Integrative Molecular Characterization of Malignant Pleural Mesothelioma

Malignant pleural mesothelioma (MPM) is a highly lethal cancer of the lining of the chest cavity. To expand our understanding of MPM, we conducted a comprehensive integrated genomic study, including the most detailed analysis of BAP1 alterations to date. We identified histology-independent molecular prognostic subsets, and defined a novel genomic subtype with TP53 and SETDB1 mutations and extensive loss of heterozygosity. We also report strong expression of the immune-checkpoint gene VISTA in epithelioid MPM, strikingly higher than in other solid cancers, with implications for the immune response to MPM and for its immunotherapy. Our findings highlight new avenues for further investigation of MPM biology and novel therapeutic options.

**SIGNIFICANCE:** Through a comprehensive integrated genomic study of 74 MPMs, we provide a deeper understanding of histology-independent determinants of aggressive behavior, define a novel genomic subtype with TP53 and SETDB1 mutations and extensive loss of heterozygosity, and discovered strong expression of the immune-checkpoint gene VISTA in epithelioid MPM. Cancer Discov; 8(12): 1548–65. © 2018 AACR.

See related commentary by Aggarwal and Albelda, p. 1508.
RESULTS

Cohort Description

We studied 74 samples of primary MPM from patients with no prior systemic therapy. This cohort was predominantly male (82%), with a median age of 64 years, and tumors were of mostly epithelioid histology (70%), a typical profile for MPM. Asbestos exposure history was positive in 62%, negative in 18%, and unavailable or unknown in the remainder. Demographic and clinical details are provided in Supplementary Tables S1A and S1B, as well as Supplementary Fig. S1.

We performed comprehensive molecular profiling, including exome sequencing, copy-number arrays (Supplementary Fig. S2), mRNA sequencing (Supplementary Fig. S3), noncoding RNA profiling, DNA methylation (Supplementary Fig. S4), and reverse-phase protein arrays (RPMA; Supplementary Fig. S5). Methods and detailed results of individual analyses are provided in Supplementary Sections 1–13.

Landscape of Somatic Mutations and Copy-Number Alterations

Whole-exome sequencing (WES) revealed a somatic mutation rate of <2 nonsynonymous mutations per megabase in all samples except for an outlier case with a mutation rate of 8 nonsynonymous mutations per megabase (Fig. 1A). This places MPM at the low end of somatic mutation burden among cancers (8). The outlier tumor with a 10-fold higher mutation rate showed a distinctive pattern of C>T mutations occurring almost exclusively at CpG dinucleotides (Fig. 1B). This relatively hypermutated tumor harbored a homozygous nonsense mutation in MSH2, which would suggest that the tumor lacked mismatch-repair capacity, but the observed mutational spectrum was atypical for mismatch-repair deficiency (9). Otherwise, the observed mutational spectrum was similar across patients and lacked distinctive or novel features (Fig. 1C). Signatures of smoking- or APOBEC-induced mutagenesis were not observed. Asbestos has been proposed...
to cause genotoxicity via DNA breaks and secondary oxidative damage (10). However, the mutational spectrum and local sequence context was not significantly different between cases with or without known asbestos exposure ($\chi^2$, $P = 0.3$); although this negative finding should be viewed with caution given the limitations of the data set, it is in agreement with prior studies (5).

We sought to identify genes mutated significantly above the background rate using MutSig2CV. Significantly mutated genes (SMG) at a false discovery rate (FDR) of $<0.05$ included BAP1, NF2, TP53, LAT2, and SETD2, all known cancer genes in MPM (Fig. 1A; Supplementary Table S1C). All five genes showed high rates of nonsense, frameshift, and splice-site mutations, highlighting them as targets of inactivation, consistent with their functions as tumor suppressors. Mutations in these five SMGs did not show associations with histories of asbestos exposure or smoking. For validation in an independent cohort that also used a different algorithm to define SMGs, we compared the results of our SMG analysis with the SMGs previously identified in 99 MPM exomes using the MUSIC algorithm (5) and found a strong overlap for SMGs at an FDR of $<0.05$ in both studies (Fig. 1D). Notably, among lower confidence SMGs (FDR $<0.15$ but $>0.05$), only SETDB1, which may define a novel subtype of MPM (discussed below), was identified in both analyses (Fig. 1D).

