

RESEARCH ARTICLE

Open Access



# Tumor mutational burden assessment and standardized bioinformatics approach using custom NGS panels in clinical routine

Célia Dupain<sup>1†</sup>, Tom Gutman<sup>2†</sup>, Elodie Girard<sup>2</sup>, Choumouss Kamoun<sup>2</sup>, Grégoire Marret<sup>1</sup>, Zahra Castel-Ajgal<sup>1</sup>, Marie-Paule Sablin<sup>1</sup>, Cindy Neuzillet<sup>3</sup>, Edith Borcoman<sup>1</sup>, Ségolène Hescot<sup>1</sup>, Céline Callens<sup>4</sup>, Olfa Trabelsi-Grati<sup>4</sup>, Samia Melaabi<sup>4</sup>, Roseline Vibert<sup>4</sup>, Samantha Antonio<sup>4</sup>, Coralie Franck<sup>4</sup>, Michèle Galut<sup>5</sup>, Isabelle Guillou<sup>1</sup>, Maral Halladjian<sup>1</sup>, Yves Allory<sup>6</sup>, Joanna Cyrta<sup>5</sup>, Julien Romejon<sup>2</sup>, Eleonore Frouin<sup>4</sup>, Dominique Stoppa-Lyonnet<sup>4,7,8</sup>, Jennifer Wong<sup>4</sup>, Christophe Le Tourneau<sup>1,9,10</sup>, Ivan Bièche<sup>4,7,11</sup>, Nicolas Servant<sup>2†</sup>, Maud Kamal<sup>1†</sup> and Julien Masliah-Planchon<sup>4\*†</sup>

## Abstract

**Background** High tumor mutational burden (TMB) was reported to predict the efficacy of immune checkpoint inhibitors (ICIs). Pembrolizumab, an anti-PD-1, received FDA-approval for the treatment of unresectable/metastatic tumors with high TMB as determined by the FoundationOne<sup>®</sup>CDx test. It remains to be determined how TMB can also be calculated using other tests.

**Results** FFPE/frozen tumor samples from various origins were sequenced in the frame of the Institut Curie (IC) Molecular Tumor Board using an in-house next-generation sequencing (NGS) panel. A TMB calculation method was developed at IC (IC algorithm) and compared to the FoundationOne<sup>®</sup> (FO) algorithm.

Using IC algorithm, an optimal 10% variant allele frequency (VAF) cut-off was established for TMB evaluation on FFPE samples, compared to 5% on frozen samples. The median TMB score for MSS/POLE WT tumors was 8.8 mut/Mb versus 45 mut/Mb for MSI/POLE-mutated tumors. When focusing on MSS/POLE WT tumor samples, the highest median TMB scores were observed in lymphoma, lung, endometrial, and cervical cancers. After biological manual curation of these cases, 21% of them could be reclassified as MSI/POLE tumors and considered as “true TMB high.” Higher TMB values were obtained using FO algorithm on FFPE samples compared to IC algorithm (40 mut/Mb [10–3927] versus 8.2 mut/Mb [2.5–897],  $p < 0.001$ ).

**Conclusions** We herein propose a TMB calculation method and a bioinformatics tool that is customizable to different NGS panels and sample types. We were not able to retrieve TMB values from FO algorithm using our own algorithm and NGS panel.

<sup>†</sup>Célia Dupain and Tom Gutman contributed equally to this work and should be co-first authors.

<sup>†</sup>Nicolas Servant, Maud Kamal, Julien Masliah-Planchon contributed equally to this work and should be co-last authors.

\*Correspondence:

Julien Masliah-Planchon  
julien.masliahplanchon@curie.fr

Full list of author information is available at the end of the article



**Keywords** Tumor mutational burden, Calculation, Immunotherapy, Precision medicine, Molecular Tumor Board

## Background

Over the past decade, immunotherapy, and especially immune checkpoint inhibitors (ICIs), has revolutionized the management of several cancer types. Given the durable benefit limited to a minority of patients, the potential toxicities related to ICIs, and the high economic cost of these treatments, predictive biomarkers of response to ICIs are urgently needed.

PD-L1 expression on tumor and/or immune cells using immunohistochemistry has been demonstrated to correlate with ICI efficacy in different cancer types [1–5]. However, PD-L1 expression as a predictive biomarker of efficacy has several limitations, including the lack of sensitivity and specificity, the poor uniformity in the PD-L1 antibody clones, the different scoring methods, and positivity cut-off used [6–9].

Microsatellite instability (MSI) is caused by defects in the mismatch repair genes (therefore also called dMMR and as opposed to microsatellite stable MSS=proficient pMMR) *MSH2*, *MLH1*, *MSH6*, or *PMS2*, leading to an increased rate of mismatch errors [10–12]. Pan-cancer studies have demonstrated the predictive value of MSI (dMMR) on the response to ICIs [13, 14]. However, only 40% of patients with MSI (dMMR) tumors experience an objective response to ICIs. MSI (dMMR) tumors remain rare outside of colorectal and endometrial cancers [15, 16].

*POLE* pathogenic mutations result in ultramutated genomes and were shown to predict response to ICIs [13, 14, 17]. Specifically, mutations in the *POLE* proof-reading domain were shown to induce a high tumor mutational burden (TMB). *POLE* mutations remain extremely rare.

TMB is defined as the total number of nucleotidic variants acquired in a tumor and expressed as a number of variants per megabase (Mb). The predictive value of TMB on ICIs efficacy was retrospectively evaluated in the KEYNOTE-158 phase II basket trial of pembrolizumab [18]. High overall response rate was reported in patients with TMB-high tumors defined as  $\geq 10$  mutations per Mb using the FoundationOne<sup>®</sup>CDx assay, leading to FDA-approval of pembrolizumab across cancer types in TMB-high tumors. Besides the number of variants/Mb, the type of variants taken into account when estimating the TMB is crucial, because all mutations might not necessarily induce the release of immunogenic peptides and should reflect as close as possible the overall neoantigen load [19]. So far, no consensus exists on TMB calculation method. Besides variations in bioinformatics processing, including variant calling methods and variants filtering,

many other factors could influence the TMB estimation [20, 21]. These variations limit the harmonization of TMB calculation and robust effective cut-offs [22–24].

In this study, we aimed to estimate the TMB values from next generation sequencing (NGS) data generated from both FFPE and frozen samples using our own panel and bioinformatics algorithm and to compare the values using the FoundationOne<sup>®</sup> (FO) algorithm [25, 26]. We eventually propose customizable bioinformatics tool that allows estimating TMB values using other assays than the FO one.

## Results

### Patient characteristics

Tumor samples from 763 patients with various cancer types sequenced through the IC Molecular Tumor Board of using an in-house NGS panel were analyzed in this study. After removing the samples that did not fit the quality criteria ( $n=78$ ), 685 samples including 390 FFPE and 295 frozen samples from 43 different cancer types were assessed for estimation of the TMB (Table 1 and Fig. 1). In total, 28 samples were MSI high (dMMR) and four samples had a *POLE* mutation (Table 1).

