


BMJ Open Artificial intelligence-derived gut microbiome as a predictive biomarker for therapeutic response to immunotherapy in lung cancer: protocol for a multicentre, prospective, observational study

Fumihiro Shoji ¹, Takanori Yamashita,² Fumihiko Kinoshita,¹ Shinkichi Takamori,¹ Takatoshi Fujishita,¹ Ryo Toyozawa,¹ Kensaku Ito,¹ Koji Yamazaki,³ Naoki Nakashima,² Tatsuro Okamoto¹

To cite: Shoji F, Yamashita T, Kinoshita F, *et al.* Artificial intelligence-derived gut microbiome as a predictive biomarker for therapeutic response to immunotherapy in lung cancer: protocol for a multicentre, prospective, observational study. *BMJ Open* 2022;**12**:e061674. doi:10.1136/bmjopen-2022-061674

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-061674>).

Received 02 February 2022
Accepted 30 May 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Thoracic Oncology, National Kyushu Cancer Center, Fukuoka, Japan

²Medical Information Center, Kyushu University, Fukuoka, Japan

³Department of Thoracic Surgery, National Hospital Organisation Kyushu Medical Center, Fukuoka, Japan

Correspondence to

Dr Fumihiro Shoji;
fumshojifumshoji@gmail.com

ABSTRACT

Introduction Immunotherapy is the fourth leading therapy for lung cancer following surgery, chemotherapy and radiotherapy. Recently, several studies have reported about the potential association between the gut microbiome and therapeutic response to immunotherapy. Nevertheless, the specific composition of the gut microbiome or combination of gut microbes that truly predict the efficacy of immunotherapy is not definitive.

Methods and analysis The present multicentre, prospective, observational study aims to discover the specific composition of the gut microbiome or combination of gut microbes predicting the therapeutic response to immunotherapy in lung cancer using artificial intelligence. The main inclusion criteria are as follows: (1) pathologically or cytologically confirmed metastatic or postoperative recurrent lung cancer including non-small cell lung cancer and small cell lung cancer; (2) age ≥ 20 years at the time of informed consent; (3) planned treatment with immunotherapy including combination therapy and monotherapy, as the first-line immunotherapy; and (4) ability to provide faecal samples. In total, 400 patients will be enrolled prospectively. Enrolment will begin in 2021, and the final analyses will be completed by 2024.

Ethics and dissemination The study protocol was approved by the institutional review board of each participating centre in 2021 (Kyushu Cancer Center, IRB approved No. 2021-13, 8 June 2021 and Kyushu Medical Center, IRB approved No. 21-076, 31 August 2021). Study results will be disseminated through peer-reviewed journals and national and international conferences.

Trial registration number UMIN000046428.

INTRODUCTION

Immunotherapies such as immune checkpoint inhibitors (ICIs) targeting programmed cell death-1 (PD-1), programmed cell death-ligand 1 (PD-L1) and cytotoxic T-lymphocyte-associated protein-4 (CTLA-4) are widely used to treat various malignancies, including lung cancer. Immunotherapy has changed therapeutic

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ The present study is a multicentre, prospective, observational study.
- ⇒ This is a study to discover the specific composition of the gut microbiome or combination of gut microbiome comprising a true predictive biomarker of therapeutic response to immunotherapy in patients with lung cancer using artificial intelligence.
- ⇒ This study could also perform the subgroup analysis for the different population enrolled such as non-small cell lung cancer versus small-cell lung cancer or mono-immunotherapy versus combined-immunotherapy.
- ⇒ As with all observational studies, any association may not be causal, and it will require evaluation in randomised controlled studies.