The somatic copy-number alteration (SCNA) landscape was characterized by frequent recurring focal and arm-level deletions, but no recurring amplifications, consistent with the notion that MPM development is driven primarily by loss of tumor suppressor genes, rather than by activation of classic oncogenic driver genes (Supplementary Section 4; Supplementary Fig. S2). Focal deletions affected several tumor suppressor genes known to be altered in MPM (5), most notably CDKN2A with deletions (defined as deep, likely homozygous) in 36 (49%) and losses (defined as shallow, possibly single-copy) in 5 (7%) samples. Likewise, NF2 deletions were confirmed in 25 (34%) samples and losses in 30 (40%) samples; as many of the latter harbored mutations of the remaining allele, evidence of biallelic NF2 inactivation was common (Fig. 1A). CDKN2A deletions often encompass MTAP, the adjacent gene on 9p21 (11), which encodes methylthioadenosine phosphorylase, whose deficiency has recently been reported to lead to reduced PRMT5 enzymatic activity and heightened sensitivity to its pharmacologic inhibition (12, 13). Cedeletion of CDKN2A and MTAP, associated with low levels of mRNA expression of both genes, was observed in 20 cases (Supplementary Fig. S2). Although no correlation with overall survival was observed for BAP1 status, loss of CDKN2A was strongly associated with shorter overall survival (Cox proportional hazards, $P = 7.3 \times 10^{-4}$), as previously shown (14, 15).

Finally, an analysis for genomically integrated viral sequences, including SV40, was negative (Supplementary Section 8), as was a screen of exome and RNA-sequencing data for evidence of EWSR1 and ALK fusions, recently reported in rare cases of MPM (16, 17) and peritoneal mesothelioma (18), respectively. As well, no activating mutations in the canonical MAPK or PI3K/AKT signaling pathways were identified in this cohort.

**Comprehensive Analysis of BAP1 Status in MPM**

BAP1, encoding a nuclear deubiquitinase, is the most frequently mutated cancer gene in MPM, in both our data set and others (5), and is also recurrently inactivated in clear cell renal carcinoma, uveal melanoma, and cholangiocarcinoma (19). Somatic BAP1 mutations have germline counterparts that define the BAP1 hereditary cancer syndrome (20). In contrast to other cancers, BAP1 inactivation in MPM does not correlate with adverse outcomes. Because its role in MPM biology remains unclear, we compared BAP1-inactivated and wild-type cases across multiple platforms. First, to better segregate cases according to BAP1 status, we performed a comprehensive analysis of inactivating alterations through a detailed review of single-nucleotide variants, small and large indels, whole gene deletions, and structural variants; this showed the overall prevalence of BAP1 alterations to be 57% (Fig. 1A; Supplementary Table S2A), in line with recent studies (21). Most MPM with inactivating mutations in BAP1 (25/26, 96%) also had concurrent loss of heterozygosity (LOH) on chromosome 3p21.1, supporting a classic two-hit tumor suppressor mechanism. Overall, BAP1 status was defined as follows: 32 samples with no evidence of BAP1 alteration, 6 with a single (heterozygous) mutation or deletion, and 36 with biallelic inactivation. No germline mutations in BAP1 were identified in this cohort.

BAP1 alterations showed nonrandom patterns of occurrence with mutations in other key MPM cancer genes (Fig. 2A). As expected, mRNA expression levels of BAP1 itself were reduced in the presence of genomic BAP1 inactivation or loss (Fig. 2B). We also observed an inverse correlation of BAP1 alterations and SCNAs. BAP1-altered tumors had fewer chromosome arm gains and losses (Fig. 2C, median 9.5 vs. 15.5, Mann–Whitney, $P < 0.01$), as well as fewer focal SCNAs (Supplementary Fig. S2).

Because BAP1-mediated deubiquitination of histones and transcriptional proteins is thought to regulate gene expression, we compared mRNA expression patterns between tumors that were wild-type for BAP1 and those with inactivation of one or both BAP1 alleles. We identified 1,324 differentially expressed genes with an FDR of $<0.01$, of which 75% were downregulated in the BAP1-inactivated samples. As previously noted in experimental models (22), BAP1-inactivated tumors in our cohort had lower mRNA expression of several HOXA genes, including HOXAS and HOXAX6, but no difference in EZH2 mRNA expression or its PRC2 partners EED and SUZ12.

As BAP1 is known to regulate the ubiquitination and hence stability of several transcriptional proteins, we examined the association of BAP1 status with inferred activity of transcription factors (TF) using a recently developed computational strategy that integrates phosphoproteomic and transcriptomic data with predicted TF binding sites (23). We used this algorithm to assess significant differences in inferred TF activities between BAP1-inactivated and wild-type MPM (satisfying FDR-corrected $P < 0.01$, $t$ test), linking BAP1 status to altered activity of TFs. Indeed, many TFs identified in this analysis had highly significant associations with BAP1 status (Fig. 2D and E; Supplementary Table S2B). In particular, YY1 had significantly reduced inferred activity in BAP1-inactivated samples (Fig. 2F). BAP1 forms a
**A**

In-frame indel
Frameshift
Missense
Nonsense
Splice site
Other
Inactivated
BAP1 exclusive
BAP1 co-occurrence
* FDR<0.25

**B**

\[ P = 6.9e^{-13} \]

BAP1 wild-type  BAP1 any hit

**C**

\( P < 0.01 \)

BAP1 wild-type  BAP1 2 hits

**D**

TF activity

\( 0.4 \)
\( 0.2 \)
\( 0 \)
\( -0.2 \)
\( -0.4 \)