### Development of the in-house TMB estimation algorithm (IC algorithm)

In order to select only potential immunogenic somatic variants, we only considered high-quality, coding, non-synonymous, nonsense, driver variants, and small insertion/deletions (indels), absent from the known polymorphisms/germline database (Fig. 2 and the “Methods” section). For the same reason, we also decided to determine the minimum VAF to take into account to avoid false positives. To study this parameter, we assessed the evolution of all TMB scores based on the VAF and the sample type (FFPE or frozen), among the MSS/*POLE* WT cases (Fig. 3). The TMB score inversely correlated with the minimum VAF (Fig. 3 and Additional file 1: Table S1). Higher TMB high scores were observed in FFPE samples compared to frozen samples. TMB scores in frozen tumors rapidly decreased, reaching a plateau for a minimal VAF value around 5%, whereas much heterogeneous results were observed in FFPE tumors with a decrease of TMB scores in much higher VAF cut-offs (Fig. 3). With a minimal VAF threshold fixed at 5%, only 114/362 (31%) FFPE samples had a TMB score between 0 and 10 mut/Mb compared to 147/291 (50%) for frozen samples. Similarly, 44/362 (12%) FFPE samples had a TMB score

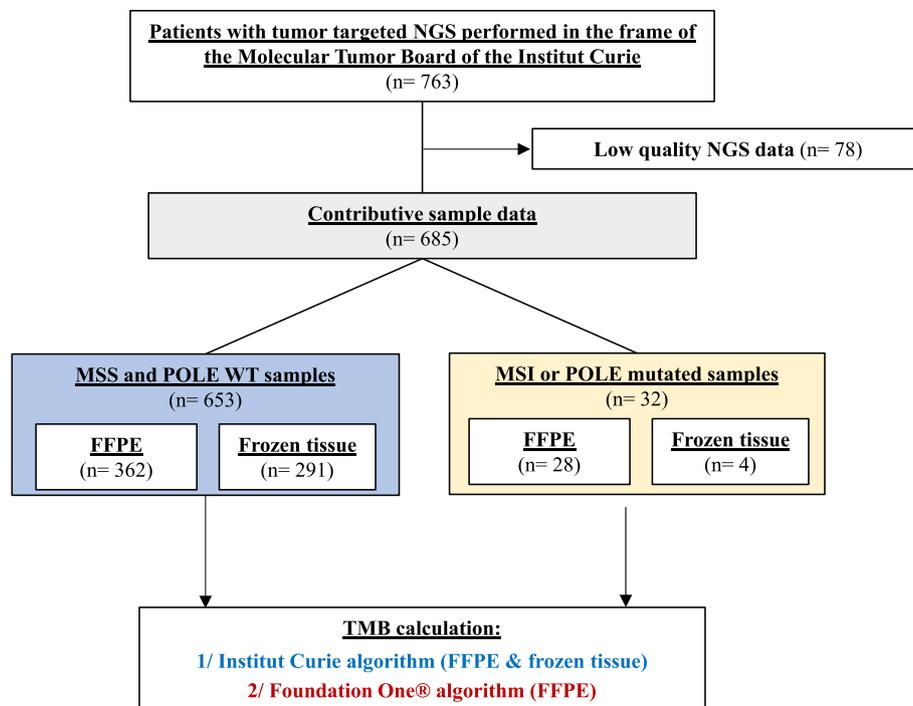
**Table 1** Cohort characteristics

| Cancer type                     | <i>n</i> | Sample type | <i>n</i> | MSI/MSS status | <i>n</i> | POLE status | <i>n</i> |
|---------------------------------|----------|-------------|----------|----------------|----------|-------------|----------|
| <b>Breast</b>                   | 126      | FFPE        | 60       | MSI            | 3        | POLE mut    | 1        |
|                                 |          | Frozen      | 66       | MSS            | 123      | POLE WT     | 125      |
| <b>Colorectal</b>               | 72       | FFPE        | 52       | MSI            | 14       | POLE mut    | 1        |
|                                 |          | Frozen      | 20       | MSS            | 58       | POLE WT     | 71       |
| <b>Sarcoma</b>                  | 72       | FFPE        | 40       | MSS            | 72       | POLE WT     | 72       |
|                                 |          | Frozen      | 32       |                |          |             |          |
| <b>Ovarian</b>                  | 59       | FFPE        | 41       | MSI            | 3        | POLE mut    | 1        |
|                                 |          | Frozen      | 18       | MSS            | 56       | POLE WT     | 58       |
| <b>CNS tumor</b>                | 58       | FFPE        | 19       | MSS            | 58       | POLE WT     | 58       |
|                                 |          | Frozen      | 39       |                |          |             |          |
| <b>Pancreatic carcinoma</b>     | 42       | FFPE        | 35       | MSI            | 2        | POLE WT     | 42       |
|                                 |          | Frozen      | 7        | MSS            | 40       |             |          |
| <b>Endometrial</b>              | 28       | FFPE        | 17       | MSI            | 4        | POLE WT     | 28       |
|                                 |          | Frozen      | 11       | MSS            | 24       |             |          |
| <b>Thyroid</b>                  | 20       | FFPE        | 18       | MSS            | 20       | POLE WT     | 20       |
|                                 |          | Frozen      | 2        |                |          |             |          |
| <b>HNSCC</b>                    | 19       | FFPE        | 15       | MSS            | 19       | POLE WT     | 19       |
|                                 |          | Frozen      | 4        |                |          |             |          |
| <b>Lymphoma</b>                 | 18       | Frozen      | 18       | MSS            | 18       | POLE WT     | 18       |
| <b>Cholangiocarcinoma</b>       | 17       | FFPE        | 12       | MSS            | 17       | POLE WT     | 17       |
|                                 |          | Frozen      | 5        |                |          |             |          |
| <b>Lung</b>                     | 17       | FFPE        | 8        | MSS            | 17       | POLE WT     | 17       |
|                                 |          | Frozen      | 9        |                |          |             |          |
| <b>Cervical</b>                 | 13       | FFPE        | 10       | MSS            | 13       | POLE WT     | 13       |
|                                 |          | Frozen      | 3        |                |          |             |          |
| <b>ACUP</b>                     | 12       | FFPE        | 5        | MSI            | 1        | POLE WT     | 12       |
|                                 |          | Frozen      | 7        | MSS            | 11       |             |          |
| <b>Rhabdoid tumor</b>           | 11       | FFPE        | 2        | MSS            | 11       | POLE WT     | 11       |
|                                 |          | Frozen      | 9        |                |          |             |          |
| <b>Anal</b>                     | 10       | FFPE        | 7        | MSS            | 10       | POLE WT     | 10       |
|                                 |          | Frozen      | 3        |                |          |             |          |
| <b>Gastric</b>                  | 10       | FFPE        | 5        | MSS            | 10       | POLE WT     | 10       |
|                                 |          | Frozen      | 5        |                |          |             |          |
| <b>Adenoid cystic carcinoma</b> | 8        | FFPE        | 5        | MSS            | 8        | POLE WT     | 8        |
|                                 |          | Frozen      | 3        |                |          |             |          |
| <b>Bladder</b>                  | 7        | FFPE        | 2        | MSS            | 7        | POLE WT     | 7        |
|                                 |          | Frozen      | 5        |                |          |             |          |
| <b>Mesothelioma</b>             | 7        | FFPE        | 3        | MSS            | 7        | POLE WT     | 7        |
|                                 |          | Frozen      | 4        |                |          |             |          |
| <b>Vaginal</b>                  | 6        | FFPE        | 3        | MSS            | 6        | POLE WT     | 6        |
|                                 |          | Frozen      | 3        |                |          |             |          |
| <b>Prostate</b>                 | 5        | FFPE        | 3        | MSS            | 5        | POLE WT     | 5        |
|                                 |          | Frozen      | 2        |                |          |             |          |
| <b>Uveal melanoma</b>           | 5        | Frozen      | 5        | MSS            | 5        | POLE WT     | 5        |
| <b>Cutaneous melanoma</b>       | 4        | FFPE        | 2        | MSS            | 4        | POLE WT     | 4        |
|                                 |          | Frozen      | 2        |                |          |             |          |
| <b>Sex chord tumor</b>          | 4        | FFPE        | 3        | MSS            | 4        | POLE WT     | 4        |
|                                 |          | Frozen      | 1        |                |          |             |          |
| <b>Appendix</b>                 | 3        | FFPE        | 3        | MSS            | 3        | POLE WT     | 3        |

