Data in Brief 44 (2022) 108556

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Datasets for gene expression profiles of head and neck squamous cell carcinoma and lung cancer treated or not by PD1/PD-L1 inhibitors

Jean-Philippe Foy[a,b,c,†,#,*], Andy Karabajakian[c,d,†], Sandra Ortiz-Cuaran[c,†], Maxime Boussageon[d,†], Lucas Michon[c], Jebrane Bouaoud[a,c], Dorssafe Fekiri[c], Marie Robert[c], Kim-Arthur Baffert[c], Geneviève Hervé[a,e], Pauline Quilhot[a,e], Valéry Attignon[c], Angélique Girod[a,b], André Chaine[a,b], Mourad Benassarou[a,b], Philippe Zrounba[f], Christophe Caux[c], François Ghiringhelli[g], Sylvie Lantuejoul[h], Carole Crozes[h], Isabelle Brochériou[a,e], Maurice Pérol[d], Jérôme Fayette[d], Chloé Bertolus[a,b,c], Pierre Saintigny[c,d,i,#,*]

[a] *Sorbonne Université, Paris, France*
[b] *Department of Maxillo-Facial Surgery, Hôpital Pitié-Salpêtrière, Assistance Publique des Hôpitaux de Paris, 47-83 boulevard de l'Hôpital, Paris 75013, France*
[c] *Univ Lyon, Université Claude Bernard Lyon 1, INSERM 1052, CNRS 5286, Centre Léon Bérard, Centre de recherche en cancérologie de Lyon, Lyon 69008, France*
[d] *Department of Medical Oncology, Centre Léon Bérard, 28 rue Laennec, 69373 Lyon cedex 08, Lyon 69008, France*
[e] *Department of Pathology, Hôpital Pitié-Salpêtrière, Assistance Publique des Hôpitaux de Paris, Paris, France*
[f] *Department of Surgery, Centre Léon Bérard, Lyon 69008, France*
[g] *Department of Medical Oncology, Centre Georges-François Leclerc, Dijon 21000, France*
[h] *Department of Pathology, Centre Léon Bérard, Lyon 69008, France*
[i] *Department of Translational Medicine, Centre Léon Bérard, Lyon 69008, France*

DOI of original article: 10.1016/j.ejca.2022.06.034
* Corresponding authors.
E-mail addresses: jean-philippe.foy@aphp.fr (J.-P. Foy), pierre.saintigny@lyon.unicancer.fr (P. Saintigny).
Social media: @LYriCAN_SIRIC (J.-P. Foy)
† equally contributed to this work
# Co-corresponding authors.

https://doi.org/10.1016/j.dib.2022.108556

ARTICLE INFO

Article history:
Received 3 August 2022
Revised 19 August 2022
Accepted 22 August 2022
Available online 27 August 2022

Keywords:
Head and neck cancer
Non-small cell lung cancer
Transcriptome profile
Biomarker
Immunotherapy

ABSTRACT

Identification of tumors harboring an overall active immune phenotype may help for selecting patients with advanced head and neck squamous cell carcinomas (HNSCC) and non-small cell lung cancer (NSCLC) who may benefit from immunotherapies. In this context, we generated targeted gene expression profiles in three and two independent cohorts of patients with HNSCC or NSCLC respectively, treated or not by PD-1/PD-L1 inhibitors. Notably, we generated two datasets including 102 and 82 patients with HNSCC or NSCLC treated with PD-1/PD-L1 inhibitors. Clinical information, including detailed survival raw data, is available for each patient, allowing to test association between gene expression data and patient survival (overall and progression-free survival). Moreover, we also generated gene expression datasets of 27 paired HNSCC samples from diagnostic biopsies and versus surgically resected specimens as well as 33 paired HNSCC samples at initial diagnosis (untreated) and at recurrence. Those datasets may allow to test the stability of a given biomarker across paired samples.

Specifications Table

Subject	Oncology
Specific subject area	Gene expression profiles were generated in head and neck or lung cancer samples in order to develop a biomarker to identify patients who may benefit from PD-1/PD-L1 inhibitors.
Type of data	Table Figure
How the data were acquired	Gene expression profiles of each sample were generated with the HTG EdgeSeq technology using the Oncology Biomarker panel. Sequencing was performed on an Illumina MISeq instrument using 150-cycle V3 kit. Clinical data were collected prospectively for some patients included in clinical trials testing PD-1/PD-L1 antibodies alone or in combination with other drugs [1] as well as retrospectively from medical records.
Data format	Raw (deposited in Gene Expression Omnibus) Analyzed (figures and tables included in this publication)
Description of data collection	Five cohorts of patients with HNSCC and NSCLC patients treated at Centre Léon Bérard (CLB, Lyon), at Centre Georges Francois Leclerc (CGFL, Dijon) and at Groupe Hospitalier Pitié-Salpêtrière (GHPS, Paris) in France, named CLB-OSCC, CLB-IHN, NIVOBIO, CLB-TUMADOR and GHPS, were used for gene expression profiling using the EdgeSeq Technology. Sequencing was performed on an Illumina MiSeq instrument. Log transformed Counts Per Million Standardization [log2(CPM)] was used to scale gene level data within a sample.
Data source location	Institution: Centre Léon Bérard • City/Town/Region: Lyon • Country: France • Latitude and longitude for collected samples/data: 45.7412392 and 4.8786326

