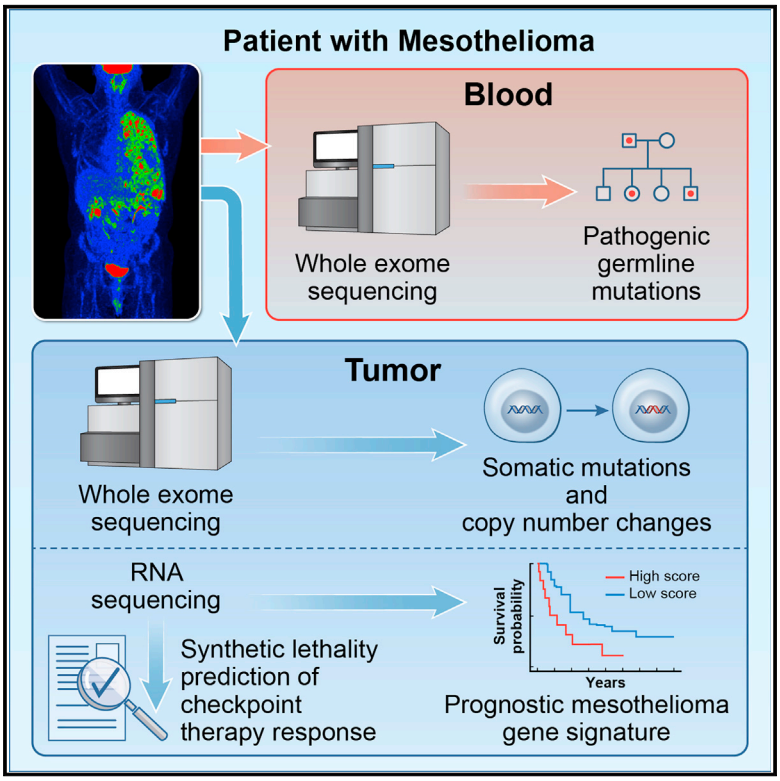
Cell Reports
Medicine

Article

Genomic and transcriptomic analyses identify a prognostic gene signature and predict response to therapy in pleural and peritoneal mesothelioma

Graphical abstract



Highlights

- A 122-patient mesothelioma cohort with genomics, transcriptomics, and phenotypic data
- 48-gene signature highly predictive of patient survival
- Evaluated tumor immune microenvironment using transcriptomics
- Transcriptomic analysis to predict response to therapy

Authors

Nishanth Ulhas Nair, Qun Jiang, Jun Stephen Wei, ..., Javed Khan, Eytan Ruppin, Raffit Hassan

Correspondence

eytan.ruppin@nih.gov (E.R.), hassanr@mail.nih.gov (R.H.)

In brief

Nair et al. present a 122-patient pleural and peritoneal mesothelioma cohort with genomics, transcriptomics, and phenotypic data. They identify a 48-gene transcriptomics-based signature that is highly predictive of mesothelioma patient survival, including the *CCNB1* gene. Transcriptomics analysis is done to study the tumor immune microenvironment and predict drug response in patients.

Nair et al., 2023, Cell Reports Medicine *4*, 100938
February 21, 2023
https://doi.org/10.1016/j.xcrm.2023.100938

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Genomic and transcriptomic analyses identify a prognostic gene signature and predict response to therapy in pleural and peritoneal mesothelioma

Nishanth Ulhas Nair,[1,7] Qun Jiang,[2,7] Jun Stephen Wei,[3,7] Vikram Alexander Misra,[2] Betsy Morrow,[2] Chimene Kesserwan,[3] Leandro C. Hermida,[1,4] Joo Sang Lee,[1,5] Idrees Mian,[2] Jingli Zhang,[2] Alexandra Lebensohn,[3] Markku Miettinen,[6] Manjistha Sengupta,[2] Javed Khan,[3] Eytan Ruppin,[1,*] and Raffit Hassan[2,8,*]

[1]Cancer Data Science Laboratory, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD 20892, USA

[2]Thoracic and GI Malignancies Branch, CCR, NCI, NIH, Bethesda, MD 20892, USA

[3]Genetics Branch, CCR, NCI, NIH, Bethesda, MD 20892, USA

[4]Tumor Microenvironment Center, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA

[5]School of Medicine and Department of Artificial Intelligence, Sungkyunkwan University, Suwon 16419, Republic of Korea

[6]Laboratory of Pathology, CCR, NCI, NIH, Bethesda, MD 20892, USA

[7]These authors contributed equally

[8]Lead contact

*Correspondence: eytan.ruppin@nih.gov (E.R.), hassanr@mail.nih.gov (R.H.)

https://doi.org/10.1016/j.xcrm.2023.100938

SUMMARY

Malignant mesothelioma is an aggressive cancer with limited treatment options and poor prognosis. A better understanding of mesothelioma genomics and transcriptomics could advance therapies. Here, we present a mesothelioma cohort of 122 patients along with their germline and tumor whole-exome and tumor RNA sequencing data as well as phenotypic and drug response information. We identify a 48-gene prognostic signature that is highly predictive of mesothelioma patient survival, including *CCNB1*, the expression of which is highly predictive of patient survival on its own. In addition, we analyze the transcriptomics data to study the tumor immune microenvironment and identify synthetic-lethality-based signatures predictive of response to therapy. This germline and somatic whole-exome sequencing as well as transcriptomics data from the same patient are a valuable resource to address important biological questions, including prognostic biomarkers and determinants of treatment response in mesothelioma.

INTRODUCTION

Malignant mesothelioma is an aggressive cancer arising from the mesothelial cell linings of the pleura, peritoneum, pericardium, or tunica vaginalis with an annual incidence of 3,300 new cases in the United States.[1] Malignant pleural mesothelioma (MPM) comprises 80% of the cases, while malignant peritoneal mesothelioma (MPeM) includes the remaining 15%–20%.[2] Pericardial and tunica vaginalis mesothelioma are very rare. Mesothelioma is a difficult-to-treat cancer with a median overall survival of approximately 12 months with a treatment regimen of pemetrexed and cisplatin.[3] Recently, the US Food and Drug Administration (FDA) has approved the combination of immune checkpoint inhibitors, ipilimumab and nivolumab, as first-line therapy for MPM, with a modest but statistically significant increase in overall survival of 18.1 months over 14.1 months in the chemotherapy group.[4] MPeM has a better prognosis than pleural mesothelioma, with median overall survival longer than 5 years, when amenable to cytoreductive surgery and hyperthermic intraperitoneal chemotherapy.[2]

Predisposing factors of mesothelioma are exposure to asbestos fibers and prior radiation therapy.[5–8] In addition, germline mutations in *BRCA1*-associated-protein 1 (*BAP1*) increase the risk of developing malignant mesothelioma as well as other common cancers.[9] The somatic mutations that have been identified in MPM include loss-of-function mutations in the tumor suppressor gene *BAP1* as well as the epigenetic regulatory genes *DDX3X* and *SETD2*.[10] In addition, recurring deletions of chromosomal 3p21 (target gene *BAP1*), 9p21 (*CDKN2A*), and 22q12 (*NF2*) have been identified in malignant mesothelioma.[11]

Because current therapies for mesothelioma lead to a good response only in a small set of patients,[12–22] it is important to identify predictive biomarkers for patient response. Several prior studies with small number of patients have evaluated predictors of response to therapy and patient survival. These include MesoNet, a deep convolutional network approach, which uses whole-slide digitized images to predict the overall survival of mesothelioma patients.[23] Using neural networks, a gene-expression-based classifier was obtained from a small dataset of 21 MPM patients to

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1. Mesothelioma patient characteristics

Patient characteristics	n (%) 122 (100)
Gender	
Male	69 (57)
Female	53 (43)
Age at diagnosis, years	
<60	76 (62)
≥60	46 (38)
Asbestos exposure[a]	
Yes	67 (54)
No	33 (27)
Unknown	22 (19)
Mesothelioma site	
Pleura	59 (48)
Peritoneum	61 (50)
Tunica vaginalis	2 (2)
Radiation prior to mesothelioma diagnosis	
Yes	9 (8)
No	113 (92)
Prior mesothelioma treatment	
None	55 (45.1)
Platinum based chemotherapy	54 (44.3)
Chemotherapy plus immunotherapy	12 (9.8)
Other	1 (0.8)

[a]Asbestos exposure by self-report.

predict survival.[24] Another study developed a gene expression ratio-based predictor derived from 17 mesothelioma tumors for determining treatment outcome.[25] Additionally, overexpression of aurora kinases A and B was observed to be associated with more aggressive mesotheliomas.[26] More recently, loss of the tumor suppressor gene *BAP1* has been proposed as a candidate biomarker for immunotherapy treatment in mesothelioma.[27]

A deeper knowledge of the genetic, transcriptomic, and immunogenic events involved in malignant mesothelioma is critical for successful development of prognostic biomarkers and personalized therapeutic modalities. In this study, we have three main aims. Our first goal is to present a mesothelioma dataset of 122 patients with their genomic and transcriptomic profiles, as well as phenotypic and drug response information. Unlike previous large-scale studies that focused on MPM patients,[10,28] our dataset contains an approximately equal representation of MPM and MPeM patients, which allows us to identify differences between them. Our second goal is to come up with a transcriptome-based gene signature to predict mesothelioma patient survival in large-scale independent cohorts and to help identify important genes that could be potential drug target candidates. Finally, as a proof of concept, we apply a precision oncology framework that uses transcriptomic data and synthetic lethality (SL) predictions to predict drug response and suggest potential treatment options for mesothelioma patients.

RESULTS

National Cancer Institute (NCI) mesothelioma dataset and patient characteristics

We performed whole exome sequencing (WES) of paired blood and tumor tissues from 122 malignant mesothelioma patients who participated in the clinical trial at the Thoracic and Gastrointestinal (GI) Malignancies Branch, Center for Cancer Research, NCI, NIH (ClinicalTrials.gov: NCT01950572) to detect somatic and germline mutations (STAR Methods). The patient population was enriched for those who carried a previously described germline mutation in BROCA V10 panel target genes.[29] Our study cohort includes an equal proportion of patients with pleural and peritoneal mesothelioma compared with published mesothelioma cohorts, which only included patients with pleural mesothelioma.[10,28] We also performed RNA sequencing (RNA-seq) analysis on 100 tumor samples for which RNAs were available (STAR Methods). Phenotypic characteristics and drug response information are also presented. This data resource is available (Tables 1, S1, S2, S3, and S4; Data and code availability).

Clinical characteristics of patients in the NCI mesothelioma cohort are shown in Table 1. The majority of patients are male (57%) with an average age at diagnosis of 54.2 years (range, 12–80 years). A large proportion of patients was diagnosed early (62% <60 years old). Peritoneal (50%) and pleural (48%) mesothelioma cases existed in almost equal proportions. Most tumors had epithelial histology (91%); the average age at diagnosis of patients with pleural mesothelioma was higher than that of those with peritoneal mesothelioma (58.9 ± 10.9 years vs. 50.0 ± 14.4 years, $p < 0.0001$), and they were more likely to self-report asbestos exposure (61% vs. 47.5%, $p = 0.02$). The patients in our study were largely treatment naive at the time of tissue sequencing (45.1%) or had received platinum-based chemotherapy (44.3%), with a small proportion of patients who received a combination of chemotherapy and immune therapy (9.8%).

