consistently higher mean α-diversity compared to R, which could indicate a genetically diverse microbial community that, however, might be less responsive or adaptable to immunotherapy. These findings highlight the complex interplay between microbial diversity and treatment outcomes, suggesting that while microbial diversity may be higher in NR, it may not necessarily correlate with a better therapeutic response.

Machine learning models, particularly LR classifier, effectively prioritized key SNPs that were linked with treatment response at baseline in NR, providing robust predictive power with high accuracy, precision, and recall. Important genes such as *sufD*, *lpdA*, *rhaT*, and *xylE*, along with significant missense variants like p.Ala168Val, p.His233Arg, p.Met418Leu, p.Tyr233His, and p.Ala311Thr, were consistently highlighted, underscoring their relevance in classifying R and NR. These prioritized SNPs may determine the suitability of the patients for ICI therapy by profiling them based on their baseline characteristics.

The structural and stability effects of SNPs prioritized through ML in the normal *obgE* and *lpdA* proteins, may play important roles in the normal functioning of the different bacteria by enabling energy production, metabolism and helping it cope with oxidative stress [41–43]. The presence of SNPs at key residues, such as Histidine-233 in *obgE* and Tyrosine-233 in *lpdA*, highlights important considerations regarding protein function

and stability. While these SNPs may not completely disrupt the involvement of the active sites in the respective protein functions, they can significantly impact the overall stability of the proteins. Specifically, the substitution of Histidine-233 with Arginine-233 in *obgE* results in a reduction of polar contacts, leading to decreased protein stability, which could compromise its functional integrity. Similarly, the change of Tyrosine-233 to Histidine-233 in *lpdA*, along with alterations to residues like Leucine-330, further destabilizes the protein structure, potentially influencing its enzymatic activity. Although the active sites may still remain engaged after the variation, the reduced stability of the variant proteins is a crucial factor that may inhibit their normal functionality. This destabilization could ultimately have an impact on treatment response, as the altered proteins may not perform optimally, potentially affecting treatment outcomes in clinical settings. These findings suggest that even small genetic variations can significantly impact protein stability and, as a result, therapeutic efficacy. This underscores the importance of further research to understand how these variations affect treatment responses in NSCLC patients.

The SNPs identified in this study hold significant clinical potential as biomarkers for patient stratification and treatment decision-making in NSCLC. Machine learning models prioritized key SNPs, including p.Ala168Val, p.His233Arg, p.Met418Leu, p.Tyr233His, and p.Ala311Thr, that were predictive of NR status at baseline. These prioritized variants may serve as early indicators of a patient's likelihood to respond to ICI therapy, enabling a more personalized approach to treatment selection. By integrating SNP profiling into clinical workflows, clinicians may be able to identify high-risk patients who may not benefit from standard ICI therapy and explore alternative or combination treatments.

This study has several limitations. First, the small sample size may limit the generalizability of the reported findings. Second, it focuses on a single cohort, which may not fully capture the variability and diversity across different populations. Additionally, while the study employs a longitudinal design, the limited number of timepoints prevents the establishment of reliable links between microbiome patterns and treatment outcomes. Furthermore, it is important to note that this work is based solely on in silico analyses, and the results need to be experimentally validated to confirm their biological relevance.

Future research should focus on experimental validation of these findings through in vitro and in vivo studies to confirm the biological impact of the identified SNPs and microbial strains on immune responses and treatment efficacy. Additionally, larger multi-cohort studies with extended follow-up periods are needed to establish reliable associations between microbiome shifts and treatment outcomes. Furthermore, the potential of microbiome-targeted interventions, such as probiotics, prebiotics, and fecal microbiota transplantation (FMT), in modulating immune responses remains underexplored. Evaluating whether these interventions can enhance immunotherapy efficacy and be integrated into NSCLC treatment protocols will be crucial in translating microbiome and SNP-based findings into clinical applications for improved patient outcomes.

Conclusion

This study underscores the association between gut microbial strains and genomic variants with the efficacy of immunotherapy in NSCLC patients. By integrating metagenomics, machine learning, and structural bioinformatics, we successfully prioritized microbial genes and variants that hold potential as biomarkers for determining NSCLC patient's suitability for ICI-based treatments. Nevertheless, these findings warrant validation through carefully designed experimental studies in the future.