(*continued on next page*)

	Institution: Groupe Hospitalier Pitié-Salpêtrière • City/Town/Region: Paris • Country: France • Latitude and longitude for collected samples/data: 48.840183 and 2.363221
Data accessibility	Raw and normalized data were deposited in Gene Expression Omnibus (GEO) (CLB-OSCC=GSE162519; CLB-IHN=GSE159067; NIVOBIO=GSE161537; CLB-TUMADOR=GSE162520 and GHPS=GSE159141). Repository name: Gene Expression Omnibus Data identification number: GSE162519; GSE159067; GSE161537; GSE162520 and GSE159141 https://www.ncbi.nlm.nih.gov/gds/?term=GSE162519 https://www.ncbi.nlm.nih.gov/gds/?term=GSE159067 https://www.ncbi.nlm.nih.gov/gds/?term=GSE161537 https://www.ncbi.nlm.nih.gov/gds/?term=GSE162520 https://www.ncbi.nlm.nih.gov/gds/?term=GSE159141
Related research article	Our data article supports an original research article published into the European Journal of Cancer [2]

Value of the Data

- To the best of our knowledge, the two cohorts of patients with advanced head and neck (n=102) and non-small cell lung cancer (n=82), included in the current publication, are the largest ones with both gene expression and clinical data including detailed survival data, available in the public domain.
- The complete set of gene expression profiles and associated clinical information, including detailed survival raw data for each patient, has been submitted to GEO, which will allow the community to freely access the data.
- Gene expression profiles could be reused for further insights into the development of biomarkers of response to PD-1/PD-L1 inhibitors as well as prognostic markers in patients with HNSCC or NSCLC

1. Data Description

1.1. Datasets

Five independent cohorts of patients with HNSCC or NSCLC, named CLB-OSCC, CLB-IHN, NIVOBIO, CLB-TUMADOR and GHPS, included HNSCC and NSCLC patients treated at CLB (Lyon), at CGFL (Dijon) and at GHPS (Paris) in France.

For each patient, targeted RNA sequencing was performed to generate gene expression profile. Raw and normalized data were deposited in Gene Expression Omnibus (GEO) (CLB-OSCC=GSE162519; CLB-IHN=GSE159067; NIVOBIO=GSE161537; CLB-TUMADOR=GSE162520 and GHPS=GSE159141).

1.2. Tables

The following tables described clinical and pathological characteristics of patients included in each cohort Tables 1–5.

Table 1
Clinico-pathological characteristics of patients included in the CLB-OSCC cohort [3].

Variables	Overall population (n=40)	
Age (Median, range)	64	34-91
Gender (N, %)		
Male	21	53%
Female	19	47%
Habits history (N, %)		
Smokers-Drinkers	21	53%
Non-smokers – Non-drinkers	19	47%
Oral subsite (N, %)		
Floor of mouth	11	28%
Tongue	10	25%
Buccal mucosa	10	25%
Alveolar ridge	8	20%
Hard palate	1	2%
Histological stage according to AJCC 8th edition (N, %)		
pT1	11	28%
pT2	14	35%
pT3	2	5%
pT4	13	32%
pN0	23	58%
pN1	7	17%
pN2a-c	10	25%
Adjuvant treatments (N, %)		
Radiotherapy	28	70%
Chemotherapy	14	35%
No	13	33%
Oncological outcomes (N, %)		
No relapse	29	73%
Local relapse	4	10%
Regional relapse	5	12%
Local + regional relapse	2	5%
Outcomes (median in months, range)		
Overall survival	34	2-117
Progression free survival	28	2-112
Status at last follow-up (N, %)		
No evidence of disease	24	60%
Death	10	25%
Progressive disease	4	10%
Relapse	2	5%

Abbreviations: SCC: squamous cell carcinoma; CLB-OSCC: Centre Léon Bérard-Oral Cavity Squamous Cell Carcinoma; CLB-IHN: Centre Léon Bérard-Immunotherapy for Head and Neck patients; HNSCC: Head and Neck Squamous Cell Carcinomas; OSCC: oral squamous cell carcinoma; NSCLC: Non-Small Cell lung Cancer; CLB: Centre Léon Bérard; GHPS: Groupe Hospitalier Pitié-Salpêtrière.

Table 2
Clinico-pathological characteristics of patients included in the CLB-IHN cohort [1].

Variables	Overall population (N=102)	
Age (Median, range)	63	33-88
Gender (N, %)		
Male	83	81%
Female	19	19%
Smoking history (N, %)		
Current/Former	87	85%
No	15	15%
Alcohol consumption (N, %)		
Current/Former	85	83%
No	17	17%
Disease site (N, %)		
Oral cavity	34	33%
Oropharynx	40	39%
Hypopharynx	14	13%
Larynx	11	11%
Cervical node with unknown primary	1	1%
Sinus cavities	2	2%
HPV status (N, %)		
Negative	49	48%
Positive	11	11%
Unknown	42	41%
ECOG Performance Status (N, %)		
0	22	22%
1	78	76%
≥ 2	2	2%
Number of previous lines of systemic treatments (N, %)		
0	45	44%
≥ 1	57	56%
Best response on immune checkpoint inhibitors (N, %)		
Complete response	5	5%
Partial response	6	6%
Stable disease	27	27%
Progressive disease	64	63%

Table 3
Clinico-pathological characteristics of patients included in the NIVOBIO cohort.