BAP1 status

Wild-type
1 hit

Histologic type

Epithelioid
Sarcomatoid
Biphasic
NOS

**E**

\[-\log_{10} (FDR) \]

Effect difference (BAP1 any hit vs. wild-type)

**F**

\[ FDR = 2.847e^{-05} \]

**G**

\[ FDR = 2.6e^{-07} \]
Integrative Molecular Characterization of Mesothelioma

Figure 2. Supervised comparisons between BAP1-inactivated and wild-type MPM. A, BAPI inactivation by copy-number loss or mutation across the cohort, along with mutations in 6 other genes (left). These genes were selected as those with more than 3 nonsilent mutated tumors. Large genes with more than 3 kb coding regions (TTN, FAT4, and MGA) are unlikely to be functional in cancer and were excluded. The bar plot (right) shows the Fisher exact test P values for mutual exclusivity and co-occurrence relative to BAPI. Only SETD2 and NF2 approach significant co-occurrence with BAPI inactivation. The one-tail Fisher P values for co-occurrence and the Benjamini-Hochberg FDRs are P = 0.04, FDR = 0.15 for SETD2 and P = 0.05, FDR = 0.15 for NF2. We detected MALAT1 (aka NEAT2) RNAs in 4 BAPI-inactivated samples, but this does not reach significance (P = 0.05, FDR = 0.27). B, Normalized BAPI mRNA expression levels in the wild-type, 1 hit and 2 hit subgroups. C, Box plot demonstrating a significantly lower frequency of arm-level losses in BAPI-inactivated tumors (BAPI 2 hits) compared with wild-type (BAPI 0). D, Inferred transcription factor (TF) activities significantly associated with BAPI inactivation (FDR < 0.01). E, Volcano plot with mean inferred TF activity difference in BAPI-inactivated and BAPI wild-type patients plotted on the x-axis, and FDR-adjusted significance from t test plotted on the y-axis (−log10 scale). TFs significantly associated with BAPI inactivation status (FDR < 0.01) are colored in orange. F and G, Box plots with differential inferred activities of YY1 (F) and IRF8 (G), two biologically relevant transcription factors. The target genes on which inferences for YY1 (427 genes) and IRF8 (248 genes) activity were based are provided in Supplementary Table S2.
A

iCluster 1 iCluster 2 iCluster 3 iCluster 4

Chromosome

TCGA-SC-A6LP TCGA-UD-AAC1
TCGA-MQ-A6BR

B

TCGA-MQ-A6BR (TCGA case 1)

TCGA-SC-A6LP (TCGA case 2)

TCGA-UD-AAC1 (TCGA case 3)

C

ICGC-K2F2-A45 (ICGC case 1)

ICGC-K2F2-H60 (ICGC case 2)

D

SETDB1 Heterozygous, 1 copy
ACOT12: ABLIM3 Heterozygous, 2 copies
RELN Heterozygous, 1 copy
TP53 Homozygous, 2 copies
NF2 Homozygous, 2 copies
and genome duplication (Fig. 3D). Overall, these data suggest a model in which TP53 mutations occur early, and presumably permit the steady (or catastrophic) loss of chromosomes. It seems likely that genome reduplication occurs after achieving near-haploidy, with haplosufficiency of SETDB1 arising later.

**Integrative Multiplatform Analysis Defines Novel Prognostic Subsets**

Although the current classification of MPM into epithelioid, sarcomatoid, and biphasic histologies is prognostically useful, there remains variability in clinical features and patient outcomes within histologic subtypes. Previous analyses (5, 7) based on mRNA expression alone have defined unsupervised clusters largely recapitulating these histologic distinctions. To find out whether multiplatform molecular profiling might provide additional resolution to define prognostic subsets of MPM, we performed integrative clustering across multiple assay platforms using two algorithms: iCluster (32) and PARADIGM (33). Both identified four distinct integrated subtypes of MPM. There was a strong concordance in subtype assignments between the two algorithms (Fig. 4A; Supplementary Figs. S6 and S7), especially for the best (cluster 1) and worst (cluster 4) prognosis clusters, indicating that integration of molecular data can identify distinct subgroups of MPM, independent of the specific statistical methodological survival was significantly different across the 4 clusters ($P < 0.0001$ for either algorithm; Fig. 4B; Supplementary Fig. S7). This survival difference remained significant ($P = 0.008$) after adjusting for histology (epithelioid vs. nonepithelioid) and CDKN2A homozygous deletion (Fig. 4C; Supplementary Fig. S6), a known molecular prognostic factor in MPM (14, 15). Although they did provide additional independent prognostic information, the iClusters nonetheless did show correlation with consensus histology ($P = 0.002$), with iCluster cluster 1 being enriched for epithelioid histology (similar finding for PARADIGM cluster 1). They also exhibited differences in immune cell infiltrates (Supplementary Table S4A), as discussed below, but showed no significant correlation with clinical variables such as T stage, N stage, asbestos exposure history, or smoking (Supplementary Table S4B). Moleculary, these tumors had low SCNA, relatively few CDKN2A homozygous deletions (11%), and a high level of methylation (Supplementary Fig. S4). All but one (95%) had BAP1 alterations: 26% had homozygous deletions and 53% had heterozygous loss with mutations.

