Received: 10 September 2024 | Revised: 10 February 2025 | Accepted: 24 March 2025

DOI: 10.1002/cncr.35882

ORIGINAL ARTICLE

Artificial intelligence–driven microsatellite instability profiling reveals distinctive genetic features in patients with lung cancer

Quentin Dominique Thomas MD[1,2] | Julie Adèle Vendrell PhD[3] | Lakhdar Khellaf MD[4] | Sarah Cavaillon MD[1] | Xavier Quantin MD, PhD[1,2] | Jérôme Solassol MD, PhD[3] | Simon Cabello-Aguilar PhD[3,5]

[1]Department of Medical Oncology, Institute du Cancer de Montpellier, Montpellier University, Montpellier, France

[2]Oncogenic Pathways in Lung Cancer, Montpellier Cancer Research Institute, University of Montpellier, Montpellier, France

[3]Solid Tumor Laboratory, Department of Pathology and Oncobiology, Montpellier University Hospital Montpellier, Arnaud de Villeneuve Hospital, Montpellier, France

[4]Department of Pathology, Institute du Cancer de Montpellier, Montpellier University, Montpellier, France

[5]Montpellier BioInformatics for Clinical Diagnosis, Molecular Medicine and Genomics Platform, Montpellier University Hospital Montpellier, Montpellier, France

Correspondence
Quentin Dominique Thomas.
Email: quentin.thomas@icm.unicancer.fr

Abstract

Background: Microsatellite instability (MSI) has emerged as a predictive biomarker for immunotherapy response in various cancers, but its role in non–small cell lung cancer (NSCLC) is not fully understood.

Methods: The authors used the bioinformatics tool MIAmS to assess microsatellite status from next-generation sequencing (NGS) data using a tailored microsatellite score. Immunohistochemistry (IHC) assays were also performed to evaluate the correspondence between MSI and deficient mismatch repair (dMMR) status. A retrospective analysis of 1547 lung cancer patients was conducted, focusing on those with an MSI phenotype. Clinical characteristics, co-occurring molecular alterations, tumor mutation burden (TMB), and homologous recombination deficiency (HRD) status were evaluated in this subset.

Results: Of the 1547 patients analyzed, eight (0.52%) were identified as having MSI through MIAmS, with six (0.39%) of these cases also being dMMR on IHC. All patients with dMMR had an MS score ≥ 2 and a history of smoking. Most patients showed loss of MLH1 and PMS2 staining on IHC. No correlation was found between MSI status and programmed death-ligand 1 expression, although all MSI patients exhibited high TMB, averaging 21.4 ± 5.6 mutations per megabase.

Discussion: MSI/dMMR in lung cancer is exceedingly rare, affecting less than 1% of cases. NGS-based analysis combined with bioinformatics tools provides a robust method to identify MSI/dMMR patients, potentially guiding immunotherapy decisions. This comprehensive approach integrates molecular genotyping and MSI detection, offering personalized treatment options for lung cancer patients. NGS-based MSI testing is emerging as the preferred method for detecting microsatellite instability in various tumor types, including rare cancers.

KEYWORDS
microsatellite instability, neuroendocrine tumors, next-generation sequencing, non–small cell lung cancer

INTRODUCTION

Lung cancer is the main cause of cancer-related death worldwide.[1] Non–small cell lung cancer (NSCLC) is the most frequent histological type (80%–85%) and is frequently diagnosed at an advanced or metastatic stage.[2] Immunotherapy using immune checkpoint-blocking antibodies (ICIs) targeting programmed death-1 or its ligand (PD-1/PD-L1) has improved outcomes in patients with early-stage, locally advanced, and metastatic NSCLC.[3–6] PD-L1 expression in tumor cells (tumor proportion score [TPS]) has been identified as a predictive biomarker of response to immunotherapy in NSCLC.[7] However, the correlation between PD-L1 expression from TPS score and sensitivity to ICIs in NSCLC remains insufficient,[8] and the identification of additional biomarkers is a major challenge in distinguishing patients who are likely to experience a prolonged response to immunotherapy.

Microsatellite (MS) instability (MSI) has been established as a hallmark of cancer since its association with Lynch syndrome was identified in 1993.[9] The MSI phenotype is characterized by heterogeneity in the length of noncoding repeated DNA sequences. It reflects a deficiency in the DNA mismatch repair (dMMR) system, which is encoded by *MLH1*, *MSH2*, *MSH6*, and *PMS2*. Dysfunction of these genes leads to hypermutability, which is reflected by MSI. MSI is thus a tumor phenotype linked either to sporadic inactivation of the DNA mismatch repair system (methylation of the *MLH1* gene promoter inhibiting its expression), or to an inherited mutation of *MLH1*, *MSH2*, *MSH6*, or *PMS2*.[10] MSI and dMMR have been identified as predictive biomarkers of response to immunotherapy.[11,12] MSI tumors are frequently observed in colorectal and endometrial cancers but also occur in other tumor types albeit at lower frequencies.[13,14] Some studies have shown that MSI is rare in NSCLC, occurring in only a small subset (0.1%–1.0%).[15–17]

Immunohistochemistry (IHC) and polymerase chain reaction (PCR) are two commonly used techniques for identifying genetically unstable tumors. IHC examines histological slides for the absence of MSH1, MLH1, PMS2, and MLH6 expression. This technique provides an easy way of obtaining MS status, but does not take into account mutations leading to qualitative alterations in proteins. In contrast, PCR-based MSI testing requires a larger amount of tumor cells in the histological sample for DNA extraction than IHC but reporting on the genomic consequences of MMR protein loss-of-function. Although both techniques exhibit high concordance rates (90%–97%), they do not achieve 100% sensitivity.[18] Next-generation sequencing (NGS) is an emerging technique for diagnosing MSI tumors, although not currently recommended. Its advantage lies in concurrently assessing the mutational profile and MSI status.[19] Our group has previously developed and validated MIAmS,[20] a software tailored to determine the MS status of tumor samples from amplicon-based NGS assays.

In this study, we analyzed the MS status of over 1500 lung cancer samples, conducting a comprehensive examination of MSI by assessing patients' clinical characteristics and correlating NGS-based MS determination with IHC identification. Additionally, we present the co-occurring molecular alterations profiled by NGS, including the determination of tumor mutational burden (TMB) and homologous recombination deficiency (HRD) status.