Variables	Overall population (N=82)	
Age (Median, range)	67	36-84
Institution		
Centre Léon Bérard	66	80%
Centre Georges-François Leclerc	16	20%
Gender (N, %)		
Male	53	64%
Female	29	36%
Smoking history (N, %)		
Active	13	16%
Former	63	77%
Never	6	7%

(*continued on next page*)

Table 3 (*continued*)

Variables	Overall population (N=82)	
Histological subtype (N, %)		
Adenocarcinoma	55	67%
Squamous cell carcinoma	17	20%
Large-cell neuroendocrine carcinoma	3	4%
Sarcomatoid carcinoma	3	4%
Adenosquamous	1	1%
NOS	3	4%
Oncogenic driver (N, %)		
ALK	1	1%
EGFR	4	5%
KRAS	19	24%
MET (exon 14 skipping)	1	1%
MET (amplification)	1	1%
ROS1	1	1%
None	55	67%
ECOG Performance Status (N, %)		
0	11	13%
1	50	61%
≥ 2	21	8%
First-line treatment (N, %)		
Carboplatin – pemetrexed	13	16%
Cisplatin – pemetrexed	27	33%
Carboplatin – gemcitabine	8	10%
Cisplatin – gemcitabine	6	7%
Carboplatin – paclitaxel	12	15%
Cisplatin – docetaxel	2	2%
Carboplatin – etoposide	3	4%
Cisplatin – vinorelbine	3	4%
Pemetrexed	2	2%
Crizotinib	2	2%
Erlotinib	1	1%
Gefitinib	1	1%
NA	2	2%
Number of previous lines of systemic treatments (N, %)		
0	2	3%
1	47	57%
≥ 2	33	40%
Type of immune checkpoint inhibitor (N, %)		
Nivolumab	77	94%
Pembrolizumab	5	6%
Best response on immune checkpoint inhibitors (N, %)		
Complete response	1	1%
Partial response	19	23%
Stable disease	22	27%
Progressive disease	34	41%
NA	6	8%

NOS: Not otherwise specified
NA: Not available/not applicable

Table 4
Clinico-pathological characteristics of patients included in the CLB-TUMADOR cohort.

Variables	Overall population (n=92)	
Age (Median, range)	65	37-87
Gender (N, %)		
Male	58	63%
Female	34	37%
Smoking history (N, %)		
Current/Former	80	87%
No	12	13%
Other chronic toxic or professional exposition (N, %)		
Alcohol	6	7%
Asbestosis	3	3%
Chemical bonding agents	3	3%
History of previous Cancer (N, %)		
Lung	2	2%
Head and neck	10	11%
Other (bladder, breast)	15	16%
Histology of the ongoing cancer (N, %)		
Adenocarcinoma	49	53%
Squamous cell carcinoma	43	47%
AJCC disease stages at first presentation (N, %)		
IA-IB	43	55%
IIA-IIB	13	17%
IIIA	15	19%
IV	7	9%
Non available (NA)	14	
Molecular alteration (N, %)		
EGFR mutation	14	15%
KRAS mutation	20	22%
BRAF mutation	3	3%
PI3KCA mutation	3	3%
MET mutation	2	2%
ALK rearrangement	2	2%
No (wild type)	42	46%
Surgical resection (N, %)		
Pneumectomy	14	15%
Lobectomy or Segmentectomy	78	85%
Adjuvant treatments (N, %)		
Chemotherapy	27	29%
Radiotherapy	18	20%
No	53	58%
Outcomes (median in months, range)		
Overall survival	34	2-117
Progression free survival	28	2-112
Status at last follow-up (N, %)		
Death	21	23%
Progressive disease	21	23%
Complete response	50	54%

Table 5
Clinico-pathological characteristics of patients included in the GHPS cohort.

Patient	N= (%)	Tumor	N= (%)
Gender		**Site**	
Female	12 (43)	Gingivomandibular	10 (36)
Male	16 (57)	Gingivomanxillar	3 (11)
Age		Mobile tongue	11 (39)
Median/mean	58/58	Floor of mouth	2 (7)
[min;max]	[25;79]	Buccal mucosa	1 (4)
Tobacco		Hard palate	1 (4)
current	13 (46)		
reformer	3 (11)	**pT stage**	
never	11 (39)	T1-T2	13 (46)
NA	1 (4)		
Alcohol		T3-T4	15 (54)
yes	7 (25)	**pN stage**	
no	15 (54)	Nx	2 (7)
NA	6 (21)	N0	17 (61)
Adjuvant radiotherapy		≥N1	9 (32)
yes	11 (39)	**Differentiation**	
no	3 (11)	Well	13 (46)
NA	14 (50)	Moderate	13 (46)
		Poor	1 (4)
		NA	1 (4)
		Perineural invasion	7 (25)
		Lymphovascular invasion	7 (25)
		Extranodal spread	
		Yes	4 (14)
		NA	1 (4)
		Margin	
		negative	15 (54)
		Positive or close	13 (46)

1.3. Figures

The following figures showed survival distribution in relation with biomarkers of response to PD-1/PD-L1 inhibitors, including the Hot Oral Tumor (HOT) score, as previously defined [2], computed in each sample from the different cohorts Figs 1–4.