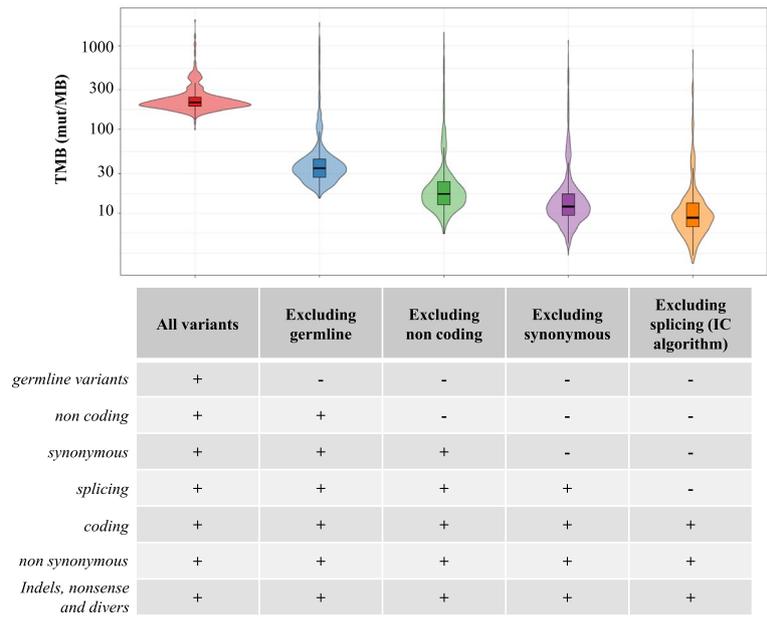
**Table 1** (continued)

| Cancer type           | n | Sample type | n | MSI/MSS status | n | POLE status | n |
|-----------------------|---|-------------|---|----------------|---|-------------|---|
| Esophageal            | 3 | FFPE        | 1 | MSS            | 3 | POLE WT     | 3 |
|                       |   | Frozen      | 2 |                |   |             |   |
| Salivary gland tumor  | 3 | FFPE        | 3 | MSS            | 3 | POLE WT     | 3 |
| UCNT                  | 3 | FFPE        | 3 | MSS            | 3 | POLE WT     | 3 |
| GIST                  | 2 | FFPE        | 2 | MSS            | 2 | POLE WT     | 2 |
|                       |   | Frozen      | 1 |                |   |             |   |
| Neuroendocrine        | 2 | FFPE        | 1 | MSS            | 2 | POLE WT     | 2 |
|                       |   | Frozen      | 1 |                |   |             |   |
| Renal                 | 2 | Frozen      | 2 | MSS            | 2 | POLE WT     | 2 |
| Vulva                 | 2 | FFPE        | 2 | MSS            | 2 | POLE mut    | 1 |
|                       |   |             |   |                |   | POLE WT     | 1 |
| Craniopharyngioma     | 1 | FFPE        | 1 | MSS            | 1 | POLE WT     | 1 |
| Cutaneous SCC         | 1 | FFPE        | 1 | MSS            | 1 | POLE WT     | 1 |
| Duodenal carcinoma    | 1 | FFPE        | 1 | MSS            | 1 | POLE WT     | 1 |
| Hepatoblastoma        | 1 | FFPE        | 1 | MSS            | 1 | POLE WT     | 1 |
| Leiomyosarcoma        | 1 | Frozen      | 1 | MSS            | 1 | POLE WT     | 1 |
| Peritoneum            | 1 | FFPE        | 1 | MSS            | 1 | POLE WT     | 1 |
| Small bowel carcinoma | 1 | FFPE        | 1 | MSI            | 1 | POLE WT     | 1 |
| Thymoma               | 1 | FFPE        | 1 | MSS            | 1 | POLE WT     | 1 |
| Waldenstrom           | 1 | Frozen      | 1 | MSS            | 1 | POLE WT     | 1 |
| Other                 | 6 | FFPE        | 4 | MSS            | 6 | POLE WT     | 6 |
|                       |   | Frozen      | 2 |                |   |             |   |

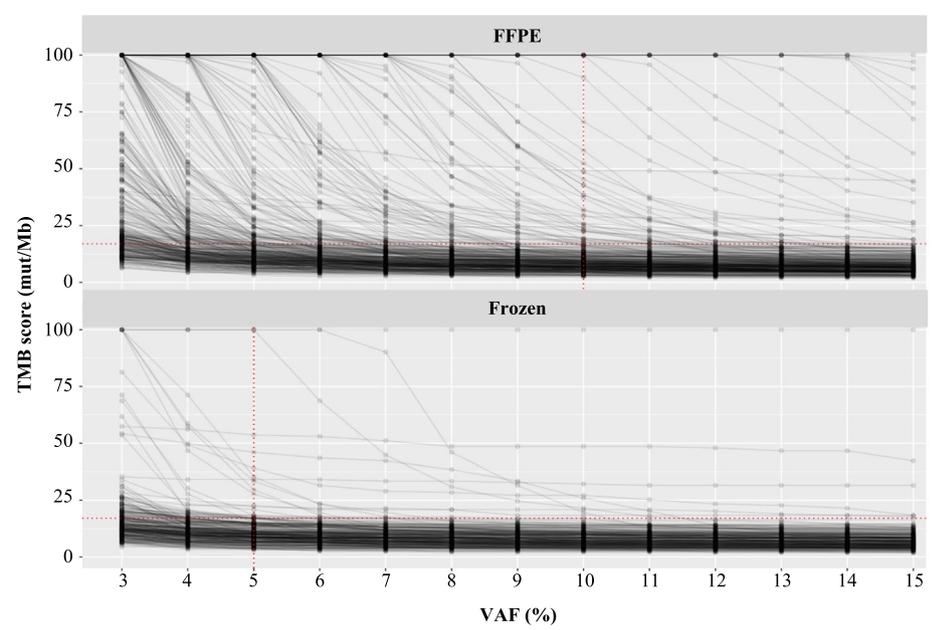
FFPE Formalin-fixed paraffin-embedded, mut mutated, WT Wild-type, SCC Squamous cell carcinoma, CNS Central nervous system, HNSCC Head and neck squamous cell carcinoma, ACUP Adenocarcinoma of unknown primary, UCNT Undifferentiated carcinoma of nasopharyngeal type, GIST Gastrointestinal stromal tumor



**Fig. 1** Analysis workflow. MSS, microsatellite stable; WT, wild-type



**Fig. 2** Distribution of TMB score variation among the cohort according to variant filters applied. IC, Institut Curie; Mut, Mutations; TMB, Tumor Mutational Burden



**Fig. 3** TMB score variation according to variant allele frequency (VAF) cut-off, and sample type (FFPE or frozen). FFPE, formalin-fixed paraffin-embedded; Mut, mutations; TMB, tumor mutational burden; VAF, variant allele frequency

greater than 100 mut/Mb compared to only 3/291 (1%) for frozen samples (Additional file 1: Table S1).