Germline and somatic mutations in mesothelioma

In the NCI mesothelioma cohort of 122 patients (n = 59 for MPM, n = 61 for MPeM, n = 2 for tunica vaginalis mesothelioma), WES analysis of germline variants was performed according to the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) guidelines for variant interpretation[30] (Figure 1; STAR Methods). A total of 43 Pathogenic (P) or likely P (LP) variants in 21 cancer predisposition genes were detected in 37 patients (Table S1). Consistent with previous studies, *BAP1* P/LP variants were the most common alterations seen in this cohort (13.1%), with a similar rate of *BAP1* P/LP in MPM (11.9%) and MPeM patients (14.8%). The percentage of *BAP1* mutation is higher in this cohort than reported previously.[29] P/LP variants in *BAP1* were found in 16 patients, accounting for 37% of the total P/LP variants seen in this cohort. Of these 16 patients with heterozygous P/LP *BAP1* variants, 3 patients were also found to be heterozygous carriers for a second P/LP variant in a gene predisposing to an autosomal recessive (AR) cancer syndrome (*ERCC2*, *SBDS*, and *XPA*), and 1 patient was heterozygous for a P variant in *SDHA*, associated

CellPress
OPEN ACCESS

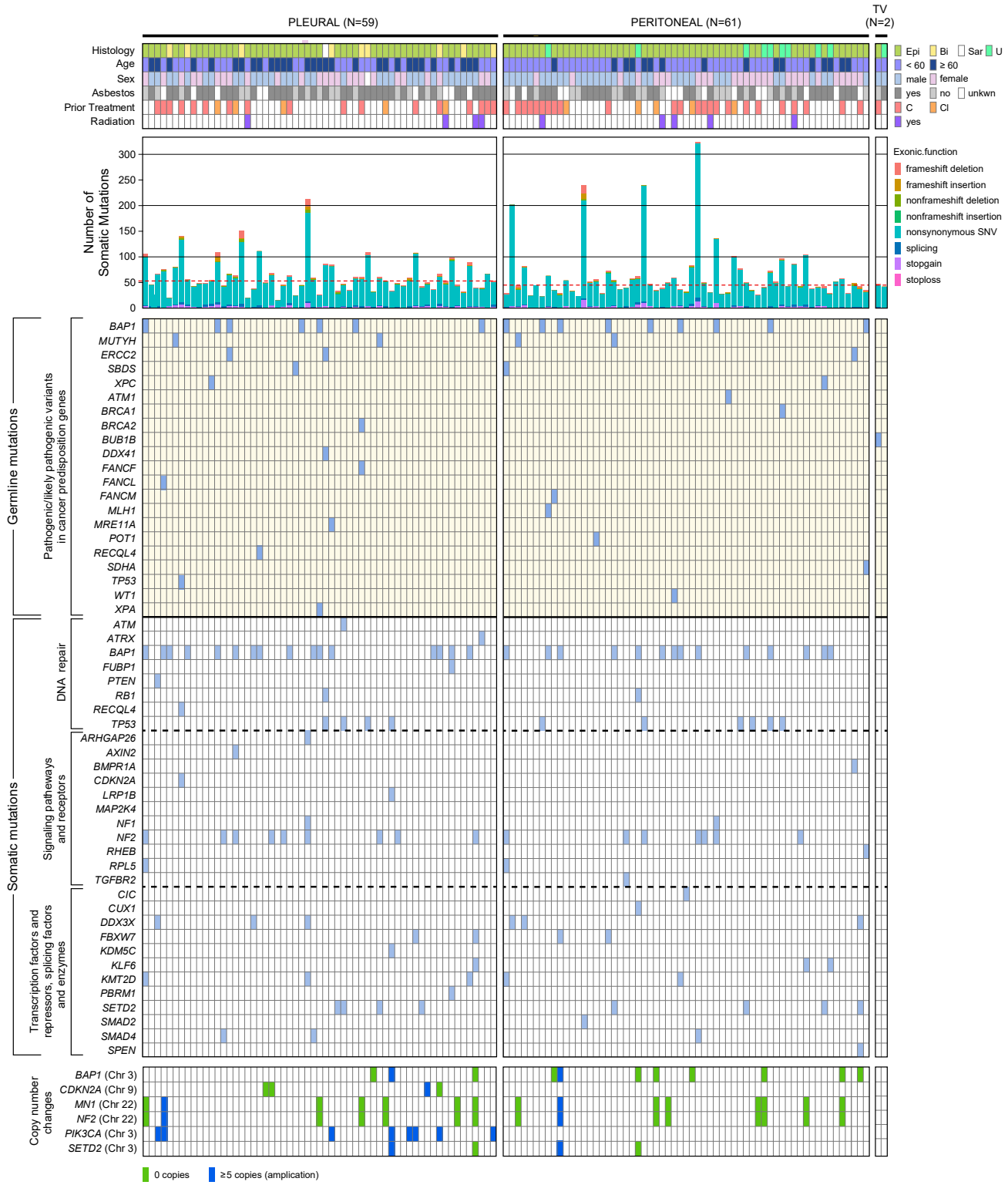


Figure 1. Overview of the NCI mesothelioma data and mutational signature analysis

Shown are demographic information, tumor histology, and mutational profiles of patients with mesothelioma. The mesothelioma cohort in this study is comprised of 122 patients, including patients with pleural (n = 59), peritoneal (n = 61), and tunica vaginalis (n = 2) forms of mesothelioma. The mutational profile consists of germline mutations in P and LP cancer predisposition genes as well as somatic mutations in genes associated with signaling pathways and receptors, DNA repair genes, TFs and repressors, splicing factors, and enzymes. Copy number changes are indicated for the relevant genes.

with an autosomal dominant (AD) predisposition to paragangliomas. Eight additional patients were positive for a P/LP variant in an AD cancer predisposition gene (*ATM*, *BRCA1*, *BRCA2*, *DDX41*, *MLH1*, *POT1*, *TP53*, and *WT1*).

For somatic mutations (also shown in Figure 1), the median number of mutation burden in a tumor is 53 mutations in MPM (a burden of 1.17 mutations/Mb, similar to what has been reported in mesothelioma),[31,32] 45 mutations in MPeM (1.00 mutations/Mb), and 44.5 mutations (0.98 mutations/Mb) in tunica vaginalis mesothelioma (STAR Methods). Insertions or deletions (indels) were more common in the pleural form than in the peritoneal form. The genes that most frequently carried somatic mutations include *BAP1* (24.59%), *NF2* (13.11%), *TP53* (8.2%), and *SETD2* (6.56%). No significant difference in the frequency of these gene mutations were observed between MPM and MPeM. We also performed Kaplan-Meier (K-M) analysis[33] to compare the overall survival times between MPM and MPeM (Figure S1A). In general, MPeM patients showed a higher overall survival than MPM patients in our cohort (log rank test, p = 0.043). Interestingly, a principal-component analysis (PCA) clustering based on the tumor gene expression data did not show any clear separation between pleural and peritoneal mesothelioma patients (Figure S1B).

We also analyzed the copy number variations (Table S2). *BAP1* copy number loss was seen in 37.3% of patients with MPM and 34.4% of patients with MPeM. Complete *BAP1* loss was seen in 9 patients. *PIK3CA* amplification with greater than or equal to 5 copies was observed in 8 of 59 MPM cases (13.56%). Although rare, *PIK3CA* amplification has been reported in 3 of 42 MPM cases in a previous study.[34] However, because *PIK3CA* amplification is commonly found in squamous cell lung carcinoma, the pathology specimens of these eight cases with *PIK3CA* amplification were re-reviewed by a pathologist (M.M.) with expertise in thoracic pathology to rule out a misdiagnosis of squamous cell lung carcinoma. In all 8 cases, the diagnosis of mesothelioma was confirmed by histologic and immunohistochemistry analysis.

Identifying a gene expression signature associated with survival in mesothelioma

We next analyzed gene expression based on RNA-seq data for 100 NCI mesothelioma patients (Table S3) and identified the genes associated with patient survival. Because some of the mesothelioma samples in the NCI mesothelioma data were acquired post treatment, we re-computed the survival of all patients from the time of biopsy sampling to the time of death or last follow-up (Table S4) and used these survival data for our gene expression and copy-number-based survival association studies.

We identified 48 genes in the NCI mesothelioma data where increased expression is associated with worse survival using Cox regression analysis after controlling for age, gender, and site of disease (the mesothelioma site could be pleural, peritoneal, or tunica vaginalis) (false discovery rate [FDR] < 0.1). We term this set of 48 genes the "mesothelioma prognostic signature" gene set, and it was used for further analysis (Table S5A). We also identified 27 genes with increased expression that were associated with better survival (Cox regression analysis after controlling for age, gender, and site of disease; FDR < 0.1; Table S5B). We did not find any genes whose copy number variation was associated with mesothelioma patient survival (after controlling for age and gender, FDR < 0.1). Gene Ontology (GO) enrichment analysis on the mesothelioma prognostic signature gene set showed a strong enrichment for GO terms related to cell cycle processes, DNA repair, chromosome organization, telomere organization, proliferation, gene silencing, and others (Figure 2A; Table S5C).

The mesothelioma prognostic signature is highly predictive of mesothelioma patient survival in other independent patient cohorts

We aimed to validate whether expression of the mesothelioma prognostic signature genes could predict patient survival in independent mesothelioma patient cohorts (RNA-seq data were processed as trimmed mean of M values [TMM] log counts per million [CPM] format; STAR Methods). We defined a combined risk score as the median expression of the 48 genes in the mesothelioma prognostic signature set and computed it for each patient. Because increased expression of these genes was associated with worse survival, we hypothesized that their risk scores would be a marker for survival in mesothelioma patients. Using Cox regression analysis in cross-validation (STAR Methods), after controlling for age, gender, and site of disease, we found, as expected, that high risk scores were indeed associated with worse survival outcome in the NCI mesothelioma data (hazard ratio [HR] = 1.76, p = 6.87e−04; Figure S2A; Table S5D). Moreover, high risk scores were also associated with worse survival outcome when we looked separately at pleural or peritoneal mesothelioma patients (Cox regression analysis in cross-validation after controlling for age and gender; pleural mesothelioma patients only: HR = 1.84, p = 0.046; peritoneal mesothelioma patients only: HR = 2.1, p = 0.0021). A K-M survival analysis also showed similar results (log rank test, p = 0.00029; Figure 2B).

To test the predictive value of the mesothelioma prognostic signature, we validated it in two large, independent mesothelioma datasets: the The Cancer Genome Atlas (TCGA) mesothelioma dataset of 85 pleural mesothelioma patients[28] and the Bueno et al.[10] dataset of 211 pleural mesothelioma patients. We computed the risk scores using the mesothelioma prognostic signature for each patient in these datasets and predicted overall survival. Cox regression analysis controlling for age and gender showed that increased risk scores were associated with worse patient survival in TCGA mesothelioma (HR = 2.6, p = 6.94e−10; Figure S2A; concordance statistic [C] obtained using the Cox regression model = 0.772) and Bueno et al.[10] dataset (HR = 1.49, p = 4.34e−07; Figure S2A; C = 0.643). The survival prediction obtained using the mesothelioma prognostic signature (C = 0.772; Cox regression using our prognostic signature, age, and gender) was significantly higher than what was recently reported by the deep-learning-based MesoNet algorithm[23] (C = 0.656 using whole-slide digitized images, age, and gender) on the TCGA mesothelioma dataset. Similar results were obtained using a K-M analysis (log rank test p = 5.37e−09 for TCGA; Figure 2C; and log rank test p = 4.59e−07 for Bueno et al.[10] data; Figure 2D). Control experiments were performed using randomized gene sets of the same size as those present