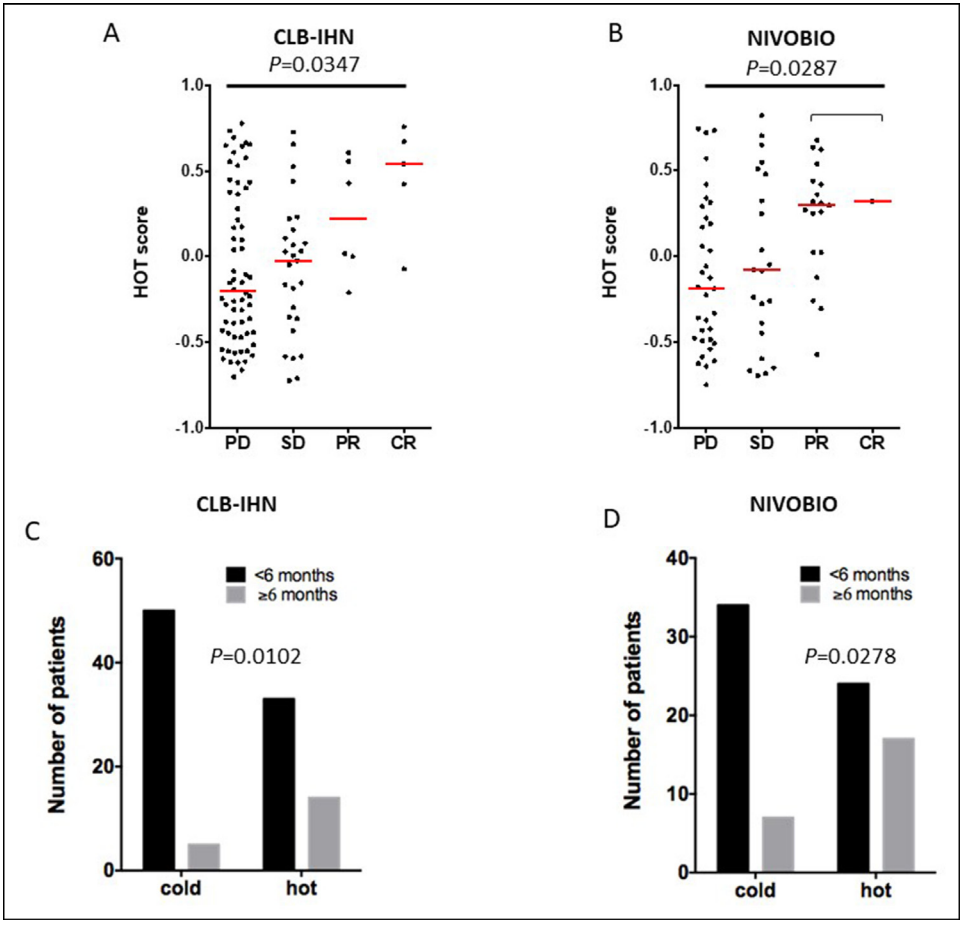

Fig. 1. HOT score and response to PD1/PD-L1 inhibitors in the CLB-IHN and NIVOBIO cohorts. The HOT score was compared between patients according to the objective response in the CLB-IHN (A) and NIVOBIO (B) cohorts (Kruskall Wallis Test). The proportion of hot/cold tumors was compared between patients with PFS<6 months and patients with PFS≥6 months in the CLB-IHN (C) and NIVOBIO (D) cohorts (Fisher's exact test). PD: Progressive disease. SD: stable disease. PR: partial response. CR: complete response.