approaches to cancer treatment. Anti-PD-1 (eg, nivolumab, pembrolizumab), anti-PD-L1 (eg, atezolizumab, durvalumab) and anti-CTLA-4 antibodies (eg, ipilimumab) enable T-cell activation and induce an immune response to cancer. Many studies reported that cancer immunotherapies including monotherapies or combination regimens featuring platinum-based chemotherapy improved patient survival in both non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC).^{1–15} Nevertheless, there is no widely accepted optimal biomarker for predicting the efficacy of immunotherapy; therefore, it is necessary to establish a definite predictive biomarker of therapeutic response to cancer immunotherapy. A previous study reported that the gut microbiome can modulate the host immune response (eg, antitumour immunity) and optimise both innate and adaptive immune responses.¹⁶ Gori *et al*¹⁷ reviewed the association between the microbiome and host immunoregulation. They introduced that

some microbiome could sustain anticancer immunosurveillance through broadening T-cell receptor repertoire or enhancing immune response; on the other hand, some microbiome might suppress host immunity by interaction with T and Natural Killer cells.

Host immunity is clearly associated with the response to ICIs, and the internal microbiome is regarded as a controlling factor of host immunity, which supported our recent finding that the pretreatment host immunonutritional condition was a prognostic biomarker for patients with NSCLC who received immunotherapy.¹⁸ In fact, preclinical analyses illustrated that the gut microbiome composition and its modification in murine models could influence the efficacy of ICIs.^{19 20} Therefore, the microbiome has been emphasised as a predictive biomarker of immunotherapy mainly based on reports from the USA or Europe. In addition, the abundance of specific gut microbiome components or the gut microbiome diversity has also been reported to be related to the efficacy of anti-PD-1 antibodies in patients with melanoma.²¹ Moreover, faecal microbiome transplantation (FMT) in murine models potentially restored the response to ICIs.^{22 23} In a recent study, FMT from ICI responders to ICI non-responders resulted in ICI efficacy in patients with melanoma.²⁴

Remarkably, there are definite differences in the microbiome composition among ethnicities.²⁵ Thus, we recently reported that high gut microbiome diversity and the presence of specific microbes such as the genus *Blautia* and order RF32 unclassified were significantly correlated with the therapeutic response to immunotherapy in Japanese patients with NSCLC. However, the study had several limitations. First, the gut microbiome was examined 'during' immunotherapy opposed to 'before' treatment. Second, the sample size was small. Therefore, additional large multicentre studies to clarify the 'genus-level gut microbiome' as a predictive biomarker of cancer immunotherapy before treatment initiation are needed. In addition, the specific composition and combination of gut microbiome species that influence the response to immunotherapy are numerous, meaning there is limited ability to clarify the specific composition or significant combinations of gut microbiome species using statistical analysis via human intelligence.

Recently, the development of artificial intelligence (AI)-based technologies for medical data has progressed, and evidence has been discovered that could not be clarified by conventional methods.^{26 27} Such technologies are expected to be used for decision-making in medical fields. It is also expected that AI will be applied in the clinical treatment of lung cancer. Compared with conventional medical statistics methods, machine-learning models can more easily capture the non-linearity of medical data, and they exhibit higher prediction performance. These models are effective for diverse and large amounts of medical data, and extract contribution variables for the purpose.

In this analysis, we will conduct a prospective observational study to clarify the specific gut microbiome composition or combination of gut microbes associated with the

response to immunotherapy in patients with lung cancer using AI.

METHODS AND ANALYSIS

Study objectives

Our objective is to discover the specific composition of the gut microbiome or combination of gut species comprising a predictive biomarker of immunotherapy efficacy in patients with lung cancer using AI.

The study was approved by institutional review board of each participating centre.

Study setting

The present study is a multicentre, prospective, observational study.

Observational points

In the present study, we will clarify the following items:

1. The specific composition of the gut microbiome or combination of gut microbes associated with clinical/graphical therapeutic responses to immunotherapy.
2. The diversity of the gut microbiome associated with clinical/graphical therapeutic responses to immunotherapy.
3. The specific composition of the gut microbiome or combination of gut microbes associated with immune-related adverse events (irAEs, \geq grade 3).
4. The diversity of gut microbes associated with irAEs (\geq grade 3).
5. The association of the composition, combination or diversity of gut microbes with clinicopathological features including PD-L1 expression, haematological data, immunonutritional indices, the use of medicines (corticosteroids, antibiotics, proton pump inhibitors, and probiotics) within 28 days before the initiation of immunotherapy and patients' dietary habits such as the consumption of milk and fermented foods.
6. Progression-free survival (PFS), overall survival, efficacy (response rate, time to treatment and disease control rate) and safety according to the following items: type of pathology, the use of medicines at baseline, sex, age at baseline, Eastern Cooperative Oncology Group performance status at baseline, patients' dietary habits, regimens, the number of treatments and regimens after the failure of immunotherapy, the number of immunotherapy cycles completed, the best response to immunotherapy, tumour proportional score (tumourous PD-L1 expression), the past history of autoimmune disease, smoking history and the status of driver mutations.