MATERIALS AND METHODS

Patient and sample cohort

A total of 1547 formalin-fixed paraffin-embedded (FFPE) specimens from lung cancer samples, submitted from August 2019 to March 2023 to the University Hospital of Montpellier (France) for variant detection analysis by NGS, were retrospectively screened for MS status determination. Only samples with tumor cell percentages equal to or greater than 30%, as determined by a trained pathologist, were included in the study. This study was approved by the institutional review board of the University Hospital of Montpellier (IRB-MTP_2023_01_202201240) and was conducted in accordance with the Declaration of Helsinki. Information regarding clinicopathological data and response to treatment was retrieved from the medical records of patients classified as MSI/dMMR by our analysis.

DNA extraction and quantification

Genomic DNA was extracted using the Maxwell RSC DNA FFPE Kit (Promega, Madison, Wisconsin) according to the manufacturer's recommendations. Extracted DNA was quantified using the Qubit dsDNA High Sensitivity Assay Kit in combination with a Qubit fluorimeter (Thermo Fisher Scientific, Waltham, Massachusetts).

NGS analysis using the Advanta Solid Tumor NGS Library Prep Assay

Two libraries were prepared per sample using the Advanta Solid Tumor NGS Library Prep Assay with the automated Juno system on integrated fluidic circuits (LP 8.8.6 IFC) (Fluidigm, San Francisco, California) following the manufacturer's instructions. The panel was developed to explore 7 MS loci (Table S1) and detect somatic alterations in 49 oncology-relevant genes (245 kb, 1581 assays)

(Table S2). Briefly, LP 8.8.6 IFCs were primed with 20 ng DNA per sample in the PCR mix. After amplification, pooled harvested samples were purified using AMPure XP beads (Beckman Coulter, Brea, California) and a second PCR was performed to integrate the sequencing adapters. Libraries were then quantified, normalized and paired-end sequenced using a NextSeq instrument (2 × 150 cycles, Illumina, San Diego, California). Single nucleotide variants (SNVs) and small insertions/deletions (indels) were detected using an in-house bioinformatics workflow as previously described.[21] Interpretation of reported variants followed the American College of Medical Genetics (ACMG) recommendations and used public databases.[22] Only variants annotated as pathogenic, likely pathogenic or of uncertain significance were included in the study.

NGS-based MS status determination

A bioinformatics tool, MIAmS,[20] was used to assess the MS status. MIAmS uses machine-learning approaches and consists of four key steps. First, *MIAmS_learn* creates a model of the length distribution corresponding to MSI and MSS for each locus based on a learning data set. Subsequently, *MIAmS_tag* predicts the MS status from two independent classifiers, MSINGS[23] and SVCpairs. The first classifier operates by comparing the deviation with the maximum-matched normal model of peak distribution. The second classifier merges pairs of reads to eliminate fragments with indel errors stemming from the sequencing process. It classifies the length distributions of each locus using models of stable and unstable distributions, employing a support vector machine (SVM) sourced from scikit-learn.[24] Each classifier inside MIAmS assigns an MSI or MSS status to each sample, along with a confidence score ranging from 0 to 1. If the score falls below 0.5, the classifier designates the MS status as "Undetermined."

Deterministic ranking algorithm

For this study, we developed a specific MS_score and implemented a deterministic ranking algorithm. This algorithm operates as a rank-by-feature method, where each item is ranked based on its feature values. In our case, the features consist of four MS status predictions per sample, and the MS_score is calculated as follows:

$$\text{score}\begin{pmatrix}\text{MSI}\\ \text{MSS}\\ \text{undetermined}\end{pmatrix}=\begin{pmatrix}1\\ -1\\ 0\end{pmatrix}$$

$$\text{MS\_score}=\sum \text{score}(X_i)$$

with X, the status from MSINGS and SVCpairs and i = lib_1, lib_2.

The samples were then sorted based on their MS_score, with the top-ranking samples scoring above 0 classified as potentially MSI.

MMR immunohistochemistry

Immunohistochemistry was performed on FFPE 3-µm tissue sections using ready-to-use monoclonal antibodies (Agilent Technologies, Santa Clara, California) against MLH1 (clone ES05), MSH2 (clone FE11), MSH6 (clone EP49), and PMS2 (clone EP51). MMR deficiency was defined as complete negative nuclear staining for any MMR protein in the presence of positive internal controls (stromal cells and/or lymphocytes).

Mismatch and homologous recombination repair gene panel analysis and HRD status determination

Libraries were prepared using the Twist library preparation kits (Twist Biosciences, San Francisco, California) according to the manufacturer's instructions. Briefly, 50 ng of DNA per sample was subjected to enzymatic fragmentation, ligation of universal adapters, and indexation using eight amplification cycles. The pre-capture libraries were purified, qualified and quantified. A total of 1.6 µg of pooled indexed libraries (eight of 200 ng each) was hybridized to the biotinylated double-stranded DNA probe panel for 16 hours at 70°C. After binding of hybridized targets to streptavidin beads, post-capture libraries were washed, amplified, purified, quantified, normalized, and paired-end sequenced on a NextSeq instrument (2 × 150 cycles, Illumina). The custom panel used allowed the detection of alterations in the full coding sequence of 21 genes involved in DNA repair pathways (Table S3). SNV and small indels were also detected using an in-house bioinformatics workflow. Briefly, reads were trimmed with cutadapt (version 3.2)[25] and aligned to the human genome GRCh37 with BWA (version 0.7.17),[26] and the variant calling was performed using VarDict (version 1.8.2),[27] freebayes (version 1.3.2),[28] and mutect2 (version 4.1.4).[29] Variants detected by the three variant callers were kept and annotated with Variant Effect Predictor (version 101)[30] and reported. Variants with a frequency of 1% or more in the population from the GnomAD database[31] were considered as polymorphisms and were excluded. For the HRD status determination, shallow whole genome sequencing was performed from the pre-capture indexed libraries and sequenced on a NextSeq instrument (2 × 150 cycles, Illumina) at a depth coverage of ~1×. HRD status was determined using the *shallowHRD* bioinformatics pipeline.[32] Variant interpretation was done using Mobidetails[22] specifically, in silico splicing prediction was done using SpliceAI[33] and SPiP (10.1002/humu.24491). The semi-automated ACMG classification was done using GeneBe.[34]

TMB status determination

Libraries were prepared using the OncoDEEP DNA Kit (OncoDEEP, Twist Biosciences, San Francisco, California) according to the manufacturer's instructions. Briefly, 100 ng of DNA per sample was

subjected to enzymatic fragmentation, ligation of universal adapters and indexation using eight amplification cycles. The pre-capture libraries were purified, qualified, and quantified. A total of 2.5 μg of pooled indexed libraries (eight of 312 ng each) was hybridized to the biotinylated double-stranded DNA probe panel for 16 hours at 70°C. After binding of hybridized targets to streptavidin beads, post-capture libraries were washed, amplified, purified, quantified, normalized, and paired-end sequenced on a NextSeq instrument (2 × 74 cycles, Illumina). The OncoDEEP comprehensive gene panel is composed of probes targeting 638 genes. TMB status determination was determined using the OncoKDM platform.