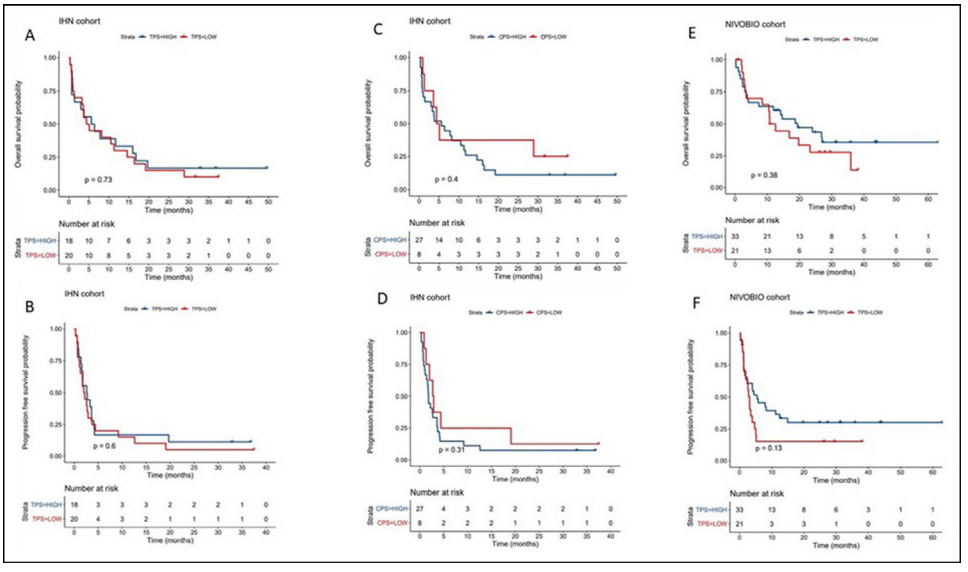

Fig. 2. Association of the CPS/TPS with survival of patients with HNSCC/ NSCLC treated with PD-1/PD-L1 inhibitor. The Tumour Proportion score (TPS) and the Combined Proportion Score (CPS) were computed in 38 and 35 recurrent/metastastic head and neck squamous cell carcinomas (HNSCC) from the CLB-IHN cohort of patients treated by anti-PD-1/PD-L1. Overall survival (OS) and progression-free survival (PFS) were compared between HNSCC with a TPS>1% (high) and tumors with TPS<1% (low) (A-B), as well as between tumors with a CPS>1% (high) and tumors with CPS <1% (low) (C-D) using a log-rank test. The TPS was also computed in 54 advanced non-small cell lung cancer (NSCLC) from the NIVOVIO cohort of patients treated by anti-PD-1/PD-L1 inhibitors. OS and PFS were compared between tumors with a TPS>1% (high) and tumors with TPS<1% (low) (E-F) using a log-rank test.

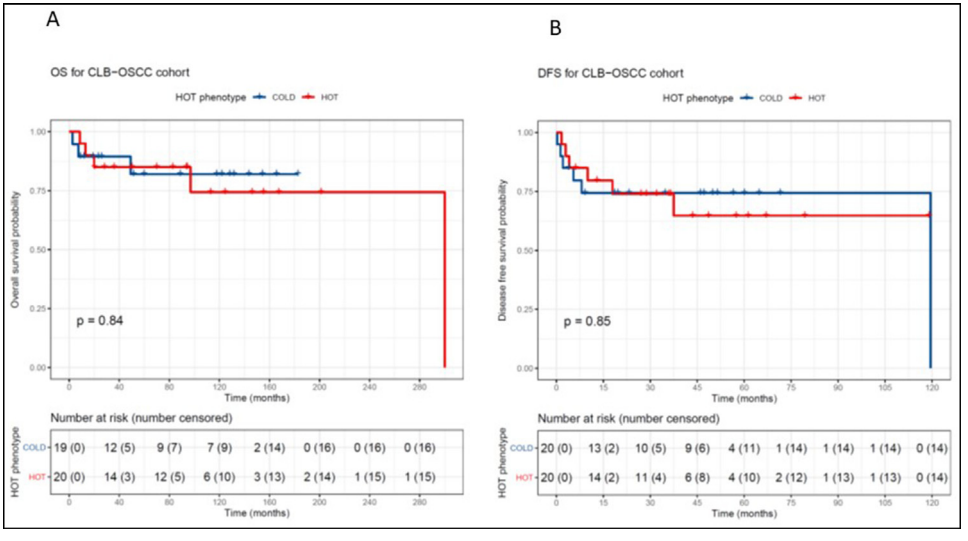

Fig. 3. HOT phenotype and survival of patients from the CLB-OSCC cohort. The HOT score was computed in 40 HPV-negative OSCC from the CLB-OSCC cohort in order to classify them as 'hot' (positive score) or 'cold' (negative score) tumors. Overall survival (OS) (A) and progression-free survival (PFS) (B) were compared between 'hot' and 'cold' tumors, using a log-rank test. OSCC: oral squamous cell carcinoma.

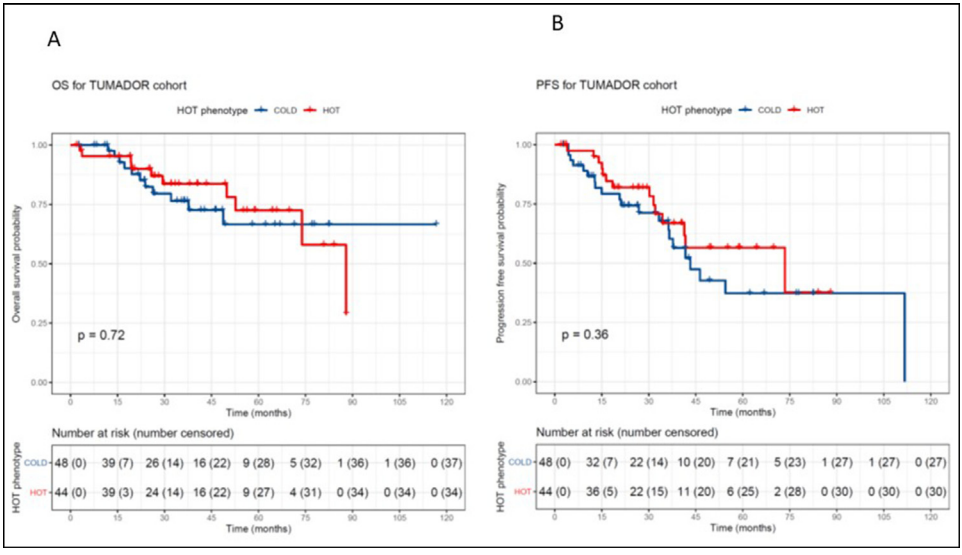

Fig. 4. HOT phenotype and survival of patients with NSCLC from the TUMADOR cohort. The HOT score was computed in 92 NSCLC from the TUMADOR cohort to classify them as either 'hot' (positive score) or 'cold' (negative score) tumors. Overall survival (OS) (A) and progression-free survival (PFS) (B) were compared between hot and cold tumors, using a log-rank test. NSCLC: non-small cell lung cancer.