Abbreviations

NSCL	Non-small cell lung cancer
ICIs	Immune checkpoint inhibitors
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
R	Responders
NR	Non-responders
GO	Gene ontology
FDR	False discovery rate
SNP	Single nucleotide polymorphism
MNP	Multiple nucleotide polymorphisms
INS	Insertions
DEL	Deletions
AUC	Area under the curve
SVM	Support vector machine
ROC	Receiver operating characteristic

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06370-0.

> Supplementary Material 1.
>
> Supplementary Material 2.

Author contributions

MFR performed the data analysis and prepared the manuscript. NK performed the structural analysis and assisted in manuscript preparation. HM and HMAT assisted in data analysis and manuscript preparation. MR guided the machine learning analysis. SR provided the resource for structural analysis. MRK conceived the research idea. MRK and LSH supervised the study. All authors reviewed and approved the final manuscript.

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 13 of 14

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Infectious Disease, Children'S Hospital, Zhejiang University School of Medicine, 3333 Binsheng Road, Binjiang District, 310052 Hangzhou, China. [2]Metagenomics Discovery Lab, School of Interdisciplinary Engineering & Sciences (SINES), National University of Sciences & Technology (NUST), Sector H-12, Islamabad 44000, Pakistan. [3]School of Interdisciplinary Engineering & Sciences (SINES), National University of Sciences & Technology (NUST), Sector H-12, Islamabad 44000, Pakistan. [4]National Clinical Research Center for Child Health, Children'S Hospital, Zhejiang University School of Medicine, 3333 Binsheng Road, Binjiang District, 310052 Hangzhou, China.