RESULTS

Retrospective analysis of MS status from lung cancer patients

From August 2019 to March 2023, a cohort of 1547 patients diagnosed with lung cancer underwent comprehensive NGS analysis and was retrospectively screened for MS status determination. The most common histological subtype within this retrospective cohort was nonsquamous NSCLC, accounting for 93.4% of cases (n = 1445). Notably, a subset of eight individuals, constituting 0.52% of the entire cohort, exhibited a MS_score >0. Tissue samples were examined through anatomopathological assays to assess the dMMR status of these patients. IHC revealed that six of these eight patients (75.0%) had a dMMR status. It is worth noting that all six of these patients exhibited an MS_score ≥2 (Figure 1; Table 1).

Further characterization of the IHC staining patterns revealed that all but one patient showed loss of staining for both MLH1 and PMS2. The remaining patient had an isolated loss of MSH6 staining (Figure 2). These combined analyses allowed for the identification of six MSI/dMMR patients among the 1547 patients (0.39%). Regarding molecular alterations of patients with MSI/dMMR phenotype, all patients presented an alteration on MLH1 or MSH6 genes (pathogenic/nonfunctional or a variant of unknown significance [VUS]).

Co-occurring molecular alterations in the MSI/dMMR samples

To explore the genetic landscape of the MSI/dMMR phenotype, we determined the prevalence of co-occurring molecular alterations in the 6 MSI/dMMR samples, including SNV-indel detection, CNV detection, HRD status determination, and TMB calculation (Figure 3; Table S4).

All patients demonstrated either a nonfunctional *TP53* alteration consistent with the guidelines set by the International Agency for Research on Cancer (five of six; P2, P4, P6, P7, and P8), or a VUS (one of six; P3). This observation is consistent with the documented smoking history for all patients.[35] Notably, three out of the four alterations observed in *MLH1* (in P2, P3, and P6), which are presumed to account for the MSI/dMMR phenotype, were identified as G:C>T: A transversions commonly associated with smoking exposure (Table S4).[36,37]

All patients had cancer-related gene alterations, including alterations in the MAP-kinase and PI3K/AKT signaling pathways (Figure 3).

HRD status determination using the shallowHRD pipeline, resulted in three patients (P2, P3, and P4) showing a clear negative profile (HRD scores of 3, 4, and 3, respectively), two displaying intermediate profiles (P7 and P8 with HRD scores of 10 and 14, respectively), and one sample (P8) returning as HRD-positive (HRD score = 22). Notably, P7 harbored a pathogenic alteration in *BRCA2* associated with an intermediate HRD score, and P3 harbored a pathogenic *BRCA1* alteration associated with a clear negative HRD score.

The patients exhibited high TMB (21.4 ± 5.6 Mut/Mb), which correlates with the elevated number of pathogenic and VUS alterations observed in various cancer-related genes (ranging from four to 17 significant alterations per sample).

Clinical characteristics of patients with the MSI/dMMR phenotype

The clinical characteristics of the six patients with MSI/dMMR are described in Table 2 and Figure 3. Among them, 66.6% were male, and the median age was 62.2 years (41–83). Notably, all patients were current or former smokers. With regard to the stage of their disease, two patients (33.3%) were stages I/II, one patient (16.7%) was stage III, and three patients (50%) were stage IV. All patients were treated according to disease staging in accordance with international guidelines. One patient (P7) had a stage III lung adenocarcinoma with PD-L1 TPS expression of 2%. NGS analysis revealed multiple alterations, and the patient was treated with concomitant radiochemotherapy followed by durvalumab (an anti-PD-L1 ICI) in consolidation for 1 year. After a follow-up of 1 year, the patient did not experience tumor relapse. Another patient (P6) had a de novo metastatic lung adenocarcinoma with 80% PD-L1 TPS expression. The patient was treated in first metastatic line by chemo-immunotherapy combination, to which the patient responded favorably, achieving disease control since 2021. The remaining patients did not receive ICIs during the course of treatment. For instance, one patient (P4) had de novo metastatic large-cell neuroendocrine carcinoma and a history of alcoholic cirrhosis. The patient was treated with platinum-based doublet chemotherapy as first-line treatment. Unfortunately, his prognosis was poor, and he died 6 months after diagnosis of the disease.

DISCUSSION

This study revealed the presence of six cases (0.39%) of MSI/dMMR patients among a cohort of 1547 lung cancer patients. Our findings align with existing literature, indicating that MSI is rare in lung cancer, identifying five of 1445 (0.34%) NSCLC patients and one of 43