CellPress
OPEN ACCESS

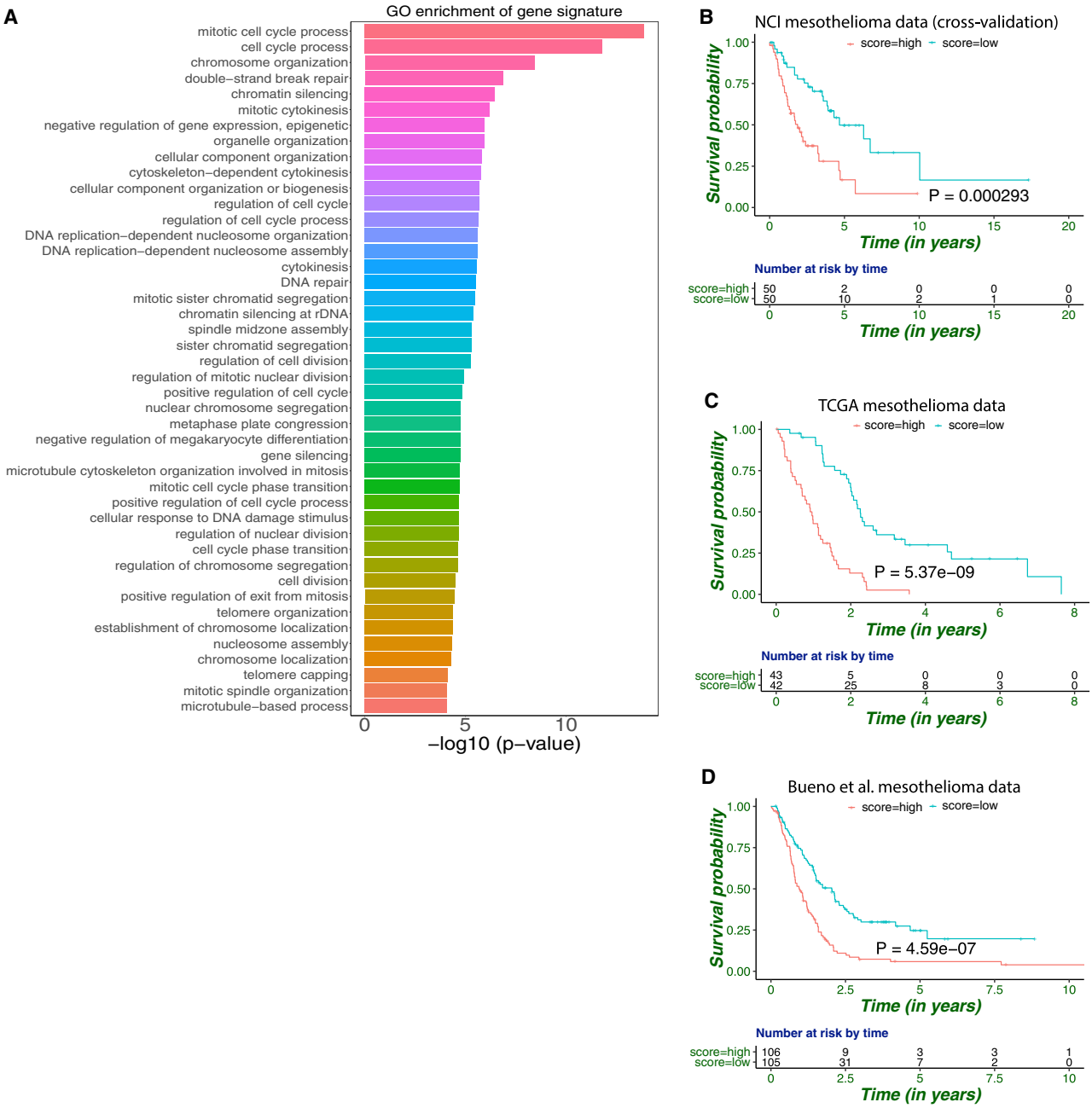


Figure 2. GO enrichment analysis and validation of mesothelioma prognostic signature

(A) GO enrichment analysis of the mesothelioma prognostic signature using the GOrilla tool. GO terms with $p < 1e{-}4$ are shown here (list of all GO terms with $p < 1e{-}3$ are listed in Table S5C).

(B–D) Survival analysis using a K-M plot of patients in the top 50th (high risk) and bottom 50th (low risk) percentiles based on their predicted risk scores in (B) the NCI mesothelioma dataset in cross-validation, (C) the TCGA mesothelioma dataset, and (D) the Bueno et al.[10] mesothelioma dataset. Log rank test p values are shown.

in mesothelioma prognostic signature gene set, and risk scores were computed. Reassuringly, these random control experiments did not show any survival association (1,000 iterations, randomization test empirical $p < 0.001$ while comparing the actual HRs with the control; see Figures S2B and S2C for details). To test the robustness of predicting survival using the 48-gene mesothelioma prognostic signature, we computed a risk score using an alternative method and found that the results were quite similar to those using computing combined risk scores (Figure S2D; STAR Methods). We also found that the list of 27 genes whose increased expression is associated with better survival did not show a survival association in the TCGA

dataset and showed only some significant survival association in the Bueno et al.[10] dataset (Figure S2E).

Of note, while we controlled for demographic factors like age and gender in the Cox regression survival analysis, we did not control for tumor histology because the overwhelming majority of tumors in the NCI mesothelioma cohort were epithelioid (Table S4). However, we note that controlling for tumor stage in the TCGA mesothelioma dataset shows that the high-risk scores (median expression of the 48 genes in the mesothelioma prognostic signature) are significantly associated with worse survival (Cox regression controlling for age, gender, and tumor stage; HR = 3.0, p = 4.8e−9). We performed the survival analysis by considering the time that elapsed from the biopsy to the time of death or last follow-up to align it with the time when the tumor molecular data were obtained. However, notably, even after explicitly controlling for prior treatment time in the Cox regression model, the increased expression in all 48 genes in the mesothelioma prognostic signature is still markedly associated with worse patient survival (Fisher's exact test for enrichment of overlapping genes between explicitly controlling or not controlling of prior treatment is extremely significant, odds ratio = infinity, p = 1.54e−103). Moreover, a PCA clustering of gene expression data did not show distinct clustering of mesothelioma patients with or without prior treatment (Figure S3), indicating that prior treatment of some of the patients is not likely to be confounding our analysis.

We checked the association of gene expression with survival for the TCGA mesothelioma and Bueno et al.[10] datasets separately, using Cox regression after controlling for age and gender. For the TCGA mesothelioma and Bueno et al.[10] datasets, there are 2,144 and 1,210 genes, respectively, whose increased expression is significantly associated with decreased survival (FDR < 0.1; 674 genes intersected between both of these datasets). We found that the mesothelioma prognostic signature of 48 genes has significant overlap with the survival-associated genes from the TCGA (p = 9.41e−15) and Bueno et al.[10] mesothelioma (p = 1.85e−14) datasets (using a hypergeometric test). There are 31 overlapping genes between the three studies whose increased expression is significantly associated with decreased survival in all three mesothelioma datasets (FDR < 0.1; Table S5E). Some of these 31 genes could be potential candidates for drug targeting (provided they are druggable) because inhibiting them could potentially improve patient survival (Table S5E). GO enrichment analysis on these 31 genes shows a strong enrichment for GO terms related to various metabolic processes (Table S5F).

One limitation of the mesothelioma prognostic signature analysis is that independent validation of the genes (risk scores) was done only on independent pleural mesothelioma cohorts (TCGA and Bueno et al.[10]) given the absence of large-scale peritoneal patient data. To study the ability of predicting the survival of such patients too, we reversed the training and test datasets; we used the median expression of the 674 genes whose increased expression was associated with worse patient survival in the TCGA mesothelioma and Bueno et al.[10] datasets to predict the survival of patients in our NCI mesothelioma cohort. We do indeed see that their median expression is predictive of pleural and peritoneal mesothelioma patient survival after controlling for age and gender (Cox regression analysis; entire patient data: HR = 1.7, p = 0.0015; pleural mesothelioma patients only: HR = 2.12, p = 0.0066; peritoneal mesothelioma patients only: HR = 1.54, p = 0.072). This further indicates that the gene expression patterns of pleural and peritoneal mesothelioma patients show a great degree of overlap in their association with patient survival.

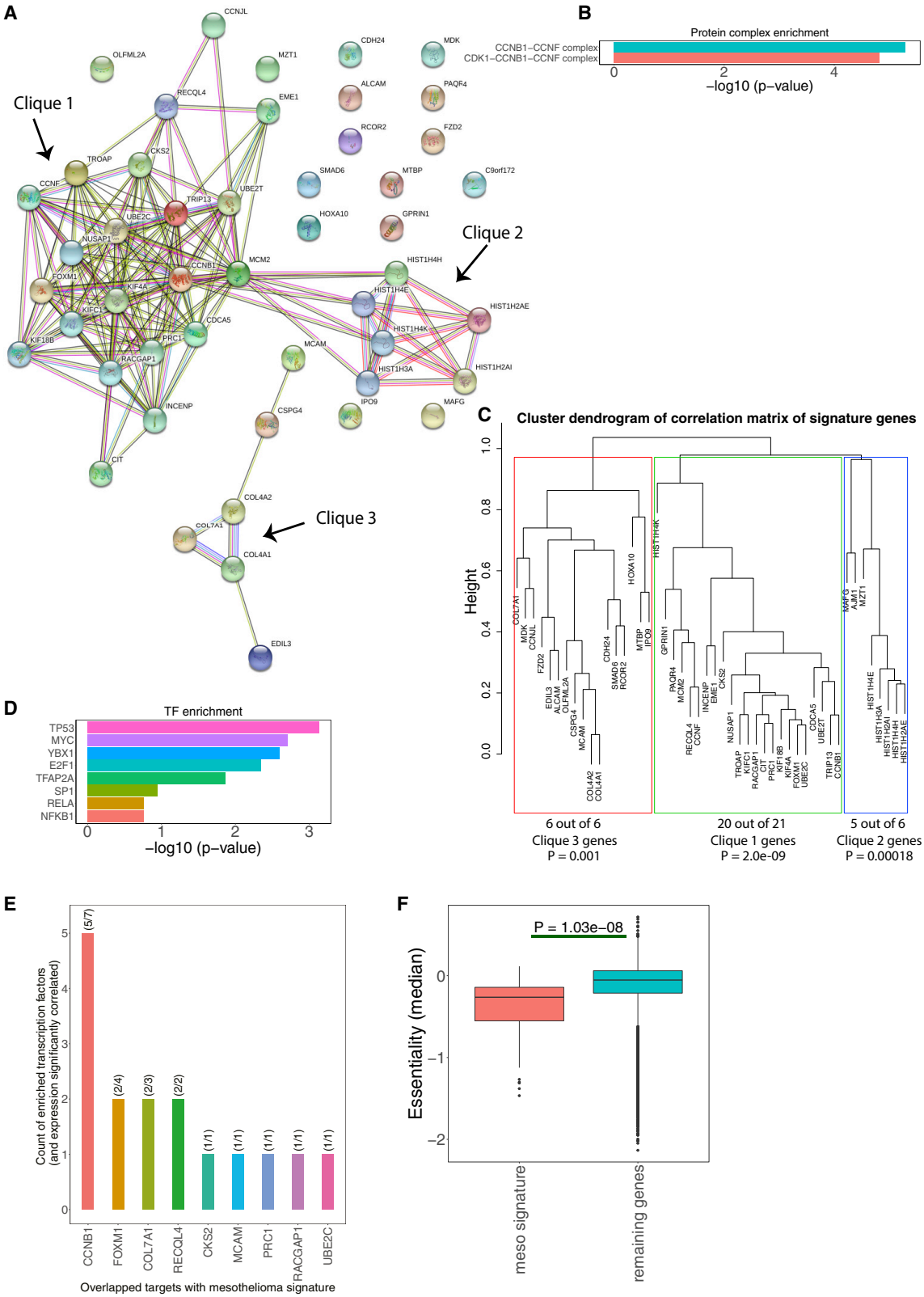
The *CCNB1* gene is strongly predictive of mesothelioma patient survival

We explored the possibility of identifying a smaller subset of genes among the mesothelioma prognostic signature genes that could possibly play an important functional role in mesothelioma with the hope that some of these genes may be potential candidates for targeted therapies. Previous work by Melaiu et al.[35] has identified 51 genes that were differentially upregulated consistently across various studies in mesothelioma vs. non-malignant mesothelial samples. Two of these genes (*CCNB1* and *NUSAP1*) were present in our 48-gene mesothelioma signature gene set.