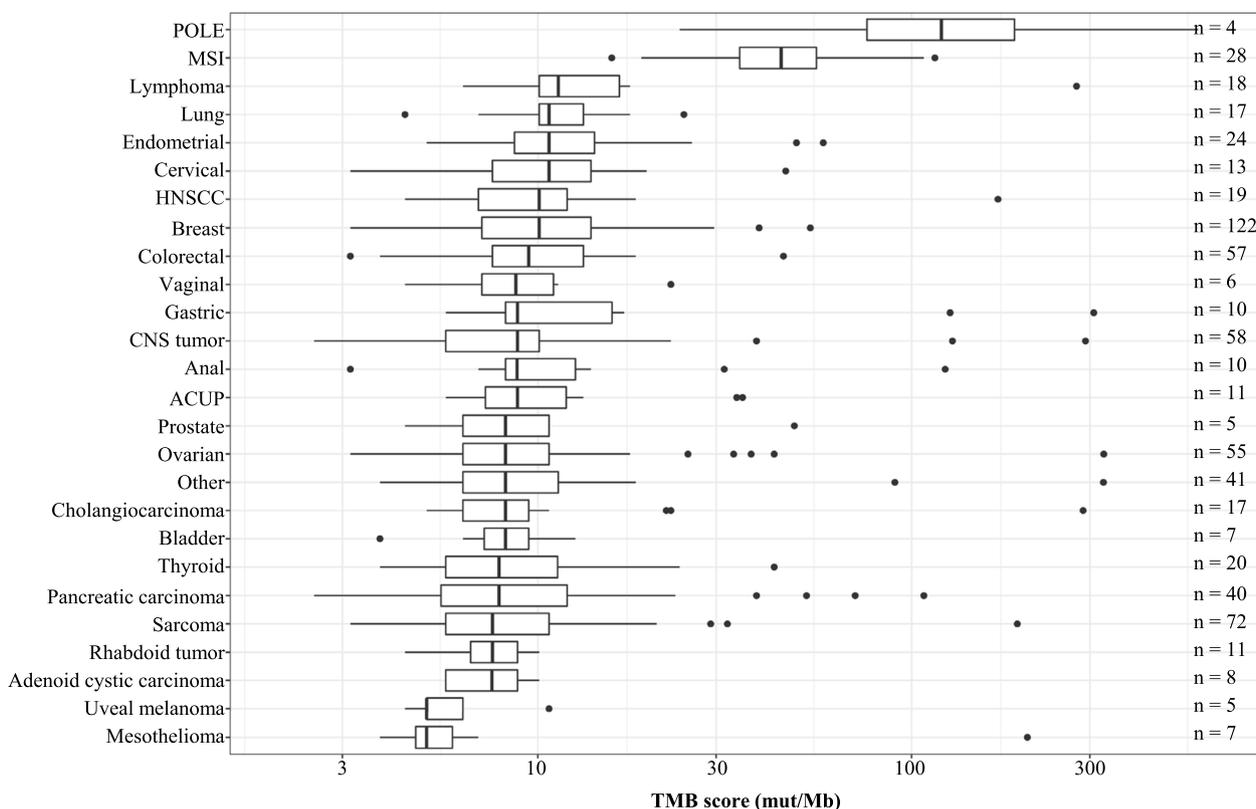
With a VAF threshold fixed at 10%, 236/362 (65%) FFPE samples had a TMB score ranging from 0 to 10 mut/Mb, compared to 209/291 (72%) for frozen samples. A total of 11/362 (3%) of FFPE samples had a TMB score greater than 100 mut/Mb compared to 1/291 (0.3%) for frozen samples. When moving the VAF threshold from 5 to 10%, 55 FFPE samples switched from a TMB score higher than 30 mut/Mb to lower than 30 compared to only 6 frozen samples (Additional file 1: Table S1).

We then focused on the tumors for which both frozen and FFPE pairs were analyzed (Additional file 2: Fig. S1). For frozen samples, a plateau (which likely represents the true TMB) was reached for a VAF at 5%. For FFPE samples, we were able to distinguish high-quality DNA and low-quality DNA based on pre-analytical parameters as defined in the “Methods” section. For high-quality FFPE, the steady state was reached with VAF below or around 10%. For low-quality FFPE, the steady state was either reached with a higher VAF or never reached.

We therefore established the minimum VAF threshold used to consider a variant in the TMB estimation to be 5% for frozen samples and 10% for FFPE samples.

### Repartition of TMB scores using IC algorithm

We then evaluated the TMB on the 685 contributive samples. The median TMB score calculated with IC algorithm of MSS/POLE-WT tumors was 8.8 mut/Mb [2.5–897] versus 45 mut/Mb [16–584] for MSI/POLE-mutated tumors (Fig. 4 and Additional file 1: Table S2). When focusing on MSS/POLE-WT tumors ( $n=653$ ), main cancer types analyzed included breast (19%), sarcoma (11%), central nervous system (CNS) (9%), colorectal (9%), and ovarian (8%) cancers. The highest median TMB scores among the MSS/POLE-WT tumors were found in lymphoma (11 mut/Mb [6.3–276]), lung (11 mut/Mb [4.4–24]), endometrial (11 mut/Mb [5.0–58]), and cervical cancer (11 mut/Mb [3.2–46]). The lowest scores among the MSS/POLE-WT tumors were found in uveal melanoma (5.0 mut/Mb [4.4–11]) and mesothelioma (5.0 mut/Mb [3.8–204]) (Fig. 4 and Additional file 1: Table S2).



**Fig. 4** Repartition of TMB scores according to tumor types using the algorithm of the Institut Curie (IC). Tumor types with less than  $n=5$  samples were grouped into “Others” in this plot which comprise the following tumor types: cutaneous melanoma, sex chord tumor, appendix, esophageal, salivary gland tumor, UCNT, GIST, neuroendocrine, renal, vulva, craniopharyngioma, cutaneous SCC, duodenal carcinoma, hepatoblastoma, leiomyosarcoma, peritoneum, small bowel carcinoma, thymoma, and Waldenstrom. HNSCC, head and neck squamous cell carcinoma; CNS, central nervous system; ACUP, adenocarcinoma of unknown primary; TMB, tumor mutational burden

### Biological curation of TMB-high cases

In order to distinguish true positive TMB-high cases from false positives and to investigate if some cases could be reclassified as MSI-high tumors (dMMR), we focused on the top 10% samples ( $n=65$ ) with the highest TMB scores among the non-MSI pMMR cases (the MSS/POLE-WT tumors). We removed 8 out of these 65 cases with a bad quality of sequence and considered them as non-contributive for TMB evaluation, leaving 57 TMB high cases. On those cases, 12/57 cases (21%) were found to have either a MSI score  $\geq 10\%$  using MSIsensor, a pathogenic variant in one of the MMR genes and/or a mutational signature suggesting a MSI profile, or POLE proofreading deficiency, or APOBEC mutational signature (Additional file 1: Table S3 and Table S4). These samples could be reclassified as MSI/POLE mutated tumors and considered as “true TMB high” cases with a high confidence. For the remaining 45 cases, the high TMB score could not be explained by an MSI status, POLE mutation, or APOBEC signature. For information, we also verified the presence of pathogenic variants (with an allelic ratio  $\geq 10\%$ ) among 3 candidate genes implicated in DNA damage repair (i.e., *TP53*, *PTEN*, and *ARID1A*). Interestingly, 17/57 cases harbored at least one pathogenic variant in these 3 candidate genes, leaving 28/57 cases (49%) with no explanation for high TMB status.

### TMB scores evaluation using FO algorithm

The TMB score using the FO algorithm was calculated on the 685 contributive samples of the cohort (Additional file 1: Table S2), with a focus on FFPE samples ( $n=390$ ) to better reproduce the FoundationOne<sup>®</sup>CDx test conditions. We observed that all TMB values exceeded 10 mut/Mb, the FDA-approved cut-off to consider a tumor TMB-high (Additional file 1: Table S2). When comparing the distribution of TMB scores obtained with the IC algorithm to the one obtained with FO algorithm on the same NGS data derived from all FFPE MSS/POLE-WT tumors ( $n=362$ ), the median TMB values obtained with IC algorithm were significantly lower compared to the one obtained with the FO algorithm (8.2 mut/Mb [2.5–897] versus 40 mut/Mb [10–3928],  $p < 0.001$ ) (Additional file 2: Fig. S2). Individually, all samples but one had higher TMB from FO algorithm compared to IC algorithm (Additional file 1: Table S2).

### Discussion

We demonstrate that both sample types (FFPE and frozen) and DNA quality (measured with Cp) had an impact on the TMB scores. False positive deamination artifacts (C>T transitions) created by formalin fixation in low-quality FFPE DNA is a well-known effect that can lead to an overestimation of the TMB [20, 24, 27, 28]. This

prevents using the same minimum VAF threshold for both FFPE and frozen samples.

Deduplication was not used in our study. Although it could have an impact on the variant calling accuracy, and thus affect the TMB score [20, 29], other studies showed that deduplication was not always mandatory [30, 31] or could be overcome by applying a 10% VAF threshold [20, 32]. We have demonstrated that the use of UMI-based deduplication did not impact our results by calculating the VAFs of all variants with or without UMI processing and computing the correlation between VAFs values for each patient. An average correlation of 0.952 for the FFPE samples and 0.983 for the frozen samples demonstrated that the UMI processing has very little impact on the VAFs (Additional file 2: Fig. S3). This is in line with other publications [30, 31, 33].