2. Experimental Design, Materials and Methods

2.1. Patients and samples

The CLB-IHN cohort is derived from a previously published cohort of 120 patients treated at CLB for a histologically confirmed R/M HNSCC in clinical trials testing PD-1/PD-L1 antibodies alone or in combination with an anti-KIR or an anti-CTLA4 antibody between March 2014 and November 2018 [1]. Targeted gene expression profiles were generated in a total of 102/120 (85%) patients who had at least one available FFPE tumor sample.

The NIVOBIO cohort retrospectively included 138 patients with locally advanced or metastatic NSCLC, which were treated with either nivolumab (≥ second-line) or pembrolizumab (front-line) between May 2016 to February 2018, at CLB (Lyon, France) and Georges-François Leclerc Cancer Center (Dijon, France). Targeted RNA-sequencing using the HTG technology was performed in 82/138 patients with at least one available FFPE-sample from a pre-treatment biopsy.

The TUMADOR cohort is a retrospective cohort of 92 patients who have undergone surgical resection for early stage NSCLC at CLB.

In the GHPS cohort, a total of 28 patients suffering from OSCC and treated at the Groupe Hospitalier Pitié-Salpêtrière (GHPS, Paris, France) between November 2017 and August 2018 were included. For each sample, the HTG EdgeSeq technology (see below) was used to generate targeted gene expression profiles in per-endoscopic tumor biopsy and paired surgical resection specimens (n=56 samples). P16 immunostaining was performed for all patients.

2.2. RNA-targeted sequencing using the HTG EdgeSeq system

In FFPE tumor samples from each cohort, we performed targeted-RNA sequencing at Centre Léon Bérard, using the HTG EdgeSeq technology [4] that allows sequencing of 2,559 oncology-related biomarker genes from only one 5µm section per sample.

2.2.1. Probe and assay design

HTG's nuclease protection probes are generated using a proprietary algorithm. Briefly, target-specific protection sequences are designed to be 50 nucleotides in length, to have optimal and balanced T_ms, and undergo sequence-based screening for uniqueness to eliminate potential cross-reactivity with other probes or other RNAs in the human genome. Each probe has 5' and 3' extensions, called wings, of 25 bases each. All 5' wings are identical and all 3' wings are identical. The final length of each nuclease protection probe in this assay is 100 nucleotides. Probes were ordered from IDT with standard desalting and were gel-purified on a 10% PAGE gel in a pool prior to use. We used the Oncology Biomarker Panel in this work (2,559 transcripts).

2.2.2. Sample preparation

All samples were received as 5 micron FFPE sections on glass slides. For each sample, tumor area was delimitated by our senior pathologist in order to perform molecular analysis in this area only. For each sample, all tumor areas were scraped into 1.7 ml tubes and Lysis Buffer (HTG) was added to a concentration of 1-1.2 mm2 per µl (for resection samples) or ~18 cells per µl (for CNBs). 500 µl Denaturation Oil (HTG) was added to each tube and tubes were heated to 95°C for 15 minutes. Tubes were cooled to room temperature and Proteinase K (Ambion) was added at a 1:20 ratio in relation to the amount of Lysis Buffer. Samples were incubated at 50°C with gentle shaking for 2 hours. Once lysed, samples were stored at -70°C until use. Following lysis, FFPE samples were diluted to a final concentration of 5 mm^2 per 25 µl in Lysis Buffer. Each sample was individually assigned to a well of a 96-well sample plate and 25 µl of sample was added per well. The sample plate was loaded into an HTG Edge Processor for the nuclease protection steps.

2.2.3. Nuclease protection assay

Nuclease protection was performed on an HTG Edge Processor. Briefly, each sample was overlain with Denaturation Oil and a probe cocktail containing nuclease protection probes, wingmen, and spike-in controls (HTG) was added to each sample well. The sample plate was heated to 85°C for 10 min for denaturation, followed by 16 hours of hybridization at 50°C. Digestion Solution (HTG), which contains S1 Nuclease (Promega) was added and incubated for 90 min at 52°C. The nuclease reaction was stopped by transferring the reaction to Termination Solution (HTG) in a V-bottom stop plate and heating the plate to 100°C for 20 minutes. The plate was allowed to cool to room temperature before continuing.

2.2.4. PCR tagging and library cleanup

Following nuclease protection, PCR was used to add sequencing adapters and barcodes to the probes in each sample. For each sample, a separate 30 µl PCR reaction was set up comprising 15 µl OneTaq 2x Master Mix with GC Buffer (New England Biolabs), 3 µl of sample, 3 µl each of 5 µM forward and reverse primers (HTG), and 9 µl dH_2O. Forward and reverse primers will prime off of the wings of each probe, and carry Illumina's paired-end sequencing adapters as well as a unique six-base barcode. PCR cycle conditions were 95°C for 4min, 20 cycles of (95°C 15sec, 56°C 45sec, 68°C 45sec), followed by a final extension for 10min at 68°C. As each sample was individually tagged during PCR, reactions for a given experiment were pooled to form a single sequencing library. The pooled PCR product was cleaned up using AMPure beads (Beckman Coulter) at 2.5x with one modification. AMPure beads were added to the cleanup at a constant 35 µl and the cleanup mixture was supplemented with PEG 8000 and NaCl to a final concentration of 9% and 1.13 M, respectively. Each cleanup was checked on a 2% agarose gel for the absence of primer. qPCR was performed to quantitate the cleaned library compared to a known standard. Each 20 µl qPCR reaction contained 10 µl of 2x SybrSelect Master Mix (ThermoFisher Scientific), 200 nM each primer, 4 µl of diluted library (1:10,000 dilution in 10 mM Tris-HCl pH 8.0, 0.1% Tween-20), and 0.25 µM additional ROX dye (Thermo Fisher Scientific). Primers amplify Illumina sequencer adapters (F primer (5′-3') AAT GAT ACG GCG ACC ACC GA and R primer (5' to 3') R: 5′-CAA GCA GAA GAC GGC ATA CGA). The standard curve was made up of five 10-fold serial dilutions of known concentrations. Cycling was performed on a StepOne Plus instrument