We calculated the risk factors associated with the increased median expression of these 51 genes and found them to be predictive of worse mesothelioma patient survival in the TCGA and Bueno et al.[10] mesothelioma cohorts (Cox regression after controlling for age and gender; HR = 1.70, p = 9.91e−05 for TCGA and HR = 1.37, p = 4.96e−04 for Bueno et al.[10] mesothelioma data; Figure S4A). The median expression of the *CCNB1* and *NUSAP1* genes by themselves is predictive of patient survival as well (Cox regression after controlling for age and gender; HR = 2.28, p = 1.67e−07 for TCGA and HR = 1.40, p = 2.37e−05 for Bueno et al.[10] mesothelioma data; Figure S4A). Among these two genes, we found that the increased gene expression of *CCNB1* alone was highly associated with worse mesothelioma patient survival outcome (Cox regression after controlling for age and gender; HR = 2.54, p = 1.89e−08 for TCGA mesothelioma and HR = 1.40, p = 1.65e−05 for Bueno et al.[10] mesothelioma data; Figure S4A). A K-M survival analysis also shows that increased gene expression of *CCNB1* alone is highly associated with worse mesothelioma patient survival outcome in the NCI mesothelioma, TCGA mesothelioma, and Bueno et al.[10] mesothelioma datasets (Figure S4B). Although these survival associations are not as strong as that those obtained using the 48-gene mesothelioma prognostic signature, they are still notable. The increased expression of the *CCNB1* gene alone is associated with worse survival outcome even when we look separately at pleural or peritoneal mesothelioma patients in the NCI mesothelioma cohort (Cox regression analysis after controlling for age and gender; pleural mesothelioma patients only: HR = 2.5, p = 0.0018; peritoneal mesothelioma patients only: HR = 1.76, p = 0.0088).

The protein-protein interaction (PPI) network connecting mesothelioma prognostic signature genes supports the functional importance of CCNB1 and the complexes it forms

To further explore the pathways in which the 48 genes in the "mesothelioma prognostic signature" are involved, we looked at the PPIs among them. The PPI network for these 48 genes was inferred from the STRING dataset and website.[36] We identified three different cliques in the network: clique 1 with 21



(legend on next page)

genes, clique 2 with 5 genes, and clique 3 with 6 genes (Figure 3A). We suggest that high-degree hub proteins (i.e., proteins with a lot of interacting partners in the PPI network) in these cliques could play a functional role in mesothelioma and could be potential drug targets. We note that *CCNB1* and *NUSAP1* are such hubs in clique 1. *CCNB1* and *NUSAP1* have degree 19 (ranked 2 of 48) and 16 (ranked 6 of 48), respectively, in the PPI network (the degree of a protein is the number of interacting partners for that protein in the PPI network). *MCM2*, with degree 22, has the highest rank.

Protein complex enrichment analysis was performed on clique 1 genes, which form the largest clique in the PPI network discussed above, using the comprehensive resource of mammalian protein complexes, the CORUM protein complex resource[37] (STAR Methods). These genes are enriched for 2 complexes, CCNB1-CCNF and CDK1- CCNB1-CCNF (Fisher's test, FDR < 0.05; Figure 3B; Table S5G), of which *CCNB1* and *CCNF* are members. These complexes' functional annotation indicates that they are involved in the cell cycle and DNA processing.

We computed the correlation between the expression of the 48 genes present in the mesothelioma prognostic signature across all patients and performed a cluster analysis on their expression patterns (STAR Methods). Again, three clusters were identified that significantly overlapped with the respective cliques obtained in the PPI network (Figure 3C, Fisher's exact test p values are shown), together providing convergent evidence of the functional importance of these genes and these two complexes.

A transcription factor enrichment analysis further points to the central role of *CCNB1*

We then aimed to identify the transcription factors that may control the expression of the mesothelioma prognostic signature genes using the transcriptional regulatory relationships unraveled by sentence-based text mining, TRRUST v.2 package[38] for transcription factor (TF) enrichment analysis. The results are summarized in Table S5H. We found significant enrichment for 8 TFs (*TP53*, *MYC*, *YBX1*, *E2F1*, *TFAP2A*, *SP1*, *RELA*, and *NFKB1*; Figure 3D; Table S5H), which together regulate 9 genes of the 48 composing the mesothelioma prognostic signature (Figure 3E; Table S5H). Notably, of these 8 TFs, 7 have binding sites and regulate *CCNB1*, again pointing to its putative central functional role in mesothelioma. Figure 3E portrays the number of signature genes whose expression is correlated with the corresponding TF gene expression (Spearman correlation, FDR < 0.2). This may provide additional evidence showing that these TFs potentially regulate these genes. Notably, *CCNB1* expression is significantly correlated with 5 of these 7 TFs.

Mesothelioma prognostic signature genes tend to be essential in CRISPR screens of mesothelioma cell lines

Because we are interested in investigating a subset of the 48-gene mesothelioma prognostic signature that could contain potential drug target candidates, we analyzed CRISPR-Cas9 gene knockout essentiality screens[39,40] from 7 pleural mesothelioma cell lines (ACCMESO1, NCIH2452, NCIH2052, MPP89, ISTMES2, MSTO211H, and NCIH28). We took the median essentiality values across the 7 cell lines for each gene. We found that the mesothelioma prognostic signature genes tend to be more essential than the rest of the genes (one-sided Wilcoxon rank-sum test, p = 1.03e−08, difference in median essentiality score = 0.21; Figure 3F). Table S5I shows the essentiality values for the signature genes. Ten genes (*MTBP*, *PRC1*, *INCENP*, *MCM2*, *RACGAP1*, *IPO9*, *CDCA5*, *KIF4A*, *CCNB1*, and *KIF18B*) have essentiality in the top 10 percentile of all genes in these cell lines. *MTBP* is the top essential gene among the 48-gene signature (ranked 98.7 percentile among all genes), and *CCNB1* is also essential (ranked 91.6 percentile among all genes).

Immune cell type abundance estimates show the differential abundance of certain immune cell types in patients with mesothelioma

Next, we looked for immune cell types that were relatively high in mesothelioma with the aim of understanding the mesothelioma immune microenvironment. We used cell-type identification by estimating relative subsets of RNA transcripts CIBERSORT[41] to estimate the abundance of 22 immune cell types in each patient in the NCI mesothelioma dataset (STAR Methods; Table S6). The estimates of immune cells of TCGA cancer patients were also obtained using a CIBERSORT analysis from Lee et al., 2019.[42] The relative immune cell abundance estimates were compared between the NCI mesothelioma dataset (NCI MESO), TCGA mesothelioma patients (TCGA MESO), TCGA pan-cancer patients (mesothelioma not included), TCGA lung adenocarcinoma (TCGA LUAD), and lung squamous cell carcinoma (TCGA LUSC) (Figure 4A). Some immune cell types, like

Figure 3. Mesothelioma prognostic signature analysis

(A) The protein-protein interaction (PPI) network of the genes in the mesothelioma prognostic signature identified 3 cliques in the NCI mesothelioma dataset by STRING.

(B) Protein complex enrichment analysis of clique 1 genes. These genes are enriched in 2 complexes, the CCNB1-CCNF complex and the CDK1-CCNB1-CCNF complex (Fisher's exact test, FDR < 0.02). The x axis represents Fisher's test negative $\log_{10}$ p values.

(C) The correlation matrix was computed on the matrix of the 48-gene mesothelioma prognostic signature across all patients. Hierarchical clustering identified three clusters that significantly overlapped with the three cliques obtained in the PPI network. Fisher's exact test p values are shown.

(D) The mesothelioma prognostic signature is enriched for 8 transcription factors (TFs) (details provided in Table S5H).

(E) Correlation between expression of mesothelioma signature genes and corresponding TFs. Expression of every gene in the mesothelioma prognostic signature cohort that is potentially regulated by one or more of the 8 enriched TFs shown in (D) was correlated with expression of the corresponding mapped TF (FDR < 0.2). The number of such enriched TFs that were significantly correlated (FDR < 0.2) with these genes is plotted as a bar graph. The count of enriched TFs that are significantly correlated vs. the total number of TFs mapped to a particular gene is shown in parentheses.

(F) Median essentiality values from CRISPR-Cas9 gene knockout essentiality screens from 7 pleural mesothelioma cell lines (ACCMESO1, NCIH2452, NCIH2052, MPP89, ISTMES2, MSTO211H, and NCIH28) for each gene are computed and compared between genes in the mesothelioma prognostic signature and the remaining genes. Less essentiality value implies that the gene is more essential. One-sided Wilcoxon rank-sum test p value is shown.

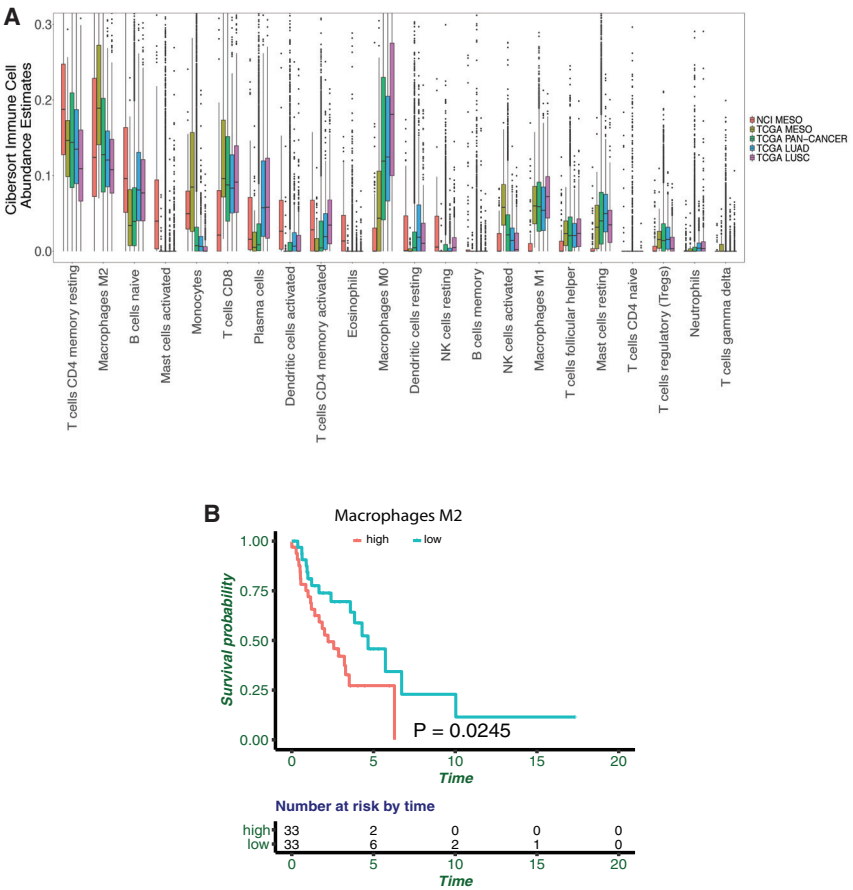


Figure 4. Immune cell abundance estimates

(A) Boxplot of the relative fractions of the immune cell abundance across all mesothelioma samples in the NCI mesothelioma dataset (NCI MESO), TCGA pan-cancer patients (TCGA PAN-CANCER; except mesothelioma), TCGA lung adenocarcinoma (TCGA LUAD), and TCGA lung squamous cell carcinoma (TCGA LUSC). The relative immune cell abundance for each immune cell type is shown as a fraction on the y axis.

(B) K-M survival plot of patients with the top 33rd and bottom 33rd percentiles of relative abundance of M2 macrophages in the NCI mesothelioma dataset. Time is shown in years. Log rank test p value is shown.

CD4$^+$ memory resting T cells, naive B cells, activated mast cells, and monocytes, had higher relative abundance in the NCI mesothelioma dataset than in other cancer types in the TCGA (one-sided Wilcoxon rank-sum test, FDR < 0.1; Figure 4A). We also observed that some immune cell types, like CD8$^+$ T cells and M0 macrophages, have lower abundance in the NCI mesothelioma dataset than in other cancer types in the TCGA database (one-sided Wilcoxon rank-sum test, FDR < 0.1; Figure 4A).