Based on our analysis of more than 750 samples and previous recommendations [20, 34], we proposed a 10% VAF cut-off for FFPE samples and a 5% cut-off for frozen samples. The high TMB scores found in FFPE samples, possibly due to fixation artifacts, represents a clinical reality to be dealt with for routine TMB calculation, across all laboratories [20, 24, 27, 28]. In this study, we propose a general algorithm with appropriate filters and threshold to limit the impact of such artifacts, but a manual curation step for this kind of samples will always be unavoidable. Using a fixed threshold allows to (i) simplify the variant calling process, making it more standardized and easier to implement across different samples and studies, (ii) provide consistency when comparing TMB across samples, and (iii) homogenize the interpretation of results. These points are particularly important in clinical settings where uniformity in methodology is required.

To overcome this problem upstream of the analysis, we applied the most rigorous possible filters to remove the false positives while preserving the true variants. Other possibilities might include the implementation of dedicated computational algorithm to rectify formalin-induced artifacts for FFPE samples [35] or optimization of the chemistry with the use of enzymes involved in base excision repair before library preparation [36].

Using the FO algorithm, all TMB scores exceeded 10 mut/Mb, which differs from what has been reported in the literature [25, 37]. These results suggest that the level of information provided by FoundationOne<sup>®</sup> does not enable to reproduce their algorithm and consequently to directly transpose the FO algorithm to other targeted NGS panels.

The choice of variants to take into account when estimating the TMB is crucial, because all mutations do not necessarily induce the release of immunogenic peptides, and should reflect as close as possible the overall neoantigen load [19]. As targeted panels include mainly cancer

genes, which are more likely to be mutated in the tumor, some methods have been proposed to filter out known cancer variants for TMB quantification. We chose to keep cancer hotspots variants in our algorithm for the TMB estimation, since they could also generate immunogenic peptides. We also chose to filter out synonym and non-coding variants as they are unlikely to generate neoepitopes and the size of the coding sequence of our in-house NGS panel is sufficient to assure TMB reliability [26]. Compared to whole exome sequencing, NGS panels are not constantly associated with the germline paired DNA sequencing. This requires a substantial methodology to filter out the polymorphisms that come from the germline and hence might not induce an immune response. Germline variants are commonly filtered using databases of known germline mutations. Some algorithms use complementary germline removal algorithm such as somatic-germline-zygosity [38]. Here, due to partially available information on the SGZ algorithm proposed by FoundationOne® as part of their commercial product (FoundationOne®CDx), we used different databases of known germline mutations as references (Exac, 1000G or GnomAD all ethnicities) to remove as many germline variants as possible and only retain private or extremely rare germline polymorphisms, which may increase TMB score [39].

Overall, several parameters including biological factors to pre-analytics, sequencing, and bioinformatics can impact the TMB scores estimation, explaining the diversity of published TMB algorithms, the heterogeneity of the results, and the complexity to harmonize methods [20]. The bioinformatics tool used in this study is freely available for the community and highly customizable to fit different targeted NGS panels and sample types (both FFPE and frozen). Other tools for TMB calculation have been developed and reported in the literature. Their applicability still needs to be tested, since they often require to have paired targeted NGS and WES data for each patient. In addition, the sample type (frozen or FFPE) and quality are not taken into account in the estimation [33, 40].

The TMB estimation using our algorithm revealed variations in the medians and ranges across tumor types, with the highest median TMB score found in MSI/*POLE*-mutated tumors. Our results are in line with previous reports in the literature [18, 25, 37, 41]. We observed that some tumors harbored very high TMB scores, although not associated with MSI status (dMMR) or *POLE* mutations at first glance. After biological manual curation of these cases, 21% of them could be reclassified as MSI/*POLE* tumors and considered as “true TMB high” with a high level of confidence, and 30% had at least one pathogenic variant among 3 candidate genes implicated in

DNA damage repair that could be related to high TMB (i.e., *TP53*, *PTEN*, and *ARID1A*) [42–44]. However, for the remaining cases, the high TMB scores could not find a biological explanation. The more detailed manual observation of TMB-high cases represents the reality of TMB status validations carried out by the experts within the framework of clinical routine use.

## Conclusions

In conclusion, we show that the TMB values obtained from the same NGS data but with different calculation methods are not comparable. In order to optimize the implementation of TMB as a robust predictive biomarker of efficacy of ICIs, the determination of the method to be used to identify the right threshold is key. Studies from cohorts of patients treated with ICIs will be needed to identify these thresholds as well as studies on larger series of matched FFPE and Frozen samples to determine the most optimal way to avoid artifacts in the calculation of TMB (i.e., using different algorithms with a possible different VAF cut-off for variant calling, or using different cut-offs on TMB values for high or low statuses according to a FFPE or frozen sample).

## Methods

### Patient selection

Patients with recurrent and/or metastatic cancers whose tumor was sequenced in the frame of Molecular Tumor Board of the Institut Curie (IC) [45] were included in this study. Informed consent with regard to the collection of tumor samples and molecular analysis was obtained from patients within the IC institutional general consent signed by every patient treated at the IC.

### In-house next generation sequencing panel

Samples were sequenced using an in-house NGS panel covering 1.6 Mb. Indexed paired-end libraries of tumor DNA were performed using the Agilent Sureselect XT-HS library prep kit. Fifty nanograms of input DNA were used to build the libraries according to manufacturer's protocol. Libraries were sequenced on the NovaSeq 6000 (Illumina) Sp 2 × 100 bp flow cell.

### Bioinformatics

After tumor DNA sequencing, bioinformatics analyses were performed as detailed below in order to detect single-nucleotide variants (SNVs) and indels, microsatellite instability statuses, mutational signatures, and TMB scores (detailed in Additional file 3: Supplementary Methods and above).

### Variant calling

Variant calling of both SNVs and indels was carried out on the aligned sequencing data as previously described [46]. Annotations from several databases [RefSeq [47], dbSNP v150 [48], COSMIC v86 [49], 1000 g project 08/2015 version [50], ESP6500 [51] gnomAD (all and ethnicities) [39], ICGC v21 [52], and dbSNP v35 [53] predictions] were provided by Annovar (04/16/2018 version, Wang et al. [54]).

### TMB calculation

After removing low NGS quality samples, i.e., samples with <20 million sequencing reads or <15% of the captured regions sequenced above 1000X, the TMB values were calculated using two different algorithms: (1) the FO algorithm on FFPE samples and (2) our IC algorithm on all samples including both FFPE and frozen (Fig. 1).

FoundationOne® (FO) TMB algorithm was reproduced based on the Summary of Safety and Effectiveness ([https://www.accessdata.fda.gov/cdrh\\_docs/pdf17/P170019S016B.pdf](https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019S016B.pdf)). Low-quality variants were removed based on the absence of “PASS” tag from varScan2 variant calling results. Germline variants were also removed from the vcf files using the somatic-germline-zygosity (SGZ) algorithm (v1.0.0) [38] as well as polymorphisms database (variants found in 1000 Genomes or Exac [55] databases for all ethnicities with a minor allelic frequency (MAF) higher than 0.1%). Non-coding variants and driver mutations found at least once in COSMIC

database were also removed. Hence, all coding variants including synonymous, splicing (defined as every intronic nucleotide within 2 bp at the exon/intron boundaries), and indels were considered for the final TMB calculation if their VAF was higher than 5% and the depth of coverage higher than 100X. Of note, with the information provided by FoundationOne®, we were not able to reproduce their exact capture regions and thus based our TMB calculation on our own design and dividing the number of variants by 1.6 Mb to obtain the number of mutations per Mb.