Data collection

Faecal samples will be collected before the initiation of immunotherapy and analysed to assess the correlations of the clinical/graphical therapeutic response to immunotherapy, severity or types of irAEs and the continuation/discontinuation of treatment with

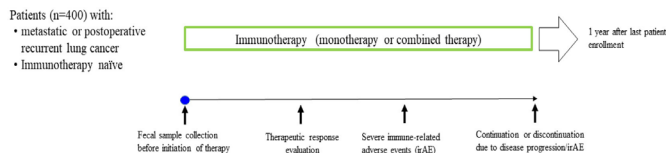


Figure 1 Study design. irAE, immune-related adverse event.

patients’ gut microbiomes. The present study design is presented in [figure 1](#) and [table 1](#).

Eligibility criteria

Inclusion criteria

- i. Pathologically or cytologically confirmed metastatic or postoperative recurrent NSCLC or SCLC.
- ii. ≥20 years old at the time of informed consent.
- iii. Planned treatment with immunotherapy including combination regimens and monotherapies as the first immunotherapy.
- iv. Ability to provide faecal samples.

- v. Provision of written informed consent.

Exclusion criteria

- i. Diseases eligible for definitive chemoradiotherapy.
- ii. Patients deemed inappropriate for the study by the investigator.

Sample size

A sample size and power calculation before study initiation were not performed because this is an exploratory observational study to clarify the predictive biomarker of therapeutic responses to immunotherapy. The present study will be conducted by two high-volume centres. Therefore, we estimate 100 participants per year in each centre. Thus, we plan to include total 400 patients over a 2-year period. This number will be sufficient to be analysed by AI, although there are few data of sample sizes recommended due to few studies similar to the present study. In total, 400 participants will be enrolled prospectively over 2 years.

Table 1 Summary of data collection and timeline

Items	Screening	Observation period		
		Pretreatment	During treatment	Discontinuation
Informed consent	●			
Patient’s background (sex, age)	●			
General conditions				
Body composition (height and weight)	●			
ECOG-PS	●			
Dietary habits	●			
Use of medicines	●			
Comorbidities	●			
Smoking status	●			
Laboratory data				
Haematology	●		●	●
Biochemistry	●		●	●
Imaging				
Chest-X ray, CT, MRI, PET	●		●	●
Pathology				
Histological type				
Gene mutation	●			
TPS (PD-L1 expression)	●			
Sample collection				
Faeces		●		
Treatment				
Regimens (immunotherapy)	●			
Immune-related adverse events			●	●
Prior Tx (regimen and response)	●			
Posterior Tx (regimen and response)			●	●

ECOG, Eastern Cooperative Oncology Group; PD-L1, programmed cell death-ligand 1; PET, positron emission tomography; PS, performance status; TPS, tumour proportion score; Tx, therapy.

Registration

Accrual started in September 2021.

Study period

The study started in September 2021 and will end 12 months after recruitment of the last participant, expected in September 2024.

Study population

We are planning to recruit eligible participants from participating hospitals. All enrolled patients will have at least one measurable target lesion on the Response Evaluation Criteria in Solid Tumours, V.1.1.28 Clinical/pathological stage will be based on the tumour node metastasis (TNM) classification established by the International Union Against Cancer.²⁹ For TNM staging, all patients will undergo CT of the thorax and upper abdomen, as well as bone scintigraphy, brain CT, MRI or fluorodeoxyglucose-positron emission tomography. Post-operative local or distant recurrence will be defined as described previously.³⁰ Immunotherapy will be continued until radiographic progression or discontinuation because of severe irAEs or the patient's request. PD-L1 protein expression will be evaluated using antibody clone 22C3 (Dako, Agilent Technologies, Santa Clara, California, USA). Adverse events will be graded according to Common Terminology Criteria for Adverse Events, V.5.0.