We tested the association of all immune cell types that have more than 5% mean relative abundance across samples in the NCI mesothelioma dataset (STAR Methods) with patient survival. We found that an increase in M2 macrophages was associated with worse mesothelioma patient survival in the NCI mesothelioma dataset (Cox regression HR = 17.42, p = 0.027, FDR = 0.16, after controlling for age and gender). We repeated this analysis using K-M survival plots and found similar results (Figure 4B; log rank p = 0.0245). Because M2 macrophages are thought to be immunosuppressive,[43,44] their association with a decrease in survival was expected. We, however, did not see any association of M2 macrophage abundance estimates (using CIBERSORT) with patient survival in the TCGA mesothelioma and Bueno et al.[10] datasets (FDR < 0.2). We also did not find any significant difference for any immune cell type abundance estimates between pleural and peritoneal mesothelioma patients (FDR < 0.2).

A synthetic lethality/rescue transcriptome-based approach predicts responses to anti-PD1 therapy and combination therapy with pemetrexed in patients with mesothelioma

To employ a precision medicine-based approach for recommending therapies in mesothelioma, we employed the concept of SL/sickness and synthetic rescue (SR) to predict patient response to drugs (STAR Methods). Lee et al.[46] used a computational pipeline called SELECT (SL and rescue-mediated precision oncology via the transcriptome), a precision oncology approach that aims to predict drug response for a given cancer patient using the whole pre-treatment transcriptome data. SELECT is based on inferring clinically relevant pan-cancer SL and SR pairs across cancer types by mining thousands of patient tumor genomics and transcriptomics data points from the TCGA pan-cancer data and various *in vitro* studies (STAR Methods).[42,45,46] For each chemotherapy or targeted therapy drug, SELECT assigns a risk score based on the number of downregulated SL partners of the target genes inhibited by the drug. For immunotherapy drugs, SELECT assigns risk scores using SR interactions in an analogous manner (STAR Methods). It is important to note that SELECT does not train any model parameters by looking at the test data, mitigating the risk of overfitting the data and loss of generalization predictive power on unseen datasets.

Among the 100 patients with expression data in the NCI mesothelioma data, 16 patients were treated with anti-PD1 immune checkpoint inhibitors (pembrolizumab or nivolumab; STAR Methods; Table S7A). DrugBank was used to map drugs to targets that are inhibited.[47] Similar to what was reported in other clinical trials,[13] only a small fraction of these patients (18.8%; 3 of 16 patients) responded to treatment (either complete or partial response). We used SELECT to predict patient response for anti-PD1 drugs in the NCI mesothelioma patient cohort, using the exact same parameters as used previously, including the decision threshold value.[46] Despite having a small

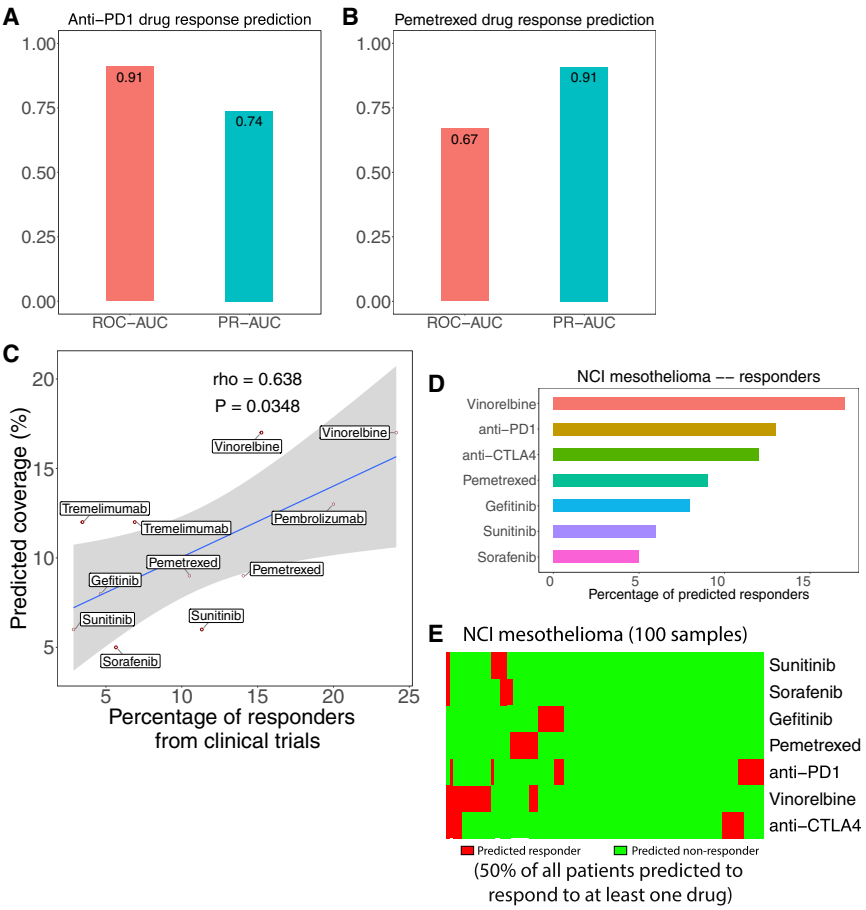


Figure 5. Overall SL transcriptomics-based response prediction across several mesothelioma clinical trials

(A) We use SELECT to predict patient response for anti-PD1 drugs in the NCI mesothelioma patient cohort using patient gene expression data. Bar plots show ROC-AUC and area under precision recall (PR-AUC) values using the SELECT-derived risk to predict responders (complete or partial response) and non-responders (stable disease or progressive disease) to anti-PD1 immunotherapy (16 samples). ROC-AUC and PR-AUC values range from 0–1, with higher values indicating higher performance of the predictor. ROC-AUC for a random predictor is expected to be 0.5. The 95% confidence interval for ROC-AUC varied from 0.7–1.

(B) Bar plots show ROC-AUC and PR-AUC values using the SELECT-derived risk to predict responders (complete or partial response) and non-responders (progressive disease) to combinations with the chemotherapy drug pemetrexed (41 samples). The 95% confidence interval for ROC-AUC varied from 0.46–0.88.

(C) Scatterplot showing the percentage of responders (objective response rate) from mesothelioma clinical trials and comparing it with the percentage of predicted responders (coverage) using SELECT in the NCI mesothelioma dataset. Spearman's ρ (rho) and p values are shown. The shaded region is the 95% confidence level interval for a linear model.

(D) Bar plot showing the percentage of predicted responders for each drug.

(E) Heatmap showing the responders in all patients and drugs for NCI mesothelioma data. The x axis corresponds to patient samples and y axis to each drug. Red indicates that the patient is predicted to respond to that drug, and green indicates non-response. 50% of the total percentage of patients are predicted to respond to at least 1 drug in the NCI mesothelioma dataset.

number of responders, we were able to accurately predict anti-PD1 drug response in mesothelioma patients (area under ROC curve, ROC-AUC = 0.91 [95% confidence interval is 0.7–1]; area under precision recall [PR-AUC] = 0.74; Figure 5A; Table S7A).

Pemetrexed is a commonly used chemotherapy drug, and it is often used in combination with other chemotherapy drugs like cisplatin or carboplatin. In the NCI mesothelioma cohort, pemetrexed was always given in combination with other drugs, and we considered the 41 patients who were given this drug (1 complete response, 23 partial response, 6 progressive disease, 11 stable disease; STAR Methods; Table S7B). We used the drug targets of pemetrexed as reported in DrugBank to predict its corresponding drug combination response using SELECT. We could not consider the response to the other drugs (cisplatin/carboplatin) given in these combinations because they could not be mapped to targets they inhibited according to DrugBank. Despite the limitations of this analysis, when we removed patients with stable disease from the analysis, we were able to predict patients with complete/partial response from those with progressive disease quite well (ROC-AUC = 0.67 [95% confidence interval is 0.46–0.88]; PR-AUC = 0.91; Figure 5B; Table S7B).

SL-based transcriptomics-based prediction of overall response to cancer drugs in numerous clinical trials in mesothelioma

We wanted to check whether the concept of SL/SRs could be used to predict the percentage of responders to various mesothelioma cancer therapies. We focused on analyzing clinical trials using various targeted therapies and immunotherapies because these categories of drugs have been shown previously to be well predicted by SELECT.[46] We additionally considered a few chemotherapies commonly used in mesothelioma, like pemetrexed and vinorelbine. We did not consider other commonly used chemotherapy drugs like cisplatin and carboplatin because they could not be mapped to drug targets they clearly inhibit (according to DrugBank). For each of the drugs, we collected information about the objective response rates (percentage of responders) in mesothelioma patients from many of the relevant published clinical trials. We succeeded in obtaining this information for 7 drugs overall, including 2 chemotherapies, 3 targeted therapies, and 2 immunotherapies: pemetrexed, vinorelbine, gefitinib, sorafenib, sunitinib, pembrolizumab, and tremelimumab[12–22] (Table S7C). The individual patients' transcriptomics data were not available for these clinical trials, so we could not predict the individual response of each patient in these trials.

However, we could still study how well the predicted SL-based coverage for each drug (based on the analysis of available mesothelioma patients' transcriptomes in our cohort) matches the actual overall response observed in these trials.

To this end, SELECT was used to predict the percentage of responders (coverage) for each drug in the NCI mesothelioma cohort, using the same parameters and thresholds used previously[46] and repeating the exact same procedure they published (STAR Methods). We found a significant correlation between the predicted and actual coverage of these drugs (Spearman's $\rho = 0.64$, $p = 0.0348$; Figure 5C; Table S7C). Notably, this also suggests that the NCI mesothelioma cohort may faithfully reflect the transcriptome of patients participating in these various clinical trials. For robustness, we performed the same analysis on TCGA mesothelioma patients and, reassuringly, obtained similar results (Figure S5A). Figures 5D and 5E shows the fraction of patients predicted to respond to each of the 7 drugs considered, using SELECT (Tables S7D and S7E). We found that 50% of the patients in the NCI mesothelioma data were predicted to respond to at least one of the 7 drugs based on the risk scores of the drugs computed for each patient (Figure 5E). This percentage is higher for the TCGA data (64.71%). Because some patients were predicted to respond to more than one drug, we ranked the effectiveness of all drugs for each patient in the NCI mesothelioma cohort. Across the chemotherapies and targeted therapies considered, vinorelbine is the highest-ranked drug (see details in Figure S5B), and this is in line with the outcomes observed in the clinical trials.[21,22] A similar analysis was done for immune checkpoint inhibitors in mesothelioma (Figure S5C).

DISCUSSION

In this study, we performed WES of blood and tumor tissues from 122 malignant mesothelioma patients and systematically studied their somatic and germline mutations. We found that key somatic mutational signatures were associated with DNA repair pathways, and *BAP1* was the most commonly mutated gene (~13% with germline mutation). We also performed RNA-seq of 100 tumor tissue samples for transcriptomics analysis. Unlike previous studies, which mainly focused on pleural mesothelioma patients, this study includes a comparable number of patients of the MPM and MPeM subtypes and, hence, gives us a better understanding of the similarities and differences that may exist in the molecular pathophysiology of the two anatomically distinct diseases. We acknowledge that the TCGA mesothelioma and Bueno et al.[10] datasets contain larger numbers of pleural mesothelioma patients than our cohort. Moreover, our dataset does not contain patient information on pathological stage. As discussed in previous studies, we observed that the overall survival of patients with MPeM was higher than that of patients with MPM. Surprisingly, we did not find too many differences between pleural and peritoneal mesothelioma samples at the level of germline or somatic mutations or at the level of copy number information or gene expression.

We further stratified the patients based on their overall survival and analyzed the differentially expressed genes in the high- and low-risk subgroups. We identified a "mesothelioma prognostic signature," a set of 48 genes that were overexpressed in a high-risk cohort with poor survival, and showed that the median expression of these genes is highly predictive of mesothelioma patient survival in two independent large-scale patient cohorts and, thus, may have high translational value in the clinic. We found that these genes were enriched for GO terms related to cell cycle processes, DNA repair, and others. Among these 48 genes, we found that the *CCNB1* gene alone was highly predictive of mesothelioma patient survival. While many of the patients in the NCI mesothelioma cohort received prior treatment, through various control analysis, we were able to show that this is not a confounding factor in our survival analysis. We performed an independent validation of the mesothelioma prognostic signature on large-scale pleural mesothelioma cohorts solely because of the absence of similar large-scale peritoneal datasets. However, we have shown in cross-validation that these signatures are predictive of peritoneal patient survival in the NCI mesothelioma cohort. Furthermore, we also found that the median expression of the genes whose expression was associated with worse survival in pleural mesothelioma (TCGA and Bueno et al.[10]) is reasonably predictive of worse peritoneal mesothelioma survival in the NCI mesothelioma cohort, further testifying to the overall similar associations of gene expression and patient survival between both disorders.