For IC TMB algorithm, recurrent variants detected in more than 15% of the samples within the same sequencing run were considered as false positive and removed from the TMB calculation. Polymorphisms found in 1000 Genomes, Gnomad, or Exac databases for all ethnicities with a MAF higher than 0.1% were also removed. Given that the goal of TMB is to identify likely immunogenic tumors that ultimately could respond to ICI, and that only somatic, acquired, coding variants encode potential neoantigens, we decided to consider in the IC algorithm the coding, non-synonymous, and indels variants but to remove non-coding, synonymous, and splice (defined as every intronic nucleotide within 2 bp at the exon/intron boundaries) variants. Finally, only variants with a VAF higher than 5% for frozen samples or 10% for FFPE samples and a depth of coverage higher than 100X were considered for TMB estimation.

In order to standardize the TMB estimation, we developed a bioinformatics tool named pyTMB that can be applied to

**Table 2** Filters applied for TMB calculation with Foundation One® (FO) algorithm and Institut Curie (IC) algorithm

|                             | FoundationOne® algorithm   | Institut Curie algorithm  |
|-----------------------------|--|---|
| Metrics                     |  |   |
| Low-quality variant removal | YES<br>PASS tag from VarScan2  | YES<br>Intra-run recurrence   |
| Minimum VAF                 | 5%   | 5% for frozen tumors<br>10% for FFPE tumors                           |
| Minimum depth of coverage   | 100X   | 100X  |
| Germline removal            | SGZ algorithm<br>And SNP from 1000 Genomes or Exac with a MAF > 0.1% were filtered | SNP from 1000 Genomes, Gnomad or Exac with a MAF > 0.1% were filtered |
| Variants types              |  |   |
| Drivers                     | Excluded<br>COSMIC   | Included  |
| Synonymous                  | Included   | Excluded  |
| Splicing                    | Included   | Excluded  |
| Indels                      | Included   | Included  |
| Nonsense                    | Included   | Included  |
| Non-coding                  | Excluded   | Excluded  |

The FO algorithm was applied on FFPE samples only as required by the FoundationOne®CDx test, while the IC algorithm was used on both frozen and FFPE samples  
FFPE Formalin-fixed paraffin-embedded, VAF Variant allele frequency, SGZ Somatic-germline-zygosity, COSMIC Catalogue of Somatic Mutations in Cancer, MAF Mutation allele frequency, SNP Single-nucleotide polymorphism

any sequencing data type. pyTMB can be easily installed with conda either directly from the source code (<https://github.com/bioinfo-pf-curie/TMB>) or from the bioconda channel. PyTMB v1.1.0 has been used by this study (<https://doi.org/10.5281/zenodo.10573735>). pyTMB requires a list of annotated variants and successively applies the different filters that can then be adapted by the users. The version 1.1.0 supports.vcf files generated with the Mutect2 and VarScan2 tools and annotated with either ANNOVAR or snpEff (Table 2 and Additional file 3: Supplementary Methods).

### Biological curation of TMB high cases

To avoid false positives related to bad quality DNA, we focused on the top 10% samples with the highest TMB scores (corresponding to a TMB > 17.5 mut/Mb using the IC algorithm) among the non-MSI (pMMR) cases (MSS/POLE WT tumors). To further investigate the high TMB cases, we individually assessed: (i) the MSI score using MSI sensor, (ii) mutations in MMR-related genes (e.g., in *MSH2*, *MSH1*, *MSH6*, or *PMS2* gene), and (iii) the presence of MMR or APOBEC-related mutational signatures (see Additional file 3: Supplementary Methods).

### Notes

#### Role of the funder

The authors are all part of the Institut Curie which provided the resources for the personnel as well as the equipment, reagents, materials, and structures needed for the Molecular Tumor Board and for the analyses. Amgen France, La Ligue Contre le Cancer, and Cancéropole Ile-de-France provided funding for reagents, sample processing, and personnel resources through grants.

### Abbreviations

|        |   |
|--------|---|
| ACUP   | Adenocarcinoma of unknown primary                 |
| COSMIC | Catalogue of Somatic Mutations in Cancer          |
| CNS    | Central nervous system                            |
| FDA    | Food and Drug Administration                      |
| FFPE   | Formalin-fixed paraffin-embedded                  |
| FO     | FoundationOne                                     |
| GIST   | Gastrointestinal stromal tumor                    |
| HNSCC  | Head and neck squamous cell carcinoma             |
| IC     | Institut Curie                                    |
| ICIs   | Immune checkpoint inhibitors                      |
| Indels | Small insertion/deletions                         |
| MAF    | Mutation allele frequency                         |
| MMR    | Mismatch repair                                   |
| MSI    | Microsatellite instability                        |
| MSS    | Microsatellite stable                             |
| Mut    | Mutated/mutations                                 |
| NGS    | Next generation sequencing                        |
| SCC    | Squamous cell carcinoma                           |
| SGZ    | Somatic-germline-zygosity                         |
| SNP    | Single-nucleotide polymorphism                    |
| TMB    | Tumor mutational burden                           |
| UCNT   | Undifferentiated carcinoma of nasopharyngeal type |
| VAF    | Variant allele frequency                          |
| WT     | Wild-Type   |

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01839-8>.

**Additional file 1: Table S1.** Detailed TMB score variation according to variant allele frequency (VAF) cut-off and to sample type (FFPE or frozen).

**Table S2.** Detailed TMB evaluation across the 685 contributive tumor samples. FFPE = Formalin-Fixed Paraffin-Embedded; mut = mutated; WT = Wild-Type; MSS = MicroSatellite Stable; MSI = MicroSatellite Instable; SCC = Squamous Cell Carcinoma; CNS = Central Nervous System; HNSCC = Head and Neck Squamous Cell Carcinoma; ACUP = AdenoCarcinoma of Unknown Primary; UCNT = Undifferentiated Carcinoma of Nasopharyngeal Type; GIST = Gastrointestinal Stromal Tumor. **Table S3.** Focus on TMB high cases including the evaluation of MSI score, MMR-related gene mutations, and MMR-related mutational signatures. \*: pathogenic variants (with an allelic ratio  $\geq 10\%$ ) among 3 candidate genes implicated in DNA damage repair (i.e., TP53, PTEN, and ARID1A). MSS = MicroSatellite Stable; MMR = Mismatch Repair. **Table S4.** MMR genes mutational variants detected in 3 samples with high TMB. MMR = Mismatch Repair; MSI = MicroSatellite Instable.

**Additional file 2: Fig. S1.** TMB score variation according to DNA sample quality and according to sample type (FFPE or frozen) in 10 sample pairs. FFPE = Formalin-Fixed Paraffin-Embedded; TMB = Tumor Mutational Burden; VAF = Variant Allele Frequency. **Fig. S2.** TMB scores according to the algorithm of the Institut Curie (IC) and FoundationOne® (FO), obtained from the same NGS data of 362 MSS/POLE WT FFPE pan-cancer samples. \*\*\*  $p < 0.001$  using Wilcoxon signed-rank test. FFPE = Formalin-Fixed Paraffin-Embedded; MSS = MicroSatellite Stable. **Fig. S3.** Computational analysis of VAFs correlation with or without UMI processing in FFPE and frozen samples for each patient.

**Additional file 3: Supplementary Methods.** TMB calculation method and parameters applied for each algorithm.

**Additional file 4.**

### Acknowledgements

The authors would like to acknowledge Amgen France and La Ligue Contre le Cancer who funded part of this study.

The authors also acknowledge all the medical oncologists of Institut Curie involved in the Molecular Tumor Board of Institut Curie: Pauline du Rusquec, Diana Bello Roufai, Maxime Frelaut, Coraline Dubot, Audrey Bellesoeur, Manuel Rodrigues, Nicolas Girard, Aude Guillemin, Perrine Vuagnat, Patricia Tresca, Amami Asnacios Leцерf, Hélène Salaun, Clélia Chalumeau, Laurence Bozec, Lorene Seguin, Pauline Vafard, Sarah Watson, Slim Bach Hamba, Sophie Frank, Valérie Laurence, and Hamid Mammari.

