Integrative transcriptome analysis of malignant pleural mesothelioma reveals a clinically relevant immune-based classification

Ania Alay,1,2 David Cordero,1,2 Sara Hijazo-Pechero,1 Elisabet Aliagas,3 Adriana Lopez-Doriga,1,2 Raúl Marín,1 Ramón Palermo,4 Roger Llatjós,5 Ignacio Escobar,6 Ricard Ramos,6 Susana Padrones,7 Víctor Moreno,2,8 Ernest Nadal1,4 Xavier Solé1,2

ABSTRACT

Background Malignant pleural mesothelioma (MPM) is a rare and aggressive neoplasia affecting the lung mesothelium. Immune checkpoint inhibitors (ICI) in MPM have not been extremely successful, likely due to poor identification of suitable candidate patients for the therapy. We aimed to identify cellular immune fractions associated with clinical outcome and classify patients with MPM based on their immune contexture. For each defined group, we sought for molecular specificities that could help further define our MPM classification at the genomic and transcriptomic level, as well as identify differential therapeutic strategies based on transcriptional signatures predictive of drug response.

Methods The abundance of 20 immune cell fractions in 516 MPM samples from 7 gene expression datasets was inferred using gene set variation analysis. Identification of clinically relevant fractions was performed with Cox proportional-hazards models adjusted for age, stage, sex, and tumor histology. Immune-based groups were defined based on the identified fractions.

Results T-helper 2 (T\textsubscript{H2}) and cytotoxic T (T\textsubscript{C}) cells were found to be consistently associated with overall survival. Three immune clusters (IG) were subsequently defined based on T\textsubscript{H2} and T\textsubscript{C} immune infiltration levels: IG1 (54.5%) was characterized by high T\textsubscript{H2} and low T\textsubscript{C} levels, IG2 (37%) had either low or high levels of both fractions, and IG3 (8.5%) was defined by low T\textsubscript{H2} and high T\textsubscript{C} levels. IG1 and IG3 groups were associated with worse and better overall survival, respectively. While no differential genomic alterations were identified among immune groups, at the transcriptional level, IG1 samples showed upregulation of proliferation signatures, while IG3 samples presented upregulation of immune and inflammation-related pathways. Finally, the integration of gene expression with functional signatures of drug response showed that IG3 patients might be more likely to respond to ICI.

Conclusions This study identifies a novel immune-based signature with potential clinical relevance based on T\textsubscript{H2} and T\textsubscript{C} levels, unveiling a fraction of patients with MPM with better prognosis and who might benefit from immune-based therapies. Molecular specificities of the different groups might be used to tailor specific potential therapies in the future.

BACKGROUND

Malignant pleural mesothelioma (MPM) is a rare and highly aggressive neoplasia arising in the pleural cavity, and its incidence is increasing worldwide. MPM tumors are classified into three distinct histological subtypes: epithelioid, biphasic, and sarcomatoid. Median overall survival in patients with MPM is approximately 12 months, although patients with epithelioid tumors usually have better prognosis than individuals with biphasic or sarcomatoid MPM.1 Furthermore, most patients are diagnosed in advanced stages, which still aggravates the burden caused by the disease. The use of surgery in patients with MPM is very limited, and current standard of care treatment for patients with MPM is chemotherapy with platinum combined with pemetrexed, which has been shown to improve overall survival and quality of life.2 However, benefits from chemotherapy are generally modest and prognosis remains dismal with 5-year survival rates lower than 5%.3 Furthermore, there is no standard second-line treatment when tumor progresses to front-line chemotherapy.

Immunotherapy based on monoclonal antibodies against programmed cell death protein 1 (PD-1) and programmed death-ligand 1 (PD-L1) have been tested in clinical trials in MPM. Several nonrandomized phase I/II trials testing single-agent immune checkpoint inhibitors (ICI) showed antitumor activity (9%–29%) and median progression-free survival ranging from 2.8 to 6.2 months.4 Preliminary results from phase II clinical trials combining in second-line inhibitors of cytotoxic T-lymphocyte-associated antigen 4 (CTLA4) plus anti-PD1/PD-L1, such as ipilimumab and nivolumab or tremelimumab...
and durvalumab, showed promising efficacy results and also significant toxicity.\textsuperscript{5} In those clinical trials, PD-L1 expression had limited value to predict benefit from ICI. Moreover, tumor mutational burden (TMB) which is generally low in MPM\textsuperscript{6} is not likely to become a predictive marker in this tumor. Therefore, it is essential to further study the role of the different components of the immune system in the evolution and prognosis of MPM, as well as the identification of additional biomarkers of immunotherapy response beyond PD-L1 immunohistochemical expression.

In this study, our aim is to capture the landscape of the immune microenvironment in a large collection of 516 MPM gene expression profiles comprised in seven different publicly available datasets.\textsuperscript{6-12} Once the immune contexture is determined for every tumor, we will identify those immune cell fractions with potential prognostic ability and classify MPM samples based on these previously identified clinically relevant immune cell types. We will subsequently use the defined MPM groups to identify genomic and transcriptomic group specificities and evaluate functional signatures of drug response to suggest possible therapeutic strategies.