Multiple analyses showed that the *CCNB1* gene may potentially play an important role in mesothelioma. Cyclin B1, a protein encoded by the *CCNB1* gene, is a key mitotic cyclin in the G2-M phase transition of the cell cycle and is overexpressed in various malignant tumors, including lung, breast, colorectal, pancreatic, and others. Previous work has shown that Cyclin B1 plays an important role in tumor development and tumorigenesis.[48,49] A meta-analysis about the prognostic role of cyclin B1 in solid tumor was conducted on 17 published studies, concluding that overexpression of cyclin B1 is a significant prognostic parameter and associated with poor survival in many solid tumors.[50] However, there is no prior work showing that expression of the *CCNB1* gene by itself is associated with poor survival of mesothelioma patients. The *CCNF* gene, which encodes Cyclin F, is also a member of the cyclin family, is known to control genome instability, and may play a role in cancer development.[51] The *NUSAP1* gene is a nucleolar-spindle-associated protein that is known to play a role in spindle microtubule organization[52] and in some other cancers[53–55] but has not been significantly discussed in mesothelioma. Looking at CRISPR-Cas9-based essentiality screens in pleural mesothelioma cell lines, we found that the mesothelioma prognostic signature genes tend to be more essential, and we found about 10 genes that were highly essential (including the *MTBP* and *CCNB1* genes). The *MTBP* gene is known to play a role in cell migration and invasion in some tumors[56–58] and has not been discussed in mesothelioma. Thus, using various analyses, we were able to identify a subset of the 48 genes in the mesothelioma prognostic signature that are likely to play an important role in mesothelioma. However, we note that our studies are associative, and further *in vitro* and *in vivo* work is

required to show any causal relationships for any of the mesothelioma prognostic signature genes to mesothelioma cancer progression.

We used CIBERSORT to determine immune cell lineages infiltrating the tumor tissue based on immune-cell-specific marker expression and compared the immune abundance estimates in mesothelioma with other cancer types. We found that some immune cell types, like $CD4^+$ memory T cells and naive B cells, are higher in relative abundance in NCI mesothelioma data than some of the other cancer types in the TCGA. We correlated the gene expression data with patient survival and found that increased tumor M2 macrophage abundance was associated with poor survival. Of note, M2 macrophages are immunosuppressive in function, and their abundance has been associated with poor survival in a few other cancers.[59]

A precision oncology-based approach called SELECT has been shown to effectively predict drug response for various cancer treatments across many different cancer patients and cancer types using patient transcriptomic data and the concepts of SL and SR.[46] Using exactly the same parameters and thresholds used by Lee et al.,[46] we applied SELECT to the NCI mesothelioma dataset and were able to effectively predict anti-PD1 immune checkpoint inhibitor response. We were also able to predict drug combination response for an important chemotherapy drug like pemetrexed. We note that SELECT does not train on the NCI mesothelioma cohort. Using SELECT to predict the percentage of responders (coverage) for various drugs in the NCI mesothelioma cohort (and in TCGA mesothelioma) for which we know the response rate based on mesothelioma clinical trials, we found a significant correlation between the predicted and actual coverage for these drugs. Obviously, the veracity of these predictions needs to be studied carefully in controlled prospective clinical trials in the future.

In conclusion, by analyzing genome and tumor transcriptomes of patients with pleural and peritoneal mesothelioma, we were able to identify gene expression profiles predictive of patient survival as well as response to available therapies. These results need to be further validated in laboratory studies and well-conducted prospective clinical trials.

Limitations of the study

Our study has several limitations. First, obviously, like so many other related studies, the expression of the 48-gene mesothelioma prognostic signature, including the CCNB1 gene, is just associated with patient survival, and any possible causal relations remain to be further studied. Second, the SELECT analysis we performed has a few limitations: (1) the number of patient samples for drug response analysis for the anti-PD1 checkpoint inhibitors and pemetrexed is small and the response classes are imbalanced, and (2) some parts of the SL analysis were performed on the cohort level in the absence of patient-specific transcriptomics data. While the results obtained given these limitations are quite encouraging, further validation in larger cohorts is warranted to evaluate the potential value of an SL-based patient stratification approach in mesothelioma.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patients
- METHOD DETAILS
  - Analysis of germline and somatic mutations and tumor mutational burden
  - RNA sequencing and data processing
  - Alternative method to compute risk scores
  - Protein complex enrichment and cluster gene expression
  - Immune cell abundance estimates
  - Drug response prediction using SELECT
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2023.100938.

ACKNOWLEDGMENTS

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute, and Center for Cancer Research (ZIA-BC-010816). This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). The results shown here are in part based on data generated by the TCGA Research Network (https://www.cancer.gov/tcga). We would like to thank Dr. Kuoyuan Cheng for help with this project.

AUTHOR CONTRIBUTIONS

Conceptualization, N.U.N., Q.J., J.S.W., J.K., E.R., and R.H.; methodology, N.U.N., Q.J., J.S.W., V.A.M., B.M., C.K., L.C.H., J.S.L., I.M., J.Z., A.L., M.M., M.S., J.K., E.R., and R.H.; software programming, N.U.N., J.S.W., L.C.H., J.K., E.R., and R.H.; validation verification, N.U.N., Q.J., J.S.W., V.A.M., B.M., C.K., L.C.H., J.S.L., I.M., J.Z., A.L., M.M., M.S., J.K., E.R., and R.H.; formal analysis, N.U.N., Q.J., J.S.W., V.A.M., B.M., C.K., L.C.H., J.S.L., I.M., J.Z., A.L., M.M., M.S., J.K., E.R., and R.H.; resource provision, J.K., E.R., and R.H.; data curation management, N.U.N., J.S.W., J.K., E.R., and R.H.; writing – original draft, N.U.N., Q.J., J.S.W., M.S., J.K., E.R., and R.H.; writing – review & editing, N.U.N., Q.J., J.S.W., V.A.M., B.M., C.K., L.C.H., J.S.L., I.M., J.Z., A.L., M.M., M.S., J.K., E.R., and R.H.; visualization, N.U.N., Q.J., J.S.W., B.M., M.S., J.K., E.R., and R.H.; supervision, J.K., E.R., and R.H.; project administration, J.K., E.R., and R.H.; funding acquisition, J.K., E.R., and R.H.

DECLARATION OF INTERESTS

E.R. is a co-founder of Medaware, Metabomed, and Pangea Biomed (divested from the latter). He serves as a non-paid scientific consultant to Pangea Biomed under a collaboration agreement between Pangea Biomed and the NCI. R.H. has received funding for conduct of clinical trials via a cooperative research and development agreement between NCI and Bayer AG and TCR2 Therapeutics. J.S.L. is a scientific consultant of Pangea Therapeutics.

Received: April 25, 2022
Revised: August 23, 2022
Accepted: January 19, 2023
Published: February 10, 2023