References
1. Lahiri A, Maji A, Potdar PD, Singh N, Parikh P, Bisht B, Mukherjee A, Paul MK. Lung cancer immunotherapy: progress, pitfalls, and promises. Mol Cancer. 2023;22(1):40. https://doi.org/10.1186/s12943-023-01740-y.
2. Li Y, Yan B, He S. Advances and challenges in the treatment of lung cancer. Biomed Pharmacother. 2023;169: 115891. https://doi.org/10.1016/j.biopha.2023.115891.
3. Mohamed E, García Martínez DJ, Hosseini M-S, Yoong SQ, Fletcher D, Hart S, Guinn B. Identification of biomarkers for the early detection of non-small cell lung cancer: a systematic review and meta-analysis. Carcinogenesis. 2024;45(1–2):1–22. https://doi.org/10.1093/carcin/bgad091.
4. Padinharayil H, Varghese J, John MC, Rajanikant GK, Wilson CM, Al-Yozbaki M, Renu K, Dewanjee S, Sanyal R, Dey A, Mukherjee AG, Wanjari UR, Gopalakrishnan AV, George A. Non-small cell lung carcinoma (NSCLC): implications on molecular pathology and advances in early diagnostics and therapeutics. Genes Diseases. 2023;10(3):960–89. https://doi.org/10.1016/j.gendis.2022.07.023.
5. Liu C, Yang M, Zhang D, Chen M, Zhu D. Clinical cancer immunotherapy: current progress and prospects. Front Immunol. 2022;13: 961805. https://doi.org/10.3389/fimmu.2022.961805.
6. Mishra V, Mishra Y. Role of gut microbiome in cancer treatment. Indian J Microbiol. 2024. https://doi.org/10.1007/s12088-024-01340-4.
7. Xin Y, Liu C-G, Zang D, Chen J. Gut microbiota and dietary intervention: affecting immunotherapy efficacy in non-small cell lung cancer. Front Immunol. 2024;15:1343450. https://doi.org/10.3389/fimmu.2024.1343450.
8. Duttagupta S, Hakozaki T, Routy B, Messaoudene M. The gut microbiome from a biomarker to a novel therapeutic strategy for immunotherapy response in patients with lung cancer. Curr Oncol. 2023;30(11):11. https://doi.org/10.3390/curroncol30110681.
9. Derosa L, Routy B, Thomas AM, Iebba V, Zalcman G, Friard S, Mazieres J, et al. Intestinal Akkermansia muciniphila predicts clinical response to PD-1 blockade in patients with advanced non-small-cell lung cancer. Nat Med. 2022;28(2):315–24. https://doi.org/10.1038/s41591-021-01655-5.
10. Ren S, Feng L, Liu H, Mao Y, Yu Z. Gut microbiome affects the response to immunotherapy in non-small cell lung cancer. Thoracic Cancer. 2024;15(14):1149–63. https://doi.org/10.1111/1759-7714.15303.
11. Zhang F, Ferrero M, Dong N, D'Auria G, Reyes-Prieto M, Herreros-Pomares A, Calabuig-Fariñas S, Duréndez E, Aparisi F, Blasco A, García C, Camps C, Jantus-Lewintre E, Sirera R. Analysis of the gut microbiota: an emerging source of biomarkers for immune checkpoint blockade therapy in non-small cell lung cancer. Cancers. 2021;13(11):11. https://doi.org/10.3390/cancers13112514.
12. Bae J, Park K, Kim Y-M. Commensal microbiota and cancer immunotherapy: harnessing commensal bacteria for cancer therapy. Immune Netw. 2022;22(1): e3. https://doi.org/10.4110/in.2022.22.e3.
13. Oh B, Boyle F, Pavlakis N, Clarke S, Eade T, Hruby G, Lamoury G, Carroll S, Morgia M, Kneebone A, Stevens M, Liu W, Corless B, Molloy M, Kong B, Libermann T, Rosenthal D, Back M. The gut microbiome and cancer immunotherapy: can we use the gut microbiome as a predictive biomarker for clinical response in cancer immunotherapy? Cancers. 2021;13(19):19. https://doi.org/10.3390/cancers13194824.
14. Zeriouh M, Raskov H, Kvich L, Gögenur I, Bennedsen ALB. Checkpoint inhibitor responses can be regulated by the gut microbiota—a systematic review. Neoplasia. 2023;43: 100923. https://doi.org/10.1016/j.neo.2023.100923.
15. Gunjur A, Shao Y, Rozday T, Klein O, Mu A, Haak BW, Markman B, Kee D, Carlino MS, Underhill C, Frentzas S, Michael M, Gao B, Palmer J, Cebon J, Behren A, Adams DJ, Lawley TD. A gut microbial signature for combination immune checkpoint blockade across cancer types. Nat Med. 2024;30(3):797–809. https://doi.org/10.1038/s41591-024-02823-z.
16. Zhang M, Liu J, Xia Q. Role of gut microbiome in cancer immunotherapy: from predictive biomarker to therapeutic target. Exp Hematol Oncol. 2023;12(1):84. https://doi.org/10.1186/s40164-023-00442-x.
17. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, Fluckiger A, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science (New York, NY). 2018;359(6371):91–7. https://doi.org/10.1126/science.aan3706.
18. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90. https://doi.org/10.1093/bioinformatics/bty560.
19. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower C, Franzosa EA, Segata N. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife. 2021;10: e65088. https://doi.org/10.7554/eLife.65088.
20. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. 2017. https://doi.org/10.1101/gr.216242.116.
21. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792. https://doi.org/10.1093/nar/gkh340.
22. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, Tickle TL, Weingart G, Ren B, Schwager EH, Chatterjee S, Thompson KN, Wilkinson JE, Subramanian A, Lu Y, Waldron L, Paulson JN, Franzosa EA, Bravo HC, Huttenhower C. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol. 2021;17(11): e1009442. https://doi.org/10.1371/journal.pcbi.1009442.
23. DeLano WL, et al. Pymol: an open-source molecular graphics tool. CCP4 Newsl Protein Crystallogr. 2002;40(1):82–92.
24. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 2006;34(Suppl_2):W116–8.
25. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Res. 2004;32(Suppl_1):D120–1.
26. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins Struct Funct Bioinf. 2006;62(4):1125–32.
27. Zhou F, Qiao M, Zhou C. The cutting-edge progress of immune-checkpoint blockade in lung cancer. Cell Mol Immunol. 2021;18(2):279–93. https://doi.org/10.1038/s41423-020-00577-5.