(ThermoFisher Scientific) in Standard mode with conditions of 95°C for 5 min followed by 30 cycles of 95°C for 30 sec, 65°C for 45 sec. Readings were taken at the end of the 65°C step. All qPCR reactions were run in triplicate and an average of resulting values was used to calculate the library concentration.

2.2.5. Sequencing and data extraction

Sequencing was performed on an Illumina MiSeq instrument using 150-cycle V3 kits. Sequencing libraries were loaded at 30 pM with a 5% PhiX spike-in. Standard Illumina protocols were used for library denaturation and loading except that denaturation of libraries was done in two steps; a standard NaOH denaturation followed by a heat-denaturation step at 98°C for 4 min and snap-chill on ice for 5 min. 50 cycles of sequencing were performed with two 6-base barcode reads. The sequencer performed demultiplexing and fastq files were returned. The HTG EdgeSeq parser was used to align the probe sequences to the results; this program is a front end for bowtie2 software [5], using a 25-base match with one mismatch allowed to the first 25 bases of sequencing information. The final data table is a compilation of all such counts per probe per sample.

Log transformed Counts Per Million Standardization [log2(CPM)] was used to scale gene level data within a sample.

2.3. Bioinformatics and statistics

The Gene Set Variation Analysis (GSVA) r package is a non-parametric unsupervised method for assessing gene set enrichment in gene expression microarray and RNA-seq data [6]. Unlike other methods that analyze differential pathways between two phenotypical groups, the GSVA tool allows for computing an enrichment score (ES) of a given gene set in each sample, with a bimodal distribution of ES. Default parameters were used (abs.ranking=false, tau=1). This method, based on the Kolmogorov-Smirnov (KS) like random walk statistic, allows to produce a bimodal distribution of enrichment scores, by generating non-zero maximum deviations under the null distribution. Contrary to alternative method based on a 'competitive' hypothesis, GSVA is based on 'self-contained null hypothesis' which analyzes each gene set in isolation, assessing differential expression of the gene set without comparing to a background [7,8]. As described by the authors, the main strength of GSVA lies in its capabilities for analyzing single samples. Compared to other methods, GSVA is also a robust method for survival analysis [6]. In the study on the HOT score [2], we performed a careful selection of highly expressed immune genes to define the 'Hot Oral Tumor' (HOT) signature, allowing us to compute the HOT score with the GSVA tool. The GSVA method produces positive and negative scores with a bimodal distribution between -1 and 1, corresponding to largest positive and negative random walk deviations from zero. Thus, based on those considerations, we defined samples with positive ($0 <$score HOT <1) and negative ($-1<$score HOT <0) scores, as hot and cold respectively.

In the CLB-OSCC and CLB-TUMADOR cohorts, overall survival (OS) time was defined by the time in months from tumor biopsy to death or loss to follow-up and progression-free survival (PFS) time was defined by the time in months from tumor biopsy to death, recurrence, or loss to follow-up. In the CLB-IHN and NIVOBIO cohorts, OS time was defined as the period from the date of initial treatment administration to the date of mortality from any cause or the last follow-up and PFS time was defined as the period from the date of initial treatment administration to the date of clinical disease progression, mortality from any cause or the last follow-up.

Ethics Statements

The study was conducted in accordance with all applicable laws, rules, and requests of French and European government authorities, including the patient's informed consents. The use of samples in research projects has been reviewed by a multidisciplinary committee. The samples were properly codified, so that in no case the Recipients are able to identify the donor's

identities, or any clinical information that may be used for the donor's identification. The CRB Centre Léon Bérard (n°BB-0033-00050) is quality certified according to the NFS96-900 French standard and ISO 9001 for clinical trials, ensuring scientific rigor for sample conservation, traceability, and quality, as well as ethical rules observance and defined rules for transferring samples for research purposes (ministry of health for activities authorization n° AC-2019-3426 and DC-2008-99). Ethical approval for the study was obtained from the Ethics committee of the CRB Centre Léon Bérard (Reference number: 2018-001 and 2018-018).