Treatment plan

The present study is allowing the immunotherapy regimens presented in [table 2](#).

Monotherapy: pembrolizumab will be administered at a dose of 200 mg intravenously every 3 weeks.^{1–3} Atezolizumab will be administered at a dose of 1200 mg intravenously every 3 weeks.^{4,5} Nivolumab will be administered at a dose of 3 mg/kg intravenously every 2 weeks.⁶

Combined therapy: KEYNOTE 189 regimen and KEYNOTE 407 regimen: carboplatin (area under the

concentration–time curve (AUC)) of 5–6 mg/mL/min plus pemetrexed 500 mg/m² (KEYNOTE 189) or paclitaxel 200 mg/m² (KEYNOTE 407). Maintenance therapy with pembrolizumab with/without pemetrexed (KEYNOTE 189) or pembrolizumab alone (KEYNOTE 407) is given every 3 weeks after induction therapy until disease progression, unacceptable toxicity, or death.^{7,8} IMpower130 regimen: atezolizumab (1200 mg intravenously every 3 weeks) plus carboplatin (AUC of 6 mg/mL/min every 3 weeks) plus nab-paclitaxel (100 mg/m² every week).⁹ IMpower132 regimen: atezolizumab (1200 mg intravenously every 3 weeks) plus cisplatin (75 mg/m²) or carboplatin (AUC of 6 mg/mL/min) plus pemetrexed 500 mg/m² every 3 weeks.¹⁰ IMpower150 regimen: atezolizumab (1200 mg intravenously every 3 weeks) plus carboplatin (AUC of 6 mg/mL/min for 4 cycles) plus paclitaxel (200 mg/m² intravenously every 3 weeks for 4 cycles) plus bevacizumab (15 mg/kg) intravenously every 3 weeks.¹¹ Maintenance therapy with atezolizumab plus bevacizumab (IMpower150), atezolizumab plus pemetrexed (IMpower132) or atezolizumab alone (IMpower130) is given every 3 weeks after induction therapy until disease progression, unacceptable toxicity, or death. CheckMate 227 regimen: nivolumab (3 mg/kg intravenously every 2 weeks) plus ipilimumab (1 mg/kg intravenously every 6 weeks).¹² CheckMate 9LA regimen: nivolumab (360 mg intravenously every 3 weeks) plus ipilimumab (1 mg/kg intravenously every 6 weeks) combined with histology-based platinum doublet chemotherapy (intravenously every 3 weeks for two cycles). The intravenous chemotherapy regimens consist of carboplatin (AUC of 6 mg/mL/min) plus paclitaxel (200 mg/m²) for patients with squamous histology or carboplatin (AUC of 5–6 mg/mL/min) or cisplatin (75 mg/m²) plus pemetrexed (500 mg/m²) for patients with non-squamous histology.¹³ Thereafter, treatment with nivolumab plus ipilimumab continues until disease progression or

Table 2 Immunotherapy regimens adopted in the present study

Study name	Regimens	References
KEYNOTE 010/024/042	○ Pembrolizumab	1–3
IMpower110/OAK	○ Atezolizumab	4 5
CheckMate 017/057	○ Nivolumab	6
KEYNOTE 189	○ Pembrolizumab + carboplatin/cisplatin + pemetrexed	7
KEYNOTE 407	○ Pembrolizumab + carboplatin + nab-paclitaxel	8
IMpower130	○ Atezolizumab + carboplatin + nab-paclitaxel	9
IMpower132	○ Atezolizumab + carboplatin/cisplatin + pemetrexed	10
IMpower150	○ Atezolizumab+ bevacizumab + carboplatin+ paclitaxel	11
CheckMate 227	○ Nivolumab+ ipilimumab	12
CheckMate 9LA	○ Nivolumab+ ipilimumab + carboplatin/cisplatin + pemetrexed	13
CheckMate 9LA	○ Nivolumab+ ipilimumab + carboplatin+ nab-paclitaxel	13
IMpower133	○ Atezolizumab + carboplatin + etoposide	14
CASPIAN	○ Durvalumab + carboplatin/cisplatin + etoposide	15