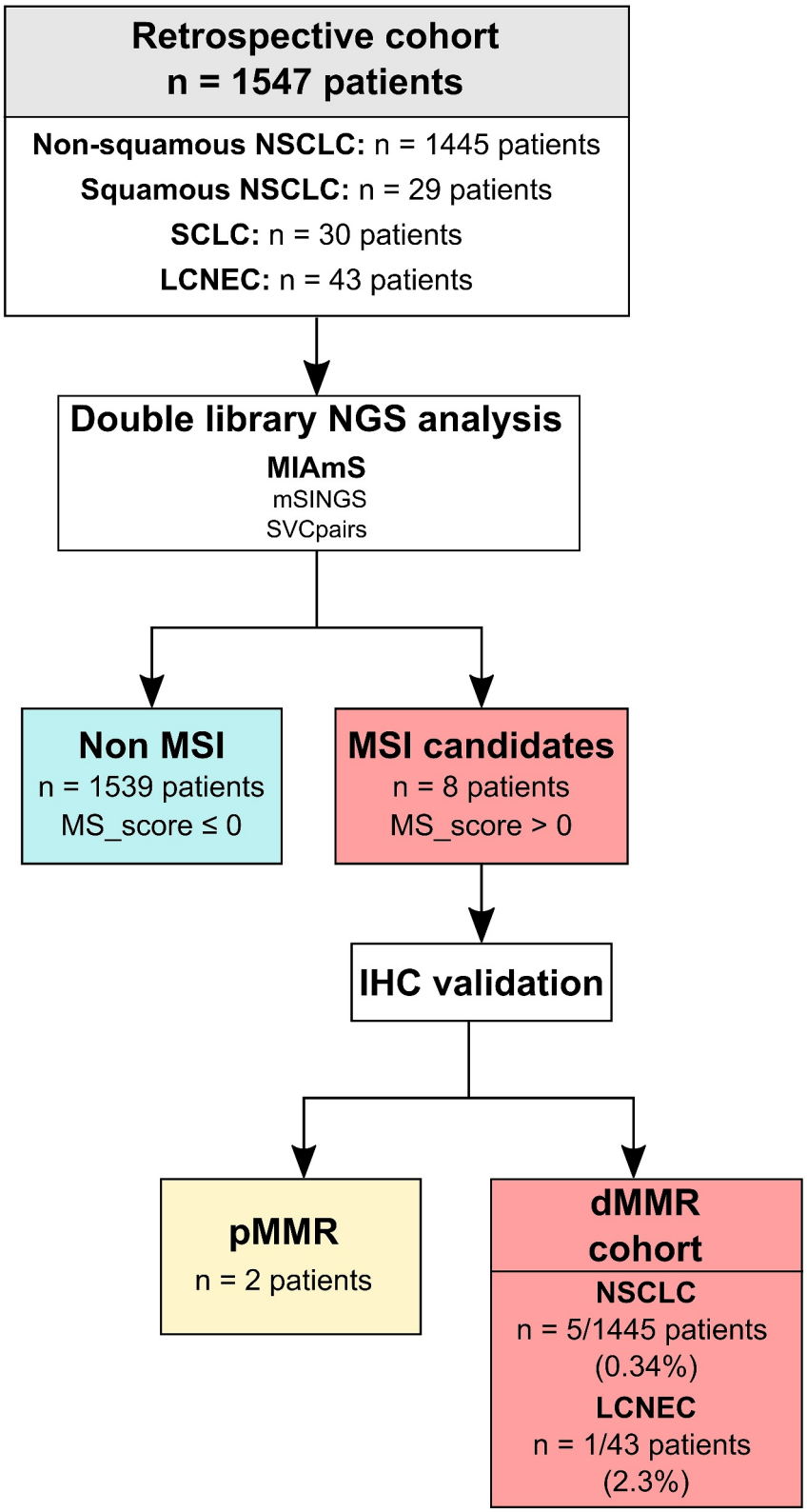

FIGURE 1 Workflow of the study. A double-library NGS analysis was conducted to assess the MSI status of 1547 lung cancer patients. High patients were identified with an MS_score >0. Among these patients, six had an MSI/dMMR phenotype, including five cases of NSCLC (0.34%) and one case of LCNEC (2.3%). dMMR indicates deficient mismatch repair; LCNEC, large cell neuroendocrine carcinoma of the lung; MSI, microsatellite instability; NGS, next-generation sequencing; NSCLC, non–small cell lung cancer.

(2.3%) neuroendocrine tumors with MSI.[15–17] Our investigation of the landscape of genetic alterations in lung cancer revealed compelling insights. In the sample set, four of six patients displayed *MLH1* alterations (P2, P3, P4, and P6), potentially elucidating the dMMR phenotype. Interestingly, three of these patients (P2, P4, and P6) exhibited alterations at a consensus splice site in *MLH1*. These alterations were classified as VUS because their impact on protein was not conclusively assessed through functional analysis. To provide more detail, in P2, the alteration affects the donor site. Although SpliceAI's prediction is inconclusive (confidence scores below 0.5), SPiP predicts a splice site alteration with a 98.41% risk. In P4, the variant is predicted to result in both acceptor and donor site loss, with confidence scores of 0.94 and 0.75, respectively. This variant is listed in ClinVar[38] as having "uncertain significance" and also has a semi-automated ACMG classification of "uncertain significance." Additionally, it is reported in LOVD[39] with a functional impact described as "probably affected." In P6, an alteration was identified that is predicted to cause acceptor site loss with a confidence score of 0.99, indicating a definitive loss of the acceptor site. The remaining patient (P3) exhibited an alteration that resulted in a stop codon in *MLH1* exon 6, which was classified as pathogenic, leading to the loss of MLH1 expression observed in IHC analysis. This MLH1 inactivation differs significantly from inactivation observed in sporadic tumors of the digestive or gynecologic tracts, where inactivation typically arises from hypermethylation of the *MLH1* gene promoter.[40] However, this mechanism was not observed in our cohort, as none of the patients exhibited hypermethylation of the *MLH1* gene promoter.

Patient 7 displayed several distinct tumor characteristics. First, the tumor sample exhibited 17 somatic alterations, the highest

TABLE 1 IHC and MS_score results.

Patient	MS_score	IHC	
P2	4	MLH1/PMS2 loss	dMMR
P3	4	MLH1/PMS2 loss	dMMR
P6	4	MLH1/PMS2 loss	dMMR
P7	4	MLH1/PMS2 loss	dMMR
P8	4	MSH6 loss	dMMR
P4	2	MLH1/PMS2 loss	dMMR
P1	1	—	pMMR
P5	1	—	pMMR

Abbreviations: dMMR, deficient mismatch repair; IHC, immunohistochemistry; MS, microsatellite instability; pMMR, proficient mismatch repair.