The authors would like to acknowledge the medical biologists of the Department of Genetics of Institut Curie for the analyses and interpretation of NGS results—Keltouma Driouch—as well as the pathologists from the Department of Pathology of Institut Curie—Anne Vincent-Salomon, Loïc Trapani, Ahmad El Sabeh Ayoun, Hrnt Ghazelian, and Sarah Nasr.

The authors would like to thank all the members of the Centre de Ressources Biologiques of Institut Curie—Odette Mariani, Nassima Mouterfi, Aurore Godard, Céline Méaudre, Sylvie Jovelin, and Cloé Pierson—the members of the SeqOIA platform, and Mario Neou and Alban Lermine for their help in the development of the TMB calculation algorithm.

Finally, the authors would like to thank the nf-core and particularly Friederike Hanssen and Alexander Peltzer who implemented pyTMB into the bioconda channel.

### Authors' contributions

Conceptualization: JMP, IB, CLT, NS, MK. Data curation: NS, TG, CK, CD, JMP. Formal analysis: NS, TG, IB, JMP, CD, EG, CK, RV, JR, EF. Funding acquisition: CLT, IB, JMP, MK, NS, CD. Investigation: JMP, CD, MK, TG, IB, CLT, CC, OTG, SM, RV, SA, CF, MG, IG, YA, JC, JR, EF, JW. Methodology: JMP, CD, MK, TG, IB, CLT. Project administration: JMP, IB, CLT, NS, MK. Resources: JMP, IB, CLT, NS, TG, MK, CD, EG, CK, GM, ZCA, MPS, CN, EB, SH, CC, OTG, SM, RV, SA, CF, MG, IG, YA, JC, JR, EF, DSL, JW. Supervision: JMP, IB, CLT, NS, MK. Validation: JMP, IB, CLT, NS, MK, CD, TG. Visualization: CD, TG, EG, CK, GM, ZCA, MPS, CN, EB, SH, CC, OTG, SM, RV, SA, CF,

MG, IG, MH, YA, JC, JR, EF, DSL, JW, CLT, IB, NS, MK, JMP. Writing—original draft: CD, TG, JMP, IB, CLT, NS, MK. Writing—review & editing: CD, TG, EG, CK, GM, ZCA, MPS, CN, EB, SH, CC, OTG, SM, RV, SA, CF, MG, IG, MH, YA, JC, JR, EF, DSL, JW, CLT, IB, NS, MK, JMP. All authors read and approved the final manuscript.

### Funding

This work was supported by Institut Curie, Amgen France, La Ligue Contre le Cancer, and Cancéropole Ile-de-France.

### Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. NGS and clinical data were obtained from tumor samples from 763 patients treated at Institut Curie. Supporting data values for  $n < 6$  individual data values reported in the figures are detailed in the [Supporting data values](#) file. Source code for pyTMB can be found on GitHub (<https://github.com/bioinfo-pf-curie/TMB>) and Zenodo (<https://doi.org/10.5281/zenodo.10573735>).

### Declarations

#### Ethics approval and consent to participate

The study was approved by Institut Curie's internal committee. Informed consent with regard to the collection of tumor samples and molecular analysis was obtained from patients within the IC institutional general consent signed by every patient treated at the IC.

#### Consent for publication

Not applicable.

#### Competing interests

EB received honoraria and nonfinancial support from Eisai, MSD, Sandoz, Daiichi Sankyo. All other authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Drug Development and Innovation (D3i), Institut Curie, Paris, France. <sup>2</sup>Bioinformatics Core Facility, INSERM U900, Mines Paris Tech, Institut Curie, Paris, France. <sup>3</sup>Department of Medical Oncology, Institut Curie, Paris & Saint Cloud, France. <sup>4</sup>Department of Genetics, Institut Curie, Paris, France. <sup>5</sup>Department of Pathology, Institut Curie, PSL Research University, Paris, France. <sup>6</sup>Department of Pathology, Université Paris-Saclay, UVSQ, Institut Curie, Saint-Cloud, France. <sup>7</sup>Paris-Cité University, Paris, France. <sup>8</sup>INSERM U830, Paris, France. <sup>9</sup>Inserm U900 Research Unit, Saint Cloud, France. <sup>10</sup>Paris-Saclay University, Paris, France. <sup>11</sup>Faculty of Pharmaceutical and Biological Sciences, INSERM U1016, Paris Descartes University, Paris, France.

Received: 29 November 2022 Accepted: 2 February 2024

Published online: 20 February 2024

### References

- Burtneff B, Harrington KJ, Greil R, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet Lond Engl*. 2019;394(10212):1915–28. [https://doi.org/10.1016/S0140-6736\(19\)32591-7](https://doi.org/10.1016/S0140-6736(19)32591-7).
- Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med*. 2015;372(21):2018–28. <https://doi.org/10.1056/NEJMoa1501824>.
- 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(14):1328–39. <https://doi.org/10.1056/NEJMoa1917346>.
- Cortes J, Cescon DW, Rugo HS, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. *Lancet*. 2020;396(10265):1817–28. [https://doi.org/10.1016/S0140-6736\(20\)32531-9](https://doi.org/10.1016/S0140-6736(20)32531-9).
- Schmid P, Cortes J, Pusztai L, et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med*. 2020;382(9):810–21. <https://doi.org/10.1056/NEJMoa1910549>.
- Mahoney KM, Sun H, Liao X, et al. PD-L1 Antibodies to its cytoplasmic domain most clearly delineate cell membranes in immunohistochemical staining of tumor cells. *Cancer Immunol Res*. 2015;3(12):1308–15. <https://doi.org/10.1158/2326-6066.CIR-15-0116>.
- Rimm DL, Han G, Taube JM, et al. A prospective, multi-institutional, pathologist-based assessment of 4 immunohistochemistry assays for PD-L1 expression in non-small cell lung cancer. *JAMA Oncol*. 2017;3(8):1051–8. <https://doi.org/10.1001/jamaoncol.2017.0013>.
- Gaule P, Smithy JW, Toki M, et al. A quantitative comparison of antibodies to programmed cell death 1 ligand 1. *JAMA Oncol*. 2017;3(2):256–9. <https://doi.org/10.1001/jamaoncol.2016.3015>.
- Torlakovic E, Lim HJ, Adam J, et al. “Interchangeability” of PD-L1 immunohistochemistry assays: a meta-analysis of diagnostic accuracy. *Mod Pathol*. 2020;33(1):4–17. <https://doi.org/10.1038/s41379-019-0327-4>.
- Bach DH, Zhang W, Sood AK. Chromosomal instability in tumor initiation and development. *Cancer Res*. 2019;79(16):3995–4002. <https://doi.org/10.1158/0008-5472.CAN-18-3235>.
- Baretti M, Le DT. DNA mismatch repair in cancer. *Pharmacol Ther*. 2018;189:45–62. <https://doi.org/10.1016/j.pharmthera.2018.04.004>.
- Bonneville R, Krook MA, Kautto EA, et al. Landscape of microsatellite instability across 39 cancer types. *JCO Precis Oncol*. 2017;2017. <https://doi.org/10.1200/PO.17.00073>.
- Overman MJ, McDermott R, Leach JL, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol*. 2017;18(9):1182–91. [https://doi.org/10.1016/S1470-2045\(17\)30422-9](https://doi.org/10.1016/S1470-2045(17)30422-9).
- Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. 2017;357(6349):409–13. <https://doi.org/10.1126/science.aan6733>.
- Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. 2015;372(26):2509–20. <https://doi.org/10.1056/NEJMoa1500596>.
- Marabelle A, Le DT, Ascierto PA, et al. Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *J Clin Oncol*. 2020;38(1):1–10. <https://doi.org/10.1200/JCO.19.02105>.
- Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19(3):133–50. <https://doi.org/10.1038/s41568-019-0116-x>.
- 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):1353–65. [https://doi.org/10.1016/S1470-2045\(20\)30445-9](https://doi.org/10.1016/S1470-2045(20)30445-9).
- Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. 2014;371(23):2189–99. <https://doi.org/10.1056/NEJMoa1406498>.
- Stenzinger A, Endris V, Budczies J, et al. Harmonization and standardization of panel-based tumor mutational burden measurement: real-world results and recommendations of the quality in pathology study. *J Thorac Oncol*. 2020;15(7):1177–89. <https://doi.org/10.1016/j.jtho.2020.01.023>.
- Budczies J, Kazdal D, Allgauer M, et al. Quantifying potential confounders of panel-based tumor mutational burden (TMB) measurement. *Lung Cancer Amst Neth*. 2020;142:114–9. <https://doi.org/10.1016/j.lungcan.2020.01.019>.
- Fancelli L, Gandini S, Pelicci PG, Mazzarella L. Tumor mutational burden quantification from targeted gene panels: major advancements and challenges. *J Immunother Cancer*. 2019;7(1):183. <https://doi.org/10.1186/s40425-019-0647-4>.
- Merino DM, McShane LM, Fabrizio D, et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. *J Immunother Cancer*. 2020;8(1):e000147. <https://doi.org/10.1136/jitc-2019-000147>.
- Stenzinger A, Allen JD, Maas J, et al. Tumor mutational burden standardization initiatives: recommendations for consistent tumor mutational burden assessment in clinical samples to guide immunotherapy treatment