The poor prognosis cluster (cluster 4; red) had a high score for epithelial-mesenchymal transition (EMT) based on gene expression ($P < 0.001$; Fig. 4D), which was distinguished by high mRNA expression of VIM, PECAM1, and TGFBI1, and low miR-200 family expression. These tumors also displayed MSLN promoter methylation and consequent low mRNA expression of mesothelin, a marker of differentiated mesothelial cells, as noted previously in sarcomatoid MPM and the sarcomatoid components of biphasic MPM (15, 34). Overall, this poor prognosis cluster also showed enrichment of LATS2 mutations (30% compared with 4% in the rest of the cohort) and CDKN2A homozygous deletions (66%). Moreover, this cluster showed higher AURKA mRNA expression, higher leukocyte fraction (based on DNA methylation), and elevated mRNA expression of EZF targets, G1-M checkpoints, and DNA damage response genes. PI3K-mTOR and RAS-MAPK signaling were upregulated, based on both mRNA and protein expression (Supplementary Fig. S7). Additionally, several miRNAs were differentially expressed between the good and poor prognostic clusters, including miR-193a-3p, which has been proposed as a potential tumor suppressor (35). Finally, a comparison of immune gene mRNA expression signatures (36) across the four clusters revealed a significantly higher score for the Th2 cell signature in the poor prognosis cluster 4 compared with the other clusters (Fig. 4E; Supplementary Table S4A). Coincidentally, it has been reported that Th2 cytokines secreted by immune cells upon exposure to asbestos may promote MPM (37). The analyses of other immune signatures are shown in Supplementary Fig. S5.

Although biphasic and sarcomatoid MPM are more aggressive, there remains a need for improved risk stratification of epithelioid MPM, for which clinical outcomes are more heterogeneous. Therefore, we conducted an integrative clustering analysis restricted to epithelioid MPM. The results for the 4-cluster epithelioid-only solution were highly similar to the 4-cluster all-MPM solution (Fig. 4A), with only 7 of the 52 epithelioid samples reassigned to other clusters. This stability indicates that the features driving the all-MPM clustering are largely independent of histology. The epithelioid-only clusters share many of the features defining the corresponding clusters in the all-MPM solution (Fig. 5B). The survival analysis also paralleled the all-MPM solution, with cluster 1 having the best outcomes and cluster 4 having the worst (Fig. 5C). PARADIGM analysis of the epithelioid-only subset confirmed upregulation of AURKA mRNA expression in the poor-prognosis epithelioid-only cluster 4 (Fig. 5D), corroborating the results from the all-MPM analysis.

Finally, we sought to independently validate the clinical correlations of clusters identified in the TCGA epithelioid cases using mRNA expression profiles from two published studies: 211 MPM analyzed by RNA sequencing (5) and 52 MPM samples analyzed by mRNA expression microarrays (14). Specifically, we assigned each mRNA expression profile to one of the integrative clusters based on the rules derived from the TCGA mRNA data set. For the larger validation cohort (hereafter referred to as Bueno), we restricted our analysis to epithelioid samples and used the epithelioid-only gene signature to cluster samples. We found that the epithelioid-only samples assigned...
Figure 4. Integrative analysis of 74 MPM. A, Concordance between integrative (PARADIGM and iCluster) and platform-specific unsupervised clustering results. Clusters are color-coded and ranked based on survival (dark blue indicates best survival, and red and orange mark the worst surviving subgroup). CNV, copy-number variation. B, Kaplan–Meier plot of the integrative subgroups reveals distinct outcomes. C, Cox regression analysis demonstrates significant associations of the molecular subtypes with patient survival, even upon adjusting for histology, age, and CDKN2A status. D, iCluster identified 4 integrative subgroups with distinct BAP1 alteration (defined as mutation and/or copy-number alteration), TP53 mutation, CDKN2A status, copy-number alteration, DNA methylation, and mRNA, lncRNA, and miRNA expression profiles. E, Comparison of Th2 cell immune gene mRNA expression signature across the four integrated clusters.
**Figure 5.** Integrative analysis of epithelioid MPM. **A**, Comparison of cluster assignments between the epithelioid-only and full cohort integrative analyses demonstrating good concordance, with only 7 cases being reassigned to another cluster. **B**, Integrative clustering analysis applied to cases with epithelioid histology. **C**, Kaplan-Meier plot of the epithelioid-only integrative subgroups. **D**, PATHMARK analysis revealed differentially active molecular pathways that define the poor-prognosis epithelioid-only subgroup. **E**, Validation of the TCGA epithelioid subtypes in an independent cohort of 141 epithelioid MPM (Bueno et al.; ref. 5) confirming the protective effect of molecular features that define iCluster 1.
to iCluster 1 (good prognosis) had significantly better survival, even after adjusting for age \( (P = 3.9 \times 10^{-4}; \text{Fig. 5E; Supplementary Fig. S6}) \). In the smaller cohort (referred to as Lopez-Rios), patient numbers were too small to split by histology. However, this analysis provided independent validation of the survival differences for the four all-MPM clusters \( (P = 0.01; \text{Supplementary Fig. S6}) \). Taken together, these results suggest that the prognostically relevant molecular profiles defined by our analysis are robust and reproducible, and could potentially be used to improve risk stratification of patients with epithelioid MPM. The core mRNA gene lists are provided in Supplementary Table S4C–S4G, which also include reduced classifier combining mRNA and methylation data to facilitate practical application of these data and independent validation of these clusters.