\section*{METHODS}

\subsection*{Data collection and processing}

A systematic search to identify publicly available datasets was performed in January 2018 using the query “pleural mesothelioma” and filtering to keep only human samples and expression profiling experiments in multiple open access repositories. We also filtered datasets with less than 30 samples and kept experiments that covered most of the transcriptome. Raw data were downloaded whenever possible and subsequently processed accordingly as summarized hereafter: for whole-exome sequencing data, Broad’s Genome Analysis Toolkit best practices were followed, RNA-sequencing data were processed according to an in-house pipeline based on ENCODE best practices, and for expression arrays, robust multiarray average algorithm was used to normalize the data. A detailed description of data identification and data processing is available in online supplemental extended methods.

\subsection*{Association of immune fractions with overall survival}

The immune infiltrate composition of 20 different cell types was deconvoluted from the obtained gene expression profiles using gene set variation analysis (GSVA) method.\textsuperscript{13} Briefly, for each sample and immune cellular fraction, we obtained a score between [-1, +1], with extreme values close to 1 or -1 indicating a relative strong presence or absence of a fraction in a specific sample, respectively. Similarly, GSVA scores close to 0 correspond to an unbiased distribution (ie, centered or randomly distributed) of the genes across the whole gene expression profile. A more detailed explanation on how this algorithm works or why its application is more robust to batch effect than other deconvolution methods can be found in Hänzelmann \textit{et al}\textsuperscript{13} and Tamborero \textit{et al}\textsuperscript{14}, respectively.

Gene signatures of the different immune fractions were obtained from a previous study\textsuperscript{15} in which signatures for a total of 24 different immune cell types were generated, including innate immune cells (eg, dendritic cells, eosinophils, mast cells, macrophages, natural killer cells, neutrophils) and adaptive immune cells (eg, B cells, T cells). Since GSVA can only deconvolute multigene signatures, we replaced two single-gene signatures (plasmacytoid dendritic cells and regulatory T cells) with their corresponding multigene signatures from another study.\textsuperscript{16} Additionally, when more specific categories were available for a cellular fraction (eg, NK CD56dim and NK CD56bright), general categories (eg, NK cells) were removed to avoid signature redundancy.

Once relative abundances for each immune fraction and sample were obtained, Cox proportional-hazards models adjusted for sex, stage, age, and histology were fitted using the largest dataset with available survival information as the discovery dataset,\textsuperscript{6} and setting the optimal cut point that maximized survival differences for each immune fraction using R packages \textit{survival} (V.3.1.7) and \textit{survMisc} (V.0.5.5). Using the GSVA cut-points defined in the discovery dataset, Cox proportional-hazards models adjusted for age, stage, sex, and histology were fitted in two additional validation datasets.\textsuperscript{7,10} After verifying there was no association of neoadjuvant therapy status with clinical outcome in Bueno \textit{et al}\textsuperscript{7} (p=0.2, data not shown), we did not adjust for neoadjuvant therapy status due to lack of information in Bott \textit{et al}\textsuperscript{10} and lack of diversity in Hmeljak \textit{et al}\textsuperscript{10} (all patients are naïve) in order to have equivalent models to test the association of clinical outcome with the immune fractions. Fractions associated to clinical outcome were determined as those significant (p<0.05) after false discovery rate (FDR) adjustment in the discovery dataset and at least one validation dataset, and with a consistent HR across the three datasets.

\subsection*{Group characterization}

\textbf{Clinicopathological variables}

Association with clinicopathological variables (age, sex, stage, tumor histology, asbestos exposure, and neoadjuvant therapy) was done using \textit{compareGroups} package for R (V.4.2.0). For categorical variables, a $\chi^2$ test was performed, and continuous variables were tested using Kruskal-Wallis test. Moreover, a multivariate analysis was performed using a multinomial logistic regression model.

\textbf{Immune checkpoints and T cell exhaustion status}

To assess immune checkpoints status and T cell exhaustion status in the different immune groups, multiple sets of genes were obtained from previous studies, including lists of immune checkpoint activators, ICI,\textsuperscript{17} and T cell exhaustion and effector markers.\textsuperscript{18} For each gene, we performed a linear model adjusted for sex, age, stage, and histology in each dataset, where the immune group was codified as a numeric variable and the gene
expression was the dependent variable. Datasets with less than three samples in any immune group were discarded. The weighted Z method from the survcomp package (V.1.36.0) was used to combine the p-values obtained for the different datasets. Bonferroni adjusted p values lower than 0.05 were considered as significant. β values were combined using the weighted mean.

Genomic alterations
Both Bueno et al.6 and Hmeljak et al.7 datasets had somatic variant data available to evaluate the mutational burden. The total number of variants per sample was calculated excluding non-exonic and synonymous mutations. Using somatic single nucleotide variants, mutational signatures were inferred using the R package deconstructSigs (V.1.8.0). Neoeptope binding affinity was also derived from somatic single nucleotide variants. More details on these analyses are available in online supplemental extended methods.