unacceptable toxicity. IMpower133 regimen: carboplatin (AUC of 5mg/mL/min intravenously on day 1) plus etoposide (100 mg/m² intravenously on days 1–3) plus atezolizumab (1200 mg intravenously every 3 weeks) for four cycles.¹⁴ Maintenance therapy with atezolizumab alone is given every 3 weeks after induction therapy until disease progression, unacceptable toxicity, or death. CASPIAN regimen: carboplatin (AUC of 5–6 mg/mL/min intravenously on day 1) or cisplatin (75–80 mg/m²) plus etoposide (80–100 mg/m² intravenously on days 1–3) plus durvalumab (1500 mg intravenously every 3 weeks) for four cycles.¹⁵ Maintenance therapy with durvalumab alone is given every 4 weeks after induction therapy until disease progression, unacceptable toxicity or death.

Sample collection, DNA extraction, gene amplification, sequencing and data analysis procedures

Faecal samples will be collected in sterile containers, immediately incubated at 4°C, and then frozen at –80°C. The preliminary treatment of faecal samples will follow a previously described method,³¹ followed by DNA extraction using an automated DNA isolation system (Gene Prep Star PI-480, Kurabo, Japan). DNA will be extracted from faeces using a Mora-Extract kit (Kyokuto Pharmaceutical, Japan). The V3–V4 regions of bacterial 16S rRNA genes will be amplified using the Pro341F/Pro805R primers³¹ and dual-index method³² under hemi-nested PCR conditions.³³ Barcoded amplicons will be paired-end sequenced on a 2×284bp cycle using the MiSeq system with MiSeq Reagent Kit chemistry, version 3 (600 cycle). Paired-end sequencing reads will be merged using the fastq-join programme with default settings.³⁴ The joined amplicon sequence reads will be processed through QIIME 2 V.2020.6.³⁵ The chimeric and low quality sequences were filtered using the DADA2 denoise-single plugin V.2017.6.0³⁶ based on default configuration. The taxonomy of representative sequences will be assigned using the Ribosomal Database Project Multi-classifier V.2.11.^{37 38} by training a naïve Bayes classifier using the q2-feature-classifier plugin. To account for compositional artefacts, we transformed relative abundances using the centred log-ratio transformation.³⁹ Alpha diversity indices (Chao1, Shannon and Simpson) will be calculated using the alpha-rarefaction plugin. The statistical significance of the Chao1, Shannon and Simpson indices among the groups will be assessed by the Kruskal-Wallis test using the alpha-group-significance plugin. Beta diversity will be analysed by the weighted UniFrac, unweighted UniFrac and Bray-Curtis distances using the core-metrics-phylogenetic plugin. The Emperor tool will be used to visualise principal coordinates analysis plots. The statistical significance of similarity of bacterial communities among the groups will be assessed by the ANalysis Of SIMilarities (ANOSIM) test using the beta-group-significance plugin. The heatmap and ward clustering from phylum to species will be presented using the feature-table heatmap plugin.

Machine-learning analysis

To develop predictive models, we will use gradient boosting decision trees (GBDTs).⁴⁰ Additionally, the feature variables obtained using GBDTs will be interpreted using SHapley Additive exPlanation (SHAP). The Shapley value is a solution concept of fairly distributing both gains and costs, and SHAP could also explain the output of any machine-learning model. These methods are the most powerful techniques for building and interpreting predictive models.^{41 42} All machine-learning analyses will be performed using the R programme.