REFERENCES

1. Carbone, M., Adusumilli, P.S., Alexander, H.R., Jr., Baas, P., Bardelli, F., Bononi, A., Bueno, R., Felley-Bosco, E., Galateau-Salle, F., Jablons, D., et al. (2019). Mesothelioma: scientific clues for prevention, diagnosis, and therapy. CA. Cancer J. Clin. *69*, 402–429.
2. Greenbaum, A., and Alexander, H.R. (2020). Peritoneal mesothelioma. Transl. Lung Cancer Res. *9*, S120–S132.
3. Vogelzang, N.J., Rusthoven, J.J., Symanowski, J., Denham, C., Kaukel, E., Ruffie, P., Gatzemeier, U., Boyer, M., Emri, S., Manegold, C., et al. (2003). Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma. J. Clin. Oncol. *21*, 2636–2644.
4. Baas, P., Scherpereel, A., Nowak, A.K., Fujimoto, N., Peters, S., Tsao, A.S., Mansfield, A.S., Popat, S., Jahan, T., Antonia, S., et al. (2021). First-line nivolumab plus ipilimumab in unresectable malignant pleural mesothelioma (CheckMate 743): a multicentre, randomised, open-label, phase 3 trial. Lancet *397*, 375–386.
5. Sluis-Cremer, G.K., Liddell, F.D., Logan, W.P., and Bezuidenhout, B.N. (1992). The mortality of amphibole miners in South Africa, 1946-80. Br. J. Ind. Med. *49*, 566–575.
6. Chang, E.T., Lau, E.C., Mowat, F.S., and Teta, M.J. (2017). Therapeutic radiation for lymphoma and risk of second primary malignant mesothelioma. Cancer Causes Control. *28*, 971–979.
7. Farioli, A., Ottone, M., Morganti, A.G., Compagnone, G., Romani, F., Cammelli, S., Mattioli, S., and Violante, F.S. (2016). Radiation-induced mesothelioma among long-term solid cancer survivors: a longitudinal analysis of SEER database. Cancer Med. *5*, 950–959.
8. Baumann, F., and Carbone, M. (2016). Environmental risk of mesothelioma in the United States: an emerging concern-epidemiological issues. J. Toxicol. Environ. Health B Crit. Rev. *19*, 231–249.
9. Carbone, M., Harbour, J.W., Brugarolas, J., Bononi, A., Pagano, I., Dey, A., Krausz, T., Pass, H.I., Yang, H., and Gaudino, G. (2020). Biological mechanisms and clinical significance of BAP1 mutations in human cancer. Cancer Discov. *10*, 1103–1120.
10. Bueno, R., Stawiski, E.W., Goldstein, L.D., Durinck, S., De Rienzo, A., Modrusan, Z., Gnad, F., Nguyen, T.T., Jaiswal, B.S., Chirieac, L.R., et al. (2016). Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. Nat. Genet. *48*, 407–416.
11. Betti, M., Aspesi, A., Biasi, A., Casalone, E., Ferrante, D., Ogliara, P., Gironi, L.C., Giorgione, R., Farinelli, P., Grosso, F., et al. (2016). CDKN2A and BAP1 germline mutations predispose to melanoma and mesothelioma. Cancer Lett. *378*, 120–130.
12. Govindan, R., Kratzke, R.A., Herndon, J.E., 2nd, Niehans, G.A., Vollmer, R., Watson, D., Green, M.R., and Kindler, H.L.; Cancer and Leukemia Group B CALGB 30101 (2005). Gefitinib in patients with malignant mesothelioma: a phase II study by the Cancer and Leukemia Group B. Clin. Cancer Res. *11*, 2300–2304.
13. Alley, E.W., Lopez, J., Santoro, A., Morosky, A., Saraf, S., Piperdi, B., and van Brummelen, E. (2017). Clinical safety and activity of pembrolizumab in patients with malignant pleural mesothelioma (KEYNOTE-028): preliminary results from a non-randomised, open-label, phase 1b trial. Lancet Oncol. *18*, 623–630.
14. Scagliotti, G.V., Shin, D.M., Kindler, H.L., Vasconcelles, M.J., Keppler, U., Manegold, C., Burris, H., Gatzemeier, U., Blatter, J., Symanowski, J.T., and Rusthoven, J.J. (2003). Phase II study of pemetrexed with and without folic acid and vitamin B12 as front-line therapy in malignant pleural mesothelioma. J. Clin. Oncol. *21*, 1556–1561.
15. Taylor, P., Castagneto, B., Dark, G., Marangolo, M., Scagliotti, G.V., van Klaveren, R.J., Labianca, R., Serke, M., Schuette, W., van Meerbeeck, J.P., et al. (2008). Single-agent pemetrexed for chemonaïve and pretreated patients with malignant pleural mesothelioma: results of an International Expanded Access Program. J. Thorac. Oncol. *3*, 764–771.
16. Papa, S., Popat, S., Shah, R., Prevost, A.T., Lal, R., McLennan, B., Cane, P., Lang-Lazdunski, L., Viney, Z., Dunn, J.T., et al. (2013). Phase 2 study of sorafenib in malignant mesothelioma previously treated with platinum-containing chemotherapy. J. Thorac. Oncol. *8*, 783–787.
17. Nowak, A.K., Millward, M.J., Creaney, J., Francis, R.J., Dick, I.M., Hasani, A., van der Schaaf, A., Segal, A., Musk, A.W., and Byrne, M.J. (2012). A phase II study of intermittent sunitinib malate as second-line therapy in progressive malignant pleural mesothelioma. J. Thorac. Oncol. *7*, 1449–1456.
18. Laurie, S.A., Gupta, A., Chu, Q., Lee, C.W., Morzycki, W., Feld, R., Foo, A.H., Seely, J., Goffin, J.R., Laberge, F., et al. (2011). Brief report: a phase II study of sunitinib in malignant pleural mesothelioma. the NCIC Clinical Trials Group. J. Thorac. Oncol. *6*, 1950–1954.
19. Calabrò, L., Morra, A., Fonsatti, E., Cutaia, O., Amato, G., Giannarelli, D., Di Giacomo, A.M., Danielli, R., Altomonte, M., Mutti, L., and Maio, M. (2013). Tremelimumab for patients with chemotherapy-resistant advanced malignant mesothelioma: an open-label, single-arm, phase 2 trial. Lancet Oncol. *14*, 1104–1111.
20. Calabrò, L., Morra, A., Fonsatti, E., Cutaia, O., Fazio, C., Annesi, D., Lenoci, M., Amato, G., Danielli, R., Altomonte, M., et al. (2015). Efficacy and safety of an intensified schedule of tremelimumab for chemotherapy-resistant malignant mesothelioma: an open-label, single-arm, phase 2 study. Lancet Respir. Med. *3*, 301–309.
21. Zucali, P.A., Perrino, M., Lorenzi, E., Ceresoli, G.L., De Vincenzo, F., Simonelli, M., Gianoncelli, L., De Sanctis, R., Giordano, L., and Santoro, A. (2014). Vinorelbine in pemetrexed-pretreated patients with malignant pleural mesothelioma. Lung Cancer *84*, 265–270.
22. Steele, J.P., Shamash, J., Evans, M.T., Gower, N.H., Tischkowitz, M.D., and Rudd, R.M. (2000). Phase II study of vinorelbine in patients with malignant pleural mesothelioma. J. Clin. Oncol. *18*, 3912–3917.
23. Courtiol, P., Maussion, C., Moarii, M., Pronier, E., Pilcer, S., Sefta, M., Manceron, P., Toldo, S., Zaslavskiy, M., Le Stang, N., et al. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nat. Med. *25*, 1519–1525.
24. Pass, H.I., Liu, Z., Wali, A., Bueno, R., Land, S., Lott, D., Siddiq, F., Lonardo, F., Carbone, M., and Draghici, S. (2004). Gene expression profiles predict survival and progression of pleural mesothelioma. Clin. Cancer Res. *10*, 849–859.
25. Gordon, G.J., Jensen, R.V., Hsiao, L.-L., Gullans, S.R., Blumenstock, J.E., Richards, W.G., Jaklitsch, M.T., Sugarbaker, D.J., and Bueno, R. (2003). Using gene expression ratios to predict outcome among patients with mesothelioma. J. Natl. Cancer Inst. *95*, 598–605.
26. López-Ríos, F., Chuai, S., Flores, R., Shimizu, S., Ohno, T., Wakahara, K., Illei, P.B., Hussain, S., Krug, L., Zakowski, M.F., et al. (2006). Global gene expression profiling of pleural mesotheliomas: overexpression of aurora kinases and P16/CDKN2A deletion as prognostic factors and critical evaluation of microarray-based prognostic prediction. Cancer Res. *66*, 2970–2979.
27. Ladanyi, M., Sanchez Vega, F., and Zauderer, M. (2019). Loss of BAP1 as a candidate predictive biomarker for immunotherapy of mesothelioma. Genome Med. *11*, 18.
28. Hmeljak, J., Sanchez-Vega, F., Hoadley, K.A., Shih, J., Stewart, C., Heiman, D., Tarpey, P., Danilova, L., Drill, E., Gibb, E.A., et al. (2018). Integrative molecular characterization of malignant pleural mesothelioma. Cancer Discov. *8*, 1548–1565.
29. Hassan, R., Morrow, B., Thomas, A., Walsh, T., Lee, M.K., Gulsuner, S., Gadiraju, M., Panou, V., Gao, S., Mian, I., et al. (2019). Inherited predisposition to malignant mesothelioma and overall survival following platinum chemotherapy. Proc. Natl. Acad. Sci. USA *116*, 9008–9013.

30. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. Genet. Med. *17*, 405–424.

31. Hiltbrunner, S., Mannarino, L., Kirschner, M.B., Opitz, I., Rigutto, A., Laure, A., Lia, M., Nozza, P., Maconi, A., Marchini, S., et al. (2021). Tumor immune microenvironment and genetic alterations in mesothelioma. Front. Oncol. *11*, 660039.

32. Shao, C., Li, G., Huang, L., Pruitt, S., Castellanos, E., Frampton, G., Carson, K.R., Snow, T., Singal, G., Fabrizio, D., et al. (2020). Prevalence of high tumor mutational burden and association with survival in patients with less common solid tumors. JAMA Netw. Open *3*, e2025109.

33. Kaplan, E.L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. *53*, 457–481.

34. Shukuya, T., Serizawa, M., Watanabe, M., Akamatsu, H., Abe, M., Imai, H., Tokito, T., Ono, A., Taira, T., Kenmotsu, H., et al. (2014). Identification of actionable mutations in malignant pleural mesothelioma. Lung Cancer *86*, 35–40.

35. Melaiu, O., Melissari, E., Mutti, L., Bracci, E., De Santi, C., Iofrida, C., Di Russo, M., Cristaudo, A., Bonotti, A., Cipollini, M., et al. (2015). Expression status of candidate genes in mesothelioma tissues and cell lines. Mutat. Res. *771*, 6–12.

36. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

37. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. Nucleic Acids Res. *47*. D559-d563.

38. Han, H., Cho, J.W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., et al. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res. *46*. D380-d386.

39. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

40. Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784.

41. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

42. Lee, J.S., and Ruppin, E. (2019). Multiomics prediction of response rates to therapies to inhibit programmed cell death 1 and programmed cell death 1 ligand 1. JAMA Oncol. *5*, 1614–1618.

43. Noy, R., and Pollard, J.W. (2014). Tumor-associated macrophages: from mechanisms to therapy. Immunity *41*, 49–61.

44. Zhao, W., Beers, D.R., Thonhoff, J.R., Thome, A.D., Faridar, A., Wang, J., Wen, S., Ornelas, L., Sareen, D., Goodridge, H.S., et al. (2020). Immunosuppressive functions of M2 macrophages derived from iPSCs of patients with ALS and healthy controls. iScience *23*, 101192.

45. Sahu, A.D., S Lee, J., Wang, Z., Zhang, G., Iglesias-Bartolome, R., Tian, T., Wei, Z., Miao, B., Nair, N.U., Ponomarova, O., et al. (2019). Genome-wide prediction of synthetic rescue mediators of resistance to targeted and immunotherapy. Mol. Syst. Biol. *15*, e8323.

46. Lee, J.S., Nair, N.U., Dinstag, G., Chapman, L., Chung, Y., Wang, K., Sinha, S., Cha, H., Kim, D., Schperberg, A.V., et al. (2021). Synthetic lethality-mediated precision oncology via the tumor transcriptome. Cell *184*, 2487–2502.e13.

47. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. *46*. D1074-d1082.

48. Soria, J.C., Jang, S.J., Khuri, F.R., Hassan, K., Liu, D., Hong, W.K., and Mao, L. (2000). Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. Cancer Res. *60*, 4000–4004.

49. Song, Y., Zhao, C., Dong, L., Fu, M., Xue, L., Huang, Z., Tong, T., Zhou, Z., Chen, A., Yang, Z., et al. (2008). Overexpression of cyclin B1 in human esophageal squamous cell carcinoma cells induces tumor cell invasive growth and metastasis. Carcinogenesis *29*, 307–315.

50. Ye, C., Wang, J., Wu, P., Li, X., and Chai, Y. (2017). Prognostic role of cyclin B1 in solid tumors: a meta-analysis. Oncotarget *8*, 2224–2232.

51. D'Angiolella, V., Esencay, M., and Pagano, M. (2013). A cyclin without cyclin-dependent kinases: cyclin F controls genome stability through ubiquitin-mediated proteolysis. Trends Cell Biol. *23*, 135–140.

52. Raemaekers, T., Ribbeck, K., Beaudouin, J., Annaert, W., Van Camp, M., Stockmans, I., Smets, N., Bouillon, R., Ellenberg, J., and Carmeliet, G. (2003). NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization. J. Cell Biol. *162*, 1017–1029.

53. Han, G., Wei, Z., Cui, H., Zhang, W., Wei, X., Lu, Z., and Bai, X. (2018). NUSAP1 gene silencing inhibits cell proliferation, migration and invasion through inhibiting DNMT1 gene expression in human colorectal cancer. Exp. Cell Res. *367*, 216–221.

54. Gordon, C.A., Gong, X., Ganesh, D., and Brooks, J.D. (2017). NUSAP1 promotes invasion and metastasis of prostate cancer. Oncotarget *8*, 29935–29950.

55. Zhang, X., Pan, Y., Fu, H., and Zhang, J. (2018). Nucleolar and spindle associated protein 1 (NUSAP1) inhibits cell proliferation and enhances susceptibility to epirubicin in invasive breast cancer cells by regulating cyclin D kinase (CDK1) and DLGAP5 expression. Med. Sci. Monit. *24*, 8553–8564.

56. Agarwal, N., Adhikari, A.S., Iyer, S.V., Hekmatdoost, K., Welch, D.R., and Iwakuma, T. (2013). MTBP suppresses cell migration and filopodia formation by inhibiting ACTN4. Oncogene *32*, 462–470.

57. Bi, Q., Ranjan, A., Fan, R., Agarwal, N., Welch, D.R., Weinman, S.A., Ding, J., and Iwakuma, T. (2015). MTBP inhibits migration and metastasis of hepatocellular carcinoma. Clin. Exp. Metastasis *32*, 301–311.

58. Lu, S., Zhou, W., Wei, H., He, L., and Li, L. (2015). MTBP promotes the invasion and metastasis of hepatocellular carcinoma by enhancing the MDM2-mediated degradation of E-cadherin. Dig. Dis. Sci. *60*, 3681–3690.

59. Luca, B.A., Steen, C.B., Matusiak, M., Azizi, A., Varma, S., Zhu, C., Przybyl, J., Espín-Pérez, A., Diehn, M., Alizadeh, A.A., et al. (2021). Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell *184*, 5482–5496.e28.

60. Roper, N., Brown, A.-L., Wei, J.S., Pack, S., Trindade, C., Kim, C., Restifo, O., Gao, S., Sindiri, S., Mehrabadi, F., et al. (2020). Clonal evolution and heterogeneity of osimertinib acquired resistance mechanisms in EGFR mutant lung cancer. Cell Rep. Med. *1*, 100007.

61. Roper, N., Velez, M.J., Chiappori, A., Kim, Y.S., Wei, J.S., Sindiri, S., Takahashi, N., Mulford, D., Kumar, S., Ylaya, K., et al. (2021). Notch signaling and efficacy of PD-1/PD-L1 blockade in relapsed small cell lung cancer. Nat. Commun. *12*, 3880.

62. Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir, A., and Mesirov, J.P. (2017). Variant Review with the integrative genomics viewer. Cancer Res. *77*, e31–e34.