Copy number alteration burden was evaluated in the samples from the Hmeljak et al. dataset. Fisher tests were computed to assess the differences between immune groups in terms of number of altered/lost/gained genes. R package copynumber (V.1.24.0) was used to plot the genome overview of altered samples.

Transcriptomic analysis
The association of the expression levels of individual genes with the immune groups was done as described in the immune checkpoint evaluation section. For pathway analysis, a preranked gene set enrichment analysis (GSEA) was run with genes ranked using the product of the slope and the negative log10(p value) from the linear model used to assess transcriptomic status at the single-gene level. Gene sets tested included hallmarks and canonical pathways (KEGG, Biocarta, Reactome) from MSigDB V.6.1.19

Assessment of response to therapy
Transcriptional signatures of sensitivity or resistance to specific treatments were used to compute GSVA scores for each sample. The tested signatures were: pemetrexed resistance,20 cisplatin resistance,21 PARP inhibitors sensitivity,22 and palbociclib resistance,23 as well as multiple signatures representing the abundance of immune cell populations by applying GSVA on a previously established set of immune-specific gene signatures (online supplemental table S2).15-16

Once the abundance of all immune fractions was quantified in the full cohort of samples, we aimed to assess whether there existed any individual fractions associated with overall patient survival using Cox proportional-hazards regression models adjusted for potential confounders (ie, age, sex, tumor stage, histology, and dataset). Thus, for each cell type, we identified the cut point that maximized the survival differences between high versus low using the largest available dataset (Bueno et al.6). The cut points obtained in the discovery dataset were further validated in the two additional cohorts with available survival information (Hmeljak et al.7 and Bott et al.30). Out of the 20 studied immune cell types, THelper 2 (TH2) and cytotoxic T cells (TC) showed a consistent association with patient survival across datasets based on our criteria (Cox p<0.05 in at least two of the three datasets, and the remaining dataset showing a consistent trend, although not necessarily significant). In our analysis, higher levels of TH2 cells (GSVA score >0.40) were associated with worse outcome (HR_Bueno =2.1, p=0.00015; HR_Hmeljak =2.6, p=0.0021; HR_Bott =2.3, p=0.066; HR_Combined =2.1, p=7.6×10−7) (figure 1A), while higher levels of TC cells (GSVA score >0.40) were associated with better prognosis (HR_Bueno =0.57, p=0.0091; HR_Hmeljak =0.73, p=0.39; HR_Bott =0.29, p=0.016; HR_Combined =0.59, p=0.00004) (figure 1B). Survival analysis results for the complete set of immune fractions can be found in online supplemental figure S1. When individuals were stratified in four groups based on the combination of the abundance (ie, low versus high) of both TH2 and TC cells, patients with MPM with low abundance of TH2 and high TC infiltration consistently showed the best overall survival when compared with patients with high levels of TH2 and low levels of TC (HR_Bueno =0.34, p=0.00022; HR_Hmeljak =0.20, p=0.0080; HR_Bott =0.12, p=0.020; HR_Combined =0.31, p=2.7×10−7) (figure 1C). Additionally, patients with either both high (IG2-H) or low abundance levels of both cell types (IG2-L) showed intermediate prognosis, suggesting some type of compensatory effect between these two immune cell types. Moreover, we do not observe statistically significant differences in survival when comparing IG2-H and IG2-L in any of the datasets. In Hmeljak et al. cohort where although IG2-H patients (n=10) have a similar survival to IG1 patients, the HR and 95% CI between IG2-L and IG2-H (HR =0.39 (0.15 to 0.9) is not statistically significant, with IG3 being the only group that shows an improved survival in that cohort. Therefore, we decided to merge these two groups into a single category to create a final classification of three
immune-based groups: TH2 High-T C Low (IG1, worse prognosis, 54.5% of the patients), TH2 High-T C High or TH2 Low-T C Low (IG2, intermediate prognosis, 37% of the patients, HR Combined =0.52), TH2 Low-T C High (IG3, better prognosis, 8.5% of the patients, HR Combined =0.32) (figure 1D).

Once the immune-based MPM groups were defined, we aimed to identify any potential differences in the remaining immune fractions among them. Groups IG1 and IG2 comprised a mixture of samples with both strongly and less immune-enriched tumors, while on average, group IG3 showed significant higher abundance for 13 of the 20 immune cell fractions (figure 2A, online supplemental figure S2). These results suggest that the combination of these two immune fractions interplays with the remaining immune cell types and contributes to identify a subset of highly immune-infiltrated MPM tumors with improved prognosis, as already reported in previous publications.14

Regarding tumors in IG2 (figure 2A), we observe that samples from the cluster on the left-side are mainly from
IG2-L, although 6.06% (n=6/105) is composed by IG2-H samples. On the right-side, consisting mainly of IG2-H samples, 41.0% of the samples (n=25/86) are from IG2-L, suggesting that samples with similar survival rates can have a heterogeneous landscape regarding immune cell abundance.