### Analysis of Noncoding RNAs in MPM

We next assessed two types of noncoding RNAs not extensively studied in MPM that may also represent a source of robust biomarkers: lncRNAs, which often show higher expression specificity for cell type than coding genes, and miRNAs, which are relatively stable in biological fluids, so potentially suitable as noninvasive biomarkers.

In both the TCGA and Bueno (5) cohorts, lncRNAs returned four unsupervised consensus subtypes that were associated with 5-year survival \( (P = 1.4 \times 10^{-10} \text{ and } 1.1 \times 10^{-10}) \), respectively. The TCGA lncRNA subtypes were highly concordant with iCluster \( (P = 8 \times 10^{-14}) \text{ and PARADIGM} \( (P = 2 \times 10^{-13}) \text{ subtypes and were associated with EMT scores} \( (P = 2.5 \times 10^{-6}; \text{Figs. 4A, 6A–E; Supplementary Table S5A.1}) \). For both cohorts, lncRNAs that were differentially abundant between the good-prognosis subtype and other samples included those associated with cancer (e.g., \( H19, LINC00152, \text{and} MEG3 \)), or with MPM in particular: \( NEAT1 \text{ and} SNHG8 \) (38) and \( GASS \) (ref. 39; Fig. 6F and G; Supplementary Table SSB.3–6). We noted that a number of lncRNAs distinguished the good-prognosis cluster in both TCGA and Bueno cohorts (Supplementary Table SSB.1.2–5.6).

Unsupervised clustering based on miRNA mature strands resolved five consensus subtypes in the TCGA cohort (Fig. 6H); these were associated with 5-year survival \( (P = 7 \times 10^{-3}) \), iCluster \( (P = 1 \times 10^{-12}) \) and PARADIGM \( (P = 3 \times 10^{-12}) \text{ clusters, and EMT scores} \( (P = 2 \times 10^{-7}; \text{Fig. 6H–K; Supplementary Table S5A.2}) \). For the good-survival cases, the miRNA subtypes were strikingly concordant across multiple analysis platforms, and many cancer-associated miRNAs were differentially abundant in the good-survival cluster (Fig. 6L; Supplementary Table SSB.3–4). Taken together, these results suggest that lncRNAs and miRNAs may be important predictors of survival in MPM.

### EMT and VISTA Expression

Because mRNA expression of EMT-associated genes was a key differentiating feature between prognostically distinct integrative clusters, we performed a detailed analysis of EMT in MPM using a previously published EMT-related mRNA signature (40). Increasing EMT scores significantly correlated with clusters defined by integrative algorithms (iCluster and PARADIGM), as well as with individual genomic features including miRNA, lncRNA, methylation, RPPA, and overall gene expression (Fig. 7A). Of all tumor types included in the TCGA pan-cancer analysis, MPM had the second-highest average EMT score (Fig. 7B; after soft-tissue sarcomas), consistent with previous reports that EMT is a frequent phenomenon in MPM (7, 41).

Although EMT score positively correlated with the mRNA expression of many immune-regulatory genes, such as \( OX40L, TGFB1, CD276, OX40, \text{and} PD-L2 \) \( (P < 0.001; \text{Fig. 7A and C}) \), mRNA expression of VISTA (42), a negative immune-checkpoint regulator primarily expressed on hematopoietic cells (43), was strongly inversely correlated with EMT score (Fig. 7C) and was expressed at levels higher than in any other TCGA tumor type analyzed (Fig. 7D). In the MPM cohort, VISTA mRNA levels were highest in the epithelioid subtype (Fig. 7C and E). Using Regulome Explorer (Supplementary Section 13), we found that VISTA mRNA expression was highly correlated with \( MSLN \) mRNA expression (Spearman correlation \( = 0.81; P = 6.3 \times 10^{-10}) \), but not with mRNA expression of PD-1 or PD-L1. Moreover, there was no significant correlation between overall mutation burden and VISTA expression levels \( (P = 0.64) \).