63. Jensen, M.A., Ferretti, V., Grossman, R.L., and Staudt, L.M. (2017). The NCI Genomic Data Commons as an engine for precision medicine. Blood *130*, 453–459.

CellPress
OPEN ACCESS

64. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

65. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678.

66. O'Neil, N.J., Bailey, M.L., and Hieter, P. (2017). Synthetic lethality and cancer. Nat. Rev. Genet. *18*, 613–623.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human PBMC from patients with malignant mesothelioma participating in "Mesothelioma Natural History" clinical trial	Thoracic and GI Malignancies Branch, Center for Cancer Research, National Cancer Institute, NIH	https://clinicaltrials.gov/ct2/show/NCT01950572
Human tissue biopsy samples from patients with malignant mesothelioma participating in "Mesothelioma Natural History" clinical trial	Thoracic and GI Malignancies Branch, Center for Cancer Research, National Cancer Institute, NIH	https://clinicaltrials.gov/ct2/show/NCT01950572
Deposited data		
Whole Exome Sequencing data	This paper	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002207.v1.p1 Individual level data available in "Molecular Datasets" tab. Release Type: Controlled access. Authorized access can be requested via https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login
RNA Sequencing	This paper	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002207.v1.p1 Individual level data available in "Molecular Datasets" tab. Release Type: Controlled access. Authorized access can be requested by: https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login
Software and algorithms		
hclust function in *R*		https://stat.ethz.ch/R-manual/R-devel/library/stats/html/hclust.html
Corum 3.0	CORUM protein complex resource. Giurgiu et al., 2019.[37]	http://mips.helmholtz-muenchen.de/corum/
CIBERSORT	Newman et al., 2015.[41]	http://cibersort.stanford.edu/
SELECT	Lee et al., 2021.[46]	https://zenodo.org/record/4661265

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Raffit Hassan, hassanr@mail.nih.gov.

Materials availability

All unique reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- The NCI mesothelioma data resource is available in Tables 1, S1, S2, S3, and S4 and dbGaP with controlled access (accession number: dbGAP: phs002207; https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002207.v1.p1). Individual level data is available in the "Molecular Dataset" Tab and requires authorized access via https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

CellPress
OPEN ACCESS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patients

All patients with malignant mesothelioma treated at the Thoracic and GI Malignancies Branch, Center for Cancer Research, National Cancer Institute (NCI), National Institutes of Health, were offered participation in a clinical trial of the natural history of malignant mesothelioma (ClinicalTrials.gov number NCT01950572; https://clinicaltrials.gov/ct2/show/NCT01950572). The study was conducted in accordance with the principles of the International Conference on Harmonisation - Good Clinical Practice (ICH-GCP) guidelines. Institutional Review Board of the NCI approved the study and all patients provided written informed consent. Between September 2013 and November 2019, 425 consecutive patients were enrolled, out of them 122 underwent whole exome sequencing of blood and tumor tissue. RNA-seq data was available for 100 patients. The diagnosis of mesothelioma in all patients was confirmed by Board-Certified Anatomic Pathologists of the NCI Laboratory of Pathology with expertise in thoracic pathology using appropriate histologic and immunohistochemical studies and characterized their origin (site of disease) as pleural, peritoneal, or tunica vaginalis. Patients were enrolled regardless of asbestos exposure, age at diagnosis, or personal or family history of cancer.

METHOD DETAILS

Analysis of germline and somatic mutations and tumor mutational burden

Whole exome sequencing (WES) was carried out on paired tumor and germline DNA, which was extracted from formalin fixed paraffin embedded (FFPE) tumor tissue and from peripheral blood mononuclear cells (PBMC), respectively, as previously described.[60,61] Exome sequencing was performed using Illumina NextSeq500 sequencers. The bcl files generated were converted to FASTQ files, which contained paired-end reads, using the bcl2fastq tool (Illumina, San Diego, CA). Germline variants found in the mesothelioma cohort (n = 122) were called using previously published methods.[60,61] High-confidence germline variants were defined by the following criteria: total coverage of greater than 20x, Fisher score <75 and variant allele frequency (VAF) $\geq$ 0.25. Germline variants were curated according to the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) guidelines for the interpretation of sequence variants.[30] Finally, pathogenic, or likely pathogenic tier 1 and 2 germline variants used in the Fisher exact test were filtered according to Phred score $\geq$20 and presence shown in Integrative Genome Viewer v. 2.3.31.[62] Somatic variant data were collected from VCF files of the genetic variants found in exome samples in tumors from the mesothelioma cohort (n = 122 tumors). The number of somatic mutations for each tumor were counted, and then divided by the number of bases in the exome of that tumor to yield the somatic mutation burden per Mb.

RNA sequencing and data processing

RNA sequencing was performed on Illumina HiSeq2000 or NextSeq500 using TruSeq3 chemistry. The bcl files generated were converted to FASTQ files, which contained paired-end reads, using the bcl2fastq tool (Illumina, San Diego, CA).

NCI mesothelioma data: The RNA-seq reads were processed using the same workflow as the NCI Genomic Data Commons (GDC) has used to process TCGA-MESO RNA-seq reads, with the same settings where appropriate (see NCI GDC mRNA Analysis Pipeline in https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/).[63] Namely, FASTQ reads were aligned to the NCI GDC reference genome GRCh38.p2 with GENCODE v22 gene annotations (see NCI GDC Reference Files in https://gdc.cancer.gov/about-data/gdc-data-processing/gdc-reference-files).[63] using STAR v2.7.3a in two-pass mode. Read counts were quantified from STAR aligned read BAMs using HTSeq v0.11.2 in reverse strand-specific mode to reflect the RNA library protocol used. Counts were further processed for downstream analysis using edgeR with default settings.[64] Genes with sufficiently large counts were filtered using the filterByExpr function, normalized using trimmed mean of M-values (TMM), and finally log counts-per-million (CPM) transformed all with default settings. There are 100 patients for which we had RNA-seq samples. TCGA mesothelioma data RNA-seq read counts (n = 85 patients) were obtained from NCI GDC.[28] Similar processing as above was done for the TCGA dataset. TCGA pan-cancer RNA-seq TPM (log transformed) data was taken from the Xena browser.[65]

Alternative method to compute risk scores

For the survival analysis, we also computed the risk scores using an alternative method. The new *fractional risk score* for a patient was computed by counting the fraction of the 48 genes in the mesothelioma prognostic signature which are highly expressed (top 33 percentile of all genes) in a patient. Survival analysis for the new risk scores is done on TCGA and Bueno et al. mesothelioma datasets after controlling for age and gender.

Protein complex enrichment and cluster gene expression

Mesothelioma prognostic signature gene-set was used for protein complex-enrichment analysis. Protein complex dataset used is 03.09.2018 Corum 3.0 current release (http://mips.helmholtz-muenchen.de/corum/). Fisher exact test was used for the enrichment studies (FDR <0.05). Hierarchical clustering of the 48 genes composing the gene mesothelioma prognostic signature was done using the hclust function in *R* (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/hclust.html).

Immune cell abundance estimates

CIBERSORT (http://cibersort.stanford.edu/) was used to estimate the immune cell abundance in each patient in the NCI mesothelioma dataset, and also in the TCGA mesothelioma (n = 85) and Bueno et al. mesothelioma (n = 211) patient datasets. CIBERSORT was run for 500 permutations using the default gene signature LM22 CIBERSORT. The results were run in the 'Relative' mode generating 'Relative' fractions of 22 immune cell types for each of the 100 patients in the NCI mesothelioma dataset. The relative values indicate the relative abundance of each immune cell type as a fraction of the bulk tumor data (which does not include cancer). The relative values for each sample add up to 1 for all cell types. The immune estimates for TCGA cancer patients were also obtained using a CIBERSORT analysis from.[42] For all the immune cell types which have more than 5% mean relative abundance across samples in the NCI mesothelioma dataset, we tested their association with patient survival (FDR <0.2).

Drug response prediction using SELECT

SELECT (SynthEtic Lethality and rescue-mediated precision onCology via the Transcriptome; https://zenodo.org/record/4661265) is a computational method to stratify patients as responders or non-responders for each drug using whole transcriptomic data.[46] SELECT first infers clinically-relevant synthetic lethal (SL) or synthetic rescue (SR) pairs across cancer types by mining thousands of patient tumor genomic and transcriptomic data in the TCGA dataset. Synthetic lethality is a genetic interaction between two genes, such that, when one of the genes is inactivated, the cells remain viable, however, when both genes are inactivated, the cells lose their viability.[42,66] Synthetic rescue on the other hand is a genetic interaction between two genes such that, when any one of the genes is inactivated, the cells become less viable, but downregulation (or upregulation) of the partner gene rescues the cell.[42,45,46] For each of the different cancer therapies, we mapped the drugs to the targets (which they inhibit) using DrugBank.[47] For chemotherapy and targeted therapy drugs, for each patient sample, SELECT assigns a risk score – called SL-score – based on the number of downregulated SL partners of the target genes inhibited by a drug. The method works under the assumption that a drug is likely to kill a tumor more effectively when its SL partners are downregulated, as the inhibition of the target by the drug will lead to active SL formations (as both the SL gene pairs will becoming jointly downregulated). For immunotherapy drugs, SELECT assigns risk scores using SRs and it works under the assumption there is an innate resistance to a drug if there are a lot of SR partners that are downregulated in the cancer sample, as the inhibition of the target by the drug will lead to active SR formations. Using the SL/SR network, SELECT was previously shown to accurately predict drug response and stratify many patients in 28 of 35 independent targeted and immunotherapy datasets ranging across various cancer types and for many different drugs,[46] using the same set of parameters. For both SL and SR based approaches, higher the risk score, more likely the patient would respond to a drug.[46] These risk scores are then used for drug response validations and classifying patients as responders or non-responders in many independent datasets. Following the findings of Lee et al., 2021,[46] for the drug coverage analysis, a risk score >0.44 was used to classify a patient as responder for chemotherapy or targeted therapy and a risk score $\geq$0.9 was used to classify a patient as responder for immunotherapy. We emphasize that our analysis uses the exact same parameters and thresholds that was used in the original SELECT pipeline,[46] to identify the pertaining SL and SR partners and to compute risk scores in the mesothelioma cohorts, without any training and parameter tuning on the latter.

QUANTIFICATION AND STATISTICAL ANALYSIS

For survival analysis, survival data was censored as of November 1, 2019. The K-M method was used to estimate the survival probability from diagnosis date to death or last follow-up (censored) date. Data were stratified either by site (peritoneal, pleural) alone, or were first analyzed according to subgroups of asbestos, smoking history, diagnosis age, and gender, and then stratified by site. Hazard ratios (and their 95% CI) were calculated by Cox proportional hazard model with pleural mesothelioma as reference. Median overall survival estimates were provided with their 95% CI (however, some were non-estimable). K-M plots are presented with pointwise 95% confidence limits and the number at risk for various points. As some of the mesothelioma samples in the NCI mesothelioma data were acquired post treatment, we computed the survival of all patients from the time of biopsy sampling to the time of death or last follow-up. This survival time estimate was used for the gene expression and copy-number based survival association studies.

Mesothelioma prognostic signature set identification and cross-validation was done based on survival analysis. Risk scores for each patient is computed as the median expression of the mesothelioma prognostic signature gene-set for that patient. For the NCI mesothelioma data, we did the analysis using a ''leave-one-out'' cross validation. For each of the training samples, we re-computed the genes whose increased expression is associated with worse survival using Cox regression analysis after controlling for age, gender, and site-of-disease (FDR <0.1), and the median expression value (risk score) for this set of genes was computed for the test data point. The risk scores were then associated with survival analysis using Cox regression (after controlling for age, gender, and site of disease).

To calculate differential gene expression between pleural and peritoneal mesothelioma, we used edgeR[64] on the RNA-seq count data to compute differential expression between the two groups, after controlling for various confounding factors like age and gender (FDR <0.1). One-sided Wilcoxon rank-sum test was used for the analysis in Figure 3F and for the immune cell abundance estimates analysis related to Figure 4A.