In order to explore potential additional clinical correlates of the three identified MPM immune groups,
we analyzed their association with available clinical covariates: age, sex, tumor stage at the time of diagnosis, tumor histology, asbestos exposure, and neoadjuvant therapy. Out of all the variables explored, we identified a significant association with age and histology (table 1). Patients’ median age in group IG1 is 66 years, compared with 63.7 years and 63.2 years for groups IG2 and IG3, respectively (p=0.021). Regarding tumor histology, IG3 showed a strong enrichment in epithelioid tumors (90.70% in IG3 vs 62.04% in IG1 and 73.68% in IG2, p=0.001).

A multivariate analysis with the observations with complete data for all the covariates (n=231) confirmed the association with histology and age, and the lack of association of our classification with sex and stage (online supplemental table S3). A slight increase of asbestos exposure in IG2 as well as a marginal association with neoadjuvant therapy in IG2 was also observed.

To further exclude a potential confounding role of the tumor histology on the differential survival among the immune groups, we evaluated whether the immune-based classification still was associated with patient survival within both epithelioid and non-epithelioid tumors. In epithelioid tumors, we observed improved survival in IG2 and IG3 patients compared with IG1 (HR<sub>IG2</sub>=0.60, p=0.074) (online supplemental figure S3B). Finally, a comparison with previously published classifications was done. While the immune-based classification was not associated to De Reyniès et al<sup>9</sup> classification, it was associated to Bueno et al<sup>6</sup> classification, where IG3 is enriched in samples from Epithelioid subtype, and Hmeljak et al<sup>7</sup> classification, where IG3 is enriched in S1 samples which is also enriched in epithelioid histology and has the best prognosis.

**Immune groups correlate with most immune checkpoint markers and markers of effector and exhausted T cells**

To further evaluate the potential clinical relevance of the identified immune groups, we evaluated the expression of a wide selection of ICI and activators. We observed a significant increasing expression pattern across MPM immune groups for most of the evaluated checkpoint markers (figure 2B). TNFRSF14, IL2RB, TNFRSF18 (GITR), CD27, and ICOS were among the strongest positively associated activator markers with the immune groups. Regarding ICI, LAG3, TIGIT, VSIR (VISTA), IDO1, CTLA4, and PDCD1 (PD-I) were the strongest correlated markers. In opposition, CD276 (B7-H3) showed negative correlation with the immune groups, being its expression higher in IG1 tumors than tumors in IG2 and IG3 groups. Detailed expression of CD274 (PD-L1), PDCD1 (PD-I), CTLA4, CD276 (B7-H3), and VSIR (VISTA) in

<table>
<thead>
<tr>
<th>N</th>
<th>IG1</th>
<th>IG2</th>
<th>IG3</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (range)</td>
<td>N=281</td>
<td>N=191</td>
<td>N=44</td>
<td></td>
</tr>
<tr>
<td>Sex, N (%)</td>
<td>63.00 (28.00–86.00)</td>
<td>63.70 (30.20–84.50)</td>
<td>63.20 (18.80–76.00)</td>
<td>0.021</td>
</tr>
<tr>
<td>Male</td>
<td>166 (83.42%)</td>
<td>120 (81.63%)</td>
<td>27 (69.23%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>33 (16.58%)</td>
<td>27 (18.37%)</td>
<td>12 (30.77%)</td>
<td></td>
</tr>
<tr>
<td>Stage, N (%)</td>
<td>63.70 (30.20–84.50)</td>
<td>27 (18.37%)</td>
<td>12 (30.77%)</td>
<td>0.888</td>
</tr>
<tr>
<td>Stage I</td>
<td>10 (5.32%)</td>
<td>9 (7.26%)</td>
<td>3 (8.33%)</td>
<td></td>
</tr>
<tr>
<td>Stage II</td>
<td>33 (17.55%)</td>
<td>21 (16.94%)</td>
<td>7 (19.44%)</td>
<td></td>
</tr>
<tr>
<td>Stage III</td>
<td>108 (57.45%)</td>
<td>68 (54.84%)</td>
<td>22 (61.11%)</td>
<td></td>
</tr>
<tr>
<td>Stage IV</td>
<td>37 (19.68%)</td>
<td>26 (20.97%)</td>
<td>4 (11.11%)</td>
<td></td>
</tr>
<tr>
<td>Histology, N (%)</td>
<td>507</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epithelioid</td>
<td>170 (62.04%)</td>
<td>140 (73.68%)</td>
<td>39 (90.70%)</td>
<td></td>
</tr>
<tr>
<td>Biphasic</td>
<td>84 (30.66%)</td>
<td>39 (20.53%)</td>
<td>4 (9.30%)</td>
<td></td>
</tr>
<tr>
<td>Sarcomatoid</td>
<td>20 (7.30%)</td>
<td>11 (5.79%)</td>
<td>0 (0.00%)</td>
<td></td>
</tr>
<tr>
<td>Asbestos exposure, N (%)</td>
<td>0.068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>41 (22.53%)</td>
<td>42 (31.82%)</td>
<td>15 (40.54%)</td>
<td></td>
</tr>
<tr>
<td>Possible</td>
<td>1 (0.55%)</td>
<td>2 (1.52%)</td>
<td>1 (2.70%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>140 (76.92%)</td>
<td>88 (66.67%)</td>
<td>21 (56.76%)</td>
<td></td>
</tr>
<tr>
<td>Neoadjuvant therapy, N (%)</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>163 (93.68%)</td>
<td>101 (87.07%)</td>
<td>30 (85.71%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11 (6.32%)</td>
<td>15 (12.93%)</td>
<td>5 (14.29%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Summary of clinicopathological variables by immune group
each dataset is available in online supplemental figure S4A, showing the distribution of tumors in each group. An overall higher expression of VI3R with respect to the other immune checkpoints is observed, as previously reported in other publications. 7,31 Importantly, immune groups still show significant differences in prognosis independently of the expression level of any of these five immune checkpoints (online supplemental figure S4B). Additionally, we also evaluated markers of effector and exhausted T cells correlated with the identified MPM immune groups (figure 2C). In our analysis, all evaluated markers—excluding CD44—showed a positive correlation with the immune group classification, with a majority being significantly enriched in IG3 (PRF1, GZB, TBX21, EOMES, PDCD1).