- decisions. *Genes Chromosomes Cancer*. 2019;58(8):578–88. <https://doi.org/10.1002/gcc.22733>.
25. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9(1):34. <https://doi.org/10.1186/s13073-017-0424-2>.
  26. Buchhalter I, Rempel E, Endris V, et al. Size matters: dissecting key parameters for panel-based tumor mutational burden analysis. *Int J Cancer*. 2019;144(4):848–58. <https://doi.org/10.1002/ijc.31878>.
  27. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol*. 2002;161(6):1961–71. [https://doi.org/10.1016/S0002-9440\(10\)64472-0](https://doi.org/10.1016/S0002-9440(10)64472-0).
  28. Jennings LJ, Arcila ME, Corless C, et al. Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn JMD*. 2017;19(3):341–65. <https://doi.org/10.1016/j.jmoldx.2017.01.011>.
  29. Hong J, Gresham D. Incorporation of unique molecular identifiers in TruSeq adapters improves the accuracy of quantitative sequencing. *Biotechniques*. 2017;63(5):221–6. <https://doi.org/10.2144/000114608>.
  30. Zhou W, Chen T, Zhao H, et al. Bias from removing read duplication in ultra-deep sequencing experiments. *Bioinformatics*. 2014;30(8):1073–80. <https://doi.org/10.1093/bioinformatics/btt771>.
  31. Ebbert MTW, Wadsworth ME, Staley LA, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics*. 2016;17(7):239. <https://doi.org/10.1186/s12859-016-1097-3>.
  32. Endris V, Buchhalter I, Allgäuer M, et al. Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in silico and real-life analysis of three larger gene panels. *Int J Cancer*. 2019;144(9):2303–12. <https://doi.org/10.1002/ijc.32002>.
  33. Vega DM, Yee LM, McShane LM, et al. Aligning tumor mutational burden (TMB) quantification across diagnostic platforms: phase II of the Friends of Cancer Research TMB Harmonization Project. *Ann Oncol*. 2021;32(12):1626–36. <https://doi.org/10.1016/j.annonc.2021.09.016>.
  34. Chen G, Mosier S, Gocke CD, Lin MT, Eshleman JR. Cytosine deamination is a major cause of baseline noise in next generation sequencing. *Mol Diagn Ther*. 2014;18(5):587–93. <https://doi.org/10.1007/s40291-014-0115-2>.
  35. Guo Q, Lakatos E, Bakir IA, Curtius K, Graham TA, Mustonen V. The mutational signatures of formalin fixation on the human genome. *Nat Commun*. 2022;13(1):4487. <https://doi.org/10.1038/s41467-022-32041-5>.
  36. Berra CM, Torrezan GT, de Paula CA, Hsieh R, Lourenço SV, Carraro DM. Use of uracil-DNA glycosylase enzyme to reduce DNA-related artifacts from formalin-fixed and paraffin-embedded tissues in diagnostic routine. *Appl Cancer Res*. 2019;39(1):7. <https://doi.org/10.1186/s41241-019-0075-2>.
  37. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–21. <https://doi.org/10.1038/nature12477>.
  38. Sun JX, He Y, Sanford E, et al. A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal. *PLoS Comput Biol*. 2018;14(2):e1005965. <https://doi.org/10.1371/journal.pcbi.1005965>.
  39. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–43. <https://doi.org/10.1038/s41586-020-2308-7>.
  40. Fancello L, Guida A, Frige G, et al. TMBler, a bioinformatic tool to optimize TMB estimation and predictive power. *Bioinforma Oxf Engl*. 2021:btab836. <https://doi.org/10.1093/bioinformatics/btab836>. Published online December 20.
  41. Goodman AM, Sokol ES, Frampton GM, Lippman SM, Kurzrock R. Microsatellite-stable tumors with high mutational burden benefit from immunotherapy. *Cancer Immunol Res*. 2019;7(10):1570–3. <https://doi.org/10.1158/2326-6066.CIR-19-0149>.
  42. Barroso-Sousa R, Keenan TE, Pernas S, et al. Tumor mutational burden and PTEN alterations as molecular correlates of response to PD-1/L1 blockade in metastatic triple-negative breast cancer. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2020;26(11):2565–72. <https://doi.org/10.1158/1078-0432.CCR-19-3507>.
  43. Okamura R, Kato S, Lee S, Jimenez RE, Sicklick JK, Kurzrock R. ARID1A alterations function as a biomarker for longer progression-free survival after anti-PD-1/PD-L1 immunotherapy. *J Immunother Cancer*. 2020;8(1):e000438. <https://doi.org/10.1136/jitc-2019-000438>.
  44. Assou S, Theou-Anton N, Nguenang M, et al. Association of TP53 mutations with response and longer survival under immune checkpoint inhibitors in advanced non-small-cell lung cancer. *Lung Cancer*. 2019;132:65–71. <https://doi.org/10.1016/j.lungcan.2019.04.005>.
  45. Basse C, Morel C, Alt M, et al. Relevance of a molecular tumour board (MTB) for patients' enrolment in clinical trials: experience of the Institut Curie. *ESMO Open*. 2018;3(3):e000339. <https://doi.org/10.1136/esmooopen-2018-000339>.
  46. Moreira A, Poulet A, Masliah-Planchon J, et al. Prognostic value of tumor mutational burden in patients with oral cavity squamous cell carcinoma treated with upfront surgery. *ESMO Open*. 2021;6(4):100178. <https://doi.org/10.1016/j.esmooop.2021.100178>.
  47. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733–745. <https://doi.org/10.1093/nar/gkv1189>.
  48. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308–11. <https://doi.org/10.1093/nar/29.1.308>.
  49. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019;47(D1):D941–7. <https://doi.org/10.1093/nar/gky1015>.
  50. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74. <https://doi.org/10.1038/nature15393>.
  51. SciCrunch | Research Resource Resolver. [https://scicrunch.org/resolver/SCR\\_012761](https://scicrunch.org/resolver/SCR_012761). Accessed 8 Feb 2022.
  52. Zhang J, Bajari R, Andric D, et al. The International Cancer Genome Consortium Data Portal. *Nat Biotechnol*. 2019;37(4):367–9. <https://doi.org/10.1038/s41587-019-0055-9>.
  53. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125–37. <https://doi.org/10.1093/hmg/ddu733>.
  54. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. <https://doi.org/10.1093/nar/gkq603>.
  55. Karczewski KJ, Weisburd B, Thomas B, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45(D1):D840–5. <https://doi.org/10.1093/nar/gkw971>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.