VISTA (V-domain Ig suppressor of T-cell activation; also known as c10orf54, PD-1H, and B7-H5) is a member of the B7 family of negative checkpoint regulators, expressed on the surface of several immune cell types. It can function as both receptor and ligand (44). As a ligand, VISTA is present on the surface of antigen-presenting cells (APC) and inhibits early-stage T-cell activation (45). The normal mesothelium has APC properties (46), which are retained upon malignant transformation (47, 48). We thus performed IHC staining of two epithelioid TCGA MPM cases, as well as normal and reactive mesothelium (Supplementary Section 14), to define the cellular compartment expressing VISTA in MPM tumor samples. Remarkably, VISTA protein expression was not restricted to infiltrating immune cells, but was present in tumor cells in MPM, as well as in normal and reactive mesothelium (Fig. 7F and G), suggesting that its expression in epithelioid MPM may reflect retention of APC properties in this more differentiated subset of MPM.

### DISCUSSION

Our comprehensive integrative analysis of 74 cases of MPM further defines a cancer driven primarily by inactivation of tumor suppressor genes. Indeed, we confirm the high frequency of \( BAP1 \) inactivation by mutation and copy-number loss, as well as recurrent inactivating alterations in \( CDKN2A, NF2, TP53, LAT52, \text{and} SETD2 \). In addition to this landscape of known loss-of-function events, we have genomicsally characterized a novel molecular subtype of MPM accounting for approximately 3% of MPM in our data sets, defined by evidence of genomic near-haploidization and recurrent \( TP53 \text{ and} SETD2B1 \) mutations, with a distinctive clinical phenotype showing female predominance and younger age at diagnosis. Our findings should facilitate systematic clinical studies of this subset to better define its survival and its association with asbestos exposure, which so far appears weak or unclear. Isolated cases with similar molecular profiles (GNH and \( SETD2B1 \) mutation) have been anecdotally reported (5, 49),...
Integrative Molecular Characterization of Mesothelioma

Figure 6. Noncoding RNA subtypes and differential abundance for IncRNAs and miRNAs in the TCGA and Bueno cohorts. A–G, IncRNA subtypes and differential abundance. A, Top to bottom: Normalized abundance heat map for a 4-subtype solution, then a silhouette width profile (Wcm) calculated from the consensus membership matrix; clinical and molecular covariates, with P values from Fisher exact, χ², or Kruskal tests; and profiles of RNA–sequencing–based EMT scores and leukocyte fraction. Red horizontal line in leukocyte fraction bar plot indicates the median value across all samples. Distribution of purity estimated by ABSOLUTE, with a Kruskal P value. C, Distribution of RNA–sequencing–based EMT scores, with a Kruskal test for overall survival, with a log-rank P value. D, Kaplan–Meier plot for overall survival, with a log-rank P value for a 4-subtype solution for the Bueno cohort. F and G, IncRNAs that were differentially abundant (SAM 2-class unpaired analysis, FDR < 0.05) between the better-survival IncRNA subtype and all other samples, for the TCGA cohort (F) and the Bueno cohort (G). The largest 15 positive and 15 negative fold changes are shown, blue triangles mark IncRNAs that were in these gene sets in results for both cohorts. Text to the right of each bar plot gives means-based fold changes, mean abundance in the target and then the other samples, and the cytoband for the gene. See also Supplementary Table S5A. (continued on next page)
Figure 6. (Continued) H-L, microRNA mature strand subtypes and differential abundance in the TCGA cohort. H, Top to bottom: Normalized abundance heat map for a 5-subtype solution, then a silhouette width profile (Wcm) calculated from the consensus membership matrix; clinical and molecular covariates, with P values from Fisher exact, $\chi^2$, or Kruskal tests; and profiles of RNA-sequencing–based EMT scores and leukocyte fraction. Red horizontal line in leukocyte fraction bar plot indicates the median value across all samples. I, with heat map for a 5-subtype solution, then a silhouette width profile (Wcm) calculated from the consensus membership matrix; clinical and molecular covariates, cytoband(s) for the mature strand. See also Supplementary Table S5B.
but had not, until now, been recognized as a distinct molecular subtype of MPM. The genomic data in our cases support a model in which early TP53 inactivation, as H3K9 methylation, reduplication and of chromosomes to a near-haploid state, followed by genome mutations are permissive for a loss of TP53. The genomic data in our cases support a model in which early TP53 inactivation, as H3K9 methylation, reduplication and of chromosomes to a near-haploid state, followed by genome mutations are permissive for a loss of TP53.