**Genomic characterization shows no differences in genomic instability among immune groups**

Once we identified the different immune-based groups in MPM tumors and established their potential clinical relevance, we wanted to characterize the different groups at the genomic level to identify any potential differential genomic patterns. These analyses were performed using Bueno et al 6 and Hmeljak et al 7 datasets.

Using whole-exome data from Bueno et al 6 and Hmeljak et al 7 datasets, we assessed if immune groups were different in terms of TMB, mutational signatures, 1 or antigen immunogenicity (figure 3A–C). Although we did not find statistically significant differences among immune groups for any of the previous analyses, we did observe a slight decrease of mutational Signature 3 in IG3 tumors compared with IG1 and IG2 patients. Signature 3 is associated with homologous recombination deficiency and BRCA1/2 deficiency. 32 Additionally, at the single-gene level, no gene was found to be preferentially mutated in any group (data not shown).

Copy number alterations were assessed for Hmeljak et al dataset. 7 There were no significant differences between immune groups in the overall number of altered genes per sample, although tumors in IG3 tend to have fewer copy number alterations overall compared with the other groups as shown in figure 3D. Figure 3E shows the overall percentage of altered samples across the genome for each immune group. Regarding MPM landmark genes, genes like BAP1, SETD2, LAT52, TP53, and NF2 did not show differences between groups. CDKN2A is altered in 81% of IG1 patients, 47% of IG2 patients, and 14% of IG3 patients, but it does not reach statistical significance after multiple testing correction, probably due to small sample size. Online supplemental figure S5 shows expression of frequent tumor suppressor genes inactivated in MPM. 3 For the case of CDKN2A, we see an increasing expression IG1<IG2<IG3, which is in agreement with the pattern observed at the genomic level.

We also evaluated differences in methylation data at the single-gene level and no significant results were found after adjusting for multiple testing (data not shown).

**Transcriptional cell cycle/immune system activity trade-off across immune groups**

Next, we sought for gene expression differences among the three immune groups, to identify those genes with an increasing or decreasing pattern among the three groups. The results are depicted in figure 4A, which shows a high number of genes with decreasing expression from IG1 to IG3 tumors related to cell proliferation such as CENPF (centromere protein also known as mitosin), BIRC5 (otherwise known as survivin, preventing apoptotic cell death), or kinesins like KIF23, involved in cell division. On the other hand, genes positively correlated with the immune groups include immune system related genes such as HSH2D, a target of T-cell activation signaling pathways, and GNL3, which encodes for a protein present in cytotoxic granules of cytotoxic T lymphocytes.

In order to capture more precisely the enriched pathways and biological functions of the genes identified in the previous analysis, we performed a preranked GSEA using a set of transcriptional signatures covering a wide range of pathways, and the genes ranked according to their linear association with the immune groups. The results are shown in figure 4B and we observed that signatures more positively correlated with the immune group (ie, more expressed in IG3 tumors) were mostly related to immune system and inflammatory response processes. Contrarily, signatures negatively correlated with the immune groups mostly relate to cell cycle and epithelial mesenchymal transition, suggesting a trade-off between these molecular processes and the immune system.

Finally, since IG1 tumors appeared to be driven by cell cycle deregulation, we assessed the potential confounder effect among cell proliferation and immune groups. Therefore, we tested the association between the surrogate cell proliferation marker MKI67 and the immune groups. Notably, regardless of MKI67 expression levels, the association of the improving survival pattern across immune groups is upheld (online supplemental figure S6), meaning that our classification is not confounded with cell proliferation.