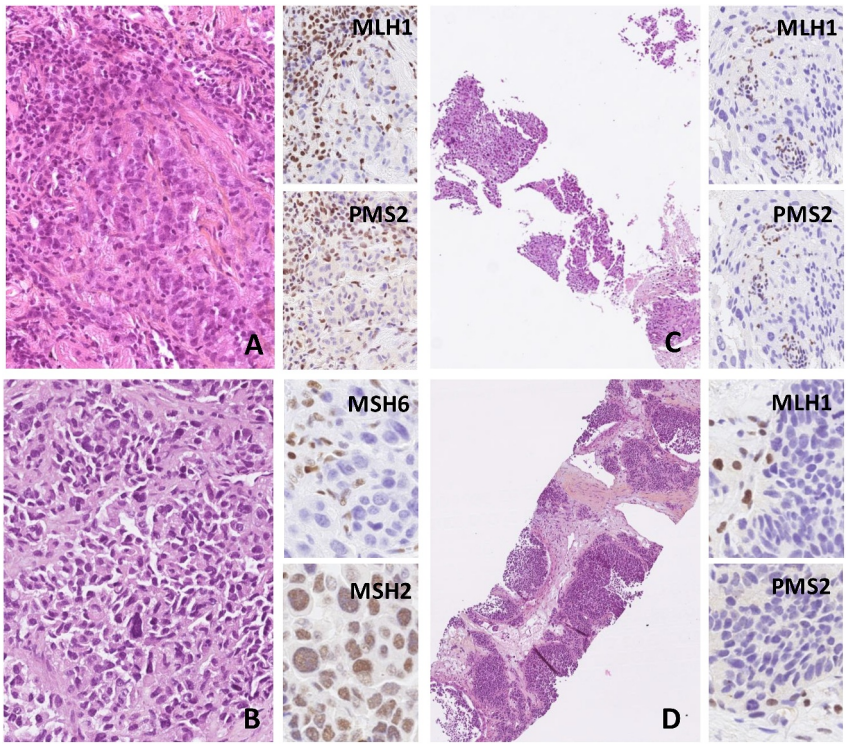

FIGURE 2 MMR immunohistochemistry results. (A) Lung adenocarcinoma, resection specimen, ×400. Loss of nuclear expression of the functional heterodimers MLH1 and PMS2. MS_score = 4. (B) Brain metastasis from lung adenocarcinoma, resection specimen, ×400. Isolated loss of nuclear MSH6 expression. Note the persistent expression of MSH2, which is a functional heterodimer. MS_score = 4. (C) Cervical lymph node metastasis from lung adenocarcinoma, biopsy needles, ×100. Loss of nuclear expression of MLH1 and PMS2. MS_score = 4. (D) Bone metastasis from a lung large-cell neuroendocrine carcinoma, biopsy needles, ×100. Loss of nuclear expression of MLH1 and PMS2. MS_score = 2. For each figure, internal controls are represented by normal immunostaining of stromal cells and lymphocytes for the MMR proteins tested. MMR indicates mismatch repair.

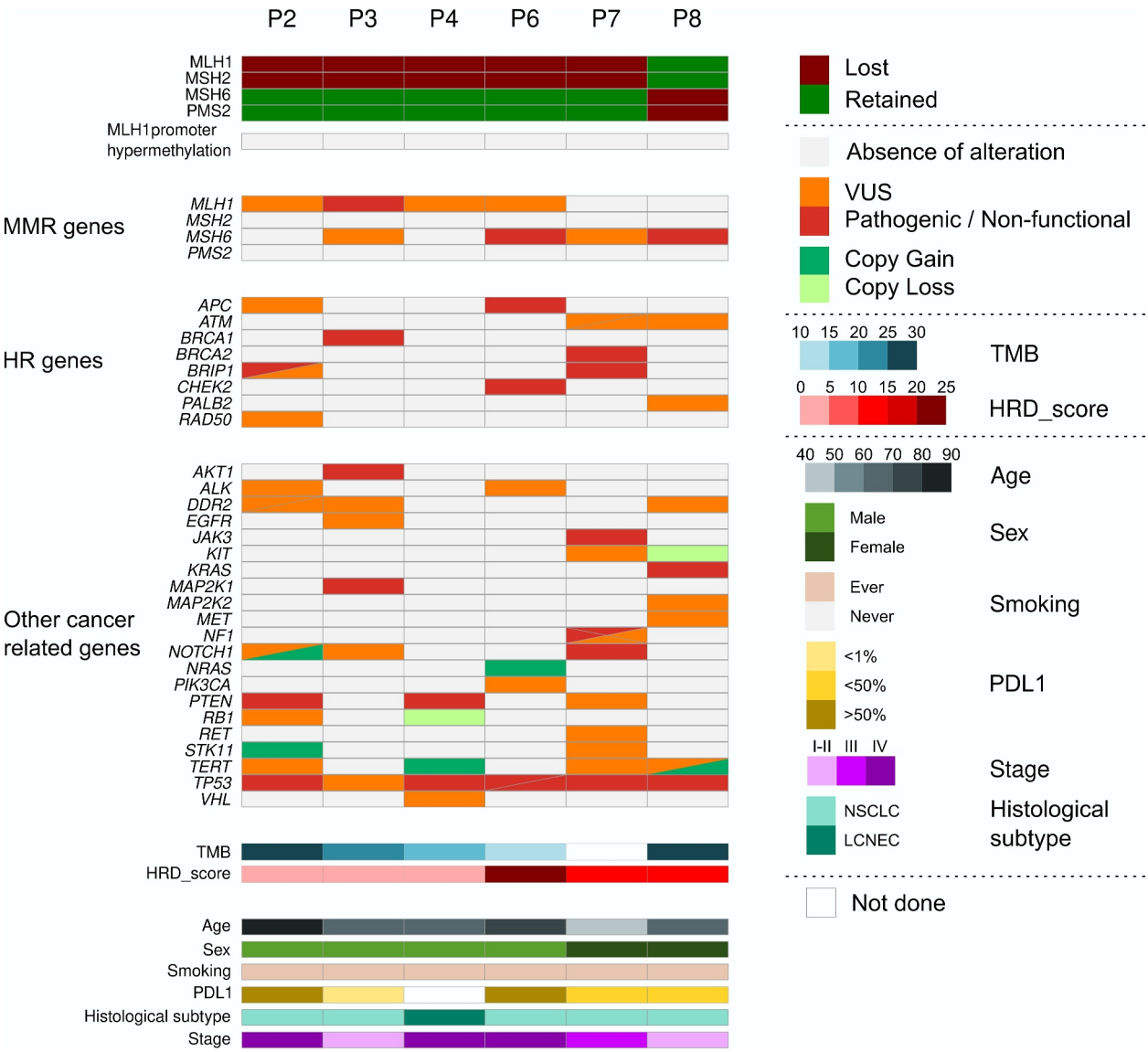

FIGURE 3 Clinicopathological and genomic characteristics. HR indicates homologous recombination; HRD, homologous recombination deficiency; LCNEC, large cell neuroendocrine carcinoma of the lung; MMR, mismatch repair; NSCLC, non–small cell lung cancer; PDL1, programmed death-ligand 1; TMB, tumor mutational burden; VUS, variant of unknown significance.