Figure 7. MPM is enriched for both EMT and mRNA expression of immune targets. A, Unsupervised analysis identifies correlations between EMT and multiple platforms. Tumors are ordered from left to right according to increasing EMT score. Numbered color bars indicate group assignments (clusters) from other data types. Statistically significant correlations are shown between EMT score and (starting at top) integrative multidimensional analyses on both iCluster and PARADIGM platforms, along with mRNA, miRNA and lncRNA clusters, methylation status, and consensus histology. Bottom plots illustrate significant correlations between these clusters and selected miRNAs, proteins, immune target genes. B, Spectra of EMT scores across different tumor types. Mesothelioma is the second most mesenchymal cancer type after sarcoma in 31 tumor types analyzed. Despite most MPM tumors having undergone EMT, a broad range of EMT scores were observed across mesothelioma cases, which corresponded to a large extent with histologic subtype. C, Waterfall plot illustrating the correlation between EMT and immune target genes. D, Plot highlighting VISTA mRNA expression, which is highest in MPM, across all TCGA tumor types. E, Box plot indicating VISTA mRNA expression levels in individual histologic types of the TCGA MPM cohort. The highest mRNA expression levels were observed in the epithelioid subtype (Wilcoxon rank-sum, \( P = 2e^{-7} \)), whereas sarcomatoid MPM had the lowest mRNA expression (\( P = 0.017 \)). Red arrows indicate two epithelioid cases that were examined by immunohistochemistry, TCGA-SC-A6LQ-01 (1) and TCGA-SC-A6LM-01 (2). (continued on next page)
expressed. Why chromosomes 5 and 7 are spared from LOH in these MPM remains mysterious—interestingly, the same two chromosomes have recently been shown to also retain heterozygosity in thyroid Hurthle cell carcinomas with evidence of genomic near-haploidization (51). Overall, the identification of this novel subset of MPM highlights how, as the proportion of asbestos-related MPM plateaus and hopefully begins declining in Western countries, cases of possibly different etiology may become more apparent.

A better understanding of the determinants of aggressive behavior and predictors of poor clinical outcomes in MPM remains an unmet clinical need. To this end, integrative analyses with iCluster and PARADIGM revealed a set of molecular features that define MPM subsets with better and worse prognosis, and might point to candidate therapeutic targets. For instance, cases in the poor-prognosis subset showed higher AURKA mRNA expression, consistent with prior studies (14, 52). Treatment of MPM cell lines with an Aurora kinase inhibitor leads to growth arrest (53), and several Aurora kinase inhibitors are under investigation in patients with MPM. AURKA mRNA expression could be used to help identify patients with poor-prognosis epithelioid MPM so that they could be directed toward experimental therapies early in their treatment course. It is presently unclear whether Aurora kinase inhibitors will be active in MPM, and the ongoing phase II study of alisertib (NCT02293005) includes all patients irrespective of molecular signature. Although the value of targeting the AURKA pathway with currently available clinical compounds in MPM remains unknown, mRNA expression of AURKA is a prognostic marker that could be used in clinical practice to help stratify patients with epithelioid MPM.

Additionally, the poor-prognosis group also exhibited upregulation of the PI3K and mTOR signaling pathways. Preclinical studies have reported this finding, leading to clinical trials with a low response rate (2%) and no significant survival benefit in the salvage setting (54). These data suggest that combination therapies with mTOR inhibition may be necessary. Our analysis of epigenetic alterations revealed some associations between BAP1 status and DNA methylation (Supplementary Table S6). A recently established functional link between BAP1 and EZH2 (22) provided the rationale for the recently completed clinical trial of the EZH2 inhibitor tazemetostat in BAP1-null MPM (NCT028602). By contrast, no relevant findings resulted from our analysis of viral and microbial sequences (Supplementary Table S7). Finally, another genomically driven potential vulnerability in MPM is very frequent CDKN2A deletion associated with codeletion of MTAP, the latter recently shown to metabolically lead to impaired activity, and therefore sensitization to further inhibition, of the arginine methyltransferase PRMT5 (12, 13), an inhibitor of which is currently being evaluated in a phase I trial (NCT02783300).

Harnessing current immunotherapy approaches to improve outcomes of patients with MPM is an area of intense clinical interest. Although our study confirms that MPMs show a low tumor mutation burden and therefore may present a more challenging setting for immunotherapy, a remarkable and novel finding of the present study is that of strong expression of the immune-checkpoint gene VISTA in epithelioid MPM, on the tumor cells themselves, unlike other cancer types where it is more often expressed on infiltrating reactive cells (42). VISTA is a member of the B7 family of negative checkpoint regulators that is expressed primarily on immune cells (43, 55). However, unique structural features mean that VISTA can repress activation of T cells both as a ligand present on the surface of APC cells and as a receptor on the surface of T cells (56). Because VISTA is expressed on MPM cells, and its mRNA expression levels do not correlate with overall mutation

Figure 7. (Continued) F, IHC staining for VISTA (rabbit monoclonal anti-VISTA antibody, clone D1L2G, 0.1 μg/mL, Cell Signaling Technology) in normal mesothelial lining from pleura (*) and benign pleuritis with reactive mesothelial proliferation (**), and 2 TCGA MPM cases, TCGA-SC-6AELQ-01 (1) and TCGA-SC-6AELM-01 (2); images captured at 100× magnification. These results confirm high protein expression of VISTA on tumor cells in epithelioid MPM. G, VISTA immunohistochemistry. VISTA protein is expressed both in tumor cells (red arrows) and in infiltrating inflammatory cells (black arrows) in the epithelioid MPM case TCGA-SC-6AELQ-01. Image captured at 200× magnification.
load, our results raise the possibility that VISTA expression may be restraining antitumor immune responses in a subset of MPM cases. As we find that nonneoplastic mesothelium also expresses VISTA protein, we speculate that VISTA expression in MPM is retained or possibly selected for by immune pressure. This is consistent with previous publications demonstrating that MPM is an “immunogenic” tumor (57), including recent trials showing some responses to immune-checkpoint blockade therapy (58, 59). Indeed, VISTA has recently been reported as a possible compensatory immune-inhibitory pathway in prostate cancers that fail to respond to ipilimumab (60). Our findings thus provide both a rationale and a candidate biomarker for clinical trials of emerging anti-VISTA therapy (refs. 42, 44; NCT02812875) in epithelioid MPM. Taken together, our findings point to new lines of investigation into the biology of MPM with the potential to lead to new therapeutic strategies.