Statistical analysis

Categorical variables will be analysed using Fisher's exact test. Continuous variables will be compared using the χ^2 test. The Mann-Whitney U test will be used to determine significant differences among the different groups using alpha diversity, which reveals the diversity in each individual sample. Logistic regression analysis will be performed to calculate ORs for the response to ICIs with respect to clinicopathological characteristics. Kaplan-Meier statistics and the log-rank test will be applied to evaluate PFS. Statistical analyses will be performed using JMP software, V.14.0 (SAS Institute). Significance will be indicated by $p < 0.05$.

Patients and public involvement

Patients and/or public were not involved in the design of the present study.

ETHICS AND DISSEMINATION

Study results will be disseminated through peer-reviewed journals and national and international conferences.

Acknowledgements We thank Joe Barber Jr, PhD, from Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript.

Contributors FS, FK, ST, TF, RT, KI, KY and TO significantly contributed to the conception and design of the study protocol. TY and NN provided statistical expertise for AI. The protocol was written by FS and critically reviewed by TO, and FS drafted the manuscript. All authors read and approved the final paper.

Funding This work was supported by Uehara Memorial Foundation, KAKENHI (No. 20K09188), Suzuken Memorial Foundation, and J-milk Foundation.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Consent obtained directly from patient(s)

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Fumihito Shoji <http://orcid.org/0000-0002-0636-8696>