number observed in our cohort. These alterations included multiple changes in *ATM*, *NF1*, *JAK3*, *KIT*, *RET*, and *STK11*, which were classified as VUS or pathogenic. Furthermore, despite extensive analyses, we were unable to identify the genetic cause of the MSI/dMMR phenotype. The only notable alteration in the DNA MMR genes was a VUS in *MSH6*, which did not result in the loss of MSH6 expression, as confirmed by IHC staining. It is plausible that this tumor sample harbored a variant in a deep intronic region not covered by our panel, potentially affecting the splicing of *MLH1*. Finally, P7 showed pathogenic alterations in *BRCA2* and *BRIP1*, along with an intermediate HRD score of 11. Notably, the presence of pathogenic alterations in *BRCA1/2* in two samples (P3 and P7), along with their low or intermediate HRD scores, suggests a distinct mechanism compared to high-grade serous ovarian cancers, where *BRCA*-altered samples consistently display an HRD-positive score.[41] Alternatively, these alterations may represent subclonal passenger mutations. P6 also exhibited a distinctive genetic profile characterized by several notable features. P6 demonstrated a high HRD score, which is indicative of genomic instability. Additionally, P6 harbored a pathogenic alteration in *MSH6* without concurrent loss of MSH6 expression, as observed by IHC staining. Notably, P6 was the sole patient in our cohort to display a pathogenic alteration in *CHEK2*, a key cell-cycle regulator and potential tumor suppressor. CHEK2 plays a crucial role in DNA damage response by interacting with and phosphorylating BRCA1, thereby enabling BRCA1 to restore cell survival after DNA damage. This mechanism may offer a potential explanation for the elevated HRD score observed in this patient. The presence of four established or emerging biomarkers associated with positive

TABLE 2 Clinicopathological features of MSI/dMMR patients.

Characteristics	MSI/dMMR patients (n = 6)
Age (years)	
Median (range)	62 (41–83)
Gender	
Male	4 (66.6)
Female	2 (33.3)
Smoking status, No. (%)	
Current or former	6 (100)
Never	0 (0)
Median (pack-year)	33.0
Stage at diagnosis	
I–II	2 (33.3)
III	1 (16.7)
IV	3 (40.0)
Histology	
Adenocarcinoma	5 (83.3)
Squamous cell carcinoma	0 (0)
Sarcomatoid carcinoma	0 (0)
Large cell neuroendocrine carcinoma	1 (16.7)
PD-L1 tumor proportion score, No. (%); (n = 5)	
<1%	1 (25.0)
1%–49%	2 (50.0)
≥50%	1 (25.0)
MIAmS score, No. (%)	
1	0 (0)
2	1 (16.7)
3	0 (0)
4	5 (83.3)

Abbreviations: dMMR, deficient mismatch repair; MSI, microsatellite instability; PD-L1, programmed death-ligand 1.

responses to ICI, namely, the MSI/dMMR phenotype,[15] high HRD score,[42] high TMB,[43,44] and elevated PDL1 expression,[45] may explain the remarkable response of this patient to ICI since 2021.

According to Yang et al.,[15] MSI NSCLC had a higher median TMB (37.4 Mut/Mb) compared to MSS NSCLC (8.5 Mut/Mb) ($p < 10^{-4}$). Our TMB calculations in the cohort were consistent with these findings (21.4 ± 5.6 Mut/Mb). This observation aligns with the hypothesis that MMR defects lead to decreased accuracy in detecting and repairing DNA damage, resulting in higher TMB and an increased frequency of somatic alterations. However, we would like to highlight some distinctions using data from the KEYNOTE-158 study[46] that evaluated the association of TMB with clinical outcomes in patients with advanced solid tumors treated with pembrolizumab. In this cohort, among TMB-high patients, only 28 of 207 (13.5%) were identified as MSI. Furthermore, there were no MSI cases observed in the non-TMB-high control group. These findings emphasize the excellent negative predictive value (NPV) of TMB for MSI status, whereas its positive predictive value remains incomplete.

Additionally, TMB is not routinely used as a predictive biomarker of ICIs in daily clinical practice for NSCLC therapeutic strategies. A possible explanation for the limited predictive sensitivity of TMB in relation to ICIs efficacy lies in the dependency of ICI response on the presence of preexisting CD8+ T cells. In some contexts, the predictive utility of TMB is more reflective of high basal immune cell infiltration rather than a direct correlation with ICI sensitivity. For instance, a cohort of more than 10,000 tumor samples from The Cancer Genome Atlas was analyzed to evaluate TMB approaches and the correlation between predicted neoantigen load and CD8+ T-cell infiltration. In cancer types where no significant relationship was observed between CD8+ T-cell levels and neoantigen load (e.g., breast cancer, prostate cancer, and glioma), TMB-high tumors failed to achieve a 20% objective response rate (ORR) (OR = 15.3%; 95% confidence interval [CI], 9.2–23.4; p = .95) and demonstrated a significantly lower ORR compared to TMB-low tumors (OR = 0.46; 95% CI, 0.24–0.88; p = .02).[47]

Importantly, although no specific clinicopathological pattern emerged for patients with MSI/dMMR, all patients had a history of current or former smoking. This association between smoking exposure and MSI/dMMR is supported by the presence of G:C>T:A transversions, commonly linked to smoking in MMR genes. This underscores the need for further investigation into the relationship between smoking and MSI pathogenesis. Moreover, we had to notice that five of six patients from our cohort presented a lung adenocarcinoma.