CRediT Author Statement

All authors critically reviewed iterations of the manuscript and approved the final draft for submission. JPF, AK, SOC, MB, CB and PS contributed to the study conception, design, or planning. JPF, AK, SOC, MB, JB, LM, MR, DF and KAB gathered data. JPF and PS did or supervised data analysis. JPF, AK, SOC, MB, CB and PS interpreted the data. AK, SOC, MB, KAB, GH, PQ, VA, AG, AC, MB, PZ, CCa, FG, SL, CCr, IB, MP, JF, provided study materials or patients. LM did data deposition on Gene Expression Omnibus. JPF and LM performed bioinformatical and statistical analysis. JPF and PS wrote sections of the initial manuscript.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

PS is a member of HTG Diagnostics Scientific Advisory Board and receives research grants from HTG Diagnostics, Inivata, ArcherDx, Bristol-Myers Squibb, Roche Molecular Diagnostics, Roche, AstraZeneca, Novartis, Bristol-Myers Squibb Foundation and Illumina; JF reports grants, personal fees and non-financial support from Bristol-Myers Squibb, personal fees and non-financial support from MSD, personal fees from Merck, personal fees and non-financial support from Astrazeneca, personal fees from Rakuten, personal fees from Biogen and personal fees from Innate Pharma; SL received advisory board honoraria from AbbVie, AstraZeneca, Bayer, Bristol-Myers Squibb, Merck Sharp and Dohme, Takeda, Roche/Genentech; FG reports personal grants consultancy for Roche, AstraZeneca, and payments for development of educational presentations for Roche, Servier, Amgen, Merck. All remaining authors have declared no conflicts of interest.

Data availability

GSE159141 (Original data) (Gene Expression Omnibus).
GSE162520 (Original data) (Gene Expression Omnibus).
GSE161537 (Original data) (Gene Expression Omnibus).
GSE162519 (Original data) (Gene Expression Omnibus).
GSE159067 (Original data) (Gene Expression Omnibus).

Acknowledgments

We would like to acknowledge Reine Guibert and Morgane Toubon, who are oncology nurse coordinators at Groupe Hospitalier Pitié-Salpêtrière (Paris, France), as well as Jessie Auclair and Carole Audoynaud for their excellent work in producing HTG gene expression profiling at Centre Léon Bérard, Platform of Cancer Genomics (Lyon, France)

Funding

This work was funded by a core grant from the Integrated Cancer Research Site LYriCAN (INCa SIRIC-LYriCAN INCa-DGOS-Inserm_12563) and Fondation d'Entreprise Bristol-Myers Squibb pour la Recherche en Immuno-Oncologie. Gene expression profiling was funded in part through the collaboration of HTG Molecular Diagnostics, Inc. with Centre Léon Bérard.

References

[1] A. Karabajakian, T. Garrivier, C. Crozes, N. Gadot, J.Y. Blay, F. Bérard, P. Céruse, P. Zrounba, P. Saintigny, C. Mastier, J. Fayette, Hyperprogression and impact of tumor growth kinetics after PD1/PDL1 inhibition in head and neck squamous cell carcinoma, Oncotarget 11 (2020) 1618–1628, doi:10.18632/oncotarget.27563.

[2] JP. Foy, A. Karabajakian, S. Ortiz-Cuaran, M. Boussageon, L. Michon, J. Bouaoud, D. Fekiri, M. Robert, K-A. Baffert, G. Hervé, P. Quilhot, V. Attignon, A. Girod, A. Chaine, M. Benassarou, P. Zrounba, C. Caux, F. Ghiringhelli, S. Lantuejoul, C. Crozes, I. Brochériou, M. Pérol, J. Fayette, C. Bertolus, P. Saintigny, Immunologically active phenotype by gene expression profiling is associated with clinical benefit from PD-1/PD-L1 inhibitors in real-world head and neck and lung cancer patients, Eur. J. Cancer. (2022), doi:10.1016/j.ejca.2022.06.034.

[3] J.-P. Foy, C. Bertolus, M.-C. Michallet, S. Deneuve, R. Incitti, N. Bendriss-Vermare, M.-A. Albaret, S. Ortiz-Cuaran, E. Thomas, A. Colombe, C. Py, N. Gadot, J.-P. Michot, J. Fayette, A. Viari, B. Van den Eynde, P. Goudot, M. Devouassoux-Shisheboran, A. Puisieux, C. Caux, P. Zrounba, S. Lantuejoul, P. Saintigny, The immune microenvironment of HPV-negative oral squamous cell carcinoma from never-smokers and never-drinkers patients suggests higher clinical benefit of IDO1 and PD1/PD-L1 blockade, Ann. Oncol. (2017) 28, doi:10.1093/annonc/mdx210.

[4] L. Girard, J. Rodriguez-Canales, C. Behrens, D.M. Thompson, I.W. Botros, H. Tang, Y. Xie, N. Rekhtman, W.D. Travis, I.I. Wistuba, J.D. Minna, A.F. Gazdar, An expression signature as an aid to the histologic classification of non-small cell lung cancer, Clin. Cancer Res. 22 (2016) 4880–4889, doi:10.1158/1078-0432.CCR-15-2900.

[5] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods. 9 (2012) 357–359, doi:10.1038/nmeth.1923.

[6] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-Seq data, BMC Bioinformatics 14 (2013), doi:10.1186/1471-2105-14-7.

[7] L. Geistlinger, G. Csaba, M. Santarelli, M. Ramos, L. Schiffer, N. Turaga, C. Law, S. Davis, V. Carey, M. Morgan, R. Zimmer, L. Waldron, Toward a gold standard for benchmarking gene set enrichment analysis, Brief. Bioinform. 22 (2021) 545–556, doi:10.1093/bib/bbz158.

[8] M. Ebrahimpoor, P. Spitali, K. Hettne, R. Tsonaka, J. Goeman, Simultaneous enrichment analysis of all possible gene-sets: Unifying self-contained and competitive methods, Brief. Bioinform. 21 (2019) 1302–1312, doi:10.1093/bib/bbz074.