**Potential therapeutic strategies**

To further evaluate the potential clinical benefit of the immune-based MPM classification, we decided to assess the behavior of different treatment-response signatures among the three groups. We first wanted to identify whether there were any differences in the response to the currently available standard chemotherapy treatment based on the combination of cisplatin and pemetrexed. We obtained two expression-based signatures derived from non-small cell lung cancer that predicted resistance to pemetrexed 20 and to cisplatin. 21 Figure 5A displays GSVA scores for the two signatures across the three MPM immune groups. Overall, the results of the analysis suggested that on average, MPM tumors in IG1 might be more resistant to pemetrexed, while IG3 tumors show a trend to lower resistance. On the other
Figure 3  Genomic characterization. (A) Tumor mutational burden among immune groups. Mean number of variants per sample. Error bars indicate SD. (B) Distribution of COSMIC mutational signatures among each immune group. (C) Immunogenicity as the mean of IC50 affinity of mutations among immune groups. (D) Copy number burden among immune groups. Mean number of altered genes per sample. Error bars indicate SD. Lighter colors show copy number gains while darker ones depict copy number losses. (E) Genomic overview showing the percentage of samples with copy number alterations among each immune group. Landmark MPM genes are depicted according to their genomic location. MPM, malignant pleural mesothelioma.
**Figure 4** Transcriptomic characterization. (A) Volcano plot of gene expression. Effect size ($\beta$) correlates with immune groups. (B) Significant pathways from preranked GSEA. Positive normalized enrichment scores correlate with an increasing expression of pathway from IG1 to IG3, while negative ones depict higher pathway expression for IG1 tumors versus IG3 tumors. Pathways are clustered using Ward’s method and an OR-based distance. GSEA, gene set enrichment analysis.
Figure 5  Assessment of gene expression signatures predictive of benefit or resistance to multiple treatments. (A) Enrichment assessment for signatures of resistance to currently established first-line chemotherapy in patients with MPM (pemetrexed—left panel, cisplatin—mid and right panels). Each MPM tumor is represented with a dot, and GSVA scores (y-axis) indicate upregulation or downregulation of the signature in each patient (as indicated in each plot title). Signature of resistance to cisplatin is divided into two subsignatures of upregulated and downregulated genes in cisplatin resistant cells. (B) Assessment of response to multiple ICI therapies. Each line represents a different ICI predictive signature. Signatures 1–9 are upregulated in ICI responders, while signatures 10 and 11 are downregulated in ICI responders, based on the literature. For each signature, we represent the $\beta$ coefficient and the 95% CI of the linear model of the GSVA scores across the three immune groups. Positive $\beta$ values indicate increasing GSVA scores (IG1<IG2<IG3) and negative $\beta$ values indicate decreasing GSVA scores (IG1>IG2>IG3). Superscripts in immunotherapy signatures correspond to bibliographic references (C) Analogously to panel A, assessment for signatures of resistance or sensitivity to targeted therapy (palbociclib—left panel, PARP inhibitors—mid and right panels). Signature of sensitivity to PARP inhibitors is divided into two subsignatures of upregulated and downregulated genes in either sensitive or resistant cells for the specified treatment.
hand, cisplatin resistance analysis showed no differences between immune groups.

Next, we tested a large set of gene-expression signatures predictive of benefit from ICI treatment\textsuperscript{24-30} to formally evaluate if MPM tumors in IG3 could benefit from an immune-based therapeutic approach (figure 5B). The overlap between these signatures and the classifier is almost non-existent (three shared genes at most). We observed that all signatures were positively correlated with the immune groups (ie, higher GSVA scores in IG3 for upregulated signatures (signatures 1 to 9) and lower GSVA scores in IG3 for downregulated signatures (signatures 10 and 11)), with IG3 tumors presenting a significant association for all but two of the signatures (signatures 9 and 10).

Given the increased activation of cell cycle and DNA repair processes in the IG1 group of MPM tumors, we wanted to explore whether any differences in sensitivity to cyclin-dependent kinase inhibitors and to drugs that inhibit DNA repair among the three groups. To do that, we collected a signature of resistance to palbociclib derived from breast cancer\textsuperscript{23} and another computationally-derived signature predicting sensitivity to DNA repair inhibition with PARP inhibitors\textsuperscript{32} (figure 5C). Despite a higher cycling nature of tumors in group IG1, they appear to be more resistant to palbociclib than tumors in other groups, based on this predictive signature. Strikingly, tumors in IG1 showed increased sensitivity to PARP inhibitors compared with the other groups.

**DISCUSSION**

In this study, we performed a comprehensive analysis of publicly available transcriptome data from more than 500 MPM tumors in order to characterize the immune landscape of this disease. By the identification of immune cells associated with clinical outcome, we classified MPM tumors according to their T-helper 2 and cytotoxic T cell abundance levels. This classification stratifies patients into three groups that represent different immune infiltration patterns and are associated with distinct survival outcomes. The group with the shortest overall survival, IG1, represents more than 50% of the analyzed tumors, while IG3, with better prognosis, accounts for only 8.5% of the tumors. We observed an increasing pattern of abundance for most immune fractions across the three immune groups. Interestingly, the survival benefit of this classification is consistent in both epithelioid and non-epithelioid tumors. We further characterized these three MPM immune groups at the genomic and transcriptomic levels and identified potential therapeutic strategies using predictive signatures and large-scale pharmacogenomics data.

Different immune-related MPM phenotypes have already been described in recent studies using comprehensive approaches.\textsuperscript{35-37} These studies use different techniques such as Nanostring technology,\textsuperscript{38} mass cytometry,\textsuperscript{39} or RNA-sequencing.\textsuperscript{40} The diversity of obtained results suggests that the composition and role of the tumor microenvironment in MPM is remarkably complex and controversial, and thus further studies will contribute to elucidate this question.