Our study presents an innovative approach to MSI that uses NGS technology to advance diagnostic methodologies in this field. Although PCR and IHC are conventional methods for MSI detection, the strength of our study lies in the development and validation of a bioinformatics tool for identifying MSI through NGS analysis. To address the limitation of the learning phase of MIAmS lacking lung cancer cases, we introduced a deterministic ranking algorithm and an MS_score tailored specifically for these patients. The robustness of our ranking feature approach was confirmed, as the top six ranked patients (MS_score ≥2) were validated as dMMR through IHC testing. Additionally, we compared our results with MSISensor-pro,[48] an upgraded version of MSISensor[49] that enables tumor-only MSI detection. MSISensor-pro does not provide a predefined cutoff, as the optimal threshold varies depending on the sequencing panel. To use MSISensor-pro with a new panel for tumor-only MSI detection, a gold standard reference is required to determine a suitable cutoff specific to that panel. For comparison purposes, we applied the same loci and the commonly accepted criterion of three MSI loci to classify a sample as MSI. Despite MSISensor-pro not being as carefully optimized as MIAmS, it produced good results: all six dMMR samples were correctly identified as MSI by MSISensor-pro. However, its

results did not fully align with the IHC findings; one of three proficient mismatch repair (pMMR) samples was misclassified as MSI by MSISensor-pro, despite having an MS_score <1. This underscores the importance of carefully parameterizing and optimizing bioinformatics tools, particularly when analyzing rare events. These findings further emphasize the robustness and relevance of our approach.

The approval by the Food and Drug Administration of pembrolizumab (anti-PD-1) for the treatment of adult and pediatric patients with unresectable or metastatic MSI or dMMR solid tumors that have progressed following prior treatment and for whom no alternative treatment options are available represents a significant milestone.[50] As observed in our study, MSI and PD-L1 expression levels were not correlated, highlighting the importance of systematic determination of MS status. This is particularly relevant for neuroendocrine lung cancers, where MSI is more frequent, and ICIs are not the standard of care. Despite the rarity of MSI/dMMR lung cancers, this straightforward implementation could enable diagnostic laboratories to adopt such a solution if MS status proves to be valuable in patient care management. Indeed, the implementation of an MS status determination tool in an "all-in-one" bioinformatics solution minimally disrupts laboratory workflows and leads to cost-effective and efficient testing.

Notwithstanding the valuable insights gained, our study has several limitations. Immunotherapy data for patients in advanced or metastatic stages were lacking, preventing a comprehensive assessment of the potential benefits of ICIs in our cohort. In addition, retrospective analysis of MSI in NGS samples from samples analyzed in daily practice, predominantly those with NSCLC, may lead to an underestimation of the prevalence of MSI in patients with other types of lung cancer.

In conclusion, the occurrence of MSI/dMMR in lung cancer is exceedingly rare, accounting for less than 1% of all cases. Nevertheless, NGS-based MSI analysis has proven to be a valuable screening method, and the MS_score is a robust bioinformatics solution for this particular context.

AUTHOR CONTRIBUTIONS

Quentin Dominique Thomas: Conceptualization; visualization; writing—original draft; writing—review and editing. **Julie Adèle Vendrell**: Conceptualization; investigation; resources; writing—review and editing. **Lakhdar Khellaf**: Visualization; investigation; resources; validation. **Sarah Cavaillon**: Writing—review and editing; supervision. **Xavier Quantin**: Writing—review and editing; supervision. **Jérôme Solassol**: Conceptualization; project administration; writing—review and editing; supervision. **Simon Cabello-Aguilar**: Conceptualization; formal analysis; investigation; methodology; software; visualization; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

The authors thank chatGPT for helping us to improve the readability and language of the article. This research received no external funding. This study was approved by the institutional review board of the University Hospital of Montpellier (IRB-MTP_2023_01_20 2201240) and was conducted in accordance with the Declaration of Helsinki.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest

DATA AVAILABILITY STATEMENT

All information is available from the corresponding author on request.

ORCID

Quentin Dominique Thomas https://orcid.org/0000-0001-9516-9797

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Howlader N, Forjaz G, Mooradian MJ, et al. The effect of advances in lung-cancer treatment on population mortality. *N Engl J Med*. 2020;383(7):640-649. doi:10.1056/nejmoa1916623
3. Paz-Ares L, Luft A, Vicente D, et al. Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. *N Engl J Med*. 2018;379(21):2040-2051. doi:10.1056/nejmoa1810865
4. Patrick M F, Jonathan S, Shun L, et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N Engl J Med* 2022;386(21):1973-1985.
5. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *N Engl J Med*. 2017;377(20):1919-1929. doi:10.1056/nejmoa1709937
6. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:10.1056/nejmoa1801005
7. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016;387(10027):1540-1550. doi:10.1016/s0140-6736(15)01281-7
8. Cabezón-Gutiérrez L, Custodio-Cabello S, Palka-Kotlowska M, Alonso-Viteri S, Khosravi-Shahi P. Biomarkers of immune checkpoint inhibitors in non-small cell lung cancer: beyond PD-L1. *Clin Lung Cancer*. 2021;22(5):381-389. doi:10.1016/j.cllc.2021.03.006
9. Boland PM, Yurgelun MB, Boland CR. Recent progress in Lynch syndrome and other familial colorectal cancer syndromes. *CA Cancer J Clin*. 2018;68(3):217-231. doi:10.3322/caac.21448
10. Goel A, Boland CR. Epigenetics of colorectal cancer. *Gastroenterology*. 2012;143(6):1442-1460. doi:10.1053/j.gastro.2012.09.032
11. Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. 2017;357(6349):409-413. doi:10.1126/science.aan6733
12. Dudley JC, Lin MT, Le DT, Eshleman JR. Microsatellite instability as a biomarker for PD-1 blockade. *Clin Cancer Res*. 2016;22(4):813-820. doi:10.1158/1078-0432.ccr-15-1678
13. Middha S, Zhang L, Nafa K, et al. Reliable pan-cancer microsatellite instability assessment by using targeted next-generation sequencing data. *JCO Precis Oncol*. 2017;2017:PO.17.00084. doi:10.1200/po.17.00084