Raziq *et al. Journal of Translational Medicine*   (2025) 23:343

Page 14 of 14

28. Reck M, Remon J, Hellmann MD. First-line immunotherapy for non-small-cell lung cancer. J Clin Oncol. 2022;40(6):586–97. https://doi.org/10.1200/JCO.21.01497.
29. Lu Y, Yuan X, Wang M, He Z, Li H, Wang J, Li Q. Gut microbiota influence immunotherapy responses: mechanisms and therapeutic strategies. J Hematol Oncol. 2022;15(1):47. https://doi.org/10.1186/s13045-022-01273-9.
30. Souza VGP, Forder A, Pewarchuk ME, Telkar N, de Araujo RP, Stewart GL, Vieira J, Reis PP, Lam WL. The complex role of the microbiome in non-small cell lung cancer development and progression. Cells. 2023;12(24):24. https://doi.org/10.3390/cells12242801.
31. Kiousi DE, Kouroutzidou AZ, Neanidis K, Karavanis E, Matthaios D, Pappa A, Galanis A. The role of the gut microbiome in cancer immunotherapy: current knowledge and future directions. Cancers. 2023;15(7):7. https://doi.org/10.3390/cancers15072101.
32. Zhang H, Xu Z. Gut–lung axis: role of the gut microbiota in non-small cell lung cancer immunotherapy. Front Oncol. 2023;13:1257515. https://doi.org/10.3389/fonc.2023.1257515.
33. Chevallier M, Borgeaud M, Addeo A, Friedlaender A. Oncogenic driver mutations in non-small cell lung cancer: past, present and future. World J Clin Oncol. 2021;12(4):217–37. https://doi.org/10.5306/wjco.v12.i4.217.
34. Frick-Cheng AE, Shea AE, Roberts JR, Smith SN, Ohi MD, Mobley HLT. Iron limitation induces motility in uropathogenic E. coli CFT073 partially through action of LpdA. MBio. 2024;15(7):e01048-e1124. https://doi.org/10.1128/mbio.01048-24.
35. Chen Y, Ying Y, Lalsiamthara J, Zhao Y, Imani S, Li X, Liu S, Wang Q. From bacteria to biomedicine: developing therapies exploiting NAD+ metabolism. Bioorg Chem. 2024;142: 106974. https://doi.org/10.1016/j.bioorg.2023.106974.
36. Gorityala N, Baidya AS, Sagurthi SR. Genome mining of Mycobacterium tuberculosis: targeting SufD as a novel drug candidate through in silico characterization and inhibitor screening. Front Microbiol. 2024. https://doi.org/10.3389/fmicb.2024.1369645.
37. Kuivanen J, Biz A, Richard P. Microbial hexuronate catabolism in biotechnology. AMB Express. 2019;9:16. https://doi.org/10.1186/s13568-019-0737-1.
38. Perkins A, Mounange-Badimi MS, Margolin W. Role of the antiparallel double-stranded filament form of FtsA in activating the Escherichia coli divisome. MBio. 2024;15:e01687-e1724. https://doi.org/10.1128/mbio.01687-24.
39. Dewachter L, Verstraeten N, Jennes M, Verbeelen T, Biboy J, Monteyne D, Pérez-Morga D, Verstrepen KJ, Vollmer W, Fauvart M, Michiels J. A mutant isoform of ObgE causes cell death by interfering with cell division. Front Microbiol. 2017. https://doi.org/10.3389/fmicb.2017.01193.
40. Trichez D, Carneiro CVGC, Braga M, Almeida JRM. Recent progress in the microbial production of xylonic acid. World J Microbiol Biotechnol. 2022;38(7):7. https://doi.org/10.1007/s11274-022-03313-5.
41. Abbas R, Sorour N. Effect of dihydrolipoamide dehydrogenase LpdA3 gene knockout in sinorhizobium meliloti metabolism. J Agric Chem Biotechnol. 2016;7(7):193–9.
42. Singh VK, Sirobhushanam S, Ring RP, Singh S, Gatto C, Wilkinson BJ. Roles of pyruvate dehydrogenase and branched-chain α-keto acid dehydrogenase in branched-chain membrane fatty acid levels and associated functions in Staphylococcus aureus. J Med Microbiol. 2018;67(4):570–8.
43. Venugopal A, Bryk R, Shi S, Rhee K, Rath P, Schnappinger D, Ehrt S, Nathan C. Virulence of Mycobacterium tuberculosis depends on lipoamide dehydrogenase, a member of three multienzyme complexes. Cell Host Microbe. 2011;9(1):21–31.

Publisher's Note