In our work, MPM tumors are classified based on $T_{H2}$ and $T_C$ cells abundance, as previously stated. $T_{H2}$ are CD4+ T cells which are induced by the presence of interleukin 4 (IL4) via STAT6 signaling and regulate humoral immune responses and responses to extracellular pathogens. $T_{H2}$ cells secrete IL4, IL5, IL10, and IL13 which promote immune suppression by inhibiting T-helper 1 cytokine production. Additionally, a recent study identified higher levels of IL5 and IL13 after exposure to asbestos in mesothelial cells.\textsuperscript{36} The role of $T_{H2}$ cells in cancer has been found to be dual and context-dependent. While they can generate antitumor immunity by recruiting specific populations of innate immune cells, they have also been found to increase tumorigenicity in certain experimental models.\textsuperscript{39} Consistent with our results, higher levels of $T_{H2}$ cells have already been associated with poor prognosis in multiple cancer types.\textsuperscript{40-42} Finally, in the analysis performed by Hmeljak et al,\textsuperscript{7} this signature has already been associated to the group with the worst prognosis.

On the other hand, CD8+ $T_C$ cells are essential actors of the effector function of adaptive cellular immune response.\textsuperscript{43} Along with natural killer (NK) cells, $T_C$ cells are ultimately responsible for targeting and attacking cancer cells by secreting cytotoxins (eg, granzymes and perforins) which reach the tumor cell cytoplasm and trigger a caspase-mediated apoptotic process.\textsuperscript{44} Additionally, in agreement with our study, a recent pan-cancer study observed that patients with tumors with higher $T_C$ infiltration tended to have better survival than patients with less $T_C$-infiltrated tumors.\textsuperscript{44}

While most immune checkpoints correlate with the defined immune groups, $CD276$ ($B7-H3$) shows an opposite pattern of expression, with decreasing expression from IG1 to IG3. $CD276$ is a member of the B7 family of immunoregulatory proteins and is overexpressed in distinct tumor types. It has been shown that $CD276$ can promote tumor proliferation, angiogenesis and metastasis and is associated with shorter survival time.\textsuperscript{45} This result ties in with the survival pattern among the groups as well as with results from functional analysis in which cell cycle shows a decreasing pattern from IG1 to IG3. Similarly, $CD44$ is the only $T_C$ exhaustion marker that shows negative correlation with the immune groups. This marker has been associated with metastasis and low survival rate in multiple cancer types, and chemoresistance in prostate and head and neck squamous cell cancer types.\textsuperscript{46} In MPM, $CD44$ has been shown to promote invasiveness when interacting with hyaluronan.\textsuperscript{47}

Regarding molecular characteristics, immune groups were not significantly correlated with specific genomic alterations in terms of gene mutations or copy number alterations, and although there seems to be a pattern of decreasing genomic instability among the groups, there is no statistically significant evidence to support this
observation. In terms of copy number, no global differences were found among MPM groups, possibly due to the smaller sample size (only one dataset with available data). However, genomic regions like 9p21, harboring landmark genes in MPM as CDKN2A or type I interferon gene cluster, were likely to be lost in the IG1 compared with the other two groups. We would require larger copy number data in order to further validate this apparent lack of differences in genomic alterations among the three immune groups given that previous studies have reported that tumors with more aneuploidy overexpress cell proliferation markers and underexpress immune infiltration markers, which are the main characteristics of IG1 tumors.

Transcriptomic characterization of these groups by gene enrichment analysis revealed a trade-off between immune response activity and cell proliferation mechanisms, showing higher activation of cell cycle in IG1 tumors, in agreement with genomic characteristics. This trade-off has already been described in other studies, reinforcing the hypothesis that cell cycle regulators and genomic instability could also have an impact on immune checkpoints. Besides, IFN activation has also been associated with better clinical outcomes in other studies.

Finally, the lack of effective therapeutic strategies beyond chemotherapy in MPM calls for assessment of potential options and tailoring for each subset of patients. Currently, several phase III clinical trials are assessing the role of ICI in combination with chemotherapy in the first-line setting of patients with advanced MPM. The results of the CheckMate-743 trial evaluating nivolumab plus ipilimumab versus platinum-based chemotherapy in previously untreated MPM showed the survival benefit from ICI treatment. This trial also reports that ICI treatment performs better than chemotherapy in non-epitheloid samples, and this might seem contradictory with our results. In fact, the survival benefit of ICI treatment was observed, regardless of histology (in non-epitheloid samples chemotherapy performs worse, thus giving an impression of higher performance for ICI treatment), and PD-L1 ≥1% was also predictive of benefit, which is higher expressed in IG3 tumors from our classification. In this sense, the treatment landscape of advanced MPM is likely to evolve and predictive markers of ICI benefit are needed. In our work, we observed that patients belonging to IG3 may benefit from pemetrexed and immunotherapy. IG3 molecular characteristics have been associated with non-progressive disease after treatment with ICI in another study. Also, no group showed any strong predisposition to benefit from chemotherapy with cisplatin treatment, although we were unable to assess this predisposition using a signature that included both therapies used in standard of care.