14. Akagi K, Oki E, Taniguchi H, et al. Real-world data on microsatellite instability status in various unresectable or metastatic solid tumors. *Cancer Sci*. 2021;112(3):1105-1113. doi:10.1111/cas.14798
15. Yang SR, Gedvilaite E, Ptashkin R, et al. Microsatellite instability and mismatch repair deficiency define a distinct subset of lung cancers characterized by smoking exposure, high tumor mutational burden, and recurrent somatic MLH1 inactivation. *J Thorac Oncol*. 2024;19(3):409-424. doi:10.1016/j.jtho.2023.10.004
16. De Marchi P, Berardinelli GN, Cavagna Rde O, et al. Microsatellite instability is rare in the admixed Brazilian population of non-small cell lung cancer: a cohort of 526 cases. *Pathobiology*. 2022;89(2):101-106. doi:10.1159/000520023
17. Yanagawa N, Yamada N, Sugimoto R, et al. The frequency of DNA mismatch repair deficiency is very low in surgically resected lung carcinoma. *Front Oncol*. 2021;11:752005. doi:10.3389/fonc.2021.752005
18. Luchini C, Bibeau F, Ligtenberg MJL, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol*. 2019;30(8):1232-1243. doi:10.1093/annonc/mdz116
19. Ratovomanana T, Cohen R, Svrcek M, et al. Performance of next-generation sequencing for the detection of microsatellite instability in colorectal cancer with deficient DNA mismatch repair. *Gastroenterology*. 2021;161(3):814-826. doi:10.1053/j.gastro.2021.05.007
20. Escudié F, Van Goethem C, Grand D, et al. MIAmS: microsatellite instability detection on NGS amplicons data. *Bioinformatics*. Published online October 24, 2019. doi:10.1093/bioinformatics/btz797
21. Blateau P, Coyaud E, Laurent E, et al. TERT promoter mutation as an independent prognostic marker for poor prognosis MAPK inhibitors-treated melanoma. *Cancers (Basel)*. 2020;12(8):2224. doi:10.3390/cancers12082224
22. Baux D, Van Goethem C, Ardouin O, et al. MobiDetails: online DNA variants interpretation. *Eur J Hum Genet*. 2021;29(2):356-360. doi:10.1038/s41431-020-00755-z
23. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. *Clin Chem*. 2014;60(9):1192-1199. doi:10.1373/clinchem.2014.223677
24. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(1532-4435):2825-2830. doi: 10.5555/1953048.2078195
25. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17(1):10-12. doi:10.14806/ej.17.1.200
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
27. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016;44(11):e108. doi:10.1093/nar/gkw227
28. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv*. 2012. http://arxiv.org/abs/1207.3907
29. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling somatic SNVs and indels with Mutect2 *bioRxiv*. 2019. https://www.biorxiv.org/content/10.1101/861054v1
30. McLaren W, Gil L, Hunt SE, et al. The Ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4
31. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7
32. Eeckhoutte A, Houy A, Manié E, et al. ShallowHRD: detection of homologous recombination deficiency from shallow whole genome sequencing. *Bioinformatics*. 2020;36(12):3888-3889. doi:10.1093/bioinformatics/btaa261
33. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535-548. doi:10.1016/j.cell.2018.12.015
34. Stawiński P, Płoski R. Genebe.net: implementation and validation of an automatic ACMG variant pathogenicity criteria assignment. *Clin Genet*. 2024;106(2):119-126. doi:10.1111/cge.14516
35. Gibbons DL, Byers LA, Kurie JM. Smoking, p53 mutation, and lung cancer. *Mol Cancer Res*. 2014;12(1):3-13. doi:10.1158/1541-7786.mcr-13-0539
36. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*. 2002;21(48):7435-7451. doi:10.1038/sj.onc.1205803
37. Husgafvel-Pursiainen K, Boffetta P, Kannio A, et al. p53 mutations and exposure to environmental tobacco smoke in a multicenter study on lung cancer. *Cancer Res*. 2000;60(11):2906-2911.
38. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980-D985. doi:10.1093/nar/gkt1113
39. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557-563. doi:10.1002/humu.21438
40. Veigl ML, Kasturi L, Olechnowicz J, et al. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci*. 1998;95(15):8698-8702. doi:10.1073/pnas.95.15.8698
41. Quesada S, Fabbro M, Solassol J. Toward more comprehensive homologous recombination deficiency assays in ovarian cancer part 2: medical perspectives. *Cancers (Basel)*. 2022;14(4):1098. doi:10.3390/cancers14041098
42. Zhou Z, Ding Z, Yuan J, et al. Homologous recombination deficiency (HRD) can predict the therapeutic outcomes of immuno-neoadjuvant therapy in NSCLC patients. *J Hematol Oncol*. 2022;15(1):62. doi:10.1186/s13045-022-01283-7
43. Ricciuti B, Wang X, Alessi JV, et al. Association of high tumor mutation burden in non-small cell lung cancers with increased immune infiltration and improved clinical outcomes of PD-L1 blockade across PD-L1 expression levels. *JAMA Oncol*. 2022;8(8):1160-1168. doi:10.1001/jamaoncol.2022.1981
44. Deng H, Zhao Y, Cai X, et al. PD-L1 expression and tumor mutation burden as pathological response biomarkers of neoadjuvant immunotherapy for early-stage non-small cell lung cancer: a systematic review and meta-analysis. *Crit Rev Oncol Hematol*. 2022;170:103582. doi:10.1016/j.critrevonc.2022.103582
45. So WV, Dejardin D, Rossmann E, Charo J. Predictive biomarkers for PD-1/PD-L1 checkpoint inhibitor response in NSCLC: an analysis of clinical trial and real-world data. *J Immunother Cancer*. 2023;11(2):e006464. doi:10.1136/jitc-2022-006464
46. Marabelle A, Fakih M, Lopez J, 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. *Lancet Oncol*. 2020;21(10):1353-1365. doi:10.1016/s1470-2045(20)30445-9
47. McGrail DJ, Pilié PG, Rashid NU, et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol*. 2021;32(5):661-672.

48. Jia P, Yang X, Guo L, et al. MSIsensor-pro: fast, accurate, and matched-normal-sample-free detection of microsatellite instability. *Genom Proteom Bioinform*. 2020;18(1):65-71. doi:10.1016/j.gpb.2020.02.001
49. Niu B, Ye K, Zhang Q, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics*. 2014;30(7):1015-1016. doi:10.1093/bioinformatics/btt755
50. Merck. FDA converts to full approval indication for Keytruda (pembrolizumab) for certain adult and pediatric patients with advanced microsatellite instability-high (MSI-H) or mismath repair deficient (dMMR) solid tumors. News release. Merck. 2023. Accessed March 29, 2023. https://www.merck.com/news/

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Thomas QD, Vendrell JA, Khellaf L, et al. Artificial intelligence–driven microsatellite instability profiling reveals distinctive genetic features in patients with lung cancer. *Cancer*. 2025;e35882. doi:10.1002/cncr.35882