REFERENCES

- 1 Reck M, Rodríguez-Abreu D, Robinson AG, *et al.* Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 2016;375:1823–33.
- 2 Reck M, Rodríguez-Abreu D, Robinson AG, *et al.* Updated Analysis of keynote-024: pembrolizumab versus platinum-based chemotherapy for advanced non-small-cell lung cancer with PD-L1 tumor proportion score of 50% or greater. *J Clin Oncol* 2019;37:537–46.
- 3 Mok TSK, Wu Y-L, Kudaba I, *et al.* Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet* 2019;393:1819–30.
- 4 Fehrenbacher L, von Pawel J, Park K, *et al.* Updated efficacy analysis including secondary population results for oak: a randomized phase III study of atezolizumab versus docetaxel in patients with previously treated advanced non-small cell lung cancer. *J Thorac Oncol* 2018;13:1156–70.
- 5 Herbst RS, Giaccone G, de Marinis F, *et al.* Atezolizumab for first-line treatment of PD-L1-selected patients with NSCLC. *N Engl J Med* 2020;383:1328–39.
- 6 Borghaei H, Gettinger S, Vokes EE, *et al.* Five-year outcomes from the randomized, phase III trials checkmate 017 and 057: nivolumab versus docetaxel in previously treated non-small-cell lung cancer. *J Clin Oncol* 2021;39:723–33.
- 7 Gandhi L, Rodríguez-Abreu D, Gadgeel S, *et al.* Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med* 2018;378:2078–92.
- 8 Paz-Ares L, Luft A, Vicente D, *et al.* Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. *N Engl J Med* 2018;379:2040–51.
- 9 West H, McCleod M, Hussein M, *et al.* Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 2019;20:924–37.
- 10 Nishio M, Barlesi F, West H, *et al.* Atezolizumab plus chemotherapy for first-line treatment of Nonsquamous NSCLC: results from the randomized phase 3 impower132 trial. *J Thorac Oncol* 2021;16:653–64.
- 11 Reck M, Mok TSK, Nishio M, *et al.* Atezolizumab plus bevacizumab and chemotherapy in non-small-cell lung cancer (IMpower150): key subgroup analyses of patients with EGFR mutations or baseline liver metastases in a randomised, open-label phase 3 trial. *Lancet Respir Med* 2019;7:387–401.
- 12 Hellmann MD, Paz-Ares L, Bernabe Caro R, *et al.* Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. *N Engl J Med* 2019;381:2020–31.
- 13 Paz-Ares L, Ciuleanu T-E, Cobo M, *et al.* First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (checkmate 9LA): an international, randomised, open-label, phase 3 trial. *Lancet Oncol* 2021;22:198–211.
- 14 Horn L, Mansfield AS, Szczesna A, *et al.* First-line atezolizumab plus chemotherapy in extensive-stage small-cell lung cancer. *N Engl J Med* 2018;379:2220–9.
- 15 Paz-Ares L, Dvorkin M, Chen Y, *et al.* Durvalumab plus platinum-etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell lung cancer (Caspian): a randomised, controlled, open-label, phase 3 trial. *Lancet* 2019;394:1929–39.
- 16 Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science* 2012;336:1268–73.
- 17 Gori S, Inno A, Belluomini L, *et al.* Gut microbiota and cancer: how gut microbiota modulates activity, efficacy and toxicity of antitumoral therapy. *Crit Rev Oncol Hematol* 2019;143:139–47.
- 18 Shoji F, Takeoka H, Kozuma Y, *et al.* Pretreatment prognostic nutritional index as a novel biomarker in non-small cell lung cancer patients treated with immune checkpoint inhibitors. *Lung Cancer* 2019;136:45–51.
- 19 Vétizou M, Pitt JM, Daillère R, *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* 2015;350:1079–84.
- 20 Sivan A, Corrales L, Hubert N, *et al.* Commensal bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* 2015;350:1084–9.
- 21 Matson V, Fessler J, Bao R, *et al.* The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* 2018;359:104–8.
- 22 Routy B, Le Chatelier E, Derosa L, *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* 2018;359:91–7.
- 23 Gopalakrishnan V, Spencer CN, Nezi L, *et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* 2018;359:97–103.
- 24 Davar D, Dzutsev AK, McCulloch JA, *et al.* Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients. *Science* 2021;371:595–602.
- 25 Nishijima S, Suda W, Oshima K, *et al.* The gut microbiome of healthy Japanese and its microbial and functional uniqueness. *DNA Res* 2016;23:125–33.
- 26 Dagliati A, Sacchi L, Tibollo V, *et al.* A dashboard-based system for supporting diabetes care. *J Am Med Inform Assoc* 2018;25:538–47.
- 27 Matsumoto K, Nohara Y, Soejima H, *et al.* Stroke prognostic scores and data-driven prediction of clinical outcomes after acute ischemic stroke. *Stroke* 2020;51:1477–83.
- 28 Eisenhauer EA, Therasse P, Bogaerts J, *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
- 29 Goldstraw P, Chansky K, Crowley J, *et al.* The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol* 2016;11:39–51.
- 30 Varlotto JM, Recht A, Flickinger JC, *et al.* Factors associated with local and distant recurrence and survival in patients with resected non-small cell lung cancer. *Cancer* 2009;115:1059–69.
- 31 Takahashi S, Tomita J, Nishioka K, *et al.* Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. *PLoS One* 2014;9:e105592.
- 32 Hisada T, Endoh K, Kuriki K. Inter- and intra-individual variations in seasonal and daily stabilities of the human gut microbiota in Japanese. *Arch Microbiol* 2015;197:919–34.
- 33 Hell K, Edwards A, Zarsky J, *et al.* The dynamic bacterial communities of a melting high Arctic glacier snowpack. *Isme J* 2013;7:1814–26.
- 34 Aronesty E. Comparison of sequencing utility programs. *Open Bioinforma J* 2013;7:1–8.
- 35 Bolyen E, Rideout JR, Dillon MR, *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37:852–7.
- 36 Callahan BJ, McMurdie PJ, Rosen MJ, *et al.* DADA2: high-resolution sample inference from illumina amplicon data. *Nat Methods* 2016;13:581–3.
- 37 Wolk K, Witte E, Witte K, *et al.* Biology of interleukin-22. *Semin Immunopathol* 2010;32:17–31.
- 38 Wu S, Li RW, Li W, *et al.* Worm burden-dependent disruption of the porcine colon microbiota by *trichuris suis* infection. *PLoS One* 2012;7:e35470.
- 39 Aitchison J. *The statistical analysis of compositional data*. London, New York: Chapman and Hall, 1986: 416 p.
- 40 Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189–123.
- 41 Nohara Y, Matsumoto K, Soejima H, *et al.* Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed* 2022;214:6584.
- 42 Yamashita T, Wakata Y, Nakaguma H, *et al.* Machine learning for classification of postoperative patient status using standardized medical data. *Comput Methods Programs Biomed* 2022;214:106583.