Moreover, our data might suggest that tumors in IG1 could be more sensitive to PARP inhibitors. However, the reasons for this increased sensitivity remain unclear, since we do not observe differences at the genomic level among the three immune groups.

It is important to stress that the presented classification in three groups is based on clinical outcome and this can generate some heterogeneity at the molecular level, as shown in IG2 group immune cell abundances. These results need further validation in a prospective cohort in order to confirm these findings. Moreover, further research in prospective cohorts could contribute to the identification of the most suitable therapeutic strategy for patients in each immune group. It is also important to point out that this study focused on gene expression signatures as proxies since there is no information regarding immunotherapy in large pharmacogenomic assays and also that even though pemetrexed and cisplatin are commonly administered in combination; we assessed their effect as single therapies due to the lack of signatures for the combined treatment.

To sum up, this study identifies a novel signature with potential clinical relevance based on T_H2 and T_C levels and unveils a small fraction of patients with MPM that show better prognosis. These groups have different molecular characteristics, mainly at the transcriptomic level. These differences might be useful to tailor potential therapies for each group in the future. Further research is needed towards the identification of a signature valid for formalin-fixed, paraffin-embedded (FFPE) samples, and validation of these results is warranted in an independent and prospective cohort of patients with MPM preferentially treated with ICI.

Author affiliations
1Unit of Bioinformatics for Precision Oncology, Catalan Institute of Oncology (ICO)—Bellvitge Biomedical Research Institute (IDIBELL), L’Hospitalet de Llobregat, Barcelona, Spain
2Biomedical Research Centre Network for Epidemiology & Public Health (CIBERESP), Madrid, Comunidad de Madrid, Spain
3Molecular Mechanisms and Experimental Therapy in Oncology Program (Oncobell), Bellvitge Biomedical Research Institute (IDIBELL), L’Hospitalet de Llobregat, Barcelona, Spain
4Department of Medical Oncology, Catalan Institute of Oncology (ICO)—Bellvitge Biomedical Research Institute (IDIBELL), L’Hospitalet de Llobregat, Barcelona, Spain
5Department of Pathology, Bellvitge University Hospital, L’Hospitalet de Llobregat, Barcelona, Spain
6Department of Thoracic Surgery, Bellvitge University Hospital, L’Hospitalet de Llobregat, Barcelona, Spain
7Department of Respiratory Medicine, Bellvitge University Hospital, L’Hospital de Llobregat, Barcelona, Spain
8Cancer Research Spain (CRES), Barcelona, Spain
9Oncology Data Analytics Program, Catalan Institute of Oncology (ICO)—Bellvitge Biomedical Research Institute (IDIBELL)—University of Barcelona (UB), L’Hospital de Llobregat, Barcelona, Spain

Twitter Ernest Nadal @NadalErnest and Xavier Solé @xsole1977

Acknowledgements We thank CERCA Programme/Generalitat de Catalunya for institutional support. We thank all the researchers who kindly put their research data for public use. Bueno et al data were accessed through a data transfer agreement with Genentech (DAT-015429). We thank the reviewers for their time and for contributing to improve the manuscript.

Contributors XS and EN conceived and designed the study. AA and DC performed the computational and statistical analysis. XS, EN, AA, DC, SHP, and EA contributed to the interpretation of the results. XS, EN, AA, and DC drafted the manuscript. ALD, RM, RP, RL, IE, RR, SP, and VM provided critical revisions of the article. All authors read and approved the final manuscript.

Open access
Funding This work is supported by the Carlos III National Health Institute funded by FEDER funds—a way to build Europe—(PI14/01109; P11/00920); the Government of Catalonia (2017SGR448). XS is supported by RTI2018-102134-A-100 grant funded by Spanish Ministry of Science and Innovation. EN received support from the SLT06/17/00127 grant, funded by the Department of Health of the Generalitat de Catalunya by the call “Acció instrumental d’internifacció de professionals de la salut”. This study has been funded by Sociedad Española de Oncología Médica (SEOM) through “Proyectos de Investigación para Grupo Emergente”.

Competing interests EN participated in advisory boards from Bristol Myers Squibb, Merck Sharp & Dohme, Lilly, Roche, Pfizer, Takeda, Boehringer Ingelheim, Amgen, and AstraZeneca. VM is consultant to Bioiberica S.A.U. and Grupo Ferrer S.A., received research funds from UVAX and is coinvestigator in grants with Anilin.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets were generated and/or analyzed for this study. Data are available under the accession IDs provided by the researchers who kindly put their research data for public use.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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

ORCID iDs
Ania Alay http://orcid.org/0000-0002-5860-2913
David Cordero http://orcid.org/0000-0003-1470-2356
Elisabet Allagàs http://orcid.org/0000-0001-5606-3252
Ernest Nadal http://orcid.org/0000-0002-9674-5554
Xavier Solé http://orcid.org/0000-0002-2197-3325

REFERENCES