METHODS

TCGA Project Management has collected necessary human subjects documentation to ensure the project complies with 45-CFR-46 (the “Common Rule”). The program has obtained documentation from every contributing clinical site to verify that Institutional Review Board (IRB) approval has been obtained to participate in TCGA. Such documented approval may include one or more of the following:

- An IRB-approved protocol with informed consent specific to TCGA or a substantially similar program. In the latter case, if the protocol was not TCGA-specific, the clinical site principal investigator provided a further finding from the IRB that the already-approved protocol is sufficient to participate in TCGA.
- A TCGA-specific IRB waiver has been granted.
- A TCGA-specific letter that the IRB considers one of the exemptions cited were that the research falls under 46.102(f)(2) or 46.101(b)(4). Both exempt requirements for informed consent, because the received data and material do not contain directly identifiable private information.
- A TCGA-specific letter that the IRB does not consider the use of these data and materials to be human subjects research. This was most common for collections in which the donors were deceased.

A detailed description of the sample acquisition and pathology review process, as well as the experimental and computational methods used for the different analyses presented in our study, is provided as Supplementary Sections 1–13.

Disclosure of Potential Conflicts of Interest

L.R. Chirieac has received remuneration for medicolegal work related to mesothelioma. V.W. Rusch reports receiving a commercial research grant from Genentech and is a consultant/advisory board member for BMS. H. Pass reports receiving a commercial research grant from Genentech. M.G. Zauderer is a member of the Board of Directors of the Mesothelioma Applied Research Foundation, reports receiving commercial research grants from Bristol-Meyers Squibb and Millennium, and is a consultant/advisory board member for AstraZeneca, Sellas Life Science, and Epizyme. D.J. Kwiatkowski is a consultant/advisory board member for Novartis. R. Bueno has ownership interest (including stock, patents, etc.) in Navigation Sciences, A.D. Cherniack reports receiving commercial research support from Bayer. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Co-chairs: Marc Ladanyi, Peter J. Campbell, Bruce W. S. Robinson

Manuscript Coordinator: Julija Hmeljak

Analysis Coordinator: Ina Felau

Data Coordinator: Francisco Sanchez-Vega

Project Manager: Ina Felau

Copy-number: Andrew D. Cherniack, Juliann Shih

mRNA: Katherine A. Hoadley

Methylation: Linda Danilova

miRNA: Reanne Bowlby, Gordon Robertson, Ewan Gibb

Exomes: Manaksi Gupta, Chip Stewart, Patrick Tarpey

RPPA: Rupa Kanchi, Rehan Akbani

Transcription factor analysis: Hatice U. Osmanbeyoglu

Integrative analyses:

PARADIGM: Yulia Newton, Kiley Graim, Firehouse: David Heiman, Michael Noble

icluster: Esther Drill, Ronglai Shen; David T. Severson, Raphael Bueno

Regulome explorer: Lisa lype, Vesteeni Thorsson, David L. Gibbs

EMT analysis: Lauren A. Byers, Carl Gay, Lixia Diao


Pathology Subgroup: William D. Travis, Sanja Dacic, Lucian Chirieac, Françoise Galateau-Sallé, Junya Fujimoto, Ahya Hussain, Marc Ladanyi

VISTA IHC: Patrice Desmeules, Achim Jungbluth


Genomic near-haploidization validation datasets: Gloria Ravegnini, Jonathan Fletcher, Hiroiway Aburatani


Acknowledgments

We thank Drs. Elaine Mardis and Matthew Meyerson for invaluable feedback and Dr. Lee Spraggon for helpful comments and assistance with figures. This work was supported by the following grants from the USA National Institutes of Health: U54 HG003273, U54 HG003067, U54 HG003079, U24 CA143799, U24 CA143835, U24 CA143840, U24 CA143843, U24 CA143845, U24 CA143848, U24 CA143858, U24 CA143866, U24 CA143867, U24 CA143882, U24 CA143883, U24 CA144025, P30 CA016672, and P30 CA008748. Whole-exome sequencing and analysis was funded by the British Lung Foundation grant APG11-7 (“The genomic architecture of mesothelioma”). Publication charges were funded by the Memorial Sloan Kettering Mesothelioma Program Fund.

Hmeljak et al.

**REFERENCES**

Integrative Molecular Characterization of Mesothelioma
